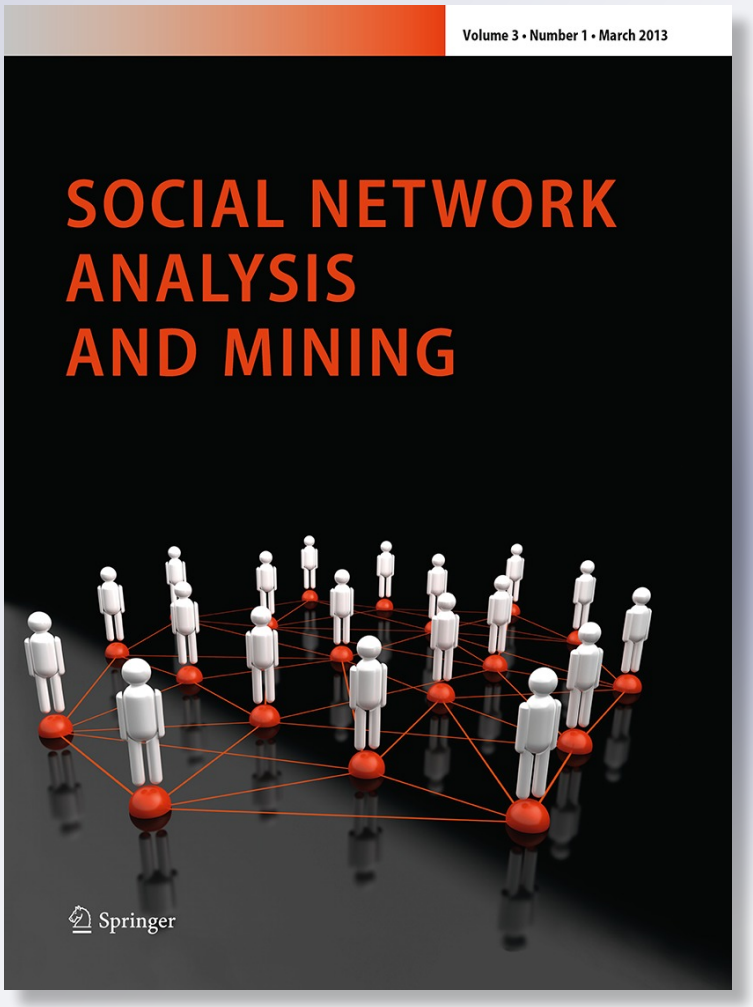
Forensic analysis of phone call networks

Salvatore Catanese, Emilio Ferrara & Giacomo Fiumara

Social Network Analysis and Mining

ISSN 1869-5450
Volume 3
Number 1

Soc. Netw. Anal. Min. (2013) 3:15-33
DOI 10.1007/s13278-012-0060-1



Springer

Soc. Netw. Anal. Min. (2013) 3:15–33
DOI 10.1007/s13278-012-0060-1

ORIGINAL ARTICLE

Forensic analysis of phone call networks

Salvatore Catanese · Emilio Ferrara · Giacomo Fiumara

Received: 13 November 2011 / Revised: 20 January 2012 / Accepted: 25 February 2012 / Published online: 11 March 2012

Abstract In the context of preventing and fighting crime, the analysis of mobile phone traffic, among actors of a criminal network, is helpful in order to reconstruct illegal activities on the basis of the relationships connecting those specific individuals. Thus, forensic analysts and investigators require new advanced tools and techniques which allow them to manage these data in a meaningful and efficient way. In this paper we present *LogAnalysis*, a tool we developed to provide visual data representation and filtering, statistical analysis features and the possibility of a temporal analysis of mobile phone activities. Its adoption may help in unveiling the structure of a criminal network and the roles and dynamics of communications among its components. Using *LogAnalysis*, forensic investigators could deeply understand hierarchies within criminal organizations, for e.g., discovering central members who provide connections among different sub-groups, etc. Moreover, by analyzing the temporal evolution of the contacts among individuals, or by focusing on specific time windows they could acquire additional insights on the data they are analyzing. Finally, we put into evidence how the adoption of *LogAnalysis* may be crucial to solve real cases, providing as example a number of case studies inspired by real forensic investigations led by one of the authors.

Keywords Social networks analysis · Forensic analysis · Phone call networks · Criminal networks

1 Introduction

The increasing usage of mobile phones in the everyday-life reflects also in their illicit adoption. For e.g., mobile communication devices are exploited by criminal organizations in order to coordinate illegal activities, to communicate decisions, etc. In order to prevent and fight crime, mobile communication service providers (according to the regulatory legislation of the state in which they operate) have to store for a given period all the data related to the phone traffic, in the shape of log files. These logs contain information about phone calls, attempted calls, short message service (SMS), multimedia messaging service (MMS), general packet radio service (GPRS) and Internet sessions. Additional information could be inferred from traffic produced by cell global identities (CGI)[1] inside their areas.

The analysis of reports supplied by mobile phone service providers makes it possible to reconstruct the network of relationships among individuals, such as in the context of criminal organizations. It is possible, in other terms, to unveil the existence of criminal networks, sometimes called rings, identifying actors within the network together with their roles. These techniques of forensic investigations are well known, and are rooted in the social network analysis (SNA). The structure of criminal networks could

S. Catanese · G. Fiumara
Department of Physics, Informatics Section,
University of Messina, Via Ferdinando Stagno
D'Alcontres, Salita Sperone, n. 31, Messina, Italy
e-mail: salvocatanese@gmail.com

G. Fiumara
e-mail: giacomo.fiumara@unime.it

E. Ferrara (✉)
Department of Mathematics, University of Messina,
Via Ferdinando Stagno D'Alcontres, Salita Sperone,
n. 31, Messina, Italy
e-mail: emilio.ferrara@unime.it

[1] CGI is a standard identifier for mobile phones cells which provides geographical positioning of mobile phones.

be efficiently formalized by means of graphs, whose nodes represent actors of the criminal organizations (or, in our case, their mobile phones), and edges represent the connections among them (i.e., their phone communications). The graphical representation of data extracted from log files is a simple task, while its interpretation may result hard, when large volumes of data are involved. In fact, it could become difficult to find anomalous values and models while browsing a large quantity of data. Moreover, visual representations of a high number of individuals and connections easily become unreadable because of nodes and edges overlapping with each other. A powerful support comes from SNA, which provides methods to evaluate the importance of particular individuals within a network and relationships among them. For example, SNA provides statistical algorithms that find those individuals/nodes in key positions and those acting as *cohesive elements*.

In this work we present a novel tool we developed, named *LogAnalysis*, for forensic visual statistical analysis of mobile phone traffic logs. *LogAnalysis* graphically represents the relationships among mobile phone users with a node-link layout. It helps to explore the structure of a graph, measuring connectivity among users and giving support to visual search and automatic identification of organizations and groups within the network. For this purpose, *LogAnalysis* integrates the graphical representation of networks with metrics and measures typical of SNA, in order to help detectives or forensic analysts to understand the structure of criminal associations while highlighting key members inside the criminal ring, and/or those members working as link among different associations, and so on. Several statistical measures have been implemented and made available to the investigators, with a seamless integration with the visual part. An additional feature is the possibility of analyzing the temporal evolution of the connections among actors of the network, for e.g., focusing on particular time windows in order to obtain additional insights about the dynamics of communications before/during/after particular criminal events. The main features of *LogAnalysis* are described together with a number of case studies, inspired to a real criminal investigation brought by one of the authors, successfully solved also by exploiting features provided by *LogAnalysis*.

2 Related work

Law enforcement and intelligence agencies frequently face the problem of extracting information from large amounts of raw data coming from heterogeneous sources, among which are phone call printouts. In the recent years, a growing number of commercial software has been developed that employ analytical techniques of visualization to help investigations. In the following we briefly describe, to the best of our knowledge, the most successful among them.

Analysts Notebook from i2 Inc.[2] provides a semantic graph visualization to assist analysts with investigations. Nodes in the graph are entities of semantic data types, such as persons, events, organizations, bank accounts, etc. While the system can import text files and do automatic layout, its primary application appears to be helping analysts in manually creating and refining case charts.

The COPLINK system (Chen et al. 2003) and the related suite of tools has a two-fold goal: to ease the extraction of information from police case reports and to analyze criminal networks. A conceptual space of entities and objects is built exploiting data mining techniques in order to help in finding relations between entities. It also provides a visualization support consisting of a hyperbolic tree view and a spring-embedder graph layout of relevant entities. Furthermore, COPLINK is able to optimize the management of information exploited by police forces integrating in a unique environment data regarding different cases. This is done in order to enhance the possibility of linking data from different criminal investigations to get additional insights and to compare them in an analytic fashion.

TRIST (Jonker et al. 2005) allows analysts to formulate, refine, organize, and execute queries over large document collections. Its user interface provides different perspectives on search results including clustering, trend analysis, comparisons, and difference. Information retrieved by TRIST then can be loaded into the SANDBOX system (Wright et al. 2006), an analytical sense-making environment that helps to sort, organize, and analyze large amounts of data. The system offers interactive visualization techniques including gestures for placing, moving, and grouping information, as well as templates for building visual models of information and visual assessment of evidence. Similarly to COPLINK, TRIST is optimized to query large databases and to analytically compare results.

Differently from COPLINK and TRIST, *LogAnalysis* adopt a different approach, which is not based on querying data, but it relies on full visual presentation and analysis of such information represented by means of network graphs. The strength of our tool is the adoption of several interactive layout techniques that highlight different aspects and features of the considered networks and it allows the inspection of elements (nodes and edges) that constitute the network itself.

Another remarkable tool is GeoTime (Kapler and Wright 2004), that visualizes the spatial interconnectedness of information over time overlaid onto a geographical

[2] i2—Analysts Notebook. http://www.i2inc.com/.

substrate. It uses an interactive 3D view to visualize and track events, objects, and activities both temporally and geo-spatially. One difference between GeoTime and *LogAnalysis* is that the feature regarding the spacial dependency of data is not yet allowed by our tool, and this makes GeoTime a useful addition to *LogAnalysis* for such type of investigations. On the other hand, the functionalities provided by *LogAnalysis* in terms of analysis of temporal dependencies of data improve those provided by GeoTime, as highlithed in Sects. 5.5–5.7.

As an example of the various general-purpose tools for analyzing social networks (differently from tools specifically designed to investigate telecom networks), we mention NodeXL (Smith et al. 2009), an extensible toolkit for network overview, discovery and exploration implemented as an add-on to the Microsoft Excel 2007/2010 spreadsheet. NodeXL is open source and was designed to facilitate learning the concepts and methods of SNA with visualization as a key component. It integrates metrics, statistical methods, and visualization to gain the benefit of all the three approaches. As for the usage of network metrics to assess the importance of actors in the network, NodeXL shares a paradigm similar to that we adopted in *LogAnalysis*, although it lacks of all the relevant features of our tools related to the temporal analysis of the networks.

