Paper

# Application of Social Network Analysis to the Investigation of Interpersonal Connections

Marcin Mincer[a] and Ewa Niewiadomska-Szynkiewicz[a,b]

[a] Institute of Control and Computation Engineering, Warsaw University of Technology, Warsaw, Poland
[b] Research and Academic Computer Network (NASK), Warsaw, Poland

**Abstract—Social network analysis (SNA) is an important and valuable tool for knowledge extraction from massive and unstructured data. Social network provides a powerful abstraction of the structure and dynamics of diverse kinds of interpersonal connection and interaction. In this paper, we address issues associated with the application of SNA to the investigation and analysis of social relationships of people. We provide a brief introduction to representation and analysis of social networks, SNA models and methods. The main objective is to investigate the application of SNA techniques to data mining in case of two social networks Facebook and Twitter. The presented simulations illustrate how social analysis can be used to determine the interpersonal connections, importance of actors in a given social network and detect communities of people. We then discuss strength and weakness of SNA techniques.**

***Keywords—centrality measures, communities detection, social network, social network analysis.***

## 1. Introduction

During the last decade social networks (SN) have become extremely popular and have been attracted attention of scientists of different disciplines, such as sociology, epidemiology, economy, computer engineering, telecommunication and many others [1]–[7]. Many systems in nature and technology are examples of social networks, i.e., systems composed of a large number of highly interconnected individuals (actors), whose structure is irregular, complex and dynamically evolving in time. Communication networks, such as the Internet and the World Wide Web, are examples of SN.

A social network is formally defined as a set of actors or social groups, and relationships such as: friendship, collaboration, business, political, etc. The first approach to capture the global properties of such systems is to model them as graphs which nodes represent the actors and links the relationships between them. Nevertheless, most of real world networks are characterized by the similar topological properties, such as relatively small characteristic path lengths, high clustering coefficients, degree correlations, which make them radically different from regular lattices and random graphs. Hence, in many cases the standard models from graph theory cannot be applied, and the dedicated techniques and methods have to be used.

In the beginning, the social network became a field of interest of sociology that did not use mathematical graph theory. It has appeared soon that merging experience of sociology and graph theory needed the dedicated formal social network analysis (SNA) methods.

Social network analysis is a group of graph theory based techniques that can be used to retrieve meaningful knowledge from networks formed by various actors. In the recent past, SNA techniques have been rapidly increasing their advance into a wide variety of applications and systems [2], [3], [8]–[13]. Due to powerful computers, emerging and widely adapted platforms such as Facebook, Twitter, Foursquare, nk.pl, and many others SNA has become commonly used approach to interpersonal connections analysis. Data about relationships of people are commonly available like never before, and applying analytical methods to them became a source of unique and valuable knowledge.

In literature, one can find an extensive survey of state of the art in SNA techniques and methods [1], [7], [14]–[18]. The topological and structural properties of social networks are considered. The major results and concepts in SN, with focus on the fundamental concept, i.e., scale-free and small-world properties, and current approaches to SN analytical analysis and simulation are described and discussed. Numerous books and papers present models demonstrating the main features of evolving networks, network topologies, and summarize software, currently used in the analysis of complex network systems.

The main aim of this paper is to present the application of SNA methods to retrieve meaningful information, from commonly used social media platforms. The goal of presented case studies is to show that SNA can be a strong technique to investigate the interpersonal connections. However, the application of SNA has some limitations and requirements for input data. The remainder of this paper is organized as follows. In Sections 2 and 3, we provide the introduction to SNA techniques. We focus on social networks properties and popular measures in SN. In Section 4, we describe two popular algorithms of communities detection. In Section 5, we present and discuss the results of simulation experiments. Two of them show the effectiveness of application of SNA techniques to data mining, in the case of the social networks Facebook and Twitter. The goal of these experiments was to illustrate how social analysis can be used to determine the social relationships of people. The next test concerned with cliques detection, show limitations of SNA techniques. The paper concludes in Section 6.

# 2. Properties of Social Networks

Various measures are used to classify a network to be the social network. They are commonly used by researchers and commercial users to analyze characteristics of social networks to be considered. The most important measures that come from the graph theory are presented below.

### *2.1. Basic Measures*

**Node degree**. The simple measure for an individual actor in a network is the degree of the corresponding node. From the graph theory the degree $k_i$ of the node $i$ is defined as follows:

$$k_i = \sum_{j=1}^{N} a_{ij} = \sum_{j=1}^{N} a_{ji}, \qquad (1)$$

where $N$ denotes a number of nodes in a network, $a_{ij}$ element of the coincidence matrix $A$ defined as follows: $a_{ij} = 1$ if the nodes $i$ and $j$ are interconnected, $a_{ij} = 0$ otherwise.

**Shortest path**. A critical primitive in large scale graph problems is the estimation of the shortest path – a path between two network nodes in a given network, such as the sum of their weights corresponding to edges is minimized. The average shortest path for the whole network is widely used in SNs to capture characteristic features of these networks. The average shortest path is calculated as follows:

$$l = \frac{1}{N(N-1)} \sum_{i \neq j} d(i,j) \approx \frac{\ln N}{\overline{k}}, \qquad (2)$$

where $N$ denotes a number of nodes in a network, $d(i,j)$ the shortest path between nodes $i$ and $j$, $\overline{k}$ the average degree of nodes calculated according to Eq. (1).

