



UNIL | Université de Lausanne

Unicentre

CH-1015 Lausanne

<http://serval.unil.ch>

Year : 2011

Méthodes de visualisation en analyse criminelle: approche générale de conception des schémas relationnels et développement d'un catalogue de patterns

Quentin Rossy

Quentin Rossy, 2011, Méthodes de visualisation en analyse criminelle: approche générale de conception des schémas relationnels et développement d'un catalogue de patterns

Originally published at : Thesis, University of Lausanne

Posted at the University of Lausanne Open Archive.
<http://serval.unil.ch>

Droits d'auteur

L'Université de Lausanne attire expressément l'attention des utilisateurs sur le fait que tous les documents publiés dans l'Archive SERVAL sont protégés par le droit d'auteur, conformément à la loi fédérale sur le droit d'auteur et les droits voisins (LDA). A ce titre, il est indispensable d'obtenir le consentement préalable de l'auteur et/ou de l'éditeur avant toute utilisation d'une oeuvre ou d'une partie d'une oeuvre ne relevant pas d'une utilisation à des fins personnelles au sens de la LDA (art. 19, al. 1 lettre a). A défaut, tout contrevenant s'expose aux sanctions prévues par cette loi. Nous déclinons toute responsabilité en la matière.

Copyright

The University of Lausanne expressly draws the attention of users to the fact that all documents published in the SERVAL Archive are protected by copyright in accordance with federal law on copyright and similar rights (LDA). Accordingly it is indispensable to obtain prior consent from the author and/or publisher before any use of a work or part of a work for purposes other than personal use within the meaning of LDA (art. 19, para. 1 letter a). Failure to do so will expose offenders to the sanctions laid down by this law. We accept no liability in this respect.



UNIL | Université de Lausanne

Faculté de Droit et des Sciences Criminelles

Ecole des Sciences Criminelles

**Méthodes de visualisation en analyse criminelle: approche
générale de conception des schémas relationnels et
développement d'un catalogue de patterns**

THÈSE DE DOCTORAT

présentée à la
Faculté de Droit et des Sciences Criminelles
de l'Université de Lausanne

pour l'obtention du grade de
Docteur ès Sciences en science forensique
par

Quentin Rossy

Directeur de thèse
Prof. Olivier Ribaux

LAUSANNE
2011



UNIL | Université de Lausanne
Ecole des sciences criminelles
bâtiment Batochime
CH-1015 Lausanne

IMPRIMATUR

A l'issue de la soutenance de thèse, le Jury autorise l'impression de la thèse de M. Quentin Rossy, candidat au doctorat en science forensique, intitulée

« Méthodes de visualisation en analyse criminelle : approche générale de conception des schémas relationnels et développement d'un catalogue de patterns »

Le Président du Jury



Professeur Pierre Margot

Lausanne, le 29 juillet 2011

Série criminalistique LII

ISBN : 2-940098-56-5

REMERCIEMENTS

Cette recherche a été menée à l'Ecole des Sciences Criminelles de l'Université de Lausanne sous la direction du Professeur Olivier Ribaux. Le jury était composé du Docteur Olivier Guéniat, commandant de la police cantonale jurassienne, de Monsieur Christian Tournié, Expert national détaché, administrateur à la Direction générale Affaires intérieures, Commission européenne et du Docteur Andrew David Barclay, Forensic Consultant et Senior lecturer à la Robert Gordon University, Aberdeen et professeur honoraire à l'Université de Hull. Je remercie l'ensemble du jury de l'intérêt qu'ils ont porté à mes travaux, de leurs encouragements et de leurs très précieux conseils et corrections.

L'effort d'introspection effectué lors de la rédaction de la thèse, m'a fait réaliser à quel point le titre de cette recherche contient un mot clé : « relationnel ». Au delà du caractère centrale de l'analyse dite relationnelle en analyse criminelle, si je ne devais retenir qu'une seule chose, c'est que l'intérêt principal de ce travail réside bien dans les relations qu'il m'a permis et me permet de créer. J'espère vivement qu'il m'offrira l'opportunité d'en réaliser de nombreuses autres.

Je remercie les personnes qui m'ont soutenu, conseillé et aidé tout au long de la thèse, ainsi que l'école et les organes de police qui m'ont offert des cadres de travail ouverts aux initiatives de recherche et développement. Un grand merci notamment à Pierre Margot, directeur de l'Ecole des Sciences Criminelles et à Olivier Guéniat, chef de la police judiciaire lors de mon activité à la police neuchâteloise. Je tiens à exprimer ma vive gratitude à Olivier Ribaux, directeur de thèse, pour la confiance qu'il m'a témoigné tout au long de mes activités de recherche et d'enseignement, pour ces très précieux conseils et pistes de réflexion et sa passion communicative.

Une pensée particulière pour tous les collègues avec qui j'ai partagé de nombreuses discussions et qui m'ont transmis un peu de leur savoir. J'aurais aimé les citer tous, j'ai une pensée pour chacun d'eux:

- Les collègues de la police neuchâteloise et tout particulièrement Jean-Claude Voirol qui m'a fait découvrir l'analyse de la délinquance sérielle.
- Les collègues des polices de suisse romande, du CICOP et les analystes criminels, plus particulièrement Julien Cartier et Sébastien Capt.
- Les collègues de l'ESC et en particulier tous les membres du groupe analyse criminelle, dont Stéphane Birrer, pour son soutien dès mes débuts en tant qu'assistant et pour tout les grands moments passé lors du développement de PICAR. Sylvain loset et Thibault Genessay, pour leur soutien permanent et les très nombreuses discussions sur différents aspects de cette recherche.

Vous avez nourri mon travail et ma curiosité, merci.

A ma famille,

A mes parents,

A ma femme,

Mille merci.

S U M M A R Y

Police work and criminal investigation require taking many decisions: *i.e.* choosing which traces should be analysed and placing under surveillance or detention a suspect are decisions that are made daily by the actors of the legal system. These decisions need to weigh competing interests based on the analysis of available information. It is the role of crime analysis to put into perspective collected information to make it intelligible to decision makers. The use of visualisations is particularly recommended to support the analysis and communication of such information by highlighting the essential features.

Link charts are used in criminal investigations in order to facilitate the processing of large-scale investigation data. The relevant elements of the inquiry are represented in the form of diagrams describing the relationships between events and entities featuring in the investigation (such as people, objects and traces). For instance, traditional uses of those graph-like techniques are: the representation of criminal networks, smuggling of goods, chronologies of events, as well as the visualisation of telephone records and financial data. In this context, visualisations are used for many objectives, such as analysing the traces and the information gathered, evaluating a cold case, helping along the categorization of a particular offence, facilitating the transmission and receipt of a case or supporting an argument at the Court's trial.

Common practice includes simple software tools that produce powerful and often elegant visualisations. However their use raises important difficulties. This research suggests that there are astonishing disparities in the use of these techniques. Reasoning and perception biases can be introduced, sometimes leading to wrong decisions with serious consequences. This observation reveals the need to consolidate available methods.

To highlight these difficulties, evaluations were conducted with practitioners and students. An empirical picture of the manifold ways in designing and interpreting visualisations has been established. The impact of this variety on decision-making is also discussed. The nature and variety of concepts to be represented, the absence of an emerging consensus on how to represent data, the diversity of visual solutions, the constraints imposed by tools and the absence of a clear formalization of the visual language, are all supposed causes of the observed difficulties.

Over the past two decades, several lines of development have been proposed to deal with these problems, such as improving automated graph building tools, exploitation of social network analysis, automation of the identification and extraction of entities in unstructured text and the definition of formal languages. This research suggests a parallel approach based on an appropriate use of graph structures and visual properties to optimize readability based on analysis objectives and nature of the information to be represented.

Solutions were sought in several areas. General recommendations from various research communities related to visualisation were gathered to provide a general approach for the design of link charts. In addition, the development of a best practices catalogue formalized as patterns was initiated. Each pattern describes a particular solution to a recurring analysis question, such as the analysis series of burglaries. Finally, the impact on tools of the proposed methodology is discussed in light of the limitations they impose. A multi-dimensional visualisation software prototype was developed.

On the whole, this research highlights the difficulties encountered in designing link charts that support inquiries and proposes methodological and technical innovations useful for practice and teaching.

RESUME

Le travail policier et l'enquête judiciaire nécessitent de prendre de nombreuses décisions : choisir quelle trace analyser, mettre sous surveillance ou en détention un suspect, sont autant de décisions qui sont prises quotidiennement par les acteurs du système judiciaire. Ces décisions font l'objet de pesées d'intérêts qui se fondent sur l'analyse de l'information accessible. C'est le rôle de l'analyse criminelle de mettre en perspective l'information colligée pour la rendre intelligible aux décideurs compétents. L'usage de représentations graphiques est notamment recommandé pour soutenir l'analyse et la communication de ces informations.

Des techniques de visualisation relationnelle sont notamment exploitées dans les enquêtes judiciaires afin de faciliter le traitement d'affaires d'envergure. Les éléments pertinents de l'enquête sont représentés sous la forme de schémas décrivant les relations entre les événements et les entités d'intérêts (tel que des personnes, des objets et des traces). Les exploitations classiques de ces techniques qui s'apparentent à des graphes, sont par exemple : la représentation de réseaux criminels, de trafics de marchandises, de chronologies d'événements, ainsi que la visualisation de relations téléphoniques et financières. Dans ce contexte, la visualisation soutient un nombre importants d'objectifs, tels qu'analyser les traces et les informations collectées, évaluer a posteriori une investigation, aider à qualifier les infractions, faciliter l'appréhension d'un dossier, voire soutenir une argumentation lors du procès.

La pratique intègre des outils logiciels simples qui produisent des graphiques élégants et souvent percutants. Leur utilisation semble néanmoins soulever des difficultés. Cette recherche tend à montrer qu'il existe des disparités étonnantes lors de l'exploitation de ces techniques. Des biais de raisonnement et de perception peuvent être induits, allant jusqu'à provoquer des

décisions aux conséquences parfois désastreuses. Ce constat révèle la nécessité de consolider les méthodes pratiquées.

Pour mettre en évidence ces difficultés, des évaluations ont été effectuées avec des praticiens et des étudiants. Elles ont permis d'établir une image empirique de l'étendue des variations de conception et d'interprétation des représentations, ainsi que de leurs impacts sur la prise de décision. La nature et la diversité des concepts à représenter, l'absence d'un consensus émergeant sur la manière de représenter les données, la diversité des solutions visuelles envisageables, les contraintes imposées par les outils exploités et l'absence d'une formalisation claire du langage, sont autant de causes supposées des difficultés.

Au cours des vingt dernières années, plusieurs axes de développement ont été proposés pour traiter les difficultés observées, tels que l'amélioration des automatismes facilitant la conception d'une représentation, l'exploitation des techniques de réseaux sociaux, l'automatisation de l'identification et de l'extraction des entités dans du texte non-structuré et la définition de langages formels. Cette recherche propose une approche parallèle fondée sur une exploitation adaptée de structures de graphe et de propriétés visuelles pour optimiser la représentation en fonction des objectifs définis et de la nature des informations à représenter.

Des solutions ont été recherchées selon plusieurs axes. Des recommandations générales, issues de diverses communautés de recherche liées à la visualisation, ont été recherchées afin de proposer une démarche générale de conception des schémas relationnels. Par ailleurs, le développement d'un catalogue de bonnes pratiques formalisées sous la forme de patterns de visualisation a été amorcé. Chaque pattern décrit une solution particulière pour un problème d'analyse récurrent, tel que l'analyse d'une série de cambriolages. Finalement, l'impact sur les outils de la méthodologie proposée est discuté en regard des limites qu'ils imposent. Un prototype de visualisation multidimensionnel a été élaboré.

Cette recherche met en évidence les difficultés rencontrées lors de l'exploitation de représentations graphiques pour soutenir le processus de l'enquête judiciaire et propose des éléments de méthode et des innovations techniques utiles tant pour l'enseignement de la discipline que pour sa pratique.

TABLE DES MATIÈRES

INTRODUCTION	1
<i>Contexte</i>	1
<i>Problématique</i>	2
<i>Structure de la thèse</i>	2
PARTIE I : CONTEXTE ET PROBLEMATIQUE	6
ANALYSE CRIMINELLE ET VISUALISATION	7
<i>Processus de l'analyse criminelle</i>	9
<i>Rôles de la visualisation dans le processus</i>	15
<i>Renseignement et formes d'analyse</i>	20
Analyse, profil et pattern	26
Analyse selon le type de données	29
<i>Dimensions d'analyse et formes de visualisation</i>	30
Dimensions d'analyse	33
Formes de visualisation	34
<i>Analyse et visualisation relationnelle</i>	36
Recommandations et standards	45
<i>Conclusion</i>	47
EXPLOITATIONS POUR L'ENQUETE JUDICIAIRE	48
<i>Enquête</i>	48
Processus de l'enquête et objectifs d'analyse	48
Intégration de l'analyse criminelle opérationnelle	52
<i>Jugement</i>	55
Objectifs d'analyse	55
Contraintes imposées sur les visualisations	56
<i>Conclusion</i>	58
MÉTHODES D'ÉVALUATION DE VISUALISATIONS	59
<i>Critères d'évaluation</i>	61
<i>Procédures d'évaluation</i>	63
EVALUATIONS PRELIMINAIRES	65
<i>Disparités de représentation</i>	65
Représentation d'un scénario simple	65
Représentation d'une vente	68
Représentation d'une affaire complexe	69
Discussion	72
<i>Impact sur l'analyse et la communication</i>	73
Représentation d'une affaire complexe	73
Représentation d'un trafic de marchandises	75
Représentation de séries de cambriolage	77
Représentation d'un cas d'abus sexuel	80

SYNTHESE	83
<i>Constat</i>	83
<i>Consolidation des méthodes</i>	87
Délimitation du cadre	87
Objectifs	88
Méthodologie	88
PARTIE II : CONTRIBUTION AUX MÉTHODES DE CONCEPTION DES SCHÉMAS RELATIONNELS	90
PROCESSUS DE CONCEPTION D'UN SCHEMA RELATIONNEL	91
<i>Description du langage visuel</i>	94
Propriétés du plan	94
Composants d'un graphe	96
Attributs graphiques	97
<i>Typologie de relations</i>	100
<i>Evaluation de la représentation de relations</i>	104
Méthode	105
Description de l'échantillon	107
Résultats	108
Tableau de synthèse	119
Discussion générale sur les résultats	119
Discussion sur la méthode d'évaluation	120
<i>Éléments de méthode</i>	122
Identification de la dimension dominante	122
Elaboration du modèle relationnel	125
Simplifications	127
Constance et définition de conventions	135
Ordonnancement des composants	137
Résumé de la méthode proposée	141
PATTERNS DE VISUALISATION RELATIONNELLE	142
<i>Analyser une série d'événements</i>	143
Représentation des liens forensiques	148
Traces de semelles et motifs	150
Traces biologiques et profil ADN	153
Saisies de stupéfiant et classes chimiques	155
Relations avec des véhicules	159
Relations spatio-temporelles	162
<i>Sélectionner des traces à analyser</i>	165
<i>Analyser des transactions</i>	169
Communications téléphoniques	172
Transactions financières	182
Ventes de marchandises	189
IMPACT SUR LA CONCEPTION D'OUTILS	193
<i>Problématique</i>	193
<i>Processus de visualisation</i>	195
<i>Modification dynamique de la structure du graphe</i>	199
<i>Intégration des dimensions d'analyse</i>	201
Prototype de visualisation multidimensionnelle	201

DISCUSSION GENERALE	207
<i>Résultats empiriques</i>	207
<i>Méthodes d'évaluation</i>	208
<i>Apports sur un plan méthodologique</i>	209
<i>Perspectives</i>	211
CONCLUSION	215
BIBLIOGRAPHIE	217
ANNEXES	232
ANNEXE 1 : DIMENSIONS D'ANALYSE ET VISUALISATION	233
<i>Analyse quantitative</i>	233
<i>Analyse temporelle</i>	239
<i>Analyse spatiale</i>	249
ANNEXE 2 : SCHEMAS DE SCENARIOS	256
ANNEXE 3 : VISUALISATION TRIDIMENSIONNELLE	259
<i>Types de représentation</i>	259
<i>Avantages</i>	261
<i>Faiblesses</i>	262
<i>Conclusion</i>	263
ANNEXE 4 : EVALUATIONS PRELIMINAIRES	265
<i>Représentation d'un scénario simple</i>	265
<i>Représentation d'une vente</i>	268
<i>Représentation d'une affaire complexe</i>	269
<i>Modification du type de représentation</i>	272
ANNEXE 5 : EVALUATION DE LA REPRESENTATION DE RELATIONS	278
<i>Influence des variables liées aux participants</i>	278
<i>Résultats détaillés</i>	281

INTRODUCTION

Contexte

Les services de police (au niveau local, national et international) ont de plus en plus recours à des modèles de prises de décisions opérationnelles et stratégiques fondées sur l'exploitation de produits de l'analyse de données. Parmi les problématiques majeures actuelles, la mise en place de structures d'analyse chargées de traiter la masse d'informations recueillies et la définition de méthodologies de travail efficaces sont des questions permanentes. De nombreuses formes de représentations graphiques sont exploitées dans ce cadre pour produire et communiquer les renseignements utiles à la prise de décision.

De nombreuses polices à travers le monde et tout particulièrement aux Etats-Unis ont investi des moyens importants pour produire des systèmes de visualisation cartographique de la criminalité. Ces systèmes permettent d'obtenir rapidement une image claire et facile d'accès de problèmes criminels souvent complexes. Par ailleurs, des logiciels de visualisation ont également été intégrés dans les enquêtes judiciaires afin de faciliter le traitement d'affaires d'envergure. Les éléments pertinents de l'enquête sont représentés sous la forme de schémas décrivant les relations entre les événements et les entités d'intérêt (tel que des personnes, des sociétés, des véhicules et des traces). Les exploitations classiques de ces techniques sont, par exemple : la représentation de réseaux criminels, de trafics de marchandises, de modes opératoires complexes, de chronologies, ainsi que la visualisation de relations téléphoniques et financières.

2 Introduction

Les nouvelles technologies apportent des solutions techniques de plus en plus variées. Mais cette apparente simplicité d'élaboration de produits graphiques élégants et souvent percutants semble toutefois cacher un nombre important de problèmes méthodologiques et pratiques.

P r o b l é m a t i q u e

Le travail policier et l'enquête judiciaire nécessitent de prendre de nombreuses décisions : une visite domiciliaire, une mise sous surveillance ou une mise en détention d'un suspect, sont autant de décisions qui sont prises quotidiennement par les acteurs du système judiciaire. Ces décisions font l'objet de pesées d'intérêts qui se fondent sur l'analyse de l'information accessible. C'est le rôle de l'analyse criminelle de mettre en perspective cette information pour la rendre intelligible aux décideurs compétents. L'usage de représentations graphiques est notamment recommandé pour soutenir l'analyse et la communication des informations. Bien que l'analyse criminelle propose une méthode générale et harmonisée, la mise en œuvre pratique de la méthode semble poser de nombreuses difficultés lors de la production de telles représentations. Cette recherche tend à montrer qu'il existe de réelles difficultés, puis proposer des solutions méthodologiques et techniques. Les questions abordées sont : quel est le contexte d'exploitation de la visualisation en analyse criminelle, quels sont ses rôles, quelles sont les formes de représentation exploitées et quels sont les problèmes rencontrés lors de leurs utilisations?

S t r u c t u r e d e l a t h è s e

P r e m i è r e p a r t i e

La première partie décrit le contexte général d'exploitation de la visualisation en analyse criminelle et plus particulièrement son usage dans l'enquête judiciaire dont sont issus les cas traités et les données analysées. Des exemples et des évaluations préliminaires sont présentés afin d'illustrer les difficultés observées et tenter d'en identifier les causes.

La méthode de travail connue sous le nom de cycle du renseignement est brièvement présentée dans le premier chapitre afin de décrire les rôles de la visualisation dans la méthode. La deuxième section du premier chapitre présente les formes d'analyse définies dans

la littérature. Elles décomposent l'analyse criminelle en fonction d'objectifs d'exploitation spécifiques. Les manuels d'analyse criminelle relèvent que certaines formes de visualisation sont plus adaptées pour certaines formes d'analyse (tel que suggéré dans le guide sur l'analyse criminelle d'Interpol (Interpol, 1997) ou, plus localement, par le manuel d'analyse criminelle opérationnelle de la police fédérale suisse (Fedpol, 2010)). Par exemple, un schéma d'événement est généralement produit lors d'une analyse de cas et des schémas de flux lors d'une analyse de données financières. De nombreuses formes ont été définies, mais la méthode définie est générale et les techniques exploitées semblent également être exploitées de façon transversale. La troisième section présente une décomposition basée sur la définition de quatre dimensions d'analyse principale : temporelle, spatiale, relationnelle et quantitative. Bien que cette classification regroupe et simplifie un ensemble de questionnements complexes, elle permet de classer les visualisations et d'orienter le choix d'une ou plusieurs d'entre elles. Les différentes formes de représentation exploitées dans le domaine sont donc classifiées et présentées selon ces quatre dimensions. La dimension relationnelle et plus particulièrement ces formes de représentations propres a été choisie comme sujet d'étude. En effet, l'exploitation des schémas relationnels qui permettent de visualiser les relations entre des entités d'intérêts (des événements, des personnes, des sociétés, etc.), a fait l'objet d'un nombre limité de recherches dans le domaine de l'analyse criminelle (Regali, 1999) (Xiang et al., 2005) (Chen et al., 2005), alors que leur usage s'est rapidement rependu depuis une vingtaine d'année. Finalement, les recommandations définies dans les manuels de l'analyse criminelle qui portent sur la conception des schémas relationnels sont présentées.

Le deuxième chapitre décrit le contexte de l'enquête judiciaire dont sont issus les exemples analysés dans cette thèse. Les objectifs attendus en matière de visualisation sont décrits ainsi que les besoins des acteurs du processus en regard de la visualisation ; qu'ils en soient les créateurs ou les destinataires. Bien que la problématique traitée dans cette thèse se base sur l'étude de cas issus de l'investigation criminelle, les problèmes identifiés et l'approche envisagée pour les traiter semblent toutefois avoir une utilité plus globale.

Il est soupçonné que les pratiques de conception des schémas relationnels sont disparates et que les visualisations produites sont plus ou moins efficaces. Le troisième chapitre définit la notion d'efficacité, les indicateurs permettant de la mesurer et les méthodologies existantes pour l'évaluer.

Les difficultés d'élaboration de représentations sont présentées dans le quatrième chapitre qui résume des évaluations préliminaires effectuées et des exemples de cas. En premier lieu, la

4 Introduction

disparité des solutions graphiques élaborées pour une même situation par différentes personnes est observée. Deux illustrations portant sur des cas simples sont présentées (représentation d'une seule phrase). La troisième illustration expose les schémas relationnels produits à partir d'une même affaire, par un ensemble d'enquêteurs, d'étudiants et d'analystes criminels. Il est alors constaté qu'il ne semble pas exister de consensus émergent en matière de représentation relationnelle des informations. Ces disparités ont-elles un impact important sur l'analyse et la communication des informations ? Des exemples d'exploitation au cours du processus d'enquête et du jugement sont présentés afin d'illustrer l'impact des choix effectués pour analyser, synthétiser et communiquer les informations d'affaires complexes.

S e c o n d e p a r t i e

Les contributions présentées dans la deuxième partie de cette thèse sont divisées en trois chapitres. Le premier chapitre est consacré à la présentation d'éléments généraux de méthode formalisés d'une part sur les travaux de recherches issus de la littérature en matière de visualisation et d'autre part sur des observations effectuées lors du traitement de situations récurrentes. Une formalisation du processus de conception d'un graphe est proposée afin d'explicitier, pour chaque étape, les causes supposées des difficultés observées lors des évaluations préliminaires. La démarche proposée repose sur une définition plus précise du langage visuel exploitable pour concevoir un schéma. Une typologie de relations est proposée afin de guider les choix de conception. Celles-ci ont été comparées empiriquement afin d'étudier l'impact du changement de structure sur l'efficacité des graphes. La dernière section du premier chapitre vise à présenter les recommandations générales identifiées. Elles sont intégrées dans une ébauche de démarche générale guidant la conception des schémas.

Les recommandations et les éléments de méthode formalisés dans le premier chapitre ont notamment été définis en regard de situations récurrentes identifiées au cours d'enquêtes. Le second chapitre présente les bonnes pratiques proposées pour représenter ces situations : tel que le traitement de séries de cambriolages ou l'analyse de données téléphoniques par exemple. La démarche est ainsi explicitée par des exemples concrets. L'objectif est d'amorcer le développement d'un catalogue de bonnes pratiques. Il ne saurait être démontré que les solutions proposées sont les meilleures possible. Elles ont toutefois montré une certaine efficacité qui est discutée pour chaque cas.

Le troisième chapitre s'intéresse à l'impact des recommandations méthodologiques identifiées sur les outils de visualisation. Ceux-ci imposent parfois leurs propres standards et les

possibilités sont évidemment cantonnées aux fonctionnalités offertes. Ils offrent généralement des solutions génériques, alors que les problèmes à résoudre sont souvent spécifiques. Ils ne favorisent pas toujours les bons choix de représentation en regard des questions d'analyse. Afin d'explicitier les difficultés liées à l'exploitation des logiciels, le processus de visualisation est décrit d'un point de vue technique. Une séparation nécessaire est identifiée entre le modèle des données récoltées et le modèle nécessaire pour produire une vue. Le développement de mécanismes permettant d'une part de faire communiquer ces modèles et d'autre part de modifier dynamiquement la structure des graphes est proposé comme une solution nécessaire pour que les techniques soutiennent plus efficacement les raisonnements effectués en cours d'analyse. Finalement, le développement de techniques de représentation intégrant l'ensemble des dimensions d'analyse est certainement l'un des défis majeurs de ces dix dernières années dans le domaine. En effet, cette problématique est identifiée depuis le début des années deux-mille et de nombreux outils sont développés dans cette perspective intégratrice facilitant l'analyse globale des données. Un prototype de visualisation multidimensionnelle a été amorcé dans le cadre de cette thèse. Il est présenté en fin de chapitre.

The power of the unaided mind is highly overrated. Without external aids, memory, thought, and reasoning are all constrained. But human intelligence is highly flexible and adaptive, superb at inventing procedures and objects that overcome its own limits. The real powers come from devising external aids that enhance cognitive abilities. How have we increased memory, thought, and reasoning? By the invention of external aids: It is things that make us smart. (Norman, 1993, p.43)

PARTIE I : CONTEXTE ET PROBLEMATIQUE

Diverses formes de représentation sont exploitées dans de nombreux domaines pour faciliter l'appréhension de grandes quantités de données, afin de détecter des tendances ou des anomalies qui sont alors rendues visuellement explicites (Ware, 2004). De nombreuses recherches ont été entreprises sur le sujet depuis les années soixante qui tendent à formaliser les principes et les méthodes de la visualisation de l'information. Dans une recherche sur la visualisation, l'analyse criminelle ne saurait être considérée autrement qu'en tant qu'application spécifique de certaines formes de représentation. L'analyste criminel, quant à lui, ne peut qu'appréhender la visualisation comme une technique particulière, certainement utile, voire nécessaire, lui facilitant l'accomplissement de certaines tâches. Il reconnaît que le fruit de son travail est souvent transmis grâce à une représentation visuelle. L'analyse criminelle semble en effet avoir réussi son intégration dans le domaine de la sécurité grâce notamment à un usage particulièrement efficace de représentations graphiques. Bien qu'elle couvre un processus bien plus large et complexe, il faut admettre que les produits de renseignement transmis sous la forme de représentations graphiques façonnent l'image de la discipline. L'analyse criminelle n'est de fait pas la seule discipline sur laquelle la visualisation a un impact important. Certains auteurs prétendent que l'essor des techniques de représentation de l'information est l'un des facteurs responsables du développement de toutes les sciences modernes (Crosby, 1997) (Friendly, 2008). Cette recherche aborde l'exploitation des représentations graphiques dans la perspective de l'analyse criminelle, tout en intégrant des principes et des recommandations formalisées dans des travaux liés à la visualisation afin de mieux comprendre les difficultés rencontrées et consolider les méthodes pratiquées.

Analyse criminelle et visualisation

Les termes d'analyse ou de renseignement criminel(le) sont parfois utilisés comme synonyme ou du moins de façon interchangeable. Dans le cadre de cette thèse, ces termes seront utilisés pour distinguer une méthode (analyse criminelle) du produit de l'analyse et de la fonction (renseignement criminel).

L'analyse criminelle est considérée en tant que méthode systématique exploitant notamment des techniques de la gestion de l'information afin de détecter, mémoriser et représenter des relations¹. L'analyse criminelle décrit un processus général qui par une exploitation structurée des informations aide à gérer les informations accessibles afin d'en maîtriser les flux et de garder la vue d'ensemble, cherche à structurer les raisonnements et en particulier à en minimiser les biais et finalement met en perspective les informations pour les présenter aux décideurs. La visualisation fait partie des moyens exploités au cours du processus pour analyser les informations collectées et communiquer les renseignements produits.

Le renseignement se définit comme le produit de l'analyse criminelle mais également comme une fonction. En tant que produit, le renseignement se distingue des notions de donnée, d'information et de connaissance. Le vocabulaire utilisé suit le modèle DIKI de Ratcliffe (2009) :

¹ L'analyse criminelle consiste en la recherche et la mise en évidence méthodiques de relations - d'une part entre les données de la criminalité elles-mêmes et, d'autre part entre des données de la criminalité et d'autres données significatives possibles - à des fins de pratiques judiciaires et policières. (Interpol, 1997)

L'analyse stratégique est l'ensemble des méthodes et des techniques d'analyse qui essayent de manière scientifique, de décrire, d'évaluer et d'expliquer la problématique de sécurité policière, en mettant en évidence des relations entre d'une part les différents aspects de ces problèmes et d'autre part des données contextuelles.

(définition de l'analyse stratégique par la police fédérale belge, http://www.polfed-fedpol.be/org/djb_environnement/djb_envir11_fr.php, dernier accès le 12.11.2010).

8 Analyse criminelle et visualisation

data, information, knowledge et intelligence. Bien que les définitions de ces termes soient sujettes à discussion et qu'il n'existe pas de consensus général, les distinguer permet de caractériser des états dans le processus de l'analyse criminelle. En effet, les informations sont en permanence en mouvement et elles se modifient. La forme et le support de communication des données peuvent varier. Une donnée est considérée comme le matériau brut, initialement récolté, dont la signification n'a pas été établie (Brodeur, 2005a). Elle devient une information lorsque sa pertinence a été évaluée en fonction des objectifs de l'analyse. Une information est généralement structurée et représentée pour être analysée puis communiquée. Une connaissance est le fruit de l'interprétation de l'information en fonction de la question d'analyse. Un renseignement intègre une notion d'utilité pour l'action de sécurité impliquée (Brodeur, 2005a). Il sert de base à la prise de décision et à la résolution de la problématique. Il doit intégrer une recommandation, une idée qui permet d'agir. Le renseignement se définit alors comme une information signifiante, raisonnée et interprétée dans un contexte particulier, utile à la prise de décision.

En tant que fonction, le renseignement se définit comme la fonction d'analyse des informations afin de soutenir les missions de répression, de prévention et de gestion de crise et d'urgence (PCSCO, 2002) (Cusson, 2008). En effet, la nécessité de disposer d'informations évaluées et interprétées en temps réel a été identifiée depuis plusieurs décennies, notamment en raison de la masse et de la complexité des données à intégrer et à analyser. La mise en place d'une étape intermédiaire entre les collecteurs de données et les décideurs a conduit à la définition de la fonction de renseignement (UNODC, 2002).

Bien que dans une certaine mesure, tous les acteurs de la sécurité peuvent produire du renseignement. Le service de renseignement se distingue des autres fonctions de sécurité, puisqu'il n'a, en principe, pas d'autres missions que d'analyser et de communiquer. L'agent de renseignement apporte une vision globale et synthétique d'un problème de sécurité en suivant une démarche analytique (démarche de l'analyse criminelle ou processus du renseignement). Le renseignement est une fonction de suivi permanent et méthodique qui contribue au passage d'une stratégie d'action policière réactive vers un modèle plus proactif guidé par le renseignement. Les analyses produites peuvent servir de base à la prise de décision et guider des choix stratégiques (à long terme) ou opérationnels (à court terme). Le renseignement peut se décomposer soit selon des types d'objectifs, comme l'analyse de phénomènes criminels, soit en fonction de la nature des informations analysées, comme le renseignement financier par exemple.

Processus de l'analyse criminelle

Le processus de l'analyse criminelle est présenté ici afin d'expliciter les rôles de la visualisation dans la démarche. L'analyse criminelle est généralement décrite en tant que démarche systématique et analytique (Atkin, 2000) (Boba, 2009) (Bruce, 2009) (Paulsen et al., 2010). Le terme d'analyse signifie la décomposition d'un ensemble en ses constituants afin de permettre l'étude et l'interprétation des informations (Atkin, 2000). D'importance égale, la synthèse quant à elle vise à regrouper un ensemble d'éléments afin d'aboutir à une nouvelle information (Peterson, 1998). L'analyse criminelle est parfois définie comme « la mise en évidence méthodique de relations » (Interpol, 1997), telles que des similitudes, des régularités ou des correspondances. La détection d'une relation sous-entend à la fois une capacité à décomposer les informations en entités (tels que des événements et des personnes) entre lesquels des relations sont recherchées et la nécessité de reconstruire une connaissance globale sur laquelle se base des décisions. L'analyse criminelle couvre donc un ensemble de méthodes et de techniques exploitées dans un processus commun afin d'accomplir un ensemble d'analyses et de synthèses. Le caractère systématique de l'analyse criminelle réside dans sa démarche, fondée sur un ensemble organisé de méthodes et de principes qui tiennent en partie du domaine de la gestion de l'information.

Le processus systématique de l'analyse criminelle est connu sous le nom de cycle du renseignement (ou processus du renseignement). Ce processus est décomposé en étapes souvent considérées comme successives : planification, acquisition, évaluation, intégration (ou traitement), analyse, communication, action et réévaluation (Atkin, 2000) (Peterson et al., 2000). Il est d'usage de parler du cycle du renseignement en raison de sa nature itérative (Innes et al., 2005). Divers modèles ont été définis pour expliciter et représenter ce processus. Les diverses formalisations du processus n'intègrent pas formellement une étape de visualisation de l'information. Elle est généralement considérée comme une technique exploitée lors des étapes d'analyse et de communication. Atkin (2000) relève d'ailleurs que ce formalisme ne décrit pas les méthodes permettant d'atteindre les objectifs de chaque étape. Le schéma présenté ci-dessous ne vise pas la description du processus dans toute sa complexité. Son objectif est de situer la place de la visualisation dans le processus. Il est inspiré du modèle décrit par Peter Pirolli et Stuart Card sous le terme de « Sensemaking loop » (Pirolli & Card, 2005), en incluant la terminologie exploitée dans cette thèse.

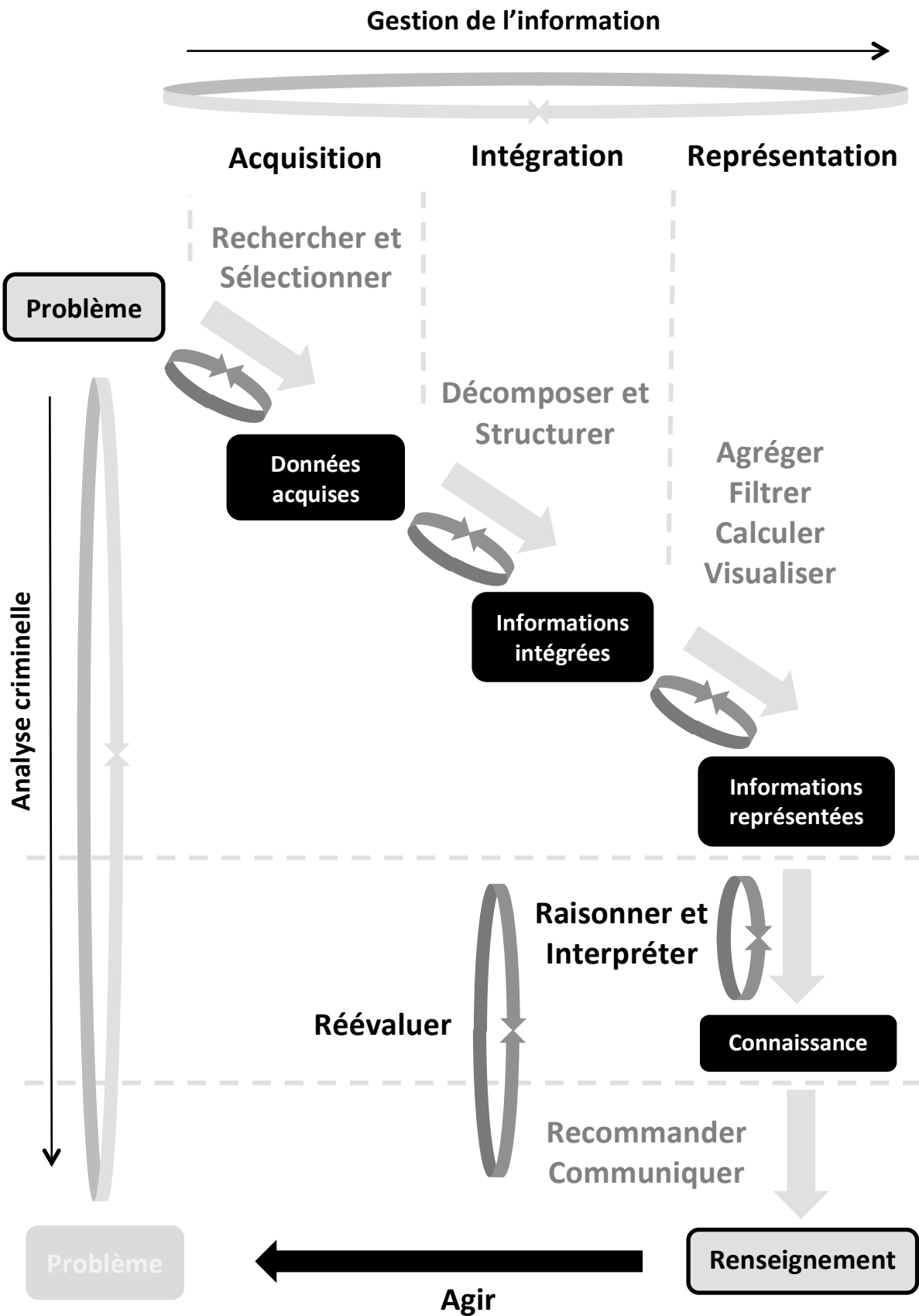





Figure 1 Processus de l'analyse criminelle

Ce schéma est structuré de la manière suivante :

-  **Quatre états** représentent les données acquises, les informations intégrées, les informations représentées et la connaissance.
-  **Cinq cycles** sont représentés pour décrire des étapes du processus : le cycle d'acquisition, d'intégration, de représentation, de raisonnement et de réévaluation.
-  **Cinq étapes** agrégées qui décrivent le processus linéaire simplifié d'élaboration du renseignement.

Le processus est itératif. Il englobe un ensemble de sous-processus représentés par des cycles sur le schéma. Des techniques spécifiques sont exploitées pour accomplir les tâches de chacune de ces étapes. Le processus est décomposé en sous-systèmes devant s'intégrer dans une architecture globale soutenant le processus de l'analyse criminelle. Toutefois, le processus n'en reste pas moins continu. L'une des difficultés majeures de la transposition technologique de ce processus est la formalisation des mécanismes de communication entre les sous-systèmes développés. Les brèves descriptions des étapes, présentées ci-dessous, soulignent certaines problématiques fondamentales du processus.

L'acquisition est une étape cruciale du processus. Elle se base sur des choix et la définition de priorités souvent imposées par le contexte (tels que les moyens à disposition ou l'urgence de la situation). La variabilité des supports sur lesquels l'information est transmise (principalement textuels, visuels et auditifs) et la diversité des formats posent de nombreux problèmes d'acquisition. Le format des données de transactions bancaires ou de communications téléphoniques récoltées lors d'une enquête par exemple, varie en fonction des prestataires de service et sont parfois transmises sous la forme papier. Une mauvaise évaluation² des besoins, de la pertinence et de la fiabilité des données récoltées peut engendrer des complications et une propagation des erreurs dans tout le continuum de l'analyse. L'accessibilité des données doit également être réglée lors de cette étape du processus. L'optimisation du processus de collecte peut faire l'objet d'une analyse complète et constitue un objectif de renseignement en soi (Ribaux et al., 2010).

L'intégration se caractérise par sa position centrale dans le processus de gestion de l'information. Elle décrit le processus de mémorisation qui comprend des étapes de décomposition des données (d'analyses) et de structuration (de synthèses). Les informations sont généralement structurées sous la forme d'entités, telles que des événements, des

² L'évaluation peut également être considérée comme une étape d'analyse entre l'acquisition et l'intégration.

12 Analyse criminelle et visualisation

personnes, des traces. L'intégration présuppose la capacité à transposer des données dans une mémoire commune en termes d'encodage, de formatage et de structuration. Les mécanismes d'alimentation, d'interrogation et de modification de la mémoire constituent les ponts entre les processus d'acquisition et de représentation décrite ci-dessous. L'intégration est l'étape globale visant la pérennisation des informations afin d'effectuer une analyse permanente. La mise en commun et le suivi du flux d'informations récoltées au cours d'une enquête ou plus globalement d'un phénomène criminel, font partie des tâches principales de l'analyste criminel.

La représentation regroupe les processus mis en œuvre pour rendre l'information perceptible. La représentation peut prendre diverses formes statiques : textuelle, imagée, graphique et symbolique. Les formes dynamiques de types olfactives, auditives et de l'image animée sont rarement exploitées dans le processus de l'analyse criminelle. La représentation présuppose en principe un encodage statique et textuel de l'information : une information intégrée (au sens du processus).

Ce travail de recherche se focalise sur l'exploitation de visualisations. Le terme de visualisation est utilisé comme synonyme de représentation³ graphique. Parmi les auteurs qui se sont intéressés de manière générale aux visualisations, Jacques Bertin est considéré comme l'un des précurseurs de la sémiologie graphique. La notion de représentation graphique décrite ici est tirée de l'un de ses ouvrages sur le sujet : *sémiologie graphique : les diagrammes - les réseaux - les cartes* (Bertin 2005, première édition en 1967).

La représentation graphique fait partie des systèmes de signes que l'homme a construits pour retenir, comprendre et communiquer les observations qui lui sont nécessaires. (Bertin, 2005)

Le terme de signe est défini comme la réunion d'un signifiant et d'un signifié. Le signifiant est la partie perceptible du signe (une image, un symbole, un mot, etc.). Le signifié est le concept mental immatériel associé au signe. Par exemple, → peut signifier flèche. Trois rôles de la représentation sont définis par Bertin. Ils sont discutés en détail dans la section suivante. L'observation est un concept général englobant les notions d'information et de connaissance.

³ Une distinction entre les termes visualisation et représentation est présentée dans la section « processus de visualisation » du chapitre « Impact sur la conception d'outils » de la deuxième partie, page 195. Ces deux termes sont toutefois utilisés comme synonymes afin d'éviter des répétitions trop lourdes.

L'information est observée ; regardée avec attention. La connaissance est une observation dans le sens où elle est le résultat de l'examen minutieux d'une information pour la comprendre. En analyse criminelle, la visualisation de l'information pourrait se décrire comme la représentation d'observations. La notion de nécessité introduite par Bertin peut probablement s'interpréter de diverses manières. Elle semble englober les notions d'intégrité et d'efficacité qui sont discutées plus loin. Bertin souligne le fait que les observations représentées doivent être utiles et requises pour atteindre un objectif défini.

La visualisation implique un certain degré de simplification et de réduction du problème analysé en regard d'une ou plusieurs dimensions particulières définies par l'objectif d'analyse (Innes et al., 2005). La simplicité de la représentation est généralement le résultat de la compréhension du problème traité par l'analyste (Peterson et al., 2000). Seuls les éléments pertinents et nécessaires pour analyser le problème sont représentés. La conception d'une représentation implique donc non seulement une habilité à produire un modèle visuel efficace, mais également une capacité d'évaluer quels sont les éléments pertinents et de les distinguer de ceux pouvant être omis. Le processus, par nature itératif, nécessite généralement de nombreux tests avant d'aboutir à une ou plusieurs représentations utiles à l'analyse. Par ailleurs, la nature évolutive des problèmes traités pose de sérieuses difficultés de représentation. Celle-ci décrit généralement une situation à un moment précis (Innes et al., 2005).

A ce stade du processus, un virage clé s'opère : la transformation d'une information en connaissance. L'analyste criminel effectue ce qu'il est courant d'appeler son analyse bien que cette terminologie engendre une confusion entre l'étape et le processus global. En fait, l'analyste raisonne sur la base des informations et les interprète pour leur donner un sens, inférer une explication, chercher des causes possibles aux effets qu'il observe. Les informations préalablement décomposées en entités d'intérêts sont examinées et comparées afin de reconstruire une vue globale du problème traité afin d'améliorer sa compréhension et formuler des hypothèses. Diverses théories sont alors exploitées. Elles relèvent de domaines variés, tels que la forensique, la criminologie et l'investigation judiciaire. Par exemple, un analyste criminel cherche à détecter des relations entre des cas, car il se base sur un ensemble de théories criminologiques ayant montré une certaine constance dans le mode opératoire des auteurs. L'analyste cherche à détecter cette régularité dans les informations qu'il traite. Il exploite alors les résultats du forensicien qui a relevé des traces, les données circonstancielle du cas (date, heure, lieu, cible, etc.), ainsi que les informations issues de l'enquête (signalement, butin, etc.).

14 Analyse criminelle et visualisation

La visualisation est exploitée lors de cette étape du processus afin de faciliter les raisonnements. Des cartes sont exploitées pour détecter des zones géographiques touchées par un phénomène criminel ou pour suivre l'activité d'un auteur particulier. Des schémas relationnels (représentation de réseaux) sont produits pour comprendre la structure d'une organisation criminelle, pour obtenir une vue d'ensemble sur un trafic de marchandises ou pour décomposer une mode opératoire complexe. Des représentations temporelles sont produites pour analyser la séquence d'une infraction particulière ou pour détecter une augmentation anormale d'un phénomène criminel. Diverses formes de représentation sont donc exploitées en analyse criminelle pour analyser les informations collectées selon divers axes qui rejoignent le questionnement Quintilien : « qui, quoi, où, avec quels moyens, pourquoi, comment et quand ? ». Les représentations cartographiques et les schémas relationnels sont les principales techniques exploitées en analyse criminelle (Innes et al., 2005). Les schémas relationnels permettent notamment d'avoir une vue d'ensemble sur les relations entre les suspects, leurs activités, les moyens utilisés et les cibles.

Finalement, **le renseignement** se distingue de la connaissance en tant que produit de l'analyse criminelle dans la mesure où il intègre une recommandation (Ratcliffe, 2008). Un renseignement vise la résolution ou au moins l'atténuation d'un problème. Il peut intégrer une proposition d'action (Atkin, 2000). Le renseignement est un « outil » de la prise de décision basé sur l'analyse des informations (dans son sens large, englobant l'ensemble du processus décrit). La visualisation des informations peut également jouer un rôle important en matière de communication entre les producteurs et les destinataires du renseignement. En effet, l'analyste n'est généralement pas celui qui prend les décisions. Il doit donc communiquer les résultats de son analyse. La représentation graphique peut alors jouer un rôle pour faciliter l'appréhension des résultats et faciliter l'intégration des renseignements dans la prise de décision. Il faut cependant relever que les produits de renseignement transmis sont souvent perçus comme des représentations précises et objectives du problème, alors qu'ils sont en fait souvent des descriptions synthétiques et simplificatrices (Innes et al., 2005). L'usage de représentation peut renforcer l'impression d'objectivité (Innes et al., 2005), conférer une plus grande crédibilité aux résultats présentés (Keim & Ward, 2007), voire concentrer l'attention sur les aspects du problème spécifiquement représentés (Kelly, 1990). Une représentation graphique peut donc influencer fortement l'appréhension d'un problème et la prise de décision. Lorsque la visualisation est inadaptée, peu efficace ou biaisée, les conséquences peuvent être dramatiques. Prendre des décisions nécessite une représentation adaptée des informations. Une bonne représentation aide à mettre en évidence des connaissances sur les mécanismes, les processus, la dynamique et les causes de problème traité (Tufte, 2005).

La visualisation est donc exploitée lors des phases de raisonnement (d'analyse) et de communication. Il faut toutefois noter qu'elle peut également être utilisée lors de l'intégration. En effet, bien que le processus formel de conception d'une représentation graphique passe par l'intégration des informations dans un format adapté (une ou plusieurs tables), certains outils⁴ exploitent la visualisation lors de l'étape de structuration des données. L'objectif de ces logiciels est de simplifier le travail de l'analyste en englobant l'ensemble des processus d'intégration et de représentation en une seule étape. La facilité d'utilisation de ces outils est remarquable. Toutefois, ils engendrent des problèmes importants. La structuration et la représentation des informations sont dynamiques. Un même type d'information peut être intégré et représenté de diverses manières, ce qui pose des problèmes de recherche, de cohérence et de lisibilité. Une telle approche bien qu'utile en terme d'efficacité (rapidité d'intégration et de représentation), ne résout pas les problématiques identifiées dans cette thèse.

R ô l e s d e l a v i s u a l i s a t i o n d a n s l e p r o c e s s u s

Les méthodes de visualisation sont exploitées dans de nombreux domaines et concentrent diverses communautés de recherche. Les rôles de la visualisation formalisés et détaillés par ces travaux sont brièvement présentés afin de compléter et préciser les rôles de la visualisation dans le processus de l'analyse criminelle.

La visualisation est l'un des supports externes développés par l'homme pour augmenter ses capacités cognitives (Norman, 1993) (Scaife & Rogers, 1996). Elle sert de mémoire et de support de raisonnement, au même titre que tous les autres supports (texte, bande son, etc.). L'une des raisons pour lesquelles les représentations graphiques sont si utiles, est que l'humain éprouve des difficultés à traiter mentalement des problèmes complexes. Les représentations graphiques facilitent la résolution de problèmes en offrant une « mémoire vive externe » (une mémoire à court terme permettant de stocker l'information pendant un raisonnement). Par exemple, des visualisations peuvent être exploitées au cours d'une enquête pour regrouper et synthétiser l'ensemble des informations collectées. L'enquêteur peut ainsi se remémorer rapidement les éléments pertinents après une absence ou lorsqu'il doit préparer une séance. Au tribunal, une représentation peut également être exploitée pour regrouper en un seul document un ensemble de données, souvent stockées dans une masse importante de

⁴ Voir notamment le logiciel TextChart de la société i2 : <http://www.i2group.com/us/products--services/analysis-product-line/textchart> (dernier accès le 26 août 2010)

16 Analyse criminelle et visualisation

rapports, procès-verbaux, etc. La visualisation sert alors de support commode évitant les efforts de mémorisation et facilitant la vue d'ensemble (Bertin, 2005).

En effet, la visualisation est particulièrement efficace par rapport à une présentation textuelle, pour diverses raisons. Dans un texte, les informations sont dispersées sur plusieurs pages, ce qui rend la vision globale des données difficile. L'information est structurée séquentiellement plutôt que simultanément (en parallèle), rendant difficile la comparaison de plusieurs variables. Une représentation visuelle est en revanche perçue selon des processus parallèles, qui facilitent la vue d'ensemble (Ware, 2004) (Atzenbeck et al., 2009a). Une représentation adéquate facilite la comparaison, la détection de différences, l'identification de régularités et de tendances, ainsi que la détection d'anomalies (Miles & Huberman, 1994) (Scaife & Rogers, 1996). Elle peut donc avoir un rôle clé lors de l'étape d'analyse.

Comme déjà mentionné, il est généralement reconnu que la visualisation peut être exploitée pour raisonner et communiquer (Card et al., 1999) (Chen et al., 2008). Certains auteurs font la distinction entre l'analyse et l'exploration des données. L'analyse est considérée comme l'évaluation d'hypothèses identifiées, alors que l'exploration vise la génération d'hypothèses et la recherche de connaissances sans idée *a priori*. Les termes d'analyse exploratoire et d'analyse confirmatoire sont également employés (Anselin & Getis, 1992). La visualisation est exploitée dans de nombreux domaines de connaissances où ces deux objectifs généraux sont reconnus.

L'exploration visuelle des données s'intègre donc dans le processus de génération d'hypothèses. Elle facilite l'appréhension, la mise en perspective et la compréhension des informations collectées afin de développer les hypothèses. Leur vérification peut également s'appuyer sur des représentations, mais également sur des techniques de la statistique, de la reconnaissance de forme, ou de l'apprentissage automatique (Keim & Ward, 2007). La visualisation présente cependant divers avantages par rapport aux techniques d'analyse automatisées ; elle permet une exploration de données non-homogènes, elle est souvent intuitive et ne nécessite pas la compréhension d'algorithmes mathématiques ou statistiques complexes, elle produit une vue d'ensemble qualitative sur les données et elle permet une analyse rapide (Keim & Ward, 2007).

A des fins de communication, il s'agira de définir quelles informations doivent être sélectionnées et quelle est la manière la plus adaptée de les représenter pour le lecteur. Les choix effectués vont généralement dépendre du contexte de la communication. Un graphique n'est en principe pas le même dans un article scientifique imprimé (où il est expliqué et

discuté), qu'un graphique présenté quelques secondes au journal télévisé. Le public, son attention, ses intérêts, le moyen de communication et le temps de lecture disponible sont autant de paramètres influençant le choix d'une méthode de visualisation dans un objectif de communication efficace (produire l'effet voulu). En effet, la visualisation offre des perspectives de communication très intéressantes dans le domaine de la rhétorique et, notamment, lorsqu'il s'agit de convaincre (de la qualité d'un produit, de la pertinence d'un propos, etc.). En se basant sur des effets de perception connus, la visualisation peut être exploitée afin de tronquer ou modifier la réalité des informations représentées. Des effets de design peuvent également être exploités afin d'accroître une certaine attractivité, notamment lorsque la visualisation est exploitée à des fins médiatiques, politiques ou commerciales.

Dans son exploitation policière et judiciaire, la visualisation doit toutefois être intègre. Elle ne doit pas biaiser l'information initiale. En effet, les décisions prises sur la base de visualisations, peuvent avoir un impact important (notamment au niveau judiciaire) ou engendrer des coûts non négligeables en termes de ressources et de temps (par exemple, lorsqu'il s'agit de mettre sur pied des opérations policières ou des mesures de prévention). Les représentations graphiques se doivent donc d'être efficaces et intègres (Tuft, 2001). Elles doivent faciliter la communication et plus particulièrement l'appréhension des renseignements en représentant la réalité des informations sans ajouter des biais.

Globalement, la visualisation peut donc être exploitée pour quatre objectifs :

Mémoriser: afin d'effectuer des comparaisons, la visualisation permet de faciliter la mémorisation en regroupant un grand nombre de données dans l'espace visuel. Un objectif de synthétisation des informations peut également être atteint à l'aide de représentations.

Explorer: l'analyse exploratoire regroupe l'ensemble des tâches de recherche et d'identification de régularités ou d'anomalies particulières dans un ensemble de données. L'objectif est de découvrir de nouvelles connaissances.

Confirmer: l'analyse confirmatoire vise à tester les hypothèses définies. Elle sert de base à la prise de décision. L'objectif est de détecter les éléments confirmant ou non les hypothèses élaborées.

Communiquer: afin de transmettre une connaissance pour faciliter la compréhension d'une problématique ou communiquer un message, convaincre.

18 Analyse criminelle et visualisation

Le tableau présenté ci-contre couvre une grande proportion des objectifs tels qu'ils ont pu être formalisés dans la littérature. Cette classification distingue les objectifs cognitifs attendus et les manières de les atteindre par le biais de processus interactifs. Une telle formalisation permet, notamment, de définir des protocoles d'évaluation. En effet, les visualisations peuvent ainsi être testées en fonction de leurs possibilités, avantages et faiblesses pour la résolution d'un objectif particulier. Bien que les types d'interactivités puissent être variés est classifiés de manières plus ou moins hiérarchisées, le formalisme des niveaux de lecture⁵ proposé par Bertin (2005) a été choisi pour les classer.

En conclusion, les rôles de la visualisation décrits dans cette section sont génériques et indépendants du problème représenté et du contexte d'exploitation. La section suivante résume les diverses formes d'analyse définies dans la littérature afin de présenter les différents contextes d'exploitation possible de la visualisation à des fins de renseignement.

⁵ Alors qu'il y a autant de types de questions que de composantes (temporelle, spatiale, etc.) dans une information, trois niveaux de lecture sont possibles : le niveau élémentaire de questionnement qui porte sur un élément particulier (à quel moment? quel date?), le niveau intermédiaire qui porte sur un groupe d'éléments (pour une telle zone, dans une telle période que s'est-il passé?) et le niveau global qui porte sur l'ensemble des éléments analysés (quelle est l'évolution générale?).

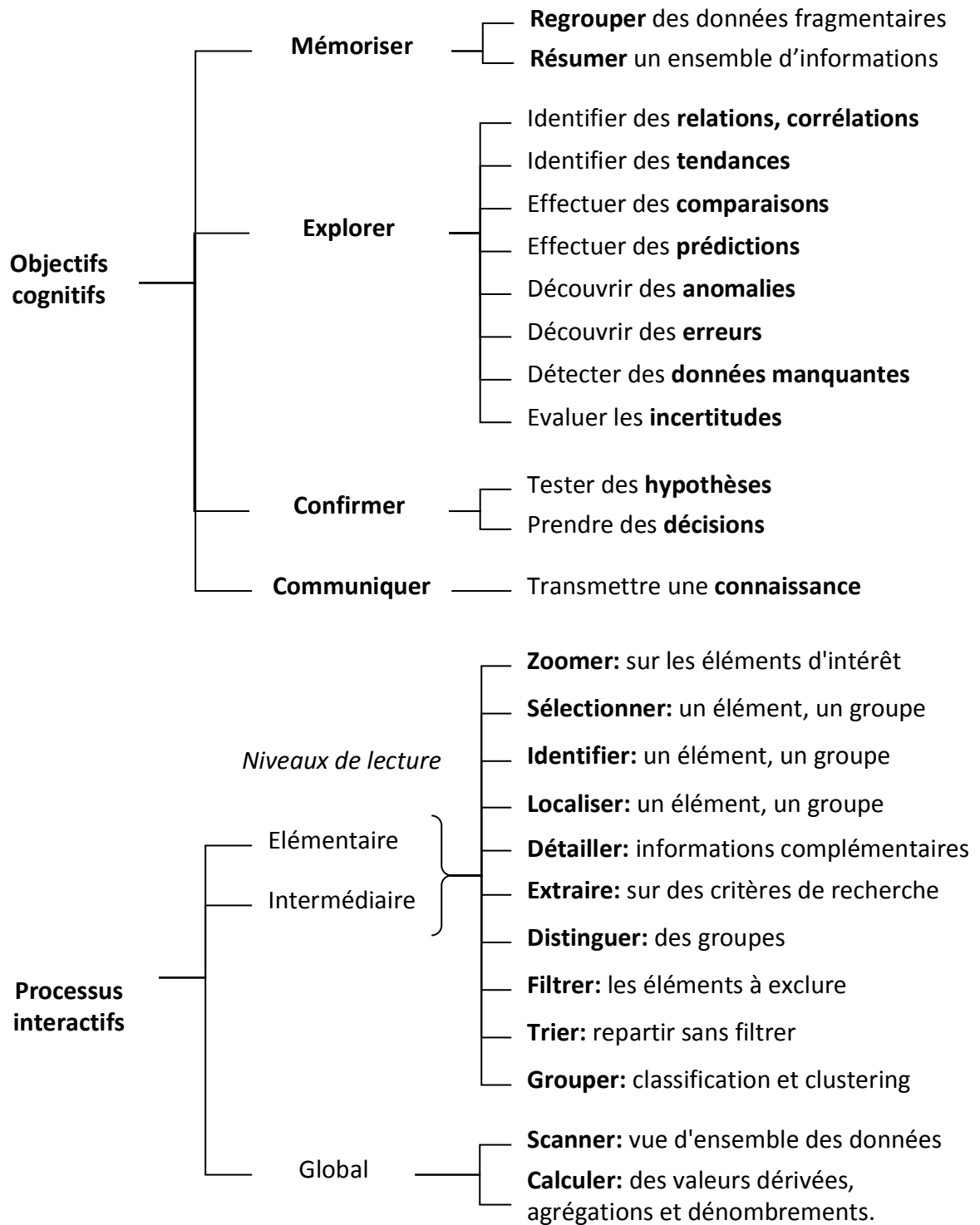


Figure 2 Classification des objectifs d'exploitation de la visualisation

Classification synthétisant les formalisations de: (Bertin, 2005), (Wehrend & Lewis, 1990), (Shneiderman, 1996), (Tweedie, 1997), (Pfitzner et al., 2003), (Amar et al., 2005), (Senator, 2005) et (Lee et al., 2006).

R e n s e i g n e m e n t e t f o r m e s d ' a n a l y s e

Le renseignement peut se décomposer selon diverses perspectives. D'une part, en fonction de la finalité de l'analyse, telle que la compréhension d'un phénomène criminel afin de mettre en œuvre des mesures de lutte ou développer de stratégies de prévention. D'autre part, en regard des informations analysées comme l'analyse financière ou l'analyse de données de facturation téléphonique par exemple. Diverses formes d'analyse ont ainsi été définies dans la littérature et reprises dans les manuels et modèles nationaux. Les produits de renseignement élaborés sont souvent accompagnés de représentations graphiques adaptées pour répondre à l'objectif d'analyse et définies en fonction de la nature des données traitées.

Certaines formes de visualisation apparaissent *a priori* plus adaptées pour certaines formes d'analyse (tel que suggéré dans le guide sur l'analyse criminelle d'Interpol (Interpol, 1997)). A titre d'exemple, l'analyse d'un phénomène criminel est souvent accompagnée d'une carte géographique représentant l'étendue du phénomène. L'analyse d'un réseau criminel s'effectue également en exploitant une représentation des relations et des rôles des protagonistes. Afin de discuter cette hypothèse, les classifications définies en analyse criminelle sont présentées. Il est globalement admis que le renseignement peut se décomposer en deux formes principales :

Le renseignement stratégique fournit une vision à long terme de l'évolution d'un problème de sécurité. Les évaluations stratégiques doivent permettre d'évaluer les risques, prévoir l'évolution à venir du problème et définir des stratégies et des priorités de lutte (en terme de prévention et de répression) (NCIS, 2000) (NCPE, 2005). Elles permettent également l'identification de lacunes à combler en matière de renseignement. En effet, l'analyse aboutit parfois à la conclusion qu'il manque des informations, que les processus et fonctionnements organisationnels sont inadaptés et qu'il est nécessaire de repenser l'action de renseignement.

Le renseignement opérationnel vise la résolution d'objectifs précis, généralement à court terme, afin de définir des actions spécifiques (comme la mise en place de surveillance, ou l'arrestation d'un suspect). Les évaluations opérationnelles proposent par l'identification de problèmes émergents, des mesures pour prévenir, perturber ou empêcher des activités criminelles (NCIS, 2000) (NCPE, 2005). De telles évaluations nécessitent un suivi permanent des problèmes et visent la définition d'opérations de police par exemple. Le renseignement stratégique se base notamment sur le renseignement opérationnel, mais dans une vision à long terme.

L'analyse criminelle a également été intégrée dans la fonction de répression au niveau de l'investigation criminelle (de l'enquête judiciaire). **Le renseignement judiciaire** peut se définir comme « *l'ensemble des informations recueillies, dans un cadre juridique déterminé, par les unités et services investis d'une mission de police judiciaire, ayant fait l'objet d'un processus élaboré dont le but est de concourir à la résolution des enquêtes ainsi qu'au démantèlement des équipes de délinquants et des réseaux relevant de la criminalité organisée* » (Carillo, 2010). En effet, la démarche de l'analyse criminelle et les techniques exploitées ont rapidement été identifiées comme efficaces et utiles pour traiter les affaires dites complexes. L'analyse criminelle n'étant que peu ou pas du tout enseignée aux enquêteurs, le métier d'analyste criminel opérationnel s'est développé en tant que fonction de soutien à l'investigation criminelle. Cette dénomination engendre de la confusion entre le rôle de soutien à l'enquête et la fonction de renseignement criminel opérationnel. Le terme d'analyse criminelle tactique est parfois utilisé pour faire cette distinction. Toutefois, il est exploité par certains auteurs et dans certains pays pour décrire la fonction de renseignement opérationnel ou le soutien à l'enquête judiciaire.

Bien que le processus de l'enquête vise la résolution de cas et la présentation des preuves à des fins judiciaires, les informations collectées par l'enquêteur, notamment lors des auditions, sont exploitables pour des objectifs plus globaux de compréhension de l'activité criminelle. A titre d'exemples, des éléments informatifs sur la structure et l'organisation des groupes criminels, sur les motivations et les processus de recel sont des informations précieuses à des fins de renseignement opérationnel, voire stratégique.

Le renseignement forensique couvre la fonction d'analyse de la trace matérielle dans un cadre plus large que l'investigation. En effet, l'exploitation classique de la trace dans l'enquête judiciaire se base sur une séquence d'inférences de base⁶ : l'identification, l'individualisation, l'association et la reconstruction (Inman & Rudin, 2001). La trace matérielle est principalement étudiée à des fins judiciaires. Elle offre des perspectives de compréhension et de reconstruction des activités criminelles, ainsi que d'évaluation du degré d'implication des suspects. La trace s'intègre donc dans un processus de collecte, d'analyse et d'évaluation visant à minimiser les erreurs judiciaires. Le renseignement forensique étend le champ d'exploitation de la trace à la détection de relations en s'appuyant sur le processus de l'analyse criminelle et selon d'autres structures d'inférences (Ribaux & Margot, 1999) (Ribaux et al., 2003) (Ribaux et al., 2006): la classification par profilage, les comparaisons de séries et de cas et la détection de patterns (de tendances et de regroupements) notamment. Cette démarche conduit les services forensiques vers l'intégration d'une nouvelle fonction de renseignement.

⁶ Un processus permettant d'aboutir à une conclusion à partir d'une observation (Kelly, 1990)

22 Analyse criminelle et visualisation

Ce changement de perspective est notamment observé en Angleterre depuis les années nonante. La contribution de la trace en termes de renseignement pour assister l'enquête judiciaire (notamment dans sa phase initiale) est dès lors considérée comme principale (Barclay, 2009). Cette vision étendue du rôle de la trace implique des changements sur la manière d'appréhender la valeur informative des traces et, parfois, de modifier leur gestion et leur analyse.

Ces quatre formes d'exploitation de l'analyse criminelle ne décrivent certainement pas de façon exhaustive et précise l'ensemble des fonctions de renseignement. La décomposition entre le niveau stratégique et le niveau opérationnel (tactique) est globalement reconnue (Interpol, 1997) (NCIS, 2000) (GIWG, 2004) (Boba, 2009). Par contre, diverses approches de classification plus fines de l'analyse criminelle ont été élaborées. Elles décomposent l'analyse en fonction de missions spécifiques. Un analyste seul ne saurait assumer l'ensemble des tâches qui nécessite de s'appuyer sur la méthode de l'analyse criminelle. La diversité des modèles définis est certainement issue des différences de structures des organisations et des stratégies de lutte. Divers modèles ont donc été formalisés et sont probablement difficilement transposables d'une organisation à l'autre. Par ailleurs, les formes actuelles décomposent l'analyse par domaine de connaissances (comme l'analyse criminelle financière ou l'analyse de communications téléphoniques).

L'ensemble de ces classifications sont issues de la littérature anglo-saxonne, à l'exception du modèle proposé par Interpol qui est présenté en introduction (Interpol, 1997) :

	Stratégique	Opérationnel
Infraction	<p>Analyse de phénomène de criminalité (nature, ampleur et développement de la criminalité)</p>	<p>Analyse de cas (exploitation dans l'investigation criminelle, reconstruction du cas)</p> <p>Analyse comparative de cas (comparaison des cas, afin de les lier à un individu ou un groupe)</p>
Auteur	<p>Analyse de profil général (identification de caractéristiques communes aux individus ayant commis le même type de cas)</p>	<p>Analyse de profil spécifique (identifier les caractéristiques du ou des auteurs d'une infraction)</p> <p>Analyse de groupes d'auteurs (comprendre la structure, les rôles d'un groupe criminel)</p>
Méthode	<p>Analyse de méthode générale (évaluation de la méthode de lutte, en vue de l'améliorer)</p>	<p>Analyse d'enquête (évaluation des tâches accomplies au cours d'une enquête, afin de la faire progresser. Identifier des problèmes dans une enquête pour ne pas les reproduire)</p>

Les formes définies dans la littérature anglo-saxonne sont nombreuses et leur classification hiérarchisée n'est pas évidente. Elles se distinguent et se recoupent selon diverses dimensions. Une représentation a été élaborée pour regrouper ces formes d'analyses et les contextualiser. Elles sont organisées au sein de cinq types généraux et positionnées en fonction des relations qu'elles entretiennent entre elles. Ces relations ne sont pas explicitées sur le schéma, car elles sont trop nombreuses et surchargeraient la représentation. Une brève description de chaque forme d'analyse est incluse.

24 Analyse criminelle et visualisation

(Bruce, 2009): Exploring Crime Analysis: Readings on Essential Skills

(Boba, 2009) : Crime Analysis and Crime Mapping

Crime analysis : étude du crime afin d'identifier et analyser des patterns, des tendances et des problèmes

Crime Intelligence analysis : étude des criminels et des structures criminelles

Criminal Investigative analysis : création de profils physiques, comportementaux et psychologiques

Police Operations analysis : études des stratégies organisationnelles internes de la police

Administrative analysis : présentation d'études à des fins administratives, politiques et de communication

Continuum de l'analyse criminelle : de l'opérationnel au stratégique

Guide de lecture

Certaines formes sont reprises dans plusieurs ouvrages, ou implicitement décrites. Pour simplifier la représentation, une seule référence est indiquée (par la couleur).

L'objectif est de montrer que les variabilités décrites par ces formes d'analyse couvrent plusieurs dimensions.

La position des formes d'analyses sur le plan n'est pas rigoureuse. Elle soutient la contextualisation des termes.

Crime analysis

Analyse de cas

Flow analysis

Event-Flow analysis
Sequence of events in a criminal activity

Suspect Identification
Attempt to identify and locate the suspect

Target/Victim Profile analysis
Types of persons, structures, vehicles

Temporal analysis
When events occurs, how long they last, their frequency and regularity

Geographic distribution analysis

Geographic analysis
To predict, disturb and disrupt criminal activity

Crime Series

A crime pattern where there are reasons to believe that same person or persons committed the crimes

Analyse comparative de cas

Case Matching
Link arrested people to other offences

Criminal Investigative analysis

Investigative analysis
Physical, behavioral and psychological profile of suspect

Target Profile analysis
Describe he criminal, criminal activity, lifestyle, associations, risks to target him

Analyse de profil spécifique

Geographic profiling
Prioritize suspect investigations based on the geographic distribution of serial crimes

Criminal Business Profile
How criminal work and choose victims

Crime Intelligence analysis

Activity-Flow analysis
Modus operandi

Content analysis

Communication a.
Oral and written communication

Bank-Record analysis

Commodity-flow a.

Financial analysis
Pattern detection and identification to discover a criminal activity

Police Operations analysis

Case analysis

Analyse d'enquête

Operations analysis
Needs and Uses of internal ressources

Visual-Investigative a.
Steps taken in the course of an investigation

(Gottlieb et al., 1994) : Crime analysis: From first report to final arrest

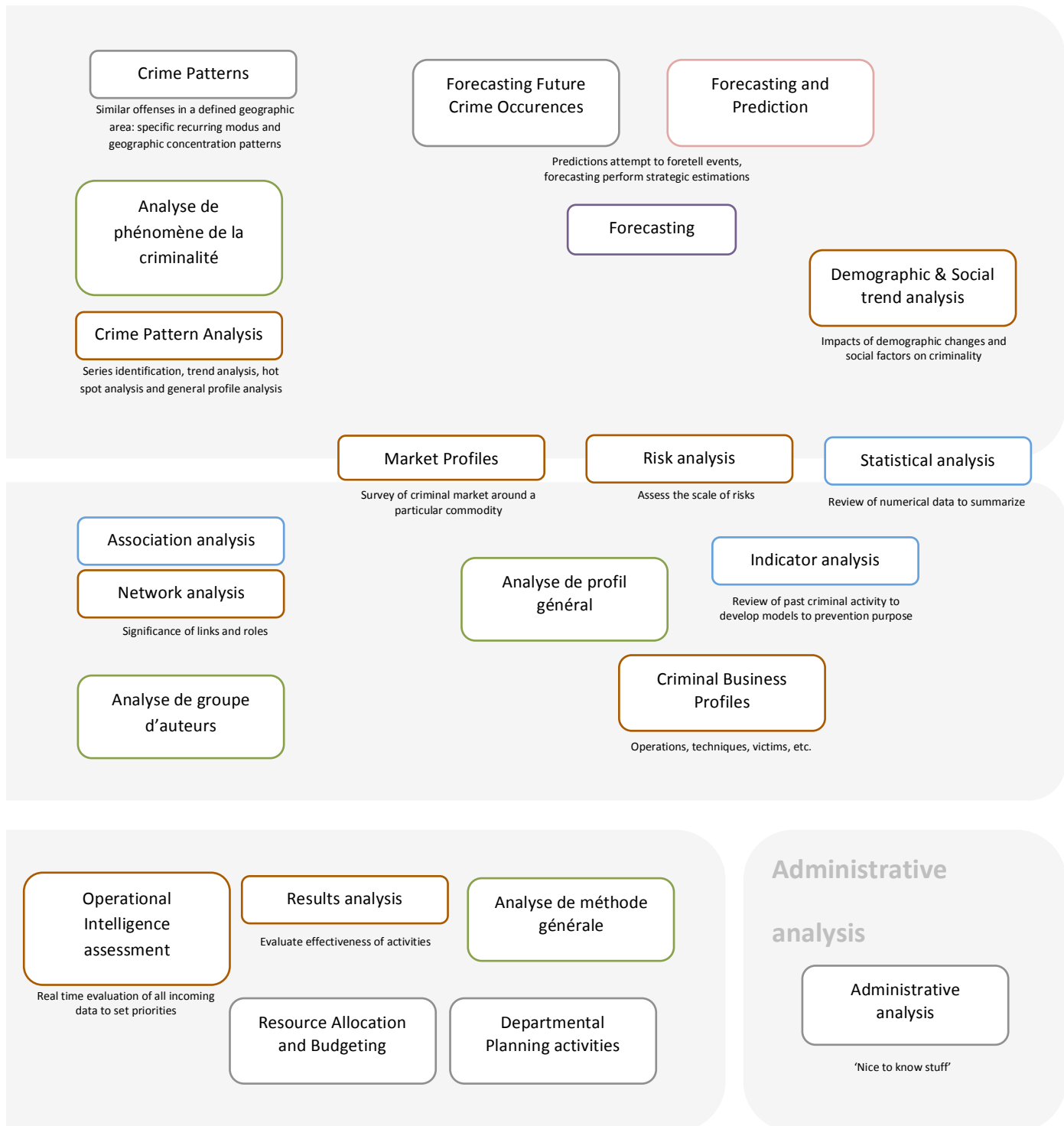
(Interpol, 1997) : **Guide sur l'analyse criminelle**

(NCIS, 2000) : **The National Intelligence Model (NIM)**

(Peterson et al., 2000) : **Intelligence 2000: Revising the Basic Elements; A Guide for Intelligence Professionals**

(IALEIA, 2004): **Law Enforcement Analytic Standards (reprend les définitions du NIM complétées par d'autres formes d'analyses)**

(Paulsen et al., 2010): **Tactical Crime Analysis: Research and Investigation**



A n a l y s e , p r o f i l e t p a t t e r n

Un vocabulaire récurrent se dégage des formes de l'analyse criminelle. En effet, les termes de profil et de pattern sont largement utilisés pour décrire ce qui semble être des produits de l'analyse. Les deux notions sont discutées ici afin de clarifier la manière dont elles seront utilisées dans cette thèse. Par ailleurs, la reconstruction d'un profil et la détection de patterns sont des tâches récurrentes de l'analyse criminelle qui peuvent être facilitées par l'utilisation de représentation graphiques. Cet aspect est brièvement introduit en fin de section.

En analyse criminelle, le terme de pattern est généralement utilisé pour décrire un ensemble de cas. Le terme de profil quant à lui est principalement exploité pour décrire un auteur ou un groupe d'auteurs. Ces nuances terminologiques apparaissent comme le résultat de la décomposition du domaine en deux communautés : les analystes du crime et les analystes du criminel. De manière plus générale, ces deux termes décrivent toutefois des concepts différents. Le terme de pattern a notamment un double sens qu'il s'agira de ne pas confondre. Un pattern décrit un schéma récurrent ou une structure particulière. Il sera toutefois également utilisé dans cette thèse pour décrire une solution adaptée pour un problème récurrent de conception⁷ (en l'occurrence, pour l'élaboration d'une représentation graphique utile pour un problème d'analyse récurrent).

Un profil peut se définir comme la description d'un ensemble de similitudes observées entre des personnes ou des objets. Il peut être utilisé pour identifier ou représenter un individu ou un groupe (Ribaux, 2008). En matière d'analyse comparative de cas, un profil décrit un ensemble d'événements pouvant être de deux types : les séries et les groupes (Ribaux, 1997). Une série décrit un regroupement de cas dont les similitudes permettent d'inférer un auteur ou un groupe d'auteurs commun, voire une cause commune en matière de sécurité (par exemple, dans le cas de l'analyse d'une série d'incendies) (Ribaux & Margot, 2008). La notion de groupe est plus générale. Un groupe est un sous-ensemble de cas partageant certaines similitudes (spatio-temporelle, mode opératoire, cible, butin, etc.). Un profil peut être spécifique à une dimension d'analyse ou en inclure plusieurs (profil géographique ou profil spatio-temporel par exemple). Lorsque l'analyse porte sur un auteur ou un groupe d'auteurs, le profil décrit un ensemble de traits caractéristiques de l'auteur ou de son activité : profils psychologiques, géographiques, etc. Dans le domaine forensique, un profil chimique ou physique représente un ensemble de composés ou de caractéristiques identifiés et quantifiés. Il apparaît que dans tous ces usages, le profil est le résultat d'un processus de reconstruction.

⁷ Cf. Patterns de visualisation relationnelle, page 144.

Les éléments constitutifs du profil sont sélectionnés et évalués selon leur pouvoir discriminatoire (leur capacité à distinguer). Ils définissent la spécificité du profil.

Un pattern décrit un schéma récurrent ou une structure particulière. En criminologie, la théorie des patterns (Brantingham & Brantingham, 1993) décrit les régularités des activités criminelles : les crimes sont regroupés en pattern, les décisions prises par les criminels suivent des patterns, et des patterns décrivent des processus mis en œuvre pour commettre les crimes. La théorie des patterns fait partie d'un ensemble de théories regroupées dans le domaine de la criminologie environnementale. La théorie des choix rationnels (le criminel raisonne en fonction des coûts et des bénéfices engendrés par la commission du crime afin d'effectuer ces choix) et la théorie des activités routinières (le crime s'intègre dans les activités habituelles de l'auteur et de la victime) en font partie également (Felson & Clarke, 1998). Comme souligné précédemment, ces théories criminologiques ont un impact important sur l'analyse criminelle. Si les crimes et les criminels suivent des schémas récurrents, il doit être possible de les détecter, formaliser et représenter. Ces patterns pouvant alors servir de base pour le développement d'évaluations opérationnelles ou stratégiques. En matière d'analyse du crime, un pattern décrit un ensemble de cas partageant des caractéristiques communes, telles que le type de délit, le mode opératoire, le type de cible, une même zone géographique, etc.

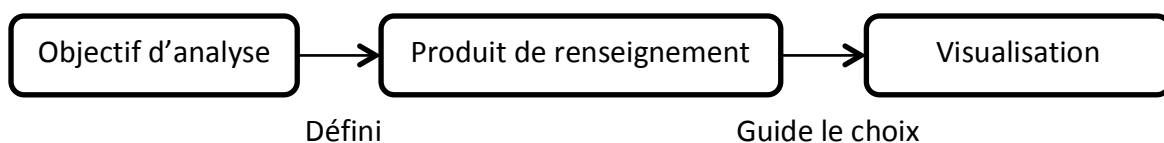
Dans le processus d'analyse, un pattern peut être détecté ou être exploité comme inférence lorsqu'il a été formalisé (Senator, 2005). D'un point de vue cognitif, un pattern décrit une connaissance *a priori* exploitée pour interpréter et catégoriser de nouvelles observations (Mennis et al., 2000). L'analyste infère qu'une observation fait partie d'un pattern lorsqu'il observe des caractéristiques constitutives de la définition du pattern. Un pattern est donc une construction, un ensemble de relations qu'il est nécessaire de détecter. Une relation est le résultat de l'observation d'une correspondance, d'une similarité ou d'une régularité particulière par exemple. Le processus d'analyse englobe le processus de détection de pattern et la comparaison d'une nouvelle information avec un pattern connu. Tel que décrit précédemment dans la section « Rôles de la visualisation dans le processus », les termes d'analyse exploratoire (ou analyse basée sur les données) et d'analyse confirmatoire (ou analyse basée sur un modèle) sont parfois utilisés pour décrire ces deux processus (Anselin & Getis, 1992).

Les notions de profil, de pattern et de relations sont donc fortement liées. Un profil se distingue toutefois d'un pattern par sa complexité intrinsèque et sa valeur informationnelle. Il peut décrire un ensemble de relations et de concepts variés. Le profil décrit un groupe dont la

28 Analyse criminelle et visualisation

spécificité est suffisante pour être pertinente en termes d'analyse de la criminalité. L'objectif est d'obtenir la meilleure représentation possible du groupe défini. La construction d'un profil est une inférence en elle-même (Ribaux & Margot, 1999), mais sa spécificité peut également conduire à inférer l'identité. C'est le cas, par exemple, d'un ensemble de loci formant un profil ADN. Dans cette thèse, lorsque le niveau de spécificité d'un profil permet d'inférer une source commune ou un type de source commun, le terme de relation est utilisé (p. ex. les relations ADN, les relations selon le motif de trace de semelles, etc.). Le terme de pattern est exploité dans son sens de description d'une certaine forme de régularité.

Diverses formes de représentation peuvent être exploitées pour représenter des profils, des patterns ou des relations⁸. Par exemple, les diagrammes à coordonnées parallèles (présentés dans l'annexe un) ont été inventés spécifiquement pour l'analyse globale d'un ensemble de variables (comme le temps, le lieu, le type de cible, etc.) pour identifier des profils particuliers. La description générale d'un mode opératoire ou le profil d'une activité criminelle particulière peuvent être représentés par des schémas d'activités. Par ailleurs, des régularités spatiales peuvent être détectées en exploitant des cartes. Lors de la représentation de transactions financières par un schéma de flux, des mécanismes financiers récurrents (des patterns) peuvent être détectés. Finalement, les relations identifiées entre les membres d'un réseau criminel sont généralement représentées au moyen d'un schéma relationnel qui facilite la compréhension de la structure du réseau et la formulation d'hypothèses sur les rôles de chaque protagoniste. Des formes de visualisation particulière peuvent donc être exploitées pour représenter des relations particulières, pour détecter des patterns, voire des profils.



L'objectif d'analyse définit le type de produits de renseignement élaborés. Celui-ci guide le choix d'une ou plusieurs formes de représentation adaptées. La visualisation en tant que telle ne saurait être considérée comme le produit de l'analyse, mais en est certainement une partie centrale sur laquelle peuvent se baser les décisions.

⁸ Les différents types de représentation sont présentés dans la section suivante et dans l'annexe un.

A n a l y s e s e l o n l e t y p e d e d o n n é e s

L'objectif d'analyse est intimement lié à la nature du problème abordé. Les formes d'analyse se décomposent d'ailleurs en fonction du type de données traité. L'ensemble des modèles classifient l'analyse selon qu'elle traite des infractions, des personnes ou des méthodes. L'analyse des méthodes (analyse d'enquête, d'opérations) ne traite pas directement de l'activité criminelle, mais de la réponse policière. Cette forme d'analyse traite des questions liées à l'allocation et l'efficacité des ressources, afin d'évaluer et définir de nouvelles stratégies organisationnelles ou de nouvelles mesures policière (Bruce, 2009). En tant qu'objet de renseignement, la méthode policière ne fait pas partie de l'analyse de la criminalité.

La décomposition de l'analyse criminelle selon qu'elle porte sur les cas ou sur les auteurs a été largement discutée dans la littérature (Ratcliffe, 2008) (Bruce, 2009). Il ressort de ces discussions que les compétences et processus de ces deux formes générales se rejoignent. Les différences se situent au niveau des objectifs et des missions d'analyse. Les structures policières et les stratégies d'action de sécurité mises en œuvre influencent notamment la définition de ceux-ci. Aux Etats-Unis, « *The Crime Analyst* » étudie les événements criminels afin de détecter des patterns permettant de définir des stratégies de prévention et des mesures de suppression ou de perturbation de la criminalité. « *The Criminal Intelligence Analyst* » étudie les organisations criminelles afin de soutenir le processus de l'investigation criminelle (Bruce, 2009). Cette décomposition reflète un choix stratégique national important. L'investigation criminelle est choisie comme mesure de lutte prioritaire lors de crimes graves. Alors que les réponses policières face aux délits sériels liés aux infractions contre le patrimoine sont préventives ou axées sur la résolution de problèmes : « *Local crime analysis identifies the location of crime problems, criminal targets and vulnerable victims to prevent and reduce crime, while investigative analysis assists with solving and the prosecution of offenders by providing information for presentation at court* » (Cope, 2003). L'analyse criminelle s'intègre donc dans un contexte complexe qui a influencé la définition des formes d'analyse. Comme en témoigne le diagramme présenté précédemment, l'image globale actuelle de ces différentes formes est difficilement structurable.

D'autres formes d'analyse sont définies selon la nature des problèmes de sécurité traités. C'est le cas par exemple, de l'analyse financière criminelle : « *L'analyse Financière Criminelle (AFC) est un ensemble de méthodes techniques et d'outils appliqués à l'investigation économique et financière. Elle consiste principalement à l'exploration, l'analyse et l'exploitation de données, d'origines et de nature différentes, destinée à donner une représentation spatio-temporelle, des liens existant et de rapprochement entre des entités et des faits.* » (AGIS, 2004) En suivant ce modèle, il serait alors possible de définir d'autres types d'analyse, tels que l'analyse criminelle des délits contre le patrimoine, de la criminalité

organisée, des délits contre l'intégrité corporelle, ou encore l'analyse criminelle de trafics de stupéfiants. Toutes ces formes d'analyse ont une réalité, mais une telle décomposition ne semble pas adéquate pour classifier les méthodes exploitées. En effet, les méthodes et techniques ne se différencient pas forcément en fonction du sujet traité. Des formes de représentation particulières sont effectivement plus utiles pour certains problèmes spécifiques. A titre d'exemples, un schéma de flux permet de représenter un trafic de stupéfiants et un schéma d'événements est exploité pour représenter une série de cambriolage. Une telle classification est utile afin de structurer des connaissances particulières liées à la démarche et à l'exploitation des techniques pour des situations d'analyses spécifiques. Toutefois, il est reconnu que ces formes d'analyse ne permettent pas de classifier les méthodes et techniques de l'analyse criminelle de manière globale (Bruce, 2009). En effet, les méthodes de traitement, de gestion et de représentation des informations sont exploitées de façon transversale. Un schéma relationnel peut notamment être utilisé pour représenter les relations existantes entre des événements, mais également pour représenter un réseau criminel. Une classification plus globale des formes d'analyse est nécessaire afin de guider les choix à effectuer en matière de visualisation.

D i m e n s i o n s d ' a n a l y s e e t f o r m e s d e v i s u a l i s a t i o n

L'objectif de cette section est de proposer une décomposition des formes d'analyse utile pour sélectionner une méthode de visualisation adaptée en vue de résoudre les questions posées. Certains auteurs suggèrent que le choix d'un graphique est essentiellement déterminé par la nature des données et parfois par l'objectif de la présentation (Pattyn & Dickens, 2001). L'approche décrite ici, quant à elle, part d'une décomposition de l'objectif d'analyse selon une ou plusieurs dimensions dominantes pour déterminer le type de représentation à concevoir. Dans le domaine de l'analyse criminelle, la définition des dimensions d'analyse est visiblement une tâche difficile :

"Event visualization can be defined as the visual representation of spatial, temporal and other dimensions of events... Events involving more than one dimension frequently occur because humans often handle multiple event dimensions concurrently with their various senses. These events require powerful visualization capabilities to display various dimensions (e.g., space, time, person, event aggregation)." (Chung et al., 2005)

"Typical crime data will have either a geographical nature, and/or contain a large number of features, and/or involve networks of offenders. Types of technologies relevant to these different aspects are: geocoded data, large feature sets, offender networks." (Oatley et al., 2006)

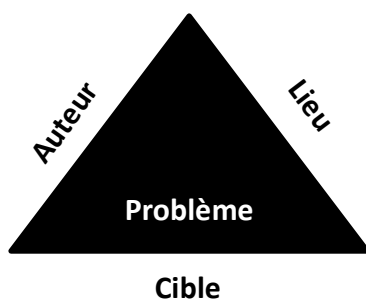
Il est intéressant de noter que les analystes spécialisés dans l'analyse de réseaux criminels ne définissent pas le temps comme une dimension dominante (le modèle de l'IALEIA (Peterson et al., 2000) ne décrit pas formellement de *temporal analysis*). La notion de dominance semble être une des clés de décomposition. Une dimension est jugée comme dominante si elle impose les plus fortes contraintes sur l'analyse (Mennis et al., 2000). Par exemple, si l'analyse porte sur la distribution géographique d'un phénomène criminel pour une période donnée, la dimension spatiale est la dimension dominante (ou principale). En effet, dans cet exemple, le temps est invariant (une seule période). En revanche, si la question porte sur la chronologie des événements, le temps devient alors la dimension dominante.

Globalement, deux dimensions sont décrites de manière récurrente et semblent dominantes : la composante temporelle et la composante spatiale. Ces deux dimensions ne sont toutefois pas suffisantes pour décrire l'ensemble des questionnements impliqués lors de l'analyse de données de criminalité. Une troisième dimension principale est définie : la composante relationnelle. De manière générale, ces trois dimensions d'analyse semblent être pertinentes et fondamentales pour traiter et interpréter les phénomènes de criminalité (Oatley et al., 2005) (Ribaux, 2008). Bien qu'une relation puisse être de nature temporelle ou spatiale, il est nécessaire de définir une dimension au sein de laquelle d'autres formes de relations se distinguent. Adrienko et Adrienko définissent cette troisième dimension sous le terme de population : un groupe d'entités pouvant être de diverses natures (Andrienko & Andrienko, 2005). D'autres auteurs parlent d'une dimension thématique complémentaire aux dimensions temporelle et spatiale (Mennis et al., 2000) ou d'une dimension sémantique (Pelekis et al., 2004), iconique (MacEachren, 2004), voire en catégories (Fredrikson et al., 1999). Elle décrit un ensemble d'observations mesurables « constitutives » d'une entité (Mennis et al., 2000). La notion d'entité est récurrente dans l'ensemble des définitions de cette dimension. Cette décomposition en trois dimensions est d'ailleurs également utilisée pour décrire la manière dont l'humain mémorise les connaissances selon trois structures : savoir-quoi, savoir-quand, savoir-où (Mennis et al., 2000).

Définir formellement la dimension relationnelle n'est pas une tâche facile. Afin de tenter de l'explicitier pour le domaine de l'analyse criminelle, les différents types de régularités détectées dans cette dimension sont décrits. Rachele Boba définit sept types de patterns principaux

32 Analyse criminelle et visualisation

(Boba, 2009): la série⁹, le « spree » (ensemble de cas perpétrés sur une période très courte par un individu ou un groupe spécifique ; plus spécifique qu'une série qui peut durer sur une période plus ou moins longue), la multi-victimisation (d'un même individu ou d'un ensemble de victimes partageant des caractéristiques communes), la récidive (auteur ou groupe d'auteurs), le « Hot spot » (point chaud : lieu spécifique ou région peu étendue où un nombre inhabituel de cas est observé sur une courte période), la « Hot target » (types de lieu, de cible fréquemment touchés) et le « Hot product » (type de bien fréquemment ciblé par un type de criminalité). Les formes de régularités formalisées ne sont pas mutuellement exclusives et elles intègrent plusieurs dimensions (notamment les dimensions spatiales et temporelles). Les régularités sont également recherchées parmi les attributs des auteurs, des cibles (victimes et objets) et des lieux (pas seulement au niveau de leurs positionnements géographiques, mais également de leurs spécificités). Cette décomposition rejoint le formalisme du « problem analysis triangle »¹⁰ (Felson & Clarke, 1998) (Eck, 2003).



La dimension relationnelle pourrait se définir comme intégrant les variabilités observées et analysées entre le problème (l'activité criminelle / l'événement), l'auteur, la cible (personne ou objet) et le lieu. Cette dimension ne couvre pas l'analyse des distributions temporelles et spatiales, mais les relations observées entre les différentes caractéristiques de ces quatre entités génériques.

Une quatrième dimension dominante semble nécessaire pour classifier les méthodes de visualisation de manière efficace en regard des types de questionnements : la composante quantitative. En effet, de nombreuses questions sont du type : « quel est le nombre de ... ». S'il fallait illustrer graphiquement ce propos, il ne serait pas nécessaire de chercher une visualisation de type relationnelle, spatiale ou temporelle. Un nombre important de visualisation a été spécifiquement développé pour l'analyse quantitative. Il est vrai que le dénombrement est très fréquemment mis en regard d'une dimension complémentaire : quel est le nombre de cas par région, quel est l'évolution temporelle du nombre de victime, etc. Dans ces cas, se pose inévitablement la question de la dominance : quelle est la dimension principale ? Mais les questions de type « combien » sont néanmoins fréquentes et impliquent des représentations propres.

⁹ Cf. Analyse, profil et pattern, page 26

¹⁰ <http://www.popcenter.org/about/?p=triangle>, dernier accès le 7 septembre 2010

Dimensions d'analyse

Les questions d'analyse semblent pouvoir se décomposer selon quatre dimensions principales. C'est-à-dire quatre composantes dans lesquelles la variabilité est observée et qui sont récurrentes dans les questions d'analyse.

La dimension temporelle couvre les questions liées à l'analyse d'une distribution temporelle. Le temps est la composante principale de l'analyse. Les réponses aux questions posées doivent se chercher dans la variabilité temporelle : *quand, sur quelles périodes, à quelle fréquence, selon quelle régularité temporelle, etc.*

La dimension spatiale couvre les questions liées à l'analyse d'une distribution géographique. L'espace est la composante principale de l'analyse. Les réponses aux questions posées doivent se chercher dans la variabilité spatiale : *où, dans quelle région, selon quel parcours, quelle étendue géographique, etc.*

La dimension relationnelle couvre les questions liées à l'analyse des relations entre les entités pertinentes, telles que les événements, les personnes, les objets et les traces. La relation est la composante principale de l'analyse. Les réponses aux questions posées doivent se chercher dans la variabilité relationnelle : *qui, quoi, entre qui, entre quoi, avec quoi, etc.*

La dimension quantitative couvre les questions liées au dénombrement. Le nombre est la composante principale de l'analyse. Les réponses aux questions posées doivent se chercher dans la variabilité quantitative : *combien de, selon quelle proportion, etc.*

Pour chaque dimension, des représentations spécifiques sont exploitées. Cette décomposition selon quatre formes de questionnement principales permet de classifier et étudier les diverses formes de représentation exploitables en analyse criminelle. En effet, les formes de représentation classifiées dans les recherches sur la visualisation rejoignent ces dimensions. Elles sont présentées dans la section suivante.

Formes de visualisation

Selon Bertin (2005), on ne saurait analyser un système de signes sans le délimiter strictement : *la graphique recouvre l'univers des réseaux, celui des diagrammes, et l'univers des cartes.*

Cette délimitation stricte exclut de la représentation graphique les écritures musicales, verbales et mathématiques, ainsi que la symbolique et l'image animée. Il apparaît inutile et certainement difficile d'expliciter les définitions sémiologiques de ces formes. Celles-ci distinguent ces trois types de représentation en fonction des possibilités d'exprimer les relations entre les éléments des variables représentées sur le plan. Des définitions plus simples sont présentées :

- *Les cartes* regroupent l'ensemble des représentations dites cartographiques, c'est-à-dire dont les dimensions du support sont exploitées pour représenter les dimensions de la localisation. En analyse criminelle, les cartes sont évidemment exploitées pour effectuer des analyses spatiales.
- *Les diagrammes* forment un ensemble de visualisations communément appelé des « graphiques », tels que les diagrammes en « barres » (histogrammes), « en camembert » (diagramme circulaire sectorisé) et les « nuages de points » (diagramme de dispersion). Les diagrammes sont utilisés à la fois pour effectuer des analyses quantitative et des analyses temporelles.
- *Les réseaux* sont des constructions basées sur la représentation de sommets reliés par des arêtes. Selon Bertin (explicitation libre de la définition), une visualisation est un réseau lorsque tous les éléments d'une des dimensions représentées sur le plan peuvent être en relation. Ils offrent donc une liberté d'expressivité, pour une dimension donnée, supérieure aux autres formes de graphique. Ils font l'objet d'études approfondies subséquentes aux travaux de Bertin. En analyse criminelle, il est d'usage de parler de schémas relationnels lorsque des réseaux sont élaborés pour effectuer des analyses dans la dimension relationnelle.

Cette classification tripartite des méthodes de visualisation proposée par Bertin (2005) est la seule taxonomie trouvée dans la littérature qui est uniquement basée sur la représentation graphique et qui permet une classification simple. En effet, les autres classifications (voir

notamment (Lohse et al., 1994), (Card & Mackinlay, 1997) et (Card et al., 1999)) décrivent des classes en fonction de la nature des données, des tâches à accomplir, ou proposent des sous-catégories de la taxonomie de Bertin. Les icônes et les images sont parfois également considérées comme des méthodes de visualisation (Lohse et al., 1994).

Effectuer un inventaire exhaustif des méthodes de visualisation est d'envergure encyclopédique (comme en témoigne l'ouvrage de Harris (Harris, 2000)). La diversité et le nombre de méthodes de visualisation existantes complique considérablement le processus de classification. A titre d'exemple, les sites Internet répertoriés ci-dessous, illustrent l'étendue (inévitablement incomplète) des possibilités et diverses approches de classification:

<http://www.math.yorku.ca/SCS/Gallery/> (*Gallery Data Visualization*)¹¹

"*The Gallery of Data Visualization*" regroupe un ensemble de visualisations liées à l'analyse statistique. Les représentations sont présentées historiquement et selon une évaluation critique de leurs performances (essentiellement basée sur les travaux de Tufte (2001)).

<http://otal.umd.edu/olive/> (*Olive Visualization*)

Le projet OLIVE "*On-line Library of Information Visualization Environments*" regroupe des produits, des projets de recherche, des citations et des vidéos, selon une classification basée sur le type de données visualisées (Shneiderman, 1996): données temporelles, uni-, bi- et tridimensionnelles, multidimensionnelles, arbres et réseaux. Le site a été développé, en 1997, par une classe du professeur Schneiderman, de l'université du Maryland.

<http://www.wikiviz.org/wiki/Techniques> (*Wikiviz*)

Wiki créé et maintenu par Riccardo Mazza, professeur à l'Institute of Innovative Technologies at the University of Applied Sciences of Southern Switzerland. Les représentations sont classifiées en sept catégories: données uni-, bi- et tri-variées, données multi-variées, données temporelles, données spatiales, données hiérarchiques, réseaux et données textuelles.

http://www.visual-literacy.org/periodic_table/periodic_table.html (*Periodic Table Visu*)

Lengler et Eppler proposent une classification complexe et dense des méthodes de visualisation en exploitant une analogie avec le tableau périodique des éléments, développé dans le domaine de la chimie (Lengler & Eppler, 2007). Une lecture interactive sur le site web

¹¹ URL : *mots-clés Google*© (premier résultat), (Derniers accès pour l'ensemble des sites, le 01 septembre 2009)

36 Analyse criminelle et visualisation

permet de découvrir des exemples pour chaque catégorie. Alors que les détails de cette taxonomie ne seront pas discutés ici, il est intéressant de relever l'objectif défini pour ce projet. Les auteurs proposent une taxonomie permettant au lecteur d'identifier rapidement une méthode adaptée pour une tâche particulière dans un domaine d'activité défini (le management). Cette taxonomie, principalement axée sur la métaphore graphique employée, combine des classifications liées à la nature des données et aux tâches à effectuer.

<http://manyeyes.alphaworks.ibm.com/manyeyes/> (*Many Eyes*)

Le projet "Many Eyes", développé par le Collaborative User Experience research group d'IBM, offre une interface en ligne de création et de partage de visualisations. Des types de visualisations sont classifiés dans six catégories principales définissant des types de tâches : voir les relations entre des données, comparer des ensembles, suivre des tendances temporelles, analyser des portions, analyser du texte et étudier l'ensemble des données.

Cette thèse se focalise sur les représentations de l'analyse relationnelle. Elle est présentée dans la section suivante. Les dimensions temporelle, spatiale et quantitative et leurs formes de représentations respectives sont présentées dans l'annexe un.

A n a l y s e e t v i s u a l i s a t i o n r e l a t i o n n e l l e

L'analyse relationnelle est centrale en analyse criminelle. Elle consiste en la détection, la compilation et l'interprétation des informations afin d'identifier la présence de relations entre des entités d'intérêts (p. ex. des personnes, des événements, des véhicules, des sociétés, etc.) (Harper & Harris, 1975) (Schroeder et al., 2007) (Heuer & Pherson, 2010). Elles permettent de rechercher et de poser des propositions telles que, quelle est la source de cette trace, qui a téléphoné avec qui, etc. Lorsque beaucoup de relations sont en jeu dans une affaire, il devient évidemment utile de les visualiser. Alors que certains auteurs définissent l'analyse relationnelle comme le processus d'identification et d'extraction des entités et relations au sein des données collectées (Pottenger et al., 2007), le terme d'analyse relationnelle est généralement exploité en analyse criminelle pour décrire l'exploitation de représentations relationnelles à des fins d'analyse. Le terme de visualisation relationnelle semble plus adapté dans ce sens.

Le processus d'analyse relationnelle présume la capacité d'identifier les entités pertinentes. Des éléments de la réalité entre lesquels l'analyste cherche à identifier et comprendre les relations. Une entité se définit comme une chose possédant une existence distincte et

identifiable¹² (littéralement Chen définit l'entité comme une chose pouvant être distinctement identifiée) (Chen, 1976). Les entités sont regroupées par type dans des classes, telles que les personnes, les traces, les moyens de communications, etc. Les termes de type et classe seront utilisés comme synonyme. Un type décrit le prédicat « est un ». Par exemple « John Doe » est une entité de la classe « personnes ».

L'étape d'identification des entités constitue l'une des étapes les plus accaparantes du processus de l'analyse relationnelle. Un domaine de connaissances est notamment souvent décrit comme un ensemble de relations entre des entités qu'il a été possible de nommer. La notion d'entité nommée (*Named entity*) fait référence à l'étape de raisonnement utilisée pour définir la relation entre une observation (une valeur, un objet) et son type (sa catégorie, sa classe). Cette étape concentre un nombre important de recherches et d'efforts afin d'extraire des entités nommées au sein de données textuelles non structurées (Mennis et al., 2000) (Chen et al., 2003) (Kolda et al., 2004) (Schroeder et al., 2007) et (Pottenger et al., 2007).

En matière d'analyse criminelle, l'analyse relationnelle est très souvent décrite comme l'analyse de relations entre des entités nommées :

*Link analysis: this type of diagram displays the associations between different entities, which may be a **person, vehicle, organisation, telephone number etc.** It will also show the flow of commodities such as money or property. (Adderley & Musgrove, 2001) p.106*

*Link analysis puts information about the relationships among entities – **individuals, organizations, locations and so on** – into a graphic format and context that will clarify relationships and aid in inference development. (UNODC, 2002) p.29*

*There are four types of searches available to the user: **person, vehicle, incident and location**. ... The system can automatically identify relationships among **Person, Organization, Location, Vehicle, and Incident and Crime type**. (Chen et al., 2003) p.30*

*Link analysis ... by visualizing associations between **entities and events**. Typically, the entities of interest include: **places, organisations, facilities, individuals, components, documents, money, weapons, vehicles, drugs**. (Mena, 2003) p.80*

*In crime analysis, it is often useful to identify the relationships among different **entities such as people, vehicles, addresses, organizations**, etc. (Xiang et al., 2005) p.70*

*Link diagrams, showing the connections between **people, places, events, and things**, are invaluable tools in these domains. (Senator, 2005) p.80*

*Relational data about world-wide **terrorist events** is available, as well as ontologies describing the organization of this data. (Barthélemy et al., 2005) p.97*

*Through link analysis, investigators draw, lay out, and link **people, facts, locations, events, objects, and data** in hopes of discovering key trends, patterns, and insights. (Wong et al., 2006) p.67*

*In law enforcement, intelligence analysts often refer to nodes and links in a criminal network as entities and relationships. The entities may be **criminals, organisations, vehicles, weapons or bank accounts**. The relationships between the entities specify how these entities are associated together. (Oatley et al., 2006) p.66*

¹² <http://www.merriam-webster.com/dictionary/entity>, dernier accès le 4 avril 2001

38 Analyse criminelle et visualisation

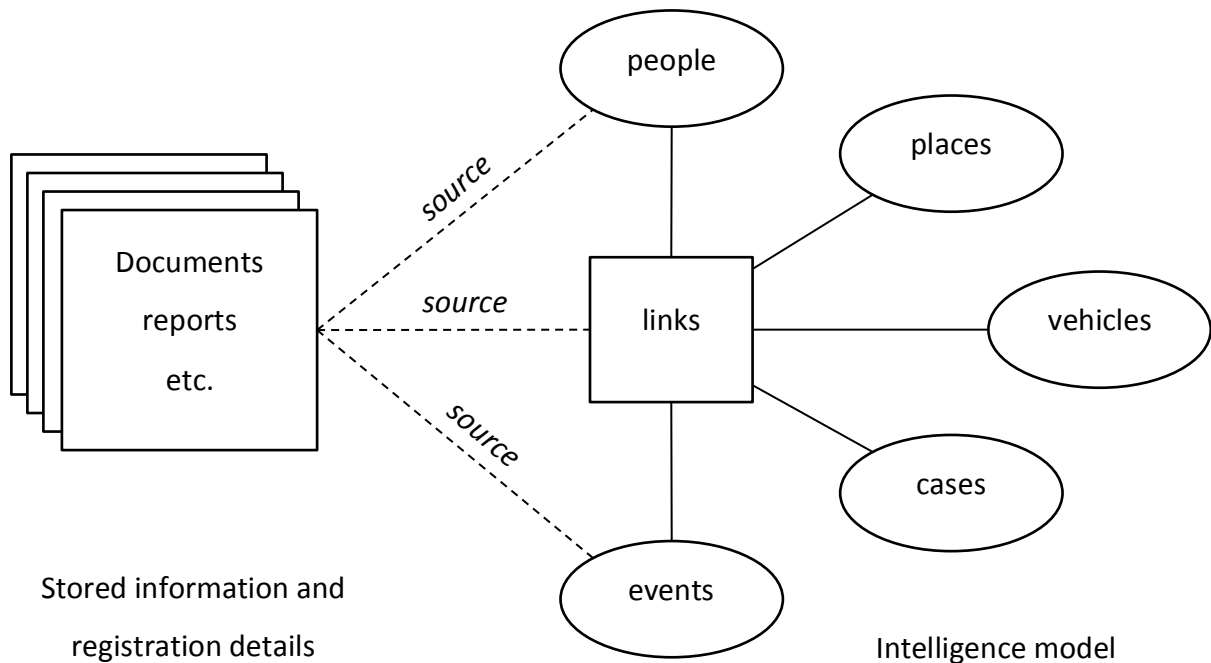
*We developed a model for organizing local data into a network of annotated relationships between **people, vehicles, and locations**. (Marshall et al., 2008) p.14*

We already applied Jigsaw to explore connections between people, places, and organizations in the Bible. (Stasko et al., 2008) p.127

*Using such tools it is possible to extract entities (**such as people, places, dates, and organizations**) from the evidence and relate them to each other through time lines or other relation schemes. These tools thus provide the analysts with a visual overview of the case at hand and the connections between various types of entities. (van den Braak, 2010) p.5*

*The essence of a criminal investigation, therefore, may be summarised as the study of actual and potential time dependent relationships between **people, organisations, objects, places and events** that may be associated with criminal activity in order to model and reason about possible patterns of activity. (Maller, 1996) p. 94*

La représentation de la démarche de l'analyse relationnelle présentée dans (Maller, 1996) p. 96 est par ailleurs très évocatrice:



De cette revue bibliographique, il ressort que des types d'entités sont généralement décrits. Ils correspondent aux types d'entités sur lesquels portent les investigations : des événements, des personnes, des sociétés, des objets, des véhicules, des lieux, des traces, etc. Il est très intéressant de relever que la description de types de relations est plus générale. A l'exception des relations de types spatio-temporelles, aucun autre type de relation n'est clairement défini. La diversité de la nature des relations pose alors la question de la définition de types génériques. Une démarche de classification en différents types de relation ne pourrait-elle pas aider à définir la manière de les représenter ?

