The Application of Social Network Analysis Algorithms in a System Supporting Money Laundering Detection

Rafał Dreżewski[a], Jan Sepielak[a], Wojciech Filipkowski[b]

[a] *AGH University of Science and Technology, Department of Computer Science, Kraków, Poland*
[b] *University of Białystok, Faculty of Law, Białystok, Poland*

Abstract

Criminal analysis is a very complex task requiring to process huge amounts of data coming from different sources such as billings and bank account transactions in order to gain knowledge useful for an investigator. In order to support human analytic capabilities, dedicated software tools are needed, and therefore Money Laundering Detection System (MLDS) was proposed as one of such tools in our previous paper. In this paper, the social network analysis component for this system is presented. The component makes it possible to use data from bank statements and the National Court Register and construct and analyze social networks during an investigation into money laundering cases. The system can assign roles to persons from the network and allows for analysis of connections between them. The paper also includes results of experiments aimed at investigating the performance of the implemented algorithms and the correctness of the analysis.

Keywords: Software supporting criminal intelligence analysis, social networks analysis, money laundering

1. Introduction

Offenders often create sophisticated organizational structures. To fight against them it is necessary to identify an entire network and detect roles of

Email address: drezew@agh.edu.pl (Rafał Dreżewski)

the members. Social Network Analysis (SNA) allows for such an investigation—it has been applied to many domains such as anthropology, biology, economics, geography, history, information science, social psychology, blogosphere analysis, and it can also be successfully utilized in the detection of criminal activities [26, 7, 21, 17].

Money Laundering Detection System (MLDS) (presented in [12]) has been enhanced with SNA algorithms in order to investigate networks consisting of holders of bank accounts and entities from National Court Register. MLDS is a module of the CAST/LINK system supporting the work of a police analyst, developed at the AGH University of Science and Technology in cooperation with the Polish National Police [9, 28].

The rest of this paper is organized as follows. Section 2 describes a legal point of view of data acquisition and utilization during the money laundering detection. Section 3 contains a short review of systems that incorporate social network analysis to detect criminal activities. In Section 4, the SNA component of the MLDS is presented. Section 5 is focused on results and experiments. Finally, the study is recapitulated in Section 6.

2. Using Different Sources of Information During Money Laundering Detection—Legal Point of View

2.1. Introductory problems

We may ask a question as to how public registers can be used in a system aimed at countering money laundering. One of the objectives of criminals perpetrating money laundering is to conceal the identify of persons involved in these illegal practices and of those who eventually control assets which come from the proceeds of crime [6, p. 83]. This is achieved by using various financial instruments, including bearer ones, as well as by hiring so-called "mules" who sell their identity for the purpose of conducting financial transactions, or to be used in official registers of business entities. We must realize, however, that the organizers of money laundering want to control their assets involved in the crime. This is why their names may appear in such registers not as the names of persons who are authorized to perform primary operations in a business entity, but rather as agents or proxies. In principle, such registers are intended for ensuring the aforementioned transparency of business relations. At the same time, if properly used by law enforcement agencies and criminal justice, they can constitute an important element of the fight against money laundering.

The objective of criminal intelligence analysis in this area is to disclose and determine the intensity of personal or capital ties between natural and legal persons [4, pp. 213–214]. The nature of such links can be both formal and informal. Formal ties can easily be verified by means of appropriate documents and public registers. Informal ties are harder to detect and prove. Their presence can be manifested in behavior and decisions made by both natural and legal persons. The identification of their presence serves detective purposes (it may enable the police to define further steps to be taken), and later it may be used in criminal procedures. However, we may not assume that the presence of the economic or private relation always means that the persons involved participate in criminal activities. The identification of the presence of such relations must be verified by competent state bodies with regard to criminal responsibility [15, pp. 106–107].

From the point of view of a criminal intelligence analyst, a register such as the Polish KRS is a source of information that is characterized by full public access to the data it contains [2, pp. 38–39]. Another feature that distinguishes it from, for example, the Internet is that it is run by state bodies. Consequently, information stored therein is credible. Entries in this register are made according to a procedure regulated by the national laws which include, among others, at least partial verification of data and documents, and ends with a decision made by a state body (e.g. a court or a competent office).

Analysis of social networks indicates that there are various ties between their nodes, for example, natural persons or legal persons [24, pp. 3ff.]. These ties can be identified on the basis of various data, like telephone calls or financial transactions. The analysis based on multiple variables also enables identification of the role of individual nodes in the entire networks [1]. On the other hand, the analysis of capital and personal ties begins at a selected starting point, for example, the first and last name of a suspect or the data of a company [15, p. 106]. The following questions must be answered during this analysis:

1. In which other companies do shareholders own shares?
2. In which other companies do members of the management board hold management positions?
3. In which other companies are certain persons present as, for example,

[1] See [22, pp. 74ff] and [30, pp. 56–58]

proxies or agents?
4. Do other business entities have the same registered office address?
5. Is the registered office located in a country that has been put on one of the "black lists" containing countries or territories that offer more advantageous fiscal conditions?
6. Are there companies with similar names, e.g. the ones with only one or several letters, a single word, or the legal or organizational form that are different?

In order to perform such an analysis, it is beneficial to have access to a register that contains relevant data and allows automatic search according to the following criteria:

1. name, registered office, registration data, register number, registered office address;
2. data concerning tax identification and payment of the VAT tax;
3. members of the management board (and the changes over time):
   (a) first name and last name;
   (b) function;
   (c) period when the function was performed;
4. proxies and authorizations granted (if any): first name and last name, period when the function was performed.

2.2. Characteristics of the National Court Register

One of the fundamental tenets of a market economy is its transparency from the point of view of its players. This principle is implemented by a number of institutions acting in the area of public commercial law. Such institutions include state-run registers of business entities, companies, associations, foundations, etc. Registers enable entities conducting business transactions to check, for example, who the owners, shareholders, or managers of other entities are, what links between various entities are, etc. The fact that registers are run by the state guarantees their reliability and is one of the manifestations of the state's active role in the market. In [20, pp. 197ff.] a Polish register operating under the name of the National Court Register (in Polish: Krajowy Rejestr Sądowy—KRS) is presented. It must be noted that national laws in many countries provide for the existence of registers of this type[2]. The differences between them pertain to the scope of

[2]See [16, pp. 25ff.]. Examples of other registers in various countries worldwide can be found in [31]

data collected and the rules for accessing the data.

