Visualizing Mobile Phone Communication Data in Criminal Investigations: the Case of Media Multiplexity

Martina Reif, Bruno Pinaud, Thomas Souvignet, Guy Melançon, Quentin Rossy

Noname manuscript No.
(will be inserted by the editor)

Visualizing Mobile Phone Communication Data in Criminal Investigations: the Case of Media Multiplexity

Martina Reif · Bruno Pinaud · Thomas Souvignet · Guy Melançon · Quentin Rossy

Abstract Given the multiple channels offered by social media apps, the analysis of communication data in criminal investigations has become a challenging task. A multivariate graph, gathering information of different types, can be inferred from communication events (calls, group discussions, etc.) and contact information (e.g. phone directory or app "friends"). Astute transformations are however required to properly associate virtual entities used by a single physical person. This paper proposes a visual analytics approach to support this task relying on graph transformations and proper visual encodings.

Keywords Mobile phone data · Communication data · Multivariate graphs · Network Visual Analytics · Crime Analysis

1 Introduction

Mobile phone communication data is frequently used in criminal investigations to reconstruct activities of interest, identify key actors and significant locations, or study the relationships between relevant entities of a case [1,2]. Visualization is an essential step in the analysis process as it facilitates information interpretation. Traditional communication data admitted a straightforward model since a phone could only be used for voice conversation, SMS or MMS, and this typically using a single phone number. As a consequence, communications could be modeled as a one-mode graph where two phone numbers (nodes) were linked whenever a communication took place between them (edges, Fig. 1, the graph on the left).

However, nowadays, users have access to a plethora of applications, such as WhatsApp, Telegram, Instagram or Facebook Messenger. The simultaneous

M. Reif, T. Souvignet, and Q. Rossy
ESC, University of Lausanne E-mail: firstname.lastname@unil.ch

B. Pinaud and G. Melançon
Univ. Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR 5800
E-mail: firstname.lastname@u-bordeaux.fr

use of such applications, a phenomenon called "Media Multiplexity" [3], turns the analysis and visualization of communication data into a subtle and complex task for several reasons. One of them is the multiplication of identifiers: as each application uses its own referencing system, traditional graph representations of combined communication data generate multiple one-mode graphs, rather than a single connected graph (Fig. 1). However, in order to reconstruct a user's activity based on phone data, a main objective in criminal investigations, all traces generated by any means of communication should be combined and visualized in a coherent manner. A simple way to link the identifiers between them is to use contact information.

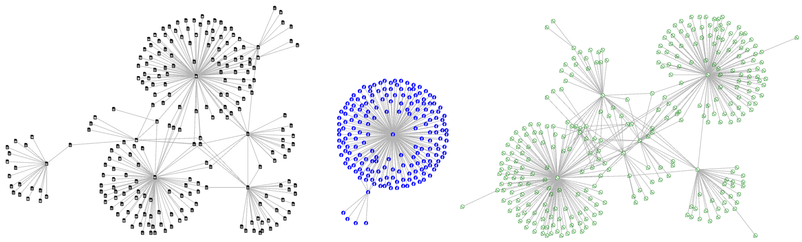

Fig. 1 "Media Multiplexity": visualizing communication data from different applications (from left to right: phone, Facebook and WhatsApp) leads to multiple one-mode networks.

This paper brings forward contributions to help solve this challenge. Communication data as well as contact information is gathered into a unique multivariate graph. Simultaneous links between nodes of different types allow to compute a quotient graph grouping nodes into "metanodes". This multilevel model thus reduces the size of the considered network making it more legible, and favors the emergence of hypotheses during the investigation. Multilevel representations of graphs are displayed as nested graph layouts. The capability to examine established associations by looking into metanodes assists analysts in assessing the relevance of the graph model.

Our approach was developed in a multidisciplinary context using data collected by a Swiss Law Enforcement Agency, anonymized and made freely available to the community [4].

2 Related Work

Visualization and analysis of phone data in criminal investigations has been studied on multiple occasions [5,6]. However, most studies focus on call detail records and data from wiretaps. As far as we know, there is no literature available on the process of visualizing multi-source communications data obtained from phone extractions. This could be due to limited access to actual case data for researchers. Additionally, the concept of media multiplexity in this process has yet to be explored.

Regarding graph representation and visualization, our work builds on past work focusing on multilevel network visualization [7,8] making use of features implemented in the Tulip graph visualization framework [9].

3 Use Case

Our starting point is data extracted from several drug trafficking offenders' seized phones [4]. Communication data such as call and messaging logs allow linking communication application identifiers (id_1 for app_1, id_2 for app_2, ...) used on the seized phones with correspondents' application identifiers. Communications are represented as directed and weighed edges between two nodes, where each node represents an application identifier ($id_i, id_j, \ldots$) and the edge weight represents the number of communications. We call this type of data "*communication data*".

A phone directory gathers contacts of the phone owner. For each contact, it lists the associated application identifiers of this contact ($id_u, id_v, \ldots$). We link them based on their co-occurrence under the same contact entry in the phone directory[1]. We call this type of data "*association data*": data that allow the establishment of relationships between nodes, other than communications.

Integrating and visualizing the two resulting data sets (see Fig. 2) considerably reduces the risk of linkage blindness, i.e., the inability to detect links between relevant entities [10]. A recurring question in criminal investigations is the identification of mutual correspondents of two or more known suspects. When dealing with media multiplexity, having association data is required. For example, in cases where suspect A communicates with person C via WhatsApp, and suspect B with the same person C via Facebook, it is crucial to establish a link between the two identifiers (id_1 of C as a WhatsApp user and id_2 of C as a Facebook user).

