Int. Journal of Intelligent Information and Database Systems , Vol. 1, No.3/4, 293-310

On Utilizing Social Networks to Discover Representatives of Human Communities

Przemysław Kazienko and Katarzyna Musiał

Institute of Applied Informatics, Wrocław University of Technology
Wyb.Wyspiańskiego 27, 50-370 Wrocław, Poland,
Phone: +48 71 3203258, fax: +48 71 3203453
Email: {Kazienko, katarzyna.musial}@pwr.wroc.pl
Home page: http://www.zsi.pwr.wroc.pl/~kazienko/eng_index.html,
http://www.zsi.pwr.wroc.pl/~musial/eng_index.html

Abstract: Virtual human communities that exist on the Internet reflect social relationships between people. They can be either directly maintained by computer systems like Frindster, LinkedIn or extracted from data about user activities in the communication network like e-mails, chats, blogs, homepages connected by hyperlinks, etc. Social network analysis can be applied to virtual communities and can deliver interesting knowledge about particular network members as well as cohesive subgroups. It appears that there is a great need to find important individuals or a set of people who would represent larger communities. These people would be able to perform specific tasks or could become a target group either for marketing or advertising purposes. The new research on representative discovery for human communities based on social network analysis is presented in this paper. Its main goal is to improve the process of target group selection by adding the social elements derived from the behaviours of people. The target group should fulfil a set of prerequisites. Simultaneously, the selection process ought to respect specific social aspects provided by the social network. The entire selection process considered in the paper is called human filtering and includes one of two separate scenarios: selection by single member or selection by social group.

Keywords: human communities; social network; group representatives; social groups; human filtering; human filtering by single member; human filtering by social group; social position.

Biographical notes

Przemysław Kazienko is an assistant professor and also the deputy director for development at the Institute of Applied Informatics, Wrocław University of Technology, Poland. He received his M.Sc. in 1991, and Ph.D. in 2000, in computer science both from the Wrocław University of Technology. He has authored over 70 scholarly and research articles on a variety of areas related to social networks, data mining, data security, knowledge management, Information Retrieval, and XML. He has also managed a number of commercial developments. His current research and scientific activities focus on the social network analysis as well as application of data mining methods to adaptive web-based systems, particularly to recommender systems.

Katarzyna Musiał is a Ph.D. student at the Institute of Applied Informatics, Wrocław University of Technology. She is especially interested in social networks and social network analysis. She received her M.Sc. in computer science in 2006 from the Wroclaw University of Technology, Poland and M.Sc. in software engineering in 2006 from the Blekinge Institute of Technology, Sweden. She focuses in her research on the social networks analysis, especially on the calculation of individuals' social position in the virtual social networks.

1 Introduction

These days there is a great need to find appropriate individuals or groups of people who meet the specified requirements and who would be suitable for the tasks specified in advance. The selection of the target groups is a common activity in almost all companies where the managers pick people for the project teams or try to find new employees. On the other hand, the target groups are the sufficient and basic component for target marketing as well as for advertising campaigns. Identifying the cohesive subgroups (Frank, 1995) is applicable in targeted advertising (Yang, *et al.*, 2006). There is also another domain – collaborative filtering, widely used in recommender systems – in which people are usually clustered into classes, nearest neighbourhoods. Each individual is suggested the appropriate services or goods based on user preferences within the group that is the closest to the given user (Adomavicius and Tuzhilin, 2005), (Montaner, Lopez, and De la Rosa, 2003).

However, people now are usually selected based on their individual abilities that match task requirements. Such a case is when hiring new members for a project team or they are selected based on the mutual profile similarities between them, as well as their past behaviours, as in the case of collaborative filtering.

The aim of the solution presented in this paper is to select the set of network representatives not only based on their demographic characteristics and interests, but also based on the position of a person within the human community. To reach this goal we should take into account that people who are in relations influence one another, so their behaviours should be analyzed in the context of their relationships (Wasserman and Faust, 1994).