Another example of a similar register is the British Companies House[3]. It contains a list of all limited companies registered in England, Wales, Northern Ireland, and Scotland. The register was established pursuant to the Companies Act of 2006, but the general system of registration of business entities in Great Britain dates back to 1844. Its basic functions are supervising the process of establishment and liquidation of companies, examining and storing legally required company documents, and providing access to these documents to the public[4]. In 2008, journalists of The Times and the Computer Weekly reported that the register contained approximately four thousand names of company managers put on official lists of persons suspected of fraud, money laundering, corruption, and financing of terrorism (see [25]). The question is why nobody has ever tried to compare records from the Companies House and the list of convicted perpetrators? Also, there have been cases of forgery of documents in the Companies House as well as cases of false identification for legally operating companies (see [27, p. 333]). Therefore, on its website the institution suggests that it is unable to verify the authenticity of all the documents it receives[5]. That simple example shows how data collected in such registers can be used in the fight against criminals.

The legal grounds for the operation of the register in question is the Act of 20 August 1997 on the National Court Register[6]. The register comprises most of all a register of business entities, a register of associations, other social and professional organizations, foundations, independent public health care institutions, and a register of insolvent debtors (Art. 1 (2) of the Act). This scope is very interesting from the point of view of an analyst as it makes it possible to search for ties not only in the private business sector, but also in the non-governmental sector.

The KRS contains the following documents pertaining to corporations[7]

[3]The official website of the Companies House: `http://www.companieshouse.gov.uk/index.shtml`

[4]According to official data, in 2011 the database was searched about 144 million times, see [1, pp. 8–9]

[5]See the official website of the Companies House: `http://www.companieshouse.gov.uk/toolsToHelp/ourServices.shtml`

[6] *The Official Journal of Republic of Poland of 2007*, no. 168, item 1186, consolidated text with amendments

[7]These include limited liability companies, joint-stock companies, partnerships limited

(Art. 8a of the Act):

- deeds of incorporation, memoranda, and articles of association, if they are separate deeds, as well as resolutions on their changes;
- resolutions on changes in the share capital, if they did not require simultaneous changes in the memoranda or the articles of association;
- resolutions on appointment and recall of members of governing bodies of companies;
- annual financial statements and annual consolidated financial statements of capital groups, subject to the laws on accounting;
- copies of resolutions on approval of annual financial statements and profit distribution or loss coverage, as well as opinions of auditors and reports on the operations of units, if an obligation to prepare such a report is set forth in special regulations;
- documents concerning formation and notification of a limited liability company whose articles of association were concluded using the model articles of association for a limited liability company available in the computer system, signed electronically and prepared in the computer system for the purpose of formation of the company.

Art. 38 of the Act defines in detail the scope of data gathered, according to the respective categories of businesses. The following data are recorded for each entity in the register:

- name or legal business name used by the entity;
- identification of its legal form;
- its registered office and address;
- if the entity has branches, then their locations and addresses;
- identification of its former court register number, or business entity register number;

---

by shares, and European companies

- if the entity was formed as a result of transformation or split of another entity, or a merger with other entities, the register must include relevant information about the way the entity was formed and identification of the former numbers in the register;
- information about the conduct of business activities with other entities pursuant to a partnership agreement;
- tax identification number (NIP);
- address of the entity's website and its email-address.

Detailed provisions concerning specific groups of business entities which may be useful in the course of analysis of capital and personal ties or social networks should also be noted. In the case of members of general partnerships, members of European commercial interest groups, members of professional partnerships, shareholders of limited partnerships, and active partners in partnerships limited by shares, the register contains information on marital status, conclusion of a marital property agreement, formation of a separate property regime between spouses, and identification of a limited capacity to perform acts in law, if any. Of course, in the case of partnerships, all partners must be listed. Whenever the Act refers to registration of data pertaining to a natural person, it means the last name, the first name, and the middle name, as well as the personal identification number (so-called PESEL) of that person (Art. 35 of the Act).

The above short presentation describes the potential range and scope of data collected by national registers.

3. Related Research in Social Network Analysis and Money Laundering

Social networks have been studied in order to examine roles and behaviors of nodes [26, 7, 21]. Criminal network analysis has also been a matter of concern over the past decade.

In [32, 8] the authors developed a crime detection systems in which a social network analysis was combined with hierarchical clustering. The clustering methodology partitioned the network into subgroups and then blockmodeling technique made it possible to identify interaction patterns between these subgroups. In this approach, however, detailed roles of offenders were not detected. The systems indicated only key network members.

In [23, 29] the authors applied SNA to destabilize terrorist networks. The aim of the system was to investigate terrorist networks in order to find out how to destabilize them and determine who was capable of carrying out terrorist activity. The approaches proposed in this paper focus on how to fight terrorist organizations. A phenomenon of money laundering is taken casually into account. Therefore there is a lack of dedicated methodology focused on this type of crime.

The actor level SNA was applied in [10, 29, 11, 8] to help detect roles of individuals who can be potentially involved in criminal activity. SNA allowed to identify key members of criminal groups. These papers do not propose an algorithm and parameter study which can be leveraged on the money laundering detection to reveal detailed roles of offenders.

This paper is fully focused on the application of SNA to money laundering detection. The authors propose an algorithm assigning roles to offenders in social networks. A selection of parameter values required by the algorithm was also described. The contributions of this work are as follows:

1. The application of advanced social network analysis to support detection of money laundering processes.
2. The proposal of intervals for roles which are utilized by the role finding algorithm.

4. Social Network Analysis Component of the Money Laundering Detection System

4.1. Architecture of the System

Money Laundering Detection System (MLDS) is part of the system supporting the police analysts (CAST/LINK). It is being developed at the AGH University of Science and Technology in cooperation with the Polish National Police.

The MLDS system is being built to facilitate the analysis of financial flows in order to fight money laundering processes.

Figure 1 presents the main components of the system architecture. There are the following modules:

- *Import module* provides data in format accepted by the system. The data come from disk files and the Web.
- *Clustering module* builds clusters from data provided by the importer using clustering algorithms.