Par ailleurs, les schémas relationnels exploités visent à répondre à des problèmes d'analyse.

Ces questionnements sont souvent exprimés au travers d'exemples spécifiques :

- Quelles sont les relations existantes ou absentes entre les individus ? (Harper & Harris, 1975) (Schroeder et al., 2007)
- Qui est la personne centrale d'une organisation ? (Harper & Harris, 1975) (Sparrow, 1991) (UNODC, 2002) (Mena, 2003)
- Quelle est la fréquence des relations ? (UNODC, 2002)
- Quelles sont les relations entre les identités des individus (les alias) ? (Sparrow, 1991) (UNODC, 2002) (Mena, 2003)
- Quelle est la structure hiérarchie d'un groupe criminel ? (Harper & Harris, 1975) (UNODC, 2002) (Senator, 2005) (Xu & Chen, 2005)
- Quelles sont les personnes les plus importantes d'un trafic (dont la suppression à l'impact le plus perturbant sur le réseau) ? (Sparrow, 1991) (UNODC, 2002) (Mena, 2003)
- Quel est le rôle de chaque individu dans l'organisation ? (Sparrow, 1991) (UNODC, 2002) (Xu & Chen, 2005)
- Quel individu fait office de « pont » entre les membres de l'organisation ? (UNODC, 2002)
- Quel est le chemin (ensemble de relations indirectes) entre deux individus ? (Schroeder et al., 2007)
- Existe-t-il des sous-réseaux dans un réseau complexe ? (Mena, 2003) (Xu & Chen, 2005)
- Quels sont les relations entre les sous-réseaux ? (Xu & Chen, 2005)
- Existe-t-il des relations ou des entités cachées ? (Mena, 2003)
- Quels sont les rôles des individus pour chaque événement ? (Senator, 2005) (Schroeder et al., 2007)
- Quelles sont les cooccurrences d'événements entre des individus ? (Schroeder et al., 2007)
- Quelles sont les régularités ou les anomalies observées dans un flux financier ? (Sparrow, 1991)
- Quelles sont les communications téléphoniques les plus pertinentes et utiles à surveiller ? (Sparrow, 1991) (UNODC, 2002)
- Quelles sont les relations de causalité entre les observations ? (Bex et al., 2003)
- Quels patterns relationnels sont observables et en quoi permettent-ils de comprendre un comportement particulier ? (UNODC, 2002)

Cette liste, évidemment non exhaustive, illustre la complexité et la diversité des questionnements possibles dans la dimension relationnelle. Des cadres conceptuels semblent avoir été plus clairement formalisés pour décrire les dimensions temporelles et spatiales¹³ que pour décrire la composante relationnelle.

Une terminologie riche existe pour décrire les représentations associées à l'analyse relationnelle. On parle parfois d'un schéma relationnel, d'un réseau, d'un arbre, d'un diagramme ou d'un graphique relationnel, ou encore d'une carte sémantique ou conceptuelle. Ces formes de représentation sont modélisées en mathématiques par la théorie des graphes qui remonterait aux travaux de Leonhard Euler en 1735¹⁴. Un graphe est une représentation abstraite d'un ensemble d'objets (des sommets), dont certaines paires sont reliées par des liens (des arêtes). La dimension mathématique n'est pas abordée dans cette thèse qui traite uniquement de la dimension visuelle des graphes. En mathématique, les arêtes d'un graphe ne peuvent relier que deux sommets. D'un point de vue purement visuel, cette limitation n'est pas forcément souhaitable. Les schémas étudiés dans cette recherche sont donc parfois des

¹³ Cf. Annexe 1: Dimensions d'analyse et visualisation, page 233

¹⁴ <http://www.math.dartmouth.edu/~euler/pages/E053.html>, dernier accès le 16 août 2010

hypergraphes (graphes où les arêtes peuvent lier plusieurs nœuds), au sens mathématique. Dans le cadre de cette thèse, nous assimilerons les schémas relationnels à des réseaux ou à des graphes.

La visualisation relationnelle est l'exploitation de graphes pour visualiser un ensemble hétérogène d'entités, de relations et d'attributs classés par types. Cette définition rejoint la définition des graphes sémantiques ou conceptuels (Kolda et al., 2004) (Barthélemy et al., 2005) (Wong et al., 2006) et des réseaux/graphes complexes (Newman, 2003).

Klerks décompose les formes de visualisation relationnelle exploitées en analyse criminelle selon trois générations de techniques (Klerks & Smeets, 2001) : *the manual approach*, *the Graph-based approach* et *the Social Network approach*.

L'approche manuelle initiée à la fin des années soixante aux Etats-Unis par le programme Anacapa (<http://www.anacapasciences.com/>, dernier accès le 13 septembre 2010) (Harper & Harris, 1975) et décrit par Morris dans son livre « The Crime Analysis Charting » (Morris, 1986) se base sur la création d'une matrice représentant les relations entre les personnes qui apparaissent dans l'affaire. Un schéma relationnel est ensuite produit sur la base de cette matrice. L'approche manuelle est encore utilisée, mais les graphes sont produits en utilisant des logiciels spécialisés qui permettent d'étendre les attributs graphiques exploitables (icônes, couleurs, épaisseur des traits, etc.). Initialement, la méthode reposait sur une représentation simplifiée de cercles (les personnes), de traits (les relations) et de cadres (les groupes : organisations criminelles et entreprises) (Harper & Harris, 1975) (Sparrow, 1991).

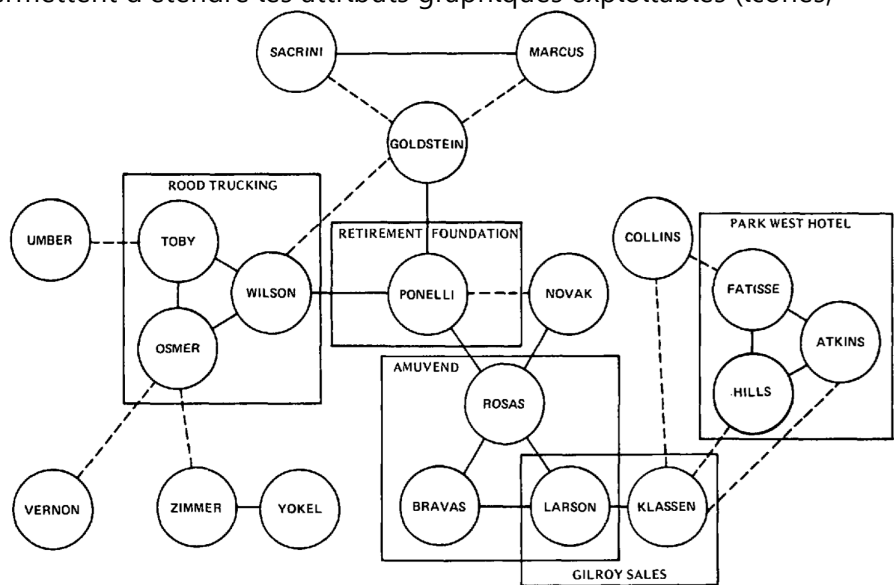


Figure 4. Criterion link diagram (solid lines represent strong links; broken lines represent weak links).

tiré de (Harper & Harris, 1975)

Harper et Harris (1975) référencent des exploitations antérieures de schémas relationnels dans le domaine de l'ingénierie où ils sont notamment utilisés pour représenter des processus de

fabrication. En fait, l'exploitation de graphes pour représenter les informations d'une enquête, se retrouve déjà dans les travaux de John Henry Wigmore (1863-1943) (Wigmore, 1913). La méthode de Wigmore appelée « argument diagram » et un résumé des travaux récents basés sur sa méthode sont présentés dans l'annexe deux.

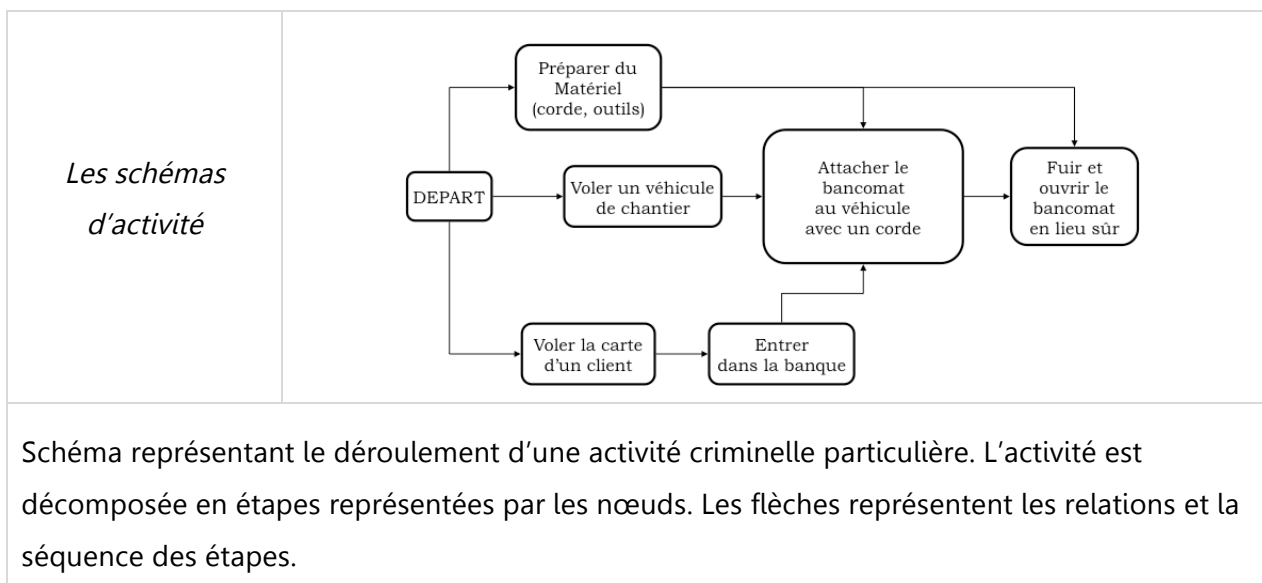
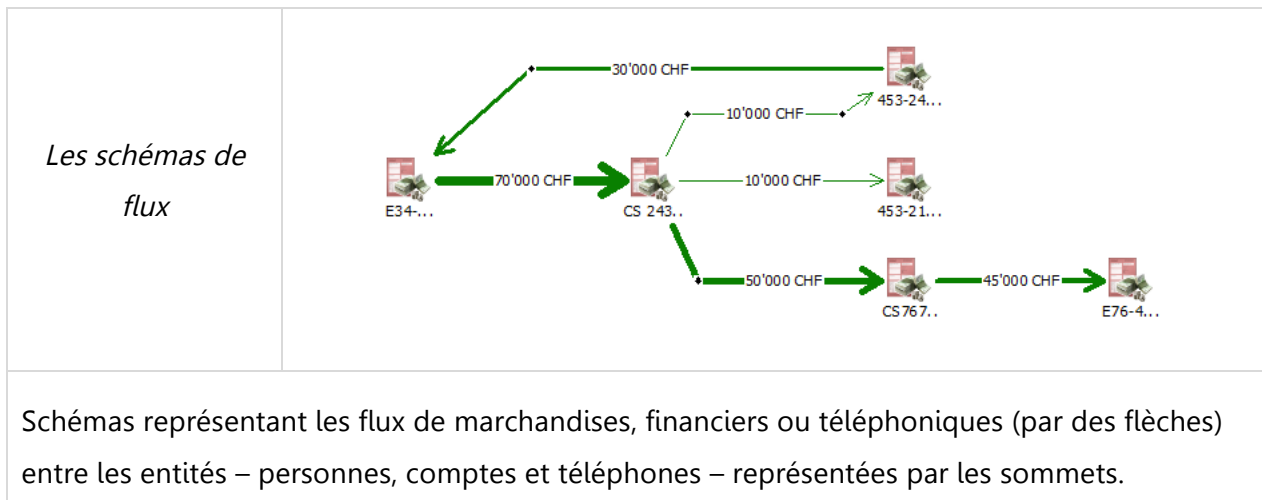
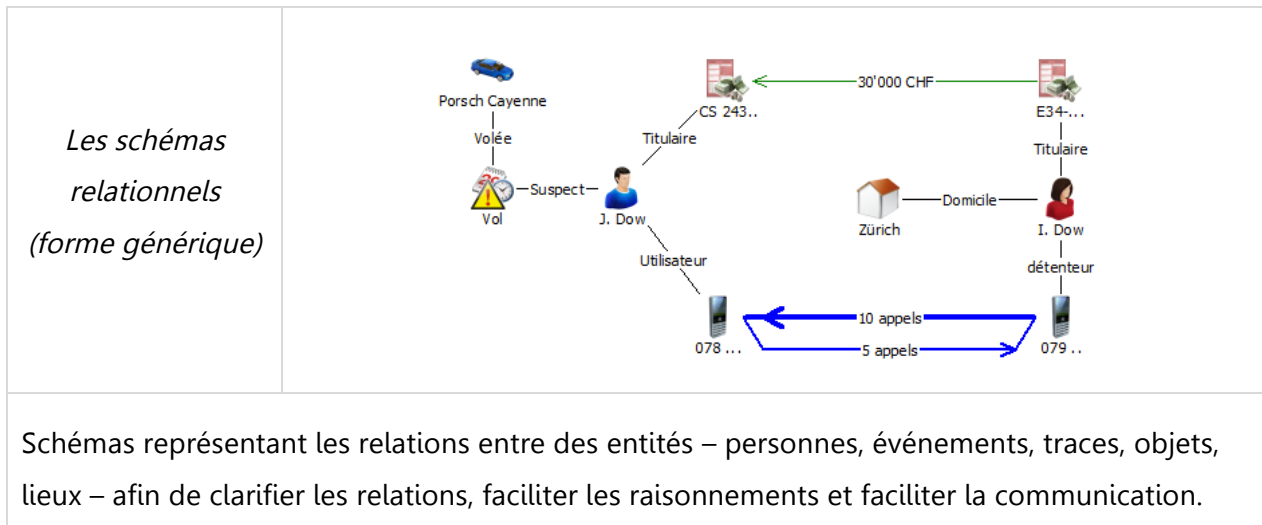
La seconde génération de techniques est basée sur l'utilisation de logiciels. Par le biais d'un filtre d'importation, les données structurées sous la forme de tableaux sont visualisées par des graphes. Certains auteurs relèvent le fait que ces approches se basent sur les capacités des utilisateurs à identifier les entités et les relations et qu'elles ne permettent pas de résoudre le problème de la surcharge d'informations (Schroeder et al., 2007). De plus, il est observé à l'usage que la conception de telle représentation nécessite beaucoup de temps (Innes et al., 2005). En effet, les visualisations relationnelles semblent bien fonctionner lorsque le nombre d'observations (événements, personnes, etc.) représentées est limité (Mena, 2003). Des approches parallèles ont été développées pour analyser et représenter des graphes de grandes tailles, comme l'amélioration des algorithmes de placement des nœuds, le développement de nouvelles méthodes d'interactions (en matière d'interface) et l'exploitation de représentations en trois dimensions. L'axe de cette recherche est différent et ces aspects du domaine ne seront pas abordés plus en détails. Les représentations en trois dimensions ont été exclues pour des raisons qui sont détaillées en annexe¹⁵. Selon Klerks (2001) ces outils n'offrent que peu de perspectives en termes d'analyse et il préconise l'exploitation des réseaux sociaux en tant qu'approche de troisième génération.

L'approche des réseaux sociaux développés dans le domaine de la sociologie se base sur la combinaison de la visualisation par les graphes afin de représenter les relations entre des personnes et le calcul de mesures mathématiques. Ces mesures, associées aux entités des graphes, décrivent notamment la centralité d'une entité au sein du réseau ou l'importance relative de chaque entité dans le réseau. L'étude de l'application des réseaux sociaux en analyse criminelle est un domaine émergent depuis une vingtaine d'années notamment en matière de criminalité organisée (Sparrow, 1991) (Klerks & Smeets, 2001) (Chen et al., 2005) (Xu & Chen, 2005). Ces mesures mathématiques sont exploitées afin de pallier les difficultés rencontrées pour analyser des réseaux complexes et particulièrement lorsque la taille des graphes ne permet plus une analyse visuelle suffisante. Cet aspect du domaine n'est pas abordé dans cette thèse. En effet, cette recherche porte sur les aspects visuels de la représentation.

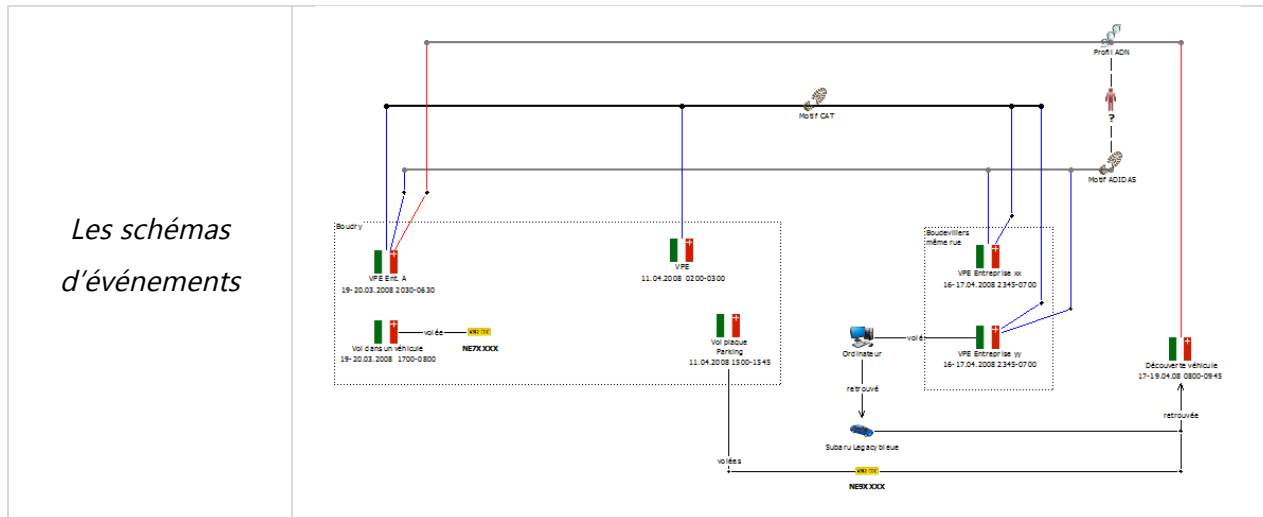
¹⁵ Cf. Annexe 3 : Visualisation tridimensionnelle, page 259

42 Analyse criminelle et visualisation

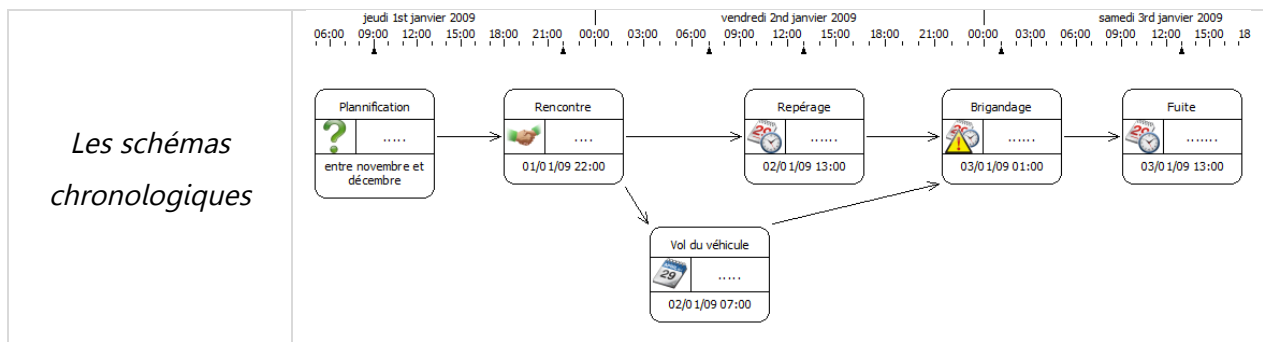
En matière d'analyse criminelle, plusieurs types de représentation relationnelle sont exploités :



Les formes de représentation ci-dessous sont des adaptations des schémas relationnels qui intègrent la composante temporelle. Ces formes de représentation sont probablement issues de visualisations développées dans les années cinquante : les diagrammes de PERT et CPM¹⁶.



Schémas représentant la séquence et les relations des événements qui sont positionnés chronologiquement. Une brève description des événements est incluse pour chaque nœud. Les liens représentent les relations entre les événements et d'autres entités (personnes, traces, véhicules, etc.).

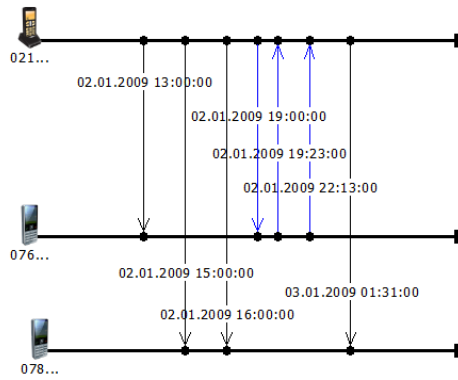


Ces schémas sont identiques aux schémas d'événements, mais les entités liées (personnes, traces, véhicules, etc.) ne sont pas représentées. L'objectif est de présenter une séquence d'événements précis sur le même modèle que les schémas d'activités, en intégrant la dimension temporelle. Les flèches et les positions des nœuds définissent la séquence.

¹⁶ Cf. Annexe 1: Dimensions d'analyse et visualisation – Analyse temporelle, page 246

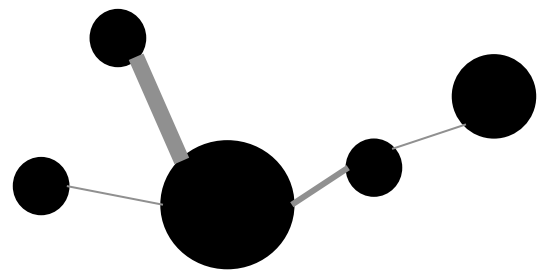
44 Analyse criminelle et visualisation

Les schémas de flux temporels



Schémas basés sur une définition des types de nœud et de lien identique aux schémas de flux. Le temps est représenté sur l'axe horizontal.

Sur un graph, la dimension quantitative est généralement représentée selon deux variables visuelles principales : l'épaisseur des liens et la taille des entités (ex. épaisseur des arcs proportionnelle au nombre de relations ou à une variable quantitative, taille des cercles proportionnelle au nombre d'autres sommets liés).



Des formes de représentation relationnelle intégrant la dimension spatiale ne sont pas définies dans les manuels du domaine. Pourtant, de telles représentations sont exploitées depuis le milieu du dix-neuvième siècle¹⁷. L'absence de ces représentations dans les manuels professionnels peut probablement s'expliquer par l'absence d'outils informatiques permettant leur élaboration de manière facile ou automatisée.

¹⁷ Cf. Annexe 1: Dimensions d'analyse et visualisation - Analyse spatiale, page 255

R e c o m m a n d a t i o n s e t s t a n d a r d s

Diverses recommandations ont été définies en matière de visualisation relationnelle, dans le domaine de l'analyse criminelle. Les premières définitions se retrouvent dans les travaux de Harper et Harris (1975) et de Morris (1986). Les recommandations ci-dessous sont issues de la littérature et des manuels de bonnes pratiques de l'analyse criminelle. Il est important de relever que les recommandations trouvées dans les sources ouvertes sont relativement rares et seul un nombre limité de manuels professionnels du domaine a pu être consulté.

E n t i t é s

- Les symboles doivent être simples et compréhensibles. Les symboles ne doivent pas prêter à confusion et ne doivent pas être interprétés de façon trop littérale. (Interpol, 1997) (Fedpol, 2010)

L i e n s

- Trois types de liens sont nécessaires : lien fort (sûr), faible (incertain) et aucun lien (Harper & Harris, 1975)
- Représenter les liens confirmés par une ligne pleine et les liens non confirmés par une ligne de pointillés. (Harper & Harris, 1975) (Morris, 1986) (Interpol, 1997) (UNODC, 2002) (Mena, 2003) (Fedpol, 2010)
- Les relations non qualifiées (sans label) portent à confusion. (Fedpol, 2010)
- Limiter l'usage des liens directionnels à la représentation de flux d'argent ou de marchandises et les appels téléphoniques. (Fedpol, 2010)

C a d r e s

- Les organisations sont représentées par des cadres. (Harper & Harris, 1975) (Morris, 1986) (Morris, 1986) (Interpol, 1997) (UNODC, 2002) (Mena, 2003) (Mowbray, 2009) (Fedpol, 2010)
- L'utilisation de cadres pour représenter des entreprises est possible, mais quand le nombre de personnes et d'entreprises est grand, il est recommandé d'utiliser des icônes. (Mena, 2003) (Fedpol, 2010)
- Les cadres peuvent être utilisés pour représenter des pays, dans ce cas il est recommandé de positionner les cadres en fonction de la position géographique des pays. (Fedpol, 2010)

C o u l e u r s

- Exploiter la couleur pour différencier des types de liens et d'entités. (Interpol, 1997) (Senator, 2005) (Fedpol, 2010)
- Toujours utiliser la même couleur pour certains processus dans un même graphique. (Fedpol, 2010)
- L'utilisation des couleurs et des icônes est libre. Une convention doit être utilisée au sein d'une même agence. (Mowbray, 2009)

S t r u c t u r e

- Lier les entités impliquées à l'événement concerné (c.-à-d. exploiter l'événement comme entité centrale pour représenter les relations). (Fedpol, 2010)
- Éviter les croisements de traits. (Sparrow, 1991) (Mena, 2003) (Mowbray, 2009)
- Maximiser l'orthogonalité. (Mowbray, 2009) (Fedpol, 2010)
- La proximité des personnes représentées sur un schéma doit être proportionnelle à leur proximité dans le réseau. (Sparrow, 1991)
- Une personne centrale sur un schéma implique qu'elle est centrale dans l'organisation. Il faut faire attention au biais de positionnement : une personne peut être centrale parce que c'est la personne sur laquelle le plus d'informations est disponible. (Sparrow, 1991)

A u t r e s

- Afficher la source de l'information pour retrouver le document initial. (Morris, 1986) (UNODC, 2002) (Fedpol, 2010)
- Représenter l'ensemble des informations. (UNODC, 2002) (Fedpol, 2010)
- Accompagner les produits d'analyse d'un rapport écrit. (UNODC, 2002)
- Un schéma doit être clair, simple, épuré et concis. (UNODC, 2002)
- Les représentations doivent être précises et correctes. Les informations à charge et à décharge doivent être incluses. (IALEIA, 2004)

Ces recommandations décrivent des éléments méthodologiques et des conseils basés sur la pratique. Le niveau de généralité des recommandations est variable. En outre, aucun des ouvrages ne présente ces recommandations dans un cadre clairement formalisé, mais plutôt sous la forme de listes.

Conclusion

Le processus systématique de l'analyse criminelle n'intègre pas formellement une étape de visualisation de l'information. La décomposition en cinq étapes présentée (acquisition, intégration, représentation, analyse et communication) permet de clarifier les rôles de la visualisation dans la méthode: (1) faciliter la mémorisation en regroupant de façon synthétique et simplifiée les informations pertinentes, (2) soutenir le processus d'exploration des informations et faciliter la découverte de nouvelles connaissances, (3) soutenir le processus de validation des hypothèses par la détection des éléments les confirmant ou non, (4) faciliter la communication et l'appréhension par les décideurs compétents du renseignement produit.

Les représentations graphiques forment certainement une partie centrale du produit de renseignement. Défini par l'objectif d'analyse, il guide le choix d'une ou plusieurs formes de représentation adaptées. Celui-ci semble par ailleurs être intimement lié à la nature du problème abordé. Les modèles classifient notamment l'analyse selon qu'elle traite des délits, des personnes ou des méthodes. Certains manuels de bonnes pratiques soutiennent que le choix d'une visualisation peut être effectué en regard des formes d'analyse. Celles-ci sont cependant définies de façon disparate et il est relevé que les méthodes se rejoignent. Une décomposition des questions d'analyse est proposée afin de faciliter la sélection d'une méthode de visualisation adaptée. L'approche retenue décompose l'objectif d'analyse selon une ou plusieurs dimensions dominantes pour déterminer le type de représentation à concevoir: (1) la dimension quantitative, (2) la dimension temporelle, (3) la dimension spatiale, (4) la dimension relationnelle qui couvre les questions liées à l'analyse des relations entre les entités pertinentes, telles que les événements, les personnes, les objets et les traces. Cette décomposition permet de classifier les types de représentation et par hypothèse facilite le choix d'une forme adaptée pour répondre à la question posée.

Les schémas relationnels exploités dans les enquêtes judiciaires ont été choisis comme sujet d'étude pour les raisons suivantes. Premièrement, les questions d'analyse sont souvent exprimées au travers d'exemples spécifiques, alors que des cadres conceptuels ont été plus clairement formalisés pour décrire les autres dimensions. Deuxièmement, la diversité de la nature des entités et des relations investiguées lors de l'enquête pose la question de la définition de types génériques. Une démarche de classification en différents types de relation ne pourrait-elle pas aider à définir la manière de les représenter ? Finalement, les recommandations définies dans la littérature et les manuels de bonnes pratiques ne sont pas intégrées dans une démarche de conception des schémas relationnels clairement formalisée.

The purpose of analytical displays of evidence is to assist thinking. Consequently, in constructing displays of evidence, the first question is, "What are the thinking tasks that these displays are supposed to serve?" Edward Tufte dans (Zachry & Thralls, 2004, p. 450)

Exploitations pour l'enquête judiciaire

L'objectif de ce chapitre est de décrire le contexte d'exploitation de la visualisation d'une part pour l'enquête judiciaire et d'autre part pour le jugement. Les objectifs d'analyse qui guident le choix d'une visualisation et les contraintes imposées par ce contexte sur la représentation de l'information sont explicités.

Le terme d'analyse criminelle opérationnelle (ACO) se définit, en Suisse, par son champ d'application dans le processus d'enquête et de jugement (Fedpol, 2010). Les étapes du processus judiciaire impliquent différents acteurs dont les besoins sont variables. La visualisation est donc exploitée de diverses manières selon les objectifs définis par les participants du processus judiciaire. Kind relève que les produits du renseignement doivent être élaborés d'après les choix et besoins des destinataires (Kind, 1990). Une représentation graphique peut notamment être exploitée afin de regrouper les informations d'une investigation, de faciliter la communication au sein d'une équipe d'enquêteurs ou comme support lors d'une plaidoirie au tribunal (Marshall et al., 2008).

Enquête

Processus de l'enquête et objectifs d'analyse

Kind décompose l'enquête criminelle en trois chapitres distincts: *"le problème de trouver* », *« la décision d'accuser »* et *« le problème de prouver »* (Kind, 1994). Le premier chapitre consiste à analyser les données de l'événement pour comprendre son déroulement, identifier, localiser et arrêter un ou plusieurs suspects. Suite à l'arrestation d'un suspect, l'enquête entre

dans un processus de structuration des indices, afin d'établir un dossier qui sera transmis au tribunal. Finalement, l'enquête se termine par la mise en accusation du prévenu et l'évaluation de sa culpabilité qui aboutit à son jugement. Brodeur décompose l'enquête initiée par la constatation d'une infraction selon un processus qui rejoint le paradigme de Kind (Brodeur, 2005b). Il définit l'enquête d'identification comme première partie du processus dont l'objectif est d'identifier « l'auteur du crime ». L'enquête de localisation subséquente vise l'arrestation du suspect. Finalement la structuration et la présentation de la preuve (ou plutôt des indices) visent la mise en accusation et la condamnation. Simms et Petersen (Simms & Petersen, 1991) décomposent le processus d'enquête en trois étapes : la détection d'infractions ou de suspects, la reconstruction des circonstances de l'activité et la résolution du problème en reliant le(s) suspect(s) à l'activité. L'enquête peut en effet être initiée sur la base d'informations sur un suspect (Simms & Petersen, 1991) (Atkin, 1998) (Brodeur, 2005b), par exemple, lorsqu'un informateur fournit des renseignements sur une possible activité délictueuse en matière de stupéfiants d'un individu ou lorsqu'une société est suspectée d'une activité illégale. Dans ces situations, l'objectif initial de l'enquête n'est pas d'identifier un suspect, mais bien de lier un suspect à une activité.

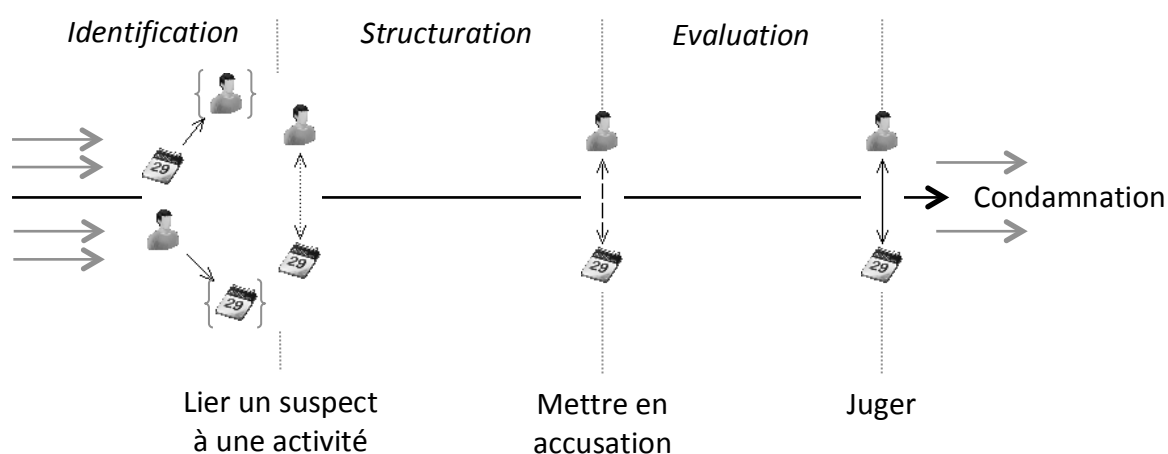


Figure 3 Processus de l'enquête judiciaire

Lors de la première phase, le raisonnement est principalement inductif (ou plus précisément abductif) (Harper & Harris, 1975) (Kind, 1987) (Kind, 1994). L'enquêteur recherche les causes des effets qu'il observe. Selon Atkin (1998), la première question qui devrait être formulée, lorsqu'une enquête débute sur un cas, est « *quelles sont les conditions qui peuvent expliquer la commission du crime ?* » ; pourquoi maintenant et ici, pourquoi cette cible et pourquoi cette manière de procéder en se plaçant dans la perspective des choix de l'auteur et des raisons sous-jacentes à ses choix. La démarche d'investigation suit le questionnement

Quintilien. Lorsque l'enquête démarre par un suspect, les questionnements sont par exemple : quelles relations le suspect entretient-il avec d'autres individus et quels sont les activités imputables. L'objectif premier de l'enquête est de relier un suspect à une activité (Atkin, 1998). Pour atteindre cet objectif, l'enquêteur raisonne également par analogie en évaluant un nouveau cas en regard de cas précédemment connus et traités (Ribaux & Margot, 2003). Un bon enquêteur construit son expérience en se basant sur ses expériences quotidiennes et sur le sens commun (Kind, 1987) (Kind, 1994). Il cherche à comprendre et reconstruire l'activité passée en variant le plus possible les points de vue sur les informations qu'il collecte (Kind, 1992). Il cherche à identifier et comprendre les relations existantes entre des événements, des personnes, des véhicules et des sociétés. Maintenir une vision holistique des informations collectées en regroupant les diverses sources d'information constitue l'un des principes essentiels à la base du processus de l'analyse criminelle (Aepli et al., 2011). Des représentations graphiques peuvent alors être conçues pour regrouper l'ensemble des informations collectées et mettre en évidence des relations afin de reconstruire et comprendre une activité criminelle.

Lors de la phase initiale de l'enquête, une ou plusieurs pistes peuvent être privilégiées au détriment d'hypothèses alternatives. En effet, les limitations des capacités cognitives humaines conduisent à mettre en œuvre des stratégies de simplification afin de prendre des décisions (Heuer, 1999). Des biais tels que l'effet tunnel, de confirmation (renforcement) et de la pensée de groupe peuvent conduire un groupe d'enquêteurs à se focaliser sur une hypothèse particulière et influencer la collecte et l'analyse des informations pour la confirmer. L'analyste criminel apporte une méthode de travail qui vise à minimiser ces biais. En effet, l'étape d'analyse survient volontairement tardivement dans le processus (après les étapes d'acquisition, d'intégration et de représentation) afin de minimiser les hypothèses hâtives et pour baser les raisonnements sur l'ensemble des informations (Aepli et al., 2011).

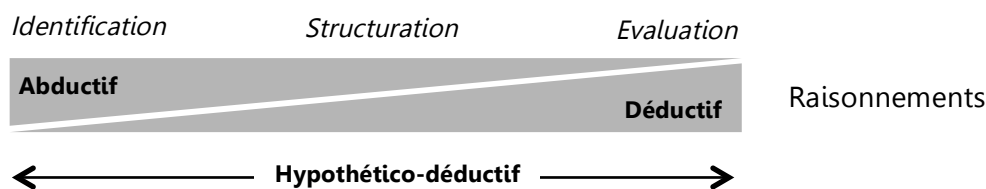


Figure 4 Formes de raisonnement au fil de l'enquête

Dans la phase de structuration, le raisonnement est essentiellement hypothético-déductif (Kind, 1994) (Ribaux et al., 2003). L'enquêteur confronte et évalue les informations collectées en regard des hypothèses formulées. L'objectif consiste à confirmer ou évacuer des pistes

d'enquête, voire d'en détecter d'autres (Carillo, 2010). Dans les premières phases d'enquêtes, le nombre de pistes et d'hypothèses formulées pour identifier les causes des effets observés sont plus nombreuses qu'en fin de processus. Ces réflexions conduisent à décider des mesures à prendre pour tester ces hypothèses, telles que la mise sous surveillance d'un individu, l'audition de témoins, l'analyse d'une trace, voir l'appréhension ou l'arrestation d'un suspect. Chacune de ces mesures a des conséquences importantes (en particulier les mesures de contraintes). Elles nécessitent des échanges permanents entre enquêteurs, forensiens, analystes criminels et magistrats. La visualisation peut alors être exploitée pour structurer les éléments pertinents pour la prise de décision et faciliter la communication entre les acteurs du processus. Un schéma relationnel peut par exemple mettre en perspective les traces matérielles collectées en fonction des lieux de prélèvement et des résultats d'analyse afin de choisir sur quelles traces des analyses complémentaires vont être entreprises. Un schéma représentant les communications téléphoniques effectuées par un suspect peut conduire à la détection d'un numéro sur lequel des contrôles pourraient être effectués (de l'identification de son détenteur à la mise sous écoute par exemple).

En fin d'enquête le raisonnement est principalement déductif. Les indices collectés sont évalués en regard des hypothèses retenues. L'acceptation des risques d'erreur diminue également au fil du processus, pour aboutir à une évaluation « au-delà de tout doute raisonnable » de la culpabilité lors du jugement (Brodeur, 2005b).

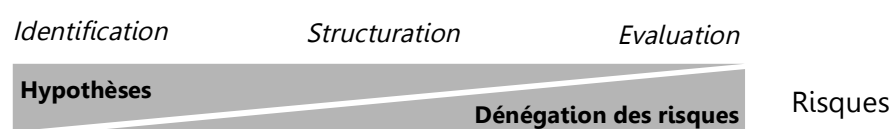


Figure 5 Acceptation des risques au fil de l'enquête

Les visualisations produites au cours de l'enquête doivent donc soutenir ces formes de raisonnement et aider à réduire les risques d'erreur. En raison de l'impact des décisions prises en cours d'enquête, l'analyse effectuée se doit d'être la plus objective possible. Les représentations graphiques exploitées doivent donc être de qualité. Elles doivent faciliter l'analyse et elles doivent être intègres par rapport aux données collectées. Idéalement, aucun biais ne doit être engendré par leur utilisation.

Intégration de l'analyse criminelle opérationnelle

Au cours de ce processus, l'enquêteur est confronté à toutes les problématiques de la gestion d'informations. L'analyse criminelle couvre donc un ensemble de méthodes et de techniques indispensables pour tout enquêteur. Elle est un moyen de l'investigation, au même titre que la surveillance, l'audition ou l'écoute téléphonique (Peterson, 1998). Deux approches parallèles d'intégration de l'analyse criminelle dans l'enquête judiciaire sont mises en œuvre :

l'intégration d'analystes criminels spécialisés au sein des équipes d'enquête et une approche de formation des enquêteurs.



Figure 6 Intégration de l'analyse criminelle dans l'enquête

L'analyste criminel « spécialiste »

La première approche consiste à considérer l'analyse criminelle comme une discipline à part entière. Des « spécialistes » viennent ainsi soutenir les enquêteurs. L'analyste peut alors être intégré directement dans l'équipe d'enquête ou travailler de manière autonome sur le cas.

Lorsque que celui-ci travaille de façon autonome, il est chargé d'élaborer des scénarios alternatifs non identifiés. Il évalue les indices afin de détecter de nouvelles pistes d'enquête ou d'identifier des informations manquantes. Il aboutit à la présentation des indices afin d'apporter de nouvelles connaissances utiles à prise de décisions, telle que prendre des nouvelles mesures d'enquête. Le processus de travail de l'analyste se décompose alors en trois étapes : la reconstruction, l'évaluation et la communication de scénarios pouvant expliquer l'activité criminelle observée. Des schémas de scénarios¹⁸ peuvent notamment être exploités pour confronter diverses versions et confronter les indices collectés (van den Braak, 2010). L'analyste criminel prend, dans une certaine mesure, un rôle d'évaluateur. Il traite

¹⁸ Cf. Annexe 2 : Schémas de scénarios, page 256

généralement l'affaire *a posteriori* et joue le rôle de « l'avocat du diable ». Il est censé apporter un regard neuf en suivant une démarche adaptée, afin d'identifier des lacunes dans l'enquête et apporter de nouvelles hypothèses. Une telle démarche d'intégration de l'analyste criminel dans l'enquête judiciaire induit toutefois un nombre important de frustrations ; tant pour l'enquêteur qui voit son travail évalué, que pour l'analyste qui a le sentiment « d'arriver après la bataille ». En effet, un tel processus d'évaluation est généralement effectué sur des affaires complexes, ayant impliqué un nombre important de mesures d'enquête. Le temps nécessaire à la lecture des informations collectées, leurs intégrations dans un système de gestion, voire la création de visualisations facilitant l'analyse, est généralement long. Les renseignements produits par l'analyste peuvent alors perdre en validité dans la mesure où de nouveaux actes d'enquêtes ne sont plus forcément possibles.

Une autre approche consiste à intégrer l'analyste criminel directement au cours de l'enquête. L'analyste endosse alors un rôle de soutien à l'investigation. L'analyste offre ses compétences en matière de traitement, de gestion et de visualisation des informations au service des enquêteurs. Intégré, dès les phases initiales de l'enquête, l'analyste peut entreprendre l'élaboration de divers produits (tableaux, banque de données, représentations graphiques, etc.) pour soutenir les tâches des enquêteurs et l'investigation dans son ensemble (Peterson, 1998). Les analystes engagés dans les services de police sont souvent des employés civils ne disposant pas d'une formation policière (Peterson, 1998). L'analyste civil bien que disposant des compétences en matière de gestion de l'information, ne possède parfois pas l'expérience ou la formation suffisante pour assimiler tous les tenants et aboutissants de l'investigation criminelle. L'analyste doit alors accepter l'image de naïveté qu'il génère. Une certaine incompréhension, voire des problèmes de communication peuvent donc survenir entre l'analyste et l'enquêteur. Ces difficultés peuvent également être renforcées par un sentiment de ne pas appartenir à la même communauté (civile ou policière). Les visualisations produites par l'analyste peuvent également générer un sentiment de frustration chez l'enquêteur. En effet, une représentation synthétique et efficace peut être appréciée et mise en valeur, au détriment de l'ensemble des tâches accomplies lors de l'investigation. La visualisation, de par sa nature à rendre explicite et compréhensible les informations, cache souvent l'ensemble d'un processus complexe. Globalement, il faut constater que le processus d'intégration de l'analyste criminel dans l'investigation n'est pas trivial (Peterson, 1998). Selon Atkin (1998), la collaboration entre analyste et enquêteur sur une même affaire reste un compromis idéal. L'expérience de l'enquêteur basée sur les cas précédemment traités contrebalance le regard externe objectif (dans le sens où il est absent de connaissances *a priori*) de l'analyste.

L' enquêteur analyste

En raison de la diversité des problèmes à traiter et des connaissances spécifiques requises pour les traiter, une approche d'intégration de l'analyse criminelle dans l'enquête judiciaire consiste à former des enquêteurs aux méthodes et techniques de l'analyse criminelle.

La formation d'enquêteur soulève toutefois un certain nombre de problèmes. Les enquêteurs formés au domaine de l'analyse criminelle acquièrent notamment un savoir-faire en matière de visualisation. Par contre, les problématiques de traitements et de gestion de l'information sont généralement survolées lors de ces formations, en raison de la difficulté de prise en main des outils. Ceux-ci sont alors considérés comme du ressort du spécialiste en analyse criminelle. *A contrario*, la prise en main des outils de visualisation semble être plus facile et leur maîtrise (du moins au niveau des fonctionnalités de base) est acquise par les enquêteurs qui peuvent ainsi les exploiter dans leurs pratiques.

Pourtant, il n'existe à leur actuelle que peu de documentations et de recommandations liées à l'exploitation de ces outils dans l'enquête judiciaire (voir chapitre « Recommandations et standards »). La maîtrise technique de l'outil peut être acquise en suivant un manuel (ou un cours), mais l'exploitation efficiente des formes de visualisation pour résoudre les problèmes rencontrés dans l'enquête reste du ressort de l'expérience et des capacités d'imagination de l'analyste enquêteur. De plus, alors que les principes méthodologiques fondamentaux de l'analyse sont indépendants du domaine traité, l'exploitation des formes de visualisation semble largement influencée par le problème spécifique. Les questions qui se posent au cours des enquêtes varient suivant la nature des infractions et les données exploitées ne sont pas les mêmes. Il semble donc *a priori* difficile de dégager des principes d'exploitation de la visualisation universellement exploitables dans l'enquête judiciaire. Toutefois, certaines tâches sont récurrentes comme la comparaison des déclarations de témoins ou de prévenus ou l'analyse de communications téléphoniques. Certains problèmes criminels spécifiques surviennent également de façon régulière, tels que la commission de plusieurs cambriolages par un auteur ou un groupe d'auteur ou le développement d'un réseau de ventes de stupéfiants. Certaines situations d'analyse étant récurrentes, des formes de représentation adaptées pour chacune d'elles peuvent par hypothèse certainement être définies.

J u g e m e n t

Pendant très longtemps, l'analyse criminelle est restée très prudente dans ses relations directes avec les magistrats. Ceux-ci sont néanmoins de plus en plus demandeur. Ce mouvement récent étend le champ d'application de l'analyse criminelle et pose des questions nouvelles. Les produits de renseignement élaborés en cours d'enquête peuvent en effet se voir intégrer comme pièces du dossier, au risque de voir de nouvelles difficultés apparaître. A titre d'exemple, lors de la phase initiale de l'enquête, des hypothèses formulées sur l'implication de suspects peuvent avoir été intégrée sur des schémas de travail. Alors même que ces pistes d'enquête peuvent être écartées en fin d'enquête, la présence de ces hypothèses dans le dossier peut mettre en difficulté l'analyste lors du jugement. Une réticence persiste donc sur l'intégration des produits de l'analyse dans le dossier de justice. Celle-ci semble néanmoins se justifier pour plusieurs raisons décrites dans cette section. Les contraintes imposées par le contexte judiciaire sur l'utilisation de représentations graphiques sont également décrites.

O b j e c t i f s d ' a n a l y s e

L'objectif de l'enquête judiciaire est d'aboutir au jugement. En conséquence, le travail de l'analyste soutient également le processus de la poursuite judiciaire (Peterson, 1998). Dans une large mesure, les objectifs et les problématiques de l'enquête rejoignent ceux de la poursuite pénale. Toutefois en regard de l'apport étendu de l'analyse criminelle, celle-ci se caractérise par certains objectifs spécifiques.

Le magistrat souhaitera qualifier les infractions en regard des lois. Il s'agira donc de connaître les éléments pertinents et nécessaires pour cette tâche, afin de les inclure dans les produits de l'analyse. Par exemple, le prix de vente est un élément capital afin d'estimer si un acheteur pouvait raisonnablement se douter qu'il faisait l'acquisition d'une contrefaçon ou du butin d'un vol. Dans les phases initiales de l'enquête, une telle question ne sera certainement pas jugée primordiale pour reconstruire la structure d'un trafic (en identifiant ses protagonistes et en évaluant leurs rôles respectifs, par exemple). Un analyste n'est ni un juriste, ni un enquêteur. Il n'a donc pas nécessairement les connaissances suffisantes (ou la sensibilité) pour identifier l'ensemble des paramètres d'importance pour la résolution d'une affaire. Dans la situation précitée, l'importance du prix est certainement évidente. Toutefois, les paramètres d'importance pour une qualification pénale ne sont pas forcément évidents à maîtriser. Ces dimensions doivent donc être discutées et explicitées dès l'intégration d'un analyste dans l'investigation criminelle. Il est du ressort de l'analyste de s'assurer qu'il a bien compris les

objectifs et les dimensions d'importance pour la poursuite pénale. Ces dimensions étant, du point de vue du magistrat instructeur et de l'enquêteur, certainement des évidences.

Du point de vue du procureur et des avocats, l'analyse criminelle est souvent perçue comme une méthode de l'investigation (Peterson, 1998). Toutefois, les produits de l'analyse et plus particulièrement les représentations graphiques peuvent également être exploités lors du jugement. En effet, la visualisation peut jouer un rôle important en facilitant la compréhension d'une affaire dite complexe. Morris relève (1993) qu'une affaire jugée complexe lors de l'investigation sera certainement presque impossible à appréhender lors du jugement (notamment par le jury). Il souligne alors l'importance d'exploiter des représentations graphiques pour soutenir l'enquête, mais également faciliter la compréhension du dossier.

C o n t r a i n t e s i m p o s é e s s u r l e s v i s u a l i s a t i o n s

Au tribunal, les représentations graphiques servent également de support rhétorique. La visualisation peut être exploitée pour faciliter la communication d'un argument et augmenter l'impact émotionnel d'un propos (Feigenson & Sherwin, 2007). Les effets rhétoriques engendrés par la visualisation devraient être évités ou au moins contrôlés, afin que la réalité des informations présentées ne soit pas détournée pour soutenir une interprétation fallacieuse des faits.

En droit américain (Joseph, 2006), les schémas (croquis), diagrammes, et graphes qui sont impartiaux et pertinents sont admissibles. Ils servent à faciliter la compréhension des questions traitées par la cour. La recevabilité est appréciée par la cour. Ils sont rejetés en cas d'abus avéré. Les représentations visuelles sont classifiées en deux catégories : les représentations à but illustratif (ne sont pas recevables en tant que preuve, mais utilisées à des fins pédagogiques et de compréhension, leur usage est à la discrétion de la cour) et les représentations recevables en tant qu'élément de preuve de fond (substantive evidence). Visant à établir un fait, elles doivent satisfaire les critères de recevabilité.

Les représentations illustratives peuvent être utilisées par tout témoin afin d'illustrer, de clarifier et de visualiser son témoignage. A des fins illustratives, une représentation ne doit pas forcément être exacte dans ces moindres détails. Des croquis sont notamment exploitables pour autant qu'ils facilitent la compréhension du propos par le jury. Le juge peut décider de refuser une représentation si elle engendre de la confusion, qu'elle induit en erreur ou que les éléments représentés sont trop simples (la rendant ainsi superflue et inutile). Même si la

représentation n'est pas recevable en tant que preuve (par exemple lorsque les sources d'informations représentées ne sont pas suffisamment documentées), elle peut être présentée à des fins illustratives dans la mesure où une explication initiale est effectuée (sur la nature des données représentées et l'usage illustratif, de compréhension).

L'usage de représentations graphiques en tant que preuve est défini dans la Federal Rule of Evidence 1006 : « *The contents of voluminous writings, recordings, or photographs which cannot conveniently be examined in court may be presented in the form of a chart, summary, or calculation, the originals, or duplicates, shall be made available for examination or copying, or both, by the other parties at reasonable time and place. The court may order that they be produced in court.* »

Pour être recevable en tant que preuve :

- une représentation doit être impartiale et précise. Des imprécisions mineures ne sont toutefois pas suffisantes pour rejeter une représentation graphique comme preuve, mais elles doivent être documentées et expliquées
- les informations initiales (sur lesquelles la visualisation est produite) doivent être accessibles
- les informations initiales doivent être recevables. Il s'agit d'une condition *sine qua non*
- les informations initiales doivent être volumineuses (c.-à-d. qu'elles ne peuvent être convenablement examinées en tant que telles).

Les représentations dont l'exactitude et la précision sont jugées insuffisantes ou qui prêtent à confusion peuvent être rejetées. L'absence de données volumineuses sous-jacentes à la représentation est rarement un critère accepté pour rejeter une représentation (Joseph, 2006). De plus, il est possible d'exploiter des représentations résumant les faits. Dans des affaires complexes, la tendance est clairement d'accepter ce genre de représentations (Joseph, 2006). Des représentations résumant uniquement les éléments à charge (ou à décharge) peuvent même être utilisées pour autant qu'elles ne soient pas présentées comme résumant l'ensemble des éléments de preuve. Les conditions d'admissibilité de ces représentations simplificatrices (de synthèse) sont définies en regard de la Federal Rule of Evidence 611(a)¹⁹ : la durée du procès, la complexité de l'affaire, et la confusion engendrées par un nombre

¹⁹ The court shall exercise reasonable control over the mode and order of interrogating witnesses and presenting evidence so as to:

1. make the interrogation and presentation effective for the ascertainment of the truth,
2. avoid needless consumption of time, and
3. protect witnesses from harassment or undue embarrassment.

important d'informations (Joseph, 2006). Le jury doit également être informé sur la manière d'appréhender le schéma.

Des dangers d'exploitation de représentations simplificatrices sont également identifiés (Joseph, 2006):

- la représentation peut être considérée comme une preuve additionnelle en tant que telle, corroborative de la vérité des informations représentées. Ce problème peut être minimisé en documentant clairement la représentation et en expliquant la manière de l'appréhender
- elle peut être exploitée comme moyen subtil d'introduire une preuve non recevable. L'admissibilité des sources d'informations doit donc être évaluée
- la représentation peut inclure des argumentaires, des jugements et des conclusions orientés. La distinction d'une visualisation représentant des preuves factuelles ou des argumentations n'est pas évidente. La règle générale est toutefois qu'une représentation devrait rester factuelle et exempte d'hypothèses ou de jugements.

C o n c l u s i o n

En regard du processus judiciaire, la visualisation peut être exploitée pour soutenir un nombre importants d'objectifs, tels qu'analyser des informations collectées au cours de l'enquête, évaluer *a posteriori* une investigation, aider à qualifier les infractions, faciliter l'appréhension d'un dossier et soutenir une argumentation lors du procès. L'exploitation des graphes pour évaluer les hypothèses alternatives formulées en cours d'enquête et lors du procès fait l'objet de recherches spécifiques qui ont abouti à la formalisation de langages graphiques spécifiques facilitant la comparaison de scénarios. En amont, ils sont également utilisés pour soutenir l'analyse des données collectées et le développement d'hypothèses.

Bien que les observations effectuées soutiennent probablement dans une certaine mesure l'ensemble de ces objectifs, les problèmes analysés et les généralisations proposées se basent en quasi-totalité sur des exemples réels d'utilisation de la visualisation au cours de l'enquête judiciaire. Les problèmes abordés et les solutions proposées traitent principalement d'une exploitation au cours de l'enquête pour faciliter l'analyse des informations collectées. Dans une large mesure, les problèmes investigués portent plus particulièrement sur l'analyse des traces. L'usage des schémas relationnels est notamment étudié dans une perspective de renseignement forensique.

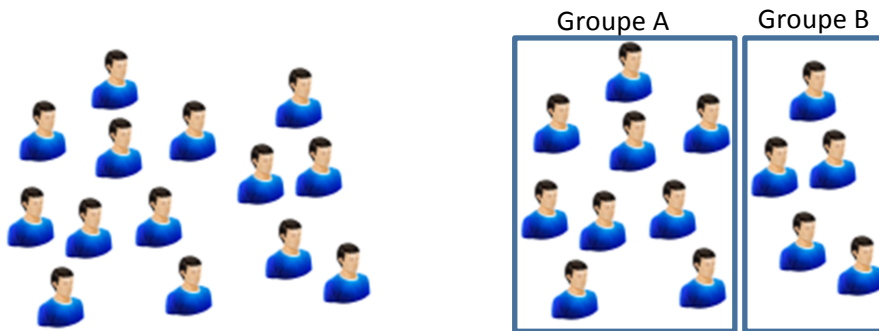
Méthodes d'évaluation de visualisations

Notre premier objectif est d'évaluer quelques pratiques en matière de visualisation. Nous soupçonnons que ces pratiques sont disparates et que les schémas produits sont plus ou moins « bons ». Que veut dire plus ou moins « bons » dans ce contexte ? Les critères qui permettent de définir et d'évaluer la qualité d'une représentation doivent être décrits. Par ailleurs, les méthodes d'évaluation empiriques développées pour évaluer des visualisations sont également présentées.

Lorsqu'une représentation est exploitée pour raisonner sur des informations, elle se doit d'être intègre : elle doit représenter la réalité des informations sous-jacentes sans ajouter des biais visuels qui peuvent fausser les observations. Par exemple, lorsque des informations quantitatives sont représentées par des différences de tailles (de barres, de symboles ou d'icônes), celles-ci doivent être proportionnelles aux variations observées au sein des données. Une visualisation est intègre lorsqu'elle est claire, détaillée, documentée et qu'elle n'engendre pas d'ambiguïté (Tufté, 2001). Bertin souligne l'importance du caractère monosémique de la graphique en tant que condition *sine qua non* de l'intégrité :

Un système est monosémique quand la connaissance de la signification de chaque signe précède l'observation de l'assemblage des signes. (Bertin, 2005)

La signification de chaque signe doit être définie à l'avance. Chaque signifiant de la graphique ne peut représenter qu'un signifié au sein d'une représentation particulière. Par exemple, un cadre peut être défini en tant que représentation visuelle d'un groupe :



La graphique se distingue des autres systèmes polysémiques de la représentation visuelle qui offrent une richesse interprétative dite artistique (telle la peinture ou l'image figurative). Le caractère monosémique de la graphique en fait un système dit rationnel (Bertin, 2005). C'est-à-dire sur lequel il est possible de raisonner. La signification du signe étant définie, elle n'est pas sujette à de multiples interprétations. Le processus de perception d'un graphique se traduit alors par la question : «*Étant donné la signification de ces éléments, quelles sont les relations qui s'établissent entre eux ?*» (Bertin, 2005). Les représentations graphiques se basent donc sur des conventions précises définissant la manière de les construire et les interpréter. Elles constituent la syntaxe et la sémantique du langage (Mackinlay, 1986).

Un système monosémique peut engendrer une connotation. Un ensemble de signes peut être perçu comme un nouveau signe. Celui-ci peut parfois être interprété de plusieurs manières. Le système est donc strictement rationnel, lorsque les structures engendrées par la combinaison des signes sont également monosémiques. Par exemple, la représentation d'un groupe peut faire apparaître des relations de hiérarchie entre eux :



Mackinlay définit le critère d'expressivité: « *un ensemble de faits est bien exprimé dans un langage visuel si les phrases (c.-à-d. les éléments de représentation) du langage expriment la totalité des faits et uniquement les faits contenus dans les données* » (traduction libre) (Mackinlay, 1986). Cette notion d'expressivité rejoint les notions d'intégrité définie par Tufte (2001) et de rationalité définie par Bertin (2005). En analyse criminelle, la monosémie d'une représentation est capitale. En effet, une mauvaise interprétation peut avoir des conséquences importantes, notamment lors du jugement. Une représentation qui engendre un biais de perception, qui induit des ambiguïtés ou qui peut être interprétée de plus manière ne satisfait pas le critère d'expressivité. Celle-ci est alors considérée comme moins « bonne » qu'une représentation respectant ce principe. Plusieurs représentations intègres peuvent toutefois être produites pour représenter une même situation. Quels sont alors les critères permettant de comparer et évaluer leurs qualités ?

Critères d'évaluation

Au-delà de la monosémie si une représentation est intègre, son efficacité n'est pas pour autant démontrée. Pour certains, l'efficacité est atteinte lorsque la visualisation produit l'effet attendu ou lorsqu'elle permet de persuader (Besson, 2004). Cette vision est probablement pertinente dans les domaines artistiques et certainement pour une exploitation rhétorique de la visualisation (journalisme, publicité, etc.). Elle n'est toutefois pas adaptée pour décrire ce que l'on attend d'une représentation graphique à des fins d'analyse. En effet, privilégier l'efficacité en termes d'effet de persuasion peut être au détriment de la condition *sine qua non* d'intégrité.

Une visualisation est plus efficace qu'une autre, si les informations transmises sont plus facilement perçues. L'efficacité est donc notamment dépendante des capacités du lecteur (Mackinlay, 1986). L'efficacité se définit alors ainsi :

Si, pour obtenir une réponse correcte et complète à une question donnée et toutes choses égales, une construction requiert un temps de perception plus court qu'une autre construction, on dira qu'elle est plus efficace pour cette question. (Bertin, 2005)

Bertin relève également que la graphique est un outil « *rationnel et efficace lorsque les propriétés de la perception visuelle sont pleinement employées* » (Bertin, 2005). Cette observation

62 Méthodes d'évaluation de visualisations

de Bertin a conduit à de nombreuses évaluations de l'efficacité de chacune des variables graphiques (tel que les dimensions du plan, la couleur, la forme, la taille, etc.).

Cette définition permet de définir clairement les indicateurs à comparer pour évaluer l'efficacité d'une représentation : l'exactitude et la complétude de la réponse, ainsi que le temps nécessaire pour y répondre. Le nombre total de réponses correctes pour une tâche particulière est un indice de l'efficacité. Combiné au temps de réponse, il permet d'évaluer l'efficience (Chen & Yu, 2000) (Xu & Chen, 2005).

Ces mesures de performance sont communément exploitées. Elles permettent de quantifier l'impact d'un choix de visualisation (p. ex. par le nombre d'erreurs), mais les raisons des différences observées ne sont pas pour autant toujours explicitées. D'autres indicateurs peuvent être exploités de façon complémentaire (Huang et al., 2008) :

- Les questionnaires : des indicateurs démographiques complémentaires sont relevés (âge, sexe, niveau de formation, etc.). Des indicateurs subjectifs peuvent également être mesurés : tel le niveau de satisfaction ou la préférence pour réaliser la tâche.
- Le suivi du regard (eye tracking) : à l'aide de caméras, le processus cognitif est détaillé en enregistrant les mouvements oculaires. L'objectif est de comprendre comment le lecteur parcourt le schéma.
- L'effort mental : le lecteur doit évaluer la représentation en utilisant une échelle verbale. Une échelle à neuf niveaux (1. effort mental très très faible à 9. effort mental extrêmement élevé) est proposée par Paas (Paas et al., 2003) (Tuovinen & Paas, 2004).

L'utilisation complémentaire de ces indicateurs peut faciliter l'interprétation des résultats. En effet, il est fréquent que l'exactitude des résultats soit proportionnelle au temps de réponse, alors qu'une représentation sera jugée plus efficace si l'exactitude est haute et le temps de réponse faible (North, 2006) (Huang et al., 2008). L'exactitude et le temps de réponse restent néanmoins les indicateurs d'efficacité les plus utilisés. Les niveaux de satisfaction ou d'effort, mesurés à l'aide d'échelles nominales, peuvent être exploités de façon complémentaire, mais ne permettent pas de généraliser les résultats. Leurs validités internes ne sont en effet pas démontrées et ils peuvent être corrélés au temps de réponse notamment (Tuovinen & Paas, 2004) (Chung et al., 2005).

Procédures d'évaluation

En se basant sur une étude menée sur une cinquantaine d'évaluations empiriques, Plaisant classifie quatre procédures distinctes (Plaisant, 2004) :

- *les expérimentations contrôlées comparant des transpositions graphiques* : par exemple, la couleur est-elle plus efficace que la forme pour encoder une variable quantitative ?
- *les expérimentations contrôlées comparant des formes de représentation* : l'objectif de ces évaluations consiste à comparer plusieurs types de représentations différents (un histogramme, un diagramme circulaire, etc.) et d'évaluer leur efficacité pour accomplir des tâches particulières (comparer des valeurs quantitatives, retrouver une information, comparer des proportions, etc.)
- *les évaluations de l'utilisabilité (ou de l'ergonomie)* : les solutions développées sont testées par des utilisateurs qui fournissent un feedback sur les problèmes rencontrés et les améliorations à apporter
- *les études de cas* : l'objectif consiste à exploiter la solution dans une situation réelle, dans un environnement opérationnel. Les possibilités d'exploitation pour résoudre des problèmes réels sont alors illustrées (afin de montrer la faisabilité et l'utilité). Induire des principes généraux sur la base d'études de cas nécessite toutefois un nombre important d'exploitations dans des contextes variés.

Une autre approche consiste à poser des questions ouvertes, nécessitant d'analyser le problème posé en fonction de connaissances spécifiques liées au domaine (évaluation basée sur la compréhension / insight-based evaluation). Le lecteur interprète ainsi la visualisation selon ses propres choix et les exprime librement. Une telle démarche engendre des difficultés pour catégoriser les résultats, mais offre une perspective d'évaluation de la compréhension de la représentation et de son impact sur les raisonnements (North, 2006) (Carpendale, 2008).

Chen et Yu définissent les conditions qu'une expérimentation empirique en matière de visualisation doit satisfaire (Chen & Yu, 2000) :

- les conditions d'expérimentation doivent être standardisées et clairement décrites
- les propriétés visuelles évaluées doivent être clairement décrites
- les tâches cognitives testées doivent être standardisées et clairement définies
- une variable dépendante de l'exactitude ou de l'efficacité doit être utilisée
- le compte rendu des résultats doit être détaillé et inclure un test statistique.

64 Méthodes d'évaluation de visualisations

Des taxonomies de tâches à évaluer sont proposées par (Ghoniem et al., 2005), (Shneiderman & Aris, 2006) et (Lee et al., 2006) pour évaluer des graphes. La lisibilité de la représentation est évaluée en fonction des performances de recherche de l'information:

- *tâches liées aux caractéristiques des sommets* : estimer le nombre total de sommets, identifier le sommet lié au plus grand (ou petit) nombre de sommets, retrouver un sommet particulier à partir de son label
- *tâches liées aux caractéristiques des liens* : estimer le nombre total de liens, retrouver un voisin commun entre plusieurs sommets, trouver l'existence d'un chemin entre deux sommets, identifier le chemin le plus court, retrouver les sommets liés par un type spécifique de lien, retrouver le sommet central
- *tâches liées aux caractéristiques des sous-graphes* : retrouver un sous-graphe particulier, retrouver les sommets liés à un sommet ou un groupe de sommets, retrouver des groupes de sommets fortement reliés (des cliques).

Par analogie, les schémas relationnels élaborés en analyse criminelle peuvent être évalués par des tâches similaires. Sur un schéma représentant des communications téléphoniques, il est fréquent de chercher à identifier des numéros en relations avec des numéros spécifiques (retrouver un voisin commun entre plusieurs sommets). Lors de l'analyse d'un trafic de marchandise, la détection des intermédiaires entre deux individus particuliers est également une question récurrente afin de comprendre la structure d'un réseau (trouver l'existence d'un chemin entre deux sommets). L'efficacité d'une visualisation est donc évaluée pour des tâches spécifiques et ne saurait être démontrée globalement.

Finalement, il est important de souligner qu'il n'y a jamais de « meilleure » solution. Une représentation est plus ou moins adéquate en fonction de ce qui est attendu. Les indicateurs mentionnés peuvent constituer des indices d'efficacité, mais son évaluation est toujours relative ; la solution est comparée à une visualisation de référence. C'est par la distance à cette représentation étalon que la qualité de la représentation est estimée. Le choix de celle-ci est donc primordial. Il s'agit de répondre à la question « comparé à quoi ? » et de définir en quoi une telle comparaison a du sens. La représentation de référence est généralement la solution la plus utilisée dans la pratique et/ou celle étant reconnue comme la plus efficace pour représenter la situation analysée. Les représentations doivent par ailleurs contenir la même information et les différences évaluées doivent être contrôlées. Modifier simultanément plusieurs variables visuelles, telles que la couleur et la position, empêche d'identifier clairement les causes des effets observés.

Evaluations préliminaires

Tel que relevé précédemment, nous soupçonnons que les visualisations produites dans la pratique de l'analyse criminelle opérationnelle sont disparates. Des évaluations préliminaires ont été effectuées avec des praticiens et des étudiants afin de tester cette hypothèse.

Disparités de représentation

Les trois exemples ci-dessous illustrent la variabilité observée lorsqu'il est demandé à différentes personnes de produire des représentations relationnelles. Les deux premiers exemples se basent sur la représentation d'une information sommaire (une phrase), afin de montrer la variabilité dans des situations simples. Le troisième exemple est tiré d'un exercice de représentation basé sur une affaire complète. L'objectif est de montrer à l'aide d'indicateurs objectifs (nombre d'entités, de liens, structures des schémas, etc.) l'étendue des variations observées sur une affaire réelle. L'impact des choix de représentation sur l'analyse et la communication est discuté dans la section suivante.

Représentation d'un scénario simple

Lors d'un examen théorique sur l'analyse criminelle opérationnelle, une des questions posées consiste à dessiner un schéma relationnel à partir du scénario suivant :

Dessinez un schéma relationnel qui représente cette situation :

« Ve 16.02.07, Lausanne, rue de bourg, un individu a menacé une dame avec une arme blanche ».

66 Evaluations préliminaires

La situation exposée est volontairement succincte et simple, pour permettre à l'étudiant de proposer une solution en quelques minutes. Dans le cadre de l'examen, l'objectif consiste à évaluer la capacité de l'étudiant à identifier les entités présentes dans un texte et leurs relations. Cette question d'examen est également utilisée lors d'examens oraux pour discuter des rôles de la représentation relationnelle en regard de différents objectifs d'analyses. Cet exemple d'exercice est présenté ici, afin de montrer la variabilité des représentations créées sur la base d'une seule phrase.

Pour cette illustration, soixante résultats d'étudiants ont été étudiés. La question a également été posée à six analystes (six collègues travaillant comme analyste criminel opérationnel, au sein d'une police).

Sur les 66 participants, il a été relevé 48 schémas différents pour représenter la même phrase constituée de six informations : un auteur (individu), une victime (dame), une arme (arme blanche), un événement (menace), une date (16.02.07), un lieu (Lausanne, rue de bourg).

Les schémas ont été regroupés en quatre catégories (Type A à D) définies selon quatre structures générales de représentation :

Type	Structure générale	Nombre de participants
A	<p>Auteur — Evénement — Victime</p>	16
B	<p>Auteur — Arme — Evénement — Victime</p>	8 (dont 2 sans l'entité victime)
C	<p>Auteur — Arme — Victime</p>	16
D	<p>Arme — Auteur — Evénement — Victime</p>	17 (dont 4 sans l'entité couteau)

Ces quatre structures ont été identifiées en comparant les 48 solutions (l'ensemble des représentations peuvent être consultées dans l'annexe quatre). Elles décrivent quatre manières récurrentes de représenter l'information. Dix schémas ne sont pas classés parmi ces quatre formes, car leurs structures sont singulières. Les plus grandes variations sont observées au niveau de l'intégration de l'information temporelle et du lieu. Globalement, les informations sont représentées de quatre façon différentes : en tant qu'entité (une icône), un attribut (un texte lié à une icône ou un lien), un lien (un trait entre deux icônes), un cadre (entourant un ensemble d'icônes et de liens).

Le tableau ci-dessous résume la prévalence des choix effectués :

Information	Représentation				Total
	Entité	Attribut	Lien	Cadre	
Auteur	64	1			65
Victime	62	2			64
Arme	54	9			63
Événement	31		18		49
Date	27	28		3	58
Lieu	19	27		5	51

Alors que l'auteur, la victime et l'arme sont généralement identifiés comme des entités, l'événement n'est représenté comme une entité que trente et une fois. Dix-huit personnes ont représenté l'événement comme une relation (principalement entre l'auteur et la victime).

D'un point de vue de la représentation des relations, il est intéressant de noter qu'aucun participant n'a tracé de lien entre plus de deux entités. Dix-huit personnes ont utilisé des flèches et un étudiant a utilisé une double-flèche. La flèche est généralement exploitée pour décrire le sens du lien entre l'auteur et la victime. Dans cet exemple, la flèche met en évidence le rôle des personnes.

Les informations spatio-temporelles sont principalement représentées comme des entités ou des attributs. En tant qu'attribut, la date et le lieu sont des caractéristiques de l'événement. Sept étudiants ont représenté la date et/ou le lieu sur une autre entité (auteur, victime et/ou arme).

R e p r é s e n t a t i o n d ' u n e v e n t e

Lors d'un cours sur la visualisation de données, les participants doivent créer un schéma relationnel représentant une vente :

Dessinez un schéma relationnel qui représente cette situation :

« Monsieur A vend n exemplaires de l'objet X à monsieur B, pour le prix de Y CHF ».

Sur les 10 participants, cinq schémas différents ont été produits pour représenter la même phrase constituée de cinq informations : deux individus, un événement, un ensemble d'objet, un prix. Les schémas ont été regroupés en deux catégories (Type A à B) définies selon deux structures générales de représentation. Les trois autres schémas ont des structures différentes²⁰:

Type	Structure générale	Nombre
A		4
B		3

Les seules informations qui sont représentées de façon analogue sur l'ensemble des schémas sont les individus qui sont représentés par des entités. A noter que dans l'exemple précédent, la quasi-totalité des participants a également représenté les deux individus (l'auteur et la victime). Deux personnes n'ont pas représenté la victime et deux autres n'ont représenté que l'événement. A noter que dans cet exemple aucun participant n'a identifié la vente (un événement) comme une entité. Ces deux exemples illustrent que même pour un scénario simple (une seule phrase) de nombreuses variations sont observées entre les différents participants. L'exemple suivant vise à quantifier l'étendue des variations pour représenter un scénario plus complexe issu d'une affaire réelle.

²⁰ Cf. Annexe 4 : Evaluations préliminaires - Représentation d'une vente, page 268

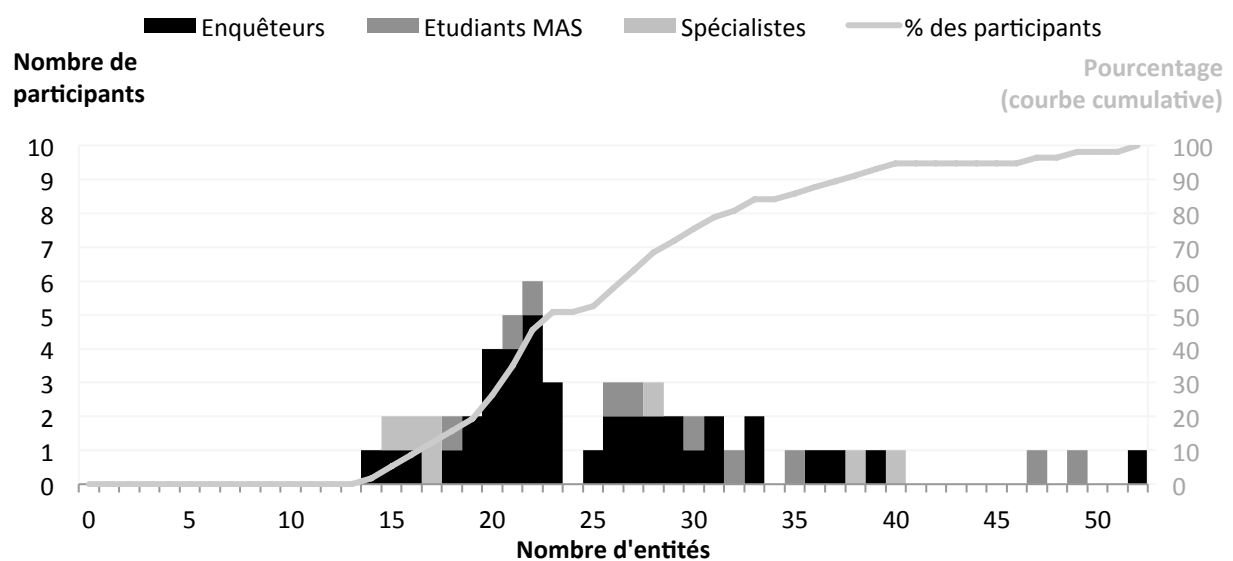
Représentation d'une affaire complexe

Lors de formations de base à l'analyse criminelle opérationnelle²¹ pour des enquêteurs de police judiciaire, un exercice de visualisation relationnelle est soumis aux participants. L'objectif de l'exercice consiste à représenter par un schéma relationnel, les informations contenues dans un rapport de deux pages (contenant environ neuf cent mots).

Les schémas produits par quarante enquêteurs ont été comparés. L'exercice a également été soumis à dix étudiants de Master professionnel²² et à sept spécialistes ayant une expérience sur la conception de schémas.

L'ensemble des cinquante-sept schémas produits n'est pas annexé. Seuls des tableaux récapitulatifs sont présentés (voir notamment l'annexe quatre). Les indicateurs de variabilités analysés sont : les nombres totaux d'entités et de liens représentés et les types d'entités et de liens exploités.

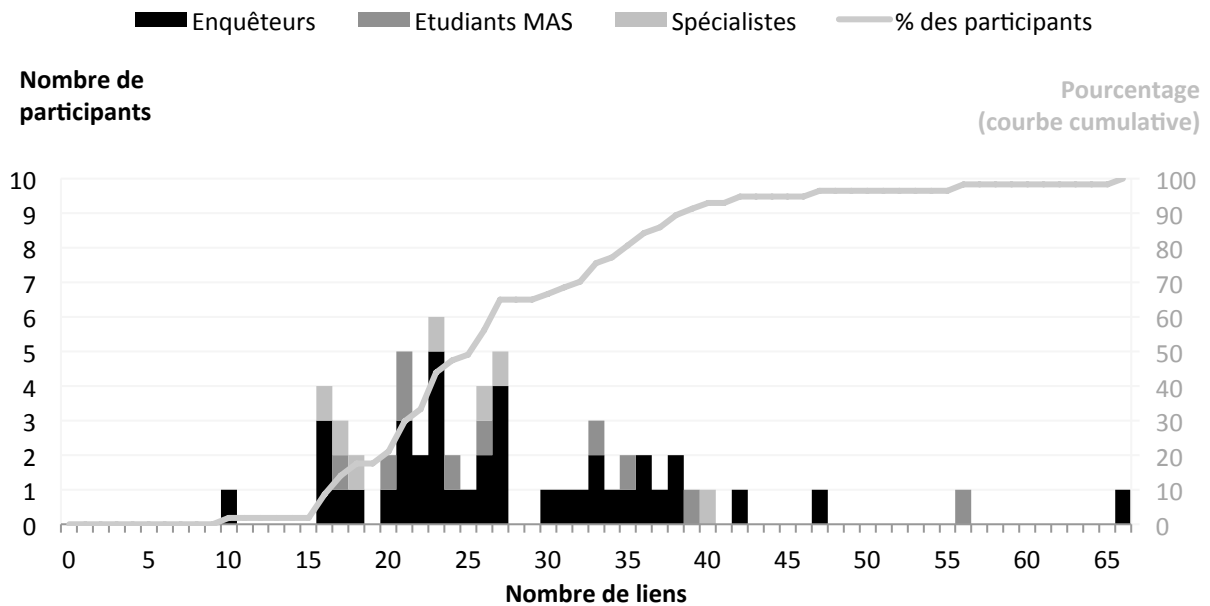
Nombre d'entités



²¹ Formation « Analyse Criminelle Opérationnelle – niveau I » organisé par l'institut suisse de police et l'école des sciences criminelles.

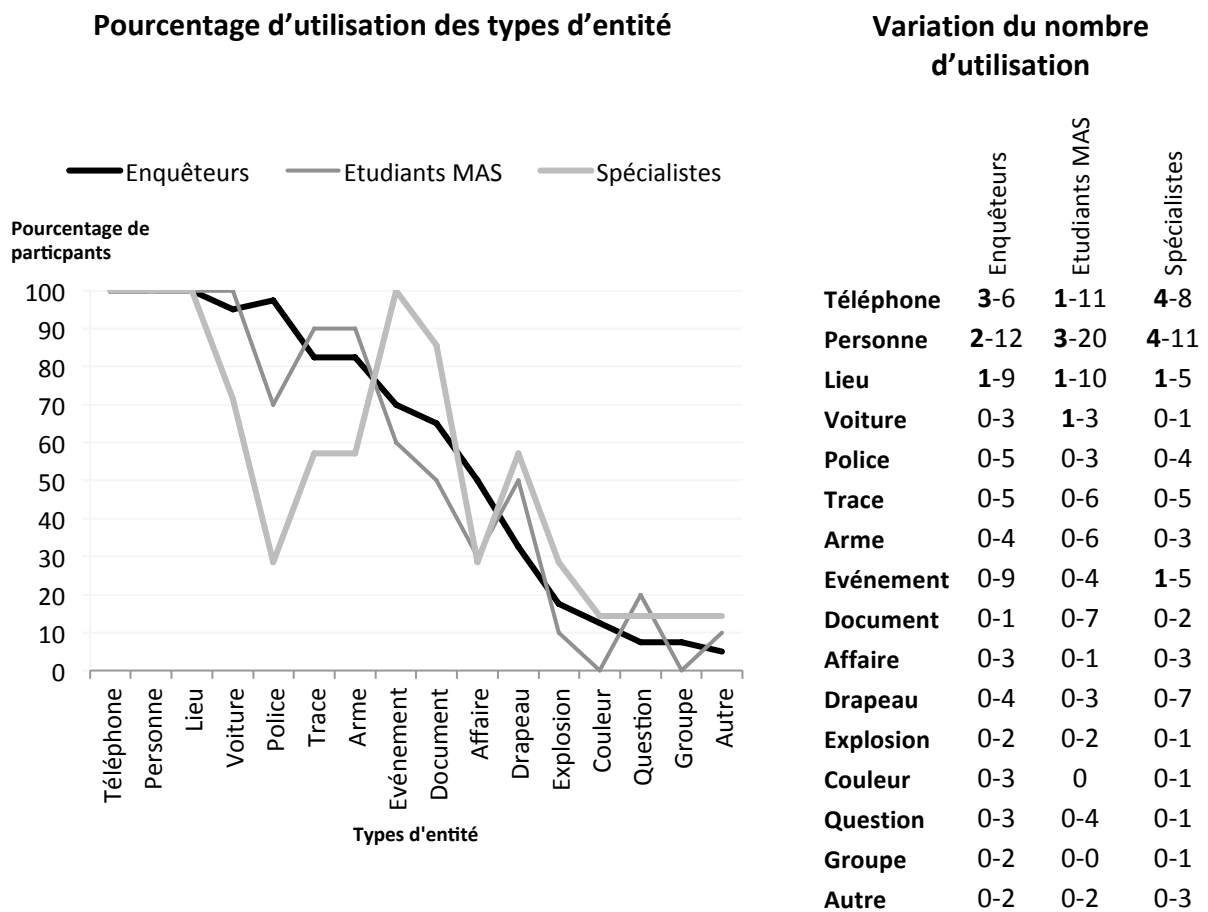
²² « Master of Advanced Studies en lutte contre la criminalité économique » de l'institut de lutte contre la criminalité économique de la haute école ARC de Neuchâtel.

N o m b r e d e l i e n s



Alors que l'ensemble des participants a reçu la même consigne et les mêmes données à représenter, les nombres d'entités et de liens exploités varient considérablement (de 13 à 52 entités et de 10 à 66 liens). Les variations sont importantes indépendamment du fait que le schéma a été produit par des personnes ayant l'habitude ou non de produire de telles représentations. La principale raison de cette variabilité semble être une différence de perception sur l'usage du schéma. Pour certains, l'objectif consiste à représenter l'ensemble des informations utiles afin de les regrouper dans un schéma unique. Alors que pour d'autres l'objectif d'un tel schéma est de synthétiser les informations de la manière la plus simple et la plus claire possible.

Types d'entité



Le programme utilisé offre la possibilité de définir des types d'entités caractérisés par des icônes spécifiques (tels que ceux présentés dans les exemples précédents). Les graphiques ci-dessus regroupent les différents types en catégories génériques qui sont définis d'un point de vue sémantique. Par exemple, les icônes «homme», «femme», «inconnu», etc. sont regroupés dans la catégorie «personne». Des tableaux, présentant le nombre d'utilisation de chaque type spécifique, sont présentés en annexe. Il faut noter qu'aucun type (c.-à-d. aucune icône) n'a été exploité par l'ensemble des participants. Au niveau des catégories d'icônes, seuls les téléphones, les personnes et les lieux ont été utilisés par tous les participants. Il est intéressant de noter que tous les spécialistes ont également exploité des icônes représentant des événements, alors que trente pourcent des enquêteurs et quarante pourcent des étudiants MAS n'ont pas exploité ce type d'entité. Par ailleurs, seuls trente pourcent des spécialistes ont représenté des icônes de type «police». Finalement, le nombre d'utilisation de chaque type d'entité varie également fortement entre les participants (cf. tableau ci-dessus). Alors que l'affaire représentée est identique, certains participants n'ont exploité, par exemple, que deux icônes de personnes et d'autres vingt. De plus, alors que le rapport ne fait mention que de quatre numéros de téléphone, de une à onze icônes de téléphones sont exploitées.

T y p e s d e l i e n s

Les détails des types de lien utilisés sont présentés en annexe. Globalement, quarante types différents ont été exploités et aucun n'a été utilisé par l'ensemble des participants. Il faut relever que les types sont prédéfinis par le programme utilisé (un seul participant a défini des types personnalisés). Pour chaque type de lien une couleur peut être choisie, le programme ne définit pas d'attributs spécifiques en fonction du type de lien. La seule raison visuelle d'exploitation des types est de faire varier la couleur des liens.

Vingt-neuf participants ont également utilisé des cadres pour représenter des relations entre les entités. Les cadres sont exploités pour décrire différents types de relations : des relations géographiques ou spatio-temporelles (15x), des groupes d'entités (groupe de personnes, de téléphones, etc.) (13x), des événements (5x) et la source des informations (6x). A noter également que huit des participants ont représenté des liens entre des cadres et d'autres entités (pour exprimer une relation commune avec l'ensemble des entités présentes dans le cadre).

D i s c u s s i o n

Lors de la conception d'un schéma relationnel, de nombreux choix doivent être effectués pour transposer l'information contenue dans du texte sous la forme d'un graphe. Comment décomposer les informations en entités et relations, quelles sont les entités pertinentes, comment représenter les relations, quels sont les attributs des entités et des relations, sont autant de questions qui sont répondues de manière variée suivant les concepteurs. Même lorsque l'information initiale est succincte, les variations observées sont importantes. Il ne semble pas exister de consensus émergent. Les choix de représentation et les objectifs sont multiples.

Dans le cadre de ces trois exemples, aucun objectif spécifique n'a été imposé aux participants. Il s'agit peut-être d'une des causes expliquant les variations observées, mais alors cela signifierait aussi une grande disparité de perception sur l'utilité d'une telle visualisation. Pour certains, le schéma doit représenter l'ensemble des informations de manière détaillée afin d'exploiter la représentation comme une mémoire de travail, pour d'autres, décomposer l'information peut servir à identifier des lacunes ou des erreurs de raisonnement. Finalement, la visualisation peut également être exploitée pour synthétiser les éléments pertinents afin de faciliter la communication d'une affaire complexe.

Impact sur l'analyse et la communication

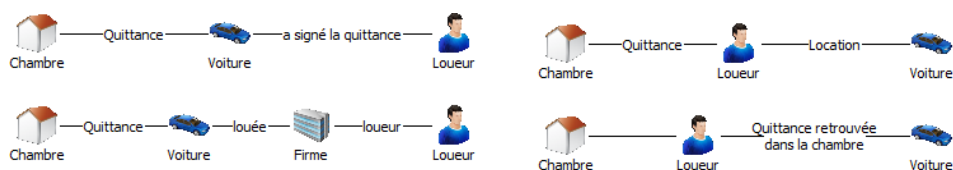
Certains choix de visualisation sont-ils plus adaptés que d'autres ? Quel est l'impact de ces choix sur l'analyse et la communication ? Cette section présente les observations effectuées sur le dernier exemple, ainsi que sur des exemples d'exploitation de la visualisation dans le cadre d'enquêtes, afin de présenter l'impact de ces choix sur l'analyse et la communication.

Représentation d'une affaire complexe

Dans la section précédente, des indicateurs objectifs ont été utilisés afin de présenter l'étendue de la variabilité observée lors de la création d'un schéma relationnel. Certaines portions particulières de l'affaire et les représentations produites sont reprises ici afin d'illustrer l'impact des choix sur l'analyse et la communication des informations.

Des ambiguïtés de représentation ont été observées sur quarante-trois schémas (75% des schémas). Ces ambiguïtés peuvent être classifiées en deux catégories. D'une part, lorsqu'une entité n'est pas identifiée et représentée comme telle, une démultiplication de liens peu clairs est observée (situation observée sur vingt-neuf schémas). D'autre part, des liens mal positionnés peuvent engendrer des difficultés de compréhension. Un lien mal placé est un lien qui aurait dû être tracé avec une autre entité pour minimiser l'ambiguïté (situation observée sur trente-cinq schémas). Les portions de schémas ci-dessous illustrent cette problématique :

Une quittance de location de véhicule a également été découverte dans la chambre... elle concernait un véhicule ... provenant de la firme... la personne ayant rempli et signé la quittance est un certain...



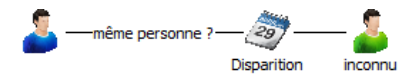
La quittance n'est pas identifiée comme une entité, les relations entre la chambre, le loueur et la voiture ne sont pas claires et peuvent porter à confusion.

74 Evaluations préliminaires

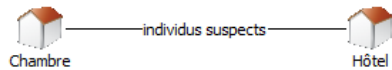
2 personnes ... se présentèrent devant les agents. Alors que ceux-ci les enjoignaient de présenter leurs papiers, ils ouvrirent immédiatement le feu sur les gendarmes au moyen d'armes de poing. Les gendarmes ripostèrent et l'un des 2 individus fut touché...



L'événement (la fusillade) n'est pas identifié comme une entité. Les liens tracés entre les entités ne sont donc pas explicitement rattachés à un événement particulier : l'ensemble des coups de feu fait-il référence à plusieurs événements, les utilisations des armes ont-elles été faites au même moment, etc.



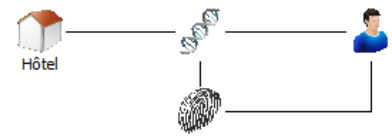
Le lien « même personne » devrait être lié à l'inconnu



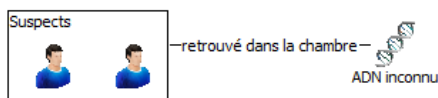
Le lien n'est pas clair (lien avec des individus)



Le rapport dit que l'individu est un activiste (le lien ne devrait donc pas être avec le corps)



La trace biologique et la trace digitale devraient être liées avec le lieu et la personne, mais pas entre elles.



Le lien devrait être avec le lieu et la trace ADN (le rapport ne dit pas que les individus sont les sources de la trace).

D'autre part, huit des schémas contiennent des liens erronés (des relations qui ne sont pas décrites dans le texte). Cinq participants ont représenté par plusieurs entités une même personne. Ce qui a pour conséquence de ne pas observer visuellement l'ensemble des informations en lien avec cette personne. Cinq participants n'ont pas différencié visuellement les hypothèses des faits. Finalement, quarante-quatre participants (77% des schémas) n'ont pas représenté une incertitude liée à l'identité d'une des personnes impliquées dans l'affaire. En effet, alors qu'il existe une convention pour représenter l'incertitude sur les liens (traits-tillés), aucun standard n'est défini pour les entités.

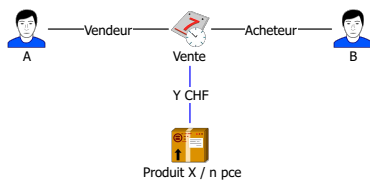
Discussion

Globalement, une mauvaise définition des entités et des relations peut engendrer des ambiguïtés sur la représentation. Les erreurs observées sont probablement liées au fait que le concepteur d'un schéma connaît bien l'affaire qu'il représente. Le schéma produit est

certainement clair à ces yeux. Pourtant pour un lecteur externe qui n'a pas connaissances des informations initiales, certains choix de représentation peuvent être sujets à interprétation et conduire à une mauvaise compréhension de l'affaire.

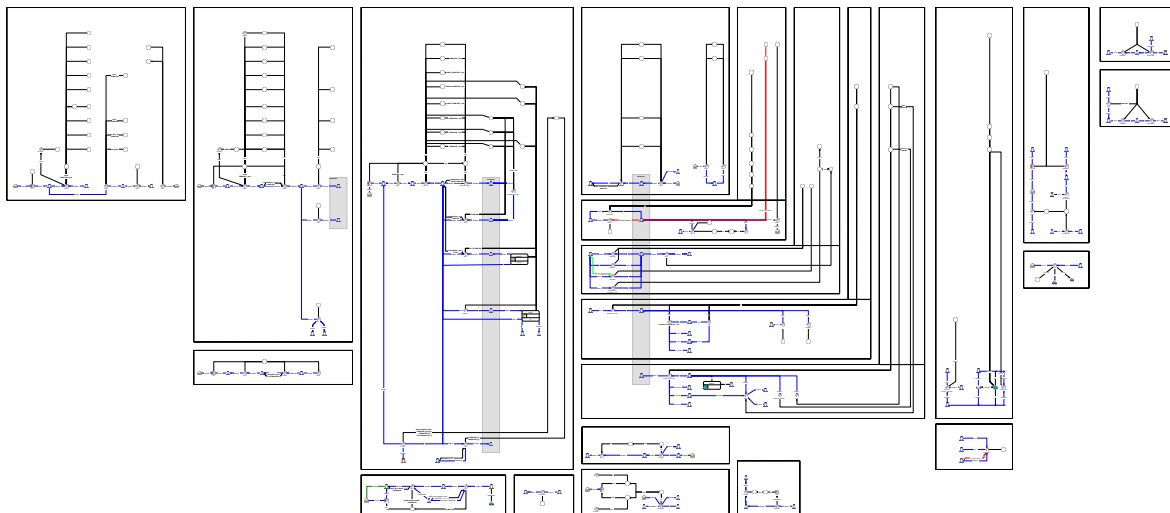
R e p r é s e n t a t i o n d ' u n t r a f i c d e m a r c h a n d i s e s

L'exemple présenté ici est tiré d'une affaire réelle qui porte sur un trafic important d'objets volés. Il est présenté afin d'illustrer l'impact des choix effectués en matière de représentation sur l'utilisabilité des schémas. En effet, le schéma ci-dessous a été effectué en se basant sur une structure relationnelle élaborée en suivant des recommandations générales définies dans les manuels de l'analyse criminelle²³ : représenter l'ensemble des informations et lier les



entités concernées lors de l'événement (la vente). Dans le cadre de cette analyse, l'ensemble des informations a été intégré dans une banque de données relationnelle structurée autour d'une table centrale contenant les événements.

L'objectif du schéma est de représenter l'ensemble des informations issues de déclarations des personnes impliquées dans l'affaire, afin de comparer à la fois les versions d'une même personne et les versions de plusieurs personnes pour un même événement.



Chaque cadre contient l'ensemble des informations issues des déclarations d'une personne entendue par les services de police. Les entités ont été positionnées afin de faciliter le plus

²³ Cf. Recommandations et standards, page 45

76 Evaluations préliminaires

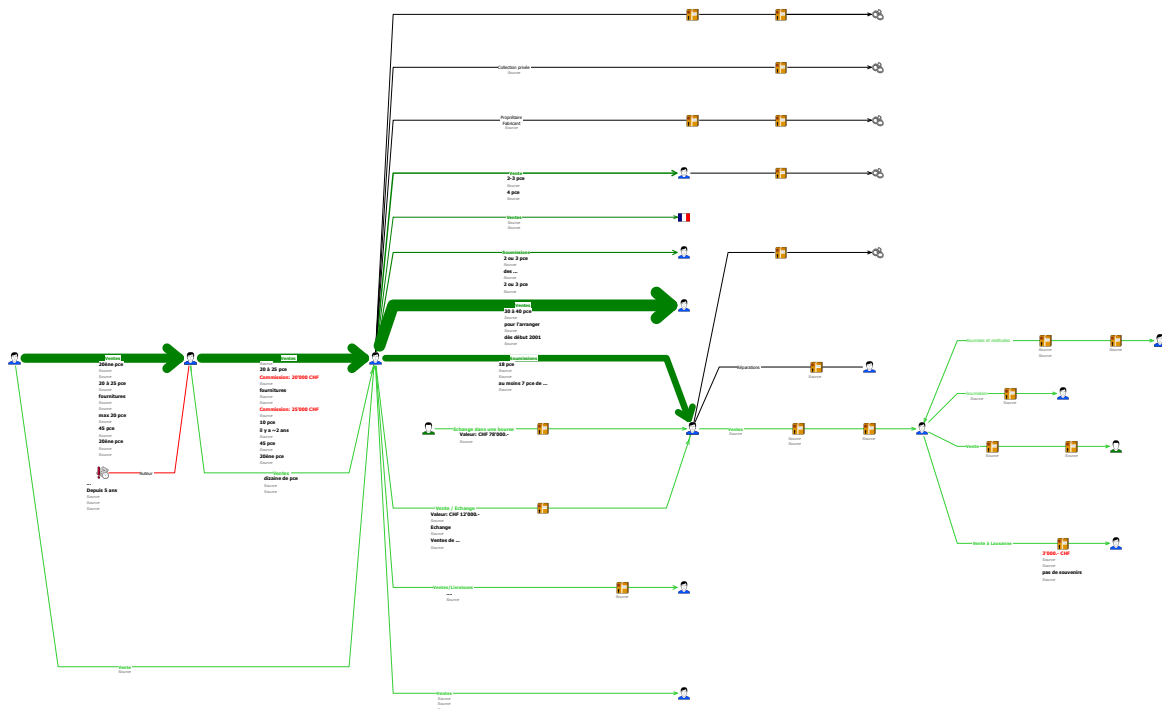
possible la comparaison des déclarations (notamment en positionnant les entités « objets » (cercles blancs) décrivant des produits similaires à la même position sur l'axe vertical). Ce schéma a été présenté aux enquêteurs en charge de l'affaire. Au vue de sa complexité, il n'a pas suscité un intérêt particulier et n'a jamais été exploité par ceux-ci pour analyser les informations à disposition (notamment lors de la préparation d'auditions ou l'écriture de rapports).

Par la suite, le schéma a été restructuré en modifiant la représentation de la vente. Celle-ci a été représentée par un trait dirigé entre l'acheteur et le vendeur.



objets vendus, les prix et les sources d'information (les références aux procès-verbaux d'auditions) ont été représentées sous la forme d'attributs.

Le schéma final ainsi produit est le suivant :



En plus du changement de structure, les déclarations portant sur une même vente (ou un ensemble de ventes) ont été regroupées. Par même vente, il est entendu une vente entre deux individus particuliers. Les déclarations sont représentées par des attributs sur les liens. Ce schéma visiblement plus efficace que le précédent offre une image claire de la structure globale du trafic. Il permet également de comparer visuellement les quantités de produits

vendus qui sont représentées par l'épaisseur des relations. Plusieurs schémas ont été produits sur ce modèle et exploités dans l'enquête. Malheureusement, il n'ont pu être utilisés qu'à la fin de l'enquête lors de la mise en accusation par le magistrat instructeur. En effet, les outils exploités pour produire ces schémas (la banque de données et l'outil de visualisation) ne permettent pas d'effectuer dynamiquement un tel changement.

D i s c u s s i o n

Cet exemple, illustre deux problèmes principaux. Premièrement, les choix effectués pour définir la structure et les attributs visuels exploités (par exemple, l'épaisseur des liens) d'un schéma relationnel peuvent avoir un impact considérable sur la lisibilité et le potentiel d'exploitation de la représentation. Des règles réutilisables peuvent-elles être définies afin de guider ces choix ? Le schéma présenté semble tout au moins constituer une solution plus efficace pour analyser cette situation particulière. Les trafics de marchandise étant des problèmes récurrents, la solution présentée est certainement exploitable pour le traitement d'affaires analogues.

Deuxièmement, il est nécessaire de disposer de mécanismes permettant de modifier dynamiquement une représentation. La structure visuelle exploitée initialement a été imposée par la structure relationnelle de la banque de données sous-jacente. En effet, les outils n'offrent généralement pas la possibilité d'aggréger des données (regrouper les déclarations portant sur une même vente) ou la possibilité de transformer la structure de la représentation (par exemple en transformant une entité en lien). Cet exemple illustre l'impact des limites imposées par les outils : l'outil ne guide pas l'utilisateur vers une solution adaptée pour son problème et il n'offre que peu de possibilités pour modifier dynamiquement une structure de graphe (notamment lorsque les choix initiaux s'avèrent inadaptés).

R e p r é s e n t a t i o n d e s é r i e s d e c a m b r i o l a g e

Des schémas relationnels sont également souvent exploités dans le cadre d'enquêtes sur des séries de cambriolages. Les relations détectées entre les différents cas de la série (cas liés par des traces, liés spatio-temporellement, liés par des véhicules, etc.) sont visualisées afin de synthétiser l'ensemble des informations et ainsi faciliter la compréhension globale de la série, tout en offrant la possibilité d'en questionner les parties.

Les deux exemples présentés ci-dessous sont des cas réels d'exploitation en cours d'enquête. Le premier schéma a été reçu par un enquêteur de la police judiciaire qui souhaitait savoir s'il

était possible de le rendre plus explicite. Le second schéma a été restructuré, car l'enquêteur l'ayant produit le jugeait difficile à comprendre. En effet, lorsque le nombre d'informations à représenter augmente, conserver la lisibilité d'un schéma n'est pas une tâche évidente.

Schéma initial

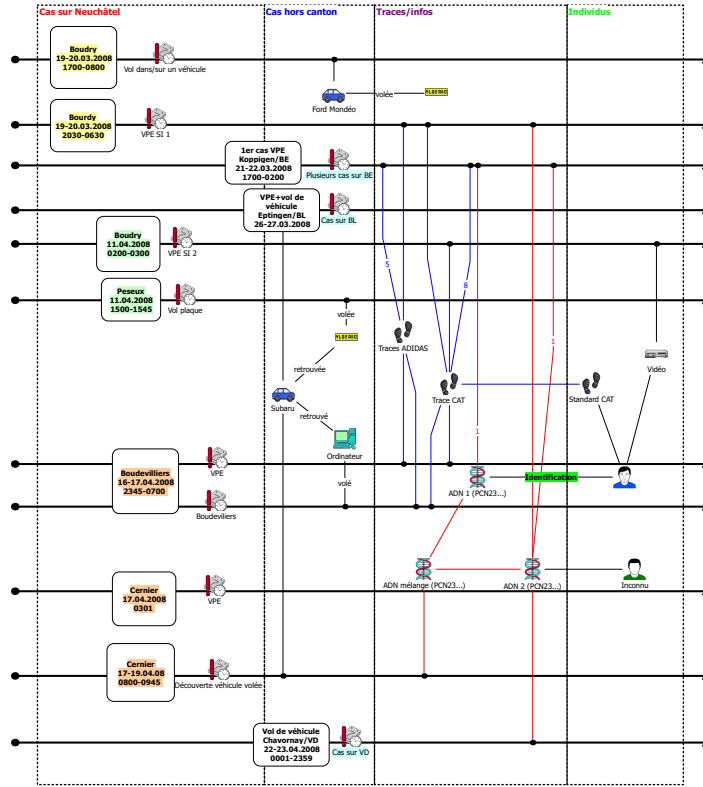
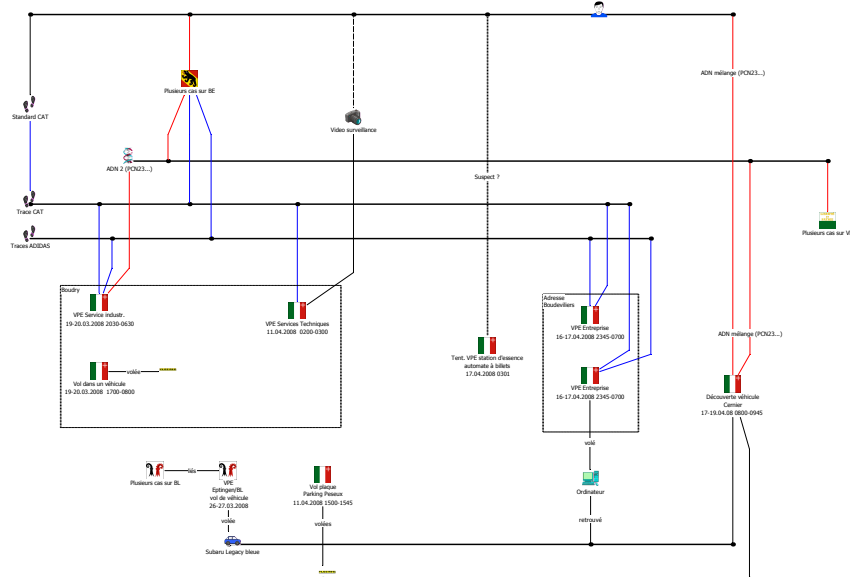
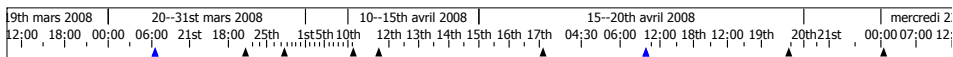


Schéma restructuré



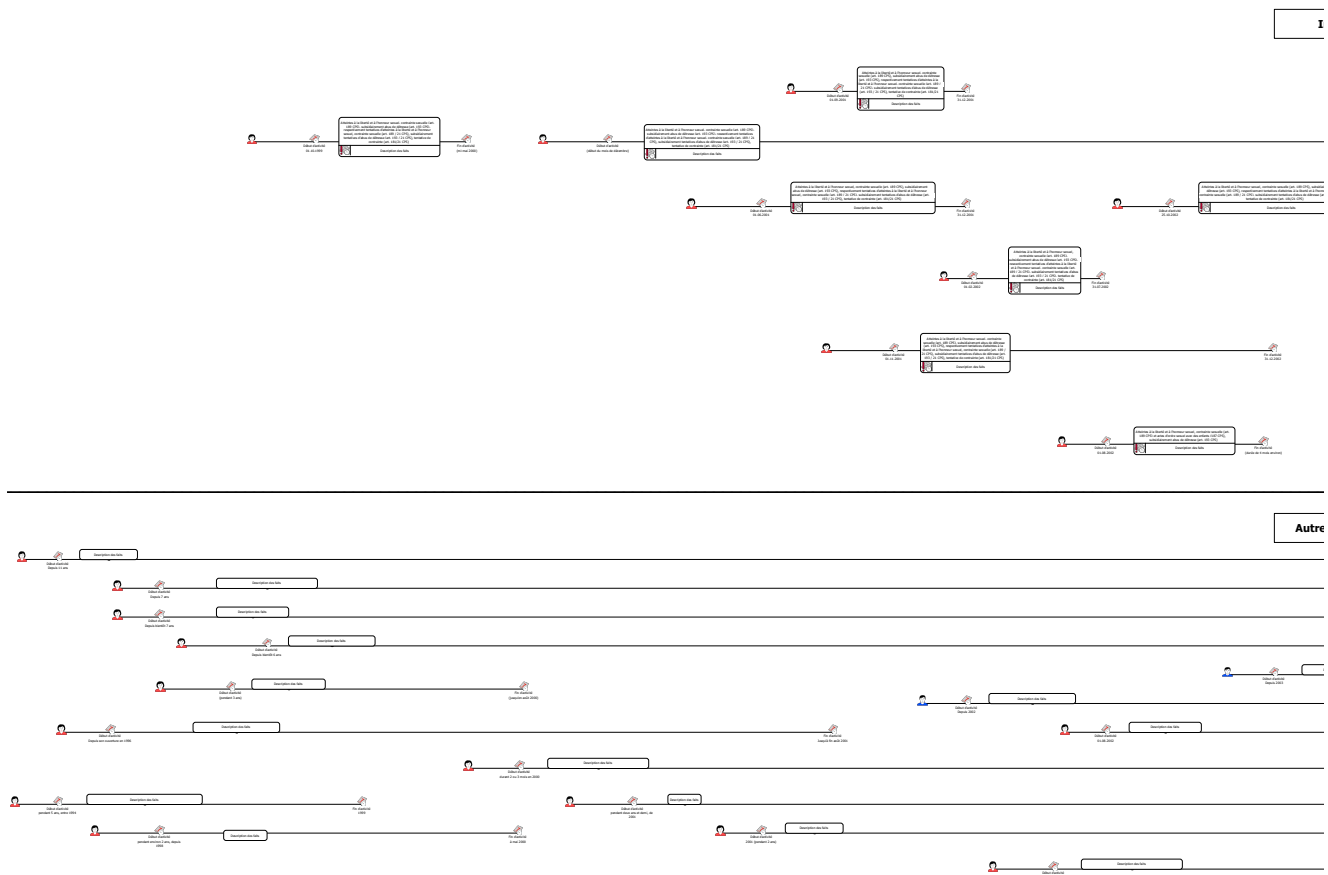
Les schémas restructurés sont plus faciles à appréhender et ont répondu aux attentes des enquêteurs. Plusieurs améliorations ont été apportées. Sur le premier, la structure globale a été revisitée. Sur le second, le placement des entités a été optimisé.