2 Related Work

Social networks have recently become one of the research areas where scientists from different fields look for inspiration. The social network analysis supported by computer science gives opportunity to develop other branches of knowledge. This cooperation of researchers from sociology and computer science helps to develop and improve both new scientific ideas and commercial solutions (Kazienko and Musiał, 2006a), (Kazienko and Musiał, 2006b), (Yang, et al. 2006). Nevertheless, before the social network analysis can be applied, some basic concepts related to social networks ought to be explained.

The main idea of social network can be easily presented as the finite set of actors who are the nodes of the network, and ties that link the nodes by one or more relations (Garton, Haythorntwaite, and Wellman, 1997), (Hanneman and Riddle, 2005), (Wasserman and Faust, 1994). Wasserman and Faust defined an actor as an individual, corporate, or collective social unit and a tie as a linkage between a pair of actors (Wasserman and Faust, 1994). The range and type of the relational tie can be extensive (Hanneman and Riddle, 2005), (Wasserman and Faust, 1994) and different depending on the type and character of the analyzed actors.

In order to analyze the social networks, the social network analysis should be performed. It focuses on understanding the connections among people and the implications of these connections (Wasserman and Faust, 1994). Thus, the main goal of social network analysis can be defined as follows: “a methodology for examining the structure among actors, groups, and organizations and aides in explaining variations in beliefs, behaviours, and outcomes” (Hatala, 2006). Note that the input data collection is required for every kind of network analysis. Social networks that exist on the Internet provide a vast amount of useful data available for this investigation. This is the reason why virtual communities based on the Internet are considered in this paper. They can be either directly maintained by computer systems like Frindster (Boyd, 2004) and LinkedIn or extracted from data about user activities in the communication networks like e-mails (Kazienko and Musiał, 2007), chats, blogs, homepages connected by hyperlinks (Adamic and Adar, 2003), etc. Some researchers identify the communities within the Web based on, e.g. link topology (Flake, Lawrence, and Giles, 2000), (Gibson, Kleinberg, and Raghavan, 1998), while others analyze the emails to discover the social network (Culotta, Bekkerman, and McCallum, 2004), (Kazienko and Musiał, 2007), (Shetty and Adibi, 2005), (Zhu, Chen, and Allen, 2006).

Social network analysis can be utilized to extract key members by determining the centrality, prestige, reachability, and connectivity measures (Carrington, Scott, and Wasserman, 2005), (Hanneman and Riddle, 2005), (Wasserman and Faust, 1994). Many approaches exist in order to evaluate the centrality of the person (Freeman, 1979). One of them is degree centrality that takes into account the number of neighbours that are adjacent from the given person (Shaw, 1954). Another example of the centrality measure is closeness centrality which pinpoints how close a member is to all the others within the social network by determining the shorter paths from user h to all other embers of the community (Sabidussi, 1966). Finally the betweeness centrality of a member pinpoints to what extend this member is between other people in the social network (Freeman, 1979). The second measure that characterizes a position of a member in the social network is prestige. Similarly to centrality, prestige can be calculated in various ways, e.g. proximity prestige, rank prestige, and degree prestige. The degree prestige takes into account the number of members that are adjacent to a particular member of the community (Wasserman and Faust, 1994).

Proximity prestige pinpoints how close all members are within the social community to a particular member (Wasserman and Faust, 1994). Finally, the rank prestige is measured based on the status of members in the network and depends not only on geodesic distance and number of relationships, but also on the prestige of members connected with the member (Katz, 1953).

The measure of centrality and prestige can be utilized not only in the regular social networks but also in the Web network, i.e. the network of web pages. Kumar *et al.* claim that the Web can be seen as a social network (Kumar, *et al.*, 2002) and this causes that the similar measures for nodes in hypertext and web-based systems can be applied (Botafogo, Rivlin, and Shneiderman, 1992).

3 Problem Description

Group selection is applicable to many domains, e.g. the choosing of people for project teams, finding potential employees, searching consumer groups for advertising campaigns, or for use in target marketing.