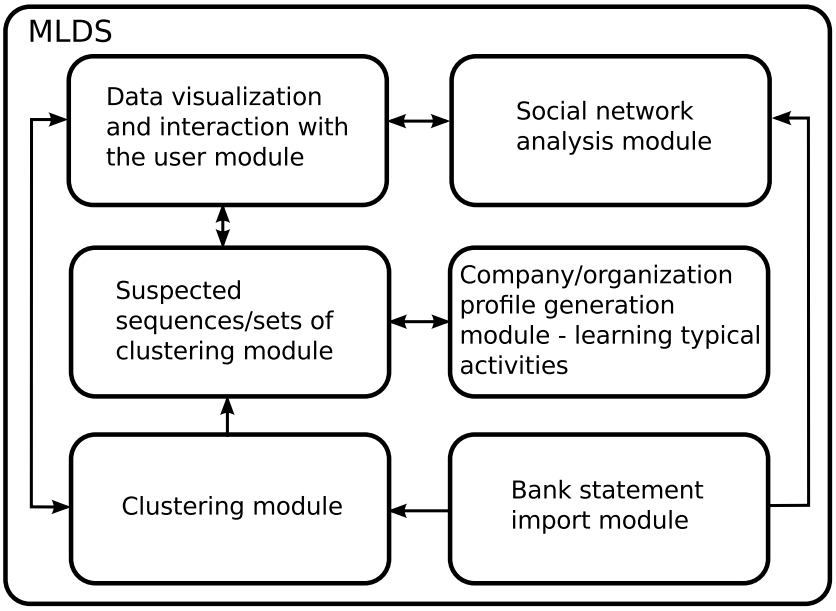

Figure 1: Architecture of the system

- *Suspected sequences/sets of clustering module* perform frequent pattern mining on data provided by the clustering algorithms.
- *Company/organization profile generation module* generates profiles of suspected companies/organizations based on their typical activities, including titles of money transfers, time/date of making money transfers, amount of money usually transferred, etc.—this module of the system is currently under development.
- *Social network analysis module* builds social networks from the imported data and then carries out analyzes of the network such as:
  - assigning roles to nodes,
  - analyzing connections between roles,
  - looking for proximity of entities,
  - comparing roles assigned to nodes in networks from different domains (e.g. bank statements, National Court Register).
- *Data visualization and interaction with the user module* visualizes the resultant data on schema and time-line diagrams.

In [12] all of the system modules were presented in details expect the social network analysis which is a topic of this paper.

The system is still actively developed and, among others, there are also plans for applying it to monitoring of online transactions and detecting of money laundering threats. In such a case we would use agent-based architecture integrated with bio-inspired artificial intelligence techniques [13, 14].

4.2. Methods of social network analysis

Evaluation of node centrality in a social network was proposed in [5] as one of the first analysis manner. Centrality measures who is the most important person in a network. Four kinds of basic measure categories can be distinguished ([19]): *degree centrality*, *betweenness centrality*, *closeness centrality* and *eigenvector centrality*.

Let us consider the graph $G = (V, E)$, where V is a set of vertexes (nodes) in the graph G and E is a set of edges in the graph G.

The system contains the following algorithms:

- *Authoritativeness* calculates the authorities. The authoritativeness of a vertex is the degree to which a vertex is pointed to by important hubs.
- *Betweenness centrality* is calculated on the basis of all shortest paths to all vertexes.

  For $G = (V, E)$ with n vertexes *betweenness centrality* $C_B(v)$ for the vertex v is formulated by:

$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}} \tag{1}$$

  where:

  - σ_{st} is the number of shortest paths from s to t;
  - $\sigma_{st}(v)$ is the number of shortest paths from s to t containing the vertex v.

- *Closeness centrality* determines how close a given vertex is to other vertexes.

For the graph $G = (V, E)$ *closeness centrality* $C_C(v)$ for the vertex v is formulated by:

$$C_C(v) = \frac{1}{\sum_{u \neq v \in V} g_{vu}} \tag{2}$$

where:

- g_{vu} is the length of the shortest path from v to u.

- *Hubness* calculates the hubness of a vertex, which is the degree to which a vertex is linked to other important authorities.

- *Degree centrality* is the number of direct connections a vertex contains.

- *Page Rank* measures the importance of a vertex in terms of the fraction of time spent at this vertex relative to all other vertexes. The graph is transformed to Markov chain and the probability of going from one vertex to another is computed.

4.3. Roles in criminal organization

John Arquilla and David Ronfeldt in [3, pp. 82–84] proposed a few roles of offenders:

- *Organizers* are individuals who constitute the core of an entire organization. Sometimes, small groups which control activity of the network.

- *Insulators* are single individuals and groups which isolate the core of the organization from infiltration. They are responsible for transmitting information from the core to the periphery of the network. Moreover, they control whether information from the periphery of the network does not compromise the core from exclusion from the network.

- *Communicators* are single individuals who control the flow of information between two nodes of the network. They are responsible for transmitting information from the core and provide feedback. In some cases, insulators and communicators will be at odds because of competing impulses inherent in their differing responsibilities.

- *Guardians* are focused on security of the network and minimization of vulnerability to external attack or infiltration. They control who is recruited and assess loyalty to the network. They prevent an individual from deserting and when it takes place, they minimize the attendant risk.

- *Extenders* deal with the extension of the network by recruiting new members and merging other networks. They encourage individuals (e.g. lawyers, policemen, politicians) from other networks to cooperate. When they are successful, the network gains new information sources. Extenders use various means such as bribery, corruption, coercion, and pressure. Their targets are generally powerful persons who are able to provide a high degree of protection.

- *Monitors* are individuals who are responsible for effectiveness of the network. They report weaknesses to the organizers who can take remedial steps, and are responsible for improvement of the network functioning. Their contribution to the network makes it adjustable to new circumstances and flexible.

- *Crossovers* are individuals recruited to the network, but they still also operate in different areas (legal, governmental, financial, or commercial institutions). Such individuals may provide additional information and protection.

The above roles describe individuals who exist within the network for a long time. They fulfill roles which require confidence in other members. This list does not contain the roles of those individuals who appear occasionally in the network or do not play any important function. By reason of this other four rules have to be proposed:

- *Soldiers* do not fulfill any important function in the network. They perform commands coming from other individuals in the network, standing higher in the hierarchy. Their superiors may be communicators, crossovers and even insulators. They can deal with basic, not sophisticated activities of the group such as car theft, drug dealing, extortion, fighting against members of other competing gangs.

- *Recruits* are new individuals in the network. They work in it, but may not be recognized as its members by the network. New individuals can

be recruited to different positions. Their most characteristic feature is a short presence in the network.

- *Outsiders* are individuals who belong to the social network but do not belong to the criminal network, e.g. families of offenders. Outsiders can become active members of the criminal group over time, but they may not even be aware of the existence of the group.

- *Occasional* are individuals with whom the network communicates very rarely. These individuals do not belong to the network. It may also happen that the network often communicates with the occasional, nevertheless they do not identify themselves with the network.

The above set of roles is standard and can be found in every criminal organization. Each of these roles has individual characteristics which are described in more details in Section 4.4.

4.4. Role finding

In order to assign roles to the nodes in the social network, Algorithm 1 was leveraged. Identifying the role is performed by adding points to particular individuals once they fulfill defined criteria. Each individual gets points on the account related to one of the nine roles presented in Tables 1–9. In the end, the individual is assigned the role which has got the most points.

Algorithm 1 presents how the main role is determined for a node. Function *RoleDetermination* calculates scores for the roles. On input it receives a node from the social network and values representing centrality measures. The values are normalized to interval [0;1] by means of *min-max* algorithm [18, pp. 71–72]. The role with the maximum score is returned. When the score is computed then parameters are compared with intervals. The greater interval length, the fewer points are granted to the role. Points are given only when the parameter is either within or beside the interval. To evaluate the former, the interval is enlarged by the value of $BESIDE\_TOLERANCE$ parameter in both directions. If the new interval encloses the algorithm's value then the parameter is beside the original interval.

Tables 1–9 present predefined role set which can be applied in the money transfer domain.