D i s c u s s i o n

Alors que les deux schémas initiaux sont très différents, les deux solutions proposées sont très similaires. En effet, les deux affaires portent sur les mêmes types d'information et l'objectif d'analyse est similaire : reconstruire et décrire une série. Les deux schémas produits sont des schémas d'événements²⁴. Est-il alors possible de formaliser une solution adaptée pour cette situation récurrente ?

R e p r é s e n t a t i o n d ' u n c a s d ' a b u s s e x u e l

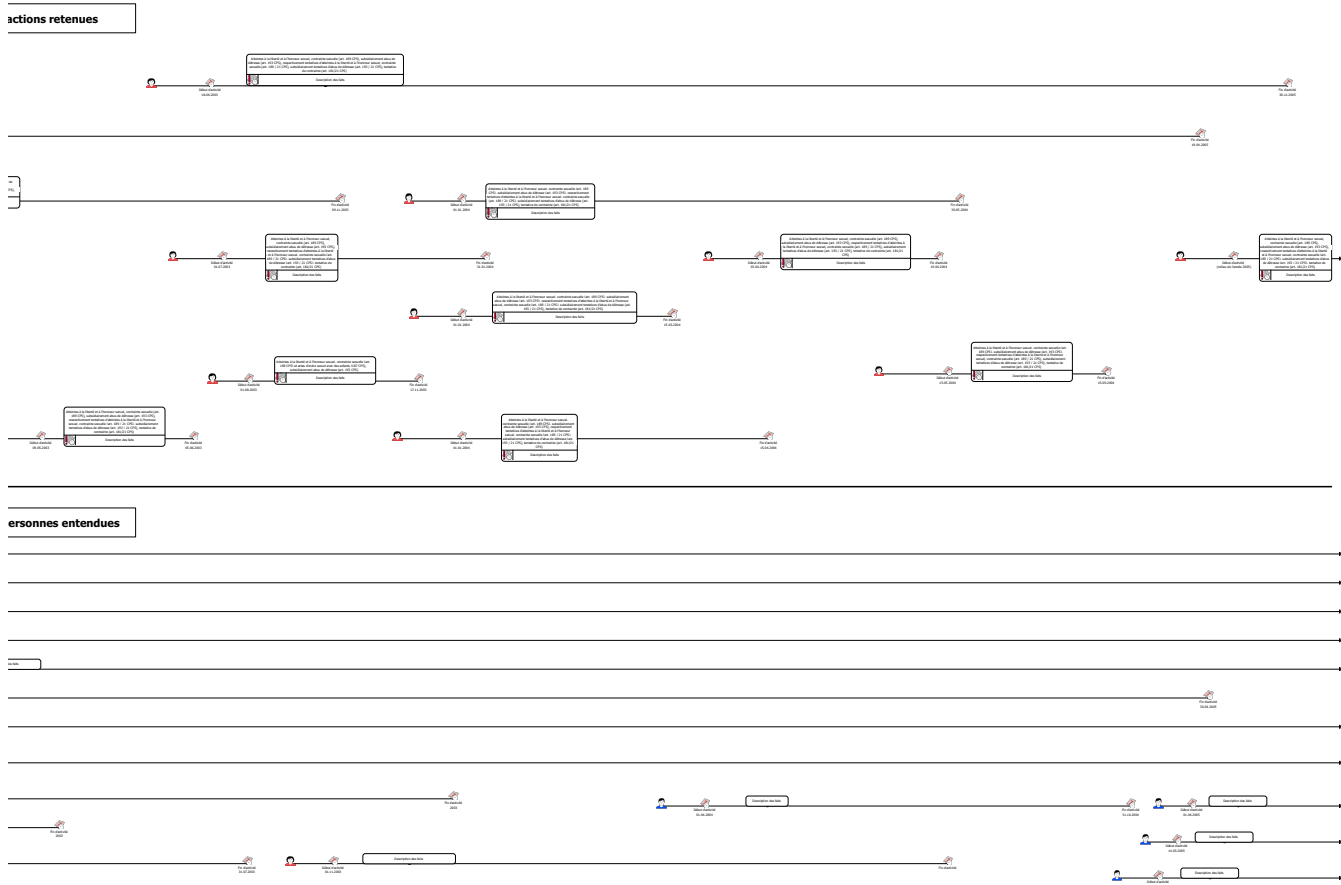
La visualisation peut également être exploitée au tribunal pour faciliter la communication d'un argument et accroître l'impact émotionnel d'un propos (Feigenson & Sherwin, 2007). Les effets psychologiques et rhétoriques engendrés par la visualisation devraient être évités ou au moins contrôlés, afin que la réalité des informations présentées ne soit pas détournée pour soutenir



²⁴ Cf. Analyse et visualisation relationnelle, page 43

une interprétation fallacieuse des faits. L'exemple de visualisation ci-dessous, a été élaboré à la fin d'une enquête pour faciliter l'appréhension du dossier par les parties et le tribunal. L'affaire porte sur la mise en accusation d'un hôtelier ayant abusé sexuellement de certaines de ces employées. Au cours des cinq années sur lesquels portent l'enquête, une trentaine d'employés a travaillé dans l'hôtel et les déclarations sont nombreuses. L'objectif de la visualisation est de résumer les informations sur les sévices subis par les victimes, ainsi que de montrer les périodes d'activités des employés afin de comparer les différences de durées entre les employés victimes et les employés non victimes.

Il s'agit d'un schéma chronologique où chaque employé est représenté par une ligne horizontale (icône de personne : rouge pour les femmes et bleu pour les hommes), les dates de début et fin d'activités sont représentées par des icônes reliées par une ligne horizontale qui décrit donc la période d'activité. Des informations textuelles sont ajoutées sur chaque ligne. Celles-ci résument les déclarations des employés concernant les sévices subis (ou l'absence de sévices). Le graphique est divisé en deux parties distinctes sur l'axe vertical : en haut, les employés ayant déclaré avoir subi des actes d'ordre sexuel et en bas, les employés non victimes.



Une telle affaire a une dimension émotionnelle relativement importante, notamment au regard du nombre de victimes. De plus, la dimension temporelle a également un impact sur l'interprétation de la gravité des infractions. Ce schéma a été présenté au tribunal. Il a également été utilisé lors de cours, comme exemple d'exploitation de la visualisation sur des affaires de mœurs. Il illustre, la possibilité de synthétiser de façon simple un nombre important d'informations et la manière de faciliter la comparaison de périodes. Toutefois, il induit également une interprétation implicite fallacieuse des périodes représentées par les lignes horizontales. En effet, la longueur des lignes semble *a priori* représenter la période pendant laquelle l'employé a subi les actes incriminés. Ce qui n'est absolument pas le cas, puisque les dates choisies sont celles de début et de fin d'activité. Deux des victimes représentées sur le schéma ont travaillé sur une période de plusieurs années, ce qui induit l'hypothèse qu'elles ont subi plus longtemps des agressions sexuelles, ce qui n'est pas forcément le cas. L'exploitation de cet exemple dans le cadre de cours a été complètement revue, suite à un lapsus, effectué inconsciemment par l'auteur, sur la signification des périodes. En effet, cet exemple illustre un biais induit par la représentation. Bien que la visualisation a été construite de façon intègre par rapport aux données initiales (la chronologie est respectée, l'ensemble des informations a été inclus, les informations sont triées de façon logique et documentée), une mauvaise interprétation reste possible. La signification induite par une visualisation peut donc dépasser la réalité des informations représentées (Feigenson & Sherwin, 2007).

Il semble *a priori* relativement difficile de se prévaloir d'un tels biais. Une discussion plus élaborée sur les objectifs attendus de la visualisation (pourquoi est-il relevant de comparer les périodes d'activités entre victimes et non victimes ?), ainsi qu'une discussion avec des lecteurs sur la manière dont ils interprètent le schéma auraient peut-être permis de détecter ce biais. Cet exemple, illustre l'importance de deux étapes importantes du processus de visualisation : la définition et l'évaluation des objectifs de l'analyse ou des observations à communiquer et l'évaluation du produit de renseignement élaboré.

Synthèse

Constat

Contexte d'exploitation

Les décisions prises dans le cadre de l'enquête judiciaire sur la base de visualisations, peuvent avoir un impact important (notamment au niveau du jugement) ou engendrer des coûts importants en terme de ressources et de temps (par exemple, lorsqu'il s'agit de mettre sur pied des opérations policières ou des mesures de prévention). La visualisation est exploitée pour soutenir diverses décisions qui peuvent également avoir un impact sur la résolution du cas, voire la condamnation d'un prévenu. Les représentations graphiques se doivent donc d'être intègres. Elles doivent faciliter l'analyse et l'appréhension des renseignements en représentant la réalité des informations sans ajouter des biais. Les choix effectués pour produire des représentations se doivent donc d'être guidés afin de concevoir des produits de renseignement efficaces.

La visualisation peut être exploitée pour soutenir un nombre importants d'objectifs ; tels qu'analyser des informations collectées au cours de l'enquête, évaluer *a posteriori* une investigation, aider à qualifier les infractions, faciliter l'appréhension d'un dossier et soutenir une argumentation lors du procès. Il n'est *a priori* pas certain qu'une forme de visualisation particulière puisse répondre à l'ensemble des attentes des différents acteurs du processus. Les objectifs et les contraintes évoluent au cours de l'enquête. Les représentations exploitées peuvent donc varier au fil du processus.

Nature des concepts à représenter

La diversité des concepts à représenter dans le cadre de l'enquête criminelle est certainement l'une des causes des difficultés observées. D'une part le domaine du discours (l'ensemble des entités nommées) n'est certainement pas universellement définissable. Une enquête financière porte sur des concepts tels que des comptes et des transactions. En matière de stupéfiants, les investigations portent sur des saisies, des échantillons, etc. Chaque champ d'investigation implique des raisonnements basés sur des concepts souvent spécifiques. Bien que certains concepts puissent être formalisés de manière apparemment générique (comme les événements, les personnes, les objets et les traces), il n'est absolument pas certain qu'il existe des manières génériques de représenter les informations. En outre, le domaine du discours change en fonction de l'évolution des problèmes traités. Une structure définie à un moment précis d'une affaire peut se voir bouleversée par l'arrivée de nouvelles informations.

Au cours du processus, diverses hypothèses sont émises et il est capital de pouvoir rattacher l'ensemble des informations représentées aux informations collectées et à leurs sources. Les différents types de sources doivent être clairement identifiables (tels que les traces physiques et les témoignages), afin de permettre l'évaluation de la fiabilité des informations. Les informations sont souvent imprécises, voire incertaines. Il est donc nécessaire de pouvoir exprimer visuellement l'incertitude. Les hypothèses doivent également être clairement distinguables des informations factuelles et testimoniales.

La gestion de la négation est, par ailleurs, une question ouverte en termes de représentation. Comment représenter par exemple le fait qu'un individu est inconnu des services de police ou représenter le fait qu'un individu nie avoir participé à un événement particulier. Ne pas représenter la négation engendre une confusion avec l'absence d'information. Selon Bertin (2005), l'absence de signe signifie l'absence d'information. La négation doit donc être représentée sur un graphe. Toutefois tracer un lien entre deux entités est généralement interprété comme la présence d'une association et non de la négation. Visuellement la première impression est donc faussée. Des conventions doivent donc être définies pour exprimer la négation et la distinguer de l'absence d'information.

Variabilité des représentations

Les exemples présentés mettent en évidence une problématique fondamentale de la représentation relationnelle des informations : il n'existe pas de consensus général pour représenter les informations sur un graphe. Suivant les perceptions, une même information

peut être représentée sous la forme d'une entité, d'un lien, d'un cadre ou d'un attribut. La transformation d'une information textuelle en représentation relationnelle ne semble en effet pas être une tâche triviale. Les graphes offrent une liberté d'expression très grande. Tous les éléments représentés peuvent être en relation (définition sémiologie d'un graphe) et le nombre de degrés de liberté disponibles est considérable. En d'autres termes, de nombreuses variables visuelles sont exploitables pour encoder une information. De plus, il n'existe que peu de recommandations et de règles sur la manière de les exploiter (c'est le cas notamment de la position des sommets sur le plan, de leur couleur, de leur forme). Le constat est d'autant plus préoccupant en regard de l'impact de ces choix de représentation sur l'analyse et la communication du renseignement.

B i a i s d ' i n t e r p r é t a t i o n e n g e n d r é s p a r l a r e p r é s e n t a t i o n

Différents biais peuvent être engendrés par un usage inadapté de représentations graphiques. Ces biais peuvent être issus d'un mauvais usage de la visualisation (par son créateur) ou d'une mauvaise lecture (par son destinataire). Lors de la création d'une représentation des biais peuvent être ajoutés intentionnellement ou non. Lors de la lecture des biais peuvent avoir de nombreux effets : engendrer de la confusion, une mauvaise interprétation, distraire, faciliter la manipulation, voire limiter la réflexion (Bresciani & Eppler, 2008). En regard du contexte d'exploitation, de tels biais doivent être évités. L'ajout intentionnel de biais ne sera pas discuté dans cette thèse. Par contre le fait qu'une représentation engendre des ambiguïtés lors de sa lecture, ou qu'elle puisse mener à des erreurs de raisonnements sont considérés comme des critères fondamentaux d'évaluation et de comparaison des solutions étudiées : le critère d'intégrité. La transposition d'un discours écrit sur un graphe présuppose que le langage visuel offre la possibilité d'exprimer les concepts de façon cohérente. Les possibilités offertes pour représenter les informations sont multiples et les choix effectués influencent d'une part l'intégrité de la représentation (le graphique représente-t-il toujours la réalité des informations sous-jacentes), mais également la facilité d'appréhension de la représentation.

I m p o r t a n c e d e l ' e f f i c a c i t é d e l a r e p r é s e n t a t i o n

L'efficacité d'une représentation graphique n'est pas définie dans l'absolu, mais toujours en regard d'une autre forme de visualisation. Une représentation est jugée plus efficace si le temps de lecture nécessaire pour répondre à une question d'analyse de façon correcte et complète est plus court qu'en exploitant une autre forme de représentation. L'optimisation de l'efficacité des représentations relationnelles est nécessaire afin d'assurer qu'elles soient

utilisables. En effet, la complexité et la quantité d'informations représentées impliquent généralement d'optimiser la manière de les représenter afin de pouvoir les analyser. Une représentation peu efficace dans le sens où elle ne permet pas une analyse rapide et fiable des informations ne remplit certainement pas son rôle. L'efficacité est donc un critère d'applicabilité important qui ne saurait être négligé et qui peut être évaluée selon des critères objectifs : l'exactitude, la complétude et le temps de réponse pour une question d'analyse déterminée.

V i d e m é t h o d o l o g i q u e

Alors que l'utilisation des techniques de visualisation est recommandée en analyse criminelle afin d'analyser et de communiquer les informations collectées, les manuels et la littérature du domaine n'offrent que peu de recommandations en la matière. Des types génériques de schémas relationnels sont définis et quelques recommandations générales sont proposées²⁵. Les outils de visualisations disponibles donnent l'impression de s'intégrer de façon harmonieuse dans une méthode standardisée. Pourtant, il ne semble pas exister à l'heure actuelle une démarche clairement formalisée guidant les choix à effectuer. Le processus de raisonnement exploité pour transposer une information sur un graphe n'est pas formalisé. Le langage visuel n'est que partiellement décrit. Finalement, les avantages et limites des propriétés visuelles exploitables sur un graphe ne sont pas clairement décrits et évalués. L'élaboration de telles représentations reste, dans une large mesure, basée sur des connaissances tacites liées à l'expérience, à la capacité de se remémorer des solutions ayant démontré leur efficacité et à transposer leur usage à de nouvelles situations (tel que présenté dans l'exemple de la représentation d'un trafic de marchandises).

L i m i t e s i m p o s é e s p a r l e s o u t i l s

Les outils exploités pour produire les visualisations limitent également les choix. La graphique offre théoriquement un éventail de possibilités uniquement limité par l'imagination. Cependant, l'avènement de l'informatique bien qu'ayant facilité l'automatisation de la production des représentations a également conduit à un appauvrissement des possibilités qui sont maintenant délimitées par les fonctionnalités proposées par les outils. Les informations sont intégrées dans un nombre restreint de structures visuelles (par exemple, une icône, un texte, un cadre, un lien), alors que les structures visuelles devraient raisonnablement être adaptées en regard de la nature des informations à représenter. Par ailleurs, afin de pouvoir gérer l'évolution des objectifs d'analyse au cours d'une enquête et la diversité des

²⁵ Cf. Recommandations et standards, page 45

concepts à représenter, les outils devraient offrir des possibilités d'adaptation dynamique des formes de représentation (notamment au niveau de la structure des graphes, des possibilités de sélection et de faire varier les niveaux d'agrégation). Les outils actuellement exploités à large échelle semblent s'être imposés en raison de leurs simplicités d'utilisation. Toutefois le traitement de l'information et les formes de raisonnements mis en œuvre pour résoudre une affaire criminelle n'en restent pas moins complexes. Les outils devraient être développés en regard des raisonnements impliqués et permettre une souplesse de représentation nécessaire lors de l'analyse d'un cas.

C o n s o l i d a t i o n d e s m é t h o d e s

D é l i m i t a t i o n d u c a d r e

Quatre dimensions d'analyse ont été définies afin de classifier les méthodes de visualisation en regard des objectifs d'analyse : les dimensions relationnelle, temporelle, spatiale et quantitative. Chaque dimension est constitutive d'une forme d'analyse et des représentations propres sont exploitées pour chacune d'elles. Cette thèse se focalise sur les représentations de l'analyse relationnelle.

De plus, l'ensemble des représentations graphiques traitées dans cette recherche sont produites dans un espace à deux dimensions. Il faut relever que de nombreuses recherches ont été effectuées en matière de visualisation à trois dimensions. L'annexe trois résume les études ayant évalué et décrit les avantages et faiblesses de la visualisation à trois dimensions. Les raisons liées au contexte d'exploitation de la visualisation en analyse criminelle qui ont motivé le choix de focaliser cette étude sur les représentations à deux dimensions sont également explicitées en annexe. Par ailleurs, l'intégration de la dimension temporelle par l'usage de représentations dynamiques n'est pas étudiée en détails. Les limites des outils exploités pour produire des graphes sont toutefois discutées (notamment en regard de la nécessité de pouvoir dynamiquement modifier la structure d'un graphe).

O b j e c t i f s

1. Identifier des éléments de méthode susceptibles d'aboutir à la réalisation de schémas efficaces et les structurer dans une démarche décrivant leur processus de conception.
2. Trouver dans la pratique des situations récurrentes qui peuvent faire appel à des visualisations typiques.
3. Initier une démarche basée sur l'élaboration d'un dictionnaire de « patterns » qui décrivent des solutions typiques pour les situations récurrentes identifiées.
4. Evaluer l'impact sur les outils des éléments de méthodes identifiés.

M é t h o d o l o g i e

Les méthodes et les recommandations générales provenant de travaux sur la visualisation en tant que domaine indépendant de ces champs d'application (par exemple les travaux de Bertin, Tufte, Ware, etc.) ont été recherchées et intégrées dans une ébauche de démarche décrivant le processus de conception d'un schéma relationnel en analyse criminelle. Les résultats de la recherche sont volontairement présentés du général au particulier afin d'illustrer la manière dont les observations générales s'intègrent dans les bonnes pratiques et de quelle manière les outils devraient intégrer les éléments de méthode proposés. La recherche a cependant également été effectuée de manière empirique à partir des situations récurrentes étudiées. Des similitudes ont été recherchées au niveau du processus de conception d'un graphe et sur les formes de représentation elles-mêmes.

Il faut d'ailleurs relever que cette recherche a été initiée suite au traitement du cas présenté comme exemple de représentation d'un trafic de marchandises²⁶. En effet, le changement de structure opéré sur le graphe a été effectué en relevant une analogie avec la représentation de données d'appels téléphoniques. La solution retenue ayant montrée une certaine efficacité à l'usage, il a été décidé de répertorier des cas d'utilisation de la visualisation, afin d'initier le développement d'un catalogue de situations récurrentes et leurs représentations propres. Cette approche vise la formalisation de bonnes pratiques liées à l'usage de représentation des informations régulièrement traitées dans l'enquête judiciaire (telles que les données de téléphonie, les données financières et les traces physiques) en fonction d'objectifs spécifiques. Des formes de visualisations adaptées pour chaque situation sont ainsi décrites. L'idée ici n'est pas formaliser des « meilleures pratiques », mais des pratiques ayant prouvé à l'usage une certaine efficacité.

²⁶ Cf. Représentation d'un trafic de marchandises, page 75

Finally, during the study of cases and the realization of relational schemas, limits imposed by the tools and technical difficulties have been identified. The method described in the first chapter and the examples presented in the second rely notably on the possibilities offered by the tools. Although the tools should integrate and support the analysis process, they sometimes create ruptures in the reasoning process and impose constraints linked to the functionalities offered. The impact of formalized methods on the tools is therefore discussed. A visualization prototype integrating several dimensions of analysis has also been elaborated.

The theories and recommendations from the work on visualization have been integrated at the level of the method, good practices and the discussion on the tools.

**PARTIE II : CONTRIBUTION AUX
MÉTHODES DE CONCEPTION DES
SCHÉMAS RELATIONNELS**

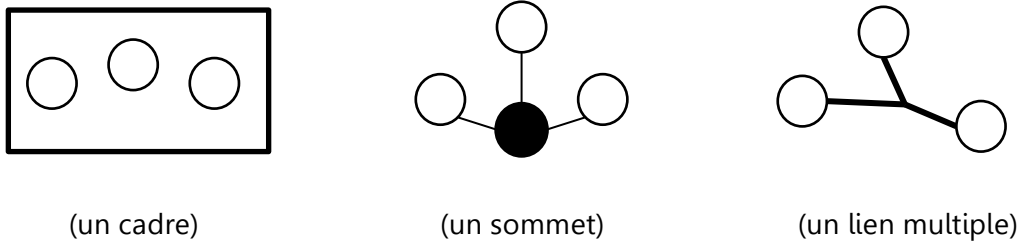
Processus de conception d'un schéma relationnel

Toute communication graphique se base sur l'utilisation de conventions entre le rédacteur et le lecteur. Elles déterminent la manière dont les messages sont produits et interprétés. En matière de représentations graphiques, les conventions définissent notamment le lien entre les composants visuels et leurs significations propres (Mackinlay, 1986). La signification du signe est définie à l'avance dans un système monosémique²⁷. Par exemple, une flèche peut représenter le signifié « direction » ou sur un graphe « relation dirigée » :

Un système de signe est donc défini par un ensemble de conventions. Toutefois, lors de la production d'un graphe, le signe flèche ●→● est généralement exploité pour représenter une autre information que « direction » ou « flèche » ou « relation dirigée ». Par exemple, les schémas de flux élaborés en analyse criminelle exploitent ce signe pour décrire une communication téléphonique, une transaction bancaire, voire une vente. Un signe adapté doit être choisi pour chaque information à représenter. Le processus n'est cependant pas direct ; une étape de raisonnement intermédiaire est nécessaire. Pour chaque information, un signifié (pour lequel un signifiant est défini) est choisi: communication téléphonique (relation dirigée), vente (relation dirigée), amis (relation non dirigée), etc.

Pour un système monosémique donné, le passage du signifié au signifiant (le signe visible) est défini par le langage visuel. En effet, si une vente est identifiée comme une relation dirigée, l'utilisation de la flèche comme signifiant est défini par le système de signe utilisé. Toutefois, différents systèmes de signe peuvent être exploités. Par exemple, une entreprise (information) peut être considérée comme un groupe (signifié) qui peut être représenté de diverses manières (signifiants):

²⁷ Cf. Méthodes d'évaluation de visualisations, page 59



Le processus de création d'un schéma relationnel peut donc se modéliser ainsi :

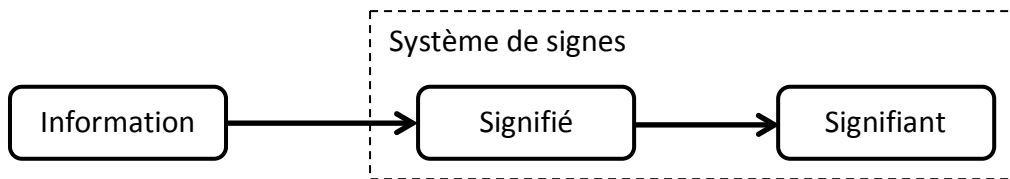


Figure 7 Processus de conception d'un graphe

Ce formalisme ne décrit pas le processus général de production d'une représentation graphique. Il permet toutefois de décomposer les étapes de raisonnements effectuées pour produire un schéma et distinguer des causes possibles expliquant certaines des difficultés observées (la diversité des manières de représenter la même information, les ambiguïtés et les erreurs de lecture) :

1. Le système de signes doit permettre de représenter l'ensemble des types d'informations à analyser.

L'une des lacunes identifiées en analyse criminelle est l'absence d'une définition formelle du langage visuel exploité sur les graphes. Une syntaxe uniforme et correcte est nécessaire afin de produire un graphe préservant la sémantique des concepts et des relations représentés (Reed & Kocura, 2005).

La démarche entreprise dans cette recherche, n'ambitionne pas de concevoir un langage formel complet. La grande diversité des situations rencontrées lors des enquêtes et la nature évolutive des problèmes criminels ou de sécurité remettent en question la possibilité de définir un langage exhaustif. L'identification de certains types d'informations génériques régulièrement investigués dans les enquêtes (p. ex. des transactions, des événements et des groupes de personnes ou de délits) restent néanmoins possible et utile pour tendre vers une formalisation plus précise du langage. De la littérature il ressort que des types d'entités sont

généralement définis et forment le domaine du discours²⁸, mais rarement des types de relations. La nature de ces relations, le degré de confiance et la fiabilité de ces relations ne sont que rarement définis et décrit par le langage (Amar & Stasko, 2005).

Hypothèse : une classification des types de relations à mettre en évidence entre les objets à analyser est nécessaire.

La démarche proposée consiste à identifier des types de relations génériques et décrire leurs propriétés. Puis de proposer et évaluer une manière efficace de les représenter. Comme en témoigne les évaluations préliminaires effectuées, choisir une manière efficace de représenter les relations semble être une tâche importante du processus.

2. Le système de signe doit être clairement défini. Pour chaque signifiant, le signifié doit être préalablement défini afin de pouvoir être interpréter et éviter les ambiguïtés.

Hypothèse : le langage visuel peut être défini en regard des types de relation identifiés.

Dans la mesure où des types de relations génériques peuvent être formalisés, le langage visuel devrait être défini en regard de cette typologie. Pour chaque type, un signe est défini afin de constituer un langage commun monosémique. La signification des signes ainsi clairement définie, le raisonnement nécessaire pour produire un graphe est simplifié. Si le langage définit par exemple le cadre comme signe pour représenter un groupe, choisir un signe adapté pour représenter une entreprise revient à déterminer si celle-ci est un groupe. Les types de relations doivent donc être clairement définis et distinguables.

3. Les composants graphiques et les propriétés visuelles exploitables pour représenter les informations sont multiples. Il convient de définir pour chaque signifié quel est le signifiant le plus adapté.

Hypothèse : pour chaque type de relation, un signifiant adapté peut être défini en fonction des critères d'intégrité et d'efficacité.

²⁸ Cf. Analyse et visualisation relationnelle, page 38

L'étape finale du processus consiste à définir un signifiant adapté pour chaque type de relation. Choisir une manière efficace et intégrée de représenter les relations est certainement l'une des étapes clé du processus pour assurer la qualité de la représentation (Senator, 2005). Le choix peut être intuitif, mais il peut aussi se baser sur une évaluation de l'efficacité relative de chaque signifiant selon des critères objectifs²⁹.

Afin de présenter les recherches et les évaluations effectuées pour tester les hypothèses présentées ci-dessus, ce chapitre est structuré en trois parties. Premièrement, le langage visuel exploité pour représenter des informations sur un graphe est décrit. Les propriétés visuelles utilisables pour représenter des relations sont présentées et une typologie de relations est proposée. Afin de faciliter la description de cette classification, les définitions sémantiques des types de relations sont présentées avec leur représentation propre. Une évaluation empirique de l'efficacité de certaines des formes de représentation proposées a été entreprise. Les résultats sont présentés et discutés. Finalement la dernière partie regroupe les éléments de méthode formalisés. Ils sont présentés en suivant le processus de conception d'un graphe.

D e s c r i p t i o n d u l a n g a g e v i s u e l

P r o p r i é t é s d u p l a n

Les composants du langage visuel s'intègrent dans le plan selon des structures visuelles qui peuvent en elles-mêmes posséder une signification. Selon la théorie de la Gestalt (théorie de la forme, du pattern), une structure perçue par un positionnement non aléatoire des composants sur le plan, induit un message souvent implicite et automatique (Card et al., 1999) (Chang et al., 2002) (Ware, 2004).

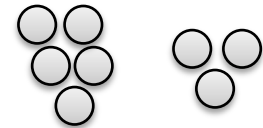
Certains principes de la Gestalt sont très connus : tel le postulat que le tout est différent de la somme de ses parties et celui que le tout est perçu avant les parties qui le composent. Ce dernier principe semble être causé par la morphologie de l'œil humain. Nous percevons les informations selon deux mécanismes distincts (liés à la différence entre les récepteurs de la rétine et de la fovéa). Premièrement, l'aspect global est perçu, puis nous avons la capacité de focaliser notre attention sur un élément plus précis. La perception globale et immédiate, alors que la focalisation est contrôlée. Une visualisation est donc perçue d'en un premier

²⁹ Cf. Critères d'évaluation, page 61

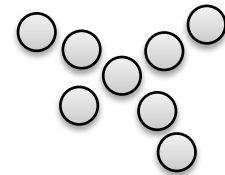
temps dans sa globalité, puis ces composants sont lus en détails (Ware, 2004). De plus, une partie d'un tout, n'est pas perçue de la même manière, si elle est représentée seule ou dans un ensemble. Le contexte global, notamment la mise en place sur le plan des composants, est un aspect important du langage visuel. En effet, les positions relatives des entités sur le plan permettent d'exprimer des relations (par exemple, de distinguer des groupes de personnes ou de décrire une chronologie d'événements).

La Gestalt décrit des mécanismes de perception affectant l'interprétation sémantique des formes. De nombreuses « lois » ont été définies. Seules celles ayant un impact sur la représentation par un graphe des informations sont résumées ci-dessous. Elles découlent du principe général de prégnance : un ensemble de composants tend à être perçu comme une forme globale régulière, symétrique et stable. Une représentation est facile à comprendre si sa structure est simple.

Loi de proximité : des composants placés à proximité, ont tendance à être perçus comme un groupe.



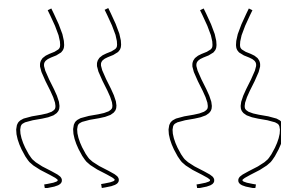
Loi de continuité : des composants sont perçus comme un groupe lorsqu'ils semblent former un prolongement.



Loi de fermeture : des composants distincts peuvent être perçus comme une forme régulière fermée. Le cerveau a tendance à grouper des composants s'arrangeant en une forme ouverte et remplit les vides.



Loi de la symétrie : des composants symétriques tendent à être perçus comme un ensemble.



Ces lois explicitent comment les positions relatives des composants génèrent des structures. Celles-ci peuvent être exploitées afin de représenter des groupes d'entités, voire des relations. En matière d'analyse criminelle, la position sur le plan des composants est parfois exploitée pour représenter une certaine proximité. Par exemple lors de la représentation d'un réseau, la proximité de deux individus sur un graphe est représentative de leur proximité dans le réseau. Une position centrale sur le graphique peut également exprimer une position centrale dans le réseau (Sparrow, 1991). Ce dernier point est toutefois sujet à interprétation.

En effet, la centralité d'une personne sur un graphe peut également être liée au fait que le nombre de données récoltées sur cette personne est simplement plus grand (notamment lorsque l'enquête a démarrée à partir d'elle). De manière générale, les entités positionnées au centre d'un schéma sont perçues plus rapidement que celles placées en périphérie (Eppler & Aeschmann, 2009). Il convient donc de positionner les informations de façon adaptée pour ne pas mettre en valeur involontairement des entités positionnées au centre d'un schéma.

Le positionnement des entités sur le plan peut également influencer l'interprétation des relations et la détection de sous-groupes indépendamment de la présence réelle de relations tracées par des traits entre les entités (McGrath et al., 1997). Des entités positionnées à proximités peuvent être considérée comme liée uniquement sur la base de leur placement. En effet, la position sur le plan des composants constitue l'une des variables visuelles la plus efficace pour encoder des relations entre les entités (Mackinlay, 1986). Par exemple une relation temporelle peut être représentée par un alignement des composants sur l'axe horizontal en fonction du temps³⁰.

C o m p o s a n t s d ' u n g r a p h e

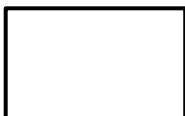
Selon Bertin (2005) toute représentation graphique est constituée de trois implantations fondamentales : *le point, la ligne et la zone*. La tâche qui rend visible un point peut varier de taille, valeur, grain, couleur, orientation et forme. En matière de visualisation relationnelle par un graphe, les composants élémentaires du langage visuel rejoignent les implantations définies par Bertin :



Le **nœud** (ou sommet) qui représente un point du plan. Le nœud est généralement exploité pour représenter une entité.



L'**arête** (ou arc) qui représente une ligne entre des nœuds. L'arête est généralement exploitée pour représenter une relation entre deux nœuds.

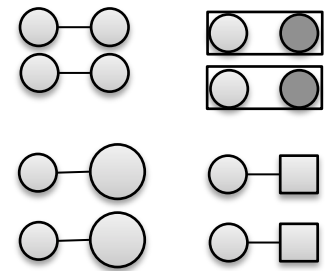


Le **cadre** qui représente une zone délimitée par un trait englobant des nœuds. Un cadre est généralement exploité pour représenter un groupe : une relation entre un ensemble d'entité.

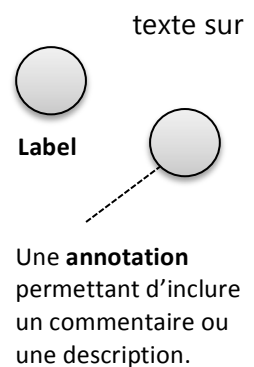
³⁰ Cf. Annexe 1: Dimensions d'analyse et visualisation - Analyse temporelle, page 239

L'une des lois de la Gestalt décrit l'utilisation combinée de ces signes (Palmer & Rock, 1994) (Ware, 2004):

Loi de connectivité : connecter des composants par un trait ou un cadre permet de représenter des relations entre des entités. La perception de la connectivité prévaut en principe sur la perception de la proximité ou de la similarité (présentée ci-dessous).



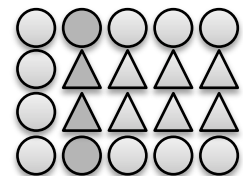
Des informations peuvent également être représentées sous la forme de un graphe. Le texte peut être inclus de deux manières: sous la forme d'un **label** (ou étiquette) directement sur l'un des composants graphiques et sous la forme d'une **annotation** (description, commentaire).



Attributs graphiques

La Gestalt définit également une loi indépendante du placement :

Loi de similarité : des composants similaires sont perçus comme des groupes. La similarité est produites par un usage commun d'un ou plusieurs attributs graphiques (forme, couleur, taille, etc.)



La loi de similarité est induite par un usage contrôlé des variables graphiques appelées variables retiennes par Bertin (2005) (p. ex. taille, valeur, grain, couleur, orientation et forme). L'efficacité de ces variables pour encoder une information a été largement étudiée dans la littérature depuis les travaux de Bertin (Mackinlay, 1986). En effet, certaines variables sont plus efficaces pour encoder des données quantitatives, nominatives ou ordinales (Bertin, 2005)³¹.

³¹ Une composante est nominale lorsque ses catégories ne s'ordonnent pas de façon universelle (par exemple, les zones géographiques qui sont notamment classables par ordre alphabétique, par superficies ou en fonction du nombre d'habitants), une composante est ordonnée lorsque ses catégories s'ordonnent d'une seule manière universelle (par exemple, le choix d'une cuisson de viande: bleu, saignant, à point, cuit, très cuit), finalement une composante est dite quantitative lorsqu'elle a comme objet de préciser une variation de distance entre deux catégories (le système métrique, le temps, etc.).

98 Processus de conception d'un schéma relationnel

Les tableaux ci-dessous résument les principales variables exploitées sur un graphe en regard du type d'information qu'elles permettent de représenter efficacement (inspiré de (Mackinlay, 1986), p. 125, (Ware, 2004), p. 214 et (Ware, 2008), p. 63-64):

A t t r i b u t s d e s n œ u d s

Variable graphique	Type de données	Description
L'icône ou la forme	Nominale	L'icône est généralement exploitée pour représenter le type d'entité (p. ex. une personne, un véhicule, etc.). L'icône permet de décrire le prédicat « est un ».
La couleur	Nominale ou ordinale	La couleur permet également de distinguer des types. La couleur peut être exploitée pour encoder des variables ordinales.
La taille	Quantitative	La taille de l'entité peut être exploitée pour représenter une variable quantitative (p. ex. le nombre d'entités liées).

A t t r i b u t s d e s a r ê t e s e t d e s c a d r e s

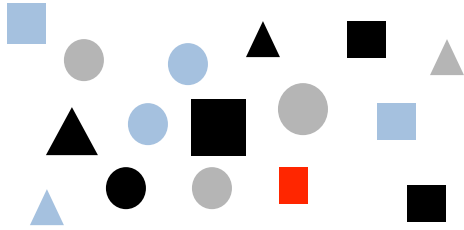
Variable graphique	Type de données	Description
La couleur	Nominale ou ordinale	La couleur permet de représenter des types de relation en fonction de leur nature : relations familiales, relations téléphoniques, financières, etc.
L'épaisseur	Quantitative	L'épaisseur des arêtes peut encoder une variable quantitative (comme le nombre de relations par exemple). Les variations d'épaisseur des liens sont plus facilement comparables que les variations de taille des nœuds (variation sur une seule dimension pour les arêtes).
Le style	Nominale ou ordinale	Le style de trait (continu, trait-tillé, etc.) est exploité pour décrire la fiabilité de la relation. La convention est que le trait-tillé décrit des relations incertaines et le trait continu des relations sûrs ³² . D'autres types de style peuvent également être exploités pour distinguer des types de relation.

Les variables rétinienne comme la couleur ou le style peuvent être exploitées pour distinguer des types de d'entités et de relations et pour exprimer la fiabilité des relations

³² Cf. Recommandations et standards, page 45

notamment. Elles doivent faire l'objet d'une convention entre le lecteur et le concepteur de la représentation (par exemple par le biais d'une légende). Les variables catégorielles sont donc généralement exploitées pour représenter les types d'entités et de relations. Elles servent à visualiser des classes, à décrire les relations « est un ... » et « est du type ... ».

Importance de la couleur



La couleur est la variable graphique la plus pré-attentive. En effet, une variation de couleur est perçue plus rapidement et facilement que les variations des autres propriétés graphiques (telles que la forme, la taille ou la texture). Il faut toutefois noter que la couleur peut généralement être

exploitée pour encoder entre six et douze catégories différentes. Au-delà, la perception des variations n'est plus pré-attentive (Ware, 2004) (Ware, 2008). La couleur rouge étant la plus pré-attentive, il est recommandé de l'utiliser pour mettre en avant le type d'informations le plus important pour la question d'analyse. Les travaux de recherches menées sur la couleur et son exploitation pour représenter l'information sont nombreux. Ils ne sont pas discutés en détails ici. Il faut cependant relever que l'utilisation d'une variation de couleur ou de contraste pour représenter une variable quantitative n'est pas un choix adapté. En effet, la perception peut être faussée en fonction des couleurs environnantes (Ware, 2008).



Les barres verticales sont de couleur identique

Typologie de relations

Les formes syntaxiques décrites dans cette section décrivent des manières de combiner et disposer les composants graphiques afin de former une structure visuelle associée à une catégorie particulière de relation.


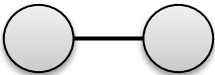
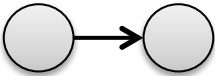
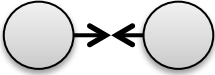
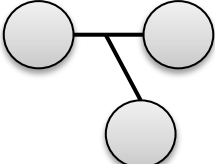
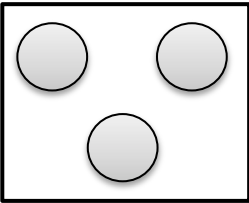

- *Le lien* : une relation entre deux entités. Elle représente le prédicat « est lié à ». Elle est définie par son type (par exemple, une relation familiale, professionnelle, etc). Par définition, le lien ne décrit pas une relation entre plusieurs entités (visuellement un lien est représenté par une arête, un trait entre deux entités).
- *Le lien dirigé* : une relation avec un sens entre une entité source et une entité cible. Par exemple, les relations de filiation ont un sens. Le rôle d'une entité dans la relation décrit la fonction qu'elle occupe. Par exemple « père » et « fils » sont des rôles d'une relation de filiation. D'autres types de relation ont également un sens : les relations de causalité et les relations définissant un ordre ou une séquence, telle que les relations chronologiques. Une relation peut donc être dirigée ou non.
- *Le lien bidirectionnel* : une relation qui se réalise dans les deux sens. Par exemple un échange de biens entre deux individus est une relation commerciale à double sens. Alors que la causalité est une relation directionnelle, l'influence par exemple peut être bidirectionnelle (deux personnes ou deux groupes d'individus peuvent s'influencer mutuellement).

Certaines relations existent entre plus de deux entités ; telle que l'appartenance à une famille qui est une relation entre plusieurs individus. Afin de distinguer les relations entre deux entités des relations entre plusieurs entités, les notions de groupe, d'ensemble et de lien multiple sont définies :

- *Le groupe* : un ensemble d'entités formant un tout. Il regroupe des entités partageant une ou plusieurs caractéristiques communes et il possède généralement un identifiant propre (le nom de famille par exemple). Un groupe décrit le prédicat « est composé de ». Deux types de groupe sont distinguables: homogène et hétérogène. Un groupe homogène décrit une relation entre des entités d'un même type (une famille qui lie des personnes, un lot qui lie plusieurs produits du même type). Un groupe hétérogène rassemble plusieurs entités dont les types peuvent varier. Afin de différencier ces deux types de groupes, le terme d'ensemble est proposé.
- *L'ensemble* : un groupe d'entités d'un même type. Des représentations graphiques spécifiques sont définies pour distinguer le groupe de l'ensemble.
- *Le lien multiple* : une relation entre plusieurs entités de différents types. Alors qu'un groupe possède un identifiant, la relation multiple n'en possède pas. Par exemple, lors de contrôle de véhicules, un lien multiple existe entre l'événement, le véhicule et la plaque. Cet exemple est présenté sous la forme d'un pattern dans le chapitre suivant. Visuellement, il n'est pas possible de représenter les relations entre les événements, les véhicules et les plaques par des liens simples sans engendrer de l'ambiguïté. Un signe spécifique est nécessaire.

Le tableau ci-contre résume l'ensemble des signes formalisés pour décrire ces relations. Ces catégories génériques ont été définies en regard des cas traités et analysés qui sont présentés dans le chapitre suivant. Le système de signe est ainsi clairement défini (pour chaque signifiant le signifié est décrit). Il est important de relever que le langage visuel défini est une extension du langage proposé par les logiciels actuellement utilisés dans la pratique. En effet, les signes décrivant le lien bidirectionnel, le lien multiple et l'ensemble ne sont pas disponibles dans l'outil principalement exploité en analyse criminel³³.

³³ Analyst's Notebook 8® de la société i2®, <http://i2.co.uk>, dernier accès le 12.01.2011

Signifiant	Signifié	Description	Exemples
	L'entité	Une chose possédant une existence distincte et identifiable	Un événement, une personne, une trace, un objet, etc.
	Le lien	Une relation entre deux entités ³⁴	Une relation matrimoniale, le rôle d'une personne pour un événement, etc.
	Le lien dirigé	Un lien dirigé est une relation contenant un sens, décrivant une séquence ou une relation de causalité	Une relation téléphonique entre deux numéros, une transaction entre des comptes, une relation père-fils, etc.
	Le lien bidirectionnel	Un lien bidirectionnel représente un échange ou un ensemble de relations bidirectionnelles	Un échange de biens entre des personnes, un ensemble de communications ou de transactions, etc.
	Le lien multiple	Un lien multiple représente une relation entre plus de deux entités	Une relation entre un événement, un véhicule et une plaque lors d'une contrôle, etc.
	Le groupe	Un cadre peut être utilisé pour représenter un groupe ou une caractéristique commune à plusieurs entités	Une société qui emploie des personnes, une zone géographique commune entre plusieurs événements, etc.
	L'ensemble	Un ensemble décrit un groupe d'entités du même type. Les accolades permettent de faire la distinction entre une entité unique et un ensemble	Un lot d'objets, une série de cambriolages, un groupe d'individus, etc.

³⁴ Un concept peut être représenté par un lien lorsqu'il est lié à deux autres entités (et uniquement deux). La possibilité de représenter un concept par un lien ou une entité est discutée et évaluée dans l'évaluation présentée dans la section suivante et dans les exemples présentés dans le chapitre suivant.

Les variables graphiques et le positionnement peut également être exploité pour représenter des relations :

Signifiant	Signifié	Description	Exemples
Icône, couleur, forme, etc.	Le type	Les variables rétinienne sont généralement exploitées pour décrire des types d'entités ou de relations définis selon des classes générales ou distinguer des concepts par leurs caractéristiques	Distinguer des événements, des personnes, des objets, etc. Distinguer des profils ADN: profil partiel, complet, de mélange, etc.
Positionnement	La proximité	Le positionnement sur le plan des composants peut être représentatif de la proximité des entités. Il peut également décrire une séquence.	Les rôles et la hiérarchie d'une organisation criminelle, etc. Une séquence temporelle

Ce langage n'est certainement pas exhaustif et pourrait être développé en fonction de nouveaux problèmes à représenter. La démarche proposée ici consiste à définir le langage en fonction de la nature des informations à représenter et non en fonction des possibilités offertes par les outils. La démarche tend vers une formalisation plus claire du langage afin de faciliter la conception des graphes et minimiser les risques d'ambiguïté lors de son utilisation. Le niveau d'abstraction est volontairement général et simple afin de l'exploiter le plus facilement possible, sans contraindre et complexifier son utilisation par des définitions rigides. Il est mis en œuvre et explicité par les exemples présentés dans le chapitre suivant.

D i s c u s s i o n s u r l ' u t i l i s a t i o n d u l a n g a g e

Sélectionner un signe adapté pour représenter un concept (sous la forme d'une entité ou d'un type de relation) n'est pas trivial. Par exemple, le fait d'être mariés décrit une relation entre deux personnes. Pourtant, le mariage peut être considéré comme une entité (un événement). Une information peut, de plus, être représentée sous la forme d'un attribut d'une entité ou comme une entité liée. La localisation d'un événement peut, par exemple, être représentée pour chaque cas en tant qu'attribut (visuellement représenté sous la forme

d'un texte, ou d'une icône par exemple). Les événements survenus dans un même lieu seront alors visuellement détectés selon le principe de similarité. Il est toutefois également possible de représenter les lieux par des entités distinctes reliées aux événements. Dans ce cas les relations de localisation sont perçues selon le principe de connectivité qui prévaut sur la perception de la similarité. Selon Chen (1976) et (Barthélemy et al., 2005), il s'agit en fait de choix guidés par l'objectif d'analyse qui définit l'importance des paramètres et des relations à représenter.

Une approche générale consiste à décomposer l'ensemble des informations en entités élémentaires, puis à représenter chacune d'elle par un type de nœud particulier. Les relations sont ensuite tracées entre les concepts par des arêtes. Cependant, une telle démarche n'est pas toujours adaptée³⁵. En effet, des simplifications et des améliorations doivent être réalisées pour exploiter pleinement les avantages de la représentation. De tels changements de structure graphique peuvent être des simplifications, telles qu'en agrégeant les données. C'est le cas notamment lors de la représentation d'un ensemble de communications téléphoniques ou la totalité des appels peuvent être regroupés sous la forme d'une relation bidirectionnelle dont l'épaisseur est proportionnelle au nombre de communications. Un changement de structure visuelle peut néanmoins également améliorer la lisibilité sans être au détriment de la complétude (p. ex. dans le cas de ventes). L'impact de ces choix a été évalué pour montrer qu'il existe des différences de perception et d'efficacité entre les diverses formes de représentation des relations.

E v a l u a t i o n d e l a r e p r é s e n t a t i o n d e r e l a t i o n s

Les choix effectués pour représenter les relations entre les entités d'intérêts pour une question d'analyse peuvent avoir un impact important sur la lisibilité et l'utilisabilité des représentations. En effet, une représentation perçue comme complexe ou difficile à lire ne remplit certainement pas son rôle de soutien aux processus d'analyse et de communication des informations. L'objectif de cette évaluation est de comparer par des indicateurs quantitatifs objectifs, plusieurs structures de graphe établies à l'aide de la typologie de relations proposée.

³⁵ Par exemple, où une vente représentée par une entité engendre une structure inadaptée rendant le schéma difficile à lire (cf. Représentation d'un trafic de marchandises, page 75).

M é t h o d e

C o n d i t i o n s d ' e x p é r i m e n t a t i o n

L'évaluation a été effectuée au moyen d'un sondage en ligne. Pour chaque question, les participants répondent en cliquant directement sur les graphes. Le questionnaire est envoyé par courriel aux participants qui sont libres d'y répondre au moment de leur choix. Il est demandé aux participants de remplir le questionnaire sans interruption. Afin de tester que chaque participant a bien compris le fonctionnement du questionnaire, une question préliminaire est posée (cliquer sur des sommets de couleurs en fonction de leurs liens avec d'autres sommets).

Pour chaque question, deux représentations différentes sont testées. Le graphe affiché pour chaque participant est sélectionné aléatoirement. Les participants ne sont pas au courant que les comparaisons sont effectuées entre deux schémas pour chaque question et non entre les questions. L'ordre des questions est identique pour tous les participants. La question et les données représentées sont identiques pour chaque paire de schéma. La méthodologie d'évaluation a été développée pour comparer les différences de perception entre deux manières de représenter les mêmes informations. Les schémas exploités sont inspirés ou directement issus de cas réels. Les données ont été simplifiées et rendues anonymes.

Un pré-test a été effectué avec une vingtaine de participants. Lors de celui-ci, il a été observé que répondre à certaines questions d'analyse réelles sur une représentation de l'affaire nécessite un temps important (entre dix et quinze minutes environ). En effet, les participants doivent comprendre le problème posé, ainsi que la nature des entités et des relations représentés. C'est pourquoi certaines questions ont été simplifiées en remplaçant les types d'entités par des formes et des couleurs.

P r o p r i é t é s v i s u e l l e s é v a l u é e s

L'évaluation porte sur la comparaison de plusieurs structures de graphe établies à l'aide de la typologie de relations présentées précédemment³⁶. Pour l'ensemble des comparaisons, les positions des entités, leurs formes, couleurs et tailles sont invariants. Seule la structure des graphes est modifiée. Les formes de représentations suivantes sont évaluées: le lien directionnel (questions 1 et 2), le lien (questions 3, 4 et 10), le lien bidirectionnel (questions 5 à 7), le groupe : questions 8 et 9.

³⁶ Cf. Typologie de relations, page 102

T â c h e s c o g n i t i v e s t e s t é e s

Pour chaque type de relation testé au moins deux questions différentes ont été posées, afin de comparer les résultats pour des questions différentes. Les questions ont été choisies en fonction des exemples et recommandations définies dans la littérature³⁷.

- Question 1 et 5 : trouver les chemins possibles entre deux sommets déterminés (cliquer sur les sommets intermédiaires)
- Question 2, 6 et 10 : trouver un ou plusieurs sommet(s) en fonction des valeurs d'un attribut sur les liens
- Question 3 et 9 : trouver les sommets liés à un sommet déterminé
- Question 4 et 7, 8 : trouver le sommet avec le plus de liens directs.

V a r i a b l e s d é p e n d a n t e s

Pour chaque question, le temps de réponse total et les positions des clics souris ont été relevés. Les clics corrects ont été déterminés en fonction de leurs positions avec une marge d'erreur variable suivant les schémas (généralement de l'ordre de 10 à 20 pixels). Certaines questions (2, 4, 6, 7 et 8) portent sur l'identification d'un sommet particulier, dans ces cas, seule l'exactitude est mesurée. Pour l'ensemble des autres questions, plusieurs sommets doivent être identifiés, la complétude est donc prise en compte dans l'évaluation.

S c h é m a s d e r é f é r e n c e s

L'évaluation de l'efficacité est toujours relative³⁸. Le choix de la représentation de référence influence donc fortement les résultats de l'évaluation. Pour chaque comparaison une représentation des concepts décomposés en entités a été choisie comme référence. Par exemple, pour l'évaluation de l'efficacité de la représentation de relations dirigées, des ventes ont été représentées par des entités entre les personnes et sous la forme de liens dirigés. Par contre pour évaluer les liens bidirectionnels, la comparaison a été effectuée avec un graphe contenant des paires de liens directionnels. Cette forme de représentation étant celle principalement exploitée pour représenter notamment des appels téléphoniques ou des transactions financières.

³⁷ Cf. Procédures d'évaluation, page 64

³⁸ Cf. Procédures d'évaluation, page 64

D e s c r i p t i o n d e l ' é c h a n t i l l o n

Le sondage a été transmis par des listes de diffusion (nombre de participants = 243) :

- Employés ESC : 106 personnes, taux de participation de 47%
- Etudiants ESC : 448 personnes, taux de participation de 19%
- Employés de la police neuchâteloise : 507 personnes, taux de participation de 20%

Participants	N	%	Formation	N	%
Etudiants	84	35	Apprentissage, CFC	55	23
Policiers	101	42	Formation professionnelle (maîtrise, brevet)	46	19
Employés ESC	33	14	Maturité (gymnasiale ou professionnelle)	9	4
Les deux	17	7	Université, EPF, Haute école spécialisée	125	51
Autres	8	3	Autre diplôme	8	3
Sexe			Cours d'analyse criminelle		
Homme	131	54	Je n'ai jamais suivi de cours	114	47
Femme	112	46	J'ai déjà suivi un cours théorique	76	31
			J'ai déjà suivi un cours théorique et pratique	53	22
Age			Schémas		
16-25	88	36	Je ne sais pas ce qu'est un schéma relationnel	60	25
26-35	80	33	J'ai déjà vu un schéma relationnel	122	50
36-45	53	22	J'ai déjà fait un ou plusieurs schéma(s) relationnel(s)	61	25
46-55	15	6			
56-65	7	3			

L'influence de ces paramètres a été testée globalement sur l'ensemble du questionnaire en comparant le temps de réponse global et le nombre total de réponses correctes. Des différences significatives ont été observées (les résultats sont présentés dans l'annexe cinq) :

- le niveau de formation et l'âge ont une influence significative sur le nombre de réponses correctes et le temps de réponse total
- le fait d'avoir suivi un cours d'analyse criminelle et d'avoir déjà vu ou fait un schéma relationnel influence le nombre de réponses correctes.

Il faut noter que l'objectif de l'évaluation n'est pas de tester ces paramètres, mais de tester l'impact d'un changement de représentation. Les versions des schémas ont donc été attribuées aléatoirement aux participants. Aucune différence significative n'a été observée entre les distributions de ces paramètres pour chaque version des schémas (cet aspect a été testé un test non paramétrique). Ces paramètres n'expliquent donc pas les différences observées lors de cette évaluation.

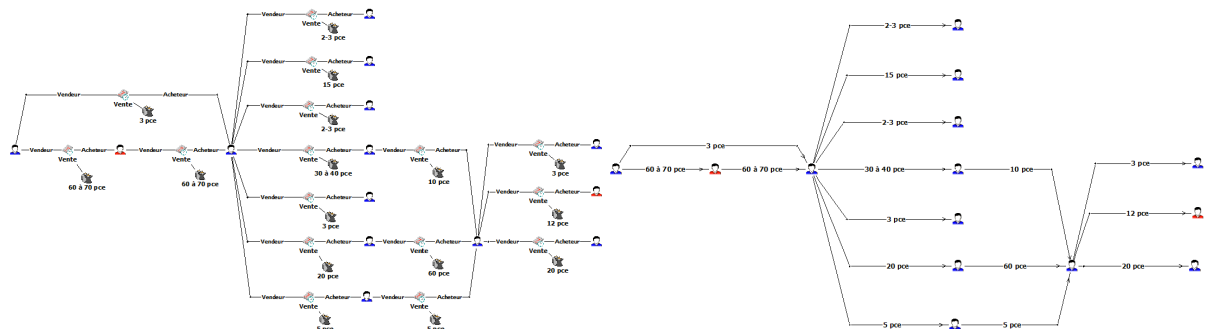
R é s u l t a t s

Les visualisations exploitées pour l'évaluation, le détail des résultats et des tests statistiques effectués sont présentés dans l'annexe cinq.

Représentation de liens dirigés



L'objectif ici est de comparer la représentation de ventes sous la forme d'une entité (liée aux personnes et à l'objet vendu) et sous la forme d'un lien directionnel. Cette évaluation a été effectuée sur la base du cas réel présenté dans la première partie³⁹.



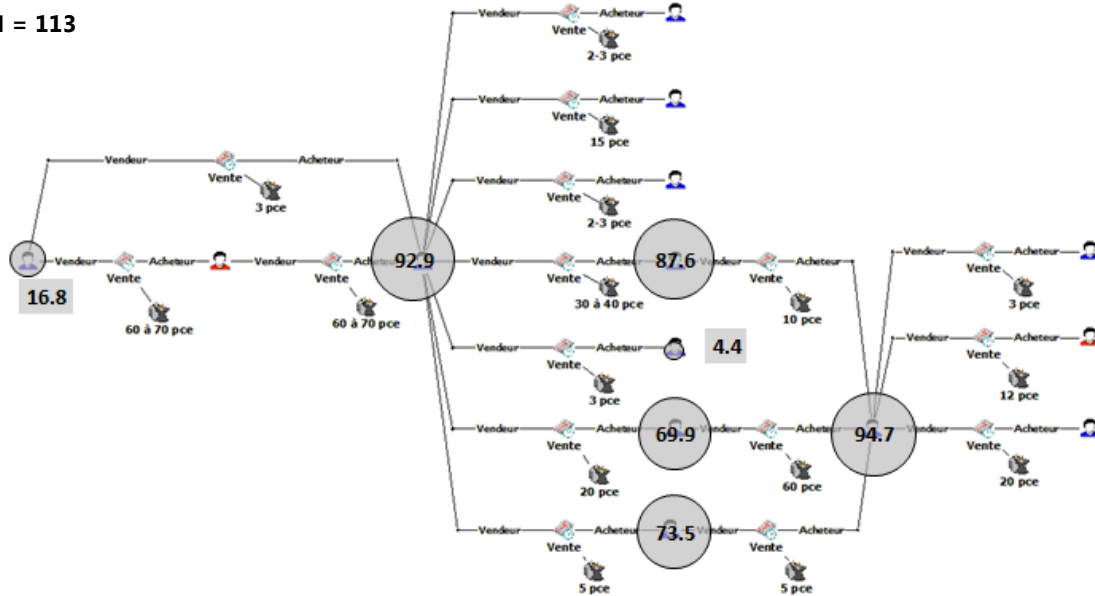
Question 1 : Cliquer sur les personnes qui peuvent avoir été des intermédiaires **entre les deux personnes en rouge**.

Question 2 : Cliquer sur la personne qui vous semble avoir vendu bien plus qu'elle n'a acheté.

Pour la question une, une augmentation de 14.8% de réponses correctes et complètes est observée avec la structure 2 ($F(1,239) = 5.08, p = 0.025$). Aucune différence significative, n'a été observée sur le temps de réponse entre les deux structures. Les schémas ci-dessous contiennent les positions des clics effectués par les participants (les chiffres affichés correspondent au pourcentage de participants).

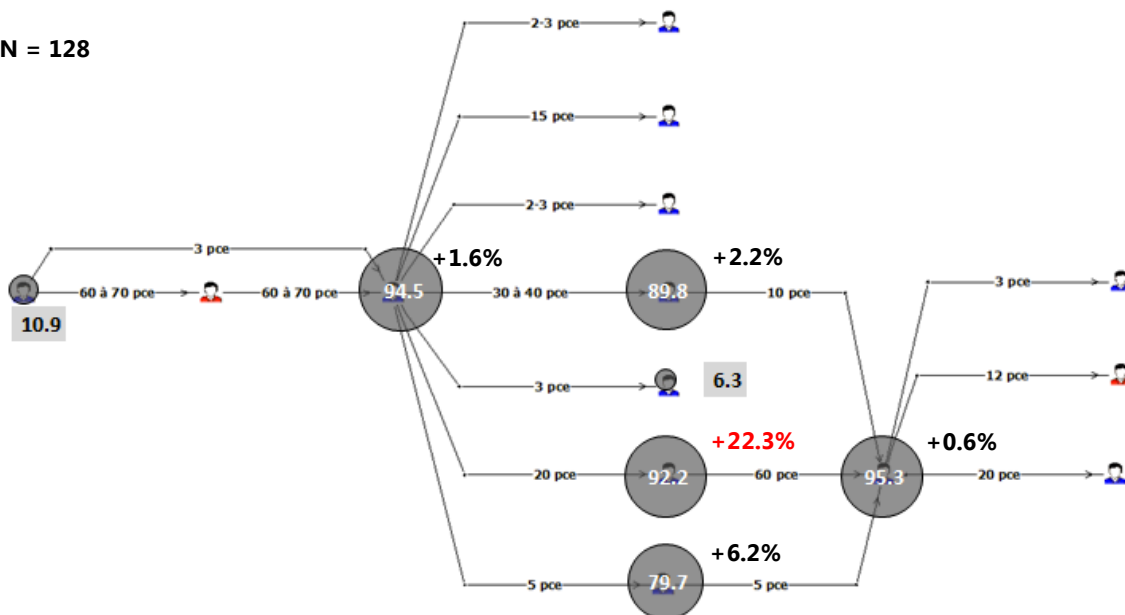
³⁹ Cf. Représentation d'un trafic de marchandises, page 75

N = 113



Les deux intermédiaires du centre en bas, ont été sélectionnés par moins de participants (une diminution d'environ vingt pourcents). Certains participants peuvent avoir mal compris la question et n'avoir sélectionnés qu'une séquence d'intermédiaires possible. Cet effet est par contre beaucoup moins marqué pour la seconde structure. Seul l'intermédiaire du bas a été sélectionné moins que les autres. Les schémas ayant été transmis aléatoirement aux participants cette observation tend à réfuter l'hypothèse d'une mauvaise compréhension. La différence pourrait être issue de la surcharge d'information (nombre d'entités) sur le premier graphe. Le positionnement des entités semblent néanmoins avoir un impact. Il est également intéressant de relever que les personnes directement liées aux personnes représentées en rouges sont plus souvent identifiées (sur les deux schémas).

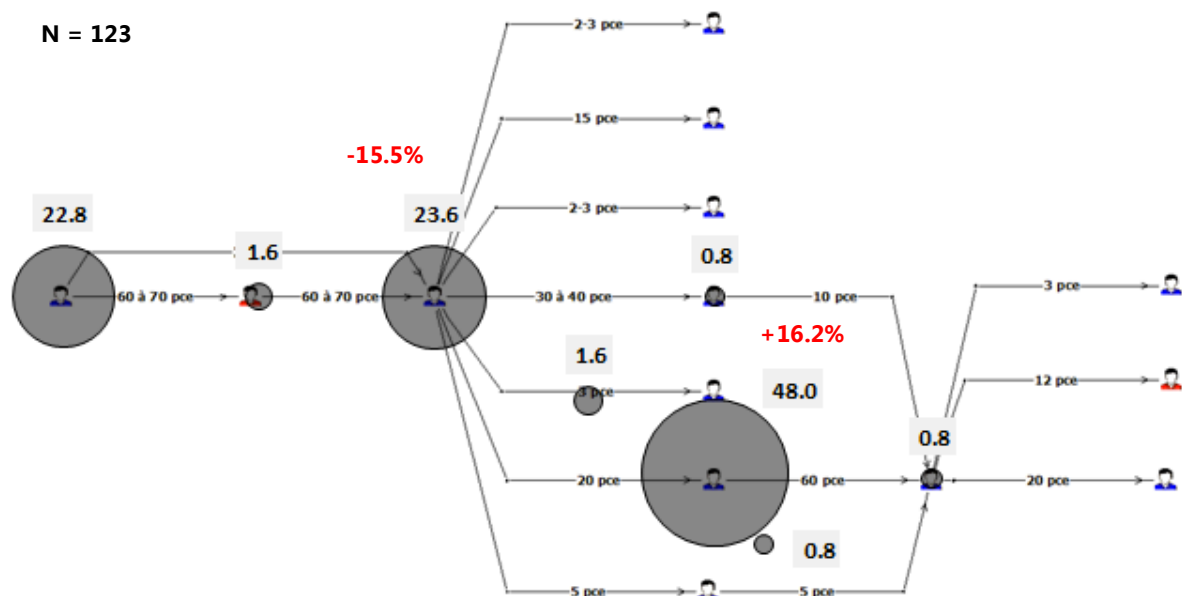
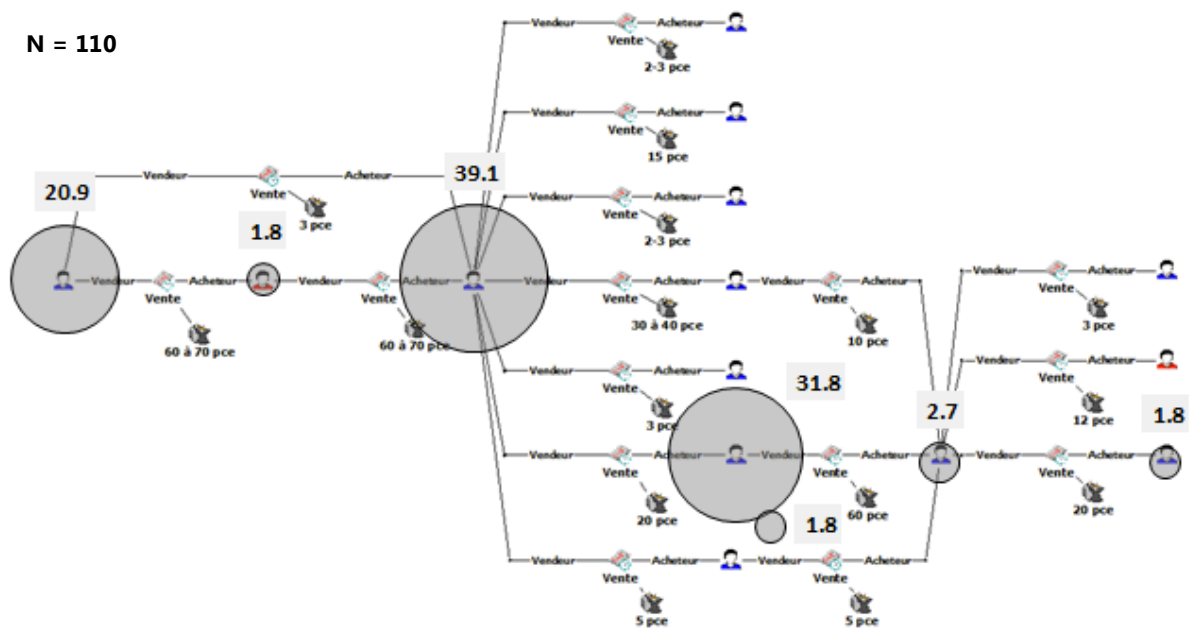
N = 128



110 Processus de conception d'un schéma relationnel

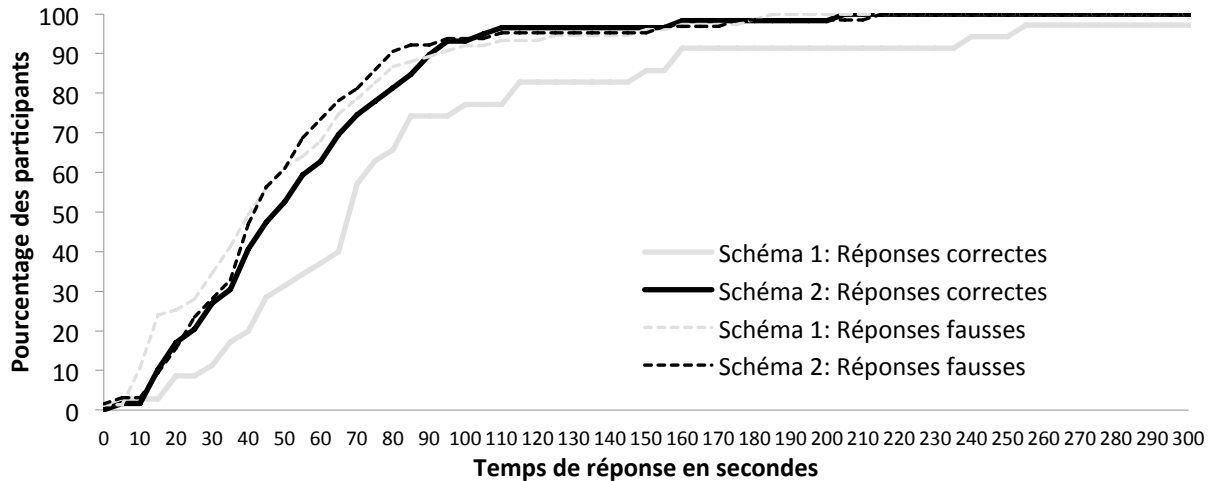
Globalement, cette structure est donc plus efficace pour répondre à la question. La position des entités semble être une variable importante au niveau de la perception et pourrait expliquer les différences au sein d'un même schéma.

Pour la question deux, une augmentation de 16.2% de réponses correctes est observée avec la structure 2 ($F(1,231) = 6.412, p = 0.012$). Les schémas ci-dessous contiennent les positions des clics effectués par les participants (les chiffres affichés correspondent au pourcentage de participants).



La réponse correcte est l'individu de droite qui a été sélectionné par 48% des participants pour la structure deux (augmentation du taux de réponse correcte de 16.2%). Pour les deux schémas, environ 20% des participants ont sélectionné l'individu de gauche qui selon les informations représentées n'a pas acheté de marchandise (dans le cadre de l'affaire, il s'agit en fait de l'individu ayant commis les vols). Le nombre d'erreurs est par contre significativement diminué pour l'individu du centre. En effet, bien qu'il a le plus de clients, le nombre total de marchandises vendues et environ égal au nombre acheté (respectivement entre 77 et 89 vendues, contre 60 à 70 achetées).

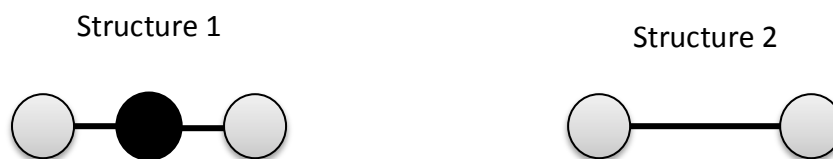
Le temps de réponse moyen pour les réponses correctes est plus court de 28.2% pour la structure 2 ($F(1,92) = 8.947, p = 0.0036$). Il passe de 89 +/- 67 secondes à 58 +/- 35 secondes. Pour les réponses incorrectes le temps de réponse est similaire entre les deux schémas. Pour le premier schéma, le temps de réponse pour les réponses incorrectes est 40.4% plus court que pour les réponses correctes ($F(1,108) = 12.731, p = 0.00054$). Il correspond au temps de réponse observé pour le deuxième schéma. Alors que pour la deuxième structure, les temps de réponses ne sont pas significativement différents entre les réponses correctes et fausses.



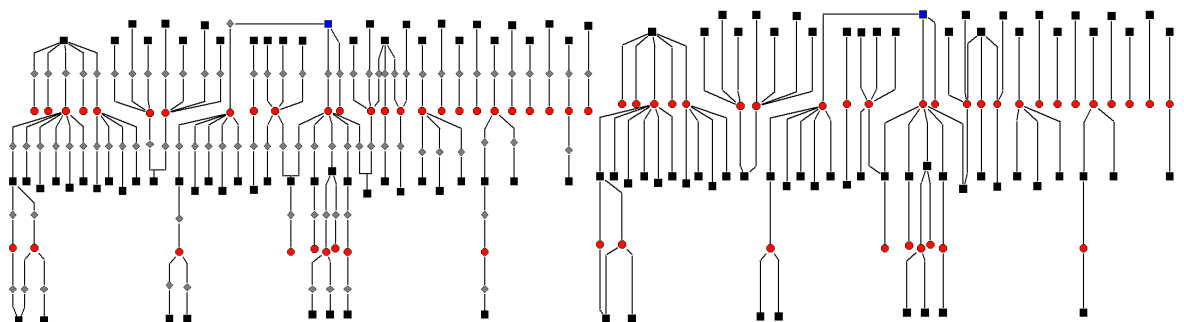
Il semble donc que la comparaison des quantités de marchandises vendues et achetées est plus difficile avec la première structure. L'effort de lecture nécessaire pour compter et comparer les proportions semble plus important avec la première structure, en regard du temps de réponse.

112 Processus de conception d'un schéma relationnel

Représentation d'une entité sous la forme d'un lien



L'objectif est de comparer la représentation d'un concept sous la forme d'une entité ou sous la forme d'un lien. Par exemple, dans le cadre de la représentation des relations entre des saisies de stupéfiants⁴⁰. Les échantillons peuvent être soit représentés comme une entité (liée à une entité « saisie » et une entité « classe chimique ») ou comme un lien (entre les deux mêmes entités). L'exemple utilisé pour les questions trois et quatre est tiré d'une affaire réelle. Les entités représentant les saisies ont été remplacées par des carrés noirs, les échantillons par des losanges gris et les classes chimiques par des ronds rouges.



Question 3 : Cliquer sur tous **les ronds rouges** liés directement et indirectement au carré bleu.

Question 4 : Cliquer sur le **rond rouge** ayant le plus de liens directs avec des carrés noirs.

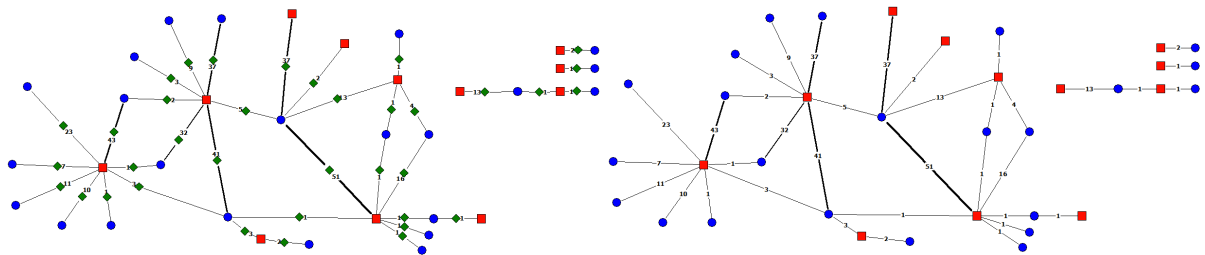
Pour la question trois, aucune différence de signification n'a été observée au niveau de l'exactitude, de la complétude et du temps de réponse.

Pour la question quatre, une augmentation de 8.9% de réponses correctes est observée avec la structure 2 ($F(1,226) = 2.973, p = 0.086$). Le temps de réponse moyen pour les réponses correctes est également plus court de 27.3% avec la structure 2 ($F(1, 183) = 8.739, p = 0.0035$). Il passe de 69 +/- 51 secondes à 50 +/- 31 secondes. Pour la question quatre,

⁴⁰ Cf. Saisies de stupéfiant et classes chimiques, page 155

l'exactitude est donc augmentée et le temps de réponse diminué lorsque la structure 2 est présentée. La structure deux est donc plus efficace que la structure 1 pour cette question.

Ce changement de structure a également été testé lors de la dernière question du test avec un autre exemple. Les relations détectées sur la base de l'analyse chimique des encres de photocopieurs sont représentées. Trois types d'entités sont représentées : les photocopieurs (losanges verts, structure 1), les marques des photocopieurs (carrés rouges) et les groupes définis sur la base de l'analyse des encres (ronds bleus). Un photocopieur ne pouvant être lié qu'à une seule marque et à un seul groupe, il peut être représenté par un lien (structure 2). Sur les schémas, l'ensemble des photocopieurs liés à une même marque et à un même groupe ont été regroupés. Le nombre de photocopieurs est représenté sous la forme d'un texte et par l'épaisseur des relations.



Question 10 : Cliquer [sur les 3 groupes](#) liés avec le plus grand nombre de photocopieurs.

Cette question a été estimée plus difficile par les participants (information transmise à la fin du questionnaire). En effet, la question porte sur l'identification de sommets liés aux photocopieurs. Ceux-ci sont représentés, sur le premier graphique, par des losanges et par des liens sur le deuxième schéma. Il n'était donc pas possible de formuler une question identique pour les deux schémas en se basant uniquement sur la représentation (par exemple, en demandant de cliquer sur les ronds bleus liés au plus grand nombre de losanges verts). Les participants ont donc dû effectuer tout le processus décrit précédemment : comprendre la problématique et identifier les types d'entité et de relations, puis comprendre l'implantation visuelle de chaque concept. Le temps de réponse est donc globalement plus long.

Aucune différence significative n'a été observée sur le taux de réponses complètes et correctes, par contre le taux de réponses fausses diminue légèrement : moyenne de 0.75 réponses fausses pour la structure 1 et de 0.47 pour la structure 2 ($F(1,192) = 4.132, p = 0.043$). Le temps de réponse passe de 136 +/- 83 secondes à 114 +/- 64 secondes avec la

114 Processus de conception d'un schéma relationnel

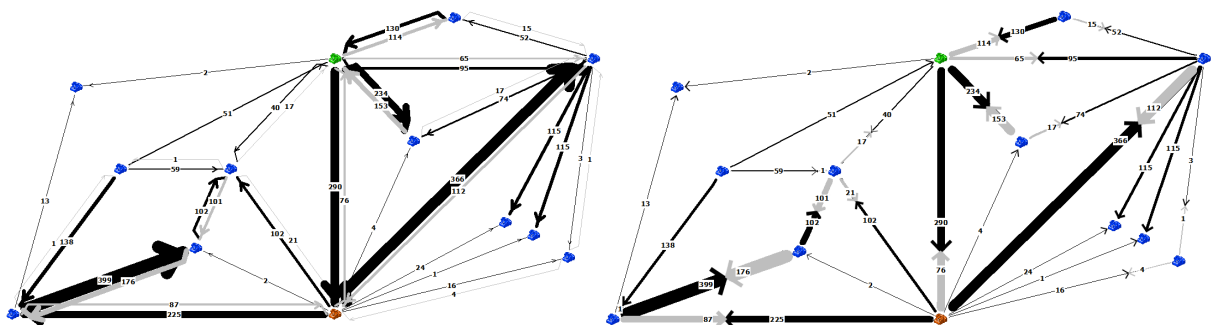
structure 2 ($F(1, 113) = 2.557, p = 0.113$). L'amélioration n'est donc pas certaine, mais un effet tendanciel est observé.

L'amélioration de lisibilité est donc légère pour ce changement de structure. Seule la question quatre a montré une amélioration en termes d'exactitude et du temps de réponse. Cet effet est probablement dû à la diminution du nombre d'entités sur le schéma. Il faut noter que les positions des sommets n'ont pas été modifiées entre les schémas comparés. Alors qu'une diminution du nombre d'entités permet généralement de réduire la taille du schéma est d'optimiser les positions des sommets. Les schémas comparés sont donc très similaires et il n'est pas étonnant que les différences soient tendanciellles. Une augmentation de 8.9% de réponses correctes et un temps moyen diminué de 27.3% ont malgré tout été observés pour la question quatre.

Représentation de liens bidirectionnels



Dans le cadre de la représentation de communications téléphoniques ou de transactions bancaires par exemple, des transactions sont souvent observés dans les deux sens entre les entités (des téléphones ou des comptes). Ces transactions peuvent être représentées par une paire de flèches ou par un lien unique avec une double flèche. L'exemple exploité est également tiré d'une affaire réelle où les communications téléphoniques entre les principaux protagonistes de l'affaire ont été représentées par un schéma de flux.



Question 5 : Cliquer sur tous [les téléphones bleus](#) ayant eu des communications avec le [téléphone orange](#) ET le [téléphone vert](#).

Question 6 : Cliquer sur le téléphone ayant reçu le plus d'appels.

Question 7 : Cliquer sur le téléphone qui est en relation avec le plus grand nombre de téléphones.

Pour les questions cinq et six, aucune différence significative n'a été observée au niveau de l'exactitude et de la complétude entre les deux structures. Il faut relever que pour la question six, le nombre de réponses correctes est faible pour les deux schémas : 22.7 % des participants pour la structure 1 et 16.5% pour la structure 2. La différence n'est cependant pas statistiquement significative ($F(1,211) = 1.297, p = 0.256$). De plus, le temps de réponse est 50% plus rapide pour les réponses fausses, indépendamment du type de schéma.

Les schémas sont donc peu efficaces pour répondre à cette question. En fait, la question porte sur le nombre d'appels total reçu. La dimension dominante est donc la dimension quantitative. L'utilisation d'un schéma relationnel pour répondre à une question de ce type ne semble pas un choix adapté. Pour répondre à la question, il est nécessaire de faire la somme des valeurs inscrites sur les relations. Un histogramme présentant le total des appels reçus par numéro est certainement une solution beaucoup plus adaptée pour ce type de question. Le choix du type de représentation en fonction de la question d'analyse et discuté plus loin.

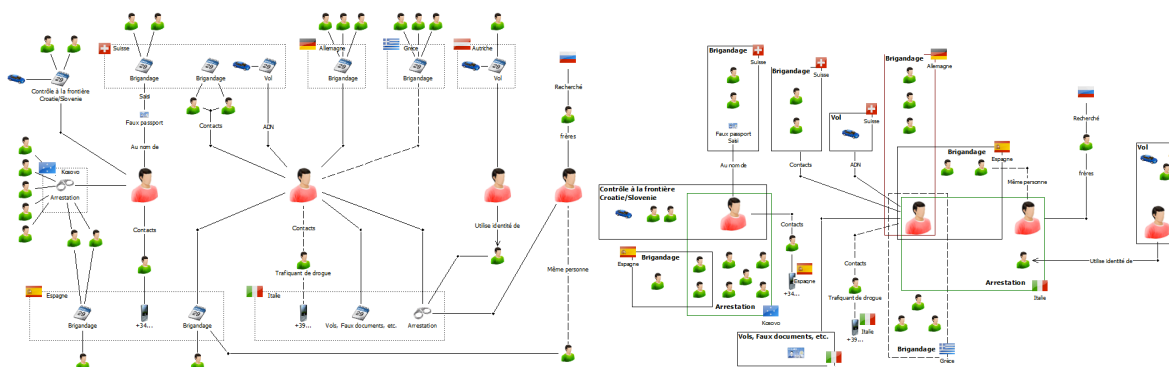
Pour la question sept par contre, le taux de réponses correctes augmente de 13.4% pour la structure 2 ($F(1, 206) = 6.526, p = 0.011$). Le temps de réponse moyen pour les réponses correctes est également plus court de 14.7% avec la structure 2 ($F(1, 169) = 3.674, p = 0.0057$). Il passe de 37 +/- 20 secondes à 31 +/- 16 secondes. Cette dernière est donc plus efficace pour répondre à cette question. Le nombre de traits nécessaires pour représenter l'ensemble des relations est diminué. Cette amélioration de la lisibilité du schéma peut expliquer l'augmentation de l'efficacité. Il faut toutefois relever qu'aucun effet n'a été observé pour la question cinq.

116 Processus de conception d'un schéma relationnel

Représentation de groupes d'entités sous la forme d'un cadre



Il est généralement d'usage de représenter des groupes d'entités (par exemple des personnes travaillant dans une même entreprise ou des cas perpétrés dans une même zone géographique) par un cadre contenant les entités liés⁴¹. L'objectif de cette question consiste à comparer la représentation d'un groupe par un cadre ou par une entité liée aux membres du groupe. Pour cette comparaison, les positions des entités ont dû être modifiées pour effectuer le changement de structure, afin de conserver une lisibilité correcte des schémas. Le schéma présenté est issu d'une affaire réelle portant sur les activités d'un groupe de suspects de brigandages agissant au niveau international.



Question 8 : Cliquer sur **le membre du groupe criminel** qui est lié au plus grand nombre d'individus en vert.

Question 9 : Cliquer sur les individus (vert et rouge) qui semblent avoir eu une activité en Espagne.

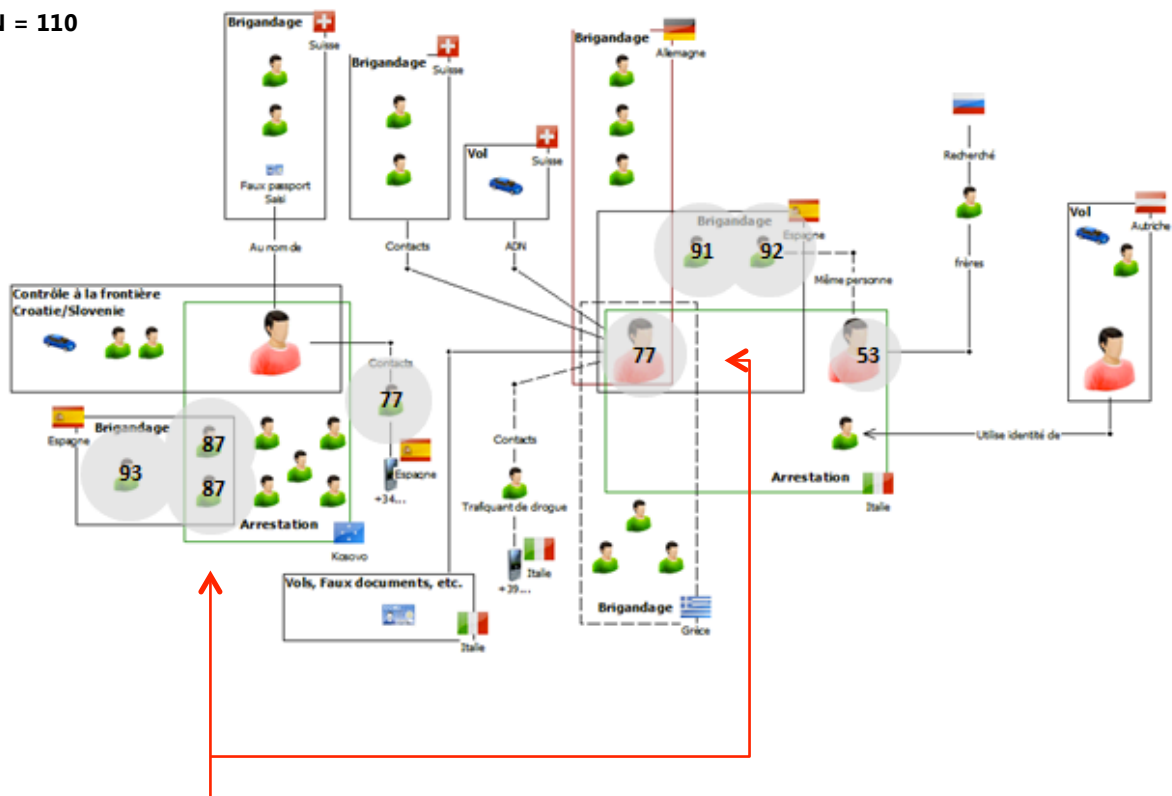
Pour la question huit, le taux de réponses correctes augmente de 9.7% pour la structure 2 ($F(1, 197) = 3.861, p = 0.051$). Aucune différence significative n'est observée pour le temps de réponse.

⁴¹ Cf. Recommandations et standards, page 45

Pour la question neuf, aucune différence n'est observée au niveau du taux de réponses complètes et correctes. Le temps de réponse pour les réponses complètes et correctes diminue de 28.6% pour la structure 1 ($F(1, 36) = 5.490, p = 0.025$). Il passe de 56 +/- 22 secondes à 40 +/- 20 secondes.

Il est très intéressant les résultats en détails pour chaque schéma (les aires sont proportionnelles au pourcentage de participants):

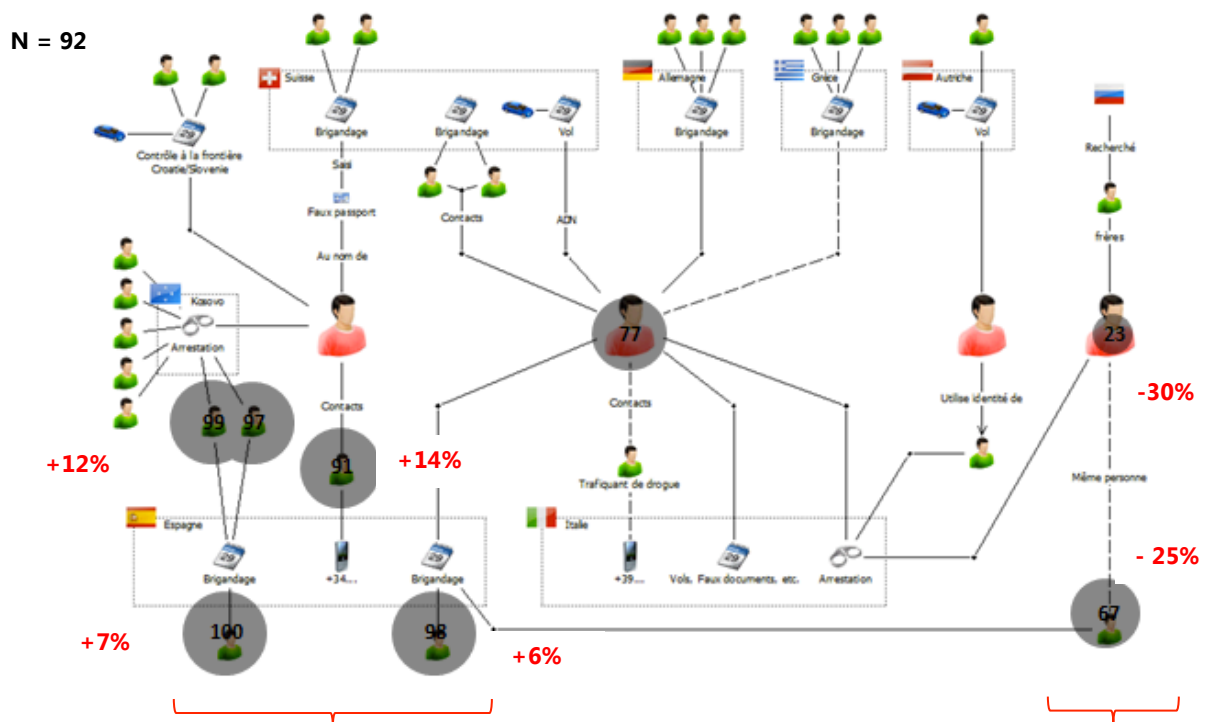
N = 110



Ces entités ont été sélectionnées moins souvent (-6% des participants pour les deux entités de gauche et -14% des participants pour l'entité centrale), par rapport aux autres entités présentes dans les mêmes cadres. Les croisements de traits avec d'autres cadres semblent être la cause de cet effet.

Cette tendance est également observée pour l'individu de droite (53% des participants l'ont sélectionné). Il faut relever qu'il n'est pas dans le cadre représentant le brigandage commis en Espagne. De plus, le lien tracé en trait-tillé peut avoir un effet (en raison de la convention d'utilisation du trait-tillé pour exprimer un lien incertain).

118 Processus de conception d'un schéma relationnel



Globalement, toutes ces entités ont été sélectionnées plus souvent par rapport à la structure 2.


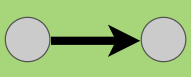





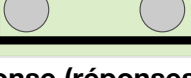
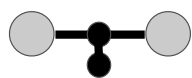
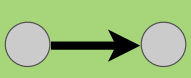






Ces deux entités ont été sélectionnées moins souvent par rapport à la structure 2. Le positionnement plus éloigné de ces entités est certainement la raison de cette diminution.

L'effet positif observé pour le groupe d'entités de gauche et donc contrebalancé par l'effet négatif de l'éloignement des deux entités de droite. Le positionnement peut expliquer globalement pourquoi aucune différence significative n'est observée entre les deux structures pour cette question. Cet effet, illustre bien la complexité de la comparaison de deux représentations graphiques. En effet, plusieurs variables visuelles peuvent avoir un impact simultané sur la perception (comme ici la structure du graphe et les positions des entités). C'est pourquoi, pour l'ensemble des autres questions les positions, les couleurs, les tailles et les icônes n'ont pas été modifiées entre les schémas. Pour cet exemple, conserver les positions n'était pas possible.

Cet exemple, a été réalisé afin de montrer l'impact sur la lisibilité de l'utilisation de cadres pour représenter deux types de relations. Pour la structure une, les cadres représentent des relations spatiales entre les activités des protagonistes qui sont représentées par des entités. Pour la structure deux, les cadres représentent les activités elles-mêmes qui entourent les

personnes impliquées. Globalement, la structure une est plus lisible que la structure deux. Lorsque les activités sont représentées par des cadres, des croisements de traits sont inévitables si des individus ont participé à plusieurs événements. Ces croisements de traits (de cadres) diminuent la lisibilité du schéma (dans le cadre de cette évaluation, une diminution de 6 à 14% de clics est observée pour les entités liées à plusieurs activités). Cet effet n'est pas observé lorsque les événements sont représentés par des entités liées par des arêtes aux individus. Le temps de réponse est également plus court pour cette structure de graphe. L'utilisation de cadres est donc possible pour exprimer des relations entre des entités, mais lorsque les cadres se croisent, l'efficacité du schéma diminue.

T a b l e a u d e s y n t h è s e

Nombre de réponses correctes et complètes			
Ventes			Q1: +14.8% $F(1,239) = 5.08, p = 0.025$ Q2: +16.2% $F(1,231) = 6.412, p = 0.012$
Traces			Q1: Non significatif Q2: +8.9% $F(1,226) = 2.973, p = 0.086$
Téléphonie			Q1 et Q2: Non significatif Q3: +13.4% $F(1,206) = 6.526, p = 0.011$
Événements			Q1: +9.7% $F(1,197) = 3.861, p = 0.051$ Q2: Non significatif
Temps de réponse (réponses correctes)			
Ventes			Q1: Non significatif Q2: -28.2% $F(1,92) = 8.947, p = 0.0036$
Traces			Q1: Non significatif Q2: -27.3% $F(1,183) = 8.839, p = 0.0035$
Téléphonie			Q1 et Q2: Non significatif Q3: -14.7% $F(1,169) = 3.674, p = 0.0057$
Événements			Q1: Non significatif Q2: -28.6% $F(1,36) = 5.490, p = 0.025$

D i s c u s s i o n g é n é r a l e s u r l e s r é s u l t a t s

Globalement des différences significatives ont été observées au niveau de l'exactitude, de la complétude et du temps de réponse en fonction de la manière de représenter les relations entre des entités d'intérêts, pour des questions spécifiques. Bien que les informations représentées soient les mêmes entre les schémas et qu'aucune des formes de représentation ne biaisent les informations, des différences d'exactitudes sont observées. Initialement, il était plutôt attendu d'observer des différences sur le temps de réponses, mais peu sur

l'exactitude. Les représentations ont été créées pour chaque structure de façon à pouvoir répondre au mieux aux questions. En effet, ces schémas ont été conçus et exploités dans le cadre d'affaires réelles⁴² (en l'occurrence les solutions les moins efficaces). Ils ont été créés avec l'intention de faciliter l'analyse et la communication. Les schémas de référence n'ont donc pas été produits pour démontrer l'efficacité des solutions proposées. Les comparaisons effectuées semblent donc intègres et confirment globalement l'importance des choix de représentation.

Par ailleurs, le temps de réponse est généralement plus court pour les réponses fausses ou incomplètes que pour les réponses correctes et complètes. L'exactitude semble globalement corrélée au temps de réponse (telle que relevée dans la littérature⁴³). Une hypothèse possible est que si pour une question donnée, l'effort nécessaire pour y répondre paraît trop important, les participants choisissent une réponse plus rapidement sans vérifier en détails l'exactitude de la réponse. Il faut toutefois noter que pour la première question l'effet est inversé avec le schéma deux (le temps de réponse moyen pour les réponses correctes et complètes est plus court). C'est également le cas pour la question huit avec les deux schémas. Globalement, il ressort également des résultats que :

1. Diminuer le nombre d'entités sur un graphe au profit d'une représentation par des liens augmente la lisibilité.
2. Les positions relatives des entités entre elles jouent un rôle important au niveau de la perception d'un graphe.
3. L'utilisation de cadres pour représenter des relations diminue la lisibilité lorsque les cadres se croisent.

Il est finalement intéressant de souligner qu'aucun résultat contradictoire n'a été observé. Même si pour certaines questions spécifiques, aucune différence significative n'a été démontrée. Lorsque des différences d'efficacité ont été observées pour plusieurs questions, elles sont toujours en faveur de la même structure.

D i s c u s s i o n s u r l a m é t h o d e d ' é v a l u a t i o n

Pour chaque comparaison effectuée, les conditions d'expérimentations étaient similaires. Toutefois l'environnement de chaque participant n'était pas contrôlé, puisque le sondage a été envoyé électroniquement aux participants qui ont effectué le sondage librement. Pour

⁴² A l'exception de l'exemple sur les photocopieurs qui est tiré d'une recherche.

⁴³ Cf. Critères d'évaluation, page 62

chaque comparaison une centaine de personnes ont répondu aux questions pour chaque schéma. Les schémas ayant été attribués aléatoirement, les différences de niveau de formation, d'âge, de connaissances en matière d'analyse criminelle et d'expérience sur la lecture et la conception de schémas n'expliquent pas les différences observées.

Il faut relever encore que les conditions de l'expérimentation effectuée ne correspondent pas aux conditions réelles d'exploitation des schémas. Au cours d'une enquête, les schémas sont exploités par des enquêteurs motivés par la résolution des cas traités et qui exploitent ces schémas dans un contexte différent et avec d'autres motivations que répondre correctement à des questions spécifiques. Les schémas sont exploités globalement pour mieux comprendre l'ensemble des informations et souvent pour répondre à de multiples questions sur différents aspects d'une affaire. Les différences d'efficacité observées en termes d'exactitude pourraient donc être plus faibles dans un contexte réel. Il faut toutefois relever que ces exemples se basent sur des cas réels où il a été parfois nécessaire d'effectuer des modifications de structures, à la demande des enquêteurs, pour que les représentations soient exploitables. En effet, l'impression de lisibilité initiale d'un schéma est très importante pour qu'il soit utilisé. Si un schéma semble compliqué ou confus, il ne suscite généralement que peu d'intérêt.

Cette évaluation se base sur des indicateurs quantitatifs pour mesurer les différences d'efficacité. Ce choix a été effectué pour confirmer par des indicateurs objectifs, les études de cas présentés dans le chapitre « Evaluations préliminaires » et pour évaluer l'impact des choix de représentation. Au vu des différences d'exactitudes observées, une évaluation qualitative complémentaire aurait été utile pour valider les hypothèses formulées pour expliquer les différences. En effet, il aurait été souhaitable de demander aux participants ayant commis des erreurs ou répondu partiellement, d'expliquer les raisons de leurs choix. Une telle démarche nécessite toutefois de développer un système de validation des réponses en temps réels. Le système utilisé pour répondre aux questions en cliquant directement sur les images a été spécifiquement développé pour ce sondage. Inclure un système automatique de contrôle des réponses nécessite donc des développements techniques complémentaires. Il faut toutefois souligner que l'objectif de cette évaluation était de montrer que choisir une manière adaptée de représenter les informations en fonction de la nature des relations accroît objectivement la lisibilité du schéma. La méthodologie d'évaluation développée a répondu à l'objectif.

E l é m e n t s d e m é t h o d e

Dans les sections précédentes, le processus de conception d'un graphe et le langage visuel ont été explicités. Des structures syntaxiques ont été formalisées pour représenter des types de relations génériques. Certaines de ces structures ont été comparées et leur efficacité a été évaluée pour représenter certains types de relations spécifiques. L'hypothèse que la manière de représenter les relations sur un graphe influence la perception et l'efficacité de la représentation a pu être confirmée pour différentes structures et différentes questions spécifiques.

Les différents éléments de méthode identifiés sur la base des cas traités, des évaluations effectuées et issus de la littérature sont présentés dans cette section. Une ébauche de démarche générale exploitable pour concevoir un schéma relationnel, en analyse criminelle, est proposée.

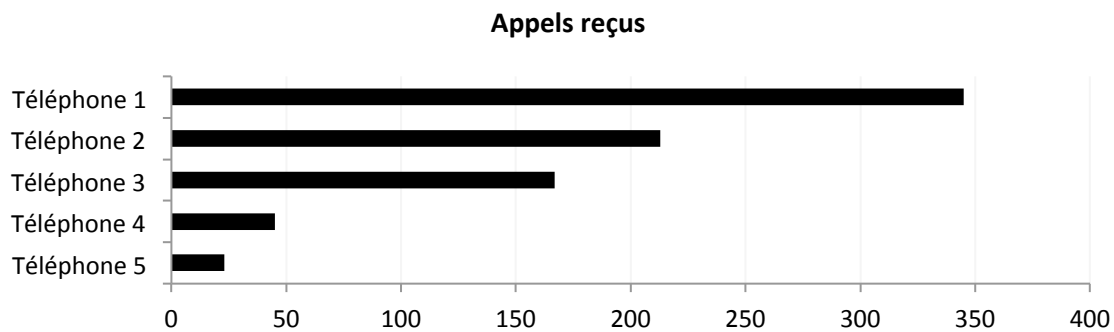
I d e n t i f i c a t i o n d e l a d i m e n s i o n d o m i n a n t e

Le nombre de visualisations existantes pour analyser des informations est considérable. Choisir quelle forme de représentation est la plus adaptée pour résoudre un problème spécifique n'est pas une tâche triviale. La décomposition présentée dans la première partie⁴⁴ permet une classification des méthodes, utile pour guider ce choix. En effet, les méthodes de visualisation peuvent être classifiées en regard des quatre dimensions définies : quantitative, temporelle, spatiale et relationnelle.

La première étape de la méthode proposée consiste donc à identifier quelle est la dimension dominante en regard de la question d'analyse posée. Dans le cadre de l'évaluation présentée dans le chapitre précédent, l'une des questions posées aux participants était : « Cliquer sur le téléphone ayant reçu le plus d'appels ». Le schéma relationnel proposé pour répondre à la question n'est certainement pas le choix le plus adapté pour répondre à cette question. En effet, pour les deux schémas proposés, moins de vingt-cinq pourcent des participants ont réussi à y répondre correctement.

⁴⁴ Cf. Dimensions d'analyse, page 33

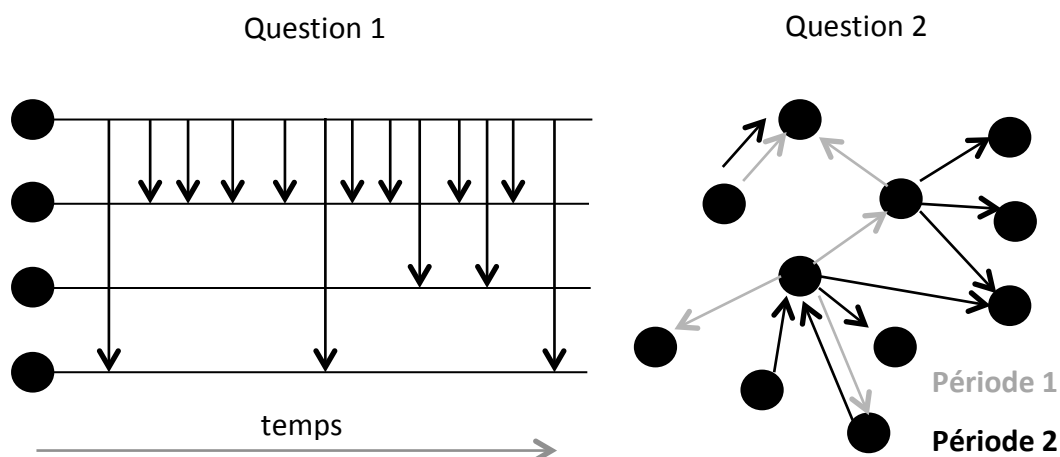
La question porte sur la dimension quantitative. Un histogramme par exemple aurait été un choix certainement plus adapté :



Lors de la décomposition d'une question d'analyse en fonction de ces quatre dimensions, il n'est pas rare que celle-ci porte sur plusieurs dimensions en même temps. Dans ce cas, il convient alors d'identifier quelle dimension est dominante. Par exemple, lors de l'analyse de données de facturation téléphonique, les questions suivantes peuvent se poser :

- Question 1 : quelle est la fréquence des contacts entre un groupe d'individus ?
- Question 2 : quelles sont les personnes en contacts sur certaines périodes définies ?

