



UNIL | Université de Lausanne

Faculté de droit
Ecole des Sciences Criminelles
Institut de Police Scientifique

**Intégration de méthodes de data mining dans le
renseignement criminel.
Analyse par des structures issues de la théorie des graphes
dans le profilage des stupéfiants.**

THESE DE DOCTORAT

Présentée à l'Institut de Police Scientifique
de l'Université de Lausanne par

Anne-Laure Terrettaz-Zufferey

Diplômée en Sciences de la Terre de l'Université de Lausanne

Jury :

Président :	Prof. Pierre Margot
Directeur :	Prof. Olivier Ribaux
Co-directeur :	Prof. Pierre Esseiva
Expert interne :	Prof. Mihail Kanevski
Expert externe :	Dr. Christian Zingg
Expert externe :	Jean-Luc Gremaud

Lausanne, 2009

Ce travail a été soutenu par le FNSRS : numéro 105211-107862

*“[L’analyse de données] est un outil pour dégager de la gangue des données le pur
diamant de la véridique nature”*

Jean-Paul Benzécri (Histoire et préhistoire de l’analyse des données, 1976)

Remerciements

Ce travail de thèse a été réalisé sous la direction de Monsieur le professeur Olivier Ribaux de l'Institut de Police Scientifique.

Le jury était composé de Monsieur le professeur Olivier Ribaux de l'Institut de Police Scientifique, directeur de thèse, de Monsieur le professeur Pierre Margot, directeur de l'Ecole des Sciences Criminelles, président, de Monsieur le professeur Pierre Esseiva de l'Institut de Police Scientifique, co-directeur, de Monsieur le professeur Mikhaïl Kanevski, de l'Institut de Géomatique et d'Analyse du Risque, expert, de Monsieur Jean-Luc Gremaud, chef de la Section d'Identité Judiciaire de la Police Cantonale du Valais et de Monsieur le Docteur Christian Zingg, chef du Service d'Identité Judiciaire de la Police Cantonale de Berne, expert.

J'aimerais remercier vivement les personnes qui ont contribué de près ou de loin à ce travail. Le professeur Olivier Ribaux qui m'a fait découvrir l'analyse criminelle et m'a permis d'utiliser mes connaissances préalables pour réaliser ce travail. Je le remercie pour sa disponibilité, sa confiance et pour les nombreuses discussions, qui ont permis de mener à bien ce travail

Le Professeur Pierre Esseiva, dont l'expérience en matière de stupéfiants et les conseils m'ont été indispensables. Le Professeur Mikhaïl Kanevski, pour ses précieux conseils méthodologiques et grâce à qui j'ai pu persévérer dans l'apprentissage des méthodes de reconnaissance de patterns.

Le Professeur Pierre Margot, directeur de l'Ecole des Sciences Criminelles, pour sa compréhension et son aide lors de mon congé maternité.

Monsieur Jean-Luc Gremaud et Monsieur le Docteur Christian Zingg pour leurs corrections du manuscrit et leurs conseils.

Frédéric Ratle, mon homologue de ce projet à l'IGAR, qui m'a beaucoup aidé pour la programmation et les méthodes d'analyse.

Le Fond National Suisse de a Recherche Scientifique grâce à qui j'ai pu être soutenue financièrement durant 4 ans. Sylvain Ioset, qui m'a permis de comprendre le contexte de ce travail en partageant ses connaissances du domaine des stupéfiants et en analyse criminel ainsi qu'Amélie Baylon avec qui j'ai pu partagé mon intérêt pour la théorie des graphes et un bureau lors de notre activité au sein de la BAAC.

Frédéric Anglada et Laetitia Gasté pour leur aide et leurs précieux conseils, le groupe stupéfiant et le groupe analyse criminelle.

Olivier Delémont, pour son aide lors du travail réalisé pour l'ECA.

Le bureau 6422, pour toute la bonne humeur qui y régnait.

Le “groupe de filles” pour leur amitié et leur soutien.

J'aimerais aussi remercier la Brigade d'Appui d'Analyse et de Coordination de la Police Cantonale Vaudoise, qui m'a permis de me former au domaine de l'analyse criminelle opérationnelle et grâce à qui j'ai pu apporter des idées nouvelles à mon travail de recherche.

La Section Stupéfiant de la Police Cantonale Valaisanne, chez qui j'ai pu effectuer un stage de plusieurs mois et qui m'a fait découvrir le domaine pratique de la lutte contre le trafic des stupéfiants.

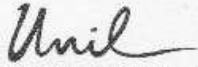
Anne Marville, grâce à qui tous les soucis administratifs pouvant survenir durant cinq années n'ont jamais été des soucis et que je remercie aussi pour ses encouragements, sa gentillesse et sa communicative bonne humeur.

Silvio Viotti, qui a toujours répondu présent lors de problèmes d'informatique.

Eric et Sarah Stauffer-Braun, pour leurs encouragements et pour m'avoir fait découvrir le monde d'Apple et pour avoir corrigé l'anglais de mes publications.

Comme dit un proverbe chinois “lorsqu'on mange un fruit il faut toujours penser à celui qui a planté l'arbre”, j'aimerais remercier le professeur Jean Guex, mon directeur de diplôme de géologue, qui m'a fait découvrir et aimer la théorie des graphes.

Ma famille qui m'a soutenue tout au long de ce travail, en particulier ma Maman avec qui je partage la passion de la recherche scientifique, mon Papa, Fabienne, mes sœurs Lucie et Théry et mon frère Jean-Christophe. Et un immense merci à Olivier, mon mari, pour son inconditionnel soutien à tout moment et Noéï, qui a égailé de son doux chant de pinson, la longue période de rédaction.



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 Madame Anne-Laure Terrettaz-Zufferey, candidate au doctorat en sciences forensiques, intitulée

« Intégration de méthodes de data mining dans le renseignement criminel.
Analyse par des structures issues de la théorie des graphes dans le profilage
des stupéfiants »

Le Président du Jury



Professeur Pierre Margot

Lausanne, le 20 février 2009

Abstract

Data mining can be defined as the extraction of previously unknown and potentially useful information from large datasets. The main principle is to devise computer programs that run through databases and automatically seek deterministic patterns. It is applied in different fields of application, e.g., remote sensing, biometry, speech recognition, but has seldom been applied to forensic case data. The intrinsic difficulty related to the use of such data lies in its heterogeneity, which comes from the many different sources of information. The aim of this study is to highlight potential uses of pattern recognition that would provide relevant results from a criminal intelligence point of view.

The role of data mining within a global crime analysis methodology is to detect all types of structures in a dataset. Once filtered and interpreted, those structures can point to previously unseen criminal activities. The interpretation of patterns for intelligence purposes is the final stage of the process. It allows the researcher to validate the whole methodology and to refine each step if necessary.

An application to cutting agents found in illicit drug seizures was performed. A combinatorial approach was done, using the presence and the absence of products. Methods coming from the graph theory field were used to extract patterns in data constituted by links between products and place and date of seizure. A data mining process completed using graphing techniques is called “graph mining”. Patterns were detected that had to be interpreted and compared with preliminary knowledge to establish their relevancy. The illicit drug profiling process is actually an intelligence process that uses preliminary illicit drug classes to classify new samples. Methods proposed in this study could be used *a priori* to compare structures from preliminary and post-detection patterns. This new knowledge of a repeated structure may provide valuable complementary information to profiling and become a source of intelligence.

Table des matières

Remerciements	iv
Abstract	vii
1 Introduction	1
2 Problématique	5
2.1 Introduction	5
2.2 “Intelligence-Led Policing”	5
2.2.1 Emergence du concept	5
2.2.2 Le modèle conceptuel	6
2.2.3 Son application	6
2.3 Le contexte criminel	7
2.4 Le data mining	8
2.5 Quels résultats espérer	10
2.5.1 Qu’est-ce qu’un “pattern”	10
2.5.2 Différenciation	10
2.6 Les connaissances <i>a priori</i>	11
2.6.1 Quelles sont-elles?	11
2.6.2 Catégorisation et utilisation	11
2.6.3 Un travail d’équipe	13
2.7 Utilisation de ces techniques	13
2.7.1 Les champs d’application	13
2.7.2 Sélection des données tests	14
2.7.3 Sélection des méthodes d’analyse	14
2.7.4 Un exemple de limite à leur application	14

3	La reconnaissance de patterns en analyse criminelle	16
3.1	Introduction	16
3.2	Reconnaissance de patterns dans le processus du renseignement	17
3.2.1	Schématisation du processus	17
3.3	Mise en pratique dans une optique de renseignement	19
3.3.1	Structure des données	19
3.3.2	Interprétation des résultats	20
3.3.3	Utilisation dans une structure de Police	22
4	Des techniques prometteuses issues de la théorie des graphes	25
4.1	Introduction	25
4.2	La théorie des graphes	27
4.2.1	Introduction	27
4.2.2	Notions théoriques	27
4.3	Techniques issues de la théorie des graphes et leurs applications	31
4.3.1	Introduction	31
4.3.2	Etat de l'art en matière d'utilisation de graphes dans le domaine du renseignement	31
4.3.3	Les algorithmes testés et leurs applications	32
4.3.4	Evaluation	37
4.4	Les outils	37
4.4.1	Les logiciels avec GUI	38
4.4.2	Les librairies	38
5	Application aux données liées au trafic des stupéfiants	39
5.1	Introduction	39
5.2	Eléments sur le trafic des stupéfiants	40
5.2.1	Généralités	40
5.2.2	Le trafic de l'héroïne	40
5.2.3	Le trafic de la cocaïne	41
5.2.4	Le profilage des stupéfiants (éléments tirés de [Ioset <i>et al.</i> , 2005])	43
5.2.5	Les données issues de saisies d'héroïne et de cocaïne	48
5.3	Analyse	51
5.3.1	Les informations à disposition	51
5.3.2	Méthodes d'analyse	58

6 Résultats et interprétation	69
6.1 Tendances générales	69
6.1.1 Introduction	69
6.1.2 Nombre de produits et nombre de cliques	70
6.1.3 Interprétation	73
6.2 Quels produits dans quels cantons	75
6.2.1 Produits les plus répandus dans les quatre cantons d'études	75
6.2.2 Produits ajoutés après le niveau de production	77
6.2.3 Relations entre les cantons	79
6.2.4 Interprétation	85
6.3 Tendances locales liées à l'espace et au temps	87
6.3.1 L'héroïne	88
6.3.2 La cocaïne	104
7 Discussion	119
7.1 Introduction	119
7.2 Application aux données issues du trafic des stupéfiants	119
7.2.1 La théorie des graphes : pertinence et originalité	119
7.2.2 Application de la théorie des ensembles à des graphes	120
7.2.3 Simplification de graphes et détection de groupes	121
7.3 Méthodes complémentaires au profilage	123
7.3.1 Quels moyens de validation	125
7.3.2 A qui les rôles d'experts ?	126
7.4 Comment intégrer ces méthodes à un processus de renseignement	127
7.4.1 Exemple	127
7.4.2 Une boîte noire dans le processus du renseignement ?	129
7.5 Perspectives	131
8 Conclusion	133
Annexes	150
A Codes des combinaisons de stupéfiants	150
A.1 Héroïne	150
A.2 Cocaïne	151
B Graphes du chapitre 6.2	159
B.1 Théorie des ensembles	159

C Graphes du chapitre 6.3	166
C.1 Patterns liés à l'héroïne	166
C.2 Patterns liés à la cocaïne	169
D Liste des abréviations	172

Table des figures

2.1	Le “modèle des trois-i” tiré de [Ratcliffe, 2008].	6
2.2	Triangle relationnel des acteurs et circonstances influençant l’occurrence d’un crime, tiré de [Felson, 1998].	8
2.3	Etapes du “data mining” [Terrettaz-Zufferey <i>et al.</i> , 2006].	9
2.4	Utilisation des connaissances <i>a priori</i> : les trois catégories interviennent particulièrement pour le choix des variables, le choix des méthodes et l’interprétation.	12
3.1	Processus d’analyse criminelle tiré du (cours ”Renseignement forensique“, O. Ribaux 2008).	17
3.2	Possibilité d’intégration des méthodes de reconnaissance de patterns dans le processus du renseignement.	18
3.3	Schématisation de l’étape d’interprétation des tendances.	21
3.4	Mise en pratique des méthodes de reconnaissance de patterns dans une coordination judiciaire (schéma élaboré en collaboration avec la BAAC de la Police Cantonale Vaudoise).	23
4.1	Graphe simple et graphe orienté.	27
4.2	Graphe simple et matrice d’adjacence.	28
4.3	Graphe d’intervalle.	28
4.4	Graphes simple et bipartis.	29
4.5	Graphe pondéré.	29
4.6	Chemin et cycle (tiré de [West, 1996]).	30
4.7	Graphe connexe et non-connexe.	30
4.8	Illustration d’un Minimum Spanning Tree mis en évidence sur le graphe initial. . .	33
4.9	Point d’articulation.	34
4.10	Point d’articulation et sommets isolés.	34
4.11	En rouge, la clique maximale : aucun autre sommet ne peut être ajouté.	35
4.12	Graphe non connexe, connexe et blocs.	36

4.13 Communautés dans un graphe [Rupnik, 2006].	37
5.1 Principales voies d'entrée en Europe de l'héroïne (d'après [ONU DC, 2006]).	41
5.2 Principales voies d'entrée en Europe de la cocaïne (d'après [ONU DC, 2006]).	42
5.3 Processus de renseignement basé sur des saisies de stupéfiants en Suisse (tiré de [Ioset <i>et al.</i> , 2005]).	43
5.4 Extraction des données à partir des chromatogrammes (tiré de [Ioset <i>et al.</i> , 2005]).	44
5.5 Processus général de l'établissement des liens chimiques (tiré de [Ioset <i>et al.</i> , 2005]).	46
5.6 Illustration des différents types de liens (tiré de [Ioset <i>et al.</i> , 2005]).	47
5.7 Caractéristiques des saisies selon les cantons.	49
5.8 Réseau général d'approvisionnement des stupéfiants (schéma inspiré de [UNODC, 2001]).	51
5.9 Hiérarchisation possible du processus de coupage. Les rectangles de couleurs correspondent à des produits de coupage.	53
5.10 Nombre de données par cantons d'étude.	56
5.11 Nombre de données par années d'étude.	56
5.12 Pureté moyenne de l'héroïne par cantons.	57
5.13 Pureté moyenne de la cocaïne par cantons.	57
5.14 Graphes du haut : les sommets correspondent à des produits et une arête les lie lorsqu'ils ont été détectés dans une région ou une année. Graphes du bas : les sommets correspondent aux combinaisons, aux trimestres ou aux localités de saisie. Une arête les lie lorsque les combinaisons y sont détectées.	60
5.15 Illustration de l'application de la théorie des ensembles	62
5.16 En haut le graphe de base, en bas le même graphe, mais simplifié par la méthode du minimum spanning tree basé sur le taux de pureté.	64
5.17 Partie d'un graphe dont les points d'articulation sont en orange.	65
5.18 Partie d'un graphe contenant deux communautés distinguables au moyen de couleurs différentes.	66
5.19 Données des combinaisons en format binaire.	67
5.20 Illustration d'un graphe d'intervalle.	67
5.21 Détection de cliques au moyen d'un graphe d'intervalle.	68
6.1 Graphes des combinaisons de l'héroïne.	71
6.2 Graphes des combinaisons de la cocaïne.	72
6.3 Nombre produits et nombre de cliques.	73
6.4 Nombre de combinaisons.	73

6.5	Grappe $G_{ProdGenH} = G_{Ti} \cap G_{Ge} \cap G_{Vd} \cap G_{Ne}$ pour l'héroïne.	76
6.6	Grappe $G_{ProdGenC} = G_{Ti} \cap G_{Ge} \cap G_{Vd} \cap G_{Ne}$ pour la cocaïne.	77
6.7	Grappe $G_{ProdCH_H} = G_{TiGeVdNeH} \setminus G_{ProdGenH}$ pour l'héroïne.	78
6.8	Grappe $G_{ProdCH_C} = G_{TiGeVdNeC} \setminus G_{ProdGenC}$ pour la cocaïne.	79
6.9	Grappe $G_{ProdEst_H} = G_{Ge} \cap G_{Ne}$ pour l'héroïne.	81
6.10	Grappe $G_{ProdEst_H} = G_{Vd} \cap G_{Ne}$ pour l'héroïne.	81
6.11	Grappe $G_{ProdEst_C} = G_{Ge} \cap G_{Ne}$ pour la cocaïne.	82
6.12	Grappe $G_{ProdEst_C} = G_{Vd} \cap G_{Ne}$ pour la cocaïne.	82
6.13	Grappe $G_{ProdEst_H} = G_{Ti} \cap G_{Ge}$ pour l'héroïne.	83
6.14	Grappe $G_{ProdEst_H} = G_{Vd} \cap G_{Ge}$ pour l'héroïne.	84
6.15	Grappe $G_{ProdEst_C} = G_{Ti} \cap G_{Ge}$ pour la cocaïne.	84
6.16	Grappe $G_{ProdEst_C} = G_{Vd} \cap G_{Ge}$ pour la cocaïne.	85
6.17	Grappe de base "Combinaisons - Localités" avec points d'articulation.	89
6.18	Illustration d'un graphe pondéré par la moyenne de la pureté pour deux localités. .	90
6.19	Détection de communautés par "walktrap" appliqué au minST "Combinaisons - Localités"	91
6.20	Détection de communautés par "walktrap" appliqué au maxST "Combinaisons - Localités"	92
6.21	Synthèse de la répartition des produits.	93
6.22	Voies de passage de l'héroïne sur le canton de Genève.	95
6.23	Grappe de base "Combinaisons - Trimestres" avec points d'articulation.	96
6.24	Détection de communautés par "walktrap" appliqué au minST "Combinaisons - Trimestres"	97
6.25	Détection de communautés par "walktrap" appliqué au maxST "Combinaisons - Trimestres"	97
6.26	Sous-graphes mis en évidence caractérisés par l'absence de P-C.	99
6.27	Illustration des relations entre groupes de sommets pour l'application du modèle des blocs.	101
6.28	Modèle des blocs appliqué au graphe "Combinaisons - Trimestres + Combinaisons - Saisies + Saisies - Trimestres"	102
6.29	Illustration de l'application du modèle des blocs à la problématique.	103
6.30	Grappe de base "Combinaisons - Localités" avec points d'articulation pour Ge-Vd-Ne.104	
6.31	Détection de communautés par "walktrap" appliqué au minST "Combinaisons - Localités" de Ge-Vd-Ne	105

6.32	Détection de communautés par “fast greedy” appliqué au maxST “ Combinaisons - Localités ” de Ge-Vd-Ne	106
6.33	Histogramme de la pureté de la cocaïne et de l’héroïne.	107
6.34	Communauté liée à Lancy (Ge) avec comme dénominateur commun : Glucose-Mannitol-Phénacétine.	108
6.35	Communauté liée à Prilly (Vd) et Anières (Ge), avec comme dénominateurs communs : Lactose - Sucrose et Lactose - Glucose.	109
6.36	Communauté liée à Tolochenaz (Vd) avec une combinaisons très proche, mais une anomalie au niveau du taux de pureté.	110
6.37	Graphe de base “Combinaisons -Trimestres” avec points d’articulation.	111
6.38	Cliques maximales pour le canton de Genève.	112
6.39	Cliques maximales pour le canton de Vaud.	112
6.40	Détection de communautés par “walktrap” appliqué au minST “ Combinaisons - Trimestres ”.	114
6.41	Communauté liée au trimestre 2, montrant des similitudes au niveau des combinaisons et des classes chimiques.	115
6.42	Analyse en composantes principales des échantillons de la communauté liée au trimestre 2.	116
6.43	Communauté liée au trimestre 24, caractérisée par des produits peu fréquemment détectés.	117
7.1	Illustration de l’utilisation de graphes simplifiés pour appliquer la théorie des ensembles.	121
7.2	Mise en évidence du minimum spanning tree d’un graphe.	122
7.3	Processus de renseignement par interprétation des liens chimiques (tiré de [Ioset <i>et al.</i> , 2005]).	124
7.4	Processus d’une demande de la part d’un enquêteur (schéma réalisé en collaboration avec les analystes de la Police Cantonale Vaudoise).	129
7.5	Processus de renseignement.	130
B.1	Graphe du canton de Genève pour l’héroïne	159
B.2	Graphe du canton de Genève pour la cocaïne	160
B.3	Graphe du canton de Vaud pour l’héroïne	161
B.4	Graphe du canton de Vaud pour la cocaïne	161
B.5	Graphe du canton du Tessin pour l’héroïne	162
B.6	Graphe du canton du Tessin pour la cocaïne	162

B.7	Graphe du canton de Neuchâtel pour l'héroïne	163
B.8	Graphe du canton de Neuchâtel pour la cocaïne	163
B.9	Graphe $G_{ProdTiNeH} = G_{Ti} \cap G_{Ne}$ pour l'héroïne	164
B.10	Graphe $G_{ProdTiNeH} = G_{Ti} \cap G_{Ne}$ pour la cocaïne	164
B.11	Graphe $G_{ProdTiVdH} = G_{Ti} \cap G_{Vd}$ pour l'héroïne	165
B.12	Graphe $G_{ProdTiVdH} = G_{Ti} \cap G_{Vd}$ pour la cocaïne	165
C.1	Détection de communautés par "fast greedy" au minST de l'héroïne Combinaisons - Localités	166
C.2	Détection de communautés par "fast greedy" au maxST de l'héroïne Combinaisons - Localités	167
C.3	Détection de communautés par "fast greedy" au minST de l'héroïne Combinaisons - Trimestres	167
C.4	Détection de communautés par "fast greedy" au maxST de l'héroïne Combinaisons - Trimestres	168
C.5	Graphe de base de la cocaïne " Combinaisons - Localités " avec points d'articulation pour Ge-Ti	169
C.6	Détection de communautés par "fast greedy" au minST de la cocaïne Combinaisons - Localités pour Ge-Ti	170
C.7	Détection de communautés par "walktrap" au minST de la cocaïne Combinaisons - Localités pour Ge-Ti	170
C.8	Détection de communautés par "fast greedy" au minST de la cocaïne Combinaisons - Localités pour Ge-Vd-Ne	171
C.9	Détection de communautés par "fast greedy" au minST de la cocaïne Combinaisons - Trimestres	171

Liste des tableaux

4.1	Connexité et adjacence (tiré de [West, 1996])	30
5.1	Format des données par produits	54
5.2	Format des données par combinaisons.	54
5.3	Catégories et répartition des produits de coupage	55
5.4	Nombre de données brutes	55
5.5	Nombre de produits et combinaisons différents détectés	55
5.6	Pureté moyenne	56
6.1	Catégories et répartition des produits de coupage (reproduction de la table 5.3) . .	80

Chapitre 1

Introduction

Mais que fait la Police ?... pour diminuer le nombre de cambriolages, le nombre de vols à l'astuce au préjudice de personnes âgées, le deal de substances dangereuses, etc.. Un des objectifs de la Police est de diminuer la criminalité. Elle le fait selon différentes approches comme par exemple, mettre en place des actions lorsqu'un type de crime (cambriolages, émergence de réseaux de stupéfiants etc..) prend de l'ampleur. Une autre possibilité consiste à prévenir une augmentation des délits avant qu'ils ne deviennent incontrôlables, en s'appuyant sur le renseignement [Ratcliffe, 2008]. Le processus du renseignement commence à être de plus en plus utilisé par les Polices [Ribaux et Birrer, 2008].

Des banques de données ont été mises en place grâce à l'évolution de la technologie, permettant ainsi de stocker des données liées à la criminalité en format numérique. Celles-ci sont en permanence alimentées et il devient toujours plus difficile pour les analystes d'en conserver une vue d'ensemble et de pouvoir agir de manière proactive.

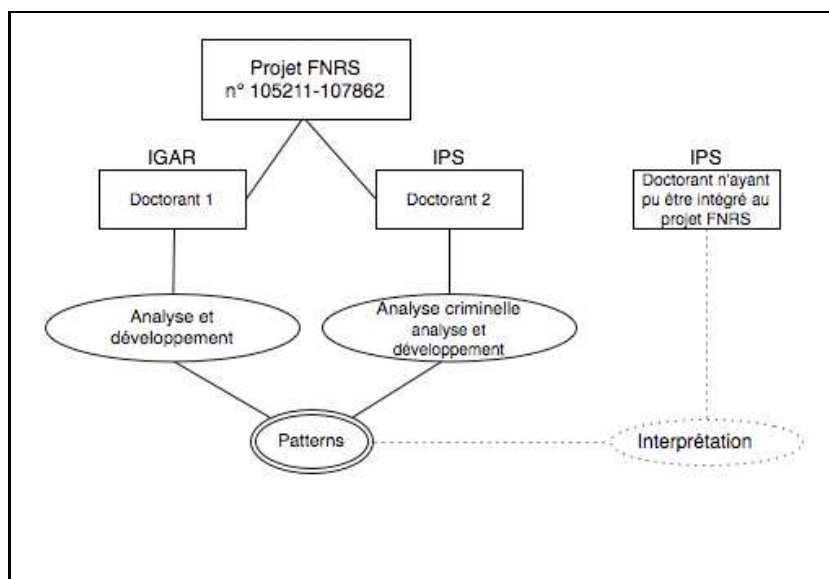
A partir de ce moment-là, il devient nécessaire pour conserver cette proactivité, d'avoir recours à d'autres méthodes pour mettre en évidence ces tendances. Un précédent travail basé sur la gestion des liens a été réalisé [Ribaux, 1997] et propose une démarche d'analyse permettant de produire de l'information.

Le "data mining" est un processus d'exploration et d'analyse de grandes banques de données informatiques [Tufféry, 2007]. Dans un cas idéal, son application devrait permettre d'extraire des informations (tendances, schémas) nouvelles et utiles pour le détenteur des données.

Ce processus semble prometteur pour détecter des schémas dans les données de criminalité, mais la manoeuvre ne s'arrête pas à presser un simple bouton et attendre que le résultat montre un nouveau phénomène criminel à endiguer. Ces méthodes sont complexes, difficiles à appliquer et demandent une interprétation soignée pour être exploitables. Qui, dans une structure de Police, pourrait utiliser ces méthodes et à qui le rôle d'interprétation. Pourquoi semblent-elles actuelle-

ment peu utilisées dans le domaine du renseignement criminel ? Est-il possible de les intégrer dans un processus de renseignement afin de les utiliser à des fins opérationnelles de Police ?

Ce travail fait partie d'un projet soutenu par le Fond National Suisse de La Recherche Scientifique (<http://www.snf.ch/F/pages/default.aspx>, numéro 105211-107862). Son objectif global est d'explorer les méthodes de data mining, en particulier de délimiter leur potentiel pour le renseignement criminel. Sa structure générale est présentée par le schéma ci-contre :

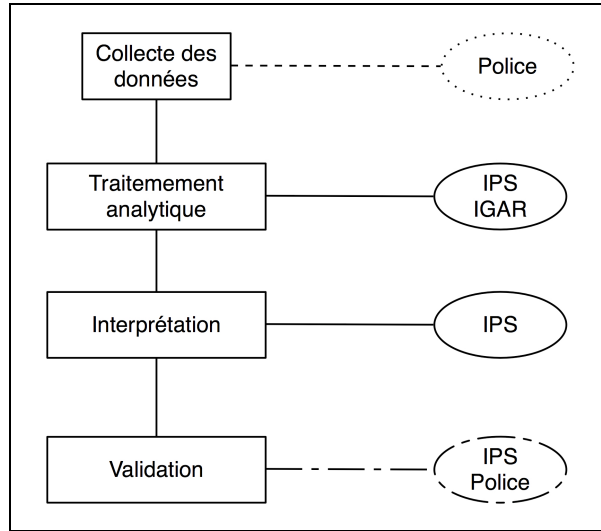


Il est mené à bien par deux doctorants ayant chacun une spécialité et un rôle complémentaire. Le premier collabore avec l'Institut de Géomatique et d'Analyse du Risque (IGAR)¹ et s'est orienté vers le développement de techniques d'analyses spécialisées pour les données liées au trafic des stupéfiants (statistiques, méthodes d'apprentissage automatisé). Le second, l'auteur de ce travail collabore avec l'Institut de Police Scientifique (IPS) et a orienté la recherche vers des méthodes pouvant être utilisées dans une unité d'analyse criminelle, pour divers types de données, ainsi que sur le cadre général dans lequel les appliquer. Une application aux données liées au trafic des stupéfiants a aussi été réalisée. Pour mener à bien l'intégralité du projet, un troisième doctorant de l'IPS aurait dû être intégré. Son rôle aurait été l'interprétation des patterns détectés au moyen de connaissances opérationnelles de Police, afin de valider les résultats. Pour des raisons de subsides non octroyés, cette partie n'aura été que partiellement faite.

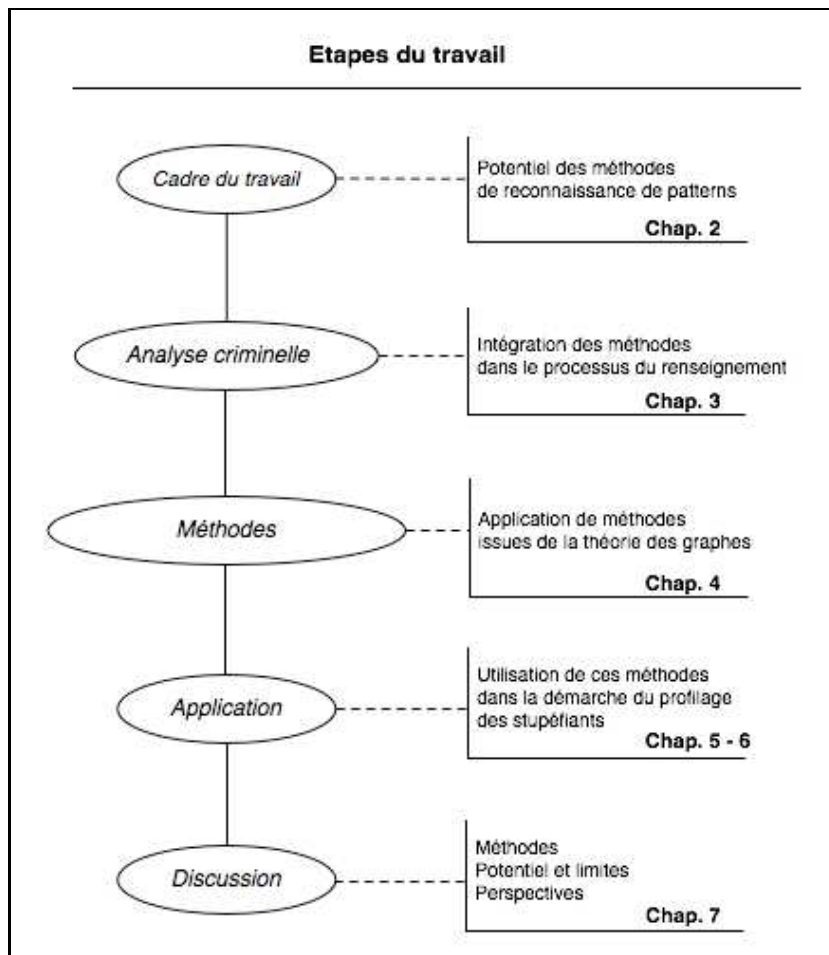
Plusieurs tâches de ce projet sont menées à bien par différents intervenants. Elles sont schématisées

1. Les abréviations sont listées en annexes.

ci-contre.



La présente recherche a pour but d'évaluer les possibilités d'intégration de ces techniques dans un processus de renseignement criminel, dans le but de pouvoir résoudre par des réflexions et solutions scientifiques, des problèmes issus du domaine opérationnel de Police. Le présent travail contient trois étapes, présentées sur le schéma suivant :



Ce travail traite la problématique à plusieurs niveaux. La partie “analyse criminelle” décrit une manière possible d’intégrer les méthodes d’analyse de patterns dans le processus du renseignement. La partie “méthodes” propose d’aborder les données liées au trafic des stupéfiants d’une manière nouvelle, et d’appliquer une technique d’analyse inédite et novatrice dans ce domaine. La partie “application” permet de réaliser des tests sur les données issues du trafic de l’héroïne et de la cocaïne en tentant de l’intégrer à un système opérationnel pour le profilage des stupéfiants.

Chapitre 2

Problématique

2.1 Introduction

Ce chapitre a pour but de placer le cadre dans lequel ce travail a été réalisé. Il va tenter de mettre en évidence le contexte dans lequel ces techniques pourraient être utilisées, leur apport pour une utilisation stratégique et opérationnelle ainsi que leur intégration à un processus de renseignement criminel.

2.2 “Intelligence-Led Policing”

2.2.1 Emergence du concept

Ce terme anglais peut être traduit par “vers une fonction de Police guidée par le renseignement”. D’après [Ratcliffe, 2008], les lacunes des systèmes traditionnels de Police, les changements au niveau de la technologie, l’évolution du crime organisé et du grand banditisme ont été des facteurs ayant mené vers l’étude de nouveaux modèles de lutte contre la criminalité. Les systèmes traditionnels de Police sont réactifs, générant des actions lorsque l’augmentation d’un type de délits ou une série de crimes s’est déjà développée. Parallèlement, les événements criminels sont enregistrés dans des banques de données. L’évolution de la technologie ainsi que des méthodes d’analyse permettraient de traiter ces données informatiquement. Le data mining interviendrait alors pour leur exploitation et mettrait en évidence de l’information qui pourrait être une aide à la décision pour la Police.

D’après [Davenport, 1997], les données analysées deviennent de l’information lorsqu’une tendance est détectée. L’information devient une connaissance et si elle est utilisée pour avoir un impact sur les décisions affectant l’environnement criminel, elle devient du renseignement. C’est pour tenter

de pallier à l'évolution de la criminalité et utiliser la technologie actuelle, qu'un modèle de Police basé sur le renseignement a été pensé.

2.2.2 Le modèle conceptuel

Dans [Ratcliffe, 2008], est proposé un modèle simplifié sur une manière de réduire le crime dans une Police guidée par le renseignement : le “modèle des trois-i” (figure 2.1).

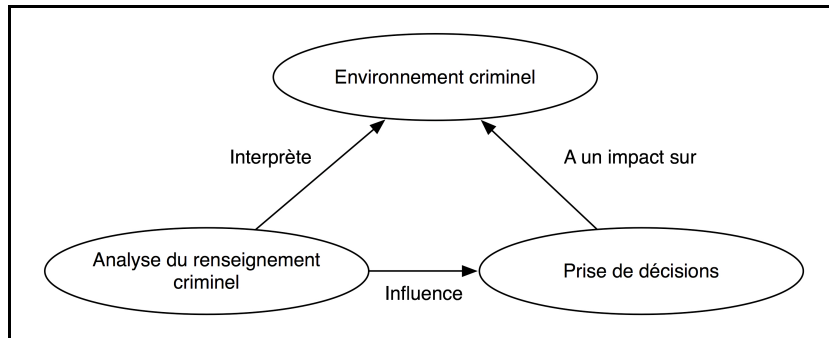


FIGURE 2.1 – Le “modèle des trois-i” tiré de [Ratcliffe, 2008].

Dans ce modèle, l'unité d'analyse d'une Police interpréterait l'environnement criminel de manière à cibler les malfaiteurs jouant les rôles les plus importants et à définir les menaces les plus sérieuses. Pour une prévention et une réduction du crime ainsi que pour perturber et désorganiser des auteurs, les analystes doivent interpréter l'environnement criminel, utiliser le renseignement pour influencer les prises de décision et ceux qui prennent les décisions doivent diriger les ressources efficacement, de manière à avoir un impact positif sur l'environnement criminel.

2.2.3 Son application

Un rapport anglais [Audit-Commission, 1993] recommande aux Polices de passer d'un système réactif vers un système proactif de lutte contre le crime. Ce changement permettrait d'augmenter l'efficacité de la Police et apporterait une meilleure utilisation des ressources financières et une meilleure offensive contre les criminels.

Le renseignement est généralement défini sur deux niveaux :

- Stratégique, en détectant des structures à partir de données, permettant d'avoir une image globale de la situation en matière de délits, de sa gravité et de son évolution.
- Opérationnel, au moyen d'enquêtes, permettant de disposer d'informations fiables, issues directement de l'environnement criminel.

Ces deux niveaux ne sont pas étanches. Les informations qu'ils produisent sont complémentaires. Un partage permettrait de valider certaines tendances détectées au niveau stratégique qui elles-mêmes pourraient guider une enquête. Aujourd'hui, certains services de Police travaillent sur ces deux niveaux. Analyser l'ensemble des données liées à la criminalité, afin de visualiser des tendances générales, permet d'obtenir du renseignement stratégique. Celui-ci permet d'avoir une vision globale qui peut être utile pour la recherche d'informations ou le traitement de données ciblées (une affaire ou un type de délit particulier), en vue de produire du renseignement opérationnel.

Le concept de Police guidée par le renseignement commence à s'imposer dans notre pays, notamment au travers d'unités d'analyse. Celles-ci génèrent des informations à un niveau stratégique, traitent l'information opérationnelle et la diffusent, ainsi que le renseignement produit, aux régions voisines.

Les informations pouvant être utilisées à un niveau stratégique sont générées par les analystes, en utilisant leurs expériences et en exploitant des banques de données. Celles-ci contiennent très certainement de l'information cachée (liée notamment au contexte criminel, par exemple), pouvant être mise à jour sous forme de patterns au moyen de techniques informatisées et utilisées à un niveau stratégique et opérationnel. C'est dans ce cadre que ce travail a été réalisé.

2.3 Le contexte criminel

La "théorie des opportunités" stipule que la problématique générale de la criminalité est liée à des personnes et à leurs comportements, donc à des habitudes de vie [Felson et Cohen, 1980; Felson et Clarke, 1998]. Ces habitudes sont régies par des activités qui se répètent fréquemment et qui font partie de la vie quotidienne des gens : manger, dormir, loisirs. Elles meublent leur existence et restreignent l'espace spatio-temporel exploitable à des fins criminelles. Un cambrioleur sera actif pour son délit à des heures différentes d'un dealer. Le premier cherchera à s'éloigner des heures d'activité humaine, alors que le second cherchera le contact de clients potentiels.

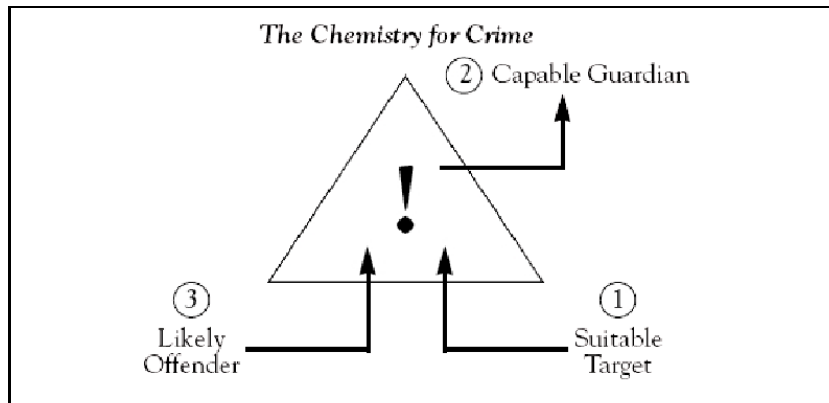


FIGURE 2.2 – Triangle relationnel des acteurs et circonstances influençant l’occurrence d’un crime, tiré de [Felson, 1998].

Une combinaison d’acteurs (un criminel motivé et une victime attractive) et de circonstances (un lieu et un moment propices) influence la réalisation d’un crime [Felson et Clarke, 1998]. Ce concept peut être représenté sous la forme d’un triangle relationnel (figure 2.2). Ainsi, si les activités des malfaiteurs sont délimitées par un ensemble de contraintes liées à l’environnement physique et social immédiat, elles devraient suivre généralement des schémas (patterns). Ce sont eux que l’on espère pouvoir détecter dans les données de criminalité ou les traces transférées par l’auteur. Lors d’une démarche d’analyse de données réelles issues de banques de données de Police, la recherche d’une combinaison de ces circonstances peut être utilisée comme outil de base pour structurer les données. Il devient ainsi plus aisé d’aborder les problématiques criminelles en structurant la démarche à partir des relations entre entités, acteurs et circonstances [Birrer, 2009].

2.4 Le data mining

Le data mining, littéralement “forage ou fouille de données”, est l’application des techniques de statistiques, d’analyse des données et d’intelligence artificielle à l’exploration et l’analyse de grandes banques de données informatiques, en vue d’en extraire des informations nouvelles et utiles [Tufféry, 2007]. C’est un processus de traitement composé de plusieurs étapes (figure 2.3). Il débute par l’intégration de données brutes et se termine par l’interprétation des résultats obtenus. Le but est de faire surgir des informations initialement invisibles car noyées dans la banque de données source. Une description du processus et de son historique est faite dans [Frawley *et al.*, 1992].

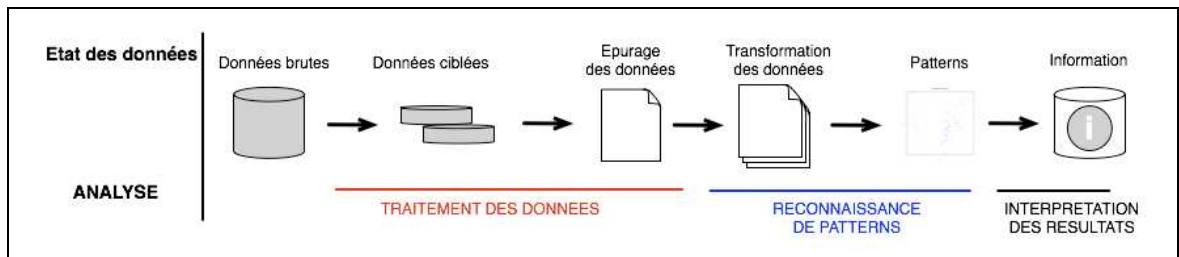


FIGURE 2.3 – Etapes du “data mining” [Terrettaz-Zufferey *et al.*, 2006].

L'étape centrale est la reconnaissance de patterns. Elle est la partie technique du data mining correspondant à une sélection et à l'application de méthodes d'analyses statistiques et mathématiques adaptées aux données et à la problématique traitée. Les résultats issus des méthodes utilisées sont ce qu'on appelle en anglais des “patterns”, sensés contenir de l'information utile au problème abordé (telle que des régularités ou des anomalies). Cette étape est suivie par une phase d'interprétation, permettant de déterminer si le pattern trouvé contient de l'information utilisable. Elle est indispensable pour valider la démarche du data mining.

Le data mining a des applications nombreuses, dont une des plus connue est l'analyse du ticket de caisse dans les grands magasins, afin de déterminer les produits souvent achetés simultanément, détermination permettant d'agencer les rayons et d'organiser les promotions en conséquences. Il est utilisé dans le domaine financier, pour la détection de fraudes, dans le domaine bancaire et les sociétés de crédit, pour prédire les risques de défaillance d'un demandeur de crédit, ou pour l'analyse sensorielle dans l'industrie agroalimentaire, et bien d'autres. Les méthodes de reconnaissance de patterns sont actuellement activement développées dans le domaine de la reconnaissance vocale et faciale, afin de faire reconnaître automatiquement une voix ou un visage par un ordinateur, au moyen d'algorithmes d'apprentissage automatisé. L'analyse d'images satellitaires exploite aussi ces méthodes, afin de reconnaître automatiquement des structures, telles que des roches en géologie ou des bases militaires dans le domaine de la défense.

En matière d'exploitation par data mining de données liées à la criminalité, des recherches sont réalisées par diverses institutions, notamment au niveau de la détection de tendances géographiques [Xue et Brown, 2006], des possibilités d'utilisation dans le domaine de la sécurité [Mena, 2002; Jesus, 2003], et dans la lutte contre la cybercriminalité au niveau de la détection d'intrusions et dans le décryptage [Chunga *et al.*, 2006]. Ces méthodes pourraient être utilisées dans d'autres domaines, comme celui de la lutte contre le trafic des stupéfiants, en appliquant le data mining aux données qui sont générées par ce type d'activité criminelle.

2.5 Quels résultats espérer

2.5.1 Qu'est-ce qu'un "pattern"

Une traduction française de ce terme anglais dans le cadre du data mining, pourrait être "tendance ou structure présente dans un jeu de données mais invisible sans traitement informatique". Des composantes géographiques, temporelles (saisonniers), humaines et sociales influencent le contexte criminel décrit précédemment. Une répétition de ces situations, enregistrées dans les banques de données, génère des régularités, des similarités et des anomalies. Celles-ci constituent des patterns que l'analyste a de la peine à détecter lors du traitement du flux d'évènements. Un exemple serait les cambriolages d'habitations en début de soirée l'automne. Une augmentation de ce type de délits est observée chaque année à la même période. Elle provient d'une situation propice créée par la tombée de la nuit à une heure où les gens ne sont pas encore rentrés chez eux. Il devient facile pour un cambrioleur de savoir si les occupants sont rentrés (si il y a ou non de la lumière dans la maison), et de réaliser son méfait discrètement dans la pénombre. Ce type de tendance est connu, donc aisément identifiable lorsque les évènements de ce type surviennent. Il en existe très probablement d'autres, non identifiés, basés sur des mécanismes similaires mais peut-être avec d'autres facteurs que les saisons (par exemple la géographie, des types d'habitations particulières etc.), qu'une analyse par data mining permettrait de détecter.

Idéalement, les méthodes de reconnaissance de patterns seraient appliquées aux banques de données pour mettre en évidence un ou des patterns dont l'interprétation devrait permettre de dégager des pistes utilisables à un niveau opérationnel et stratégique. Il est peut-être possible d'envisager que des systèmes détectent automatiquement des tendances permettant de produire du renseignement.

2.5.2 Différenciation

Avec l'aide d'analystes et dans une perspective de renseignement, des critères de catégorisation des patterns ont pu être établis et ont mené aux trois groupes suivant :

1. **Les structures utiles** : patterns détectés lors d'une analyse des données, qui ne sont pas ou trop tardivement détectés par une analyse "traditionnelle".

Ce serait par exemple une conjonction de facteurs lieu (proche d'un arrêt de bus ou en rase campagne), cible (ferme, chalet, maison de maître), jour de la semaine, qui correspondrait à un auteur, mais dont l'information serait perdue dans la masse de données et les analystes n'auraient pas la possibilité de la repérer.

2. **Les structures à interpréter** : patterns mis en évidence dans des données mais dont la signification et l'utilité doivent être déterminées avec l'aide de spécialistes.

Lorsqu'une petite partie du phénomène est connue, comme par exemple le trafic des stupéfiants,

il est difficile de pouvoir évaluer immédiatement l'utilité des patterns détectés. Des informations d'enquêtes, telles que l'auditions de suspects, permettraient de déterminer leur pertinence.

3. **Les structures inutiles** : elles peuvent être divisées en trois sous-groupes :

(a) Patterns évidents, visibles sans traitement rendant l'analyse inutile.

Ce serait le cas d'une analyse de la répartition des cambriolages sur le canton de Vaud : la région Lausannoise apparaîtrait immédiatement, ce qui est connu sans traitements.

(b) Patterns ne contenant aucune information liée à une problématique criminelle (présents par hasard).

(c) Patterns obtenus suite à une erreur de programmation ou dans le choix de la méthode d'analyse.

2.6 Les connaissances *a priori*

2.6.1 Quelles sont-elles ?

Etudier la criminalité en général, participer à des enquêtes, trier des données relatives à des délits, apprendre des techniques d'analyse, sont des bagages permettant d'appréhender des données liées à la criminalité avec un oeil familiarisé. Une connaissance du contexte criminel d'un pays ou d'une région, donne un cadre général et permet de se poser les bonnes questions lorsque les événements apparaissent.

Les mémoires humaine et informatique constitue une grande source de connaissances *a priori*. Elles permettent de reconnaître, de classer et de mettre en relation de nouveaux cas grâce aux informations historiques qu'elles contiennent. D'avoir connaissance des cambriolages de soirée en automne est une information *a priori* permettant de reconnaître immédiatement un nouveau cas et de le séparer d'autres tendances détectées. De même que pour des structures à interpréter, pouvoir s'appuyer sur des personnes en charge de l'affaire en cours d'analyse, optimisera la qualité de leur interprétation.

Ces connaissances *a priori* sont importantes pour juger de la pertinence des patterns et de leur potentiel à devenir du renseignement. Elles peuvent provenir de sources diverses, et une catégorisation est proposée dans la section suivante.

2.6.2 Catégorisation et utilisation

Trois groupes non exhaustifs de connaissances *a priori* sont proposés. Ils sont basés sur la pratique de plusieurs types de données issues d'affaires et banques de données de Police, ainsi que

sur une expérience pratique au sein de différentes brigades.

- Connaissance du contexte criminel d'un pays ou d'une région et de son histoire criminelle (mémoire).
- Connaissance technique et méthodologique (méthodes et outils d'analyse et leurs champs d'application).
- Connaissance provenant d'une enquête.

Le tableau de la figure 2.4 propose diverses possibilités d'application. Comme la liste de connaissances *a priori*, ces applications ne sont pas exhaustives. Leur but est de donner quelques pistes pour ne négliger aucune source d'information susceptible d'être intégrée à une démarche d'analyse de données à caractère criminel.

		Types de connaissances <i>a priori</i>		
		Du contexte criminel et mémoire	Technique et méthodologique	De l'affaire traitée
Utilisation	Choix des variables	Indispensables	Utiles	Indispensables
	Prétraitement des données et choix des méthodes	Pas nécessaires	Indispensables	Utiles
	Interprétation	Nécessaires	Utiles	Indispensables

	Pas nécessaires
	Utiles
	Nécessaires
	Indispensables

FIGURE 2.4 – Utilisation des connaissances *a priori* : les trois catégories interviennent particulièrement pour le choix des variables, le choix des méthodes et l'interprétation.

Une connaissance du contexte criminel et la mémoire (banques de données et expériences des analystes) sont indispensables pour le choix des variables, c'est-à-dire pour déterminer quels sont les facteurs à analyser, susceptibles de délivrer des tendances utiles. Elles sont aussi d'une grande utilité pour l'interprétation des données, car permettent de séparer les nouvelles tendances de celles qui sont déjà connues.

Une connaissance technique et méthodologique est indispensable pour choisir une méthode adaptée aux données et au contexte et éviter des erreurs d'analyse pouvant délivrer des patterns inutiles.

Une connaissance de l'affaire traitée, comme par exemple dans l'analyse de données liées au

trafic des stupéfiants est indispensable pour le choix des variables et l'interprétation. Suivant le problème posé par les enquêteurs, les variables à prendre en compte seront différentes. Si un danger lié à un médicament ajouté à un stupéfiant apparaît, une analyse d'où et quand il a été détecté et avec quels autres produits pourrait être faite. Les variables seront entre autres géographiques et temporelles. Une connaissance de l'affaire est aussi indispensable pour interpréter les patterns, notamment pour ceux dont la pertinence est à évaluer. Une collaboration avec les enquêteurs permettrait de disposer d'informations permettant des les contextualiser et ainsi optimiserait leur interprétation.

2.6.3 Un travail d'équipe

Le traitement de données issues de divers types de criminalité nécessitera probablement des connaissances *a priori* provenant de compétences différentes : expérience, enquêtes, techniques.

Une coordination des données et des compétences permettrait de disposer de moyens nécessaires à une analyse dont le but est la production de renseignements. Elle favoriserait une vue d'ensemble de la problématique, ce qui est important pour éviter que les connaissances *a priori* ne jouent un rôle d'oeillères, influençant l'analyste vers ce qu'il s'attend à trouver, plutôt que d'appréhender les résultats de manière plus "naïve". Sans une vision globale, obtenue au moyen de banques de données, de l'expérience de tendances passées et de celle d'enquêteurs, il sera difficile à l'analyste de contextualiser correctement les résultats. Son interprétation risque d'être faussée, orientée uniquement par ses seules connaissances *a priori*.

Coordonner compétences et données permettrait de mettre les forces présentes là où elles sont nécessaires, afin d'interpréter au mieux l'environnement criminel. C'est dans ce sens que les unités d'analyse de certaines Polices Cantoniales semblent se diriger.

2.7 Utilisation de ces techniques

2.7.1 Les champs d'application

Dans une perspective de renseignements, ces techniques devraient être applicables à tout type de données à caractère criminel. Il est possible que ce soit le cas, mais le manque d'applications pratiques ne permet pas de l'affirmer. L'application de ce travail a porté sur des données liées au trafic des stupéfiants, et l'expérience pratique réalisée dans une unité d'analyse a permis de faire des tests sur des données de délits et de technologies de localisation géographique.

2.7.2 Sélection des données tests

Un jeu de données issues du trafic de l'héroïne et de la cocaïne provient de l'analyse chimique des échantillons saisis lors d'opérations de Police dans plusieurs cantons suisses. Ces données ont été choisies pour tester des méthodes de data mining, car étaient, d'une part disponibles (stockées dans une banque de données à l'Institut de Police Scientifique) et d'autre part de format homogène [Esseiva, 2004; Ioset *et al.*, 2005].

Peu d'information est à disposition concernant la production de cocaïne et d'héroïne en amont du trafic observable dans nos régions. Les connaissances *a priori* portent sur le phénomène général du trafic des stupéfiants et sur les événements locaux, traités par des enquêteurs. Le trafic est différent d'un canton à l'autre. Cette propriété rend ce jeu de données très intéressant car tout pattern détecté peut refléter des régularités liées au trafic ou au mode de fabrication et de diffusion des stupéfiants. Une évaluation des tendances détectées avec l'aide de spécialistes permettra de déterminer leur pertinence et peut-être de les valider.

2.7.3 Sélection des méthodes d'analyse

Leur choix dépend, entre autres, du problème à traiter, de l'homogénéité des données et de leur nature. Des intervalles de temps ou des informations de relations entre des personnes seront traités avec des méthodes différentes que pour l'analyse de données chimiques quantitatives, par exemple.

Le but est de pouvoir disposer de méthodes permettant de produire de l'information, susceptible de devenir du renseignement. Il serait intéressant de savoir si il est possible d'élaborer une "boîte à outils" adaptée aux données liées à la criminalité, ou si il est nécessaire de choisir une méthode au cas par cas.

2.7.4 Un exemple de limite à leur application

Un exemple extrême des limites du data mining dans le domaine de la sécurité est le "Total Information Awareness" développé par l'Information Awareness Office (IAO) aux Etats-Unis, suite aux attentats du 11 septembre, afin de maîtriser les informations qui pourraient relever du terrorisme et permettre des interventions plus rapides [Brodeur et Leman-Langlois, 2004]. Les sources d'informations sont nombreuses et de toutes sortes (transactions bancaires, conversations téléphoniques, données basées sur les nationalités, cours de pilotages). Les dimensions prises en compte sont très nombreuses : dimensions temporelle, spatiale, relationnelle, sociologique, linguistique... Ces données contiennent non seulement beaucoup d'informations relatives à des activités courantes et légales mais sont aussi extrêmement hétérogènes. L'ambition de ce projet amène à douter de son efficacité ce qui a pour effet de discréditer ces méthodes qui pourtant sont appliquées

avec succès dans d'autres domaines. Le projet « Total Information Awareness » a été abandonné, mais des composantes de ce projet subsistent dans des programmes de recherche parallèles [Brodeur et Leman-Langlois, 2004].

Les chapitres suivants présentent une description de l'application du data mining dans une optique de renseignement ainsi que des possibilités de mise en pratique et des exemples d'application. Les divers types de données à disposition d'une Police seront décrits ainsi que les méthodes de data mining pouvant leur être associées.

Chapitre 3

La reconnaissance de patterns en analyse criminelle

3.1 Introduction

L'évolution de la criminalité et de la technologie, la croissance des banques de données, nécessitent d'adapter l'analyse avec les moyens d'aujourd'hui, tels que le data mining. Celui-ci permettrait d'aider les analystes à détecter des schémas inconnus, cachés dans des banques de données, en complément de leur expérience.

Beaucoup de publications sur le data mining utilisé dans le domaine de l'analyse criminelle montrent qu'il y a un intérêt croissant pour ces méthodes [McCue, 2006; Terrettaz-Zufferey *et al.*, 2006] et des conférences internationales traitant de ce sujet se mettent en place ("Mining Massive Data-sets for Security" organisée par le North Atlantic Treaty Organisation, <http://mmdss.jrc.it/>, dernier accès le 6 octobre 2008). Ces méthodes ont été appliquées notamment dans le domaine de la lutte contre le terrorisme [Stevens, 2003; Chen *et al.*, 2003] ou la détection et la prédiction de crimes au moyen d'analyses géographiques [Nath, 2006]. Paradoxalement, il semble qu'il y ait peu d'applications et d'exemples d'une utilisation routinière dans des structures de Police (en excluant le domaine bancaire et de téléphonie mobile). Il est possible qu'une technologie adaptée au domaine de l'analyse criminelle existe et puisse être utilisée, mais que son champ d'application ne soit pas bien délimité. Le transfert de technologies à partir de recherches scientifiques vers une application pratique est délicat et demande une adaptation des structures pour être efficace, tant au niveau des futurs utilisateurs et bénéficiaires que de la part des développeurs.