The intervals were established on the basis of characteristics of the roles. They were not verified by real world data since the authors did not possess them.

Algorithm 1: Schema of the role determining algorithm

```
1  RoleDetermination(N, CV, RS, R)
   Data: N—node in a social network, CV—normalized values from algorithms
         which measure centrality, RS—role set
   Result: the determined role for the node, R
2  maxScore = negative_infinity;
3  foreach r in roles do
4  |   score = CalculateScore(r, CV);
5  |   if score > maxScore then
6  |   |   maxScore = score;
7  |   |   R = role;
8  |   end
9  end
10 return R;
11 CalculateScore(R, CV)
   Data: R—role, CV—normalized values from algorithms which measure centrality
   Result: S—score for the role
12 POINTS = 5.0;
13 DEFAULT_LENGTH = 0.2;
14 BESIDE_TOLERANCE = 0.2;
15 S = 0;
16 foreach i in role intervals do
17 |   v = get value for interval's algorithm(CV);
18 |   if v is within i then
19 |   |   S = S + POINTS/(length(i)/DEFAULT_LENGTH);
20 |   end
21 |   else if v is beside i then
22 |   |   S = S + POINTS/((length(i) +
   |   |   BESIDE_TOLERANCE)/DEFAULT_LENGTH);
23 |   end
24 |   return S;
25 end
```

Table 1: Default intervals for the organizer role

Algorithm	Interval	Interval Description
ClosenessC	[0.0, 0.2]	low since the nodes of this role are separated by the insulators
BetweennessC	[0.0, 0.2]	low since the nodes of this role do not mediate between other nodes
PageRank	(0.8, 1.0]	high since the nodes of this role are core of the network
Degree	[0.0, 0.2]	low since the nodes of this role are separated by the insulators
Authoritativeness	(0.8, 1.0]	high since the nodes of this role are the leaders of the network
Hubness	[0.0, 0.2]	low since the nodes of this role do not point out other nodes with authority

Table 2: Default intervals for the insulator role

Algorithm	Interval	Interval Description
ClosenessC	(0.6, 0.8]	moderately high since the nodes of this role protect the core of the organization
BetweennessC	(0.6, 0.8]	moderately high since the nodes of this role mediate between the organizers and other nodes
PageRank	(0.6, 0.8]	moderately high since the nodes of this role separate the organizers
Degree	(0.8, 1.0]	high since the nodes of this role are often related to other nodes
Authoritativeness	[0.0, 0.2]	low since the nodes of this role are rarely pointed out by hubs
Hubness	(0.8, 1.0]	high since the nodes of this role point out the organizers

Table 3: Default intervals for the communicator role

Algorithm	Interval	Interval Description
ClosenessC	(0.6, 0.8]	moderately high since the nodes of this role are often related to a large number of other nodes
BetweennessC	(0.6, 0.8]	moderately high since the nodes of this role mediate between other nodes
PageRank	(0.6, 0.8]	moderately high since the nodes of this role can point out the organizers
Degree	(0.8, 1.0]	high since the nodes of this role can be related to other nodes
Authoritativeness	[0.0, 0.2]	low since the nodes of this role are rarely pointed out by hubs
Hubness	(0.4, 0.6]	moderate since the organizers can be pointed out by the nodes of this role

Table 4: Default intervals for the guardian role

Algorithm	Interval	Interval Description
ClosenessC	[0.0, 0.2]	low since the nodes of this role work on peripherals
BetweennessC	(0.4, 0.6]	moderate since the nodes of this role can mediate between other nodes
PageRank	[0.0, 0.2]	low since the nodes of this role rarely point out important nodes
Degree	(0.2, 0.4]	moderately low since the nodes of this role are rarely related to other nodes
Authoritativeness	[0.0, 0.2]	low since the nodes of this role are not pointed by important hubs
Hubness	(0.2, 0.4]	moderately low since the nodes of this role rarely point out the nodes with authority

Table 5: Default intervals for the extender role

Algorithm	Interval	Interval Description
ClosenessC	(0.4, 0.6]	moderate since the nodes of this role recruit new members or look for cooperating individuals in other networks
BetweennessC	(0.4, 0.6]	moderate since the nodes of this role can mediate between other nodes
PageRank	(0.2, 0.4]	moderately low since the nodes of this role rarely point out important nodes
Degree	(0.4, 0.6]	moderate since the nodes of this role are related to other nodes in order to encourage new individuals
Authoritativeness	[0.0, 0.2]	low since the nodes of this role are not pointed by important hubs
Hubness	(0.2, 0.4]	moderately low since the nodes of this role rarely point out the nodes with authority

Table 6: Default intervals for the monitor role

Algorithm	Interval	Interval Description
ClosenessC	(0.8, 1.0]	high since the nodes of this role monitor the entire network
BetweennessC	(0.8, 1.0]	high since the nodes of this role interact with many other nodes
PageRank	(0.2, 0.4]	moderately low since the nodes of this role rarely point out important nodes
Degree	(0.4, 0.6]	moderate since the nodes of this role are related to other nodes in order to find weaknesses of the network
Authoritativeness	(0.2, 0.4]	moderately low since the nodes of this role are not pointed out by many important hubs
Hubness	[0.0, 0.2]	moderate since the nodes of this role do not point out many nodes with authority

Table 7: Default intervals for the crossover role

Algorithm	Interval	Interval Description
ClosenessC	[0.0, 0.2]	low since the nodes of this role can work outside the network
BetweennessC	[0.0, 0.2]	low since the nodes of this role do not mediate between many nodes
PageRank	(0.2, 0.4]	moderately low since the nodes of this role rarely point out important nodes
Degree	(0.2, 0.4]	moderately low since the nodes of this role are rarely related to other nodes
Authoritativeness	(0.4, 0.6]	moderate since the nodes of this role are not pointed out by hubs
Hubness	[0.0, 0.2]	low since the nodes of this role do not point nodes with authority

Table 8: Default intervals for the soldier role

Algorithm	Interval	Interval Description
ClosenessC	[0.0, 0.2]	low since the nodes of this role work on peripherals
BetweennessC	(0.2, 0.4]	moderately low since the nodes of this role do not mediate between other nodes
PageRank	[0.0, 0.2]	low since the nodes of this role do not point out important nodes
Degree	(0.2, 0.4]	moderately low since the nodes of this role are rarely related to other nodes
Authoritativeness	[0.0, 0.2]	low since the nodes of this role are not pointed out by hubs
Hubness	[0.0, 0.2]	low since the nodes of this role do not point out nodes with authority

Table 9: Default intervals for "other" (recruit, outsider, occasional) role

Algorithm	Interval	Interval Description
ClosenessC	[0.0, 0.2]	low since the nodes of this role work on peripherals
BetweennessC	[0.0, 0.2]	low since the nodes of this role do not mediate between other nodes
PageRank	[0.0, 0.2]	low since the nodes of this role do not point out important nodes
Degree	[0.0, 0.2]	low since the nodes of this role are rarely related to other nodes
Authoritativeness	[0.0, 0.2]	low since the nodes of this role are not pointed out by hubs
Hubness	[0.0, 0.2]	low since the nodes of this role do not point out nodes with authority

4.5. Analysis of connections between roles

Police analysts usually have knowledge of the structure of connections between the roles. This knowledge allows them to verify whether these roles are correctly assigned and indicates which entities have to be monitored since they may fulfill important roles in an organization.