The information available in the human community can improve the process of finding the appropriate employees for a specific task. On the one hand, the head-hunters can find new employees who will suit their expectations by social network analysis. On the other hand, the analysis of an individual’s characteristics and their acquaintances within the social network support the development of target marketing by searching for the appropriate target group. For example, some specific products or services can be offered to the single and relatively small clique that has been identified in the community. Moreover, these goods may be suggested only to the carefully selected representatives of the group, who are the most important in the population as well as those who have the greatest influence on others. In all of these cases, the main goal is the same – to find the group of representatives that will fulfil fixed requirements with respect to social features of network’s members. The new solution proposed below combines the traditional way of people selection with some new elements of social network analysis.

In this paper, an actor is defined as a discrete individual, i.e. a single person, while ties reflect the behavioural interactions between actors (Figure 1). For example, a couple of people who exchange email messages, who talk to each other or who comment on the same internet blogs are in a mutual relation. The social network that includes only individuals as actors is called a human community or human social network, and this is an important assumption for the selection of the target group considered in this paper. Based on these assumptions the definition of human social network can be created.

Definition 1. A human social network *HSN* is a tuple $HSN=(H,R)$, where H is the finite set of members – humans, and R is the set of social relationships (ties) that join pairs of distinct humans, $R{:}H\times H$, i.e. $R=\{(h_i,h_j){:}\ h_i,\in H,\ h_j\in H,\ i\neq j\}$.

Note that according to Definition 1 a network member cannot be in relation with himself. Moreover, only networks with card(H)>1 are reasonable to consider.

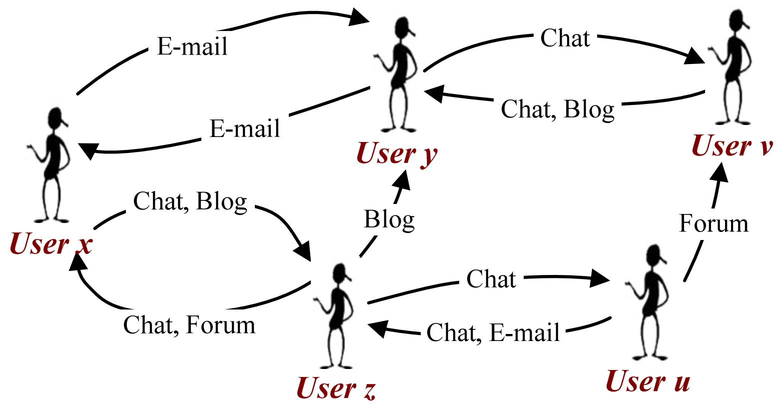

Figure 1 A social network

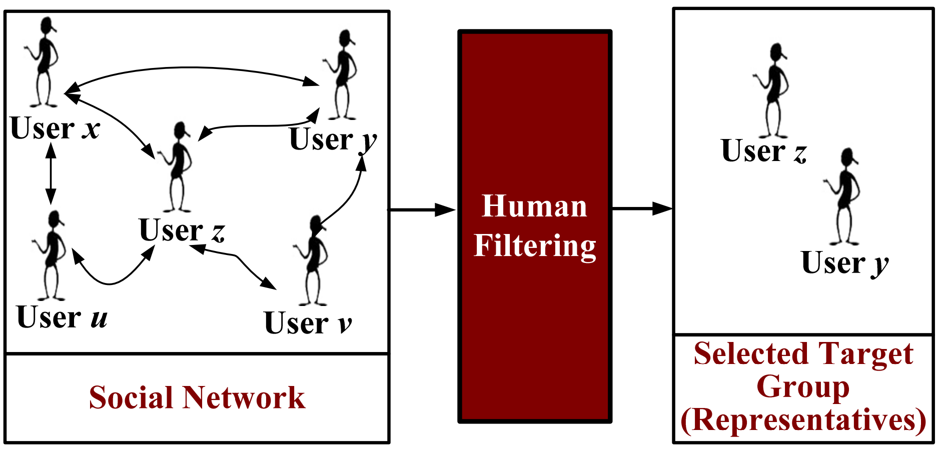

Figure 2 The process of selecting people for specific tasks

4 Representative Selection Based On Social Network

The main concept of human group selection from a social network that could represent the larger population is quite simple. The general aim is to pick the appropriate group of human community representatives that fulfil established requirements. Simultaneously, the social aspects of peoples' behaviours should be taken into consideration during the selection. The entire process, which is further precisely described, is called human filtering (Figure 2).

