# A combined approach for suspicious networks detection in graphs

**Charles PEREZ, Babiga BIRREGAH, Patrick LACLEMENCE, Eric CHATELET**
charles.perez@utt.fr, babiga.birregah@utt.fr, patrick.laclemence@utt.fr, eric.chatelet@utt.fr

Charles Delaunay Institute, UMR (CNRS) 6279 STMR,
University of Technology of Troyes, 12 Rue Marie Curie, 10010 Troyes, France. Tel: +33(0) 325717600

**Mots clefs :**
réseaux sociaux de l'Internet, analyse des réseaux sociaux, détection d'anomalies, familiar stranger, décomposition en valeurs singulières, réseaux suspects
**Keywords:**
Internet social network, social network analysis, anomalies detection, familiar stranger, singular value decomposition, suspicious network

## Abstract

The recent growth of the social networks of the Internet was followed by the emergence of new forms of threats and vulnerabilities. Users of these social networks sites are not immune to these new challenges. Several studies are devoted to users' vulnerability in these social networks and their entities. The threat via web 2.0 (soon web 3.0) takes the form of communities of dealers, dormant networks of extremists, organizations and individuals with malicious intent (pedophilia, piracy, hacking, etc.). Social networks have become a large platform for the activities of these groups such as propaganda, recruitment, training, selecting and hitting their targets without crossing any countries borders. All countries are concerned by this potential threat through Web 2.0 and thus the detection of malicious networks on the Internet is a central concern of all state authorities. It is therefore a major objective to detect these malicious networks. In this article we propose to combine anomaly detection with the detection of Familiar Stranger to identify networks that may be vectors of threats (malicious, hidden, dormant networks, etc). In our approach we combine a concept based on the topology and the configuration of the network with a purely sociological concept. Indeed the anomaly detection is devoted to identify individuals whose behaviors (in relation to their interactions with others) are abnormal. Therefore the awareness of these anomalies is crucial because an abnormal behavior is often synonymous with abnormal activity. This is the starting point of our work. Once these anomalies are detected, we propose to explore the network to find the familiar strangers of each of them. At this stage we have a family of nodes, named suspicious nodes, consisting of abnormalities and their respective Familiar Strangers. This family will help us to rebuild a part of the original network including individuals located on the shortest paths connecting suspicious nodes. Our work is a first attempt to build an efficient algorithm to detect suspicious nodes and links in a social network.

# 1 Introduction

Social Networks can be defined as a set of elements (persons, organizations...) that are interacting through relations. Social Networks Analysis (SNA) have been used since several decades in social science, marketing, studies of nonhuman social life, some branches of mathematics, computer science, etc. In 1953 the work of Moreno [1] contributed to lay the foundations of SNA such as sociometry. In early studies in social networks the major challenge was to access information on the nodes (persons or groups) and links (relations) of the networks under study. In this last decade social networks on the Internet has grown with the progress of the web 2.0 ([2], [3]) and web 3.0 ([4]). It is now possible to access a wide range of data on the nature, and the evolution of a given social network on the web. The popularity of social networking web sites, such as Facebook, Twitter and Linkedin, is now well established ([5], [6]). Within a few years social networking has become the communication medium of choice for Internet users. Most people prefer to use them to communicate with their list of contacts (friends, work colleagues, customers and even strangers). Social networks have become the platform used by people on the Internet to exchange information and data on their lives, their projects... Simultaneously, the diversity of the actors of social networks provides support to networks of criminals (pedophile, drug dealers, terrorists...) to communicate with their members, to recruit new members and to share their ideas. To cope with these new forms of threats, we need more decision support tools which can monitor the evolution of web communities, decrypt and detect the formation of new communities at risk and provide a global overview of the sociological challenges inherent to social networks on the web ([7]). Several studies have tried to handle these issues and propose some algorithms to extract suspicious networks from large-scale networks' data. The Anomalies detection ([8]) is one of these studies. Its main Goal is to detect individuals, who have an abnormal behavior in a given network. In the present work we make the assumption that not all abnormal behavior in a network is suspicious, but all suspicious behavior should be abnormal ([9]). However detecting suspicious nodes or links in a given network can be the first step of SNA strategy. In the remainder of this paper a node will be considered suspect:

- if it is an anomaly (in the sense of [9]) and possesses a specific subset of (suspicious) attributes (named suspicious anomalies).
- or if it is a familiar stranger ([10]) of a suspicious anomaly.

We propose in this work a combined approach to detect suspicious networks in graphs. The remainder of this paper is structured as follows. Section 2 presents the approach of anomaly detection in social networks represented by graphs. The concept of Familiar Strangers is then introduced in section 3 and some algorithms to detect them are presented. In the section 4 we propose a combined approach to detect some suspicious networks in graphs using anomalies and familiar stranger detection. This work ends in section 5 by an implementation of our algorithm on a case study and some perspectives in the conclusion section.