To analyze the structure of connections between the roles, several algorithms were implemented. Some of them handle the aggregate number of edges and some—non-aggregate. When the number is aggregate then many connections between vertexes are treated as single.

The system currently contains the following algorithms:

- Computing aggregate number of all edges (incoming and outgoing) for vertexes.
- Computing non-aggregate number of all edges (incoming and outgoing) for vertexes.
- Computing aggregate number of outgoing edges from vertexes.
- Computing non-aggregate number of outgoing edges from vertexes.
- Computing percentage values using the following formula:

$$P_{ag} = 100\frac{E_{av}}{E_g} \tag{3}$$

where:

  - E_{av} is the aggregate number of all edges between vertexes;
  - E_g is the number of all edges in the graph.

- Computing percentage values using the following formula:

$$P_{nag} = 100\frac{E_{nav}}{E_g} \tag{4}$$

  where:

  - E_{nav} is the non-aggregate number of all edges between vertexes;
  - E_g is the number of all edges in the graph.

- Computing percentage values using the following formula:

$$P_{oa} = 100\frac{E_{ov}}{E_{av}} \tag{5}$$

  where:

  - E_{ov} is the aggregate number of outgoing edges from a vertex;
  - E_{av} is the aggregate number of all edges between vertexes.

- Computing percentage values using the following formula:

$$P_{nona} = 100\frac{E_{nov}}{E_{nav}} \tag{6}$$

  where:

  - E_{nov} is the non-aggregate number of outgoing edges from a vertex;
  - E_{nav} is the non-aggregate number of all edges between vertexes.

The algorithms, for each node, compute statistical data concerning the roles of neighbors. By including this knowledge, the system can assign roles to the nodes with higher precision, and advice which nodes should be monitored in more details, as they may fulfill important roles in the organization.

Table 10: Exemplary values from the data set with role comparison

Node Label	Roles (Bank Statement)	Roles (KRS)
Ryszard Paszka	Crossover	Crossover
Alicja Osysko	Guardian	Guardian
Janusz Motas	Guardian	Guardian
Remigiusz Melson	Insulator	Crossover
Jan Lichowski	Monitor	Crossover

4.6. Proximity of entities

Police analysts frequently have to find out which bank accounts belong to the same persons. Offenders often use more than one bank account to hide criminal activities and mislead investigators.

When holders of bank accounts represent nodes in a social network, then two kinds of vectors identifying the holders can be obtained:

- vector of values coming from algorithms calculating centrality measures,
- vector of values coming from algorithms analyzing connections between the roles.

The cosine of the angle between the vectors describing different nodes is calculated. The closer the result to 1.0, the greater the similarity of holders. When the proximity value of two nodes is exceptionally high, then it may mean that two holders in the network represent the same entity.

4.7. Role comparison

Once the roles for entities from National Court Register and bank statements are found, then there can be the necessity of comparing the results. Role comparison allows the user to easily find out which entities from different domains have the same roles in a social network. This kind of operation makes it possible to verify whether the roles found can be matched.

Table 10 presents a comparison between the roles from a bank statement data set and the KRS data set. Only the first three nodes have the same roles.

These results show that Ryszard Paszka, Alicja Osysko and Janusz Motas have correct matches because their roles are the same in the two domains (bank statements and KRS). Remigiusz Melson and Jan Lichowski have different roles thus disproving their correctness.

5. Results of Experiments

5.1. Constants used in the Role Finding Algorithm

The algorithm 1 contains three constants: $POINTS$, $BESIDE\_TOLERANCE$, $DEFAULT\_LENGTH$. Values of these constants were established during consecutive experiments. During these experiments, firstly, the graphs and roles settings were generated at random, then the values of constants were changed and finally the algorithm was run.

In order to generate a graph, in the beginning, nodes were generated and then connected with random edges. The graphs always had 20 nodes and 20 edges. Role set had always 8 roles and each role had 4 intervals, which were selected at random. When the roles for nodes were found, variations were calculated for scores of the assigned roles.

The results indicate that the constant values affect the diversity of role scores. When the values of $BESIDE\_TOLERANCE$ and $DEFAULT\_LENGTH$ constants are smaller and smaller then many roles get the same or similar scores and roles assigned by the algorithm are uncertain. If the values of these constants are greater than 1.0 then the results are unpredictable. The proposed values of the constants provide results that are stable enough.

5.2. Functioning of the System

The tests which were carried out were intended to verify the correctness of the system functioning. Verification was done manually. An organizer located on the periphery of the network can be an example of detected errors.

As it was said above, the networks during experiments were generated at random: firstly, nodes were generated and then connected with random edges.

Tables 11 and 12 show a few nodes with the roles assigned to them. The tables also contain values which were utilized to compute scores for the roles.

Figure 2 displays a sample schema diagram created for holders of bank accounts. Colored borders and subtitles indicate the roles the nodes fulfill.

The network in Figure 2 does not contain a distinct organizer. In fact, the role should not have been assigned to this node because an organizer should not be connected with extenders, soldiers or monitors. A node *Huwonolo Kames* has also an uncertain role (soldier) assigned. A soldier should interact mainly with extenders, guardians and monitors. In this case it is also connected with an organizer.

Table 11: Exemplary nodes with assigned roles (part 1)

Node Name	Role Name	Role Score	Close-nessC	Between-nessC
Cilygade	Communicator	20	0.689	1
Mariusz Cyfka	Crossover	22.5	0.545	0.021
Anna Jamioł	Extender	20	0.291	0.686
Piotr Mamcarczyk	Guardian	20	0.567	0.474
Jadwiga Pyza	Insulator	20	0.746	0.711
Jarosław Dyjach	Monitor	17.5	0.844	0.121
Marek Morajko	Organizer	20	0.069	0.4
Jarosław Szturo	Other	25	0.385	0
Huwonolo Kames	Soldier	25	0.162	0.389

Table 12: Exemplary nodes with assigned roles (part 2)

Node Name	Degree	Authorita-tiveness	Hubness	Page Rank
Cilygade	0.714	0.619	0.555	0.716
Mariusz Cyfka	0.286	0.555	0.003	0.195
Anna Jamioł	0.571	0.005	0.395	1
Piotr Mamcarczyk	0.143	0.164	0.003	0.146
Jadwiga Pyza	1	0.732	0.798	0.503
Jarosław Dyjach	0.429	0.294	0.865	0.127
Marek Morajko	0.714	0.829	0.01	0.919
Jarosław Szturo	0	0	0.299	0
Huwonolo Kames	0.286	0.008	0.004	0.877