The human filtering process includes one of two separate scenarios: selection by single member and by social group. Both of them are presented below.

4.1. Selection by Single Member

The single member approach is composed of three main stages: simple matching, social position assignment and ranking creation (Figure 3). As a result of the human filtering, the ranking list of all humans H from human social network $HSN=(H,R)$ is created and top N representatives form the target group.

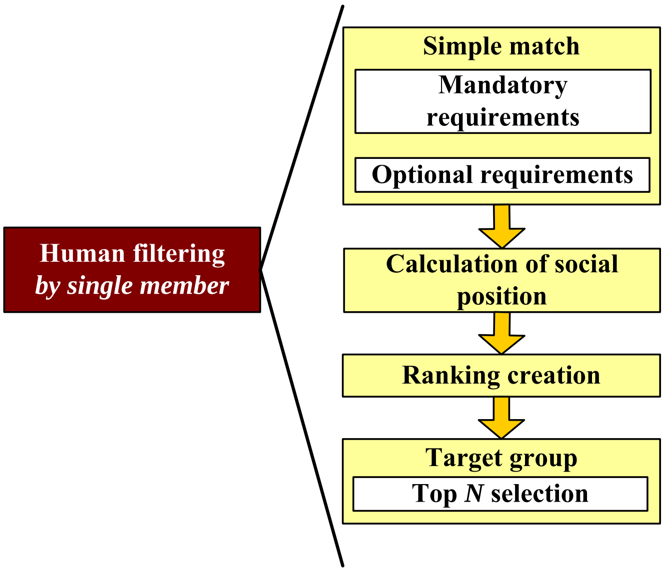

Figure 3 Elements of the process in human filtering by single member

This simple matching technique consists of testing members of the human social network *HSN* against the set of requirements to be fulfilled, such as gender, education level, age, etc. These requirements are established by the people who are looking for a person who suits their expectations.

Two groups of requirements: mandatory and optional can be identified. The former has to be fulfilled, whereas the latter does not necessarily. Language and particular ability can be considered as the mandatory requirements. Let's assume that we search for person who is going to be hired in Poland as a Java developer. Such a person should definitely speak Polish or English and can programme in Java. In such case mandatory requirements are: $language \in \{Polish, English\}$ and $occupation \in \{Java\ developer\}$. On the other hand, some optional requirements may also exist, e.g. additional skills. Assume that we are looking for Java developer in Poland and it will make a good impression if the person has some knowledge about databases or data mining then this skill is optional requirement, so $skill \in \{databases, data\ mining\}$.

The more prerequisites the user meets the higher position in the final ranking list one obtains. First, every user is tested against the mandatory requirements. If even one of these requirements is not fulfilled then such person is rejected from the further calculations. If none of the community members meets all of the mandatory requirements then these requirements should be weakened and all the calculation repeated. Then, each user is checked against the optional requirements. After the simple matching process, the social position value should be assigned to every member of the network and based on its value the ranking is created. Top N members from the ranking form the final group of representatives.

The whole process of human filtering for an individual h from human social network $HSN=(H,R)$, $h \in H$ is presented in Figure 4.