Afin de ne pas reproduire l'exemple inapplicable en l'état du Total Information Awareness [Brodeur et Lemay-Langlois, 2004], une réflexion sur l'applicabilité de ces méthodes dans le domaine pratique

est nécessaire. Où dans un processus de renseignement pourraient-elles s'intégrer, quelles pourraient être les données à disposition, les techniques adaptées, comment utiliser ces nouvelles informations, comment les faire interagir avec l'information opérationnelle et stratégique déjà connue, qui doit manipuler ces techniques ?

3.2 Reconnaissance de patterns dans le processus du renseignement

3.2.1 Schématisation du processus

La figure 3.1 schématise un processus d'analyse criminelle dans le but de produire du renseignement, tel que décrit dans [Ribaux et Margot, 2003]. Le but du modèle est de traiter l'information en sélectionnant des jeux de données dans lesquels des patterns auraient le plus de chance d'être détectés et d'en alimenter une mémoire. La figure 3.2 montre, selon le schéma de la figure 3.1 et de manière plus détaillée, où ces méthodes ont été placées. Son but est de cadrer là où elles pourraient intervenir et avec quelles autres informations elles pourraient interagir.

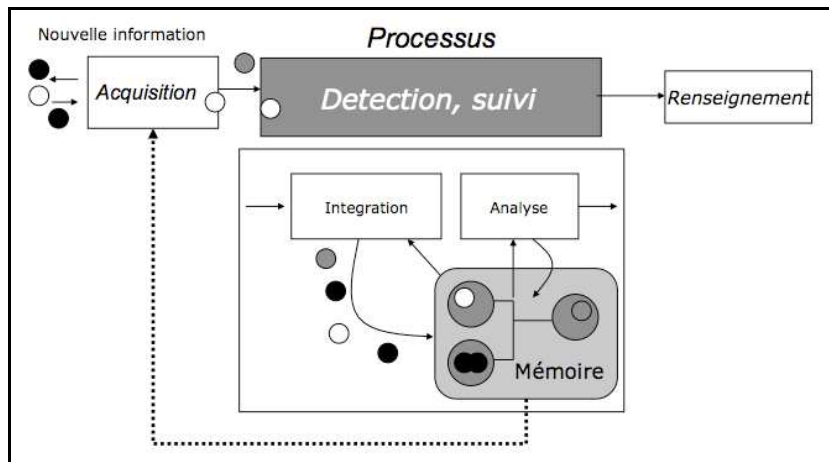


FIGURE 3.1 – Processus d'analyse criminelle tiré du (cours "Renseignement forensique", O. Ribaux 2008).

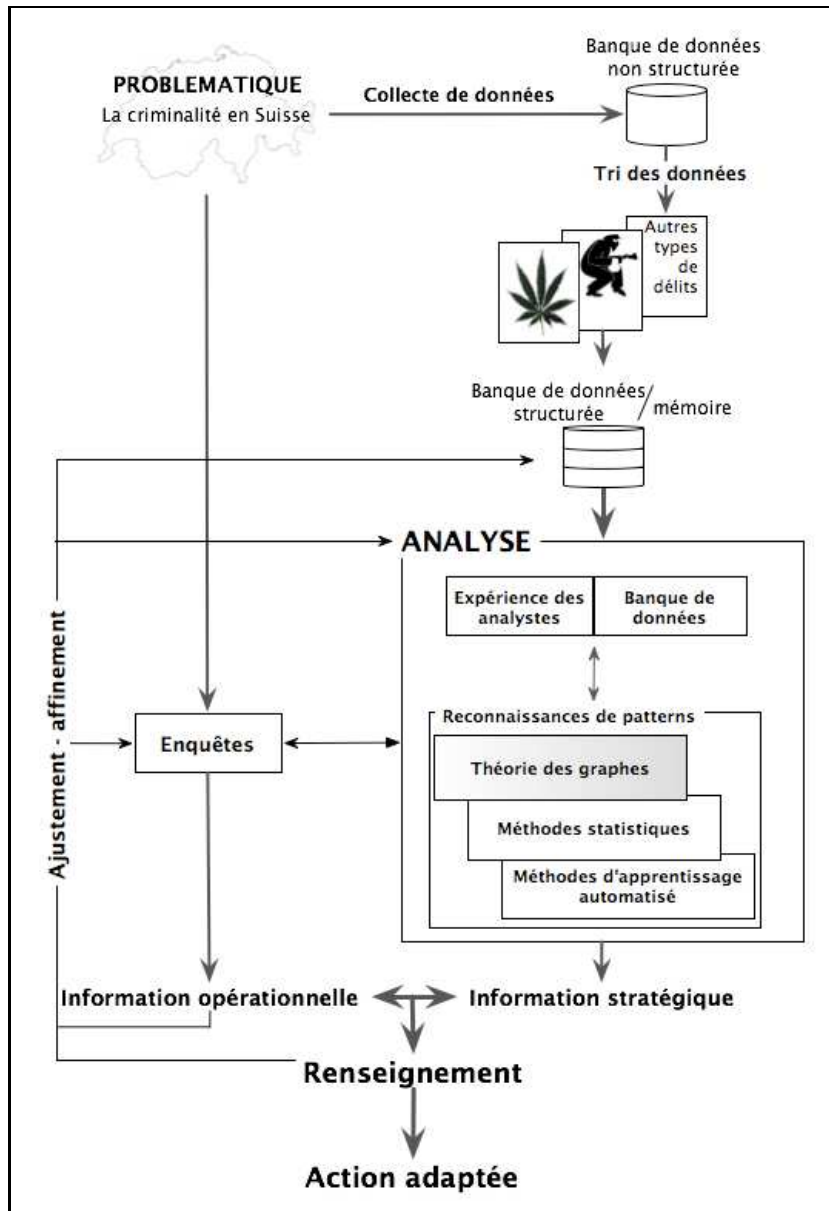


FIGURE 3.2 – Possibilité d'intégration des méthodes de reconnaissance de patterns dans le processus du renseignement.

Des événements liés à la criminalité sont enregistrés en continu dans des banques de données. Ils sont ensuite triés et stockés dans des banques structurées ou mémoire, en fonction des types de délits. Cette mémoire est composée de l'ensemble des expériences connues. Ces données sont traitées par une unité d'analyse, dont le but est de mettre en évidence un nouveau phénomène ou une nouvelle série. Un phénomène est un recoupement d'activités délictueuses basé sur un contexte et un mécanisme utilisés par les auteurs pour commettre leurs délits [Ribaux *et al.*, 2003a]. Une série est une répétition d'un type de délits, pouvant être perpétré par un auteur ou un groupe d'auteurs.

Les événements composant une série peuvent avoir une caractéristique commune, comme un *modus operandi* par exemple, qui pourrait permettre de la détecter. Les analystes utilisent leur expérience (leur mémoire d'une série dont les caractéristiques pourraient correspondre à un nouveau cas), des recherches dans les banques de données ainsi que des statistiques comparatives. L'information produite est en grande partie stratégique, permettant de prendre des décisions afin d'avoir un impact sur la criminalité. Parallèlement, des informations provenant d'enquêtes peuvent amener à la validation de certaines tendances observées. Par leurs questions et leurs connaissances *a priori*, les enquêteurs amènent matière à réflexion sur l'environnement criminel du moment.

Les méthodes de reconnaissance de pattern pourraient être un outil au même titre que l'expérience et la mémoire des analystes, permettant de produire de l'informations stratégique et opérationnelle. L'interprétation des patterns se ferait au moyen des connaissances *a priori* des analystes et des enquêteurs, de même que sa validation, si elle est possible. L'information produite, si elle s'avère utile, deviendrait du renseignement.

3.3 Mise en pratique dans une optique de renseignement

3.3.1 Structure des données

Selon [Ribaux, 2008], la structure des données générées par la criminalité et à disposition des Polices peut être divisée en deux catégories :

- Les données **nominales**, composées de personnes, d'objets (récidivistes, objets ou véhicules volés) ainsi que de leurs relations. Elles sont obtenues au moyen d'enquêtes ou de surveillance, comme par exemple des listes d'appels téléphoniques, ou des personnes correspondant à un certain profil.
- Les données **provenant de traces résultant d'une activité criminelle**, peuvent être des traces physiques ou récoltées sur une scène de crime, des informations provenant de témoins ou des *modus operandi* par exemple.

Cette distinction influence les méthodes d'analyse pouvant être utilisées pour les traiter.

Les données nominales sont caractérisées par des informations de **relations** entre personnes ou objets obtenues lors d'enquêtes. Elles ne sont pas quantitatives et ne reflètent pas l'entièreté de la criminalité, car dépendantes du travail de la Police. Elles relèvent en partie des moyens mis en oeuvre pour les obtenir ainsi que de contraintes d'horaires, d'environnement et de charge de travail (vacances, maladie, évènements particuliers nécessitant les forces de Police pour une période). L'analyse de réseaux sociaux (ARS) traite des relations entre des personnes ou des biens, afin de réaliser un modèle permettant de mettre en évidence des entités-clefs. Des données liées au crime organisé et au terrorisme ont notamment été traitées par des techniques issues de l'ARS [Morselli

et al., 2007].

Les données provenant de traces résultant d'activités criminelles sont, par exemple pour un cambriolage, enregistrées suite à une plainte et les traces récoltées sur le lieu du délit viennent ensuite les compléter. Il est alors possible d'avoir une représentation de la quantité de cambriolages survenus en une année, par exemple. Des méthodes statistiques d'analyse permettraient de mettre en évidence des tendances temporelles, spatiales, des régularités ou une augmentation anormale d'un type de délits.

Des données intermédiaires à ces deux catégories existent, telles que les résultats d'analyse de saisies de stupéfiants. Elles ne proviennent pas de traces relevées sur une scène de crime, mais de saisies réalisées lors d'opérations de Police. Le nombre de saisies réalisées par la Police correspond à une petite partie de la quantité de stupéfiants en circulation et une statistique sur le nombre de saisies ne montrerait pas une image réelle de la situation de ce type de délit.

Ces données sont liées à des personnes (dealers, consommateurs) et contiennent des informations chimiques de la substance saisie. Leur analyse ne permet pas d'identifier un dealer, mais pourrait mettre en évidence des tendances, telles que des similitudes entre substances [Esseiva *et al.*, 2003; Ioset *et al.*, 2005].

La théorie des graphes regroupe des méthodes qui conviendraient à des données intermédiaires aux deux catégories, comme celles issues du trafic des stupéfiants. Ces méthodes permettent de traiter les relations entre tout type d'entité. Les techniques utilisées par l'analyse de réseaux sociaux sont issues de la théorie des graphes mais ont été développées pour réaliser des modèles sociaux en se basant sur des interactions entre personnes. La théorie des graphes permet de s'abstraire du contexte de l'ARS, et ainsi d'élargir son champ d'application. Elle est décrite au chapitre 4.

3.3.2 Interprétation des résultats

Elle constitue l'étape permettant d'évaluer la pertinence des tendances détectées (voir section 2.5.2). La figure 3.3 propose une manière de réaliser cette étape.

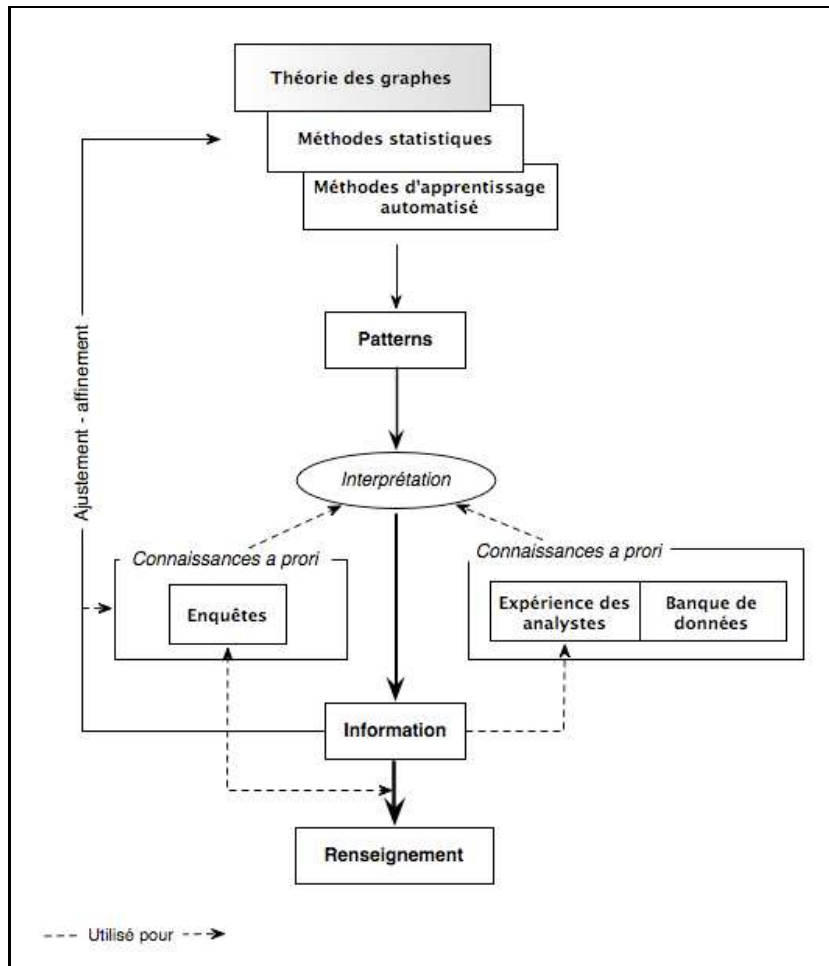


FIGURE 3.3 – Schématisation de l'étape d'interprétation des tendances.

Placer les patterns dans le contexte de l'analyse, évaluer leur pertinence au moyen des connaissances *a priori* et mettre en évidence de l'information susceptible de produire du renseignement, permet, par exemple, de déterminer si un phénomène ou une série apparaît ou réapparaît, si la tendance est insignifiante (inutile) ou si il faut réajuster l'analyse en tenant compte d'autres informations *a priori*.

L'aboutissement devrait, dans le cas idéal, aider à décider d'une stratégie policière afin d'influencer le phénomène ou la série, par exemple par le biais de l'enquête ou d'une action. L'information produite, si elle ne permet pas d'opération directe, peut alimenter la mémoire et permettre de comparer d'autres tendances passées ou de faire des croisements d'informations. Par exemple, vérifier si des régions avec un trafic actif de stupéfiants seraient anormalement touchées par des vols ou des cambriolages.

Le recours à l'analyse par reconnaissance de patterns n'aboutira pas toujours à de l'information

utile, mais apporterait probablement des éléments nouveaux, qui mis en relation avec les connaissances *a priori* pourraient être utiles pour réajuster l'analyse. Durant la période de temps que prend la démarche, des événements criminels en lien avec notre analyse se produisent et apportent leur lot d'informations. L'une d'elles peut être l'élément nouveau qui permettra d'obtenir le renseignement nécessaire pour des mesures opérationnelles préventives ou répressives optimales.

3.3.3 Utilisation dans une structure de Police

Une expérience pratique au sein de deux brigades de Police (une unité d'analyse et une section de lutte contre le trafic des stupéfiants), a permis d'observer de quelle manière ces méthodes pourraient être utilisées.

L'unité d'analyse de la Police Cantonale Vaudoise traite des données provenant de rapports saisis par des intervenants sur des scènes de crime ou lors de plaintes. Des relations avec des informations déjà répertoriées peuvent être établies par des tris, des recherches ou des comparaisons. Un premier tri est réalisé à ce niveau, afin de séparer les événements non significatifs pour l'unité d'analyse (chien écrasé par exemple). Suivant un système de codification harmonisé, les données sont ensuite codées dans le but de détecter des phénomènes répétitifs et d'avoir un suivi de la délinquance sérielle. Ce système se base sur la théorie des opportunités [Felson et Cohen, 1980; Felson et Clarke, 1998], qui consiste à considérer à la fois l'auteur et sa cible dans un environnement contraint physiquement. Ces éléments forment ensemble une situation. Pour qu'un cambriolage puisse se réaliser, ces dimensions doivent présenter des caractéristiques très particulières. La manière de procéder de l'auteur ainsi que le gain espéré sont conditionnés par l'environnement et la cible. Les situations sont donc spécifiques et leur description doit permettre de séparer les données relevant d'activités différentes [Ribaux et Birrer, 2008]. Après codification, les données sont stockées dans une banque de données structurée et les analyses s'opèrent en interaction avec celle-ci. Elles sont initiées par des demandes provenant de deux sources : l'une externe provenant d'enquêteurs et d'officiers et la seconde des analystes. Lors de l'étape de tri des nouveaux événements, certains peuvent être un vecteur d'information sur la possible émergence d'une nouvelle série par exemple. Dans ce cas, l'analyse permettrait de la mettre en évidence.

La mise en pratique de la détection de patterns au sein de cette unité d'analyse pourrait intervenir selon le schéma de la figure 3.4.

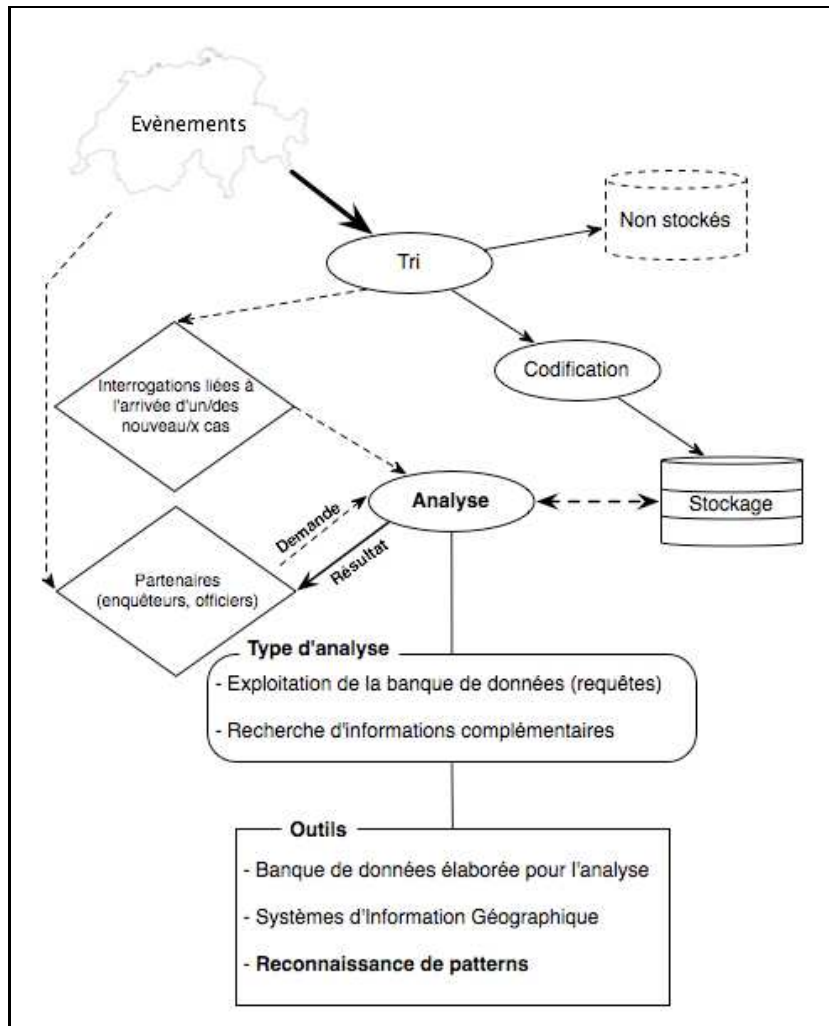


FIGURE 3.4 – Mise en pratique des méthodes de reconnaissance de patterns dans une coordination judiciaire (schéma élaboré en collaboration avec la BAAC de la Police Cantonale Vaudoise).

Pour répondre aux sollicitations venant des enquêteurs, des officiers et des analystes, il existe divers outils, telle que ceux dédiés à la gestion de donnée élaborée pour l'analyse criminelle opérationnelle, les Systèmes d'Information Géographique et les méthodes de reconnaissance de patterns. Ces méthodes pourraient aussi avoir comme fonction de détecter des patterns sans que le point de départ ne vienne d'un analyste ou d'un enquêteur. Le but idéal serait de mettre en évidence de nouvelles tendances (séries ou phénomènes), avant que les analystes ne les aient détectées par une accumulation anormale de cas. Cette application n'a pas été testée, seules les données liées au trafic des stupéfiants ont été traitées dans cette optique.

Une expérience pratique au sein de la section stupéfiants de la Police Cantonale Valaisanne a

été réalisée. Le fonctionnement de cette brigade semble refléter un processus de renseignement. Le travail des enquêteurs est proactif. Des informations sont récoltées afin de pouvoir se représenter l'environnement criminel régional lié au trafic des stupéfiants et ainsi l'influencer par des actions. Cette production de renseignement leur permet de suivre l'évolution de ce type de criminalité.

Dans ce cadre, les méthodes de reconnaissance de patterns ont porté sur des données techniques liées à l'enquête. Des outils ont été développés afin d'exploiter au maximum ces données et de mettre en évidence de l'information inconnue jusqu'alors. Ceci a été fait dans une démarche d'analyse criminelle opérationnelle, en collaboration directe avec un enquêteur de la section.

Chapitre 4

Des techniques prometteuses issues de la théorie des graphes

4.1 Introduction

Le fonctionnement de la grande majorité des méthodes de reconnaissance de patterns est de déterminer la similarité existant entre les données afin de mettre en évidence des groupes, les patterns. Ces méthodes sont classées en deux familles :

1. Les méthodes non supervisées :
 - les méthodes dites exploratoires telles que *l'analyse en composantes principales*
 - Le clustering tel que *le kmeans*
2. Les méthodes supervisées :
 - Les méthodes d'apprentissage automatique (plus connues par leur nom anglais de “machine learning”), telles que *la SIMCA, et les réseaux de neurones*

Les méthodes non supervisées ont pour but d'explorer les données afin de visualiser leur structure (complexité, tendance à former des groupes, etc...) et de déterminer des groupes (le clustering). Les méthodes supervisées permettent, à partir d'une classe connue, de classer les données dans ces dernières de manière automatique. L'ordinateur “apprend” la structure des données afin de pouvoir replacer automatiquement les données dans les catégories initiales au moyen d'algorithmes basés sur les similarités entre les données.

Parmi les méthodes existantes, celles qui apparaissent dans la littérature comme les plus couramment utilisées dans le domaine forensique chimique, sont les méthodes statistiques utilisant

différentes mesures de proximité, telles que l'analyse en composantes principales (ACP) [Pearson, 1901], le clustering ascendant hiérarchique, le kmeans [McQueen, 1967], le Soft Independent Modeling of Class Analogies (SIMCA) [Wold et Sjostrom, 1977] et les réseaux de neurones [Rosenblatt, 1958]. Les méthodes citées ne sont pas exhaustives mais apparaissent dans des travaux liés à l'analyse de données de stupéfiants. L'ACP est une méthode utilisée dans le profilage des stupéfiants [Guéniat et Esseiva, 2002; Guéniat, 2004; Guéniat et Esseiva, 2005; Esseiva, 2004; Esseiva *et al.*, 2002, 2003, 2007; Ioset *et al.*, 2005]. La SIMCA a été utilisée dans [Locicero, 2007] pour aboutir à l'harmonisation de méthodes d'analyses entre deux laboratoires. La méthode a permis de classer automatiquement les données selon des classes basées sur des lots de production de cocaïne. Les réseaux de neurones ont été testés sur des échantillons d'héroïne afin de les classer selon leur provenance géographique [Esseiva, 2002].

Des méthodes de classification automatique ont été testées dans [Madden et Ryder, 2002] et [O'Connell *et al.*, 2005] sur des données chimiques, issues du même type d'appareil que celui analysant les stupéfiants à l'Institut de Police Scientifique (la spectrométrie de masse).

De nombreuses publications existent sur les méthodes de clustering et de classification automatique. Les algorithmes développés sont de plus en plus performants et spécifiques à un domaine. Dans le cadre de ce projet, un algorithme de clustering adapté aux données de stupéfiants a été développé, [Ratle *et al.*, 2006a], [Ratle *et al.*, 2006b], [Ratle *et al.*, 2008].

Les méthodes statistiques ne seront pas détaillées dans ce travail. Elles font partie d'un travail de thèse mené en parallèle de celui-ci par l'auteur des dernières publications citées.

Une dimension des données n'a actuellement pas été exploitée dans le domaine des stupéfiants. Il s'agit de la relation "présence-absence" d'un produit. Les données de type ordinal (vrai-faux, présence-absence) peuvent être traitées au moyen de méthodes combinatoires. Des éléments de la théorie des graphes [Euler, 1759; Berge, 1958] ont été utilisés. Les techniques mises en oeuvre permettent de réaliser de la reconnaissance de patterns dans des données structurées sous forme de graphes. Il est alors possible de donner au processus général de data mining le nom de "graph mining".

Ces méthodes sont présentées dans la section suivante et seront contextualisées selon la problématique du trafic des stupéfiants.

4.2 La théorie des graphes

4.2.1 Introduction

Cette branche des mathématiques voit ses débuts en 1736 par la caractérisation des graphes Eulériens. Euler s'était penché sur le problème des sept ponts de Königsberg¹ [Euler, 1759]. Des notions de graphes ont ensuite été utilisées par Gustave Kirchhoff pour sa loi des circuits. Depuis les années 1950, l'informatique utilise des algorithmes issus de cette théorie. Dès les années 1960, Claude Berge développe la théorie des graphes telle que nous la connaissons aujourd'hui [Berge, 1958].

4.2.2 Notions théoriques

Seules les définitions et notations nécessaires pour la compréhension des méthodes décrites plus loin sont présentées dans cette section.

Soit $G = (E, V)$ un **graphe non-orienté**, dont $V = (v_1, v_2, \dots, v_v)$ est l'ensemble de ses *sommets* ou *noeuds* et $E = (e_1, e_2, \dots, e_e)$ l'ensemble de ses *arêtes*. Les éléments de E d'un **graphe orienté** sont appelés les *arcs* de G (figure 4.1).



FIGURE 4.1 – Graphe simple et graphe orienté.

Soit $A(G) = a_{ij}$, la matrice d'adjacence du graphe G , dont a_{ij} est le nombre d'arêtes ou d'arcs joignant les sommets v_i et v_j .

1. La ville de Königsberg est construite autour de deux îles reliées entre elles par un pont et six ponts relient le continent à l'une ou l'autre des deux îles. Le problème consiste à déterminer s'il existe ou non une promenade dans les rues de Königsberg permettant, à partir d'un point de départ au choix, de passer une et une seule fois par chaque pont, et de revenir à son point de départ, sachant qu'on ne peut traverser la rivière Pregolya qu'en passant sur les ponts.

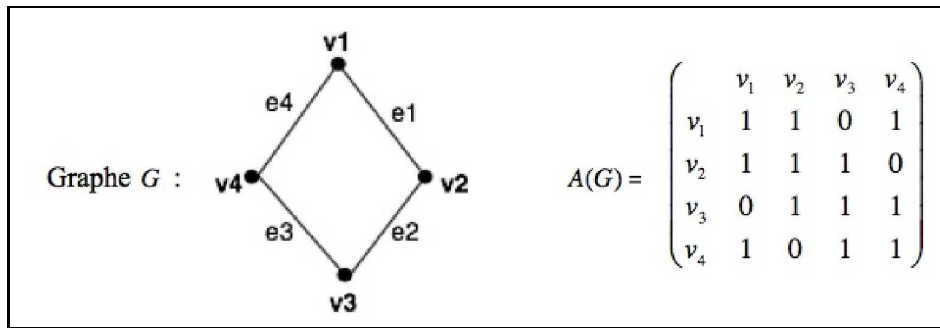


FIGURE 4.2 – Graphe simple et matrice d'adjacence.

Soit $G = (E, V)$ un graphe. G est un **graphe d'intervalle**, si à chaque sommet v de V , peut être associé un intervalle I_v ouvert de l'axe des réels, tel que pour toute paire $u, v \in V$ de sommets, $uv \in E$ si et seulement si $I_u \cap I_v \neq \emptyset$ [Gardi, 2005] (figure 4.3).

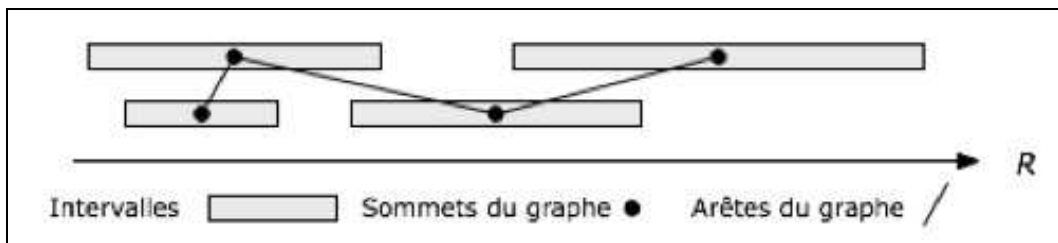


FIGURE 4.3 – Graphe d'intervalle.

Soit $G = (E, V)$ un graphe. G est un **graphe biparti**, si et seulement si l'ensemble des sommets $V(G)$ est l'union de deux ensembles indépendants disjoints, appelés **set biparti** de G . [West, 1996]

Un exemple de graphe simple et biparti est présenté par la figure 4.4. Autrement expliqué :

- **Les graphes simples** : tous les sommets peuvent être potentiellement liés.
- **Les graphes bipartis** : les sommets peuvent être divisés en deux groupes et aucun lien n'existe au sein d'un groupe.

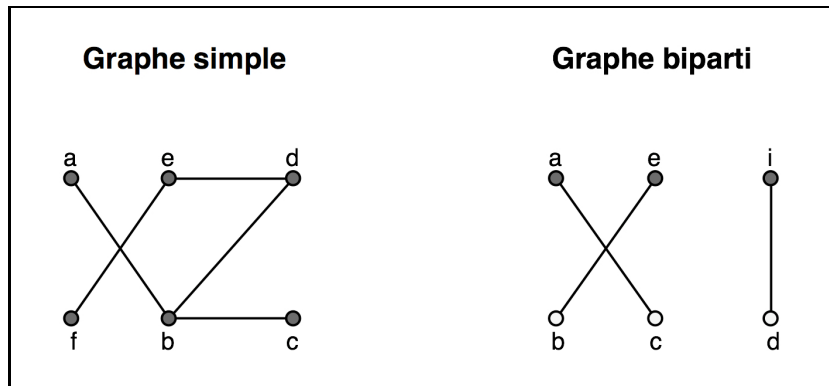


FIGURE 4.4 – Graphes simple et bipartis.

Soit $G = (E, V, w)$, un **graphe pondéré** et $w = (w_1, w_2, \dots, w_n)$, l'ensemble des labels numériques de ses arêtes (figure 4.5).

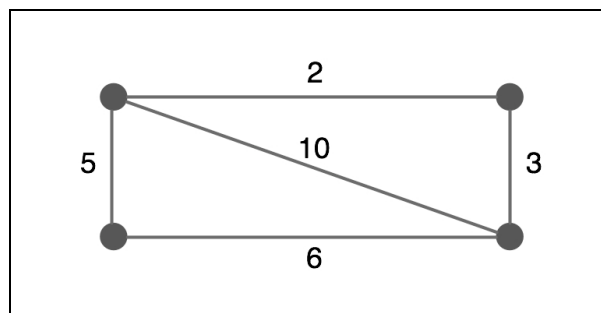


FIGURE 4.5 – Graphe pondéré.

Un **chemin** est un graphe dont les sommets peuvent être ordonnés de manière à ce qu'ils soient adjacents deux à deux si et seulement si ils sont consécutifs dans la liste. Un **cycle** est un graphe avec un nombre égal de sommets et d'arêtes, dont les sommets peuvent être placés autour d'un cercle de manière à ce qu'ils soient adjacents deux à deux si et seulement si ils sont placés de manière consécutive sur le cercle [West, 1996].

Dans l'exemple ci-dessous (figure 4.6), en listant les sommets dans l'ordre x, b, a, z, y , un chemin et un cycle apparaissent. Si une arête d'un cycle est enlevée, un chemin est obtenu.

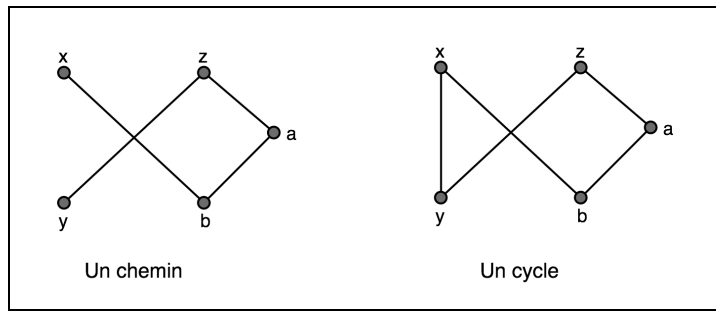


FIGURE 4.6 – Chemin et cycle (tiré de [West, 1996]).

Un **sous-graphe** de G est un graphe H tel que $V(H) \subseteq V(G)$ et $E(H) \subseteq E(G)$ et tel que les sommets des arêtes de H soient les même que dans G . Lorsqu'il est écrit $H \subseteq G$, cela veut dire "G contient H" [West, 1996].

Un graphe est **connexe** si il possède un chemin u, v pour tout sommet $u, v \in V(G)$ (autrement G est non-connexe). Si G a un chemin u, v , alors u est connexe à v dans G . Les relations connexes de $V(G)$ correspondent à des paires ordonnées de sommets (u, v) de manière à ce que u soit connexe à v .

La figure 4.7 montre la différence entre un graphe connexe et non-connexe. Celui de droite ne possède pas de chemin, contrairement à celui de gauche.

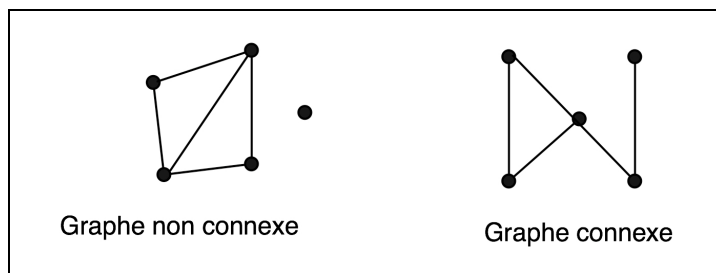


FIGURE 4.7 – Graphe connexe et non-connexe.

Connexe est un adjectif appliqué uniquement aux graphes et aux paires de sommets. Le tableau 4.1 a pour but de clarifier la distinction entre **connexité** et **adjacence** [West, 1996].

TABLE 4.1 – Connexité et adjacence (tiré de [West, 1996])

G contient un chemin u, v	$u, v \in E(G)$
u et v sont connexes	u et v sont adjacents
u est connexe à v	u est lié à v
	u est adjacent à v

Un **sous-graphe connexe maximal** de G est un sous-graphe connexe qui n'est contenu dans aucun autre sous-graphe connexe de G .

Les **composantes** d'un graphe G sont leurs sous-graphes connexes maximaux [West, 1996].

4.3 Techniques issues de la théorie des graphes et leurs applications

4.3.1 Introduction

La théorie des graphes est vaste et de très nombreux algorithmes ont été développés pour des utilisations liées à plusieurs domaines. Des problématiques ont été posées, telle que le chemin le plus court d'un sommet à l'autre dans un graphe pondéré, la détection de groupes dans un graphe complexe basée sur le nombre d'arêtes liées à un sommet, la détection de structures liées à une position particulière de sommets dans un graphe, et bien d'autres, qui ont été résolues par l'implémentation d'algorithmes. Certains ont été développés pour répondre à une problématique spécifique à un domaine, comme par exemple les algorithmes de cliques maximales appliqués à des données paléontologiques pour résoudre des problèmes de chronologies [Guex, 1984] ou appliqués à des problématiques liées à l'analyse de réseaux sociaux [Wasserman *et al.*, 1994].

Dans ce chapitre les méthodes décrites sont les plus prometteuses pour l'analyse de données ordinales liées au trafic des stupéfiants. Un survol de techniques décrites dans la littérature des sciences forensiques et de criminologie y est présenté.

4.3.2 Etat de l'art en matière d'utilisation de graphes dans le domaine du renseignement

L'**analyse de réseaux (ARS)** est issue de la théorie des graphes par la nomenclature et les analyses réalisées. Les réseaux sont des graphes basés sur des relations entre personnes dont les méthodes d'analyse tentent de répondre à des problématiques relatives aux sciences sociales en particulier. [Wasserman *et al.*, 1994].

Ce type d'analyse a été utilisé pour la modélisation de réseaux liés au terrorisme [Marcus *et al.*, 2006] et à des réseaux criminels [Morselli *et al.*, 2007].

L'**algorithme "Subdue"** [Cook *et al.*, 2006] permet une exploration globale de graphes (ainsi que la détection de groupes et des analyses supervisées), en recherchant des sous-structures. Ce sont des entités créant des patterns pouvant être traitées comme des sous-graphes. L'algorithme a été testé sur des données chimiques, génétiques et sur des données liées au terrorisme [Potts *et al.*,

2005]. Une modélisation des informations relationnelles issues de la sécurité avant et après les attentats du 11 septembre a été faite sous forme de graphes et traitée par l’algorithme Subdue. Des tendances structurelles ont été détectées permettant de mettre en évidence des entités initialement invisibles.

Une évaluation des méthodes de data mining appliquées à des données de crime est décrite dans [Chen *et al.*, 2003]. Les cas d’études sont basés sur l’analyse du renseignement suite aux attentats du 11 septembre. Les analyses ont été réalisées sur des données liées au trafic des stupéfiants, au terrorisme et au problème des gangs aux États-Unis. Les techniques présentées sont notamment le Support Vector Machine et le “Criminal Network Analysis” basé sur l’analyse de réseaux sociaux dans le cadre du projet COPLINK (<http://ai.bpa.arizona.edu/research/coplink/index.htm>, dernier accès le 26 février 2009).

L’analyse de liens est très utilisée en analyse criminelle comme outil de réflexion, car elle permet de visualiser tous les liens existant entre entités d’une affaire par exemple. Des notions de distance entre les sommets du graphe peuvent être ajoutées [Shimbo et Ito, 2007].

Le design de graphes regroupe des techniques de mise en forme automatique des entités par l’application d’algorithmes. Il permet de mettre en évidence des patterns grâce à une structuration visuelle optimale du graphe. Ces algorithmes peuvent être adaptés à une problématique, telle qu’une visualisation topologique ou en 3 dimensions [Didimo et Liotta, 2006].

4.3.3 Les algorithmes testés et leurs applications

Méthode d’optimisation

Minimum spanning tree est un calcul d’optimisation appliqué à un graphe pondéré (les arêtes du graphe ont une valeur numérique). Le but est de mettre en évidence les sommets reliés par des arêtes dont les poids sont minimaux (figure 4.8).

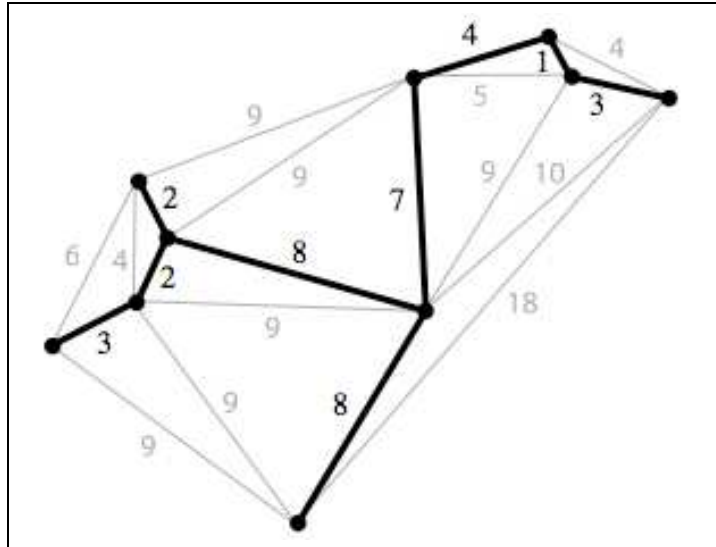


FIGURE 4.8 – Illustration d'un Minimum Spanning Tree mis en évidence sur le graphe initial.

Application : A partir d'un graphe complexe (composé de nombreux sommets et arêtes), un algorithme de minimum spanning tree permettrait de simplifier le graphe en ne mettant en évidence que les liens entre sommets dont la valeur des poids serait minimale.

Détection d'entités particulières

Les points d'articulation : *sommets d'un graphe dont la suppression augmente le nombre de composantes connexes* [West, 1996].

Dans l'exemple de la figure 4.9, il s'agit du sommet ayant le plus de liens. S'il est supprimé, deux composantes connexes apparaissent.

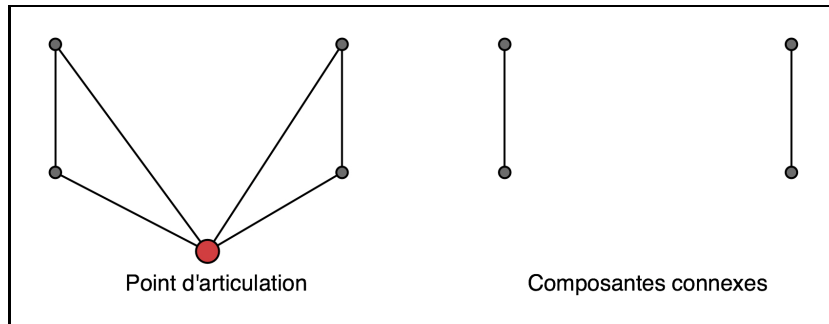


FIGURE 4.9 – Point d'articulation.

Un sommet dont la suppression crée des sommets isolés (sans arêtes) ou en augmente le nombre est aussi un point d'articulation (figure 4.10).

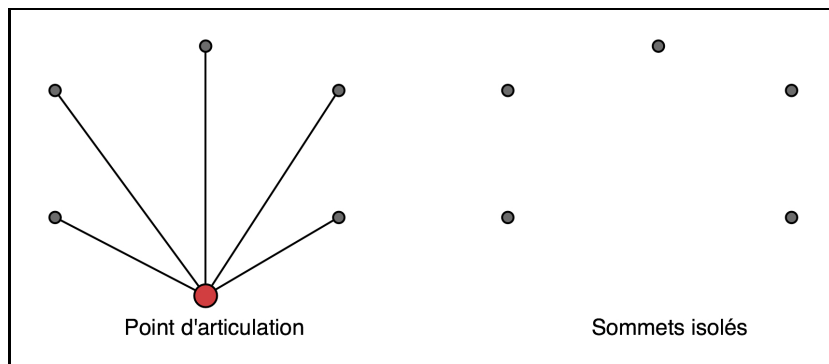


FIGURE 4.10 – Point d'articulation et sommets isolés.

Application : Ils permettent de visualiser les sommets d'un graphe avec le plus de liens, ce qui mettrait en évidence des noeuds ayant une importance pour la structure du graphe.

Méthodes de clustering

La recherche de groupes dans des graphes met en évidence des entités partageant les mêmes relations ou ayant les mêmes propriétés. L'application de ce type d'algorithmes permet de détecter des structures, dont la pertinence dépendra des données et de la problématique.

La détection de sous-groupes par k -cores [Seidman, 1983] se base sur le degré des noeuds (nombre de liens qui lui sont connectés). Un k -core est un sous-graphe dont chaque sommet est adjacent à un minimum de k autres sommets du sous-graphe.

La détection des cliques maximales [Luce et Perry, 1949; Harary *et al.*, 1965] permet de mettre en évidence des groupes. Une clique d'un graphe est un sous-graphe dont les sommets sont adjacents les uns aux autres. La clique est maximale si aucun autre noeud du graphe n'est adjacent à ceux de la clique (figure 4.11). Il existe divers algorithmes de calcul. Celui testé dans ce travail est l'algorithme de Bron et Kerbosch [Bron et Kerbosch, 1973].

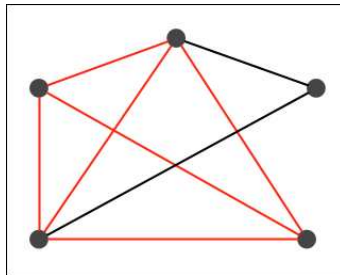


FIGURE 4.11 – En rouge, la clique maximale : aucun autre sommet ne peut être ajouté.

Le modèle des blocs [White *et al.*, 1976] a beaucoup été développé par l'analyse de réseaux sociaux (ARS), qui propose plusieurs algorithmes adaptés aux problématiques des sciences sociales. La méthode de base permet de décomposer un graphe en sous-composantes maximales biconnexes : les blocs. Un graphe connexe possède des arêtes reliant chaque sommet, donc possède un chemin entre ses sommets. Un graphe est biconnexe si en lui supprimant un sommet, il devient connexe (figure 4.12).

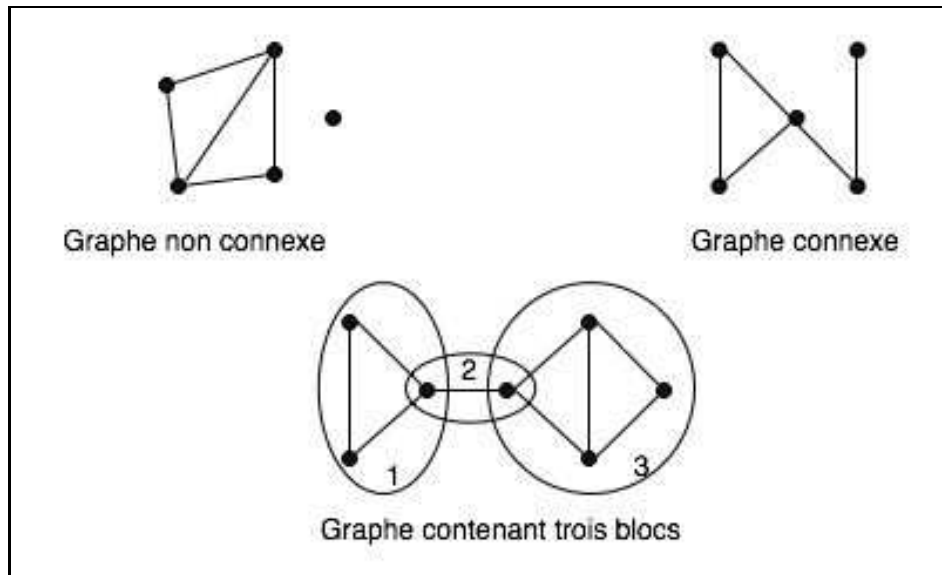


FIGURE 4.12 – Graphe non connexe, connexe et blocs.

Certains algorithmes utilisant le modèle des blocs appliquent des permutations de lignes et de colonnes sur la matrice d’adjacence afin de mettre en évidence des zones de la matrice contenant un agencement significatif de “1”, indiquant des relations particulières entre entités. Ces blocs peuvent être une absence de “1”, ou au contraire une présence exclusive ou des symétries. Le jeu “rubicscube” est une illustration de cette technique. Des plaques sont permutées afin d’obtenir des faces de couleur unifiée. Le jeu étant en trois dimensions, les plaques correspondent aux lignes et aux colonnes d’une matrice d’adjacence.

Un graphe biparti est, par définition, constitué de composantes connexes. L’application du modèle des blocs sur ce type de graphe sera inutile. Il existe cependant d’autres méthodes de détection de groupes dans ce type de graphes, telles que la détection de communautés.

Les algorithmes de “communauté” [Rupnik, 2006], [Clauset *et al.*, 2004], [Pons et Latapy, 2006] traitent des graphes constitués de groupes distincts plutôt que d’une masse de sommets. Ces groupes sont séparés par peu de liens, alors qu’à l’intérieur, les sommets sont reliés par de nombreuses arêtes. De tels groupes sont appelés “communautés” (figure 4.13). Le terme de “communauté” est une notion à caractère social, mais les techniques s’appliquent à n’importe quel type de graphe et en particulier à de grands graphes.

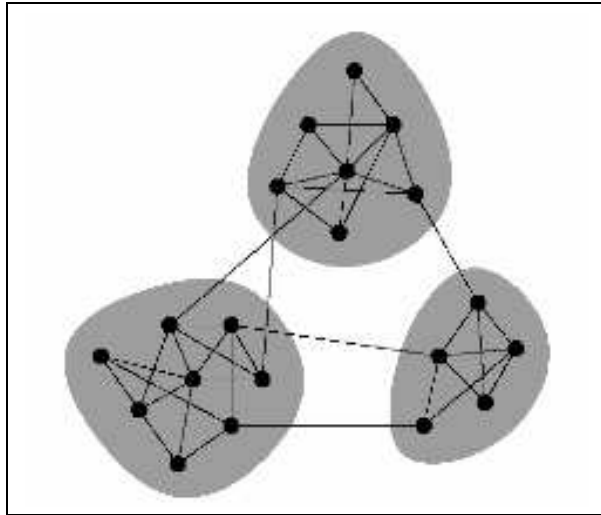


FIGURE 4.13 – Communautés dans un graphe [Rupnik, 2006].

Deux algorithmes ont été testés. Le premier, appelé “walktrap” [Pons et Latapy, 2006], est basé sur une probabilité proportionnelle aux poids des arêtes et au degré des sommets voisins. Le second, appelé “fast greedy” [Clauset *et al.*, 2004] se base sur une décomposition hiérarchique des arêtes du graphes pour mettre en évidence les communautés.

L’algorithme “Subdue” [Cook *et al.*, 2006] permet d’étudier la structure d’un graphe et de détecter des clusters. A partir des sous-graphes principaux, l’algorithme recherche les plus petits possibles. Les groupes détectés correspondent aux plus petits sous-graphes.

4.3.4 Evaluation

Comme décrit à la section 3.3.1, la théorie des graphes semble bien s’adapter à la structure des données nominales et liées à une activité criminelle [Ribaux, 2008]. Les sommets correspondraient à des entités liées à l’enquête, telles que types de délits, lieux, dates, personnes, objets et événements, qui seraient mis en relation par les arêtes du graphe. Le modèle ainsi créé pourrait être exploité par des techniques de graph mining.

4.4 Les outils

Il existe plusieurs outils permettant de réaliser des traitements par théorie des graphes et analyse de réseaux. Ils sont divisés en deux catégories : les logiciels avec interface graphique (GUI) et les bibliothèques. Une “bibliothèque” est une sorte de bibliothèque de fonctions codées dans un langage

de programmation, permettant de simplifier des calculs relatifs à un domaine. Par exemple, une librairie de géométrie contiendra des fonctions de calculs géométriques.

4.4.1 Les logiciels avec GUI

- Le logiciel **Past** [Hammer *et al.*, 2001] (<http://folk.uio.no/ohammer/past/>), développé pour traiter des données et des problématiques paléontologiques, permet d’appliquer des méthodes de statistiques exploratoires, de data mining et de réaliser des analyses basées sur des graphes d’intervalle.
- Les logiciels **UCINET**[©] [MacEvoy et Freeman, 2000] (<http://www.analytictech.com/>) et **Pajek** [Batagelj et Mrvar, 2003, 1998] (<http://vlado.fmf.uni-lj.si/pub/networks/pajek/>) sont des outils performants d’analyse de réseaux et de graphes, cités en référence dans de nombreuses publications.
- Le logiciel **Graphviz** [Low, 2004] (<http://www.graphviz.org/>), permet d’appliquer des algorithmes de mise en page performants, permettant une interprétation optimale.
- Le logiciel **Analyst’s Notebook**[©] de la suite I2[©], est un outil puissant de représentation de graphes complexes et de schémas relationnels. Quelques algorithmes d’analyse, tels que le plus court chemin et la détection de cliques sont implémentés.

4.4.2 Les librairies

L’utilisation d’un langage de programmation permet de réaliser des applications “maison” adaptées aux problématiques traitées. Le paramétrage des fonctions utilisées est souvent plus fin et peut être calibré pour le type de données traitées.

Le langage de programmation Python [Van Rossum, 1990] (<http://www.python.org/>) a été utilisé ainsi que les librairies suivantes :

- **Networkx** [Hagberg *et al.*, 2004] (<http://networkx.lanl.gov>).
- **IGraph** [Csárdi et Nepusz, 2003] (<http://cneurocvs.rmki.kfki.hu/igraph/>).

Chapitre 5

Application aux données liées au trafic des stupéfiants

5.1 Introduction

Cette application a pour but de tester les méthodes décrites précédemment et d'évaluer la pertinence des résultats dans une optique de renseignement. Elle va tenter de mettre en évidence des tendances cachées dans un jeu de données. Ces méthodes novatrices pourraient s'intégrer dans le processus du profilage qui sera décrit dans la section 5.2.4, afin d'apporter des informations complémentaires pouvant alimenter le renseignement.

Les données proviennent de l'analyse d'échantillons de cocaïne et d'héroïne et seront décrites à la section 5.2.5.

Des hypothèses peuvent être établies :

1. Le phénomène du coupage lié au trafic des stupéfiants évolue dans le temps.
2. Une modélisation globale d'échantillons saisis sur la base de leur contenu en produits de coupage et selon des dimensions spatiales et temporelles, peut révéler des tendances utiles sur le trafic des stupéfiants d'une échelle nationale à régionale.
3. L'analyse des relations existant entre les caractéristiques physico-chimiques associées aux informations circonstancielles (temps et géographie) des saisies, permet de produire du renseignement opérationnel.

Afin de le comprendre, une description du phénomène basée sur le rapport mondial des drogues 2006 de l'ONU [ONUDD, 2006] est présentée dans la section suivante. Ce rapport a été utilisé car il correspond à la période où les données ont été traitées et semble mieux refléter la situation du

trafic des stupéfiants lors de l'analyse des données que celui de 2008 [UNODC, 2008].

5.2 Eléments sur le trafic des stupéfiants

5.2.1 Généralités

En Suisse, l'activité des réseaux de distribution s'est modifiée en fonction des lois fédérales et cantonales en vigueur. Les activités de la Police ayant pour but de limiter l'étendue du phénomène et protéger de nouveaux consommateurs potentiels, ont contribué à lutter contre ce trafic.

Depuis la fin des années 1990, le prix de la cocaïne a chuté jusqu'à atteindre le prix de l'héroïne [ONUDC, 2006; OFSP, 2006]. La consommation de ce produit s'est intensifiée et donc aussi sa vente. Les enquêtes de Police ont montré que le marché des stupéfiants est alimenté principalement par des ressortissants des Balkans pour l'héroïne et d'Afrique pour la cocaïne. Ce phénomène est très probablement lié au fait que des axes principaux de transit passent par ces régions, facilitant l'embauche de dealers.

5.2.2 Le trafic de l'héroïne

La morphine, une composante de l'opium, est transformée en diacétylmorphine qui est l'héroïne [Guéniat et Esseiva, 2005]. L'opium brut est tiré de la capsule du pavot (*Papaver Somniferum*), cultivé pour la production d'héroïne principalement en Afghanistan, au Pakistan (Croissant d'Or) et en Birmanie, au Laos et en Thaïlande (Triangle d'Or).

Des laboratoires situés principalement en Afghanistan et en Turquie transformaient l'opium en héroïne. Cette observation semble s'être modifiée, car d'après le rapport mondial sur les drogues 2006 de l'ONU, la majorité des démantèlements de laboratoires clandestins s'est opérée en Russie, en Moldavie et en Afghanistan. Ce changement est corroboré par le rapport mondial sur les drogues 2008 [UNODC, 2008], qui décrit une intensification de l'utilisation des routes passant par les ex-républiques d'Asie Centrale et la Russie. En 2004, un seul laboratoire a été démantelé en Turquie. Les événements politiques survenus en Afghanistan ont influencé la production et le marché de l'héroïne. L'interdiction de cultiver du pavot sur terre afghane, décrétée en 2000, s'est ressentie par une diminution des saisies sur les routes de transit et dans les pays consommateurs. Une reprise des cultures s'est produite en 2003 et une augmentation des saisies, notamment en Asie Centrale, s'est ressentie.

Selon le rapport mondial sur les drogues de l'ONU, il y a principalement deux routes partant de l'Afghanistan en direction de l'Europe :

- Les routes du Nord par l'Asie Centrale
- Les routes du Sud par l'Iran

Toujours selon ce rapport, la plus grande partie de l'héroïne transitant par l'Asie Centrale est destinée à la Russie. Seuls 15% seront acheminés vers les pays d'Europe occidentale. En 2005, la culture du pavot au nord de l'Afghanistan a fortement diminué, ce qui a eu pour conséquence de ralentir l'exportation par la route passant en Asie Centrale. L'essentiel des exportations d'Afghanistan passe par l'Iran.