**Social network graphs of Internet**

A simple way to represent social networks is by graphs in which nodes represent the entities and edges describe links, relations or other characteristics that reflect information collected on entities. Obviously the type and properties of the resulting graphs closely depends on the knowledge that one has about the social network analyzed. For example the graph resulting from an ISN can be oriented or not depending on the meaning of the relation between the actors of de SN. In Internet social networks we assume in this paper that one works with non oriented graphs. That means if person A is related to the person B then B is also related to A (symmetric relation). Most of Internet social networks are working based on symmetric relation. In some cases the strength of links can be important. In other cases the state of the traffic (number of messages exchanged, etc) can be useful to obtain a weighted network ([11]). Nodes in social networks can possess a set of attributes such as name, surname, address, interests and other data that can be collected on a person. One of the advantages of graph representation is that a matrix can uniquely represent the graph. The main challenge in modeling social networks by graphs is to take into account the concept of hierarchy between nodes. For example, visualizing terrorist groups as simple graphs hide the fact that they are composed by leaders, intermediaries and followers ([12]). To analyze a hierarchical

network the partially ordered Set (poset) theory has been proposed to represent terrorist networks ([13], [14]).

# 2 Detecting Anomalies on Internet social networks

The field of anomalies detection is related to decision support strategy to address the problem of finding patterns in data that do not conform to expected behavior ([8]). With the exponential growth of networks anomalies detection cover a wide range of application fields such as fraud detection (credit cards, Stock Exchange), phone listings monitoring, intrusion detection, etc ([15]). [16] proposes six types of anomalies for web data that can be organized as graphs. (i) A vertex that exists is unexpected, (ii) An edge that exists is unexpected, (iii) The label on a vertex is different than was expected, (iv) The label on an edge is different than was expected, (v) An expected vertex is absent, (vi) An expected edge between two (or a self-edge to a vertex) is absent. In Social Networks represented by graphs, the aim of anomaly detection is to extract suspicious nodes that have an unexpected behavior. The anomaly detection is a central concern in cyber-security in the 21st century ([17], [18], [19]). Several methods are used to detect anomalies in data such as statistical techniques, K-Nearest Neighbor, clustering approaches and spectral-based approaches ([20], [21], [22], [23]). The latter techniques can also be applied to data that are represented by graphs. Most of these algorithms give as output a score reflecting how far a given node can be considered as abnormal. Figure 1 gives an overview of the classification of different approaches ([24], [8]). Recent developments discuss about the possibility to implement a comprehensive behavioral anomaly detection platform that will auto-analyze large amounts of data ([25], [26], [27]).

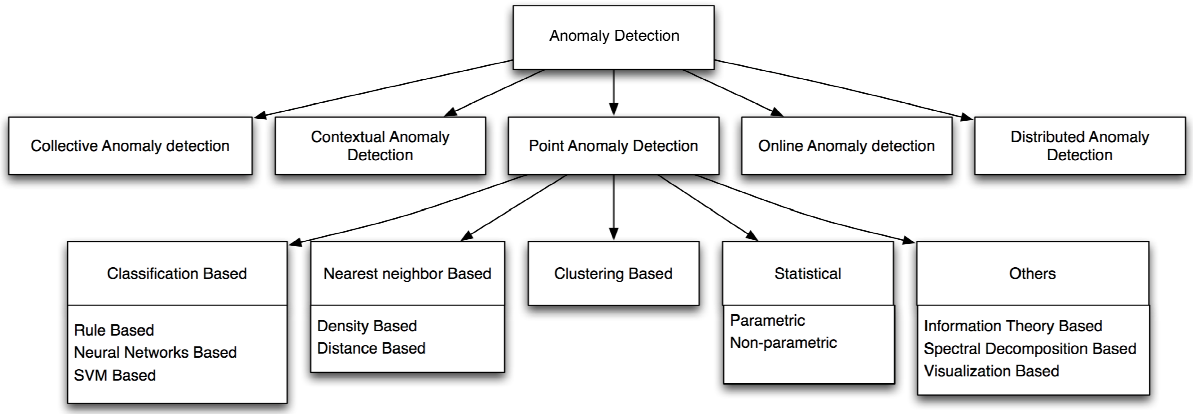



*Figure 1: Overview of anomaly detection fields*

In the remainder of this section we briefly present two of the four detection approaches mentioned above.

## 2.1 K-Nearest Neighbor approach (KNN)

This method is classified as proximity based approach ([28]). Thus a given element will be considered abnormal if it is located "far" from others while normal node posses a compact neighborhood. An element is declared abnormal whenever the distance to his $K^{th}$ Nearest Neighbor is high. The KNN method gives for each node

a score that is equal to the distance to the $K^{th}$ nearest neighbor. Nodes whose scores are greater than the threshold are ranked as anomalies. One refers to the 1-NN (KNN for K = 1) as the Nearest Neighbor method. In this case the score affected to each node is simply the distance to his nearest neighbor. To determine the scores one must define the notion of distance. The definition of distance is one of the challenging issues of this method. It is common that based on specific criteria; one proposes a new way to evaluate this distance ([29], [8]). That is the case especially when the notion of Euclidean distance between nodes is not relevant. In social networks (represented by graphs) this distance can depend on attributes and links between nodes and therefore KNN will be implemented after a preliminary analysis. The score of an anomaly is sensitive to the parameter K: a too small value of K can lead to a noisy outlier score while a too high score can lead to smooth set of scores. [30] analysis suggests a method to compare and evaluate the efficiency of the parameter K. For more details on the KNN methods the reader can refer to ([31], [32]).