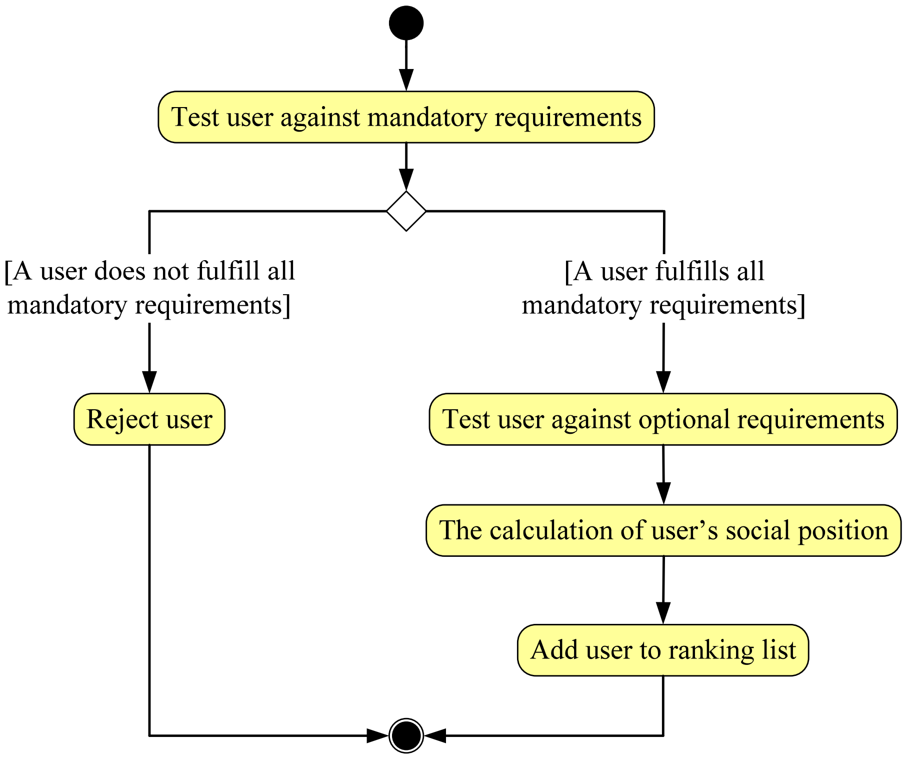

Figure 4 Human filtering process for an individual

This ranking list of $h \in H$ is created based on the value of the match function $M(h)$, as follows:

$$M(h) = \alpha \cdot GOR(h) + \beta \cdot \frac{SP(h)}{SP_{\max}} \qquad (1)$$

where:
$GOR(h)$ – the general optional requirements' function that pinpoints to what extend the user matches the optional requirements. The range of the function is [0,1];
$SP(h)$ – the social position of user h within the community, the range is from $[0,C]$; C – the number of members within the community;
$SP_{\max}$ – the maximum value of social position within the given network;
α, β – the coefficient that indicates the importance of the functions $OR(h)$ and $SP(h)$ respectively.

4.1.1. Optional Requirements

The optional requirements indicate the conditions that the person does not have to fulfil. Nevertheless, the more optional prerequisites the user meets the higher the position in the final ranking list one obtains.

The value of the general optional requirements' function $GOR(h)$ reflects to what extend user h matches optional requirements. It is calculated by using:

$$GOR(h) = \sum_{i=1}^{K} w_i OR_i(h) \qquad (2)$$

where:
K – the number of optional requirements;
$OR_i(h)$ – the function that pinpoints whether the i-th optional requirement is fulfilled or not. If so then $OR_i(h)$=1; otherwise $OR_i(h)$=0;
w_i – the weights assigned to every single optimal requirement, $w_i \in [0;1]$ and $\sum_{i=1}^{K} w_i = 1$.

Note that the weights, which are assigned to an optional requirement, enable us to adjust the importance of particular optional requirements. For example, if the weight for age is greater than for gender then the user had better be of the proper age rather than of proper gender.

After calculation of $OR(h)$ for all members, their social position is computed.