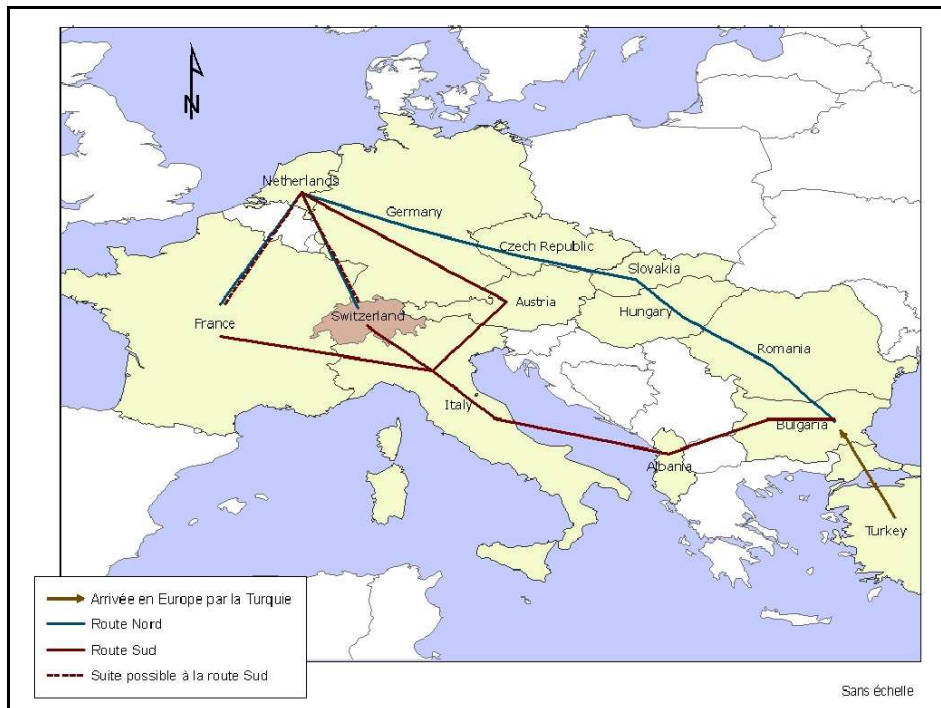


FIGURE 5.1 – Principales voies d'entrée en Europe de l'héroïne (d'après [ONU DC, 2006]).

L'entrée en Europe de l'Est pour atteindre l'Europe occidentale se fait par deux voies principales. Elles sont représentées sur la carte de la figure 5.1 qui a été réalisée selon les informations tirées du rapport mondial sur les drogues de l'ONU. Depuis 2004, le trafic passant par la voie Sud passant par l'Albanie puis l'Italie au moyen de camions chargés sur des ferries a augmenté. Cette route passe essentiellement par voie maritime (Adriatique) pour atteindre l'Italie.

5.2.3 Le trafic de la cocaïne

La cocaïne est extraite de la feuille de coca (*Erythroxylum Coca*), plante poussant principalement dans les régions Andines d'Amérique du Sud. Le premier producteur de coca est la Colombie, où la culture est autorisée pour un usage traditionnel (mastication des feuilles de coca, tisane). En Bolivie, la culture est autorisée sur un territoire limité, celui des Yungas près de La Paz. Le Pérou est un autre producteur important, mais la culture n'y est pas autorisée. La transformation

en cocaïne à partir de la feuille de coca est réalisée dans des laboratoires clandestins, localisés dans la plupart des pays d'Amérique du Sud [Guéniat et Esseiva, 2005]. La cocaïne est exportée depuis l'Amérique du Sud vers les Etats-Unis et l'Europe selon différentes routes. D'après le rapport mondial sur les drogues 2006 de l'ONU et schématisé sur la carte (figure 5.2), plusieurs routes sont suivies pour acheminer ce stupéfiant en Europe :

- Transit par les Caraïbes (et les Antilles Néerlandaises) puis vers l'Europe par la Hollande, l'Espagne et le Portugal.
- Transit par l'Afrique, particulièrement le Ghana, le Kenya et le Nigeria en direction de l'Afrique du Sud et de l'Europe.
- Transit par les îles du Cap Vert en direction de l'Europe par le Portugal et l'Espagne.
- Exportation directe vers l'Europe par le Portugal et l'Espagne.

Selon le rapport de l'ONU, le pays comptant le plus grand nombre de saisies de cocaïne est la Colombie, suivie des Etats-Unis et de l'Espagne. Une augmentation significative des saisies réalisées dans les pays d'Europe, montre que les organisations de trafiquants ciblent de plus en plus les pays européens. Les pays ayant signalé les plus grosses saisies en 2004 sont par ordre décroissant, l'Espagne, les Pays-Bas et le Portugal.

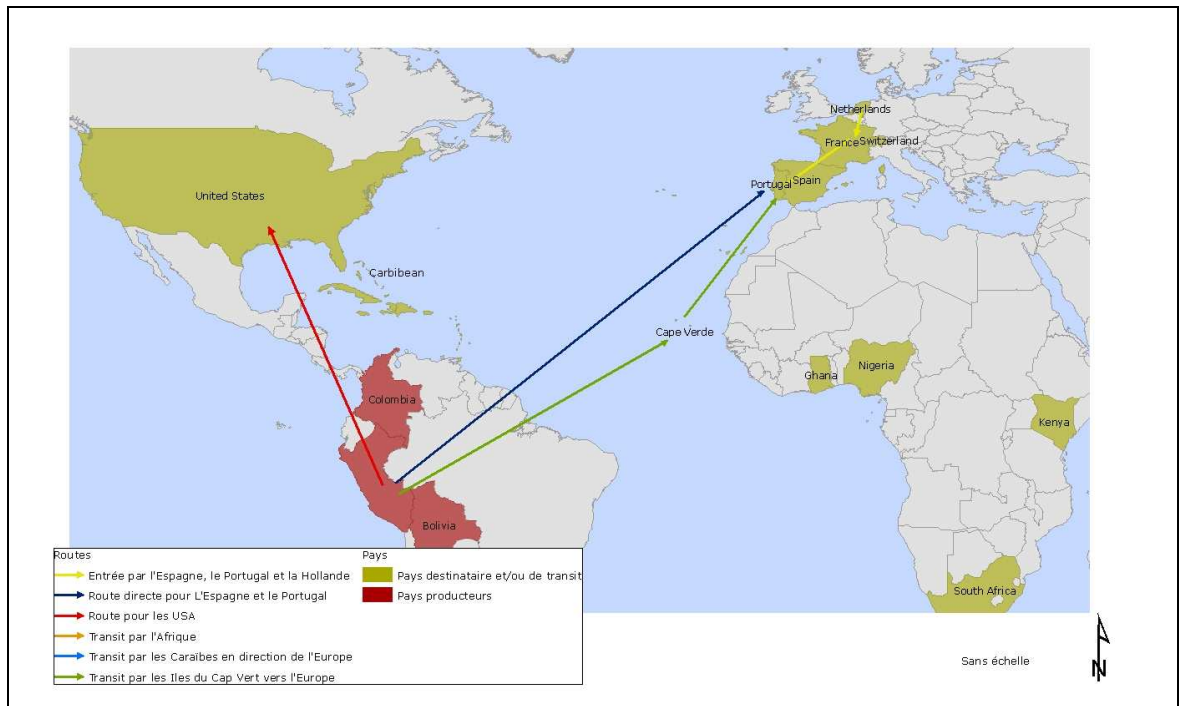


FIGURE 5.2 – Principales voies d'entrée en Europe de la cocaïne (d'après [ONU DC, 2006]).

5.2.4 Le profilage des stupéfiants (éléments tirés de [Ioset *et al.*, 2005])

Introduction

Un processus de renseignement basé sur des saisies de stupéfiants a été mis au point à l'Institut de Police Scientifique (IPS) de l'Université de Lausanne et permet de réaliser un profilage de plusieurs types de stupéfiants saisis en Suisse.

Une systématique a été mise en place depuis l'échantillonnage jusqu'à l'interprétation des résultats. Cette démarche a pour but de pouvoir apporter du renseignement opérationnel et stratégique aux services de Police luttant contre le trafic des stupéfiants. Elle permet aussi une veille du phénomène et de l'émergence de nouvelles tendances au niveau des stupéfiants circulant en Suisse. Ceci permet d'étendre ses compétences aux questions posées par le domaine de la santé.

Le schéma général est présenté par la figure 5.3 .

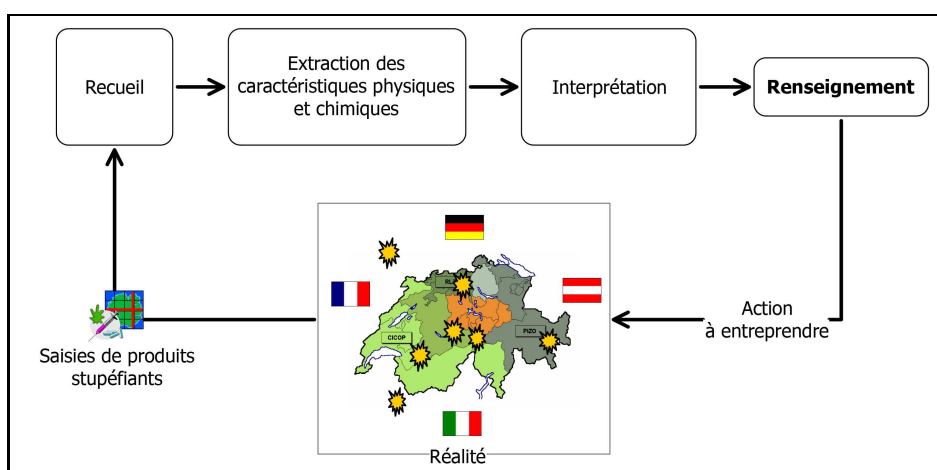


FIGURE 5.3 – Processus de renseignement basé sur des saisies de stupéfiants en Suisse (tiré de [Ioset *et al.*, 2005]).

Description du processus

Les saisies réalisées dans les cantons de Genève, Vaud, Neuchâtel, Jura et Tessin, par les Polices Cantonales, la Police Judiciaire Fédérale et la Police communale de Lausanne sont analysées par l'IPS. Malgré le fait que toutes les saisies de stupéfiants réalisées en Suisse ne soient pas analysées au même endroit, les données issues de l'IPS sont suffisamment représentatives pour être exploitées à des fins de renseignement opérationnel.

Les échantillons sont transmis par l'enquêteur chargé de l'affaire ou le service d'identité judiciaire de la Police concernée. Des informations circonstanciées de dates et lieux de saisies sont enregistrées lorsque c'est possible. Ces données sont stockées dans une banque de données et seront

complétées après l'étape d'analyse.

L'échantillonnage est réalisé par un expert issu du laboratoire de l'IPS ou d'un service d'identité judiciaire, afin de garantir l'homogénéité de l'échantillon. Le taux de pureté (proportion de substance active par rapport à la quantité totale) est nécessaire pour déterminer la gravité du délit et l'échantillonnage permet de disposer de constituants chimiques et physiques représentatifs de l'ensemble de l'échantillon et pas seulement de la zone de prélèvement. Des données de base valables permettront une analyse optimale.

Les variables chimiques sont extraites de l'échantillon. Elles ne correspondent pas uniquement à la substance active du stupéfiant mais à d'autres éléments provenant de la plante, de son étape de fabrication ainsi que de produits utilisés comme agents coupe-glace. Ces variables sont identifiées au moyen d'une méthode analytique, la chromatographie en phase gazeuse couplée à une détection par spectrométrie de masse, permettant en une seule analyse de fournir un maximum d'informations sur la nature et le contenu du stupéfiant. Le résultat est un chromatogramme, dont les pics correspondent aux substances détectées.

Leur quantité relative peut être calculée en mesurant l'aire du pic correspondant à la variable détectée. La structure générale de la quantité relative de chaque variable peut être représentée sous forme de matrice (figure 5.4).

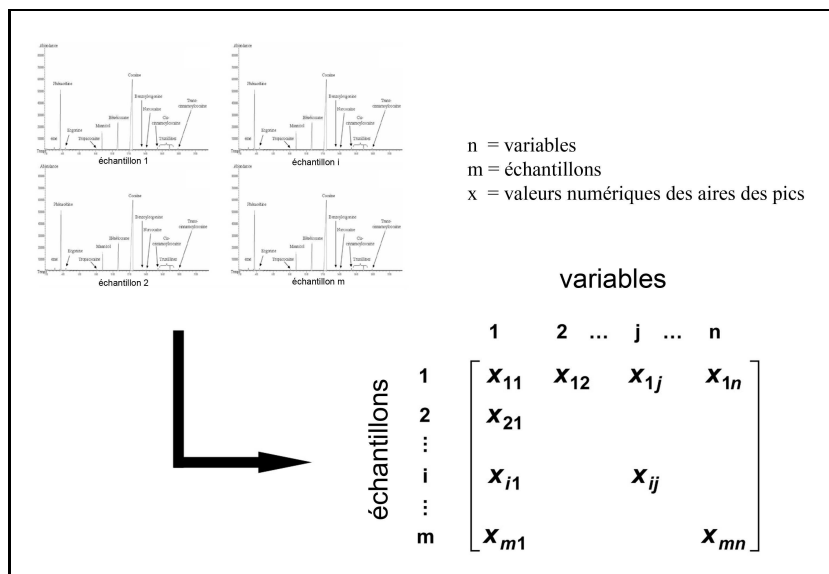


FIGURE 5.4 – Extraction des données à partir des chromatogrammes (tiré de [Ioset *et al.*, 2005]).

Les colonnes représentent la quantité relative de chaque variable et les lignes, les échantillons

analysés. C'est à partir d'une telle matrice que les profils chimiques sont établis.

Les méthodes utilisées ont pour but de comparer un échantillon-candidat à l'ensemble de la banque de données afin de déterminer à quelle autre classe ou profil chimique (groupes d'échantillons dont les caractéristiques chimiques sont similaires), il correspond. L'analyse en composantes principales (ACP) permet de réaliser une présélection d'échantillons de la banque de données, dont les proportions de variables chimiques sont proches de celles du candidat. La présélection est ensuite comparée en détail au candidat en utilisant des mesures de corrélation afin d'établir une similarité. Une analyse au cas par cas est finalement réalisée par un expert de l'IPS afin de déterminer si un lien chimique peut être établi entre le candidat et les échantillons sélectionnés.

Les résultats peuvent être de trois types. Soit l'échantillon n'a aucune similitude chimique avec les échantillons pré-sélectionnés et il va alimenter la banque de données, soit il est similaire à des échantillons correspondant à une classe chimique et peut ainsi la rejoindre. Une troisième possibilité est que le candidat soit similaire à des échantillons non classés, alimentant la banque de données. Dans ce cas, une nouvelle classe chimique est établie.

Une classification automatique des échantillons sur la base de leur classe chimique respective peut être réalisée comme validation. Cette étape est réalisée par des méthodes statistiques supervisées de classification (SIMCA, réseaux de neurones). Un modèle des classes chimiques peut être établi et si le candidat est classé de manière optimale selon ce modèle, sa classe chimique est validée. Le modèle doit ensuite être mis à jour avec l'inclusion du nouvel échantillon classé.

Un schéma global de l'établissement des liens chimiques est présenté par la figure 5.5.

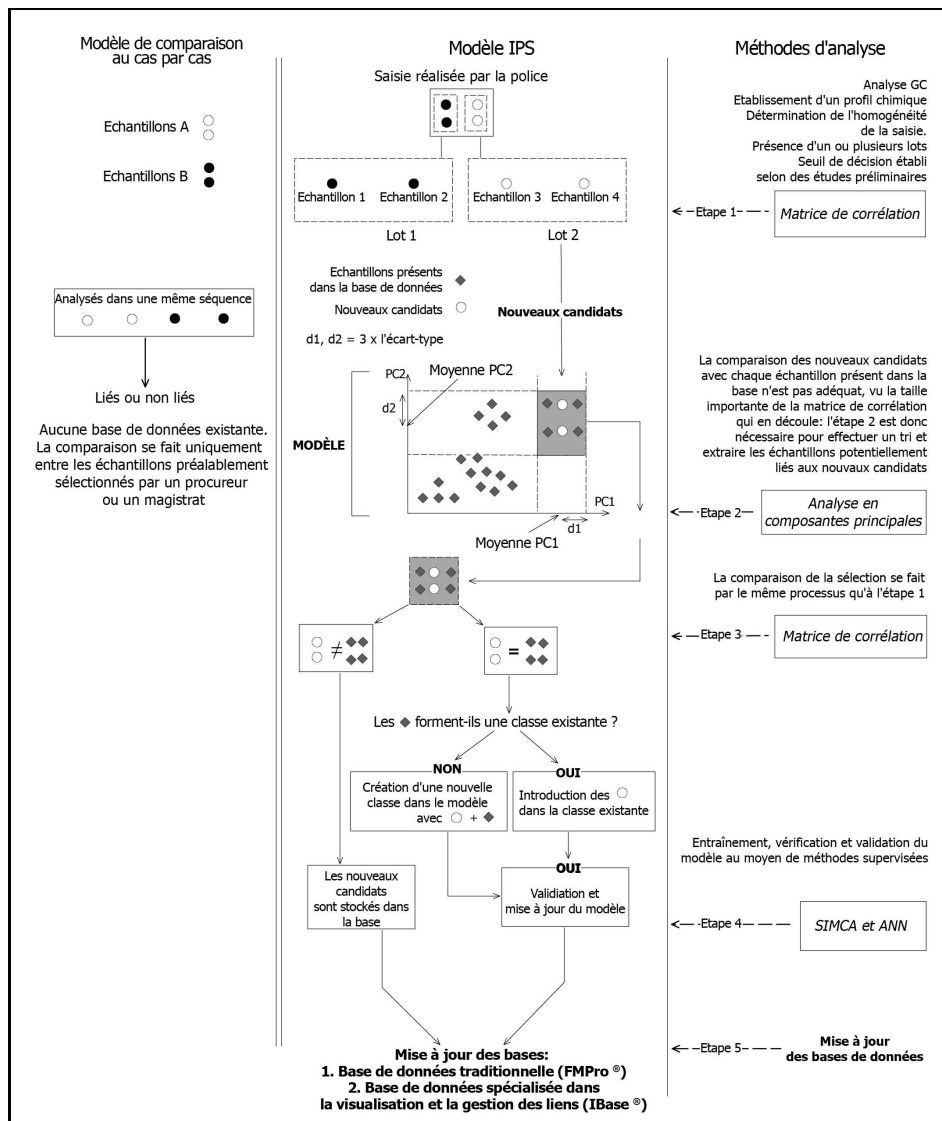


FIGURE 5.5 – Processus général de l'établissement des liens chimiques (tiré de [Ioset *et al.*, 2005]).

Une autre partie du processus du profilage est le traitement, la visualisation et l'interprétation de ces liens chimiques. Elle fait l'objet d'un travail de thèse réalisé à l'IPS par le premier auteur de l'article [Ioset *et al.*, 2005].

La complexité de la gestion des liens provient notamment du fait que des saisies peuvent contenir des échantillons pouvant être de classes chimiques différentes. Un **lien chimique** ne peut être établi qu'entre échantillons. Il n'est pas possible de lier chimiquement des saisies, en raison de la reconnaissance de classes chimiques différentes dans une même prise. En revanche, ces échantillons peuvent être l'objet de liaisons entre différentes saisies (figure 5.6), ou avoir des relations avec

les informations de lieux et dates, par exemple. L'exploitation de ces liens permet de décrire des phénomènes et des séries. Il faut toutefois garder à l'esprit que l'étape entre la production et la distribution d'un stupéfiant à l'origine de la similitude entre échantillons d'une même classe, est inconnue.

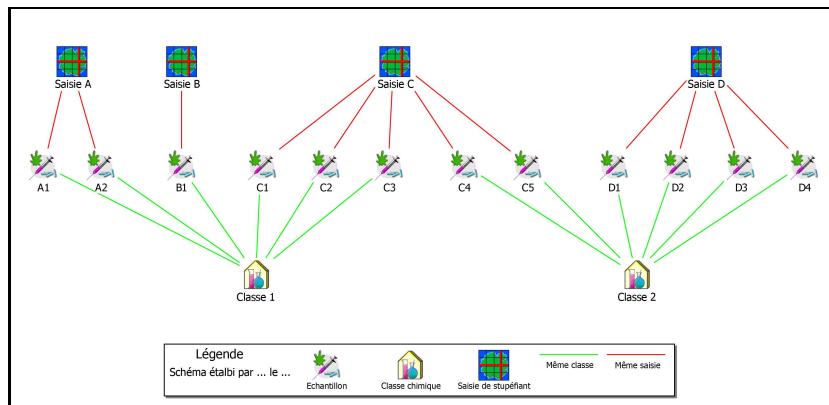


FIGURE 5.6 – Illustration des différents types de liens (tiré de [Ioset *et al.*, 2005]).

En se basant sur les informations circonstancielles obtenues durant une enquête, les enquêteurs peuvent établir et définir des liens entre des personnes actives dans un réseau de stupéfiants. Les liens chimiques permettent de corroborer ces liens et peuvent aussi mettre en évidence des relations inconnues jusqu'alors.

L'utilisation des liens chimiques pour lier les saisies sur la base d'échantillons de même classe chimique permet d'orienter une enquête vers des directions inexploitées. Ceci permet de générer de nouvelles investigations ou de déterminer des priorités d'actions lorsqu'un individu est lié à d'autres individus, jouant un rôle dans un réseau de stupéfiants. L'exploitation des liens chimiques permet de mettre en évidence des associations de personnes ou d'affaires de Police qu'une méthode traditionnelle d'enquête ne verrait pas.

Mise en pratique

La difficulté principale d'un tel processus (de même que les méthodes de reconnaissance de patterns qui pourraient lui être intégrées), est de disposer d'un expert pour gérer et interpréter les nouvelles données. Ce travail ne peut pas être réalisé par les enquêteurs en raison de leurs tâches d'enquêtes chargées, ainsi qu'en raison de la complexité de l'information. Il nécessite une formation spécifique et prend beaucoup de temps pour disposer pleinement du potentiel de cet outil. Actuellement, le profilage des stupéfiants est réalisé à l'Institut de Police Scientifique, où se situent les compétences techniques et analytiques, ainsi que les laboratoires et les machines

d'analyse. Il serait néanmoins utile de mettre en commun les connaissances *a priori* des enquêteurs et celles de scientifiques spécialisés en analyse criminelle et dans le domaine de la lutte contre le trafic des stupéfiants, afin d'exploiter et de diffuser l'information opérationnelle issue du profilage. Une unité d'analyse spécialisée au sein d'une Police pourrait coordonner l'information entre un laboratoire appliquant la méthode du profilage et les enquêteurs. Cette perspective sera abordée dans la discussion au chapitre 7.

5.2.5 Les données issues de saisies d'héroïne et de cocaïne

Géographie des saisies

Les données utilisées pour cette étude proviennent de saisies faites dans les cantons suisses de Genève (Ge), Vaud (Vd), Neuchâtel (Ne) et Tessin (Ti).

Les cantons d'étude peuvent être caractérisés par leur position géographique et stratégique. Leur situation influence très probablement l'origine et les lieux de saisies de stupéfiants, ce qui permet de poser quelques hypothèses quant à leur provenance.

Le canton de Genève, localisé à l'extrême Ouest de la Suisse, fait frontière avec la France sur la grande majorité de sa limite. Il possède un aéroport international et sa capitale est la plus grande ville de Suisse romande. Les stupéfiants saisis dans cette région proviennent des frontières, de l'aéroport et de la rue. La situation stratégique de ce canton, sa faible étendue et son aéroport en font une plaque tournante pour la Suisse romande. Il est possible que des produits en provenance du Nord (Zürich, Hollande, Allemagne) et du Sud (Italie) passent par cette région (information obtenue par des enquêteurs).

Le canton de Vaud se situe sur un axe routier menant au Nord et à l'Est de la Suisse. Sa situation géographique favorise très probablement le transit et la distribution de stupéfiants dans les villes principales. Les saisies sont essentiellement réalisées dans la rue. L'information d'une prise dans un train ou sur un axe routier important n'est actuellement pas disponible.

Le canton de Neuchâtel est situé au Nord-Ouest du pays et fait frontière avec la France. Une source d'approvisionnement provient probablement du Nord (France, Hollande). Les saisies sont mixtes (frontière et rue).

Le canton du Tessin est situé au Sud-Est de la Suisse et fait frontière avec l'Italie. Il est fort possible qu'il soit une des voies d'entrée d'une route du Sud. Des saisies de rue, d'axe routier et de frontière y sont réalisées et la majorité des échantillons disponibles pour cette étude provient de prises faites aux frontières.

Il est possible que les sources d'approvisionnement soient différentes en fonction de la situation géographique et stratégique des cantons; ceci se refléterait alors au niveau du contenu du produit

vendu. Dans ce sens, une classification basée sur la situation des cantons est proposée et illustrée par la figure 5.7.

- **Les saisies faites aux frontières** : elles pourraient être des substances en transit entre deux pays (qualité moyenne à faible), ou une importation de produits plus ou moins bruts (qualité supérieure).
- **Les saisies faites dans l’enceinte d’un aéroport** : elles correspondraient à une importation de produits plus ou moins bruts (qualité supérieure).
- **Les saisies faites sur un axe routier** : elles pourraient être de qualités diverses, supérieure à faible, si le processus de coupage a été réalisé avant son transport.
- **Les saisies faites dans la rue** : elles seraient caractérisées par une qualité moyenne à faible et pourraient correspondre à l’étape de diffusion d’un réseau local.

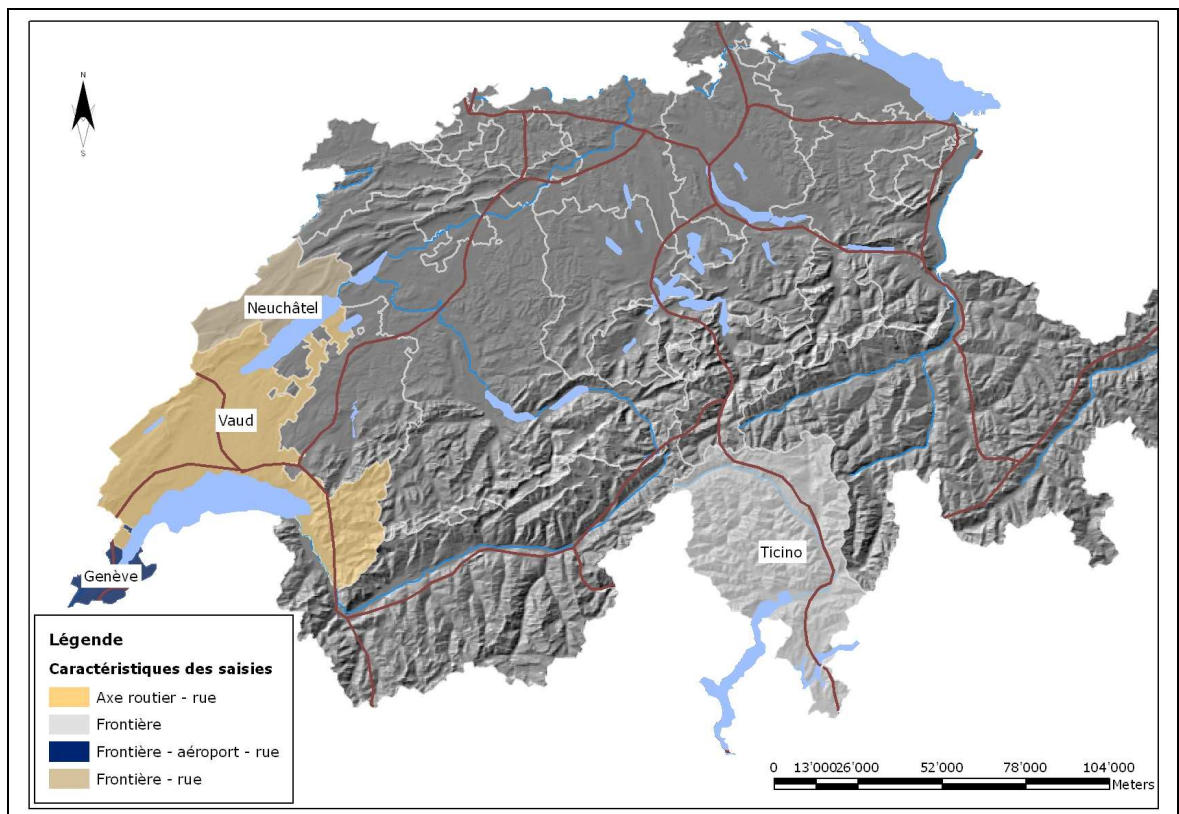


FIGURE 5.7 – Caractéristiques des saisies selon les cantons.

Le problème rencontré dans cette étude est que l’information de saisies faites dans la rue n’est pas disponible. Les données issues d’échantillons saisis aux frontières, dans un aéroport, dans la rue ou à la fin d’une longue enquête sont mélangées. Seules les données du Tessin peuvent être caractérisées comme étant issues de nos frontières. Les caractéristiques géographiques et stratégiques

de chaque canton sont les seules observations permettant de pouvoir expliquer certaines tendances liées aux lieux de saisie. Elle seront présentées dans le chapitre 6.

Rythme des saisies

La proportion de stupéfiants saisis est faible par rapport à ce qui est en circulation. Les saisies, dont le rythme est dépendant de l'activité de la Police, sont l'aboutissement d'une enquête. Si cette dernière est de longue durée, avec pour but est de stopper la diffusion d'une grande quantité de stupéfiants, la masse de la saisie sera importante, de même que le nombre d'échantillons prélevés. Le trimestre et la localité où la prise aura été effectuée seront caractérisés par un grand nombre d'échantillons, ce qui ne sera pas représentatif de l'activité criminelle mais de celle de la Police. Des périodes dans l'année telles que les vacances, peuvent aussi être la source de diminutions du nombre de saisies.

Ces observations sont importantes pour l'interprétation des résultats, afin de garder à l'esprit que les tendances observées ne représentent que la partie émergée de l'iceberg.

Les données

Les composés majeurs

Les données proviennent du laboratoire d'analyses des stupéfiants de l'IPS, après que les échantillons aient été transmis par un enquêteur pour analyse et profilage (cf. section 5.2.4 et [Ioset *et al.*, 2005]). La quantité relative des composés majeurs de l'héroïne et de la cocaïne ainsi que des produits de coupage est établie, de même que le taux de pureté (cf. section 5.2.4). Ces données quantitatives, stockées dans une banque de données, sont disponibles pour des analyses, telles que l'établissement des liens chimiques dans le processus du profilage, ou une analyse par reconnaissance de patterns, telle que réalisée dans [Ratle *et al.*, 2006b,a, 2008]. Les données liées aux composés majeurs n'ont néanmoins pas été utilisées pour l'application réalisée dans ce travail.

Les produits de coupage

Le jeu de données utilisé pour une analyse par data mining est composé de l'information de présence ou d'absence¹ de substances de coupage. Celles-ci peuvent être des diluants, dont le but est de diminuer la pureté, des adultérants, utilisés pour leurs propriétés pharmacologiques augmentant les effets du stupéfiant ou des solvants. Les produits de coupage peuvent être ajoutés au stupéfiant à n'importe quel moment entre le produit illicite fini sortant du laboratoire et la vente pour consom-

1. Cette information d'absence ou de présence correspond à la détection ou à la non-détection des substances par la méthode et la machine utilisée

mation dans la rue. Le schéma de la figure 5.8 indique les différentes étapes de transformation du stupéfiant.

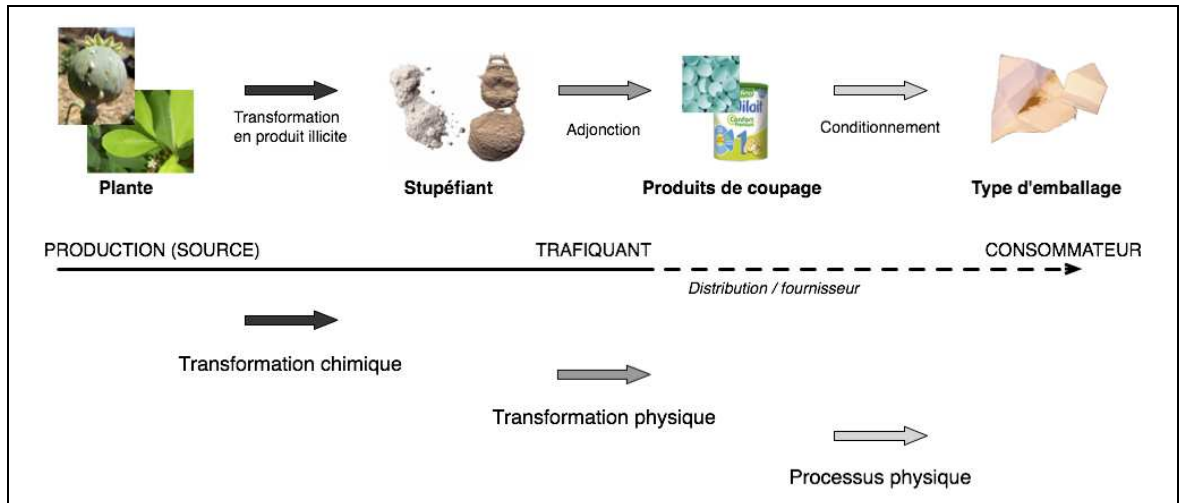


FIGURE 5.8 – Réseau général d’approvisionnement des stupéfiants (schéma inspiré de [UNODC, 2001]).

Les données circonstancielles

Des informations de date et de lieu de saisies accompagnent les échantillons lors de leur transmission à l’IPS. Ces données circonstancielles ne sont pas systématiquement utilisées dans la démarche du profilage. Il est possible que des tendances liées au temps et à l’espace existent, car le trafic de l’héroïne et de la cocaïne semble évoluer. L’utilisation de données circonstancielles pour une recherche de tendances liées à des régions (lieux de saisies) et au temps (dates de saisies), a été réalisée dans ce travail.

5.3 Analyse

5.3.1 Les informations à disposition

D’après [Ioset *et al.*, 2005], les liens chimiques entre des échantillons de stupéfiants peuvent s’établir n’importe où entre la production et la distribution des stupéfiants. Leur origine exacte est inconnue. Leur utilisation conjointe avec des informations d’enquête permet de lier des saisies, de confirmer des hypothèses de l’enquêteur et d’ouvrir de nouvelles perspectives quant aux relations possibles entre des personnes en possession des stupéfiants et des réseaux. Il est cependant difficile d’identifier des réseaux locaux de petite étendue géographique. La méthode du profilage n’a, jusqu’ici, pas permis d’en mettre en évidence.

Les informations circonstanciées (lieu et date de saisie) concernent des événements dont l'échelle géographique est cantonale à régionale. L'analyse de ces données dans une démarche d'analyse criminelle pourrait apporter diverses informations sur l'évolution et la répartition de l'état physico-chimique de l'héroïne et de la cocaïne dans le temps et par région, telles que :

- La possibilité de mettre en évidence des caractéristiques cantonales et régionales en associant le contenu de ces stupéfiants à la configuration géographique.

Les cantons frontières, à axes de communication majeurs, ou possédant un aéroport international, peuvent avoir une dynamique différente au niveau des réseaux d'approvisionnement et de distribution. Les saisies en provenance du Tessin et celles faites aux douanes genevoises pourraient renseigner sur l'approvisionnement. Celles réalisées dans la rue, sur l'état du stupéfiant lors de sa distribution.

- L'analyse des données circonstanciées et l'analyse physico-chimique des échantillons, apporterait des informations sur leur "voyage" et leur destination finale.

Le nombre d'intermédiaires pouvant modifier l'état de la cocaïne et de l'héroïne de son entrée dans le pays à sa distribution dans la rue est actuellement inconnu. Les connaissances actuelles laissent supposer que ce nombre varie beaucoup pour la cocaïne, moins pour l'héroïne. Le nombre total de produits de coupage détectés pour l'héroïne est plus petit que celui pour la cocaïne. Une modélisation de ces intermédiaires permettrait une meilleure vision de la distribution des stupéfiants.

Date et lieu de saisie

Le jeu de données utilisé dans cette recherche a été récolté durant les années 2000 à 2005 dans les cantons de Vaud, Tessin, Genève et Neuchâtel.

La date de saisie ainsi que la commune où elle a été faite sont indiquées. Les échantillons provenant du Tessin sont essentiellement issus de saisies douanières. Ceux réalisés dans les autres cantons sont hétérogènes (douane à rue).

Les produits de coupage

L'adjonction de produits à divers stades d'un réseau, dépend du type de stupéfiant (les substances de coupage de la cocaïne sont en général différentes de celles de l'héroïne) et probablement du laboratoire duquel il est issu, du fournisseur, du nombre d'intervenants et du dealer. Il est possible que le choix des produits dépende aussi des substances disponibles (elles peuvent varier selon les pays où le coupage est réalisé), ainsi que du fournisseur ou du dealer. Selon cette hypothèse, le choix d'un produit plutôt que d'un autre et leur disponibilité, se traduiraient par des **combinaisons différentes**.

Pour l'héroïne, les échantillons saisis contiennent presque tous les même produits qualifiables de “

de base”, qui sont le paracétamol et la caféine. Dans cette optique, il est probable que des produits fortement répandus spatialement et temporellement aient été ajoutés à un niveau élevé du réseau. Cette hypothèse est illustrée par la figure 5.9.

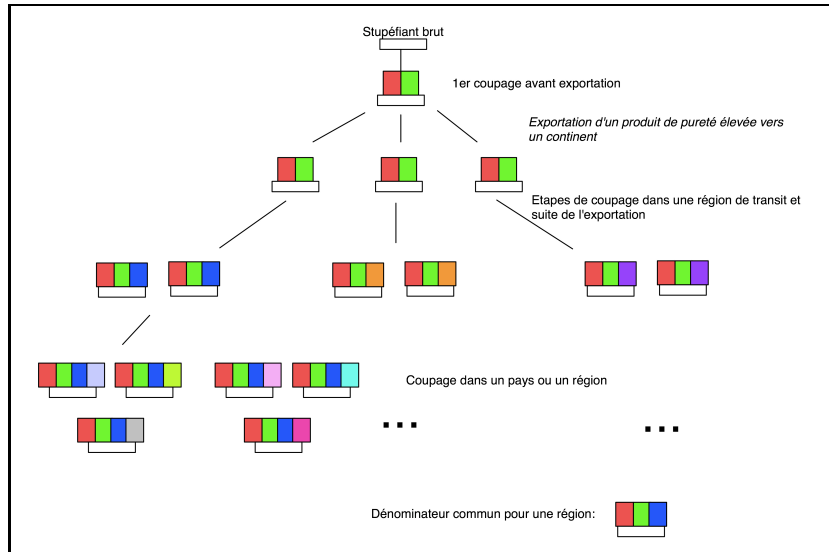


FIGURE 5.9 – Hiérarchisation possible du processus de coupage. Les rectangles de couleurs correspondent à des produits de coupage.

Selon cette hypothèse, si des substances sont un dénominateur commun d'échantillons de qualité élevée avec peu de produits différents et d'échantillons de faible qualité et plusieurs fois coupés, il est possible qu'elles aient été ajoutées en amont du réseau. En transposant cette hypothèse à un niveau régional, si une combinaison semble être le dénominateur commun d'échantillons saisis dans une même région, il est possible qu'elle reflète l'activité d'un réseau local.

L'étude de ces combinaisons selon un axe spatio-temporel permettrait de modéliser le phénomène du coupage et d'apporter des informations sur une possible activité criminelle affectant la population régionale.

Les mélanges de produits de coupage détectés sont appelés “combinaisons” ou “assemblages” dans ce travail.

On distingue parmi ces produits les diluants, dont la présence a pour but de diminuer la pureté, les adultérants présents pour augmenter les effets du stupéfiant et les produits issus du conditionnement que l'on appellera “divers”. Le nombre de produits de coupage détectés à l'IPS entre 2000 et 2005 sont les suivants :

- **Cocaïne** : 46 substances
- **Héroïne** : 22 substances

Ces produits ont été catégorisés en fonction de leur fréquence d'apparition durant les années

considérées. La classification proposée se base sur les observations faites au travers des analyses :

1. Les produits les plus fréquents, détectés dans presque tous les échantillons.
2. Les produits fréquents, caractéristiques d'un des deux stupéfiants mais parfois absents.
3. Les produits peu fréquents mais néanmoins connus pour être utilisés comme coupage.
4. Les produits très peu fréquents, détectés une seule fois.

Le tableau 5.3 énumère les produits de coupage par catégories. Des indices indiquent le type de produits : a = adultérant, d = diluant, div = divers.

Pour faciliter la manipulation des données, celles-ci ont été traitées selon un format binaire (la détection de la substance de coupage = "1" et la non-détection = "0") par produit et selon les combinaisons (tableaux 5.1 et 5.2). L'exemple suivant illustre cette mise en forme pour trois échantillons d'héroïne provenant de deux saisies. Les combinaisons de substances seront appelées C1, C2, C3 :

- **Saisie 1** : deux échantillons contenant Paracétamol - Caféine - Griséofulvine = C1
- **Saisie 2** : un échantillon contenant Paracétamol - Caféine = C2
- **Saisie 2** : un échantillon contenant Paracétamol - Caféine - Glycérol = C3

TABLE 5.1 – Format des données par produits

		Paracétamol	Caféine	Glycérol	Griséofulvine
Saisie 1	Echantillon 1	1	1	0	1
Saisie 1	Echantillon 2	1	1	0	1
Saisie 2	Echantillon 1	1	1	1	0
Saisie 2	Echantillon 2	1	1	0	0

TABLE 5.2 – Format des données par combinaisons.

Saisie 1	Echantillon 1	C1
Saisie 1	Echantillon 2	C1
Saisie 2	Echantillon 1	C3
Saisie 2	Echantillon 2	C2

Le nombre de combinaisons observées pour les années considérées sont :

- **Cocaïne** : 254 combinaisons
- **Héroïne** : 36 combinaisons

La correspondance des codes des combinaisons est placée en annexe.

TABLE 5.3 – Catégories et répartition des produits de coupage

Catégories	1	2	3	4
Cocaïne	lactose ^d mannitol ^d glucose ^d phénacétine ^a inositol ^d lidocaïne ^a sucrose ^d	caféine ^a procaine ^a diltiazem ^a 1,4-dmt ^{*div} paracétamol ^a acide citrique ^a amidon ^d hydroxisine ^a piractam ^a benzocaïne ^a	lévamisole ^a sucres ^d glycine ^d acétylsalicylate de lysine ^a aspirine ^a atropine ^a lysine ^d maltose ^d phosphate ^d acide salicylique ^a antipyrine ^a créatine ^a éphédrine ^a fentanyl ^a ibuprofène ^a acide citrique ^a	acétate de propyl ^a acide ascorbique ^a acide glutamique ^a acides gras ^d aminogluthétimide ^a amphétamine ^a aspégic ^a glycérol ^d glycol ^d nicotinamide ^a nicotine ^a
Héroïne	caféine ^a paracétamol ^a griséofulvine ^a glycérol ^d sucrose ^d mannitol ^d	lactose ^d glucose ^d procaine ^a phénobarbital ^a diazepam ^a	acide citrique ^a fructose ^d aspirine ^a phosphate ^d	naphtol ^a acides gras ^d maltose ^d méthadone ^a piracetam ^a

*1,4dmt = diméthylterephthalate, a = adultérant, d = diluant, div = divers

Description synthétique des données

Les informations du nombre de données utilisées sont les suivantes :

TABLE 5.4 – Nombre de données brutes

Héroïne	6365
Cocaïne	7237

TABLE 5.5 – Nombre de produits et combinaisons différents détectés

Stpéfiant	Nombre de produits	Nombre de combinaisons
Héroïne	22	36
Cocaïne	46	254

TABLE 5.6 – Pureté moyenne

Héroïne	14.9%
Cocaïne	51.4%

Les figures suivantes indiquent la répartition du nombre de données par cantons et années.

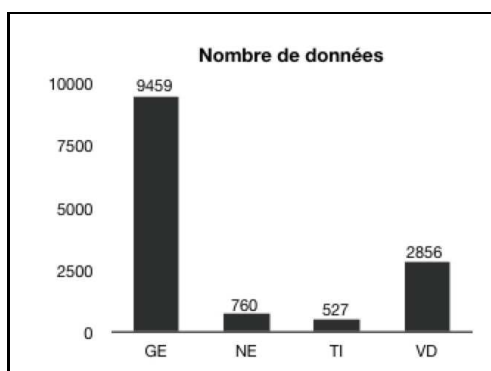


FIGURE 5.10 – Nombre de données par cantons d'étude.

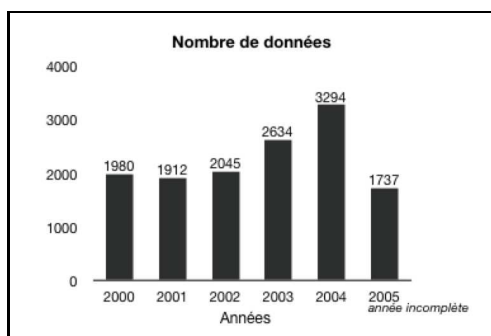


FIGURE 5.11 – Nombre de données par années d'étude.

Les figures suivantes montrent la pureté moyenne de l'héroïne et de la cocaïne par cantons.

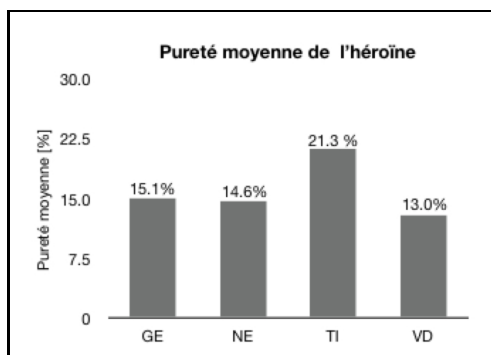


FIGURE 5.12 – Pureté moyenne de l'héroïne par cantons.

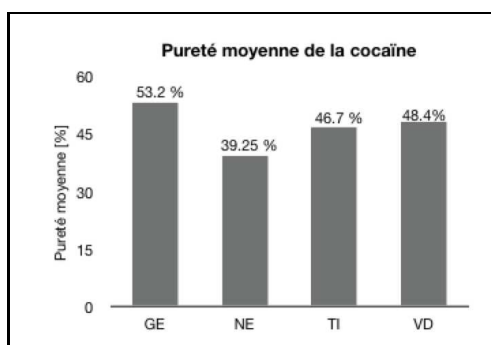


FIGURE 5.13 – Pureté moyenne de la cocaïne par cantons.

5.3.2 Méthodes d'analyse

Hypothèses

Des hypothèses de travail ont été formulées à la section 5.1. Il est possible de les affiner, afin de mieux cadrer l'analyse et l'interprétation des tendances :

1. Le phénomène du coupage lié au trafic des stupéfiants évolue dans le temps.
2. Une modélisation globale d'échantillons saisis sur la base de leur contenu en produits de coupage et selon des dimensions spatiales et temporelles, peut révéler des tendances utiles sur le trafic des stupéfiants d'une échelle nationale à une échelle régionale.
 - (a) *Existe-t-il des réseaux de distribution identifiables sur la base des données de produits de coupage et des informations circonstancielles ?*
 - (b) *Est-il possible de mettre en évidence un transit de stupéfiants entre régions ?*
3. L'analyse des relations existant entre les caractéristiques physico-chimiques associées aux informations circonstancielles (temps et géographie) des saisies, permet de produire du renseignement opérationnel.

Pour répondre à ces hypothèses, un modèle global du coupage a été créé, en se basant sur les informations de lieux et dates ainsi que sur les divers produits détectés. Les méthodes d'analyse statistiques permettent d'obtenir des informations globales sur la répartition et les tendances fréquentistes des données. Les relations entre une période ou un lieu de saisie et une combinaison particulière de produits pourront être difficilement mises en évidence par des méthodes statistiques. Le choix a été de se baser sur les liens existant entre les produits et les lieux et dates de saisies. L'analyse combinatoire regroupe des techniques permettant de réaliser de tels modèles.

Les méthodes utilisées sont issues de la théorie des graphes [West, 1996] et ont pour but de traiter l'information relationnelle existant entre les données (produits de coupage et données circonstancielles). Une première étude a été réalisée pour les produits de coupage de l'héroïne [Terrettaz-Zufferey *et al.*, 2007]. Ce domaine permet de représenter graphiquement une matrice contenant l'information numérique des relations entre entités (arêtes et sommets ou noeuds du graphe).

Les tests ont été appliqués aux produits de coupage et aux combinaisons, de manière séparée :

- **Pour les produits** : deux modèles ont été créés, l'un pour les produits détectés dans les cantons d'étude, l'autre pour les produits détectés durant les années d'étude. Le modèle

temporel (selon les années) va permettre de tester la première hypothèse ; le modèle spatial (selon les cantons) permettra de tester la deuxième hypothèse.

Le fichier de base se compose d'une matrice de 1 et de 0, indiquant les produits détectés dans un canton ou une année et avec quels autres produits. Les sommets correspondent aux substances de coupage et un lien les lie lorsqu'elles ont été détectées ensemble dans un même échantillon. Un graphe simple par canton et par année a été construit.

Dans le première exemple de la figure 5.14, à gauche, tous les produits représentés par les sommets du graphe ont été détectés dans un même échantillon, au moins une fois durant l'année 2004, pour tous les cantons confondus. Pour celui de droite, dans le canton de Neuchâtel et durant les années 2000 à 2005, le Glycérol et la Griséofulvine n'ont jamais été détectés dans un même échantillon, contrairement aux autres qui l'ont été au moins une fois.

- **Pour les combinaisons**, les fichiers sont des listes d'arêtes de graphes bipartis caractérisés par deux ensembles distincts de sommets. L'un correspond aux combinaisons de produits de coupage et l'autre à l'information circonstancielle qui est :
 - **les dates de saisies** : les trimestres (le trimestre est considéré comme une unité de temps)
 - **le lieu de saisie** : les localités

Dans le second exemple de la figure 5.14, des graphes bipartis sont utilisés. Les deux types de sommets sont les combinaisons de produits, en foncé et les trimestres (trim1, trim2, ..., trim.x) ou les localités (Lausanne, Genève) en clair. Cette représentation permet de visualiser quelles combinaisons ont été détectées durant quel trimestre, respectivement quelle localité.

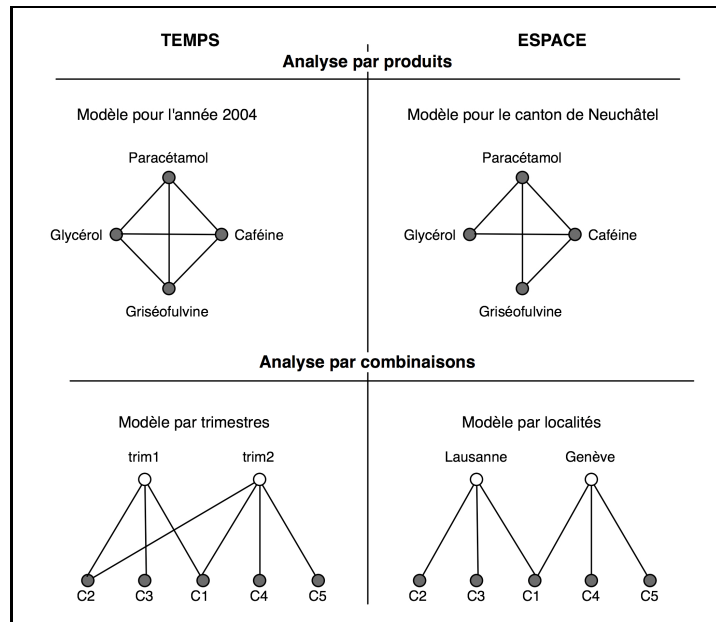


FIGURE 5.14 – Graphes du haut : les sommets correspondent à des produits et une arête les lie lorsqu'ils ont été détectés dans une région ou une année. Graphes du bas : les sommets correspondent aux combinaisons, aux trimestres ou aux localités de saisie. Une arête les lie lorsque les combinaisons y sont détectées.

Exploitation de l'information numérique des graphes

La création d'un modèle de toutes les données et leurs relations selon l'espace et le temps permettent d'explorer les informations de manière globale. L'analyse du graphe au niveau exploratoire permet de mettre en évidence des tendances selon les axes spatio-temporels choisis.

L'information numérique suivante a été extraite :

1. **Nombre de produits et de combinaisons** différents par année et par canton.
2. **Nombre de cliques**².

L'évolution du nombre de produits et de combinaisons selon les années permet d'avoir une estimation de la variabilité des substances de coupage dans le temps. De même pour les cantons, cette information permet d'avoir une idée des différences de variabilité entre les régions étudiées.

Les cliques sont des sous-graphes dont les sommets sont tous adjacents. Elles correspondent dans notre analyse à des produits détectés une même année dans un même canton, mais pas forcément dans un même échantillon. Connaître le nombre de cliques différentes par région et selon le temps est une indication complémentaire à la variabilité. Pour le graphe des produits, une augmentation

² Une clique d'un graphe est un sous-graphe dont les sommets sont adjacents les uns aux autres. La clique est maximale si aucun autre noeud du graphe n'est adjacent à ceux de la clique. Voir chap. 4.3.3

du nombre de cliques indique un changement au niveau de l'utilisation des produits et peut correspondre à une augmentation du nombre de substances utilisées par coupage.

Les résultats obtenus permettraient une vision d'ensemble et généralisée du phénomène du coupage ainsi que d'éventuels changements au cours du temps et selon les régions. La mise en évidence d'une évolution permettrait de répondre à la première hypothèse. Il faut toutefois garder à l'esprit que les informations numériques des graphes sont en grande partie dépendantes de l'activité de la Police. Certaines variations peuvent exprimer une activité policière particulière plutôt que des changements au niveau du phénomène du trafic des stupéfiants.

Différences de coupage entre les cantons

Selon les connaissances actuelles, les stupéfiants traversant nos frontières contiennent déjà des adjuvants et des diluants. Un processus de coupage intervient entre le pays de production et le pays de livraison. Il est ainsi possible de supposer que les substances relatives à cette intervention soient assez fréquentes dans les saisies réalisées dans les différents pays. C'est le cas pour le Paracétamol et la Caféine présents dans presque toutes les saisies d'héroïne.

Une partie des produits varie beaucoup selon la géographie. Il serait intéressant de connaître quelles substances sont communes à deux cantons ou plus.

En recherchant les produits communs à tous les cantons, il serait possible ensuite de les séparer des autres afin de cibler les substances ajoutées après le niveau de production, voire à un niveau local.

Une application de la théorie des ensembles aux graphes des produits détectés dans chaque canton a été réalisée, afin de mettre en évidence des produits présents dans deux cantons ou plus (**intersection**) et ceux exclusifs à une région (**différence**). L'intersection conserve les sommets communs à deux graphes, la différence soustrait un graphe B d'un graphe A, et le résultat ne contient que les sommets exclusifs au graphe A (figure 5.15).

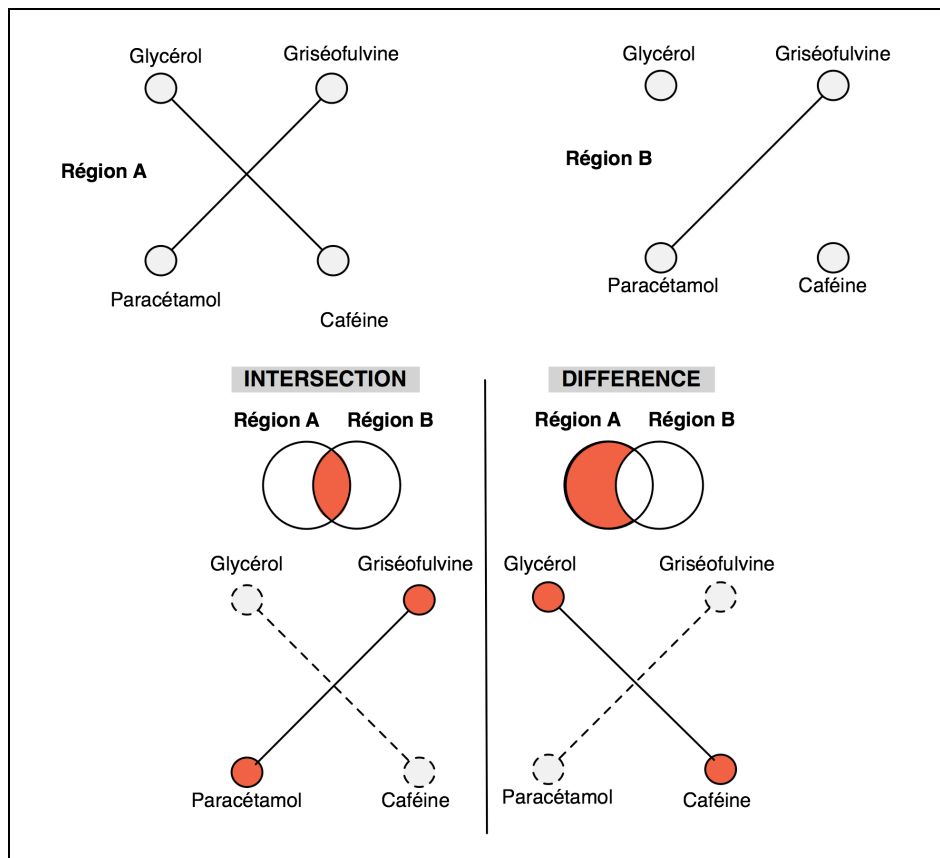


FIGURE 5.15 – Illustration de l'application de la théorie des ensembles

Cette application a été réalisée au moyen des bibliothèques **IGraph** et **NetworkX** [Csárdi et Nepusz, 2003; Hagberg et al., 2004] en langage de programmation Python [Van Rossum, 1990].