Regarding those researches that apply SNA to relevant topics related to this work, recently von Landesberger et al. (2011) surveyed the available techniques for the visual analysis of large graphs. Graph visualization techniques are shown and various graph algorithmic aspects are discussed, which are useful for the different stages of the visual graph analysis process. In this work we received a number of challenges proposed by von Landesberger et al. (2011), trying to address for example the problem of large-scale network visualization for ad-hoc problems (in our case, to study phone telecom networks).

Also the analysis of phone call networks has been a subject of intensive study. Mellars (2004) investigated the principal ways a phone call network operates and how data are processed. Particular attention has been given to the methodology of investigation of data about the phone activity that it is possible to collect directly from the devices.

More recently, different works (Blondel et al. 2008; Onnela et al. 2007a, b; Palla et al. 2007) used mobile phone call data to examine and characterize the social interactions among cell phone users. They analyze phone traffic networks consisting of the mobile phone call records of million individuals.

In detail, in Onnela et al. (2007a, b) the authors present the statistical features of a large-scale Belgian phone call network constituted by 4.6 millions users and 7 millions links. That study highlights some features typical of large social networks (Ferrara and Fiumara 2011) that characterize also telecom networks, such as the fission in small clusters and the presence of strong and weak ties among individuals. In addition, in Palla et al. (2007) the authors discuss an exceptional feature of that network, which is the division in two large communities corresponding to two different language users (i.e., English and French speakers of the Belgian network).

The community structure of phone telecom networks has been further investigated in Blondel et al. (2008). The authors exploited an efficient community detection algorithm called *Louvain method* (Blondel et al. 2008; De Meo et al. 2011) to assess the presence of the community structure and to study its features, in a large phone network of 2.6 millions individuals.

In conclusion, during the latest years Eagle et al. (2008, 2009) investigated the possibility of inferring a friendship social network based on the data from mobile phone traffic of the same individuals. This problem attracted the attention of other recent studies (Candia et al. 2008; Sundsøy et al. 2010), particularly devoted to understand the dynamics of social connections among individuals by means of mobile phone networks.

2.1 Contribution of this work

LogAnalysis has been originally presented in a preliminary version during late 2010 (Catanese and Fiumara 2010) and has received a positive critique by the research community of *forensic analysts* and *social network analysts*.

We argue that the further developments of this tool have increased its potential and performance. In particular, the research direction that we are following with *LogAnalysis* is devoted to include the possibility of analyzing temporal information from *phone call networks*, and the tool has been specifically optimized to study *mobile phone telecom networks*, whose analysis has attracted relevant research efforts in the recent period (Saravanan et al. 2011). Additional efforts have been carried out so as to improve the possibilities provided by *LogAnalysis* to unveil and study the community structure of the networks, whose importance has been assessed during latest years in a number of works (Porter et al. 2009; Gilbert et al. 2011), by means of different community detection techniques (Coscia et al. 2011; De Meo et al. 2011; Fortunato 2010).

Our tool introduces a number of novelties with respect to similar platforms existing as to date. In detail, *LogAnalysis* primarily differs from the systems described above as we focused on the visual representation of the relationships among entities in phone calls. We adopted different state-of-the-art view layouts for promoting fast exploration and discovery of the analyzed networks.

Furthermore, our tool provides a system model which aims at improving the quality of the analysis of social relationships of the network through the integration of visualization and SNA-based statistical techniques, which is a relevant topic in the ongoing research in SNA (Scott 2011).

To this purpose, *LogAnalysis* has been assessed as an invaluable support during real investigations carried out by professional *forensic analysts*, in particular, in the context of analyzing large-scale *mobile telecom networks* exploited for criminal purposes.

One of the merits of this work, in fact, is to analyze several different real-world use cases inspired by forensic investigations carried out by one of the authors. During these investigations, *LogAnalysis* has been exploited to examine the structural features of criminal phone call networks with a systematic methodology adopting a unique tool, differently from previous cases in which a combination of different SNA-based and digital forensic tools had to be adopted to reach similar results. Some relevant information about the usage of *LogAnalysis* in the context of real-world investigations have been reported in this work. To the best of our knowledge, this is the first work to present critical information from real forensic investigations in mobile phone call networks, dealing with real data acquired from actual criminal cases. As a relevant fact, we provide with some clues that support our claim about the advantages of adopting *LogAnalysis* to unveil possible criminal connections among actors of mobile telecom networks.

3 Analysis of mobile phone traffic networks

The relationships established by means of phone calls may be explored using different techniques and approaches. Sometimes, forensic analysis relates to phone traffic made by international mobile subscriber identity (IMSI)[3] and by international mobile equipment identity (IMEI).[4] Detectives generally distinguish three main types of analysis of phone traffic logs: (1) *relational*, in order to show links (and hence acquaintance) among individual users; (2) *spatial*, helpful to show geographical displacements of a mobile phone in order to assess location of an individual before, during and after a crime has been committed; and (3) *temporal*, useful to discover, for e.g., at what time a phone call has been made or a SMS has been sent, which

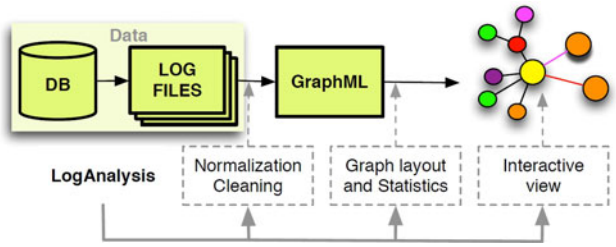


Fig. 1 Architecture of *LogAnalysis*

contacts were involved in a phone conversation or how long an Internet connection lasted. *LogAnalysis* provides some tools to investigate relational and temporal aspects of phone call networks.

The architecture of *LogAnalysis*, shown in Fig. 1, is designed by extensible levels: (1) import of data provided by informative systems of mobile phone service providers (usually, under the form of *textual log files*); (2) conversion of data to the GraphML[5] format, a structured XML format, more suitable for graphical representation and portability among several different graph drawing applications; (3) visualization and dynamic exploration of the obtained mobile phone traffic network.

An example of the usage of this tool is the research of particular elements in the network. For this purpose, it is possible to visually discover subsets and gangs (or rings), by measuring their cohesion in terms of the density of internal connections. Thus, from the overall structure of the network are extracted those elements of interest for investigations. In fact, some nodes are prominent due to their high degree of connection with others, other nodes for their strategic position of centrality in terms of connections, etc. A number of complete case studies has been analyzed in Sect. 5 describing some features provided by *LogAnalysis* that have been exploited in the context of a real-world investigation.

3.1 Implementation

Our system is implemented in Java and integrates several open-source toolkits. In particular, *Prefuse*[6] provides the underlying node-link data structures and has been used to support the dynamic exploration of networks, according to force-directed and radial models, and to identify communities. *JUNG*[7] has been used to implement some of the SNA ranking algorithms, for the computation and visualization of the shortest path(s) connecting a pair of nodes, and for the network visualizations and clustering.

[3] IMSI is a unique number associated with all GSM and UMTS network mobile phone users. It is stored in the SIM inside the phone and is sent by the phone to the network.

[4] IMEI is a unique 17 or 15 digit code used to identify an individual mobile station to a GSM or UMTS network.

[5] http://graphml.graphdrawing.org/.

[6] http://prefuse.org/.

[7] http://jung.sf.net/.

3.2 Data import

In the context of real-world investigations, mobile phone service providers, upon request by judiciary authorities, release data logs, normally in textual file format, with space or tab separation (CSV format). A typical log file contains, at least, the values shown in Table 1.

Similarly, information about the owners and dealers of SIM cards, and operations like activation, deactivation, number portability are provided by the service providers as additional material in order to ease and support the investigation activities. Log file formats produced by different companies are heterogeneous. *LogAnalysis*, first of all, parses these files and converts data into GraphML format. It is an XML-valid and well-formed format, containing all nodes and weighted edges, each weight representing the frequency of phone calls between two adjacent nodes. GraphML has been adopted both because of its extensibility and ease of import from different SNA toolkits and graph drawing utilities.

3.3 Data normalization/cleaning

Data clean-up usually means the deletion of redundant edges and nodes. This step is very important since datasets often contain redundant information, that crowds graph visualization and biases statistical measures. In these circumstances, redundant edges between the same two nodes are collapsed and a coefficient—i.e., a edge weight—is attached, which expresses the number of calls. Our tool normalizes data after reading and parsing log files whichever format they have been provided among the standard formats (i.e., *fixed width text*, *delimited*, CSV, and more) used by mobile service providers.

4 Eyes on some features of *LogAnalysis*

In this section we put into evidence some of the main features of *LogAnalysis* that have been inspired both by forensic analysis and the social network analysis. In particular, in Sect. 4.1 we point out those data exploration features provided by our tool. Subsequently, in Sect. 4.2 we discuss the role and functioning of a set of centrality measures implemented in *LogAnalysis* that can be exploited to assess the importance of actors of mobile telecom networks. Furthermore, in Sect. 4.3 the layout models adopted in our tool are described, focusing on the novelties introduced by *LogAnalysis* in respect to general purpose SNA tools.

Table 1 An example of the structure of a log file

Field	Description
IMEI	IMEI code MS
Called	Called user
Calling	Calling user
Date/time start	Date/time start calling (GMT)
Date/time end	Date/time end calling (GMT)
Type	SMS, MMS, voice, data, etc
IMSI	Calling or called SIM card
CGI	Lat. long. BTS company

4.1 Data exploration

The main goal of *LogAnalysis* is to support the forensic detectives into the exploration of data provided by mobile phone service providers about the phone traffic activity of particular individuals of interest for the forensic investigations. This support is given by means of an interactive visual representation of the phone traffic network. For this purpose, individuals are identified by means of their phones and are represented by nodes of a graph. The phone calls, instead, represent the interactions among actors and for this reason they are captured as the edges of the same graph. More formally, the structure of phone traffic is described in terms of directed graphs $G = (V, E)$, where V is the set of telephone numbers (nodes) and E is the set of calls (edges) among the nodes. The edges, directed and weighted, show the direction (incoming or outgoing) and the number of phone calls between the various pairs of adjacent nodes.

LogAnalysis is able to manage phone traffic networks up to hundreds of thousands of elements and log files up to millions of entries. However, in our experience, a meaningful interactive visual representation of these data is viable analyzing networks up to some thousands of elements. To this purpose, in Sect. 5 we describe a number of case studies inspired by real investigations whose network includes thousands of elements, and in which *LogAnalysis* played a fundamental role in the successful conclusion of the investigation.