Ces deux questions portent à la fois sur la dimension relationnelle (les contacts entre des individus) et sur la dimension temporelle (la fréquence / des périodes temporelles définies). Ces deux questions nécessitent pourtant deux formes de représentation différentes pour y répondre :

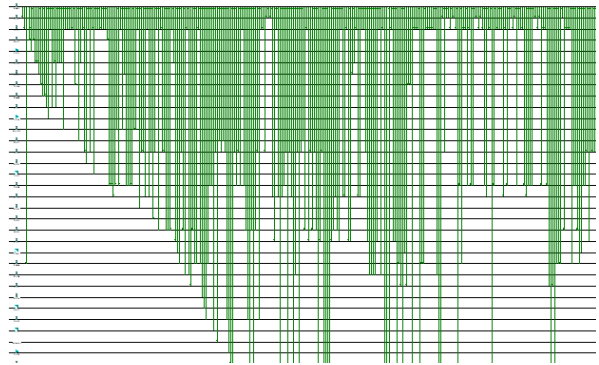


Pour la première question, la dimension temporelle est dominante. Un schéma de flux temporel est donc exploité. Alors que pour la deuxième question, un schéma de flux relationnel est utilisé, car la dimension relationnelle est dominante. En effet, pour cette

124 Processus de conception d'un schéma relationnel

question, la variabilité temporelle est limitée à quelques périodes définies (par exemple deux sur le schéma). De plus, la question porte sur l'identification de contacts liés à un ou plusieurs numéros particuliers. Le degré de variabilité des informations analysées pour chaque dimension permet dans ce cas, d'identifier la dimension dominante : le nombre de périodes est beaucoup plus restreint que le nombre de contacts. Par contre pour la première question, le nombre de contacts est limité (par exemple à un nombre restreint de numéros d'intérêt) et l'ensemble des communications sont représentées temporellement afin de visualiser globalement la fréquence et le rythme des contacts.

Les schémas de flux temporel sont d'ailleurs rarement exploités pour analyser des communications entre un grand nombre de numéros, car le résultat est très souvent difficile à lire :



L'importance de l'identification de la dimension dominante a également été mise en évidence lors de l'évaluation du changement de type de graphique pour analyser la distribution géographique des envoyeurs de spams⁴⁵. Détecter des zones géographiques plus fortement touchées, implique de comparer le nombre de courriels envoyé en fonction de la distribution spatiale. La dimension spatiale est dominante. La carte permet donc d'effectuer visuellement une telle analyse de manière plus efficace. Lors de l'évaluation, il a été observé que le nombre d'observations effectuées par les étudiants en se basant sur la carte est environ dix fois supérieur qu'avec les graphiques pour le niveau de lecture intermédiaire (qui correspond à une lecture aux niveaux de régions géographiques).

Décomposer la question d'analyse en regard des dimensions quantitative, temporelle, spatiale et relationnelle afin d'identifier la dimension dominante qui guide le choix d'une représentation efficace.

La sélection d'un type de représentation adapté a un impact important sur les possibilités d'analyse. La décomposition proposée en fonction des quatre dimensions d'analyse, semble être un guide utile lors de ce choix. Cette proposition initiale est générale. Elle est certainement utile lors de l'élaboration de toutes les formes de représentation. Les éléments

⁴⁵ Cf. Annexe 4 : Evaluations préliminaires - Modification du type de représentation, page 272

de méthode décrits dans les sections suivantes concernent par contre uniquement la dimension relationnelle. Des descriptions des formes de représentation exploitables pour l'analyse quantitative, temporelle et spatiale en regard de différents objectifs d'analyse sont présentées dans l'annexe un.

E l a b o r a t i o n d u m o d è l e r e l a t i o n n e l

La première étape à effectuer lors de la conception d'un schéma relationnel consiste à définir le modèle relationnel du problème traité : une décomposition sous la forme d'entités (des personnes, des événements, des objets, des traces, etc.) et de leurs relations. Parce que le modèle défini omet les éléments non essentiels pour la question d'analyse, il facilite le traitement des informations en réduisant leur complexité. L'abstraction définie par le modèle relationnel permet d'isoler les aspects importants pour la question d'analyse. Il est donc par nature incomplet et imprécis. Il n'y a cependant pas un modèle « correct » pour une situation, uniquement des modèles adéquats ou non en fonction de l'objectif d'analyse (Rumbaugh et al., 1991).

Pour chaque entité les attributs d'intérêts doivent également être identifiés. Par exemple, dans le cadre de l'analyse d'un trafic de marchandises, la date et le lieu des ventes, ainsi que le prix et la quantité des produits vendus sont des éléments généralement pertinents pour chaque transaction investiguée. La démarche générale recommandée consiste à faire l'inventaire le plus exhaustif possible des entités pertinentes à représenter, puis de tracer les relations existantes entre celles-ci.

Identifier les informations pertinentes en fonction de la question d'analyse et élaborer un modèle relationnel décomposant les concepts en entités, attributs et relations.

En effet, il est fréquemment observé que des ambiguïtés de représentation sont issues d'une mauvaise identification d'une entité. Dans le cadre de l'exemple présenté dans le chapitre « Evaluations préliminaires » de telles ambiguïtés sont observées sur un peu plus de cinquante pourcent des schémas analysés⁴⁶. Lors des formations de base sur la conception de schémas relationnelles, il est très souvent observé notamment que les événements sont apparemment les entités les plus difficiles à identifier. Il s'agit du moins du type d'entité le plus fréquemment omis. De multiples liens sont alors tracés entre les entités impliquées lors

⁴⁶ Avec un total de cinquante-sept participants (cf. Représentation d'une affaire complexe, page 73)

126 Processus de conception d'un schéma relationnel

de l'événement (comme des personnes, des véhicules ou des objets). Le schéma s'en trouve alors beaucoup plus complexe, voire dans certains cas ambigu.

Une fois les entités d'intérêts identifiés, leurs relations doivent être définies. Par exemple, les personnes sont généralement reliées aux événements par des relations de type « rôle ». Les traces sont collectées sur des scènes de crimes, des objets ou des personnes. Lors de l'évaluation effectuée, dans plus de soixante pourcent des schémas, certaines relations sont mal positionnées par les participants.

Ces deux types d'erreur semblent être les causes principales des ambiguïtés observées sur des schémas relationnels. Il ressort globalement des évaluations que la décomposition d'une situation en entités et relations n'est pas une tâche évidente. En effet, dans les exemples traités, aucune décomposition émergente n'est observée. Cette étape du processus nécessite donc un certain entraînement et une bonne compréhension de la notion d'entité et de l'importance d'une décomposition adaptée en regard des objectifs d'analyse.

La nature des informations représentées doit également être comprise et maîtrisée. Lors du pré-test réalisé lors de l'évaluation de la section précédente, des difficultés de compréhension ont été observées. Les schémas qui portaient sur des entités peu familières des participants, telles que les saisies, les échantillons et les classes chimiques étaient visiblement plus difficiles à comprendre. Les participants ont passé en moyenne une dizaine de minutes de plus pour répondre aux questions. Cette observation souligne l'importance d'une définition claire des entités en jeu.

Un schéma nécessite d'être accompagné d'une légende et d'une explication guidant sa lecture.

S i m p l i f i c a t i o n s

A readable sociogram is a good sociogram. (Moreno 1953)

Tracer l'image la plus efficace, la plus simple possible. (Bertin, 1967/2005)

Il a été montré⁴⁷ que les représentations qui sont visuellement les plus simples sont généralement les plus efficaces. Le nombre d'entités et de liens est notamment un facteur de complexité. Il est en effet reconnu que plus le nombre d'entités est grand, plus le graphe est difficile à lire. Moreno (1953) recommande de construire des réseaux d'une trentaine d'entités au maximum et de simplifier le schéma si le nombre de sommets est trop important. D'autres auteurs soutiennent également que les graphes les plus efficaces sont de petites tailles (de dix à cinquante entités et entre vingt à cent liens) (Shneiderman & Aris, 2006) (Teyseyre & Campo, 2009).

Optimiser la représentation pour la rendre la plus simple et la plus lisible possible en réduisant le nombre d'entités et de liens.

Simplifier le schéma en représentant certains concepts par des cadres ou par des attributs permet notamment de réduire le nombre d'entités et de liens et ainsi d'augmenter sa clarté. Lorsque deux choix de représentation sont possibles, il est recommandé d'exploiter la forme la plus simple : celle offrant la meilleure lisibilité.

I d e n t i f i c a t i o n d e s e n t i t é s p r i n c i p a l e s

Afin d'effectuer des simplifications sur le schéma, il est nécessaire d'identifier qu'elles sont les entités principales sur lesquelles porte la question d'analyse et entre lesquelles des relations doivent être cherchées et représentées, par exemple :

1. Si l'objectif d'analyse consiste à reconstruire une série d'événements, la représentation est élaborée afin de visualiser les relations entre les cas. L'événement est le type d'entité principal. Lors de la reconstruction de la série, il est notamment nécessaire de décider quels cas font partie de la série en fonction des relations détectées, telles que des relations basées sur les traces ou la proximité spatiotemporelle.

⁴⁷ par les exemples présentés dans la première partie (page 73 et suivantes) et l'évaluation effectuée sur la représentation des relations (pages 104 et suivantes).

2. Dans le cas de la reconstruction d'un réseau criminel, les personnes sont les entités principales entre lesquelles l'enquêteur cherche à identifier et comprendre les relations et les rôles respectifs de chaque protagoniste.
3. Lors de l'évaluation de la pertinence des traces, un schéma relationnel peut être élaboré afin d'avoir une vue d'ensemble des informations. La trace est l'entité principale. La décision porte sur le choix des traces à analyser en priorité.

Identifier les entités principales sur lesquelles portent la question d'analyse et la prise de décision permet de guider les choix de simplification.

Cette étape de processus est importante, car elle permet de guider les choix de représentation, notamment lorsque des simplifications de représentation sont nécessaires pour augmenter la lisibilité. En principe, les concepts sur lesquels se base la prise de décision sont représentés par des entités. Les simplifications sont en priorité effectuées sur les entités liées, telles que les sociétés qui peuvent être représentées par des cadres pour exprimer des relations professionnelles entre des individus sur le schéma d'un réseau criminel.

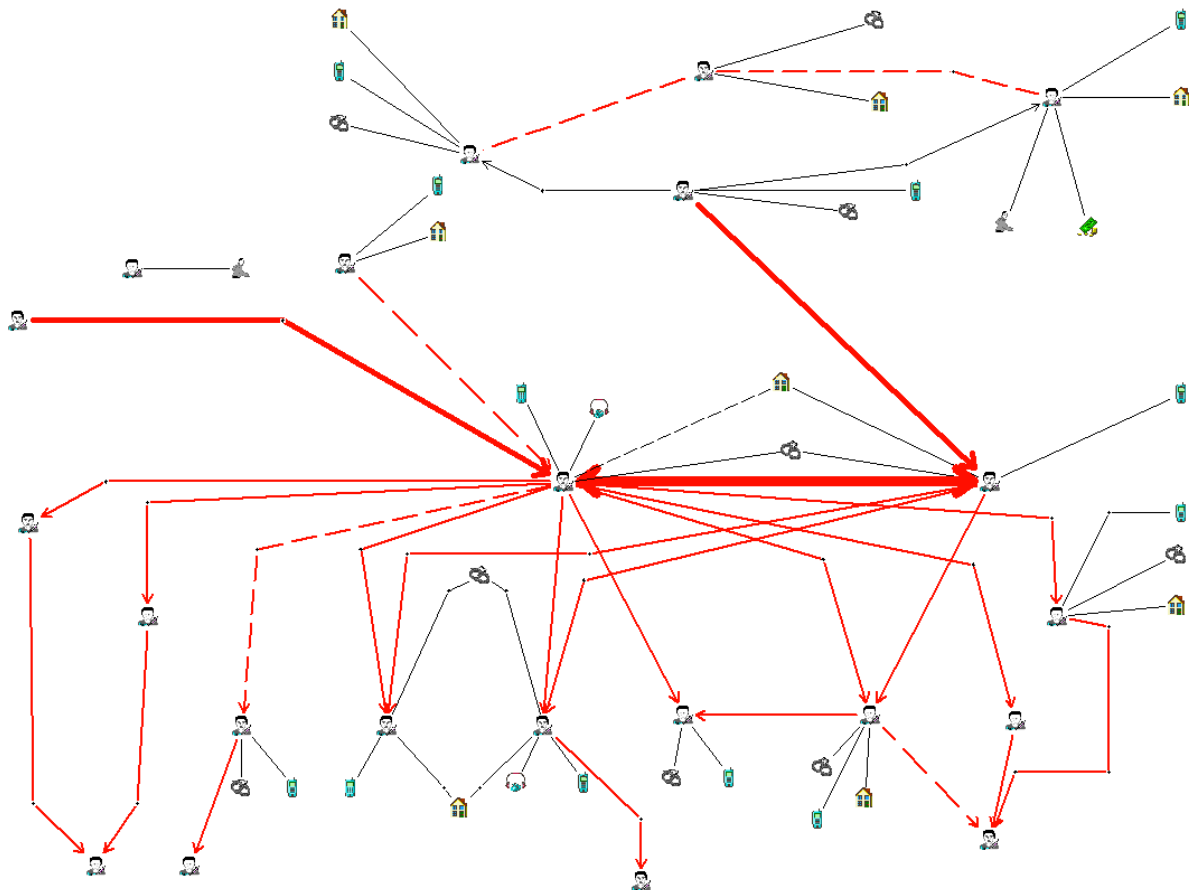
S i m p l i f i c a t i o n p a r s é l e c t i o n

La première stratégie de simplification à mettre en œuvre consiste à ne représenter que ce qui est pertinent et de s'abstraire des éléments inutiles. En fonction de la question d'analyse, il est possible d'identifier les éléments moins pertinents qui peuvent être représentés comme des attributs. Ils peuvent également être retirés de la représentation, s'ils ne sont pas utiles pour répondre à l'objectif d'analyse.

Réduire le nombre d'entités et de liens par sélection. En fonction de l'objectif d'analyse, s'abstraire des éléments inutiles.

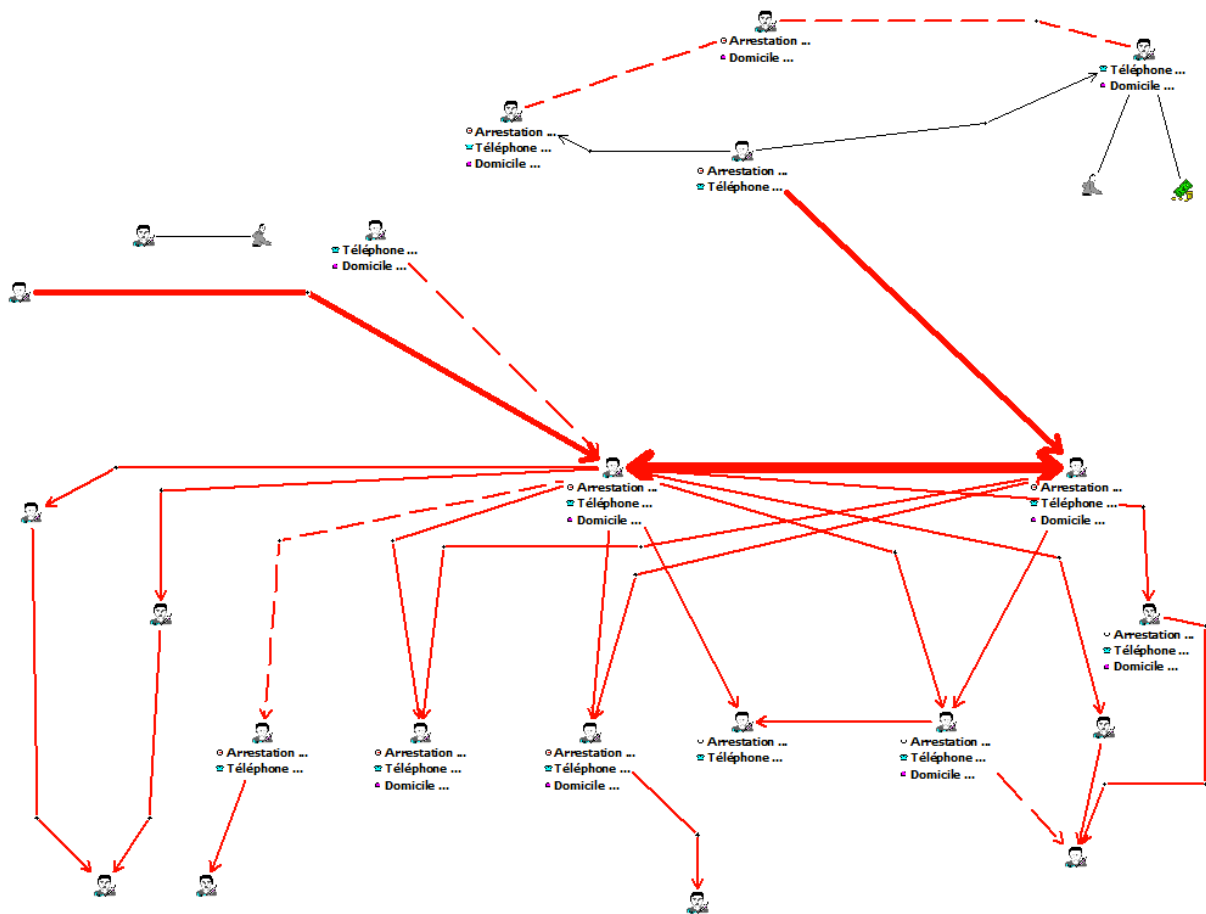
Bertin (2005) décrit la conception d'une représentation comme un processus logique de simplification. Le schéma doit être simplifié par réduction afin d'aboutir à un message clair et efficace. Il souligne néanmoins l'importance de conserver l'ensemble des relations pertinentes pour la question d'analyse.

L'exemple ci-dessous est tiré d'une affaire réelle. L'objectif était de représenter la structure d'un trafic de stupéfiants. Sur ce schéma les informations ont été décomposées en cinq types entités principales : des personnes, des numéros de téléphones, des arrestations et des adresses. Les relations en rouge représentent des ventes de stupéfiants. Les personnes sont les entités principales. Leurs relations mutuelles et le rôle de chaque individu dans le trafic sont investigués.



Les relations téléphoniques entre les personnes n'étant pas représentées, il n'est pas utile de représenter les numéros de téléphone sous la forme d'entités. En outre, à l'exception de deux arrestations, tous les individus ont été arrêtés seul. La question de l'utilité de représenter les relations établies sur la base des arrestations peut également se poser. Finalement, la plupart des informations sur les domiciles des personnes ne permettent pas de mettre en évidence des relations. Il a donc été choisi de représenter ces informations sous la forme d'attributs plutôt que sous la forme d'entités. Le nombre global d'entités est ainsi diminué rendant le schéma plus lisible.

130 Processus de conception d'un schéma relationnel



Globalement, il convient donc d'établir si représenter certains types d'informations par des entités permet de mettre en évidence des relations utiles pour l'analyse. Dans le cas contraire ces informations peuvent être supprimées de la représentation ou représentées sous la forme d'attributs. D'autres simplifications peuvent être apportées à ce schéma, elles sont présentées dans les sections suivantes.

Modification de la structure

Une fois l'ensemble des entités et des relations identifiées, il convient de définir la manière de représenter les relations entre les concepts en fonction des objectifs d'analyse définis. En effet, la représentation des informations décomposées en entités élémentaires et en l'ensemble de leurs relations ne constitue pas toujours une solution efficace. Tel qu'il a été montré dans le cas de la représentation d'un trafic de marchandise⁴⁸, représenter les ventes sous la forme d'entités liées aux acheteurs, vendeurs et objets n'est pas une solution adaptée lorsque le nombre de transactions est important. Un changement de structure est parfois nécessaire pour pouvoir exploiter tout le potentiel de la représentation graphique. La démarche proposée repose sur un usage adapté des formes de représentation des relations définies : le lien, le lien dirigé, le lien bidirectionnel, le lien multiple, le groupe, l'ensemble ou le type.

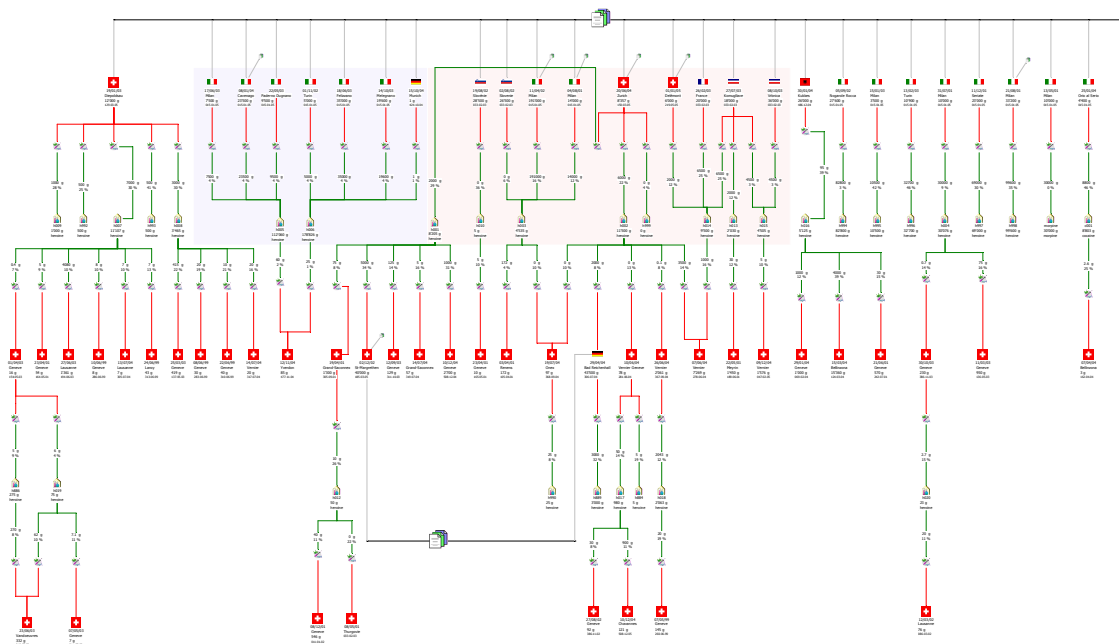
Identifier les types de relation entre les entités principales et sélectionner une forme de représentation adaptée pour chacun d'eux.

Afin d'explicitier la démarche, les schémas relationnels produits pour représenter les liens établis entre des saisies de stupéfiants par l'analyse chimique est présenté (deux échantillons ont le même profil chimique, lorsqu'ils ont été dans la même masse à un moment du trafic (Guéniat & Esseiva, 2005))⁴⁹. La représentation des relations entre les saisies se base sur trois types d'entités : les saisies, les échantillons et les classes chimiques. Pour chaque saisie, un ou plusieurs échantillons sont prélevés des masses de stupéfiant saisies, puis analysées par un laboratoire. Pour chaque échantillon, une classe chimique est déterminée. Des relations avec d'autres saisies sont établies avec les autres échantillons de la même classe. Les saisies, les échantillons et les classes chimiques sont des entités qui peuvent être représentées comme telles. L'exemple présenté est tiré d'une affaire réelle.

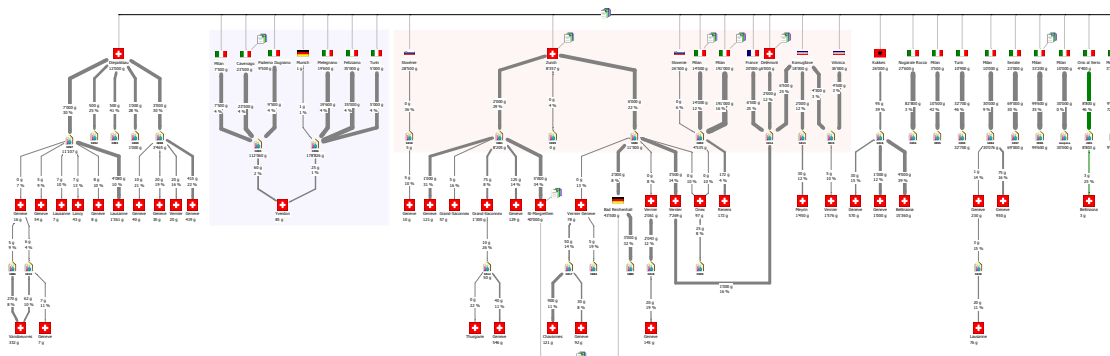
⁴⁸ Cf. Représentation d'un trafic de marchandises, page 75

⁴⁹ Cf. Saisies de stupéfiant et classes chimiques, page 155

132 Processus de conception d'un schéma relationnel



La représentation peut cependant être simplifiée en représentant l'échantillon sous la forme d'un lien entre la saisie et la classe chimique. En effet, un échantillon ne pouvant être lié qu'à une seule saisie et à une seule classe chimique, ce concept correspond à la définition graphique du lien⁵⁰.

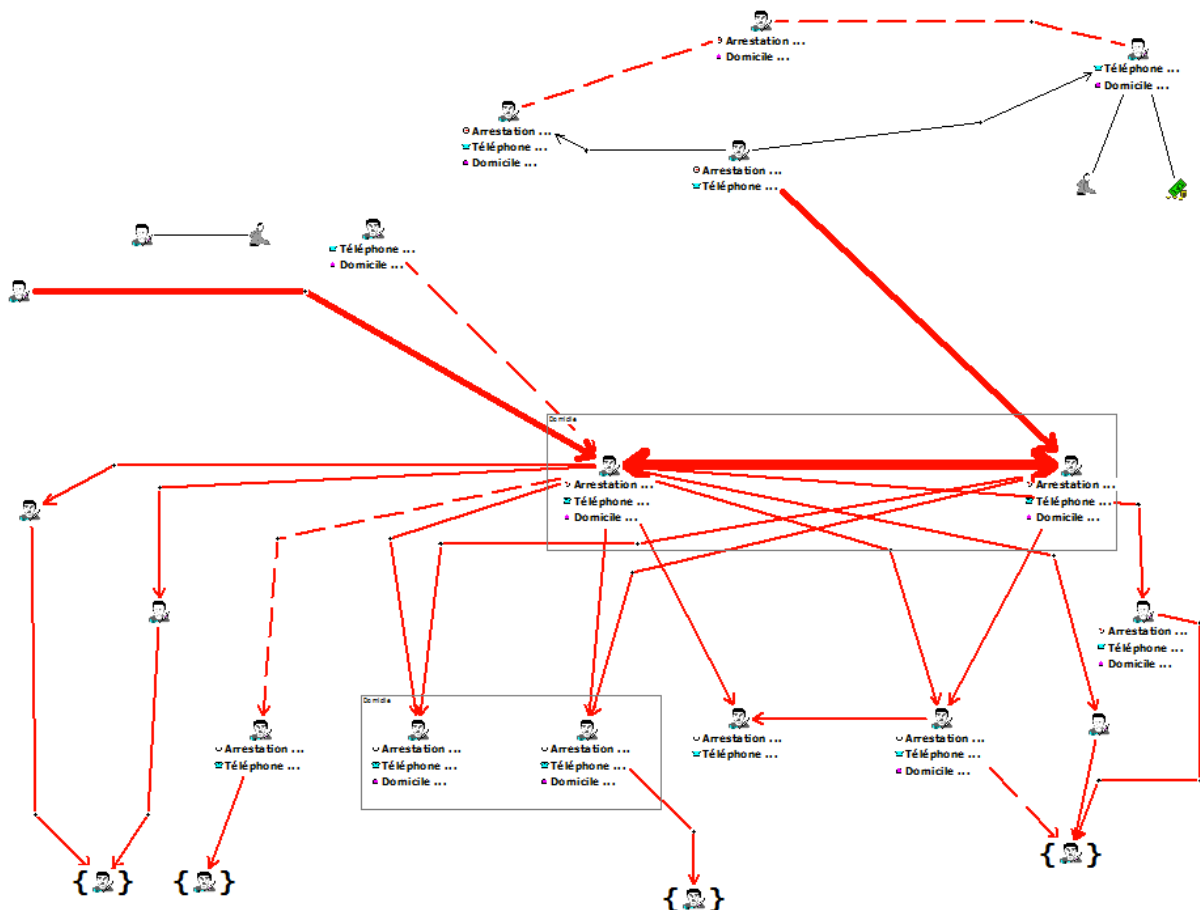


Une telle modification de la structure du schéma améliore sa lisibilité⁵¹. La représentation des relations entre les saisies de stupéfiants fait l'objet d'un pattern détaillé dans le chapitre suivant.

⁵⁰ Cf. Typologie de relations, page 102

⁵¹ Cf. Représentation d'une entité sous la forme d'un lien, page 112

Sur le schéma du trafic de stupéfiant présenté dans la section précédente, les liens établis sur la base du domicile des personnes peuvent être représentés en exploitant des cadres par exemple.



D'autre part, des ensembles de personnes sont également mis en évidence par l'utilisation des accolades (signifiant défini pour représenter un groupe d'entités d'un même type). En effet, les quatre icônes de personnes en bas du schéma représentaient, sur le schéma initial, des groupes de clients toxicomanes. L'usage des accolades permet d'effectuer cette distinction rapidement. Alors que sur le premier schéma, la lecture du label des entités est nécessaire pour faire cette distinction.

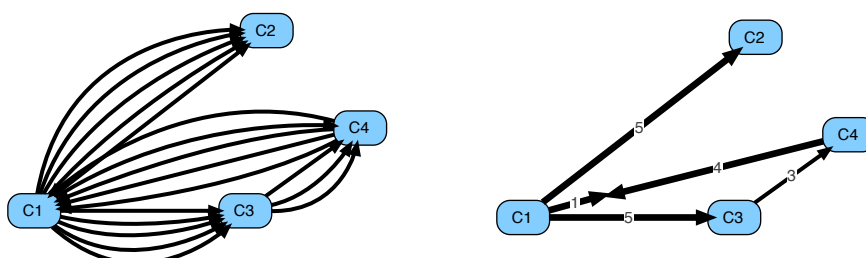
Simplification par agrégation

L'agrégation vise la réduction du nombre d'informations représenté en regroupant les éléments (des entités ou des liens) en groupes : des agrégats (Andrienko & Andrienko, 2005). Ceux-ci sont caractérisés par des attributs calculés qui décrivent l'ensemble, tels que le nombre d'entité ou le nombre de liens. L'utilisation d'agrégats permet de simplifier la représentation, mais elle engendre une perte d'information sur les caractéristiques spécifiques de chacun des éléments regroupés. Varier le niveau d'agrégation permet néanmoins de changer la perspective et peut mettre en évidence des structures ou des régularités difficilement perceptibles au niveau élémentaire (des patterns).

Procéder par agrégation afin de varier la perspective sur les informations et réduire le nombre d'informations par regroupement

Le processus d'agrégation est caractérisé par trois aspects principaux : la manière de regrouper les informations en agrégats, leurs caractéristiques et la manière de les visualiser (Andrienko & Andrienko, 2005).

A titre d'exemple, lors de la représentation de transactions financières, il n'est pas rare de devoir analyser un nombre important de transactions entre un nombre important de comptes bancaires. Plusieurs stratégies d'agrégation peuvent être mises en œuvre : regrouper les comptes (les entités) par leur détenteurs (une personne physique ou morale), regrouper les transactions par types (retraits cash, versements, emprunts, etc.) ou par des périodes temporelles, etc. Dans le cas de l'agrégation des liens, diverses caractéristiques du groupe peuvent être définies : leur nombre, la somme ou la moyenne des montants, la période temporelle totale, etc. Les caractéristiques propres à chaque transaction sont alors supprimées au profit d'une description globale du groupe. Finalement, des choix de représentations doivent être faits pour choisir quelle variable visuelle sera exploitée pour représenter les caractéristiques du groupe, telle que l'épaisseur du lien pour représenter le nombre total de transactions :



C o n s t a n c e e t d é f i n i t i o n d e c o n v e n t i o n s

Conserver une représentation constante de chaque type d'informations est recommandé pour préserver la cohérence du schéma et pour en faciliter la lecture. Moreno (1953) relève : « *For the comparison of two or more groups the sociograms must be constructed after the same principle in order to be readable* ». Pour Bertin (2005), « *dans une image, un même concept ne peut avoir deux implantations différentes* ». Il souligne également que « *toute variation visuelle apparaît comme significative* ». Ce principe de constance est important afin de faciliter la lecture des schémas (Eppler & Aeschmann, 2009).

Définir des conventions de représentation : l'ensemble des informations d'un même type doivent être représentés de façon similaire.

A titre d'exemple, la couleur des liens peut être exploitée pour distinguer des types de relations, tels que le bleu pour les relations familiales et le vert pour les relations professionnels entre des personnes, etc. Si sur un même schéma ces couleurs sont également exploitées pour représenter d'autres types de relations, sa lecture est rendue plus difficile, voire même ambigu si les liens ne sont pas documentés par des labels.

Des conventions doivent également être définies pour exprimer l'incertitude, pour distinguer les faits des hypothèses et pour représenter la négation (Atzenbeck et al., 2009b). Les choix peuvent être divers, mais ils doivent être appliqués de manière constante et documentés par une légende.

R e p r é s e n t e r l ' i n c e r t i t u d e

Une convention existe pour distinguer des relations confirmées de relations incertaines. En effet, l'usage du trait-tillé semble globalement admis pour exprimer l'incertitude.



Par contre, aucune convention n'existe pour exprimer l'incertitude sur les nœuds. Pourtant, il n'est pas rare que des informations descriptives des entités soient incertaines (comme l'identité d'une personne par exemple). De ces cas, il est utile d'exprimer visuellement cette incertitude. Le style de la police d'écriture peut être utilisé : par exemple, l'italique.



Identité établie

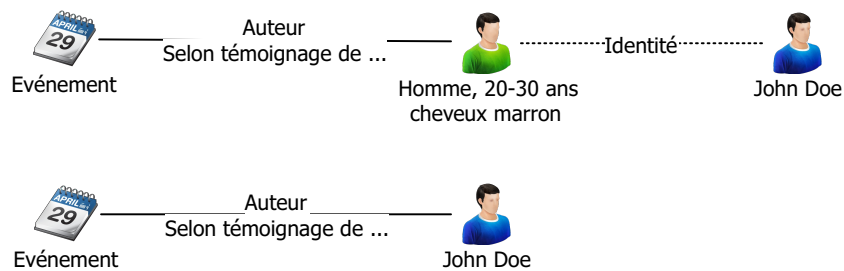


Identité incertaine

Dans le cadre de l'enquête judiciaire, l'incertitude est permanente, notamment lorsque les informations représentées sont issues de déclarations de victimes, de témoins ou de prévenus. L'incertitude est alors représentée en fonction de la source. Si celle-ci déclare avoir des doutes, l'information est représentée comme incertaine. Si par contre elle affirme un fait, l'information est représentée comme confirmée. Il est donc important de représenter également la source de l'information pour en évaluer la fiabilité.

Distinguer les faits des hypothèses

Les notions d'incertitude et d'hypothèse sont très proches. Il est toutefois utile de distinguer les hypothèses formulées sur la base des données collectées de l'incertitude ou des hypothèses faites par les sources d'informations. Par exemple, l'analyste peut faire l'hypothèse d'un lien entre une personne inconnue (liée à une activité particulière) et un suspect potentiel. Sur le second schéma, l'incertitude est issue du témoignage :



Une hypothèse étant par définition incertaine, la convention du trait-tillé est exploitée. La police de caractère est alors utilisée pour distinguer l'hypothèse de l'incertitude. Des points d'interrogation peuvent également être ajoutés pour renforcer la distinction.

Représenter la négation


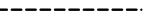

Représenter la négation sur un graphe n'est *a priori* pas une difficulté :



D'un point de vue visuel, toutefois, tracer un lien entre deux entités exprime la présence d'une relation. Une lecture détaillée des labels des liens est alors nécessaire pour distinguer la négation. Sur un graphe de taille plus importante, ne pas distinguer visuellement la négation peut engendrer une perception visuelle faussée des relations. Distinguer visuellement la négation en exploitant la couleur, le style de trait, voire la police s'avère nécessaire :



Il est donc important de définir trois conventions différentes pour distinguer ces dimensions.

Légende				AaBbYyZz	AaBbYyZz
	Lien confirmé	Lien incertain	Hypothèse	Négation	Information incertaine

O r d o n n a n c e m e n t d e s c o m p o s a n t s

Lorsque la structure du graphe est définie et simplifiée par sélection, par l'exploitation de représentations adaptées des relations ou par agrégation, le positionnement des entités peut être optimisé afin d'accroître la lisibilité du schéma. Des algorithmes automatiques ont été développés pour optimiser ce placement (cet aspect technique fait l'objet de nombreuses recherches qui ne sont pas discutées dans cette thèse).

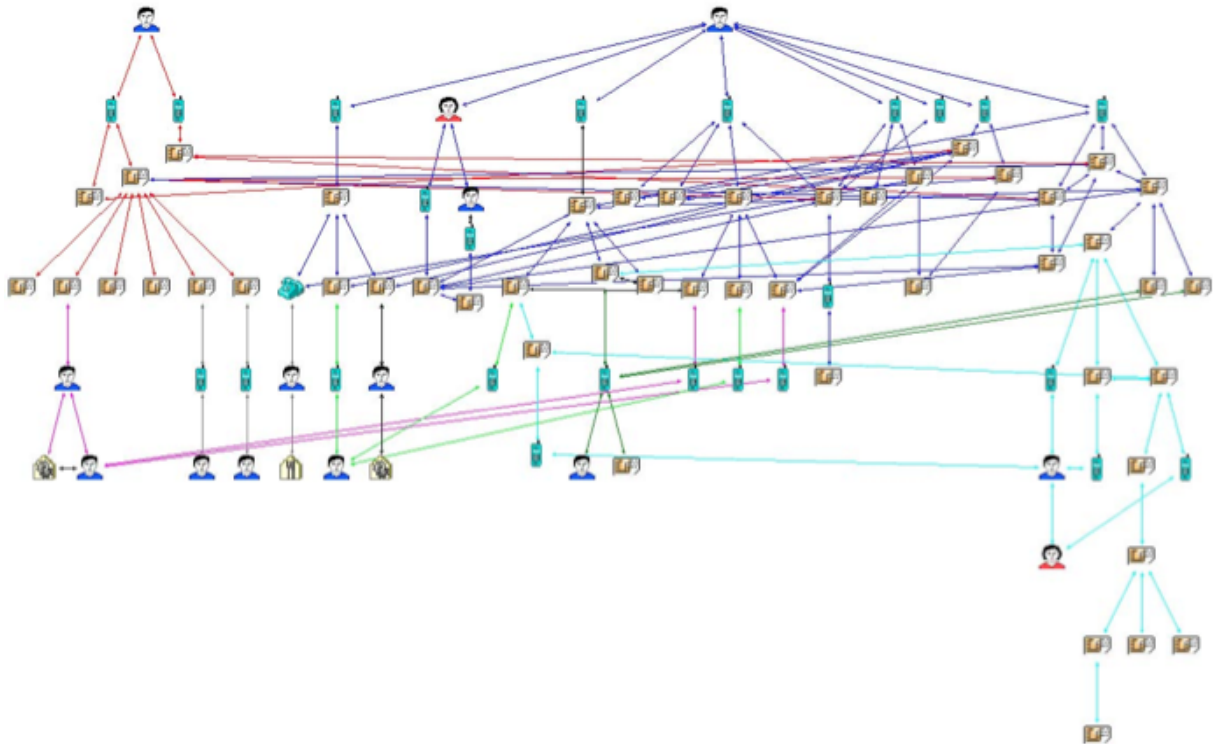
Globalement l'optimisation du positionnement se base sur deux principes généraux :

1. Le placement doit minimiser le nombre de croisement de traits
2. Les nœuds doivent être positionnés afin de rendre les traits des relations orthogonaux ou parallèles.

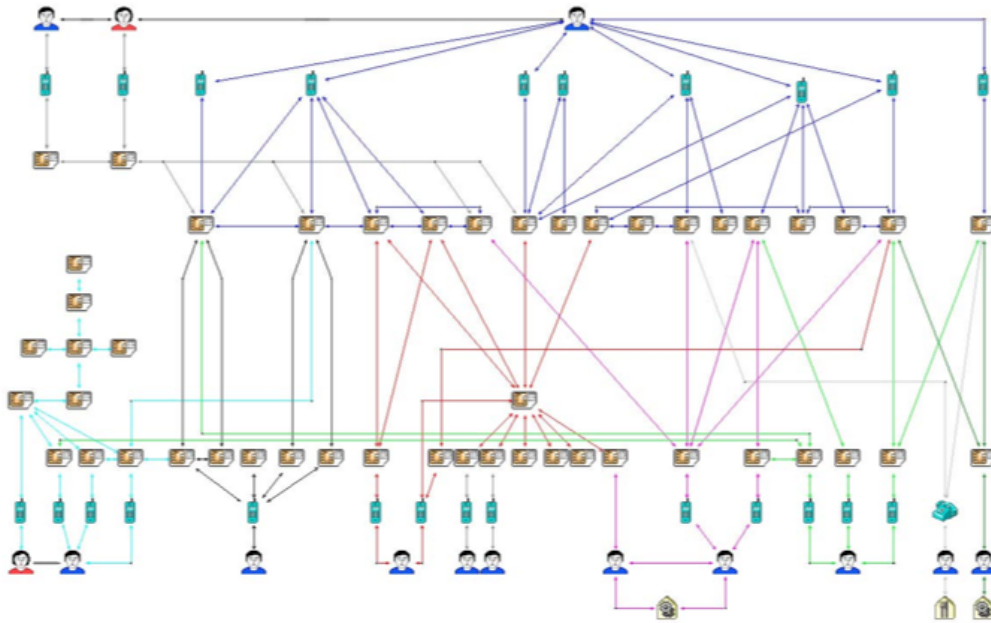
L'origine exacte de ces recommandations est difficile à établir. En effet, de nombreux auteurs ont discutés et évalués ces recommandations sans toujours en citer une origine. Bertin (2005) relevait déjà qu'il est nécessaire de « *chercher la disposition qui offre le minimum de croisements ou la figure la plus simple, tout en conservant les groupements, oppositions ou ordres éventuels* ». Moreno (1953), dont les travaux semblent être à l'origine des réseaux sociaux, recommande également de minimiser les croisements de traits. Les travaux de recherches menés sur l'impact du placement des nœuds sur la lisibilité d'un graphe ont montré que minimiser les croisements de traits améliore significativement la lisibilité. L'orthogonalité et le parallélisme des relations semblent avoir par contre un impact plus faible. L'orthogonalité et le parallélisme augmente toutefois l'esthétisme du graphe et rend sa lecture plus agréable et facile (Purchase, 2000).

138 Processus de conception d'un schéma relationnel

L'exemple ci-dessous illustre l'impact du positionnement des entités sur la lisibilité et son impact sur l'utilisabilité du schéma. L'exemple présenté est issu d'une affaire réelle. Il résume les nombreuses relations téléphoniques entre les protagonistes de l'affaire. Quatre types d'entités sont représentés : des personnes, des sociétés, des téléphones et des cartes SIM. Les téléphones et les cartes SIM sont séparés en deux types d'entités, car certains individus ont utilisé plusieurs cartes dans un même boîtier.



La première observation faite au niveau du positionnement des nœuds est l'apparente structure hiérarchique. Il a donc été demandé au créateur de la représentation si ce positionnement était voulu (par exemple, afin de représenter par le placement une structure entre les individus ou mettre en avant certains protagonistes). Cette structure était en fait apparue au fil des tentatives effectuées pour minimiser les croisements de traits et rendre le schéma plus lisible. Elle n'était pas particulièrement souhaitée. Ce concepteur ne trouvait pas de solution efficace et avait perdu la maîtrise de son schéma. Le schéma ci-dessous a alors été élaboré en optimisant le placement des entités.



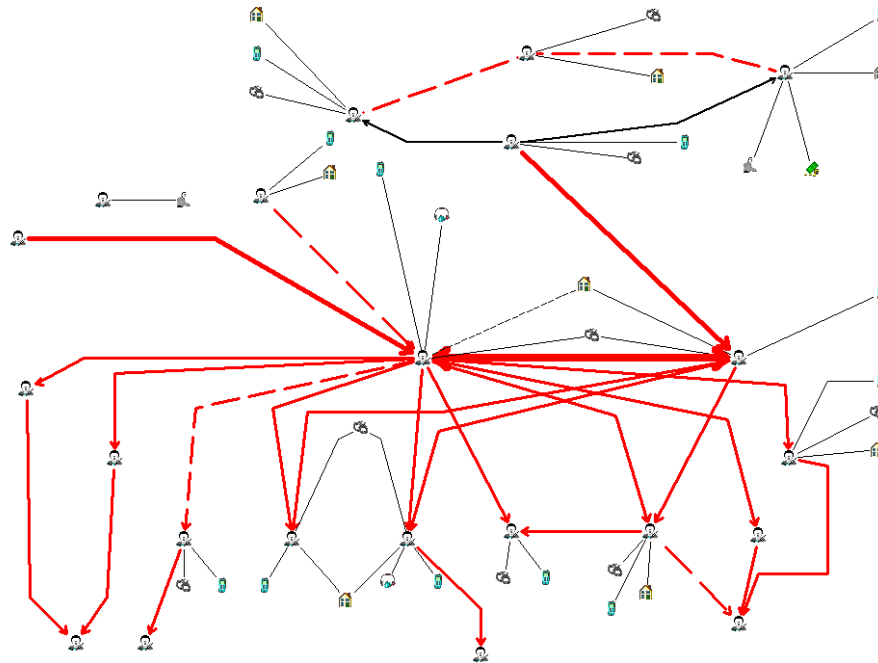
Le schéma final est visiblement plus clair et simple à lire. Les modifications apportées se basent sur les deux principes décrits précédemment (minimiser les croisements de traits et maximiser l'orthogonalité), mais également sur un placement ordonné des nœuds en fonction de leurs types. En effet, les personnes, les boîtiers et les cartes SIM sont repartis sur l'axe vertical sur des lignes bien séparées.

Par ailleurs, les couleurs des relations n'ont pas été modifiées par rapport au schéma initial. Pourtant sur le schéma final, les couleurs créent un effet souhaité de couches qui permet de distinguer plus facilement les relations et minimise l'impact sur la lisibilité des croisements de traits. Dans cet exemple, la couleur est définie en regard des personnes liées (une couleur par personne ou groupe de personnes). Il est également intéressant de relever que l'individu en haut au centre du schéma est mis en évidence par rapport aux autres personnes. Cet effet visuel est connu (Eppler & Aeschmann, 2009). Moreno (1953) relevait déjà que la position centrale en fait un groupe privilégié qui « écrase » les autres groupes. Un ordonnancement linéaire des nœuds peut néanmoins légèrement atténuer ce biais visuel (tel que dans l'exemple ci-dessus).

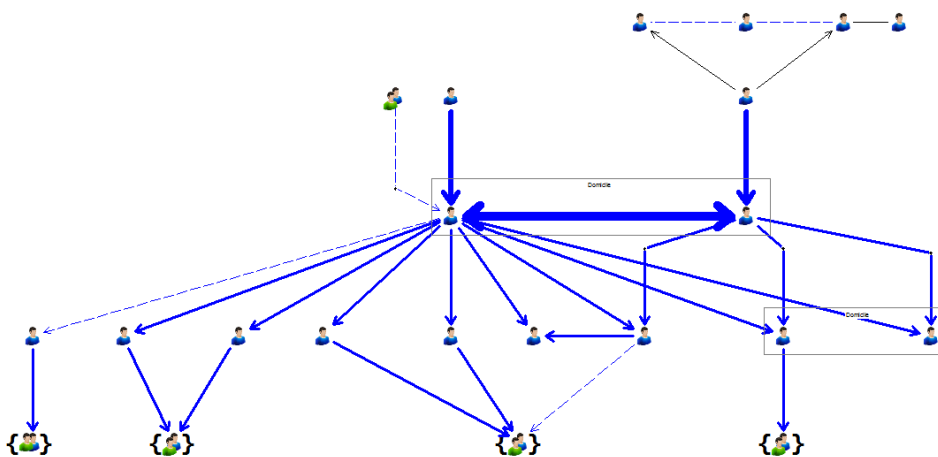
Un positionnement adapté des nœuds est donc important pour accroître la lisibilité d'un schéma. Il peut également être exploité pour représenter des relations particulières entre les entités. Représenter une chronologie dans le cas d'un schéma temporel ou séparer des groupes de personnes lors de la représentation d'un réseau sont des exemples d'exploitation du positionnement pour représenter des relations.

140 Processus de conception d'un schéma relationnel

Positionner les sommets sur le plan pour exprimer un type de relation particulier tel que la chronologie et/ou en fonction des types d'entités, tout en veillant à minimiser les croisements de traits et à maximiser l'orthogonalité afin d'accroître la lisibilité.



La lisibilité du schéma représentant un trafic de stupéfiant précédemment présenté peut également être améliorée en optimisant les placements des entités en fonction de l'implication des protagonistes dans le réseau :



Les deux individus centraux sont les trafiquants sur lesquels l'enquête c'est focalisée. Les individus au-dessus sont les fournisseurs identifiés et les personnes en dessous sont les dealers auxquels ils ont vendu des stupéfiants. Les entités en bas représentent des groupes de clients. Le positionnement est donc défini en fonction du rôle des personnes afin de représenter la structure du réseau.

R é s u m é d e l a m é t h o d e p r o p o s é e

1. Définir clairement le ou les objectif(s) de la représentation. Quelles sont les questions auxquelles elle doit répondre ?
2. Décomposer la question d'analyse en regard des dimensions quantitative, temporelle, spatiale et relationnelle afin d'identifier la dimension dominante qui guide le choix d'une représentation efficace.
3. Identifier les informations pertinentes en fonction de la question d'analyse et élaborer un modèle relationnel décomposant les concepts en entités, attributs et relations. De plus, le schéma nécessite d'être accompagné d'une légende et d'une explication qui définit notamment les types d'entités et de relations afin de guider sa lecture.
4. Optimiser la représentation pour la rendre la plus simple et la plus lisible possible en réduisant le nombre d'entités et de liens:
 - a. Identifier les entités sur lesquelles portent la question d'analyse et la prise de décision permet de guider les choix de simplification. Les concepts sur lesquels se base la prise de décision sont en principe représentés par des entités.
 - b. Réduire le nombre d'entités et de liens par sélection. En fonction de l'objectif d'analyse, s'abstraire des éléments inutiles. Ceux-ci peuvent être par exemple représentés comme des attributs ou retirés de la représentation.
 - c. Identifier les types de relation entre les entités principales et sélectionner une forme de représentation adaptée pour chacun d'eux : le lien, le lien dirigé, le lien bidirectionnel, le lien multiple, le groupe, l'ensemble ou le type.
 - d. Procéder par agrégation afin de varier la perspective sur les informations et réduire le nombre d'informations par regroupement.
5. Définir des conventions de représentation : l'ensemble des informations d'un même type doivent être représentés de façon similaire. Spécifier les conventions utilisées pour décrire les types d'entités et de liens, mais également pour exprimer l'incertitude, pour distinguer les faits et les hypothèses et pour représenter la négation.
6. Positionner les sommets sur le plan pour exprimer un type de relation particulier tel que la chronologie et/ou en fonction des types d'entités, tout en veillant à minimiser les croisements de traits et à maximiser l'orthogonalité afin d'accroître la lisibilité.

A pattern language is nothing more than a precise way of describing someone's experience. (Alexander, 1979)

P a t t e r n s d e v i s u a l i s a t i o n r e l a t i o n n e l l e

Les recommandations et les éléments de méthode formalisés dans le chapitre précédent sont en partie issus de recherches empiriques et de principes issus de la littérature. Mais ils ont également été formalisés sur la base d'observations effectuées au cours d'enquêtes. Ce chapitre présente les bonnes pratiques identifiées pour représenter certaines situations récurrentes traitées dans le cadre d'enquêtes judiciaires, telles que le traitement de séries de cambriolages ou l'analyse de données téléphoniques. La démarche est ainsi explicitée par des exemples concrets. L'objectif est d'amorcer le développement d'un catalogue de bonnes pratiques qui répertorie des situations qui se répètent et des solutions qui sont appropriées. Il ne saurait être démontré que les solutions proposées sont les meilleures possible. Elles ont toutefois montré une certaine efficacité lors de l'analyse des cas ou pour la communication des résultats d'enquête lors de la transmission de dossiers aux magistrats instructeurs. Inventorier des formes de représentations adaptées pour des situations fréquemment traitées devrait être utile pour faciliter l'échange de connaissances entre praticiens et également en matière d'enseignement.

Le terme de pattern est utilisé pour décrire ces bonnes pratiques. La notion de pattern est issue des travaux de Christopher Alexander qui, dans les années septante, a développé des solutions pour des problèmes spécifiques de conception architecturale (Alexander, 1979). La formalisation proposée par Alexander a par la suite été reprise dans le domaine de la conception logiciel où des patterns ont été définis pour décrire des solutions spécifiques pour des problèmes de développement informatique récurrents (Gamma et al., 1995). Un pattern peut se définir comme la description d'une solution adaptée pour un problème récurrent de conception. Dans le cadre de cette thèse, les patterns formalisés décrivent des solutions efficaces pour représenter des situations récurrentes. Chaque pattern est composé de quatre parties : (1) une description du contexte d'exploitation, (2) une description du ou des objectifs

d'analyse, (3) une description des concepts à représenter et (4) une description de la solution graphique proposée. Pour chaque pattern, au moins un exemple d'exploitation est présenté et les choix de représentation sont discutés. Seuls trois patterns ont été formalisés sur la base des cas traités par l'auteur et de discussions avec des praticiens : analyser une série d'événements, évaluer la pertinence de traces et analyser des transactions. L'idée de ce chapitre est d'amorcer le développement d'un catalogue de patterns plus large.

A n a l y s e r u n e s é r i e d ' é v é n e m e n t s

C o n t e x t e

De nombreuses enquêtes portent sur le traitement de séries de délits perpétrés par des auteurs dits prolifiques. Il n'est pas rare, par exemple, que des cambrioleurs commettent de nombreux cas. Savoir détecter, reconstruire et analyser une série d'événements offre de nombreuses perspectives opérationnelles. Détecter des régularités spatio-temporelles pour la mise en place de surveillances, confirmer l'hypothèse qu'un ensemble de cas a été commis par le même individu ou groupe d'individus à des fins de poursuites judiciaires ou encore identifier des modes opératoires émergeant à des fins de prévention, sont autant de renseignements qui découlent de l'analyse d'une série. Dans ce contexte, plusieurs formes de visualisation peuvent être exploitées pour soutenir l'analyse : représenter les distributions temporelles et spatiales par des diagrammes et des cartes⁵² ou représenter l'ensemble des relations entre les cas sur des schémas d'événements⁵³ afin notamment de tester l'hypothèse de la série. La conception de ce type de schémas est explicitée dans ce pattern.

O b j e c t i f s d e l a r e p r é s e n t a t i o n

Les schémas d'événements sont exploités au cours de l'enquête pour reconstruire la série et faciliter sa vue d'ensemble. En fin d'enquête, ces schémas servent également à faciliter l'appréhension du dossier par les acteurs du processus judiciaire.

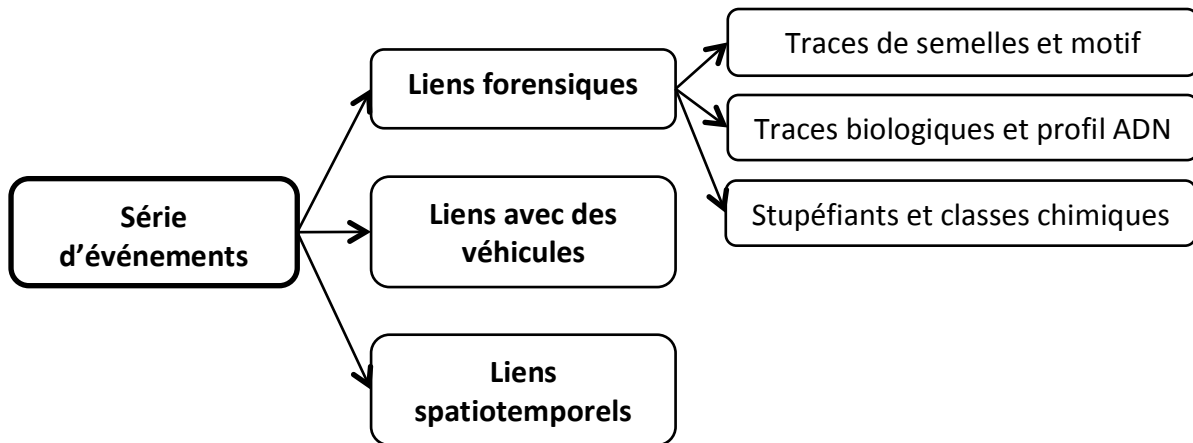
⁵² Cf. Annexe 1: Dimensions d'analyse et visualisation, page 233

⁵³ Cf. Analyse et visualisation relationnelle, page 43

Concepts à représenter

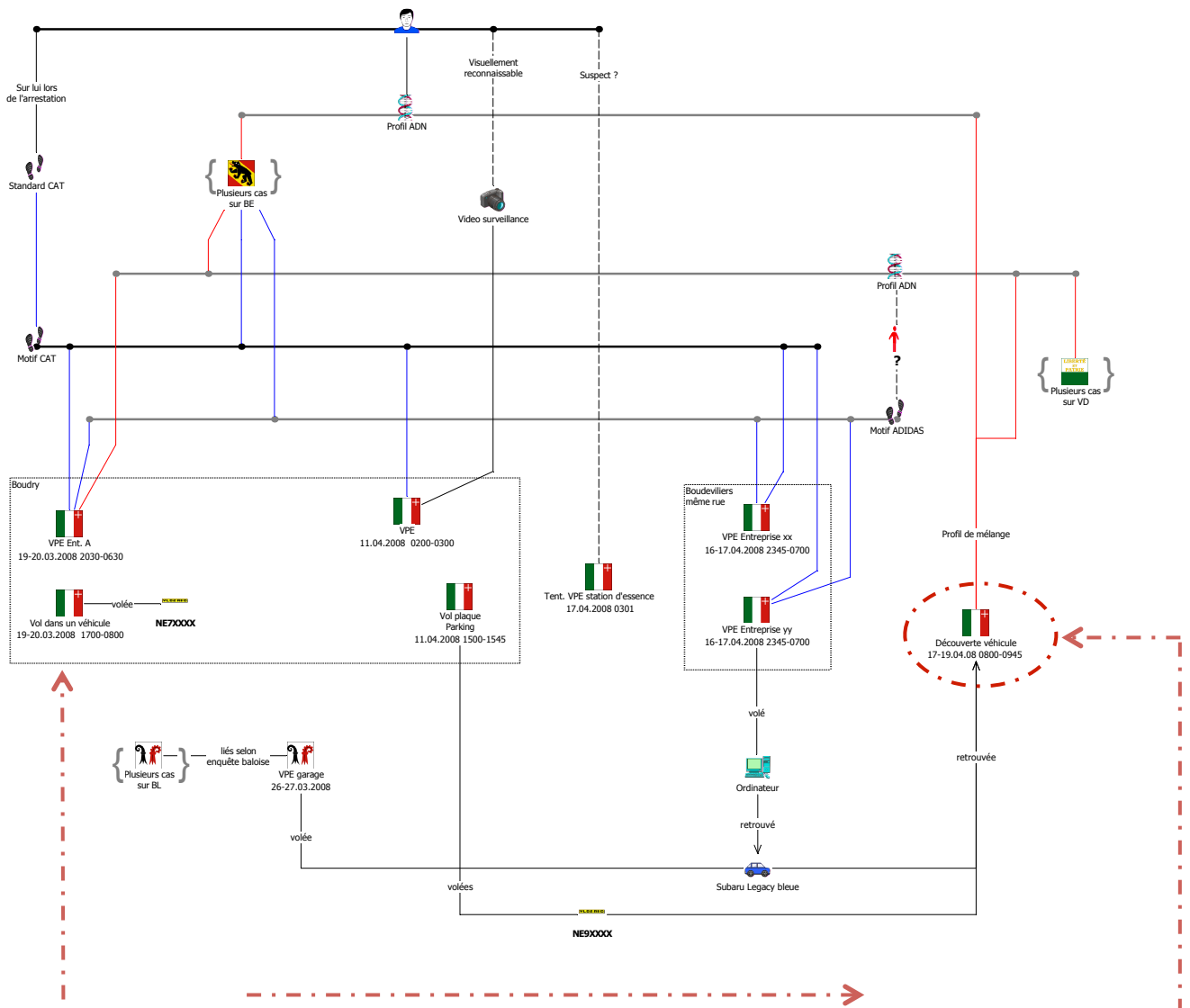
Différents types de relations entre les cas peuvent être représentés : les relations forensiques, les liens spatio-temporelles, les liens avec des véhicules volés et retrouvés lors de la commission des cas ou à proximité ou des objets (tels que le butin des vols. Chaque type de relation doit être clairement distinguable et les incertitudes associées à chaque lien (par exemple, lorsque les traces sont partielles) doivent également être représentées afin de pouvoir évaluer la fiabilité des relations.

Les principaux types de liens sont décrits dans les sections suivantes. La manière de représenter chacun d'eux est présentée et discutée dans des sections séparées.



Exemple de cas

L'exemple ci-dessous intègre, au sein d'une même affaire, l'ensemble des types de relations fréquemment observés lors de l'analyse d'une série de cambriolages.



Positionnement des événements le long de l'axe horizontal pour représenter la **chronologie** (relations représentées selon le principe de proximité)

Représentation par un cadre des **relations spatiotemporelles** (relations représentées selon le principe de connectivité)

Représentation du **type d'événement** par une icône (ou ici du canton, par exemple) (relations représentées selon le principe de similarité)

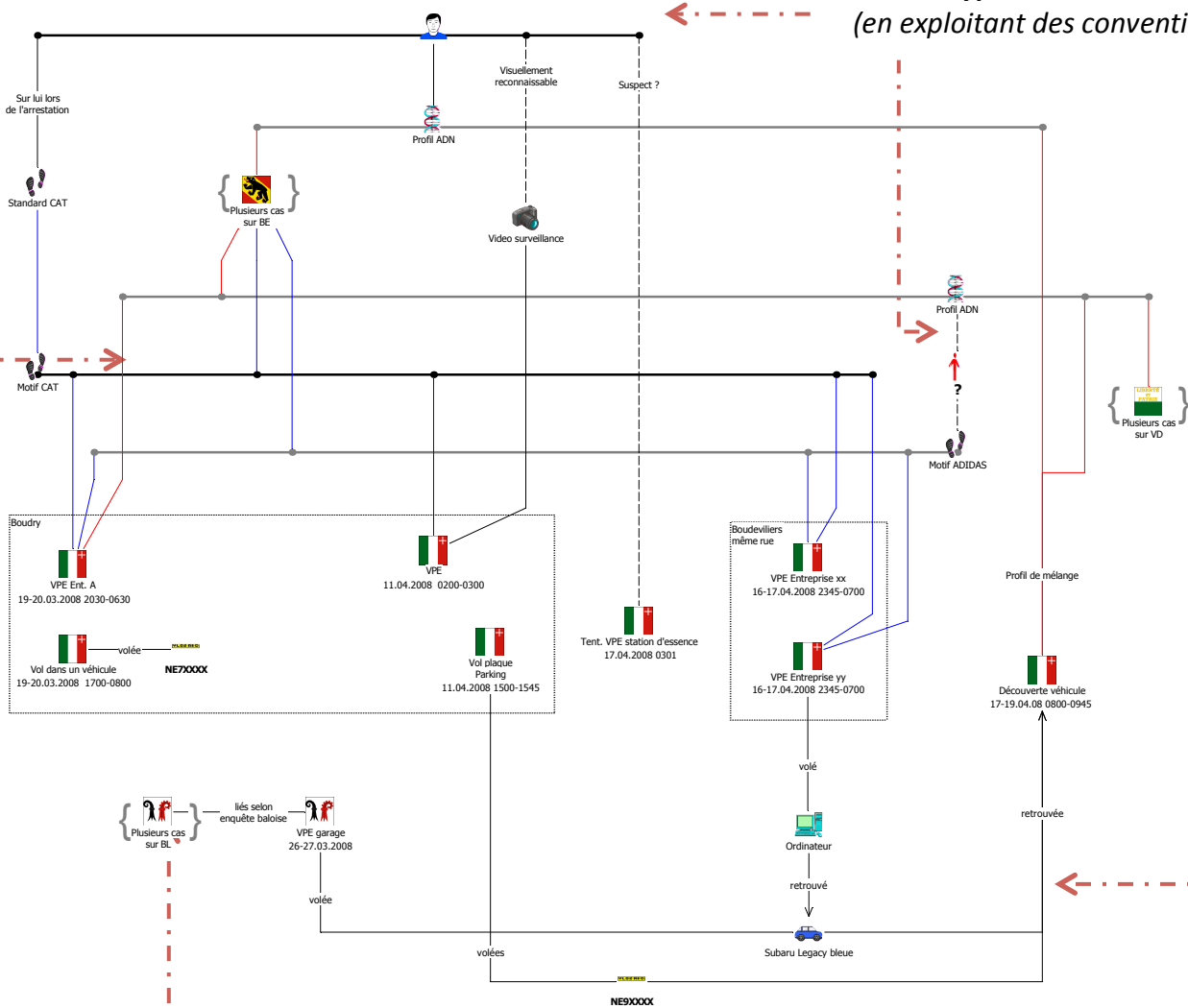
Distinction des **types de relations** par des couleurs

(bleu : motif de semelle / rouge : profil adn)

(représentation des types selon le principe de similarité)

Représentation de **l'incertitude**
ou des **hypothèses**

(en exploitant des conventions)



Représentation de **sous-groupes d'événements** par des ensembles

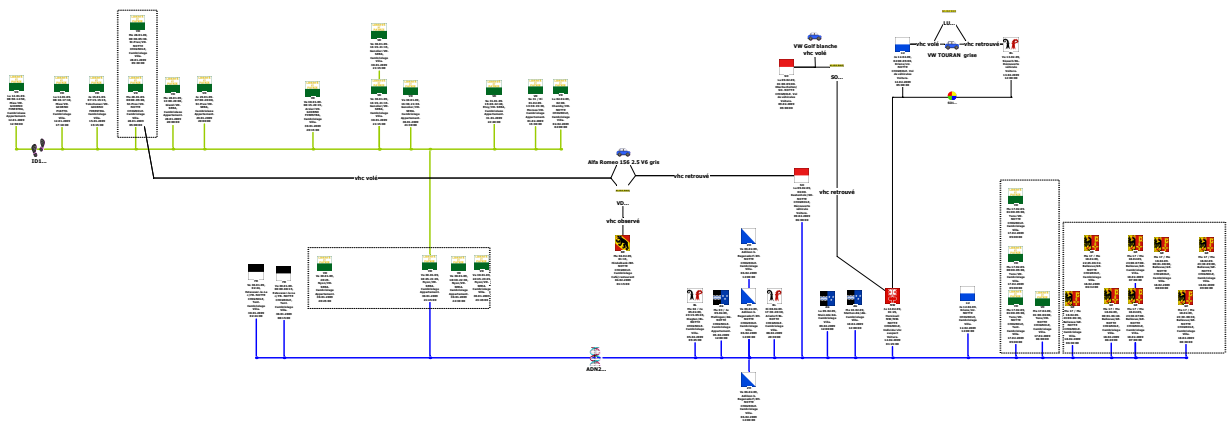
(distinction avec les événements uniques)

Les relations avec des **véhicules et plaques** sont représentées par des liens

multiples (relations représentées selon le principe de connectivité)

Sur ce schéma, six types d'entité sont représentées : des événements, des personnes, des profils (les motifs de semelles et les profils ADN), des véhicules, des plaques et des objets. Ces types sont distingués par des icônes symbolisant les concepts (selon le principe de similarité). Les traces sont visuellement représentées par des liens entre les événements et les profils. Ce choix de représentation est explicité dans les sections suivantes. La représentation est volontairement simplifiée (au niveau des labels des entités et des liens) afin de faciliter sa lecture et mettre en évidence sa structure.

De tels schémas peuvent également être exploités pour représenter des séries plus importantes, voire mettre en évidence des relations particulières entre plusieurs séries. Dans le cas présenté ci-dessous, un cas vaudois et un véhicule permettent de relier deux séries initialement perçues comme séparées: une série détectée par des traces de semelles (en vert) et une série détectée par un profil ADN (en bleu).



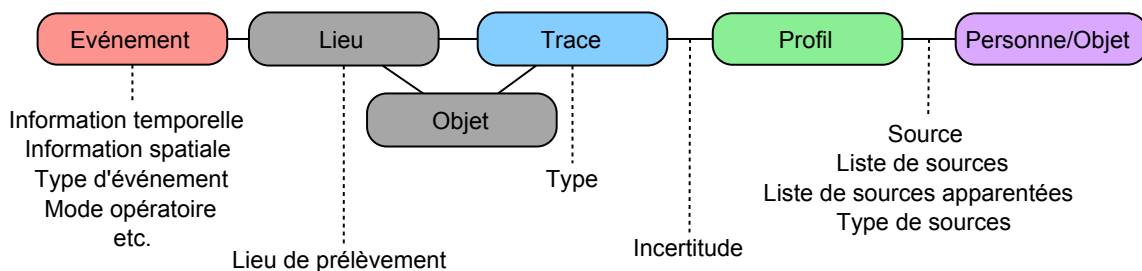
R e p r é s e n t a t i o n d e s l i e n s f o r e n s i q u e s

La trace physique offre de nombreuses perspectives pour détecter des relations entre des événements. De manière générale, différents types de liens peuvent être établis sur la base des traces : un lien avec la source de la trace (une personne ou un objet), des liens vers plusieurs sources (notamment lorsque la trace est partielle), voire des sources apparentées dans le cas de l'ADN, ainsi que des liens établis sur la base du type de source, comme le motif d'une semelle (Ribaux & Margot, 1999). L'ensemble de ces relations se basent sur la détection d'une correspondance de profils (un ensemble de caractéristiques observées et mesurées⁵⁴). Par exemple, un ensemble de minuties concordantes (le profil) est observé entre une trace digitale relevée sur une scène de crime et l'empreinte d'un suspect.

C o n c e p t s à r e p r é s e n t e r

Dans le cadre de l'analyse d'une série, il est nécessaire de pouvoir représenter pour chaque trace :

- l'endroit de prélèvement de la trace (sur les lieux ou sur un objet) afin de pouvoir évaluer la pertinence de la trace par rapport à l'activité
- le type de trace afin de connaître la nature de la relation
- l'incertitude entre la trace et le profil sur la base duquel des relations sont établies
- le type de relation entre le profil et la ou les sources possibles.

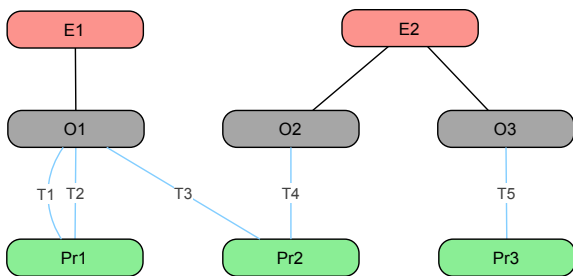
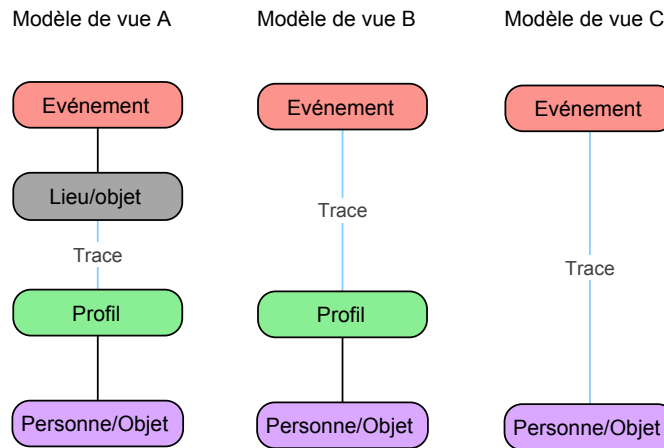


R e p r é s e n t a t i o n d e s r e l a t i o n s

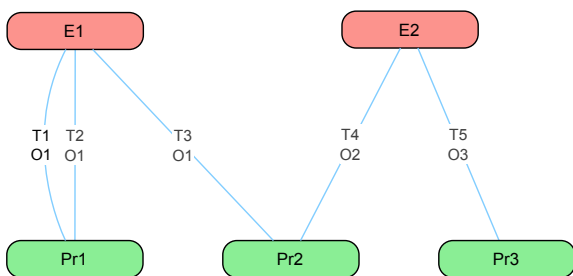
Chaque trace est prélevée sur un lieu ou un objet particulier. Elle n'est donc liée qu'à une seule entité lieu ou objet. Dans la majeure partie des cas, la trace n'est également liée qu'à un seul profil (des exceptions sont présentées plus loin). En regard de la typologie des relations définies⁵⁵, la trace peut donc être représentée par un lien :

⁵⁴ Cf. Analyse, profil et pattern, page 26

⁵⁵ Cf. Typologie de relations, page 102

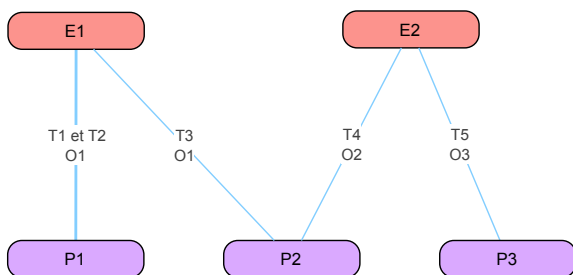


Modèle A : la trace est liée à son support. Ce modèle est utile lorsque pour un même objet ou un même lieu plusieurs traces ont été relevées. Les relations spatiales (lieux de prélèvement) sont alors mises en évidence et représentées selon le principe de connectivité.



Modèle B : la trace est directement liée à l'événement. Selon l'objectif, il est possible de s'abstraire du support à des fins de simplification. Le lieu de prélèvement est représenté sous la forme d'un attribut de la trace (un texte sur le lien). Ce modèle est recommandé lorsque chaque trace a été

prélevée à un endroit ou sur un support différent. Si plusieurs traces ont été prélevées sur le même support, il est toutefois possible de conserver cette structure. En effet, les relations spatiales entre les traces peuvent être représentées à l'aide de la couleur du texte.

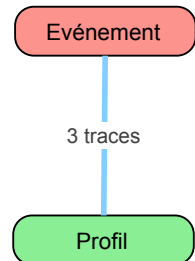


Modèle C : la trace est directement liée à sa source (une personne ou un objet) lorsque le profil correspond à une seule personne ou à un seul objet. Il n'est cependant pas recommandé de fusionner des entités objets et personnes afin de ne pas cacher l'indirection. En effet, une trace de semelle peut lier

une chaussure à l'endroit du prélèvement, mais certainement pas une personne. Le lien entre la personne et la chaussure peut également être sujet à discussion. Il est donc nécessaire de le représenter pour ne pas minimiser une possible incertitude. Celle-ci peut alors être représentée sur le lien entre l'objet et la personne.

Lorsque la trace est représentée sous la forme d'un lien :

- la couleur du lien peut représenter le type de trace
- le style du lien permet d'exprimer l'incertitude de la relation (trait-tillé pour un profil partiel, p. ex.)
- Par agrégation, plusieurs traces liées à un même événement et à un même profil peuvent être regroupées sous la forme d'un seul lien. L'épaisseur du lien est alors exploitée pour représenter le nombre de traces. Une telle agrégation simplifie le schéma en réduisant le nombre de liens.

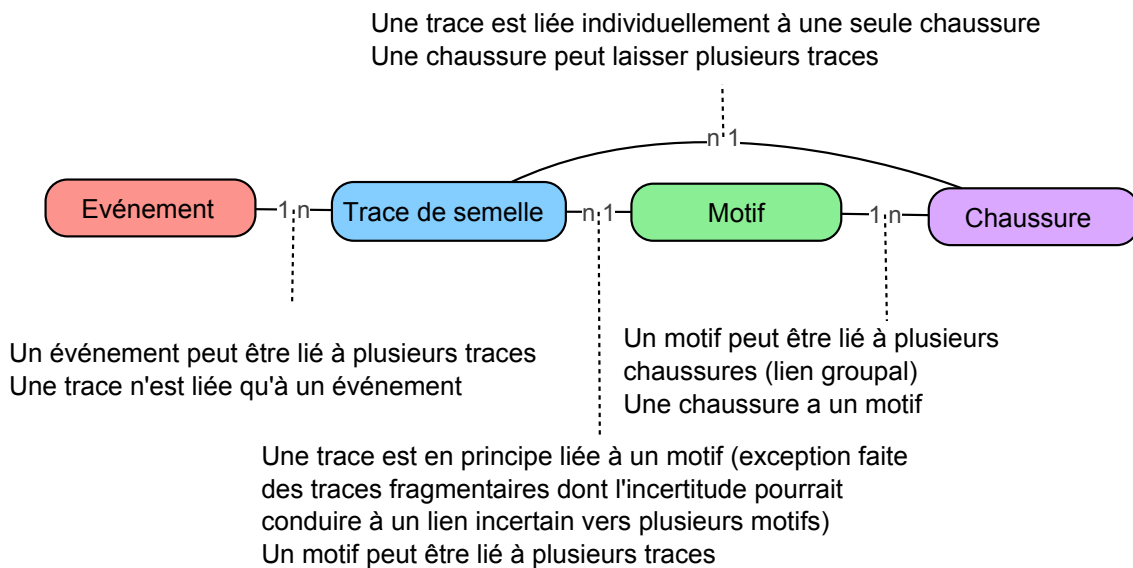


Dans les sections suivantes, la représentation de différents types de trace est décrite. Certains cas particuliers, où l'utilisation de ce modèle est discutable voire n'est pas souhaitable, sont également présentés.

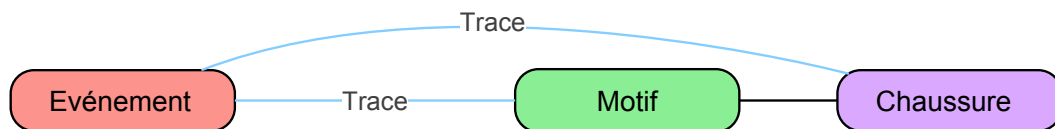
Traces de semelles et motifs

Des traces de semelles sont régulièrement recherchées et relevées sur des scènes de crimes afin de reconstituer des cheminements et comprendre l'activité, d'estimer le nombre de personnes impliquées, voire de guider la collecte d'autres traces. En suisse plus particulièrement, ces traces sont recherchées en routine sur des cas de cambriolages notamment en raison de leur potentiel pour lier des cas. En effet, il n'est pas rare que des auteurs sériels conservent leurs chaussures sur des périodes plus ou moins importantes. Des relations sont alors détectées entre les cas sur la base des motifs de semelle et peuvent contribuer à détecter des séries particulières (Girod et al., 2008). L'utilisation de représentations spatiotemporelles facilite notamment la détection de regroupements inhabituels, signes d'une possible activité sérielle (Ribaux et al., 2003). Sur un schéma relationnel, ces relations sont misent en perspective avec l'ensemble des relations détectées, afin de tester l'hypothèse de la série.

Concepts à représenter



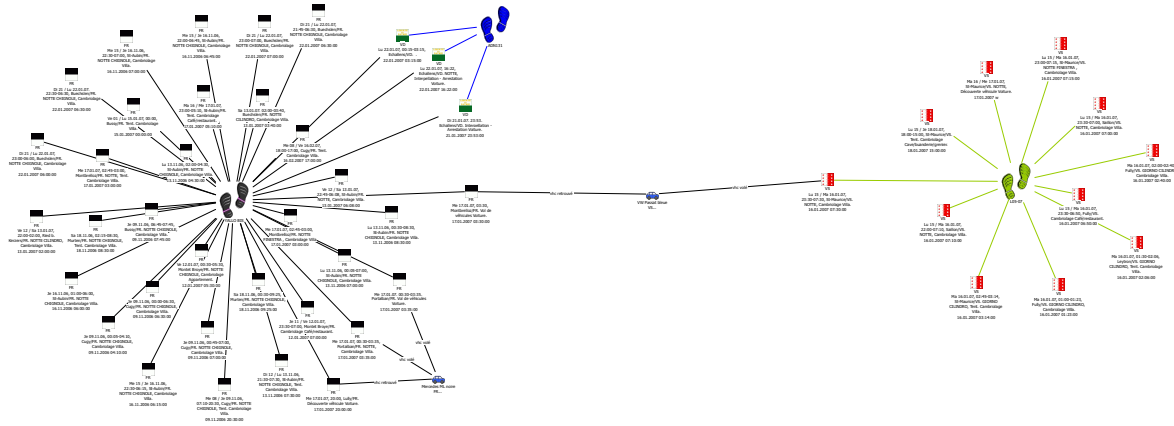
Représentation des relations



Lorsque la qualité de la trace est suffisante, il est possible d'identifier le motif correspondant au dessin de la semelle. La trace peut alors être représentée par un lien, puisqu'elle ne peut provenir que d'un événement et qu'une semelle n'a qu'un seul motif. Le lien établi est groupal. Il se base sur le type de trace. La comparaison des caractéristiques acquises des traces peuvent conduire en inférer une source. La relation est alors établie avec la source (une chaussure) (Girod et al., 2008). Dans ce cas de figure également la trace peut être représentée par un lien.

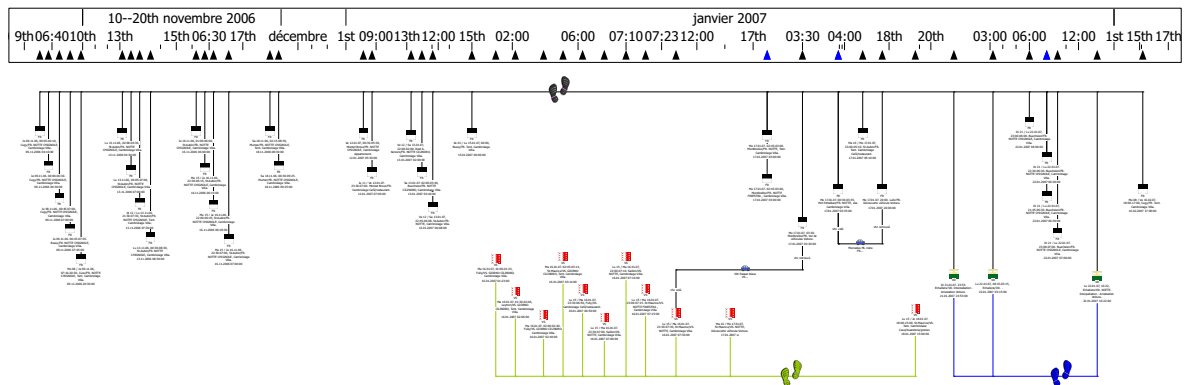
Exemple de cas

Schéma relationnel



Ce schéma tiré d'une affaire réelle représente les relations établies sur la base de trois motifs de traces de semelles. Deux séries supposées sur la base des motifs sont liées par un véhicule volé lors d'un cas en Valais et retrouvé à proximité d'un cas à Fribourg. Sur les cas vaudois un troisième motif (en bleu) a également été retrouvé.

Schéma temporel

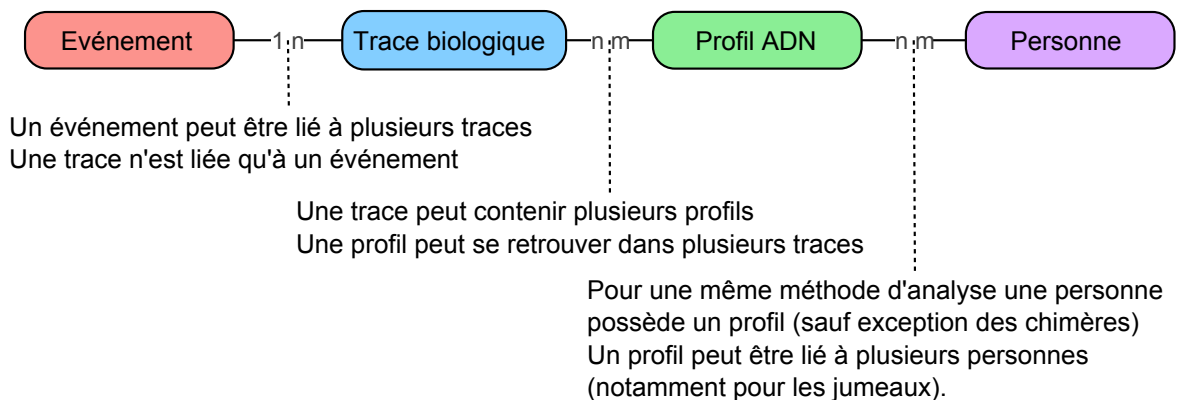


La série peut également être représentée chronologiquement. Chaque cas est positionné sur l'axe horizontal en fonction de sa date. Les relations établies sur la base du motif sont représentées par des lignes horizontales liées aux cas.

Traces biologiques et profil ADN

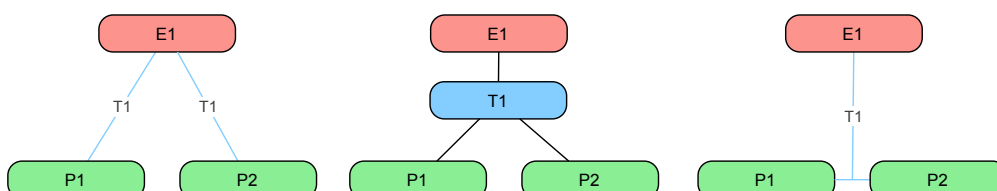
Les traces biologiques, telles que des traces de sang et de résidus de peau, sont très largement exploitées pour lier des cas sur la base du profil extrait de l'ADN. Contrairement aux systèmes automatiques exploités pour la gestion des traces digitales, les bases de données stockant les profils ADN issus de traces permettent la détection automatisée de relations entre cas (indépendamment de l'identification du profil). Deux traces biologiques sont liées lorsque leurs profils sont concordants (sur la base de la comparaison des locus analysés).

Concepts à représenter



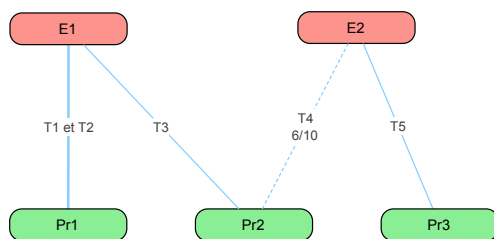
Représentation des relations

Une trace biologique ne peut être liée qu'à un seul événement. Par contre, il arrive parfois que certaines traces contiennent plusieurs profils (par exemple lors de prélèvements sur des victimes ou lorsque plusieurs personnes ont touché un même objet). Dans ce cas, la trace ne saurait être représentée par un lien, puisqu'elle peut être liée à plusieurs profils. Trois solutions graphiques sont alors exploitables :



La première solution a le désavantage de ne pas mettre en évidence visuellement le fait qu'il ne s'agit du même spécimen collecté⁵⁶. Elle a cependant l'avantage de conserver un modèle de représentation constant pour toutes les traces. Le deuxième modèle engendre une démultiplication du nombre d'entités, notamment si l'ensemble des traces sont représentées sous la forme d'entités pour conserver un modèle commun sur la globalité du schéma. La troisième solution permet de conserver une structure comparable par rapport aux traces liées à un seul profil tout en évitant de représenter la même trace deux fois (par deux liens).

Des liens sont également établis sur la base de profils dits partiels issus de traces dont l'ensemble des loci n'ont pas pu être détectés. Des relations plus ou moins certaines peuvent donc être établies en fonction du nombre de loci correspondants. Visuellement, il est possible



de distinguer les liens établis sur la base de profils complets et partiels en exploitant la convention du trait-tillé pour les liens incertains. Le nombre de loci correspondants peut alors être indiqué sous la forme d'un attribut de la trace.

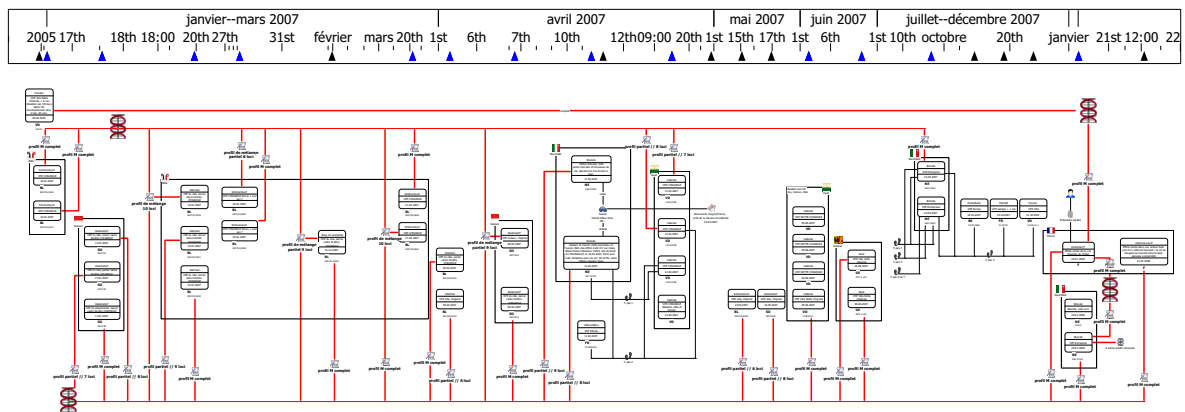
Simplifier la représentation en regroupant le profil et la personne n'est pas souhaitable. En effet, un profil ADN similaire peut être détecté pour plusieurs personnes différentes (notamment dans le cas de jumeaux univitellins, mais également en cas de coïncidences fortuites qui peuvent survenir pour des profils de mélange ou des profils partiels plus particulièrement). Une même personne peut également avoir plusieurs profils (les chimères). Il paraît donc important de ne pas supprimer cette indirection pour ne pas cacher l'incertitude possible entre le profil et la personne.

⁵⁶ Un profil de mélange peut toutefois être considéré comme le résultat de plusieurs traces. Dans cette optique cette structure visuelle explicite la présence de plusieurs traces (vestiges des activités de plusieurs personnes).

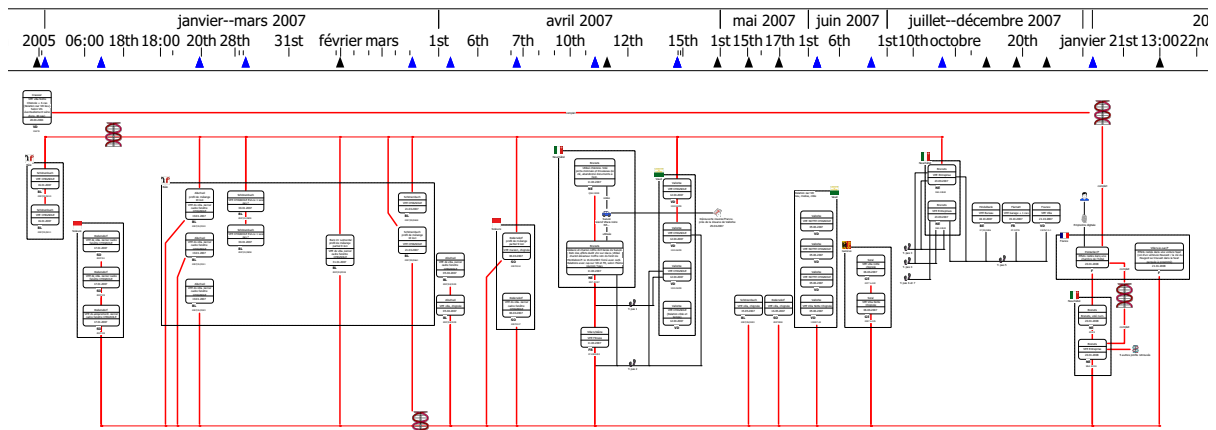
Exemple de cas

La série représentée ci-dessous a été détectée sur la base de trois profils ADN. Le premier schéma représente les traces par des entités liées aux cas et aux profils. Le second exploite la simplification proposée.

Représentation de la trace par une entité



Représentation de la trace par un lien

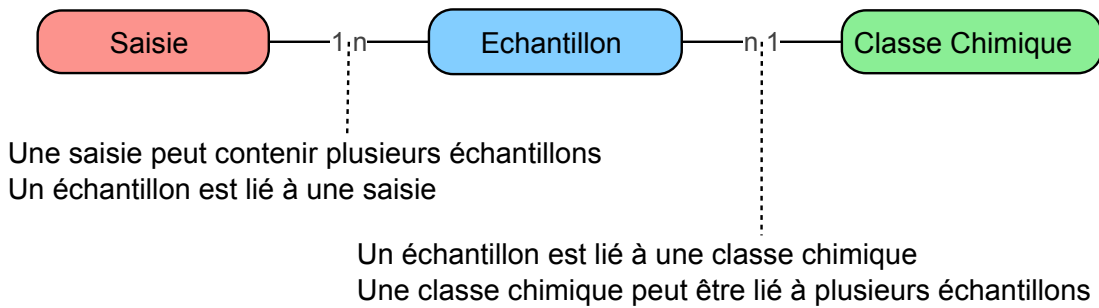


Saisies de stupéfiant et classes chimiques

L'établissement de liens entre des échantillons de stupéfiants saisis peut être réalisé par différents moyens, tels que la comparaison des propriétés physiques (formes, tailles ou couleurs de pilules par exemple) ou l'analyse chimique du contenu des stupéfiants (tels que des principes actifs, des impuretés ou des produits de coupage). Les liens détectés à partir de la composition chimique du stupéfiant se basent sur l'hypothèse que des variations importantes sont observables entre différents lots et qu'il est possible de définir des profils dont la spécificité permet d'inférer que deux échantillons dont les profils sont similaires proviennent d'une même masse (Guéniat & Esseiva, 2005).

La détermination du profil chimique d'un échantillon de stupéfiants permet d'établir des liens entre les saisies et par là-même de détecter des groupes de cas. Les relations issues du profilage peuvent alors être interprétées en regard des données circonstancielle de l'enquête et permettent ainsi de reconstruire l'activité de suspects.

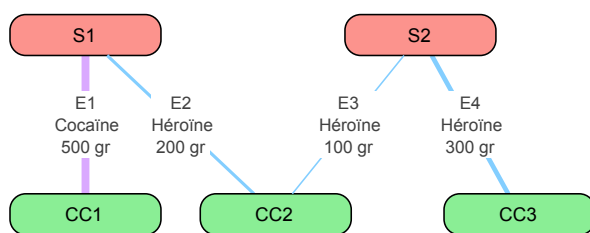
Concepts à représenter



Représentation des relations



Un échantillon de stupéfiants ne peut être lié qu'à un seul événement et à une seule classe chimique (par définition de la classe chimique). Il peut donc être représenté par un lien.



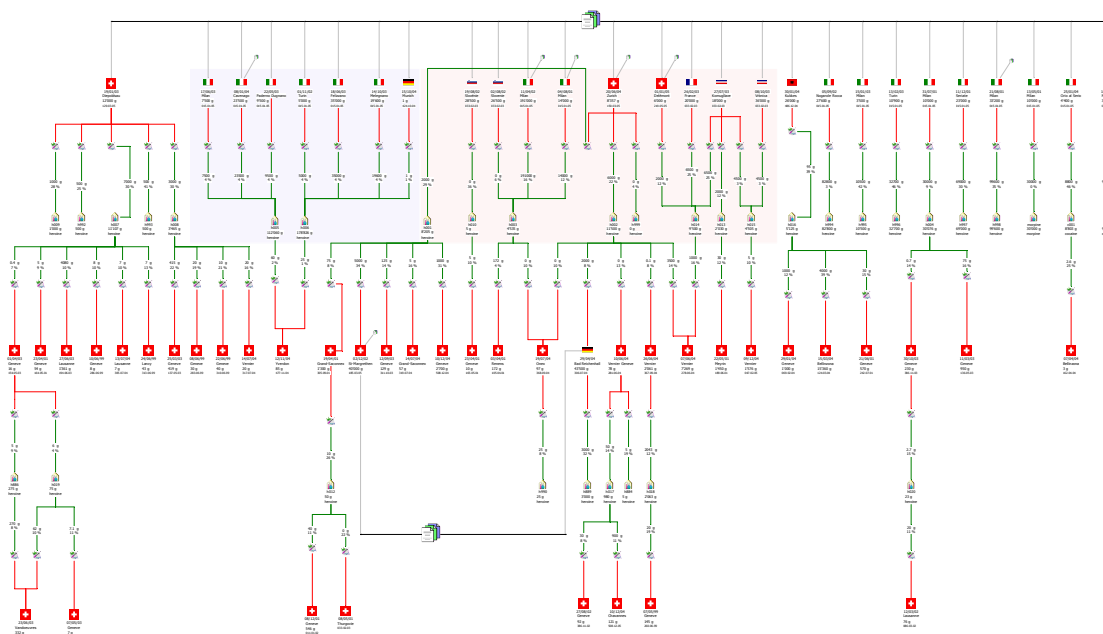
Représenter l'échantillon sous la forme d'un lien permet d'exploiter l'épaisseur du trait pour représenter la masse de l'échantillon (information utile d'un point de vue de la qualification pénale, puisque la distinction entre le cas grave et le cas simple se base sur la quantité de stupéfiant⁵⁷). De plus, le type de stupéfiant peut être distingué par la couleur (selon le principe de similarité).

⁵⁷ ATF 109 IV 143 (Art. 19 ch. 2 lettre a LStup : dès 12g d'héroïne, 18g de cocaïne, 4 kg de haschich ou 200 trips de LSD) et ATF 113 IV 32 (dès 36g d'amphétamines).

Exemple de cas

Le cas présenté est tiré d'une affaire réelle ayant une portée internationale. L'objectif du schéma était de représenter l'ensemble des relations établies par le profilage entre les différentes saisies. Le schéma initialement produit a été élaboré en décomposant les informations en trois entités principales : les saisies, les échantillons et les classes chimiques. Le second schéma illustre la simplification apportée en représentant les échantillons par des liens. Ce cas a été utilisé lors de l'évaluation de la représentation des relations⁵⁸. Pour rappel, le changement de structure proposé a engendré une augmentation du taux de réponses correctes de 8.9% et une réduction du temps de réponse moyen de 27.3% pour identifier la classe chimique liée au plus grand nombre de saisies⁵⁹.

Représentation des échantillons par des entités

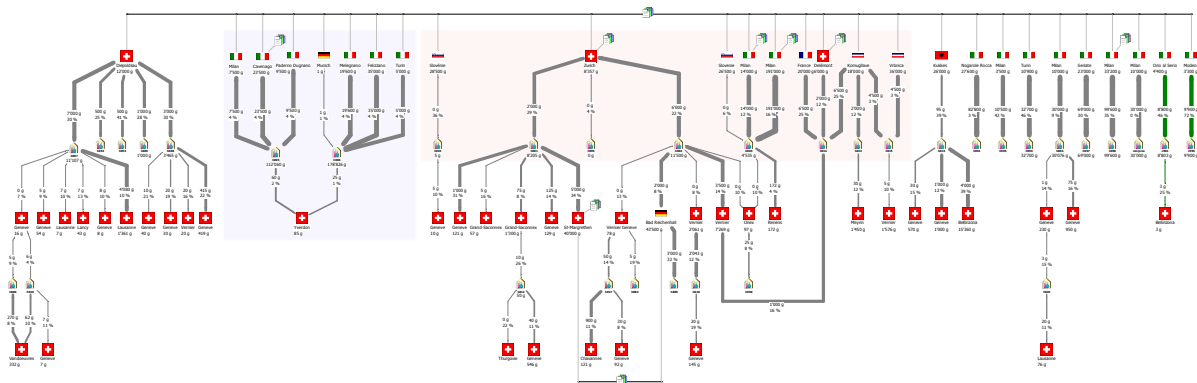


En raison du nombre important d'échantillons, tous ceux faisant partie d'une même saisie et liés à la même classe chimique ont été regroupés et représentés par une seule icône. Des drapeaux ont été utilisés comme icônes des saisies afin de mettre en évidence les liens géographiques.

⁵⁸ Cf. Représentation d'une entité sous la forme d'un lien, page 112

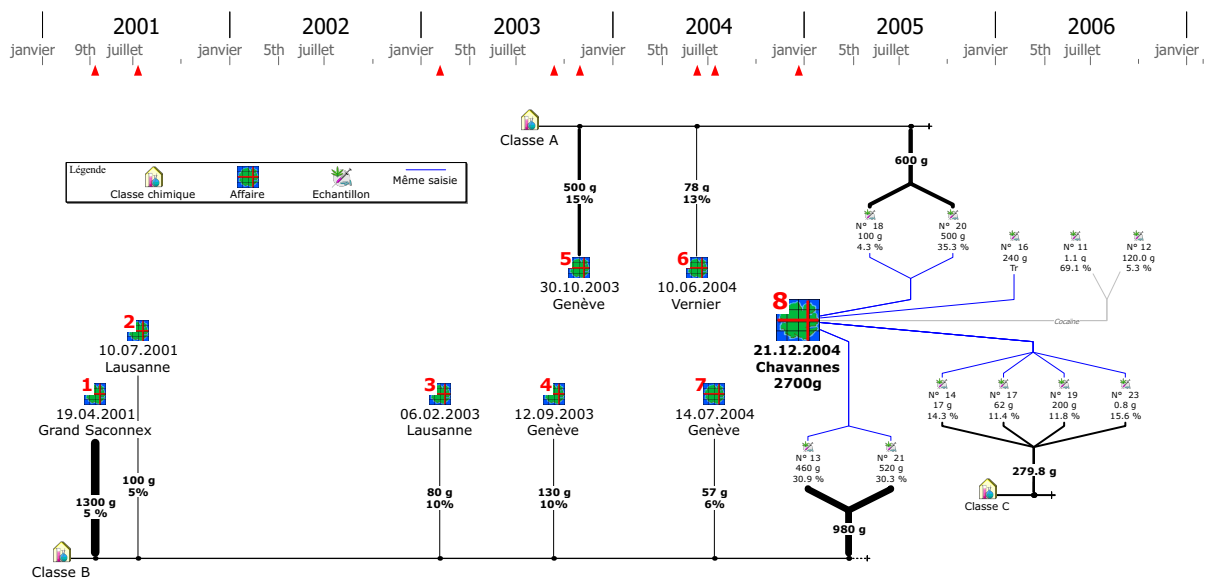
⁵⁹ Dans l'évaluation: le rond rouge ayant le plus de liens directs avec des carrés noirs

Représentation des échantillons par des liens entre les saisies et les classes chimiques



Les échantillons sont représentés par des liens entre les saisies et les classes chimiques. L'épaisseur des liens est proportionnelle à la masse de stupéfiant saisie. La couleur du lien décrit le type de stupéfiant (gris pour l'héroïne et vert pour la cocaïne). La quantité et la pureté sont affichées sur le lien. Le nombre d'échantillons ainsi que leurs numéros pourraient également y être inscrit. L'espace utilisé pour représenter la même information est réduit. Les relations entre les saisies sont mises en évidence et des groupes de saisies apparaissent de manière plus évidente. Un repositionnement global des icônes améliorerait encore cet aspect. De nouvelles informations apparaissent plus rapidement, telles que le type et la masse des échantillons (ici regroupés par classe chimique). La répartition géographique des échantillons est plus facilement comparable (essentiellement grâce au nombre réduit d'icônes).

L'exemple présenté ci-dessous a été effectué pour représenter les relations établies avec une saisie particulière (l'affaire numéro huit sur le schéma) :

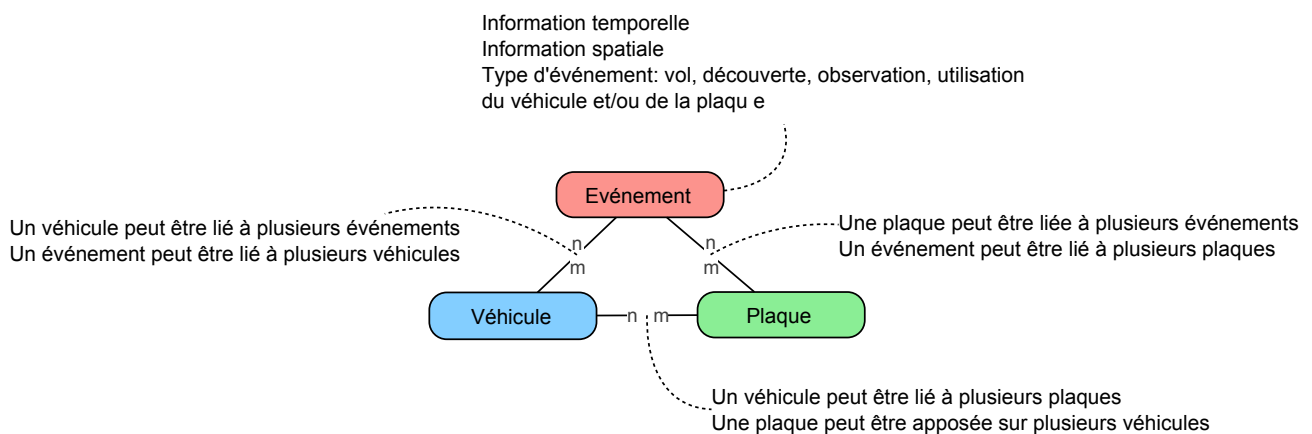


Sur ce schéma les deux structures ont été exploitées afin de représenter deux niveaux de détails différents. Pour l'affaire en cours, l'ensemble des échantillons a été représenté par des entités. Les échantillons analysés pour les saisies liées sont quant à eux représentés de manière simplifiée. L'objectif ici était de faciliter la communication des résultats d'analyse. Chaque échantillon est représenté pour l'affaire en cours afin que les enquêteurs puissent mettre en relation ces informations avec les autres éléments de l'enquête (notamment les données issues des auditions).

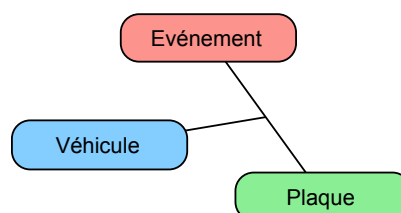
R e l a t i o n s a v e c d e s v é h i c u l e s

Lors de la commission de délits, il n'est pas rare que les auteurs volent des véhicules pour se déplacer entre des lieux de cambriolages ou pour transporter leurs butins par exemple. Certains véhicules sont volés directement lors de cambriolage (par exemple en volant les clés du véhicule), tandis que d'autres sont parfois retrouvés à proximité de cambriolages. Ces véhicules permettent donc de mettre en évidence des relations entre différents cas de cambriolages. De plus, les auteurs changent souvent les plaques d'immatriculations des véhicules afin d'éviter les contrôles et d'exploiter les véhicules sur de plus longues périodes.