**Clustering coefficient**. Clustering coefficient $G_i$ of the node $i$ is defined as a fraction of existing edges between neighbors of the node $i$, and all edges that are possible between those neighbors. In undirected network, the maximal number of edges is computed as $\frac{k_i(k_i-1)}{2}$, where $k_i$ denotes the degree of the $i$-th node. Clustering coefficient of the node $i$ is computed as follows:

$$G_i = \frac{2|\{e_{jk}\}|}{k_i(k_i-1)}, \qquad (3)$$

where $\{e_{jk}\}$ denotes a set of edges connecting neighbors of the node $i$.
We can calculate average clustering coefficient ($\overline{G}$) in a network:

$$\overline{G} = \frac{1}{N} \sum_{i=1}^{N} G_i, \qquad (4)$$

where $N$ denotes a number of nodes.

### *2.2. Properties of SN*

The common properties of social networks are:

- scale-free networks,
- clusterability,
- small-world networks.

Social networks are conjectured to be *scale-free*. The typical scale-free network consists of a few nodes with high degree, and long tail of nodes with low degree. It is a common structure of most networks encountered in nature that was investigated by R. Albert and A.-L. Barabási and described in [19] and [20]. R. Albert and A.-L. Barabási observed that in case of social networks a degree distribution follows a power law:

$$P(k) \propto k^{-\alpha}, \qquad (5)$$

where $P(k)$ denotes a probability that a degree of randomly selected node will be equal to $k$.
Typical social network consists of a set of communities, grouping strongly connected actors – the value of $\overline{G}$ defined in Eq. (4) is usually high (close to 1). Hence, we can say about high clusterability of SN [7].
The small-world networks were investigated by D. Watts, and described in [18]. It was proved that networks that widely occur in nature, especially communities of people are small-world networks. The typical feature of so-called *small-world* networks is that an average shortest path $l$ defined in Eq. (2) is very small relative to the number of nodes $N$ forming a network. It can be observed that in social networks $l \approx \ln(N)/\overline{k}$.

# 3. Centrality Measures

In many social network applications, the main objective of data analysis is to identify the *most important actors* in a network. We consider a network node (an actor) to be a prominent one, if it is extensively involved in relationships with other nodes that form a social network. Moreover, an importance of a node relies on the number of prominent nodes that are connected to this node. A variety of statistical parameters – centrality measures were designed to show differences in the importance of actors. They are described in details in literature [7], [17]. To calculate these measures direct and indirect, inter-node connections have to be considered. In this section we present definitions of the most noteworthy and popular measures.

### *3.1. Betweenness Centrality*

A betweenness centrality is a very important measure, while considering flows in a network. The large betweenness value means that a given actor is connected with many other actors (directly and indirectly). The betweenness centrality for the $i$-th node is calculated as follows [15]:

$$Cb_i = \sum_{j}^{N}\sum_{k}^{N}\frac{g_{jik}}{g_{jk}}, \qquad i \neq k \neq j, \tag{6}$$

where $g_{jik}$ denotes a number of shortest paths linking nodes $j$ and $k$ passing through the node $i$, $g_{jk}$ a number of paths not including the node $i$.
Usually, $Cb_i$ is normalized to values from $[0,1]$ by multiplying through $\frac{2}{(N-1)(N-2)}$, where $N$ denotes a number of nodes.

### 3.2. Closeness Centrality

A view of a node centrality can be based on closeness or distance. The question is how close is a node to all other nodes in a network. This measure is very important and commonly used in the graph theory. In general, a closeness $Cc_i$ of the node $i$ is defined as the inverse of the sum of distances between the node $i$ and all other nodes in a network:

$$Cc_i = \frac{1}{\sum_{j\neq i} d(i,j)}, \tag{7}$$

where $d(i,j)$ denotes the shortest path between node $i$ and $j$.
This closeness measure can be viewed as a time required to spread information from a given node to all other reachable nodes in a network [15].
Another definition of closeness was proposed by M. E. J. Newman in [5]. $Cc_i$ is defined as the average shortest path from the node $i$ to all other reachable nodes

$$Cc_i = \frac{\sum_{j\neq i} d(i,j)}{N-1}, \tag{8}$$

where $N \geq 2$ denotes a number of nodes in a network.

### 3.3. Eigenvector Centrality

The eigenvector centrality measure highlights the importance of the node $i$ within a social network. The value of this measure relies on a number of other prominent nodes that are linked to the node $i$. The eigenvector centrality corresponds to the network coincidence matrix $A$. According to formula (9), the centrality of the node $i$ is proportional to the sum of centralities of all nodes that are connected to the $i$-th node.

$$Ce_i = \frac{1}{\lambda}\sum_{j=1}^{N} a_{ij} Ce_j, \tag{9}$$

where $Ce_j$ is the eigenvector centrality of the $j$-th node, $N$ is a number of nodes in a network and $\lambda$ is the constant value, $a_{ij}$ an element of the coincidence matrix $A$.

## 4. Community Detection

Social networks are usually formed by smaller subnetworks (communities). It is obvious that community consists of subset of actors (nodes) with dense inter-node connections within this subset. The links to nodes from other communities are less dense. The communities detection, which idea is to divide a network into communities is one of the most interesting and important problem in the investigation, and analysis of networks. It is a challenging task, especially when consider overlapping communities and the dynamics of networks. In social networks, the overlapping is natural, as people usually belong to many communities. Many algorithms of communities detection in complex systems have been developed and described in literature [14], [21], [16], [22]. Two common techniques, i.e., an algorithm developed by M. Girvan, M. E. J. Newman and modified by A. Clauset, and an algorithm proposed by V. D. Blondel *et al.* are described below.