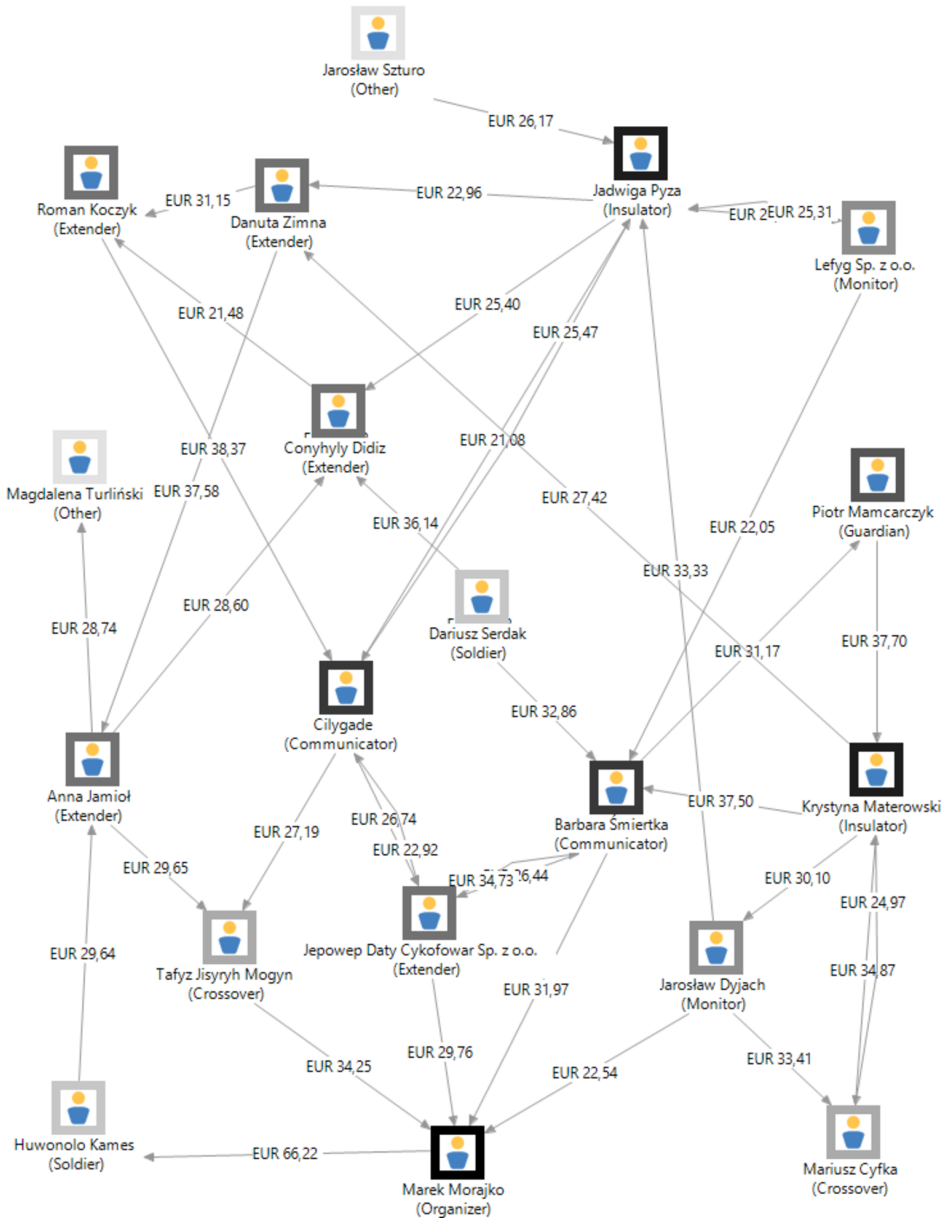

Figure 2: Sample diagram with holders of bank accounts and with marked roles

Cilygade is an example of a node with the role correctly assigned (communicator). It does not appear on the periphery and interacts with nodes such as crossover, extenders and insulator. A node *Anna Jamioł* also has a correct role assigned (extender). It is connected with a crossover, an extender, a soldier and "other".

Table 13: Exemplary values from the analysis of connections between roles

Source Element Name	Marek Morajko	Marek Morajko
Source Role Name	Organizer	Organizer
Target Role Score	Communicator	Crossover
Edge Count (Aggregated)	1	1
Outgoing Edge Count (Aggregated)	0	0
Percent of Aggregated Outgoing Edge Count	0	0
Percent of All Aggregated Edge Count	2.5	2.5
Edge Count (non-Aggregated)	1	1
Outgoing Edge Count (non-Aggregated)	0	0
Percent of non-Aggregated Outgoing Edge Count	0	0
Percent of All non-Aggregated Edge Count	2.5	2.5

Table 13 shows values coming from the analysis of connections between the roles. A node *Marek Morajko* is an organizer connected with one communicator and one crossover node. Both connections are incoming. Rows *Percent of All Aggregate Edge Count* and *Percent of All non-Aggregate Edge Count* have value of 2.5, as the network consists of 40 edges.

Table 14 shows sample entities with proximity values. The third and the fourth column contain the cosine of the angles between the vectors. The vectors for the third column contain centrality values. The vectors for the fourth column come from the analysis of connections between the roles.

5.3. Performance

Using the implemented social network analysis algorithms, a series of experiments was carried out. The aim of these experiments was to compare

Table 14: Exemplary values from the analysis of entity proximity

Element 1	Element 2	Roles	Role Pairs	Average
Nasevi Sp. z o.o.	Marek Dyzio	0.99340	0.99987	0.99663
Antoni Ciupke	Renata Dostal	0.98989	0.99831	0.99410
Scott Cirka	Irena Chabelski	0.98781	1.00000	0.99390
Roman Pasiciel	Hans Cywicki	0.98422	1.00000	0.99211

the role finding algorithms and to analyze the connections between the roles. Graphs for the experiments were generated randomly. In the beginning, a specified number of nodes were generated. Then the nodes were connected, using a required number of edges. Roles and intervals for them were also generated randomly.

All of the experiments were carried out on the machine with AMD Phenom II X4 965 processor.

Work time was measured with the use of the following values of parameters:

- number of roles—11,
- number of nodes in graphs—1000,
- number of edges in graphs—1000.

During the examination of the impact of the number of roles on the algorithms work-time, the graphs were generated, using the following parameters:

- number of nodes—120,
- number of edges—200.

Algorithms were run 10 times for each parameter value coming from the established range, and average results were computed.

On the basis of other publication ([33]) it was assumed that in real cases social networks have about 2500 edges and 7000 nodes. Since the experiments demonstrated that the number of nodes and edges reached those values, the results simulate the performance of the algorithms in real scenarios.