Le graphe obtenu en intersectant ceux de chaque canton montrera les produits détectés dans toutes la région étudiée. Il est fortement possible que ces produits soient déjà présents avant leur entrée en Suisse. Les produits mis en évidence en soustrayant ce graphe à celui de chaque canton correspondront très probablement à des adjonctions faites à l'intérieur de nos frontières.

En soustrayant entre eux les graphes de chaque canton, les produits "typiques" d'un canton seront mis en évidence. Cela signifie qu'ils n'ont encore jamais été détectés ailleurs et non qu'ils sont absents d'autres régions.

L'analyse permettrait de répondre à la deuxième hypothèse et en particulier à la 2.b. L'utilisation des opérations de la théorie des ensembles permettrait de mettre en évidence des produits communs à plusieurs cantons, pouvant indiquer un transit de stupéfiants. Ce traitement ne permet pas de mettre en évidence des adjonctions locales de produits déjà présents dans le stupéfiant. Par exemple, pour différencier du Paracétamol ajouté sur le lieu de production de celui ajouté dans notre pays,

il sera nécessaire de réaliser un profilage chimique de cette substance.

Détection de tendances

Le graph mining regroupe des techniques d'analyse permettant la détection de patterns dans des données relationnelles complexes. Le premier pas pour mettre en oeuvre une telle analyse est de structurer les données pour construire le graphe en utilisant les informations circonstancielles. Il est possible notamment d'utiliser la pureté comme facteur de pondération.

Plusieurs méthodes d'analyse ont été décrites au chapitre 4.3. Leur utilisation permettrait de mettre en évidence des structures qui seraient susceptibles de tester la deuxième et la troisième hypothèses. L'interprétation des patterns pourra peut-être y répondre, en apportant de l'information sur des phénomènes locaux pouvant devenir du renseignement opérationnel. Chaque méthode a été testée et les résultats obtenus sont décrits et interprétés dans le chapitre 6.

- **Le minimum spanning tree**³ permet d'obtenir un graphe simplifié montrant uniquement les sommets reliés par un chemin dont les poids sont minimaux. Il a été appliqué aux graphes bipartis "combinaisons - date" et "combinaisons-lieux", pondérés par la pureté. L'inverse de la pureté a aussi été utilisé comme poids, afin d'obtenir un "maximum" spanning tree lié à des puretés élevées.

L'utilisation de cette méthode permet de simplifier le graphe de base en le divisant en deux, mettant en évidence des combinaisons, des lieux et des dates de saisies concernés par des puretés basses, respectivement élevées. La figure 5.16 illustre ce type de résultat.

3. Le minimum spanning tree est un calcul d'optimisation appliqué à un graphe pondéré (les arêtes du graphe ont une valeur numérique). Le but est de mettre en évidence les sommets reliés par des arêtes dont les poids sont minimaux. Voir chap. 4.3.3.

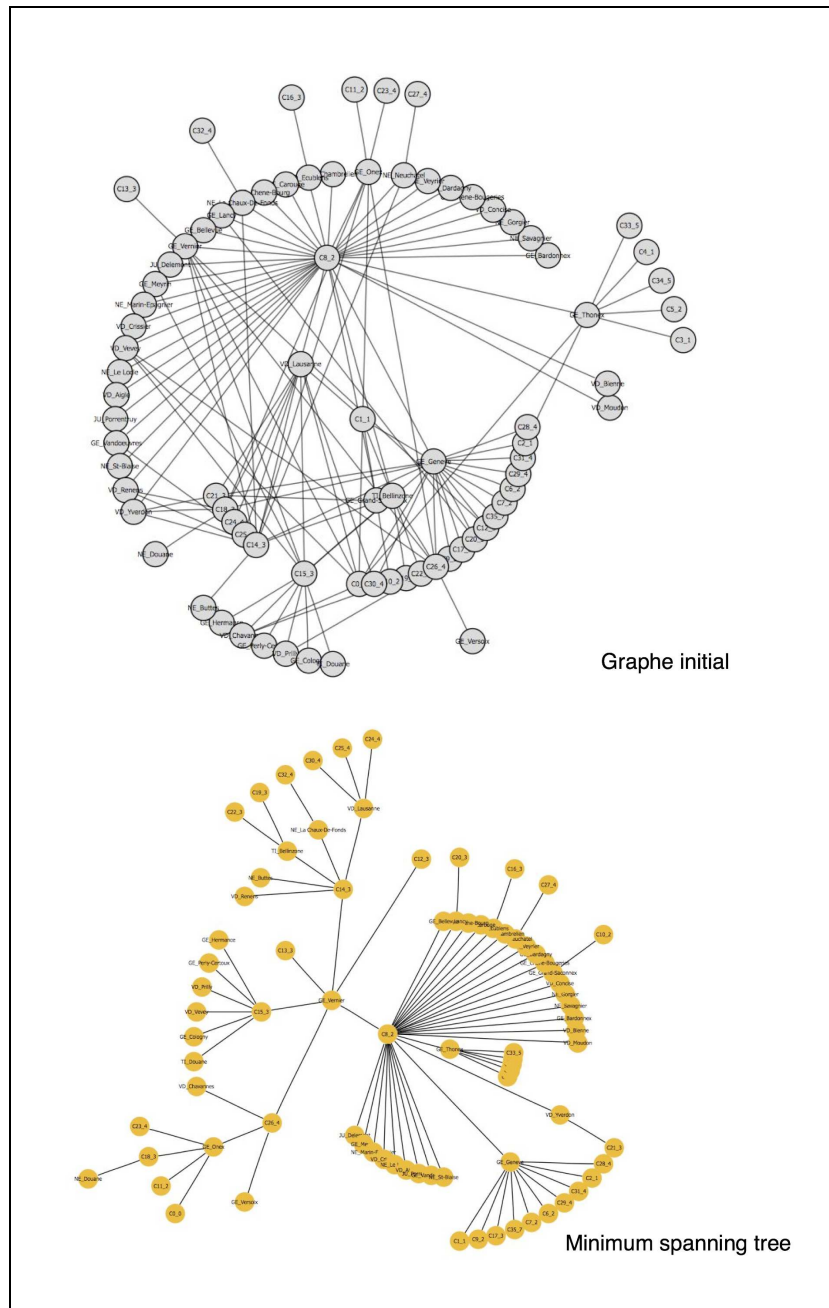


FIGURE 5.16 – En haut le graphe de base, en bas le même graphe, mais simplifié par la méthode du minimum spanning tree basé sur le taux de pureté.

- **Les points d’articulation** mettent en évidence des entités “importantes” pour le graphe, dans le sens où elles sont liées à plusieurs sommets différents. Dans l’exemple de la figure 5.17, les points d’articulation sont en orange. La combinaison C8_2 a été détectée dans plusieurs localités, elle est liée à plusieurs sommets et est donc un point d’articulation du graphe. De même pour les localités dans lesquelles plusieurs combinaisons différentes ont été trouvées,

tel que le sommet représentant GE_Onex.

Ce procédé permet de faire ressortir des sommets-clés du graphe, facilitant sa lecture, les graphes de base étant le plus souvent complexes, comme le montre la figure 5.17.

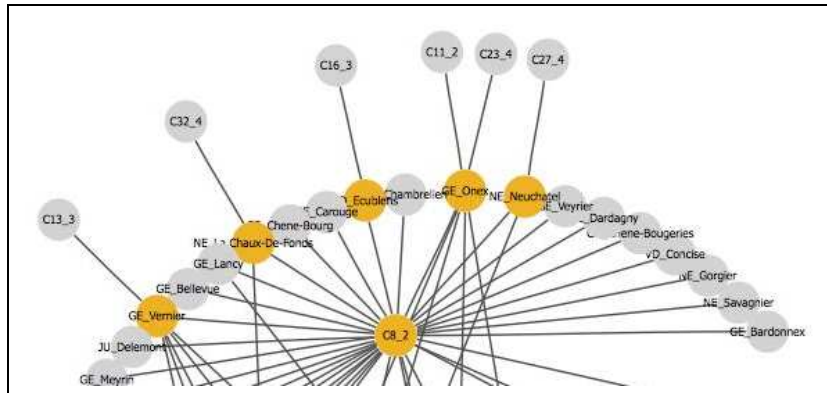


FIGURE 5.17 – Partie d'un graphe dont les points d'articulation sont en orange.

- **Le modèle des blocs** permet de détecter des groupes biconnexes caractérisés par des équivalences structurelles au sein du graphe (cf. chapitre 4.3.3). Cette méthode donne de bons résultats pour des graphes non bipartis. Pour un graphe biparti, un grand nombre de blocs insignifiants seront détectés (un bloc est un sous-graphe biconnexe *i.e.* si en enlevant un sommet il devient connexe ; un graphe biparti est composé de composantes connexes, elles seront détectées par la méthode des blocs, sans pour autant avoir une réelle signification). Cette méthode nécessite un nombre de sommets peu élevé afin d'obtenir un résultat lisible et exploitable. Elle a tout de même été testée. Un fichier contenant les relations entre les combinaisons, les dates de saisies et l'identifiant des saisies de cocaïne a été utilisé.

Le but de cette application est uniquement de tester la méthode. Celle-ci n'est pas adaptée à la structure des graphes bipartis, c'est pourquoi un jeu de données différent a été créé. Si les résultats mettent en évidence des groupes, leur interprétation permettra de juger si l'information obtenue va dans le sens d'une des hypothèses.

- **Les algorithmes de communautés** ont la propriété de détecter des groupes distincts reliés entre eux par un faible nombre d'arêtes (voir chap. 4.3.3). Un réseau local de distributions caractérisé par des combinaisons de produits de coupage et une pureté similaire, pourrait correspondre à une communauté et être détecté par ce type d'algorithme. Les graphes "combinaisons - cantons" décrits plus loin par la figure 6.2 ont la même structure que le graphe illustré dans [Clauset *et al.*, 2004] utilisé pour tester un algorithme appelé "fast greedy". Les résultats sont obtenus sous forme de graphes dont les sommets faisant partie d'une communauté possèdent une couleur distincte. Les couleurs sont attribuées en fonction du

nombre de sommets composant la communauté (figure 5.18). La couleur grise correspond aux sommets ne faisant pas partie d'une communauté, une communauté de deux sommets est en vert, de trois sommets en orange, de quatre en rouge etc...

Chaque communauté mise en évidence a des caractéristiques propres, générant le regroupement de combinaisons et lieux et trimestres de saisies. Une communauté pourrait être composée, par exemple, de plusieurs combinaisons détectées dans une même localité ou un même trimestre et avec une pureté similaire. En détaillant les produits composant les combinaisons, il serait possible de mettre en évidence un dénominateur commun, ou des similitudes pouvant refléter un réseau local.

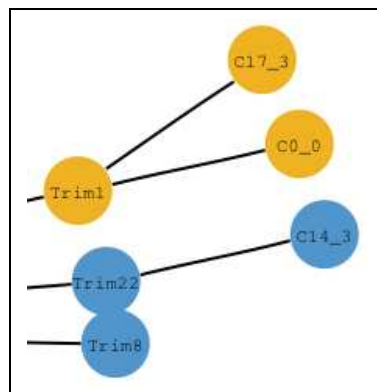


FIGURE 5.18 – Partie d'un graphe contenant deux communautés distinguables au moyen de couleurs différentes.

*Ces quatre méthodes ont été appliquées au moyen des bibliothèques **IGraph** et **NetworkX**.*

- **Les cliques maximales**⁴. sont des sous-graphes maximaux (tous les sommets sont reliés entre eux). La méthode a été appliquée aux “combinaisons-dates” pour chaque canton. Les données de base ont été transformées en une matrice binaire (figure 5.19), afin de disposer des combinaisons détectées durant une même période de temps. Le graphe résultant est un graphe d'intervalle, permettant de traiter les co-occurrences des combinaisons. La figure 5.20 illustre ce type de graphe. Chaque sommet représente un intervalle de temps. Une arête les relie lorsque les intervalles se croisent. Par exemple, les intervalles de temps **B**, **C** et **A** sont concomitants. Sur le graphe, des arêtes relient ces trois sommets.

4. Une clique d'un graphe est un sous-graphe dont les sommets sont adjacents les uns aux autres. La clique est maximale si aucun autre noeud du graphe n'est adjacent à ceux de la clique. Voir chap. 4.3.3

A	B	C	D	E	F	G
	C1_0	C10_4	C11_3	C12_3	C13_4	C14_5
0_GE-nov05	0	0	0	0	0	0
0_GE-oct05	0	0	0	0	0	0
0_GE-sept05	0	0	0	0	0	0
0_GE-août05	0	0	0	0	0	0
0_GE-juil05	0	0	0	0	0	0
0_GE-juin05	0	0	0	0	0	0
0_GE-mai05	0	0	0	0	0	0
0_GE-avr05	1	0	0	0	0	0
0_GE-mars05	0	0	0	0	0	0
0_GE-févr05	1	0	0	0	0	0
0_GE-janv05	0	0	0	1	0	0
0_GE-nov06	0	0	0	0	0	0
0_GE-oct06	0	0	0	0	0	0
0_GE-sept06	0	0	0	0	0	0
0_GE-août06	0	0	0	0	0	0
0_GE-juil06	0	0	0	0	0	0
0_GE-juin06	0	0	0	0	0	0
0_GE-mai06	0	0	0	0	0	1
0_GE-avr06	1	0	0	0	0	0
0_GE-mars06	0	0	0	0	0	0
0_GE-févr06	0	0	0	0	0	0
0_GE-janv06	0	0	0	0	0	0
0_GE-nov07	0	0	0	0	0	0
0_GE-oct07	0	0	0	0	0	0
0_GE-sept07	0	0	0	0	0	0
0_GE-août07	1	0	0	0	0	0
0_GE-juil07	0	0	0	0	0	0
0_GE-juin07	0	0	0	0	0	0
0_GE-mai07	0	0	0	0	0	0
0_GE-avr07	0	0	0	0	0	0
0_GE-mars07	1	0	0	0	0	0
0_GE-févr07	0	0	0	0	0	0
0_GE-janv07	0	0	0	0	0	0
0_GE-nov08	1	0	0	0	0	0

FIGURE 5.19 – Données des combinaisons en format binaire.

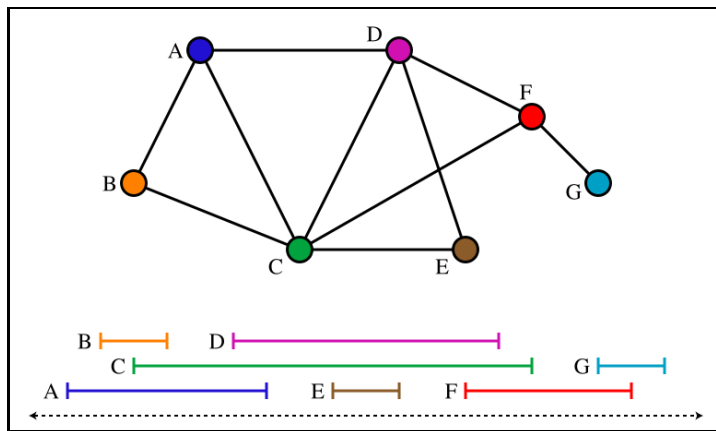


FIGURE 5.20 – Illustration d'un graphe d'intervalle.

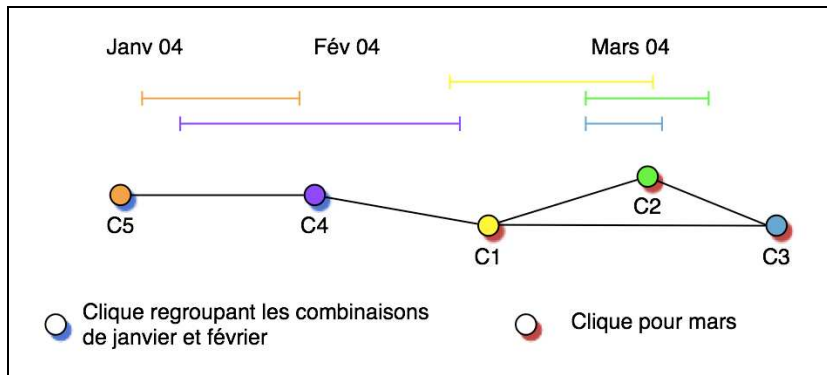


FIGURE 5.21 – Détection de cliques au moyen d'un graphe d'intervalle.

La méthode va permettre de mettre en évidence les combinaisons détectées durant une même période dans un même canton (figure 5.21). Les mois de saisies ont été utilisés comme unité de temps et le résultat montre les périodes “d'activité” des combinaisons. Chaque clique regroupe les assemblages détectés ensemble.

Le but de l'analyse est de voir si des combinaisons ne joueraient pas un rôle de traceur ou au contraire si certaines cesseraient d'être détectées et que d'autres apparaîtraient. Les résultats permettraient d'avoir en quelque sorte un calendrier des périodes de détection et de non-détection des combinaisons.

*Cette analyse a été réalisée au moyen du logiciel **Past**.*

Chapitre 6

Résultats et interprétation

6.1 Tendances générales

6.1.1 Introduction

Deux méthodes différentes ont été utilisées dans le but de mettre en évidence des tendances générales :

- Des graphiques statistiques sur les valeurs numériques des graphes afin de disposer d'une vue d'ensemble de leur évolution dans le temps.
- Une application de la théorie des ensembles afin de dissocier des caractéristiques de coupage selon les cantons de saisies.

Il est possible que les valeurs de fréquence d'apparition de certaines substances ou combinaisons reflètent davantage l'activité policière durant une période de temps ou une région, plutôt que l'activité criminelle. Les stupéfiants non saisis pourraient modifier passablement les résultats, si ils étaient à disposition dans les données.

Pour mener une interprétation rigoureuse, des informations d'enquêtes devraient être utilisées. Ceci a pu se faire pour une petite partie de ce travail. Des problèmes de financement ont été rencontrés, ce qui n'a pas permis à un expert de traiter dans son travail de recherche [Ioset *et al.*, 2005] toutes les tendances détectées et de faire le lien avec l'activité policière. Ces données n'étant pas à disposition, l'interprétation laisse des zones d'ombre quant à certains pics mis en évidence dans les statistiques des graphes.

6.1.2 Nombre de produits et nombre de cliques

Cette analyse a pour but de visualiser des tendances générales présentes dans les données relationnelles liées aux produits de coupage et à leur combinaisons. Les graphes des assemblages de substances sont présentés dans les figures 6.1 et 6.2. Ceux des produits apportent peu d'informations visuelles et ne sont pas mentionnés.

Les paramètres traités sont le nombre de produits de coupage et de combinaisons détectés, le nombre de cliques et le degré des noeuds des réseaux.

La période de temps d'étude couvre les années 2000 à 2005 et les cantons de Genève, Vaud, Neuchâtel et Tessin. Les données du canton du Valais ont été occasionnellement traitées.

Les résultats obtenus sont présentés par les graphiques des figures 6.3 et 6.4.

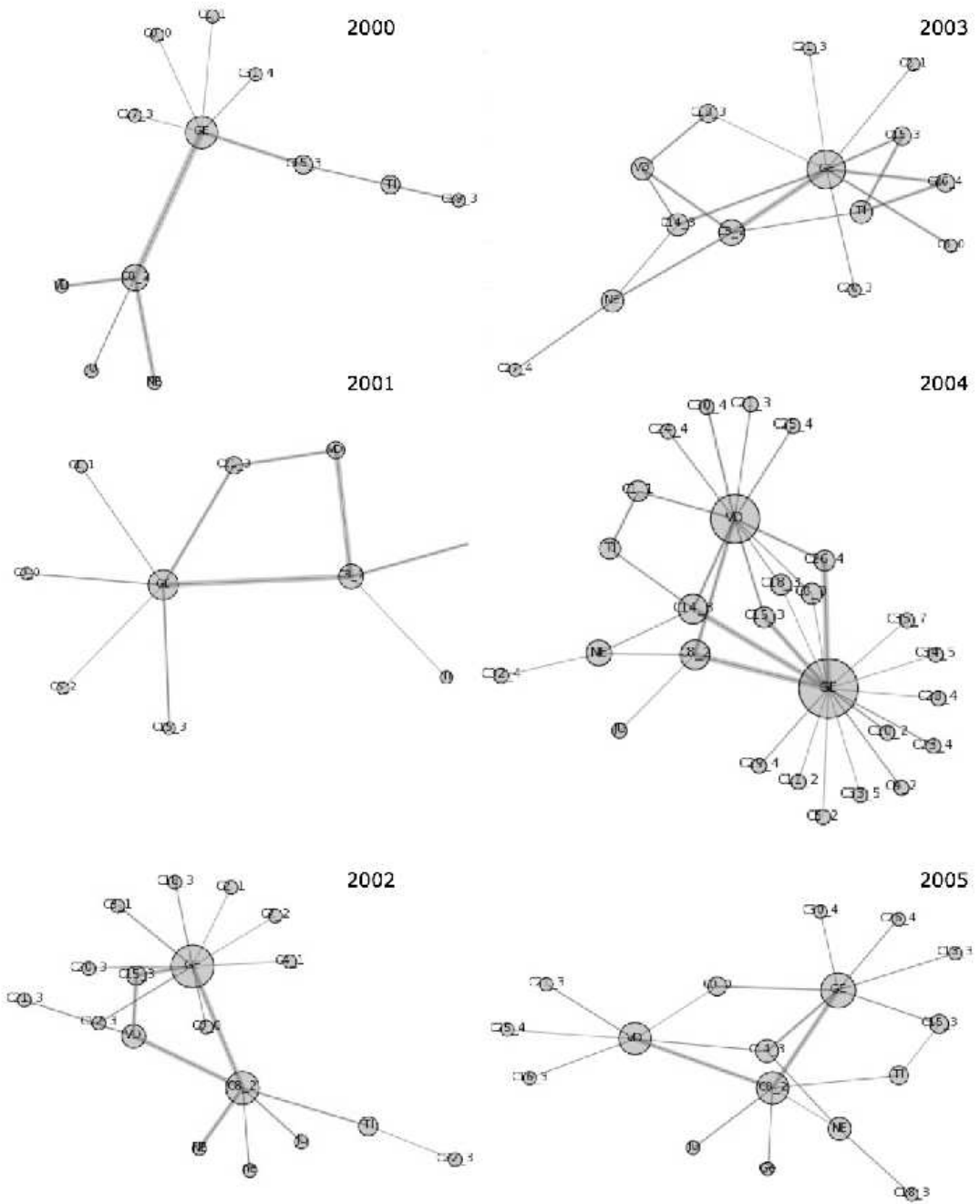


FIGURE 6.1 – Graphes des combinaisons de l'héroïne.

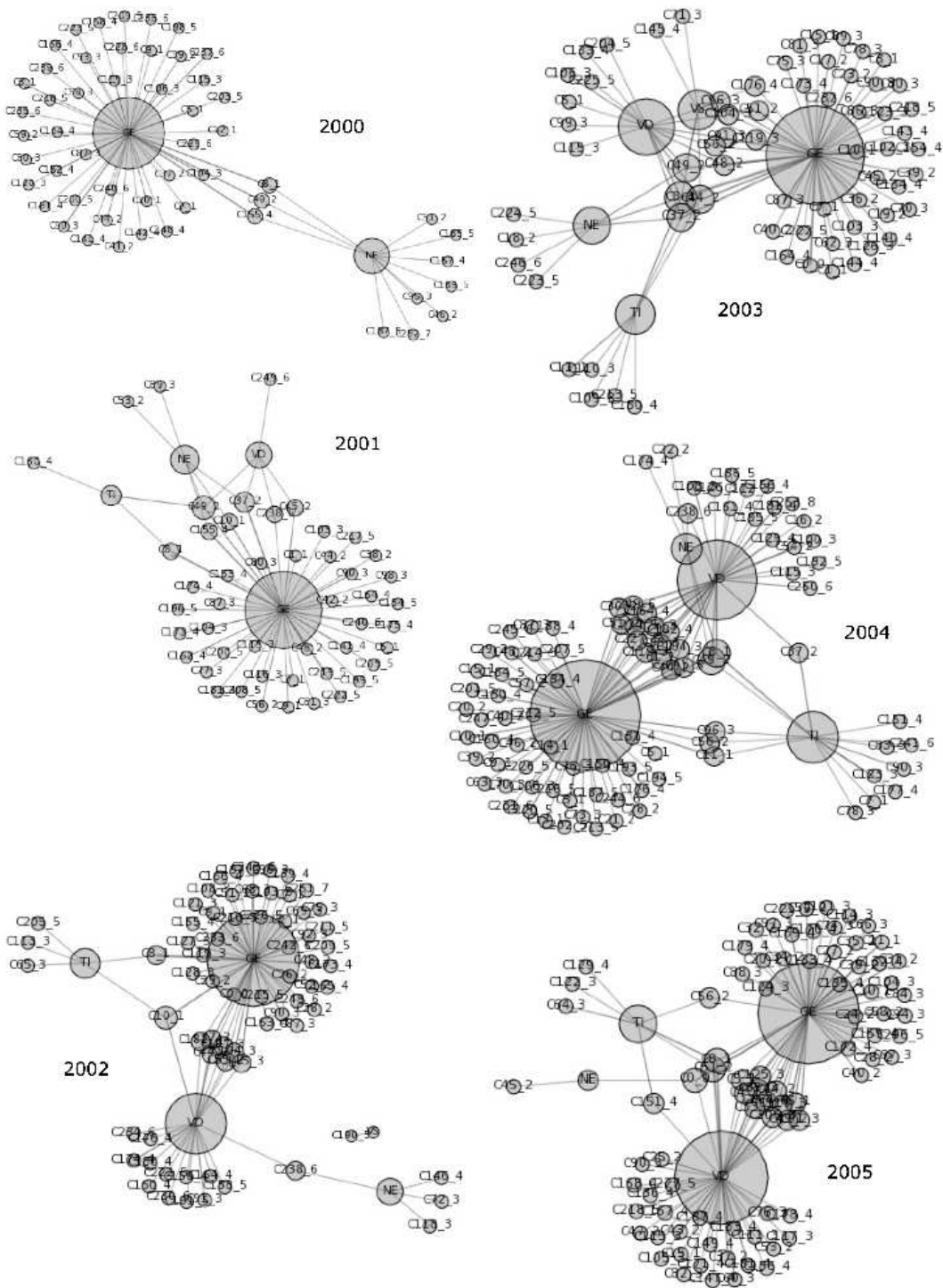


FIGURE 6.2 – Graphes des combinaisons de la cocaïne.

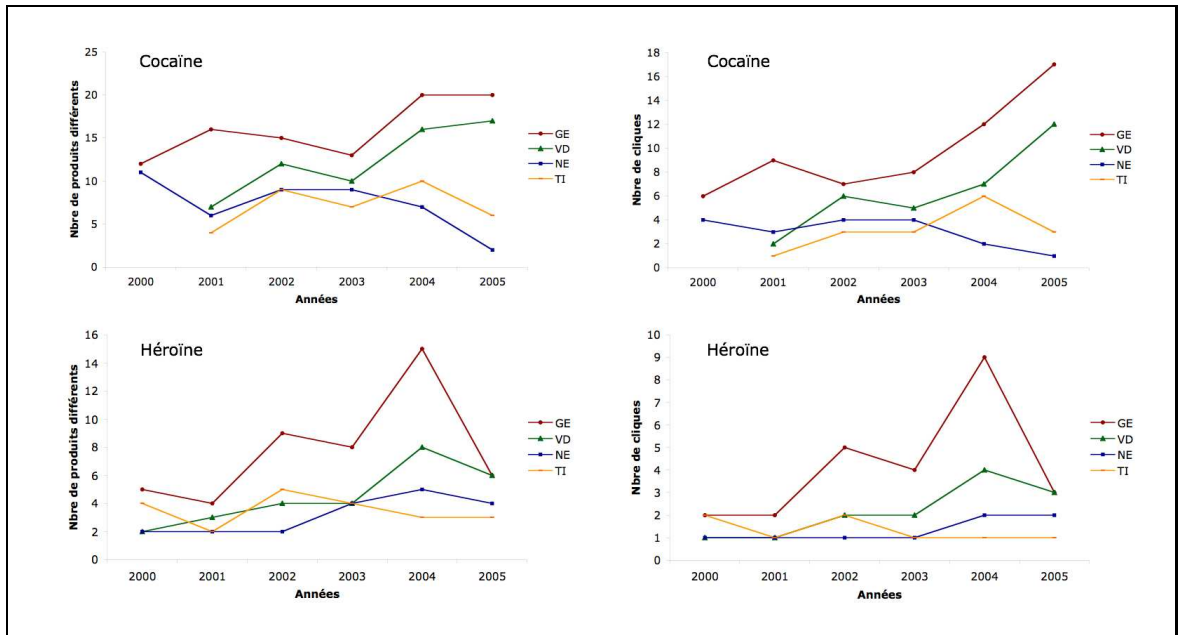


FIGURE 6.3 – Nombre produits et nombre de cliques.

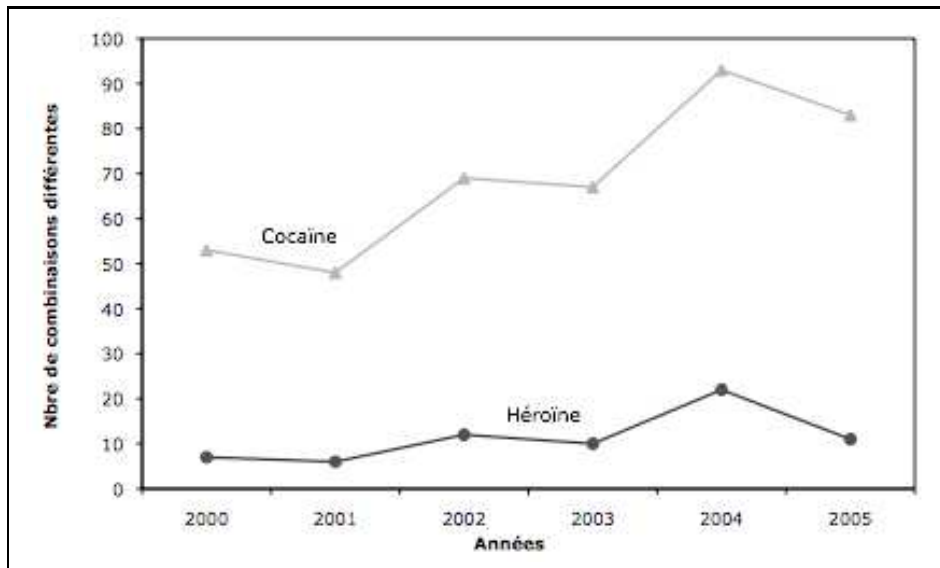


FIGURE 6.4 – Nombre de combinaisons.

6.1.3 Interprétation

Les graphes des combinaisons (figures 6.1 et 6.2) ont une taille de noeuds proportionnelle à leur degré (nombre de connexions) et l'épaisseur des liens correspond au nombre de fois qu'une combinaison a été détectée dans un canton pour l'année indiquée.

Pour les deux types de stupéfiants, les réseaux se complexifient d'année en année, avec une augmentation des échantillons saisis pour les cantons de Genève, Vaud et Tessin.

La variation du nombre de produits de coupage différents enregistrés par année et par canton (figure 6.3, graphiques de gauche), montre des courbes en phase pour les années 2000 à 2004, sauf pour Neuchâtel dont l'évolution semble être indépendante. Une séparation en 2005 entre Tessin - Neuchâtel et Vaud - Genève est observée. Pour l'héroïne, les courbes des cantons de Vaud et Neuchâtel diffèrent, alors que celles du Tessin et Genève sont en phase pour les années 2000 à 2003. On distingue un rapprochement des courbes de Genève et Vaud en 2004 et 2005. Une extension du trafic entre ces deux cantons a pu se produire, ce qui expliquerait cette séparation des courbes pour les deux types de stupéfiants et l'augmentation des combinaisons communes à ces cantons, visible sur les graphes. En 2000 et 2001, il est possible que le coupage de Genève et du Tessin corresponde à une même ramification de réseaux. Les stupéfiants se sont ensuite répandus à l'intérieur du pays, expliquant l'augmentation des produits détectés, probablement ajoutés sur le territoire suisse.

Les graphiques de droite montrent l'évolution dans le temps du nombre de cliques. Celles-ci sont une indication des possibilités de combinaisons des produits par canton. Les courbes devraient intuitivement être en phase avec celles du nombre de produits. C'est le cas pour l'héroïne. Pour la cocaïne, on observe une augmentation des possibilités de combinaisons en 2005 pour les cantons de Genève et Vaud contre une élévation modérée du nombre de produits. Ceci peut être expliqué par un mélange de plusieurs stocks de cocaïne ou/et par une nouvelle tendance de coupage ou le démarrage de nouveaux réseaux locaux. Les effets seraient une augmentation des possibilités de combinaisons, chaque dealer amenant sa "recette de cuisine", le nombre de produits de coupage resterait stable, car limité par la disponibilité des substances.

L'évolution du nombre de combinaisons par année (figure 6.4) montre des courbes en phase et en hausse générale pour la cocaïne et l'héroïne, traduisant une même tendance pour les deux types de stupéfiants. Cette observation est identique à celle faite pour l'évolution des graphes des figures 6.1 et 6.2. Ces résultats sont interprétés avec la seule vision des données. Les hypothèses émises sont peut-être loin de la réalité, la dimension de l'enquête de Police étant inconnue. De plus, les résultats doivent être lus en gardant à l'esprit qu'une partie seulement du phénomène est visible.

Synthèse

1. Pour les deux types de stupéfiants, une augmentation des produits est observée depuis 2000.

2. L'augmentation du nombre de cliques depuis 2000 traduit une augmentation des possibilités de combinaisons de produits, illustrée aussi par les graphes des combinaisons (figures 6.1 et 6.2).
3. Les cantons de Neuchâtel et du Tessin se démarquent de ceux de Genève et Vaud, montrant un pattern de coupage différent. Le nombre de produits diffère, de même que les possibilités de combinaisons, pouvant être expliquées par une position frontalière du Tessin et l'influence de réseaux issus du nord de l'Europe pour Neuchâtel.
4. Les cantons de Vaud et Genève, à partir de 2003, montrent une tendance similaire (courbes en phase). Une augmentation des échanges entre les deux cantons peut expliquer ce résultat.
5. En 2005, pour la cocaïne, une augmentation du nombre de combinaisons mais pas du nombre de produits a été observée. Elle peut indiquer l'émergence d'une nouvelle tendance de coupage, d'un mélange de stocks ou du démarrage de nouveaux réseaux. Les échantillons créant cette diversité ont pu être l'aboutissement d'une enquête ayant permis la saisie de cocaïne coupée de manière différente.

6.2 Quels produits dans quels cantons

Dissocier les produits liés à des régions différentes a été tenté en appliquant deux opérations de la théorie ensembliste aux graphes des produits. L'**intersection** met en évidence les substances communes à deux cantons ou plus ; le résultat obtenu par la **différence** montre les produits présents dans un canton et absents dans l'autre. Les résultats sont présentés sous la forme de graphes.

6.2.1 Produits les plus répandus dans les quatre cantons d'études

Les symboles mathématiques suivant sont utilisés :

- \cap pour l'**intersection**
- \setminus pour la **différence**

Une intersection successive des graphes des quatre cantons a été appliquée :

$$G_{ProdGen} = G_{Ti} \cap G_{Ge} \cap G_{Vd} \cap G_{Ne}$$

Les graphes résultants sont présentés dans les figures 6.5 et 6.6.

Pour l'héroïne, les triplets Paracétamol-Caféine-Griséofulvine et Paracétamol-Caféine-Glycérol sont les combinaisons les plus répandues. Cela signifie qu'elles ont été trouvées dans tous les cantons. Il est intéressant de remarquer que le lien Griséofulvine-Glycérol est absent, ce qui veut dire qu'il y a au moins un canton dans lequel ces deux produits n'ont jamais été rencontrés

ensemble dans un échantillon saisi par la Police. Si cette observation ne touche pas seulement les données à disposition mais tout le phénomène du trafic des stupéfiants, elle pourrait indiquer deux sources différentes d’approvisionnement d’héroïne. L’une contenant de la Griséofulvine et l’autre du Glycérol.

Selon les connaissances actuelles du trafic de l’héroïne, ces triplets sont très certainement les combinaisons “de base”, et correspondent à l’état dans lequel l’héroïne atteint l’Europe.

Concernant la cocaïne, le nombre de substances de coupage est beaucoup plus élevé, composé en grande partie des diluants Glucose, Lactose, Inositol, Mannitol, Sucrose, Caféine, des adultérants Phénacétine, Procaïne, Lidocaïne et de 1,4dmt (cf § 5.3.1, tableau 5.3).

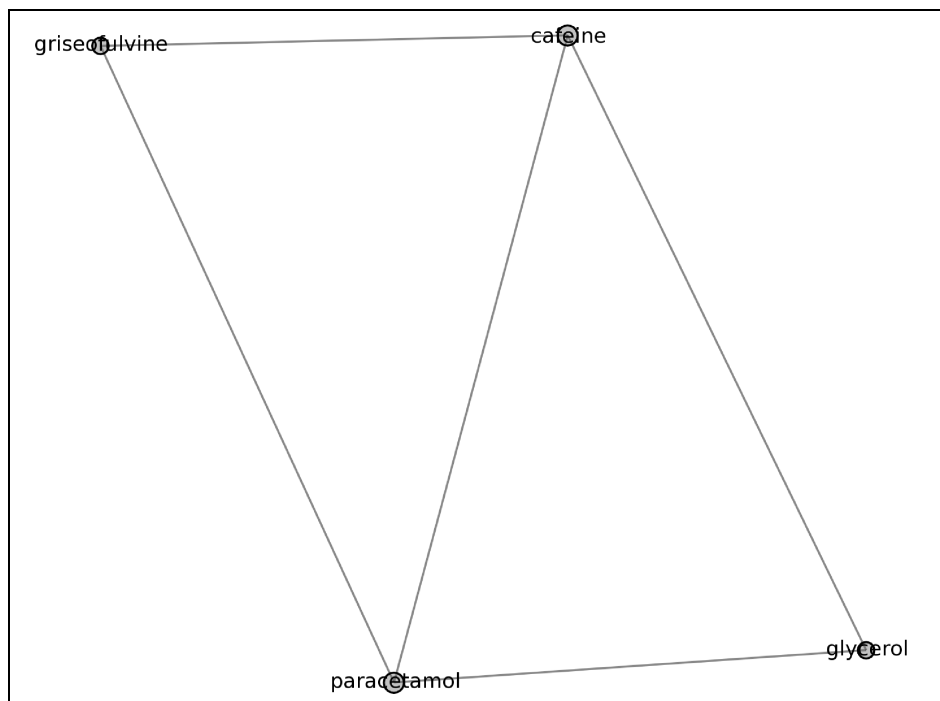


FIGURE 6.5 – Graphe $G_{ProdGenH} = G_{Ti} \cap G_{Ge} \cap G_{Vd} \cap G_{Ne}$ pour l’héroïne.

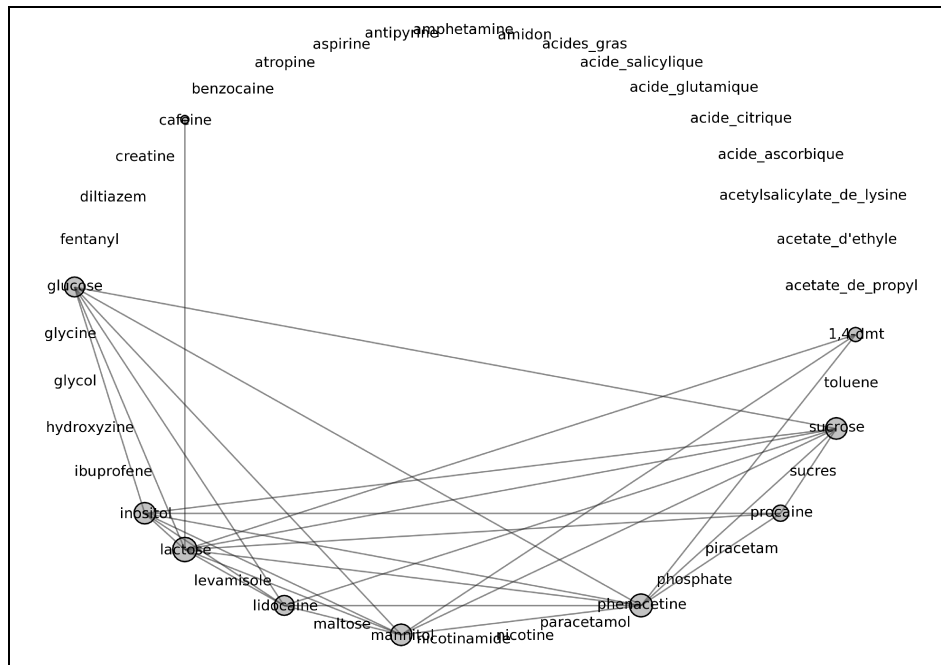


FIGURE 6.6 – Graphe $G_{ProdGenC} = G_{Ti} \cap G_{Ge} \cap G_{Vd} \cap G_{Ne}$ pour la cocaïne.

6.2.2 Produits ajoutés après le niveau de production

Les graphes des figures 6.7 et 6.8 ont été réalisés en appliquant la différence entre le graphe des produits de tous les cantons (les données ont toutes été réunies dans un même fichier d'entrée) et celui résultant de l'intersection successive des graphes de chaque canton. Le premier graphe, $G_{TiGeVdNe}$, représente la totalité des produits détectés dans les cantons d'étude. Le second, $G_{ProdGen}$ correspond aux produits communs à toutes les régions étudiées, donc aux produits qui pourraient être déjà présents dans le stupéfiant avant son entrée en Suisse. Le but de la différence est d'obtenir un graphe indiquant, selon l'hypothèse que les produits de $G_{ProdGen}$ soient présents avant de passer nos frontières, des substances ajoutées sur notre territoire :

$$G_{ProdCH} = G_{TiGeVdNe} \setminus G_{ProdGen}$$

Les graphes résultants montrent les produits détectés deux à deux dans trois cantons au maximum. Ils peuvent correspondre à des produits ajoutés à l'intérieur de nos frontières, mais il est impossible à ce stade de pouvoir déterminer où, en aval du niveau de production. Selon les connaissances actuelles du trafic, il est probable qu'une partie de ces adjonctions aient été faites dans des pays de transit et/ou de stockage des stupéfiants et une autre partie dans le pays où ils sont distribués (en Suisse).

Pour l'héroïne (figure 6.7), le graphe permet de visualiser les substances ajoutées aux triplets de base. Pour la cocaïne (figure 6.8), il montre une grande variété de produits et une grande complexité. Il est difficile d'extraire plus d'information de ce graphe. D'autres techniques issues du graph mining ont été appliquées afin d'extraire des informations sur les tendances locales de coupage et les résultats sont exposés plus loin à la section 6.3.

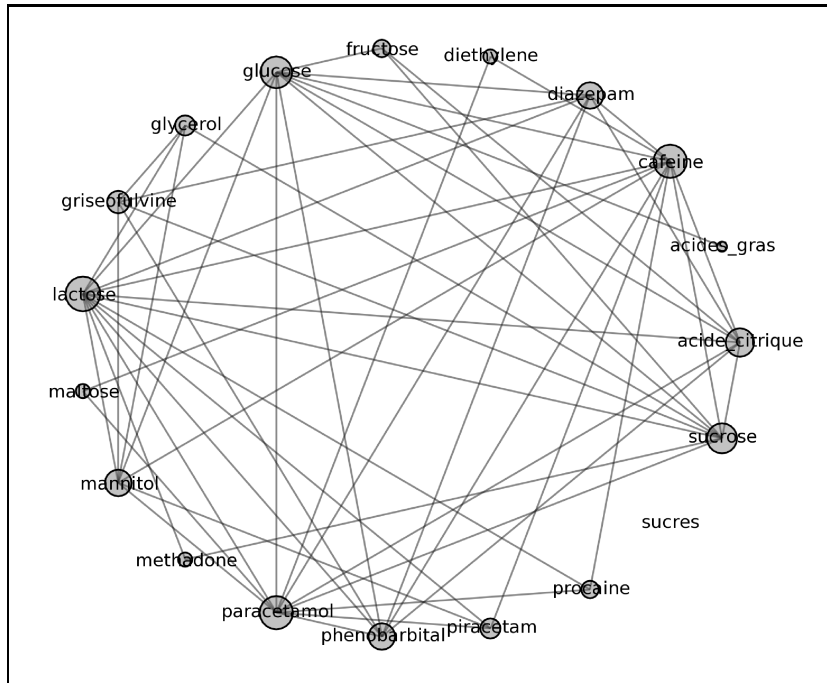


FIGURE 6.7 – Graphe $G_{ProdCH_H} = G_{TiGeVdNeH} \setminus G_{ProdGenH}$ pour l'héroïne.

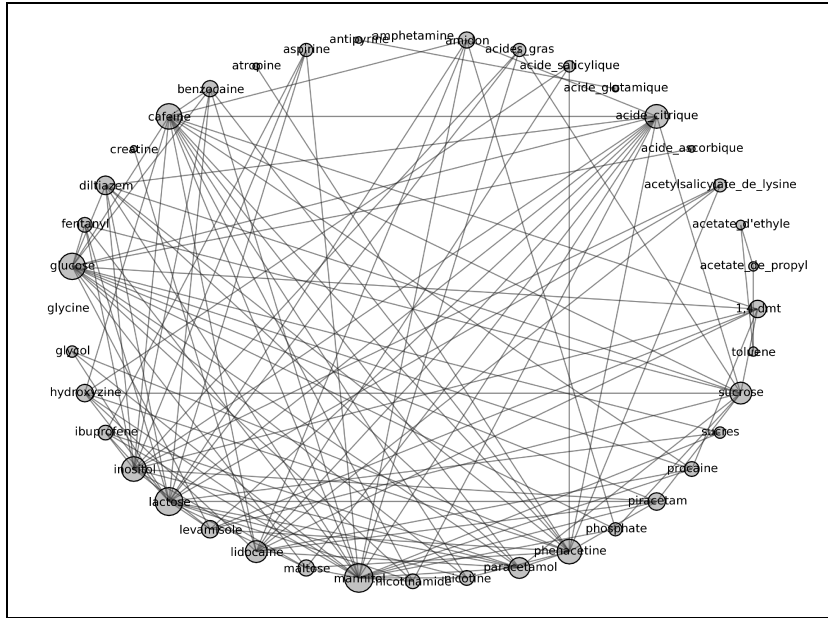


FIGURE 6.8 – Graphe $G_{ProdCHC} = G_{TiGeVdNeC} \setminus G_{ProdGenC}$ pour la cocaïne.

6.2.3 Relations entre les cantons

Les sections précédentes ont permis de mettre en évidence les produits les plus répandus et ceux pouvant correspondre à des adonctions faites en Suisse ou à un niveau en aval de celui de production. Dans cette section, les graphes des produits communs à deux et trois cantons ont été réalisés au moyen de l'intersection.

$$G_{CantonABC} = G_A \cap G_B \cap G_C.$$

Les substances détectées dans deux cantons, si elles n'appartiennent pas aux produits les plus répandus, pourraient correspondre à un trafic régional. Les régions testées sont l'Est (Ge-Ne et Vd-Ne) et le Sud (Ge-Ti et Ge-Vd). Les graphes d'intersection (Ti-Ne) et (Ti-Vd) n'ont pas été inclus dans ce chapitre car les résultats sont peu pertinents. Ils sont néanmoins placés en annexe. Les graphes des produits par canton, aussi placés en annexe, peuvent être utiles comme outil de comparaison.

Afin de faciliter la différenciation des substances rares de celles répandues, le tableau 5.3 du chapitre 5.3.1 a été repris ci-après. L'information des catégories a été ajoutée aux noeuds des graphes. Les catégories sont basées sur la fréquence d'apparition des produits et ne tiennent pas compte de leur répartition spatiale.

TABLE 6.1 – Catégories et répartition des produits de coupage (reproduction de la table 5.3)

Catégories	1	2	3	4
Cocaïne	lactose ^d mannitol ^d glucose ^d phénacétine ^a inositol ^d lidocaïne ^a sucrose ^d	caféine ^a procaïne ^a diltiazem ^a 1,4-dmt ^{*div} paracétamol ^a acide citrique ^a amidon ^d hydroxisine ^a piractam ^a benzocaïne ^a	lévamisole ^a sucres ^d glycine ^d acétylsalicylate de lysine ^a aspirine ^a atropine ^a lysine ^d maltose ^d phosphate ^d acide salicylique ^a antipyrine ^a créatine ^a éphédrine ^a fentanyl ^a ibuprofène ^a acide citrique ^a	acétate de propyl ^a acide ascorbique ^a acide glutamique ^a acides gras ^d aminogluthétimide ^a amphétamine ^a aspégic ^a glycérol ^d glycol ^d nicotinamide ^a nicotine ^a
Héroïne	caféine ^a paracétamol ^a griséofulvine ^a glycérol ^d sucrose ^d mannitol ^d	lactose ^d glucose ^d procaïne ^a phénobarbital ^a diazepam ^a	acide citrique ^a fructose ^d aspirine ^a phosphate ^d	naphtol ^a acides gras ^d maltose ^d méthadone ^a piracetam ^a

*1,4dmt = diméthylterephthalate, a = adultérant, d = diluant, div = divers

Les régions de l'Est de la Suisse

Pour l'héroïne, les figures 6.9 et 6.10 correspondant à l'intersection des graphes de Neuchâtel et Genève ainsi que Neuchâtel et Vaud, sont identiques. L'omniprésence du Paracétamol et de la Caféine donne aux graphes une structure en triplets. L'absence d'arête entre les sommets de la Griséofulvine et du Glycérol indique que ces produits n'ont jamais été détectés ensemble dans ces deux cantons confondus. En comparant les graphes de chaque canton (cf. annexes), c'est dans le canton de Neuchâtel que ces substances n'ont jamais été présentes dans un même échantillon.

Les graphes de la cocaïne (figures 6.11 et 6.12) sont très proches. On peut notamment relever que le Maltose a été détecté dans les cantons de Genève et Neuchâtel. Il s'agit, selon la classification proposée par le tableau ci-contre, d'un produit de catégorie 3. La Caféine et la Lidocaïne, de même que le 1,4dmt et l'Inositol, ont été détectés ensemble dans ces deux cantons mais pas sur Vaud.

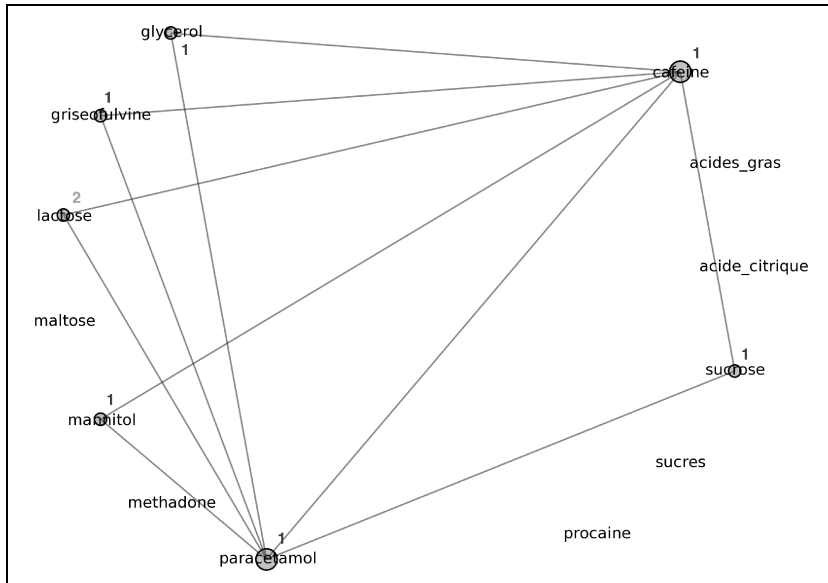


FIGURE 6.9 – Graphe $G_{ProdEst_H} = G_{Ge} \cap G_{Ne}$ pour l'héroïne.

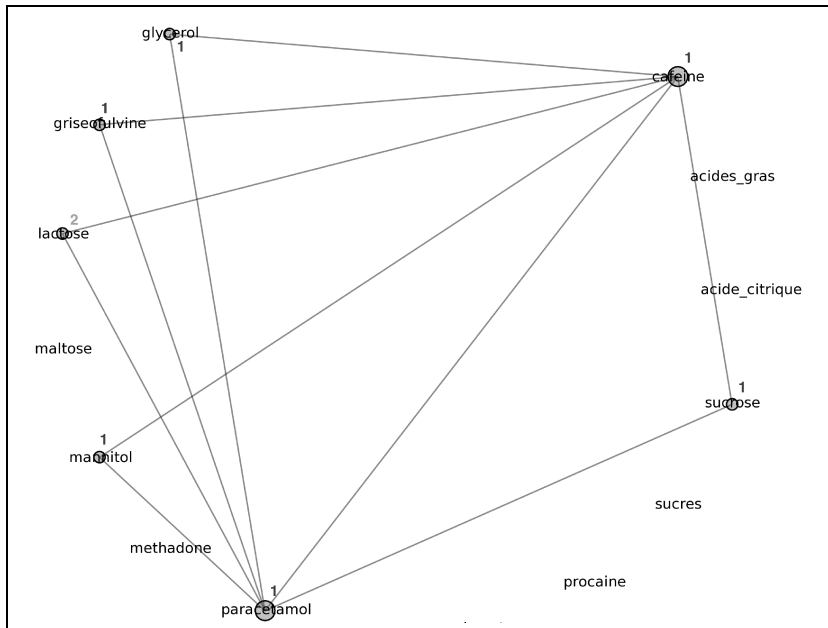


FIGURE 6.10 – Graphe $G_{ProdEst_H} = G_{Vd} \cap G_{Ne}$ pour l'héroïne.

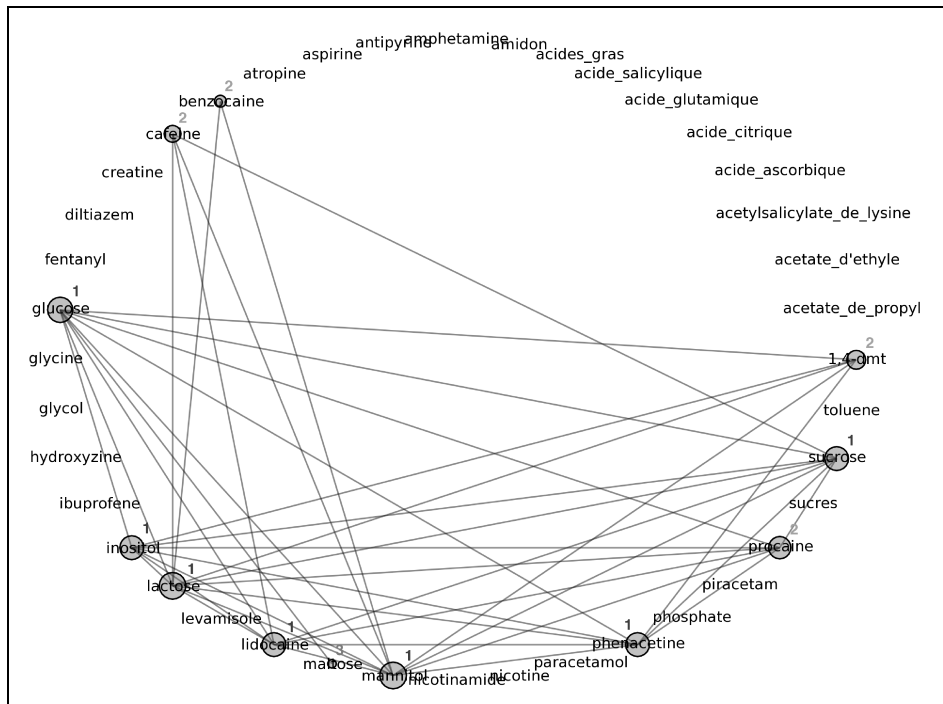


FIGURE 6.11 – Graphe $G_{ProdEst_C} = G_{Ge} \cap G_{Ne}$ pour la cocaïne.