### 4.1. Clauset&Newman Algorithm

The first algorithm of communities detection was developed by Girvan and Newman, and described in [22]. It was improved by Clauset [16]. The idea of the Clauset&Newman algorithm is to identify the edges in a network, which links different communities. This identification is based on the betweenness centrality measure Eq. (6) that is extended to the case of edges. The communities detection is performed in two phases. In the first phase, the betweenness centrality measures are calculated for all edges in the network. Next, the edge with the highest betweenness is identified and removed from the set of edges. A high value of the betweenness centrality is typical to nodes connecting two communities – many shortest paths linking nodes from different communities pass through such edge. In this way, we can split our network into subnetworks. In every iteration, a dendrogram is produced to illustrate how the network splits into communities with the successive removal of edges. The first phase stops when all edges are removed from the set. The final result is the dendrogram that demonstrates the clustering structure of a network. The algorithm switches to the second phase. In the second phase, the calculated dendrogram is analyzed, and a number of communities forming the network is estimated based on the value of a modularity coefficient $Q$. The modularity coefficient $Q$ defined in Eq. (10) is calculated for all splits performed in successive iterations in the first phase, and demonstrated in the dendrogram.

$$Q = \sum_{l=1}^{M} e_{ll} - p_l^2, \qquad p_l = \sum_{m=1}^{M} e_{lm}, \tag{10}$$

where $M$ denotes the number of groups, $e_{lm}$ denotes the fraction of edges linking two groups $l$ and $m$, $e_{ll}$ the fraction of edges linking nodes from the same community $l$, $p_l$ the fraction of edges with at least one end vertex inside the community $l$.

### 4.2. Blondel Algorithm

There are many alternative methods for communities detection. One of them was developed by V. D. Blondel *et al.*

and is described in [14]. It is a simple heuristic technique based on modularity optimization that calculates a network partition in a short computation time. The authors claim in [14] that their algorithm outperforms many other methods in terms of quality of communities detection and computation time.

The algorithm is composed of two phases that are repeated iteratively. It starts from the assumption that every node is assigned to a different communities, hence the initial number of communities is equal the number of nodes $N$ in a network. Next, for each node $i$ and all its neighbors $j$ values of modularity coefficient $Q$ (10) are detected under the assumption that the node $i$ is moved to the community of $j$. The calculations are repeated for all neighbors of $i$. Finally, the node $i$ is moved to the community, for which the gain of $Q$ is the highest one. The calculations are repeated for all nodes in a network, until no further improvement can be achieved. The algorithm switches to the second phase. A new aggregated network is built. In every community detected during the first phase, all nodes from this community are aggregated into one *super-node*. The weights of the edges between super-nodes are equal to the sum of weights of edges, linking two communities corresponding to these super-nodes. Hence, a new network is formed by these super-nodes. The second phase is completed and the first phase of the algorithm is executed for the aggregated network. Then, both phases are repeated iteratively, until no further improvement in the modularity coefficient can be achieved. The result of the algorithm is the partition of the original network into communities. Moreover, the algorithm also computes division inside computed groups.

## 5. Numerical Experiments

Multiple experiments were performed for data acquired from widely used social networks. The goal was to verify the results of application of SNA methods to knowledge extraction from massive commonly available data. In our tests, we validated and compared two techniques for community detection, described in the previous section: Clauset&Newman and Blondel *et al*. algorithms. Four series of experiments were performed for data acquired from the social platforms. The objective of the first set of tests was to compare the performance of described grouping techniques. Next, two series of experiments were performed for data about interpersonal connections, acquired from two commonly used platforms Facebook and Twitter. Different kinds of social networks were considered. The last series, of tests was performed for data acquired from the thesixtyone.com web page. The objective was to detect cliques of malicious voters.

### 5.1. Comparison of Algorithms of Communities Detection

We validated the communities detection algorithms through simulation. The accuracy and performance of three algorithms were compared, two presented in Section 4 and MCL (Markov Clustering) technique described in [23]. All experiments were performed for data containing members of Karate club from San Francisco. The results of calculations, i.e., discovered communities are presented in Figs. 1–3.