In detail, one of the most useful features of *LogAnalysis* is that it is able to identify and visually put into evidence those actors in the network that play a crucial role in the communication dynamics. This is done by exploiting the centrality measures provided by the SNA (described in the next section). On the other hand, a visual layout only could not be sufficient to put into evidence all the required information. For e.g., different visual representation would help detectives to reach additional insights about data, the dynamics of the phone traffic network and the activities of the actors of the network. For this reason, *LogAnalysis* provides different interactive visual representations, by adopting several algorithms.

4.2 Centrality measures

LogAnalysis takes into account the concept of *centrality measure* to highlight actors that cover relevant roles inside the analyzed network. Several notions of centrality have been proposed during the latest years in the context of SNA.

There are two fundamentally different class of centrality measures in communication networks. The first class of measures evaluates the centrality of each node/edge in a network and is called point centrality measure. The second type is called graph centrality measure because it assigns a centrality value to the whole network. These techniques are particularly suited to study phone traffic and criminal networks.

In detail, in *LogAnalysis* we adopted four point centrality measures (i.e., *degree*, *betweenness*, *closeness* and *eigenvector* centrality), to inspect the importance of each node of the network.

The set of measures provided in our tool is a selection of those provided by SNA (Wasserman and Faust 1994). It could be not sufficient to solve any possible task in phone call network analysis. In fact, for particular assignments it could yet be necessary to use additional tools in support to *LogAnalysis*, and in further evolutions we plan to incorporate new centrality measures (De Meo et al. 2012; Abdallah 2011) if necessary.

For each centrality measure, the tool gives the possibility, to rank the nodes/edges of the network according to the chosen criterion. Moreover, *LogAnalysis* allows to select those nodes that are central, according to the specified ranking, highlighting them and putting into evidence their relationships, by exploiting the node-link layout techniques (discussed in the following). This approach makes it possible to focus the attention of the analysts on specific nodes of interest, putting into evidence their position and their role inside the network, with respect to the others.

In the following, we formally describe the centrality measures used in *LogAnalysis*.

They represent the centrality as an indicator of the activity of the nodes (degree centrality), of the control on other nodes (betweenness centrality), of the proximity to other nodes (closeness centrality) and of the influence of a node (eigenvector centrality).

4.2.1 Degree centrality

The degree centrality of a node is defined as the number of edges adjacent to this node. For a directed graph $G = (V, E)$ with n nodes, we can define the in-degree and out-degree centrality measures as

$$C_D(v)_{\text{in}} = \frac{d_{\text{in}}(v)}{n-1}, \quad C_D(v)_{\text{out}} = \frac{d_{\text{out}}(v)}{n-1} \tag{1}$$

where $d_{\text{in}}(v)$ is the number of incoming edges adjacent to the node v, and $d_{\text{out}}(v)$ is the number of the outgoing ones.

Since a node can at most be adjacent to $n-1$ other nodes, $n-1$ is the normalization factor introduced to make the definition independent on the size of the network and to have $0 \leq C_D(v) \leq 1$.

In and out-degree centrality indicates how much activity is going on and the most active members. A node with a high degree can be seen as a hub, an active nodes and an important communication channel.

We chose to include the degree centrality for a number of reasons. First of all, calculation is computationally even on large networks. Furthermore, in the context of phone call networks it could be interpreted as the chance of a node for catching any information traveling through the network.

Most importantly, in this type of directed networks, high values of in-degree are considered a reliable indicator of a form of popularity/importance of the given node in the network; on the contrary, high values of out-degree are interpreted as a form of gregariousness of the given actor with respect to the contacted individuals.

4.2.2 Betweenness centrality

The communication between two non-adjacent nodes might depend on the others, especially on those on the paths connecting the two nodes. These intermediate elements may wield strategic control and influence on many others.

The core issue of this centrality measure is that an actor is central if she lies along the shortest paths connecting other pairs of nodes. The betweenness centrality of a node v can be defined as

$$B_C(v) = \sum_{s \neq v \neq t} \frac{\sigma_{\text{st}}(v)}{\sigma_{\text{st}}} \tag{2}$$

where σ_{st} is the number of shortest paths from s to t and $\sigma_{\text{st}}(v)$ is the number of shortest paths from s to t that pass through a node v.

The importance of the betweenness centrality regards its capacity of identifying those nodes that vehiculate information among different groups of individuals.

In fact, since its definition due to Freeman (1977) the betweenness centrality has been recognized as a good indicator to quantify the ability of an actor of the network to control the communication between other individuals and, specifically for this reason it has been included in *LogAnalysis*.

In addition, it has been exploited by Newman (2004) to devise an algorithm to identify communities within a network. Its adoption in the phone traffic networks is crucial in order to identify those actors that allow the communication among different (possibly criminal) groups.

4.2.3 Closeness centrality

Another useful centrality measure that has been adopted in *LogAnalysis* is called *closeness centrality*. The idea is that an actor is central if she can quickly interact with all the others, not only with her first neighbors (Newman 2005). The notion of closeness is based on the concept of shortest paths (geodesic) $d(u, v)$, the minimum number of edges traversed to get from u to v. The closeness centrality of the node v is define as

$$C_C(v) = \frac{1}{\sum_{u \in V} \mathrm{d}(u, v)} \tag{3}$$

Such a measure is meaningful for connected graphs only, assuming that $d(u, v)$ may be equal to a finite value.

In the context of criminal networks, this measure highlights entities with the minimum distance from the others, allowing them to pass on and receive communications more quickly than anyone else in the organization. For this reason, the adoption of the closeness centrality is crucial in order to put into evidence inside the network, those individuals that are closer to others (in terms of phone communications).

In addition, high values of closeness centrality in such type of communication networks are usually regarded as an indicator of the ability of the given actor to quickly spread information to all other actors of the network. For such a reason, the closeness centrality has been selected to be included in the set of centrality measures adopted by *LogAnalysis*.

4.2.4 Eigenvector centrality

Another way to assign the centrality to an actor of the network in *LogAnalysis* is based of the idea that if a node has many central neighbors, it should be central as well. This measure is called *eigenvector centrality* and establishes that the importance of a node is determined by the importance of its neighbors.

The eigenvector centrality of a given node v_i is

$$C_E(v_i) \propto \sum_{u \in N_i} A_{ij} C_E(u) \tag{4}$$

where N_i is the neighborhood of the given node v_i, and $x \propto Ax$ that implies $Ax = \lambda\, x$. The centrality corresponds to the top eigenvector of adjacency matrix A.

In the context of telecom networks, eigenvector centrality is usually regarded as the measure of influence of a given node. High values of eigenvector centrality are achieved by actors who are connected with high-scoring neighbors, which in turn, inherited such an influence from their high-scoring neighbors and so on.

This measure well reflects an intuitive important feature of communication networks i.e., the influence diffusion and for such a reason we decided to include the eigenvector centrality in *LogAnalysis*.

4.2.5 Clustering coefficient (transitivity)

The clustering (or transitivity) coefficient of a graph measures the degree of interconnectedness of a network or, in other words, the tendency of two nodes that are not adjacent but share an acquaintance, to get themselves in contact. High clustering coefficients mean the presence of a high number of triangles in the network.

The local clustering coefficient C_i for a node v_i is the number of links among the nodes within its neighborhood divided by the number of links that could possibly exist among them

$$C_i = \frac{|\{e_{jk}\}|}{k_i(k_i - 1)} : v_j, v_k \in N_i, e_{jk} \in E \tag{5}$$

where the neighborhood N of a node v_i is defined as $N_i = \{v_j : e_{ij} \in E \wedge e_{ji} \in E\}$, while $k_i(k_i - 1)$ is the number of links that could exist among the nodes within the neighborhood.

It is well-known in the literature (Wasserman and Faust 1994) that communication networks show high values of clustering coefficient since they reflect the underlying social structure of contacts among friends/acquaintances. Moreover, high values of local clustering coefficient are considered a reliable indicator of of nodes whose neighbors are very well connected and among which a substantial amount of information may flow. For such a reason, *LogAnalysis* provides the possibility of computing both the global clustering coefficient for any given phone call network and the local clustering coefficient of any given node.

4.3 Layout algorithms

In this section we introduce the strategies of interactive visual representation of the phone traffic networks adopted in *LogAnalysis*. In detail, the graphical representation of phone relationships in *LogAnalysis* exploits features provided by two well-known toolkits, *Prefuse* and *JUNG*.

4.3.1 Force-directed model

The main visual representation strategy adopted in *LogAnalysis* is the called *force-directed model*. It is computed using the Fruchterman–Reingold algorithm (Fruchterman and Reingold 1991), in which nodes repel each other and

edges act as springs. The consequent displacement of nodes and links shows users clustered in groups which can be identified on the base of their increase of connectivity. The Barnes-Hut algorithm (Barnes and Hut 1986) simulates a N-body repulsive system in order to continuously update positions of elements. Optimization of visualization is interactively obtained by modifying parameters relative to the tensions of springs. Nodes with minor connectivity have greater tension, resulting in a displacement of the elements of a group in *orbital* position with respect to the central group. In *LogAnalysis* it is possible to modify different parameters, for e.g., spring constant of force, gravitation force and viscosity/drag of forces. In Fig. 2 it is possible to appreciate an example of the force-directed visualization model.

4.3.2 Edge betweenness clusterer

We have found that the Fruchterman–Reingold layout in conjunction with the edge betweenness clusterer (Girvan and Newman 2002) allows the interactive discovery of groups (henceforth, called *clans*) existing inside the network and those individuals acting as links among groups (hereafter, called *referents*). This feature is crucial because it allows to forensic analysts to highlight with low efforts those clans whose activity may be suspect inside the phone call network. Moreover, it leads to additional insights in particular regarding the interconnection of these referents among each other and among clans.

More generally, the edge betweenness clusterer, introduced by Newman (2004), is instrumental in the discovery of groups (called *communities* in SNA). This algorithm takes into account the weights of the edges in the network. In the particular scenario of the phone traffic networks, the concept of weights has already been defined as the number of phone communications among individuals. To highlight the clans, *LogAnalysis* exploits this technique according to a specific visualization strategy, called *visual aggregation*.

4.3.3 Visual aggregation

LogAnalysis adopts two algorithms to detect aggregations inside the network which represents the phone traffic. The first algorithm called *Edge Betweenness Clusterer* has been previously introduced. To this purpose, instead of considering the *betweenness centrality* associated to a node, we consider the *betweenness centrality* of an edge, which is defined as the number of shortest paths connecting pairs of nodes traversing it.