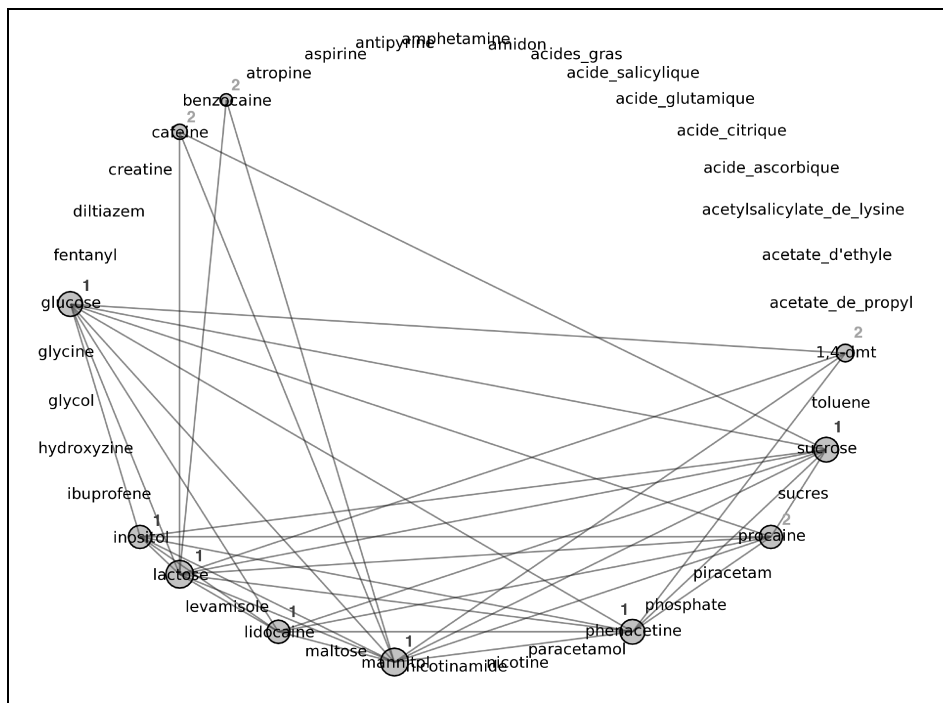


FIGURE 6.12 – Graphe $G_{ProdEst_C} = G_{Vd} \cap G_{Ne}$ pour la cocaïne.

Les régions du Sud de la Suisse

Le graphe relatif à l'héroïne pour Genève et le Tessin (figure 6.13) montre peu de produits en commun ; à relever toutefois la présence de Phénobarbital.

Les produits communs aux cantons de Vaud et Genève (figure 6.14) sont plus nombreux que ceux des autres graphes. Une hypothèse serait qu'il existe une plus grande activité liée aux stupéfiants dans ces deux cantons.

Pour la cocaïne, le graphe concernant le Tessin et Genève (figure 6.15) ne contient que le Sucre. La facilité d'accès à ce produit comme diluant ne permet pas de faire des hypothèses quand à un marqueur de réseau.

Le graphe de Vaud et Genève (figure 6.16) est plus complexe. La présence commune du Lévamisol de catégorie 3 selon la classification proposée, est à relever. L'Acide Citrique comme la présence du Sucre est difficile à expliquer en raison de sa facilité d'accès et sa présence dans beaucoup de substances.

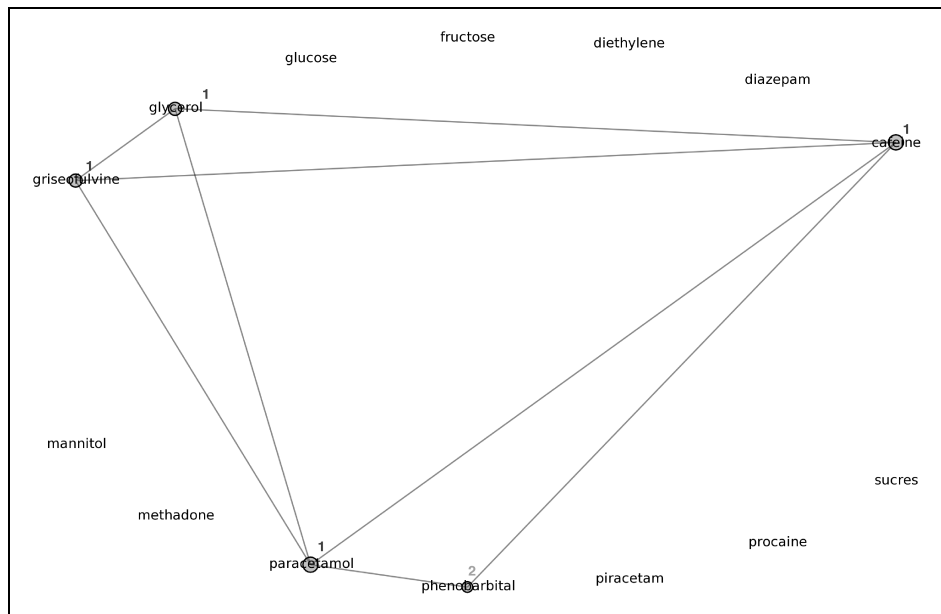


FIGURE 6.13 – Graphe $G_{ProdEst_H} = G_{Ti} \cap G_{Ge}$ pour l'héroïne.

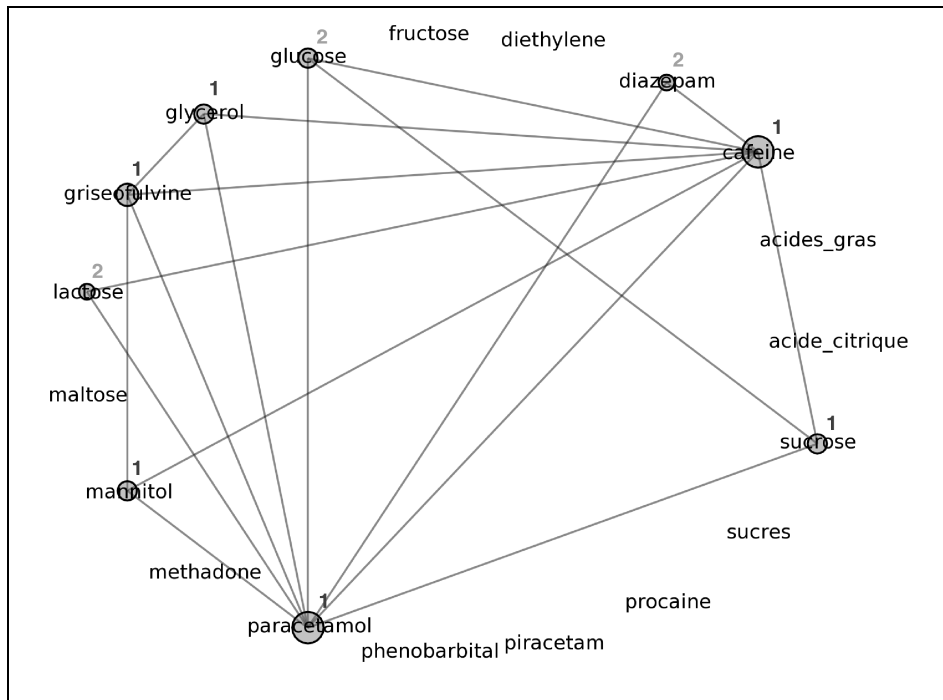


FIGURE 6.14 – Graphe $G_{ProdEst_H} = G_{Vd} \cap G_{Ge}$ pour l'héroïne.

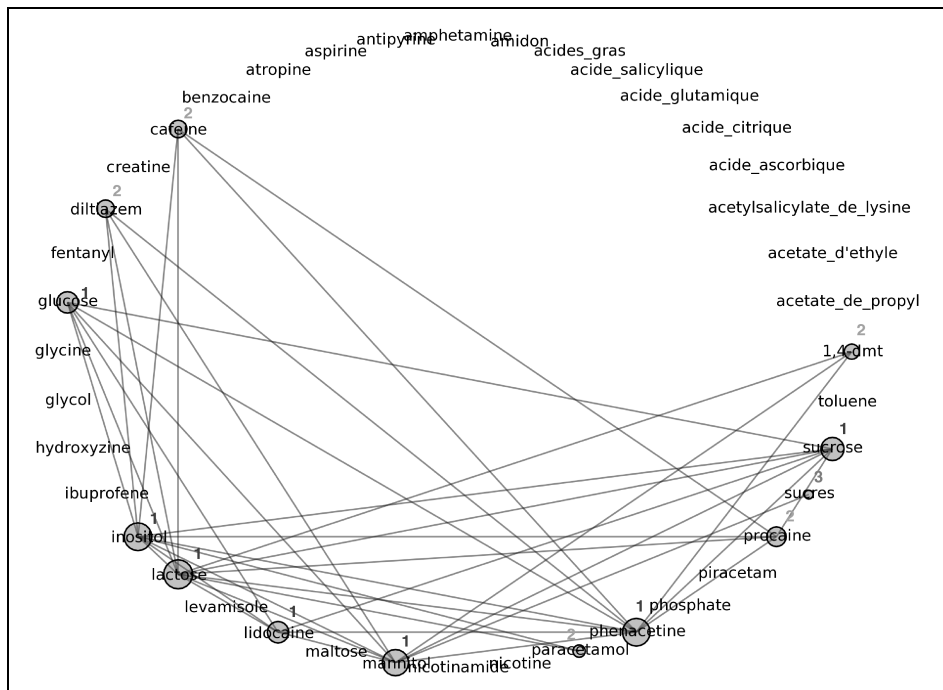


FIGURE 6.15 – Graphe $G_{ProdEst_C} = G_{Ti} \cap G_{Ge}$ pour la cocaïne.

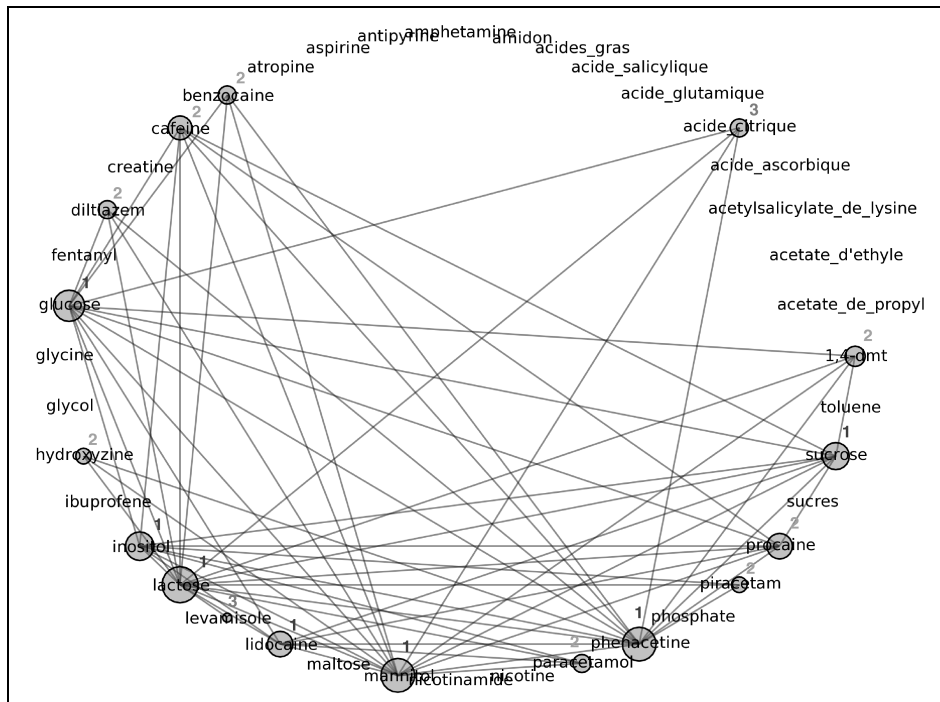


FIGURE 6.16 – Graphe $G_{ProdEst_C} = G_{Vd} \cap G_{Ge}$ pour la cocaïne.

6.2.4 Interprétation

La recherche des produits les plus répandus spatialement met en évidence les deux triplets de base des produits de coupage de l'héroïne (Paracétamol-Caféine-Griséofulvine et Paracétamol-Caféine-Glycérol). Il n'y a que dans le canton de Neuchâtel que la Griséofulvine n'a jamais été détectée en même temps que le Glycérol. Cette observation suggère la présence de deux filières distinctes. En contrôlant les données de base, le Glycérol est plus fréquemment présent dans de l'héroïne saisie à Neuchâtel et la Griséofulvine dans de l'héroïne saisie au Tessin. Selon le rapport mondial sur les drogues de l'ONU [ONUDD, 2006], il existe deux voies principales d'entrée en Europe; une voie Sud et une voie Nord. Comme il n'y a pas de raison chimique (incompatibilité) à ce que ces produits cohabitent dans un échantillon, il est possible que les triplets correspondent chacun à l'une de ces voies d'entrée. Il est aussi possible que des filières provenant de la même région initiale soient actives dans des zones géographiques indépendantes des pays de distribution. Les filières peuvent couper différemment leur stupéfiant et les distribuer dans des régions différentes.

Du Phénobarbital a été détecté dans des échantillons issus de Genève et du Tessin exclusivement. Ce médicament était, dans les années 1990, l'adultérant le plus fréquent après la Caféine [Gomez et Rodriguez, 1989], le Paracétamol étant absent. Cette tendance s'est modifiée au profit

du Paracétamol. Trois saisies ont été faites contenant du Phénobarbital : une à Genève et une au Tessin en septembre 2000, dont les échantillons ont tous la même classe chimique et une troisième à Genève en novembre 2004, de classe chimique indéterminée. Les échantillons de cette dernière saisie contiennent sept produits de coupage différents et une faible pureté. Un coupage local s'est très certainement produit. Il est possible que des stocks anciens et plus récents aient été mélangés et distribués en Europe par un réseau particulier, ce qui expliquerait la rareté du produit et sa présence conjointe avec le Paracétamol.

Les cantons de Vaud et Genève ont un nombre de produits en commun plus élevé que celui des cantons du Tessin et Neuchâtel. Des hypothèses peuvent être formulées pour expliquer ce résultat :

- Des stupéfiants issus de différents réseaux (plusieurs manières de couper l'héroïne) entrent en Suisse par Genève et des échanges ont lieu avec le canton de Vaud.
- Il existe plusieurs réseaux genevois, coupant l'héroïne, qui après coupage transite vers le canton de Vaud.
- Un même réseau principal alimente les deux régions. Selon les connaissances actuelles, de l'héroïne en provenance du canton de Zürich alimente les cantons de Vaud et Genève.

Les analyses, la situation géographique et stratégique des cantons de Genève et Vaud, ainsi que les connaissances du trafic corroborent les trois hypothèses. Tous les cantons partagent les triplets de base et le Tessin est aussi une voie d'entrée montrant une plus faible variabilité. Il est probable que les frontières genevoises soient des voies d'entrée d'héroïne issue de plusieurs horizons (grands axes routiers nord et sud, aéroport), apportant des caractéristiques différentes de coupage. Par la suite, elle serait coupée par des réseaux locaux puis une partie transiterait vers le canton de Vaud, dont la situation géographique le rend dépendant d'autres cantons (dont Zürich) pour son approvisionnement (la voie lacustre entre la France et la Suisse est, semble-t-il, peu utilisée).

Les produits de coupage de la cocaïne communs aux quatre cantons sont plus nombreux que ceux de l'héroïne et n'ont pas de structures particulières telles que les triplets. Plusieurs produits fréquemment utilisés pour couper la cocaïne sont mis en évidence. Les possibilités combinatoires sont élevées (noeuds des graphes avec de nombreux liens) et il est difficile de mettre en évidence des "combinaisons de base" pouvant correspondre à des tendances particulières liées à des réseaux d'approvisionnement différents. Certains adjuvants sont très fréquents et répandus, tels que la Phénacétine (souvent associée au Diltiazem) ou la Lidocaïne.

L'analyse des produits détectés dans une même région (Est et Sud) montre que les substances les plus fréquentes le sont aussi spatialement. La présence de produits de niveau 3 (rares) selon la classification proposée, a été mise en évidence. Le Maltose est commun à Genève et Neuchâtel, en présence notamment de Lactose et de Glucose. Ces trois substances constituent la base du lait

pour bébé. La facilité d'accès à ces produits suggère que la présence commune du maltose dans ces deux cantons est fortuite ; de plus les classes chimiques ne correspondent pas.

Du Lévamisol a été détecté sur Vaud et Genève. Il a aussi été signalé en France dans une cocaïne provenant du Venezuela (http://www.ofdt.fr/BDD/sintes/ir_050121_leva.pdf) saisie fin 2004. En Suisse, les saisies en contenant ont été faites en 2005. Depuis cette année, cette substance est plus fréquemment rencontrée. Une nouvelle tendance de coupage a pu se mettre en place ou une nouvelle filière écoule de la cocaïne contenant du Lévamisol en Europe.

Synthèse :

1. *Les triplets Paracétamol-Caféine-Glycérol et Paracétamol-Caféine-Griséofulvine de l'héroïne peuvent être issus de deux filières différentes Nord et Sud. Ils peuvent aussi être des marqueurs de réseaux, sans qu'une origine géographique différente soit la cause de la séparation Nord-Sud.*
2. *La présence occasionnelle de Phénobarbital, un ancien produit de coupage remplacé par le Paracétamol, avec ce dernier, peut correspondre à un mélange de stocks d'héroïne.*
3. *Le nombre de produits de coupage de l'héroïne communs aux cantons de Vaud et Genève est plus élevé que celui des cantons du Tessin et Neuchâtel. Les hypothèses émises sont :*
 - *une entrée par Genève d'héroïne issue de réseaux différents et un approvisionnement du canton de Vaud par Genève*
 - *un coupage local à Genève puis un transit vers le canton de Vaud*
 - *un approvisionnement commun des deux cantons par Zürich (connaissance actuelle).*
4. *Les produits de coupage de la cocaïne sont plus nombreux et variables, semblant être régis plus par un mode d'utilisation de certains produits que par des combinaisons "de base".*
5. *L'utilisation du Lévamisol comme adultérant pourrait être interprétée comme une nouvelle tendance initialisée en 2005.*

6.3 Tendances locales liées à l'espace et au temps

Les méthodes décrites à la section 5.3.2 ont été appliquées aux données liées au trafic de l'héroïne et de la cocaïne. Elles ont été testées dans le but de détecter des patterns "utiles" ou "à interpréter" (cf § 2.5.2), et de répondre aux questions posées précédemment, qui sont :

- Existe-t-il une évolution des produits de coupage au cours du temps et par région ?
- Existe-t-il des réseaux de distribution identifiables sur la base des données du coupage et des informations circonstanciées ?

- Est-il possible de mettre en évidence un transit de stupéfiants entre régions?

Les résultats obtenus suite à l'application des méthodes sélectionnées devraient permettre de répondre à ces questions. Seuls les graphes les plus pertinents sont illustrés dans ce chapitre; les autres sont placés en annexe. Ils ont été calculés au moyen de fonctions écrites en langage Python et avec l'aide des bibliothèques IGraph et Networkx. Les résultats graphiques ont été faits avec le logiciel Graphviz [Low, 2004] selon une mise en page radiale (les noeuds sont placés du centre vers la périphérie selon la décroissance de leur degré).

L'analyse débute par des observations faites sur un graphe "de base", contenant toutes les relations et dont les sommets importants sont mis en évidence au moyen de la méthode des points d'articulation. Il est la représentation graphique du jeu de données dans lequel des patterns vont être cherchés.

Les jeux de données ont été testés selon les dimensions spatiales (les localités) et temporelles (décomposition du temps en trimestres) :

- Combinaisons - Localités
- Combinaisons - Trimestres

6.3.1 L'héroïne

Les informations liées au contenu en produits de coupage ont été traitées sous forme de combinaisons. Les correspondances des codes pour l'héroïne sont placées en annexe.

Relations spatiales

Le graphe de base est présenté par la figure 6.17 et les points d'articulation apparaissent en orange.

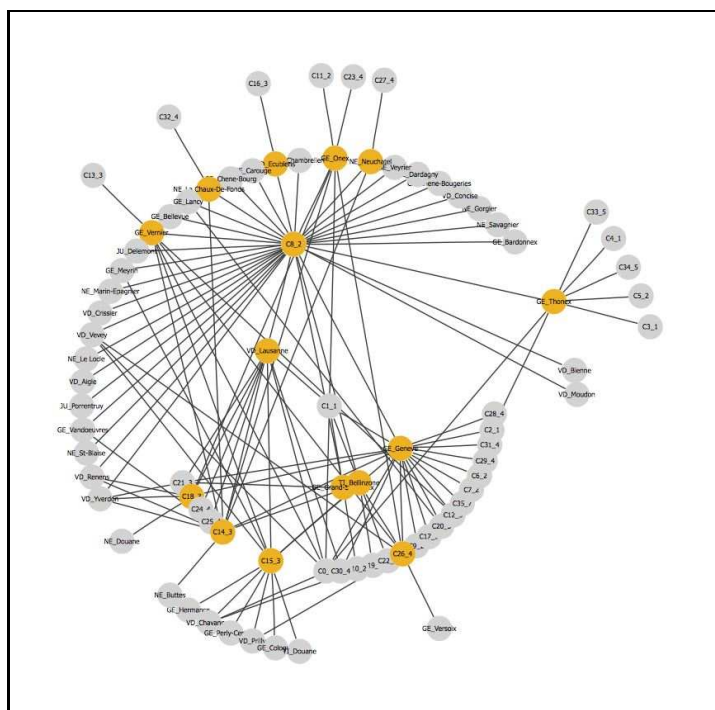


FIGURE 6.17 – Graphe de base “Combinaisons - Localités” avec points d’articulation.

Les sommets mis en évidence correspondent à des localités dans lesquelles le plus de combinaisons différentes ont été détectées et les combinaisons les plus répandues (ayant beaucoup de liens). Les localités avec la plus grande variabilité de combinaisons sont les grandes villes (Genève, Lausanne, Bellinzone). Les combinaisons les plus répandues sont celles mises en évidence précédemment à la section 6.2.

Un tel résultat n’est pas aisé à lire ni à interpréter en raison de sa complexité. Afin de diminuer le nombre de sommets et de liens, un traitement a été apporté aux graphes au moyen d’un algorithme de “minimum spanning tree”. Comme décrit aux sections 4.3.3 du chapitre 4 et 5.3.2 du chapitre 5, ce type d’algorithme d’optimisation se base sur les poids des arêtes d’un graphe pondéré. Les valeurs de pondération ont été créées à partir de l’information de pureté.

Chaque échantillon saisi est accompagné dans la banque de données de sa pureté, des composés majeurs, des produits de coupage traités sous forme de combinaison, de sa date et de son lieu de saisie. Le graphe de base à simplifier se compose des combinaisons détectées dans les localités pour la dimension spatiale et durant les trimestres entre les années 2000 et 2005 pour la dimension temporelle. Chaque arête du graphe relie une combinaison à une localité ou un trimestre. Des échantillons différents contenant un même assemblage de produits mais dont la pureté diffère, peuvent être saisis plusieurs fois dans un même lieu ou une même période de temps. Il est nécessaire

de disposer d'une seule valeur de pureté par arête. Il a été choisi de calculer la pureté moyenne des échantillons contenant une même combinaison, détectée dans une localité ou un trimestre. Le graphe résultant est illustré par la figure 6.18.

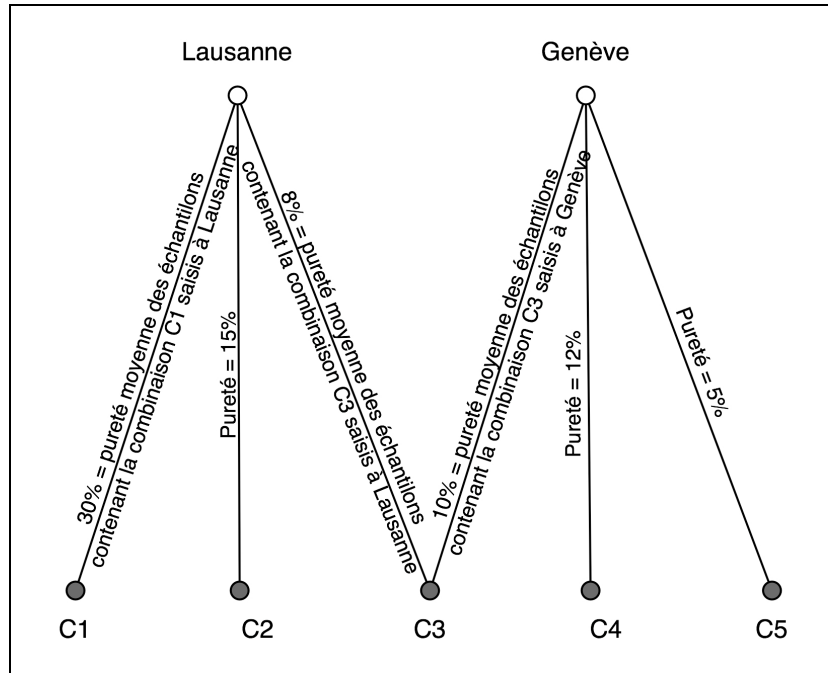


FIGURE 6.18 – Illustration d'un graphe pondéré par la moyenne de la pureté pour deux localités.

Le *minimum spanning tree* a pour but d'optimiser les arêtes d'un graphe dont les poids sont minimaux. Pour notre problématique, il serait intéressant de traiter non seulement les puretés minimales mais aussi les maximales. Pour se faire, deux calculs ont été faits :

- un *spanning tree* dont le poids des arêtes = **la pureté moyenne** par localité ou trimestre appelé "**minST**" (*minimum Spanning Tree*)
- un *spanning tree* dont le poids des arêtes = **l'inverse de la pureté moyenne** par localité ou trimestre appelé "**maxST**" (*maximum Spanning Tree*)

Cette dissociation permet de disposer de graphes simplifiés et plus lisibles et de leur appliquer d'autres tests. Il faut cependant garder à l'esprit que tous les liens et sommets ne sont pas présents. Les algorithmes appelés "walktrap"¹ et "fast greedy"² de détection de communautés ont été appliqués aux graphes "minST" et "maxST". Les deux algorithmes donnent des résultats proches, mais seuls les résultats du "walktrap" sont présentés, car la méthode est mieux adaptée à notre

1. L'algorithme "walktrap" utilise une probabilité proportionnelle aux poids des arêtes.

2. L'algorithme "fast greedy" enlève progressivement les arêtes avec le plus haut degré d'intermédiaires jusqu'à n'obtenir plus qu'un seul lien séparant deux communautés.

problématique au niveau du traitement des poids. Les résultats obtenus avec l'algorithme "fast greedy" sont placés en annexe. Ces traitements ont aussi été appliqués aux graphes des relations temporelles.

Les figures présentant les résultats sous forme de graphes ne permettent pas toujours de lire le nom des sommets, en raison de la taille du graphe. Pour y remédier, lorsque des patterns sont détectés, des extraits agrandissant certaines zones ont été faits afin de visualiser les entités composant les tendances.

Les résultats obtenus par détection de communautés sont présentés par les figures 6.19 et 6.20.

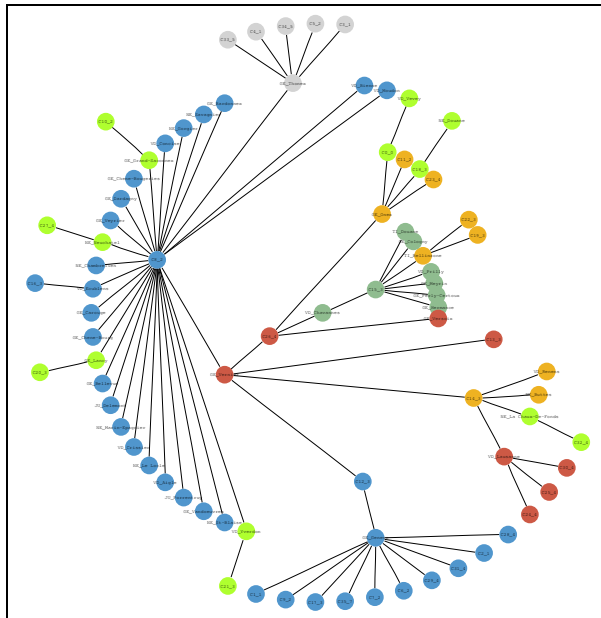


FIGURE 6.19 – Détection de communautés par "walktrap" appliqué au **minST "Combinaisons - Localités"**.

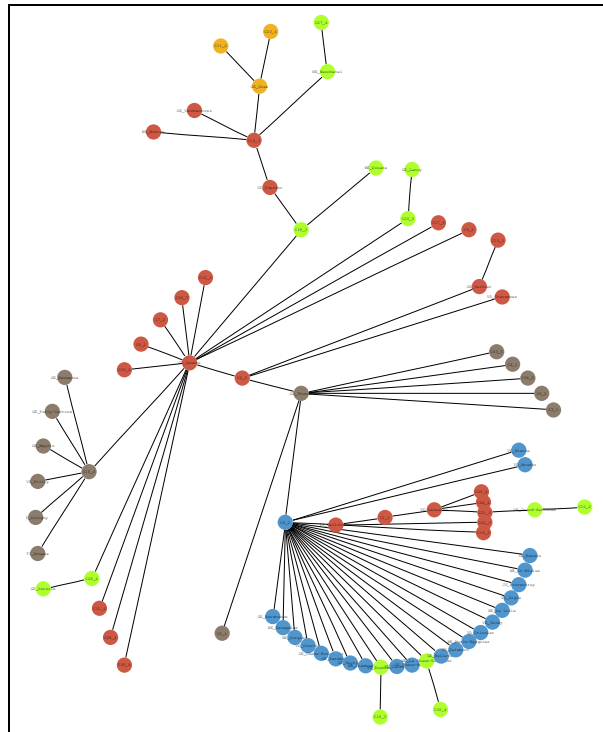


FIGURE 6.20 – Détection de communautés par “walktrap” appliqué au **maxST “Combinaisons - Localités”**.

Interprétation

Pour simplifier la lecture, la combinaison de base Paracétamol-Caféine est abrégée “P-C”.

Les spanning trees créés en fonction des puretés minimales et maximales ont une structure assez proche, montrant les relations entre localités et combinaisons en fonction de leur degré de pureté. Les communautés détectées correspondent aux lieux de saisies, ainsi qu’aux combinaisons rencontrées en plusieurs endroits. Celles-ci sont les combinaisons de base P-C (C8_2), détectées partout, P-C-Griséofulvine (C15_3) et P-C-Glycérol (C14_3). En analysant le contenu des combinaisons liées aux localités, il est possible de mettre en évidence une régionalisation de certains produits, additionnés à la combinaison P-C ou détectés seuls. Dans chaque canton, des combinaisons détectées uniquement dans cette région sont mises en évidence. Le canton de Genève est le seul à posséder des échantillons ne contenant ni Paracétamol ni Caféine. Il est possible que la présence d’un aéroport international en soit la cause.

La figure 6.21 ci-dessous est une synthèse de ces observations. Seuls les produits ajoutés à la combinaison de base sont mentionnés. Lorsqu’il n’y a pas de Paracétamol ou de Caféine, le symbole \emptyset est ajouté devant le nom des substances.

Code	Produits	Nord (Neuchâtel + Yverdon)	Vaud	Genève	Tessin	Localités
C0_0	ø					
C1_1	øcafeine					
C2_1	øglucose					Thônex
C3_1	ølactose					Thônex
C4_1	øsucres					Thônex
C5_2	øacides_gras-glucose					Thônex
C6_2	øcafeine-griseofulvine					Genève
C7_2	øcafeine-lactose					Genève
C10_2	ølactose-mannitol					Grand-Saconnex
C11_2	ølactose-procaine					Onex
C22_3	ølactose-sucrose-methadone					Bellinzone
C8_2	cafeine-paracetamol					
C14_3	glycerol					
C26_4	glycerol-griseofulvine					
C33_5	glycerol-griseofulvine-mannitol					Thônex
C27_4	glycerol-lactose					Neuchâtel
C28_4	glycerol-mannitol					Genève
C29_4	glycerol-sucrose					Genève
C15_3	griseofulvine					
C30_4	griseofulvine-mannitol					Genève - Lausanne
C32_4	griseofulvine-sucrose					La Chaux-de-Fond
C31_4	griseofulvine-phenobarbital					Genève
C24_4	griseofulvine-diazepam					Lausanne
C16_3	lactose					Ecublens
C34_5	lactose-mannitol-piracetam					Thônex
C18_3	mannitol					
C21_3	sucrose					Yverdon-Lausanne-Genève
C17_3	maltose					Genève
C20_3	procaine					Genève-Lancy
C19_3	phenobarbital					Bellinzone
C12_3	diazepam					Vernier
C13_3	diethylene					Vernier
C9_2	glucose-mannitol					Genève
C25_4	glucose-sucrose					Lausanne
C23_4	acide_citrique-fructose-glucose-sucrose					Onex
C35_7	acide_citrique--diazepam-glucose-lactose-phenobarbital					Genève

FIGURE 6.21 – Synthèse de la répartition des produits.

L'analyse met en évidence les produits très répandus qui, ajoutés à P-C, pourraient faire office de dénominateurs communs à certaines régions : le Glycérol (détecté partout), la Griséofulvine (Centre et Sud), le Mannitol et le Sucrose (Nord, Vd, Ge). La Griséofulvine n'est pas détectée au Nord et il est intéressant de remarquer qu'un mélange P-C-Griséofulvine-Glycérol a été détecté dans de l'héroïne sur Vaud et Genève uniquement.

Neuchâtel et Yverdon ont six assemblages différents dont les plus fréquents sont P-C, P-C-Glycérol, P-C-Mannitol. Deux assemblages seulement sont exclusifs à cette région : P-C-Glycérol-Lactose et P-C-Griséofulvine-Sucrose. Leur exclusivité suggère une adjonction locale du Lactose et du Sucrose. La présence de Griséofulvine dans cette région est inhabituelle. Après vérification dans les données de base, cette combinaison a été observée dans un seul échantillon en 2004 dans le canton de Neuchâtel. La combinaison P-C-Glycérol-Lactose a été détectée dans cinq échantillons d'une saisie.

Le canton de Vaud a douze assemblages différents. Par son emplacement central, son approvisionnement dépend des cantons frontières. Les connaissances actuelles indiquent une provenance de Zürich, mais il est possible, comme décrit à la section 6.2.4, que Genève soit une voie

d'entrée et un lieu de coupage et que l'héroïne transite vers le canton de Vaud. Trois combinaisons ne sont observées que dans ces deux cantons (\emptyset , P-C-Glycérol-Griséofulvine, P-C-Griséofulvine-Mannitol). Trois autres sont exclusives au canton de Vaud : P-C-Lactose, P-C-Griséofulvine-Diazepam, détectées une fois dans une affaire et P-C-Glucose-Sucrose, détectée dans deux affaires en 2004 et 2005. Cette dernière combinaison est intéressante car détectée deux fois dans la même ville. Elle pourrait être un marqueur de réseau.

Le Tessin a sept combinaisons, dont deux sont endémiques. La présence dominante du Glycérol au Nord et de la Griséofulvine au Sud peut indiquer un mélange d'héroïne provenant de deux filières principales lorsque les deux produits sont présents. La combinaison P-C-Phénobarbital peut correspondre, comme discuté précédemment, à un mélange de stocks, le Phénobarbital étant un ancien produit de coupage remplacé par le Paracétamol. La combinaison Lactose-Méthadone-Sucrose a été détectée une seule fois. Selon les informations d'enquête, la masse d'héroïne saisie était faible (0,18 g) ainsi que sa pureté (6,3 %). Elle pourrait correspondre à un achat fait par un consommateur en Italie et saisi lors d'un contrôle à la frontière.

Genève est le canton détenant le plus de combinaisons endémiques. L'héroïne ne contenant pas de P-C a probablement été saisie peu après son entrée en territoire helvétique. La présence de ces produits uniques sur différents axes d'entrée en Suisse, ainsi que dans des combinaisons saisies plus au centre de Genève, peut indiquer deux voies différentes, comme illustré par la figure 6.22. Au sud par Thônex entrerait une héroïne contenant du Lactose, du Glucose et des Sucres, au nord par le Grand-Saconnex, une héroïne avec Lactose-Mannitol. A Onex et Lancy, sur un axe routier passant par les frontières en direction du centre du canton, de l'héroïne contenant Lactose-Procaïne et P-C-Procaïne a été saisie. Dans la ville de Genève ainsi que dans les localités périphériques, un mélange de plusieurs de ces produits avec les combinaisons de base a été rencontré.

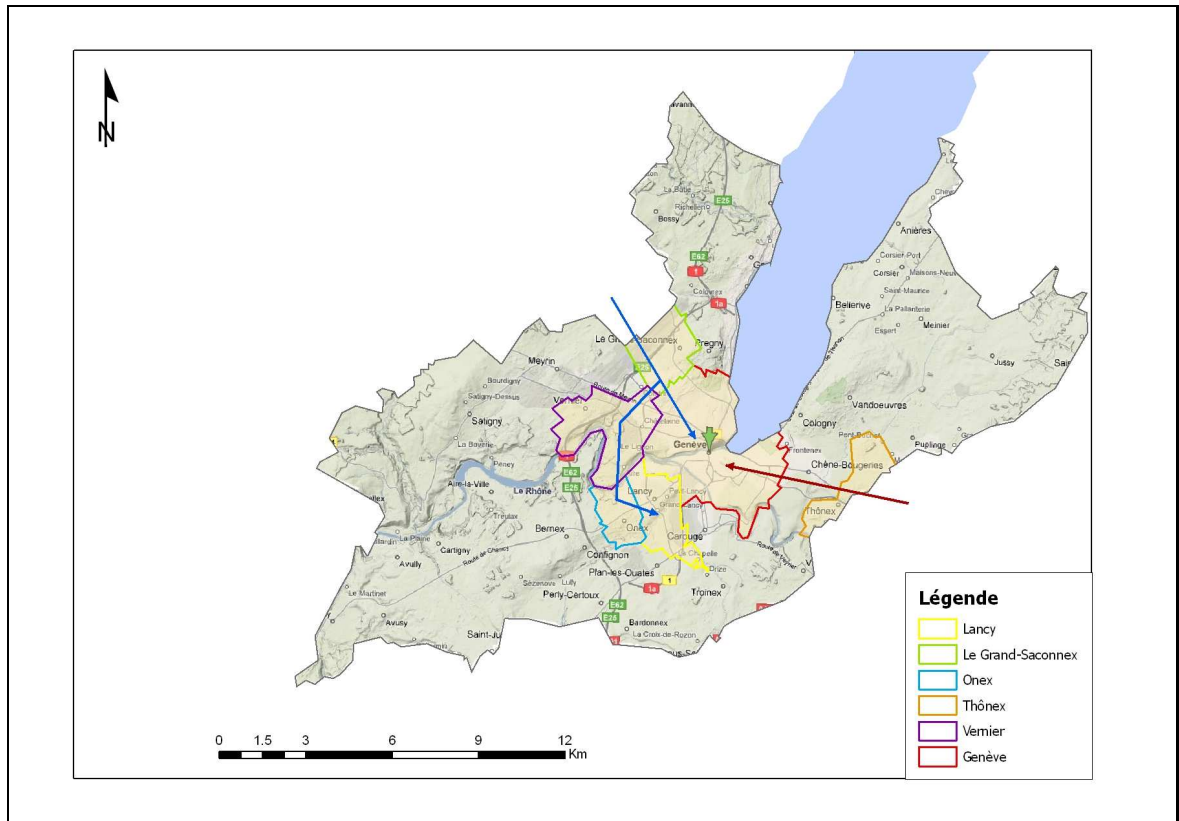


FIGURE 6.22 – Voies de passage de l'héroïne sur le canton de Genève.

Synthèse :

1. *L'utilisation du minimum spanning tree avec la pureté et l'inverse de la pureté comme poids, permet une bonne simplification du graphe initial. Il faut toutefois garder à l'esprit que tous les sommets ne sont pas présents, lorsqu'une combinaison a une pureté variable selon sa localité de saisie.*
2. *L'application des algorithmes de détection de communautés aux spanning trees a mis en évidence des groupes dont certains correspondent à des combinaisons endémiques et d'autres à des combinaisons répandues spatialement.*
3. *Chaque canton d'étude possède des combinaisons qui lui sont propres. Malheureusement peu d'entre elles ont été détectées plusieurs fois; seule une combinaison du canton de Vaud pourrait être liée à un réseau local.*
4. *Le canton de Genève possède beaucoup de combinaisons endémiques, notamment sans P-C. Celles-ci pourraient être liées à des voies d'entrée de l'héroïne en Suisse par le canton de Genève.*

5. Les combinaisons du canton de Vaud sont pour la majorité un mélange des produits détectés sur le canton de Genève.

Relations temporelles

Le graphe de base est présenté par la figure 6.23.

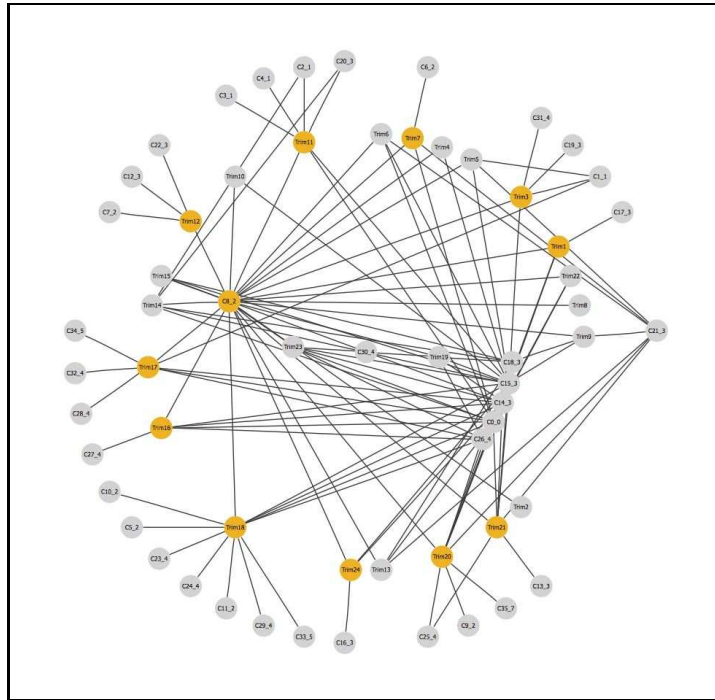


FIGURE 6.23 – Graphe de base “Combinaisons -Trimestres” avec points d’articulation.

La majorité des points correspondent aux trimestres. La combinaison Paracétamol-Caféine (C8_2) mise à part, il y a une plus grande variabilité de combinaisons détectées durant un trimestre que de combinaisons présentes de manière continue dans le temps. Ceci se voit sur le graphe par des sommets représentant les combinaisons ayant un seul lien les connectant à un trimestre. Les résultats obtenus par détection de communautés sont illustrés par les figures 6.24 et 6.25.

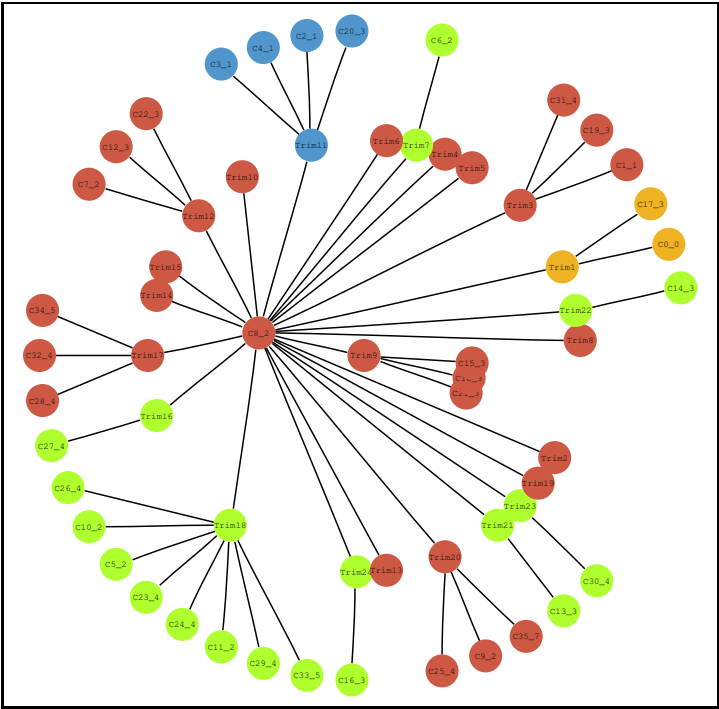


FIGURE 6.24 – Détection de communautés par “walktrap” appliqué au **minST** “Combinaisons - Trimestres”.

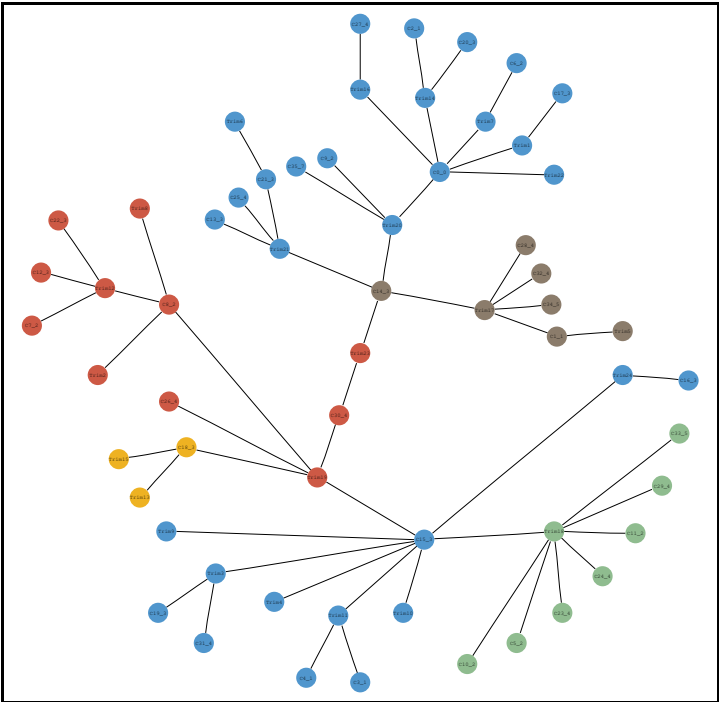


FIGURE 6.25 – Détection de communautés par “walktrap” appliqué au **maxST** “Combinaisons - Trimestres”.

Interprétation

La différence de structures entre minST et maxST est plus accentuée que pour les localités.

Le minST montre les combinaisons détectées durant un seul trimestre et dont la pureté tend à être minimale, le tout disposé autour de P-C détecté partout lorsque la pureté est faible. En majorité, les assemblages limités dans le temps sont aussi peu fréquents, ce qui indique qu'ils se situent probablement au bout du trafic. Si c'est bien le cas, la répartition temporelle serait alors identique à la répartition spatiale étudiée : les échantillons à faible pureté correspondant à l'extrémité de la chaîne du trafic sont en majorité peu fréquents voire uniques.

Les communautés détectées sont peu intéressantes car évidentes : elles correspondent aux combinaisons et à leur trimestre respectif.

Le maxST a une ramification plus complexe, dépourvue d'un sommet central (la combinaison P-C avec une pureté à tendance élevée est moins répandue dans le temps). Cette structure semble cohérente : une héroïne dont la pureté est plus élevée que la moyenne (qui est de 10-15%) n'est probablement pas au bout de la chaîne du trafic. Une autre phase de coupage devra probablement avoir lieu avant qu'elle ne soit vendue dans la rue. Elle se situerait au niveau d'un réseau national à cantonal, écoulée à des périodes différentes avant de subir un dernier coupage et d'être distribuée dans la rue.

Les communautés correspondent aux combinaisons détectées durant un ou plusieurs trimestres et dont les puretés sont similaires. Ce résultat permet de connaître les combinaisons peu répandues dans le temps et dont la pureté relativement élevée peut indiquer une tendance particulière. Il met aussi bien en évidence les combinaisons à pureté élevée, persistantes dans le temps.

L'analyse permet une visualisation facilitée, montrant des sous-graphes à explorer plus en détail, dont l'interprétation peut apporter des informations utiles pour la compréhension d'une dynamique temporelle. Ces sous-graphes sont indiqués sur la figure 6.26.

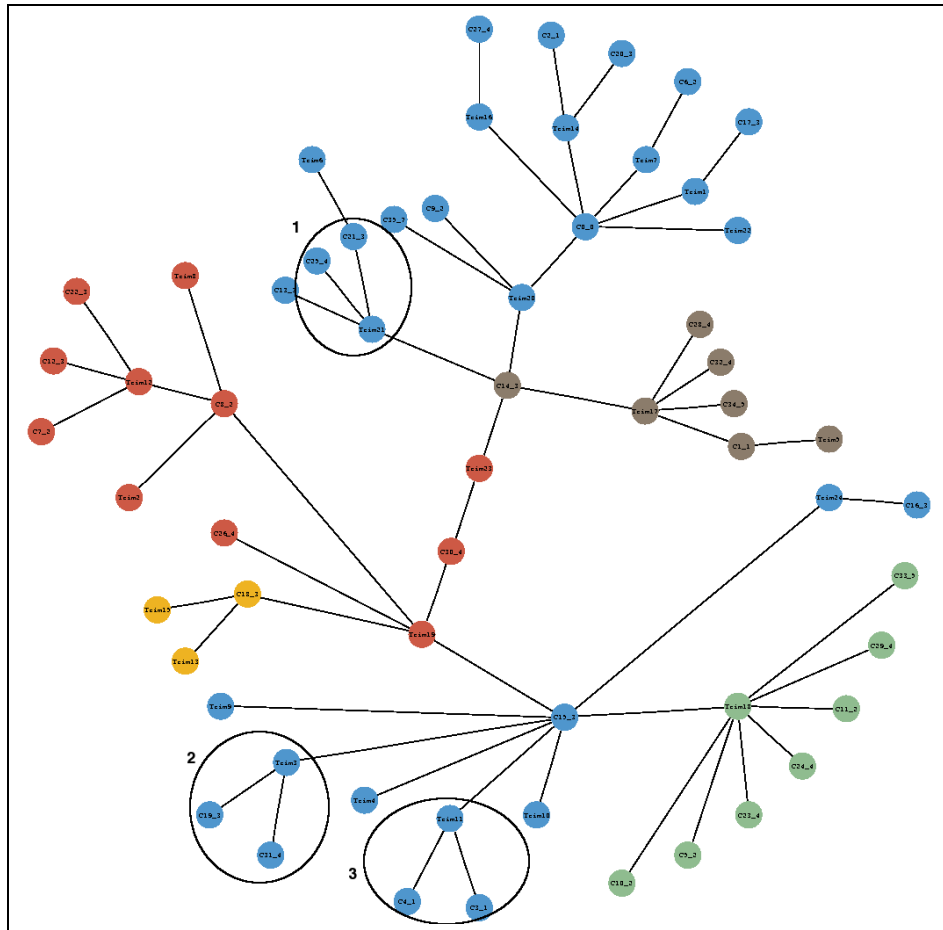


FIGURE 6.26 – Sous-graphes mis en évidence caractérisés par l'absence de P-C.

Les triplets sont composés de :

1. Les sommets C21_3 et C25_4 liés au trimestre 21 sont composés de P-C-Sucrose et P-C-Glucose-Sucrose, sans classe chimique.
2. Les sommets C19_3 et C31_4 liés au trimestre 3 sont composés de P-C-Phénobarbital et P-C-Griséofulvine-Phénobarbital, de classe chimique cl9.
3. Les sommets C3_1 et C4_1 liés au trimestre 11 sont composés de Lactose et de Sucres seuls, dont certains échantillons sont liés chimiquement.

Chacun de ces triplets est composé de deux combinaisons dont un produit est en commun (l'absence de P-C pour le dernier). Les échantillons ont une pureté proche, légèrement plus basse pour P-C-Glucose-Sucrose comparé à P-C-Sucrose. L'héroïne contenant P-C-Phénobarbital et P-

C-Griséofulvine-Phénobarbital a été saisie à Bellinzone le 29.09.2000 et à Genève le 28.09.2000. Celle-ci est intéressante, car elle pourrait correspondre à un mélange de stocks d'héroïne coupée différemment mais de même classe chimique. Selon les informations d'enquête, des dealers de même nationalité ont été arrêtés. Ils pourraient provenir d'un même réseau écoulant une marchandise mélangée. Les échantillons ne contenant pas de P-C ont une pureté supérieure à la moyenne et leurs seuls produits de coupage sont du Sucre et du Lactose, des substances très faciles d'accès. Il est possible que le stupéfiant soit issu d'un réseau particulier ne coupant pas l'héroïne avec P-C. Les informations d'enquête indiquent que les échantillons contenant les combinaisons C3.1 et C4.1 font partie d'une procédure contre une seule personne. Ces échantillons ont été saisis à Thônex, où passe l'autoroute A40, menant au carrefour de voies rapides reliant le Nord, le Sud et l'Ouest de la France. Dans la banque de données étudiée se trouvent cinq combinaisons ne contenant ni Paracétamol ni Caféine avec un ou deux produits, et dont la pureté est supérieure à 30% :

- Glucose : Thônex (pureté de 66%) et Genève (pureté de 35%)
- Lactose : Thônex (pureté de 35%)
- Sucres : Thônex (pureté de 52%)
- Acides gras - Glucose : Thônex (pureté de 60%)
- Glucose-Mannitol : Genève (pureté de 58%)

La saisie d'échantillons contenant du Glucose et Lactose le même trimestre est une indication supplémentaire aux corrélations géographiques. D'un point de vue opérationnel, ce résultat met en évidence le trajet possible d'un réseau d'approvisionnement marqué par l'absence de P-C, passant par Thônex et atteignant Genève. La personne mise en cause pourrait être un relais principal pour la Suisse.

Synthèse :

1. *La pureté modifie la structure des spanning trees selon son degré :*
 - *Le minST montre que lorsque la pureté tend à être basse, les combinaisons varient dans le temps, semblant indiquer un niveau du trafic proche de la rue.*
 - *Le maxST montre que lorsque la pureté tend à être élevée, les combinaisons répandues le sont aussi dans le temps, pouvant correspondre à un niveau du trafic en amont de celui de la rue.*
2. *L'application de la détection de communautés au minST est inutile car les groupes sont directement liés aux trimestres et visibles sans traitement.*
3. *Les communautés du maxST sont des sous-graphes dont une analyse plus approfondie des sommets a permis de mettre en évidence des tendances intéressantes :*

- Un possible mélange d'héroïne marqué par un produit particulier (Phénobarbital et Sucrose)
- La saisie d'héroïne sans P-C à pureté élevée un même trimestre corrobore l'hypothèse d'une voie d'entrée en Suisse par Thônex. L'héroïne liée à ce réseau d'approvisionnement serait caractérisée par l'absence de P-C.

Relations temporelles entre saisies et combinaisons : test du modèle des blocs

Le modèle des blocs est “à la mode”, notamment dans le domaine de l'analyse de réseaux sociaux. La structure du jeu de données à disposition ainsi que le phénomène du trafic des stupéfiants pourraient s'apparenter à un réseau social. Les échantillons de stupéfiants saisis étaient pour la majorité, possédés par des personnes. Un réseau de stupéfiants est généré non seulement par les substances en trafic, mais aussi par les acteurs du réseau. C'est selon l'hypothèse qu'un réseau de stupéfiants peut être vu comme un réseau social, que le test du modèle des blocs a été fait dans une optique d'analyse de réseaux sociaux (ARS). Il est cependant nécessaire de disposer d'un graphe non biparti (cf. 4.3.3 et 5.3.2). Pour ce faire, un groupe de sommets a été ajouté afin de disposer d'un graphe simple. Ce sont les saisies faites par la police qui contiennent plusieurs échantillons. Il devient alors possible de lier les sommets du groupe “combinaisons” à ceux du groupe “trimestres” et “saisies”, ainsi que les sommets du groupe “trimestres” à ceux du groupe “saisies”. Le schéma de la figure 6.27 illustre ces relations.

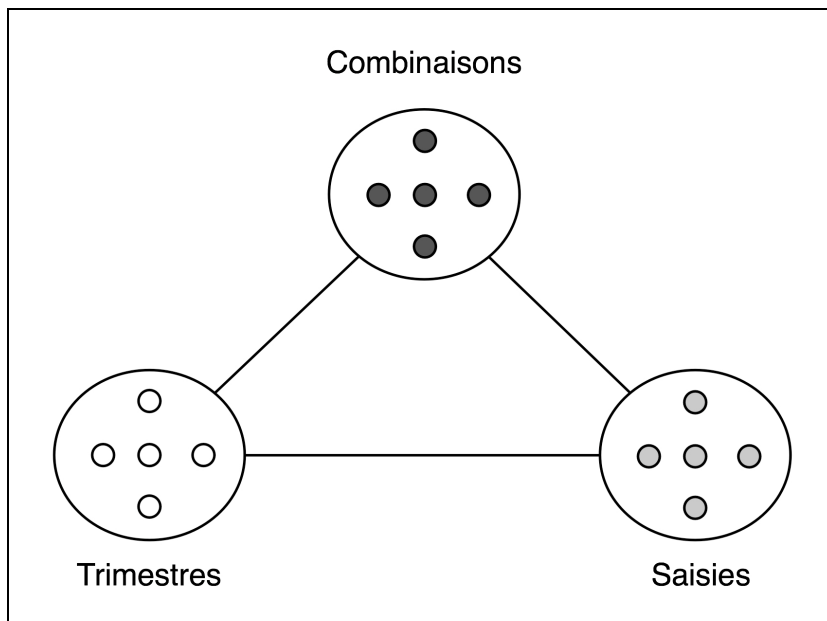


FIGURE 6.27 – Illustration des relations entre groupes de sommets pour l'application du modèle des blocs.