4.1.2. Social Position

The social position refers to the standing and potential social capital of the user in the network (Kazienko and Musiał, 2006b). The concept of social capital has been studied by many sociologists (Boyd, 2004), (Coleman, 1990), (Putnam, 2000). In general, the greater social position one possesses the more valuable this member is to others. It's often the case that we only need to extract the highly important persons, i.e. with the greatest social position. Such people surely have the biggest influence on others. As a result, we can focus our activities like advertising or marketing solely on them and we would expect that they would entail their acquaintances. The value of social position depends on both the strength of the relationships the person maintains in the network as well as the social positions of their acquaintances. In other words, the social position of a person is inherited from others but the level of inheritance depends on the activities of the acquaintances directed to this person. Thus, the social position depends also on the number and quality of relationships. The same social position can be achieved by member h if h has many relationships with people who have medium social position or if h has only few relationships but with participants with high social position.

Social position function $SP(h)$ of user h respects both the value of social positions of user's h acquaintances as well as their activities in relation to h:

$$SP(h) = (1-\varepsilon) + \varepsilon \cdot (SP(y_1) \cdot C(y_1 \to h) + ... + SP(y_m) \cdot C(y_m \to h)) \qquad (3)$$

where:
ε – the constant coefficient from the range [0,1];
$y_1, \ldots, y_m$ – acquaintances of h, i.e. members who are in the direct relation to h;
m – the number of acquaintances of user h;
$C(y_1 \to h), \ldots, C(y_m \to h)$ – the function that denotes the contribution in activity of $y_1, \ldots, y_m$ directed to h.

To calculate the social position of the person within the social network the convergent, iterative algorithm is used. This means that there have to be a fixed appropriate stop condition. The formal proof of the algorithm convergence is in (Kazienko and Musiał 2007).

Firstly, people who do not have any relationships within the network are rejected and their social position equals 0. Next, the contribution of the activity of the members who are in some relationships but are not active within these relationships at all is distributed equally among all their acquaintances. The reason being is that the sum of all contributions of one person in their relationships ought to be equal 1.

Research shows that the initial value of social position does not influence its final value, but the number of iterations. Moreover, the sum of all the users' social positions within the network is convergent to the number of network's members. In consequence, the average social position is convergent to 1. The formal proof of this feature is in (Kazienko and Musiał 2007).

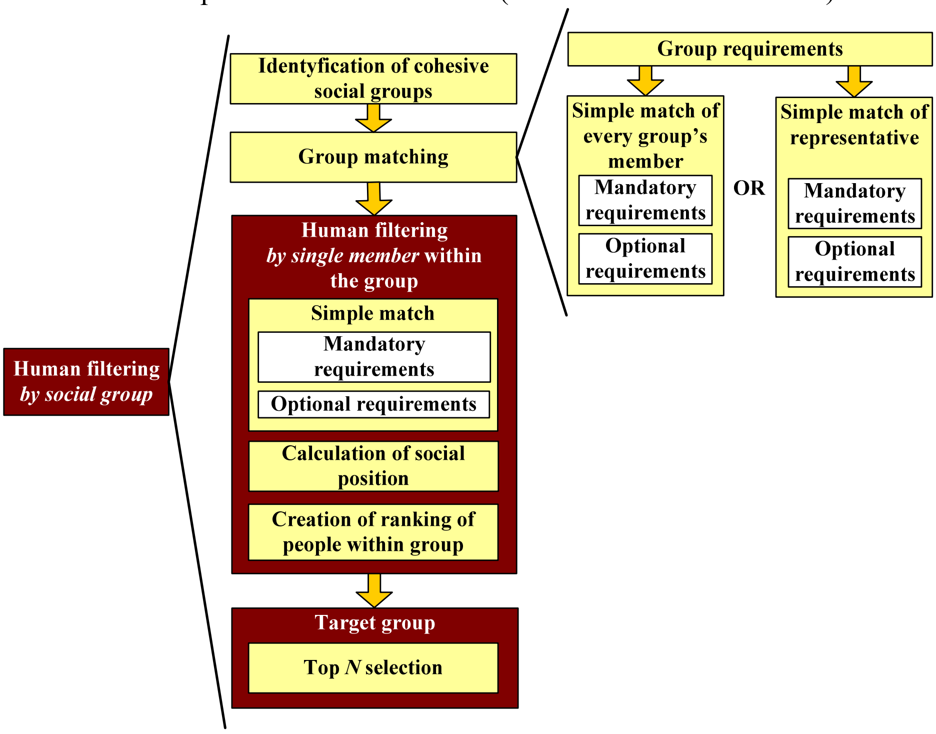

Figure 5 Elements of the process in human filtering by social group

4.2. Selection by Social Group

The second approach to human filtering is called selection by social group (Figure 5). First, a social clustering method that originates from social network analysis (SNA) is utilised to extract cohesive social subgroups, i.e. groups of people close to one another.