## 2.2 The clustering methods

The clustering methods aim to organize nodes of a graph into different clusters based on a similarity measure ([33]). One can classify these methods into two groups: those who put all nodes into clusters ([34], [35]) and those who isolate nodes who do not belong to any community. The second type is closely related to the definition of the concept of anomaly. The assumption is that a normal node must belong to a cluster. Thus anomalies do not belong to any cluster. To extend the basic method sometimes the distance from each node to the centroid of the nearest community is used as anomaly score. This approach can be viewed as a particular application of 1-NN. Basically a node that does not belong to any community has a high score ([36]). Obviously these methods have the disadvantage that they are not principally built to detect anomalies but to detect clusters. However some exceptions can be encountered in literature ([37], [38]). Though these techniques are commonly encountered in several works dedicated to our subject, we chose in the present approach to use one of the spectral-based methods applied to social networks' graphs as presented below.

## 2.3 Spectral-based approach using SVD

The spectral method is not based on the notion of distance between nodes and therefore can be used for any type of graphs whose distance has no particular meaning. The assumption is that data can be represented in a subspace of smaller dimension in which normal entities (nodes) and anomalies appear significantly different. This approach is commonly used in image processing for compression and characteristics' extraction ([39]). Firstly data are represented in an eigenspace composed of eigenvectors of the adjacency or Laplacian matrix. The main advantage of this space is that vectors are orthogonal and then the notion of distance (in its classical sense) has a meaning. The term spectral is used because it is based on spectral theory of graphs. Spectral methods are based on matrix decomposition to extract the eigenspace corresponding. One can find in [40] a comparison between spectral methods for adjacency matrix and Laplacian matrix. In the present work we use the approach of [9] that applies Singular Value Decomposition (SVD) on the walk Laplacian matrix of a graph. One of the advantages of this method is that the SVD extracts a singular space with singular values that are sorted. Singular values can be understood as scores affected to each singular vector. This ranking helps to choose an appropriate two-dimensional subspace in which the data will be represented (algorithm 1). When the subspace is well chosen anomalies are the nodes that are far from the center. Let us consider a Social Network represented by the graph G(N,E,A), where N is the set of nodes (entities),E represents the set of edges (links, interactions, relations) and A the set of nodes' attributes. The anomaly detection algorithm based on SVD can be implemented as shown in Algorithm 1.

**Input**:
Graph *G(N,E,A)*,
Anomalies threshold *r*,
**Output:**
Set of Anomalies *A*

**1** $L_w \leftarrow$ walk Laplacian of *G*
**2** Apply the SVD to the $L_w$ matrix to generate the matrices *U* and *D*
**3** $\{d_1, d_2\} \leftarrow$ Select the two smallest non-zero eigenvalues *U* and *D*
**4** $\{ \vec{v}_1 , \vec{v}_2 \} \leftarrow$ Select the two Singular vectors from *U* corresponding to $\{d_1, d_2\}$
**5** Plot each node according to $\{ \vec{v}_1 , \vec{v}_2 \}$
**6 foreach** node $N_k$ from *N*
**7** **if** distance from origin to node $N_k$ is greater than *r* **then**
**8** Add $N_k$ to A
**9 end**
**10 return** *A*

*Algorithm 1: Spectral Based Anomaly Detection (SBAD)*

On the step 2 of algorithm 1 we must specify that other types of decompositions can replace SVD and the rest of the algorithm should be adapted to the chosen decomposition. The steps 3 and 4 are possible because the singular value decomposition leads to singular values that are positive and sorted in the diagonal matrix D. SVD can be time consuming ([41]). Some recent works propose alternative decompositions such as Compact Matrix Decomposition - CMD ([42]), CUR ([43], [44]) to gain calculation time. The plot of the graph with the two lowest singular vectors shows global disparities while choosing others shows more local disparities. Steps 3 and 4 concern the selection of the most interesting subspace. Notice that plotting the mean of the absolute value of each vector and selecting the two closest to zero can be also a good choice. The plot of the network's nodes in the subspace determined by the two lowest singular vectors shows global disparities while choosing others shows more local disparities ([45]). In our study we are interested in global behavior and therefore the choice of the subspace is every time based on the two lowest singular vectors. The step 7 is crucial because it gives the limit between abnormal and normal nodes. The choice of *r* is discussed later in section 4.

# 3 Familiar Strangers

The Familiar Stranger concept has been introduced by S. Milgram [46]([47]). This definition states that a familiar stranger is that person whom we encounter repeatedly, but never interacts with us ([47], [48], [49]). The notion of familiar stranger is the intermediary step in our everyday bipartite view (in social life) that only includes familiar and stranger persons. Such a concept also exists in the social network communities ([50], [51], [52]). In this paper we define them as individuals who do not interact with each other (no links between them) but who share some characteristics (stored in their attributes). These characteristics can be location, interests, job, hobbies or some other information collected (through the web) about them (set of facilitators ([53])). In the area of social network security, discovering familiar strangers can be a very interesting issue.

Indeed, it is observed that any person of a narcotic cell (embedded in a larger social network) came into the network through facilitators ([53]). These facilitators can most of the time be found as subsets of attributes' set. A typical example of facilitator that can be used in discovering suspicious individuals (drug dealers) is the fact that they have shared the same prison cell ([54]). Let us consider a social network represented by G(N,E,A). The attributes in A can be information filled in by users in their profile or collected information about them by an authorized third party on the web. We assume that each person Ni posses a list of attributes Ai. [10] defines the concept of familiar stranger as follows:

**Definition 1: Familiar Stranger**

The set of familiar stranger $T_u$ of a node u should respect two conditions:

- stranger condition (H1) : $\forall n \in T_u, edge(n,u) = 0$
- familiar condition (H2) : $\forall n \in T_u, \{A_n \mid \gamma \neq \varnothing, \gamma \subseteq A_i\}$, $\gamma$ is the Goal for the detection.