C o n c e p t s à r e p r é s e n t e r

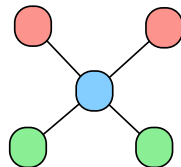
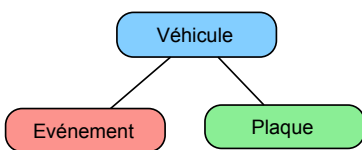


R e p r é s e n t a t i o n d e s r e l a t i o n s



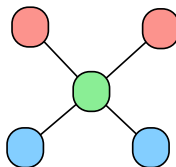
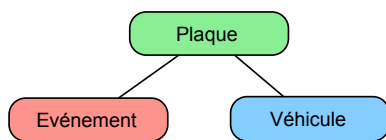
Les relations entre les événements, les véhicules et les plaques d'immatriculation sont représentées par des liens multiples. Bien que diverses structures de graphe soient envisageables, le lien multiple est la seule représentation qui permet de représenter ces relations efficacement et sans ambiguïté.

Les structures basées sur l'utilisation de liens simples (entre deux entités) sont présentées ci-dessous afin d'expliquer les difficultés de représentation :



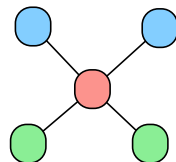
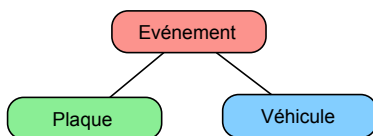
Lorsqu'un véhicule est lié à plusieurs événements et plaques, il n'est plus possible de déterminer quelle plaque était apposée sur le véhicule lors de chaque événement.

Exemple : un véhicule est volé avec une plaque, puis retrouvé avec une autre.



Lorsqu'une plaque est liée à plusieurs événements et véhicules, il n'est plus possible de déterminer quel véhicule est lié à chaque événement.

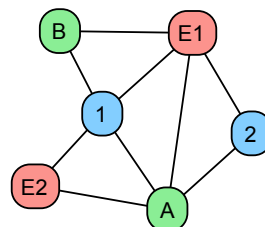
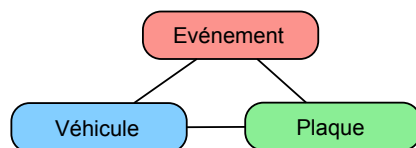
Exemple : une plaque est volée sur un véhicule et retrouvée sur un autre.



Lorsqu'un événement est lié à plusieurs véhicules et plaques, il n'est plus possible de déterminer quelle plaque est apposée sur quel véhicule.

Exemple : plusieurs véhicules sont volés, contrôlés ou retrouvés en même temps.

Afin d'éviter les ambiguïtés présentées ci-dessus, les trois types d'entités peuvent être liées par trois liens :



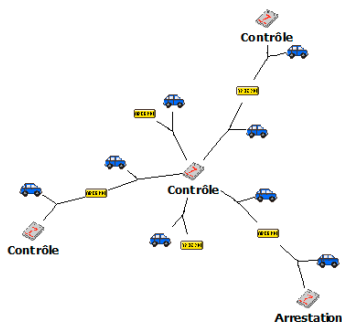
Certaines situations peuvent toutefois porter à confusion. Dans cet exemple, deux véhicules (1 et 2) sont liés à un événement (E1). Pour déterminer à quel véhicule les plaques A sont liées

lors de l'événement E1, il est nécessaire de se baser sur les relations des plaques B. Si le véhicule 1 n'avait pas de plaques lors de l'événement E1, il ne serait pas possible de le voir visuellement. Dans cette situation, la représentation est donc ambiguë. De telles relations nécessitent donc une forme adaptée de représentation : les liens multiples. Il faut relever que la plupart des logiciels utilisés actuellement ne permettent pas de créer de telles relations. Lier l'ensemble des entités est alors recommandé en prenant garde aux situations particulières présentées ci-dessus.

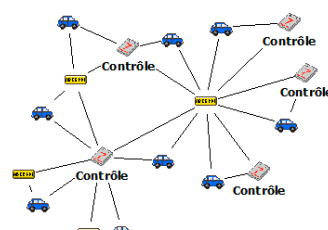
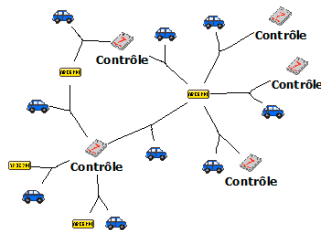
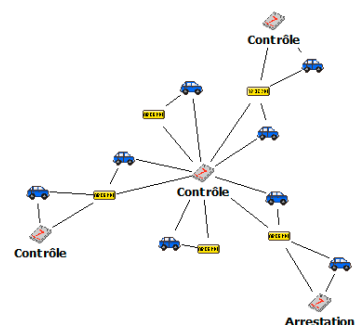
Exemple de cas

Les schémas ci-dessous représentent des relations issues d'une banque de données de police. Le premier schéma a été réalisé en utilisant des liens multiples. Le second schéma représente les mêmes relations avec des liens simples. L'efficacité des liens multiples par rapport aux liens simples n'a pas été testée dans le cadre de l'évaluation présentée dans le chapitre précédent, puisque l'utilisation de liens simples peut engendrer des ambiguïtés. L'intégrité de la représentation n'est donc pas garantie en exploitant des liens simples. Visuellement, l'utilisation des liens multiples semble rendre la représentation plus lisible. Cet aspect pourrait également être évalué formellement.

Représentation des relations avec des liens multiples

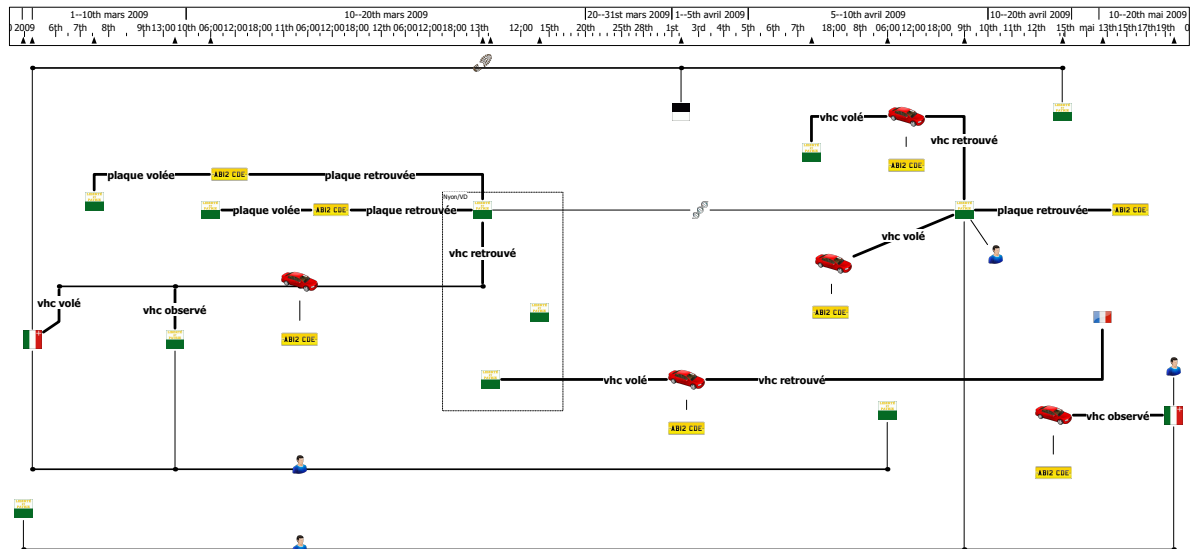


Représentation des relations avec des liens simples (entre deux entités)



162 Patterns de visualisation relationnelle

Dans la majeure partie des séries analysées, les relations établies sur la base de véhicules volés, retrouvés ou observés sont plus simples. Les situations où l'utilisation de liens multiples est nécessaire pour éviter de l'ambiguïté sont assez rares. Il est donc possible d'exploiter des liens simples (entre deux entités) pour représenter les relations :

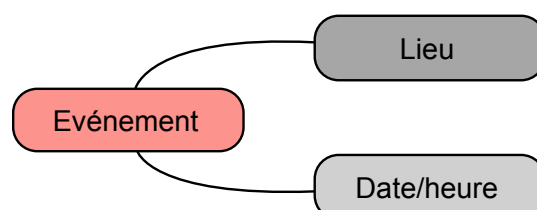


Les relations avec des véhicules et des plaques ont été volontairement mises en évidence pour cet exemple (contrairement au schéma transmis dans le cadre de l'affaire).

Relations spatio-temporelles

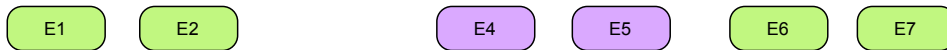
Lors de la représentation d'une série d'événements, la proximité spatio-temporelle entre les événements constitue également une information pertinente. En effet, plusieurs cas ne peuvent avoir été commis par un même auteur ou un même groupe d'auteurs que si les distances spatio-temporelles entre les cas sont compatibles avec les temps de déplacement nécessaires pour parcourir les distances entre les cas.

Concepts à représenter

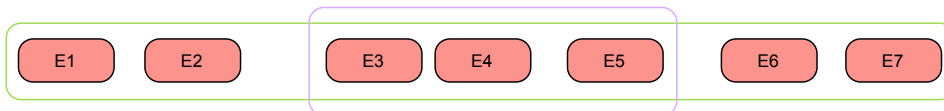


R e p r é s e n t a t i o n d e s r e l a t i o n s

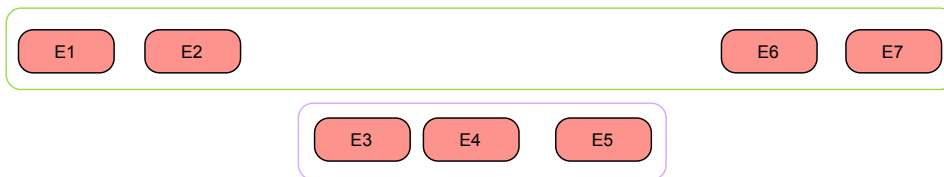
Comme déjà mentionné, il est d'usage de représenter le temps par la position des événements sur l'axe horizontal. Les relations spatiales peuvent être représentées de diverses manières : par des entités spécifiques, par les icônes ou les couleurs des événements ou encore par des cadres. Dans l'exemple de la représentation des relations entre des saisies de stupéfiant présenté précédemment, des icônes sont utilisées pour distinguer des pays ou des cantons. Les relations spatiales peuvent également être représentées par la couleur :



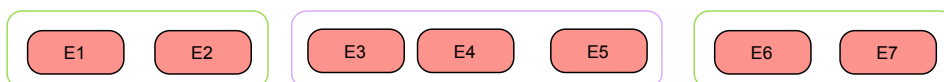
Des cadres peuvent également être exploités. Toutefois, lorsque de nombreux événements sont représentés, il n'est pas rare que des événements aient lieu dans les mêmes régions sur plusieurs périodes distinctes. Des croisements de cadres peuvent alors être créés :



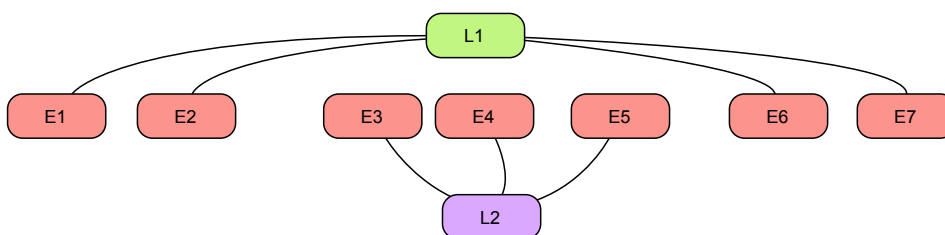
Les croisements peuvent être évités en exploitant le positionnement sur l'axe vertical :



Une modification de placements sur l'axe vertical n'est cependant pas toujours suffisante pour éviter tous les croisements de traits (notamment en raison des autres relations représentées entre les cas). Dans une telle situation, les cadres peuvent être divisés et représentent alors une proximité spatio-temporelle :



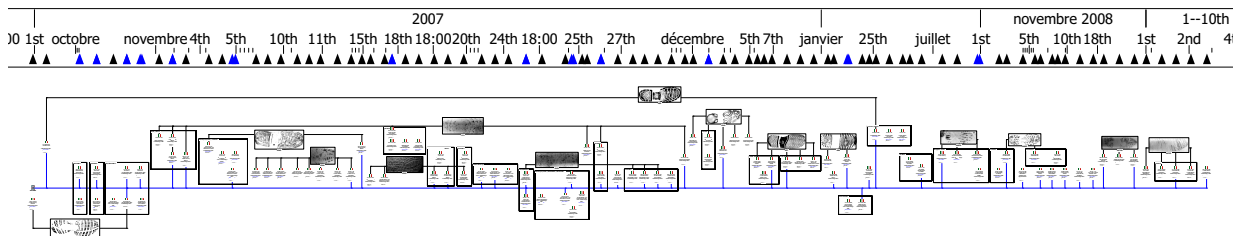
Les relations peuvent aussi être représentées par des entités « lieux » liées à chaque événement :



Afin de minimiser le nombre d'entités et de liens et ainsi réduire les risques de croisements de traits, un tel choix n'est généralement pas adapté. En effet, lors de la représentation de grandes séries (plusieurs dizaines, voire centaines de cas), ajouter des entités et des liens supplémentaires complexifie inutilement la représentation. Exploiter des cadres est, dans ce contexte, particulièrement efficace. Un effet de couches est engendré. Celui-ci permet de distinguer les différents types de relations.

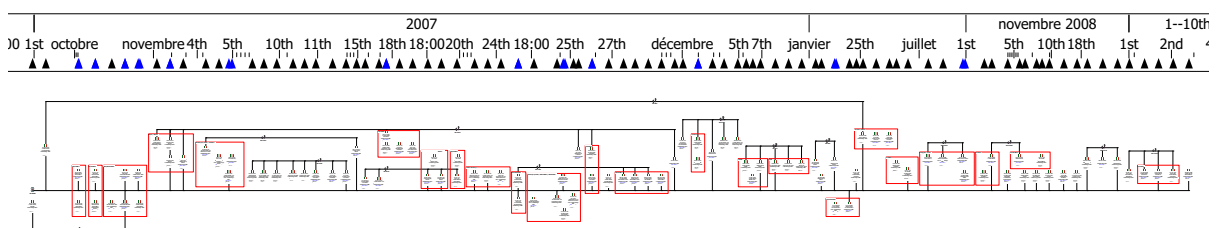
Exemple de cas

La série représentée regroupe plus de cent cas de cambriolages du soir supposés avoir été commis par un même groupe d'auteurs.



La ligne horizontale bleue représente un profil ADN détecté sur plus de cinquante cas. Des relations établies sur la base de motifs de traces de semelle sont également représentées. Les images des motifs ont été insérées pour les représenter.

Les relations établies sur la base des traces ont permis de lier près de quatre-vingt pourcent des cas de la série. Les autres vingt pourcent ont été liés sur la base de proximités spatio-temporelles. Un schéma d'événement a donc été produit. La chronologie est représentée par la position sur l'axe horizontal et la proximité spatiale est visualisée par des cadres. Le schéma ci-dessous met en évidence ces relations qui sont surlignées en rouge :



S é l e c t i o n n e r d e s t r a c e s à a n a l y s e r

Dans le précédent pattern, la représentation visait à soutenir l'analyse des relations entre un ensemble d'événements afin de reconstruire une série. L'événement était l'entité principale de l'analyse et les traces étaient considérées parmi un ensemble de relations, telles que les relations spatio-temporelles, les relations avec des véhicules, du butin, etc. Dans le cadre de la situation présentée ici, la trace est l'entité centrale de l'analyse.

C o n t e x t e

Lors d'infractions graves, comme des brigandages ou des meurtres par exemple, il n'est pas rare qu'un nombre important de traces soient prélevées sur la ou les scènes de crimes, ainsi que sur l'ensemble des lieux ou des objets en relation avec l'affaire. Dans de tels cas, des choix doivent être effectués afin de déterminer quelles traces seront analysées et dans quel ordre de priorité. En effet, l'analyse des traces nécessite un temps plus ou moins important et parfois des coûts élevés. C'est le cas plus particulièrement des analyses ADN qui peuvent prendre plusieurs jours, voire quelques semaines et coûter plusieurs centaines de francs par trace. Des choix doivent être effectués et les priorités sont définies et réévaluées à différents stades de l'enquête. Dans la phase initiale de l'enquête, ces choix sont parfois évidents, notamment lorsque les traces sont prélevées sur des armes, sur des victimes ou sur des voies d'accès. Par contre, lorsque certains résultats sont obtenus et qu'il s'agit de définir quelles sont les autres traces qui seront analysées en regard des informations déjà disponibles, les choix peuvent être plus compliqués.

O b j e c t i f d e l a r e p r é s e n t a t i o n

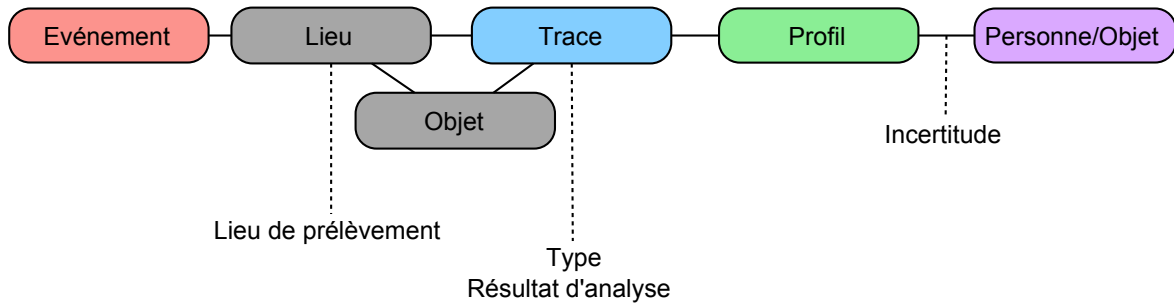
Des schémas relationnels peuvent être produits afin de faciliter et guider la prise de décision pour sélectionner quelles sont les traces les plus pertinentes à analyser.

C o n c e p t s à r e p r é s e n t e r

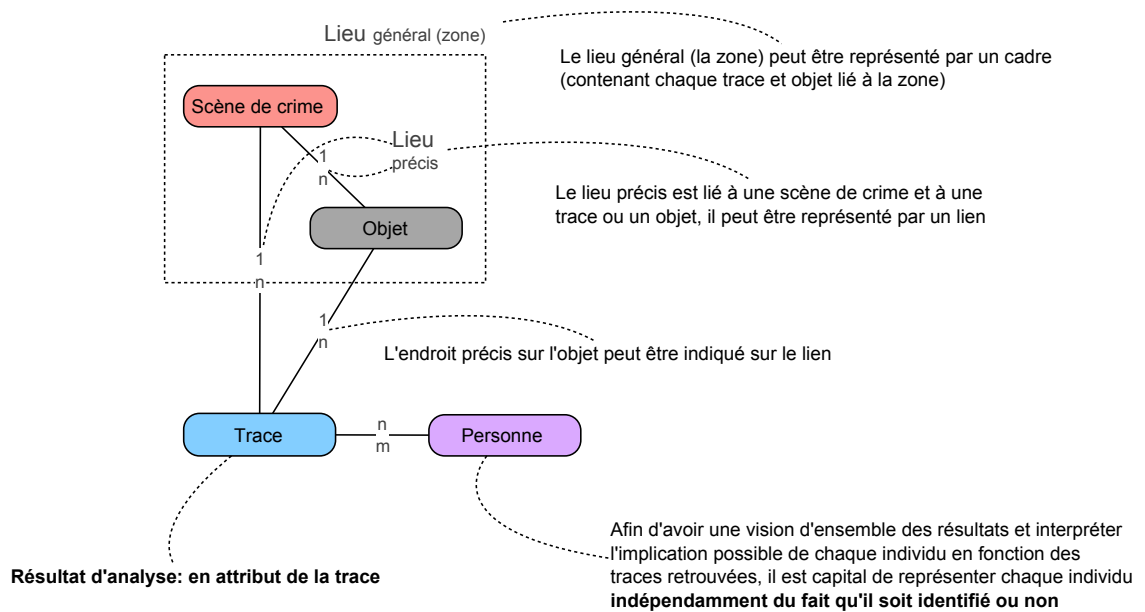
Pour évaluer la pertinence des traces, plusieurs paramètres doivent être considérés en parallèle :

- chaque scène doit être clairement distinguée
- l'endroit de prélèvement de la trace (sur les lieux ou sur un objet) est représenté afin d'évaluer la pertinence en fonction des chances de retrouver une trace. Celles-ci sont évaluées en regard du support des traces (transfert, affinité, etc.) et des éléments circonstanciels sur l'activité (issus notamment des déclarations de témoins).

- le résultat d'analyse doit être inclus. Pour l'analyse ADN, il peut se classifier : pas analysé, négatif, profil établi (complet, partiel, mélange) et profil non-interprétable
- les relations avec des individus (identifié ou non) et le degré de certitude des relations



Représentation des relations



C'est la variable principale de la question d'analyse. Afin de la mettre en évidence, la couleur est utilisée. Le rouge est exploité pour représenter les traces non analysées (celles sur lesquelles il faut se déterminer)

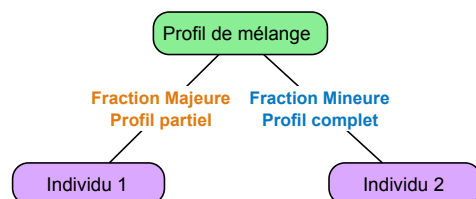
- Non-analysée
- Profil complet
- Profil partiel
- Profil de mélange

- Négatif
- Non-interprétable

Les profils négatifs ou non-interprétable n'ont pas besoin d'être distingués pour l'objectif d'analyse

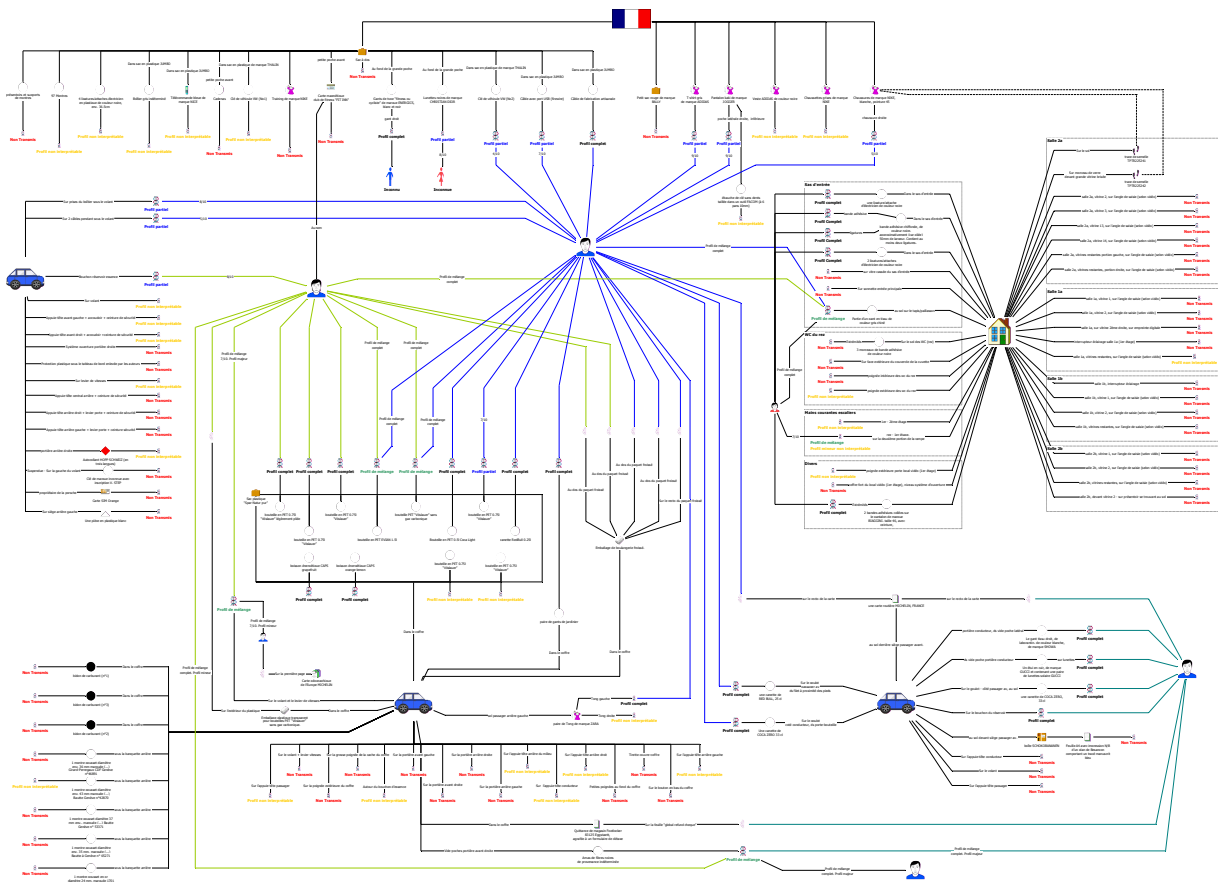
Remarque: si l'objectif est de décider sur quelles traces effectuer des comparaisons locales, l'attribut "profil partiel" peut être mis en évidence par la couleur rouge

En cas de **mélange**, les informations sur les profils peuvent être représentées sur le lien entre la trace et l'individu (en exploitant le même code de couleur)



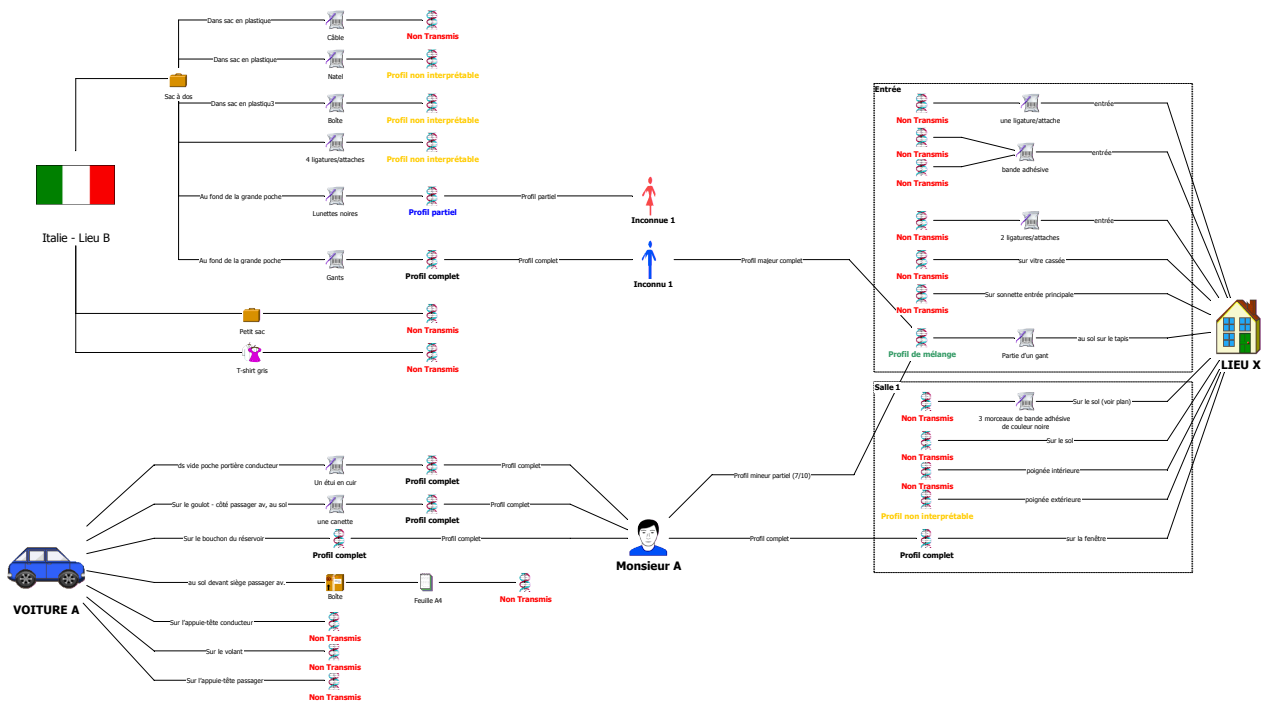
Exemple de cas

A la suite d'un brigandage survenu dans un musée, de nombreux prélèvements ont été effectués sur les lieux du crime, dans les véhicules utilisés par les auteurs et à l'endroit de la découverte du butin. Le schéma a été produit plusieurs semaines après le cas. Sur la centaine de prélèvements effectués, la moitié environ avait déjà été analysée. Le schéma produit ci-dessous a été exploité lors d'une séance avec le magistrat instructeur pour définir quelles autres traces devaient être analysées.

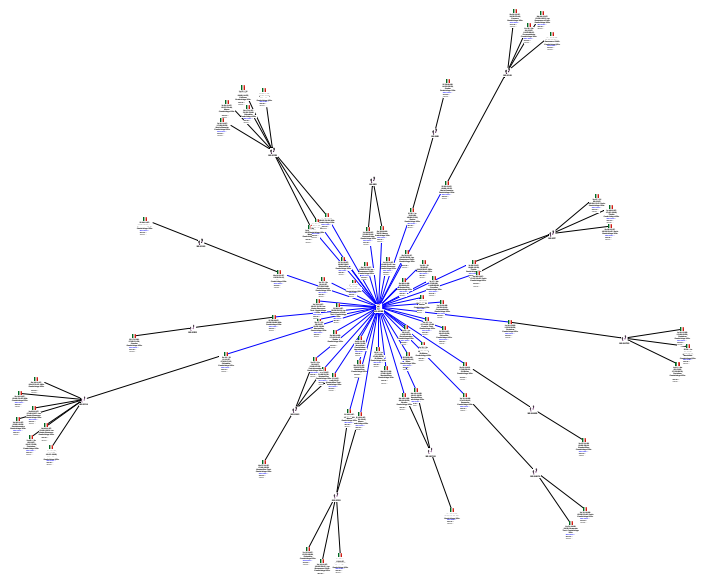


Bien qu'apparemment complexe, le schéma a facilité les discussions sur la pertinence des traces. En moins d'une heure de séance, l'ensemble des traces non transmises avaient été discutées et directement annotées sur le schéma.

Le schéma simplifié ci-dessous permet de voir plus clairement la structure proposée :



Le schéma relationnel ci-contre issu du même cas que présenté dans la section précédente (« Représentation des relations spatio-temporelles ») met en évidence une stratégie intéressante. Pour chaque groupe de cas liés par un motif, un nombre limité de cas sont également liés par l'ADN. A la lecture de ce schéma, il est possible de penser que les liens établis par les motifs de semelles ont été exploités pour guider la prise de décision lors de la sélection des traces biologiques à envoyer pour analyse. Une telle sélection est souvent nécessaire pour minimiser les coûts d'analyse. Une stratégie pourrait être de sélectionner un cas par groupe (défini par les relations établies sur la base des traces de semelles), puis d'envoyer un prélèvement pour analyse. Le choix du cas et plus particulièrement du prélèvement envoyé est, dans une deuxième phase, effectué en fonction des critères présentés précédemment. Si le résultat est négatif, une autre trace du groupe est envoyée. Dans le cadre de cette affaire, deux traces par cas avaient en fait été analysées en raison de l'importance de la série. Une telle stratégie pourrait néanmoins être exploitée.



Anal yser des trans actions

C o n t e x t e

Les appels téléphoniques entre des numéros de téléphones, les opérations financières entre des comptes et les ventes entre des personnes sont des relations fréquemment rencontrées dans le cadre d'enquêtes. Ces relations sont dirigées : un appel est effectué depuis un numéro appelant vers un numéro appelé, une transaction financière s'opère entre un compte débiteur et un compte créditeur, finalement une vente est réalisée entre un vendeur et un acheteur. Bien qu'issues de sources différentes, toutes ces relations peuvent être représentées de façon similaire – par un lien dirigé.

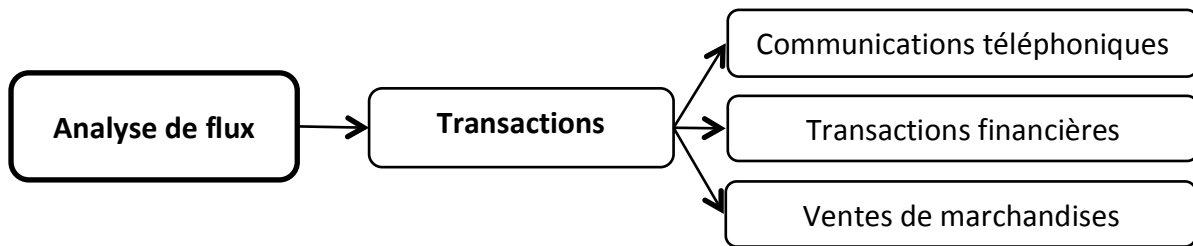
Un pattern d'analyse de transactions est formalisé sur la base des similitudes observées entre ces trois situations d'analyse. Dans certains manuels, ce type d'analyse est appelée « analyse de flux » (Interpol, 1997) (Peterson, 1998) (UNODC, 2002) (Fedpol, 2010). Celle-ci est généralement décomposées en plusieurs sous catégories. L'analyse de flux de marchandises se focalise sur le transfert de biens et d'argent entre des personnes et/ou entre des sociétés. Elle vise la reconstruction de la structure d'un trafic et l'identification des rôles des protagonistes. L'analyse de flux d'événements (Event Flow Analysis) et l'analyse de flux d'activités (Activity Flow Analysis) sont parfois également définies (Peterson, 1998) (UNODC, 2002). Celles-ci visent ce pendant des objectifs différents : décomposer une séquence d'événements pour mieux comprendre leurs relations (temporelles et de causalité) et respectivement comprendre une activité criminelle par la description des étapes nécessaires à sa réalisation. Elles exploitent les schémas chronologiques et respectivement les schémas d'activités précédemment décrits⁶⁰.

Le pattern proposé intègre l'analyse de flux de marchandises, mais se différencie des analyses de flux d'événements et d'activités qui exploitent des formes de représentation différentes. Il inclut par contre l'analyse de données de téléphonie et l'analyse de données financières. Ces trois « formes » d'analyse sont souvent abordées séparément et des formes de représentations propres sont définies (Peterson, 1990) (Peterson, 1998) (UNODC, 2002) (AGIS, 2004). Bien que chacune d'elles réponde à des besoins spécifiques, plusieurs similitudes sont observées entre elles. Une abstraction plus globale est donc proposée sous la forme d'un pattern d'analyse de transactions (ou de flux). Son application est également décrite pour chaque type de

⁶⁰ Cf. Analyse et visualisation relationnelle, page 42

170 Patterns de visualisation relationnelle

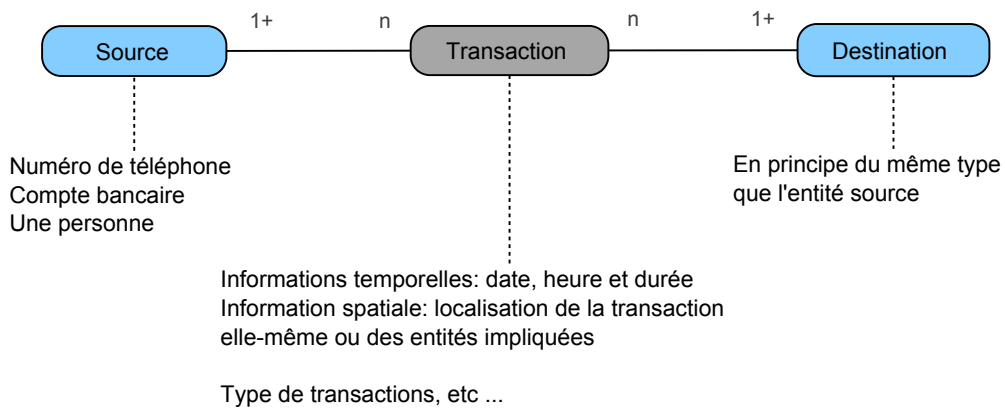
transactions : les communications téléphoniques, les transactions financières et les ventes de marchandises.



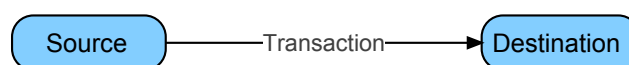
Objectifs de la représentation

Les schémas de flux relationnels représentent des transactions afin de détecter des régularités particulières signes d'une activité illicite, de reconstruire la structure d'un trafic de marchandises, d'identifier des entités (des personnes, des sociétés, des téléphones) intéressantes pour l'enquête, notamment par leur position centrale dans le flux.

Concepts à représenter



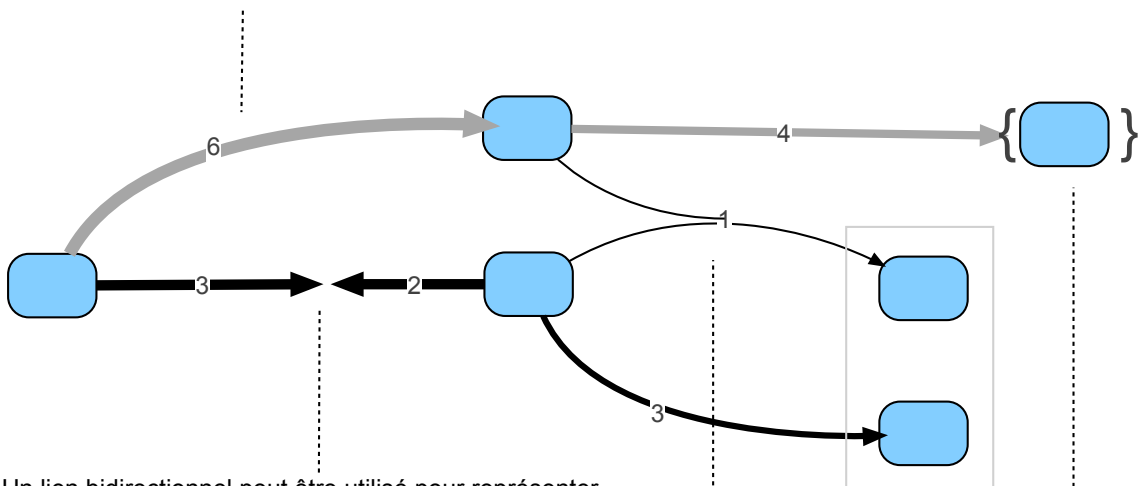
Représentation des relations



Que cela soit des ventes, des communications ou des transactions financières, les transactions peuvent être représentées par des liens entre des entités (respectivement des personnes, des comptes ou des numéros de téléphones).

Globalement plusieurs types de relations et leurs représentations associées peuvent être intégrées sur les schémas de flux relationnels. De plus, représenter les transactions par des liens permet de regrouper par agrégation un ensemble de transactions sous la forme d'un seul trait. L'épaisseur de celui-ci est alors exploitée pour représenter une variable quantitative utile pour l'objectif d'analyse, telle que le nombre de transactions, ou la quantité de produits vendue.

Un lien unique peut représenter, par agrégation, un ensemble de transactions.



Un lien bidirectionnel peut être utilisé pour représenter un ensemble de transactions effectuées dans les deux sens ou une transaction bidirectionnelle, telle qu'un échange.

Un lien multiple permet de représenter des transactions impliquant plusieurs sources ou plusieurs destinations.

Un cadre représentant l'appartenance de plusieurs entités à un même groupe, tel qu'une société par exemple.

Plusieurs entités peuvent être regroupées sous la forme d'un ensemble pour simplifier la représentation. Les transactions vers d'autres entités sont alors agrégées.

C o m m u n i c a t i o n s t é l é p h o n i q u e s

C o n t e x t e

L'analyse de données issues de communications téléphoniques est une technique courante de l'investigation criminelle. En effet, certaines activités criminelles nécessitent l'utilisation de ces moyens de communication ou sont facilitées grâce à eux. L'avènement de la téléphonie mobile a conduit à l'intégration complète de la téléphonie dans nos activités routinières et évidemment celles des criminels.

L'intérêt du contenu des communications interceptées par le biais d'écoutes actives est évident. De manière complémentaire, les métadonnées techniques (tels que les numéros des correspondants, la date et l'heure des appels) sont également précieuses. L'historique des appels contenant ces données techniques peut être obtenu rétroactivement sur des périodes plus ou moins longues en fonction des législations nationales. Les représentations discutées ici portent sur l'analyse de ces informations techniques indépendamment du contenu des communications.

Ces données sont dites « massives ». En effet, les données récoltées pour un seul numéro peuvent contenir plusieurs centaines, voire milliers de communications sur quelques mois. Des techniques particulières sont donc nécessaires pour les analyser. L'apport de la visualisation pour traiter ces informations est reconnu depuis plusieurs décennies (Peterson, 1990). Lors d'une investigation sur un trafic de stupéfiants, l'analyse des contacts d'un suspect peut, entre autres, permettre d'identifier des fournisseurs (lors d'appels internationaux notamment) ou des transporteurs sur la base de déplacements spécifiques connus.

Les données de facturations téléphoniques sont particulièrement intéressantes et complexes à traiter car elles offrent des possibilités d'analyse dans l'ensemble des dimensions. En effet, chaque appel est défini dans le temps par une date, une heure et une durée. Des représentations particulières peuvent être établies d'une part pour représenter le profil de l'activité téléphonique par jour de la semaine ou par heure notamment (voir l'annexe un pour des exemples de représentation) et d'autre part pour visualiser la séquence temporelle des appels (par exemple au moyen de diagrammes ou de schémas de flux temporels). Lorsque l'analyse porte sur les données issues d'un téléphone portable, l'antenne par laquelle la communication a transité est également mémorisée. Une analyse spatiale peut donc être effectuée afin d'identifier des déplacements particuliers ou des patterns d'activités suspects.

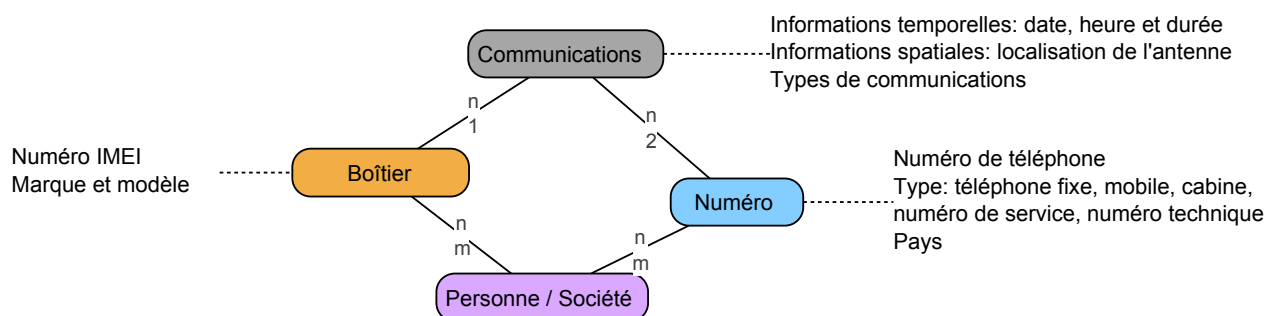
Dans le cadre de séries de cambriolages, comparer spatio-temporellement les appels d'un suspect à des événements permet parfois de trouver des correspondances intéressantes.

D'un point de vue quantitatif, le nombre et la fréquence des appels pour chaque correspondant sont, par exemple, des questions récurrentes. Ces aspects ne sont pas détaillés dans cette section qui présente les formes de représentations exploitées pour analyser ces données dans la dimension relationnelle afin de représenter les contacts.

Objectifs de la représentation

Les schémas de flux relationnels sont produits sur la base des communications téléphoniques entre différents numéros. Des schémas représentant les relations entre les numéros et les boîtiers peuvent également être produits lors de l'utilisation de téléphones portables. En effet, il est parfois observé que des individus s'échangent ou se prêtent ponctuellement leurs boîtiers, alors même qu'aucune communication n'est détectée entre leurs numéros.

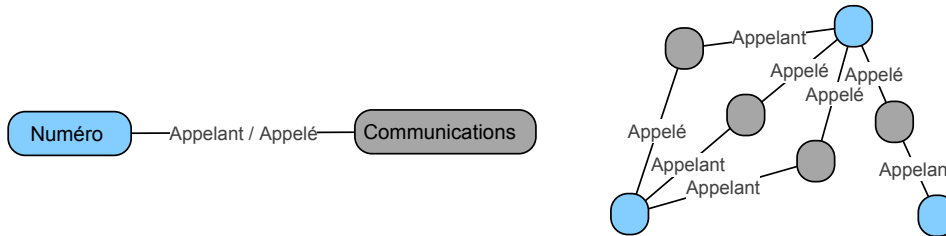
Concepts à représenter



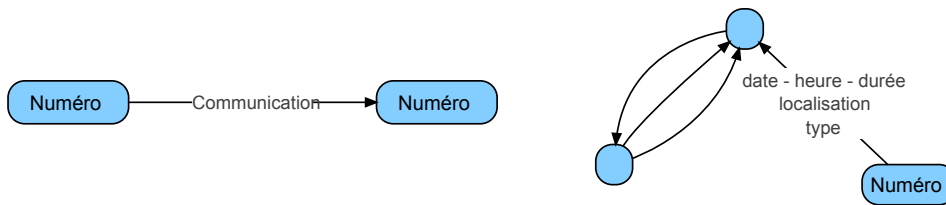
De manière générale, les relations établies sur la base des données de téléphonie permettent de lier des raccordements, voire des appareils et non directement des individus. Les relations avec des personnes physiques ou morales sont établies sur la base de l'enregistrement du numéro auprès des fournisseurs de services.

Représentation des relations

La relation principale sur laquelle porte l'analyse est la communication établie entre les numéros de téléphone:



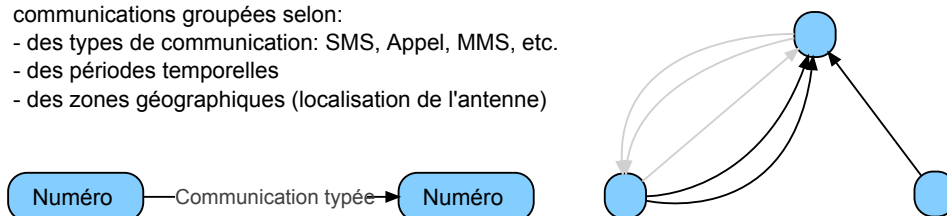
Une communication étant par définition liée à un numéro appelant et à un numéro appelé, peut être représentée par un lien (à l'exception des conférences téléphoniques). Le lien a un sens entre le numéro appelant et le numéro appelé :



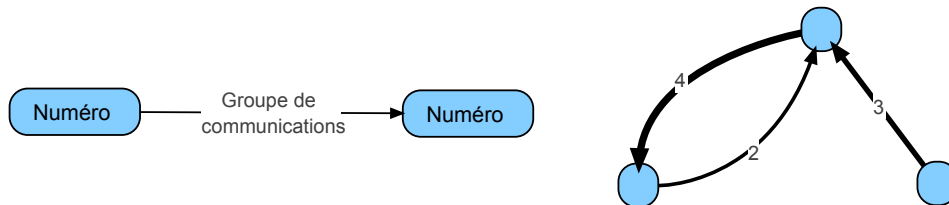
Sur chaque relation, la date, l'heure, la localisation et le type de communication sont des informations qui peuvent être représentées sous forme de texte notamment. Chaque entité représente un numéro. L'icône et la couleur peuvent être exploitées pour différencier des types de numéro (fixe, mobile, cabine téléphonique), ou pour distinguer des pays d'origine ou des cantons par exemple. La couleur des relations peut également être exploitée pour différencier des types de communications :

La couleur peut être exploitée pour représenter des communications groupées selon:

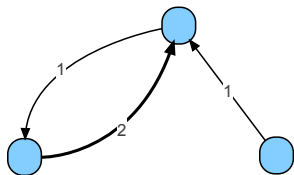
- des types de communication: SMS, Appel, MMS, etc.
- des périodes temporelles
- des zones géographiques (localisation de l'antenne)



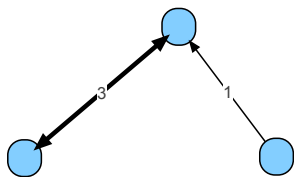
Le nombre de communications entre deux numéros peut être très grand. La représentation devient rapidement illisible si un lien est tracé pour chaque communication. C'est pourquoi, il est d'usage de regrouper les communications (le nombre de communications est généralement inscrit sur le label du lien) :



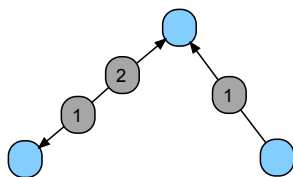
Plusieurs stratégies peuvent être exploitées pour représenter les communications de façon agrégée :



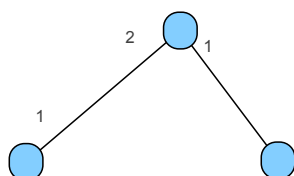
Deux liens dirigés sont exploités. Les épaisseurs sont proportionnelles au nombre de communications. Il s'agit du modèle le plus couramment utilisé actuellement.



L'utilisation d'un seul lien permet de simplifier la représentation, mais rend impossible le dénombrement des communications en fonction du sens.

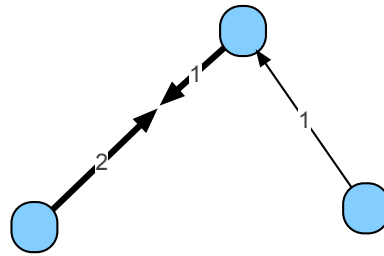


Historiquement, Andrews et Peterson (1990) proposent un modèle basé sur la représentation des communications par de petits cercles dont le label correspond au nombre de communications. Le nombre d'entité est fortement augmenté au détriment de la lisibilité générale.



Morris (1986) propose une représentation simple basée sur le positionnement de deux labels pour décrire le sens des relations.

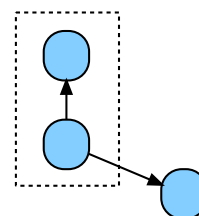
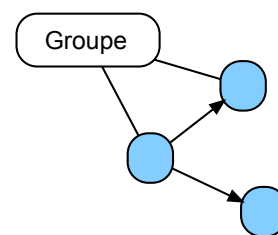
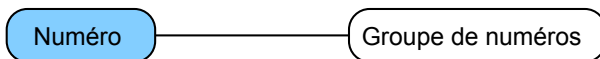
Le modèle proposé se base sur une représentation spécifique des relations bidirectionnelles⁶¹ :



La largeur globale du lien représente le nombre total de communications. Un seul lien permet de représenter les deux sens de communication. La position des flèches permet d'évaluer visuellement la proportion de communications dans chaque sens. Cette forme de représentation a été évaluée dans le chapitre précédent par rapport à l'utilisation de deux liens dirigés. Pour rappel, une amélioration significative a été observée pour identifier le numéro ayant le plus de contacts, tant au niveau du taux de réponses correctes que du temps de réponse⁶².

Comme mentionné ci-dessus, il est souvent nécessaire de distinguer des types de numéros tels que des cabines, des téléphones mobiles, etc. Certains numéros sont parfois liés entre eux et forment des groupes particuliers. C'est le cas notamment d'un ensemble de numéros appartenant à une même société ou à une même personne. A titre d'exemple, des cabines téléphoniques situées dans une même région géographique forment également des groupes particuliers. Ces groupes peuvent être représentés soit par le biais d'une entité liée ou par un cadre :

- Des numéros peuvent être regroupés selon diverses catégories:
- même entreprise / personne
 - zone géographique (ex: cabines d'un aéroport)

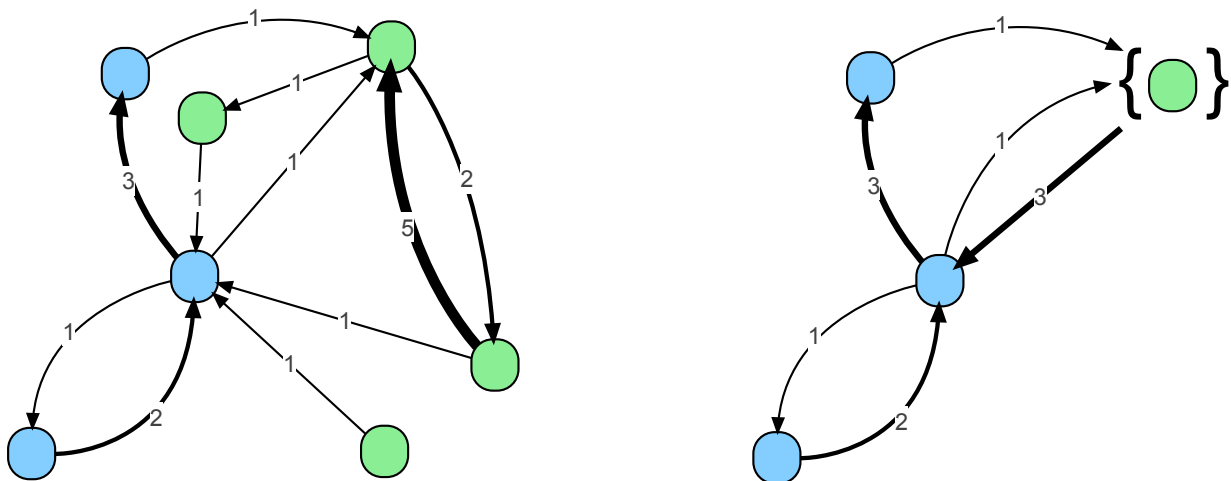


⁶¹ Cf. Typologie de relations, page 103

⁶² Cf. Représentation de liens bidirectionnels, page 116

Les numéros de téléphones et les communications étant généralement nombreux, l'ajout d'entités et de liens supplémentaires aurait tendance à complexifier le schéma. Il est donc recommandé d'utiliser des cadres qui peuvent se distinguer des liens par des effets de couches (notamment en variant le type de trait ou la couleur). Les cadres évitent la démultiplication des entités et liens. L'utilisation de cadres ne permet toutefois pas de commenter les relations spécifiques entre chaque numéro et le groupe. Le placement des entités peut également être exploité pour représenter des groupes selon le principe de proximité (des entités placées à proximité tendent à être perçue comme un groupe).

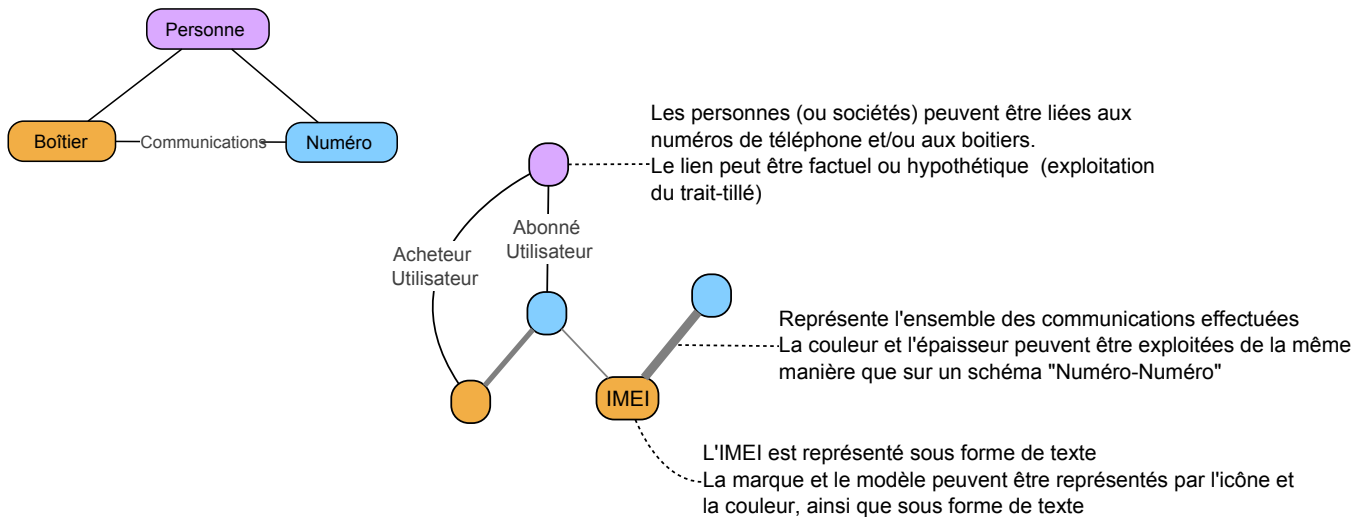
Des groupes de numéros peuvent également être représentés sous la forme d'un ensemble⁶³ afin de diminuer le nombre d'entités et simplifier la représentation. Une telle simplification peut être utile notamment pour regrouper l'ensemble des numéros d'une même société :



Les relations entre les numéros ainsi fusionnés ne sont évidemment plus visibles. Mais en fonction de la question d'analyse une telle simplification peut être très efficace. Les outils actuellement disponibles ne permettent cependant pas d'exploiter ce type de représentation.

⁶³ Cf. Typologie de relations, page 102

Des relations sont également établies sur la base d'échanges de boîtiers entre les utilisateurs. En effet, pour chaque communication le numéro IMEI⁶⁴ est présent pour le numéro à partir duquel les données ont été obtenues. Des schémas relationnels peuvent être produits pour représenter les relations entre les numéros et les boîtiers.



Finalement, pour chaque numéro, le détenteur est identifiable auprès des opérateurs. Des informations d'enquête peuvent également être disponibles sur les utilisateurs des téléphones. Il est alors possible de représenter ces relations sur les schémas relationnels (sous la forme d'entités liées, telles que présentées ci-dessus, ou sous la forme d'un attribut de chaque numéro). Lorsque le nombre de numéros de téléphone utilisés par chaque individu s'accroît, une simplification de visualisation peut être effectuée en ne représentant que les relations entre les personnes :

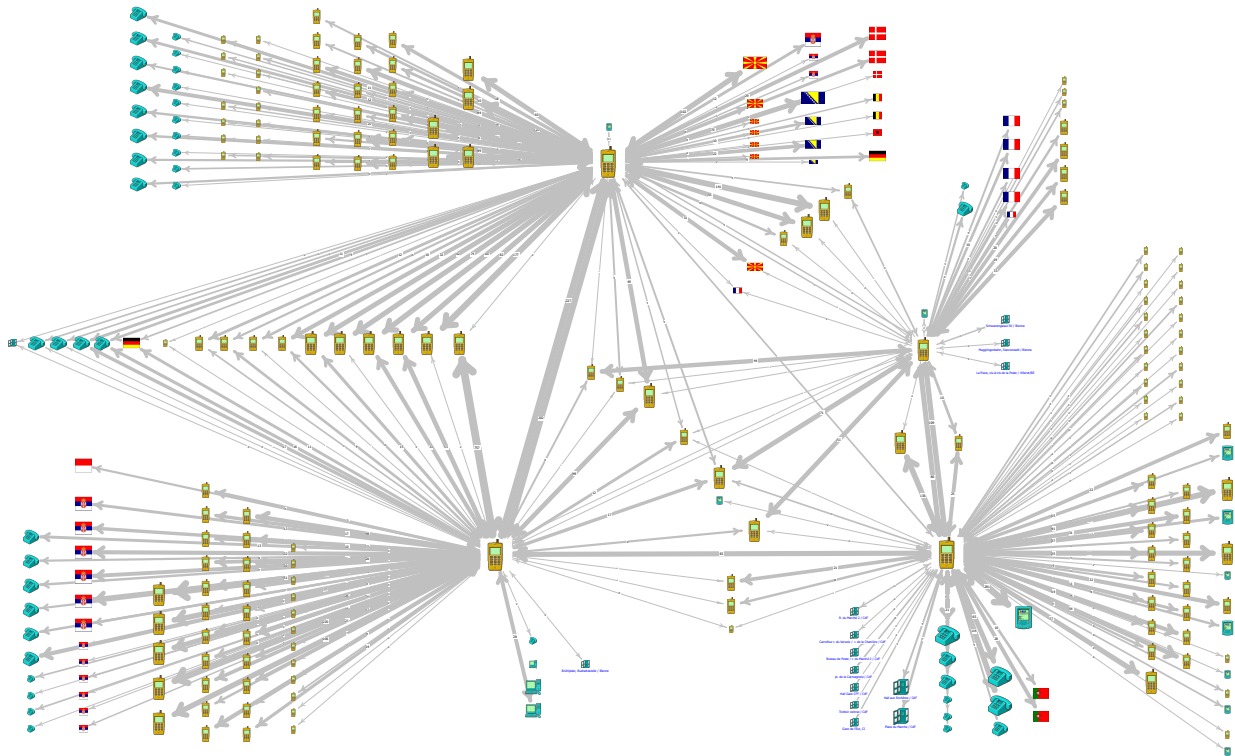


Toutefois des incertitudes sur les liens entre les personnes et les numéros existent. Le fait qu'un individu soit bel est bien l'utilisateur d'un téléphone est très souvent incertain. Il n'est d'ailleurs pas rare de voir des communications entre des numéros attribués à une même personne. Ceci peut s'expliquer par la vente ou le prêt de cartes SIM ou simplement par l'achat d'une carte effectué sous un faux nom. Il n'est donc pas recommandé d'effectuer cette simplification qui cache cette indirection importante.

⁶⁴ l'International Mobile Equipment Identity est un numéro d'identification des appareils mobiles. Il permet notamment d'identifier le fabricant et le modèle de l'appareil.

Exemple de cas

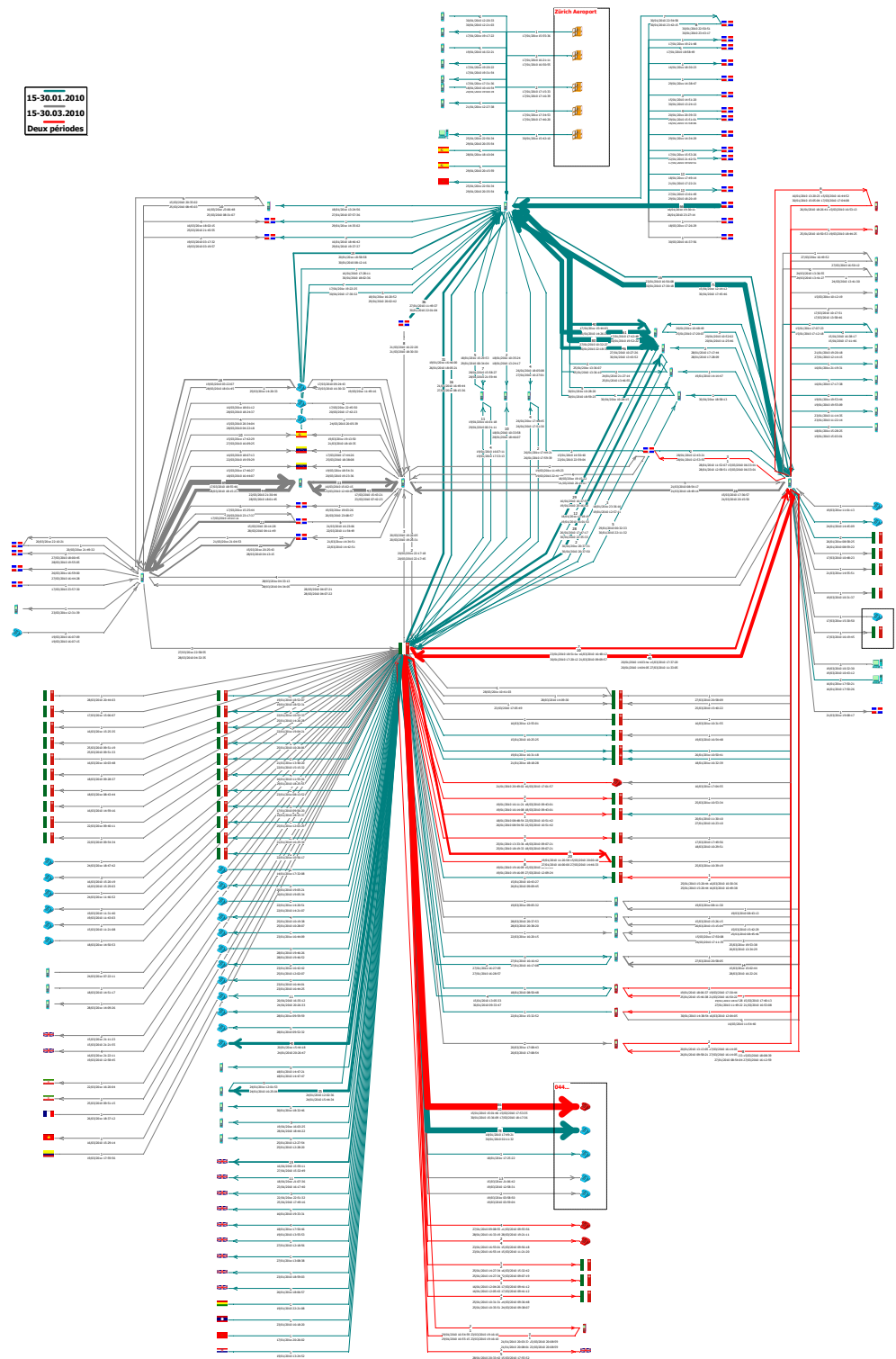
Dans l'exemple présenté ci-dessous, la forme de représentation des liens bidirectionnels n'a pas pu être exploitée, car les outils disponibles n'offrent pas cette possibilité. Des liens simples regroupant les communications entrantes et sortantes sont exploités. Un exemple de cas exploitant le pattern proposé a été présenté dans l'évaluation⁶⁵.



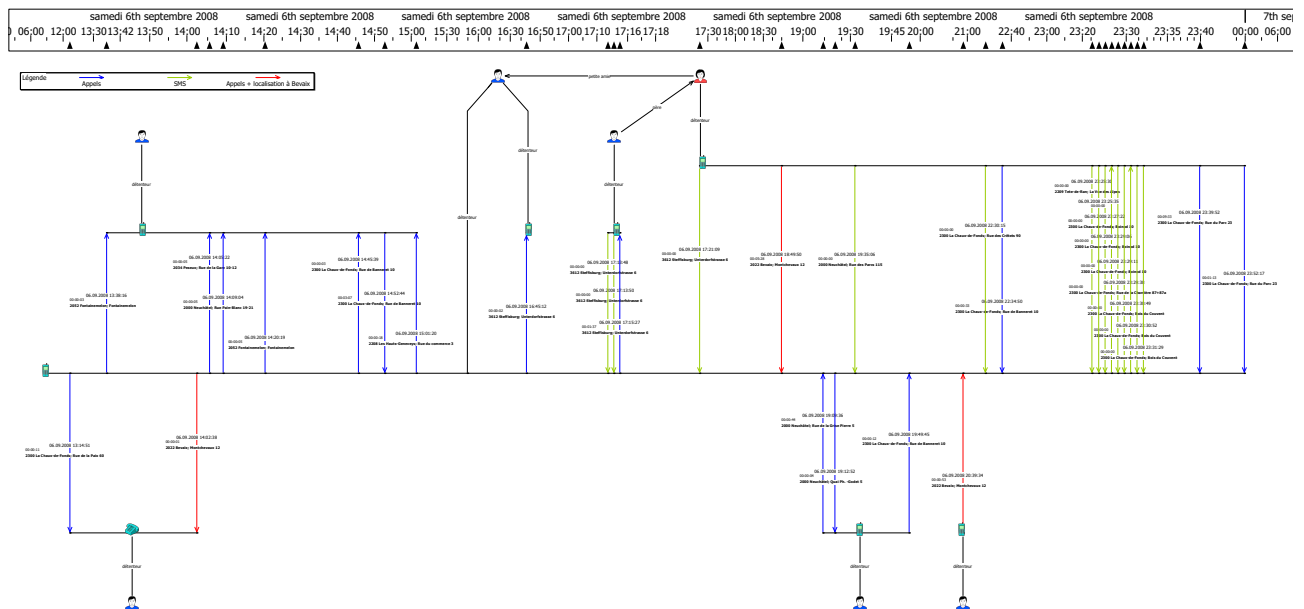
Cet exemple met en évidence l'usage des icônes et des couleurs pour distinguer des types de téléphone (fixe, mobile, cabine, etc.), ainsi que des pays. Le placement des nœuds a également été réalisé en fonction des types de numéros pour faciliter la lecture du schéma, tel que discuté dans le chapitre précédent. Cet exemple illustre l'utilité de positionner les entités en fonction de leurs types. Malheureusement, il n'existe à l'heure actuelle aucun algorithme automatique facilitant un tel positionnement. Le travail est donc complètement manuel. Finalement la taille des icônes est proportionnelle au nombre de communications.

⁶⁵ Cf. Représentation de liens bidirectionnels, page 114

Le schéma ci-contre présente une stratégie particulière d'intégration de la dimension temporelle sur un schéma relationnel. En effet, la couleur des liens est exploitée pour distinguer des périodes temporelles précises : en gris pour la première période, en turquoise pour la seconde et en rouge lorsque des communications ont été effectuées sur les deux périodes. Dans cette affaire l'objectif était d'identifier un numéro particulier en fonction de deux activités supposées de son détenteur sur deux périodes particulières. Il était également supposé que le détenteur de ce numéro avait eu des contacts avec les protagonistes de l'affaire pendant ces deux périodes. L'analyse porte principalement sur les contacts. La dimension relationnelle est donc dominante. Dans la dimension temporelle, la variabilité est définie par trois périodes.

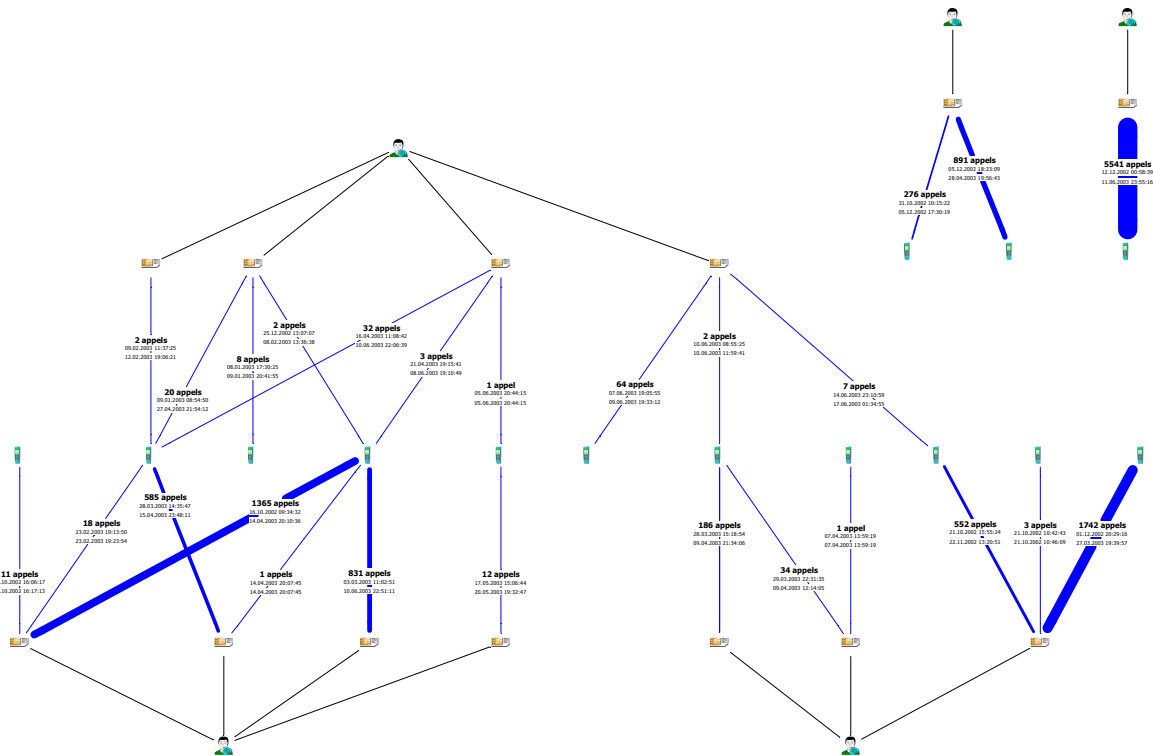


Le schéma ci-dessous a été réalisé pour analyser les communications du numéro de téléphone portable utilisé par le suspect d'une agression. L'agression a eu lieu dans une ville du littoral neuchâtelois et le suspect habitait dans une autre région du canton. L'objectif du schéma était de représenter la chronologie des appels effectués par le suspect en regard de la position géographique du relais utilisé lors de chaque appel et représenter également les personnes contactées. Cette question couvre trois des dimensions d'analyse : relationnelle (les contacts), temporelle (la chronologie des appels) et spatiale (la position des relais). Un schéma de flux temporel a été établi. En effet, d'un point de vue spatial, les appels ont été effectués dans deux régions uniquement : le lieu de domicile du suspect (en bleu et vert) et le lieu de l'agression (en rouge). Les liens verts sont des messages et les liens bleus des appels vocaux. Le nombre de contacts est relativement petit (sept numéros contactés). Finalement d'un point de vue temporel, l'analyse porte sur une journée particulière.



Ce schéma a permis de comparer le scénario décrit par le suspect et les traces laissées par son activité téléphonique. Il a ainsi été possible d'identifier des incohérences.

Le schéma ci-dessous représente les relations établies sur la base d'échange de téléphones portables :



De tels schémas sont notamment utiles pour détecter des relations entre deux cartes SIM, même si aucun appel n'a été effectué. En effet, il peut arriver que des complices veillent à ne pas se contacter directement, mais s'échangent leurs boîtiers.

Transactions financières

Contexte

L'analyse de transactions financières issues d'enregistrements bancaires, consiste en l'examen des dépôts, des paiements et des retraits effectués par une personne (physique ou morale) suspectée d'une activité illégale. L'analyse des flux financiers peut être exploitée pour estimer les dépenses et le train de vie d'un individu, pour identifier la présence éventuelle de revenus illégaux, et pour détecter des activités de skimming, de fraude ou de blanchiment notamment.

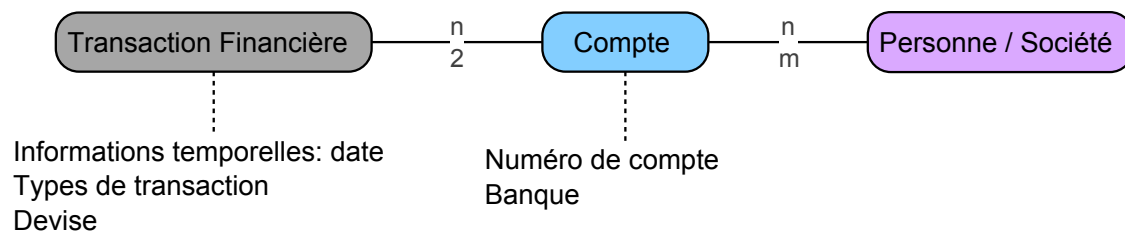
Les informations sont généralement fournies directement par des organismes financiers, voire saisies lors de perquisitions. L'intégration de ces informations pose en Suisse notamment de nombreux problèmes liés aux variations de formats et au support de transmission (la plus grande partie des banques transmettent les informations sous forme papier). Par ailleurs, ces données sont généralement fragmentaires. En effet, les relevés de comptes, bien qu'intégrant

la date, le montant et la devise de chaque transaction, ne contiennent pas toujours des informations précises sur la source ou la destination des fonds. De telles transactions ne peuvent donc plus être représentées précisément entre un compte source et un compte de destination. De plus, le nombre de comptes analysés dans le cadre d'une affaire peut être important. Lors de l'analyse des comptes d'une société (par exemple en cas de suspicion d'une fraude), il n'est pas rare que la société soit divisée en plusieurs entités disposant de nombreux comptes (clients, etc.). Des stratégies particulières de représentation sont alors exploitées pour obtenir une vue d'ensemble des informations.

Objectifs de la représentation

Les schémas de flux relationnels de flux sont produits sur la base de transactions financières afin d'explicitier un mécanisme financier particulier ou de détecter une activité de fraude ou de blanchiment par exemple. Ils permettent de détecter des patterns correspondant à des modes opératoires particuliers (par exemple afin de dissimuler la provenance de fonds).

Concepts à représenter



Représentation des relations

La relation principale sur laquelle porte l'analyse est la transaction financière effectuée entre un compte débité et un compte crédité. Les transactions financières peuvent donc être représentées par des liens dirigés.

L'ensemble des observations et des recommandations de représentation décrites pour l'analyse de données de téléphonie s'applique aux données financières. Afin d'éviter des redondances, elles ne sont pas réexpliquées ici. Seuls certains problèmes spécifiques complémentaires sont développés.

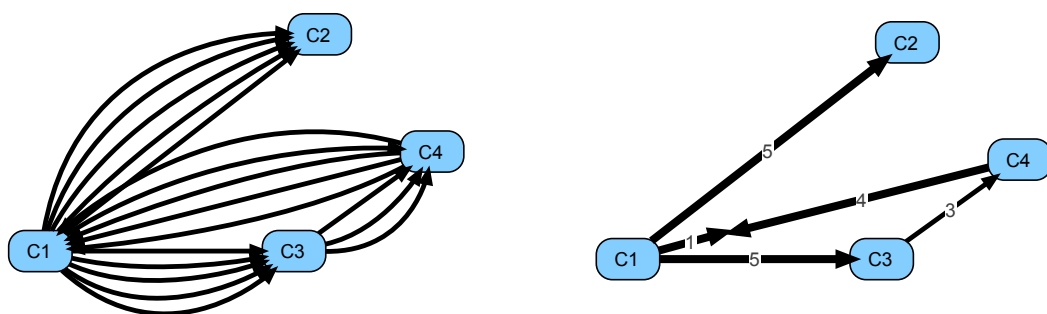
Produire un schéma relationnel efficace est une tâche difficile pour les raisons suivantes :

- Le nombre global de transactions
- Le nombre de comptes différents pour une même société ou personne
- La diversité des types de transactions : paiements de factures, de salaires, dépôts ou retraits cash, bonifications, placements, etc.
- L'incomplétude des données : absence du compte source pour une transaction au crédit ou du compte de destination pour une transaction au débit.

Lors de la création d'un schéma de flux temporel, plusieurs transactions entre deux mêmes comptes peuvent avoir été effectuées le même jour. Contrairement aux communications téléphoniques dont l'heure et la durée sont définies, seule la date de la transaction est disponible en matière de finance. Les transactions doivent donc être au minimum agrégées par jour pour éviter des superpositions visuelles.

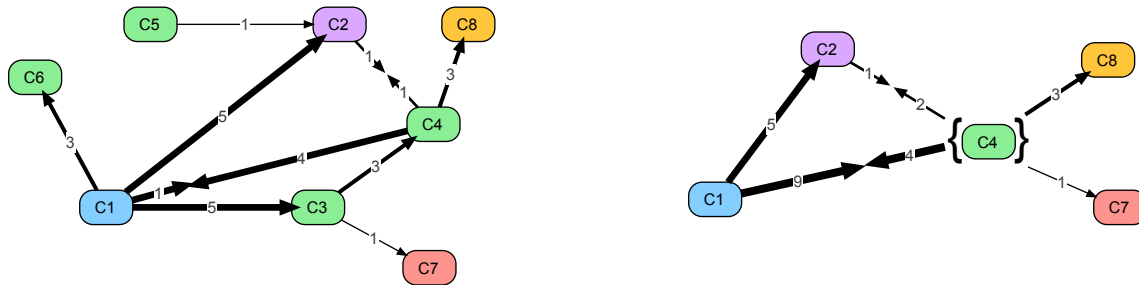
Pour exploiter au mieux la représentation, des sélections et des agrégations peuvent être effectuées. Un sous-groupe de transactions peut être filtré pour représenter une portion particulière des données. De telles sélections sont généralement effectuées pour répondre à des questions spécifiques : pour telles périodes, tels comptes, telles personnes, etc. quelles sont les relations ? Des agrégations particulières peuvent également être effectuées. Elles sont décrites dans ce pattern.

L'ensemble des transactions entre deux comptes sont regroupées. Le lien représente alors un groupe de transactions avec comme attributs: le nombre et le montant total des transactions.



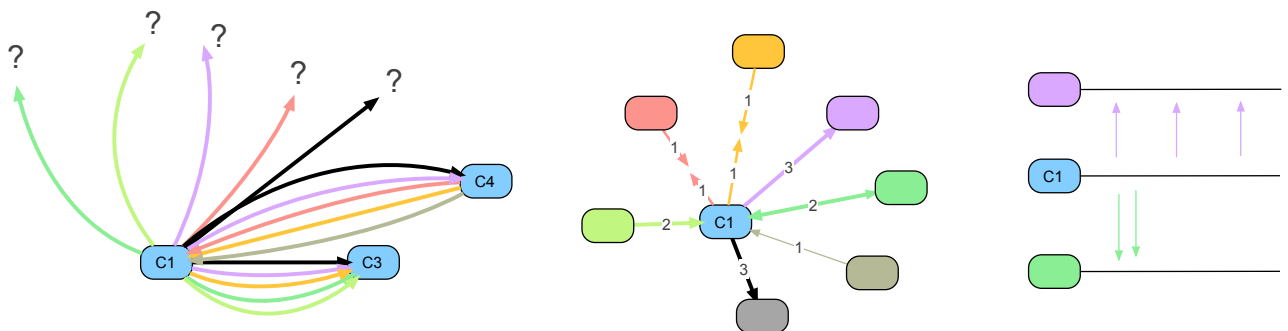
Agrégation de l'ensemble des transactions entre deux comptes. L'épaisseur des liens est alors proportionnelle au nombre de relations regroupées ou à la somme des montants.

L'ensemble des comptes d'une même personne ou d'une même société peut être fusionné. Les transactions entre ces comptes ne sont plus visibles, mais les entrées et sorties d'argent peuvent être analysées de manière globale.



Agrégation de l'ensemble des comptes d'une même personne ou d'une même société (en vert).

Une autre stratégie d'agrégation consiste à regrouper des transactions en fonction de leurs types (retraits cash, paiements, etc.). Cette méthode permet notamment de gérer l'incomplétude des données, puisqu'il n'est plus nécessaire de connaître précisément la source ou la destination de la transaction, mais uniquement son type (représenté ci-dessous par la couleur) :

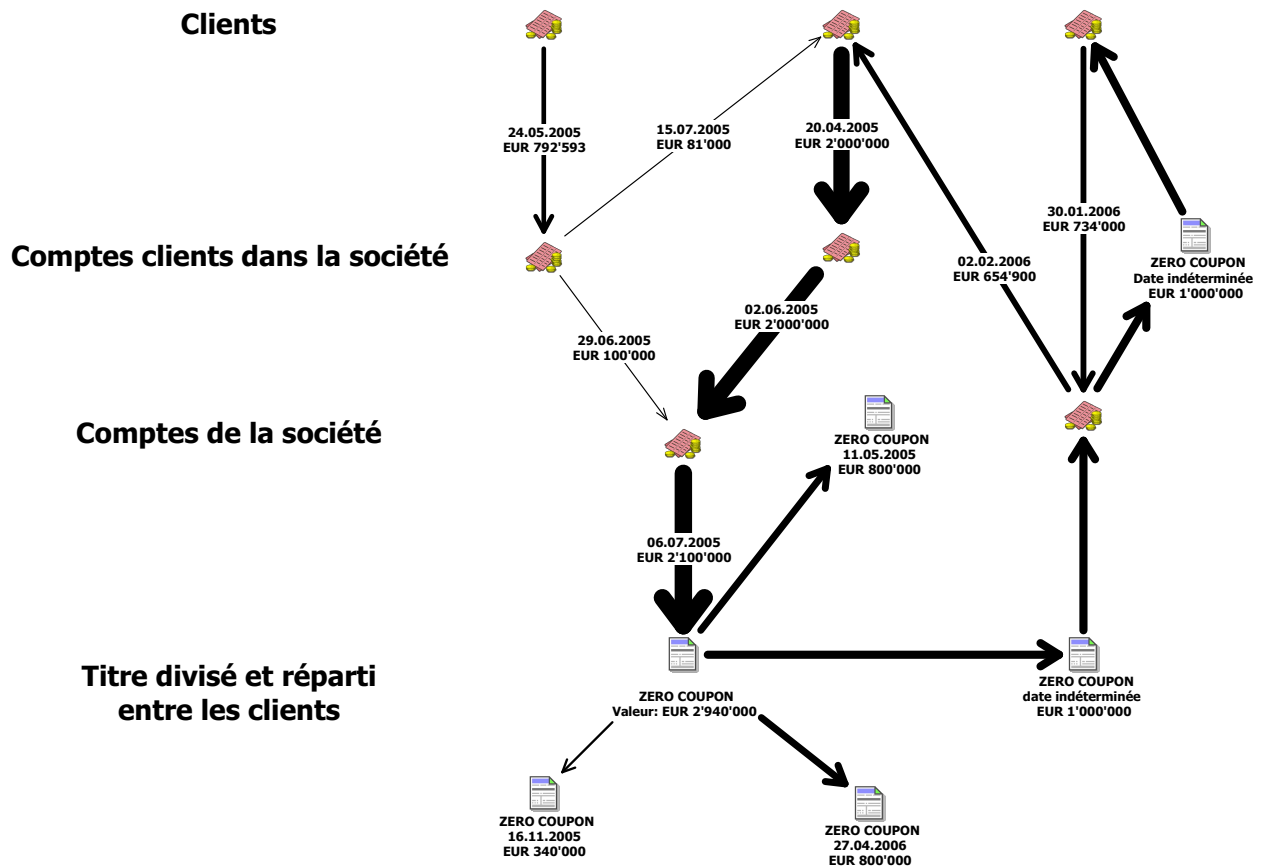


Sur le schéma de gauche, la couleur distingue des types de transaction et les nœuds représentent des comptes (« ? » lorsque celui est inconnu). Sur les deux schémas de droite, les nœuds représentent les types de transaction. L'agrégation par type permet de simplifier le schéma et d'effectuer des analyses particulières, telles que l'analyse des entrées et sorties des comptes d'une société.

Sur le schéma temporel où chaque transaction est représenté par un lien vertical, le compte lié à la transaction peut être représenté sous la forme d'un attribut du lien. La couleur peut également être exploitée pour distinguer des personnes ou des sociétés (voir l'exemple de cas présenté plus loin).

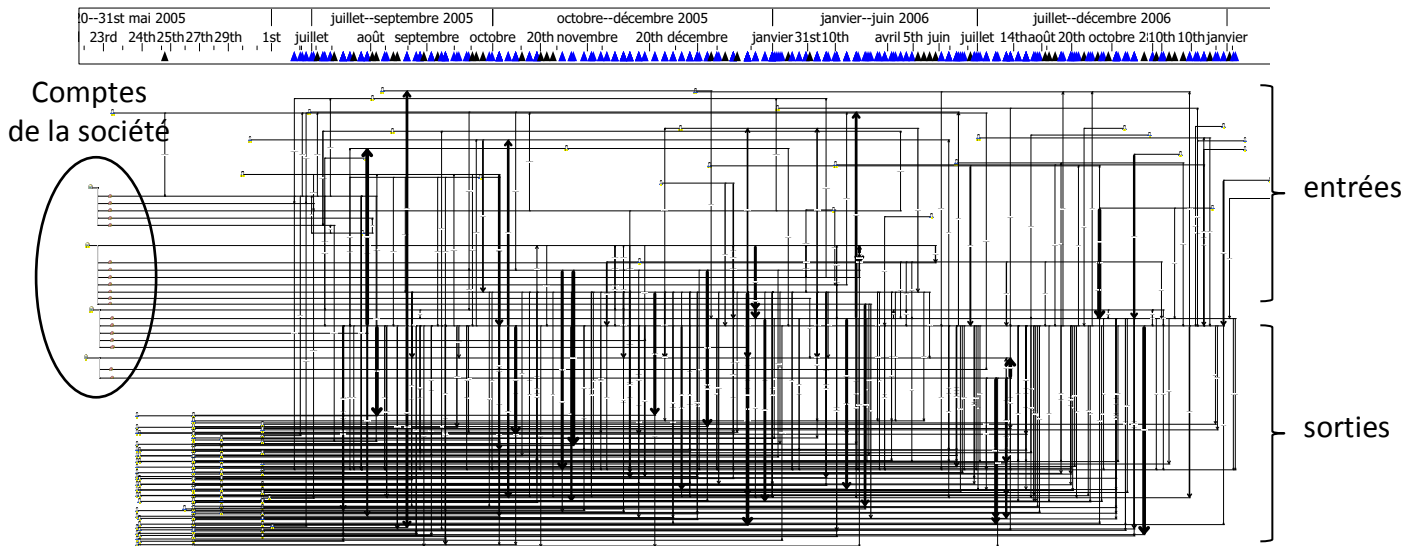
Exemple de cas

Des schémas relationnels peuvent être produits pour représenter un mécanisme financier particulier ou un ensemble de transactions suspectes. Le cas présenté porte sur l'analyse des comptes d'une société de placement suspectée d'avoir une activité frauduleuse. En effet, sur l'ensemble de la période d'activité de cette société, seul un placement financier a été retrouvé dans les comptes. Le schéma ci-dessous a été produit pour décrire ce placement.

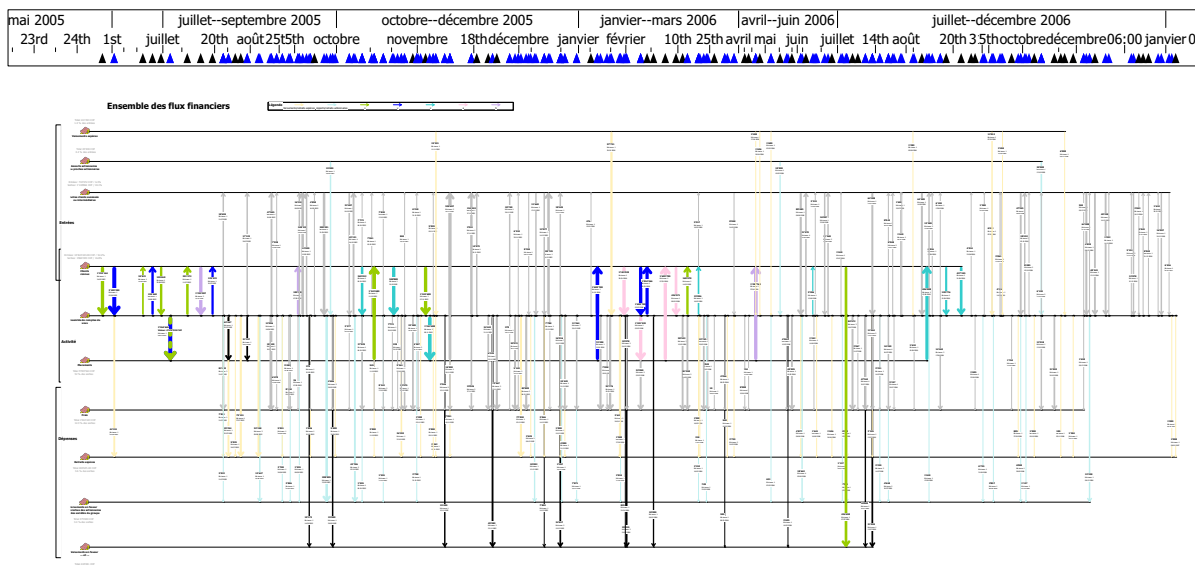


Le schéma décrit l'ensemble des transactions depuis les comptes des clients ayant apporté des avoirs jusqu'à l'achat de titres. Les entités ont été positionnées en fonction des types de comptes. De tels schémas sont utiles pour faciliter la description des mouvements bancaires.

Un schéma chronologique a également été produit pour représenter l'activité totale de la société. L'objectif était de représenter les entrées et les sorties bancaires. Le schéma ci-dessous a été produit en représentant l'ensemble des transactions pour tous les comptes de la société :



Le résultat est peu lisible. Des stratégies de simplification ont été appliquées pour pouvoir analyser les données. Le schéma final produit est le suivant :



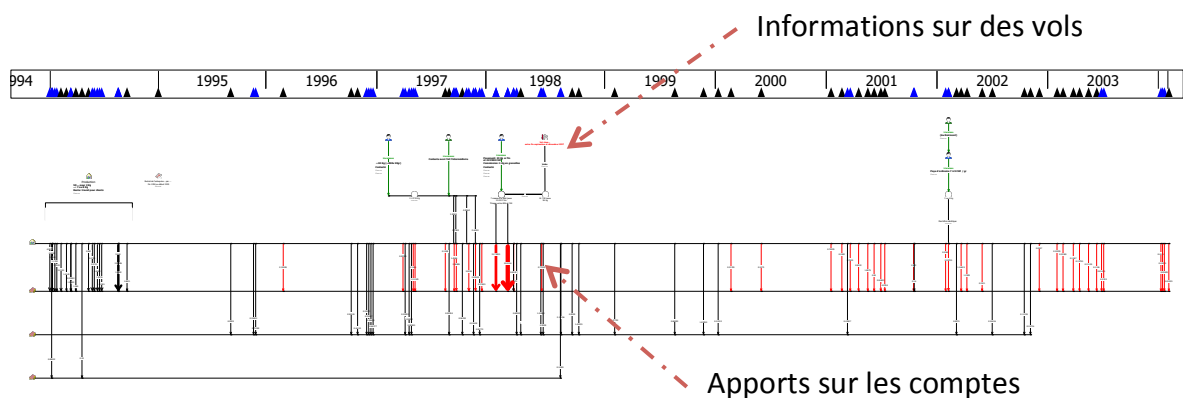
Les comptes de la société ont été regroupés et représentés par une seule ligne horizontale (au centre, car les mouvements internes ne sont pas pertinents pour l'objectif d'analyse). De plus, seules les transactions d'un montant supérieur à dix mille francs ont été sélectionnées. Ce choix a été effectué par le magistrat instructeur. Une telle sélection permet de réduire le nombre d'informations à représenter et donc la complexité du schéma. Un tel choix doit toutefois être justifié en fonction du contexte de l'affaire. Finalement l'ensemble des entrées et

188 Patterns de visualisation relationnelle

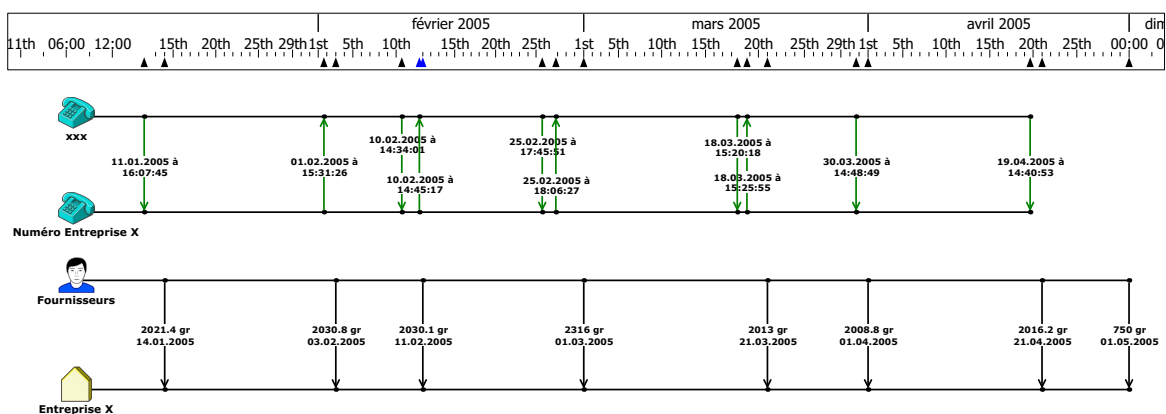
des sorties ont été regroupées selon leurs types (apports de clients, d'actionnaires et cash, frais divers, retraits cash, retraits au profit des actionnaires, etc.). La couleur a été exploitée pour distinguer les différents clients.

Ce schéma met en évidence l'absence d'une activité de placement sur l'ensemble de la période (sauf un placement en début d'activité). Il montre comment l'argent des clients a été dépensé. Finalement, il met en évidence le mécanisme particulier de la fraude. Lorsqu'un client mécontent souhaite récupérer son argent, une somme comparable provenant d'un autre client est créditée.

Des représentations peuvent également être produites pour mettre en relation des transactions financières avec d'autres informations. Sur le schéma ci-dessous, les apports effectués sur des comptes sont mis en relation avec des vols particuliers.



Le schéma ci-dessous représente la relation établie entre des communications téléphoniques et des transactions financières :



L'affaire porte sur la détection d'apports supposés frauduleux d'or auprès d'une société.

L'analyse des communications téléphoniques de cette société a permis de détecter un numéro de téléphone ayant effectué des communications avec la société de façon régulière ; deux ou trois jours avant chaque apports. Cette analyse a permis de remonter à la source des dépôts.

V e n t e s d e m a r c h a n d i s e s

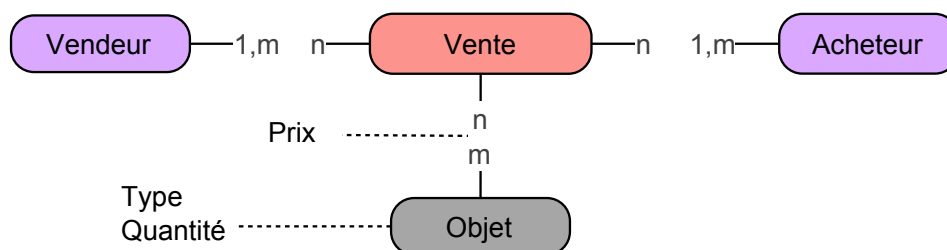
C o n t e x t e

Des transactions sont également fréquemment analysées lors de l'analyse d'un trafic (par exemple en matière de stupéfiants). A la différence des analyses effectuées sur la base de données de téléphonie ou de transactions financières, l'analyse d'un trafic se base généralement sur les déclarations de suspects ou de témoins.

O b j e c t i f s d e l a r e p r é s e n t a t i o n

Les schémas de flux relationnels représentent un ensemble de ventes entre des personnes, afin d'obtenir une vue d'ensemble du trafic et afin notamment de comparer les versions des différentes sources d'information.

C o n c e p t s à r e p r é s e n t e r

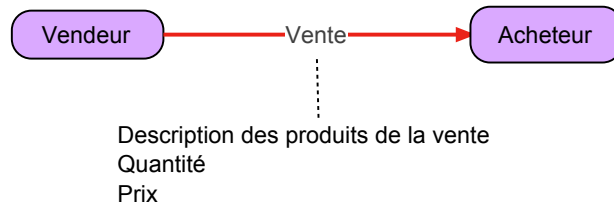


R e p r é s e n t a t i o n d e s r e l a t i o n s

Une vente est un type particulier d'événement où chaque individu possède un rôle particulier : le vendeur, l'acheteur, voire un intermédiaire. Bien qu'une vente puisse être réalisée entre plus de deux personnes, la majeure partie d'entre elles est effectuée entre un acheteur et un vendeur.

190 Patterns de visualisation relationnelle

Les ventes peuvent alors être considérées comme des transactions entre deux personnes, c'est-à-dire des relations dirigées entre le vendeur et l'acheteur :



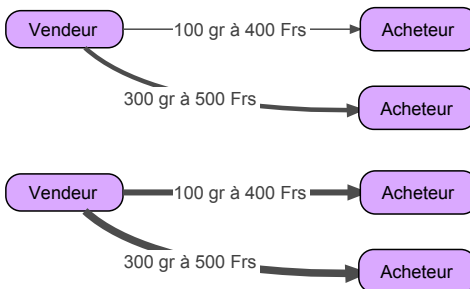
Les informations sur la nature, la quantité et le prix des objets vendus peuvent être représentées comme attributs sur les liens :



Pour chaque vente particulière (ou ensemble de ventes entre deux mêmes individus), les déclarations des différentes sources d'informations peuvent être comparées en ajoutant des attributs pour chaque source.



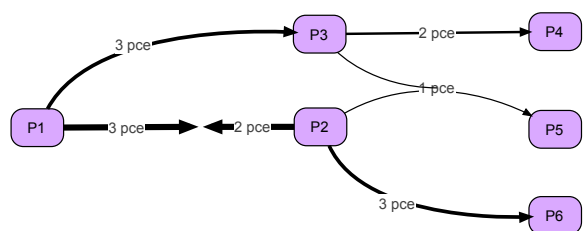
La couleur du lien peut être exploitée pour distinguer des types d'objets vendus par exemple.



L'épaisseur des liens peut être proportionnelle à la quantité ou au nombre d'objets vendus, voire au montant de la transaction en fonction de l'objectif d'analyse. Si la quantité est l'élément principal pour la qualification pénale (notamment en matière de stupéfiants), il peut être souhaité de la représenter visuellement. Si au contraire le prix est une variable

plus importante (par exemple afin d'estimer si l'acheteur pouvait se douter qu'il faisait l'acquisition d'un bien volé ou d'une contrefaçon), il pourra être privilégié.

Représenter une vente sous la forme d'un lien dirigé présuppose que la transaction ne concerne que deux personnes et soit dirigée dans un seul sens. Les échanges et les ventes entre plusieurs individus sont des cas particuliers qui peuvent être représentés en exploitant les liens bidirectionnels et les liens multiples.



Exemple de cas

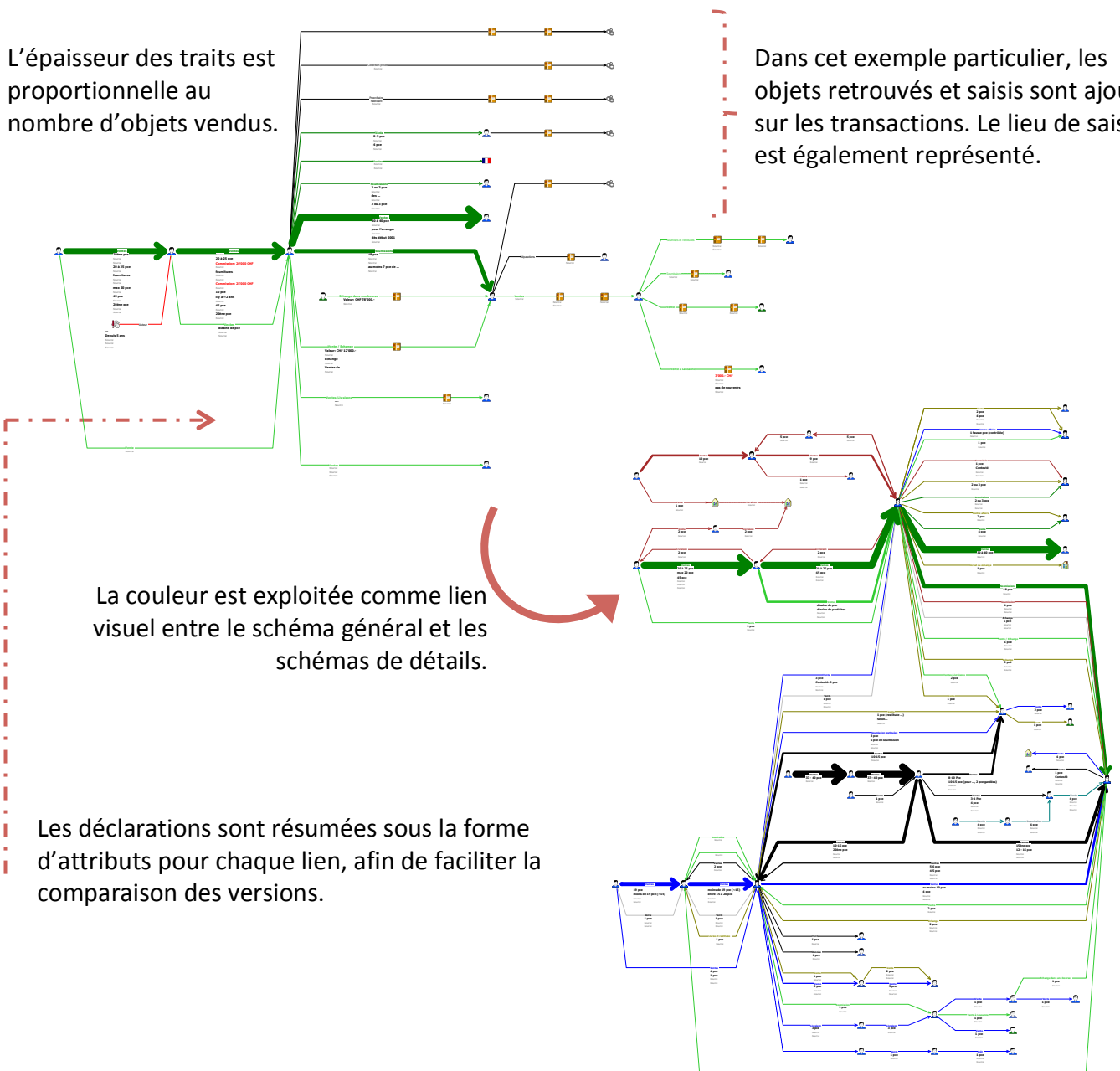
Cet exemple (déjà présenté dans la section « Représentation d'un trafic de marchandises » du chapitre « Evaluations préliminaires ») porte sur l'analyse d'un trafic d'objets volés ou contrefaits. Le schéma ci-dessous résume les informations tirées d'une cinquantaine de procès-verbaux effectués au cours de l'enquête. Il décrit l'ensemble des transactions pour une marque particulière d'objets. Les personnes impliquées dans cette affaire ont effectué un nombre très important de ventes et d'échanges d'objets de diverses marques. Le deuxième schéma représente l'ensemble du trafic. La couleur des liens a été exploitée pour différencier les marques et permettre de lier visuellement les schémas de détails (tel que celui présenté ci-dessous) et le schéma général.

L'épaisseur des traits est proportionnelle au nombre d'objets vendus.

Dans cet exemple particulier, les objets retrouvés et saisis sont ajoutés sur les transactions. Le lieu de saisie est également représenté.

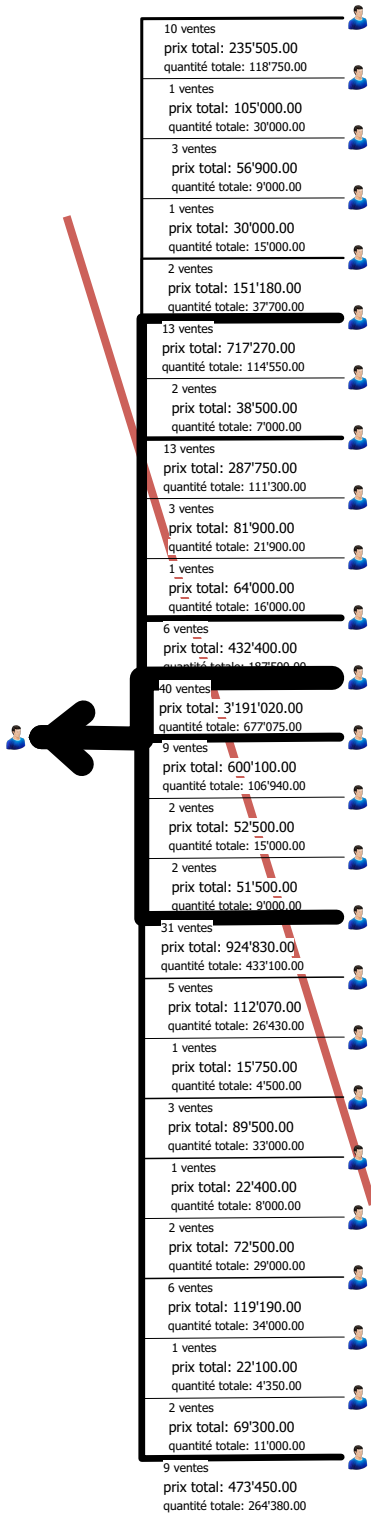
La couleur est exploitée comme lien visuel entre le schéma général et les schémas de détails.

Les déclarations sont résumées sous la forme d'attributs pour chaque lien, afin de faciliter la comparaison des versions.



192 Patterns de visualisation relationnelle

Ce dernier exemple de représentation de transactions est présenté afin d'illustrer un usage peu efficace des schémas relationnels. Les données représentées ici sont issues d'un fichier tenu à jour par un trafiquant de stupéfiants sur l'ensemble des achats qu'il a réalisés au cours de plusieurs années :



Représenter un ensemble de transactions liées à un seul individu (ou compte, ou téléphone) par un schéma relationnel n'est pas efficace. L'utilisation de tableaux ou de graphiques est par contre beaucoup plus utile.