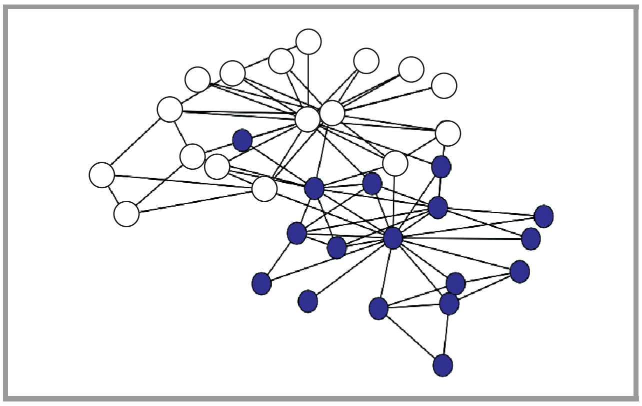

***Fig. 1.*** Detected communities (Clauset&Newman aglorithm).

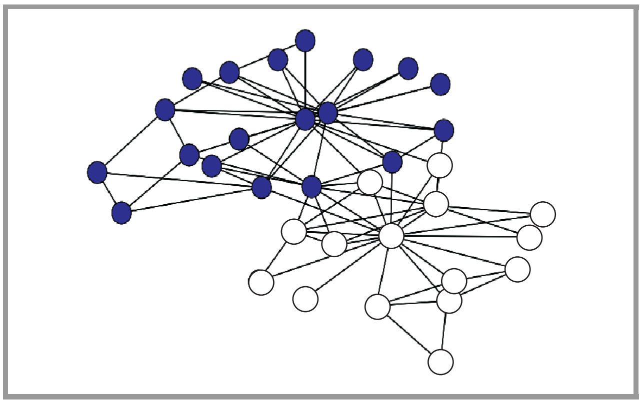

***Fig. 2.*** Detected communities (MLC aglorithm).

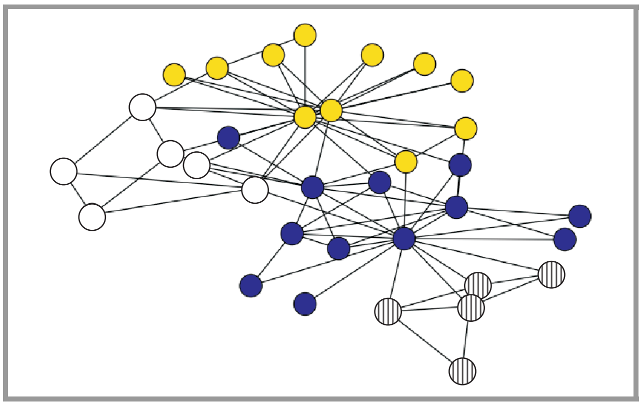

***Fig. 3.*** Detected communities (Blondel aglorithm).

We obtained similar groupings with these three algorithms, although with some differences. Both Clauset&Newman and MLC algorithms discovered two groups, which differed only in two nodes. The disadvantage of the MLC algorithm is that it has to be tuned manually, so it is difficult to use. The Blondel algorithm identified four groups, but after dividing them into two pairs and merging members of these pairs, the results were the same as calculated using the Clauset&Newman algorithm.

Table 1
Communities detection – calculation time

| Algorithm | Calculation time [s] |
|---|---|
| Griewan&Newman | 5.051 |
| MLC | 4.979 |
| Blondel | 2.680 |

The goal of the second series of experiments was to compare the performance and efficiency of the algorithms. The network formed by 931 nodes and 73 228 edges was considered. The calculation times of communities detection using different algorithms are collected in Table 1. The results presented in this section indicate that the Blondel algorithm produced more accurate results, and it was about 2 times faster than the other methods. However, from the perspective of community detection accuracy, the suggestion is to use more than one algorithm and compare the results.

### *5.2. Social Network from Facebook*

Facebook is a social networking service and website that connects people with other people, and share data between people. A user can create a personal profile, add other users as friends, exchange data, create and join common interest communities. The objective of our experiment was to validate the theorems formulated in SN domain on real network. We extracted a subnetwork from the Facebook database. Next, we calculated the centrality measures described in this paper, and finally divided this network into smaller communities. The test network was a special-kind network, so called ego-network. In such network, all nodes are connected to the *central node* (apart from being connected among themselves). In our case, the central node was one of the authors and the rest of the network was formed by his friends that were registered in Facebook.

We started our experiment from calculating the centrality measures of all nodes in our network. The results – values of degree, closeness, betweenness and eigenvector centrality measures are depicted in Figs. 4–7. From the experimental results, we can observe that for most nodes in the test network the calculated centrality measures are similar, low values. The results confirm theory about free-scale nature of social networks. It means that in SNs, usually only a few nodes are important and the other nodes are similarly not so much important. Moreover, it can be noticed that the correlation between centrality measures calculated for all nodes in the network is positive, i.e., in case of all nodes, a high value of one measure for a given node involves high values of other measures for this node.

Next, measures for the whole network were computed, i.e., a shortest path and a clustering coefficient. We ob-