Depending on the purpose, the familiar stranger detection can be time consuming, especially if one wants to detect all the familiar strangers of any node without any limitation. The basic common approach is to fix a Goal that is a subset of attributes and to look on the graph for nodes that share this attributes (H2) but that are not directly connected (H1). Each approach will therefore focalize on the way to search on the graph with some extensions of the concept. Two main strategies are proposed in the literature: Random approaches and exhaustive approaches. ([10]). In this work we choose an exhaustive approach based on the Algorithm 2.

---

**Input**:
Graph *G(N,E,A)*,
Node *u*,
Goal $\gamma$ (Here a set of attributes),
List of contacts of *u*: $\vec{C}_u$

**Output:**
Set of nodes $F_u$ familiar strangers of node *u*

**1** $F_u \leftarrow \varnothing$
**2 foreach** node $N_k$ of $N$-{$u$}
**3** **if** $N_k$ respects $\gamma$ and $N_k \notin \vec{C}_u$ **then**
**4** $F_u \leftarrow N_k$
**5 end**
**6 return** $F_u$

---

*Algorithm 2: Exhaustive Familiar Stranger Detection (EFSD) of node u*

This algorithm is a first attempt to implement the familiar stranger algorithm. In this paper we don't address the (crucial) issue of its complexity, though a more efficient one can replace it. We focus on the fact that the algorithm must be able to find all familiar strangers of a given node.

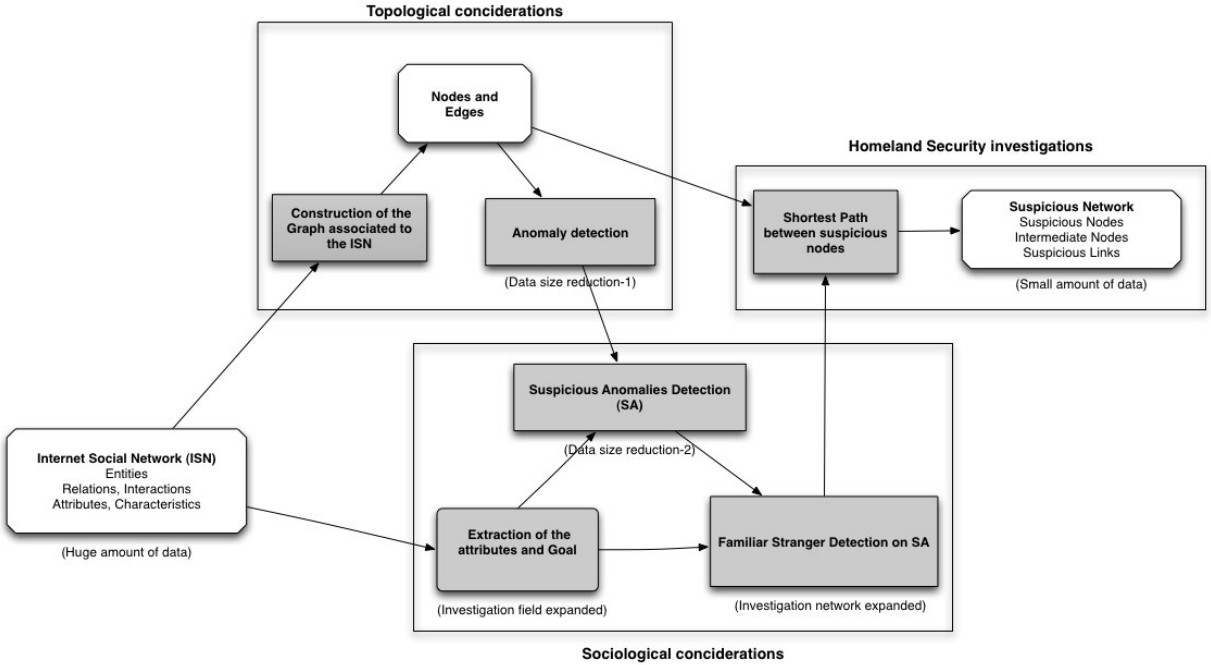



*Figure 2: Suspicious network detection framework*

# 4 Combined approach to detect some suspicious networks in graphs

The last decade has seen the emergence of new forms of threats that use the Internet (especially on exchange platforms) as a support. It is a further proof that social networks are full of information on the preparation and occurrence of malicious activities. Internet has been and remains the starting point for the work of investigators in suicide attacks, collective suicides, sniper's actions, etc. ([55], [56]). One of the main challenges faced by social networks' investigators is the fact that Internet provides huge amount of multivariate data. It is then important to perform a first exploration to exclude parts without much interest. Figure 2 presents an overview of the proposed framework for detecting suspicious networks in Internet social networks. This framework combines sociological assumptions with topological aspects of graphs. The stage concerning the topological aspects is dedicated to the extraction of a graph representing the social network and the application of the singular value decomposition for spectral-based anomaly detection. On the other side, depending on the expected result, one will fixe a Goal (here a set of some critical attributes related to suspicious activities) and inspect the detected set of anomalies to retain nodes that have any relation with suspicious activities. Then we use familiar stranger detection on these suspicious anomalies to find possible nodes belonging to the same subnetwork. Finally we propose to apply a shortest path algorithm to extract what we call suspicious subgraph.