Fournisseur	Nombre	Quantité	Prix
M	40	677'075	3'191'020
I	31	433'100	924'830
?	9	264'380	473'450
N	6	187'500	432'400
X	10	118'750	235'505
S	13	114'550	717'270
Q	13	111'300	287'750
L	9	106'940	600'100
T	2	37'700	151'180
C	6	34'000	119'190
F	3	33'000	89'500
W	1	30'000	105'000
D	2	29'000	72'500
H	5	26'430	112'070
P	3	21'900	81'900
O	1	16'000	64'000
K	2	15'000	52'500
U	1	15'000	30'000
A	2	11'000	69'300
J	2	9'000	51'500
V	3	9'000	56'900
E	1	8'000	22'400
R	2	7'000	38'500
G	1	4'500	15'750
B	1	4'350	22'100
Total	169	2'324'475	8'016'615

Le tableau peut être trié en fonction du prix, de la quantité de marchandise ou du nombre de transactions. Les histogrammes permettent de comparer les différents ratios (quantité/prix, quantité/nombre, etc.).

The methods and technologies that are supposed to support data analysis and reporting – known as business intelligence – have failed so far to deliver on their promise of intelligence. Despite great technical progress in data acquisition, data integration, data improvement through cleansing and transformation, and the construction of huge data warehouses that we can access at incredible speeds, the business intelligence industry has largely ignored the fact that intelligence resides in human beings, and that information only becomes valuable when it is understood, not just when it's made available. (Few, 2009, p.3)

Impact sur la conception d'outils

Problématique

Avant l'avènement du micro-ordinateur, les représentations graphiques étaient produites à la main. Actuellement elles sont produites, quasiment en totalité, à l'aide de logiciels de visualisation spécialisés. Ceux-ci imposent parfois leurs propres standards et les possibilités sont évidemment cantonnées aux fonctionnalités offertes. De plus, les logiciels sont très souvent développés pour répondre aux attentes du plus grand nombre. Ils offrent donc des solutions génériques, alors que les problèmes à résoudre sont souvent spécifiques. D'un côté l'analyste dispose de données propres à la situation traitée qui sont souvent volumineuses (telles que des données téléphoniques, des procès-verbaux d'auditions ou des listings de cas). De l'autre, il exploite des outils qui, malgré un nombre impressionnant de fonctionnalités, ne le guident que rarement vers des choix de représentation adaptés aux questions d'analyse.

Le processus de conception d'une représentation se base sur un ensemble de choix et de traitements qui doivent être effectués par l'analyste (Xu & Chen, 2005). Pourtant, les éléments méthodologiques et les recommandations présentés dans le premier chapitre sont rarement intégrés dans les outils. Les moyens techniques sont mis à disposition de l'utilisateur, mais il est rarement guidé pour effectuer ses choix (Sparrow, 1991). Cette liberté est certainement souhaitable. Elle offre de nombreuses perspectives d'exploitation et une large possibilité pour l'utilisateur d'user d'imagination. Par ailleurs, lorsqu'une méthodologie de travail est reconnue comme efficace, il serait certainement profitable que les outils l'intègrent - surtout lorsque l'on sait que de mauvais choix peuvent engendrer des biais d'interprétation.

La méthode décrite dans le premier chapitre et les exemples présentés dans le second reposent sur les possibilités offertes par les outils. Il a été montré par les exemples que le langage visuel actuellement disponible est malheureusement encore insuffisant. Par exemple,

les notions de groupe, d'ensemble et de relation multiple sont rarement intégrées de façon formelle dans les outils de visualisation relationnelle. Sélectionner une structure de graphe adaptée n'est pas toujours une tâche évidente : il est souvent nécessaire d'effectuer plusieurs essais avant d'aboutir à une solution efficace. De plus, pour un même problème et un même jeu de données, plusieurs questions peuvent se poser. Elles peuvent porter sur plusieurs dimensions d'analyse, sur l'ensemble des données, une partie ou un élément spécifique. Afin de faciliter la conception d'un graphe, il devrait techniquement être possible d'intégrer des données indépendamment de leurs structures et par le biais de transformations, représenter les informations selon divers modèles de graphe (Senator, 2005). Les outils exploités pour soutenir le processus d'analyse doivent donc permettre de modifier dynamiquement la représentation afin de pouvoir tester différentes solutions et changer les points de vue sur les données. Certains auteurs relèvent même que les outils actuellement utilisés sont plus utiles pour l'élaboration de comptes-rendus, qu'en tant qu'outils d'analyse facilitant les raisonnements et le développement d'hypothèses alternatives (Wright et al., 2006).

Les outils de visualisation sont également largement exploités pour simplifier la communication de problèmes complexes. Simplifier un problème, en écartant des éléments de détails, n'est pas toujours une bonne solution (Tillers, 2007). Les outils de visualisation doivent permettre de naviguer entre différents niveaux d'agrégation, afin d'effectuer l'analyse des informations la plus fine possible au niveau de détails souhaité.

La première partie de ce chapitre décrit le processus détaillé de conception d'un graphe d'un point de vue technique et tel qu'il devrait être intégré dans les outils. La seconde section présente des exemples de transformations dynamiques qui sont généralement utiles pour faciliter l'analyse des données par un graphe et modifier les points de vue sur les données. La question de l'intégration de l'ensemble des quatre dimensions dans un outil global est abordée dans la troisième partie. Finalement un prototype de visualisation multidimensionnel est présenté. Le prototype développé est une extension d'un logiciel de visualisation spatiale au sein duquel un module temporel a été intégré. L'intégration de représentations relationnelles n'a pu être réalisée dans le cadre de cette recherche, mais la manière de la réaliser est présentée.

Processus de visualisation

Une adaptation du modèle élaboré par Ed H. Chi (Chi & Riedl, 1998), (Chi, 2000) et (Chi & Panizzi, 2002) sous le nom de "*data state model*" puis repris dans (Card et al., 1999) sous le nom de "*information visualization reference model*" est proposée. La description du processus présentée ci-dessous intègre des aspects décrits par (Tang et al., 2004) et (Heer & Agrawala, 2006). Étonnamment, aucun de ces modèles ne formalisent le processus global selon une décomposition en cinq étapes, telle que décrite ici.

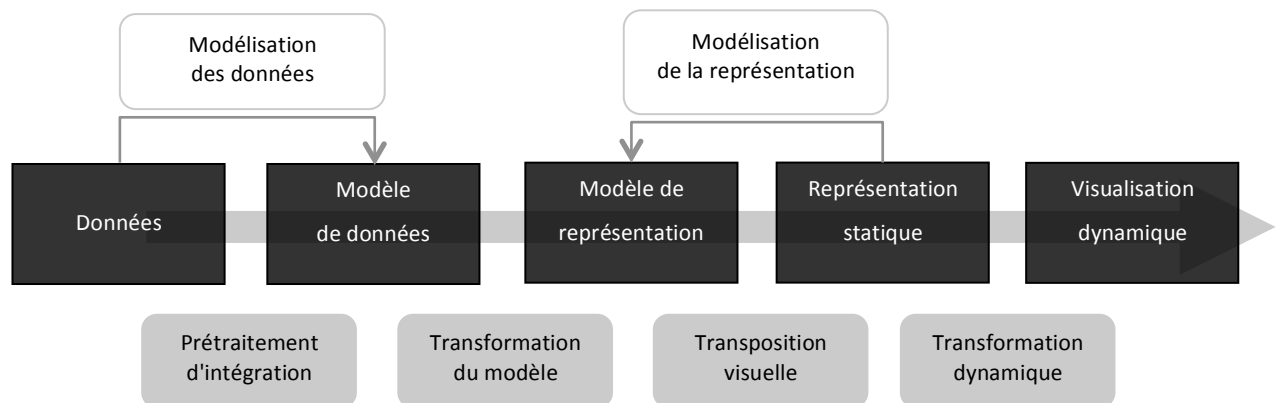
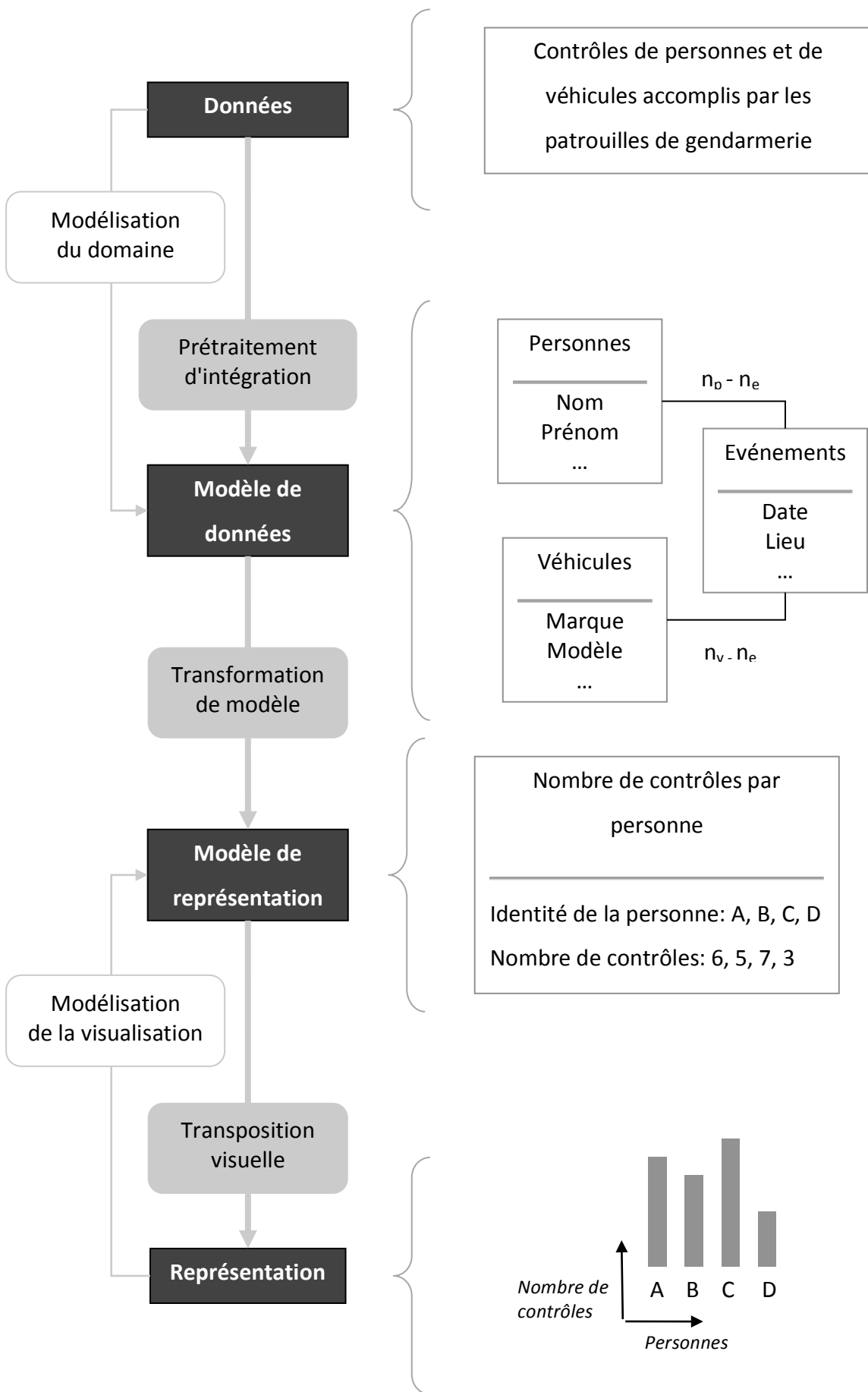


Figure 8 Processus de conception d'une visualisation

Ce processus décrit la relation qui lie les données et les vues par une combinaison de transformations entre **différents états**. Il permet une décomposition des problématiques rencontrées lors de l'élaboration d'une représentation graphique et met en évidence l'existence d'opérations (combinaisons de transformations) liées soit à la nature des données traitées, soit aux caractéristiques de la vue (Chi & Riedl, 1998). Afin d'explicitier le modèle, la visualisation des données issues des contrôles policiers de personnes et de véhicules est prise comme exemple plus loin. Cette formalisation met en évidence l'une des difficultés du processus : les représentations graphiques nécessitent une structuration de l'information qui n'est pas toujours en adéquation avec le modèle de données conçu lors de l'intégration. Cette problématique est identifiée, dans le domaine de la conception logicielle, sous le nom d'*Impedance Mismatch*⁶⁶. Les termes de représentation et de visualisation sont parfois utilisés pour différencier les représentations statiques des visualisations dynamiques⁶⁷ (Card et al., 1999).

⁶⁶ <http://c2.com/cgi/wiki?ImpedanceMismatch> (dernier accès le 14 août 2010)

⁶⁷ Des techniques de transformations dynamiques sont utilisées pour faciliter les interactions avec l'utilisateur ; telles que l'utilisation de popups, d'outils de sélection ou des distorsions de la vue.

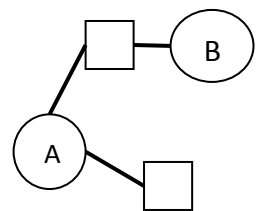
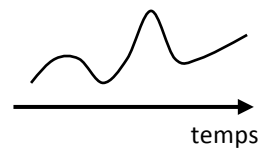
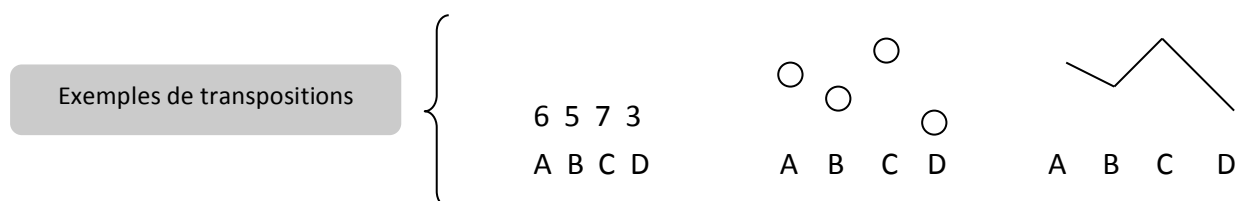


La première étape du processus consiste à intégrer les données sources brutes dans une ou plusieurs tables formant le **modèle des données**. Cette structure de données est définie par la modélisation du domaine pour le problème à traiter. Dans le processus d'analyse, cette étape est effectuée dans le processus d'intégration.

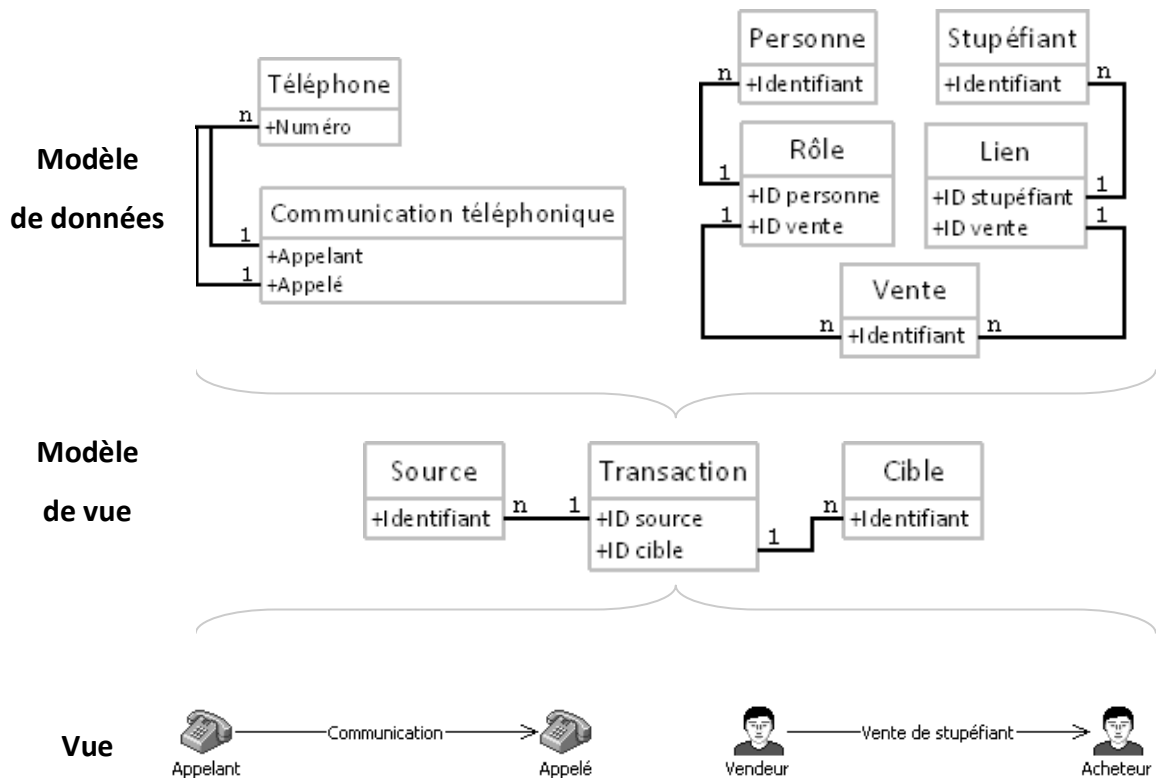
Toutefois ce modèle ne permet pas toujours une transposition directe vers la vue. Par exemple, afin de produire un histogramme du nombre de contrôles par personne, une table contenant les identités des individus et leurs nombres respectifs de contrôles est nécessaire. Un processus de transformation du modèle (constitué d'opérations spécifiques) est nécessaire pour transférer les données dans le **modèle de la représentation**. Ces opérations sont, pour cet exemple, une sélection des attributs d'intérêt (nom, prénom, date de naissance des personnes) et le calcul du nombre de cas par individu. Cette décomposition entre le modèle de données et le modèle de la vue, permet de décrire les visualisations indépendamment de la nature des données. Par ailleurs, pour un même modèle de données, plusieurs modèles de vue peuvent être exploités.

Le processus de modélisation de la visualisation définit la structure de la représentation (son modèle). Pour un histogramme il s'agit d'une table contenant une valeur numérique associée à chaque catégorie d'une variable. Le modèle est fonction de la vue (une carte, un schéma relationnel ou un graphique par exemple) et peut requérir la présence de données spécifiques (comme des données géocodées pour la production d'une carte).

Finalement, une **représentation** peut être produite sur la base du modèle de la représentation. Il est important de noter que plusieurs vues peuvent partager la même structure. En effet, sur la base d'un même modèle (de vue), plusieurs vues peuvent être produites:



Cette décomposition du processus de visualisation définie dans le domaine informatique a de fortes implications sur la conception d'un logiciel de visualisation. Malheureusement, la majeure partie des logiciels n'intègre pas ou mal la transformation de modèles. Des logiciels spécialisés sont parfois développés pour traiter cet aspect spécifiquement. Les problématiques informatiques liées à cette étape du processus ne seront pas abordées plus en détail. Afin d'illustrer le problème, un exemple est présenté. Il a été montré dans le chapitre précédent que la visualisation relationnelle de communications téléphoniques peut être établie de façon analogue à la visualisation de ventes de stupéfiants, par exemple. Ces deux problématiques possèdent toutefois des modèles de données différents.



Le modèle de vue est identique, mais pas les modèles des données sous-jacents. Une transformation du modèle doit donc être effectuée pour produire la représentation. Cette incompatibilité de modèles implique donc d'effectuer des traitements sur les données. Dans l'idéal ce processus devrait être automatisé et intégré dans l'outil.

M o d i f i c a t i o n d y n a m i q u e d e l a s t r u c t u r e d u g r a p h e

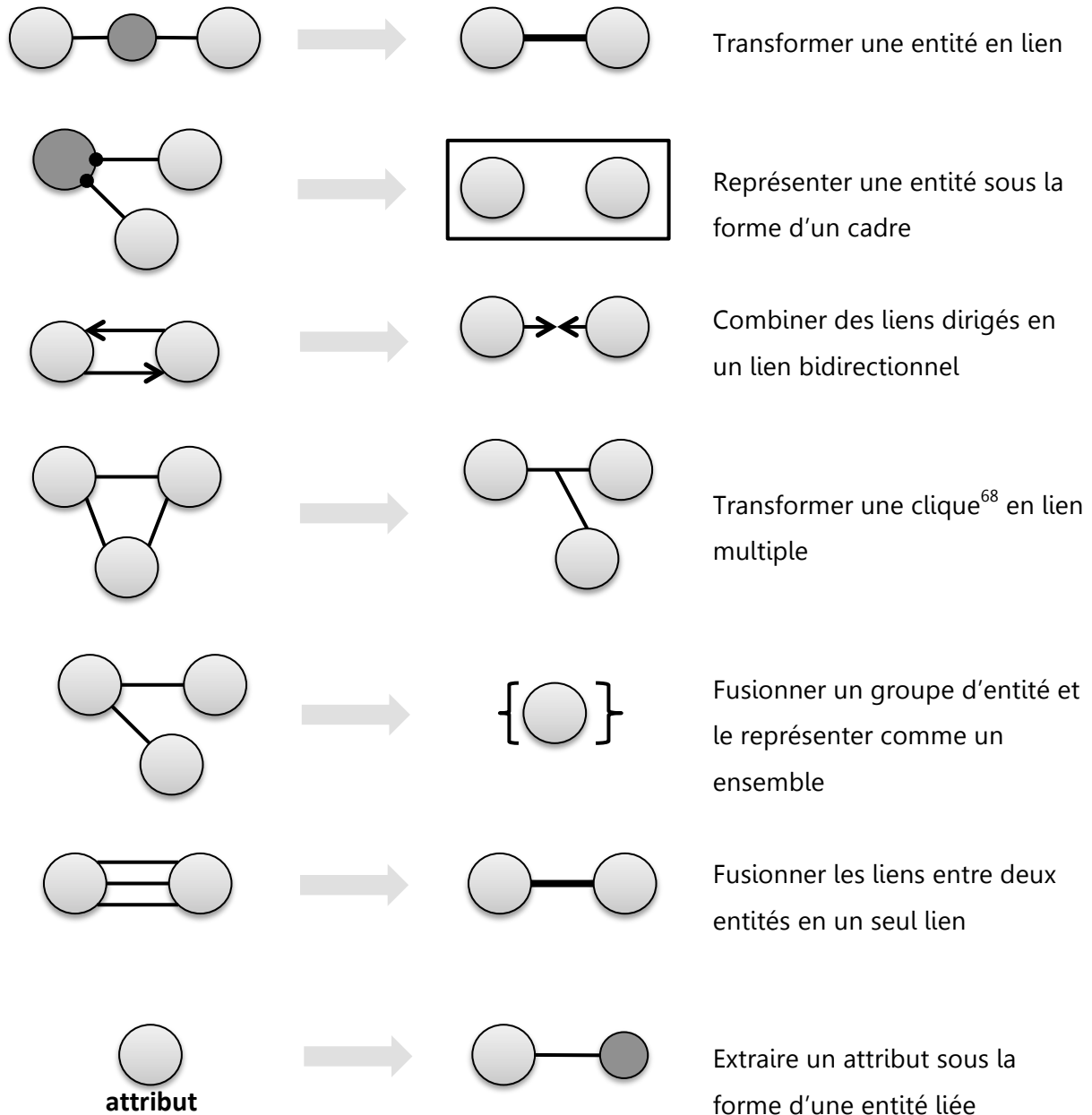
Afin de faciliter l'exploration des données, il serait très utile que les outils de visualisation relationnelle disposent de mécanismes de modification de la structure des graphes.

Actuellement, lorsque l'utilisateur souhaite effectuer des modifications, il peut soit les effectuer manuellement sur le graphe, soit réimporter les données préalablement modifiées.

Dans ce deuxième cas, si le modèle des données n'est pas adaptée pour la nouvelle structure visuelle souhaitée, des transformations manuelles et fastidieuses doivent être effectuées pour reconstruire un modèle de données adapté pour la vue.

Lors du traitement des différents cas présentés dans cette thèse, un certain nombre de transformations génériques fréquemment utiles ont été identifiées. La nécessité de telles transformations a été confirmée lors de discussions avec des praticiens. L'absence de cette fonctionnalité génère de nombreuses frustrations et des pertes de temps parfois importantes. Il est d'ailleurs intéressant de relever que, lors de formations, il est fréquent que les utilisateurs débutants identifient cette lacune et s'interrogent sur cette absence tant elle leur paraît frappante.

Le tableau ci-dessous décrit les principales transformations identifiées comme utiles pour explorer visuellement les informations sur un graphe. La liste n'est certainement pas exhaustive. Elle regroupe des types génériques de transformation et permet d'explicitier la nature des modifications fréquemment effectuées. Ces transformations sont évidemment définies en regard de la typologie de relations proposées. Elles permettraient notamment de faciliter la deuxième étape de la méthode qui consiste à définir qu'elle est la manière la plus adaptée de représenter les relations entre les entités pertinentes. En effet, cette étape n'est pas toujours évidente et elle pourrait être facilitée si l'outil permettait d'effectuer ces changements dynamiquement. Il serait alors possible de modifier la représentation en cours d'analyse et mettre en avant certaines relations plutôt que d'autres au fil des questionnements. Par exemple lors de l'analyse d'une série de cambriolages, extraire le lieu de commission du délit sous la forme d'une entité liée permet de visualiser rapidement les relations spatiales selon le principe de connectivité décrit précédemment.



Exemples de modifications dynamiques de la structure d'un graphe utiles lors de l'analyse des informations

La séparation du modèle de données et du modèle de vue constitue l'une des clés du développement de telles interactions. Les transformations devraient également être bidirectionnelles. En effet, les opérations inverses sont également utiles. Lors de la simplification d'un schéma, transformer une entité liée (comme le lieu de résidence d'une personne) en attribut permet de réduire le nombre d'entités et ainsi accroître la lisibilité globale de la représentation.

⁶⁸ Portion d'un graphe dont l'ensemble des sommets sont liés deux à deux.

Intégration des dimensions d'analyse

L'analyse simultanée des informations dans les quatre dimensions d'analyse constitue l'une des problématiques majeures de la visualisation de l'information. De nombreuses recherches ont été menées sur le sujet (notamment en matière d'analyse spatio-temporelle) depuis le début des années deux-mille. A l'époque, il était constaté que peu d'outils offraient des possibilités d'analyse dans plusieurs dimensions simultanément (Buetow et al., 2003). Depuis lors, de nouveaux outils performants ont été développés dans ce sens (voir notamment (Guo et al., 2006) et (Brunsdon et al., 2007)). Les logiciels actuels, tels que « Tableau Software », « Spotfire » et « Analyst's Notebook »⁶⁹ intègrent déjà plusieurs des dimensions d'analyse. La tendance actuelle est clairement de développer des outils intégrateurs, décomposant les problèmes selon ces dimensions afin de faciliter l'exploration des données. La démarche proposée rejoint cette tendance actuelle.

Un prototype de visualisation multidimensionnelle a été développé dans le cadre de cette recherche. La technique de représentation exploitée pour intégrer l'ensemble des dimensions dans un système commun et faciliter ainsi l'analyse dynamique des informations est présentée dans la section suivante.

Prototype de visualisation multidimensionnelle

Le développement d'un outil informatique n'était certainement pas un objectif fondamental de cette recherche. Plusieurs prototypes ont toutefois été développés pour tenter d'intégrer les dimensions d'analyse dans un seul logiciel et pour intégrer des mécanismes de modification dynamiques des vues. L'un des composants développés a montré son utilité dans le cadre d'affaires réelles. Il porte sur la combinaison des dimensions spatiales et temporelles. En effet, la visualisation spatio-temporelle constitue un sujet de recherche spécifique qui soulève de nombreuses questions et problèmes de représentation. Naviguer entre une représentation géographique et des vues temporelles cause des ruptures dans les processus de raisonnement. En effet, il n'est pas rare que l'analyste souhaite investiguer la distribution temporelle d'une série d'événements pour une région déterminée où comparer ces distributions pour un ensemble de sous-régions. Diverses stratégies de représentation spatiale et temporelle sont présentées dans l'annexe un, ainsi que des formes classiques de questionnements et diverses formes de relations temporelles et spatiales. Toutefois la

⁶⁹ <http://www.tableausoftware.com/>, <http://spotfire.tibco.com/> et <http://i2.co.uk/>, derniers accès le 22 janvier 2011

combinaison de ces deux types de représentation nécessite des efforts particuliers et le développement de nouvelles stratégies de représentation.

Soutenir l'interprétation des informations spatiotemporelles en visualisant les deux dimensions sur une même représentation ou en établissant un lien visuel entre plusieurs vues est crucial en analyse criminelle. En effet, les crimes sont commis dans un environnement social et physique qui définit des situations spécifiques (Felson & Clarke, 1998). Les contraintes imposées par le temps (notamment le moment de la journée) et l'espace influencent les activités des criminels. Ces comportements se manifestent selon des patterns spatio-temporels que l'analyste tente de discerner au sein des données collectées (telles qu'un historique de cas, des données de facturation téléphonique ou un relevé GPS). De nombreuses questions surviennent au cours de l'analyse pour comprendre et reconstruire ces activités. Par exemple, existe-t-il des régularités temporelles ou spatiales pour un phénomène criminel particulier? Quelles sont les relations entre les crimes formant une série particulière? Est-il possible de lier l'activité téléphonique d'un suspect avec la commission d'infractions? Quels étaient les déplacements d'un individu pour une période particulière? A chaque question posée correspond un ou plusieurs produit(s) de renseignement spécifique(s). Sur la base des régularités détectées, l'analyste peut alors rechercher des causes possibles, explicatives de la situation et sur lesquelles des recommandations peuvent être élaborées (par exemple, pour la mise en place de surveillances ou pour développer des stratégies de prévention).

Diverses techniques de représentation spatio-temporelle ont été développées. Elles sont présentées ci-dessous. Il faut relever que chaque dimension soulève déjà de nombreux problèmes spécifiques liés à la qualité et la quantité d'informations accessibles : comment représenter l'imprécision et l'incertitude temporelle, comment gérer la superposition de symbole sur une carte, comment choisir des couleurs appropriées et les niveaux d'agrégation adaptés pour ne pas biaiser la perception? Combiner plusieurs dimensions amplifie ces difficultés et les visualisations produites sont parfois inexploitables. L'approche développée dans le cadre de cette recherche se base sur l'utilisation de représentation en deux dimensions (pour des raisons explicitées dans l'annexe deux). Une approche évidente d'intégration de la dimension temporelle sur une carte, consiste à représenter le temps par la couleur des symboles. Une autre approche consiste à exploiter des signes particuliers comme des flèches par exemple. Des techniques d'interpolations sont également exploitables pour représenter le temps par une variation de couleurs. Ces techniques ne fonctionnent cependant pas lorsque les événements représentés se superposent spatialement. En effet, les symboles se chevauchent et l'interprétation est alors difficile voire impossible. Trois approches principales

ont été développées pour gérer la superposition spatiale des points : les animations, les treillis et lier des vues par le symbole.

A n i m a t i o n s

Certaines études ont montré que l'utilisation d'animations pour représenter des données spatio-temporelles semble être un moyen efficace offrant de nombreuses perspectives (Brunsdon et al., 2007). En effet, une animation est souvent intuitive et elle peut être directement associée au temps. Cependant, les animations se basent sur les capacités de mémorisation de l'utilisateur. Cette contrainte pose problème lorsque celles-ci sont longues. De plus les animations nécessitent d'être contrôlées par l'utilisateur (notamment au niveau de la vitesse de défilement des cas) et des supports de transmission adaptés doivent être développés. Dans le cadre de l'enquête judiciaire, des représentations annexées à un rapport écrit sont souvent souhaitables. L'usage d'animations est donc relativement peu fréquent et moins pratique que l'usage de représentations statiques.

T r e i l l i s (c o m a p , s m a l l - m u l t i p l e s)

Un treillis est la juxtaposition de plus cartes en parallèle. Chaque carte peut alors représenter la répartition spatiale pour une période temporelle définie (Keim et al., 2005). L'analyse spatio-temporelle est effectuée par comparaison des cartes. Tufte (1990) décrit cette technique sous le nom de « small-multiples ». Il relève notamment que l'ensemble des diagrammes alignés côte-à-côte doit pouvoir être observé en une seule fois, afin de d'effectuer des comparaisons sur l'ensemble des données. Selon lui, cette technique fait partie des meilleures solutions exploitables pour effectuer des comparaisons entre plusieurs groupes d'intérêts. Cette technique n'est cependant efficace que pour un nombre limité de cartes et donc pour un nombre de limité de périodes. Dans le cas contraire, le nombre de comparaisons et le temps de lecture nécessaires pour interpréter les résultats sont généralement trop importants. Certains chercheurs rejettent d'ailleurs l'hypothèse que cette technique est efficace, car ils estiment que le processus de comparaison est difficile voire imprécis (MacEachren, 2004).

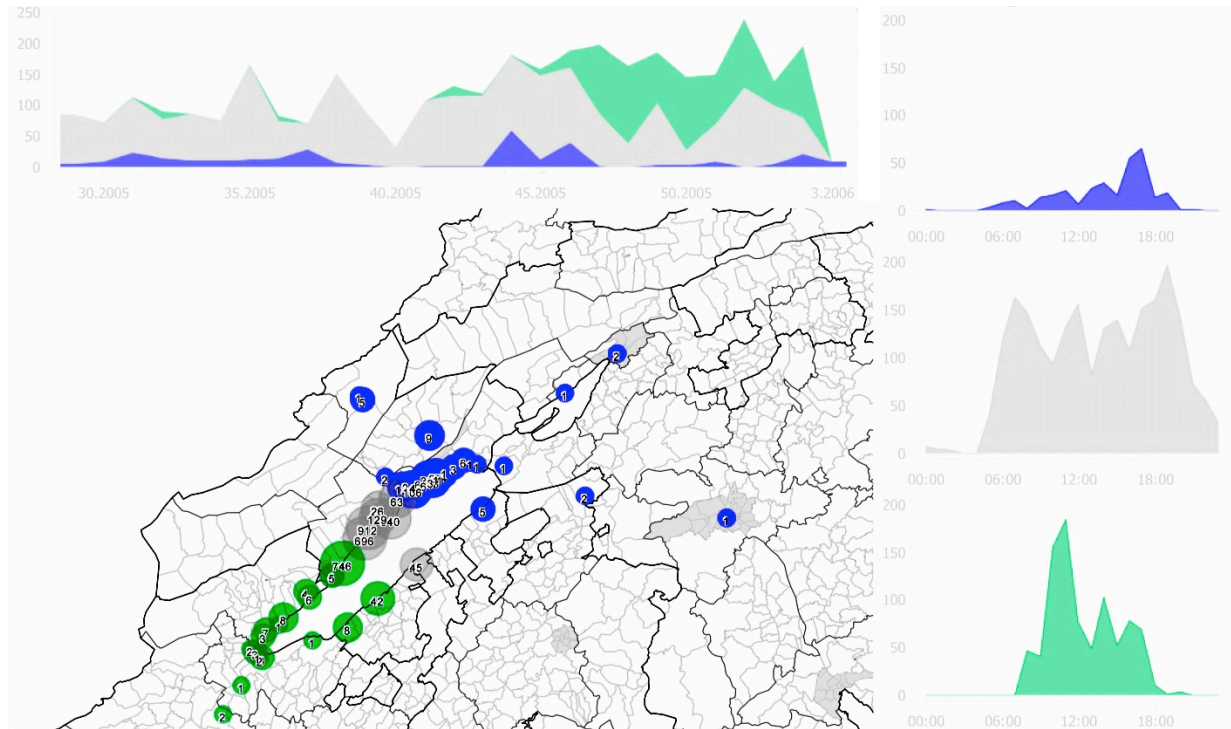
L i e r d e s v u e s (f o c u s i n g , l i n k i n g a n d a r r a n g i n g)

Une troisième technique consiste à exploiter plusieurs vues séparées pour explorer les données en parallèle pour chaque dimension. Diverses techniques peuvent être utilisées pour lier les vues suivant que la représentation est dynamique ou non, et que les vues sont présentées simultanément ou en séquence (Fredrikson et al., 1999). Par exemple, un moyen courant consiste à utiliser des techniques de sélections. Lorsque l'utilisateur sélectionne une zone particulière sur l'une des vues, les données sont filtrées ou mises en valeur sur les autres vues (technique dite du Brushing (Keim et al., 2005) (Chen & Czerwinski, 2000)). Une autre approche consiste à exploiter la couleur pour représenter un attribut particulier. Les couleurs ainsi définies sont reportées dans chaque vue (cette technique est notamment exploité dans (Guo et al., 2006)). Lier des vues à l'aide de la couleur est généralement fait sur la base d'un attribut de la dimension relationnelle (comme le type d'événement, l'âge d'une personne, etc.)

Cette technique est reconnue comme efficace et exploitée par de nombreux outils (Brunsdon et al., 2007). L'hypothèse fondamentale de cette technique est qu'il est plus efficace de chercher à combiner plusieurs vues spécifiques et adaptées pour représenter des dimensions particulières, plutôt que de tenter de développer une vue unique optimale (MacEachren, 2004). L'approche proposée se base sur cette technique de représentation combinée à la méthode des treillis.

Approche développée

L'approche imaginée pour développer le prototype est issue du travail de master en sciences forensiques effectué par Bénédicte Heidl (Heidl, 2007). L'outil développé est disponible en ligne⁷⁰.



La méthode de représentation développée est basée sur l'idée que l'analyse visuelle est facilitée en intégrant la dimension temporelle sur des vues spécifiques plutôt que directement sur une carte. Le lien entre les vues est réalisé par un usage spécifique de la couleur. La distribution géographique est divisée en sous-régions et une couleur est attribuée à chacune d'elles. Pour chaque sous-groupe d'observations (des cas, des positions GPS ou dans le cas présenté ci-dessus des appels téléphoniques), la couleur est reportée sur les vues temporelles afin de représenter chaque distribution. Dans cet exemple, le diagramme du haut est une ligne de temps linéaire (les données sont représentées par semaine). Le diagramme de droite exploite la technique des treillis pour représenter les distributions par heure de la journée pour chaque sous-groupe.

⁷⁰ <http://www.analysecriminelle.org/visualist/>, dernier accès le 23 janvier 2011.

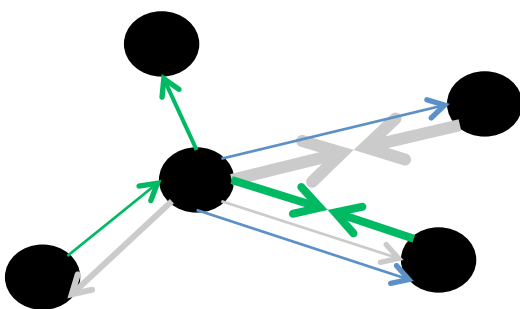
Afin de faciliter l'analyse, le prototype permet d'effectuer diverses modifications dynamiques (pour plus de détails, voir les descriptions des dimensions temporelle et spatiale de l'annexe un, ainsi que les formes de représentations associées):

1. Les groupes peuvent être personnalisés par sélection des cas sur la carte.
2. Les niveaux d'agrégations sont modifiables dynamiquement sur les vues temporelles. Les données peuvent être représentées linéairement (par jour, semaine, mois et par années) et cycliquement (par heure de la journée, jour de la semaine, jour du mois et par jour de l'année).
3. Les niveaux d'agrégations peuvent être modifiés sur la carte : au niveau des observations ou par regroupement selon des régions administratives telles que les communes, districts, pays, etc. Différents type de représentation peuvent être exploités (tels que décrit dans l'annexe un) : carte à points, à symboles proportionnels ou choroplèthe.

Le prototype nécessite cependant des développements afin de pouvoir effectuer des sélections dynamiques sur les représentations temporelles et créer des groupes en fonction du temps.

P e r s p e c t i v e

Finalement, la démarche développée permet également d'intégrer les informations spatiales sur des représentations relationnelles :



Les couleurs correspondant aux zones géographiques définies sur la carte peuvent notamment être intégrées sur les liens d'un schéma relationnel. Par exemple, lors de l'analyse de données de téléphonie les communications effectuées dans chaque région peuvent être investiguées dans la dimension relationnelle. Le schéma permet

ainsi d'identifier rapidement les correspondants d'un numéro en fonction des régions où celui-ci était localisé. La démarche permet ainsi d'analyser un jeu de données dans les trois dimensions simultanément et dynamiquement. A noter que l'intégration de la dimension relationnelle n'a pas encore été implémentée.

DISCUSSION GÉNÉRALE

Résultats empiriques

Les évaluations préliminaires et les études de cas réalisées mettent en évidence de réelles difficultés de conception des schémas relationnels. Concevoir un schéma efficace pour répondre à une question d'analyse requiert des capacités spécifiques en termes de modélisation. L'élaboration d'un schéma nécessite en effet de savoir isoler les aspects essentiels du problème et d'identifier clairement les entités pertinentes sur lesquelles portent les questionnements et la recherche de relations. La réalisation d'une visualisation ne relève pas uniquement de l'utilisation d'un outil informatisé, mais constitue essentiellement une activité de modélisation. Cela confirme la nécessité d'étendre la formation au-delà de la maîtrise des outils informatisés.

Les trois cas présentés pour illustrer les variabilités de représentation ont été sélectionnés parmi un ensemble d'exercices effectués par des étudiants et des praticiens. Plusieurs autres exercices portant sur diverses thématiques (telles que l'analyse de trafics de stupéfiants, de série de cambriolages ou d'analyse de spams par exemple) ont également mis en exergue les difficultés observées. L'échantillon présenté dans cette thèse n'a pas été sélectionné pour amplifier le constat. Au contraire, les exemples portant sur une seule phrase ont été choisis pour leur simplicité, afin de montrer que la complexité des informations n'est pas la seule cause des variations. Le troisième cas a été choisi en raison du nombre de participants afin de présenter les difficultés sur un échantillon le plus représentatif possible.

L'absence d'un consensus global semble notamment être le résultat de différences de perception sur les objectifs d'exploitation de la visualisation. Mémoriser l'ensemble des faits, détecter des incohérences, chercher de nouvelles pistes d'enquête, résumer les éléments essentiels sont autant de perspectives différentes pouvant expliquer les variations. Les difficultés observées pour identifier les entités et définir leurs relations indiquent également que le processus n'est pas évident et nécessite un certain entraînement. L'absence d'une méthode clairement formalisée est certainement l'une des causes de ce constat.

Finalement, l'impact des choix de représentation sur l'utilisabilité des représentations a été discuté par une série d'études de cas. Les exemples présentés sont des cas particuliers, mais ils sont issus de cas réels et illustrent des problèmes récurrents. Ils attestent de l'importance de la lisibilité et de l'intégrité des représentations. Des ambiguïtés peuvent, en effet, engendrer des difficultés et des erreurs d'interprétation, voire l'impossibilité d'exploiter la représentation de façon efficace. Bien que largement exploitée pour soutenir les raisonnements et faciliter la communication, la visualisation soulève de réelles questions méthodologiques. L'absence d'une formalisation claire du langage (Reed & Kocura, 2005), la masse d'information à représenter (Schroeder et al., 2007) et l'identification des concepts pertinents et de leurs relations (Stasko et al., 2008) sont autant de causes supposées des difficultés. La réalité de ces difficultés reste néanmoins implicite dans ces études. Dans ce contexte, les évaluations effectuées fournissent une image empirique de l'étendue des variations observables et présentent, par des exemples réels, l'impact de choix inadaptés sur l'analyse et la communication.

M é t h o d e s d ' é v a l u a t i o n

Lors de la présentation du plan de recherche, la question de l'évaluation des recommandations méthodologiques et techniques proposées a été soulevée. Un effort particulier avait déjà été effectué pour récolter des données empiriques afin d'explicitier et évaluer la réalité et l'ampleur de la problématique. L'utilisation des exercices rendus lors de formations a été choisie comme stratégie.

Cette démarche a permis d'obtenir plusieurs échantillons sans mettre en place des expérimentations spécifiques trop difficile à mettre en œuvre. Les représentations ont été produites sans pression de temps et selon des objectifs pédagogiques. Elles n'ont donc pas été produites dans l'objectif spécifique des évaluations, mais dans des conditions réelles

d'apprentissage. Il faut toutefois relever que les conditions n'étaient pas formellement contrôlées. En effet, les participants pouvaient librement discuter entre eux et consulter le travail de leurs voisins. Une grande diversité de représentation a néanmoins été observée.

Par ailleurs, l'évaluation effectuée sur les modifications du type de représentation (une carte, un histogramme et un diagramme circulaire) s'est avérée très intéressante et utile d'un point de vue qualitatif. En effet, la comparaison des constatations et des analyses effectuées par les étudiants a permis d'illustrer l'impact de choix de représentation sur l'analyse, mais également de profiter des commentaires des participants sur les représentations elles-mêmes. En effet, la totalité des observations discutées sur les avantages et limites de chaque représentation a été relevée par les participants. Une telle démarche facilite la discussion en profitant des commentaires de l'ensemble des participants. Une approche similaire aurait certainement pu être mise en place lors de l'évaluation de la typologie de relations proposée. Cette dernière a cependant répondu à l'objectif défini. Elle a confirmé l'impact sur l'efficacité de la structure des schémas. Les différences d'exactitude observées n'étaient d'ailleurs *a priori* pas attendues. Les représentations ayant été conçues en respectant l'intégrité des données sous-jacentes, des différences de temps de réponse étaient plutôt supposées. La lisibilité d'un graphe semble ainsi avoir un impact important sur l'utilisabilité d'une représentation.

Apports sur un plan méthodologique

Dimensions d'analyse

Afin d'exploiter de manière efficace la visualisation, l'analyste doit choisir le type de représentation le plus adapté pour répondre à la question d'analyse. Il s'agit certainement de la première difficulté rencontrée lors de la conception d'une représentation. La décomposition des questions d'analyse selon quatre dimensions principales semble offrir une méthode utile pour guider ces choix. L'analyste identifie quels sont les paramètres importants dans chaque dimension et identifie la dimension dominante qui guide alors la sélection d'une représentation particulière. La validité de cette décomposition n'a pas été formellement évaluée. Il faut cependant relever qu'une telle approche est également exploitée dans d'autres domaines que l'analyse criminelle, comme en témoigne les nombreux ouvrages et articles sur l'analyse temporelle, spatiale, relationnelle ou quantitative. Lors de formations, il a également été constaté que décomposer les problèmes ainsi permet, d'une part, d'explicitier les causes de

mauvais choix de représentation et, d'autre part, de présenter en détails les choix qui ont mené aux solutions proposées.

Processus de conception d'un schéma relationnel

La formalisation proposée a permis de mettre en évidence d'autres sources de difficulté. Choisir un signe (une entité, un lien ou un cadre par exemple) pour représenter un concept n'est pas un processus direct. En effet, pour chaque symbole une signification spécifique est définie (telle qu'une relation dirigée pour une flèche). Choisir la manière de représenter un concept revient donc à identifier quel signifié est le plus adapté. La définition d'une typologie de relations est proposée afin de définir clairement le langage est ainsi faciliter ces choix. L'évaluation empirique effectuée a montrée sur des cas précis que certaines formes de représentation des relations sont plus efficaces et efficientes pour des tâches précises. Ces choix sont donc importants est doivent être guidés.

L'approche globale proposée pour concevoir un graphe et gérer les difficultés liées à la complexité des informations à représenter se fonde sur un usage adapté de structures graphiques et des propriétés visuelles pour améliorer la lisibilité de la visualisation par diverses stratégies de simplification. Cette approche est certainement novatrice en analyse criminelle. Elle se distingue des approches proposées dans la littérature. En effet, de nombreux travaux ont été menés pour améliorer les algorithmes de placement des nœuds, sans prendre en compte la sémantique des concepts représentés et la possibilité d'exprimer des relations par le placement. L'approche proposée repose notamment sur un placement des entités en regard de leurs types ou afin de distinguer des groupes. L'approche est certes manuelle, mais elle permet d'obtenir un résultat visuel efficace. Par ailleurs, l'approche des réseaux sociaux qui connaît un essor particulier pour résoudre le problème de la surcharge d'information, n'améliore pas la qualité visuelle et l'utilisabilité des représentations. Elle soutient l'analyse en facilitant la détection des entités centrales d'un réseau par exemple, mais n'améliore pas la représentation à des fins de communication. L'approche proposée offre quant à elle des pistes pour exploiter de manière efficace la visualisation tant pour l'analyse que la communication en améliorant la lisibilité du graphe.

Afin de compléter la validation la démarche proposée, des expérimentations subséquentes pourraient être effectuées. L'utilisabilité de la typologie de relations proposée repose sur la capacité de l'analyste à identifier les types d'entités et de liens investigués. Il n'est pas certain que cette étape soit évidente. C'est pourquoi, le développement d'un catalogue de bonnes

pratiques est proposé. Il faut relever ici que la démarche générale proposée a été formalisée sur la base des situations récurrentes présentées et des solutions identifiées pour les traiter. Même si la méthode est probablement applicable à d'autres situations, il n'est absolument pas certain que les catégories de relations identifiées soient exhaustives. Il semble toutefois que le niveau d'abstraction auquel elles ont été définies devrait permettre de les exploiter pour la résolution de nouveaux problèmes, voire d'étendre ultérieurement la classification.

P a t t e r n s d e v i s u a l i s a t i o n

Les solutions proposées bien qu'ayant démontré leur utilité dans le cadre d'affaires réelles, ne sauraient être considérées comme les meilleures possible. La découverte de nouvelles méthodes ou l'application de meilleures stratégies de représentation peuvent remettre en question les solutions proposées et enrichir les résultats. L'optimisation de la méthode de représentation des informations est donc une question permanente. Dans cette perspective, le regroupement de solutions graphiques ayant montré une certaine efficacité est nécessaire. L'élaboration d'un catalogue de situations récurrentes a été amorcée dans cette optique. Cette approche fixe un cadre qui ouvre la voie à des développements ultérieurs.

P e r s p e c t i v e s

D é v e l o p p e m e n t s e n t e r m e s d e r e c h e r c h e

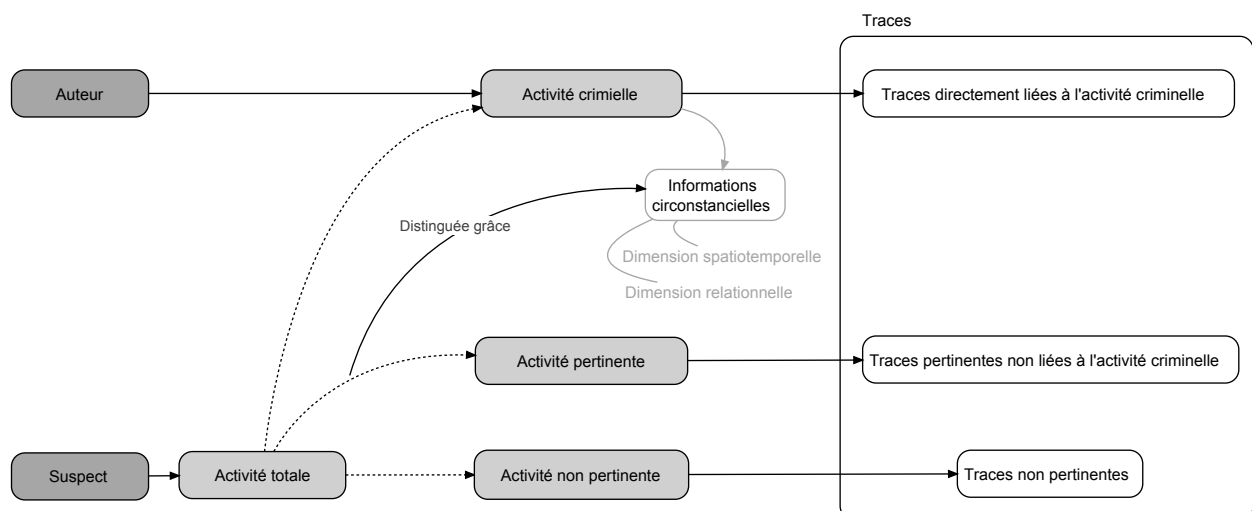
Cette recherche s'est essentiellement focalisée sur l'exploitation de la visualisation pour l'analyse relationnelle dans le cadre très opérationnel des enquêtes criminelles. La visualisation a un rôle plus large en matière d'analyse de la criminalité, dans l'ensemble des dimensions définies et tant d'un point de vue opérationnel que stratégique. Diverses perspectives complémentaires sont donc envisageables pour mieux comprendre les rôles, les difficultés et les apports de la visualisation dans le domaine des sciences criminelles. D'un point de vue des méthodes, les dimensions d'analyse définies forment un cadre certainement utile pour explorer et mieux comprendre les exploitations possibles de la visualisation pour soutenir les processus d'analyse des données collectées et de communication des renseignements produits.

La visualisation s'intègre en effet dans de nombreux processus du traitement de l'information.

A titre d'exemples :

- La lecture et l'interprétation des résultats fournis par électrophorèse lors de l'analyse d'un profil ADN est facilitée par des représentations graphiques.
- Le profilage des stupéfiants exploite également des visualisations. Des chromatogrammes sont utilisés pour vérifier que l'analyse c'est bien déroulée (que les substances se sont bien séparées). La visualisation est également exploitée pour ajuster le calcul de l'intégrale des pics en ajustant manuellement sur la représentation la base des pics.
- L'interprétation de la valeur indiciaire des traces peut être effectuée au moyen de réseaux bayésiens qui exploitent des graphes pour décrire les relations de causalités entre des variables.
- La visualisation joue un rôle certainement central lors de la présentation des indices. Investiguer l'impact de l'usage de représentations sur la prise de décision fait également l'objet de recherche dans le domaine judiciaire.
- L'analyse globale de phénomène criminel en criminologie profite également des possibilités de la visualisation pour mieux appréhender et comprendre l'ampleur, l'évolution et l'étendue spatio-temporelle des problèmes.

Dans le cadre de l'analyse criminelle opérationnelle, les patterns définis ont été formalisés en regard d'objectifs d'analyse précis. Ils sont cependant largement influencés et orientés en fonction de la nature des informations représentées. Il serait certainement souhaitable d'étendre la démarche et de formaliser les pratiques en regard des types de raisonnement exploités. A titre d'exemple, lors de l'analyse de données de téléphonie, les traces récoltées sont issues de l'activité globale de suspects. Les indices en lien avec une activité suspecte doivent donc être distingués de l'ensemble des communications non pertinentes qui forment un bruit de fond important. L'analyse des traces s'effectue au niveau de l'activité en exploitant notamment les informations circonstancielles du cas :



Certains exemples d'analyse présentés exploitent ce genre de raisonnement. La visualisation est exploitée pour détecter les traces d'activités, directement ou indirectement liées aux activités criminelles, en se basant sur des informations circonstanciées du cas telles que des données spatio-temporelles ou des relations supposées avec des individus particuliers. Il serait certainement utile d'identifier et de formaliser les formes d'exploitation de la visualisation en fonction des formes de raisonnement effectués lors des analyses.

E c h a n g e d e c o n n a i s s a n c e s

Le catalogue de patterns amorcé dans le cadre de cette recherche est fondé principalement sur les cas traités par l'auteur au cours des six années passées à collaborer sur des enquêtes. Les situations présentées ne sont toutefois pas spécifiques à un cas particulier, mais décrites à un niveau d'abstraction qui devrait permettre de ré-exploiter les solutions sur des cas analogues. Il faut noter qu'il a été difficile de récolter des exemples traités par des collègues en raison notamment de la sensibilité des informations représentées. De nombreuses discussions avec des collègues analystes et enquêteurs ont toutefois permis d'enrichir le contenu des solutions proposées. A l'exception de certaines propositions liées au développement du langage visuel, les patterns présentés ne sont certainement pas novateurs pour des analystes expérimentés. Ils ont probablement une utilité plus grande pour des analystes débutants. La prochaine étape est certainement de mettre à disposition ces propositions afin de stimuler l'échange d'expériences et faire évoluer les propositions vers des solutions peut-être plus efficaces et innovantes et étendre le recueil de bonnes pratiques. Une plateforme en ligne intégrant des patterns de visualisations est certainement un développement nécessaire pour stimuler ces échanges.

I m p a c t s u r l ' e n s e i g n e m e n t

Les résultats obtenus et les recommandations formalisées ont une utilité directe en matière d'enseignement. Ils ont d'ailleurs déjà été intégrés lors de formations à l'analyse criminelle opérationnelle. Les méthodes de visualisation sont cependant exploitées de manière plus large. En criminologie par exemple, il est intéressant de relever que les méthodes d'évaluations statistiques sont très courantes et largement enseignées. La visualisation en tant que méthode complémentaire reste quant à elle principalement du domaine de l'apprentissage personnel. Alors que les statistiques permettent de tester des hypothèses, la visualisation peut jouer un rôle important en amont lors de l'élaboration d'hypothèses. Il est d'ailleurs très fréquent de trouver des chapitres sur le sujet dans des ouvrages sur la fouille de données et l'analyse quantitative notamment. La représentation de l'information fait partie des méthodes facilitant

l'exploration des données. Connaître les possibilités qu'elle offre, savoir évaluer ses limites et comprendre comment l'exploiter de manière efficace pour analyser des informations sont autant de motivations qui conduisent d'ailleurs au développement de programmes d'enseignement en la matière. Ils restent cependant pour la plupart intégrés dans des cursus spécifiques, tels que le développement informatique ou la pédagogie par exemple.

D é v e l o p p e m e n t s i n f o r m a t i q u e s

L'exploitation de logiciels de visualisation a visiblement été perçue au cours des vingt dernières années comme une solution aux problèmes de gestion et d'analyse des informations récoltées au cours des enquêtes et pour l'analyse des phénomènes de criminalité. De nombreuses bases de données et divers outils de visualisation se sont multipliés dans les services de police. Ils ont certainement amélioré l'efficacité des processus de travail, mais ils ont également imposé des standards de travail et ajouté des contraintes définies par leurs fonctionnalités. Les observations effectuées dans cette recherche et les recommandations définies devraient permettre de mieux exploiter ces techniques. Elles présentent également des pistes de développement certainement utiles voire nécessaires. Les outils exploités devraient notamment être repensés afin de mieux guider les choix de conception et pourraient intégrer les éléments de méthode décrits.

Le prototype développé a d'ailleurs répondu à une lacune existante au niveau de l'analyse « multidimensionnelle » des informations. Il a prouvé son utilité dans le cadre d'affaires criminelles et est toujours exploité. De tels développements poussent notamment à clarifier les pratiques et formaliser les méthodes.

CONCLUSION

Dans ce travail, nous avons montré que l'exploitation des méthodes de visualisation en analyse criminelle soulève de nombreuses difficultés méthodologiques. Le processus systématique de l'analyse criminelle se veut global et n'intègre pas formellement une étape de visualisation de l'information, alors que celle-ci forme certainement une partie centrale des produits de renseignement.

Les schémas relationnels exploités dans les enquêtes judiciaires ont été choisis comme sujet d'étude. Des manuels de bonnes pratiques existent, mais les niveaux de généralité des recommandations définies sur la conception de tels schémas sont variables et elles ne sont pas présentées dans un cadre clairement formalisé.

Des évaluations préliminaires ont été effectuées afin de mesurer l'ampleur des difficultés. Elles fournissent une image empirique de l'étendue des variations observables lors de la conception des schémas et présentent, par des exemples réels, l'impact de choix inadaptés sur l'analyse et la communication. Elles révèlent l'importance de consolider les méthodes. La lisibilité des schémas produits semble notamment avoir un impact important sur leur utilisabilité.

Dans ce contexte, une double approche a été entreprise. D'une part, des règles et des recommandations générales issues notamment d'études sur la visualisation ont été recherchées afin de proposer une démarche de conception des schémas relationnels:

- Une description du processus de l'analyse criminelle en cinq étapes (acquisition, intégration, représentation, analyse et communication) permet de clarifier les rôles de la visualisation dans la méthode : (1) faciliter la mémorisation, (2) soutenir le processus

d'exploration des informations, (3) soutenir le processus de validation des hypothèses, (4) faciliter la communication du renseignement produit.

- Une décomposition des questions d'analyse est proposée afin de faciliter la sélection d'une méthode de visualisation adaptée. L'approche retenue décompose l'objectif d'analyse selon une ou plusieurs dimensions dominantes : temporelle, spatiale, relationnelle et quantitative.
- Plusieurs méthodes réduisant la complexité des schémas et ainsi augmentant leur lisibilité sont présentées.

D'autre part, l'étude de cas spécifiques a permis d'identifier des situations récurrentes pour lesquelles des représentations efficaces ont pu être définies. Le terme de pattern est utilisé pour décrire ces bonnes pratiques. Chaque pattern est composé de quatre parties : le contexte d'exploitation, le ou les objectifs d'analyse, les concepts à représenter et la solution graphique proposée. L'objectif est d'amorcer le développement d'un catalogue de patterns.

La démarche effectuée a donc permis de définir des recommandations générales facilitant la conception des schémas relationnels. Elle a également permis d'identifier des bonnes pratiques répondant à des problèmes spécifiques, dans l'optique de développer un catalogue de patterns facilitant l'analyse de situations récurrentes de l'analyse criminelle opérationnelle. Par l'échange de connaissances entre praticiens, une telle démarche pourrait s'étendre à l'ensemble des méthodes de la discipline et ainsi améliorer l'enseignement notamment.

Cette recherche s'est essentiellement focalisée sur l'exploitation de la visualisation dans le cadre très opérationnel des enquêtes criminelles. Diverses perspectives complémentaires sont envisageables pour mieux comprendre les rôles, les difficultés et les apports de la visualisation pour analyser les données de la criminalité, tant d'un point de vue opérationnel que stratégique. D'un point de vue des méthodes, les dimensions d'analyse définies forment un cadre certainement utile pour explorer et mieux comprendre les exploitations possibles de la visualisation dans le domaine des sciences criminelles.

BIBLIOGRAPHIE

- ADDERLEY, R. W. & MUSGROVE, P. (2001). Police crime recording and investigation systems - A user's view. *Policing* 24, no. 1: 100-114.
- AEPLI, P., RIBAUX, O. & SUMMERFIED, E. (2011). *Decising making in policing*. Lausanne, Switzerland: EPFL Press.
- AGIS. (2004). *Le guide méthodologique et des bonnes pratiques de l'analyse financière criminelle*. Commission européenne (direction générale justice et affaire intérieures).
- AIGNER, W., BERTONE, A., MIKSCH, S., TOMINSKI, C. & SCHUMANN, H. (2007). Towards a conceptual framework for visual analytics of time and time-oriented data. Dans *Proceedings of the 2007 Winter Simulation Conference*, éd. S. G. Henderson, B. Biller, M.-H. Hsieh, J. Shortle, J. D. Tew, & R. R. Barton, 721-729. WSC '07. Piscataway, NJ: IEEE Press.
- ALEXANDER, C. (1979). *The timeless way of building*. New York, NY: Oxford University Press.
- AMAR, R., EAGAN, J. & STASKO, J. (2005). Low-level components of analytic activity in information visualization. Dans *Proceedings of the 2005 IEEE Symposium on Information Visualization*, 111–117. INFOVIS '05. Minneapolis, MN: IEEE Computer Society.
- AMAR, R. & STASKO, J. (2005). A knowledge task-based framework for design and evaluation of information visualizations. Dans *Proceedings of the 2004 IEEE Symposium on Information Visualization*, 143–150. INFOVIS '04. Austin, TX: IEEE Computer Society.
- ANDERSON, T. J. (2007). Visualization tools and argument schemes: a question of standpoint. *Law, Probability and Risk* 6, no. 1-4: 97-107.

- ANDRESEN, M. A., WUSCHKE, K., KINNEY, J. B., BRANTINGHAM, P. J. & BRANTINGHAM, P. L. (2009). Cartograms, crime, and location quotients. *Crime Patterns and Analysis* 2, no. 1: 31-46.
- ANDRIENKO, N. & ANDRIENKO, G. (2005). *Exploratory analysis of spatial and temporal data: a systematic approach*. 1er éd. Berlin Heidelberg: Springer.
- ANSELIN, L., COHEN, J., COOK, D., GORR, W. & TITA, G. (2000). Spatial analyses of crime. *Criminal Justice* 4: 213–262.
- ANSELIN, L. & GETIS, A. (1992). Spatial statistical analysis and geographic information systems. *The Annals of Regional Science* 26, no. 1: 19-33.
- ATKIN, H. (1998). Keep it simple: a practitioners approach to applying the intelligence process, from formulating premises to recommendations. *IALEIA Journal* 11, no. 2: 1-11.
- ATKIN, H. (2000). Criminal intelligence analysis: a scientific perspective. *IALEIA Journal* 13, no. 1: 1-15.
- ATZENBECK, C., HICKS, D. L. & MEMON, N. (2009a). Supporting emergent knowledge and team communication in police investigations. Dans *Proceedings of the 2009 IEEE international conference on Intelligence and Security Informatics*, 95-100. ISI '09. Richardson, TX: IEEE Press.
- ATZENBECK, C., OZGULYZ, F. & HICKS, D. L. (2009b). Linking and organising information in law enforcement investigations. Dans *Proceedings of the 2009 13th International Conference Information Visualisation*, 443-449. IV '09. Barcelona, Spain: IEEE Computer Society.
- BARCLAY, D. (2009). Using forensic science in major crime inquiries. Dans *Handbook of forensic science*, éd. J. Frazer & R. Williams, 337–358. Cullompton, Willan.
- BARTHÉLEMY, M., CHOW, E. & ELIASSI-RAD, T. (2005). Knowledge representation issues in semantic graphs for relationship detection. Dans *Proceedings of the AAAI Spring Symposium on AI Technologies for Homeland Security*, 91-98. Stanford, CA: AAAI Press.
- BERTIN, J. (2005). *Sémiologie graphique: les diagrammes - les réseaux - les cartes*. 4 éd. Paris: Les ré-impressions des Editions de l'Ecole des Hautes Etudes en Sciences Sociales.
- BESSON, J.-L. (2004). *Les cartes du crime*. 1er éd. Paris: Presses Universitaires de France.
- BEX, F., VAN DEN BRAAK, S. W., VAN OOSTENDORP, H., PRAKKEN, H., VERHEIJ, B. & VREESWIJK, G. (2007). Sense-making software for crime investigation: how to combine stories and arguments? *Law, Probability and Risk* 6, no. 1-4: 145–168.
- BEX, F., PRAKKEN, H., REED, C. & WALTON, D. (2003). Towards a formal account of reasoning about evidence: argumentation schemes and generalisations. *Artificial Intelligence and Law* 11, no. 2: 125–165.

- BOBA, R. (2009). *Crime analysis and crime mapping*. 2 éd. Thousand Oaks, CA: Sage Publications, Inc.
- VAN DEN BRAAK, S. W. (2010). Sensemaking software for crime analysis. Utrecht, Netherlands: PhD thesis, Utrecht University.
- VAN DEN BRAAK, S. W., VAN OOSTENDORP, H., PRAKKEN, H. & VREESWIJK, G. (2006). A critical review of argument visualization tools: do users become better reasoners. Dans *Workshop Notes of the ECAI-06 Workshop on Computational Models of Natural Argument*, éd. F. Grasso, R. Kibble, & C. Reed, 67-75. CMNA '06. Riva del Garda, Italy.
- BRANTINGHAM, P. L. & BRANTINGHAM, P. J. (1993). Environment, routine and situation: toward a pattern theory of crime. Dans *Routine activity and rational choice: advances in criminological theory*, éd. R. Clarke & M. Felson, 5:259–94. New Brunswick, NJ: Transaction Publishers.
- BRESCIANI, S. & EPPLER, M. (2008). *The risks of visualization. A classification of disadvantages associated with graphic representations of information*. ICA Working Paper 1/2008. Lugano, Switzerland: University of Lugano.
- BRODEUR, J.-P. (2005a). Le renseignement: distinctions préliminaires. *Revue canadienne de criminologie et de justice pénale* 47, no. 1: 15-43.
- BRODEUR, J.-P. (2005b). L'enquête criminelle. *Criminologie* 38, no. 2: 39-64.
- BRUCE, C. (2009). Fundamentals of crime analysis. Dans *Exploring crime analysis: readings on essential skills*, éd. S. Gwinn, C. Bruce, J. P. Cooper, & S. Hick. 2 éd. Overland Park, KS: BookSurge Publishing.
- BRUNSDON, C., CORCORAN, J. & HIGGS, G. (2007). Visualising space and time in crime patterns: a comparison of methods. *Computers, Environment and Urban Systems* 31, no. 1: 52-75.
- BUETOW, T., CHABOYA, L., O'TOOLE, C., CUSHNA, T., DASPIT, D., PETERSEN, T., ATABAKHSH, H. & CHEN, H. (2003). A spatio temporal visualizer for law enforcement. Dans *Proceedings of the 1st NSF/NIJ conference on Intelligence and Security Informatics*, 181-194. ISI'03. Tucson, AZ: Springer.
- CANTER, P. (2000). Using a geographic information system for tactical crime analysis. Dans *Analyzing crime patterns: frontiers of practice*, éd. V. Goldsmith, P. G. McGuire, J. H. Mollenkopf, & T. A. Ross. Thousand Oaks, CA: Sage Publications, Inc.
- CARD, S. K. & MACKINLAY, J. (1997). The structure of the information visualization design space. Dans *Proceedings of the 1997 IEEE Symposium on Information Visualization*, 92-99. INFOVIS '97. Phoenix, AZ.
- CARD, S. K., MACKINLAY, J. & SHNEIDERMAN, B. (1999). *Readings in Information Visualization: Using Vision to Think*. San Francisco, CA: Morgan Kaufmann.

- CARILLO, C. (2010). Police judiciaire et renseignement. *Cahiers de la Sécurité* 13: 39-50.
- CARLIS, J. V. & KONSTAN, J. A. (1998). Interactive visualization of serial periodic data. Dans *Proceedings of the 11th annual ACM symposium on User Interface Software and Technology*, 29-38. UIST '98. San Francisco, CA: ACM Press.
- CARPENDALE, S. (2008). Evaluating information visualizations. Dans *Information visualization: human-centered issues and perspectives*, éd. A. Kerren, J. T. Stasko, J.-D. Fekete, & C. North. Berlin Heidelberg: Springer.
- CHAINEDY, S. & RATCLIFFE, J. (2005). *GIS and crime mapping*. London, England: John Wiley and Sons.
- CHANG, D., DOOLEY, L. & TUOVINEN, J. E. (2002). Gestalt theory in visual screen design: a new look at an old subject. Dans *Proceedings of the Seventh World Conference on Computers in Education: Australian topics*, 8:5–12. CRPIT '02. Darlinghurst, Australia: Australian Computer Society, Inc.
- CHEN, C. (2006). *Information Visualization: Beyond the Horizon*. 2 éd. London: Springer.
- CHEN, C. & CZERWINSKI, M. (2000). Empirical evaluation of information visualization: an introduction. *International Journal of Human-Computer Studies* 53, no. 5: 631-635.
- CHEN, C., HÄRDLE, W. & UNWIN, A. (2008). *Handbook of data visualization*. 1er éd. Berlin Heidelberg: Springer.
- CHEN, C. & YU, Y. (2000). Empirical studies of information visualization: a meta-analysis. *International Journal of Human-Computer Studies* 53, no. 5: 851-866.
- CHEN, H., ATABAKHSH, H., TSENG, C., MARSHALL, B., KAZA, S., EGGERS, S., GOWDA, H., SHAH, A., PETERSEN, T. & VIOLETTE, C. (2005). Visualization in law enforcement. Dans *Extended abstracts on Human factors in computing systems*, 1268–1271. CHI '05. New York, NY: ACM Press.
- CHEN, H., ZENG, D., ATABAKHSH, H., WYZGA, W. & SCHROEDER, J. (2003). COPLINK: managing law enforcement data and knowledge. *Communications of the ACM* 46, no. 1: 28-34.
- CHEN, P. P.-S. (1976). The entity-relationship model: toward a unified view of data. *ACM Transactions on Database Systems* 1: 9-36.
- CHI, ED H & PANIZZI, E. (2002). Expressiveness of the data flow and data state models in visualization systems. Dans *Proceedings of the Working Conference on Advanced Visual Interfaces*, éd. M. De Marsico & S. Levialdi, 375-378. AVI '02. Trento, Italy: ACM Press.
- CHI, ED H. (2000). A taxonomy of visualization techniques using the data state reference model. Dans *Proceedings of the 2000 IEEE Symposium on Information Visualization*, 69-75. INFOVIS '00. Salt Lake City, UT: IEEE Computer Society.

- CHI, ED H. & RIEDL, J. (1998). An operator interaction framework for visualization systems. Dans *Proceedings of the 1998 IEEE Symposium on Information Visualization*, 63–70. INFOVIS '98. Research Triangle Park, NC: IEEE Computer Society.
- CHUNG, W., CHEN, H., CHABOYA, L. G., O'TOOLE, C. D. & ATABAKHSH, H. (2005). Evaluating event visualization: a usability study of COPLINK spatio-temporal visualizer. *International Journal of Human-Computer Studies* 62, no. 1: 127-157.
- CLARKE, R. V. (2004). Technology, criminology and crime science. *European Journal on Criminal Policy and Research* 10, no. 1: 55-63.
- CLEVELAND, W. S. & TERPENNING, I. J. (1982). Graphical methods for seasonal adjustment. *Journal of the American Statistical Association* 77: 52-62.
- COCKBURN, A. & MCKENZIE, B. (2001). 3D or not 3D?: evaluating the effect of the third dimension in a document management system. Dans *Proceedings of the SIGCHI conference on Human factors in computing systems*, 434-441. CHI '01. Seattle, WA: ACM press.
- COCKBURN, A. & MCKENZIE, B. (2004). Evaluating spatial memory in two and three dimensions. *International Journal of Human-Computer Studies* 61, no. 3: 359-373.
- COCKBURN, A. & MCKENZIE, B. (2000). An evaluation of cone trees. Dans *People and Computers XIV: British Computer Society Conference on Human Computer Interaction*, 425-436. HCI '00. University of Sunderland, UK: Springer.
- COPE, N. (2003). Crime analysis: principles and practice. Dans *Handbook of policing*, éd. T. Newburn, 340-362. 1er éd. Cullompton, Devon: Willan Publishing.
- CROSBY, A. W. (1997). *The measure of reality: quantification and western society, 1250-1600*. New York: Cambridge University Press.
- CUSSON, M. (2008). De l'action de sécurité. Dans *Traité de sécurité intérieure*, éd. M. Cusson, B. Dupont, & F. Lemieux, 43-57. Lausanne, Switzerland: Presses polytechniques et universitaires romandes.
- ECK, J. (2003). Police problems: the complexity of problem theory, research and evaluation. *Crime Prevention Studies* 15: 79–114.
- EPPLER, M. J. & AESCHIMANN, M. (2009). A systematic framework for risk visualization in risk management and communication. *Risk Management* 11, no. 2: 67-89.
- FEDPOL. (2010). Manuel d'analyse criminelle opérationnelle. Office fédéral de la police (Groupe de travail suisse analyse criminelle opérationnelle).
- FEIGENSON, N. & SHERWIN, R. K. (2007). Thinking beyond the shown: implicit inferences in evidence and argument. *Law, Probability & Risk* 6, no. 1-4: 295-310.

- FELSON, M. & CLARKE, R. V. (1998). *Opportunity makes the thief: practical theory for crime prevention*. Police Research Series. London: Home Office: Policing and Reducing Crime Unit (Research, Development and Statistics Directorate).
- FEW, S. (2009). *Now you see it: simple visualization techniques for quantitative analysis*. 1er éd. Oakland, CA: Analytics Press.
- FREDRIKSON, A., NORTH, C., PLAISANT, C. & SHNEIDERMAN, B. (1999). Temporal, geographical and categorical aggregations viewed through coordinated displays: a case study with highway incident data. Dans *Proceedings of the 1999 workshop on new paradigms in information visualization and manipulation*, 26-34. NPIV '99. Kansas City, MI: ACM press.
- FRIENDLY, M. (2008). A brief history of data visualization. Dans *Handbook of computational statistics: data visualization*, éd. C. Chen, W. K. Härdle, & A. Unwin, 15-56. Berlin Heidelberg: Springer.
- GAMMA, E., HELM, R., JOHNSON, R. & VLISSIDES, J. (1995). *Design patterns: elements of reusable object-oriented software*. Boston, MA: Addison-Wesley Longman Publishing Co., Inc.
- GHONIEM, M., FEKETE, J.-D. & CASTAGLIOLA, P. (2005). A comparison of the readability of graphs using node-link and matrix-based representations. Dans *Proceedings of the 2004 IEEE Symposium on Information Visualization*, 17-24. INFOVIS '04. Austin, TX: IEEE Computer Society.
- GIROD, A., CHAMPOD, C. & RIBAUX, O. (2008). *Traces de souliers*. Lausanne, Switzerland: Presses polytechniques et universitaires romandes.
- GIWG. (2004). *National criminal intelligence sharing plan*. 1er éd. Washington, DC: Department of Justice (Global Intelligence Working Group).
- GOODWIN, J. & FISHER, A. (2000). Wigmore's chart method. *Informal Logic* 20, no. 3: 223-243.
- GOTTLIEB, S., ARENBERG, S. & SINGH, R. (1994). *Crime analysis: from first report to final arrest*. Montclair, CA: Alpha Publishing.
- GUENIAT, O. & ESSEIVA, P. (2005). *Le profilage de l'héroïne et de la cocaïne: une méthodologie moderne de lutte contre le trafic illicite*. Lausanne, Switzerland: Presses polytechniques et universitaires romandes.
- GUERRY, A.-M. (1833). *Essai sur la statistique morale de la France*. Paris, France: Crochard.
- GUO, D., CHEN, J., MACEACHREN, A. M. & LIAO, K. (2006). A visualization system for space-time and multivariate patterns (VIS-STAMP). *IEEE Transactions on Visualization and Computer Graphics* 12, no. 6: 1461-1474.
- HARPER, W. R. & HARRIS, D. H. (1975). The application of link analysis to police intelligence. *Human Factors* 17, no. 2: 157-164.

- HARRIS, R. L. (2000). *Information graphics: a comprehensive illustrated reference*. 1er éd. New York, NY: Oxford University Press.
- HEER, J. & AGRAWALA, M. (2006). Software design patterns for information visualization. *IEEE Transactions on Visualization and Computer Graphics* 12, no. 5: 853–860.
- HEIDL, B. (2007). La visualisation de données temporelles et élaboration de nouvelles représentations spatio-temporelles de données téléphoniques. Université de Lausanne, Switzerland: travail de master à Institut de Police Scientifique.
- HELMS, D. (2009). Temporal analysis. Dans *Exploring crime analysis: readings on essential skills*, éd. S. Gwinn, C. Bruce, J. P. Cooper, & S. Hick, 214-257. 2 éd. Overland Park, KS: BookSurge Publishing.
- HEUER, R. J. (1999). *Psychology of intelligence analysis*. 1er éd. Washington, DC: Central Intelligence Agency (Center for the Study of Intelligence).
- HEUER, R. J. & PHERSON, R. H. (2010). *Structured analytic techniques for intelligence analysis*. Washington, DC: CQ Press.
- HICKS, M., O'MALLEY, C., NICHOLS, S. & ANDERSON, B. (2003). Comparison of 2D and 3D representations for visualising telecommunication usage. *Behaviour and Information Technology* 22, no. 3: 185-201.
- HUANG, W., EADES, P. & HONG, S.-H. (2008). Beyond time and error: a cognitive approach to the evaluation of graph drawings. Dans *Proceedings of the 2008 conference on BEyond time and errors: novel evaluation methods for Information Visualization*, 1-8. BELIV '08. Florence, Italy: ACM press.
- HUBONA, G. S., SHIRAH, G. W. & FOUT, D. G. (1997). 3D object recognition with motion. Dans *extended abstracts on Human factors in computing systems: looking to the future*, 345-346. CHI '97. Atlanta, GA: ACM press.
- IALEIA. (2004). *Law enforcement analytical standards*. Richmond, VA: U.S. Department of Justice (Bureau of Justice Assistance).
- INMAN, K. & RUDIN, N. (2001). *Principles and practices of criminalistics: the profession of forensic science*. Boca Raton: CRC Press.
- INNES, M., FIELDING, N. & COPE, N. (2005). The appliance of science? The theory and practice of crime intelligence analysis. *British Journal of Criminology* 45: 39-57.
- INSELBERG, A. (1985). The plane with parallel coordinates. *The Visual Computer* 1, no. 2: 69-91.
- INSELBERG, A. (1997). Multidimensional detective. Dans *Proceedings of the 1997 IEEE Symposium on Information Visualization*, 100-107. INFOVIS '97. Phoenix, AZ: IEEE Computer Society.

- INTERPOL. (1997). *Guide sur l'analyse criminelle*. 2 éd. Lyon: Interpol (Groupe de travail sur l'analyse criminelle).
- IRANI, P. & WARE, C. (2003). Diagramming information structures using 3D perceptual primitives. *ACM Transactions on Computer-Human Interaction* 10, no. 1: 1-19.
- JOHNSON, S. D. (2010). A brief history of the analysis of crime concentration. *European Journal of Applied Mathematics* 21, no. 4-5: 349-370.
- JOSEPH, G. P. (2006). *Modern visual evidence*. New York, NY: Law Journal Press.
- KEIM, D. & WARD, M. (2007). Visualization. Dans *Intelligent data analysis*, éd. M. Berthold & D. J. Hand, 403-427. 2 éd. Berlin Heidelberg: Springer.
- KEIM, D. A., PANSE, C. & SIPS, M. (2005). Information visualization: scope, techniques and opportunities for geovisualization. Dans *Exploring Geovisualization*, éd. Dykes, A. M. MacEachren, & M.-J. Kraak, 23-52. Amsterdam: Elsevier.
- KELLEY, J. E. JR & WALKER, M. R. (1959). Critical-path planning and scheduling. Dans *Papers presented at the December 1-3, 1959, eastern joint IRE-AIEE-ACM computer conference*, 160-173. IRE-AIEE-ACM '59. New York, NY: ACM press.
- KELLY, R. J. (1990). The development of inferences in the assessment of intelligence data. Dans *Criminal intelligence analysis*, éd. P. Andrews & M. B. Peterson, 149-180. Loomis, CA: Palmer Press.
- KIND, S. S. (1987). *The scientific investigation of crime*. Harrogate: Forensic Science Services Ltd.
- KIND, S. S. (1990). What makes a good forensic scientist? *Journal of the Forensic Science Society* 30, no. 4: 247-256.
- KIND, S. S. (1992). Mental images in crime investigation. *The Criminologist* 16, no. 1: 15-24.
- KIND, S. S. (1994). Crime investigation and the criminal trial: a three chapter paradigm of evidence. *Journal of the Forensic Science Society* 34, no. 3: 155-164.
- KLERKS, P. & SMEETS, E. (2001). The network paradigm applied to criminal organizations: theoretical nitpicking or a relevant doctrine for investigators? Recent developments in the Netherlands. *Connections* 24, no. 3: 53-65.
- KOLDA, T., BROWN, D., CORONES, J., CRITCHLOW, T., ELIASSI-RAD, T., GETOOR, L., HENDRICKSON, B. ET AL. (2004). *Data sciences technology for homeland security information management and knowledge discovery*. Lawrence Livermore Nat'l Laboratory: Report of the DHS Workshop on Data Sciences.
- LAXMAN, S. & SASTRY, P. S. (2006). A survey of temporal data mining. *Sadhana* 31, no. 2: 173-198.

- LEE, B., PLAISANT, C., PARR, C. S, FEKETE, J. D & HENRY, N. (2006). Task taxonomy for graph visualization. Dans *Proceedings of the 2006 AVI workshop on beyond time and errors: novel evaluation methods for information visualization*, 1–5. BELIV '06. New York, NY: ACM press.
- LEGLER, R. & EPPLER, M. J. (2007). Towards a periodic table of visualization methods for management. Dans *Proceedings of the Conference on Graphics and Visualization in Engineering*, 1-6. GVE '07. Clearwater, FL: ACTA press.
- LOHSE, G. L., BIOLSI, K., WALKER, N. & RUETER, H. H. (1994). A classification of visual representations. *Communications of the ACM* 37, no. 12: 36-49.
- MAC EACHREN, A. M. (1992). Visualizing uncertain information. *Cartographic Perspective* 13, no. 3: 10–19.
- MAC EACHREN, A. M. (2004). *How maps work: representation, visualization, and design*. New York, NY: The Guilford Press.
- MACKINLAY, J. (1986). Automating the design of graphical presentations of relational information. Dans *Readings in intelligent user interfaces*, éd. M. T. Maybury & W. Wahlster, 5:177-193. San Francisco, CA: Morgan Kaufmann Publishers Inc.
- MALLER, V. A. J. (1996). Criminal investigation systems: the growing dependence on advanced computer systems. *Computing and Control Engineering Journal* 7, no. 2: 93-100.
- MARSHALL, B., CHEN, H. & KAZA, S. (2008). Using importance flooding to identify interesting networks of criminal activity. *Journal of the American Society for Information Science and Technology* 59, no. 13: 2099–2114.
- MCGRATH, C., BLYTHE, J. & KRACKHARDT, D. (1997). The effect of spatial arrangement on judgments and errors in interpreting graphs. *Social Networks* 19, no. 3: 223-242.
- MENA, J. (2003). *Investigative data mining for security and criminal detection*. Burlington, MA: Butterworth-Heinemann.
- MENNIS, J. L., PEUQUET, D. J. & QIAN, L. (2000). A conceptual framework for incorporating cognitive principles into geographical database representation. *International Journal of Geographical Information Science* 14, no. 6: 501–520.
- MILES, M. B. & HUBERMAN, A. M. (1994). *Qualitative data analysis: an expanded sourcebook*. 2 éd. Thousand Oaks, CA: Sage Publications, Inc.
- MORRIS, JACK. (1986). *The Crime Analysis Charting*. Loomis, CA: Palmer Press.
- MOWBRAY, C. (2009). Analytical charting. Dans *Exploring crime analysis: readings on essential skills*, éd. S. Gwinn, C. Bruce, J. P. Cooper, & S. Hick. 2 éd. Overland Park, KS: BookSurge Publishing.

- MÜLLER, W. & SCHUMANN, H. (2003). Visualization for modeling and simulation: visualization methods for time-dependent data - an overview. Dans *Proceedings of the 35th conference on Winter simulation: driving innovation*, 1:737-745. WSC '03. New Orleans, LA.
- MULLET, K. (1995). 3D or not 3D: "more is better" or "less is more"? (panel session). Dans *Conference companion on Human factors in computing systems*, 174-175. Denver, CO: ACM press.
- NCIS. (2000). The National Intelligence Model. National Crime Intelligence Service.
- NCPE. (2005). *Guidance on the national intelligence model*. Wyboston, England: Association of Chief Police Officers (National Centre for Policing Excellence).
- NEWMAN, M. E. J. (2003). The structure and function of complex networks. *SIAM Review* 45, no. 2: 167-256.
- NORMAN, D. A. (1993). *Things that make us smart: defending human attributes in the age of the machine*. New York, NY: Basic Books.
- NORTH, C. (2006). Toward measuring visualization insight. *IEEE Computer Graphics and Applications* 26, no. 3: 6-9.
- OATLEY, G., EWART, B. & ZELEZNIKOW, J. (2006). Decision support systems for police: Lessons from the application of data mining techniques to soft forensic evidence. *Artificial Intelligence and Law* 14, no. 1-2: 35-1000.
- OATLEY, G., ZELEZNIKOW, J., LEARY, R. & EWART, B. (2005). From links to meaning: a burglary data case study. Dans *International Conference on Knowledge-Based Intelligent Information and Engineering Systems*, 813-822. Melbourne, Australia: Springer.
- OCAGNE, M. D'. (1885). *Coordonnées parallèles et axiales. Méthode de transformation géométrique et procédé nouveau de calcul graphique déduits de la considération des coordonnées parallèles*. Paris, France: Gauthier-Villars.
- PAAS, F., TUOVINEN, J. E., TABBERS, H. & VAN GERVEN, P. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational psychologist* 38, no. 1: 63-71.
- PALMER, S. & ROCK, I. (1994). Rethinking perceptual organization: the role of uniform connectedness. *Psychonomic Bulletin & Review* 1, no. 1: 29-55.
- PARKER, G., FRANCK, G. & WARE, C. (1998). Visualization of large nested graphs in 3D: navigation and interaction. *Journal of Visual Languages and Computing* 9, no. 3: 299-317.
- PATTYN, M & DICKENS, C. (2001). Formation analyse stratégique. Police fédérale belge (CGC/AS).

- PAULSEN, D. J., BAIR, S. & HELMS, D. (2010). *Tactical crime analysis: research and investigation*. Boca Raton, FL: CRC Press.
- PCSCO. (2002). Crime organisé - Etude des meilleures pratiques no. 4. Conseil de l'Europe, Direction Générale I (Groupe de Spécialistes sur les aspects de droit pénal et les aspects criminologiques du crime organisé).
- PELEKIS, N., THEODOULIDIS, B., KOPANAKIS, I. & THEODORIDIS, Y. (2004). Literature review of spatio-temporal database models. *The Knowledge Engineering Review* 19, no. 3: 235-274.
- PETERSON, M. B. (1990). Telephone record analysis. Dans *Criminal intelligence analysis*, éd. P. Andrews & M. B. Peterson, 85-116. Loomis, CA: Palmer Press.
- PETERSON, M. B. (1998). *Applications in criminal analysis: a sourcebook*. Westport, CT: Praeger.
- PETERSON, M. B., MOREHOUSE, B. & WRIGHT, E. (2000). *Intelligence 2000: revising the basic elements*. Lawrenceville, NJ: Law Enforcement Intelligence Unit and International Association of Law Enforcement Intelligence Analysts.
- PEUQUET, D. J. (1984). A conceptual framework and comparison of spatial data models. *Cartographica: the International Journal for Geographic Information and Geovisualization* 21, no. 4: 66-113.
- PEUQUET, D. J. (1994). It's about time: a conceptual framework for the representation of temporal dynamics in geographic information systems. *Annals of the Association of American Geographers* 84, no. 3: 441-461.
- PFITZNER, D., HOBBS, V. & POWERS, D. (2003). A unified taxonomic framework for information visualization. Dans *Proceedings of the Asia-Pacific symposium on Information visualisation - Volume 24*, 24:57-66. APVis '03. Darlinghurst, Australia: Australian Computer Society, Inc.
- PHAN, D., XIAO, L., YEH, R., HANRAHAN, P. & WINOGRAD, T. (2005). Flow map layout. Dans *Proceedings of the 2005 IEEE Symposium on Information Visualization*, 219-224. INFOVIS '05. Minneapolis, MN: IEEE Computer Society.
- PIROLI, P. & CARD, S. (2005). The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. Dans *Proceedings of International Conference on Intelligence Analysis*. McLean, VA.
- PLAISANT, C. (2004). The challenge of information visualization evaluation. Dans *Proceedings of the Working Conference on Advanced Visual Interfaces*, 109-116. AVI '04. New York, NY: ACM press.
- PLAYFAIR, W. (2005). *Commercial and political atlas and statistical breviary*. Éd. H. Wainer & I. Spence. New York, NY: Cambridge University Press.

- POTTENGER, W. M., YANG, X. & ZANIAS, S. V. (2007). *Link analysis survey status update - january 2006*. U.S. Department of Justice (National Criminal Justice Reference Service).
- PURCHASE, H. C. (2000). Effective information visualisation: a study of graph drawing aesthetics and algorithms. *Interacting with Computers* 13, no. 2: 147–162.
- QUETELET, A. (1842). *A treatise on man and the development of his faculties*. Edinburgh, England : W. and R. Chambers.
- RATCLIFFE, J. H. (2000). Aoristic analysis: the spatial interpretation of unspecific temporal events. *International Journal of Geographical Information Science* 14, no. 7: 669-679.
- RATCLIFFE, J. H. (2008). *Intelligence-led policing*. Cullompton, Devon: Willan Publishing.
- REED, C. & ROWE, G. (2007). A pluralist approach to argument diagramming. *Law, Probability and Risk* 6, no. 1-4: 59–85.
- REED, R. N. & KOCURA, P. (2005). Conceptual graph based criminal intelligence analysis. Dans *Common semantics for sharing knowledge: contributions to ICCS 2005 13th International Conference on Conceptual Structures*, éd. F. Dau & M.-L. Mugnier, 146-159. ICCS 2005. Kassel, Germany: Kassel university press.
- REGALI, J. E. (1999). Visual investigative analysis: an underutilized case management tool. *IACLEIA Journal* 12, no. 1: 48-52.
- RIBAU, O. (1997). La recherche et la gestion des liens dans l'investigation criminelle: le cas particulier du cambriolage. Thèse de doctorat, Université de Lausanne, Switzerland.
- RIBAU, O. (2008). Chapitre 3. Définitions. Dans *D6.7c: Forensic Profiling*, éd. Z. Geradts & P. Sommer, 12-30. FIDIS Future of Identity in the Information Society.
- RIBAU, O., BAYLON, A., LOCK, E., DELÉMONT, O., ROUX, C., ZINGG, C. & MARGOT, P. (2010). Intelligence-led crime scene processing. Part II: Intelligence and crime scene examination. *Forensic Science International* 199, no. 1-3: 63-71.
- RIBAU, O., GIROD, A., WALSH, S. J., MARGOT, P., MIZRAHI, S. & CLIVAZ, V. (2003). Forensic intelligence and crime analysis. *Law, Probability and Risk* 2, no. 1: 47-60.
- RIBAU, O. & MARGOT, P. (1999). Inference structures for crime analysis and intelligence: the example of burglary using forensic science data. *Forensic Science International* 100, no. 3: 193–210.
- RIBAU, O. & MARGOT, P. (2003). Case based reasoning in criminal intelligence using forensic case data. *Science and Justice* 43, no. 3: 135–144.
- RIBAU, O. & MARGOT, P. (2008). La trace matérielle vecteur d'information au service du renseignement. Dans *Traité de sécurité intérieure*, éd. M. Cusson, B. Dupont, & F. Lemieux, 300-321. Lausanne, Switzerland: Presses polytechniques et universitaires romandes.

- RIBAUX, O., WALSH, S. J. & MARGOT, P. (2006). The contribution of forensic science to crime analysis and investigation: forensic intelligence. *Forensic Science International* 156, no. 2-3: 171-181.
- RUMBAUGH, J., BLAHA, M., LORENSEN, W., EDDY, F. & PREMERLANI, W. (1991). *Object-oriented modeling and design*. New York, NY: Prentice-Hall International, Inc.
- SCAIFE, M. & ROGERS, Y. (1996). External cognition: how do graphical representations work? *International Journal of Human-Computer Studies* 45: 185–213.
- SCHROEDER, J., XU, J., CHEN, H. & CHAU, M. (2007). Automated criminal link analysis based on domain knowledge. *Journal of the American Society for Information Science and Technology* 58, no. 6: 842-855.
- SENATOR, T. E. (2005). Link mining applications: progress and challenges. *SIGKDD Explorations* 7, no. 2: 76-83.
- SHNEIDERMAN, B. (1996). The eyes have it: a task by data type taxonomy for information visualizations. Dans *Proceedings of the 1996 IEEE Symposium on Visual Languages*, 336–343. VL '96. Boulder, CO: IEEE Computer Society.
- SHNEIDERMAN, B. (2003). Why not make interfaces better than 3D reality? *IEEE Computer Graphics and Applications* 23, no. 6: 12-15.
- SHNEIDERMAN, B. & ARIS, A. (2006). Network visualization by semantic substrates. *IEEE Transactions on Visualization and Computer Graphics* 12, no. 5: 733–740.
- SIMMS, B. W. & PETERSEN, E. R. (1991). An information processing model of a police organization. *Management Science* 37, no. 2: 216–232.
- SPARROW, M. K. (1991). The application of network analysis to criminal intelligence: an assessment of the prospects. *Social networks* 13, no. 3: 251–274.
- SPRINGMEYER, R. R., BLATTNER, M. M. & MAX, N. L. (1992). A characterization of the scientific data analysis process. Dans *Proceedings of the 3rd conference on Visualization*, 235-242. VIS '92. Los Alamitos, CA: IEEE Computer Society.
- ST JOHN, M., COWEN, M. B., SMALLMAN, H. S. & OONK, H. M. (2001). The use of 2D and 3D displays for shape-understanding versus relative-position tasks. *Human Factors* 43, no. 1: 79-98.
- STASKO, J., GÖRG, C. & LIU, Z. (2008). Jigsaw: supporting investigative analysis through interactive visualization. *Information Visualization* 7, no. 2: 118–132.
- STASKO, J. & WEHRLI, J. F. (1993). Three-dimensional computation visualization. Dans *Proceedings of the 1993 IEEE Symposium on Visual Languages*, 100-107. VL '93. Bergen, Norway: IEEE Computer Society.
- TANG, D., STOLTE, C. & BOSCH, R. (2004). Design choices when architecting visualizations. *Information Visualization* 3, no. 2: 65-79.

- TAVANTI, M. & LIND, M. (2001). 2D vs 3D, implications on spatial memory. Dans *Proceedings of the 2001 IEEE Symposium on Information Visualization*, 139-145. INFOVIS '01. San Diego, CA: IEEE Computer Society.
- TAYLOR, B., KOWALYK, A. & BOBA, R. (2007). The integration of crime analysis into law enforcement agencies: an exploratory study into the perceptions of crime analysts. *Police Quarterly* 10, no. 2: 154-169.
- TEYSEYRE, A. R. & CAMPO, M. R. (2009). An overview of 3D software visualization. *IEEE Transactions on Visualization and Computer Graphics* 15, no. 1: 87-105.
- TILLERS, P. (2007). Introduction: visualizing evidence and inference in legal settings. *Law, Probability and Risk* 6, no. 1: 1-4.
- TILLERS, P. & SCHUM, D. (1991). A theory of preliminary fact investigation. *UC Davis Law Review* 24: 931-1012.
- TORY, M., KIRKPATRICK, A. E., ATKINS, M. S. & MÖLLER, T. (2006). Visualization task performance with 2D, 3D, and combination displays. *IEEE Transactions on Visualization and Computer Graphics* 12, no. 1: 2-13.
- TUFTE, E. R. (1990). *Envisioning information*. Cheshire, CT: Graphics Press.
- TUFTE, E. R. (1997). *Visual explanations: images and quantities, evidence and narrative*. Cheshire, CT: Graphics Press.
- TUFTE, E. R. (2001). *The visual display of quantitative information*. 2 éd. Cheshire, CT: Graphics Press.
- TUFTE, E. R. (2005). *Visual and statistical thinking: displays of evidence for making decisions*. 3 éd. Cheshire, CT: Graphics Press.
- TUOVINEN, J. E. & PAAS, F. (2004). Exploring multidimensional approaches to the efficiency of instructional conditions. *Instructional Science* 32, no. 1: 133-152.
- TWEEDIE, L. (1997). Characterizing interactive externalizations. Dans *Proceedings of the SIGCHI conference on Human factors in computing systems*, 375-382. CHI '97. Atlanta, GA: ACM press.
- UNODC. (2002). *Criminal intelligence training: manual for analysts*. United Nations Office on Drugs and Crime.
- WARE, C. (2004). *Information visualization: perception for design*. Burlington, MA: Morgan Kaufmann.
- WARE, C. (2008). *Visual thinking for design*. Burlington, MA: Morgan Kaufmann.
- WARE, C. & FRANCK, G. (1994). Viewing a graph in a virtual reality display is three times as good as a 2D diagram. Dans *Proceedings of the 1994 IEEE Symposium on Visual Languages*, 182-183. St. Louis, MO: IEEE Computer Society.

- WARE, C. & MITCHELL, P. (2008). Visualizing graphs in three dimensions. *ACM Transactions on Applied Perception* 5, no. 1: 1-15.
- WEHREND, S. & LEWIS, C. (1990). A problem-oriented classification of visualization techniques. Dans *Proceedings of the 1st IEEE conference on Visualization '90*, 139–143. VIS '90. Los Alamitos, CA: IEEE Computer Society.
- WIGMORE, J. H. (1913). *The principles of judicial proof as given by logic, psychology, and general experience, and illustrated in judicial trials*. Boston: Little, Brown, and company.
- WONG, P.C., CHIN, G., FOOTE, H., MACKAY, P. & THOMAS, J. (2006). Have green - a visual analytics framework for large semantic graphs. Dans *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology*, 67-74. VAST '06. Baltimore, MD: IEEE Computer Society.
- WRIGHT, W., SCHROH, D., PROULX, P., SKABURSKIS, A. & CORT, B. (2006). The Sandbox for analysis: concepts and methods. Dans *Proceedings of the SIGCHI conference on Human Factors in computing systems*, éd. R. Grinter, T. Rodden, P. Aoki, E. Cutrell, R. Jeffries, & G. Olson, 810-810. CHI '06. Montréal, Quebec: ACM press.
- XIANG, Y., CHAU, M., ATABAKHSH, H. & CHEN, H. (2005). Visualizing criminal relationships: comparison of a hyperbolic tree and a hierarchical list. *Decision Support Systems* 41, no. 1: 69–83.
- XU, J. & CHEN, H. (2005). Criminal network analysis and visualization. *Communications of the ACM* 48, no. 6: 100-107.
- ZACHRY, M. & THRALLS, C. (2004). An interview with Edward R. Tufte. *Technical Communication Quarterly* 13, no. 4: 447–462.

ANNEXES

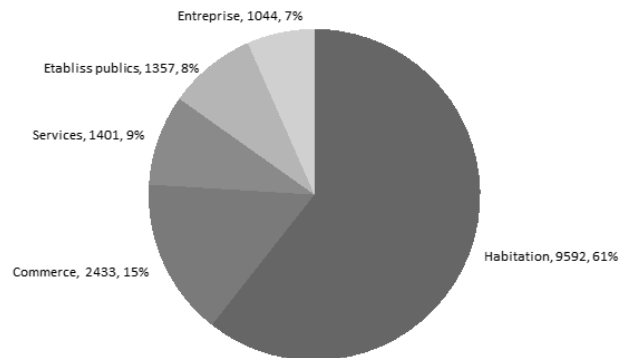
Annexe 1: Dimensions d'analyse et visualisation

Analyse quantitative

De nombreuses représentations différentes ont été inventées au cours de l'histoire pour analyser des données quantitatives. William Playfair (1759-1823) est considéré comme l'inventeur de la plupart des types de diagrammes utilisés de nos jours (linéaires, histogrammes, circulaires, etc.) (Playfair, 2005) (Friendly, 2008). Ces représentations ont été largement étudiées et développées depuis les travaux de Playfair, comme en témoigne l'ouvrage de Robert L. Harris qui référence une liste encyclopédique de représentations (Harris, 2000). Les travaux d'Edward Tufte (Tufte, 2001),(Tufte, 1997),(Tufte, 1990) ont largement contribué au développement de l'étude moderne de ces représentations. L'accent est mis, en particulier sur les effets visuels exploités pour tronquer des données et par opposition sur les techniques efficaces et intègres. Parallèlement, depuis les travaux de Bertin sur la sémiologie graphique (Bertin, 2005), de nombreuses études ont explicité divers aspects de la représentation de données quantitatives. A titre d'exemple, citons les travaux sur le processus de la visualisation (Card et al., 1999), sur l'impact du processus de perception et la sémiotique visuelle (MacEachren, 2004) (Ware, 2008), ou encore sur les types d'analyses effectuées à l'aide de représentations quantitatives (Few, 2009). Selon Stephen Few, l'analyse quantitative vise l'étude des relations entre des valeurs. Les méthodes de visualisation exploitées varient en fonction de la nature de ces relations : temporelles, classements et proportions, déviations, distributions, corrélations et multi-variées (Few, 2009). La décomposition proposée par Few est utilisée pour présenter les principales techniques. Les représentations temporelles font l'objet de la section suivante.

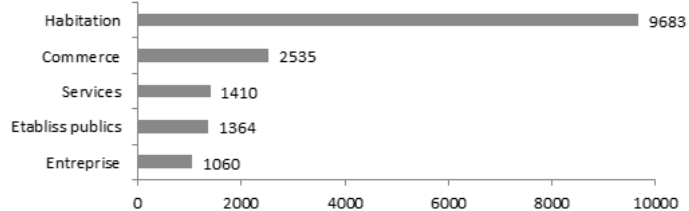
Représentations de classements et de proportions

Les diagrammes circulaires sectorisés



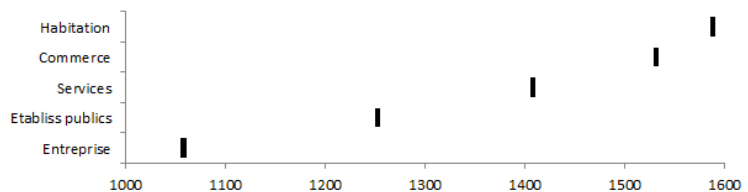
Ces diagrammes visent la **comparaison de proportions** et à répondre aux questions du type : « Quelle est la proportion d'une valeur par rapport à l'ensemble ? ».

Les histogrammes



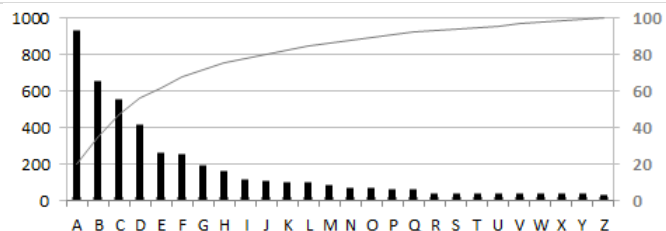
Ces diagrammes visent la **comparaison de proportions, le classement et la comparaison de valeurs**.

Les diagrammes en points



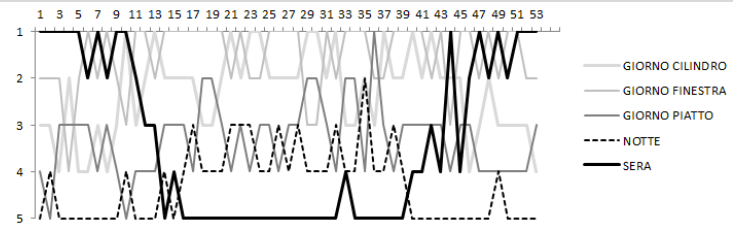
Ces diagrammes visent la **comparaison de différences entre les catégories**, principalement lorsque l'ensemble des valeurs sont grandes. Ils permettent d'adapter l'échelle sans ajouter de biais visuels liés à la comparaison de la longueur des barres.

Les diagrammes de Pareto



Ces diagrammes permettent de représenter la distribution par ordre décroissant des valeurs et **incluent une courbe cumulative** des valeurs, généralement exprimées en pourcentage.

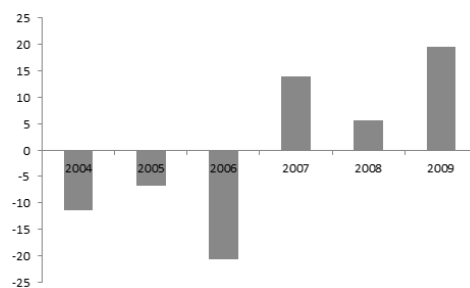
*Les diagrammes
de Bumps*



Ces diagrammes en lignes représentant **l'évolution du classement** (axe vertical) au cours du temps (axe horizontal).

Représentations de différences / de déviations

*Les diagrammes de
déviations*



Variation annuelle en pourcent par rapport à la moyenne

Les diagrammes de déviation représentent les **valeurs quantitatives en fonction d'une valeur définie** : la moyenne, une valeur du passé, actuelle ou du futur, une valeur standard (seuil acceptable, maximum, minimum, etc.). Les différences peuvent être exprimées en pourcentage et représentées par des barres ou des lignes.

Représentations de distributions

Les strip plot

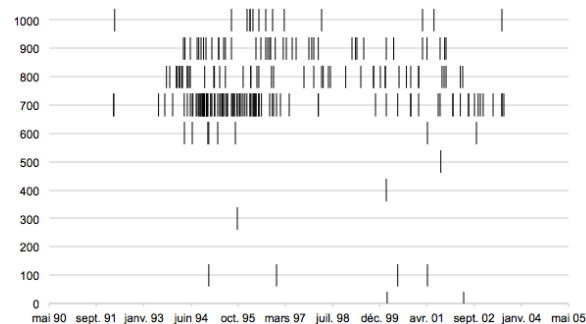


Diagramme représentant la **distribution d'une variable sur un seul axe** (dans cet exemple, le temps). L'exemple ci-dessus est un « multiple strip plot » puisque les distributions de plusieurs catégories sont représentées (axe vertical).