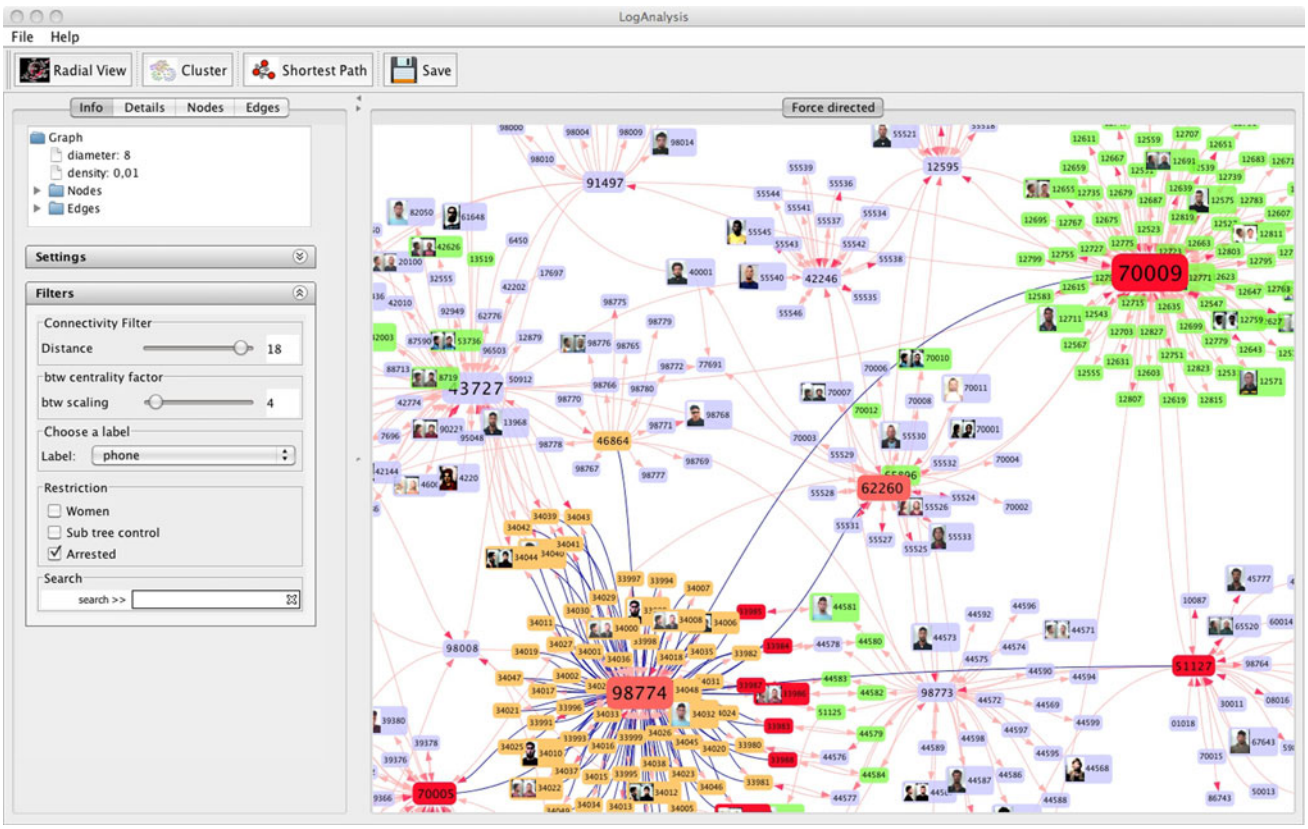


Fig. 2 Example of the force-directed visualization model. It is possible to put into evidence that groups of nodes repeal each other with respect to the weights of the edges connecting them

In the context of the visual aggregation, once the ranking of the edges is calculated, the algorithm simulates the deletion of those edges with the highest centrality, one by one, obtaining the effect of clustering the network in different groups (i.e., clans) that are weakly coupled with each other but densely interconnected within them. The functioning of this algorithm is based on the intuition that edges with high centrality connect groups characterized by high interconnectedness among their members and low outgoing connections. The edge betweenness clusterer has been proved to work well in the context of social networks. To the best of our knowledge, this is the first attempt to adopt this strategy to identify clans inside phone traffic networks.

The second algorithm, known as *Newman's community identification algorithm* (Newman 2004), is a variant of the hierarchical agglomerate clustering (it is also adopted in Vizster (Heer and Boyd 2005)). Regardless the adopted algorithm, *LogAnalysis* visually presents the identified clans by surrounding all members with a translucent convex hull (see Fig. 3).

By expanding the action of filtering one can obtain interesting visualizations. Groups comprised of a single node (i.e., monadic clusters) which satisfy the filtering condition are compressed and shown as a star.

Moreover, interactive cluster discovery is available in *LogAnalysis*. Users can suppress an arbitrary number of edges to discover strategic groups, together with links among groups. Target edges are chosen according to the algorithm known as edge betweenness clusterer. Labels of nodes belonging to the same cluster bear the same color. Different colors identify elements not belonging to a cluster.

4.3.4 Radial tree layout

The third layout algorithm introduced in *LogAnalysis* is called radial tree. It allocates the elements of a graph in

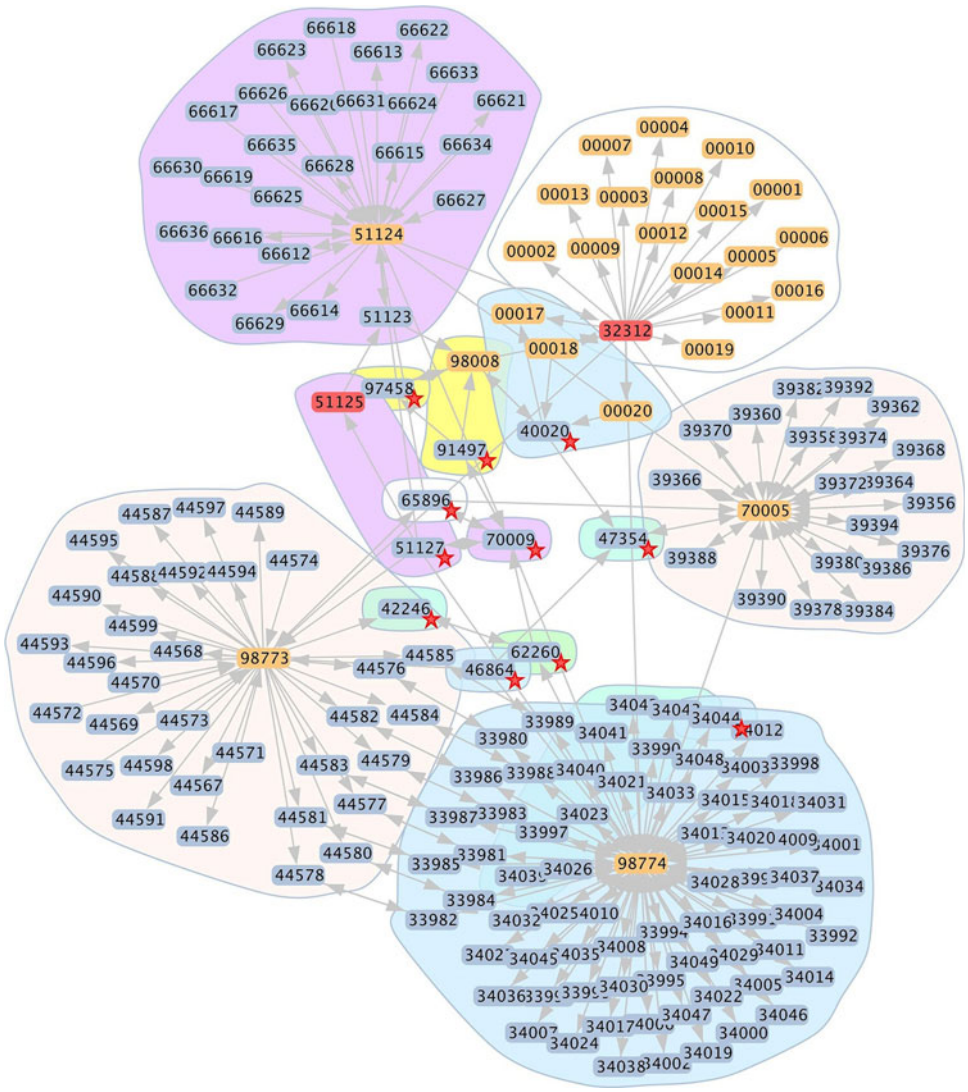


Fig. 3 Example of visual aggregation layout. It is based on the edge betweenness clusterer to divide the network in different clans on the base of the interactions among members

radial positions and defines several levels upon concentric circles with progressively increasing radii. The algorithm developed by Yee et al. (2001) also puts nodes in radial positions but gives the possibility of varying positions while preserving both orientation and order.

According to that technique, a selected element is placed at the center of the canvas and all the other nodes are subsequently placed upon concentric circles with radii increasing outwards. This visualization strategy is instrumental in the context of the forensic analysis because it allows to focus the attention of detectives on a suspect, and to have a close look to its connections.

The interactive visualization by using the radial tree layout is shown in Fig. 4. The interface supports filtering and searching elements within the network; a forensic analyst could select a specific node, which is placed at the center of the canvas. Nodes lying on the circumference of concentric circles, centered on that node, could be also progressively displaced from the selected one. Moreover, edges are visualized by using different thickness, calculated with respect to the number of calls among the given connected nodes. The overlapping of nodes may be avoided by superimposing a force-directed visualization to the radial tree algorithm.

A useful extension that we implemented, shown in Fig. 5, is called radial exploded layout. Selecting a specific node, the analysts can focus on its acquaintances that are displayed by using a radial layout. The characteristic of this exploded strategy is that if focuses only on a specific suspect and puts into evidence its links.

5 Case study

5.1 Aim of the experimentation

The aim of the current section is two-fold: first of all, in order to highlight the potential of *LogAnalysis* and the features provided to forensic analysts by the adoption of this tool, we discuss more in detail a number of examples, including the applicability of some centrality measures discussed above in the assessment of the importance of actors in phone call networks, the application of visualization techniques to highlight patterns of interactions among individuals in the networks, etc.