## 4.1 The algorithm

Our key assumption is that all suspicious behavior of a given individual in a social network should appear as an abnormal node. In other words it is almost impossible for an individual to be engaged into illicit activities in a social network without generating atypical patterns. The construction of a good Goal needs a wide range of expertise and a thorough exploration of the set of anomalies. In general the chosen attributes for γ are those that are closely related to the characteristics of an individual with the matching profile. Once we get the anomalies that match the Goal, the algorithm performs the search of the set of the familiar stranger associated with these anomalies. It is observed that in some situations malicious networks have a big average distance between nodes and so persons can be far from each other while belonging to the same cell. One can find a good illustration in [57] which analyzes the network of the 19 Hijackers of September 11th 2001. Moreover, these malicious people are not necessarily directly related. To detect the cells to which they belong one can explore their relationship (interactions) with their neighborhoods. To do this several criteria can be used. In this paper we assume that suspicious nodes are more likely to contact each other through the shortest path. A shortest path algorithm is thus performed between each suspicious node. We propose to apply the well-known Dijkstra shortest path algorithm between each selected nodes ([58], [11]). Algorithm 3 summarizes the steps described above.

---

**Input**:
Graph *G(N,E,A)*,
Goal $\gamma$,
Anomalies threshold *r*,
**Output:**
Suspicious subnetwork *G'(N',E',A')* of *G*,

**1** $A \leftarrow$ result of algorithm SBAD(*G*,*r*)
**2** $A_\gamma \leftarrow$ Set of anomalies respecting the Goal
**3** $N' \leftarrow A_\gamma$
**4 foreach** node $u \in A_\gamma$

**5** $T_u \leftarrow$EFSD(G, $u, \gamma$)
**6** $N' \leftarrow N' \cup T_u$
**7 end**
**8 foreach** couple $(n_1, n_2) \in N' \times N'$
**9** $N' \leftarrow N' \cup Dijkstra(G,n_1,n_2)$ *(Dijkstra return a set of nodes)*
**10 end**
**11 foreach node** $(u, v) \in N' \times N'$