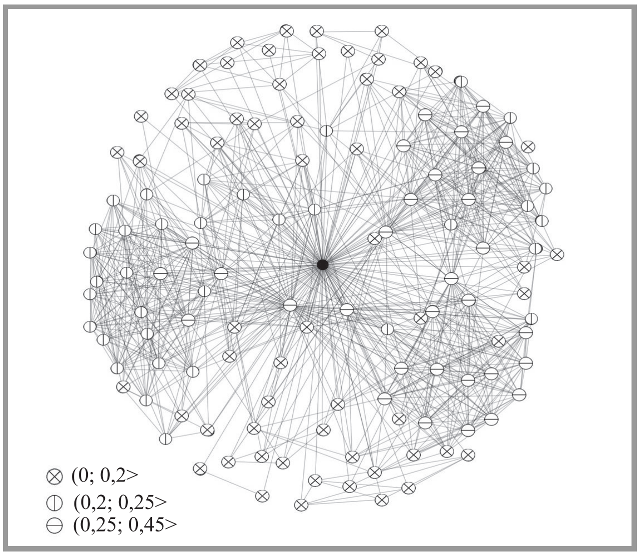


***Fig. 4.*** Degree centrality of nodes; the Facebook network.

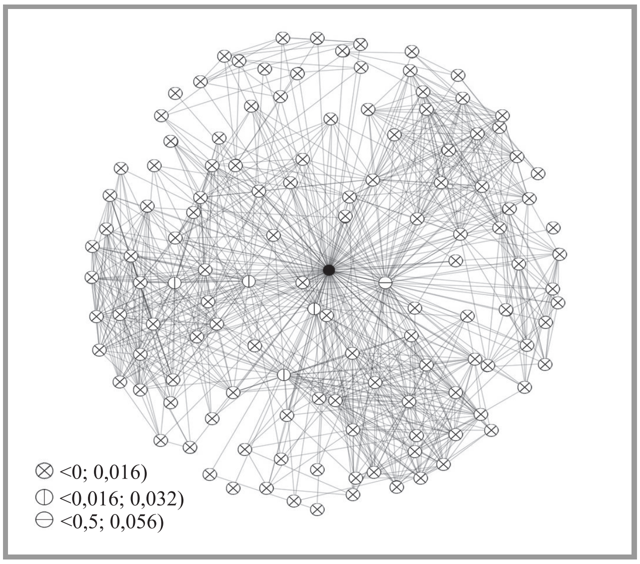


***Fig. 5.*** Betweenness centrality of nodes; the Facebook network.

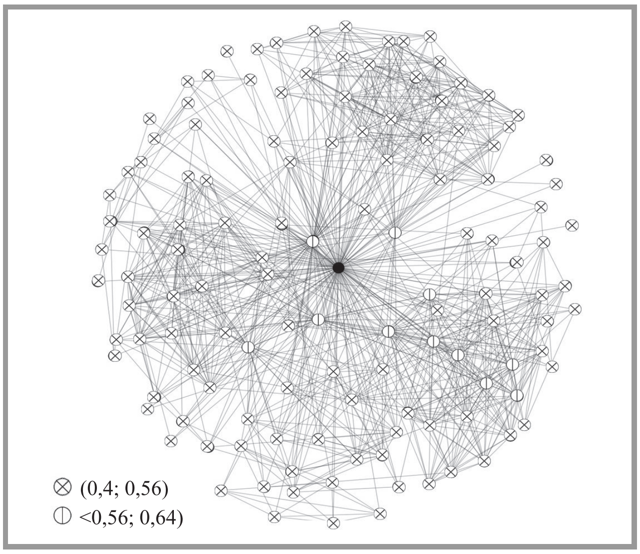


***Fig. 6.*** Closeness centrality of nodes; the Facebook network.

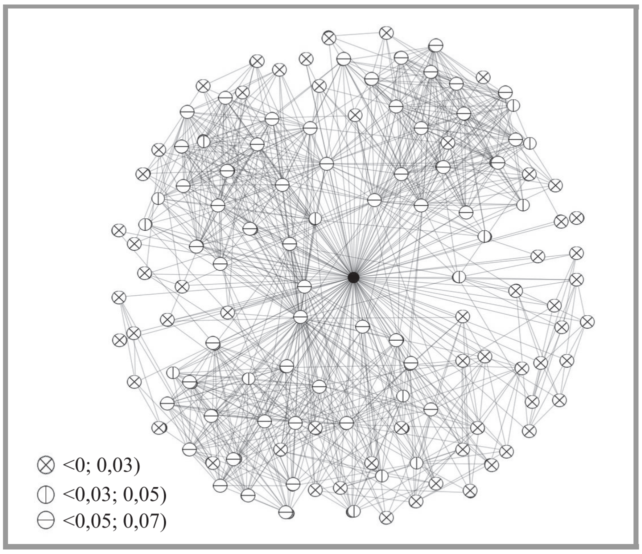


*Fig. 7.* Eigenvector centrality of nodes; the Facebook network.

tained the following values: the average clustering coefficient was quite high and equal to 0.7487, the average shortest path length was rather low and equal to 1.887. Such values of these measures are typical to small-world networks. Hence, the results of our experiments confirmed that our test network is a typical SN.

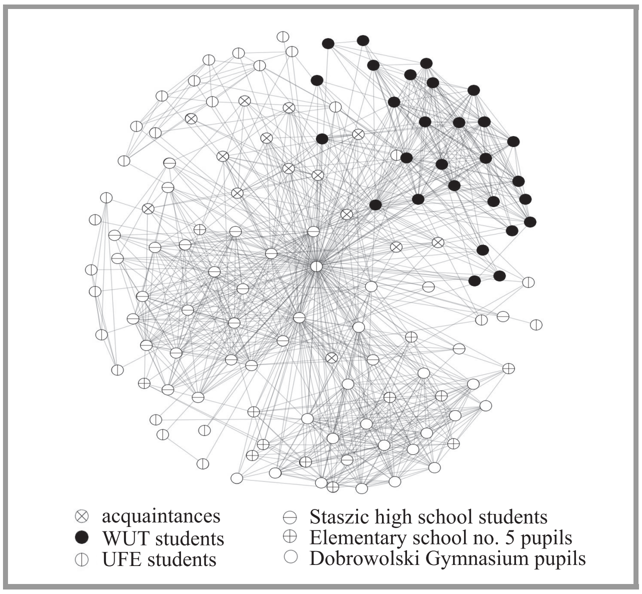


*Fig. 8.* Community detection using Blondel *et al.* algorithm; the Facebook network.

Finally, we used the Blondel *et al.* algorithm of communities detection in our test network. The results – communities extracted from the network are presented in Fig. 8. To verify the results of the experiment, we manually (based on our knowledge) detected the communities that were formed by friends of the author from different periods of his life (primary school, high school and university). After comparison of calculated and manually detected groups, we obtained the accuracy of the Blondel *et al.* algorithm equal to 68%.

### *5.3. Social Network from Twitter*

The next series of experiments was performed for data acquired from the Twitter platform. Twitter is a social networking and microblogging service. The users of Twitter can exchange text-based posts called *tweets*. A tweet is a maximum 140 characters long but can be augmented by pictures or audio recording. The main concept of Twitter was to build a social network formed by friends and followers. Friends are people who you follow, followers are those who follow you. Hence, the person who has many followers in Twitter is recognized as an important actor in a given network. The Twitter system collects not only data about people who send tweets but also those who decide to forward these tweets to other users of Twitter. Moreover, the tweets are aggregated to speed up Twitter. Hashtag (a word included in a tweet preceded with a hash # symbol) is added to some tweets. Next, tweets with the same hashtag are aggregated into one stream. Therefore, persons who are interested in a popular topic have an easy and fast access to information concerned with this selected topic.