En utilisant toutes les données présentes dans le set initial, le graphe résultant est illisible. Une sélection aléatoire de 350 relations a été faite.

Il existe différents algorithmes de détection de blocs, permettant de mettre en évidence des structures. Les logiciels avec ces algorithmes implémentés nécessitent un graphe avec un nombre limité de sommets. Nos données dépassent largement ce nombre limite, ainsi l'algorithme utilisé est celui de la librairie IGraph. Les blocs mis en évidence sont les composantes biconnexes du graphe³.

Le graphe résultant est présenté par la figure 6.28. La lisibilité n'est pas aisée, il est cependant possible avec le logiciel Graphviz de représenter le graphe selon une autre mise en page, permettant d'analyser le résultat à l'écran.

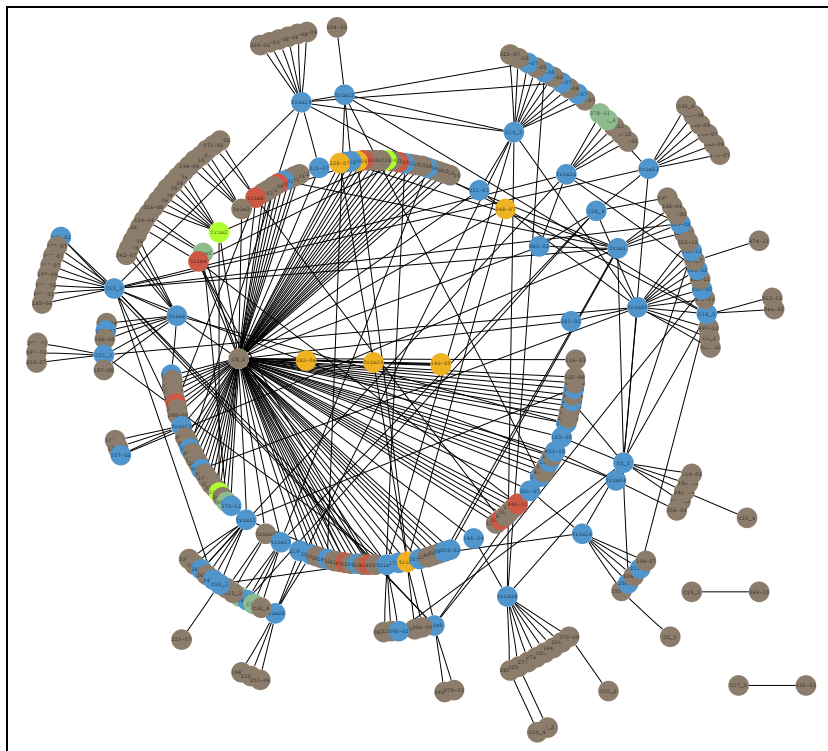


FIGURE 6.28 – Modèle des blocs appliqué au graphe “**Combinaisons - Trimestres + Combinaisons - Saisies + Saisies - Trimestres**”.

Le résultat met en évidence des sous-graphes dont les sommets sont structurellement équivalents, tels que toutes les saisies faites un même trimestre ou toutes les saisies contenant la combinaison P-C. La méthode a un potentiel mais n'est pas adaptée à la problématique. D'une part en raison du nombre de sommets trop élevé et d'autre part en raison de la relative “simplicité” du graphe.

3. Composante biconnexe = composante devenant connexe si on lui enlève un sommet ; composante connexe = sous-graphe dont les arêtes relient chaque sommet, donc possède un chemin (cf figure 4.7, chapitre 4.2.2).

Il n'y a que trois types de relations : Saisies-Trimestres, Combinaisons-Saisies et Combinaisons-Trimestres. Le modèle des blocs est potentiellement utile pour un graphe avec un nombre de sommets limité, mais dont les relations sont complexes et de tout type. Les sous-graphes avec une équivalence structurelle (les composantes biconnexes du graphe) seront alors beaucoup moins nombreux et interprétables. Dans notre cas, la structure de base du graphe crée de nombreuses composantes biconnexes sans information, provenant d'une part du fait que la majorité des saisies est ponctuelle dans le temps (liée à un seul trimestre) et d'autre part que certaines combinaisons sont détectées partout et en tout temps. Les blocs détectés correspondent à des groupes triviaux, dépourvus d'information nouvelle et utile. La figure 6.29 illustre le problème.

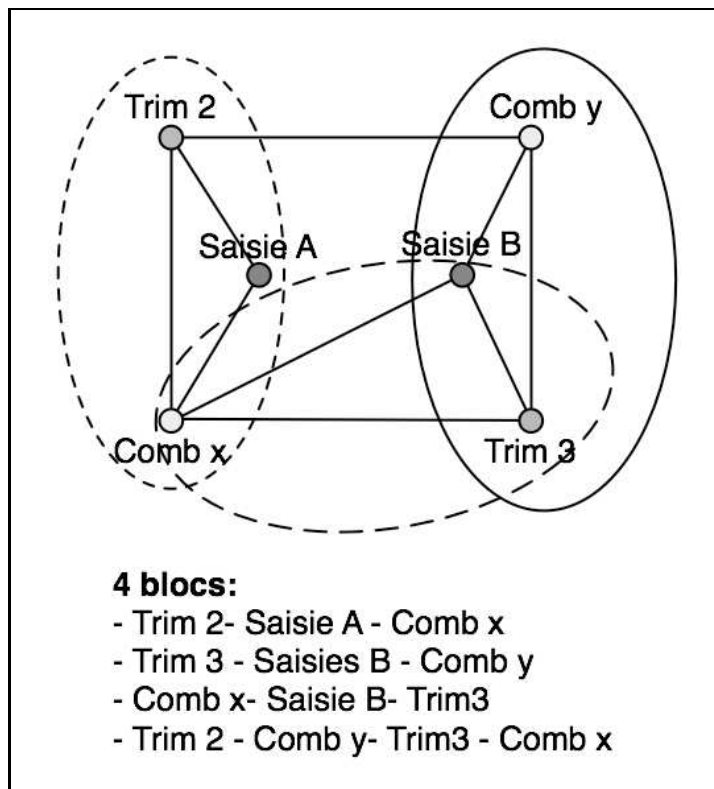


FIGURE 6.29 – Illustration de l'application du modèle des blocs à la problématique.

Pour la problématique du trafic des stupéfiants traitée au moyen des produits de coupage et des informations circonstancielles, cette méthode n'est pas optimale. Dans le cadre d'une analyse d'un cas d'enquête, en intégrant à la structure du graphe présenté dans ce test des informations de personnes, de téléphonie mobile etc..., l'analyse s'orienterait plus vers une ARS. Le nombre de données serait probablement plus restreint et il serait possible d'utiliser le modèle des blocs à des fins opérationnelles.

6.3.2 La cocaïne

Le nombre de combinaisons est beaucoup plus élevé que pour l'héroïne; la correspondance des codes est placée en annexe.

Relations spatiales

Le nombre de combinaisons des produits de coupage détectés dans des échantillons de cocaïne est très grand. Afin de disposer de graphes lisibles, les localités ont été traitées de manière séparée :

- Les communes des cantons de Genève-Vaud-Neuchâtel
- Les communes des cantons de Genève-Tessin

Les graphes “Combinaisons - Localités” de Genève et du Tessin apportent peu d'informations en raison du petit nombre de combinaisons communes entre le Tessin et Genève. Les graphiques sont placés en annexe. Seuls les résultats pour Genève, Vaud et Neuchâtel sont présentés dans cette section.

Le graphe de base avec les points d'articulation mis en évidence est présenté par la figure 6.30. Il n'est pas aisément lisible et sa présence ici a pour but de visualiser la complexité des relations.

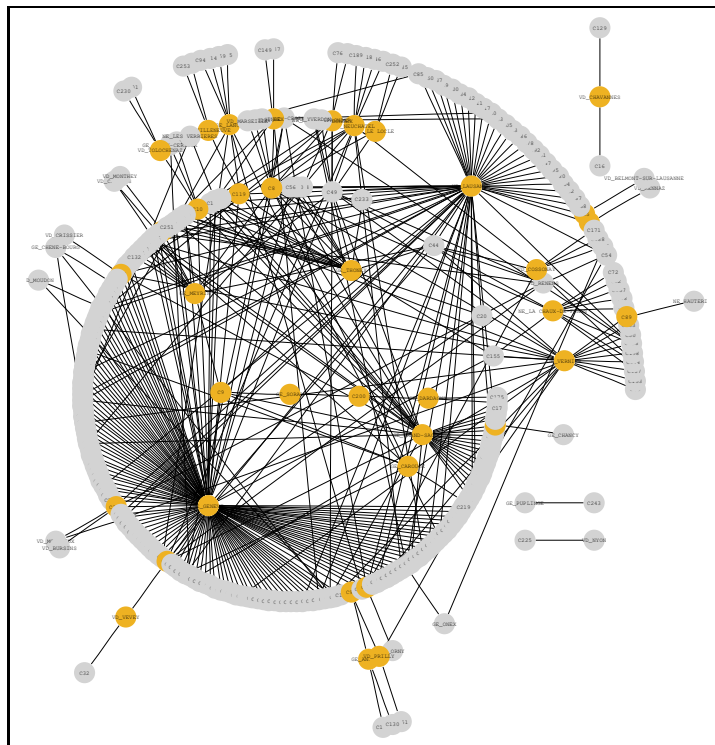


FIGURE 6.30 – Graphe de base “Combinaisons - Localités” avec points d'articulation pour Ge-Vd-Ne.

Les spanning trees permettent de simplifier le graphe initial (figures 6.30) et les algorithmes de communautés ont été appliqués aux minST et maxST. Les algorithmes “walktrap” et “fast greedy” ont été utilisés et les résultats obtenus avec le “fast greedy” montrent parfois plus de communautés qu’avec le “walktrap”. Lorsque c’est le cas, le résultat de l’algorithme “fast greedy” est placé dans le chapitre et celui du “walktrap” en annexe.

Les résultats sont présentés par les figures 6.31 et 6.32. Les noms des sommets ne sont pas lisibles, mais un agrandissement de certaines zones des graphes sera présenté dans l’interprétation.

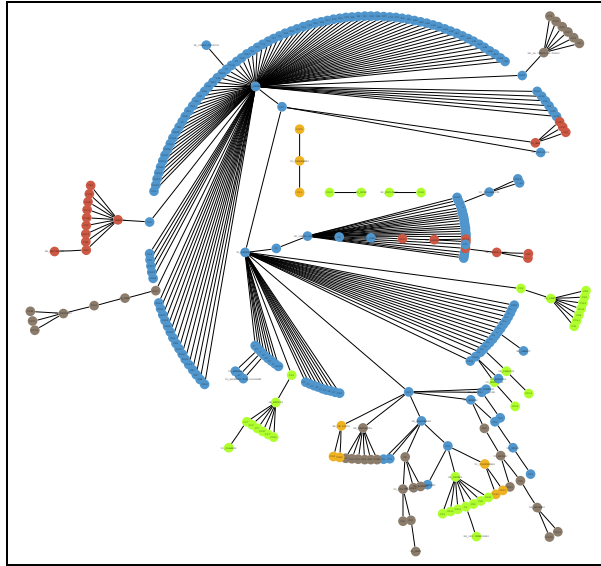


FIGURE 6.31 – Détection de communautés par “walktrap” appliqué au minST “Combinaisons - Localités” de Ge-Vd-Ne.

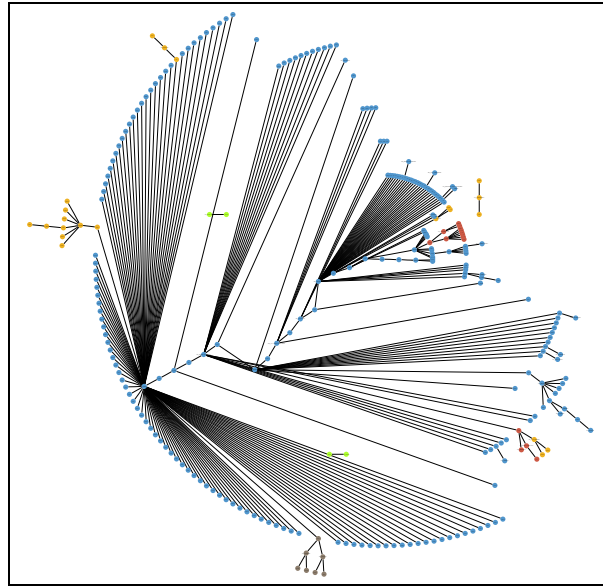


FIGURE 6.32 – Détection de communautés par “fast greedy” appliqué au **maxST** “**Combinaisons - Localités**” de **Ge-Vd-Ne**.

Interprétation

Les graphes minST et maxST ont des structures proches mais le maxST présente peu de communautés. Les spanning trees sont construits en optimisant les valeurs de pureté. Les structures générales semblables des minST et maxST ainsi que la difficulté à trouver des communautés dans le maxST, doivent provenir du fait que les méthodes n’arrivent pas à dissocier les puretés maximales des minimales. Ceci n’était pas le cas avec l’héroïne. Afin de mettre en évidence une différence entre les valeurs de pureté de l’héroïne et de la cocaïne, un histogramme a été calculé. Le but est de visualiser quelles tranches de pureté d’échantillons sont les plus fréquemment saisies. Si les valeurs de pureté les plus fréquentes avoisinent les 50%, ceci expliquerait que l’algorithme du spanning tree n’arrive pas à les dissocier. Le résultat est présenté par la figure 6.33.

Les histogrammes sont différents au niveaux des fréquences maximales. Pour l’héroïne, il s’agit d’une pureté à tendance basse rencontrée le plus souvent ; l’algorithme arrive donc facilement à dissocier le graphe en minST et maxST. Pour la cocaïne, c’est une pureté moyenne caractérisant la majorité des échantillons. Ceci explique le fait que les spanning trees aient des structures similaires, la pureté minimale et maximale étant difficile à dissocier. L’algorithme de détection de communautés fonctionne néanmoins correctement pour minST. Ceci pourrait être expliqué en supposant que les données à faible pureté génèrent plus de groupes que celles à pureté élevée. En imaginant qu’une cocaïne de basse qualité corresponde à un niveau du réseau proche de la rue, elle pourrait avoir tendance à être plus coupée qu’une cocaïne de plus haute qualité. Elle contiendra alors plus

de produits de coupage et donc disposera d'une plus grande variabilité combinatoire et ainsi aura une plus grande tendance à former des communautés que des données à valeurs élevées de pureté. Le but étant de détecter des patterns et comme il semble que ce soit plus le cas avec un minST, l'interprétation sera faite sur ce graphe uniquement.

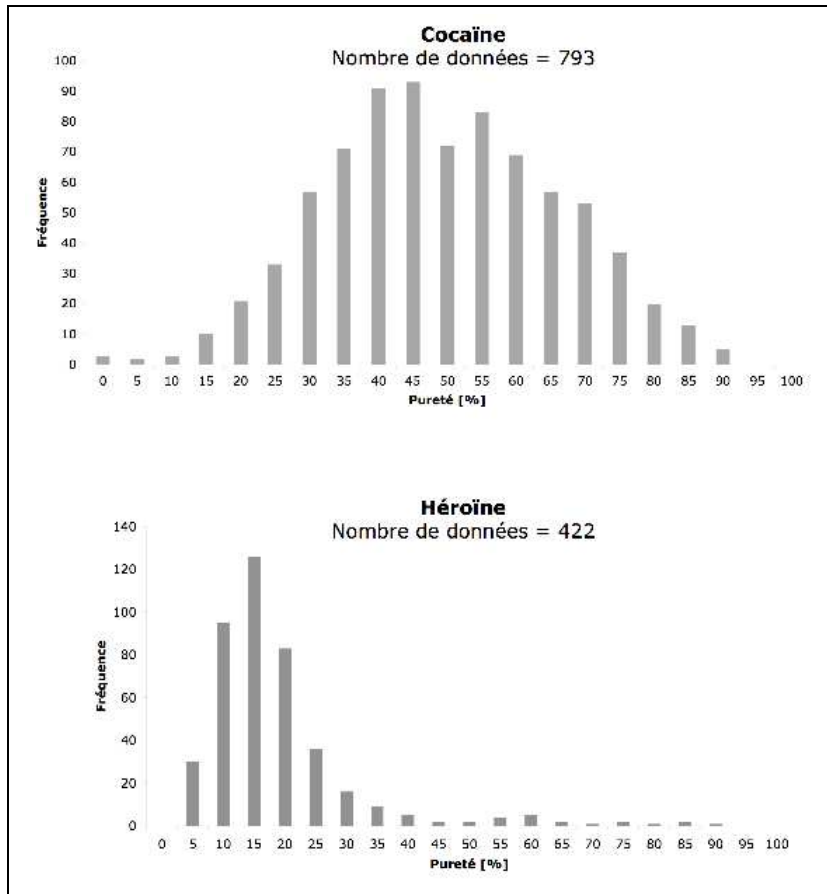


FIGURE 6.33 – Histogramme de la pureté de la cocaïne et de l'héroïne.

Les communautés mises en évidence sont le plus souvent liées à une région. Elles sont composées de la localité et des combinaisons qui y ont été détectées, dont les valeurs de pureté sont proches. Une analyse approfondie a été réalisée, ayant permis d'évaluer la pertinence de ces communautés. Cette évaluation a montré une tendance répétée dans plusieurs localités : **le contenu en coupage d'échantillons différents saisis dans une localité et dont la pureté est proche, ont plusieurs produits en commun (pouvant être des dénominateurs communs)**. Les cas suivants illustrent cette observation.

Deux saisies faites à Lancy le 16.09.04 (figure 6.34) contenant :

- **C96** : Glucose-Mannitol-Phénacétine, une pureté de 61,5% et de classe chimique "cl2".

- **C134** : Glucose-Mannitol-Phénacétine-Acide citrique, une pureté de 58% et de classe chimique "cl2".

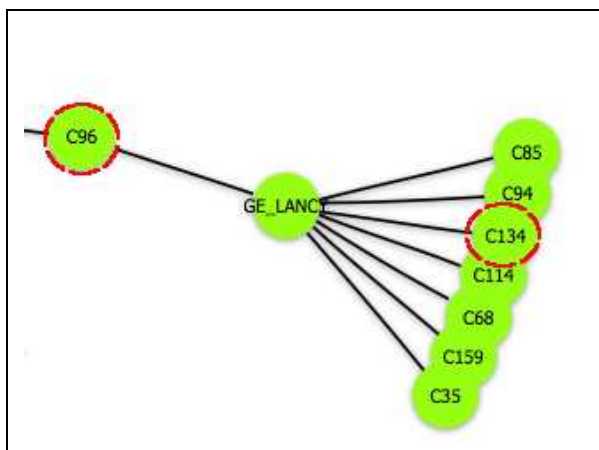


FIGURE 6.34 – Communauté liée à Lancy (Ge) avec comme dénominateur commun : Glucose-Mannitol-Phénacétine.

Ces deux saisies ont un dénominateur commun, Glucose-Mannitol-Phénacétine, une pureté proche et sont toutes deux de la même classe chimique. De l'Acide citrique a très probablement été ajouté localement à une cocaïne issue du même lot de production.

Des saisies faites à Prilly sur le canton de Vaud et à Anières sur le canton de Genève (figure 6.35) :

- Prilly, le 14.01.04 :
 - **C54** : Lactose - Sucrose, une pureté de 66,4 %, sans classe chimique.
 - **C93** : Lactose - Sucrose - Glucose, une pureté de 51 à 54% (deux échantillons), sans classe chimique.
- Anières, le 16.02.04 :
 - **C93** : Lactose - Sucrose - Glucose, une pureté de 37,4%, sans classe chimique.
 - **C130** : Lactose - Inositol - Glucose - Acide citrique, une pureté de 20,3% et de classe chimique "cl29"
 - **C194** : Lactose - Sucrose - Inositol - Glucose - Acide citrique, une pureté de 17,3%, et de classe chimique "cl29".
 - **C151** : cette combinaison ne semble pas faire partie de la tendance détectée (Diltiazem - Lactose - Mannitol - Phénacétine).

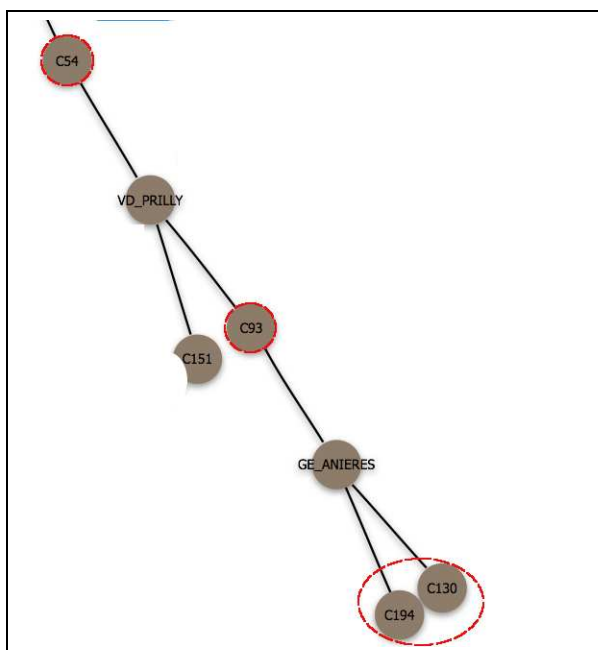


FIGURE 6.35 – Communauté liée à Prilly (Vd) et Anières (Ge), avec comme dénominateurs communs : Lactose - Sucrose et Lactose - Glucose.

La cocaïne saisie contient la même combinaison de base Lactose-Sucrose pour Prilly et Lactose-Sucrose-Glucose pour Anières. L'éloignement des échantillons au niveau de leur contenu en éléments majeurs réfute l'hypothèse d'une cocaïne issue d'un même lot de production. Cette hypothèse est aussi réfutée pour les échantillons saisis à Anières, excepté ceux contenant les combinaisons C130 et C194, faisant partie de la même classe chimique. Il est néanmoins possible pour la saisie de Prilly que des cocaïnes différentes aient été mélangées ; l'une contenant la combinaison C54, l'autre la C93.

Les échantillons d'Anières ont la même combinaison de base, dont deux la même classe chimique et le Sucrose en plus pour la C194. Ceci expliquerait les 17,3% de pureté comparée aux 20,3% de l'échantillon à combinaison C130. Dans ce cas, la cocaïne viendrait du même lot de production et du Sucrose a été ajouté, probablement par le dealer local. La pureté relativement basse de cette cocaïne suggère qu'elle se situe à un niveau du trafic proche de la rue.

Une saisie de cocaïne a été effectuée à Tolochenaz le 06.02.02 (figure 6.36) dont deux échantillons contenaient les produits suivant :

- **C191** : Acide citrique - Caféine - Glucose - Mannitol - Nicotinamide, une pureté de 32,9%, sans classe chimique.
- **C230** : Acide citrique - Caféine - Glucose - Mannitol - Nicotinamide - Lactose, une pureté

de 38,8%, sans classe chimique.

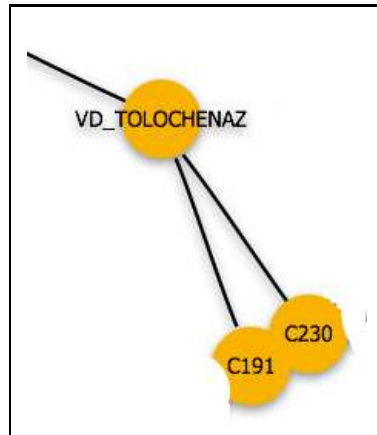


FIGURE 6.36 – Communauté liée à Tolochenaz (Vd) avec une combinaison très proche, mais une anomalie au niveau du taux de pureté.

Ces deux échantillons ont presque le même contenu en coupage, au Lactose près. Ils sont particuliers car celui contenant le plus de produits de coupage (le Lactose) a une pureté plus élevée que l'autre. Il est possible qu'un stock d'une pureté supérieure à 38,8% contenant du Lactose ait été ajouté à une partie de la cocaïne contenant la combinaison C191. Il en résulterait une cocaïne contenant les deux combinaisons de puretés légèrement différentes. La "recette de cuisine" sans Lactose serait alors un marqueur de réseau.

Synthèse :

1. La classe de pureté de la cocaïne la plus fréquente se situe entre 40 et 60%, ce qui ne permet pas à l'algorithme de minimum spanning tree de dissocier les graphes. Les algorithmes de détection de communautés montrent plus de groupes pour de faibles valeurs de pureté. Ceci tend à montrer que de la cocaïne de basse qualité contient plus de produits de coupage. Les combinaisons et la pureté formeraient alors plus de patterns que de la cocaïne de haute qualité pouvant provenir d'un niveau plus élevé d'un réseau.
2. Les communautés détectées semblent représenter des mélanges de "recettes de cuisine" pouvant indiquer des mélanges de stocks de cocaïne.
3. Des échantillons de cocaïne de mêmes classes chimiques peuvent avoir des contenus en coupage légèrement différents, en raison d'un possible mélange de stocks ou de l'adjonction d'un diluant.

Relations temporelles

Le graphe de base avec les points d'articulation est présenté par la figure 6.37.

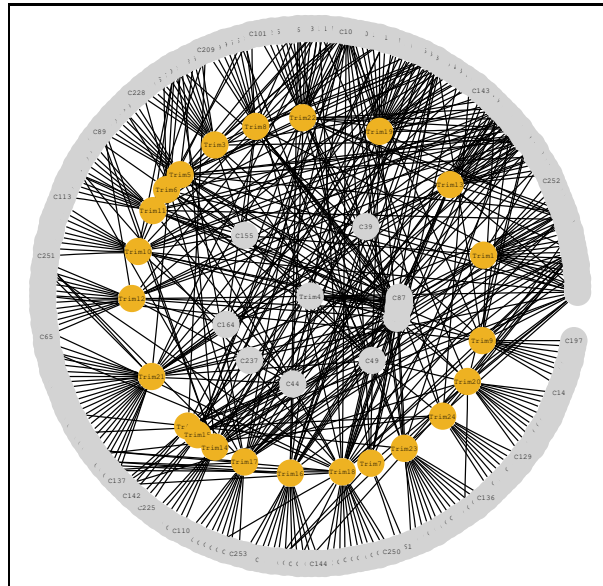


FIGURE 6.37 – Graphe de base “Combinaisons -Trimestres” avec points d'articulation.

Le résultat a une structure particulière dont les échantillons semblent rayonner autour des trimestres, comme si les combinaisons de produits évoluaient dans le temps. Ce résultat montre un pattern lié au temps. La création d'un graphe d'intervalle afin d'appliquer une détection de cliques a été réalisée afin de visualiser quelles combinaisons ont été détectées ensemble créant ainsi des groupes, les cliques⁴. La méthode permet aussi d'identifier un assemblage qui “disparaît” et un autre qui “apparaît”. L'analyse a été appliquée aux cantons de Genève et Vaud. Le résultat est présenté par les figure 6.38 et 6.39 . Afin d'avoir un aperçu de l'ensemble, l'image est présentée à grande échelle, ce qui ne permet pas de lire les détails. Les combinaisons sont indiquées verticalement à gauche et triées en fonction du nombre croissant de produits constituant la combinaison. En haut est mentionné horizontalement le temps. Les traits gris correspondent aux périodes où une combinaison a été observée lors de saisies de cocaïne.

4. Dans ce contexte, une clique est constituée d'un laps de temps et de combinaisons détectés ensemble dans ce laps de temps. Le sous-graphe correspondant est maximal (voir les sections 4.2.2 et 4.3.3).

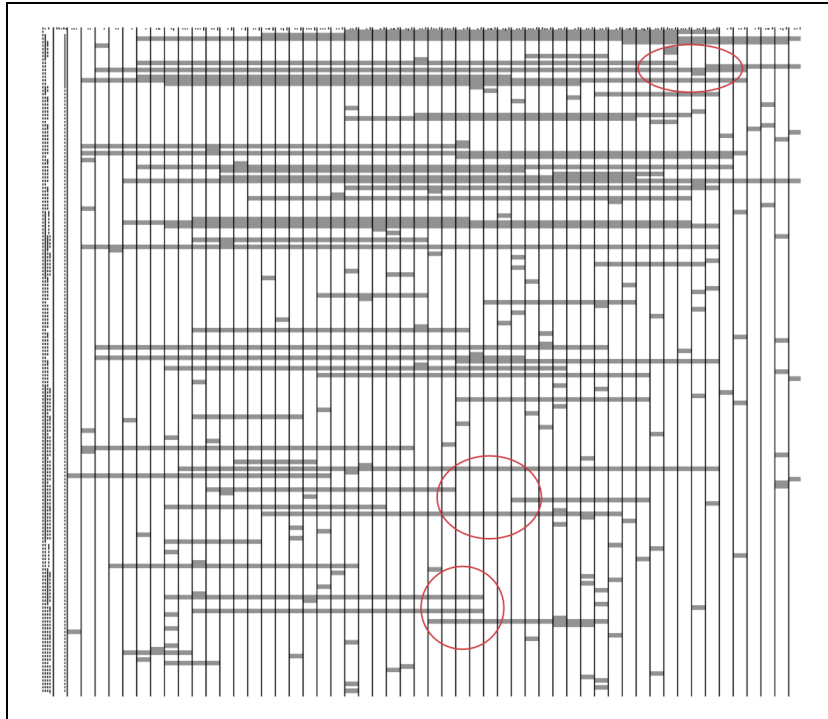


FIGURE 6.38 – Cliques maximales pour le canton de Genève.

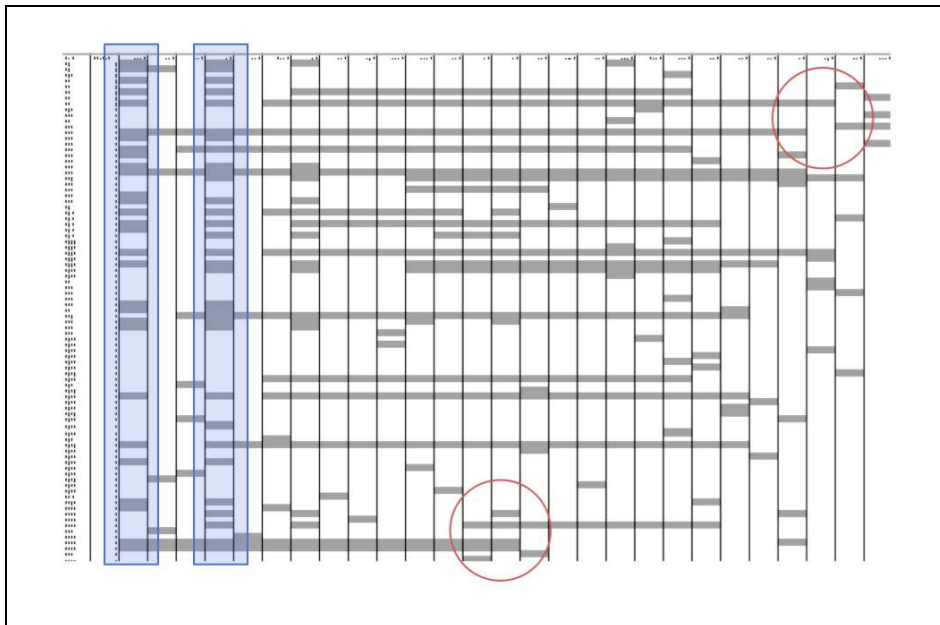


FIGURE 6.39 – Cliques maximales pour le canton de Vaud.

Le canton de Genève a des combinaisons qui perdurent durant presque toute la période étudiée

et d'autres qui apparaissent et disparaissent brièvement. Ce sont ces combinaisons de courtes durées qui donnent au graphe de base son aspect "en rayon". Mises à part les combinaisons les plus courantes, il n'y a pas de véritables groupes (ce qui se traduirait par la présence de plusieurs assemblages durant une même période). Néanmoins des "relais" de produits sont observés et ils sont marqués d'un cercle rouge sur les figures. Parmi ces successions, une tendance liée à la Phénacétine mérite d'être relevée. D'une manière générale, la présence de ce produit est en augmentation depuis 2000. L'analyse montre que plusieurs combinaisons sans Phénacétine s'arrêtent et que d'autres apparaissent, en contenant.

- **Glucose - Phénacétine** succède à **Glucose - Lactose**
- **Inositol - Phénacétine** succède à **Inositol - Mannitol**
- **Glucose - Lactose - Phénacétine** succède à **Glucose - Lactose - Mannitol**
- **Inositol - Lactose - Mannitol - Phénacétine** succède à **Inositol - Lactose - Mannitol - Lidocaïne**

Les résultats obtenus par la méthode des cliques pourraient indiquer que l'augmentation de la Phénacétine la rend plus disponible lorsque des stocks sont mélangés, apparaissant ainsi dans plusieurs combinaisons. Dans le cas du dernier exemple, elle "remplace" la Lidocaïne qui est comme la Phénacétine un adjuvant et qui est utilisée comme anesthésique.

Le résultat obtenu pour le canton de Vaud (figure 6.39) montre, au début de la période étudiée, deux groupes constitués de produits détectés ensemble (en bleu sur la figure 6.39). L'analyse du contenu des combinaisons présentes dans les groupes n'a pas abouti à une tendance pertinente. Des relais de combinaisons apparaissent aussi, mais ceux-ci n'ont rien révélé de pertinent non plus.

Les graphes min et maxST ayant les mêmes structures, le minST seul a été utilisé pour la détection de communautés. L'algorithme "walktrap" est celui montrant le plus de structures, le graphe résultant est présenté par la figure 6.40. Les autres sont placés en annexe.

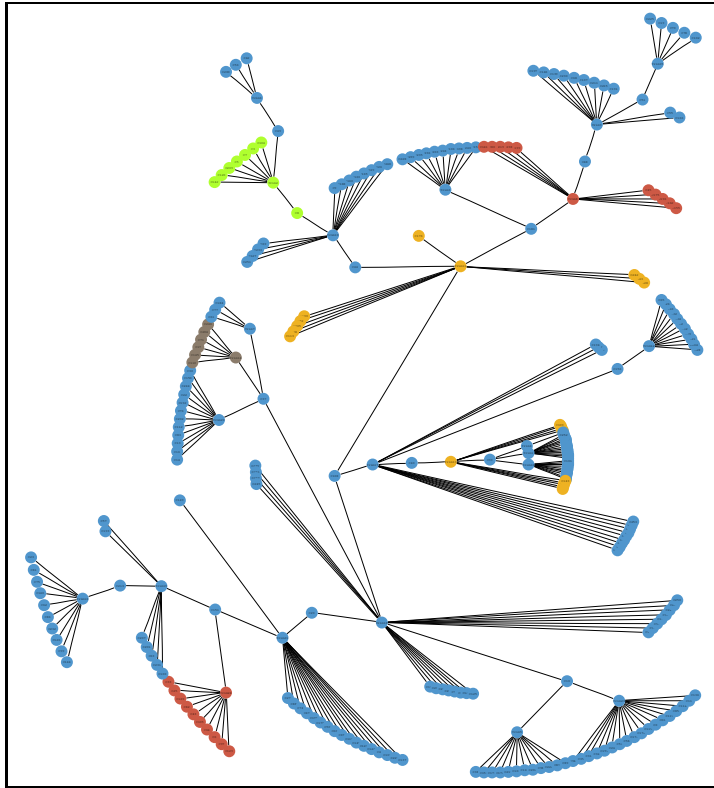


FIGURE 6.40 – Détection de communautés par “walktrap” appliqué au minST “Combinaisons - Trimestres”.

Interprétation

Les communautés détectées liées à un trimestre semblent indiquer, comme pour les localités, un mélange de substances ou une même cocaïne mais coupée différemment. Deux cas montrent une tendance particulière où, durant le trimestre, des échantillons de cocaïne ne contenant qu'un seul produit ont été détectés ainsi que des échantillons contenant ces produits additionnés d'autres substances. Au trimestre 2 (avril, mai et juin 2000), ont été détectées les combinaisons suivantes (figure 6.41) :

- **C8** : Lactose, d'une pureté de 45,8%, sans classe chimique.
- **C3** : Caféine, d'une pureté de 64%, sans classe chimique.
- **C5** : Glucose, d'une pureté de 39%, sans classe chimique.
- **C7** : Inositol, d'une pureté de 70%, sans classe chimique.
- **C104** : Inositol - Lactose - Mannitol, d'une pureté de 28,5%, de classe chimique “cl2”.
- **C120** : Lactose - Mannitol - Piracetam, d'une pureté de 15,8%, de classe chimique “cl2”.
- **C142** : Caféine - Inositol - Lactose - Mannitol, d'une pureté de 22,1%, sans classe chimique.
- **C209** : Glucose - Inositol - Lactose - Mannitol - Lidocaïne, d'une pureté de 28,2%, sans

classe chimique.

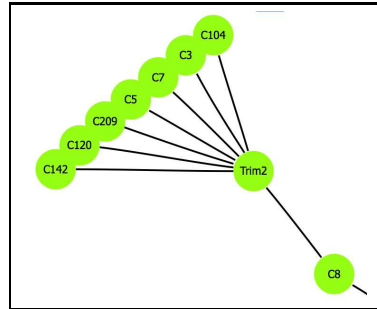


FIGURE 6.41 – Communauté liée au trimestre 2, montrant des similitudes au niveau des combinaisons et des classes chimiques.

Les échantillons contenant plus d'un produit ont tous une pureté plus faible que ceux à une seule substance de coupage. De plus, la combinaison C142 est l'addition de C104 et C3. C209 correspond à l'addition de C5 et C104 avec de la Lidocaïne en plus. Les échantillons contenant C104 et C120 sont de la même classe chimique. Il est possible qu'un stock de cocaïne de classe "c12" contenant les deux "recettes de cuisine" C104 et C120 ait circulé durant ce trimestre et qu'une partie contenant la combinaison C104 ait été mélangée à une autre cocaïne, perdant ainsi sa classe chimique. Si un mélange est survenu, il est possible que les échantillons résultant aient conservé une certaine similitude au niveau chimique avec le stock de base. Si c'est le cas, les cocaïnes contenant les combinaisons C104, C209, C120 et C142 devraient être assez proches au niveau de leur contenu en éléments majeurs. Une analyse en composantes principales (ACP) permettrait de visualiser si similitudes il y a.

Un jeu de données a été créé, contenant les valeurs chimiques des éléments majeurs des échantillons ciblés, comme variables. Des données chimiques sans classe et hors de la période d'étude ont été utilisées pour diversifier le set. Le résultat est présenté par la figure 6.42. La légende des combinaisons permet de repérer les points projetés selon les deux premières composantes. Les échantillons de classe chimique "c12" sont marqués en bleu (C104 et C120). Ceux contenant la combinaison C142 sont passés en vert, la C209 en rouge et la C8-S1 (échantillons contenant C8 issu d'une première saisie) en orange. Les points sans label correspondent aux données de diversification.

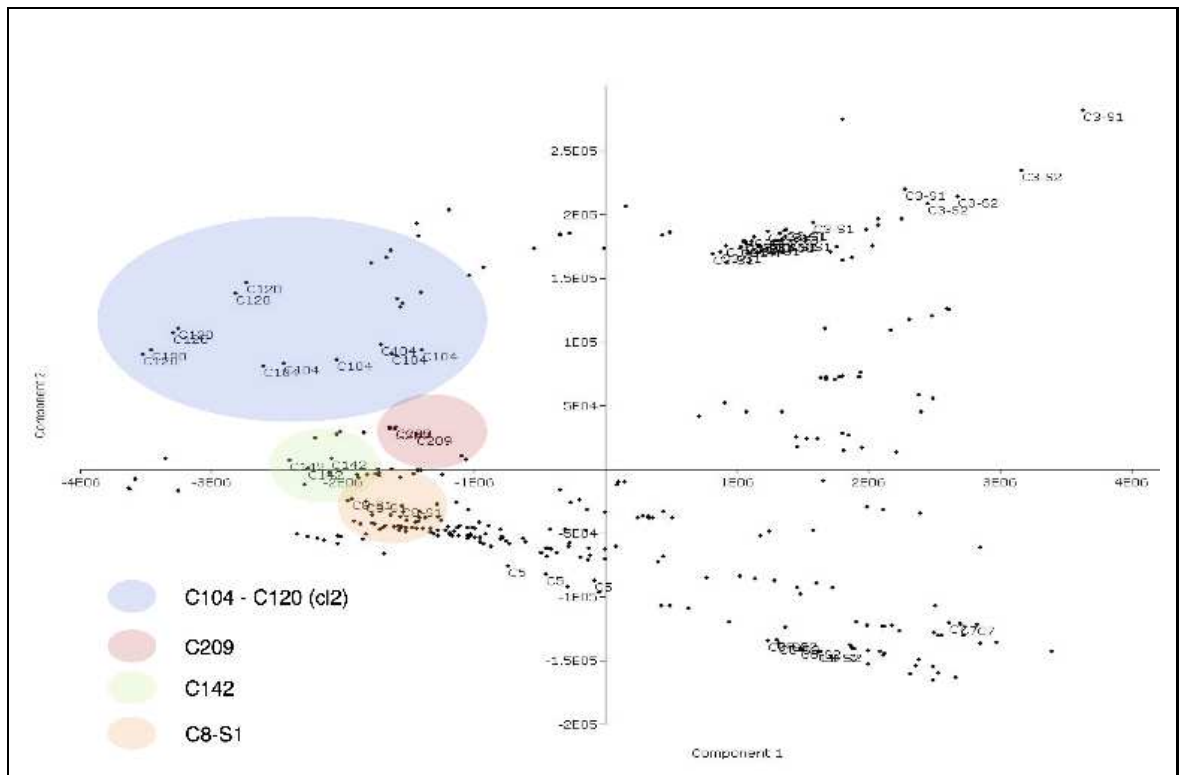


FIGURE 6.42 – Analyse en composantes principales des échantillons de la communauté liée au trimestre 2.

Le résultat est intéressant, car les échantillons dont le coupage semble être un mélange mais dont les classes chimiques ne correspondent pas, sont relativement proches. Celui ne contenant que du Lactose (C8-S1) est aussi similaire. Il est possible qu’il y ait de la cocaïne commune à ces échantillons, auquel cas il s’agirait d’une même substance mais d’un coupage différent. Ce dernier aurait pu être l’objet d’adjonctions successives à partir d’une cocaïne à un seul produit de coupage, dans le but de diminuer le taux de pureté.

Ce jeu de données a été testé “à l’aveugle” par le doctorant de ce projet traitant du développement d’algorithmes d’apprentissage automatisé. La méthode du **k-means** ainsi qu’un algorithme basé sur les **réseaux de neurones** (développé par lui pour le clustering) ont été testés. Les résultats obtenus montrent des similitudes entre les échantillons contenant la combinaison C8-S1 (correspondant à une première saisie) et les échantillons contenant C104, C120, C209 et C142. Les données correspondant à ces assemblages se regroupent dans le même cluster, indiquant une similitude. Ce résultat va dans le sens de l’hypothèse d’une cocaïne commune.

Au trimestre 24 (octobre, novembre, décembre 2005) ont été détectées les combinaisons suivantes (figure 6.43) :

- **C6** : Hydroxyzine, d'une pureté de 64%, sans classe chimique.
- **C25** : Atropine - Phénacétine, d'une pureté de 50,8%, sans classe chimique.
- **C33** : Diltiazem - Lactose, d'une pureté de 52,8%, sans classe chimique.
- **C43** : Hydroxyzine - Lactose, d'une pureté de 58,6%, sans classe chimique.
- **C64** : 1,4_dmt - Mannitol - Phénacétine, d'une pureté de 46,3%, sans classe chimique.
- **C82** : Diltiazem - Lactose - Glucose, d'une pureté de 75,5%, sans classe chimique.
- **C97** : Glucose - Mannitol - Phosphate, d'une pureté de 59,9%, sans classe chimique.
- **C105** : Inositol - Lactose - Paracétamol, d'une pureté de 62,3%, sans classe chimique.
- **C129** : 1,4_dmt - Lactose - Mannitol - Phénacétine, d'une pureté de 29,4%, sans classe chimique.
- **C169** : Glucose - Mannitol - Phosphate - Sucrose, d'une pureté de 60,7%, sans classe chimique.

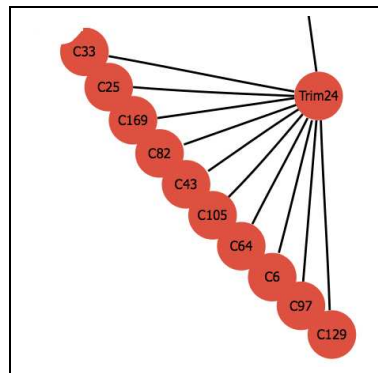


FIGURE 6.43 – Communauté liée au trimestre 24, caractérisée par des produits peu fréquemment détectés.

Cette communauté ne contient aucun échantillon avec une classe chimique. En revanche, des produits peu courants sont détectés dans plusieurs échantillons dont l'assemblage de produits de coupage est différent. L'Hydroxyzine est présente dans les échantillons contenant la combinaison C6 et C43; le Diltiazem dans les échantillons à combinaisons C33 et C82 et le Phosphate est contenu dans C97 et C169. Ces substances correspondent à la catégorie 2 de la classification proposée par le tableau 5.3, excepté le Phosphate qui serait de catégorie 3. Ils sont néanmoins peu communs et semblent ici avoir un rôle de signature. Cette tendance est intéressante car ces produits particuliers ont été observés dans de la cocaïne saisie en 2004 et 2005 uniquement et cette tendance est liée au trimestre 24 (le dernier de la période d'étude) pour toutes les régions étudiées. Elle pourrait indiquer l'émergence de nouveaux produits de coupage ou qu'un réseau local particulier mélange des stocks avec une cocaïne contenant ces produits.

Synthèse :

1. *De nombreuses combinaisons ne sont détectées que durant un laps de temps restreint, inférieur au trimestre, pouvant indiquer des changements constants dans la manière de couper la cocaïne.*
2. *Des échantillons saisis dont les combinaisons contiennent de la Phénacétine succèdent dans le temps à des échantillons dont les combinaisons en sont exemptes. Cette observation peut correspondre aux effets d'une généralisation de l'utilisation de ce produit.*
3. *Les communautés détectées ont mis en évidence deux tendances possibles :*
 - (a) *Le possible mélange de stocks de cocaïne, marqué par une même combinaison de base de produits de coupage et parfois un même profil chimique, ainsi que l'utilisation de produits peu fréquents comme signature.*
 - (b) *L'émergence d'une nouvelle tendance de coupage avec des produits peu courants.*

Chapitre 7

Discussion

7.1 Introduction

Ce travail a pour but d'évaluer les possibilités d'intégration des méthodes de reconnaissance de patterns à un processus de renseignement criminel. Elles ont été testées sur des données issues du trafic des stupéfiants, dans le cadre d'un système opérationnel de renseignement basé sur l'analyse des stupéfiants, le profilage [Ioset *et al.*, 2005]. Les méthodes de détection de patterns ont été mises à l'épreuve, afin d'essayer de déterminer des causes à leur apparente faible utilisation au niveau opérationnel de Police. Cet exercice devrait permettre de proposer des pistes pour y remédier.

Ce chapitre va tenter de mettre en évidence les avantages, les inconvénients, les limites et le potentiel d'utilisation de ces méthodes dans une optique de renseignement.

La discussion va tout d'abord porter sur une critique générale des méthodes testées et des résultats. Elle va ensuite mettre en évidence comment les méthodes pourraient être complémentaires à la démarche du profilage et enfin discuter des possibilités d'utilisation des technologies du data mining dans un processus de renseignement.

7.2 Application aux données issues du trafic des stupéfiants

7.2.1 La théorie des graphes : pertinence et originalité

Le terme de data mining regroupe beaucoup de techniques, en constant développement. Le choix de méthodes se fait, entre autres, en fonction de la problématique et du format des données (cf chapitre 3.3.1). Les données ont été utilisées dans un format binaire, de manière à pouvoir utiliser des méthodes d'analyse combinatoire. L'information traitée correspond à l'absence et la présence des produits de coupage (plus précisément, leur détection ou non-détection par la méthode et la

machine d'analyse). Le jeu de données ainsi mis en forme nécessitait l'utilisation de méthodes permettant de détecter des schémas en tenant compte de leur format combinatoire. Le graph mining est un terme regroupant des techniques de data mining appliquées à de grands graphes, dans le but de détecter des schémas (sous-graphes, sous-structures répétitives etc...). Le jeu de données des produits de coupage de l'héroïne et de la cocaïne, structuré en fonction de leur présence ou de leur absence dans une saisie, dans un lieu et à une période donnée, peut être modélisé sous forme de graphe. Les relations entre produits ainsi que les combinaisons détectées en un lieu et un temps peuvent contenir des schémas et refléter une activité particulière liée au trafic des stupéfiants, permettant d'apporter de l'information. C'est dans cette optique que la théorie des graphes a été choisie comme méthode d'analyse.

Les résultats obtenus montrent que les méthodes aident à détecter de tels schémas. A partir de ces derniers, des hypothèses peuvent expliquer leur présence, mais elles doivent être testées, confirmées ou réfutées, en fonction des connaissances sur le trafic des stupéfiants ainsi que d'autres informations accessibles. Il est possible d'affirmer que ces méthodes permettent d'attirer l'attention vers des schémas particuliers, mais qu'en soit, elles ne suffisent pas pour développer un système complet d'analyse. Leur intégration à la démarche du profilage permettrait néanmoins d'apporter des informations complémentaires. Un aperçu des méthodes testées va permettre de mettre en évidence quelles sont ces informations complémentaires ainsi que leur originalité, leurs limites et lacunes.

7.2.2 Application de la théorie des ensembles à des graphes

Plusieurs graphes ont été réalisés, modélisant les relations entre produits par année et par canton. Une première représentation de ce type de modèles avait été faite dans [Terrettaz-Zufferey *et al.*, 2007]. La théorie des ensembles a été utilisée, afin d'obtenir une visualisation globale des relations entre produits. L'apport de cette technique réside dans la mise en évidence de différences ou de similitudes entre des graphes, donc entre des modèles de l'activité liée au trafic des stupéfiants, de régions ou de périodes de temps.

La compréhension des opérations utilisées, la différence et l'intersection, est relativement intuitive. En revanche, les calculs réalisés dans le but de dissocier des produits probablement ajoutés en Suisse de ceux déjà présents avant leur entrée, demande plus de réflexion.

Cette application a néanmoins apporté des informations inédites sur certaines tendances de coupage, notamment sur les produits "typiquement suisses", ainsi que sur des différences et similitudes géographiques. Elle a un potentiel d'utilisation original et non négligeable car elle permet de réaliser les intersections et soustractions en une seule action, en évitant le traitement "au cas par cas". La majeure partie du travail réside dans la création du modèle, les produits par cantons ou localités, par années ou par mois, etc... Il est imaginable qu'elle permette d'exploiter une banque de données

en recherchant des similitudes et des différences, à partir de connaissances *a priori* de certaines tendances de coupage. Elle permettrait d'écartier des échantillons et de tester la méthode du profilage sur des données restreintes.

Il faut toutefois garder à l'esprit que l'utilisation de la théorie des ensembles dans cette optique ne se base pas sur le nombre de fois que des produits sont détectés. Pour réaliser les calculs de différence, il est nécessaire que tous les produits soient comptabilisés une seule fois. Autrement dit, les graphes ne doivent pas être pondérés et il faut donc les simplifier systématiquement afin de ne pas oublier des poids qui fausseraient les résultats. Voici un exemple illustré par la figure 7.1 : un analyste souhaite connaître les produits détectés sur Neuchâtel qui ne l'ont pas été au Tessin. Il va soustraire le graphe du Ti à celui de Ne. Il obtiendra pour résultat un graphe indiquant les produits uniquement présents dans le canton de Ne.

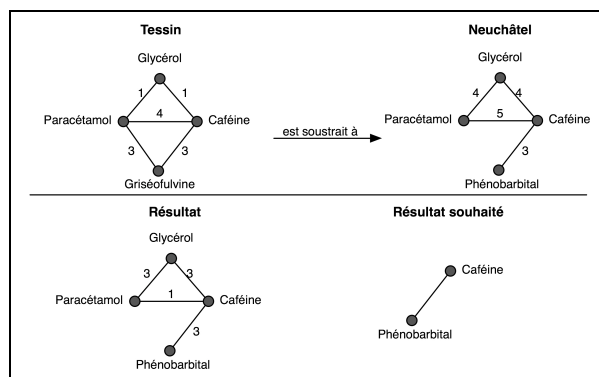


FIGURE 7.1 – Illustration de l'utilisation de graphes simplifiés pour appliquer la théorie des ensembles.

Si le graphe est pondéré, il y aura des arêtes résiduelles indiquant des produits présents aussi au Ti. Les seules relations présentes à Ne qui sont absentes au Ti sont celle de la Caféine et du Phénobarbital. Avec des graphes pondérés, les relations Paracétamol, Caféine et Glycérol seront conservées.

Cette limite empêchera de voir si la détection d'un produit dans une seule région est de l'ordre de l'exception ou si il s'agit d'un phénomène répétitif. Pour s'en assurer, il sera nécessaire d'aller vérifier les données de base.

7.2.3 Simplification de graphes et détection de groupes

Les minimum et maximum spanning trees (minST et maxST)

Le point de départ, souvent déroutant en reconnaissance de patterns appliqués à des données réelles, est de les aborder par une vue d'ensemble. Ceci est réalisé, en général, au moyen de l'analyse

en composantes principales pour le traitement de quantités et par un graphe général pour le présent travail. Dans les deux cas, seul un nuage de points ou de traits est observé. C'est dans un amas de données que des structures doivent être mises en évidence. Ainsi, une des premières étapes consisterait à les simplifier en les séparant en jeux distincts. Dans ce travail, l'application d'un algorithme de minimum spanning tree (figure 7.2) à des graphes pondérés a été réalisée. Les poids correspondaient à la pureté ainsi qu'à son inverse; l'algorithme de spanning tree mettant en évidence un sous-graphe dont les poids sont minimaux. En utilisant la pureté, le sous-graphe mis en évidence est minimal (minST). Pour l'inverse de la pureté, il est maximal (maxST). Cette démarche a permis, en une analyse, de rendre lisibles les graphes et de mettre en évidence des structures différentes liées à des taux de pureté différents.

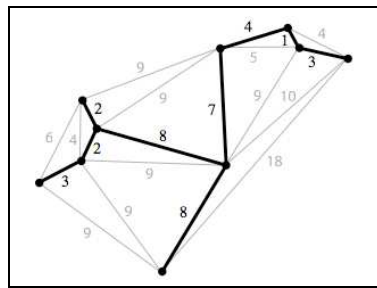


FIGURE 7.2 – Mise en évidence du minimum spanning tree d'un graphe.

Cette application a été faite afin de **pouvoir** visualiser des patterns. Des informations ont peut-être été perdues, mais sans cette étape, les résultats obtenus par la détection de communautés et les tendances liées à la pureté auraient été noyés dans la masse de données. Il aurait été très difficile d'analyser les résultats car ils auraient été illisibles.

Le choix a été fait de s'intéresser à la pureté pour séparer les graphes. Il a été motivé par l'hypothèse qu'intuitivement, un échantillon de stupéfiant saisi dans la rue sera moins pur, donc plus coupé qu'une saisie faite aux frontières. La manœuvre permettait déjà de séparer en quelque sorte la problématique entre combinaisons proches de la rue et celles plus en amont d'un réseau.

Il s'agit d'une hypothèse qui est peut-être loin de la réalité. Elle a cependant été prise comme point de départ, afin de disposer d'une piste pour l'analyse.

Il aurait aussi été possible d'utiliser le nombre de fois qu'une combinaison a été détectée durant un trimestre ou une localité. Ceci n'a pas été réalisé en raison de la grande variabilité de combinaisons et de l'impact de l'activité de la Police. Beaucoup d'assemblages n'ont été observés qu'une seule fois et d'autres sont très fréquents et détectés plusieurs fois lors de l'aboutissement

d'une enquête, par exemple. Ceci n'aurait pas permis une bonne séparation des graphes.

Cette technique a une limite, illustrée par le problème rencontré pour la cocaïne, lorsque la tranche de pureté dont la fréquence est maximale se situe autour de 50%. La méthode ne peut pas séparer les valeurs maximales des minimales.

La détection de communautés plutôt que le modèle des blocs

Le modèle des blocs est une technique "à la mode", souvent mentionnée dans des publications dédiées à la lutte contre la criminalité [Chen *et al.*, 2003]. L'application faite dans ce travail utilise des graphes bipartis "combinaisons de produits - lieux" et "combinaisons de produits - temps". Le test du modèle des blocs à un graphe simple "combinaisons - trimestre + combinaisons - saisies + saisies - trimestres" a permis d'obtenir des blocs, mais il y en avait trop et pour la majorité triviaux (cf. section 6.3.1). Cette méthode s'applique mal à notre problématique. De plus, elle ne tient pas compte des poids (ici la pureté) ajoutés au graphe de base.