**12** **if** $\{(u,v)\} \in E$ **then**

**13** $E' \leftarrow E' \cup \{(u,v)\}$

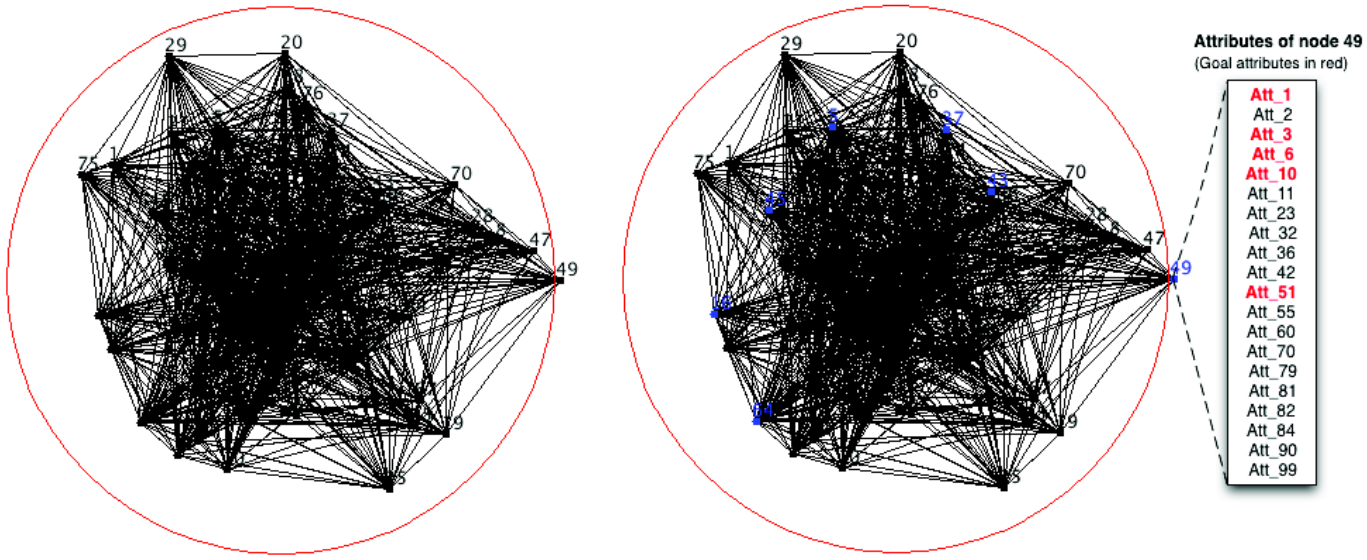
**14 end**

**15 return** *G'(N',E',A')*

*Algorithm 3: Suspicious subnetwork detection*

## 4.2 Application

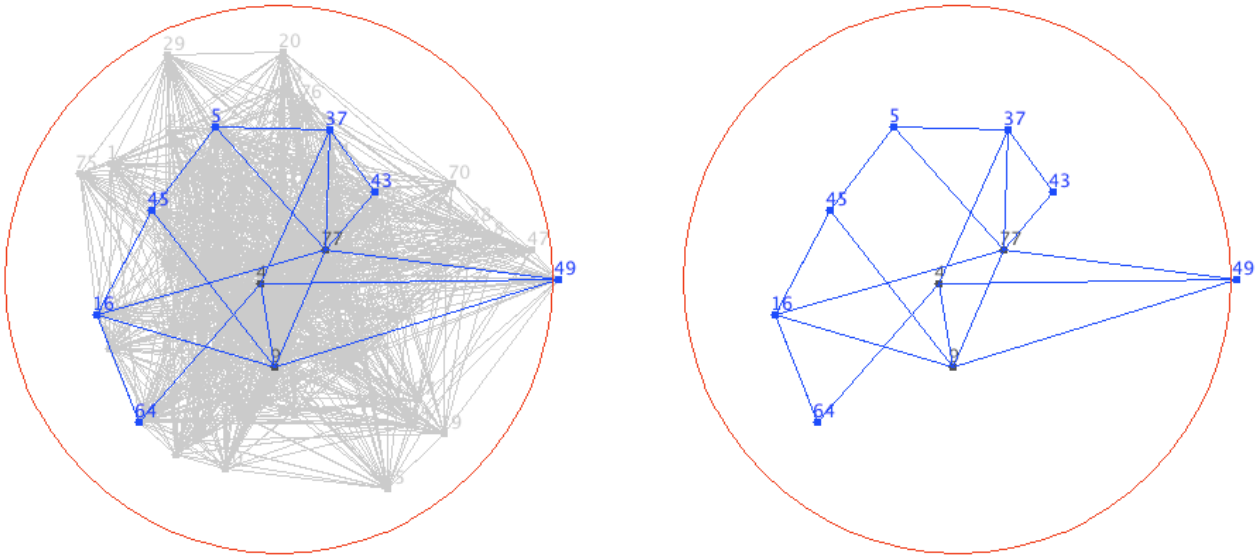
We propose to test our algorithm on a simple graph from a case study of a network of 80 persons. For each person one has collected 20 attributes from a list of 100 attributes. The attributes are chosen randomly but in practice they should result from the analysis of the behavior of persons on the network. For example: social group membership, web site visited, types of posts or comments sent. Data mining algorithms can be powerful tools to perform attributes generation ([59], [60], [61]). The 100 attributes are firstly ranked in a minimum of two groups: likely suspicious and non suspicious. The Goal for the suspicious network detection should thus be a subset of the set of attributes that are ranked as likely suspicious. In our case study $\gamma = \{1,3,10,51\}$. The Algorithm is firstly applied to the data sample, leading to similarity graph ([9]). On this representation the anomalies appear far from the center. By convenience a circle is put in this representation to highlight the position of the nodes. One consider as anomalies all nodes located outside the circle. In our case node 49 appears as an anomaly as depicted in Figure 3-(a).




*(a)* *(b)*

*Figure 3: (a) Similarity graph with node 49 ranked as anomaly, (b) Anomaly and its familiar strangers that match the Goal* $\gamma$

The Figure 3-(b) shows the result of the familiar stranger algorithm applied to the anomaly 49 previously detected after checking if this node matches the Goal. Figures 4-(a) and 4-(b) show the result of the proposed algorithm.




*(a)* *(b)*

*Figure 4: (a) Suspicious network highlighted, (b) Extraction of the suspicious network*

Comparing 3-(b) and 4-(a) one can notice the presence of nodes 4, 9 and 77, which were revealed by the construction of the shortest path between the suspicious nodes. These nodes are not suspicious but can be key persons in the understanding of the cell. They have (sometimes unintentionally) a strategic position in information diffusion between the suspicious persons. In the case of narcotic networks these nodes can be the intermediates (deliverers) between dealers and consumers.

# 5 Conclusion

We have proposed a framework that combines a topological method for anomalies detection with a sociological concept, named familiar stranger, to detect suspicious networks in social networks of the Internet. We applied our approach on a case study to demonstrate that it is possible to extract from a social network a sub-network of nodes ranked as suspicious. The key criterion used is that the node must be an anomaly that matches the Goal or a familiar stranger (with respect to the same Goal) of an anomaly. If no anomaly matches the Goal, the algorithm stops and returns empty suspicious network. However in such a case it can be interesting to review the Goal in order to find more relevant information. This method is original because it takes into account the possibility that suspicious actors

do not share information through direct links but often use intermediates. Although our work did not include the dynamic (time dimension), it is possible to adapt the algorithm using an intelligent learning approach.

# 6 References


[1] MORENO J.L.. *Who Shall Survive: Foundations of Sociometry, Group Psychotherapy, and Sociodrama.* Beacon House 1953.

[2] O'REILLY Tim. *What Is Web 2.0? Design Patterns and Business Models for the Next Generation of Software,* COMMUNICATIONS & STRATEGIES. 2007;65:17-37.

[3] KOSTAKOS Vassilis. *Social networking 2.0* in CHI 08 Human factors in computing systems(New York, NY, USA):3381-3386ACM 2008.

[4] LASSILA Ora, HENDLER James A.. *Embracing "Web 3.0"* IEEE Internet Computing. 2007;11:90-93.

[5] PERER Adam. *Making sense of social networks* in CHI '06: CHI '06 extended abstracts on Human factors in computing systems(New York, NY, USA):1779–1782ACM 2006.

[6] HEER Jeffrey, BOYD Danah. Vizster: *Visualizing Online Social Networks* in IEEE Symposium on Information Visualization(Minneapolis, Minnesota) 2005.

[7] TRIAS Maj Eric D., M.BELL Capt Bryan. *Cyber This, Cyber That . . . So What?* Air & Space Power Journal. 2010.

[8] CHANDOLA Varun, BANERJEE Arindam, KUMAR Vipin. *Anomaly detection*: A survey ACM Comput. Surv.. 2009;41:1–58.