Similarly to the previous set of tests, we tried to extract knowledge about examined social networks. We compared two social networks formed by two different groups of users tweeting about two topics: a pop starlet Justin Bieber and July Oslo massacre on Utoya island. In the first step of our experiment we collected tweets with the hashtags #justinbierber and #oslo, and formed two groups corresponding to two hastags. Then information about senders of all collected tweets were downloaded from the Internet. Two social networks (one for each tag) were built with nodes corresponding to the senders and edges linking nodes that followed one another. The measures described in Section 3 were calculated for both networks. The computed values of an average clustering coefficient, node degree and shortest path length are presented in Table 2. From the results we can observe that people twitting about Justin form a community with higher connectivity. It seems credible because this group consists of young people – typical users of social networking platforms such as Twitter, Facebook etc. and fans of the singer. The "oslo" network was formed by loosely connected people just as a response to one event

Table 2
Twitter networks characteristics

| Measure | #justinbieber | #oslo |
|---|---|---|
| Number of nodes | 1470 | 519 |
| Number of edges | 7081 | 636 |
| Maximal degree of node | 1414 | 403 |
| Avg. clustering coefficient | 0.137 | 0.1017 |
| Avg. node degree | 9.38 | 2.45 |
| Avg. shortest path length | 3.06 | 5.95 |

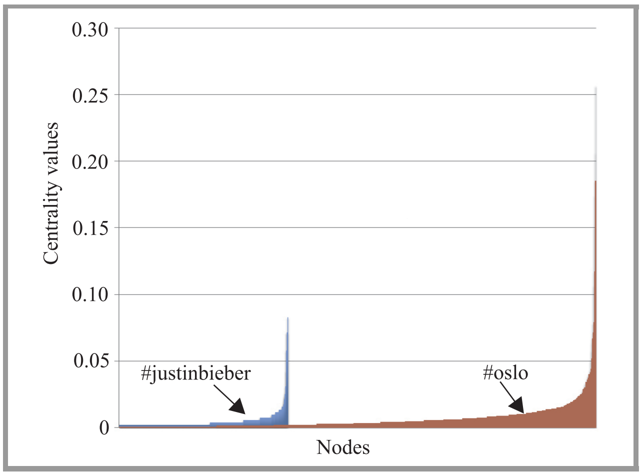


*Fig. 9.* Degree centrality of nodes; the Twitter network.

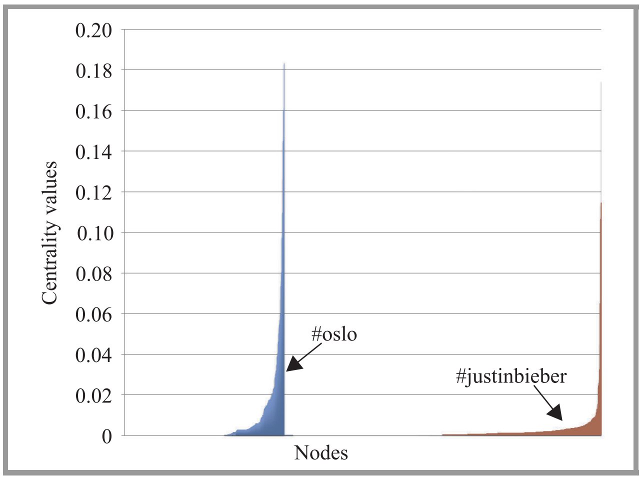


*Fig. 10.* Betweenness centrality of nodes; the Twitter network.

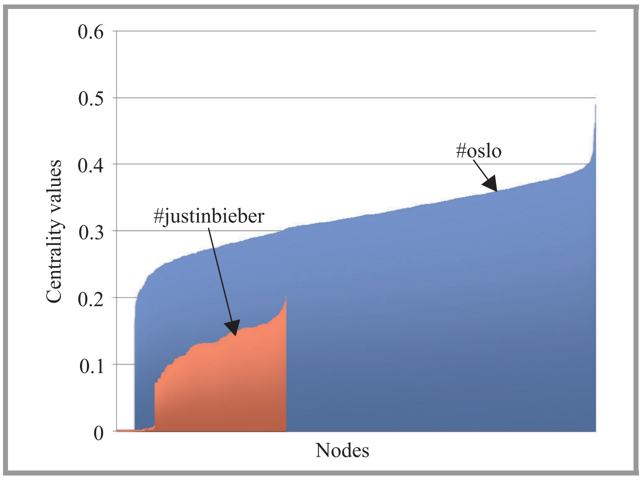


*Fig. 11.* Closeness centrality of nodes; the Twitter network.

that had suddenly happened. Next, we calculated the centrality measures. The results are depicted in Figs. 9–12. Finally, we checked the scale-free structure of the Twitter network. The calculated degree distribution is presented in Fig. 13, and the retweet distribution using approximately 65% of tweets in Fig. 14. Our experiments proved that the topic-based networks, as Twitter, are typical scale-free networks.

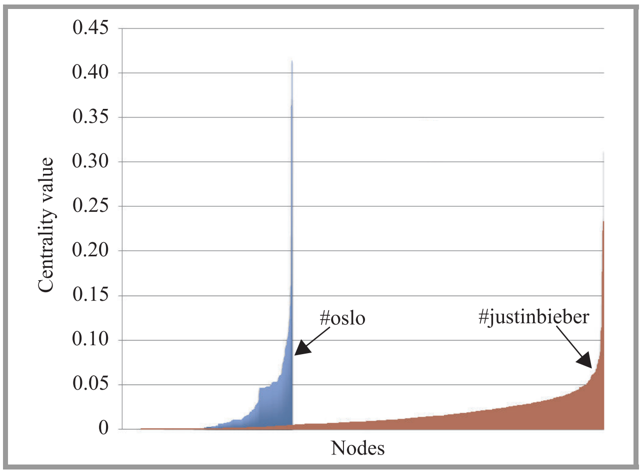


*Fig. 12.* Eigenvector centrality of nodes; the Twitter network.