During the experiments the total run-times needed to process the network were composed of work-times of the algorithms computing centrality measure listed in Section 4.2 (for the role finding process) and the algorithms analyzing the structure of connections between the roles listed in Section 4.5.

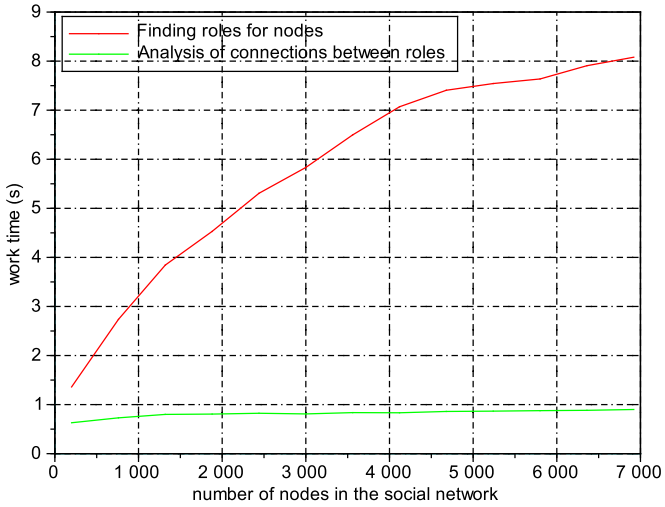

Figure 3: Work-time of the algorithms when the node count changes

Figure 3 presents the work-time of the algorithms when the number of nodes changes. The performance of the role finding algorithm is much worse than the performance of the algorithm analyzing connections between the roles. What is more, the work-time of the former grows faster in conjunction with the node count. When node the count is equal to 200, then the role finding is by 0.73 s. slower than the analysis of connections, but at the value of 6920 it is by 7.18 s. slower.

Figure 4 presents the work-time of the algorithms when the number of edges changes. The algorithm analyzing connections between the roles also outperforms the role finding algorithm. In the beginning (edge count equal to 200), the role finding is by 0.3 s. slower. In the end (edge count equal to 2480), it is by 15.59 s. slower.

Figure 5 shows the work-time of the algorithms when the number of roles changes. This parameter has no impact on the work-time of the algorithm for analyzing connections between the roles, whereas the work-time of the role finding algorithm goes up with an increase in the number of roles.

The experiments indicate that the analysis of connection between the roles is much more scalable than the role finding analysis.

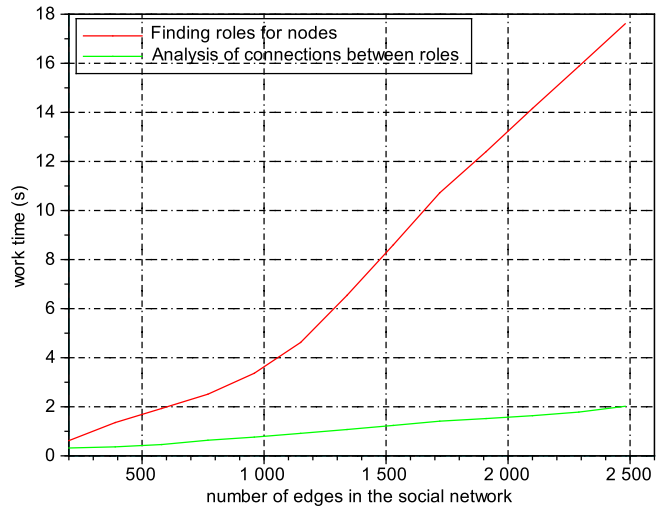

Figure 4: Work-time of the algorithms when the edge count changes

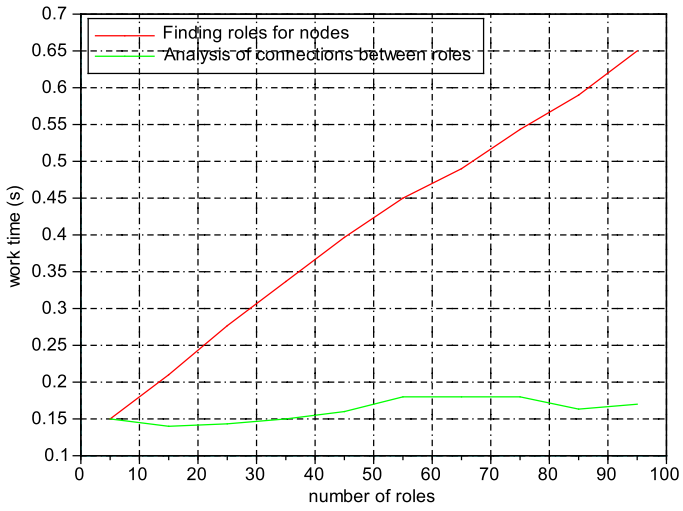

Figure 5: Work-time of the algorithms when the role count changes

6. Conclusions and Future Work

In this paper we explored an approach applying the social network analysis to money laundering detection. The system presented here constructs social networks from the bank statements and National Court Register. The assignment of the roles to persons from the network allows us to reveal true leaders and vulnerabilities inside the network. Having detected the roles of the persons within the network, an analysis of connections between them can be conducted. A node proximity module allows the analyst to detect which bank accounts are possessed by the same persons. The module for finding nodes with the same roles in data sets, coming from different domains (bank statements, National Court Register) allows us to confirm correctness of the assigned roles.

Clustering technique and frequent pattern mining (described in [12]) also allow us to find the roles of persons. SNA is not limited to a few roles but provides the user with a wide range of possibilities of configuration. SNA in conjunction with these techniques is a solid method for detecting suspicious transactions and uncovering groups of offenders.

Performance verification of the algorithms was also carried out. It indicated that the role finding algorithm is always faster than the interconnection analysis. Experiments also showed that the SNA is a feasible and effective technique in money laundering detection domain. At the same time, there are many problems to be further studied, such as on-line money laundering detection, learning from decisions of a human analyst, and an automatic creation of a company/organization model of typical activities.

The above techniques for automated criminal network analysis and visualization allow the identification of interaction patterns of offenders and their roles in criminal groups. To achieve better efficiency, the system does not rely only upon a singe data domain. The input data can come from bank statements and National Court Register. The data are obtained from disk files and Web pages.

Social network analysis is important for us to understand the structure of criminal organization. Also, other advanced techniques are needed to extract knowledge about criminal networks such as data mining, machine learning, and data clustering. Such tools could help a police analysts enhance public safety and national security by developing strategies to prevent organized crime of money laundering.

Acknowledgments

This research was partially supported by a grant "Advanced IT techniques supporting data processing in criminal analysis" (No. 0008/R/ID1/2011/01) from the Polish National Centre for Research and Development, and by Polish Ministry of Science and Higher Education under AGH University of Science and Technology Grant No. 11.11.230.124 (statutory project).