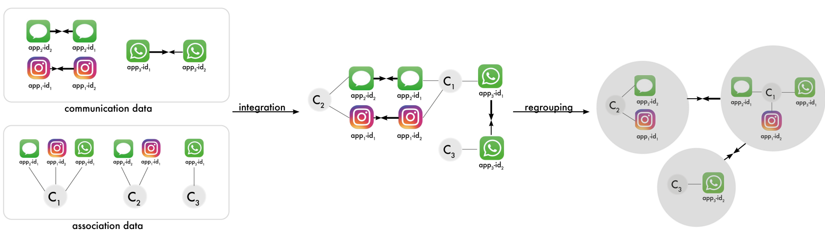

Fig. 2 Communication and association data are integrated into a combined view, then clustered with metanodes to improves its readability.

The resulting graph representation becomes multivariate. Nodes represent phone numbers or application identifiers. Edges can represent communica-

[1] In reality, the situations to be taken into account are numerous. We limit ourselves to this description for the sake of simplicity and because of space constraints.

tions (directed) or associations (not directed). The integration of associations allow to visually link different communications attributed to a same user and thus better understand the activity of a user in its entirety. However, such multivariate graph quickly become dense, cluttered, and difficult to read and analyze.

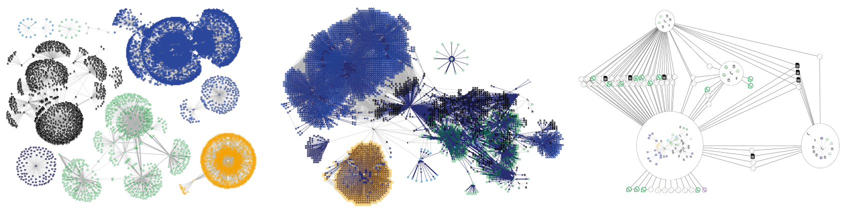

Fig. 3 From left to right: (1) graphs resulting from visualization of communication data, (2) graph integrating communication and association data, (3) simplified graph with metanodes

To improve graph legibility, we propose clustering associated entities, without merging them, using metanodes from the Tulip visual analytics platform [9]. Nodes being linked via association data thus become one single metanode, and communications carried out by identifiers used by the same person are regrouped on the metanode level. This reduces the number of visible nodes on the graph, without losing the underlying information as illustrated in Fig. 3. This has two major advantages for crime analysts. On the one hand, metanodes give analysts the possibility to verify established associations, and modify them if needed. Most criminal analysis visualization tools require users to work with static data structures, without much possibility to dynamically adapt the latter. In order to regroup identifiers used by a single person, a user is therefore compelled to merge the entities in question. On the contrary, Tulip allows for temporary regrouping of entities, without actually merging them. Analysts can thus properly evaluate established associations between identifiers and easily adapt the structure if necessary. On the other hand, original data are kept untouched. A chain of custody is a requirement for anyone working with any kind of trace in law enforcement context. Tulip is not merging entities when creating metanodes. Creating metanodes implies the creation of a graph hierarchy allowing to easily find how each node of the quotient graph was created.

4 Conclusion

An approach to visualize communication and association data for criminal investigations more effectively is promoted through the use of graph transformations and metanodes. Integrating communication with association data, such as contact information extracted from mobile phones, allow for data interpretation on a person-level rather than an application-level. Grouping associated entities improves legibility, helps rendering the visualization more efficient and supports the establishment of hypotheses. The use of metanodes allows analysts to verify established associations and, if necessary, adjust them in light of new clues.

References

1. O. Ribaux, O. Delémont, C. Roux, F. Crispino, *La science forensique au service de l'action de sécurité* (2019), pp. 114–125
2. S. Keating, Q. Rossy, P. Esseiva, Criminologie **53**(2), 171 (2020). DOI 10.7202/1074192ar
3. K.V. Cleemput, Bulletin of Science, Technology & Society **30**(2), 75 (2010). DOI 10.1177/0270467610363143
4. M. Reif, B. Pinaud, Q. Rossy, G. Melançon. Multivariate phone communication network [data set] (2023). DOI 10.5281/zenodo.7648495
5. E. Ferrara, P. De Meo, S. Catanese, G. Fiumara, Expert Systems with Applications **41**(13), 5733 (2014)
6. D. McAndrew, in *The Social Psychology of Crime*, ed. by L. Alison, D. Canter (Routledge, London, 2000), p. 44
7. D. Archambault, T. Munzner, D. Auber, IEEE Transactions on Visualization and Computer Graphics **13**(2), 305 (2007)
8. C. Rozenblat, G. Melançon, *Methods for Multilevel Analysis and Visualisation of Geographical Networks*, *Methodos Series*, vol. 11 (2013)
9. D. Auber, D. Archambault, R. Bourqui, M. Delest, J. Dubois, A. Lambert, P. Mary, M. Mathiaut, G. Melançon, B. Pinaud, B. Renoust, J. Vallet, in *Encyclopedia of Social Network Analysis and Mining*, ed. by R. Alhajj, J. Rokne (Springer, New York, NY, 2017), pp. 1–28. DOI 10.1007/978-1-4614-7163-9_315-1
10. S.A. Egger, Journal of Police Science & Administration (1984)