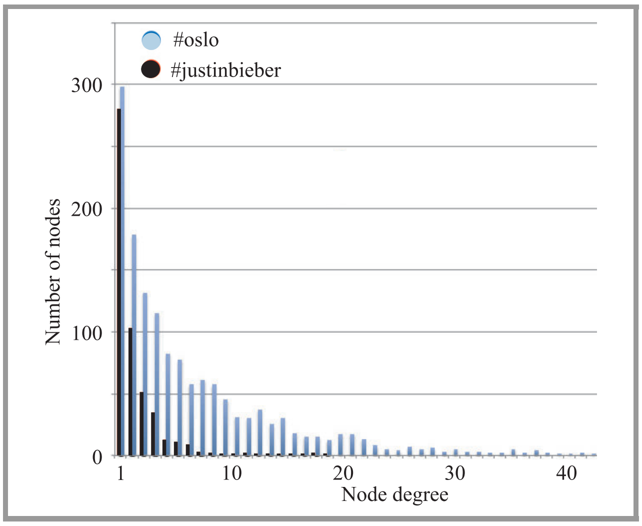


*Fig. 13.* Degree distribution; the Twitter network.

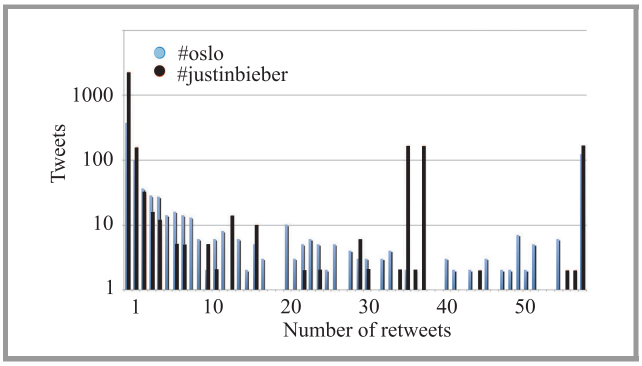


*Fig. 14.* Retweet distribution using approximately 65% of tweets.

The untypical result is a peak in histograms describing "oslo" network, Fig. 14. It can be caused by a variety of actors with high centrality forming this network. We can expect that participants of this network are typical leaders, i.e., news agencies or newspaper accounts, and another leaders – *information brokers* such as journalists.

### *5.4. Detecting Cliques of Malicious Voters*

The last series of experiments was performed for data acquired from the thesixtyone.com web page. This page is owned by the record company. It presents pieces of songs done by young, rather unknown artists that try to release their first album. The company founds a recording of such an album for the group, which get the highest number of votes given by the web users. In general, democratic voting is widely adopted by many web applications. Unfortunately, in case of such a voting there is an obvious room for abuses, such as bribing the voters or using voting bots in order to get the highest number of votes. For this reason, the owner of the thesixtyone.com is interested in discovering such cliques of malicious voters.
The formulation of the problem was as follows: given the list of objects and lists of users, who voted on these objects, identify cliques of malicious voters. Every voter could vote on many objects, but on one object each voter can vote only once. The SNA techniques were employed to solve the problem. First, the social network formed by voters was generated, and then the communities detection algorithms were employed to identify groups of voters. Finally, we tried to recognize suspicious groups – cliques of malicious voters. The network was created under following assumptions.

- The network was formed by voters (network nodes). Each edge linked two voters voted on the same object.
- The weights were assigned to each edge; $weight = \frac{1}{L}$, where $L$ denoted a number of times that connected voters voted on the same object.
- The edges with values of *weight* grater than an assumed threshold value *cut-off level* were removed from considerations (only persons who often vote on similar objects were suspected).

The Blondel algorithm was used to detect cliques. The network was divided into groups. The smallest one consisting of 106 voters (11.38% of network nodes) was recognized as a clique of malicious voters.
In order to verify the performance of the proposed method, we performed several experiments for data generated by our network simulator. The simulator applies NetworkX library. It was used to generate networks with properties similar to the thesixtyone network. Next, the list of malicious voters was generated. Using different parameters we generated networks with different properties (number of cliques, size of cliques, etc.). Multiple experiments were performed for a network formed by 500 nodes, *cut-off level*=1/3, and different input parameters. In general, the results were unsatisfactory. On average, only 5% of voters recognized as suspected persons were among real malicious voters.

The results of this experiment show the limitations of application of simple grouping techniques to social networks analysis. It is often difficult to divide actors who behave in a similar way into groups. The key issue is to define the adequate criterion or measure for the selecting procedure when strong differences between actors can not be observed. In such cases other methods of analysis applied to larger set of data should support the simple SNA techniques (see [12]). In case of our experiment we probably could reduce the number of badly classified voters considering data from not one but series of voting records.

## 6. Summary and Conclusion

The paper provides the short overview of social network analysis techniques. The common properties of social networks were summarized. By performing experiments for real life social networks available in two different types of popular social services Facebook and Twitter, we tried to show that SNA is a valuable tool for extracting knowledge from networks encountered in nature, especially networks formed by people. Our results confirm that both Facebook and Twitter are typical social networks, i.e., scale-free and small-world networks. It is worth mentioning that SNA techniques are based on data processing, and unfortunately, they may fail for more complex problems when network properties and available data are not enough to make a decision and solve a task.