[9] SKILLICORN David B.. *Detecting Anomalies in Graphs*, Proceedings of the IEEE Intelligence and Security Informatics. 2007:209-216.

[10] AGARWAL Nitin, LIU Huan, MURTHY Sudheendra, SEN Arunabha, WANG Xufei. *A social identity approach to identify familiar strangers in a social network* in Proceedings of the 3rd International AAAI Conference of Weblogs and Social 2009.

[11] DREYFUS Stuart E.. *An Appraisal of Some Shortest-Path Algorithms,* Operations Research.1969;17:395–412.

[12] FARLEY Jonathan David. *Breaking al qaeda cells: A mathematical analysis of counterterrorism operations (a guide for risk assessment and decision making)* in Studies in Conflict & Terrorism:2003.

[13] FREESE Ralph. *Automated lattice drawing in Concept Lattices*. Proc. 2nd International Conf. on Formal Concept Analysis:112–127 Springer 2004.

[14] KLERKS Peter, SMEETS Eysink. *The Network Paradigm Applied to Criminal Organizations: Theoretical nitpicking or a relevant doctrine for investigators?* Recent developments in the Netherlands Connections. 2001;24:53–65.

[15] HOFMEYR Steven A., FORREST Stephanie, SOMAYAJI Anil. *Intrusion Detection using Sequences of System Calls,* Journal of Computer Security. 1998;6:151–180.

[16] EBERLE William, HOLDER Lawrence. Discovering *Structural Anomalies in Graph-Based Data*, Data Mining Workshops, International Conference on. 2007;0:393-398.

[17] SHEN Haipeng, HUANG Jianhua Z.. *Analysis of call centre arrival data using singular value decomposition*: Research Articles Appl. Stoch. Model. Bus. Ind.. 2005;21:251–263.

[18] PATTI Jeff, BELOV Nadya, CRAVEN Patrick, THAYER Timothy. *Applying Adaptive Anomaly Detection to Human Networks* 2009.

[19] PATTI Jeff, BELOV Nadya, CRAVEN Patrick, THAYER Timothy. *Applying Adaptive Anomaly Detection to Human Networks in Human Behavior-Computational Modeling* Intelligence Modeling Workshop (Oak Ridge) 2009.

[20] BARBARÁ Daniel, KAMATH Chandrika. , eds.Proceedings of the Third SIAM International Conference on Data Mining, San Francisco, CA, USA, May 1-3, 2003SIAM 2003.

[21] Proceedings of the SIAM International Conference on Data Mining, SDM 2009, April 30 - May 2, 2009,Sparks, Nevada, USASIAM 2009.

[22] GHOSH Joydeep, LAMBERT Diane, SKILLICORN David B., SRIVASTAVA Jaideep. , eds.Proceedings of the Sixth SIAM International Conference on Data Mining, April 20-22, 2006, Bethesda, MD,USASIAM 2006.

[23] JAKKULA V., COOK D. J.. *Anomaly detection using temporal data mining in a smart home environment*. Methods of information in medicine. 2008;47:70–75.

[24] CHANDOLA Varun, BANERJEE Arindam, KUMAR Vipin. Outlier Detection: A Survey 2007.

[25] H0 Shuyuan Mary, TREGLIA Joseph Vincent. *Feasibility Discussion on Identifying Possibility for a National Behavioral Anomaly Detection Platform* in iSociety: Research education engagement(University of north carolina) 2009.

[26] HO L. Lawrence, MACEY Christopher J., HILLER Ronald. *A Distributed and Reliable Platform for Adaptive Anomaly Detection in IP Networks* in DSOM '99: Proceedings of the 10th IFIP/IEEE International Workshop on Distributed Systems: Operations and Management(London, UK):33–46Springer-Verlag 1999.

[27] OVERILL Richard E.. *Computational immunology and anomaly detection,* Information Security Technical Report. 2007;12:188 - 191.

[28] MANIKAS Konstantinos. *Outlier Detection* in Online Gambling Master's thesis, IT University of Goteborg 2008.

[29] WANG Jigang, NESKOVIC Predrag, COOPER Leon N.. *Improving nearest neighbor rule with a simple adaptive distance measure* Pattern Recognition Letters. 2007;28:207 - 213.

[30] MULLIN Matthew, SUKTHANKAR Rahul. *Complete Cross-Validation for Nearest Neighbor Classifiers* in 17th International Conference on Machine Learning (ICML 2000.

[31] RAMASWAMY Sridhar, RASTOGI Rajeev, SHIM Kyuseok. *Efficient algorithms for mining outliers from large data sets* in SIGMOD '00: Proceedings of the 2000 ACM SIGMOD international conference on Management of data(New York, NY, USA):427–438ACM 2000.

[32] ZHANG Ke, HUTTER Marcus, JIN Warren. *A New Local Distance-based Outlier Detection Approach for Scattered Real-World Data* in Proc. 13th Pacific-Asia Conf. on Knowledge Discovery and Data Mining (PAKDD'09);5467 of LNAI(Bangkok):813–822Springer 2009.

[33] SCHAEFFER S.. *Graph clustering* Computer Science Review. 2007;1:27–64.

[34] ESTER Martin, KRIEGEL Hans, SANDER Jörg, XU Xiaowei. A *density-based algorithm for discovering clusters in large spatial databases with noise* :226–231AAAI Press 1996.