Les diagrammes de distribution

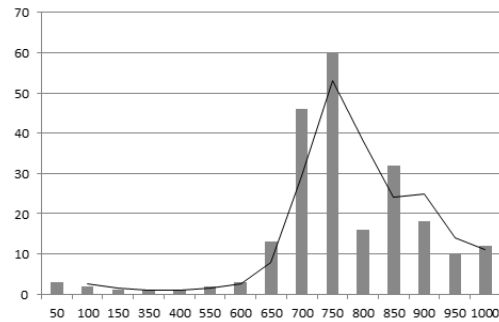
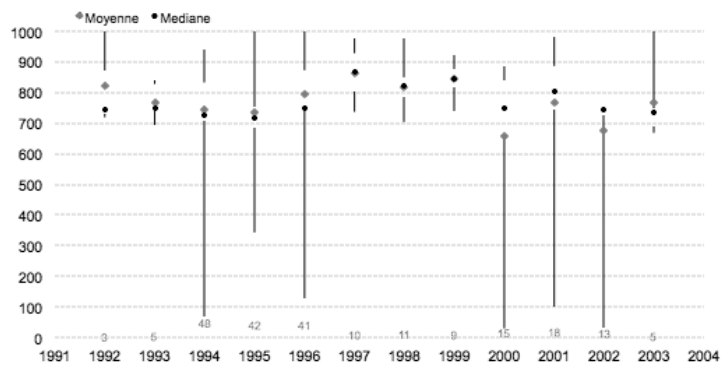


Diagramme représentant la distribution d'une variable en indiquant la **quantité pour chaque classe sur l'axe vertical**.

Les diagrammes en boîte (Box and Whiskers plot)



Les box plot représentent **l'étendue, la médiane et les quartiles de la distribution**. Dans l'exemple ci-dessus, la moyenne et le nombre d'observations (en bas) sont également représentés. Ce type de diagramme a été inventé par John Tukey (1915-2000), dans les années septantes.

Les diagrammes à feuilles (Stem-and-Leaf plot)

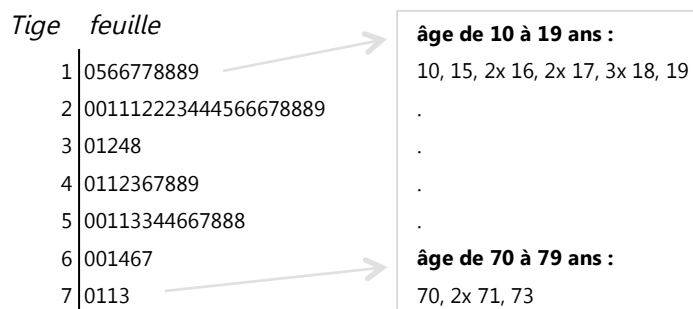


Diagramme divisé en deux parties : chaque ligne représente une tranche d'une échelle (ci-dessus les dizaines) affichée à gauche de la ligne verticale. A droite, chaque observation est représentée par le reste (ici l'unité). **La longueur de chaque ligne est donc proportionnelle au nombre d'observations par tranche**. Ce type de diagramme a également été inventé par John Tukey (1915-2000), dans les années septantes.

Représentations de corrélations

*Les nuages de points
(Scatter plot)*

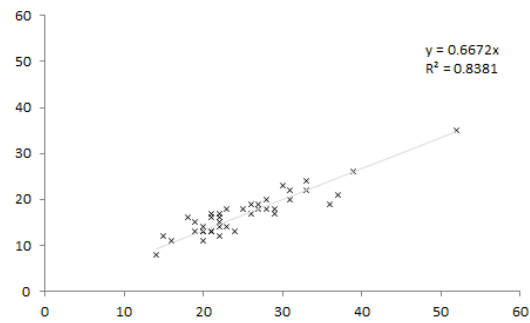


Diagramme représentant **la variation d'une variable quantitative en fonction d'une autre.**

Représentations de multiples variables

*Les cartes de chaleur
(Heatmap)*

	Lu	Ma	Me	Lu	Ma	Me	Lu
GIORNO	23	26	35	23	19	21	11
GIORNO CILINDRO	73	54	55	53	13	62	14
GIORNO EPAULEE	9	23	5	8	2	13	2
GIORNO FINESTRA	44	35	61	43	18	32	13
GIORNO PIATTO	56	39	50	54	22	49	27
HALL	24	7	23	18	12	23	12
NOTTE	15	11	13	12	25	20	15
NOTTE CHIGNOLE	11	8	3	5	9	16	3
NOTTE CILINDRO	6	5	10	12	10	6	4
NOTTE FINESTRA	15	17	16	15	25	7	10
SERA	51	33	38	36	70	61	65

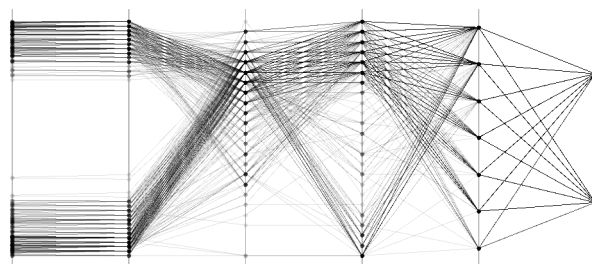
Diagramme représentant les variations par une **échelle de couleurs ou de niveaux de gris.**

*Les tables de données
(Table lens)*

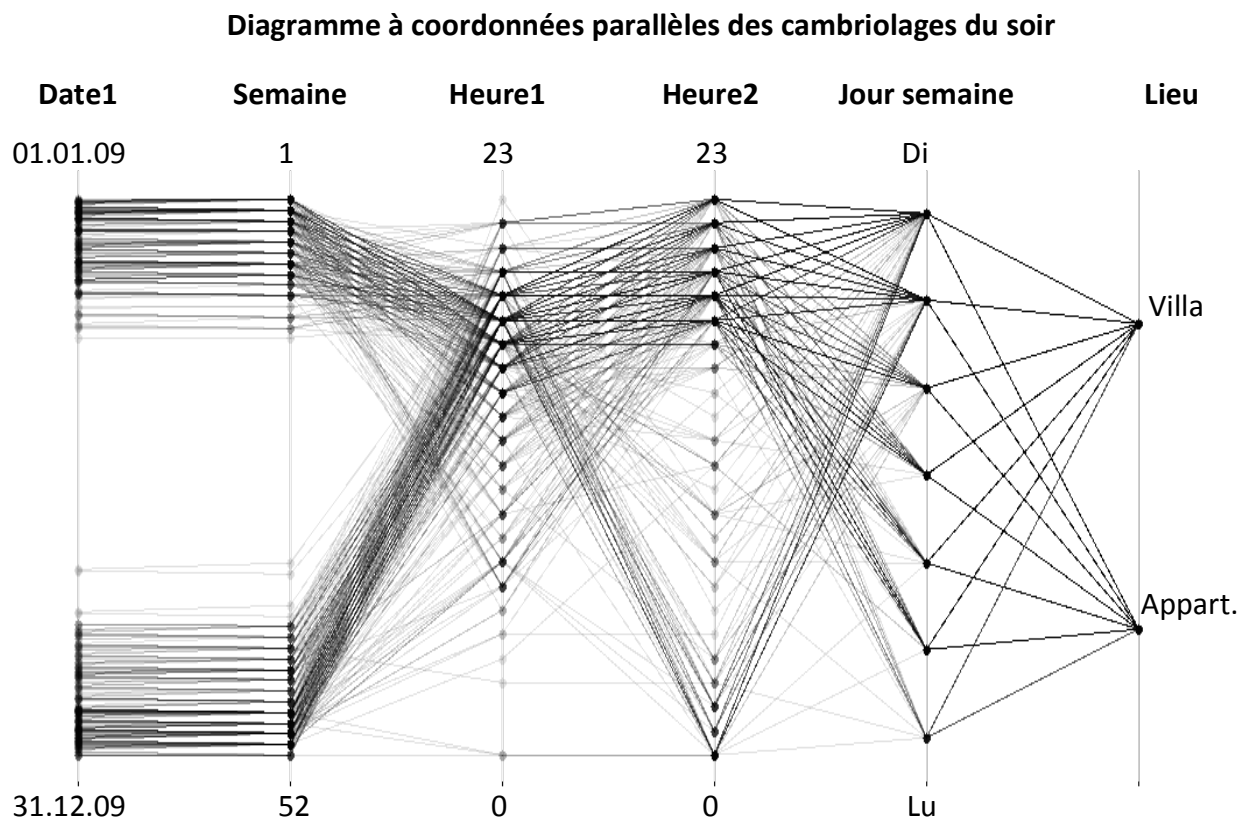
A	B	C	D	E	F	G	H	I	J	K
23	73	9	44	56	24	15	11	6	15	51
26	54	23	36	39	7	11	8	5	17	33
35	55	5	61	50	23	13	3	10	16	38
23	53	8	43	54	18	12	5	12	15	36
19	13	2	18	22	12	25	9	10	25	70
21	62	13	32	49	23	20	16	6	7	61
11	14	2	13	27	12	15	3	4	10	65

Représentation exploitant une forme de **tableau où les valeurs sont représentées par des diagrammes** pour remplacer le texte (en plus de celui-ci dans l'exemple ci-dessus)

*Les diagrammes à
coordonnées parallèles*



Alors que la majorité des diagrammes développés pour l'analyse quantitative sont largement utilisés en analyse criminelle, les graphiques à coordonnées parallèles semblent être sous exploités. Ils ont pourtant été spécifiquement développés afin de faciliter la détection et la formalisation de profils⁷¹. En effet, la représentation par un graphe des profils peut être très difficile voire inefficace si le nombre de relations du profil est important. Dans une telle situation, l'emploi d'un diagramme à coordonnées parallèles est recommandé. Ces représentations sont d'ailleurs également désignées sous le nom de diagramme de profil. Dans les ouvrages récents (Card et al., 1999) (Few, 2009), l'invention de cette forme de représentation est attribuée à Alfred Inselberg (Inselberg, 1985) (Inselberg, 1997). Leur exploitation remonterait en fait à la fin du dix-neuvième siècle, période à laquelle, Maurice d'Ocagne les auraient inventés (Ocagne, 1885). La raison principale de la sous-utilisation de cette forme de représentation réside dans la difficulté initiale de les appréhender : « *The first time I laid eyes on a parallel coordinates plot, I laughed and cringed simultaneously because it struck me as a ridiculously complex and ineffective display. ... However, this frenzy of intersecting lines, when used properly for multivariate analysis, can actually lead to great insight* » (Few, 2009).



Source : banque de données du CICOP, année 2009, canton de Vaud

⁷¹ Cf. Analyse, profil et pattern, page 26

Le choix de l'ordre des variables (et leur nombre) sur l'axe horizontal et le nombre d'observations représentées influence fortement la lisibilité de ces graphiques (Few, 2009). Leur production nécessite des outils dynamiques afin de modifier les arrangements, d'effectuer des sélections et de modifier les variables visuelles (telles que la couleur, l'épaisseur des traits et la transparence), afin d'identifier des profils. L'exemple ci-dessus représente un profil particulier de cambriolages d'habitations (cf. variable « lieu ») commis pendant la période hivernale (cf. variables « date1 » et « semaine »), le soir (cf. « heure1 » et « heure2 »). Il est notamment possible de détecter une corrélation entre la variable « heure2 » (heure à laquelle les victimes détectent le cambriolage, c'est-à-dire lorsqu'elles rentrent chez elles) et le jour de la semaine. En effet, le week-end les victimes s'absentent visiblement plus fréquemment jusque tard le soir (en l'occurrence tôt le lendemain matin). Les variables représentées dans ce cas sont principalement de type temporel. D'autres formes de représentation temporelle peuvent être exploitées pour détecter un tel profil (cf. section suivante). Ces diagrammes sont particulièrement efficaces pour regrouper dans une seule vue un nombre important de variables différentes. Les diagrammes dit « en étoiles » sont basés sur le même paradigme visuel, mais les axes sont positionnés de façon radiale plutôt qu'en parallèle.

A n a l y s e t e m p o r e l l e

L'analyse temporelle couvre un ensemble de questionnements complexes. Les objectifs de cette forme d'analyse peuvent être classifiés selon six tâches principales : la prédiction, la classification, le regroupement (clustering), la recherche et la récupération, ainsi que la détection de pattern (Laxman & Sastry, 2006). En matière d'analyse criminelle, deux objectifs principaux sont définis : la compréhension des événements qui se sont passés et la prédiction de ce qui va se produire (Boba, 2009) (Helms, 2009).

Contrairement aux autres variables quantitatives, la composante temporelle possède une structure sémantique particulière qui augmente sa complexité (Aigner et al., 2007). En effet, le temps possède une structure hiérarchique. Il est décrit selon des niveaux d'agrégation complexes dont les divisions varient (soixante minutes, vingt-quatre heures, sept jours de la semaine, douze mois de l'année, etc.). De plus le temps est perçu à la fois comme linéaire, mais également selon des cycles et des répétitions, tels que les jours de la semaine ou les saisons. Ces cycles peuvent également être irréguliers, notamment ceux liés à des facteurs sociologiques (comme les vacances par exemple).

240 Annexe 1: Dimensions d'analyse et visualisation

Un cadre conceptuel est globalement reconnu pour décrire la variable temporelle (Peuquet, 1994) (Müller & Schumann, 2003) (MacEachren, 2004) (Aigner et al., 2007) :

Le temps peut être organisé selon deux structures conceptuelles primaires :

- *Linéaire (séries temporelles)* : le temps est observé selon une séquence linéaire d'événements du passé jusqu'au présent
- *Cyclique (cycles temporels)* : le temps est traité selon des cycles répétitifs liés à la rotation et à la révolution de la terre. Des cycles sociologiques sont également observés.

Le temps est décrit selon deux types d'échelles :

- *Discrète (ordinaire)* : lorsque le temps est décrit selon une échelle discrète, il n'est possible de répondre qu'aux questions de type avant/après
- *Continue* : lorsque l'échelle est continue, les différences temporelles peuvent être quantifiées et comparées.

Le temps peut être décomposé en entités de base :

- *L'événement* : un point identifiable dans le temps à un moment spécifique
- *La période (la scène)* : durée entre deux états identifiants (deux événements)
- *L'épisode* : un élément d'une séquence de scènes ordonnées de façon linéaire
- *La phase* : un élément identifiable d'un cycle (un épisode de la structure cyclique du temps).


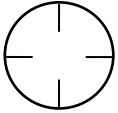



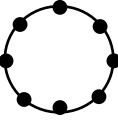
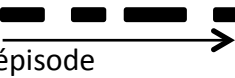
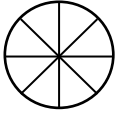

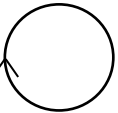
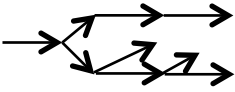
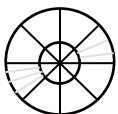
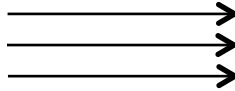
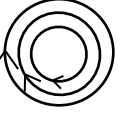

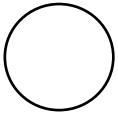

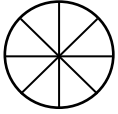

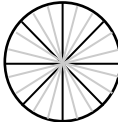
Le temps peut être observé selon trois perspectives :

- *Ordonnée* : les événements se produisent les uns après les autres
- *Ramifiée* : lorsque plusieurs alternatives sont considérées, que plusieurs séquences d'événements peuvent succéder à chaque événement
- *Multiple* : lorsque plusieurs alternatives sont considérées, que plusieurs séquences d'événements sont observées en parallèle.

Le temps peut être décrit selon trois niveaux d'agrégation:

- *Aucune* : lorsque le temps est considéré de façon continue
- *Simple* : tel que les secondes d'une minute, les jours de la semaine, etc.
- *Multiple* : la combinaison de plusieurs granularités engendre une structure hiérarchique, comme les calendriers qui intègrent les jours, les semaines, et les mois.

Cadre conceptuel de la dimension temporelle

		Structure	
		Linéaire	Cyclique
Echelles	Discrète		
	Continue		
Entités	Événement		
	Période		
Perspectives	Ordonnée		
	Ramifiée		
	Multiple		
Niveaux d'agrégation	Aucune		
	Simple		
	Multiple		

Représentations inspirées de (Aigner et al., 2007)

Les questions d'analyse se décomposent également en catégories (MacEachren, 2004):

Question d'analyse	
<i>L'existence</i>	Est-ce que l'élément existe à un moment précis ?
<i>La position temporelle</i>	Quand ? La position décrit le moment de l'observation.
<i>L'intervalle temporel</i>	Combien de temps ? L'intervalle décrit la durée de vie de l'observation.
<i>La fréquence temporelle</i>	Avec quelle fréquence ? la fréquence décrit un intervalle de temps entre les observations en fonction d'une unité temporelle définie.
<i>Le taux de changement</i>	A quelle vitesse, selon quelle différence ? Le taux de changement décrit le ratio entre le nombre d'observations entre deux événements et leur intervalle temporel.
<i>La séquence</i>	Selon quel ordre ?
<i>La synchronisation</i>	Les observations se produisent-elles simultanément ?

De nombreuses formes de visualisations ont été développées pour analyser les données dans la dimension temporelle, afin de détecter des patterns et répondre aux questions présentées ci-dessus. Certaines permettent de répondre à plusieurs questions, alors que d'autres sont spécifiques à une question particulière. Sans être exhaustifs, les tableaux des pages suivantes résument les différentes approches de visualisations exploitées pour faciliter l'analyse temporelle des données.

Représentations avec un système de coordonnées cartésiennes

Les lignes de temps

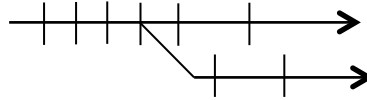


Diagramme à **une dimension** : les événements sont représentés le long de l'axe horizontal. Des périodes peuvent également être représentées.

Les séries chronologiques

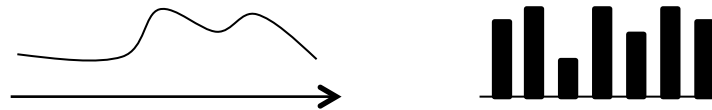
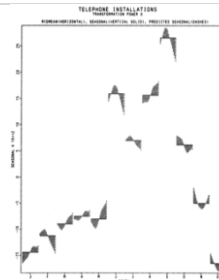


Diagramme à **deux dimensions** : l'axe horizontal représente le temps. La variable peut être continue ou discrète et organisée selon une structure linéaire ou cyclique. **L'axe vertical représente une variable quantitative**. De nombreuses sous-catégories de représentations peuvent être définies.

Les diagrammes saisonniers (Seasonal Subseries Plot)



tiré de (Cleveland & Terpenning, 1982), p.55

Diagramme à **deux dimensions** : l'axe vertical représente une variable quantitative. **Le temps est ordonné en « saison » (structure cyclique) sur l'axe horizontal**. Chaque sous-diagramme contient la distribution pour une période définie de chaque phase.

Les diagrammes d'activité (de Gantt)

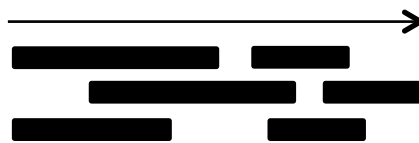


Diagramme à **deux dimensions** : visant la comparaison d'événements, d'activités ou d'actions. **L'axe vertical est catégorisé** selon des activités, des tâches, des personnes, etc.

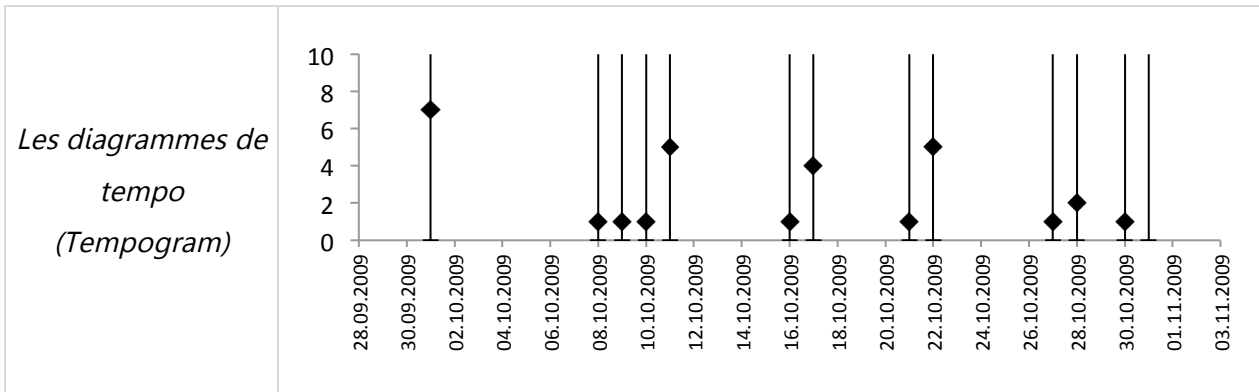
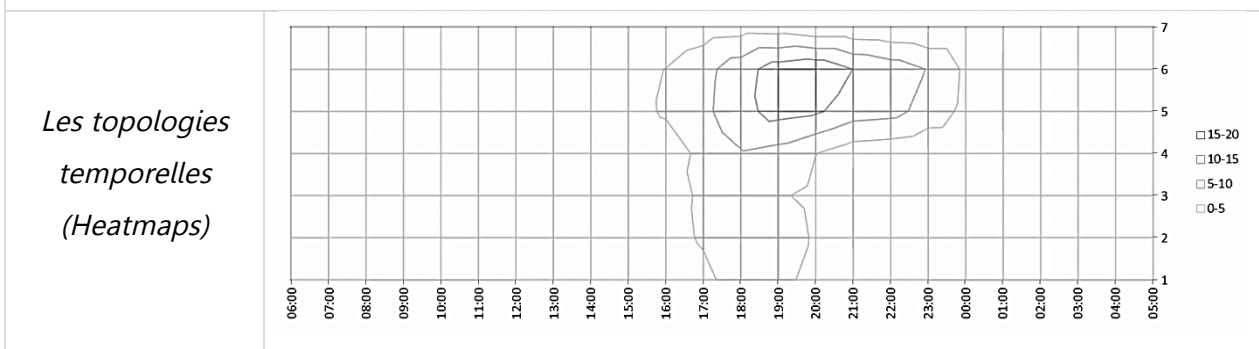
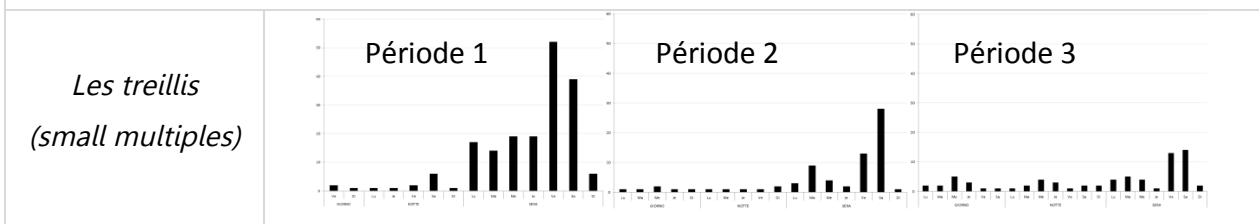


Diagramme à **deux dimensions** : l'axe horizontal représente le temps et **l'axe vertical l'intervalle de temps avec l'occurrence suivante.**



Les deux axes du diagramme représentent la dimension temporelle : les heures de la journée sur l'axe horizontal et les jour de la semaine sur l'axe vertical⁷². Une structure linéaire peut également être représentée sur l'un des axes (par exemple, les semaines de l'année pour observer la saisonnalité d'un phénomène).



Ensemble de diagrammes similaires (à une ou deux dimensions). Chaque diagramme illustre une situation à un moment différent. **L'invariant de chaque représentation couvre la dimension temporelle.**

⁷² Les données utilisées sont les mêmes que celle présentées dans la section « analyse quantitative ». Le nombre de cas est plus important les vendredis et samedis soirs (5 et 6 sur l'axe vertical). L'étendue est plus grande sur ces soirées du week-end, comme précédemment expliqué.

Représentations avec un système de coordonnées polaires

*Les diagrammes
circulaires
(Radars)*

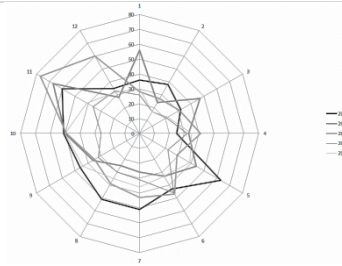
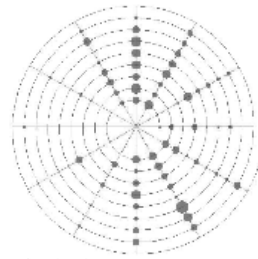


Diagramme **représentant le temps selon une structure cyclique discrète**. L'objectif de ces graphiques est de comparer plusieurs itérations d'un cycle (défini par le niveau d'agrégation) afin de détecter des régularités et des anomalies.

*Les diagrammes
en spirale*



tiré de (Carlis & Konstan, 1998), p.32

Diagramme **représentant le temps selon une structure cyclique continue**. L'objectif est similaire à celui des diagrammes circulaires.

Représentations temporelles intégrant une dimension relationnelle

Méthode de chemin critique
(Critical path method)

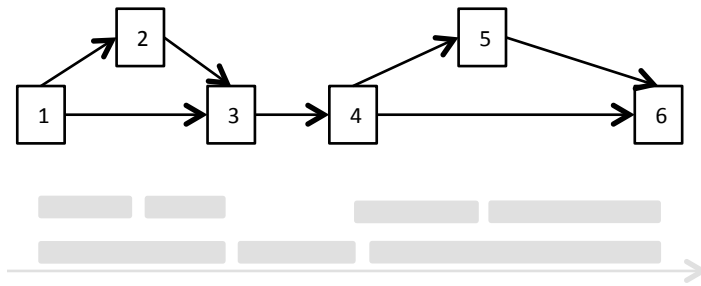


Diagramme exploité pour décrire une séquence d'activités (de périodes). **Chaque flèche représente une activité** dont la longueur sur l'axe horizontal est proportionnelle à sa durée. Les relations entre les activités sont illustrées par des sommets numérotés décrivant la séquence (Kelley & Walker, 1959).

Réseau PERT
(PERT chart : Program Evaluation and Review Technique)

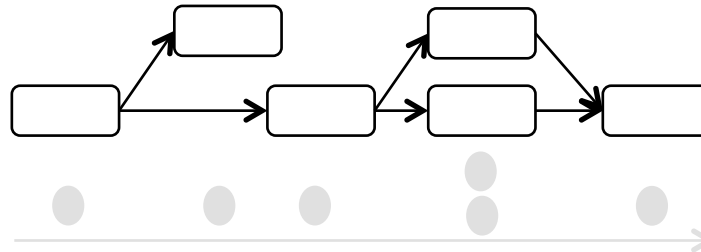
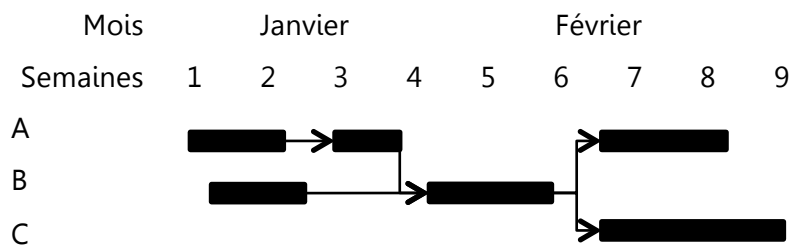


Diagramme exploité pour décrire une séquence d'événements. **Chaque noeud représente un événement.** La durée n'est pas représentée, mais uniquement le moment de l'événement.

Diagramme d'activité

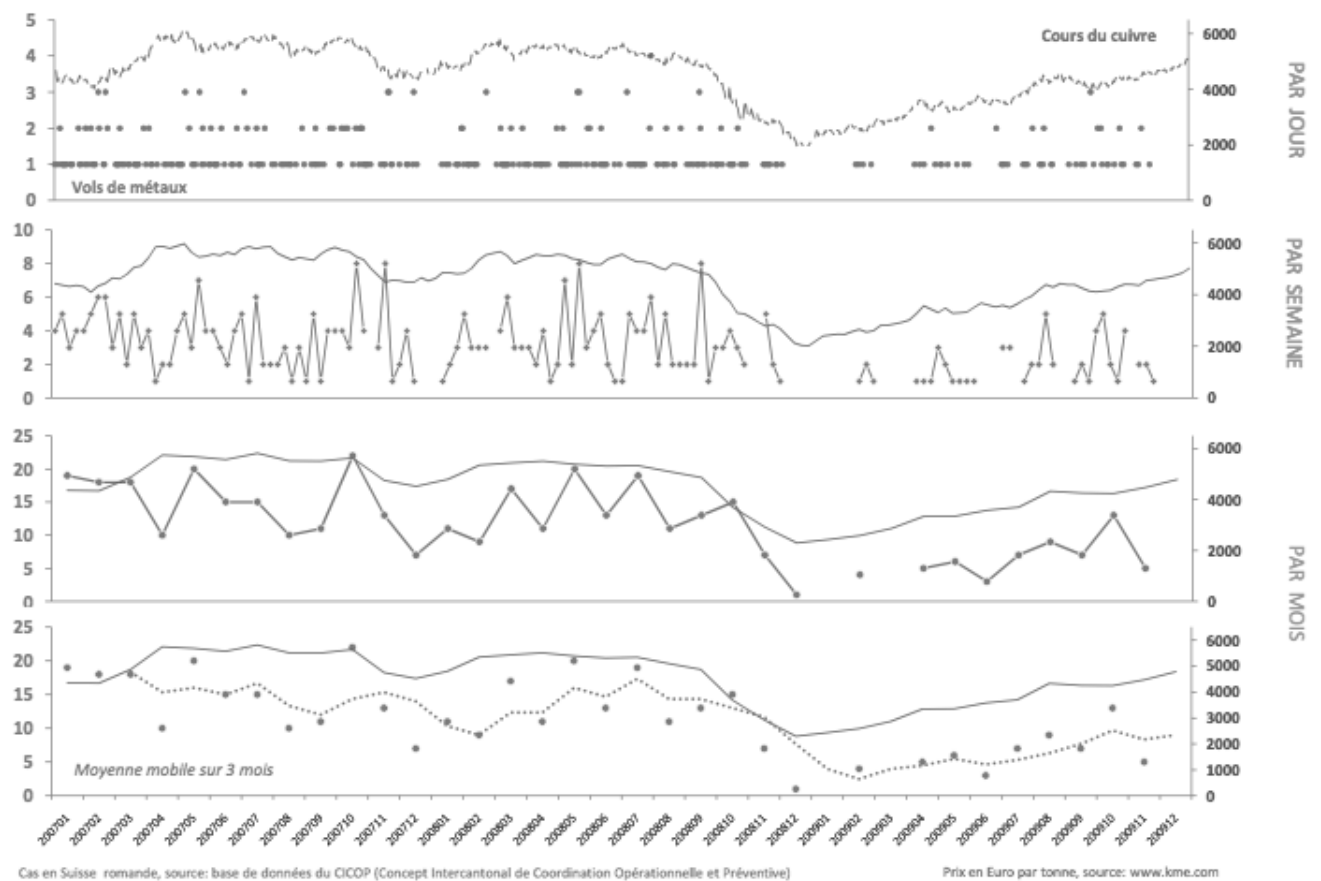


Chaque activité est représenté par une barre horizontale dont la longueur est proportionnelle à sa durée. Les relations entre les activités (par exemple, si l'une ne peut commencer qu'à la fin de certaines autres) sont représentées par des flèches. L'axe vertical est exploité pour décomposer soit des types d'activités, de personnes, etc.

Différents types de patterns sont observables lors de l'analyse temporelle des données (Boba, 2009) (Helms, 2009) (Few, 2009) (Paulsen et al., 2010) :

- **Les tendances** décrivent un comportement général. Elles peuvent être de trois types : décroissante, stable ou croissante⁷³.
- **La variabilité** décrit le degré de changement. Elle est généralement décrite par l'écart-type et l'étendue de la variable.
- **Le taux de changement** décrit la différence de valeur entre deux observations successives.
- **La covariance** décrit la relation entre deux variables temporelles

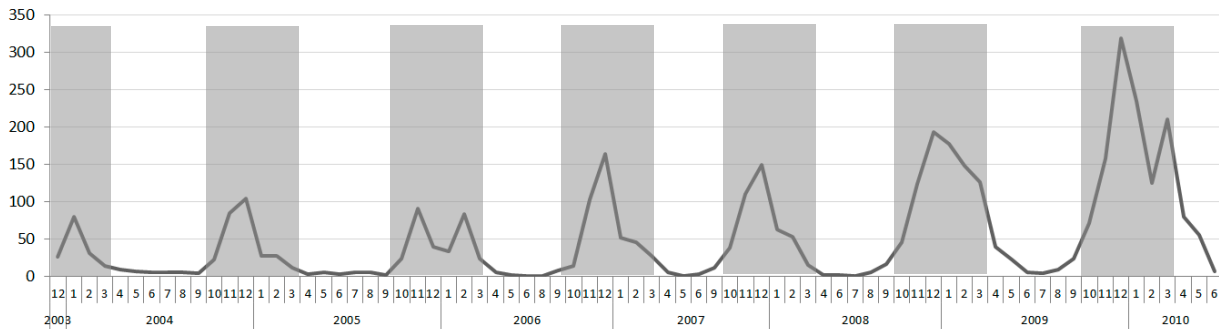
Covariance entre le nombre de vols de métaux et le cours du cuivre.



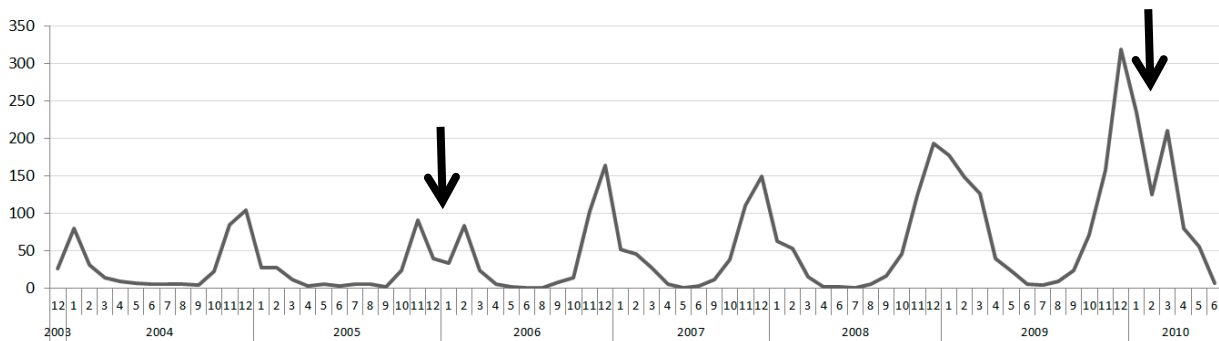
⁷³ Le tempo est également classifié selon trois types de tendances : accélération, stabilisation et décélération.

- **Les régularités cycliques** sont observées lorsque le temps est analysé selon une structure cyclique (jour de la semaine, mois de l'année, etc.)

Répétitions saisonnières des cambriolages du soir



- **Les exceptions** sont des valeurs anormales qui ne suivent pas la tendance générale.

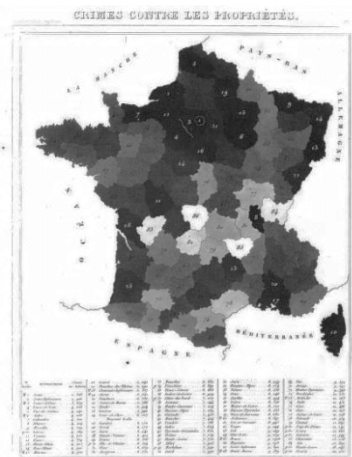


La gestion de l'incertitude liée au temps est particulièrement critique en analyse criminelle. En effet, l'information temporelle collectée pour décrire une activité criminelle est souvent imprécise. Les événements criminels sont généralement décrits par un intervalle entre deux événements (la date et l'heure de début et la date et l'heure de fin du délit). Cet intervalle peut dans certains cas décrire avec précision la durée du délit (par exemple lors de brigandages), mais il est plus souvent le fait d'une incertitude (le délit s'est produit à un moment indéterminé entre deux événements). C'est généralement le cas des cambriolages commis pendant l'absence des lésés. L'intervalle est alors défini par le moment où la victime a quitté le lieu de commission du délit et le moment où elle est revenue. Différents types d'estimateurs, tels que la moyenne, le médiane ou un calcul de fréquences peuvent alors être exploités pour gérer l'incertitude temporelle (Ratcliffe, 2000).

Analyse spatiale

L'usage de représentations cartographiques de l'environnement semble, pour la plupart des civilisations, coïncider avec l'invention de l'écriture. La plus ancienne représentation géographique découverte pourrait être une carte babylonienne du monde (gravé entre 500 et 700 av. J.-C.). La cartographie s'est largement développée au cours de l'histoire, notamment pour la navigation. Dès le XIXe siècle apparaissent les premières cartes dites thématiques visant à représenter des données qualitatives et quantitatives en relation avec leurs distributions spatiales.

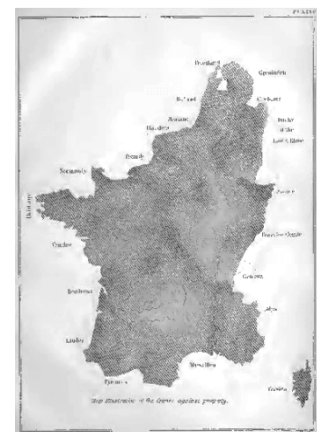
Le crime, comme tout événement, possède évidemment une réalité géographique. De très nombreuses études ont montré la nature non aléatoire des phénomènes de criminalité dans la dimension spatiale. Il ne semble toutefois pas exister de consensus sur les causes des distributions observées (Canter, 2000).



Les premières cartes de la criminalité sont attribuées aux travaux d'André-Michel Guerry (1802-1866) sur la statistique morale en France (Guerry, 1833). Guerry a montré que les taux de criminalité (décomposés par département et par types de criminalité : contre la personne et contre le patrimoine) variaient fortement selon la région géographique et également selon les saisons. Ces travaux sont considérés comme fondateurs des sciences sociales modernes (Friendly, 2008).

A la même période, Adolphe Quetelet (1796 – 1874) effectue également des cartes de la criminalité (Quetelet, 1842).

Ces premières analyses spatiales de la criminalité furent possibles grâce aux réformes entreprises, au début du XIXe siècle, en France et en Angleterre, en matière de collecte et d'enregistrement des données de la criminalité (Anselin et al., 2000) (Chainey & Ratcliffe, 2005).



Au début du XXe siècle, l'école de sociologie de Chicago, exploitera largement les cartes pour mettre en évidence les relations socioculturelles et socio-économiques de la criminalité, en

fonction des quartiers de la ville de Chicago, notamment lors d'études longitudinales de la délinquance juvénile (Canter, 2000) (Chainey & Ratcliffe, 2005).

Un regain d'intérêt est observé en matière de cartographie du crime depuis les années nonante. Clarke relève que « *Crime mapping will become as much an essential tool of criminological research as statistical analysis is at present* » (Clarke, 2004) p. 60. Des efforts particulièrement importants en termes de recherche et de développement d'outils facilitant l'analyse spatiale ont été effectués. La cartographie est largement exploitée afin d'analyser la distribution géographique du crime et de l'activité policière. Elle sert de base à la prise de décision opérationnelle en identifiant des zones particulièrement touchées, afin d'allouer les ressources et définir des stratégies d'action (Anselin et al., 2000) (Chainey & Ratcliffe, 2005) (Boba, 2009) (Taylor et al., 2007).

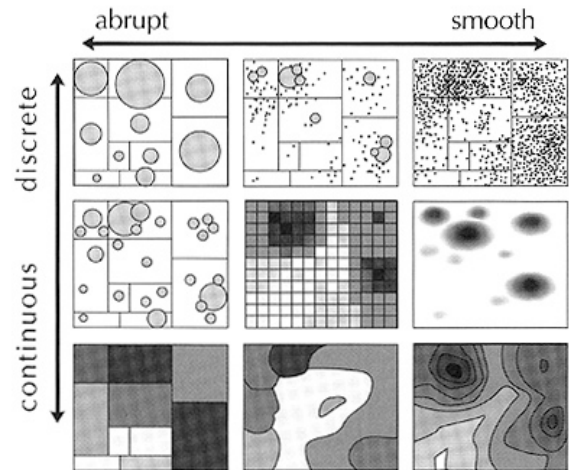
En 1994, la police de New-York instaure une nouvelle approche managériale basée sur l'analyse des informations notamment présentées sous forme de cartes, afin d'identifier les problèmes et définir les réponses policières. La composante géographique est également exploitée afin d'effectuer du profilage géographique. C'est-à-dire de mettre en relation les événements criminels et les activités routinières des criminels et des victimes afin d'identifier notamment le lieu de résidence des auteurs, des points de rencontre ou de passage (Paulsen et al., 2010). La première exploitation de cartes géographiques à des fins de profilage, a probablement été effectuée dans l'affaire du « Yorkshire Ripper » (1976-81) (Kind, 1987) (Kind, 1994) (Barclay, 2009).

La définition d'un cadre conceptuel décrivant la variable spatiale présenté ici est tiré de (Peuquet, 1984). Les données spatiales peuvent être classifiées selon quatre types principaux :

- *Les points* : une position unique est associée à chaque observation
- *Les lignes* : la composante géographique est définie par un ensemble de positions pouvant être représentées par des lignes isolées (telles que des failles), des structures en arbre (comme les rivières) ou des réseaux complexes (les routes)
- *Les polygones* : la composante géographique est définie par des positions formant un ensemble fermé (des zones). Les polygones peuvent être isolés (ne partageant pas de « frontière » avec d'autres zones), adjacents ou imbriqués.

Une typologie de la variable spatiale (et des formes de visualisation spatiale) est également définie selon deux dimensions (MacEachren, 2004):

- *La continuité spatiale* décrit la distribution du phénomène. Elle est continue lorsque le phénomène est observé sur l'ensemble du territoire et discrète lorsqu'il a lieu à des endroits séparés
- *La dépendance spatiale* décrit le niveau de variation spatiale du phénomène. Si le phénomène possède une indépendance spatiale, les observations peuvent varier fortement entre deux lieux proches (la distribution est abrupte). Les variations spatiales sont faibles (lisses), si le phénomène possède une forte dépendance spatiale.



tiré de (MacEachren, 1992) p. 17

La composante spatiale d'une observation (d'une trace) peut être de différente nature (Reed & Kocura, 2005) :

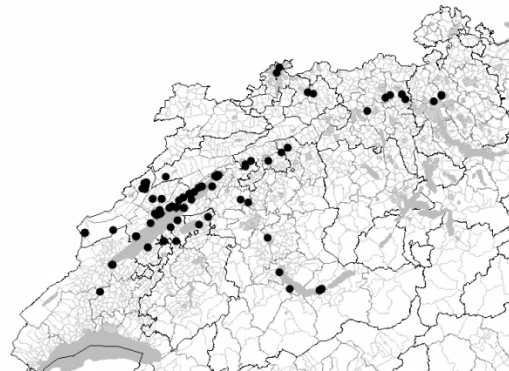
- *Un point unique* (comme un événement)
- *Un ensemble de points* (comme un appel téléphonique caractérisé par la position de l'appelant et de l'appelé)
- *Un chemin* (tel le déplacement d'un véhicule enregistré par une balise GPS)
- *Une zone* (telle la couverture d'une antenne GSM).

Diverses techniques sont au service de l'analyse spatiale de la criminalité. Besson les regroupe en neuf catégories (Besson, 2004):

- *L'analyse du point* (formation et dispersion : Weisburd, Anselin)
- *Le groupement de points* (hotspot : Sherman, Eck, Gersh, Taylor, Jerrifs)
- *L'analyse des distances* (distance décroissante, zone tampon : Brantingham, Rossmo)
- *Le quotient relatif local de criminalité* (Brantingham)
- *Le quotient relatif de temps* (Ratcliffe, Bair)
- *L'analyse des facteurs d'influence*
- *La régression spatiale*
- *L'auto-corrélation spatiale* (Anselin, Jefferis)
- *Les mesures de contrôles et de tests*

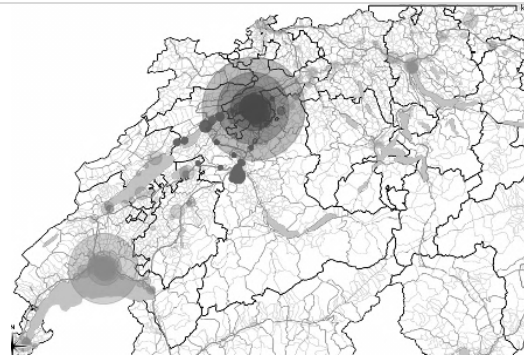
Représentations spatiales

Les cartes en points



Cartes représentant **chaque observation par un point (ou une forme)**. Une analyse quantitative est généralement difficile à effectuer, en raison de la superposition des points. Seule la distribution spatiale est observable.

Les cartes à symboles proportionnels



Cartes exploitant **l'aire des cercles (ou de formes) pour représenter une variable quantitative**. La problématique de la superposition est atténuée en exploitant la transparence.

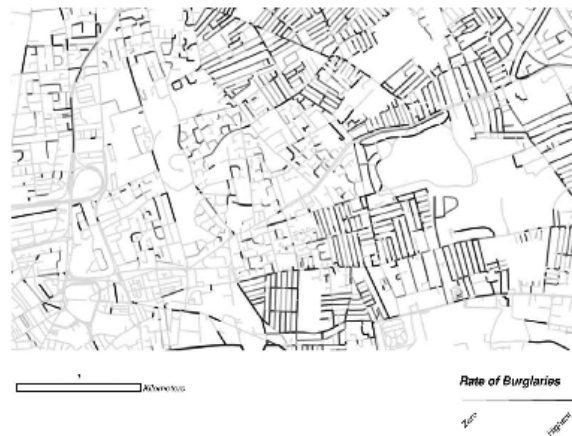
Les cartes statistiques



tiré de <http://sitn.ne.ch/geoclip/carto.php>, dernier accès le 21.09.2010

Cartes exploitant des **diagrammes quantitatifs** (histogrammes, circulaires sectorisés, etc.) pour représenter la distribution spatiale de plusieurs variables. Les cartes à symboles proportionnels sont parfois considérées comme un sous-type des cartes statistiques.

*Les cartes à
lignes graduées*



tiré de (Johnson, 2010)

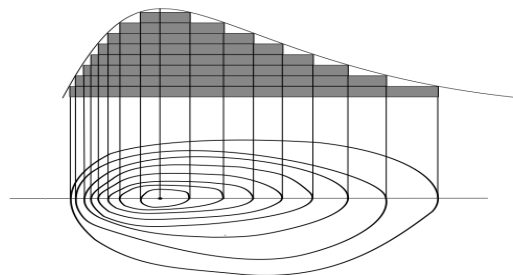
Cartes exploitant **l'épaisseur des lignes** pour représenter une variable quantitative.

Les cartes choroplèthes



Cartes exploitant **une échelle de niveau de gris ou de couleur** des polygones pour représenter une variables quantitatives. Les polygones sont généralement définis en fonction de limites administratives (communes, districts, pays, etc.).

*Les cartes
isolignes/isoplèthes*



tiré de http://upload.wikimedia.org/wikipedia/commons/d/d1/Courbe_niveau.svg, dernier accès le 21.09.2010

Cartes utilisant des **lignes de points dont la valeur associée est constante**. Contrairement aux cartes choropleth, les lignes ne dépendent pas de polygones prédéfinis (telles des frontières administratives).

Les cartes de densité
(hotspots)



Cartes de **type raster** dont la valeur de chaque maille de la grille est **basée sur un calcul de densité** (généralement par densité de Kernel avec un estimateur gaussien).

Les cartes à déformations
(Cartograms)

Figure 3
Burglary Rate per 1000 housing units, Planimetric Map

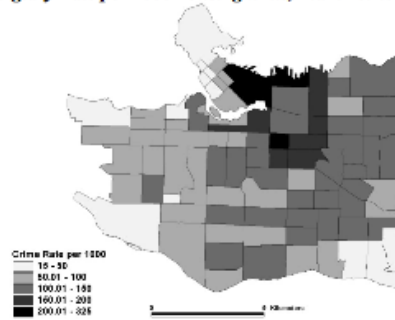
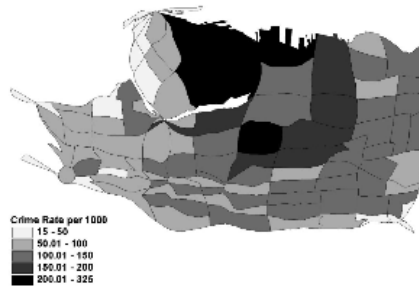


Figure 4
Burglary Rate per 1000 housing units, Cartogram

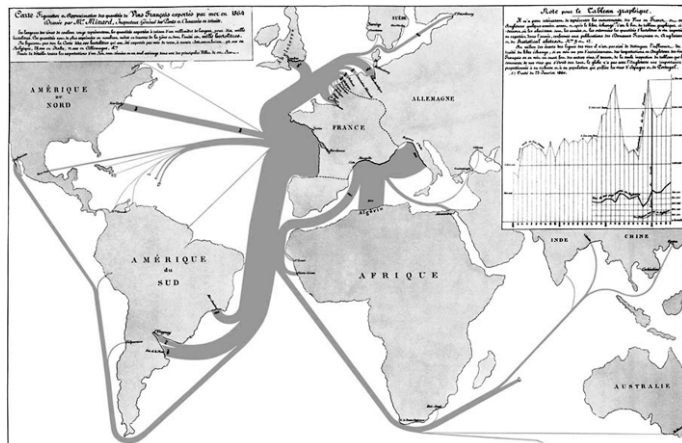


tiré de (Andresen et al., 2009)

Cartes de type choroplèthe **où la zone (les dimensions du plan) est proportionnelle à une variable quantitative**, et non plus directement fixée selon la composante spatiale. Deux variables quantitatives différentes peuvent donc être représentées respectivement par une échelle de niveaux de gris et par la taille des polygones.

Représentations spatiales intégrant une dimension relationnelle

Les cartes de flux



tiré de <http://datavisualization.ch/tools/flow-map-layout>,
dernier accès le 21.09.2010

Cartes représentant **le flux de marchandises ou de personnes**. L'avantage de ce type de carte réside dans la simplification de l'ensemble des relations en les regroupant par des chemins (Phan et al., 2005). Les premières cartes de flux sont attribuées à Charles Joseph Minard (1781 – 1870), auteur notamment d'une carte des pertes successives en hommes de l'armée de Napoléon et de la carte de l'exportation de vin depuis la France, présentée ci-dessus (Friendly, 2008).

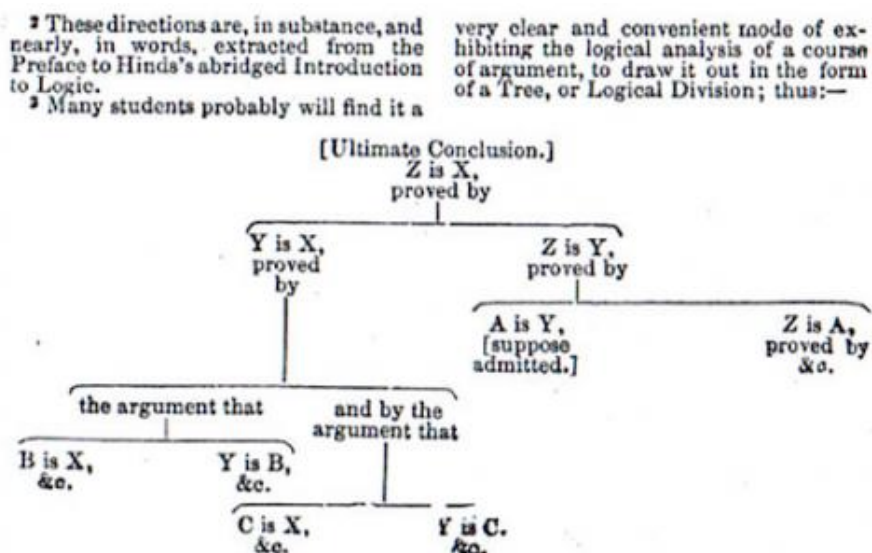
Les graphes géo-référencés



Cartes et graphes juxtaposés. La position des nœuds du graphe est déterminée par la position spatiale de chaque observation.

Annexe 2 : Schémas de scénarios

L'exploitation de graphes pour représenter les informations d'une enquête, se retrouve déjà dans les travaux de John Henry Wigmore (1863-1943) qui pourraient être basés sur les diagrammes développés par Richard Whately (1787-1863) publiés en 1826⁷⁴ (Reed & Rowe, 2007) :



La méthode de Wigmore appelée « argument diagram » est constituée de sommets et d'arcs (Wigmore, 1913) p.751 et suivantes. Les sommets représentent des « faits » (des observations, des portions d'une séquence d'événements constituant l'histoire/le scénario de l'affaire, des portions de déclarations). Les arcs (liens) représentent des inférences reliant les différentes portions d'information. Puisque les liens représentent des inférences entre des prémisses et des conclusions, ils sont généralement accompagnés d'une flèche qui décrit la causalité de la relation (Bex et al., 2003) (van den Braak et al., 2006).

⁷⁴ Illustration tirée de <http://timvangelder.com/page/5/?archives-list&archives-type=months>, dernier accès le 31 janvier 2011.

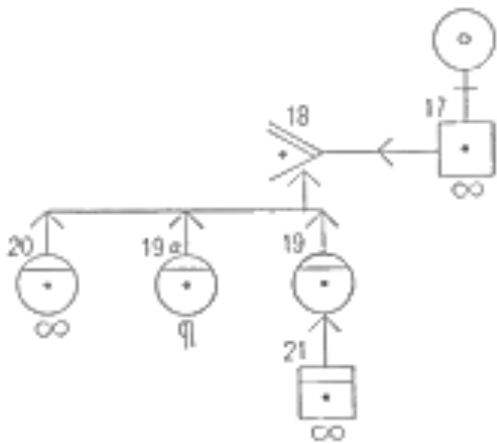


Illustration tirée de (Wigmore, 1913) p. 754.

Diverses notations sont exploitées afin de distinguer différents types d'informations ⁷⁵: les preuves circonstancielles (cercles), les preuves testimoniales (carrés) et les hypothèses (triangles). Les lignes verticales représentent des relations entre des permises confirmant les conclusions. Les lignes horizontales marquées d'une croix représentent des relations réfutant les conclusions. La crédibilité des informations est également représentée par une annotation des nœuds (« ? » : incertain, « • » :

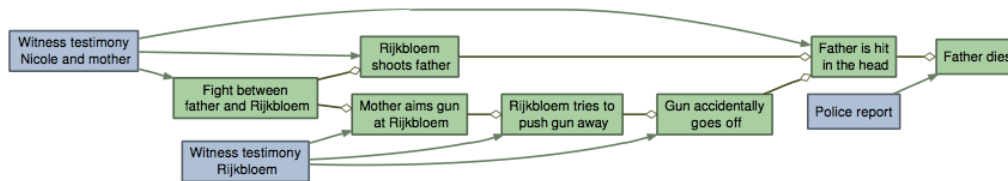
crédible, « •• » : plus crédible, « ° » ou « °° » : non-crédible). Les chiffres sont des indexes qui font référence aux informations (présentées sous la forme d'une liste accompagnant le schéma) (Bex et al., 2003) (Anderson, 2007).

L'un des aspects essentiels de la méthode de Wigmore réside dans l'objectif d'exploitation de ces représentations : raisonner de manière critique sur les informations (Bex et al., 2003). Le graphique explicite le processus de raisonnement effectué pour aboutir à une conclusion et prendre une décision (Goodwin & Fisher, 2000). Cette forme d'exploitation des graphes est initialement appréhendée comme un support de raisonnement pour la prise de décision lors de la phase de jugement (afin d'évaluer les diverses hypothèses définies) (Tillers & Schum, 1991). Selon Tillers (2007), de telles représentations devraient faciliter l'appréhension de la logique bayésienne pour le jury, les avocats et les étudiants en droits notamment. Il émet également l'hypothèse que la visualisation peut aider à améliorer nos limites cognitives et expliciter les raisonnements formels effectués sur les traces. La représentation facilite l'analyse des informations et l'évaluation des hypothèses, mais leur développement reste du ressort de ses utilisateurs (Bex et al., 2007). Un graphe n'est toutefois pas uniquement un support commode pour représenter des raisonnements complexes, mais une aide efficace lors de l'élaboration d'hypothèses et pour faciliter les raisonnements (Tillers, 2007).

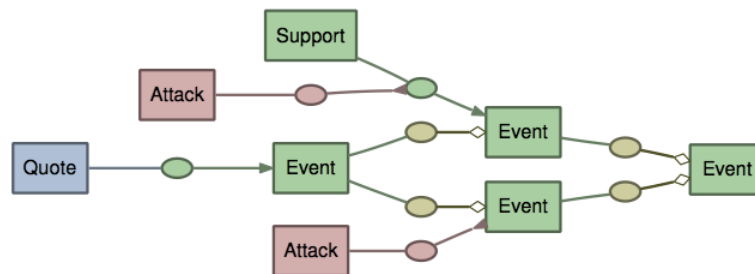
Dans les années cinquante, deux approches complémentaires ont été développées : les diagrammes de Toulmin et les diagrammes de Beardsley (Reed & Rowe, 2007). Depuis les années nonante, plusieurs développements parallèles ont été effectués et différents logiciels développés pour représenter les raisonnements effectués sur la base des traces en regard de

différents scénarios (Tillers & Schum, 1991) (Goodwin & Fisher, 2000) (Bex et al., 2003) (Reed & Rowe, 2007).

De nouveaux outils sont développés afin de représenter les informations d’une enquête, d’exprimer les inférences effectuées, de comparer les versions et les hypothèses alternatives possibles pour interpréter les traces collectées (van den Braak, 2010). Van den Braak (2010) définit un langage visuel formel basé notamment sur les travaux de Bex (2007), afin de concevoir des représentations relationnelles explicitant d’une part, des scénarios (par exemple en regard d’information testimoniales) et, d’autre part, les relations entre les indices et ces scénarios. Le langage intègre sous la forme de nœuds : les informations descriptives des cas (les informations circonstanciées, les témoignages et les traces collectées), les hypothèses formulées sur la base d’observations (physiques ou testimoniales) et les sources d’informations (un rapport de police, un témoignage, un constat, etc.). Les types d’entités sont distingués par la forme et la couleur des nœuds. Les relations décrivent les relations entre les différents types de nœuds.



(a) the Rijkbloem case



(b) support and attack

tiré de (van den Braak, 2010)

Selon Van Den Braak (2010), le langage défini sert de base à la décomposition des informations d’enquête pour évaluer les indices et comparer et évaluer des scénarios alternatifs explicatifs des informations collectées. L’ensemble des informations est donc décomposé sous la forme de scénarios sur lesquels sont intégrées les inférences effectuées sur la base des informations collectées.

Annexe 3 : Visualisation tridimensionnelle

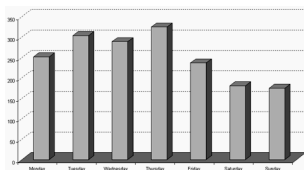
It doesn't seem to matter whether we perform our tasks well or not; many of us like 3D visualizations anyway. Interestingly, many of us do not like 3D by instinct. (Chen, 2006).

Les questions liées à l'usage et l'efficacité de la troisième dimension sur une représentation graphique font débats. Les performances des représentations en trois dimensions varient en fonction des tâches impliquées et des choix de représentation (Cockburn & McKenzie, 2004).

Une visualisation à trois dimensions (3D) est une projection sur le plan de l'environnement tridimensionnel (Tavanti & Lind, 2001). L'image créée est une perspective ou une vue oblique d'un objet ou d'une scène sur un espace à deux dimensions. La perception de la troisième dimension est produite par la perspective résultant de l'angle de vision et de l'usage de repères visuels décrivant la profondeur (tel un éclairage produisant des ombres) (St John et al., 2001).

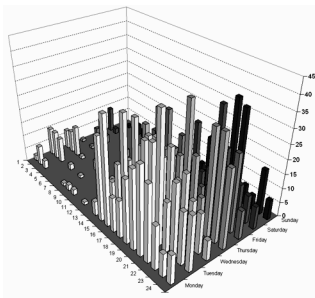
Types de représentation

La classification des méthodes de représentation proposée par (Stasko & Wehrli, 1993) est utilisée pour discuter leurs usages et efficacités :

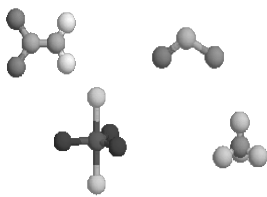


Vues 2D augmentées: la troisième dimension est ajoutée à une représentation en deux dimensions pour des raisons d'esthétisme (par exemple sur des histogrammes), afin d'accroître l'attractivité. Tufte soutient que l'usage de la troisième dimension dans ce type de représentation est superflu (deux dimensions suffisent pour transmettre la connaissance) (Tufte, 2001) (Irani & Ware, 2003). D'autres auteurs pensent que l'ajout d'éléments esthétiques

stimule la perception visuelle et facilite la compréhension qualitative et la présentation des idées (Springmeyer et al., 1992). Toutefois, certaines études (Aigner et al., 2007) (Hicks et al., 2003) ont montré que l'usage de la troisième dimension sur des histogrammes ralentissent le processus de lecture et peuvent induire en erreur le lecteur. En effet, une perspective trop importante peut déformer les tailles des formes. La comparaison des tailles peut alors être trompeuse, dans la mesure où les différences ne correspondent plus aux variations des données sources.



Vues 2D adaptées : plusieurs représentations bidimensionnelles sont intégrées dans une vue à trois dimensions, afin de représenter une information supplémentaire. Le principal désavantage de ces représentations est l'*occlusion* : certaines parties du schéma sont cachées par des éléments placés en premier plan. La lecture de chacune des distributions présentées est donc incomplète.



Vue de structures en 3D : la vue représente la réalité physique des objets (comme la représentation de structures moléculaires en chimie ou la représentation d'immeuble en architecture). Il a été démontré empiriquement que ces représentations sont très efficaces pour comprendre des formes et structures tridimensionnelles (St John et al., 2001).

(Images tirées de « Molecular database without transition elements », <http://www.faidherbe.org/site/cours/dupuis/banque.htm>, dernier accès le 4 février 2010)

L'avantage apparemment évident de profiter d'une dimension supplémentaire pour encoder des informations doit être discuté. Généralement ces représentations sont exploitées dans un environnement où l'utilisateur peut naviguer dynamiquement au sein de la vue (Stasko & Wehrli, 1993). La conception d'interfaces permettant une telle navigation implique des problèmes complexes. Paradoxalement, l'interface peut engendrer des difficultés et des frustrations pour l'utilisateur, mais également un certain enthousiasme et une appréhension plus agréable de la représentation. Certains évaluations relèvent que l'efficacité des représentations 3D semble dépendre plus de la qualité de l'interface que de la nature de la visualisation (St John et al., 2001). Pour Robertson (dans (Mullet, 1995)) et Schneiderman (Shneiderman, 2003), l'évaluation de l'efficacité des représentations 3D devrait se focaliser sur des indicateurs de performance pour accomplir des tâches spécifiques. Des études empiriques ont montré qu'il n'y a pas de différence significative en termes de performances, mais une

préférence (subjective) pour les représentations à trois dimensions (Cockburn & McKenzie, 2000) (Cockburn & McKenzie, 2001). Les utilisateurs tendent à préférer la 3D pour communiquer et pour mémoriser les informations (Aigner et al., 2007). D'autres études (voir notamment (St John et al., 2001), (Teyseyre & Campo, 2009) et (Tory et al., 2006)) ont identifié des avantages et des inconvénients spécifiques qui sont présentés dans les sections suivantes.

Avantages

Description de formes : l'avantage principal des visualisations 3D réside dans son application pour représenter de formes tridimensionnelles.

Reconstruire un volume complexe sur la base de représentation

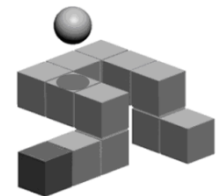
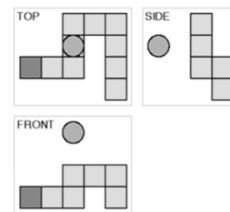
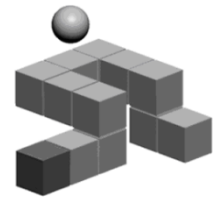
bidimensionnelle nécessite un entraînement

particulier et s'avère souvent bien plus difficile qu'observer une représentation tridimensionnelle

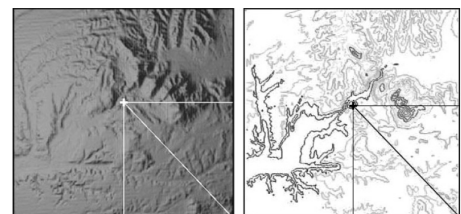
(St John et al., 2001). Toutefois pour que la troisième dimension soit visible, l'usage de repères visuels décrivant la profondeur est nécessaire (tels

que des ombres, des variations de tailles, etc.). Sinon, la projection est ambiguë. Dans

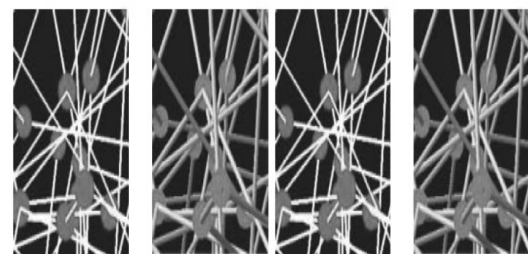
l'exemple ci-contre, sans l'ombre, il n'est pas possible d'identifier à la verticale de quel cube, la sphère est positionnée (exemple tiré de (St John et al., 2001)).



Représentation de topographie : les vues 3D sont très efficaces pour représenter la structure des terrains. La représentation symbolique de l'altitude dans une vue à deux dimensions nécessite plus de temps à interpréter (St John et al., 2001).



Représentation de grands graphes : certains auteurs prétendent que la 3D offre de nombreux avantages pour la représentation de graphes de grande taille (Ware & Franck, 1994) (Parker et al., 1998) (Ware & Mitchell, 2008). De ces travaux, il ressort que les



représentations 3D stéréoscopiques (couplées à des lunettes permettant une immersion) sont plus efficaces que les représentations 2D pour explorer des graphes complexes (en l'occurrence pour trouver le chemin le plus court entre deux nœuds). La possibilité de naviguer

dynamiquement à l'intérieur d'un graphe dans toutes les dimensions peut être une cause explicatrice des performances observées (Chen, 2006).

Toutefois, les relations entre les nœuds et les positions des nœuds ont été définies aléatoirement dans les conditions expérimentales de ces études. Les cas traités ne sont donc pas vraiment représentatifs de situations réelles où la structure des graphes et le positionnement des nœuds sont rarement aléatoires. De plus, l'utilisation d'une telle technique dans la pratique semble difficile en raison du matériel requis pour créer une représentation stéréoscopique de haute résolution. Finalement, l'interface de navigation (translation, rotation and zoom) et les algorithmes de placement (pour minimiser les croisements de traits) sont également des sources de difficultés (Parker et al., 1998) (Teyseyre & Campo, 2009).

Augmentation de l'espace : les représentations 3D offrent un espace plus grand qui permet d'augmenter le nombre d'informations représentées dans une même vue (Mullet, 1995). La structure générale d'une représentation devrait donc être plus facilement perceptible, puisque l'ensemble des données sont visualisées dans un seul espace. Toutefois, la superposition souvent inévitable des éléments représentés sur une représentation 3D est en contradiction avec cet argument.

Accroissement de la mémoire spatiale : la mémorisation des positions d'objets semblent être améliorée par la 3D (Tavanti & Lind, 2001). D'autres études n'ont cependant pas observé de différences significatives (Cockburn & McKenzie, 2004). Ces résultats contradictoires sont certainement liés au fait que les variables visuelles qui favorisent la mémorisation spatiale ne sont pas clairement identifiées. En effet, dans une vue 3D la taille des éléments changent, des textures et des couleurs sont parfois ajoutées (notamment dans (Irani & Ware, 2003)). Lors des comparaisons, plusieurs variables visuelles sont donc modifiées. La relation entre les performances et le passage à la 3D n'est donc pas toujours évidente.

F a i b l e s s e s

Occlusion (superposition) : des parties de la représentation peuvent être cachées par des éléments positionnés au premier plan (voir l'exemple des vues 2D adaptées). Cet effet est partiellement maîtrisable en ajoutant une interaction dynamique (offrant ainsi à l'utilisateur la possibilité de varier les angles de vues) (Hubona et al., 1997).

Ambiguïté due à la projection : en l'absence de repères de profondeur, la position des objets peut-être ambiguë. De plus, alors que les effets de perspective permettent de visualiser la troisième dimension, les possibilités d'effectuer des mesures de distances, d'angles et d'évaluer les positions relatives sont réduites (Tory et al., 2006).

Non-linéarité : l'espace représenté sur une vue 3D est distordu de façon non linéaire. Les distances et les angles sont donc également distordus.

Dynamisme : la possibilité de déplacer la vue est généralement proposée comme solution au problème d'occlusion et de l'ambiguïté due à la projection. Le support de transmission du résultat est alors limité à l'informatique ou à l'image animée.

Représentation des textes : les polices nécessaires pour atteindre le même degré de lisibilité que sur une représentation 2D, doivent être plus grandes sur une représentation 3D (Card et al., 1999).

Finalement, les ressources de calculs nécessaires pour produire une représentation 3D sont plus importantes et l'interface nécessaire est généralement plus complexe à concevoir et à utiliser (Teyseyre & Campo, 2009).

C o n c l u s i o n

Les représentations en trois dimensions offrent des solutions efficaces pour la représentation de scènes, d'objets et de la topographie. En matière de représentation de données abstraites, l'apport de la troisième dimension n'est pas clairement établi. Les représentations en deux dimensions offrent de meilleures performances pour analyser en précision des données et comparer des distances notamment. De plus, une interface généralement plus complexe est requise pour faciliter la navigation dans la vue.

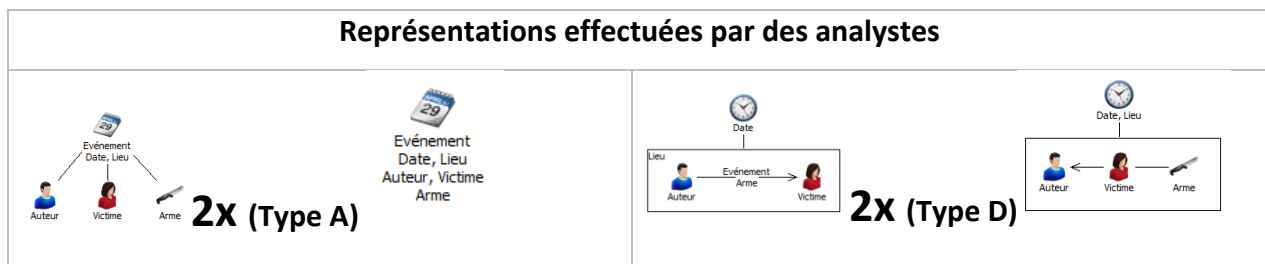
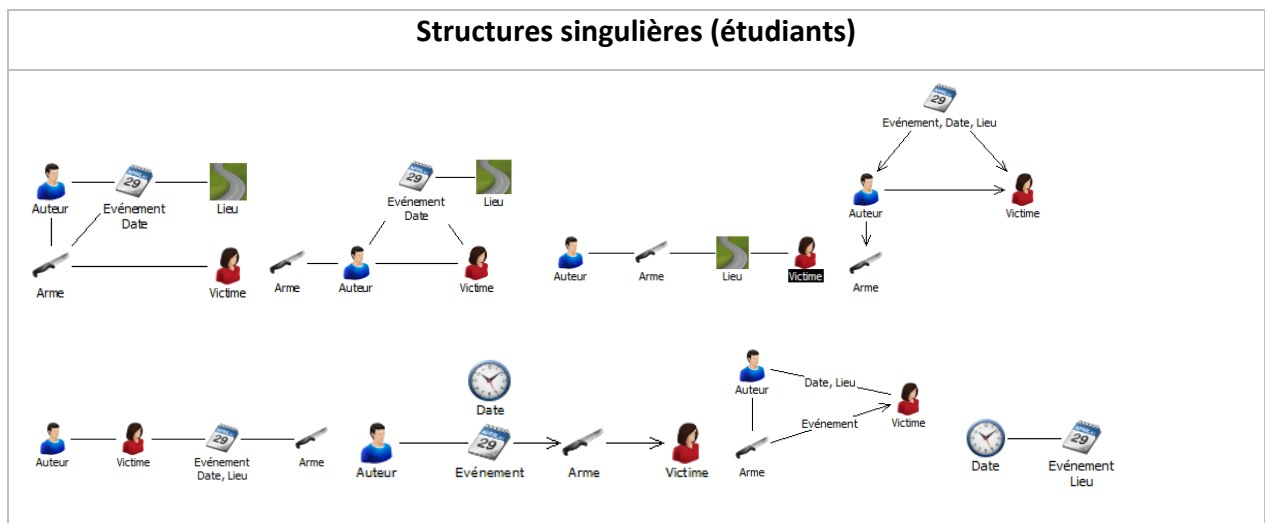
Pour conclure ce rapide survol de la littérature sur le sujet, il semble que deux avis divergents existent en matière de représentation 3D. D'une part, certains chercheurs, notamment influencée par les travaux d'Edward Tufte, pensent que les représentations 2D sont suffisantes. L'usage de la troisième dimension pour augmenter une impression subjective d'esthétisme ne devrait pas être préconisé au détriment de l'efficacité et de la simplicité d'une représentation. Si les informations à représenter peuvent l'être par le biais d'une représentation à deux

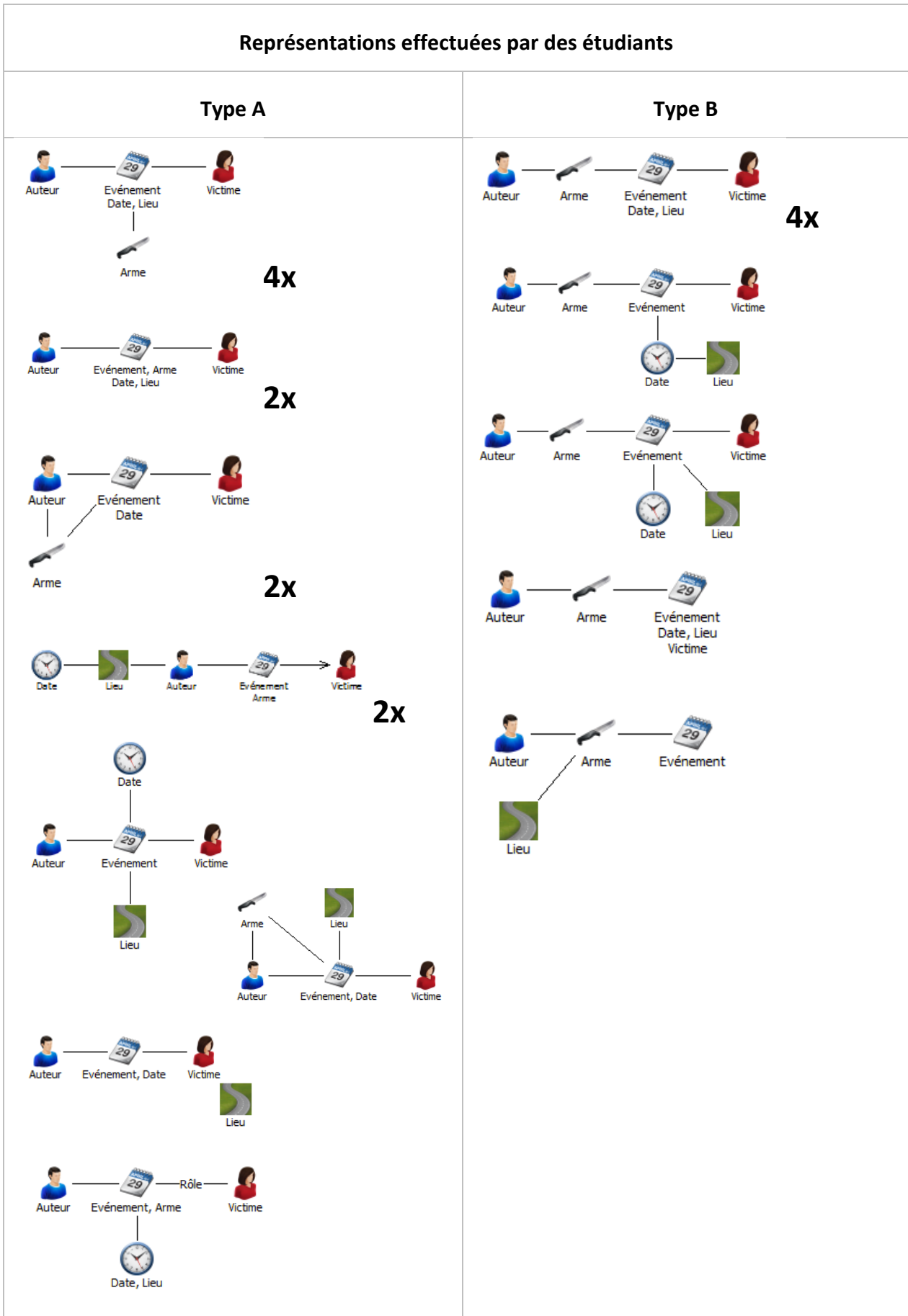
dimensions, il s'agit de la méthode recommandée. D'autres chercheurs (principalement issus du domaine de la conception d'interfaces graphiques et du design) suggèrent d'exploiter la troisième dimension à des fins de communication. En effet, un usage adapté de la 3D semble avoir un potentiel pour faciliter l'interaction entre l'homme et les représentations informatisées, ainsi qu'accroître une certaine forme d'attractivité. Les évaluations en la matière montrent toutefois des résultats contradictoires, dont les causes ne sont pas clairement établies. En effet, les mécanismes complexes de la perception humaine et le nombre important de variables dépendantes impliquées en matière de visualisation (comme la 3D, la couleur, la taille, le positionnement, etc.) rendent difficile les comparaisons et les évaluations. Les représentations à trois dimensions ont démontré leurs utilités pour représenter la réalité physique d'objets (en médecine, en architecture et en chimie notamment), mais la question de l'apport de la 3D en matière de représentation d'informations abstraites reste ouverte. Les désavantages en termes d'intégrité (occlusion et ambiguïté de projection) ne semblent toutefois pouvoir être ignorés au profit d'avantage liés à la communication.

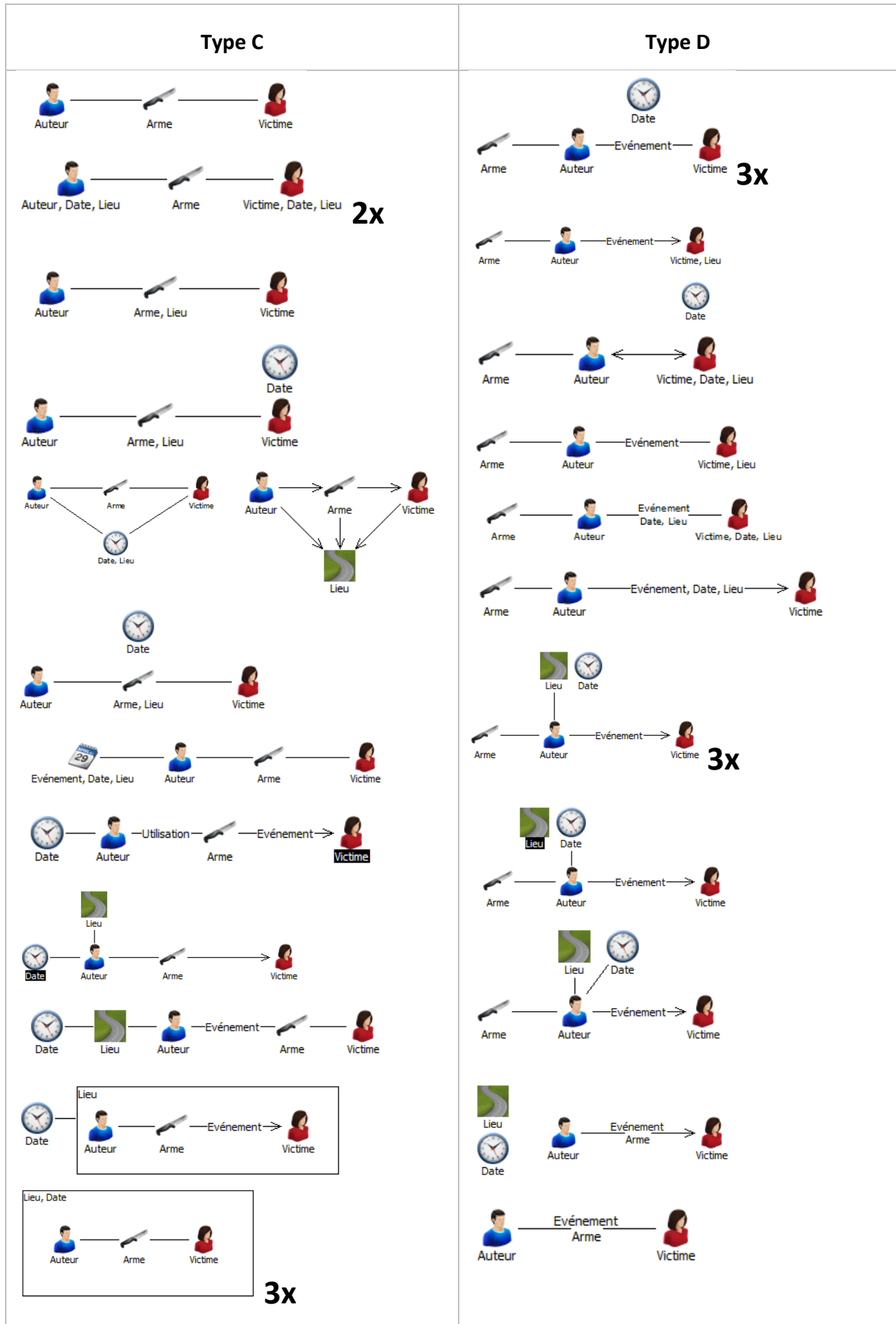
Annexe 4 : Evaluations préliminaires

Représentation d'un scénario simple

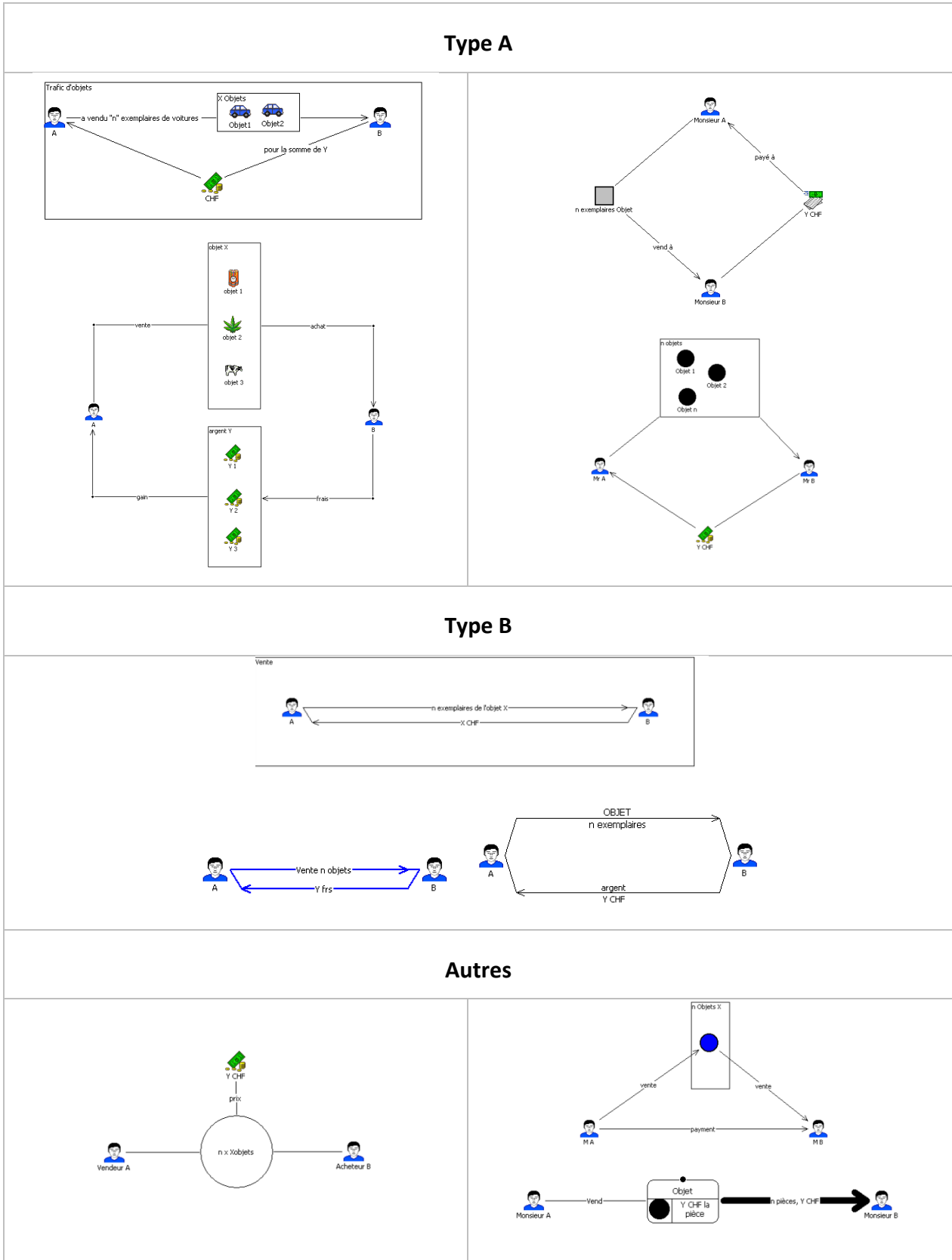
L'ensemble des solutions sont présentées dans les tableaux des pages suivantes. Afin de faciliter la comparaison et la lecture des schémas, ils ont été recréés de façon uniforme.





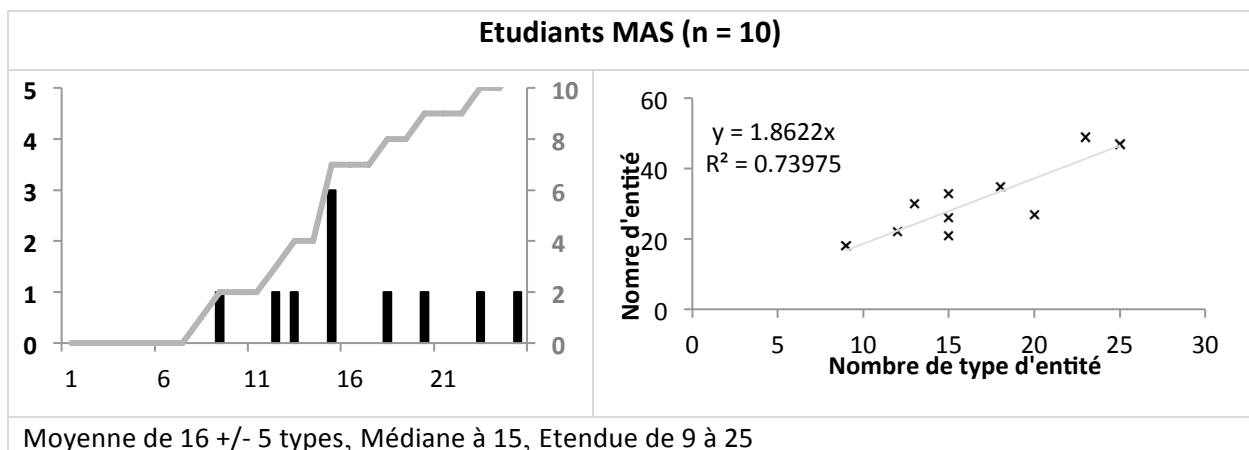
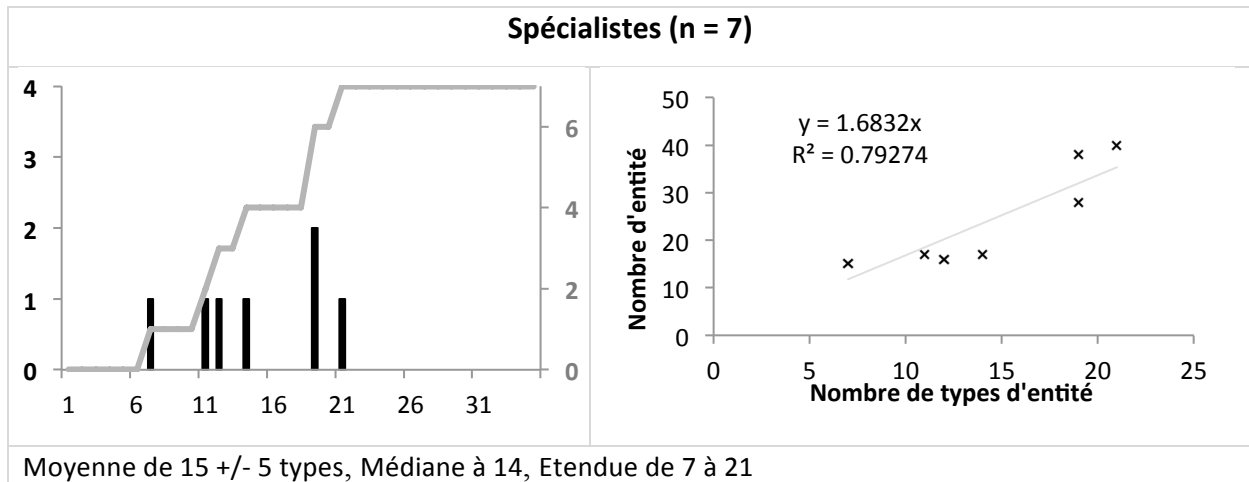
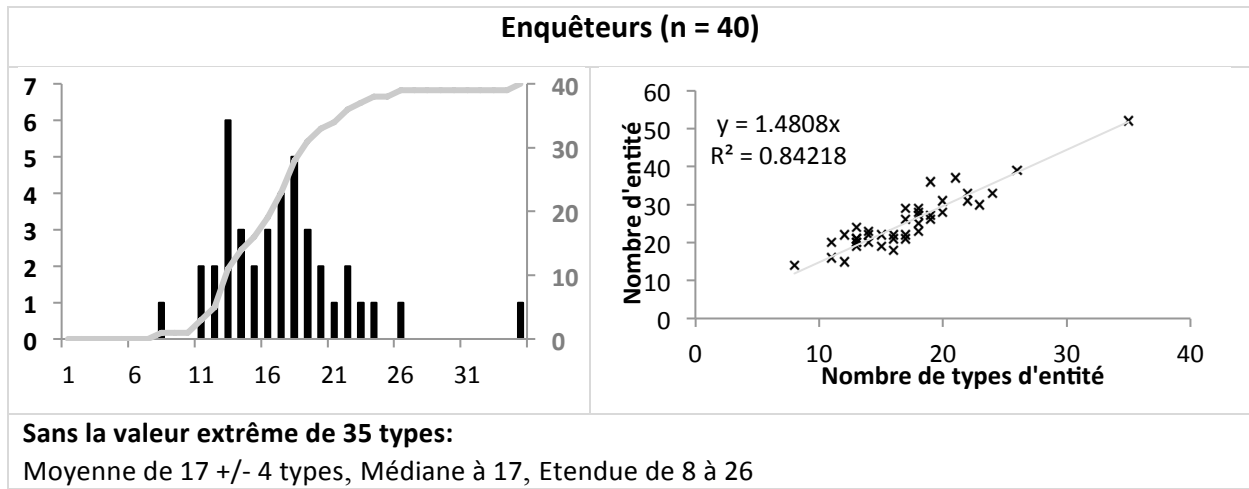


Représentation d'une vente



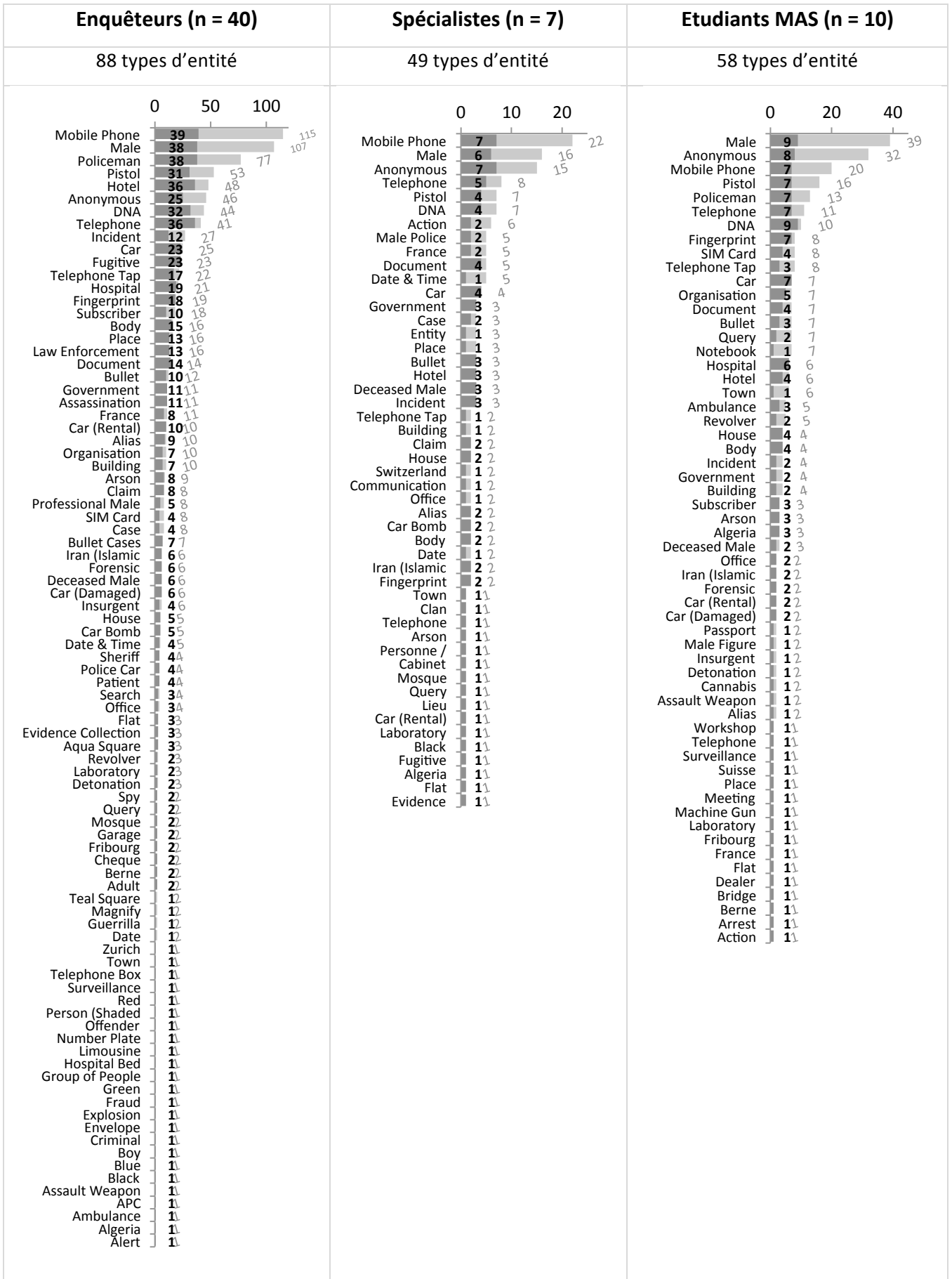
Représentation d'une affaire complexe

Nombre d'entité et de types d'entité exploités :

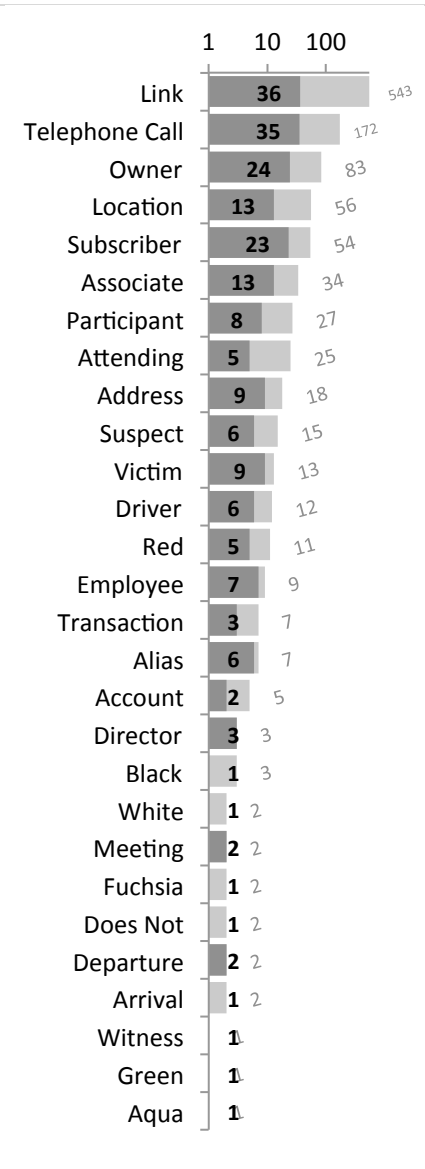
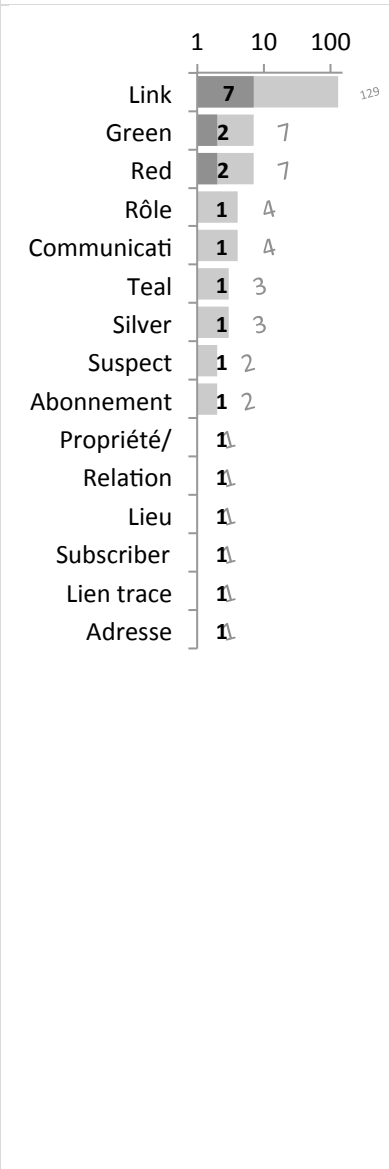
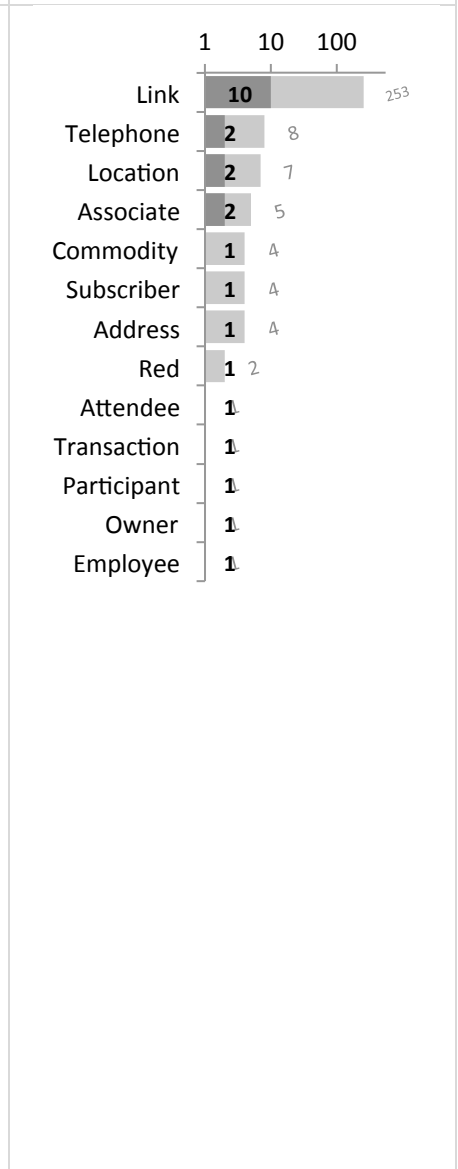


En regroupant les trois populations, la régression linéaire est $Y=1.5663X$, $R^2 = 0.7277$

270 Annexe 4 : Evaluations préliminaires



Types de liens exploités :

Enquêteurs (n = 40)	Spécialistes (n = 7)	Etudiants MAS (n = 10)
28 types de lien	15 types de lien	13 types de lien
		
<p>A ---- B 905 (40)</p> <p>A ---> B 195 (37)</p> <p>A <--> B 12 (7)</p>	<p>A ---- B 124 (7)</p> <p>A ---> B 43 (7)</p> <p>A <--> B 0 (0)</p>	<p>A ---- B 254 (10)</p> <p>A ---> B 35 (5)</p> <p>A <--> B 3 (1)</p>
<p>Trait plein 1052 (40)</p> <p>Trait tillé 60 (25)</p>	<p>Trait plein 153 (7)</p> <p>Trait tillé 7 (4)</p>	<p>Trait plein 285 (10)</p> <p>Trait tillé 7 (3)</p>
<p>Simple 1005 (40)</p> <p>Multiple 97 (22)</p> <p>Double 10 (5)</p>	<p>Simple 160 (7)</p> <p>Multiple 6 (2)</p> <p>Double 1 (1)</p>	<p>Single 283 (10)</p> <p>Multiple 9 (2)</p>

Modification du type de représentation

Ce dernier exemple ne porte pas sur la représentation relationnelle spécifiquement, mais sur l'impact des choix effectués en matière de visualisation (notamment le type de représentation) sur l'analyse.

Contexte d'évaluation

Lors de formations, il a été demandé à des étudiants de Master⁷⁶ de décrire leurs constatations et de formuler des hypothèses sur la base d'une représentation graphique. Les données représentées proviennent d'un échantillon de 5914 spams liés à la contrefaçon horlogère (courriels collectés pendant le mois de janvier 2008). Les données sont issues d'une banque de données accessible en ligne (<http://untroubled.org/spam/>, dernier accès le 28 septembre 2010).

La consigne de l'exercice est la suivante :

Soit, la répartition géographique des adresses IP des envoyeurs de spams:

1. Que constatez-vous ?

Formuler des hypothèses sur l'envoi des emails et décrivez en quoi la représentation graphique ci-dessus vous permet de confirmer ou infirmer vos hypothèses.

2. Quelles autres informations souhaiteriez-vous avoir à disposition pour affiner votre analyse ?

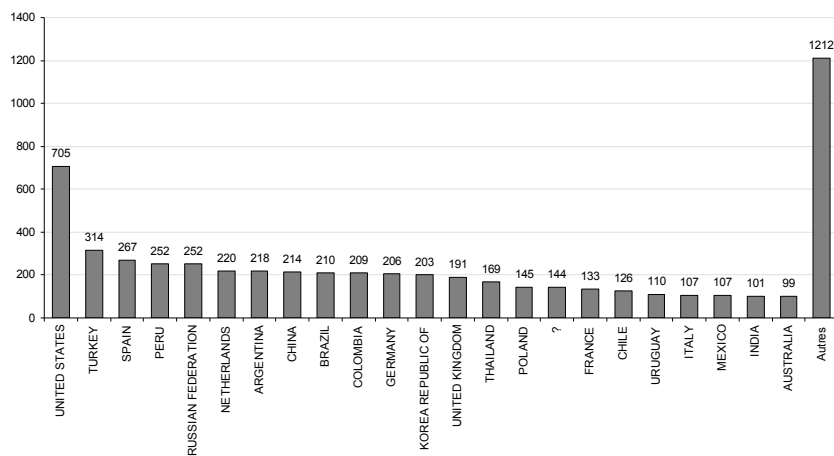
⁷⁶ L'expérience a été réalisée avec 57 étudiants de Master des Universités de Lausanne et d'Aix-Marseille.

Les trois représentations suivantes ont été testées (Les trois graphiques ont été présentés en niveau de gris) :

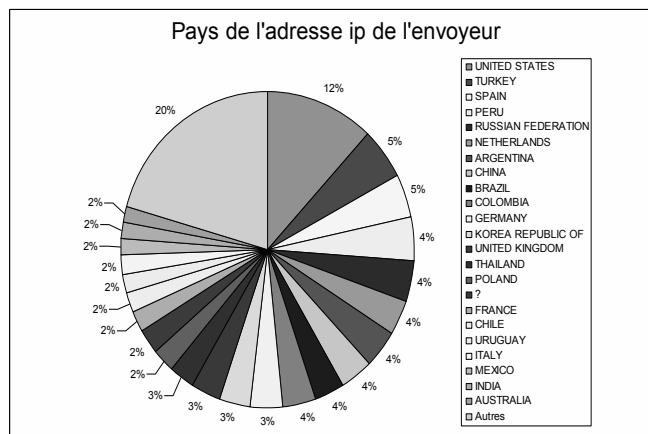
Une carte à symboles proportionnels



Un histogramme



Un diagramme circulaire sectorisé



O b j e c t i f s d e l ' é v a l u a t i o n

L'objectif principal de cette évaluation est de présenter par un exemple, l'impact des choix effectués en termes de représentation de l'information sur l'analyse de données. La méthodologie d'évaluation consiste à poser une question ouverte pour obtenir des résultats en termes de compréhension⁷⁷. Chacune des représentations propose un point de vue différent sur les données. La carte permet d'analyser la distribution spatiale dans son ensemble. L'histogramme permet une comparaison de la prévalence par pays. Le diagramme circulaire permet de comparer la proportion d'envois pour chaque pays en regard du nombre total (pourcentage).

L'évaluation comparative de l'efficacité des représentations ne constitue pas un objectif de cette expérience. En effet, l'évaluation de l'efficacité nécessite la mise en place d'une procédure expérimentale stricte basée sur des critères précis (exactitude, complétude et temps de réponse). En outre, chaque représentation a ses propres limitations et des informations sont volontairement non représentées. Sur la carte, l'échelle est absente et la variable quantitative représentée n'est pas décrite (en l'occurrence le nombre d'emails envoyés). Sur les diagrammes, l'ensemble des pays dont le nombre d'emails envoyés est inférieur à 2% du total sont regroupés. Les adresses ip dont il n'a pas été possible de déterminer le pays sont regroupées dans la catégorie « ? ». Elles ne sont pas représentées sur la carte.