In addition, we underline that *LogAnalysis* has been already adopted by one of the authors during several real-world criminal investigations. To this purpose, in this

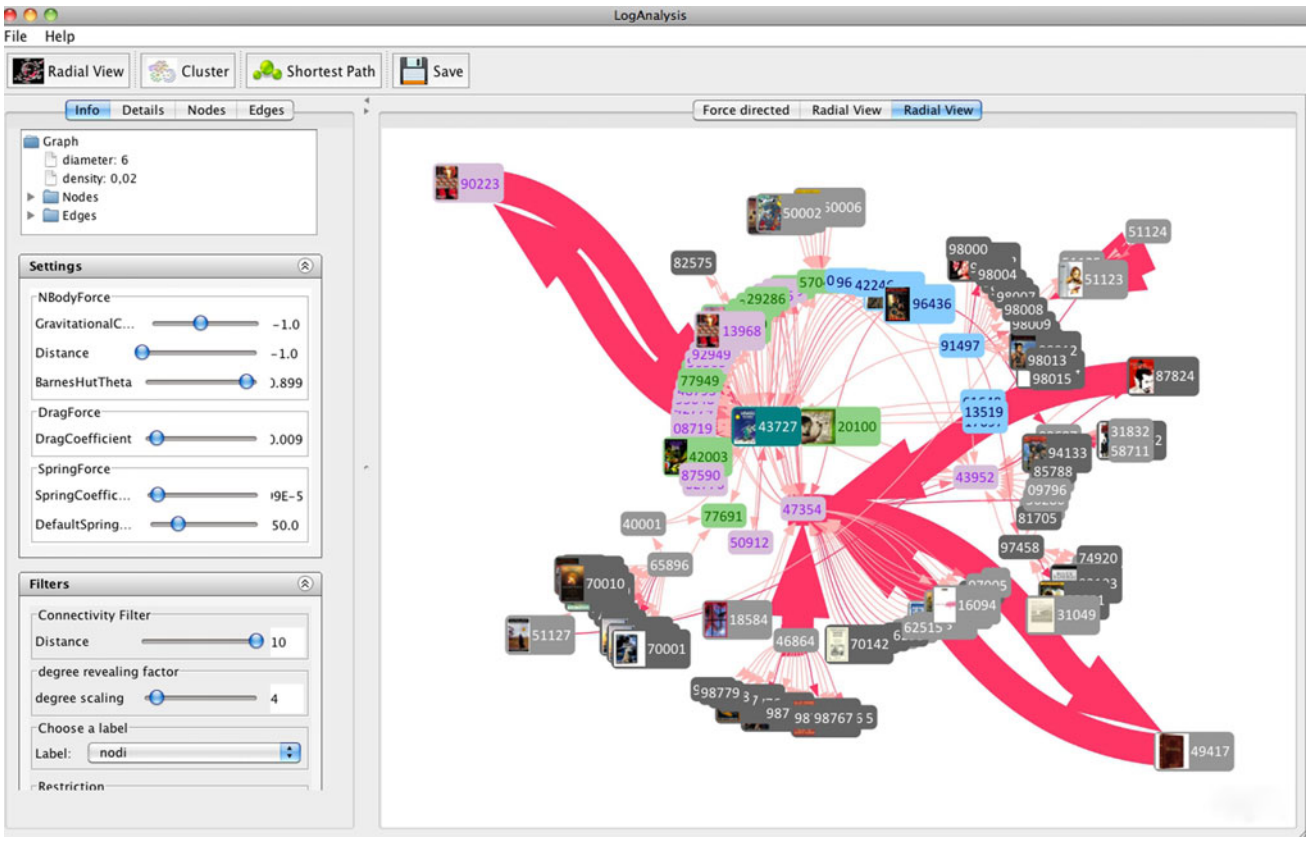


Fig. 4 Example of radial view layout. A node specified by the analyst is put into the center of the visualization. In addition, nodes represented with different colors belong to different groups (i.e., clans)

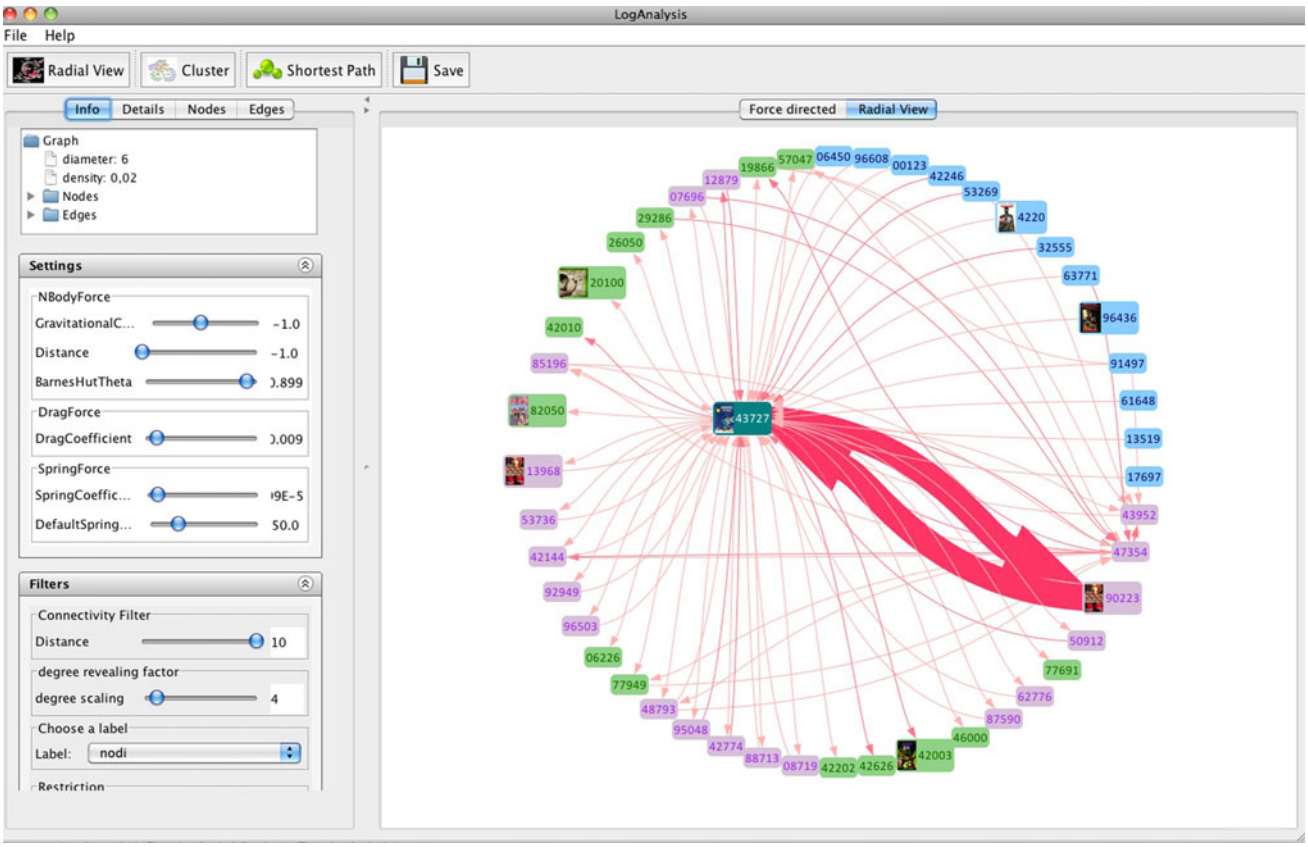


Fig. 5 Example of exploded radial view layout. A selected node is put into the center and the neighborhood is presented in a radial layout. Nodes with different colors belong to different groups

section we also report some data about these cases. In particular, we provide details regarding several small and large case studies (including the details about the datasets of phone call networks adopted during the investigations) in which *LogAnalysis* has been adopted to obtain additional insights regarding the networks structure. Finally, this section is instrumental to introduce some additional features, voluntarily not discussed before, in order to understand their usage in the context of a real investigation.

We tested our tool against different datasets (reported in Table 2), whose size was comprised between about 4,000 and 8 millions mobile communications.

One important feature we discovered in the criminal phone call networks is the growth rate. We found that, even though the number of entries in the log files grows, the corresponding size of the network grows more slowly. In fact, the analysis performed by forensic investigators is focused on the study of the network related to individuals who are already suspected of being involved in criminal activities, or being part of criminal organizations (i.e., the *clans*) or terroristic groups. This reflects in a network whose structure grows slowly and comprises a relatively small number of nodes/edges (with different weights) with respect to the number of phone calls reported by the log files.

In our real phone call network criminal investigations, usually the analyst started with the study of the *ego networks* of the individuals already suspect, those whose involvement in the criminal activities has been previously proved. The main goal of the analyst was to disclose

Table 2 Datasets adopted during real forensic investigations using *LogAnalysis*. We highlight that the number of entries of the log files grows at a very different rate with respect to the number of nodes and edges in the network. This is a typical feature of the criminal phone call networks

Case no.	1	2	3	4	5	6	7	8
No. nodes	148	170	381	461	320	543	912	702
No. edges	204	212	688	811	776	1,229	2,407	1,846
No. log entries	4,910	8,447	125,679	250,886	280,466	589,512	7,567,119	8,023,945

additional information about the underlying criminal organization. For example, one important task was to put into evidence other individuals, whose activity was suspect, in order to hypothesize their complicity with actors whose involvement in the criminal organization was ensured. This step was fundamental because, by identifying a small number of additional possible suspects, it has been possible to proceed with other "traditional" investigation methodologies, which would be not possible (in terms of time and cost constraints) otherwise.

5.2 Further details and simple use case

In the following we discuss an example use case that describes the usage of *LogAnalysis* during criminal investigations. As introduced above, analyzed data represent the phone call network of individuals suspected of belonging to criminal organizations. The period of analysis usually coincides with the commission of certain serious crimes. The adoption of our tool is instrumental to prove that those criminal facts have been planned and committed by the considered suspects.

Upon request by judiciary authorities, mobile phone service providers release all data logs about a certain set of suspected actors to the police force. After the import of phone call data in *LogAnalysis*, the process of analysis may start with the visualization of the phone call network by using the force-directed layout. This is helpful to get a picture of the phone call network and the connections of suspected actors among each other and with other external individuals. Unless the number of individuals exceed thousands of actors, which requires a manual process of filtering, we remark that our tool is able to provide with a graphical meaningful visualization of the phone call network. One advantage of the force-directed layout is the possibility of easily identifying clusters of actors within the network.

In order to improve the visualization, it is possible to apply some simple filters. For example, once the forensic investigator identifies an actor of interest, just clicking on it, *LogAnalysis* highlights those individuals with which the given actor is connected to, and those with which it shares the most of the contacts. In that case, the number of in-going connections represent the *popularity* of a certain actors and the number of out-going connections represents its *gregariousness*. It is easily possible to identify who are the individuals with respect to this actor is a gregarious, and who are those of which he/she exercise any influence.

Double-clicking on a given actor, the layout manager exploits the force-directed radial layout (see Fig. 4). In such away it is possible, not only to have a picture of all the contacts of a given actor, but also to highlight the intensity with which those communications occur. In addition, it is possible to put into evidence the *affiliation* of each actor to a given cluster, identified by different colors (see Fig. 5).

LogAnalysis is particularly suited to assess the presence of clusters in the given phone call network and to visually put into evidence their structure (see Fig. 3). This functionality is helpful to establish the role of a given set of actors inside a given group and to understand the structural and hierarchical organization of a possible criminal network. To assess certain hypotheses on the hierarchical structure of given criminal network, the forensic analyst may exploit the tool depicted in Fig. 6, that is helpful to have an immediate picture of the intensity of the communications among a set of actors, highlighting those connections whose relevance for the investigation is higher.