[35] ERTZ Levent, STEINBACH Michael, KUMAR Vipin. Finding *Topics in Collections of Documents: A Shared Nearest Neighbor Approach* in In Proceedings of Text Mine 01, First SIAM International Conference on Data Mining 2001.

[36] MÜNZ Gerhard, LI Sa, CARLE Georg. *Traffic Anomaly Detection Using K-Means Clustering* in Proceedings of Leistungs-, Zuverlässigkeits- und Verlässlichkeitsbewertung von Kommunikationsnetzen und Verteilten Systemen, 4. GI/ITG-Workshop MMBnet 2007(Hamburg, Germany) 2007.

[37] CHAN Philip K., MAHONEY Matthew V., ARSHAD Muhammad H.. *A machine learning approach to anomaly detection* tech. rep. 2003.

[38] JAING M. F., TSENG S. S., SU C. M.. *Two-phase clustering process for outliers detection* Pattern Recogn. Lett.. 2001;22:691–700.

[39] VILLEGAS Osslan Osiris Vergara, ELIAS Raul Pinto, SANCHEZ Vianey Guadalupe Cruz. *Singular value decomposition image compression system for automatic object recognition* in ACST'06: Proceedings of the 2nd IASTED international conference on Advances in computer science and technology (Anaheim, CA, USA):95–100ACTA Press 2006.

[40] ZUMSTEIN Philipp. *Comparison of Spectral Methods Throught the Adjacency Matrix and the Laplacian of a Graphs.* PhD thesisETH Zürich 2005.

[41] SUN Jimeng, XIE Yinglian, ZHANG Hui, FALOUTSOS Christos. *Less is more: Sparse Graph Mining with Compact Matrix Decomposition Statistical Analysis and Data Mining.* 2008;1:6–22.

[42] SUN Jimeng, XIE Yinglian, ZHANG Hui, FALOUTSOS Christos. *Less is more: Compact matrix decomposition for large sparse graphs* in In Proc. SIAM Intl. Conf. Data Mining 2007.

[43] TONG Hanghang, PAPADIMITRIOU Spiros, SUN Jimeng, YU Philip S., FALOUTSOS Christos. *Colibri: fast mining of large static and dynamic graphs*. in KDD (Li YING, Liu BING, Sarawagi SUNITA., eds.): 686-694ACM 2008.

[44] DRINEAS Petros, KANNAN Ravi, MAHONEY Michael W. *Fast Monte Carlo algorithms for matrices III: Computing a compressed approximate matrix decomposition.* SIAM Journal on Computing.2004; 36:2006.

[45] SKILLICORN David B. *Finding Unusual Correlation Using Matrix Decompositions* in ISI: 83-99 2004.

[46] MILGRAM S. *The Familiar Stranger: An aspect of the urban anonymity Newsletter.* 1972; Division 8.

[47] LAWRENCE Jamie, PAYNE Terry. *Exploiting Familiar Strangers: Creating a Community Content Distribution Network by Co-Located Individuals in 1st Workshop on Friend of a Friend*, Social Networking and the Semantic Web 2004.

[48] NICOLAI Tom, YONEKI Eiko, BEHRENS Nils, KENN Holger. *Exploring Social Context with the Wireless Rope* in OTM Workshops (1):874-883 2006.

[49] PAULOS Eric, GOODMAN Elizabeth. *The familiar stranger: anxiety, comfort, and play in public places in CHI '04*: Proceedings of the SIGCHI conference on Human factors in computing systems (New York, NY, USA):223–230ACM 2004.

[50] AGARWAL Nitin, LIU Huan, SALERNO John, YU Philip. *Searching for Familiar Strangers on Blogosphere: Problems and Challenges* in NGDM. 2007.

[51] AGARWAL N., LIU H., SALERNO J. J., SUNDARAJAN S. *Understanding group interaction in blogosphere: A case study in 2nd International Conference on Computational Cultural Dynamics*. Washington D.C. 2008.

[52] AGARWAL Nitin. *Social Computing in Blogosphere.* PhD thesis. Arizona State University 2009.

[53] HU Daning, KAZA Siddharth, CHEN Hsinchun. *Identifying significant facilitators of dark network evolution* J. Am. Soc. Inf. Sci. Technol.. 2009;60:655–665.

[54] KAZA Siddharth, HU Daning, ATABAKHSH Homa, CHEN Hsinchun. *Predicting criminal relationships using multivariate survival analysis in dg.o '07*: Proceedings of the 8th annual international conference on Digital government research:290–291Digital Government Society of North America 2007.

[55] KREBS Valdis E.. Mapping Networks of Terrorist Cells 2001.

[56] Mc ILWAIN Jeffrey Scott. Organized crime: *A social network approach Crime, Law and Social Change*.1999;32:301-323.

[57] KREBS Valdis E. *Uncloaking Terrorist Networks First*. 2002;7.

[58] DIJKSTRA E. W. *A note on two problems in connexion with graphs.* Numerische Mathematik.1959;1:269–271.

[59] TAN Ah-hwee. Text Mining: *The state of the art and the challenges*. In Proceedings of the PAKDD 1999 Workshop on Knowledge Disocovery from Advanced Databases:65–70 1999.

[60] FELDMAN Ronen, SANGER James. The Text Mining Handbook: *Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press. 2006.

[61] RUBINSTEIN Ira, LEE Ronald D., SCHWARTZ Paul M., Data Mining and Internet Profiling*: Emerging Regulatory and Technological Approaches*. University of Chicago Law Review. 2008;75:261.