L'algorithme Subdue a aussi été testé, mais aucun groupe n'a été détecté car il recherche des structures répétées dans un graphe et les modèles utilisés dans ce travail en sont dépourvus. De plus, il fonctionne mal avec des graphes bipartis.

La détection de communautés dispose de plusieurs algorithmes pour détecter des groupes, dont deux ont été testés : les algorithmes "walktrap" et "fast greedy", tous deux tenant compte des poids des arêtes du graphe. Ils s'appliquaient très bien aux minST et maxST.

Parmi les deux algorithmes, le "walktrap" semblait correspondre le mieux à la problématique et a été le plus utilisé. Les problèmes rencontrés sont le nombre de groupes détectés. Ils ont tous été analysés, afin de déterminer leur pertinence. Ceci peut prendre beaucoup de temps, mais lorsqu'ils s'avèrent intéressants, leur interprétation pourrait apporter du renseignement utile.

7.3 Méthodes complémentaires au profilage

Le schéma 3.2, présenté au chapitre 3, illustre l'emplacement possible des méthodes de reconnaissance de patterns dans un processus de renseignement. Afin de discuter plus en détail d'une utilisation opérationnelle de ces techniques, leur utilisation dans la démarche du profilage sera en premier abordée.

Les méthodes d'analyse proposées ont permis de mettre en évidence des patterns. Leur but est d'offrir une forme d'exploitation supplémentaire aux méthodes d'analyse et de regroupement des échantillons de stupéfiants sous forme de classes, appliquées dans la démarche du profilage. Elles complètent ces techniques car utilisent un minimum *d'a priori* sur des structures préalablement

mis en évidence, permettant d'intégrer le flux de nouveaux échantillons. Ceux-ci sont intégrés à la mémoire au moyen des méthodes traditionnelles, en fonction de ce qui est déjà connu (les classes chimiques). Cela peut amener à se focaliser sur un processus de classification par "effet tunnel", empêchant l'analyste de voir d'autres structures. L'application des méthodes proposées, en abordant les données de manière globale et sans *a priori* sur des patterns préexistants, permet de découvrir d'autres structures. Le défi consiste donc à combiner les deux approches dans un processus de renseignement.

L'interprétation des liens chimiques est un processus de renseignement ; l'interprétation de patterns détectés par des méthodes de graphes pourrait s'intégrer à celui-ci. Les données ont la même origine, seules les techniques diffèrent et cette étape pourrait être placée en parallèle de l'étape d'évaluation et de classement des informations.

Un schéma issu de [Ioset *et al.*, 2005] (figure 7.3) illustre à quelle étape elles pourraient être intégrées.

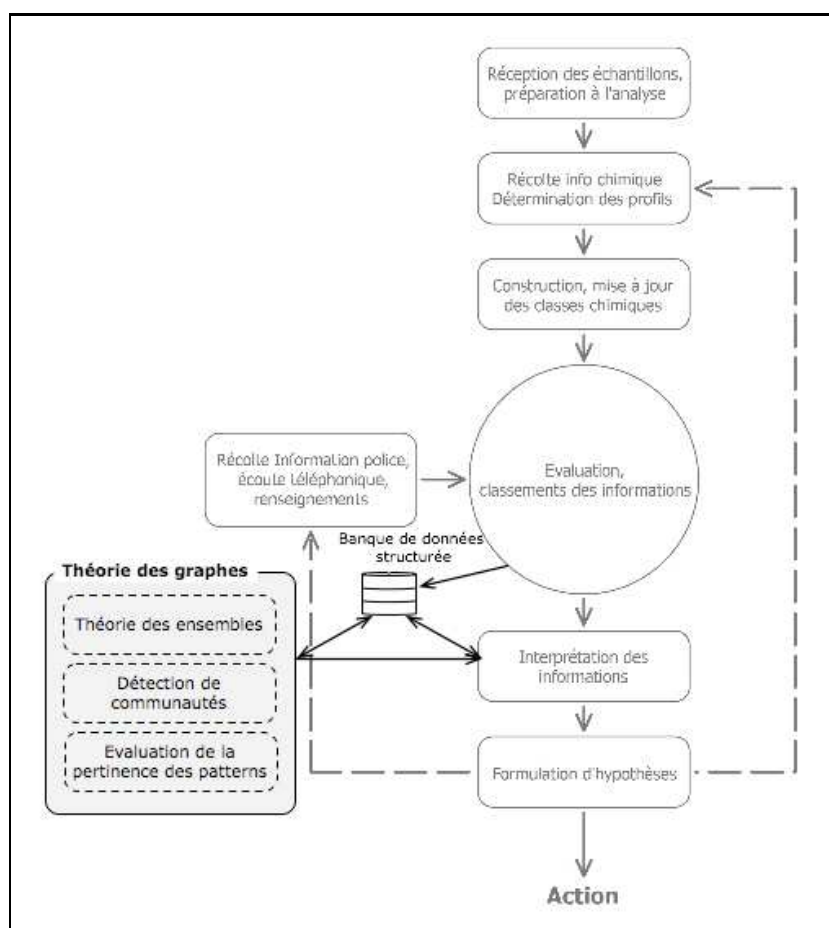


FIGURE 7.3 – Processus de renseignement par interprétation des liens chimiques (tiré de [Ioset *et al.*, 2005]).

Les tendances détectées, après une évaluation de leur pertinence, pourraient être interprétées de la même manière que les profils chimiques. Ecarter les tendances inutiles et conserver celles évaluées comme utiles ou à interpréter, constitueraient de l'information à intégrer au processus d'interprétation permettant de produire du renseignement. Intégrées à la banque de données structurée (la mémoire) comme connaissances, elles permettraient, lorsque de nouveaux patterns sont détectés, de les comparer au contenu de la mémoire, de les reconnaître et de les classer dans une catégorie de connaissance similaire. Autrement dit, d'adapter la structure des connaissances actuelles sur les activités des malfaiteurs. L'exemple suivant illustre la démarche : une analyse sans aucun *a priori* est réalisée sur les données des produits de coupage de la cocaïne. Le pattern résultant montre qu'une combinaison de substances est liée à une période de temps et à une pureté particulière. Dans la mémoire, ce même pattern a été détecté quelque temps auparavant. En le comparant à la mémoire, il est possible de lier le résultat à une structure préexistante. Selon sa signification au niveau du trafic des stupéfiants (un réseau local qui refait surface ou une cocaïne coupée selon ces produits qui redevient active) et les mesures prises en fonction de cette connaissance, du renseignement sera produit.

7.3.1 Quels moyens de validation

Que ce soit le profilage ou les méthodes de reconnaissance de patterns, les informations de viennent du renseignement après interprétation, faisant intervenir des compétences analytiques du phénomène et de l'affaire traitée. Une description de comment la combinaison de ces champs de connaissance permettrait de tester les hypothèses de travail, a été réalisée. Son but est d'illustrer le processus, afin de mettre en évidence l'efficacité de cette association de compétences :

Le phénomène du coupage lié au trafic des stupéfiants évolue dans le temps.

En se basant sur les méthodes testées, l'hypothèse peut être confirmée. Les graphes des figures 6.1 et 6.2 du chapitre 6 montrent une évolution des combinaisons dans le temps. Du point de vue d'un enquêteur ou d'un analyste connaissant bien le phénomène, une partie des changements pourrait aussi provenir de l'aboutissement d'une longue enquête, du début d'une collaboration avec un canton (par exemple, l'arrivée d'échantillons résultant de cette nouvelle collaboration ferait croître rapidement la banque de données), ou d'un changement au niveau d'une loi sur le trafic de stupéfiants.

Une modélisation globale d'échantillons saisis sur la base de leur contenu en produits de coupage

et selon des dimensions spatiales et temporelles, peut révéler des tendances utiles sur le trafic des stupéfiants d'une échelle nationale à régionale.

Un expert en méthodes d'analyse sans connaissances du phénomène pourra proposer plusieurs méthodes d'analyse possibles, mais ne pourra pas déterminer laquelle mettra en évidence des patterns utiles. Un analyste connaissant le phénomène pourra proposer certaines méthodes, aiguiller leur utilisation et réaliser une première évaluation de la pertinence des structures détectées. L'enquêteur, connaissant bien des phénomènes locaux et actuels, pourra interpréter la pertinence des structures, en valider certaines et peut-être s'en servir à des fins opérationnelles.

L'analyse des relations existant entre les caractéristiques physico-chimiques associées aux informations circonstancielle (temps et géographie) des saisies, permet de produire du renseignement opérationnel.

Cette hypothèse relève de l'analyse en particulier. Le rôle des unités d'analyse est justement de produire du renseignement sur la base de cette hypothèse. Actuellement, en Suisse, les unités traitent divers types de phénomènes, tels que les cambriolages, les vols et d'autres types de délits pouvant être sériels et diffusent l'information dans leur canton. Cette information est partagée avec celle d'autres cantons au moyen notamment du CICOP (Concept Intercantonal de Coordination Opérationnelle et Préventive, [Ribaux, 2005b]), une plateforme d'échange et d'analyse entre cantons.

Pour répondre à cette hypothèse, ne serait-il pas envisageable de mettre en place une unité d'analyse traitant des données liées au trafic de stupéfiants? Elle participerait aux analyses (profilage, détection de patterns) demandées par les enquêteurs, aidée d'experts (IPS), les interpréterait avec l'aide des enquêteurs, produirait du renseignement et le partagerait avec les unités d'analyse d'autres cantons. Il serait imaginable qu'un CICOP des stupéfiants soit un bon moyen pour partager cette information et mettre à jour le renseignement d'une manière globale.

7.3.2 A qui les rôles d'experts ?

Dans [Ioset *et al.*, 2005], il est mentionné qu'une difficulté d'implémentation du profilage réside auprès des opérateurs pour traiter et interpréter les données. Il en va de même pour les méthodes de détection de patterns. Les analyses sont faites pour les enquêteurs, mais ceux-ci n'ont pas le temps ni la formation pour les réaliser eux-mêmes. Les analystes? Il n'y a actuellement pas d'unité d'analyse de données issues du trafic des stupéfiants, ni de partage d'informations, mais en imaginant qu'il y en ait une, un analyste criminel pourrait-il réaliser l'analyse à partir de données provenant de l'IPS? Les méthodes de reconnaissance de patterns nécessitent plusieurs compétences : techniques (pour développer et appliquer les méthodes), analytiques (pour l'interprétation), opérationnelles

(pour valider les informations délivrées par les patterns). Faut-il disposer dans une même unité d'experts pour chaque compétence ? Faut-il former des analystes au développement et à l'utilisation de méthodes de reconnaissance de patterns ? Faut-il former des experts en méthodes pour réaliser de l'analyse criminelle ? Faut-il déléguer à des structures spécialisées telles que l'IPS, les analyses, délocalisant les compétences ? Ces cas de figures ont chacun leurs avantages et leurs inconvénients : Disposer d'experts pour chaque compétence demande beaucoup de personnel et pourra créer des problèmes de ressources financières pour engager la main-d'oeuvre nécessaire.

En formant des analystes ou des experts en méthodes, le risque est de limiter les compétences de chacune des deux spécialités. L'un risque de limiter son expérience opérationnelle et l'autre pourrait ne plus avoir suffisamment de temps pour se consacrer à une veille technologique et technique.

Délocaliser les compétences en transmettant les analyses à des experts externes risquera de limiter les échanges de connaissances *a priori*. Ce risque est présent non seulement pour le traitement de données liées au trafic des stupéfiants mais aussi pour le cadre plus général du renseignement criminel.

7.4 Comment intégrer ces méthodes à un processus de renseignement

7.4.1 Exemple

Une expérience de travail dans un service de coordination judiciaire a permis d'expérimenter l'utilisation des méthodes de reconnaissance de patterns dans un processus de renseignement.

La coordination judiciaire dispose de compétences conjointes d'enquêteurs formés en analyse criminelle et de collaborateurs scientifiques de l'IPS spécialisés en analyse criminelle. Durant cette période, l'auteur de ce travail a oeuvré comme expert en méthodes d'analyses de reconnaissance de patterns et en tant qu'analyste criminel.

Pour toutes les problématiques abordées et le traitement d'informations, les enquêteurs en charge d'affaires criminelles étaient présents pour aider à l'élaboration de renseignement.

Cette collaboration de compétences semble être une solution efficace.

Un exemple concret d'application peut être relaté :

Une demande de la part d'enquêteurs consistant à déterminer dans un très grand jeu de données de délits, lesquels pourraient avoir été perpétrés par un suspect sur la base de ses relevés de téléphone mobile. Ce problème a été transmis à deux analystes (dont l'auteur de ce travail) qui ont développé un moyen de le traiter.

Diverses contraintes ont été rencontrées :

- Le grand nombre de données à considérer.
- Le temps est présenté sous forme d'intervalles (date et heure du début du délit et date et heure de fin).
- Traiter conjointement la dimension temporelle sous forme d'intervalles et la dimension spatiale.

La méthode développée utilise l'analyse combinatoire pour résoudre le problème du temps et fait l'objet d'une publication [Birrer et Terrettaz-Zufferey, 2008]. Le traitement conjoint des dimensions spatiales et temporelles est intégré en combinant des matrices d'adjacence et de distance. L'outil a été développé au moyen d'une librairie Python. Le résultat indique quels délits peuvent être compatibles avec un appel issu du relevé téléphonique d'un suspect.

Les étapes et intervenants de cet exemple peuvent être analysés :

Dans le processus du renseignement, l'exemple se situe entre la banque de données structurées, dans laquelle les données vont être puisées et l'analyse. Cette dernière va nécessiter de faire appel aux méthodes de reconnaissance de tendances pour résoudre le problème.

Un **enquêteur** pose à l' **unité d'analyse** une problématique, dont la résolution apporterait de l'information, puis éventuellement du renseignement.

Un **collaborateur scientifique spécialisé en analyse criminelle** évalue la demande et fait appel à un **spécialiste en méthodes** et à un **analyste** pour développer un outil informatisé afin de répondre à la demande de l'enquêteur (figure 7.4).

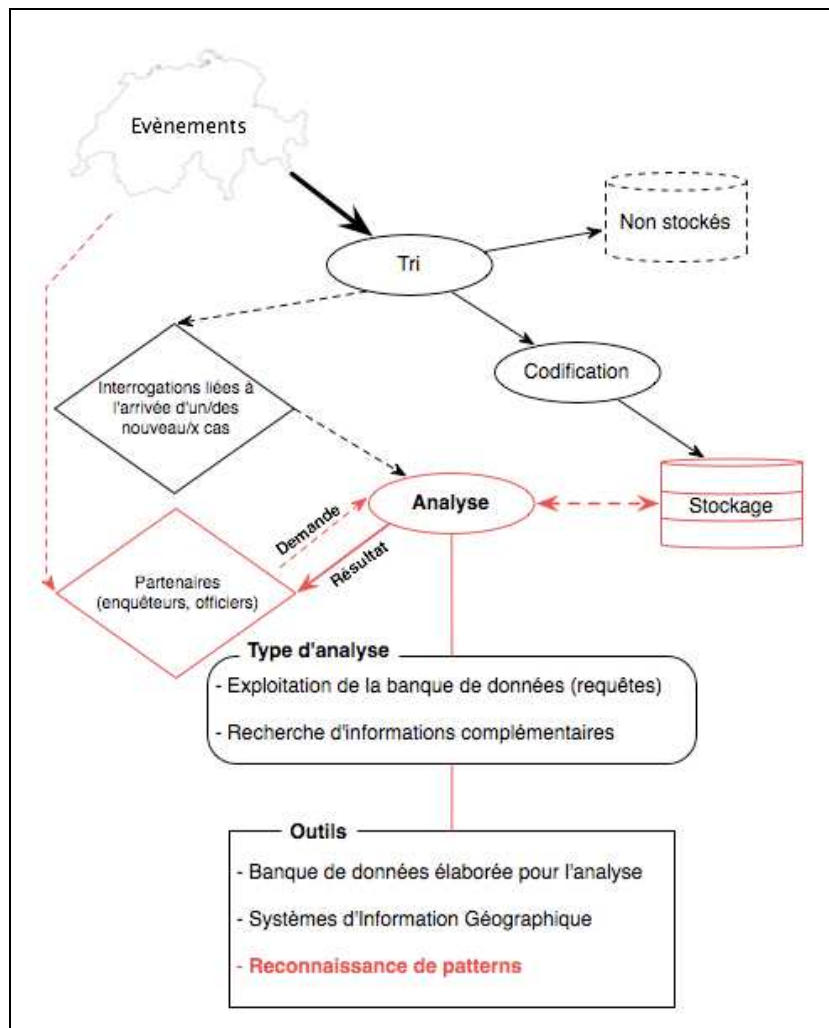


FIGURE 7.4 – Processus d’une demande de la part d’un enquêteur (schéma réalisé en collaboration avec les analystes de la Police Cantonale Vaudoise).

La collaboration et la manière de fonctionner a, dans cet exemple, mené à un résultat utile et utilisé.

La Police récolte une grande quantité de données (mémoire). L’application réalisée dans ce travail a montré qu’il est possible de réaliser des outils pratiques pour exploiter cette mémoire et mettre les éléments en connexion

7.4.2 Une boîte noire dans le processus du renseignement ?

Les méthodes de reconnaissance de patterns font partie de l’étape d’analyse du processus de renseignement (figure 7.5).

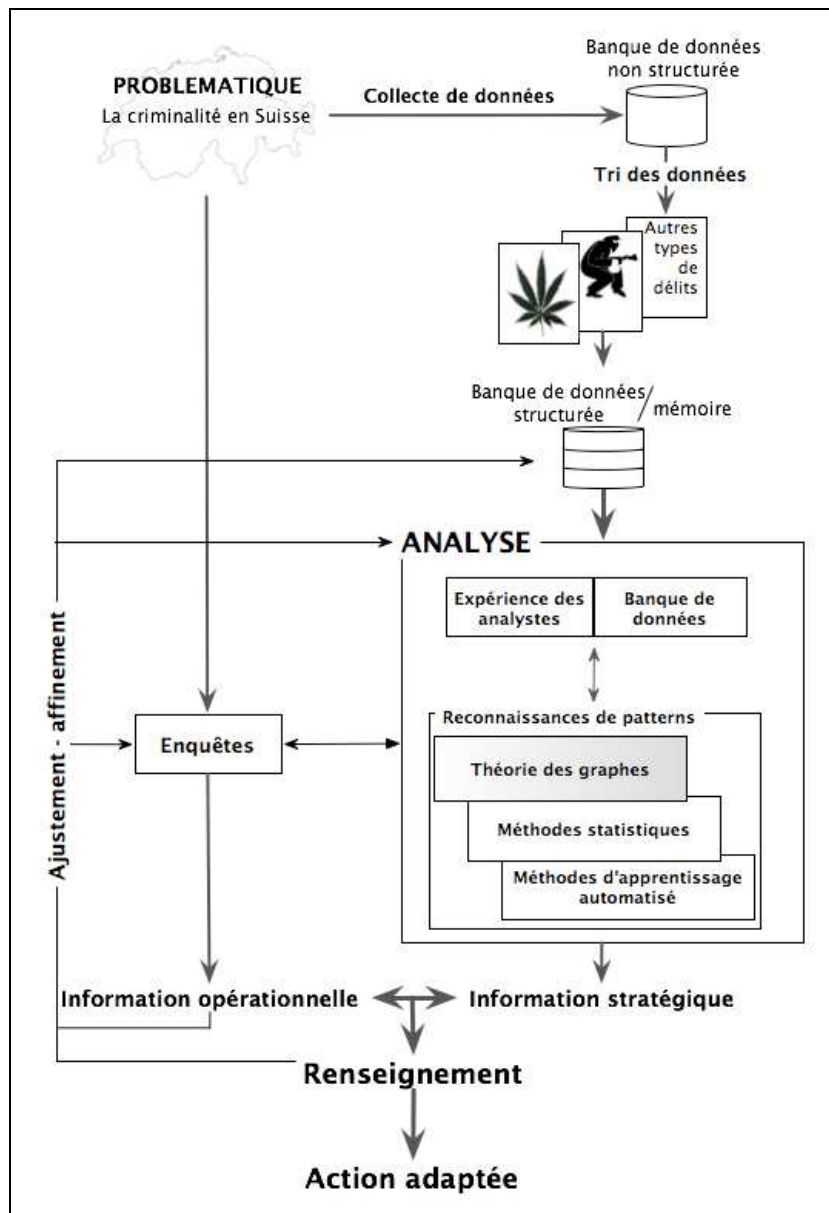


FIGURE 7.5 – Processus de renseignement.

De cette étape, font aussi partie, par exemple, la recherche d'informations dans une banque de données, ainsi que la réflexion des analystes-enquêteurs basée sur leurs expériences. Les méthodes de reconnaissance de patterns constituent une alternative, choisie lorsque une problématique nécessite leur utilisation. D'un point de vue opérationnel, elles constituent une boîte noire, dans le sens où les enquêteurs qui en bénéficient ne savent pas comment elles fonctionnent. Si un expert n'est pas disponible pour développer des méthodes et réaliser l'analyse, elles vont être difficilement applicables. La faible utilisation de ces méthodes en milieu opérationnel vient peut-être de là. Il est possible de

demander une aide externe ponctuelle, mais ceci ne garantira pas une systématique d'analyse sur le long terme. De même, une délocalisation, en demandant à une institution scientifique de réaliser l'analyse, limitera les contacts avec les enquêteurs et analystes opérationnels. L'interprétation des patterns obtenus ne sera pas optimale.

Finalement, que faire pour que ces méthodes soient utilisées? L'application a montré que suivant le cadre de leur manipulation, elles pouvaient apporter des informations opérationnelles.

Disposer des compétences d'enquêteurs, d'analystes, de collaborateurs scientifiques analystes et spécialistes en méthodes est une possibilité, qui dans l'application décrite, a bien fonctionné. L'expérience de ce travail a permis de faire des observations permettant de mettre en évidence certains points qui semblent importants pour une utilisation de ces méthodes :

1. Une collaboration directe avec les enquêteurs. Ils apportent des problématiques opérationnelles mais aussi des informations de validation.
2. Un développement au moyen de librairies plutôt que de programmes préexistants nécessitant des formations que les enquêteurs n'ont pas le temps de suivre. Il est nécessaire de présenter un outil facile d'utilisation sans fonctions superflues.
3. En conséquence, disposer d'un spécialiste en méthode ou d'un analyste avec des connaissances en reconnaissance de patterns qui puisse assurer une évolution des outils et en créer de nouveaux en fonction de la demande et de ses compétences.
4. Avoir la possibilité de faire appel à des experts d'une institution pour des problèmes ponctuels requérant une compétence extraordinaire.

Ces recommandations sont faites selon l'hypothèse qu'il est possible d'utiliser les méthodes de reconnaissance de patterns dans l'étape d'analyse du renseignement. Des problèmes de confidentialité des données, de coûts, de formation, n'ont pas été abordés mais ceux-ci sont probablement aussi une cause de leur faible utilisation.

7.5 Perspectives

Cette recherche a été basée sur deux plans. Le premier a une dimension fondamentale, traitant de l'exploration de méthodes d'analyse et de l'application novatrice de la théorie des graphes dans le domaine du profilage des stupéfiants. Le second, orienté vers l'analyse criminelle, est une réflexion sur la faible utilisation des techniques de reconnaissance de patterns, dans le but de montrer des pistes pour y remédier.

Des perspectives et priorités peuvent être proposées pour une suite à ce travail. Sur un plan fondamental, il serait intéressant de développer les méthodes issues de la théorie des graphes en analysant les structures de graphes au moyen d'algorithmes d'apprentissage automatisé. Ce type

de développement rejoindrait la recherche réalisée par le second doctorant du projet [Ratle *et al.*, 2008] et permettrait d'aller plus loin dans l'automatisation de la détection de patterns. Il pourrait répondre à la question : *Est-il possible de produire du renseignement directement à partir de patterns détectés ?* Autrement dit, est-il possible de calibrer suffisamment le modèle, de manière à extraire des banques de données, des tendances dont la signification serait du renseignement ? Cette recherche ne sera applicable qu'avec la collaboration d'une personne pouvant réaliser une interprétation des patterns, afin d'établir leur pertinence. Elle sera indispensable pour la calibration du système automatisé.

Il serait aussi intéressant d'étudier les combinaisons de l'héroïne et de la cocaïne avec une vision chimique et pharmacologique des produits qui les composent. Interpréter les résultats de cette recherche avec une vision pharmacologique, chimique et orientée analyse criminelle, permettrait peut-être de valider certaines tendances et en expliquer de nouvelles.

Un système s'utilise de manière routinière lorsque sa mise en oeuvre apporte un plus à la structure préexistante. Dans le cas qui nous concerne, si l'intégration des méthodes de reconnaissance de tendances au processus de renseignement permet d'apporter des informations nouvelles et de gagner du temps, il y a des chances que ces méthodes soient un jour utilisées. L'utilisation des nouvelles technologies à des fins criminelles nécessite parfois un recours à des experts. L'intensification de ce type de demande va peut-être ouvrir une voie à ces méthodes en les rendant de plus en plus indispensables.

Chapitre 8

Conclusion

Ce travail avait pour but d'analyser le potentiel du data mining et plus particulièrement des méthodes de reconnaissance de patterns dans une optique de renseignement. Une exploration fondamentale dans le domaine de la théorie des graphes a été réalisée dans le but de tester certaines techniques sur des données liées au trafic des stupéfiants. Ceci a permis de réaliser une application novatrice dont les données initiales proviennent de l'enquête et de cadrer le potentiel d'utilisation des méthodes dans une structure opérationnelle.

Ce travail a permis de soulever plusieurs points quant à leur utilisation dans un processus de renseignement, pouvant peut-être remédier à leur apparent faible recours actuel :

- Disposer de compétences d'experts (analyse criminelle opérationnelle et méthodes d'analyse) ainsi que d'un contact avec les enquêteurs afin de disposer de connaissances *a priori* pour l'interprétation des patterns et leur validation.
- Une méthode n'est pas meilleure qu'une autre mais peut être plus adaptée à la problématique et aux données, d'où l'importance de connaître plusieurs types de techniques.
- Pour une utilisation opérationnelle, s'orienter vers des librairies plutôt que des programmes existants, afin de proposer un outil ciblé évitant à l'utilisateur de se perdre dans les fonctionnalités du logiciel.
- Si des patterns sont mis en évidence, tous les analyser en détail afin de déterminer leur pertinence. Leur interprétation nécessitera l'aide des enquêteurs et des collaborateurs d'une coordination judiciaire par exemple.
- La criminalité évolue ainsi que la technologie et les données qui sont enregistrées. Les méthodes à proposer doivent donc aussi être évolutives et une veille technologique doit pouvoir intervenir dans le travail d'un analyste chargé de ce type de mission.

Au niveau des méthodes, il serait utile de poursuivre la recherche dans le domaine du graph mining appliqué aux données à caractère criminel, en développant des algorithmes d'apprentissage automatisé (“machine learning” en anglais) appliqués à des graphes. Il serait aussi intéressant de tester ces techniques basées sur des graphes à d'autres types de données que celles issues du trafic des stupéfiants, afin d'élargir leur champ d'application.

Les problématiques posées par les enquêteurs amènent à des réflexions pouvant aboutir à une résolution faisant intervenir des analyses fondamentales avec un intérêt scientifique concret. Cet aspect de la recherche est important car il permet d'allier recherche scientifique et pratique qui sont souvent séparées par un fossé. Ce travail montre qu'une telle collaboration est possible.

Bibliographie

Anglada F. et Esseiva P. Les dérivés amphétaminiques en Suisse et dans le monde. *Revue Internationale de Criminologie et de Police Technique et Scientifique*, 55(4) : 465–490, 2002.

Audit-Commission. Helping with Enquiries : Tackling Crime Effectively. *HMSO*, 1993.

Barabási L., Jeong H., Neda Z., Ravasz E., Schubert A. et Vicsek T. Evolution of the social network of scientific collaborations. *Physica A : Statistical Mechanics and its Applications*, 311(3-4) : 590–614, 2002.

Batagelj V. et Mrvar A. Pajek - Program for Large Network Analysis. *Connections*, 21(2) : 47–57, 1998.

Batagelj V. et Mrvar A. Pajek - Analysis and Visualization of Large Networks. Dans **Junger M. et Mutzel P.**, éditeurs. *Graph Drawing Software*, 8–11, Berlin. Springer, 2003.

Bavelas A. A mathematical model for group structure. *Human Organizations*, 7 : 16–30, 1948.

Bavelas A. Communication patterns in task-oriented groups. *Journal of the Acoustic Society of America*, 22 : 271–282, 1950.

Baylon A., 2007. Utilisation de graphes spécifiques pour représenter les évolutions des combinaisons de produits de coupages de lhérone et les possibilités d'extension de la méthode à d'autres formes d'activité criminelles. Rapport technique, Ecole des Sciences Criminelles, Université de Lausanne.

Belkin M. et Niyogi P. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6) : 1373–1396, 2003.

Berge C. Théorie des Graphes et ses Applications. *Dunod*, Paris, 1958.

Birrer S. Système de classification basé sur les approches situationnelles pour le suivi de la délinquance sérielle. Thèse de doctorat, Ecole des Sciences Criminelle, Université de Lausanne, (2009). En préparation.

- Birrer S. et Terrettaz-Zufferey A.-L.** Croisement spatial et temporel de données issues d'activités délictueuses et d'appareils permettant une géolocalisation. L'apport de la théorie des graphes dans l'automatisation de la comparaison entre des données personnelles issues de téléphones mobiles ou de GPS et des infractions enregistrées par la police. *Revue internationale de Criminologie et de Police technique et scientifique*, 2008. In Press.
- Bishop Chris M.** Novelty detection and neural network validation. *IEE Proceedings : Vision, Image and Signal Processing*, 141 : 217–222, 1994.
- Boser B., Guyon I. et Vapnik V.** A training algorithm for optimal margin classifiers. Dans *Fifth Annual Workshop on Computational Learning Theory*, 144–152, New-York, USA. ACM, 1992.
- Brodeur J.-P.**, 2006. Relier les indices - Connecting the dots. Text presented at a seminar on security.
- Brodeur J.-P. et Dupont B.** Knowledge Workers or “Knowlegde” Workers? *Policing & Socitey*, 16(1) : 7–26, 2006.
- Brodeur J.-P. et Lemay-Langlois S.** Surveillance totale ou surveillance-fiction. *Les cahiers de la sécurité intérieure*, 55 : 61–90, 2004.
- Bron C. et Kerbosch J.** Algorithm 457 : finding all cliques of an undirected graph. *Commun. ACM*, 16(9) : 575–577, 1973.
- Chen Hsinchun, Chung Wingyan, Qin Yi, Chau Michael, Xu Jennifer Jie, Wang Gang, Zheng Rong et Atabakhsh Homa.** Crime Data Mining : An Overview and Case Studies. Dans *Proceedings of the 2003 annual national conference on Digital government research*, 1–5, Boston, MA. Digital Government Society of North America, 2003.
- Chung W., Chen H., Chang W. et Chouc S.** Fighting cybercrime : a review and the Taiwan experience. *Decision Support Systems*, 41 : 669–682, 2006.
- Clauset A., Newman M. E. J. et Moore C.** Finding community structure in very large networks. *Physical Review E*, 70(066111) : 1–6, 2004.
- Cook D. J., Holder L. B. et Ketkar N.** Unsupervised and Supervised Pattern Learning in Graph Data. Dans **Cook Diane J. et Holder Lawrence B.**, éditeurs. *Mining Graph Data*, 159–182, Hoboken, New Jersey. John Wiley & Sons, 2006.
- Corcoran J.J., Wilson A.D. et Ware A.** Predicting the geo-temporal variations of crime and disorder. *International Journal of Forecasting*, 19 : 623–634, 2003.

- Corley C. D., Mikler A. R., Cook D. J. et Singh K.** Dynamic intimate contact social network and epidemic intervention. *Functional Informatics and Personalized Medicine*, 1(2) : 171–188, 2008.
- Csárdi Gábor et Nepusz Tamás**, 2003. IGraph Library. Rapport technique, KFKI RIPNP of the HAS, Budapest, Hungary.
- Cusson Maurice.** L'analyse stratégique et quelques développements récents en criminologie. *Criminologie*, 19 : 53–72, 1986.
- Davenport T. H.** Information Ecology : Mastering the Information and Knowledge Environment. *Oxford University Press*, New-York, 1997.
- Del Valle S. Y., Hyman J. M., Hethcote H. W. et Eubank S. G.** Mixing patterns between age groups in social networks. *Social Networks*, 29(4) : 539–554, 2007.
- Demyanov V., Soltani S., Kanevski M., Canu S., Maignan M., Savelieva E., Timonin V. et Pisarenko V.** Wavelet Analysis Residual Kriging vs Neural Network Residual Kriging. *Journal of Stochastic Environmental Research and Risk Assessment*, 15 : 18–32, 2001.
- Diane J. Cook Lawrence B. Holder.** Mining Graph Data. *John Wiley & Sons*, Hoboken, New Jersey, 2007.
- Didimo W. et Liotta G.** Graph Visualization and Data Mining. Dans **Cook Diane J. et Holder Lawrence B.**, éditeurs. *Mining Graph Data*, 35–63, Hoboken, New Jersey. John Wiley & Sons, 2006.
- Dubrawski Artur, Moore Andrew et Sabhnani Robin.** Monitoring Food Safety by Detecting Patterns in Consumer Complaints. Dans *Proceedings of the National Conference on Artificial Intelligence AAAI/IAAI 2006*, 2006.
- Dujourdy L., Barbati G., Taroni F., Guéniat O., Esseiva P., Anglada F. et Margot P.** Evaluation of links in heroin seizures. *Forensic Science International*, 131 : 171–183, 2003.
- Dupont B. et Ratcliffe J. H.** Juste des punaises sur une carte ? Quelques considérations critiques sur la cartographie criminelles. *Les Cahiers de la sécurité intérieure*, 41 : 229–243, 2000.
- Egger S. A.** A Working Definition of Serial Murder and the Reduction of Linkage Blindness. *Journal of Police Science and Administration*, 12(3) : 348–355, 1984.
- Esseiva P.** Application of neural network in drug intelligence based on chemical profiling (poster presentation). Dans *16th International Symposium on the Forensic Sciences, Canberra, Australia*, Canberra, Australia, 2002.

- Esseiva P.** Le profilage de l'héroïne et de la cocaïne - Mise en place d'une systématique permettant une utilisation opérationnelle des liens chimiques. Thèse de doctorat, Ecole des Sciences Criminelles. Université de Lausanne, 2004.
- Esseiva P., Anglada F., Dujourdy L. et Margot P.** A Methodology for Illicit Drug Seizures Comparison in a Drug Intelligence Perspective Using Large Databases. *Forensic Science International*, 132 : 139–152, 2003.
- Esseiva P., Anglada F., Dujourdy L., Taroni F., Margot P., du Pasquier E., Dawson M., Roux C. et Doble P.** Chemical profiling and classification of illicit heroin by principal component analysis, calculation of inter sample correlation and artificial neural networks. *Talanta*, 67 : 360–367, 2005.
- Esseiva P., Dujourdy L., Anglada F., Taroni F. et Margot P.** Les signatures chimiques à partir des saisies de produits stupéfiants. Utilité et gestion de l'information. *Revue Internationale de Criminologie et de Police Technique et Scientifique*, (1) : 104–111, 2002.
- Esseiva P. et Lock E. et al.** Identification and quantification of amphetamine and analogues by capillary zone electrophoresis. *Science & Justice*, 37(2) : 113–119, 1997.
- Esseiva P., S. Ioset, Anglada F., Gasté L., Ribaux O., Margot P., Galluser A., Biedermann A., Specht Y. et Ottinger E.** Forensic drug Intelligence : An important tool in law enforcement. *Forensic Science International*, 167 : 247–254, 2007.
- Euler L.** Solutio Problematis ad Geometrica Situs Pertinentis. *Mémoires de l'Académie des Sciences de Berlin*, 8 : 128–140, 1759.
- Felson M.** Ecology of Crime. *Encyclopedia of Crime and Justice*, 11, 1983.
- Felson M.** Crime and Everyday Life. *Pine Forge Press*, Thousans Oaks, CA, second edition, 1998.
- Felson M. et Clarke R. V.** Opportunity Makes the Thief. Practical theory for crime prevention. *Police Research Series*, London, 1998.
- Felson M. et Cohen L. E.** Human Ecology and Crime : A Routine Activity Approach. *Human Ecology*, 8 : 384–406, 1980.
- Frawley W. J., Shapiro Piatetsky G. et Matheus C. J.** Knowledge discovery in databases - an overview. *Ai Magazine*, 13 : 57–70, 1992.

- Fucci N. et De Giovanni N.** Adulterants Encountered in the Illicit Cocaine Market. *Forensic Science International*, 95 : 247–252, 1998.
- Gardi F.** Ordonnancement avec exclusion mutuelle par un graphe d'intervalles ou d'une classe apparentée : complexité et algorithmes. Thèse de doctorat, Faculté des Sciences de Luminy, Marseille, 2005.
- Girod A.** Exploitation et gestion systématiques des traces de souliers : une approche complémentaire pour l'investigation criminelle des cambriolages. Thèse de doctorat, Ecole des Sciences Criminelle, Université de Lausanne, 2002.
- Girod A., Champod C. et Ribaux O.** Les traces de souliers. *Presses Polytechniques Universitaires Romandes*, Lausanne, 2006.
- Girod A., Ribaux O., Walsh S. J. et Margot P.** Bases de données ADN : un potentiel peu exploité de mise en relation d'événements criminels. *Revue Internationale de Criminologie et de Police Technique et Scientifique*, 2 : 131–147, 2004.
- Goldstein H.** Problem Oriented Policing. *Temple University Press*, Philadelphia, 1990.
- Gomez J. et Rodriguez A.**, 1989. An evaluation of the results of a drug sample analysis. United Nations Office on Drugs and Crime (UNODC).
- Gorr W., Olligschlaeger A. et Thompson Y.** Short-term forecasting of crime. *International Journal of Forecasting*, 19 : 579–594, 2003.
- Guéniat O.** Le profilage de l'héroïne et de la cocaïne - les méthodes d'analyses, la modélisation du concept du profilage, la gestion et l'exploitation des liens. Thèse de doctorat, Ecole des Sciences Criminelles, Université de Lausanne, 2004.
- Guéniat O., Anglada F., Taroni F., Margot P., Dujourdy L. et Esseiva P.** Le profilage des saisies de produits stupéfiants. Utilité et gestion de l'information. *Revue Internationale de Criminologie et de Police Technique et Scientifique*, 104–111, 2002.
- Guéniat O. et Esseiva P.** Histoire, production et trafic de la cocaïne. *Dépendance*, (18) : 4–12, 2002.
- Guéniat O. et Esseiva P.** Le profilage de l'héroïne et de la cocaïne. Une méthodologie moderne de lutte contre le trafic illicite. Série criminalistique. *Presses Polytechniques et Universitaires Romandes*, Lausanne, 2005.

- Guex J.** Unitary association method : use of graph theory and computer algorithm. *Computers and Geosciences*, 10(1) : 69–96, 1984.
- Guex J.** Corrélations biochronologiques et associations unitaires. *Presses Polytechniques Romandes.*, 1987.
- Hagberg A., Schult D. et Swart P.**, 2004. Networkx - High productivity software for complex networks. Rapport technique, Mathematical Modeling and Analysis, Theoretical Division.
- Hammer O., Harper D.A.T. et Ryan P.D.** PAST : Paleontological Statistics Software Package for Education and Data Analysis. *Palaeontologia Electronica*, 4(1) : 1–9, 2001.
- Harary F., Norman R.Z. et Cartwright D.** Structural Models : An Introduction to the Theory of Directed Graphs. *John Wiley & Sons*, New York, 1965.
- Hastie T., Tibshirani R. et Friedman J.** The Elements of Statistical Learning. *Springer*, New York, 2001.
- House-Of-Common**, 2005. Forensic Sciences on Trial. Science and Technology, seventh report.
- Innes M., Fielding N. et Cope N.** The Appliance of Science ? The Theory and Practice of Crime Intelligence Analysis. *British Journal of Criminology*, 45 : 39–57, 2005.
- Ioset S., Esseiva P., Ribaux O., Weyermann C., Anglada F., Locicero S., Hayoz P., Baer I., Gasté L., Terrettaz-Zufferey A.-L., De la Porte C. et Margot P.** Establishment of an operational system for drug profiling : a Swiss experience. *Bulletin on Narcotics*, LVII(1 and 2) : 121–147, 2005. United Nations Office on Drugs and Crime.
- Jackson G.** The nature of forensic science opinion - a possible framework to guide thinking and practice in investigations and in court proceedings. Dans *Forensic Science Society Autumn Meeting*, Wyboston, UK. FSS, 2004.
- Jackson G., Champod C., Evett I. W. et Crossen S. M.** Investigator/Evaluator - A Possible Framework to Guide Thinking and Practice for Forensic Scientist. *Science & Justice*, 46(1) : 33–45, 2006.
- Janzen K. E., Walter L. et Fernando. A. R.** Comparison Analysis of Illicit Cocaine Samples. *Journal of Forensic Sciences*, 37 : 436–445, 1992.
- Jesus M.** Investigative Data Mining for Security and Criminal Detection. *Butterworth Heinemann*, USA, 2003.

- Kanevski M., Arutyunyan R., Bolshov L. et Demyanov V.** Artificial neural networks and spatial estimations of Chernobyl fallout. *Geoinformatics*, 7 : 5–11, 1996.
- Kanevski M. et Maignan M.** Analysis and Modelling of Spatial and Environmental Data. *EFPL Press*, Montréal, Québec, 2004.
- Kanevski M. et Maignan M.** Empirical analysis and mapping of interest rates. Dans *International Workshop on Spatial Econometrics and Statistics*. SES, 2006a.
- Kanevski M. et Maignan M.** Large Scale Mapping of Radioactively Contaminated Territories. Dans *European Geosciences Union General Assembly*. EGUGA, 2006b.
- Kanevski M., Pozdnukhov A., Canu S. et Maignan M.** Advanced Spatial Data Analysis and Modelling with Support Vector Machines. *International Journal on Fuzzy Systems*, 4(1) : 606–615, 2002a.
- Kanevski M., Pozdnukhov A., Canu S., Maignan M., Wong P.M. et Shibli S.A.R.** Support Vector Machines for Classification and Mapping of Reservoir Data. Dans **Wong P., Aminzadeh F. et Nikravesh M.**, éditeurs. *Soft Computing for Reservoir Characterization and Modeling*, 531–558, Heidelberg, N.Y. Physica-Verlag, 2002b.
- Kanevski M., Pozdnukhov A. et Maignan M.** Active Learning of Environmental Data using Support Vector Machines. Dans *International Association for Mathematical Geology*. IAMG, 2005a.
- Kanevski M., Pozdnukhov A., Tonini M., Motelica M., Savelieva E. et Maignan M.** Statistical learning theory for geospatial data. Dans *European colloquium on Theoretical and Quantitative Geography*. ECTGQG, 2005b.
- Kantz H. et Schreiber T.** Nonlinear Time Series Analysis (2nd edition). *Cambridge University Press*, Cambridge, UK, 2004.
- Killias M.** The Opening and Closing of Breaches - A Theory on Crime Waves, Law Creation and Crime Prevention. *European Journal of Criminology*, 3(1) : 11–31, 2006.
- Kind S. S.** The Scientific Investigation of Crime. *Forensic Science Services Ltd*, Harrogate, 1987.
- Kruskal J. B. et Wish M.** Multidimensional Scaling. Dans **Uslaner E. M.**, éditeur. *Quantitative Applications in the Social Sciences*, Beverly Hills and London. SAGE Publications, 1978.
- Kubica Jeremy et Moore Andrew.** Variable KD-Tree Algorithms for Spatial Pattern Search and Discovery. Dans *Advances in Neural Information Processing Systems*. ANIPS, 2005.

- Kulldorff M., Athas W.F., Feuer E.J., Miller B.A. et Key C.R.** Evaluating cluster alarms : A space-time scan statistic and brain cancer in Los Alamos, New Mexico. *American Journal of Public Health*, 88(9) : 1377–1380, 1998.
- Kulldorff M., Heffernan R., Hartman J., Asuncao R. et Mostashari F.** A space-time permutation scan statistic for disease outbreak detection. *PLoS Medicine*, 2(3) : 216–224, 2005.
- Leary D. et Pease K.**, 2002. DNA and the active criminal population. Rapport technique Ad Hoc Publications, Jill Dando Institute of Crime Science.
- Lemieux F.** Normes et pratiques en matière de renseignement criminel. Une comparaison internationale, volume 38. *Les Presses de L'Université Laval*, Saint-Nicolas, Québec, 2005.
- Lociciro S.** Profilage de la cocaine : approche inter laboratoires. Thèse de doctorat, Ecole des Sciences Criminelles, Université de Lausanne, 2007.
- Lociciro S., Hayoz P., Esseiva P., Duourdy L., Besacier F. et Margot P.** Cocaine profiling for strategic intelligence purpose, a cross-border project between France and Switzerland Part I. Optimisation and harmonisation of the profiling method. *Forensic Science International*, 167(2-3) : 220–228, 2007.
- Lociciro S., Hayoz P., Esseiva P., Duourdy L., Besacier F. et Margot P.** Cocaine profiling for strategic intelligence, a cross-border project between France and Switzerland Part II. Validation of the statistical methodology for the profiling of cocaine. *Forensic Science International*, 177(2-3) : 199–206, 2008.
- Low G.**, 2004. Mac OS X Edition of Graphviz. Rapport technique, Pixelglow.
- Luce R.D. et Perry A.D.** A method of matrix analysis of group structure. *Psychometrika*, 14 : 95–116, 1949.
- MacEvoy B. et Freeman L.**, 2000. UCINET for Windows Version 6 : Software for Social Network Analysis. Rapport technique, Analytic Technologies, Inc.
- Madden M.G. et Ryder A.G.** Machine learning methods for quantitative analysis of Raman spectroscopy data. Dans *Proceedings of the International Society for Optical Engineering Conference*, volume 4876, 1130–1139. ISOEC, 2002.
- Marcus S., Moy M. et Coffman T.** Social Network Analysis. Dans **Cook Diane J. et Holder Lawrence B.**, éditeurs. *Mining Graph Data*, 443–469, Hoboken, New Jersey. John Wiley & Sons, 2006.

- McCue Colleen.** Data Mining and Predictive Analysis : Intelligence Gathering and Crime Analysis. *Elsevier*, Newton, MA, USA, 2006.
- McQueen J.B.** Some methods for classification and analysis of multivariate observations. Dans *Proceedings of the Fifth Berkeley Symposium on mathematics, Statistics and Probability*, volume 1, 281–298, 1967.
- Mena J.** Data Mining for Security and Criminal Detection. *Elsevier Science*, Burlingto, MA, 2002.
- Mennell J.** The Future of Forensic and Crime Scene Science Part II, A UK Perspective on Forensic Science Education. *Forensic Science International*, 157(Supplement 1) : S13–S20, 2006.
- Mennell J. et Shaw I.** The Future of Forensic and Crime Scene Science Part I - A UK Forensic Science User and Provider Perspective. *Forensic Science International*, 157(Supplement 1) : S7–S12, 2006.
- Morley S.R., Forrest A.R.W et Galloway J.H.** Levamisole as a contaminant of illicit cocaine. Dans *Proceedings of the TIAFT*, 1–2, 2006.
- Morselli C., Giguère C. et Petit K.** The efficiency/security trade-off in criminal networks. *Social Networks*, 29(1) : 143–153, 2007.
- Nath Shyam Varan.** Crime Pattern Detection Using Data Mining. Dans *WI-IATW '06 : Proceedings of the 2006 IEEE/WIC/ACM international conference on Web Intelligence and Intelligent Agent Technology*, 41–44, Washington, DC, USA. IEEE Computer Society, 2006.
- NCIS, 2000.** The National Intelligence Model. Rapport technique, National Crime Intelligence Service.
- O’Connell M.L., Howley T., Ryder A.G. et Madden M.G.** Classification of target analyte in solid mixtures using principal component analysis, support vector machines, and Raman spectroscopy. Dans *Proceedings of the International Society for Optical Engineering*, 340–350. ISOE, 2005.
- OFSP, 2006.** Conférence nationale sur les drogues de synthèse et la cocaïne. Office fédérale de la Santé Publique (OFSP).
- OharaK. N P. C., Mogi A., Motoda H. et Washio T.** Constructing Decision Tree Based on Chunkingless Graph-Based Induction. Dans **Cook Diane J. et Holder Lawrence B.**, éditeurs. *Mining Graph Data*, 203–226, Hoboken, New Jersey. John Wiley & Sons, 2006.

- Olligschlaeger A.M. et Gorr W.** Artificial Neural Networks and Crime Mapping. Dans **Weisburd D. et McEwen T.**, éditeurs. *Crime Mapping and Crime Prevention*, Monsey, NY. Criminal Justice Press, 1997.
- ONU DC**, 2006. Rapport Mondial sur les Drogues. Office des Nations Unies contre la drogue et le crime (ONU DC).
- Pearson K.** On Lines and Planes of Closest Fit to Systems of Points in Space. *Philosophical Magazine*, 2(6) : 559–572, 1901.
- Pons Pascal et Latapy Matthieu.** Computing communities in large networks using random walks (long version). *Journal of Graph Algorithms and Applications*, 10(2) : 191–218, 2006.
- Potts J. T., Cook D., Holder L. B., et Coble J.** Learning Concepts from Intelligence Data Embedded in a Supervised Graph. Dans *International Conference on Intelligence Analysis*, Arlington, TX 76019 USA. u.a., 2005.
- Pozdnukhov A. et Kanevski M.** Monitoring Network Optimisation for Spatial Data Classification Using Support Vector Machines. *International Journal of Environment and Pollution*, 28 : 465–484, 2006.
- Quinlan J. M.** Learning Logical Definitions from Relations. *Machine Learning*, 5 : 239–266, 1990.
- Ratcliffe J. H.** Integrated Intelligence and Crime Analysis : Enhanced Information Management for Law Enforcement Leaders. *Police Foundation*, Washington DC, 2007.
- Ratcliffe J. H.** Intelligence-led Policing. *Willan*, Cullompton, UK, 2008.
- Ratle F., Terrettaz-Zufferey A.-L., Kanevski M., Esseiva P. et Ribaux O.** Learning Manifolds in Forensic Data. Dans *International Conference on Artificial Neural Networks*, Athens, Greece. Springer, 2006a. in press.
- Ratle F., Terrettaz-Zufferey A.-L., Kanevski M., Esseiva P. et Ribaux O.** Pattern analysis in illicit heroin seizures : a novel application of machine learning algorithms. Dans *European Symposium on Artificial Neural Networks (ESANN)*, Bruges, Belgium. d-side publications, 2006b.
- Ratle F., Weston J. et Miller M.L.** Large-scale clustering through functional embedding. Dans *Proceedings of the European Conference on Machine Learning (ECML 2008)*, Antwerp, Belgium. ECML, 2008.

- Ribaux O.** La recherche et la gestion des liens dans l'investigation criminelle : le cas particulier du cambriolage. Thèse de doctorat, Ecole des Sciences Criminelle, Université de Lausanne, 1997.
- Ribaux O.** Comprendre l'influence des NTIC sur l'évolution de la criminalité et des moyens de lutte : un champ d'étude pour les sciences criminelles. Dans **Augsburger-Bucheli I. et Bacher J.-L.**, éditeurs. *Manifestation, détection, prévention et répression de la criminalité économique*, 263– 274, Paris. L'Harmattan, 2005a.
- Ribaux O.** Le renseignement criminel pour le traitement de la délinquance sérielle dans un système fédéraliste : de l'idée à la mise en oeuvre. Dans *Colloque international francophone, la police et les citoyens*, Nicolet, Québec. CIF, 2005b.
- Ribaux O.** Les outils informatisés du renseignement criminel. Dans **Cimichella S., Kuhn A. et Niggli M. A.**, éditeurs. *Nouvelles technologies et criminalité : nouvelle criminologie ?*, 135 – 156, Zürich. Ruegger, 2006.
- Ribaux O.** Définitions. Dans **Geradts Zeno**, éditeur. *D6.7c : Forensic Profiling*, The Netherlands, 2008.
- Ribaux O. et Aepli P.** L'analyse des délits contre le patrimoine et son exploitation dans un cadre opérationnel. *Revue Internationale de Criminologie et de Police Technique et Scientifique*, 54 : 131–144, 2001.
- Ribaux O. et Birrer S.** Système de suivi et d'analyse des cambriolages appliqué dans des polices suisses. Dans *Proceedings de : Erstes Zürcher Präventionsforum - Kommunale Kriminalprävention - Crime Mapping - Einbruchskriminalität*, Université de Zürich, Europa Institute Zürich, 2008. (à paraître).
- Ribaux O., Birrer S. et Walsh S.** Three Level Architecture for the Analysis of Serial Burglary that Integrates Crime Mapping Tools and Forensic Case Data. *Forensisc Science International*, 136 : 10–11, 2003a.
- Ribaux O., Girod A., Walsh S., Margot P., Mizrahi S. et Clivaz V.** Forensic Intelligence and Crime Analysis. *Probability, Law and Risk*, 2(2) : 47–60, 2003b.
- Ribaux O., Gitz P. et Cartier J.**, 2002. L'analyse criminelle face à la complexité des données : risque pour la sphère privée ou moyen de la protéger. Rapport technique, 60^{ème} cours international de criminologie, Société Internationale de Criminologie. Helbing & Lichtenhahn, Bâle, Neuchâtel.

- Ribaux O. et Margot P.** Inference Structures for Crime Analysis and Intelligence Using Forensic Science Data : the Example of Burglary. *Forensic Science International*, 100 : 193–210, 1999.
- Ribaux O. et Margot P.** Crime-scene Investigation and Examination : Criminal Analysis. Dans **Siegel J., Saukko P. et Knupfer G.**, éditeurs. *Encyclopaedia of Forensic Sciences*, 416–422, London. Academic press, 2000.
- Ribaux O. et Margot P.**, 2001. The Analysis of Serial Crime through the Use of Different Sources of Data. Rapport technique, School of Crime Sciences, University of Lausanne.
- Ribaux O. et Margot P.** Case-Based Reasoning in Criminal Intelligence using Forensic Case Data. *Science & Justice*, 43(3) : 135–143, 2003.
- Ribaux O. et Margot P.** La trace comme vecteur d'information au service du renseignement. Dans **Cusson M., Dupont B. et Lemieux F.**, éditeurs. *Traité sur la sécurité intérieure*, Montréal. Hurtubise HMH, 2007.
- Ribaux O., Walsh S. J. et Margot P.** The Contribution of Forensic Science to Crime Analysis and Investigation : Forensic Intelligence. *Forensic Science International*, 156 : 171–181, 2006.
- Rosenblatt F.** The Perceptron : A Probabilistic Model for Information Storage and Organization in the Brain. *Psychological Review*, 65(6) : 386–408, 1958.
- Roweis S.T. et Saul L.K.** Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290 : 2323–2326, 2000.
- Rupnik Jan**, 2006. Finding Community Structure In Social Network Analysis - Overview. Rapport technique, Department of Knowledge Technologies Jozef Stefan Institute.
- Sabhnani Robin, Neill Daniel et Moore Andrew.** Detecting Anomalous Patterns in Pharmacy Retail Data. Dans *Proceedings of the KDD 2005 Workshop on Data Mining Methods for Anomaly Detection*, 2005.
- Said Y. H., Wegman E. J., Sharabati W. K. et Rigsby J. T.** Social networks of author-coauthor relationships. *Computational Statistics & Data Analysis*, 52(4) : 2177–2184, 2008.
- Saks M.J., Risinger D.M., Rosenthal R. et Thompson W.C.** Context effects in Forensic Science : A Review and Application of the Science of Science to Crime Laboratory Practice in the United States. *Science & Justice*, 43(2) : 77–90, 2003.
- Seidman S.B.** Network structure and minimum degree. *Social Networks*, 5 : 269–287, 1983.

- Shane J. M.**, 2004a. Compstat Implementation. Rapport technique June 2004, FBI Law Enforcement Bulletin.
- Shane J. M.**, 2004b. Compstat Process. Rapport technique April 2004, FBI Law Enforcement Bulletin.
- Shane J. M.**, 2004c. Comptstat Design. Rapport technique May 2004, FBI Law Enforcement Bulletin.
- Shawe-Taylor J. et Cristianini N.** Kernel Methods for Pattern Analysis. *Cambridge University Press*, Cambridge, UK, 2004.
- Sheptycki J.** Organizational Pathologies in Police Intelligence : some Contributions to the Lexicon of Intelligence-led Policing. *European Journal of Criminology*, 1(3) : 307–332, 2004.
- Shimbo M. et Ito T.** Kernels as Links Analysis Measures. Dans **Cook Diane J. et Holder Lawrence B.**, éditeurs. *Mining Graph Data*, 283–308, Hoboken, New Jersey. John Wiley & Sons, 2007.
- Stevens G. M.**, 2003. Privacy : Total Information Awareness Programs and Related Information Access, Collection, and Protection Laws. Rapport technique, Congressional Research Service - The Library of Congress.
- Terrettaz-Zufferey A.-L.**, 2005. Recognition of patterns in forensic case data : the use of chemical/physical signatures of illicit drug seizures and information collected on burglary sites in an intelligence perspective. Mémoire de thèse, School of Crime Sciences, University of Lausanne.
- Terrettaz-Zufferey A.-L., Ratle F., Ribaux O., Esseiva P. et Kanevski M.** Assessment of Data Mining Methods for Forensic Case Data Analysis. *Journal of Criminal Justice and Security*, 8 : 350–355, 2006.
- Terrettaz-Zufferey A.-L., Ratle F., Ribaux O., Esseiva P. et Kanevski M.** Pattern Detection in Forensic Case Data Using Graph Theory : Application to Heroin Cutting Agents. *Forensic Science International*, 167 : 242–246, 2007.
- Tufféry S.** Data mining et statistiques décisionnelle. L'intelligence des données. *TECHNIP*, France, 2007.
- Tuia D., Kanevski M. et Kaiser C.** Clustering of Environmental Monitoring Networks : Dimensional Resolutions and Patterns Detection. Dans *International Conference on Geostatistics for Environmental Applications*, 2006.