References

[1] Companies House Annual Report and Accounts 2011/12, 2011. `http://tinyurl.com/bpfdwd2`.

[2] Poland. Company Laws and Regulations Handbook, International Business Publications, Washington, 2012.

[3] J. Arquilla, D.F. Ronfeldt (Eds.), Networks and Netwars. The Future of Terror, Crime, and Militancy, RAND Corporation, 2002.

[4] T.E. Baker, Intelligence-led Policing: Leadership, Strategies & Tactics, Looseleaf Law Publications Inc., New York, 2011.

[5] A. Bavelas, Communication patterns in task-oriented groups, Journal of the Acoustical Society of America 22 (1950) 725–30.

[6] M.E. Beare, S. Schneider, Money Laundering in Canada: Chasing Dirty And Dangerous Dollars, University of Toronto Press, Toronto, 2007.

[7] S.P. Borgatti, M.G. Everett, J.C. Johnson, Analyzing Social Networks, SAGE Publications, 2013.

[8] H. Chen, J. Jie, Y. Qin, M. Chau, Crime data mining: A general framework and some examples, IEEE Computer 37 (2004) 50–6.

[9] J. Dajda, R. Dębski, A. Byrski, M. Kisiel-Dorohinicki, Component-based architecture for systems, services and data integration in support for criminal analysis, Journal of Telecommunications and Information Technology (2012) 67–73.

[10] K. Dawoud, R. Alhajj, J. Rokne, A global measure for estimating the degree of organization of terrorist networks, in: N. Memon, R. Alhajj (Eds.), Advances in Social Networks Analysis and Mining (ASONAM), 2010 International Conference on, IEEE Computer Society, Los Alamitos, California, Washington, Tokyo, 2010, pp. 421–7.

[11] M.J. Dombroski, K.M. Carley, Netest: Estimating a terrorist network's structure, Comput. Math. Organ. Theory 8 (2002) 235–41.

[12] R. Dreżewski, J. Sepielak, W. Filipkowski, System supporting money laundering detection, Digital Investigation 9 (2012) 8–21.

[13] R. Dreżewski, J. Sepielak, Evolutionary system for generating investment strategies, in: M. Giacobini, et al. (Ed.), Applications of Evolutionary Computing, EvoWorkshops 2008: EvoCOMNET, EvoFIN, EvoHOT, EvoIASP, EvoMUSART, EvoNUM, EvoSTOC, and EvoTransLog, Naples, Italy, March 26-28, 2008. Proceedings, volume 4974 of *LNCS*, Springer-Verlag, Berlin, Heidelberg, 2008, pp. 83–92.

[14] R. Dreżewski, L. Siwik, Co-evolutionary multi-agent system for portfolio optimization, in: A. Brabazon, M. O'Neill (Eds.), Natural Computing in Computational Finance, volume 1, Springer-Verlag, Berlin, Heidelberg, 2008, pp. 271–99.

[15] J. D'Souza, Terrorist Financing, Money Laundering, and Tax Evasion: Examining the Performance of Financial Intelligence Units, CRC Press, Taylor & Francis Group, Boca Raton, London, New York, 2012.

[16] V. Giannella, Computerised commercial registers in europe and the planned "european business register", in: Computerised registers in the public sector (in civil, penal and administrative law), Proceedings 12th Colloquy on Legal Data Processing in Europe, Ljubljana (Slovenia), 2-4 October 1995, Council of Europe, Strasbourg, 1998, pp. 25–46.

[17] B. Gliwa, J. Kozlak, A. Zygmunt, K. Cetnarowicz, Models of social groups in blogosphere based on information about comment addressees and sentiments, in: K. Aberer, A. Flache, W. Jager, L. Liu, J. Tang, C. Guéret (Eds.), Social Informatics - 4th International Conference, SocInfo 2012, Lausanne, Switzerland, December 5-7, 2012. Proceedings,

volume 7710 of *Lecture Notes in Computer Science*, Springer, 2012, pp. 475–88.

[18] J. Han, M. Kamber, Data Mining: Concept and Techniques, Morgan Kaufmann, 2006.

[19] R. Hanneman, M. Riddle, Introduction to Social Network Methods, University of California Press, 2005.

[20] J. Kielbowicz, A. Kreczmar, Report on the bahalf of the delegation of the polish republic. the polish national computer registers centre (cors), in: Computerised registers in the public sector (in civil, penal and administrative law), Proceedings 12th Colloquy on Legal Data Processing in Europe, Ljubljana (Slovenia), 2-4 October 1995, Council of Europe, Strasbourg, 1998, pp. 197–206.

[21] D. Knoke, S. Yang, Social Network Analysis, SAGE Publications, 2007.

[22] R. Milne, Forensic Intelligence, CRC Press, Taylor & Francis Group, Boca Raton, London, New York, 2013.

[23] M. Nasrullah, Detecting terrorist activity patterns using investigative data mining tool, International Journal of Knowledge and System Science 3 (2005) 43–52.

[24] C.A.R. Pinheiro, Social Network Analysis in Telecommunications, John Wiley & Sons Inc., Hoboken, New Jersey, 2011.

[25] A. Savvas, UK companies house register contains 3,994 high-risk individuals, datanomic finds, 2008. `http://tinyurl.com/cgsxnp7`.

[26] J.G. Scott, Social Network Analysis, SAGE Publications, Los Angeles, London, New Delhi, Singapore, Washington DC, 2012.

[27] L.S. Spedding, The Due Diligence Handbook: Corporate Governance, Risk Management and Business Planning, CIMA Publishing, Oxford, 2009.

[28] A. Świerczek, R. Dębski, P. Włodek, B. Śnieżyński, Integrating applications developed for heterogeneous platforms: Building an environment

for criminal analysts, in: Multimedia Communications, Services and Security. 4th International Conference, MCSS 2011, Krakow, Poland, June 2-3, 2011. Proceedings, Springer Berlin Heidelberg, 2011, pp. 19–27.

[29] M.A. Tayebi, L. Bakker, U. Glasser, V. Dabbaghian, Locating central actors in co-offending networks, in: Advances in Social Networks Analysis and Mining (ASONAM), 2011 International Conference on, IEEE Computer Society, Los Alamitos, California, Washington, Tokyo, 2011, pp. 171–9.

[30] U.S. Congress, Office of Technology Assessment, Information technologies for the control of money laundering, U.S. Government Printing Office, Washington, 1995. OTA-ITC-630.

[31] Wikipedia, List of company registers, 2013. `http://en.wikipedia.org/wiki/List_of_company_registers`.

[32] J.J. Xu, H. Chen, Crimenet explorer: a framework for criminal network knowledge discovery, ACM Trans. Inf. Syst. 23 (2005) 201–26.

[33] A. Zygmunt, P. Bródka, P. Kazienko, J. Kozlak, Key person analysis in social communities within the blogosphere, J. UCS 18 (2012) 577–97.