5.3 Overall metric tool

In the following, we are going to introduce additional features of *LogAnalysis* that are instrumental in the context of real-world criminal investigations.

An important and useful feature provided by our tool discussed in this case study is related to the possibility of calculating global quantitative metrics on the nodes/edges of the network. In particular, it is possible to evaluate some features usually adopted in SNA (Perer and Shneiderman 2006) such as: (1) overall network metrics (i.e., number of nodes and edges, density, diameter); (2) node rankings (i.e., degree, betweenness, closeness and eigenvector centrality); and finally (3) edge rankings (by means of weights).

This first step has been helpful for the analyst to gain a first insight about the structure of the network, in particular putting into evidence individuals whose centrality values were suspect, with respect to the others. Similarly, the same quantitative evaluation puts into evidence those connections (i.e., phone communications) that occur more frequently and those actors that are more active in the network. In Table 3 we report all the metrics calculated on the case studies 1–4.

5.4 Data visualization

Figures 2, 4, and 5 show some details about the *LogAnalysis* user interface. Once imported, data about the phone traffic network are visually represented by using the default view (i.e., by means of the aforementioned force-directed layout). Each node represents a cell phone, and edges indicate communications among them. On the left, a control panel provides tools and filters in order to tune the visualization of the network.

Using the available dynamic filters, it is possible to hide or highlight those nodes (or connections) which satisfy specific criteria. Moreover, analysts could interact with

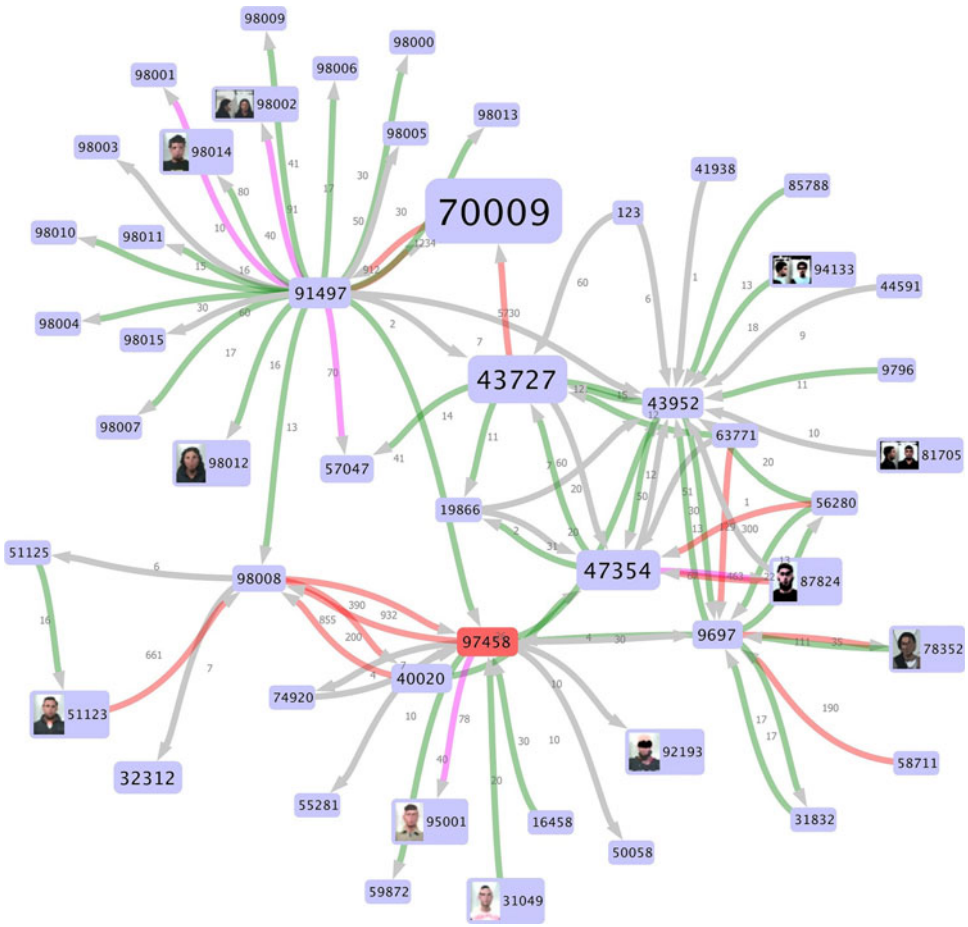


Fig. 6 Example of edge decorated. This feature classifies the edges with respect to a weight function and adopts different colors according to these weights. It is fundamental in order to give to the analyst an immediate picture that summarizes the intensity of the communications through the network

the graph, for e.g., moving, hiding or emphasizing specific elements, in order to dynamically re-arrange the structure of the graph.

The visualization algorithm adopts a weighted representation of edges, drawing those edges with higher weights by means of thicker lines. Standard nodes are represented by using light-blue as default color. Color filters could be defined by users, accordingly to specific conditions. For e.g., in this case study, "light-green" nodes reflect the "arrested" condition, "light-red" nodes accord to the "sub tree control" filter.

All these tools are provided in order to produce more readable network graphs. It is additionally possible to adopt "distance" filters, excluding from the visualization all the nodes far from the selected one more than the user-defined threshold. This is particularly helpful if the network that is under investigation is very large, constituted by more than 1,000 elements.

The optimal network visualization is a combination of both manual and automatic arrangements. First, it is possible to automatically pan and zoom, so as the whole network fits the display area (this particular approach may or may not be appropriate, depending on the size of the network or on the specific task the analyst would like to perform).

However, the display automatically pans when a new node is expanded, centering on the newly expanded network. In addition, the tool provides manual panning and zooming features in order to better satisfy user needs. Moreover, it is possible to choose which kind of "labels" should be visualized, among the "node-id", the picture (if available), or both. Even if the case study presented in this paper is inspired by a real investigation, for privacy reasons in the figures displayed in this work the pictures are fictitious. In the real investigations, these pictures represent the mugshots of suspects (for those who are available).

5.4.1 *Edge decorated*

In this section we focus the attention on a specific feature regarding the visualization, called *edge decorator*. This technique we propose is optimal in the case both of phone traffic networks and, in our opinion, more widely in SNA.

In detail, this strategy that has been introduced in *Log-Analysis* produces graphs not only according to the force-directed layout, but also by adopting different colors not

Table 3 Overall metrics calculated on the datasets of our four case studies

Metric	Case 1	Case 2	Case 3	Case 4
No. log entries	4,910	8,447	125,679	250,886
No. nodes	148	170	381	461
No. edges	240	212	688	811
No. connected components	1	1	1	1
Diameter	6	7	7	6
Average geodesic	3	3.418	3.898	3.514
Graph density	0.002	0.010	0.005	0.004
Minimum in-degree	0	0	0	0
Maximum in-degree	32	48	80	83
Average in-degree	1.419	1.533	1.806	1.765
Median in-degree	1	1	1	1
Minimum out-degree	0	0	0	0
Maximum out-degree	33	38	78	81
Average out-degree	1.414	1.438	1.806	1.765
Median out-degree	1	1	1	1
Minimum betweenness centrality	0	0	0	0
Maximum betweenness centrality[a]	12,975.267	14,345.581	63,244.345	85,132.261
Average betweenness centrality	358.932	443.231	1,105.139	1,154
Median betweenness centrality	0	0	0	0
Minimum closeness centrality	0.001	0.001	0.001	0
Maximum closeness centrality	0.003	0.005	0.001	0.001
Average closeness centrality	0.002	0.003	0.001	0.001
Median closeness centrality	0.002	0.003	0.001	0.001
Minimum eigenvector centrality	0	0	0	0
Maximum eigenvector centrality	0.077	0.004	0.033	0.040
Average eigenvector centrality	0.007	0.005	0.003	0.002
Median eigenvector centrality	0.494	0.376	0.001	0.001
Minimum clustering coefficient	0	0	0	0
Maximum clustering coefficient	1	1	1	1
Average clustering coefficient	0.069	0.088	0.027	0.036
Median clustering coefficient	0	0	0	0

[a] In *LogAnalysis* the betweenness centrality is not normalized

only for nodes but even for edges. To this purpose, we recall that the node color is given by the clan each node belongs to. Instead, the edge color is calculated by means of a weight function (in our case, the number of calls between a pair of nodes). Edges are annotated with weights associated to both directions (in- and out-degree). The interval in which the weights lie is normalized, depending on the characteristics of the network. However, this strategy results in the adoption an *edge color code*, that in our case study has been calculated as follows: (1) gray for $weight < 10$, (2) green for $10 \leq weight \leq 60$, (3) fuchsia for $61 \leq weight \leq 100$, and (iv) red for $weight > 100$.

The main advantage of introducing color code for nodes and edges is the possibility of easily identifying the strongest relationships, among hundreds of, or even thousands of, nodes and edges. During the real investigation, this feature has been proved to be helpful in order to give to the analyst a clear picture of the intensity of the communications among different actors of the network, with the only effort to give a overall glance on the network itself. Finally, the possibility of visually putting into evidence those communications paths that occur more frequently with respect to the average is instrumental because it allows to highlight in a visual way those information provided by the *overall metric tool*.

5.4.2 Shortest path finder

Another useful visualization tool provided by *LogAnalysis* is the *shortest path finder*. The usage of the shortest path finder is crucial to highlight those paths that are optimal in order to spread information through the network. In detail,

the tool is useful to highlight nodes and edges involved in the shortest path between any given pair of nodes of the network. This representation allows to highlight relationships among individuals belonging to distant groups in the graph. In Fig. 7 the usage of the tool is presented. In this specific case, the analyst was interested in understanding the most efficient way of communication that intervenes between nodes 289 and 379, two possible suspects. Even though these nodes appears to be distant, it exists in the considered network a relatively short path, constituted only by four hops that connects these suspects. Another essential information that it is possible to put into evidence by using this tool is that, usually, information can efficiently flow through those nodes that are more central in their respective clans, and that there exist usually a small number of *referents* that vehiculate the most of the communications.