## References


[1] S. N. Dorogovtsev and J. F. F. Mendes, "Evolution of networks", *Advances Phys.*, vol. 51, no. 4, pp. 1079–1187, 2002.

[2] W. Gruszczyński and P. Arabas, "Application of social network to improve effectiveness of classifiers in churn modelling", in *Proc. 3rd Int. Conf. Comput. Aspects of Social Netw. CASoN'11*, Salamanca, Spain, 2011.

[3] M. Kamola, B. C. Piech, and E. Niewiadomska-Szynkiewicz, "Reconstruction of a social network graph from incomplete call detail records", in *Proc. 3rd Int. Conf. Comput. Aspects of Social Netw. CASoN'11*, Salamanca, Spain, 2011.

[4] M. E. J. Newman, "Modularity and community structure in networks", *Proc. Nat. Academy Sci. USA*, vol. 103, no. 23, pp. 8577–8582, 2006.

[5] M. E. J. Newman, A. L. Barabasi, and D. J. Watts, *The Structure and Dynamics of Networks*. USA: Princeton University Press, 2006.

[6] M. E. J. Newman, "Communities, modules and large-scale structure in networks", *Nature Phys.*, vol. 8, pp. 25–31, 2011.

[7] S. Wasserman and K. Faust, *Social Network Analysis*. USA: Camridge University Press, 2009.

[8] S. Eubank, H. Guclu, V. S. A. Kumar, M. V. Marathe, A. Srinivasan, Z. Toroczkaj, and N. Wang, "Modelling disease outbreaks in realistic urban social networks", *Nature*, vol. 429, pp. 180–183, 2004.

[9] A. D. Henry, "Belief-oriented segregation in policy networks", *Procedia – Social Behav. Sci.*, vol. 22, pp. 14–26, 2011.

[10] B. Karrer and M. E. J. Newman, "Competing epidemics on complex networks", *Phys. Rev.*, vol. 84, pp. 1–14, 2011.

[11] S. L. Magsino, *Applications of Social Network Analysis for Building Community Disaster Resilience*. USA: The National Academies Press, 2009.

[12] M. A. Porter, P. J. Mucha, M. E. J. Newman, and A. J. Friend, "Community structure in the united states house of representatives", *Physica*, vol. 386, pp. 414–438, 2007.

[13] Z. Tarapata and R. Kasprzyk, Graph-based optimization method for information diffusion and attack durability in networks. *Lecture Notes Artif. Intel.*, vol. 6086, pp. 698–709, 2010.

[14] V. D. Blondel, J. L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in lagre networks", *J. Statistical Mechanics: Theory and Experiment*, no. 10, pp. 1–12, 2008.

[15] S. P. Borgatti, "Centrality and network flow", *Social Netw.*, vol. 27, pp. 55–71, 2005.

[16] A. Clauset, M. E. J. Newman, and C. Moore, "Finding community structure in very large networks", *Rev. Modern Phys.*, vol. 70, pp. 66–111, 2004.

[17] A. Fronczak and P. Fronczak, *Świat sieci złożonych. Od fizyki do Internetu*. Warsaw: PWN, 2009 (in Polish).

[18] D. Watts, "Networks, dynamics and the small-world phenomenon", *The American J. of Sociol.*, vol. 105, no. 2, pp. 493–527, 1999.

[19] R. Albert and A. L. Barabási, "Statistical mechanics of complex networks", *Rev. Modern Phys.*, vol. 47, pp. 47–97, 2002.

[20] A. L. Barabási and R. Albert, "Emergence of scaling in random networks", *Science*, vol. 286, pp. 509–512, 1999.

[21] S. Boccaletti, M. Ivanchenko, V. Latora, A. Pluchino, and A. Rapisarda, "Detecting complex network modularity by dynamical clustering", *Phys. Rev.*, vol. 75, no. 4, pp. 1–4, 2007.

[22] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks", *Proc. Nat. Academy Sci. USA*, vol. 99, no. 12 pp. 7821–7826, 2002.

[23] A. J. Enright, S. Van Dongen, and C. A. Ouzounis, "An efficient algorithm for large-scale detection of protein families", *Nucleic Acid Res.*, vol. 30, no. 7, pp. 1575–1584, 2002.



**Marcin Mincer** received his B.Sc. in Computer Science from the Warsaw University of Technology, Poland, in 2011. Currently he is a M.Sc. student in the Institute of Control and Computation Engineering at the Warsaw University of Technology. Since 2011 he is involved in the ECONET project of EU. His research area focuses on social network analysis applied on emerging Internet social media platforms.

E-mail: M.Mincer@stud.elka.pw.edu.pl
Institute of Control and Computation Engineering
Warsaw University of Technology
Nowowiejska st 15/19
00-665 Warsaw, Poland



**Ewa Niewiadomska-Szynkiewicz,** D.Sc. (2005), Ph.D. (1995), M.Eng., Professor of Control and Information Engineering at the Warsaw University of Technology, head of the Complex Systems Group. She is also the Director for Research of Research and Academic Computer Network (NASK). She is the author and co-author of three books and over 120 journal and conference papers. Her research interests focus on complex systems modeling and control, computer simulation, global optimization, parallel computation, computer networks and ad hoc networks. She was involved in a number of research projects including EU projects, coordinated the Groups activities, managed organization of a number of national-level and international conferences.

E-mail: ens@ia.pw.edu.pl
Institute of Control and Computation Engineering
Warsaw University of Technology
Nowowiejska st 15/19
00-665 Warsaw, Poland

E-mail: ewan@nask.pl
Research and Academic Computer Network (NASK)
Wąwozowa st 18
02-796 Warsaw, Poland