- United-States**, 2004. The 9/11 Commission Report. Rapport technique, National Commission on Terrorist Attacks Upon the United States.
- UNODC**, 2001. Drug Characterization Impurity Profiling. United Nations Office on Drugs and Crime (UNODC).
- UNODC**, 2008. World Drug Report 2008. United Nations, Office on Drugs and Crime.
- Van Rossum G.**, 1990. Python. Rapport technique, Python Software Foundation.
- Walsh S. J., Ribaux O., Buckelton J. S., Ross A. et Roux C.** DNA profiling and criminal justice - a contribution to a changing debate. *The Australian Journal of Forensic Science*, 36(1) : 34–43, 2004.
- Walsh S. J., Roux C., Ross A., Ribaux O. et Buckleton J. S.** Forensic DNA Profiling : Beyond Identification. *Law Enforcement Forum*, 2(3) : 13–21, 2002.
- Wasserman Stanley, Faust Katherine, Iacobucci Dawn et Granovetter Mark.** Social Network Analysis : Methods and Application. *Cambridge University Press*, wasserman & faust, 1994.
- West D. B.** Introduction to Graph Theory. *Prentice Hall*, Upper Saddle River, New Jersey, 2nd, 1996.
- White H.C., Boorman S.A. et Breiger R.L.** Social structure from multiple networks. I. Blockmodels of roles and positions. *American Journal of Sociology*, 81 : 730–779, 1976.
- Wiggett A. E., Walters A.E., O’Hanlon L. et Ritchie F. H.** Forensic Science Society Spring Meeting 2002 : Intelligence. *Science & Justice*, 43(2) : 109–118, 2003.
- Witzig E. W.**, 2003. The New VICAP. Rapport technique Juin, FBI Law Enforcement Bulletin.
- Wold S. et Sjostrom M.** SIMCA : A method for analysing chemical data in terms of similarity and analogy. Dans *Chemometrics Theory and Application*, 52, 243–282, Washington D.C. American Chemical Society Symposium, Kowalski, B.R, ed., 1977.
- Wylie J. L., Shah L. et Jolly A.** Incorporating geographic settings into a social network analysis of injection drug use and bloodborne pathogen prevalence. *Health & Place*, 13(3) : 617–628, 2007.
- Xue Y. et Brown D. E.** Spatial analysis with preference specification of latent decision makers for criminal event prediction. *Decision Support System*, 41 : 560–573, 2006.

Zingg C. et Esseiva P., 1998. Les composantes de l'héroïne saisie par les services de police - rapport intermédiaire. Rapport technique, Ecole des Sciences Criminelles, Université de Lausanne. Volume 1 : recherche fondamentale.

Annexe A

Codes des combinaisons de stupéfiants

A.1 Héroïne

C0.0 0

C1.1 caféine

C2.1 glucose

C3.1 lactose

C4.1 sucres

C5.2 acides_gras-glucose

C6.2 caféine-griseofulvine

C7.2 caféine-lactose

C8.2 caféine-paracétamol

C9.2 glucose-mannitol

C10.2 lactose-mannitol

C11.2 lactose-procaïne

C12.3 caféine-diazépam-paracétamol

C13.3 caféine-diéthylène-paracétamol

C14.3 caféine-glycérol-paracétamol

C15.3 caféine-griseofulvine-paracétamol

C16.3 caféine-lactose-paracétamol

C17.3 caféine-maltose-paracétamol

C18.3 caféine-mannitol-paracétamol

C19.3 caffeine-paracetamol-phenobarbital
C20.3 caffeine-paracetamol-procaine
C21.3 caffeine-paracetamol-sucrose
C22.3 lactose-methadone-sucrose
C23.4 acide_citrique-fructose-glucose-sucrose
C24.4 caffeine-diazepam-griseofulvine-paracetamol
C25.4 caffeine-glucose-paracetamol-sucrose
C26.4 caffeine-glycerol-griseofulvine-paracetamol
C27.4 caffeine-glycerol-lactose-paracetamol
C28.4 caffeine-glycerol-mannitol-paracetamol
C29.4 caffeine-glycerol-paracetamol-sucrose
C30.4 caffeine-griseofulvine-mannitol-paracetamol
C31.4 caffeine-griseofulvine-paracetamol-phenobarbital
C32.4 caffeine-griseofulvine-paracetamol-sucrose
C33.5 caffeine-glycerol-griseofulvine-mannitol-paracetamol
C34.5 caffeine-lactose-mannitol-paracetamol-piracetam
C35.7 acide_citrique-caffeine-diazepam-glucose-lactose-paracetamol-phenobarbital

A.2 Cocaïne

C0 0
C1 amidon
C2 antipyrine
C3 caffeine
C4 diltiazem
C5 glucose
C6 hydroxyzine
C7 inositol
C8 lactose
C9lidocaine
C10 mannitol
C11 phenacetine
C12 piracetam
C13 procaine
C14 sucres
C15 sucrose

- C16 1,4_dmt-mannitol
- C17 1,4_dmt-phenacetine
- C18 acide_ascorbique-glucose
- C19 acide_citrique-amidon
- C20 acide_citrique-glucose
- C21 acide_glutamique-antipyrine
- C22 amidon-lactose
- C23 amidon-mannitol
- C24 amidon-phosphate
- C25 atropine-phenacetine
- C26 cafeine-glucose
- C27 cafeine-hydroxyzine
- C28 cafeine-lactose
- C29 cafeine-lidocaine
- C30 cafeine-mannitol
- C31 creatine-lidocaine
- C32 diltiazem-inositol
- C33 diltiazem-lactose
- C34 diltiazem-paracetamol
- C35 diltiazem-phenacetine
- C36 glucose-inositol
- C37 glucose-lactose
- C38 glucose-lidocaine
- C39 glucose-mannitol
- C40 glucose-phenacetine
- C41 glucose-procaine
- C42 glucose-sucre
- C43 hydroxyzine-lactose
- C44 inositol-lactose
- C45 inositol-mannitol
- C46 inositol-phenacetine
- C47 inositol-procaine
- C48 lactose-lidocaine
- C49 lactose-mannitol
- C50 lactose-paracetamol

- C51 lactose-phenacetine
- C52 lactose-piracetam
- C53 lactose-procaïne
- C54 lactose-sucrose
- C55 lidocaine-mannitol
- C56 mannitol-phenacetine
- C57 mannitol-piracetam
- C58 mannitol-sucres
- C59 procaïne-sucrose
- C60 1,4_dmt-lactose-levamisole
- C61 1,4_dmt-lactose-lidocaine
- C62 1,4_dmt-lactose-mannitol
- C63 1,4_dmt-lactose-phenacetine
- C64 1,4_dmt-mannitol-phenacetine
- C65 acetate_de_propyl-acetate_d'ethyle-toluene
- C66 acide_citrique-glucose-lactose
- C67 acide_citrique-glucose-levamisole
- C68 acide_citrique-glucose-mannitol
- C69 acide_citrique-glucose-sucrose
- C70 amidon-cafeine-lidocaine
- C71 aspirine-lactose-paracetamol
- C72 benzocaine-lactose-mannitol
- C73 cafeine-glucose-lactose
- C74 cafeine-inositol-lactose
- C75 cafeine-inositol-mannitol
- C76 cafeine-lactose-mannitol
- C77 cafeine-lactose-paracetamol
- C78 cafeine-lactose-phenacetine
- C79 cafeine-lactose-sucrose
- C80 cafeine-lidocaine-mannitol
- C81 cafeine-mannitol-phenacetine
- C82 diltiazem-glucose-lactose
- C83 diltiazem-inositol-lactose
- C84 diltiazem-lactose-mannitol
- C85 diltiazem-lactose-phenacetine

- C86 dimethylterephthalate-lactose-mannitol
- C87 glucose-inositol-lactose
- C88 glucose-inositol-phenacetine
- C89 glucose-lactose-lidocaine
- C90 glucose-lactose-mannitol
- C91 glucose-lactose-phenacetine
- C92 glucose-lactose-piracetam
- C93 glucose-lactose-sucrose
- C94 glucose-levamisole-phenacetine
- C95 glucose-maltose-mannitol
- C96 glucose-mannitol-phenacetine
- C97 glucose-mannitol-phosphate
- C98 glucose-mannitol-sucrose
- C99 glucose-phenacetine-sucrose
- C100 hydroxyzine-lactose-mannitol
- C101 hydroxyzine-lidocaine-phenacetine
- C102 ibuprofene-lactose-paracetamol
- C103 inositol-lactose-lidocaine
- C104 inositol-lactose-mannitol
- C105 inositol-lactose-paracetamol
- C106 inositol-lactose-phenacetine
- C107 inositol-lactose-procaine
- C108 inositol-lactose-sucrose
- C109 inositol-mannitol-phenacetine
- C110 inositol-paracetamol-phenacetine
- C111 inositol-phenacetine-piracetam
- C112 inositol-phenacetine-sucrose
- C113 inositol-procaine-sucrose
- C114 lactose-levamisole-sucrose
- C115 lactose-lidocaine-mannitol
- C116 lactose-lidocaine-paracetamol
- C117 lactose-lidocaine-phenacetine
- C118 lactose-lidocaine-sucrose
- C119 lactose-mannitol-phenacetine
- C120 lactose-mannitol-piracetam

- C121 lactose-mannitol-procaine
- C122 lactose-mannitol-sucres
- C123 lactose-mannitol-sucrose
- C124 lactose-phenacetine-procaine
- C125 lactose-phenacetine-sucrose
- C126 lidocaine-mannitol-phenacetine
- C127 lidocaine-phenacetine-procaine
- C128 lidocaine-procaine-sucrose
- C129 1,4_dmt-lactose-mannitol-phenacetine
- C130 acide_citrique-glucose-inositol-lactose
- C131 acide_citrique-glucose-inositol-mannitol
- C132 acide_citrique-glucose-lactose-lidocaine
- C133 acide_citrique-glucose-lactose-phenacetine
- C134 acide_citrique-glucose-mannitol-phenacetine
- C135 acide_salicylique-hydroxyzine-lidocaine-phenacetine
- C136 benzocaine-cafeine-lactose-mannitol
- C137 benzocaine-glucose-lactose-phenacetine
- C138 benzocaine-inositol-lactose-mannitol
- C139 benzocaine-lactose-mannitol-phenacetine
- C140 cafeine-dimethylterephthalate-lactose-phenacetine
- C141 cafeine-glucose-lactose-mannitol
- C142 cafeine-inositol-lactose-mannitol
- C143 cafeine-inositol-mannitol-paracetamol
- C144 cafeine-inositol-mannitol-phenacetine
- C145 cafeine-lactose-mannitol-phenacetine
- C146 cafeine-lactose-mannitol-sucrose
- C147 cafeine-lactose-phenacetine-sucrose
- C148 cafeine-lidocaine-mannitol-sucrose
- C149 diltiazem-glucose-lactose-phenacetine
- C150 diltiazem-inositol-lactose-phenacetine
- C151 diltiazem-lactose-mannitol-phenacetine
- C152 fentanyl-glucose-lactose-mannitol
- C153 fentanyl-glucose-lactose-sucrose
- C154 glucose-inositol-lactose-lidocaine
- C155 glucose-inositol-lactose-mannitol

-
- C156 glucose-inositol-lactose-phenacetine
C157 glucose-inositol-lactose-sucrose
C158 glucose-inositol-phenacetine-procaine
C159 glucose-lactose-levamisole-phenacetine
C160 glucose-lactose-lidocaine-mannitol
C161 glucose-lactose-lidocaine-phenacetine
C162 glucose-lactose-lidocaine-sucrose
C163 glucose-lactose-mannitol-paracetamol
C164 glucose-lactose-mannitol-phenacetine
C165 glucose-lactose-mannitol-procaine
C166 glucose-lactose-mannitol-sucrose
C167 glucose-lactose-phenacetine-sucrose
C168 glucose-lidocaine-mannitol-phenacetine
C169 glucose-mannitol-phosphate-sucrose
C170 hydroxyzine-lactose-lidocaine-mannitol
C171 hydroxyzine-lactose-phenacetine-sucrose
C172 inositol-lactose-levamisole-phenacetine
C173 inositol-lactose-lidocaine-mannitol
C174 inositol-lactose-lidocaine-phenacetine
C175 inositol-lactose-lidocaine-sucrose
C176 inositol-lactose-mannitol-phenacetine
C177 inositol-lactose-paracetamol-phenacetine
C178 inositol-lactose-phenacetine-piracetam
C179 inositol-lactose-phenacetine-sucrose
C180 inositol-lidocaine-mannitol-phenacetine
C181 lactose-lidocaine-mannitol-phenacetine
C182 lactose-lidocaine-phenacetine-sucrose
C183 lactose-mannitol-phenacetine-piracetam
C184 1,4_dmt-glucose-inositol-lactose-mannitol
C185 1,4_dmt-glucose-lactose-mannitol-phenacetine
C186 1,4_dmt-glucose-lactose-mannitol-sucrose
C187 1,4_dmt-inositol-lactose-mannitol-phenacetine
C188 acetylsalicylate_de_lysinе-inositol-lactose-lidocaine-phenacetine
C189 acetylsalicylate-glucose-lactose-mannitol-phenacetine
C190 acetylsalicylate-glycine-lactose-mannitol-phenacetine

- C191 acide_citrique-cafeine-glucose-mannitol-nicotinamide
C192 acide_citrique-diltiazem-glucose-lactose-mannitol
C193 acide_citrique-glucose-inositol-lactose-mannitol
C194 acide_citrique-glucose-inositol-lactose-sucrose
C195 acides_gras-inositol-lactose-mannitol-sucrose
C196 aspirine-ibuprofene-inositol-lactose-mannitol
C197 benzocaine-glucose-inositol-mannitol-phenacetine
C198 cafeine-glucose-inositol-lactose-mannitol
C199 cafeine-glucose-lactose-lidocaine-mannitol
C200 cafeine-glucose-lactose-lidocaine-sucrose
C201 cafeine-glucose-lactose-mannitol-phenacetine
C202 cafeine-hydroxyzine-lidocaine-mannitol-phenacetine
C203 cafeine-inositol-lactose-lidocaine-mannitol
C204 cafeine-inositol-lactose-mannitol-phenacetine
C205 cafeine-inositol-lactose-phenacetine-procaine
C206 diltiazem-glucose-lactose-mannitol-phenacetine
C207 diltiazem-inositol-lactose-phenacetine-sucrose
C208 fentanyl-glucose-inositol-lactose-mannitol
C209 glucose-inositol-lactose-lidocaine-mannitol
C210 glucose-inositol-lactose-lidocaine-phenacetine
C211 glucose-inositol-lactose-lidocaine-procaine
C212 glucose-inositol-lactose-lidocaine-sucrose
C213 glucose-inositol-lactose-mannitol-phenacetine
C214 glucose-inositol-lactose-mannitol-sucrose
C215 glucose-inositol-lidocaine-mannitol-phenacetine
C216 glucose-inositol-mannitol-phenacetine-sucrose
C217 glucose-lactose-lidocaine-mannitol-paracetamol
C218 glucose-lactose-lidocaine-mannitol-phenacetine
C219 glucose-lactose-lidocaine-mannitol-piracetam
C220 glucose-mannitol-phenacetine-procaine-sucrose
C221 glycol-lactose-mannitol-phenacetine-triethylene
C222 inositol-lactose-lidocaine-mannitol-phenacetine
C223 inositol-lactose-lidocaine-mannitol-sucrose
C224 inositol-lactose-lidocaine-phenacetine-sucrose
C225 inositol-lactose-mannitol-phenacetine-procaine

- C226 inositol-lactose-mannitol-phenacetine-sucrose
C227 lactose-lidocaine-mannitol-phenacetine-sucrose
C228 1,4_dmt-caffeine-glucose-inositol-lactose-procaine
C229 1,4_dmt-inositol-lactose-lactose-mannitol-phenacetine
C230 acide_citrique-caffeine-glucose-lactose-mannitol-nicotinamide
C231 acide_citrique-glucose-inositol-lactose-maltose-phenacetine
C232 caffeine-glucose-inositol-lactose-lidocaine-phenacetine
C233 caffeine-glucose-lactose-mannitol-nicotine-phenacetine
C234 caffeine-glucose-lactose-phenacetine-procaine-sucrose
C235 caffeine-inositol-lactose-lidocaine-mannitol-sucrose
C236 glucose-inositol-lactose-lactulose-mannitol-piracetam
C237 glucose-inositol-lactose-lidocaine-mannitol-paracetamol
C238 glucose-inositol-lactose-lidocaine-mannitol-phenacetine
C239 glucose-inositol-lactose-lidocaine-mannitol-procaine
C240 glucose-inositol-lactose-lidocaine-mannitol-sucrose
C241 glucose-inositol-lactose-lidocaine-phenacetine-sucrose
C242 glucose-inositol-lactose-mannitol-phenacetine-procaine
C243 glucose-lactose-lidocaine-mannitol-phenacetine-procaine
C244 glucose-lactose-lidocaine-mannitol-phenacetine-sucrose
C245 glucose-lactose-lidocaine-phenacetine-piracetam-sucrose
C246 glucose-lactose-lidocaine-phenacetine-procaine-sucrose
C247 glucose-lidocaine-mannitol-phenacetine-procaine-sucrose
C248 inositol-lactose-lidocaine-mannitol-phenacetine-procaine
C249 inositol-lactose-lidocaine-mannitol-phenacetine-sucrose
C250 lactose-lidocaine-mannitol-paracetamol-phenacetine-sucrose
C251 glucose-inositol-lactose-lidocaine-mannitol-phenacetine-procaine
C252 glucose-inositol-lactose-mannitol-phenacetine-procaine-sucrose
C253 glucose-inositol-lactose-lidocaine-mannitol-phenacetine-procaine-sucrose

Annexe B

Graphes du chapitre 6.2

B.1 Théorie des ensembles

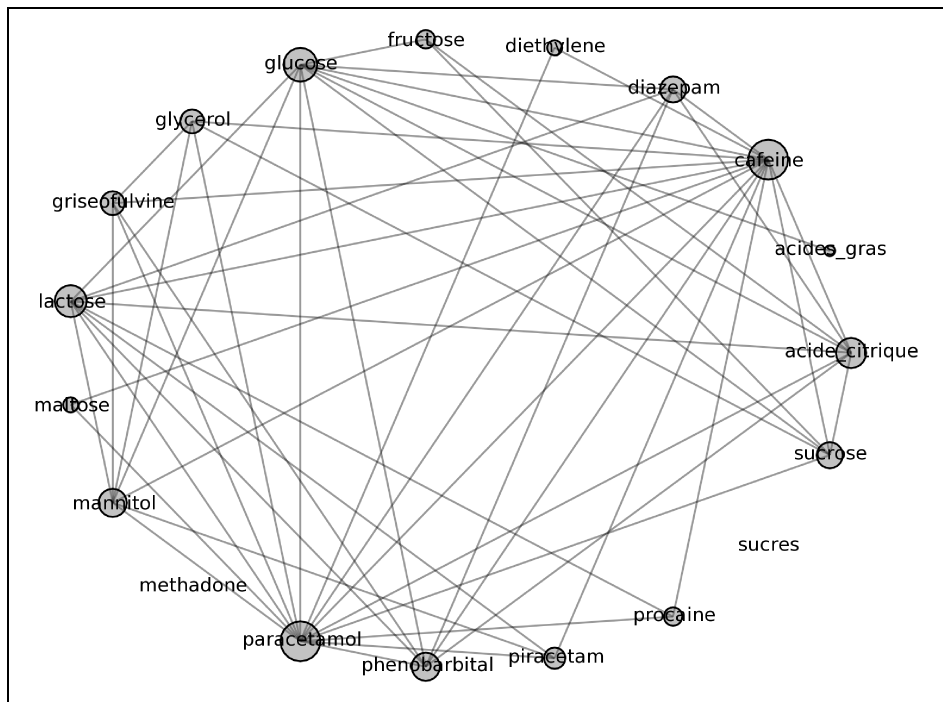


FIGURE B.1 – Graphe du canton de Genève pour l'héroïne

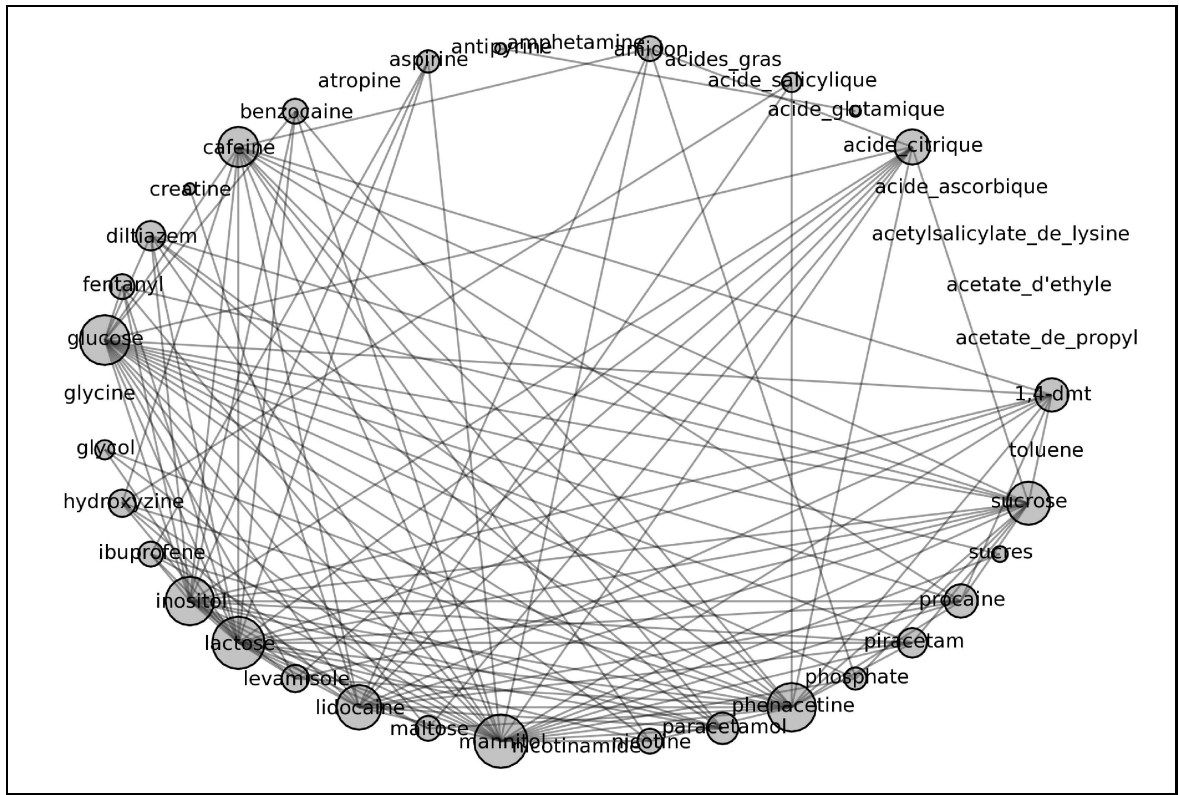


FIGURE B.2 – Graphe du canton de Genève pour la cocaïne

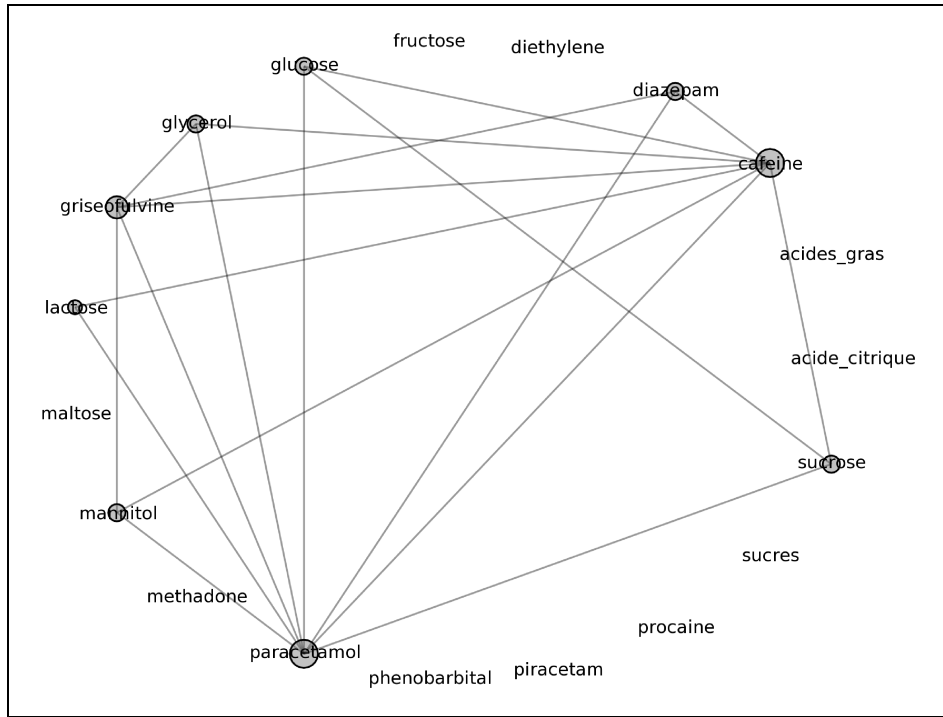


FIGURE B.3 – Graphe du canton de Vaud pour l'héroïne

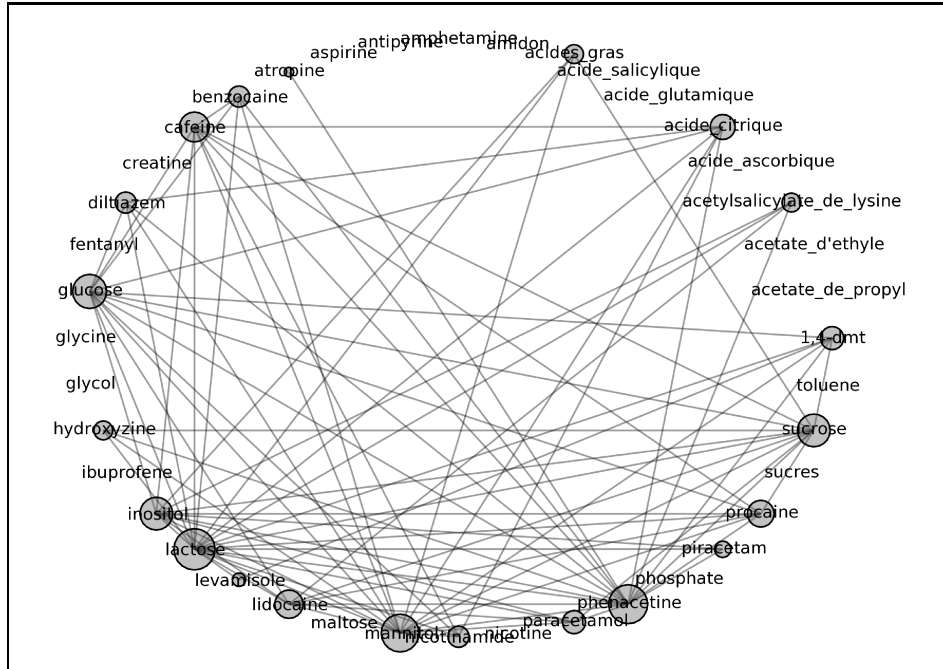


FIGURE B.4 – Graphe du canton de Vaud pour la cocaïne

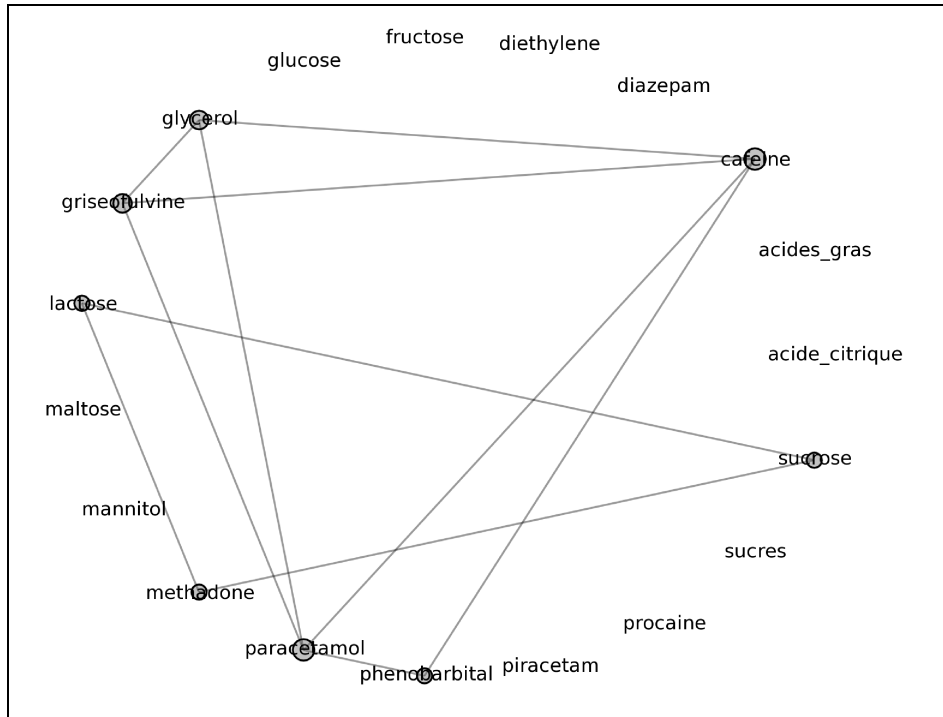


FIGURE B.5 – Graphe du canton du Tessin pour l'héroïne

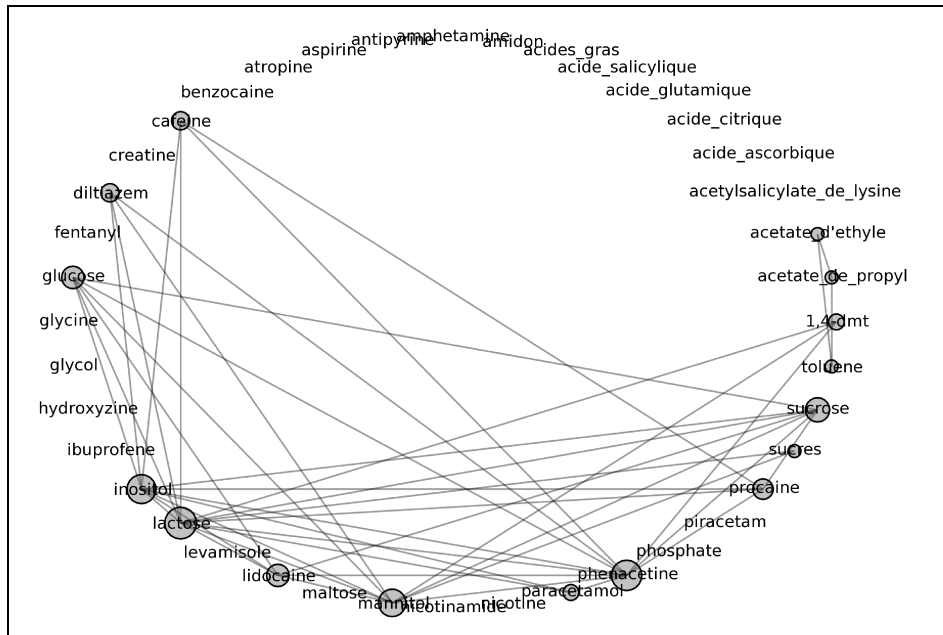


FIGURE B.6 – Graphe du canton du Tessin pour la cocaïne

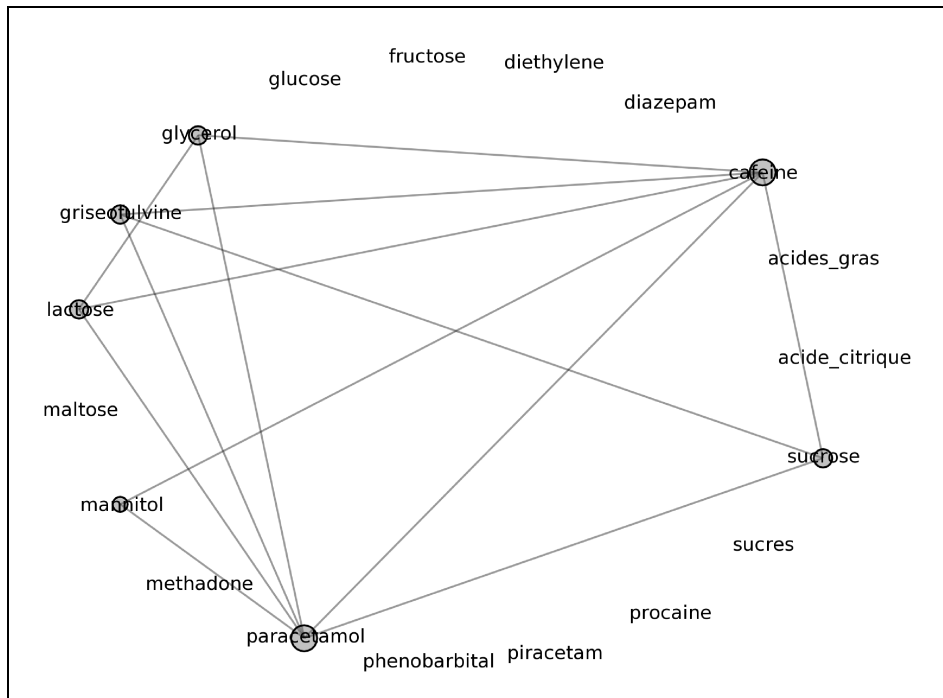


FIGURE B.7 – Graphe du canton de Neuchâtel pour l'héroïne

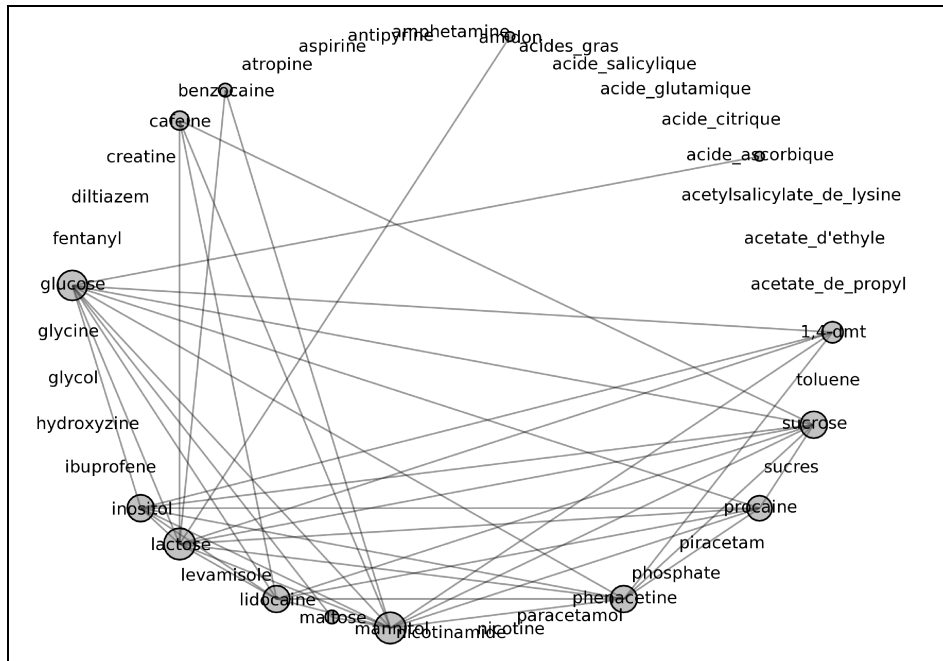


FIGURE B.8 – Graphe du canton de Neuchâtel pour la cocaïne

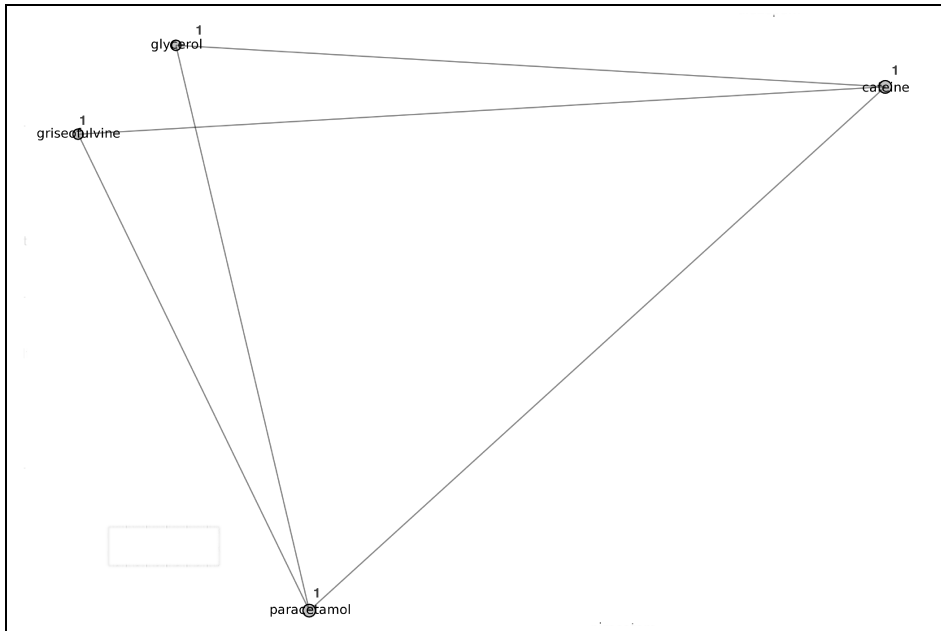


FIGURE B.9 – Graphe $G_{ProdTiNeH} = G_{Ti} \cap G_{Ne}$ pour l'héroïne

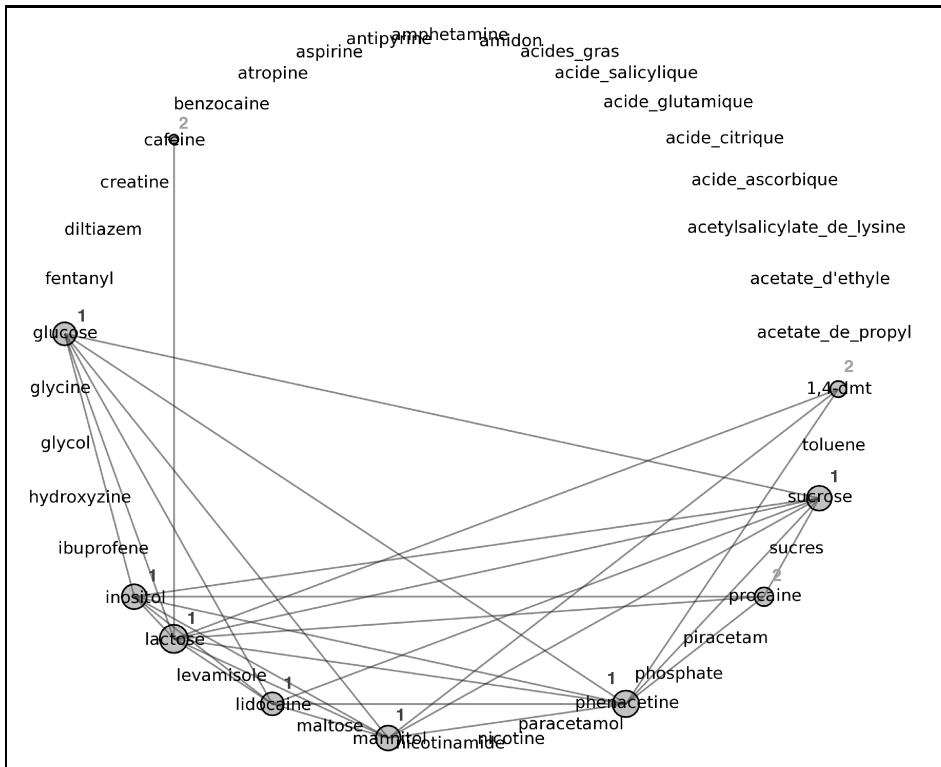


FIGURE B.10 – Graphe $G_{ProdTiNeH} = G_{Ti} \cap G_{Ne}$ pour la cocaïne

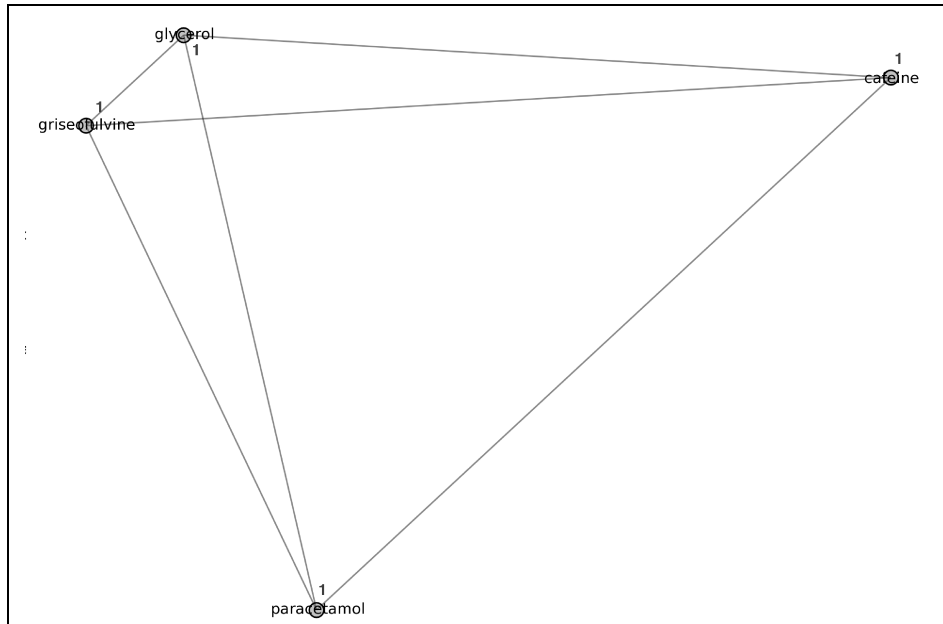


FIGURE B.11 – Graphe $G_{ProdTiVdH} = G_{Ti} \cap G_{Vd}$ pour l'héroïne

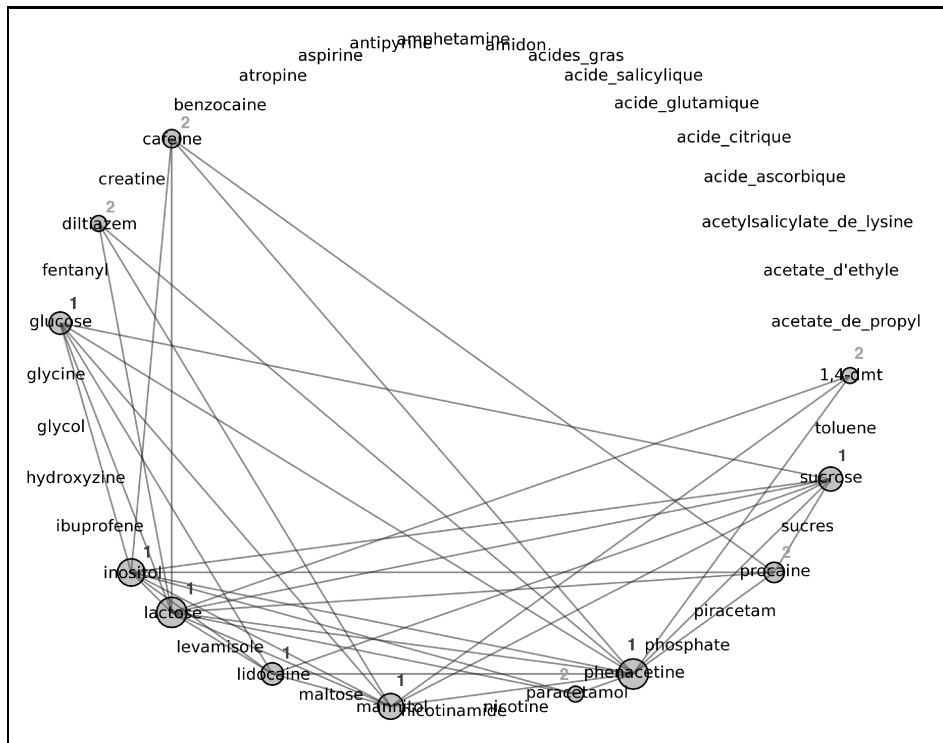


FIGURE B.12 – Graphe $G_{ProdTiVdH} = G_{Ti} \cap G_{Vd}$ pour la cocaïne

Annexe C

Graphes du chapitre 6.3

C.1 Patterns liés à l'héroïne

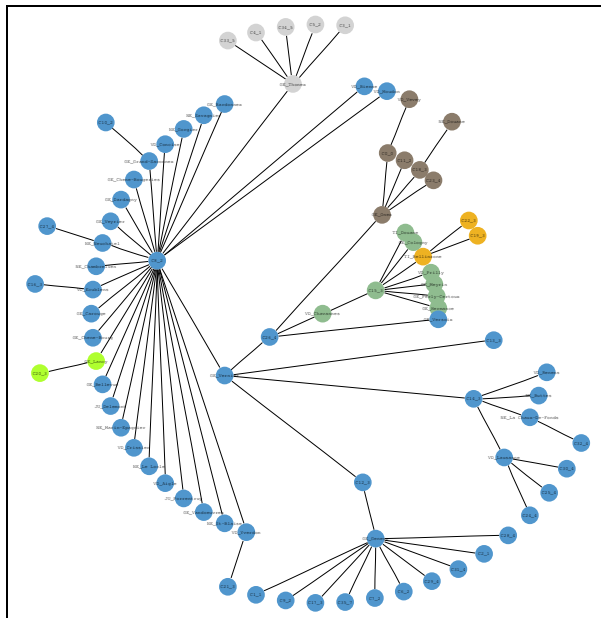


FIGURE C.1 – Détection de communautés par “fast greedy” au **minST** de l’héroïne
Combinaisons - Localités

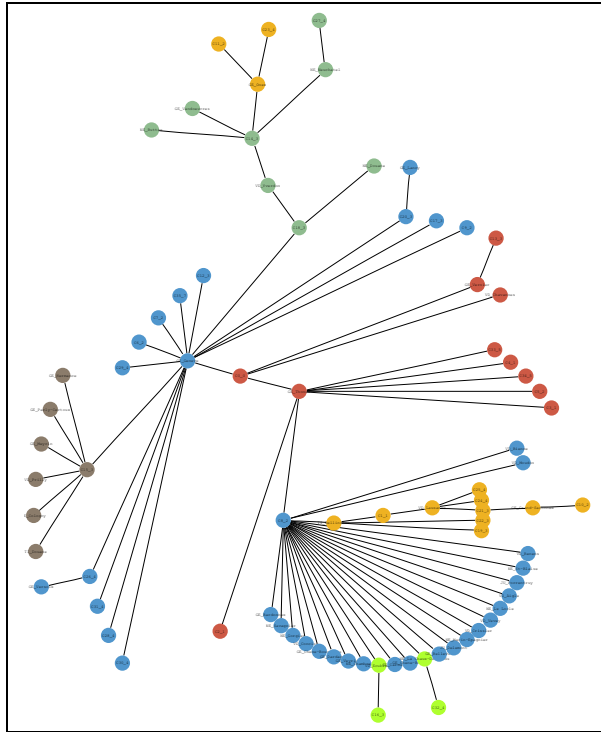


FIGURE C.2 – Détection de communautés par “fast greedy” au **maxST** de l'héroïne
Combinaisons - Localités

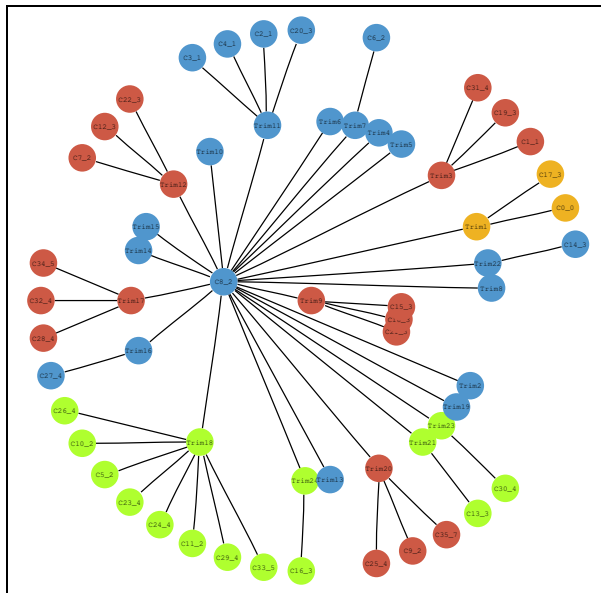


FIGURE C.3 – Détection de communautés par “fast greedy” au **minST** de l'héroïne
Combinaisons - Trimestres

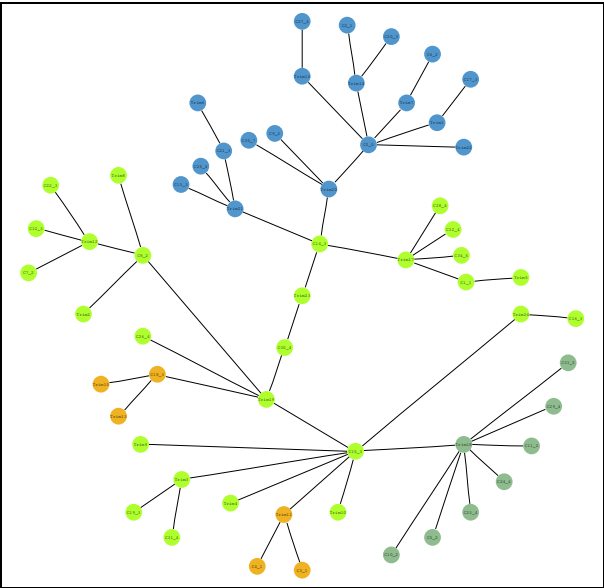


FIGURE C.4 – Détection de communautés par “fast greedy” au **maxST** de l’héroïne
Combinaisons - Trimestres

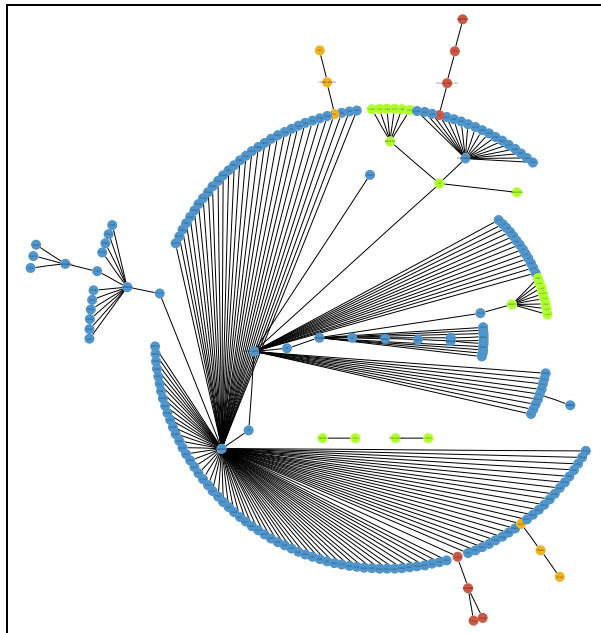


FIGURE C.6 – Détection de communautés par “fast greedy” au **minST** de la cocaïne
Combinaisons - Localités pour Ge-Ti

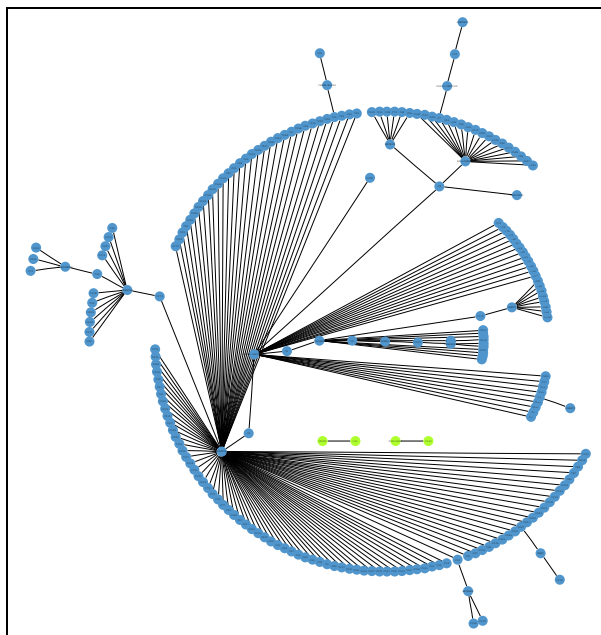


FIGURE C.7 – Détection de communautés par “walktrap” au **minST** de la cocaïne
Combinaisons - Localités pour Ge-Ti

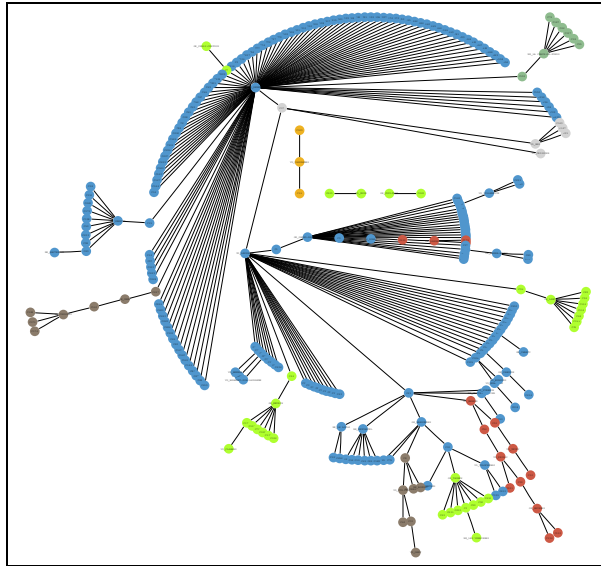


FIGURE C.8 – Détection de communautés par “fast greedy” au minST de la cocaïne
Combinaisons - Localités pour Ge-Vd-Ne

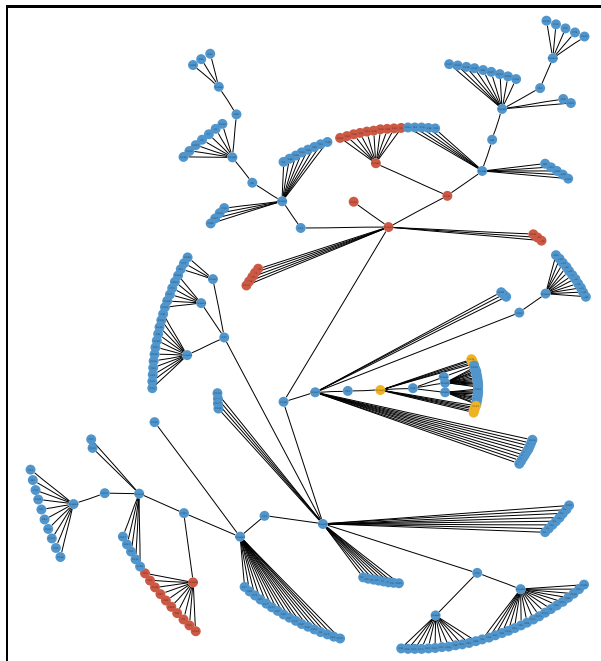


FIGURE C.9 – Détection de communautés par “fast greedy” au minST de la cocaïne
Combinaisons - Trimestres

Annexe D

Liste des abbréviations

ACP ou PCA	Analyse en Composantes Principales ou Principal Component Analysis
ARS	Analyse de réseaux sociaux
FNSRS	Fond National Suisse de la Recherche Scientifique
GUI	Graphical User Interface
IGAR	Institut de Géomatique et d'Analyse du Risque
IPS	Institut de Police Scientifique
minST	minimum Spanning Tree
maxST	maximum Spanning Tree
SIMCA	Soft Independent Modelling of Classe Analogies
SVM	Support Vector Machine