5.5 Time filtering

A powerful filter included in *LogAnalysis*, which deserves a specific explanation, is the *time filter*. Starting from the assumption that phone call networks are time-dependent, and the structure of the network could change accordingly, we introduced in our tool the possibility of "filtering" the structure of the network with respect to specific temporal constraints. As shown in Fig. 8, it is possible to select a time interval, by using a slider which comprises the whole temporal range covered by the log file. The structure of the network is filtered accordingly, removing all the edges representing connections (i.e., phone calls) which did not take place in that specific time window, and insulating (or hiding) those nodes not involved in the network at that given time. In addition, if the user modifies the time interval, nodes involved are automatically "engaged" or detached and, thanks to the force-directed algorithm, are attracted or rejected inside/outside the network. The *time filter* is a feature that has been proved to be incredibly powerful. Its adoption helps the analyst in identifying those communications that happened in a specific time window (say, for e.g., a day) and the structure of the graph during the given interval. Such a possibility heightens the capability of the detective to understand the structure of a criminal organization and its evolution over time. In fact, as the connections may

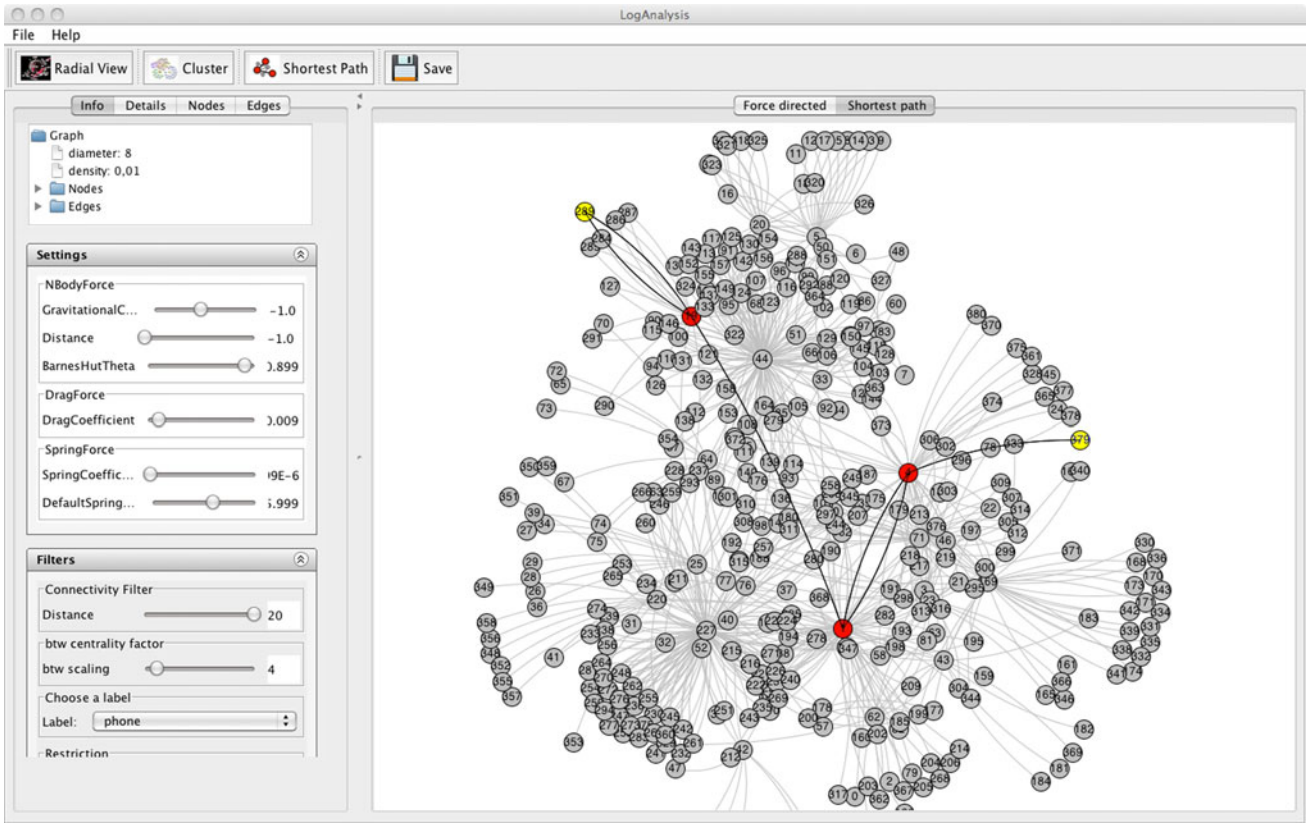


Fig. 7 The shortest path viewer. It is fundamental when the analyst would like to understand the shortest ways of communications in the network. Usually, criminal organizations are structured in order to optimize the number of communications among members to efficiently disseminate information. This is possible by following short paths of communications that can be discovered by using this tool

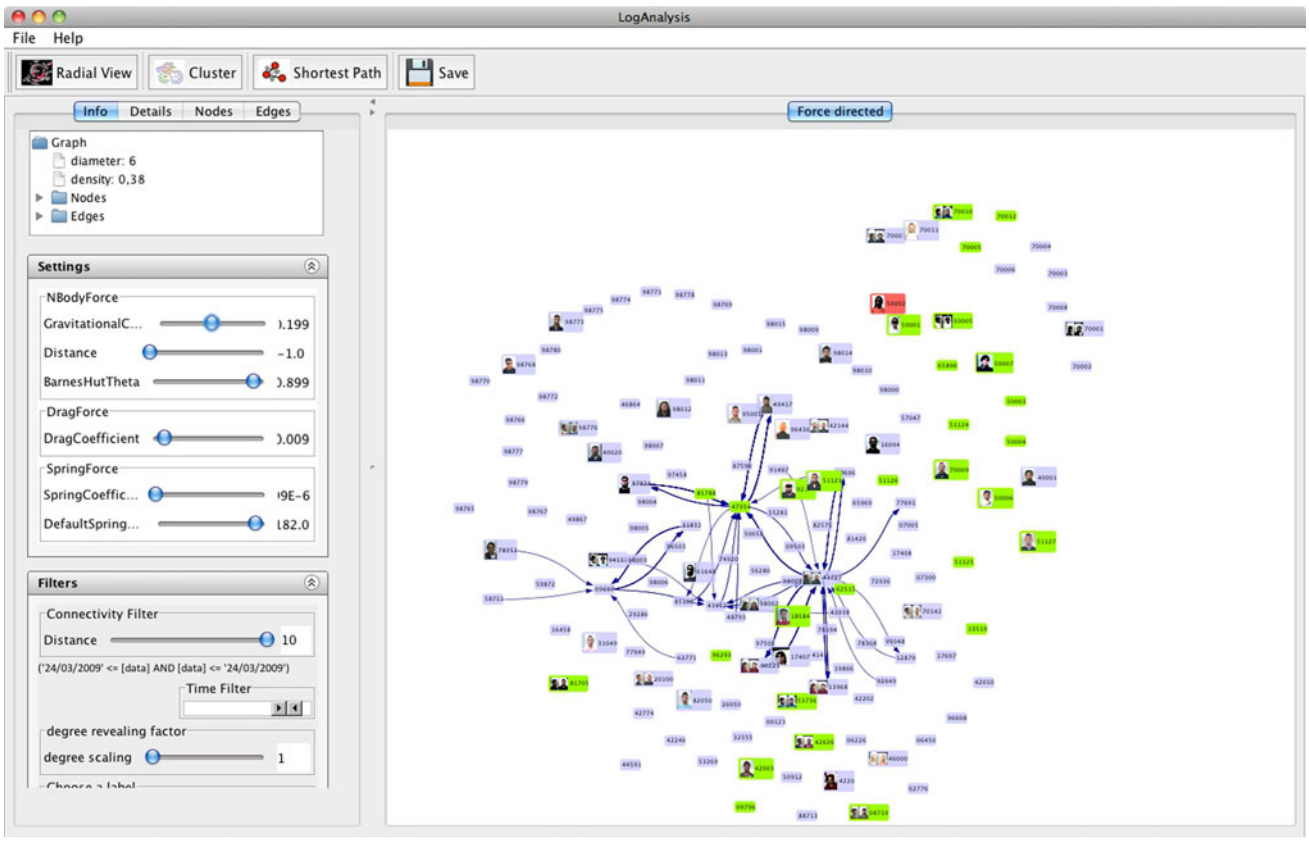


Fig. 8 The time filter feature. This tool is able to improve the capabilities of an analyst because it allows to specify a particular time window and to investigate how the structure of the network changes accordingly. Nodes are dynamically engaged or detached according to the time information about the phone call, dynamically altering the structure of the network

spread during a long time interval, it is fundamental for the investigator to understand at what time the given graph was already reflecting, for e.g., the structure of a clan or the presence of a particular *referent* in the network. Similarly, the possibility of dynamically visualize the effect of engaging or detaching nodes according to the modification of the time filter is crucial in order to highlight those nodes that are involved, during a specific time window, in the phone traffic network.

5.6 Time flow analyzer

The last visual tool which has been included in *LogAnalysis* is related to the time filtering features previously presented, but it is also detached from the representation by means of a graph of the phone traffic networks. In fact, the *Time Flow Analyzer* (see Fig. 9) considers each single phone call as an *event*, graphically represented in a timeline which covers a specific, user-defined, interval of time. The advantage of a time-dependent visualization is crucial in the scenario of the forensic investigations. In fact, it allows to organize information and event-flows in a visual manner in order to put into evidence the degree of correlation of specific events (in our case the phone connections).

In the *Time Flow Analyzer* we included in *LogAnalysis*, the visual representation of a bi-dimensional space presents the days on the x-axis and the hours on the y-axis. Each event is presented by a colored square, whose color depends on the type of communication represented (i.e., sent/received calls and SMS and other type of communications, etc.). It is possible to apply several filters, in order to select only specific events:

All, all the phone events;
1–2, sent/received calls;
6–7, Sent/received SMS;
0, All the other type of communications.

Moreover, it is possible to zoom in/out the time interval in order to obtain additional insights about connections of events. Finally, the *Time Flow Analyzer* allows the analysts to query the data in order to retrieve information about specific events or even about specific phone numbers, etc. The adoption of this tool during real

Fig. 9 The time flow analyzer tool. This tool is helpful to consider the time-dependence of events (i.e., phone calls) in a specific time window and it is crucial in order to highlight phone call cascades during criminal events

investigations is crucial to identify single events that set off to cascades of related events. In detail, the time-dependent visualization allows the analyst to highlight those communications that triggered, in cascade, additional communications to other actors. For e.g., it is possible to specify small time windows that may coincide to specific criminal events, in order to emphasize those phone connections that happened during the that interval and the involved actors, with an heightened probability of finding additional suspects or individuals involved in the criminal organization.

The aspect of temporal analysis in the context of phone call investigations has an extreme relevance. The time flow analyzer feature of *LogAnalysis* allows to forensic analysts to highlight those fundamental communications that happened in critical periods of interest for a given investigation. For example, from Fig. 9 it is possible to put into evidence that an important amount of phone calls happened before, during and after the commission of a serious crime, among those components of the criminal organizations highlighted by means of the network structure of the phone calls. The temporal analysis, although not directly represented by means of networks, is closely interconnected to the structure and the evolution of the phone call network itself, and the time flow analyzer tool is instrumental to highlight and understand this critical dependency.