⁷⁷ Cf. Procédures d'évaluation, page 64

R é s u l t a t s

Le tableau ci-dessous résume les constatations effectuées par les étudiants⁷⁸ :

	(nombre d'étudiants)	17	19	21	57
		Carte	Histogramme	Circulaire	Total
Observations effectuées					
Niveau de lecture global		5	14	10	29
La répartition géographique est très grande		4 (23.5)	2 (10.5)	5 (23.8)	11 (19.3)
Plusieurs continents			1 (5.3)	3 (14.3)	4 (7.0)
La distribution est relativement uniforme/homogène			5 (26.3)		5 (8.8)
Certaine régularité pour les autres pays (hors USA)			6 (31.6)		6 (10.5)
Certaine régularité pour les autres pays (hors USA et Turquie)				1 (4.8)	1 (1.8)
La répartition géographique n'est pas uniforme		1 (5.9)			1 (1.8)
Trop de pays représentés				1 (4.8)	1 (1.8)
Niveau de lecture intermédiaire		47	5	3	55
Beaucoup d'envois depuis l'Europe		16 (94.1)			16 (28.1)
Peu d'envoi depuis l'Afrique		13 (76.5)	2 (10.5)		15 (26.3)
Beaucoup d'envois depuis l'Amérique du Sud		9 (52.9)			9 (15.8)
Beaucoup d'envois depuis l'Asie		6 (35.3)			6 (10.5)
7 pays d'Amérique Latine et 7 pays européens			1 (5.3)		1 (1.8)
Envois principalement depuis les pays industrialisés		3 (17.6)	2 (10.5)		5 (8.8)
Envois principalement de l'hémisphère nord				1 (4.8)	1 (1.8)
Principalement de pays où la contrefaçon est en vente "libre"				1 (4.8)	1 (1.8)
Une grande partie des mails provient de 8 pays				1 (4.8)	1 (1.8)
Niveau de lecture élémentaire		16	38	18	72
Beaucoup d'envois depuis les Etats-Unis		13 (76.5)	17 (89.5)	5 (23.8)	35 (61.4)
Une grande partie des mails provient de pays inconnus			11 (57.9)	3 (14.3)	14 (24.6)
Une colonne à un label "?" (absence d'information)			6 (31.6)	2 (9.5)	8 (14.0)
Peu d'envoi depuis la Chine		1 (5.9)	2 (10.5)		3 (5.3)
Peu d'envoi depuis L'Australie			2 (10.5)		2 (3.5)
Peu d'envoi depuis Le Canada		1 (5.9)			1 (1.8)
Peu d'envoi depuis La Russie		1 (5.9)			1 (1.8)
Les spams proviennent principalement de 2 pays (20% et 12%)				3 (14.3)	3 (5.3)
Beaucoup d'envois depuis Chine (sans certitude)				1 (4.8)	1 (1.8)
Beaucoup d'envois depuis Chine (produit beaucoup de contrefaçon)				1 (4.8)	1 (1.8)
20% des envois proviennent d'un pays, peut-être la chine				1 (4.8)	1 (1.8)
20% des envois proviennent d'un pays, on ne peut pas dire lequel				1 (4.8)	1 (1.8)
La Chine a la plus grande proportion de spams, suivi des USA				1 (4.8)	1 (1.8)
Total		68	57	31	156

⁷⁸ Les constatations dont le sens est jugé similaire sont regroupées, même si la formulation varie suivant les étudiants.

De ces résultats, il ressort que :

1. Le nombre total de constatations est deux fois moins important lorsque le graphique circulaire est présenté.

Le diagramme ayant été présenté en niveau de gris, un nombre important d'étudiants ne s'est pas exprimé car ils n'arrivaient pas à faire le lien entre le graphique et la légende (remarque écrite par les étudiants). Seul un nombre restreint d'étudiants s'est exprimé, en observant que les catégories étaient triées (sur le graphique : dans le sens des aiguilles de la montre depuis midi et sur la légende : de haut en bas).

2. Au niveau de lecture global, 58% des étudiants, ayant reçu l'histogramme, constatent que la distribution est relativement uniforme.

Cette observation est probablement engendrée par l'agrégation de l'ensemble des pays dont le taux de prévalence est inférieur à 2%. Dix-huit étudiants (13 pour l'histogramme et 5 pour le diagramme circulaire) ont déclaré souhaiter connaître les détails de la catégorie « Autres ».

3. Les observations au niveau intermédiaire de lecture, ont été principalement formulées par des étudiants ayant reçu la carte.

Dans le cadre de cette analyse, le niveau intermédiaire se définit par les observations portant sur un groupe de pays. Sur les diagrammes, les pays sont triés en fonction du nombre de courriels envoyés. Un classement par zone géographique (par exemple par continent), aurait pu faciliter une telle lecture du schéma (trois étudiants ayant reçu l'histogramme ont fait cette observation).

4. Sur les diagrammes, des observations sont relevées sur la forme et non sur le contenu.

Le fait que des pays sont regroupés dans une catégorie unique est relevé sur les diagrammes. Ce choix de regroupement est inapproprié pour dix-huit étudiants qui ont écrit avoir besoin de ces informations pour s'exprimer. L'attention se focalise donc sur un choix de représentation. Huit étudiants ont également mis en avant la présence de la catégorie « ? » et un étudiant déclare que trop de pays sont représentés.

5. Le diagramme circulaire a engendré des observations formulées sur la base d'un *a priori*.

La catégorie « Autres » est représentée par un niveau de gris similaire à la catégorie « Chine ». Huit étudiants font l'observation que cette catégorie (dont la proportion est la plus grande) est un pays. Quatre d'entre eux émettent l'hypothèse qu'il s'agit de la Chine. Des commentaires écrits par ces étudiants, il ressort que cette hypothèse vient du fait que la Chine est selon eux le principal producteur de contrefaçons.

D i s c u s s i o n

Globalement, les étudiants ont relevé de manière très pertinente que chacun de ces graphiques offre des possibilités d'analyse limitées. Sur la carte, il a été relevé qu'il manque une indication sur la variable quantitative représentée (prévalence, ratio, etc.). La couleur a été très largement plébiscitée par les lecteurs du diagramme circulaire, dont la plupart n'a pas souhaité s'exprimer sur les données représentées. Sur l'histogramme, l'agrégation des données en une catégorie « Autres » a été relevée et jugée problématique pour l'analyse. Un nombre important d'étudiants a également souligné la nécessité de pondérer les résultats en fonction soit de la population des pays, soit du nombre d'ordinateurs disposant d'une connexion Internet.

De cette expérimentation, il ressort que les choix effectués en matière de représentation de l'information influencent fortement les observations qu'il est possible de faire sur un ensemble de données. Certains choix peuvent également engendrer des erreurs.

Annexe 5 : Evaluation de la représentation de relations

Influence des variables liées aux participants

Niveau de formation

Réponses correctes					Intervalle de confiance à 95% pour la moyenne			
Formation	N	Moyenne	Ecart-type	Erreur standard	Borne inférieure	Borne supérieure	Min	Max
Apprentissage, CFC	55	21	9	1,2	18,7	23,7	3	35
Formation professionnelle (maîtrise, brevet)	46	28	8	1,2	25,4	30,1	4	37
Maturité (gymnasiale ou professionnelle)	9	25	8	2,8	18,3	31,2	6	31
Université, EPF, Haute école spécialisée	125	31	4	0,3	30,7	32,0	21	37
Total	235	28	8	0,5	27,0	29,0	3	37

$F(3, 231) = 31.705, p = 3.3 \cdot 10^{-17} \rightarrow$ différence significative, corrélation de Pearson = 0.509

Temps de réponse total (sec)					Intervalle de confiance à 95% pour la moyenne			
Formation	N	Moyenne	Ecart-type	Erreur standard	Borne inférieure	Borne supérieure	Min	Max
Apprentissage, CFC	55	435	274	37	361	509	36	1220
Formation professionnelle (maîtrise, brevet)	46	505	252	37	430	580	66	1334
Maturité (gymnasiale ou professionnelle)	9	631	511	170	238	1024	116	1512
Université, EPF, Haute école spécialisée	125	588	253	23	543	633	161	1721
Total	235	538	277	18	502	573	36	1721

$F(3, 231) = 4.645, p = 0.0036 \rightarrow$ différence significative, corrélation de Pearson = 0.229

Expérience professionnelle

Réponses correctes					Intervalle de confiance à 95% pour la moyenne			
Expérience	N	Moyenne	Ecart-type	Erreur standard	Borne inférieure	Borne supérieure	Min	Max
Etudiants	84	31	4	0,4	30	32	21	37
Employés ESC	33	32	3	0,5	31	33	23	37
Policiers	101	23	9	0,9	22	25	3	36
Les 2	17	3	3	0,7	32	35	27	36
Total	235	28	8	0,5	27	29	3	37

$F(3, 231) = 29.955, p = 2.1 \cdot 10^{-16} \rightarrow$ différence significative

Temps de réponse total (sec)					Intervalle de confiance à 95% pour la moyenne			
Expérience	N	Moyenne	Ecart-type	Erreur standard	Borne inférieure	Borne supérieure	Min	Max
Etudiants	84	517	171	19	479	554	161	981
Employés ESC	33	647	235	41	563	730	214	1292
Policiers	101	498	313	31	436	560	36	1512
Les 2	17	714	423	103	496	931	272	1721
Total	235	541	277	18	505	578	36	1721

$F(3, 231) = 5.088, p = 0.002 \rightarrow$ différence significative

Age

Réponses correctes					Intervalle de confiance à 95% pour la moyenne			
Age	N	Moyenne	Ecart-type	Erreur standard	Borne inférieure	Borne supérieure	Min	Max
16-25	88	30	5	0,5	29	33	12	37
26-35	80	28	8	0,9	26	30	3	37
36-45	53	24	9	1,2	21	26	3	36
46-55	15	26	10	2,5	20	31	4	35
56-65	7	27	10	3,9	17	37	7	34
Total	243	28	8	0,5	27	29	3	37

$F(4, 238) = 6.619, p = 4.6 \cdot 10^{-5} \rightarrow$ différence significative, corrélation de Pearson = -0.264

Temps de réponse total (sec)					Intervalle de confiance à 95% pour la moyenne			
Age	N	Moyenne	Ecart-type	Erreur standard	Borne inférieure	Borne supérieure	Min	Max
16-25	88	496	188	20	456	536	94	981
26-35	80	507	234	26	455	559	36	1199
36-45	53	596	374	51	492	699	66	1721
46-55	15	731	416	107	501	962	59	1512
56-65	7	615	269	102	365	864	198	930
Total	243	540	277	18	504	574	36	1721

$F(4, 238) = 3.418, p = 0.010 \rightarrow$ différence significative, corrélation de Pearson = 0.202

Cours d'analyse Criminelle

Réponses correctes					Intervalle de confiance à 95% pour la moyenne			
Cours	N	Moyenne	Ecart-type	Erreur standard	Borne inférieure	Borne supérieure	Min	Max
Je n'ai jamais suivi de cours	114	26	9	0,8	24	27	3	37
J'ai déjà suivi un cours théorique	76	29	6	0,7	28	31	6	37
J'ai déjà suivi un cours théorique et pratique	53	31	5	0,7	29	32	16	36
Total	243	28	8	0,5	26,91	28,88	3	37

$F(2, 240) = 10.618, p = 3.8 \cdot 10^{-5} \rightarrow$ différence significative, corrélation de Pearson = 0.277

Schémas

Réponses correctes					Intervalle de confiance à 95% pour la moyenne			
Schémas	N	Moyenne	Ecart-type	Erreur standard	Borne inférieure	Borne supérieure	Min	Max
Je ne sais pas ce qu'est un schéma relationnel	60	25	10	1,2	22	27	3	37
J'ai déjà vu un schéma relationnel	122	28	7	0,6	27	30	4	37
J'ai déjà fait un ou plusieurs schéma(s) relationnel(s)	61	30	7	0,8	28	32	3	36
Total	243	28	8	0,5	27	29	3	37

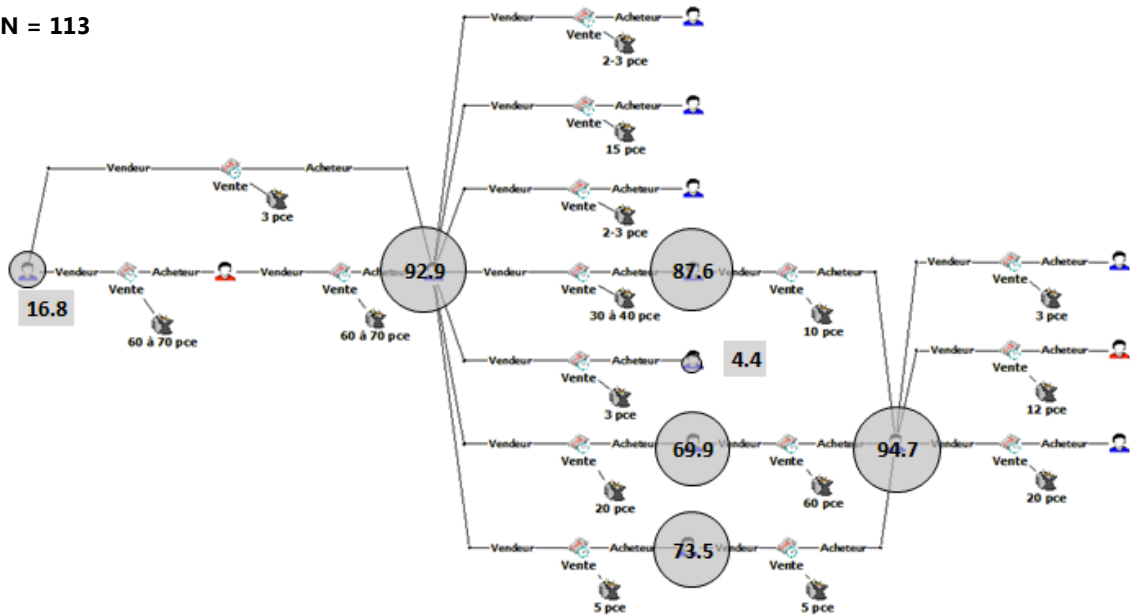
$F(2, 240) = 8.076, p = 4.0 \cdot 10^{-4} \rightarrow$ différence significative, corrélation de Pearson = 0.241

Résultats détaillés

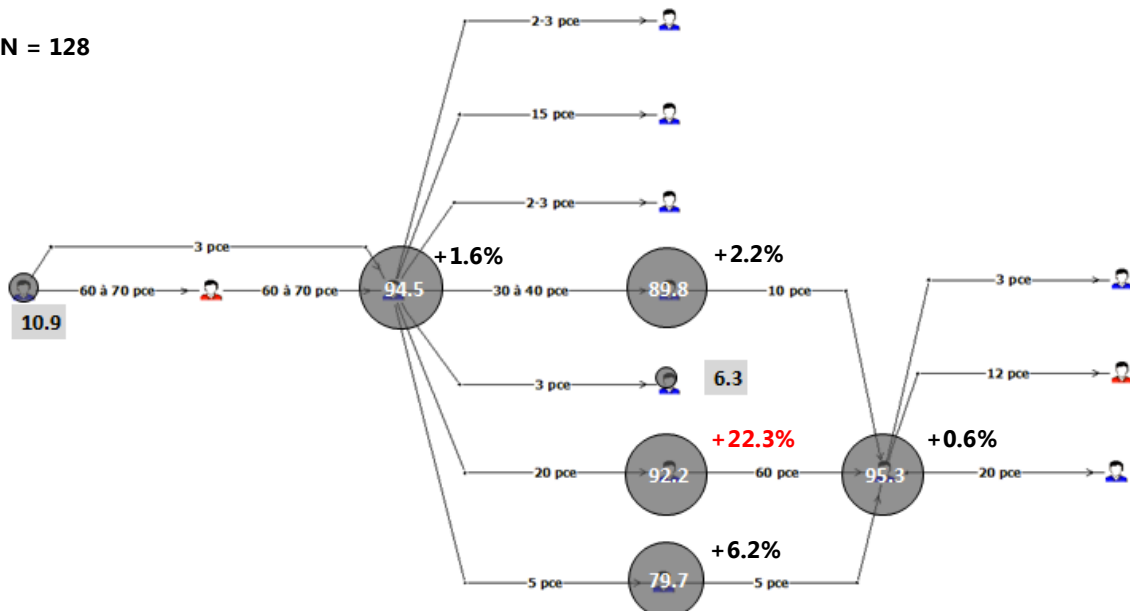
Question 1 : Le schéma ci-dessous représente un trafic. Cliquer sur les personnes qui peuvent avoir été des intermédiaires **entre les deux personnes en rouge**.

Positions des clics : aires proportionnelles au pourcentage de participants

N = 113

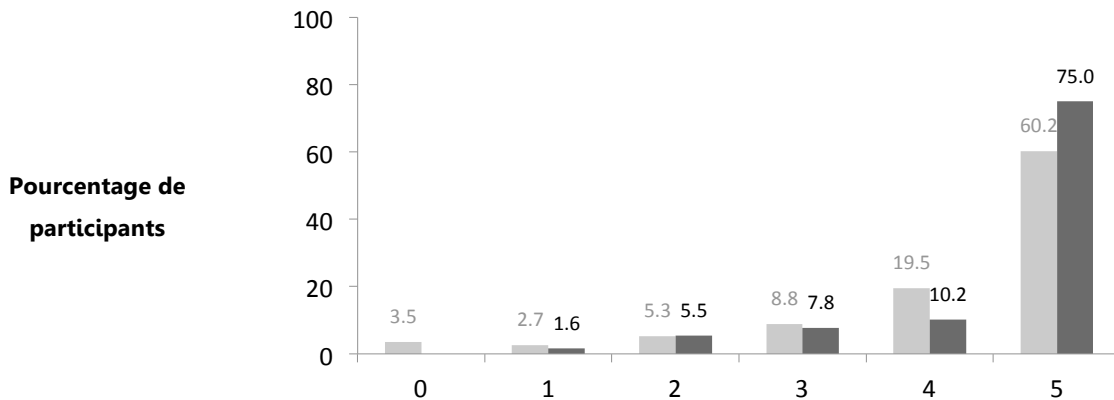


N = 128

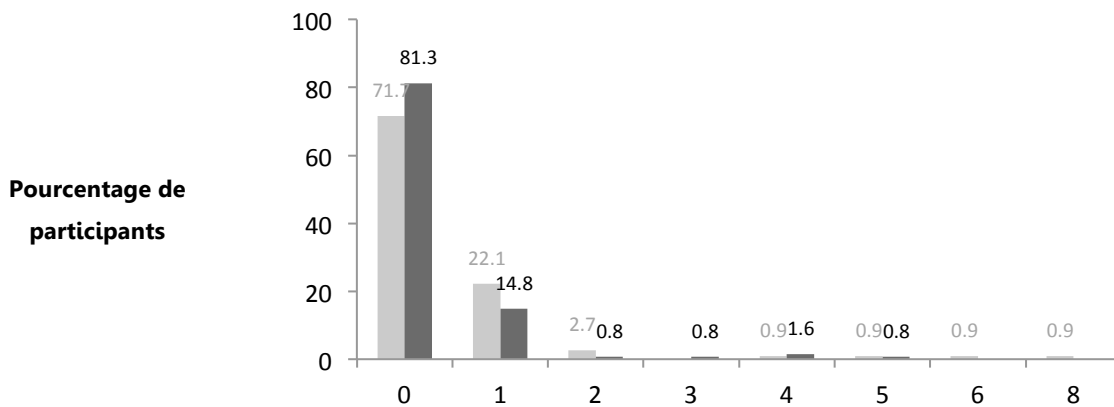


L'ensemble des réponses incorrectes ne sont pas représentées sur les schémas, à l'exception des deux erreurs faites par plus de 2% des participants (identiques sur les deux versions du schéma).

Nombre de réponses correctes



Nombre de réponses fausses



Tests statistiques sur l'exactitude et la complétude

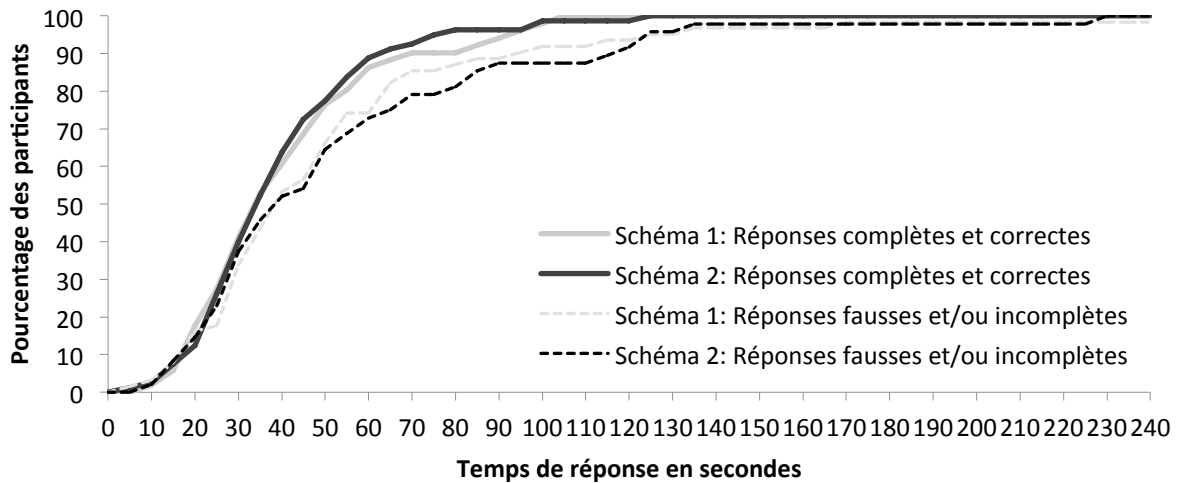
Réponses correctes					Intervalle de confiance à 95% pour la moyenne			
Schéma	N	Moyenne	Ecart-type	Erreur standard	Borne inférieure	Borne supérieure	Min	Max
1	113	4,2	1,2	0,1	3,9	4,4	0	5
2	128	4,5	1,0	0,1	4,4	4,7	1	5

$F(1,239) = 5.08, p = 0.025 \rightarrow$ différence significative, corrélation de Pearson = 0.144

Augmentation du taux de réponse complète et correcte de 14.8%

Réponses fausses					Intervalle de confiance à 95% pour la moyenne			
Schéma	N	Moyenne	Ecart-type	Erreur standard	Borne inférieure	Borne supérieure	Min	Max
1	113	0,5	1,2	0,1	0,3	0,7	0	8
2	128	0,3	0,8	0,1	0,2	0,4	0	5

$F(1,239) = 2.219, p = 0.138 \rightarrow$ pas de différence significative



Tests statistiques sur le temps de réponse

Réponses complètes et correctes					Intervalle de confiance à 95% pour la moyenne		Min	Max
Schéma	N	Moyenne	Ecart-type	Erreur standard	Borne inférieure	Borne supérieure		
1	51	44	22	3	38	51	14	108
2	80	42	21	2	38	47	9	126

$F(1, 129) = 0.212, p = 0.646 \rightarrow$ pas de différence significative

Réponses fausses et/ou incomplètes					Intervalle de confiance à 95% pour la moyenne		Min	Max
Schéma	N	Moyenne	Ecart-type	Erreur standard	Borne inférieure	Borne supérieure		
1	62	54	40	5	44	64	9	250
2	48	56	41	6	44	68	13	234

$F(1, 75) = 0.092, p = 0.762 \rightarrow$ pas de différence significative

Schéma 1 : différence entre les réponses complètes et correctes et les autres :

$F(1, 111) = 2.442, p = 0.121 \rightarrow$ pas de différence significative

Schéma 2 : différence entre les réponses complètes et correctes et les autres :

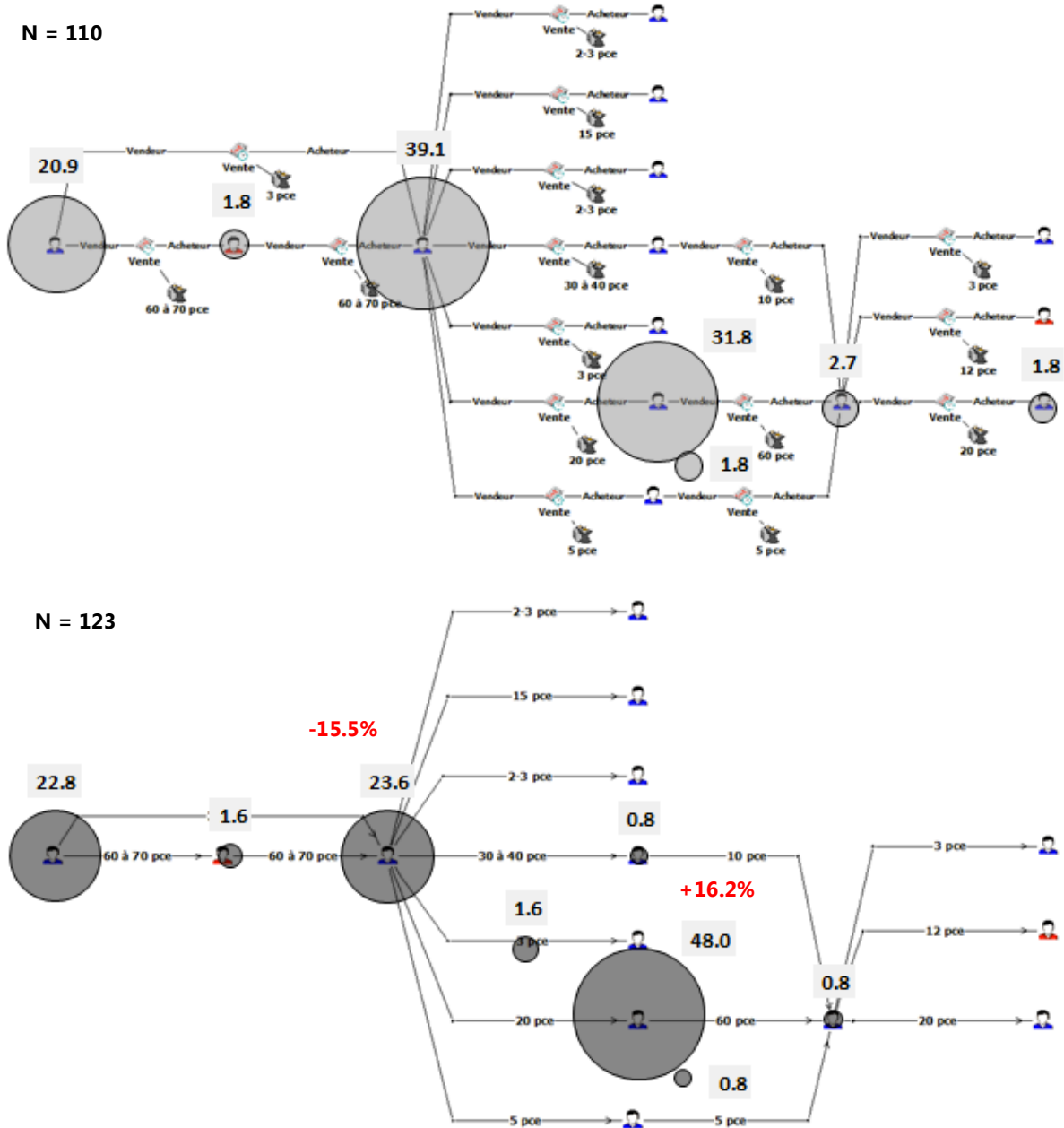
$F(1, 126) = 6.442, p = 0.012 \rightarrow$ différence significative, corrélation de Pearson = -0.221

Temps de réponse moyen 24.7% plus court pour les réponses complètes et correctes

284 Annexe 5 : Evaluation de la représentation de relations

Question 2 : Le schéma ci-dessous représente le même trafic. Cliquer sur la personne qui vous semble avoir vendu bien plus qu'elle n'a acheté.

Positions des clics : aires proportionnelles au pourcentage de participants

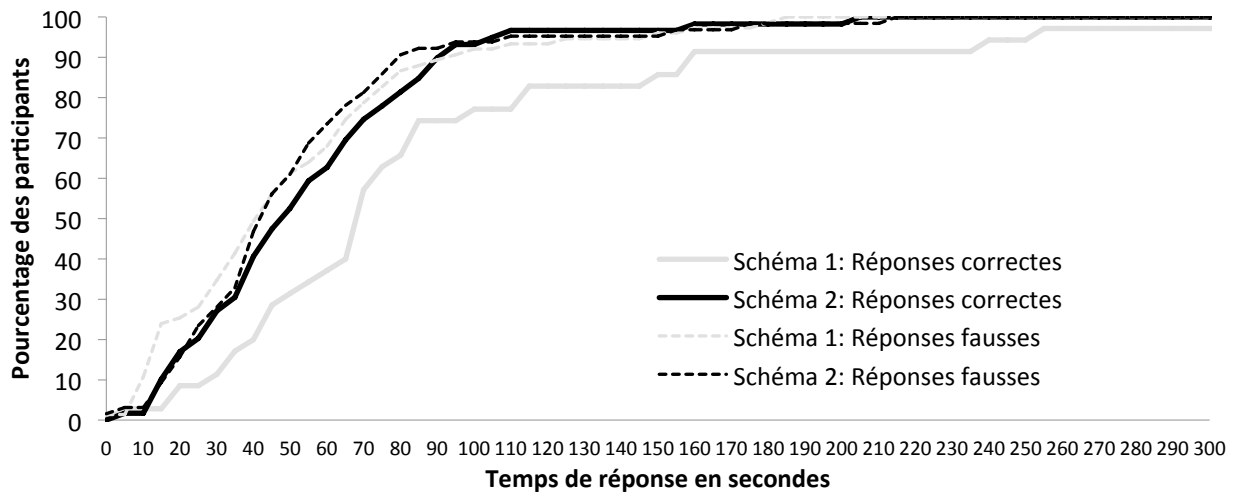


Tests statistiques sur l'exactitude

Réponses correctes					Intervalle de confiance à 95% pour la moyenne			
Schéma	N	Moyenne	Ecart-type	Erreur standard	Borne inférieure	Borne supérieure	Min	Max
1	110	0,32	0,47	0,05	0,23	0,41	0	1
2	123	0,48	0,50	0,05	0,39	0,57	0	1

$F(1,231) = 6.412, p = 0.012 \rightarrow$ différence significative, corrélation de Pearson = 0.164

Augmentation du taux de réponse correcte de 16.2%



Tests statistiques sur le temps de réponse

Réponses correctes					Intervalle de confiance à 95% pour la moyenne		Min	Max
Schéma	N	Moyenne	Ecart-type	Erreur standard	Borne inférieure	Borne supérieure		
1	35	89	67	11	66	112	8	310
2	59	58	35	5	48	67	6	205

$F(1, 92) = 8.947, p = 0.0036 \rightarrow$ différence significative, corrélation de Pearson = -0.298

Temps de réponse correcte moyen 28.2% plus court avec le schéma 2

Réponses fausses					Intervalle de confiance à 95% pour la moyenne		Min	Max
Schéma	N	Moyenne	Ecart-type	Erreur standard	Borne inférieure	Borne supérieure		
1	75	53	39	4,4	44	62	4	187
2	64	54	37	4,6	45	64	4	219

$F(1, 137) = 0.036, p = 0.849 \rightarrow$ pas de différence significative

Schéma 1 : différence entre les réponses complètes et correctes et les autres :

$F(1, 108) = 12.731, p = 0.00054 \rightarrow$ différence significative, corrélation de Pearson = 0.325

Temps de réponse moyen 40.4% plus court pour les réponses incomplètes et/ou contenant des erreurs

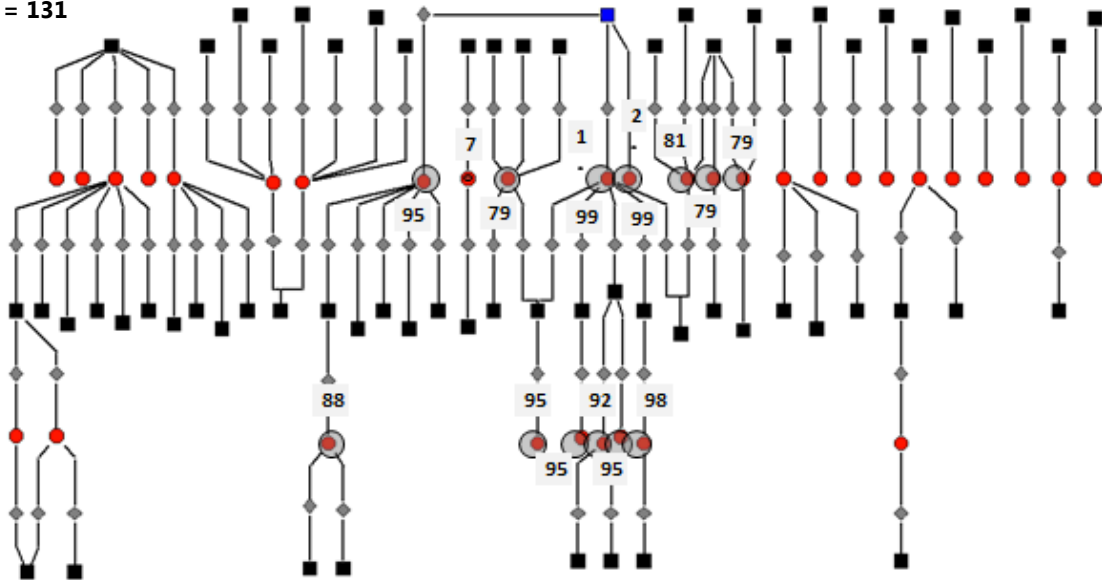
Schéma 2 : différence entre les réponses complètes et correctes et les autres :

$F(1, 121) = 0.246, p = 0.621 \rightarrow$ pas de différence significative

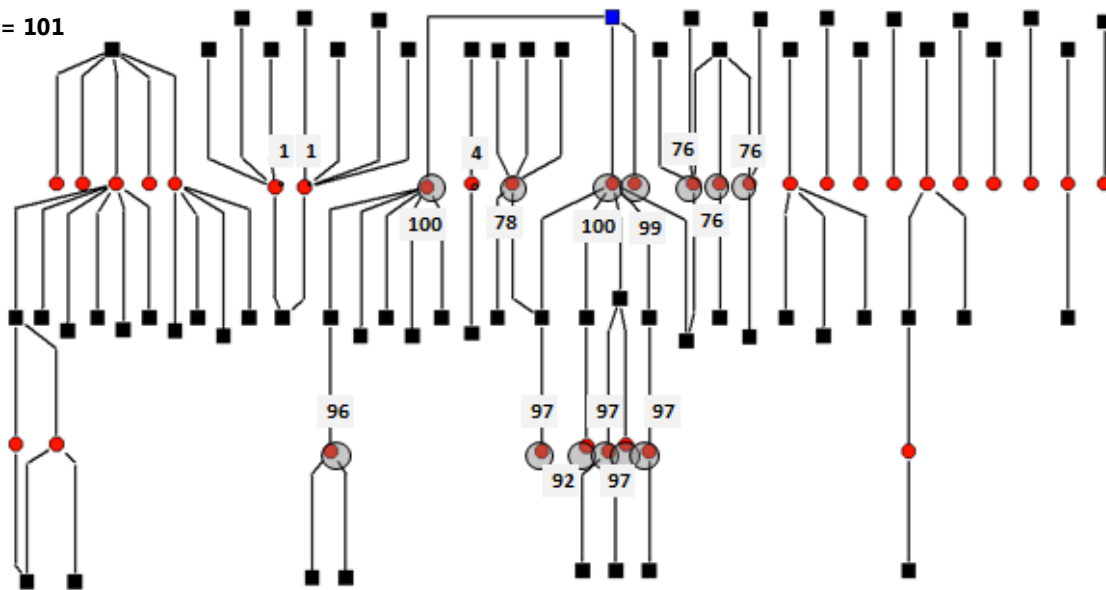
Question 3 : Cliquer sur tous **les ronds rouges** liés directement et indirectement au carré bleu.

Positions des clics : aires proportionnelles au pourcentage de participants

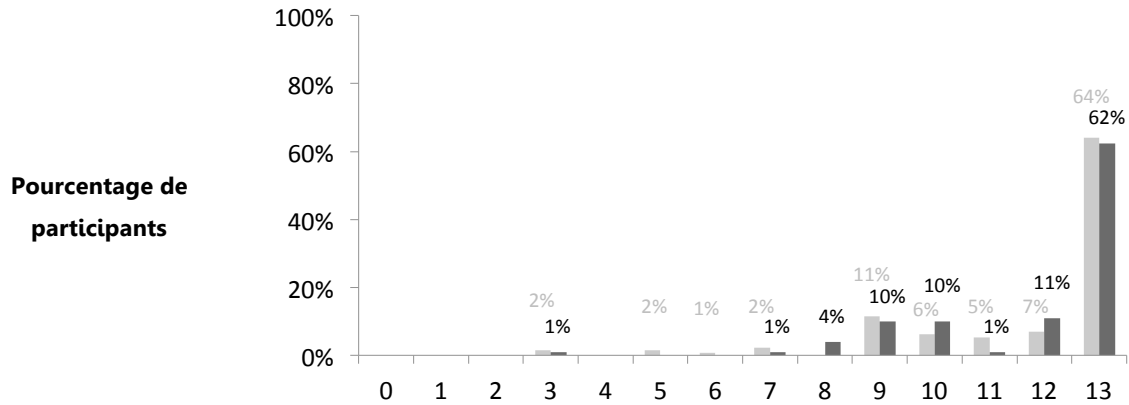
N = 131



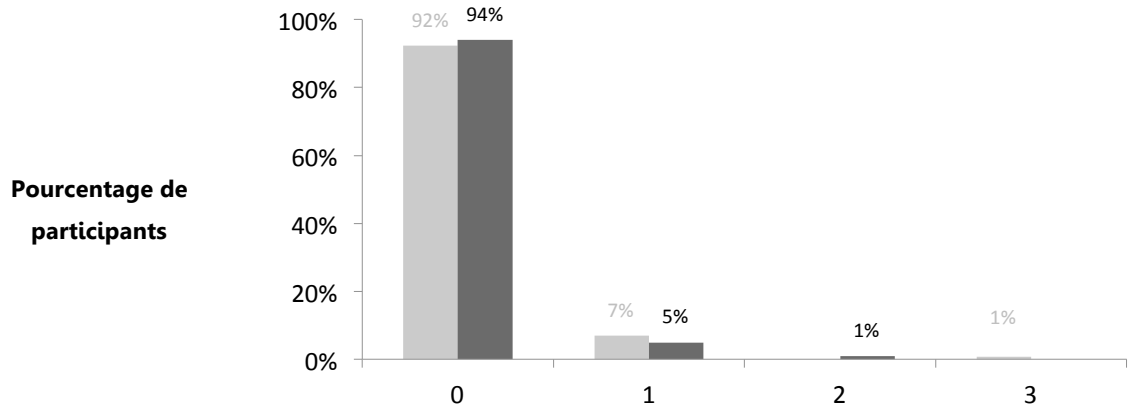
N = 101



Nombre de réponses correctes



Nombre de réponses fausses



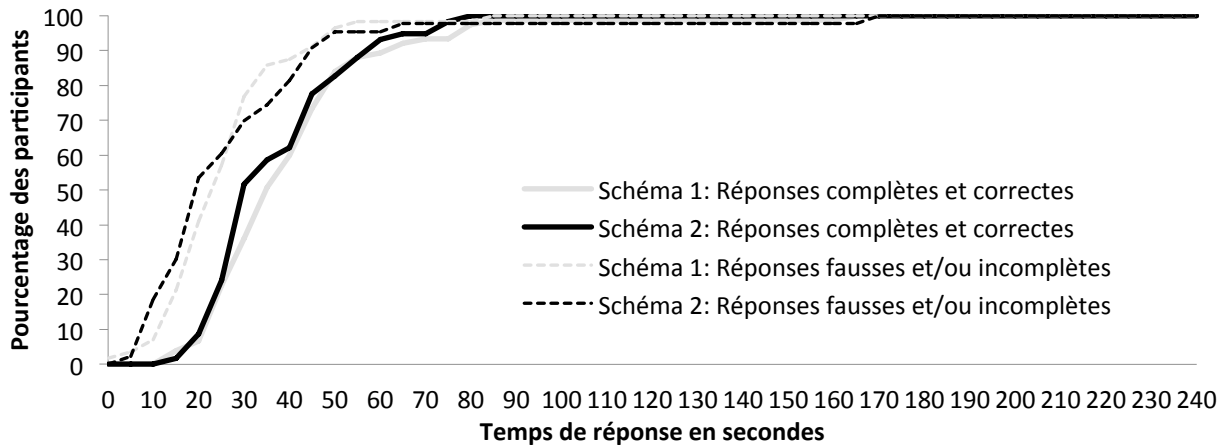
Tests statistiques sur l'exactitude et la complétude

Réponses correctes					Intervalle de confiance à 95% pour la moyenne			
Schéma	N	Moyenne	Ecart-type	Erreur standard	Borne inférieure	Borne supérieure	Min	Max
1	131	12	2	0,2	11	12	3	13
2	101	12	2	0,2	11	12	3	13

$F(1, 230) = 0.145, p = 0.703 \rightarrow$ pas de différence significative

Réponses fausses					Intervalle de confiance à 95% pour la moyenne			
Schéma	N	Moyenne	Ecart-type	Erreur standard	Borne inférieure	Borne supérieure	Min	Max
1	131	0,1	,4	0,03	0,03	0,15	0	3
2	101	0,1	,3	0,03	0,01	0,13	0	2

$F(1, 230) = 0.257, p = 0.613 \rightarrow$ pas de différence significative



Tests statistiques sur le temps de réponse

Réponses complètes et correctes					Intervalle de confiance à 95% pour la moyenne			
Schéma	N	Moyenne	Ecart-type	Erreur standard	Borne inférieure	Borne supérieure	Min	Max
1	75	60	25	2,8	54	66	21	181
2	58	55	21	2,8	49	61	27	153

$F(1, 131) = 1.605, p = 0.849 \rightarrow$ pas de différence significative

Réponses fausses et/ou incomplètes					Intervalle de confiance à 95% pour la moyenne			
Schéma	N	Moyenne	Ecart-type	Erreur standard	Borne inférieure	Borne supérieure	Min	Max
1	56	50	33	4,4	41	59	18	207
2	43	49	35	5,4	38	60	14	189

$F(1, 97) = 0.034, p = 0.855 \rightarrow$ pas de différence significative

Schéma 1 : différence entre les réponses complètes et correctes et les autres :

$F(1, 129) = 4.023, p = 0.047 \rightarrow$ différence significative, corrélation de Pearson = 0.174

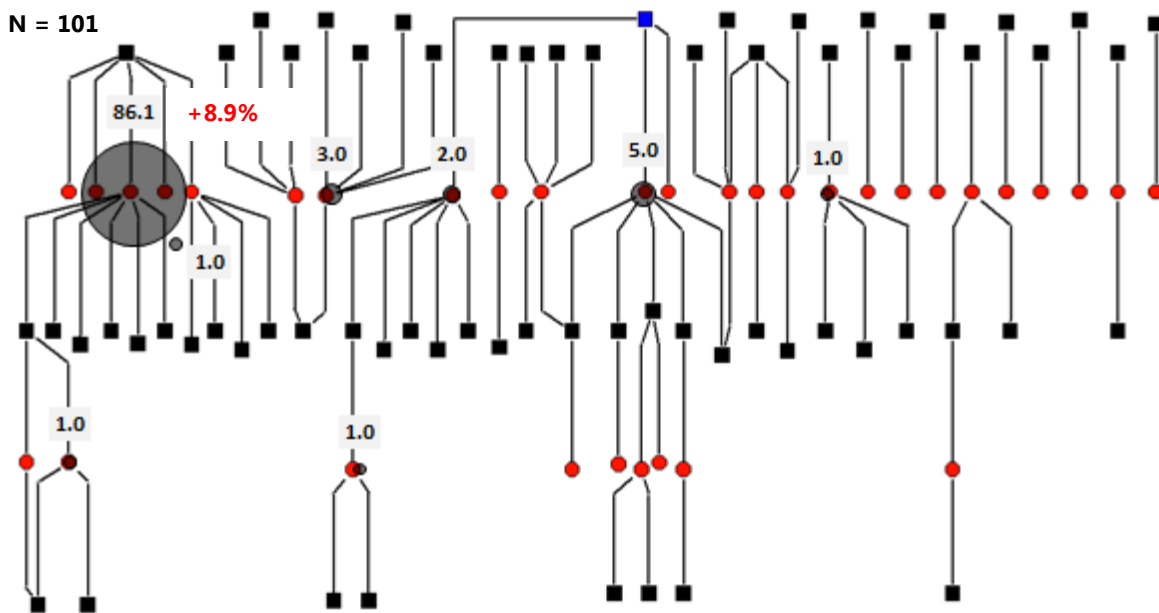
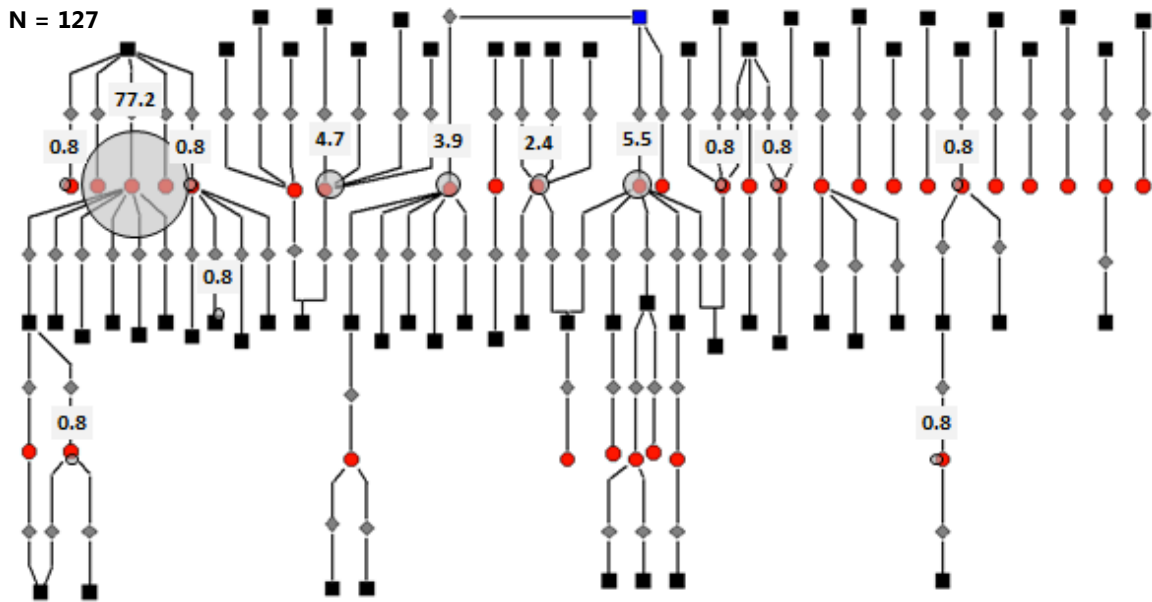
Temps de réponse moyen 16.8% plus court pour les réponses incomplètes et/ou contenant des erreurs

Schéma 2 : différence entre les réponses complètes et correctes et les autres :

$F(1, 99) = 1.212, p = 0.274 \rightarrow$ pas de différence significative

Question 4 : Cliquer sur le **rond rouge** ayant le plus de liens directs avec des carrés noirs.

Positions des clics : aires proportionnelles au pourcentage de participants



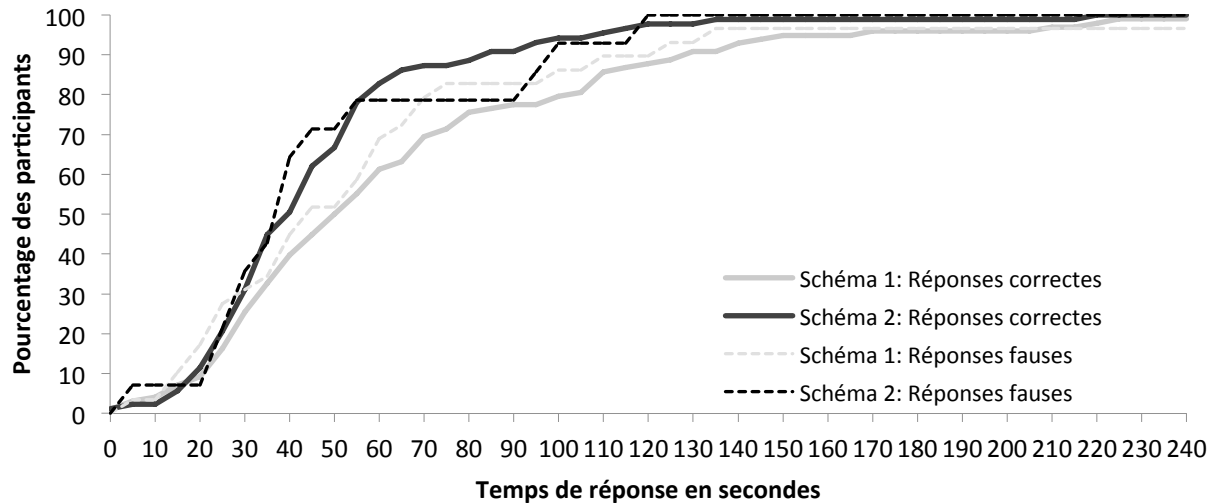
Tests statistiques sur l'exactitude

Réponses correctes					Intervalle de confiance à 95% pour la moyenne		Min	Max
Schéma	N	Moyenne	Ecart-type	Erreur standard	Borne inférieure	Borne supérieure		
1	127	0,77	0,42	0,037	0,70	0,85	0	1
2	101	0,86	0,35	0,035	0,79	0,93	0	1

$F(1,226) = 2.973, p = 0.086 \rightarrow$ effet tendanciel, corrélation de Pearson = 0.114

Augmentation du taux de réponse correcte de 8.9%

290 Annexe 5 : Evaluation de la représentation de relations



Tests statistiques sur le temps de réponse

Réponses correctes					Intervalle de confiance à 95% pour la moyenne			
Schéma	N	Moyenne	Ecart-type	Erreur standard	Borne inférieure	Borne supérieure	Min	Max
1	98	69	51	5,1	58	79	4	298
2	87	50	31	3,4	43	57	4	223

$F(1, 183) = 8.739, p = 0.0035 \rightarrow$ différence significative, corrélation de Pearson = -0.213

Temps de réponse correcte moyen 27.3% plus court avec le schéma 2

Réponses fausses					Intervalle de confiance à 95% pour la moyenne			
Schéma	N	Moyenne	Ecart-type	Erreur standard	Borne inférieure	Borne supérieure	Min	Max
1	29	63	56	10	42	84	9	298
2	14	51	33	9	32	70	6	121

$F(1, 41) = 0.541, p = 0.466 \rightarrow$ pas de différence significative

Schéma 1 : différence entre les réponses correctes et les autres :

$F(1, 125) = 0.291, p = 0.591 \rightarrow$ pas de différence significative

Schéma 2 : différence entre les réponses correctes et les autres :

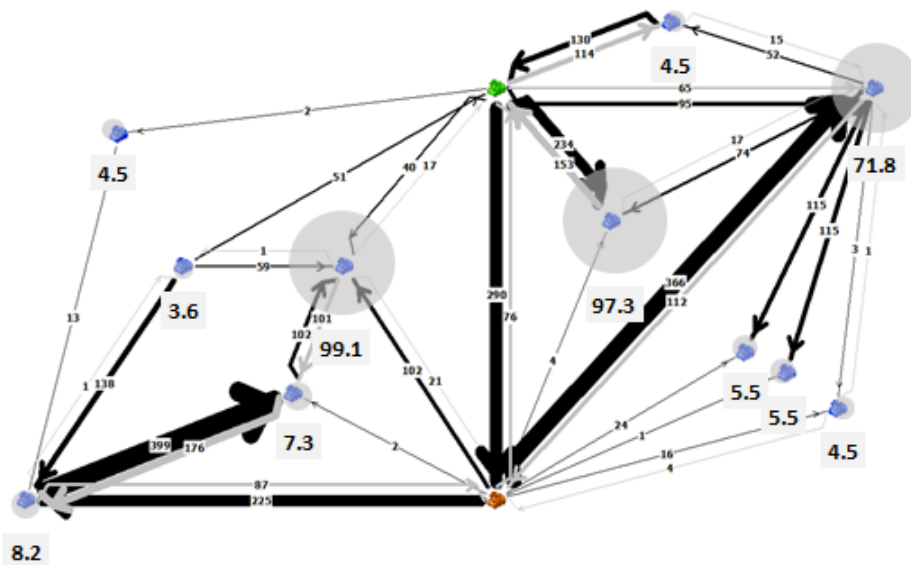
$F(1, 99) = 0.009, p = 0.924 \rightarrow$ pas de différence significative

Le graphe ci-dessous représente un ensemble de communications entre différents téléphones. Les flèches indiquent la direction des communications. Les trois questions qui suivent portent sur ce même schéma.

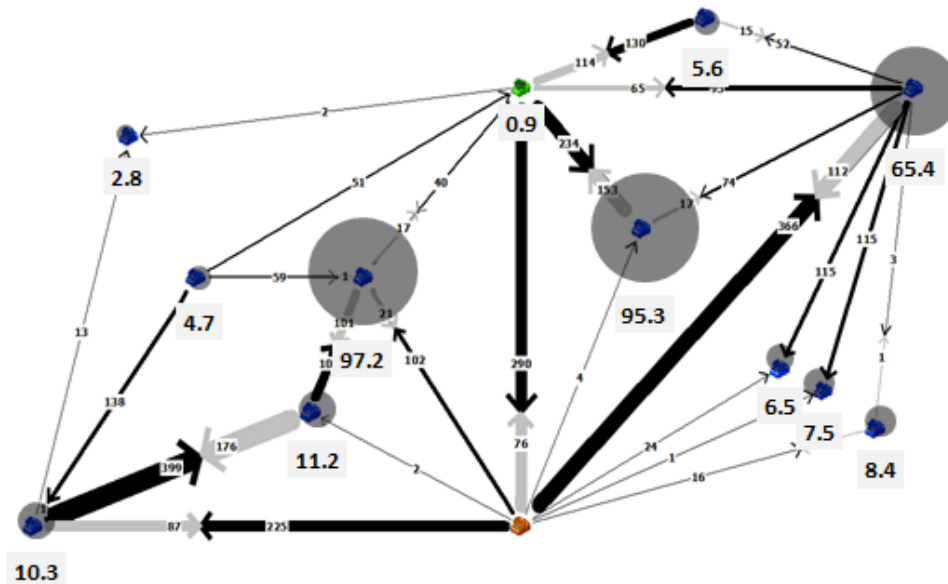
Question 5 : Cliquer sur tous [les téléphones bleus](#) ayant eu des communications avec le [téléphone orange](#) ET le [téléphone vert](#).

Positions des clics : aires proportionnelles au pourcentage de participants

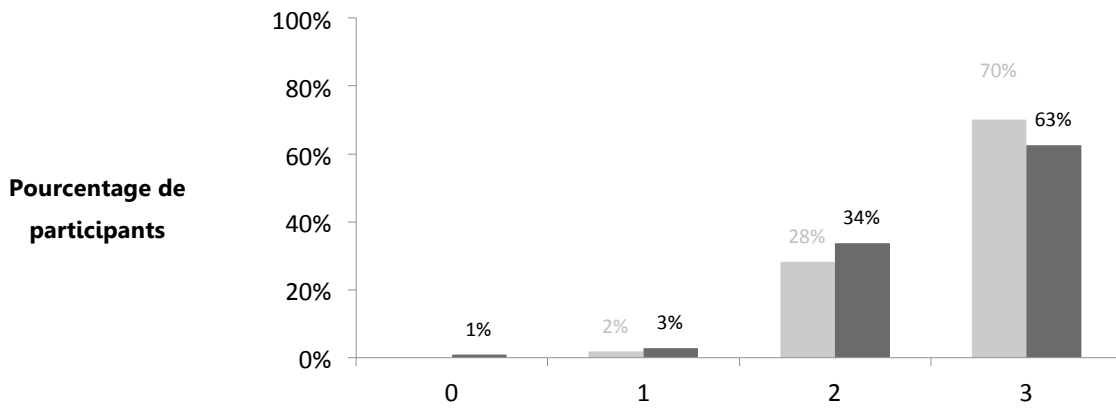
N = 110



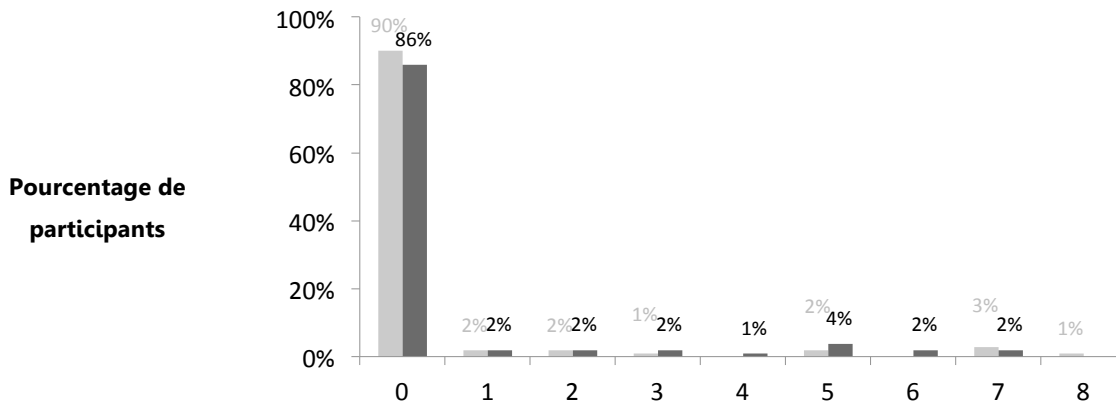
N = 107



Nombre de réponses correctes



Nombre de réponses fausses



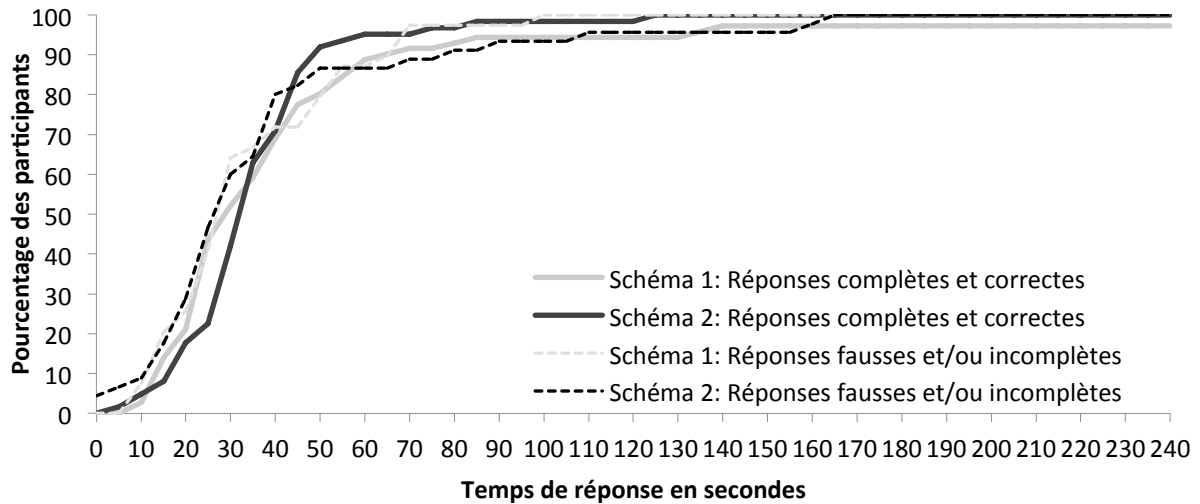
Tests statistiques sur l'exactitude et la complétude

Réponses correctes					Intervalle de confiance à 95% pour la moyenne		Min	Max
Schéma	N	Moyenne	Ecart-type	Erreur standard	Borne inférieure	Borne supérieure		
1	110	2,7	0,5	0,05	2,6	2,8	1	3
2	107	2,6	0,6	0,06	2,5	2,7	0	3

$F(1,215) = 1.854, p = 0.175 \rightarrow$ pas de différence significative

Réponses fausses					Intervalle de confiance à 95% pour la moyenne		Min	Max
Schéma	N	Moyenne	Ecart-type	Erreur standard	Borne inférieure	Borne supérieure		
1	110	0,4	1,5	0,15	0,2	0,7	0	8
2	107	0,6	1,6	0,16	0,3	0,9	0	7

$F(1,215) = 0.444, p = 0.506 \rightarrow$ pas de différence significative



Tests statistiques sur le temps de réponse

Réponses complètes et correctes					Intervalle de confiance à 95% pour la moyenne			
Schéma	N	Moyenne	Ecart-type	Erreur standard	Borne inférieure	Borne supérieure	Min	Max
1	71	66	52	6	53	78	22	367
2	62	59	24	3	52	65	13	141

$F(1, 131) = 0.945, p = 0.333 \rightarrow$ pas de différence significative

Réponses fausses et/ou incomplètes					Intervalle de confiance à 95% pour la moyenne			
Schéma	N	Moyenne	Ecart-type	Erreur standard	Borne inférieure	Borne supérieure	Min	Max
1	39	52	28	4	42	61	14	155
2	45	54	43	6	41	67	3	199

$F(1, 82) = 0.117, p = 0.733 \rightarrow$ pas de différence significative

Schéma 1 : différence entre les réponses complètes et correctes et les autres :

$F(1, 108) = 2.452, p = 0.120 \rightarrow$ pas de différence significative

Schéma 2 : différence entre les réponses complètes et correctes et les autres :

$F(1, 105) = 0.421, p = 0.518 \rightarrow$ pas de différence significative

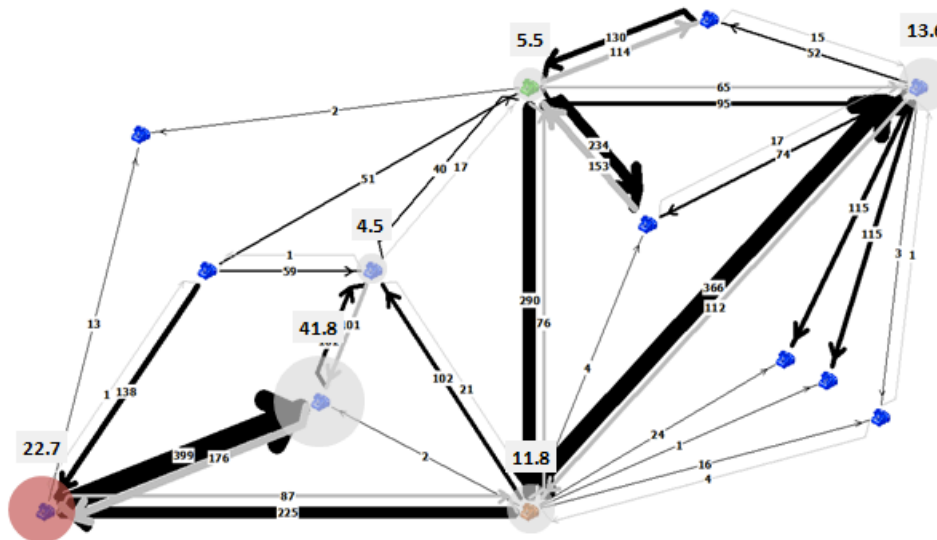
294 Annexe 5 : Evaluation de la représentation de relations

Le graphe ci-dessous représente un ensemble de communications entre différents téléphones. Les flèches indiquent la direction des communications. Les trois questions qui suivent portent sur ce même schéma.

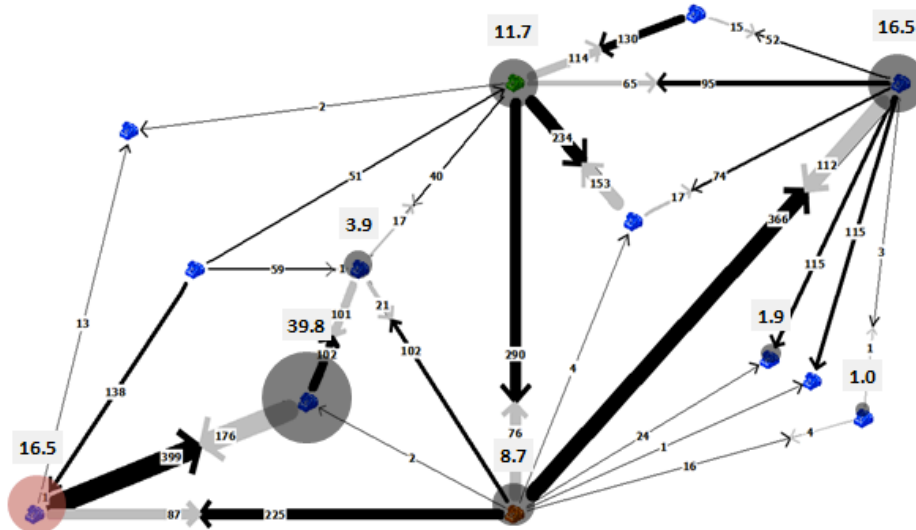
Question 6 : Cliquer sur le téléphone ayant reçu le plus d'appels.

Positions des clics : aires proportionnelles au pourcentage de participants

N = 110



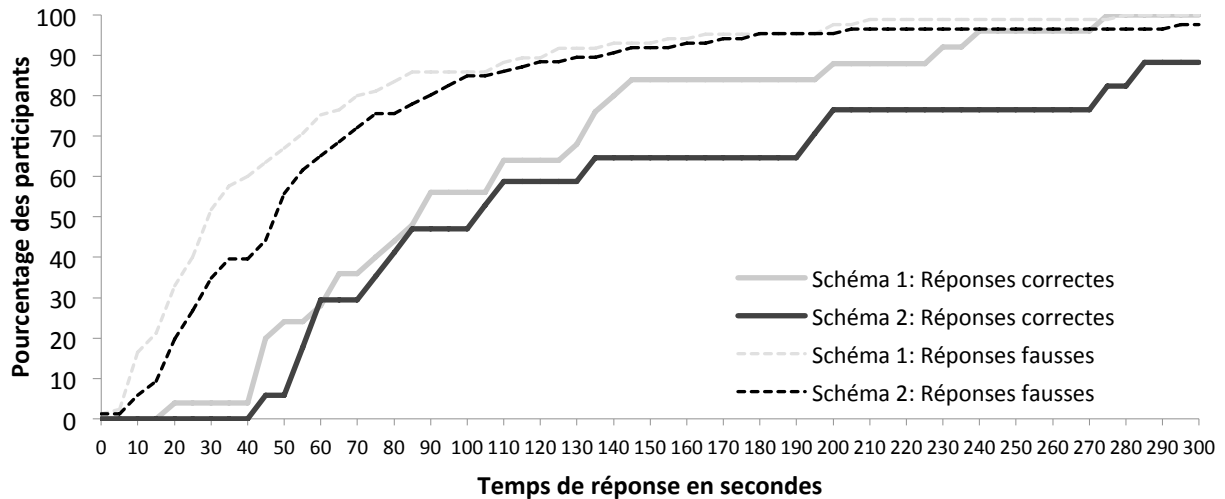
N = 103



Tests statistiques sur l'exactitude

Réponses correctes					Intervalle de confiance à 95% pour la moyenne			
Schéma	N	Moyenne	Ecart-type	Erreur standard	Borne inférieure	Borne supérieure	Min	Max
1	110	0,23	0,42	0,04	0,15	0,31	0	1
2	103	0,17	0,37	0,04	0,09	0,24	0	1

$F(1,211) = 1.297, p = 0.256 \rightarrow$ pas de différence significative



Tests statistiques sur le temps de réponse

Réponses correctes					Intervalle de confiance à 95% pour la moyenne			
Schéma	N	Moyenne	Ecart-type	Erreur standard	Borne inférieure	Borne supérieure	Min	Max
1	25	110	68	14	82	138	23	279
2	17	153	114	28	94	212	47	427

$F(1, 40) = 2.315, p = 0.136 \rightarrow$ pas de différence significative

Réponses fausses					Intervalle de confiance à 95% pour la moyenne			
Schéma	N	Moyenne	Ecart-type	Erreur standard	Borne inférieure	Borne supérieure	Min	Max
1	85	53	52	6	42	65	3	281
2	86	68	62	7	55	81	4	315

$F(1, 169) = 2.791, p = 0.097 \rightarrow$ effet tendanciel, corrélation de Pearson = 0.127

Temps de réponse correcte moyen 21.6% plus court avec le schéma 1

Schéma 1 : différence entre les réponses correctes et les autres :

$F(1, 108) = 19.865, p = 0.00002 \rightarrow$ différence significative, corrélation de Pearson = 0.394

Temps de réponse moyen 51.6% plus court pour les réponses incorrectes

Schéma 2 : différence entre les réponses correctes et les autres :

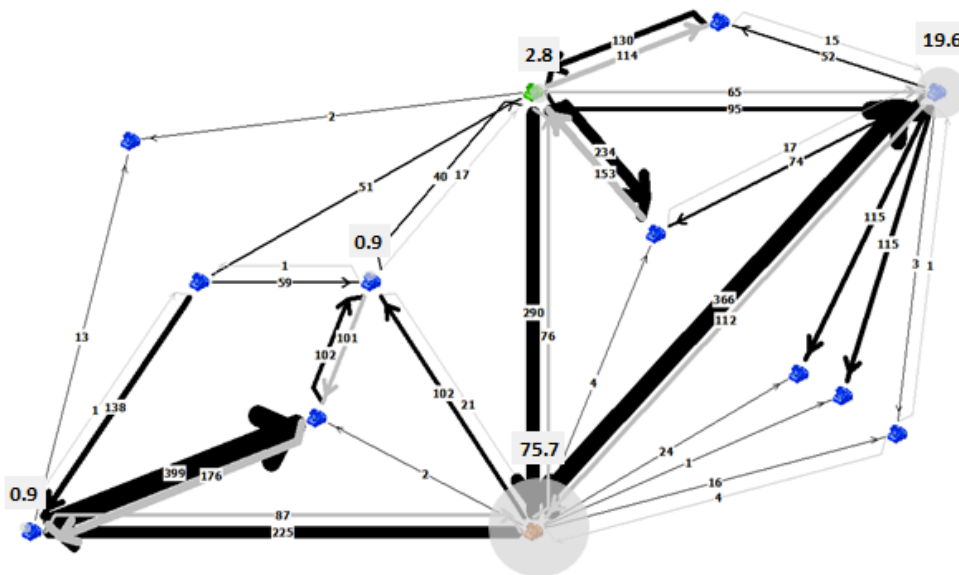
$F(1, 101) = 19.325, p = 0.00003 \rightarrow$ différence significative, corrélation de Pearson = 0.401

Temps de réponse moyen 55.5% plus court pour les réponses incorrectes

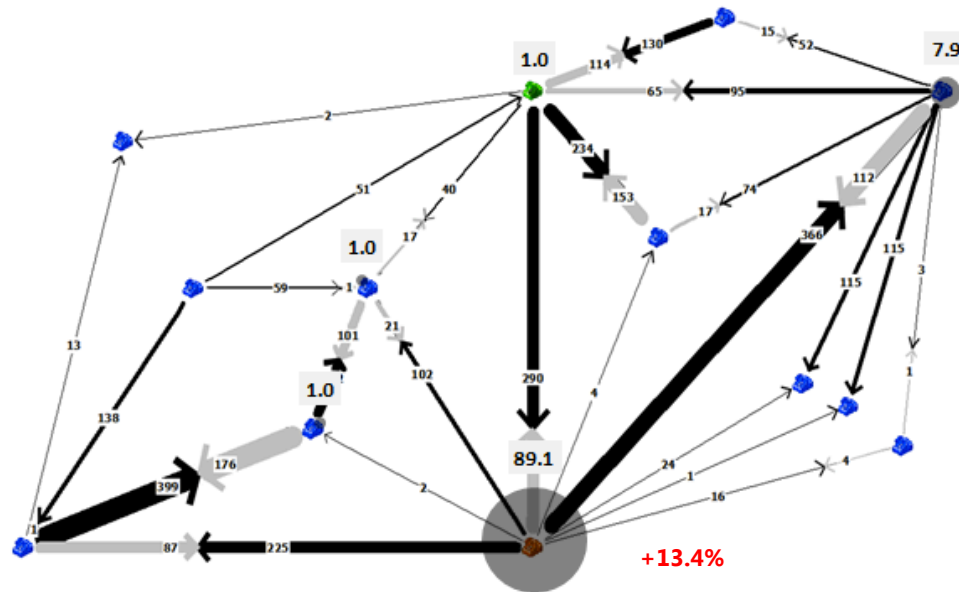
Question 7 : Cliquer sur le téléphone qui est en relation avec le plus grand nombre de téléphones.

Positions des clics : aires proportionnelles au pourcentage de participants

N = 107



N = 101

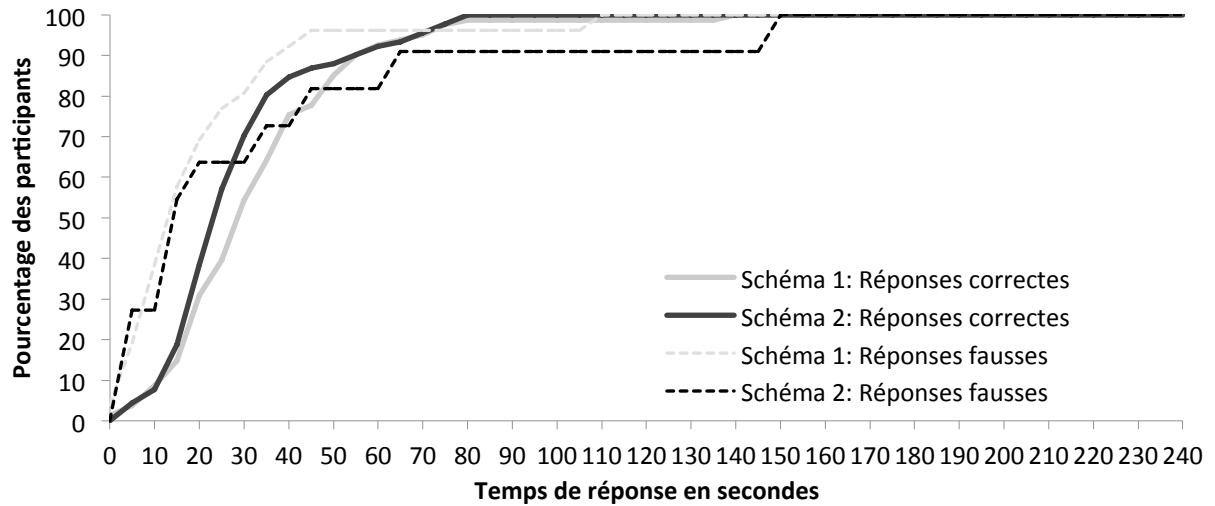


Tests statistiques sur l'exactitude

Réponses correctes					Intervalle de confiance à 95% pour la moyenne		Min	Max
Schéma	N	Moyenne	Ecart-type	Erreur standard	Borne inférieure	Borne supérieure		
1	107	0,76	0,43	0,04	0,67	0,84	0	1
2	101	0,89	0,31	0,03	0,83	0,95	0	1

$F(1,206) = 6.526, p = 0.011 \rightarrow$ différence significative, corrélation de Pearson = 0.175

Augmentation du taux de réponse complète et correcte de 13.4%



Tests statistiques sur le temps de réponse

Réponses correctes					Intervalle de confiance à 95% pour la moyenne			
Schéma	N	Moyenne	Ecart-type	Erreur standard	Borne inférieure	Borne supérieure	Min	Max
1	81	37	20	2	32	41	4	142
2	90	31	16	2	28	35	5	81

$F(1, 169) = 3.674, p = 0.057 \rightarrow$ différence significative, corrélation de Pearson = -0.147

Temps de réponse correcte moyen 14.7% plus court avec le schéma 2

Réponses fausses					Intervalle de confiance à 95% pour la moyenne			
Schéma	N	Moyenne	Ecart-type	Erreur standard	Borne inférieure	Borne supérieure	Min	Max
1	26	23	22	4	14	32	2	114
2	11	36	42	13	7	64	6	151

$F(1, 35) = 1.412, p = 0.243 \rightarrow$ pas de différence significative

Schéma 1 : différence entre les réponses correctes et les autres :

$F(1, 106) = 8.263, p = 0.005 \rightarrow$ différence significative, corrélation de Pearson = 0.270

Temps de réponse moyen 36.7% plus court pour les réponses incorrectes

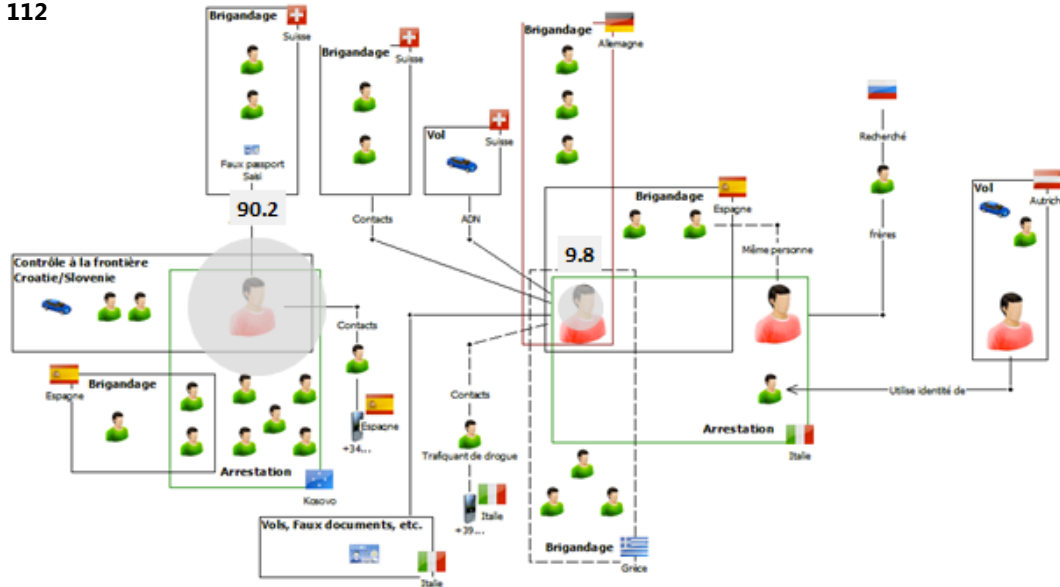
Schéma 2 : différence entre les réponses correctes et les autres :

$F(1, 99) = 0.457, p = 0.501 \rightarrow$ pas de différence significative

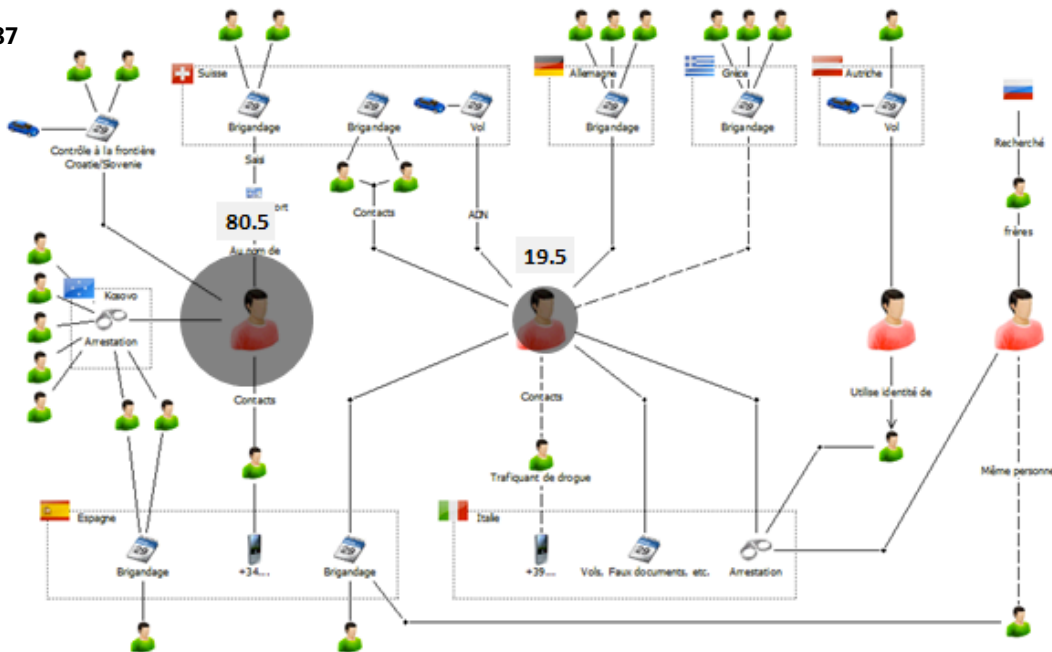
Question 8 : Cliquer sur **le membre du groupe criminel** qui est lié au plus grand nombre d'individus en vert.

Positions des clics : aires proportionnelles au pourcentage de participants

N = 112



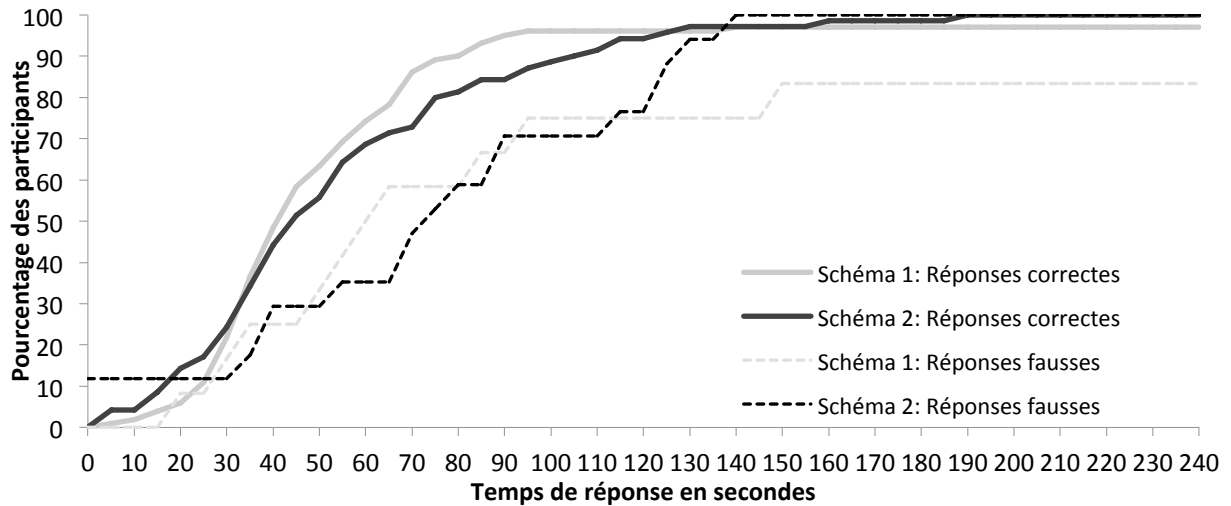
N = 87



Tests statistiques sur l'exactitude

Réponses correctes					Intervalle de confiance à 95% pour la moyenne			
Schéma	N	Moyenne	Ecart-type	Erreur standard	Borne inférieure	Borne supérieure	Min	Max
1	112	0,90	0,30	0,03	0,85	0,96	0	1
2	87	0,80	0,40	0,04	0,72	0,89	0	1

$F(1,197) = 3.861, p = 0.051 \rightarrow$ différence significative, corrélation de Pearson = -0.139



Tests statistiques sur le temps de réponse

Réponses correctes					Intervalle de confiance à 95% pour la moyenne			
Schéma	N	Moyenne	Ecart-type	Erreur standard	Borne inférieure	Borne supérieure	Min	Max
1	101	57	50	5	48	67	5	336
2	70	58	36	4	49	67	8	191

$F(1, 169) = 0.005, p = 0.942 \rightarrow$ pas de différence significative

Réponses fausses					Intervalle de confiance à 95% pour la moyenne			
Schéma	N	Moyenne	Ecart-type	Erreur standard	Borne inférieure	Borne supérieure	Min	Max
1	11	112	99	30	45	179	23	307
2	17	77	43	10	55	99	2	143

$F(1, 26) = 1.648, p = 0.211 \rightarrow$ pas de différence significative

Schéma 1 : différence entre les réponses complètes et correctes et les autres :

$F(1, 110) = 9.297, p = 0.003 \rightarrow$ différence significative, corrélation de Pearson = -0.279

Temps de réponse moyen 48.7% plus court pour les réponses correctes

Schéma 2 : différence entre les réponses complètes et correctes et les autres :

$F(1, 85) = 3.627, p = 0.060 \rightarrow$ effet tendanciel, corrélation de Pearson = -0.202

Temps de réponse moyen 24.9% plus court pour les réponses correctes

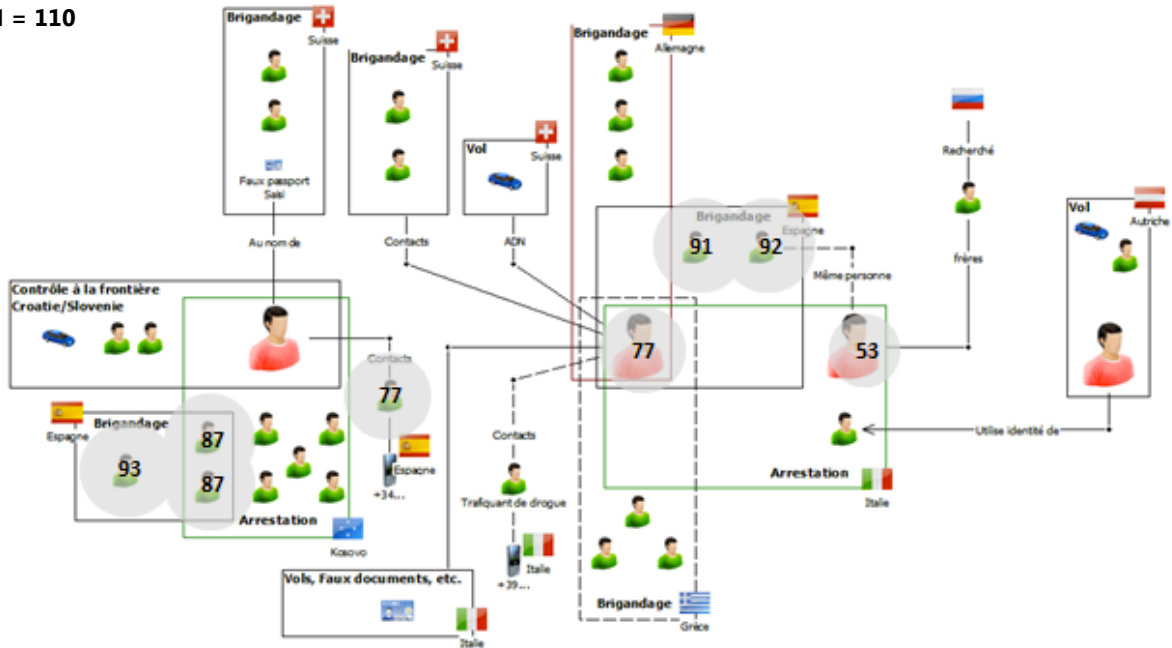
300 Annexe 5 : Evaluation de la représentation de relations

Ce schéma représente les relations d'un groupe criminel constitué des quatre individus en rouge. Il s'agit du même schéma que précédemment.

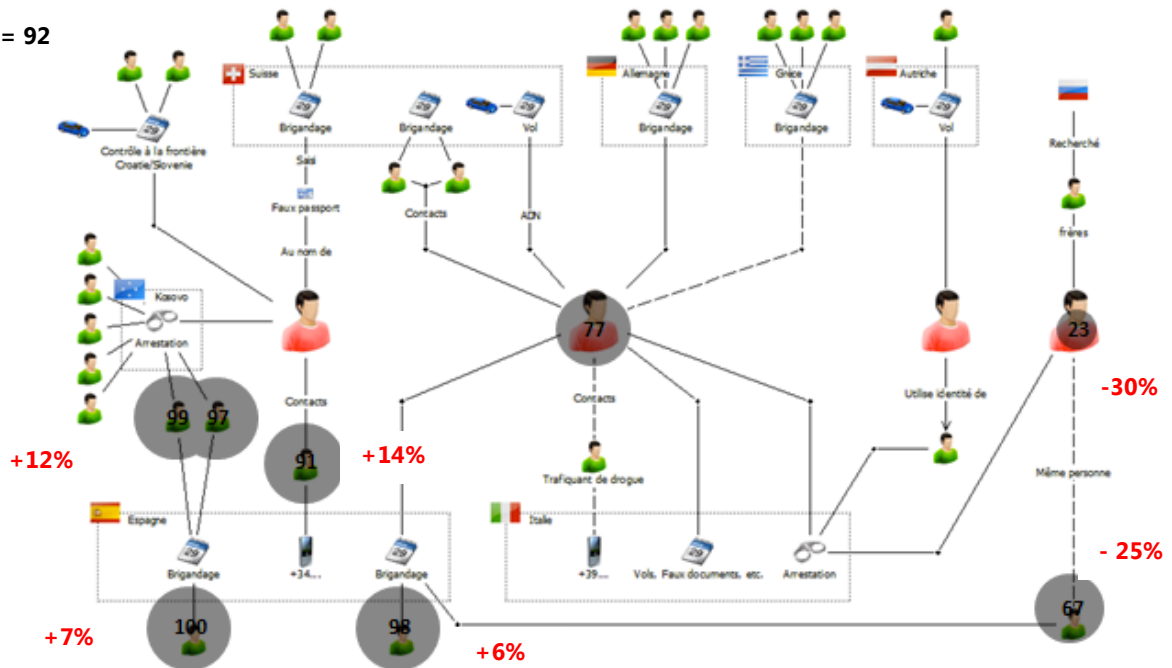
Question 9 : Cliquer sur les individus (vert et rouge) qui semblent avoir eu une activité en Espagne.

Positions des clics : aires proportionnelles au pourcentage de participants

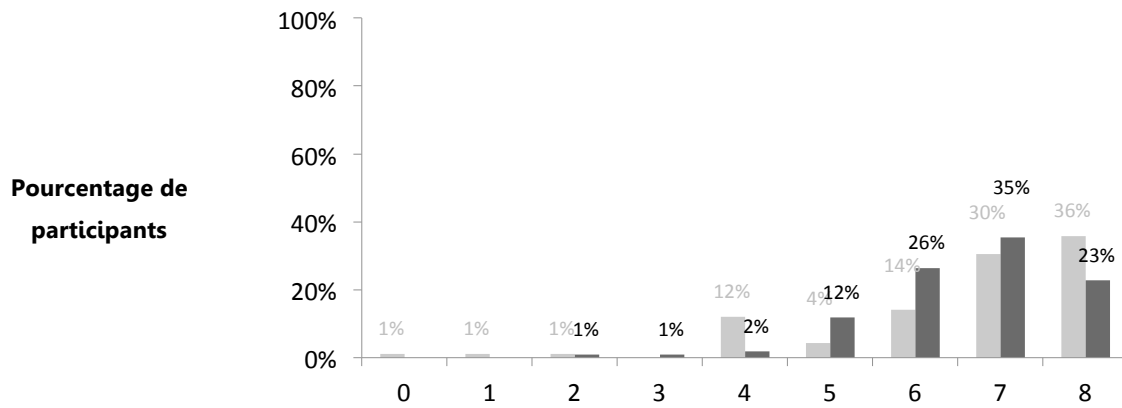
N = 110



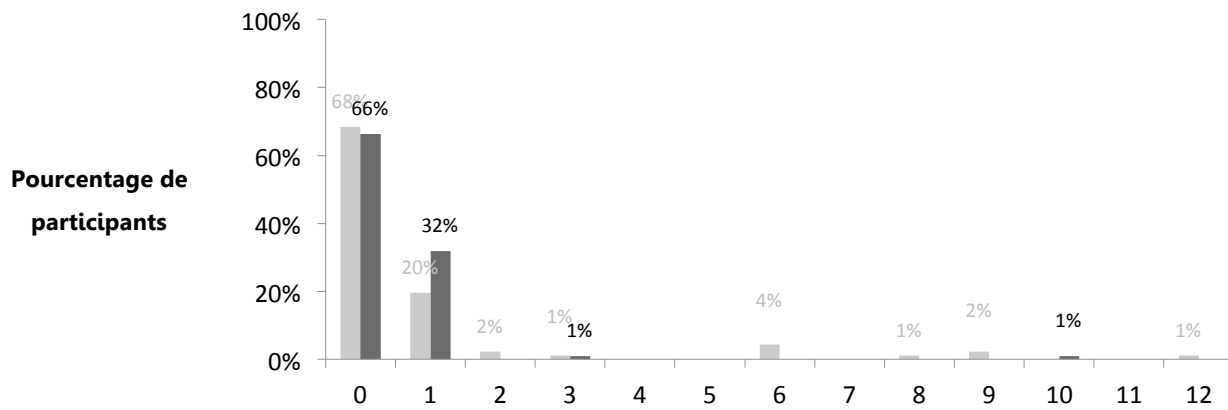
N = 92



Nombre de réponses correctes



Nombre de réponses fausses



Tests statistiques sur l'exactitude et la complétude

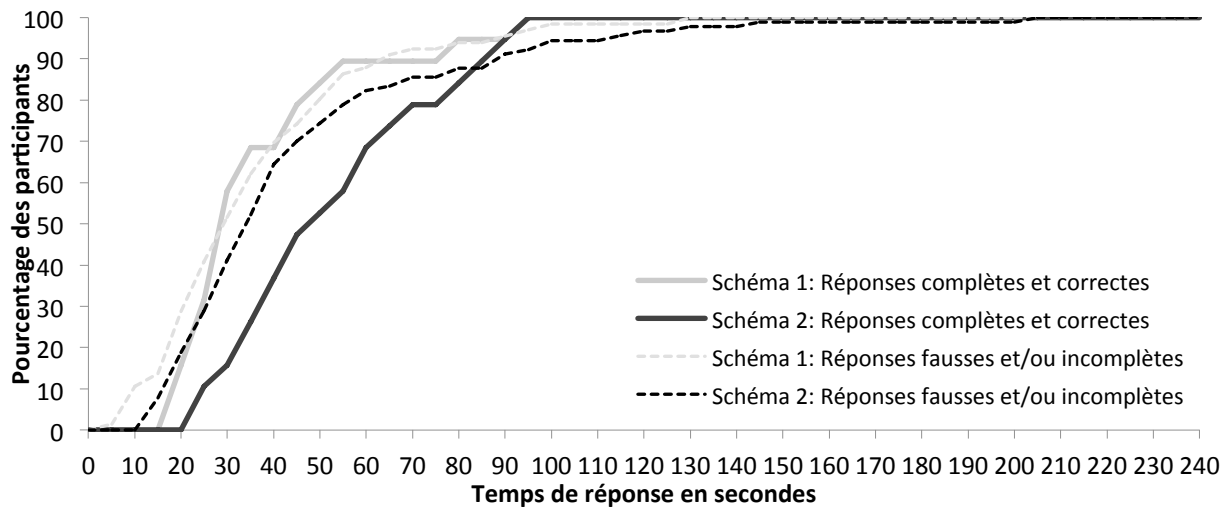
Réponses correctes					Intervalle de confiance à 95% pour la moyenne			
Schéma	N	Moyenne	Ecart-type	Erreur standard	Borne inférieure	Borne supérieure	Min	Max
1	110	6,6	1,2	0,1	6,4	6,8	2	8
2	92	6,6	1,7	0,2	6,2	6,9	0	8

$F(1,200) = 0.000, p = 0.984 \rightarrow$ pas de différence significative

Réponses fausses					Intervalle de confiance à 95% pour la moyenne			
Schéma	N	Moyenne	Ecart-type	Erreur standard	Borne inférieure	Borne supérieure	Min	Max
1	110	0,4	1,1	0,1	0,2	0,6	0	10
2	92	0,9	2,2	0,2	0,5	1,4	0	12

$F(1,200) = 4.214, p = 0.041 \rightarrow$ différence significative, corrélation de Pearson = -0.144

302 Annexe 5 : Evaluation de la représentation de relations



Tests statistiques sur le temps de réponse

Réponses complètes et correctes					Intervalle de confiance à 95% pour la moyenne			
Schéma	N	Moyenne	Ecart-type	Erreur standard	Borne inférieure	Borne supérieure	Min	Max
1	19	40	20	5	30	50	20	96
2	19	56	22	5	45	67	27	98

$F(1, 36) = 5.490, p = 0.025 \rightarrow$ différence significative, corrélation de Pearson = -0.364

Temps de réponse correcte moyen 28.6% plus court avec le schéma 1

Réponses fausses et/ou incomplètes					Intervalle de confiance à 95% pour la moyenne			
Schéma	N	Moyenne	Ecart-type	Erreur standard	Borne inférieure	Borne supérieure	Min	Max
1	73	41	25	3	36	47	7	132
2	91	48	32	3	41	54	15	208

$F(1, 162) = 1.880, p = 0.172 \rightarrow$ pas de différence significative

Schéma 1 : différence entre les réponses complètes et correctes et les autres :

$F(1, 108) = 1.218, p = 0.272 \rightarrow$ pas de différence significative

Schéma 2 : différence entre les réponses complètes et correctes et les autres :

$F(1, 90) = 0.041, p = 0.840 \rightarrow$ pas de différence significative

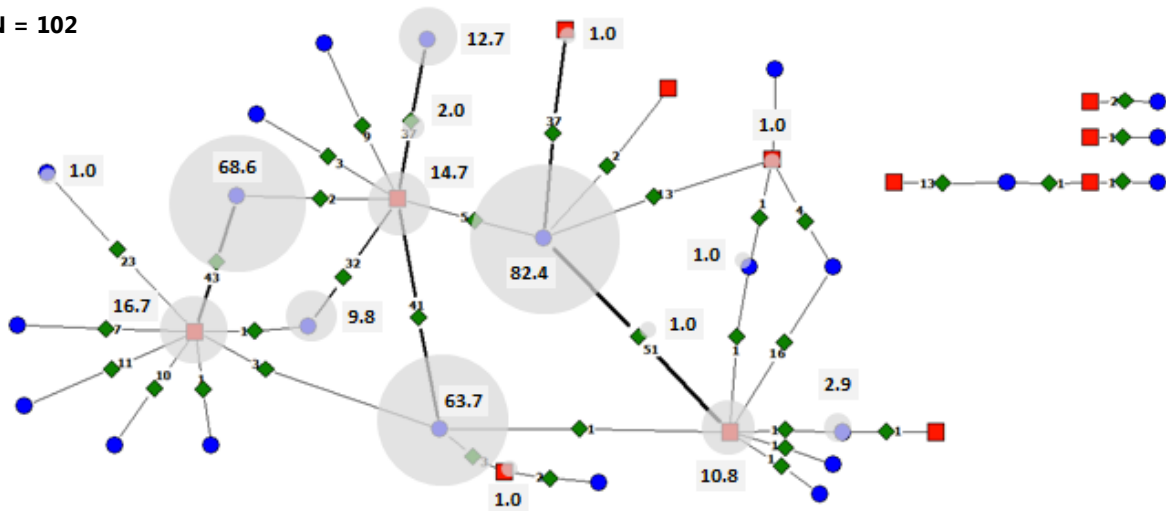
Les carrés rouges représentent **des marques** de photocopieurs.

Les ronds bleus représentent **des groupes** (de photocopieurs qui impriment de la même manière). Pour chaque groupe, le nombre de photocopieurs liés à une marque particulière est indiqué par un chiffre.

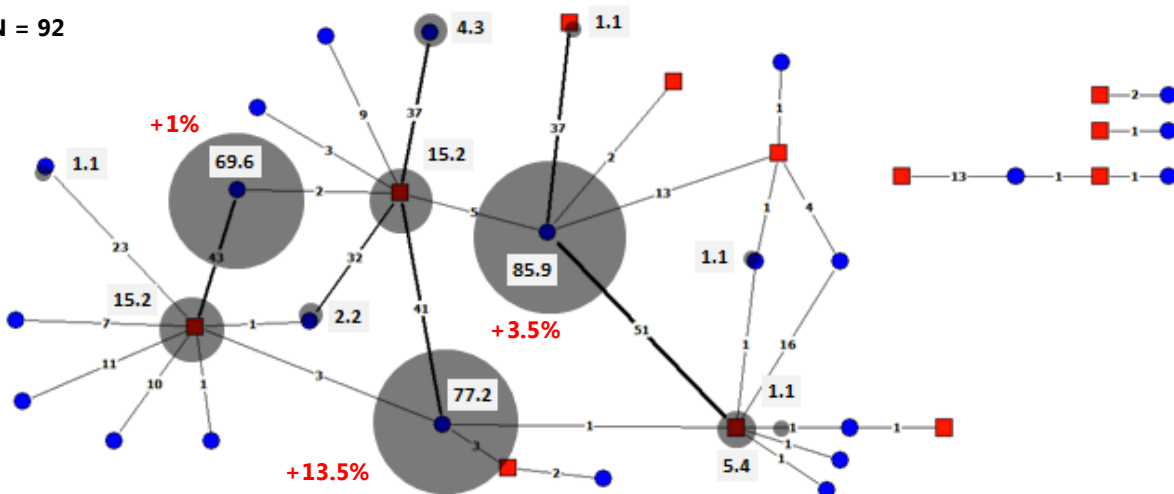
Question 10 : Cliquer [sur les 3 groupes](#) liés avec le plus grand nombre de photocopieurs.

Positions des clics : aires proportionnelles au pourcentage de participants

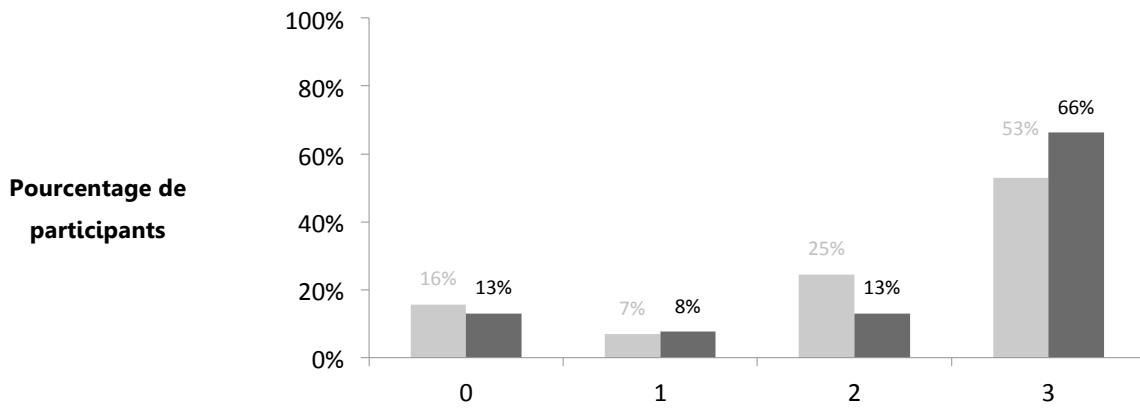
N = 102



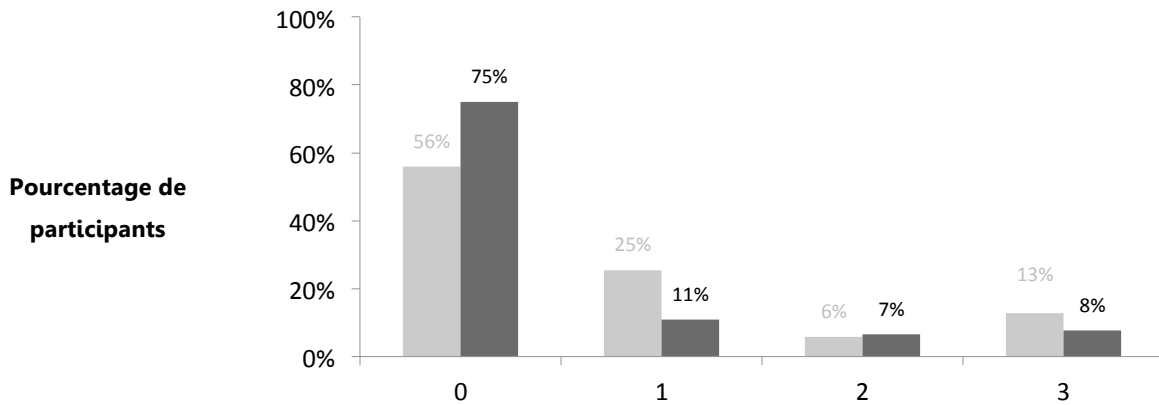
N = 92



Nombre de réponses correctes



Nombre de réponses fausses



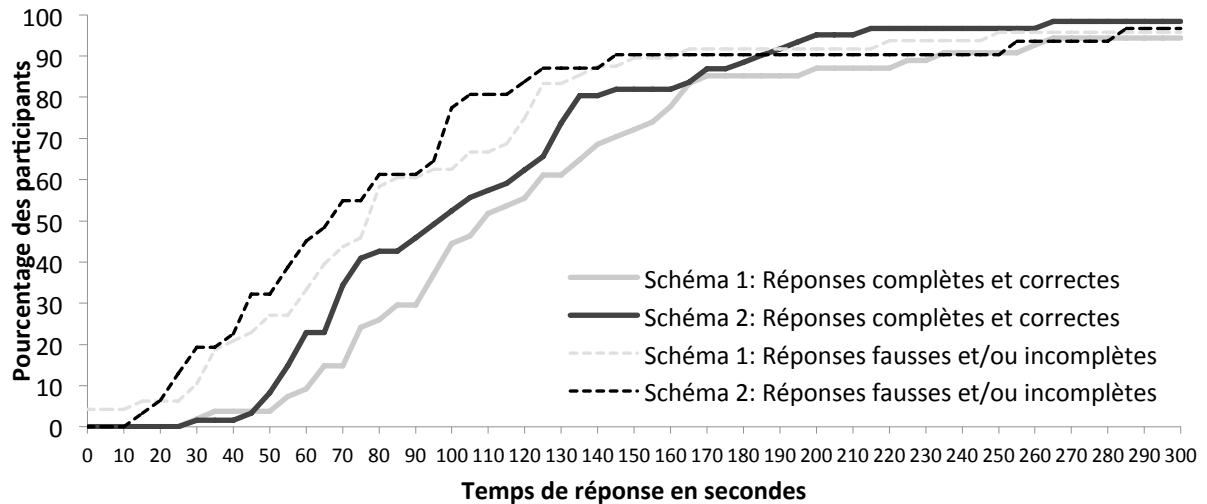
Tests statistiques sur l'exactitude et la complétude

Réponses correctes					Intervalle de confiance à 95% pour la moyenne		Min	Max
Schéma	N	Moyenne	Ecart-type	Erreur standard	Borne inférieure	Borne supérieure		
1	102	2,2	1,1	0,1	1,9	2,4	0	3
2	92	2,3	1,1	0,1	2,1	2,6	0	3

$F(1,192) = 1.300, p = 0.256 \rightarrow$ pas de différence significative

Réponses fausses					Intervalle de confiance à 95% pour la moyenne		Min	Max
Schéma	N	Moyenne	Ecart-type	Erreur standard	Borne inférieure	Borne supérieure		
1	102	0,8	1,0	0,1	0,5	1,0	0	3
2	92	0,5	0,9	0,1	0,3	0,7	0	3

$F(1,192) = 4.132, p = 0.043 \rightarrow$ différence significative, corrélation de Pearson = -0.145



Tests statistiques sur le temps de réponse

Réponses complètes et correctes					Intervalle de confiance à 95% pour la moyenne			
Schéma	N	Moyenne	Ecart-type	Erreur standard	Borne inférieure	Borne supérieure	Min	Max
1	54	136	83	11	113	158	32	450
2	61	114	65	8	97	130	34	434

$F(1, 113) = 2.557, p = 0.113 \rightarrow$ effet tendanciel

Réponses fausses et/ou incomplètes					Intervalle de confiance à 95% pour la moyenne			
Schéma	N	Moyenne	Ecart-type	Erreur standard	Borne inférieure	Borne supérieure	Min	Max
1	48	107	109	16	75	138	3	580
2	31	93	81	15	63	123	15	383

$F(1, 77) = 0.366, p = 0.547 \rightarrow$ pas de différence significative

Schéma 1 : différence entre les réponses complètes et correctes et les autres :

$F(1, 100) = 2.336, p = 0.130 \rightarrow$ pas de différence significative

Schéma 2 : différence entre les réponses complètes et correctes et les autres :

$F(1, 91) = 1.792, p = 0.184 \rightarrow$ pas de différence significative