5.7 Stacked histograms

The last tool of *LogAnalysis* described in this work is called stacked histograms. This tool empowers the temporal analysis features provided by *LogAnalysis* and it is shown in Fig. 10. Its functioning is explained as follows. Similar to the time flow analyzer tool, in the stacked histograms on the x-axis it is represented the time flow, but on the y-axis there is the amount of phone calls in the given interval. In the stacked histograms, each actor has assigned a stack, whose color and intensity is proportional to the number of phone calls related to the given individual, during the specific period of interest taken in consideration from the forensic analyst. In detail, the intensity of the color with which the stack histograms are represented is related to the absolute number of phone calls (in-coming and out-coming contacts) of each actor, while the thickness of the histogram may represent the in-degree or the out-degree of the given user at that day (highlighting those actors who are more popular and those who are more gregarious). The stacked histogram tool is helpful to get a picture of the phone call activity of the set of considered actors elapsed during a specific time window. Finally, it is particularly instrumental to understand in which proportion the phone activity of a given actor is with respect to the other individuals in its network who are in contact with him/her (i.e., its ego-network), in that specific time period.

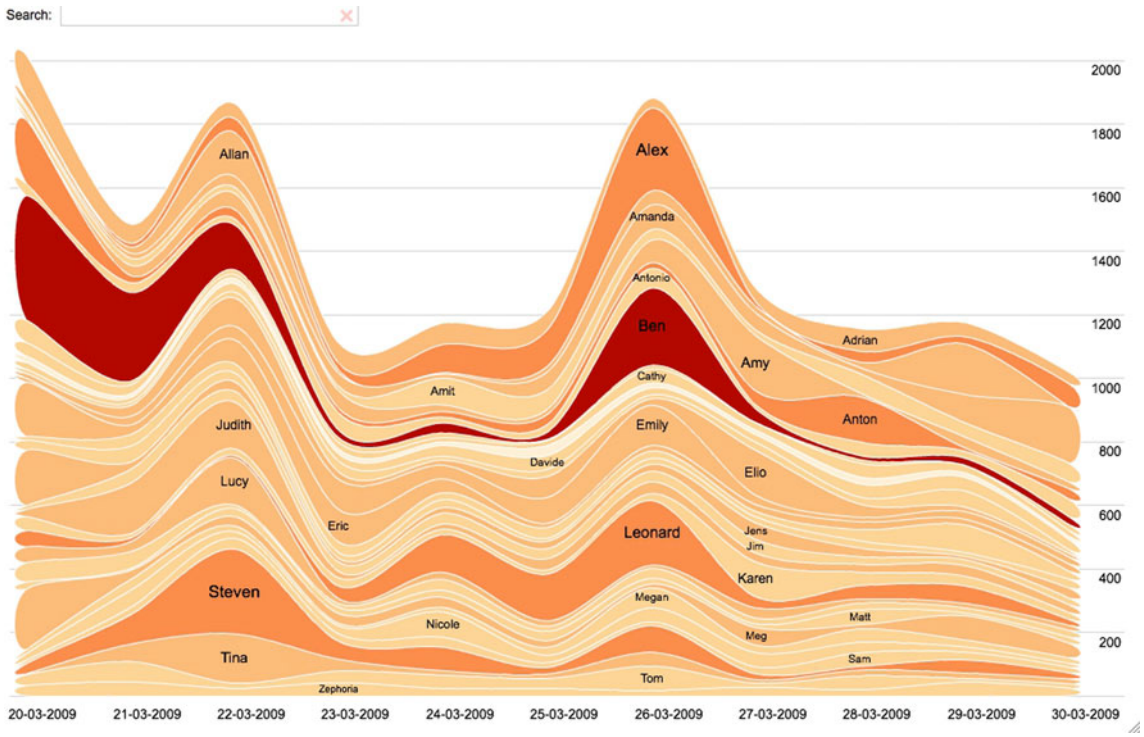


Fig. 10 The stacked histogram tool is helpful to visually summarize the communications among actors elapsed in a given time interval

6 Conclusions

The analysis of networks of phone traffic for investigative and forensic activities, aimed at discovering the relational dynamics among individuals belonging to criminal associations is a hard task. Our goal was to develop a systematical model of analysis oriented to simplify exploration of networks whose elements are large collections of mobile phone traffic data. Our approach is based on SNA studies, which developed useful techniques to tackle the problem. Nevertheless, few useful tools hitherto support this type of network analysis. The tool we developed, *LogAnalysis*, supports the exploration of networks representing mobile phone traffic networks. It employs visual and statistical features in order to help in discovering cohesive groups, *key figures* and individuals acting as link. *LogAnalysis* helps in systematically and flexibly obtaining measures typical of SNA in order to find outlier/anomalous values. Users can interactively identify sub-groups and focus on interesting actors of the network. In addition, the tool includes the possibility of exploring the temporal evolution of the network structure and the temporal information flow.

Future improvements to *LogAnalysis* will concern the geo-spatial analysis of phone traffic networks and the implementation of novel measures of centrality (De Meo et al. 2012; Abdallah 2011), community detection algorithms and graph visualization techniques.

Acknowledgments The authors would like to thank the editor and the anonymous reviewers whose comments helped us to greatly improve the quality of the work.

References

Abdallah S (2011) Generalizing unweighted network measures to capture the focus in interactions. Soc Netw Anal Min 1(4):255–269

Barnes J, Hut P (1986) A hierarchical 0(N log N) force-calculation algorithm. Nature 324:4

Blondel V, Guillaume J, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. J Stat Mech Theory Exp 10:P10008

Candia J, González M, Wang P, Schoenharl T, Madey G, Barabási A (2008) Uncovering individual and collective human dynamics from mobile phone records. J Phys A Math Theor 41:224,015

Catanese S, Fiumara G (2010) A visual tool for forensic analysis of mobile phone traffic. In: Proceedings of the second ACM workshop on multimedia in forensics, security and intelligence, ACM, pp 71–76

Chen H, Zeng D, Atabakhsh H, Wyzga W, Schroeder J (2003) Coplink: managing law enforcement data and knowledge. Commun ACM 46:28–34

Coscia M, Giannotti F, Pedreschi D (2011) A classification for community discovery methods in complex networks. Stat Anal Data Min 4(5):459–563

De Meo P, Ferrara E, Fiumara G, Provetti A (2011) Generalized louvain method for community detection in large networks. In: Proceedings of 11th international conference on intelligent systems design and applications IEEE, pp 88–93

De Meo P, Ferrara E, Fiumara G, Ricciardello A (2012) A novel measure of edge centrality in social networks. Knowledge-based Systems. doi: 10.1016/j.knosys.2012.01.007

Eagle N, Pentland A, Lazer D (2008) Mobile phone data for inferring social network structure. Social Computing, Behavioral Modeling, and Prediction 79–88

Eagle N, Pentland A, Lazer D (2009) Inferring friendship network structure by using mobile phone data. Proc Natl Acad Sci 106(36):15,274

Ferrara E, Fiumara G (2011) Topological features of online social networks. Commun Appl Ind Math 2(2):1–20

Fortunato S (2010) Community detection in graphs. Phy Rep 486(3–5):75–174

Freeman L (1977) A set of measures of centrality based on betweenness. Sociometry 40:35–41

Fruchterman T, Reingold E (1991) Graph drawing by force-directed placement. Softw Pract Exp 21(11):1129–1164

Gilbert F, Simonetto P, Zaidi F, Jourdan F, Bourqui R (2011) Communities and hierarchical structures in dynamic social networks: analysis and visualization. Soc Netw Anal Min 1(2):89–95

Girvan M, Newman M (2002) Community structure in social and biological networks. Proc Natl Acad Sci 99(12):7821

Heer J, Boyd D (2005) Vizster: visualizing online social networks. In: Proceedings IEEE symposium on information visualization, p 5

Jonker D, Wright W, Schroh D, Proulx P, Cort B (2005) Information triage with trist. In: Proceedings of 2005 international conference on intelligence analysis, pp 2–4

Kapler T, Wright W (2004) Geotime information visualization. In: Proceedings of IEEE symposium on information visualization, pp 25–32

Mellars B (2004) Forensic examination of mobile phones. Digit Investig 1(4):266–272

Newman M (2004) Fast algorithm for detecting community structure in networks. Phys Rev E 69(6):066,133

Newman M (2005) A measure of betweenness centrality based on random walks. Soc Netw 27(1):39–54

Onnela J, Saramäki J, Hyvönen J, Szabó G, Lazer D, Kaski K, Kertész J, Barabási A (2007a) Structure and tie strengths in mobile communication networks. Proc Natl Acad Sci 104(18):7332

Onnela JP, Saramäki J, Hyvönen J, Szabó G, de Menezes MA, Kaski K, Barabási AL, Kertész J (2007b) Analysis of a large-scale weighted network of one-to-one human communication. New Journal of Physics 9(6):179+

Palla G, Barabási A, Vicsek T (2007) Quantifying social group evolution. Nature 446(7136):664–667

Perer A, Shneiderman B (2006) Balancing systematic and flexible exploration of social networks. IEEE Trans Visual Comp Graph 12:693–700

Porter M, Onnela J, Mucha P (2009) Communities in networks. Notices AMS 56(9):1082–1097

Saravanan M, Prasad G, Karishma S, Suganthi D (2011) Analyzing and labeling telecom communities using structural properties. Soc Netw Anal Min 1(4):271–286

Scott J (2011) Social network analysis: developments, advances, and prospects. Soc Netw Anal Min 1(1):21–26

Smith M, Shneiderman B, Milic-Frayling N, Mendes Rodrigues E, Barash V, Dunne C, Capone T, Perer A, Gleave E (2009) Analyzing (social media) networks with NodeXL. In: Proceedings 4th international conference on communities and technologies, ACM, pp 255–264

Sundsøy P, Bjelland J, Canright G, Engø-Monsen K, Ling R (2010) Product adoption networks and their growth in a large mobile phone network. In: Proceedings of 2010 international conference on advances in social networks analysis and mining, IEEE, pp 208–216

von Landesberger T, Kuijper A, Schreck T, Kohlhammer J, van Wijk JJ, Fekete JD, Fellner DW (2011) Visual analysis of large graphs: state-of-the-art and future research challenges. In: Computer graphics forum

Wasserman S, Faust K (1994) Social network analysis: methods and applications. Cambridge University Press, Cambridge

Wright W, Schroh D, Proulx P, Skaburskis A, Cort B (2006) The sandbox for analysis: concepts and methods. In: Proceedings of SIGCHI conference on human factors in computing systems, ACM, pp 801–810

Yee K, Fisher D, Dhamija R, Hearst M (2001) Animated exploration of dynamic graphs with radial layout. In: Proceedings of IEEE symposium on information visualization, p 43