Definition 2. A cohesive social subgroup $CSG=(H^G,R^G)$ in the human social network $HSN=(H,R)$, is the coherent partition of the network $HSN=(H,R)$ that exceeds the given threshold of cohesiveness. $H^G \subset H$, $R^G \subset R$, $R^G=\{(h_i,h_j): h_i, \in H^G, h_j \in H^G, i \neq j\}$.

The measure of cohesiveness can be defined in many different ways. Wasserman and Faust distinguish four general properties of cohesive subgroups that can be used to evaluate the group cohesiveness (Wasserman and Faust, 1994):

a) The frequency of relationships always exceeds the given threshold τ – requires that every human $h_i \in H^G$ in the subgroup $CSG=(H^G,R^G)$ is related to at least τ humans $h_j \in H^G$ in this subgroup.
b) The mutuality of the relationship – requires that every human $h_i \in H^G$ in the subgroup $CSG=(H^G,R^G)$ is adjacent to all others in this subgroup. As a result, we obtain a clique. This is a stronger version of a).
c) The closeness or reachability of the subgroup members – requires that every human $h_i \in H^G$ in the subgroup $CSG=(H^G,R^G)$ is reachable to all others, but not necessarily adjacent. In other words, there exists a path between any pair of humans.
d) The relative high frequency of relationships within the subgroup – requires that every human $h_i \in H^G$ in the subgroup $CSG=(H^G,R^G)$ has at least as many relationships with subgroup members $h_j \in H^G$ as with other social network members $h_k \in H \backslash H^G$.

Cohesive subgroups can be extracted using one of the well-known clustering algorithms (Tan, Steinbach, and Kumar, 2006).

After the cohesive subgroups are identified, group matching is performed. The first step of this activity is to test each group against a set of additional group requirements that concerns the overall characteristic of the searched group. The cardinality of the group or cohesion level can be seen as the additional requirements. For example, we are interested only in groups that exceed a certain number of members and simultaneously, that the cohesion of the group is at least of a given level.

The second step of group matching is the simple match known from the selection by single member approach. However, here we can either test all users from the group against the requirements or create a representative of each group to test against the specified requirements. The group representation can be created based on the calculations of average characteristics of all of the group's members. In another approach, the representation is the person who is the closest to the centre of the group. As the result of group matching only one group that best suits the specific needs is selected. After that, human filtering by single member is performed to the selected group (see the previous section). Note that the input for human filtering by single member is one group instead of the entire social network. Finally, the ranking list of the selected members is created in order to identify the most valuable persons within the matched group.

4.3 Comparison of Both Approaches

Two variants of human filtering were presented in the previous sections: by single member and by social group. In the social group approach we obtain a list of the most potentially valuable members from one previously identified social group, i.e. those who are close to one another (Figure 6a). We can also assume that they are in the direct relationship, have common acquaintances or at least the shortest paths between them are relatively short. On the other hand, the single member method provides the distributed group of people from the entire network who not necessarily are related or close to one another (Figure 6b).

5. Case Study

The following case study presents how to calculate the social position of individuals within the social network. Let us assume that we have a human community as in Figure 7. The arc values indicate the contributions of activities between a pair of members. The sum of contributions per each member equals 1.

In order to calculate the social position the iterative algorithm is utilized. The calculation of social position is repeated until the defined stop condition is reached. In the presented case study the stop condition is defined as the precision to the fifth decimal place between two following iterations.

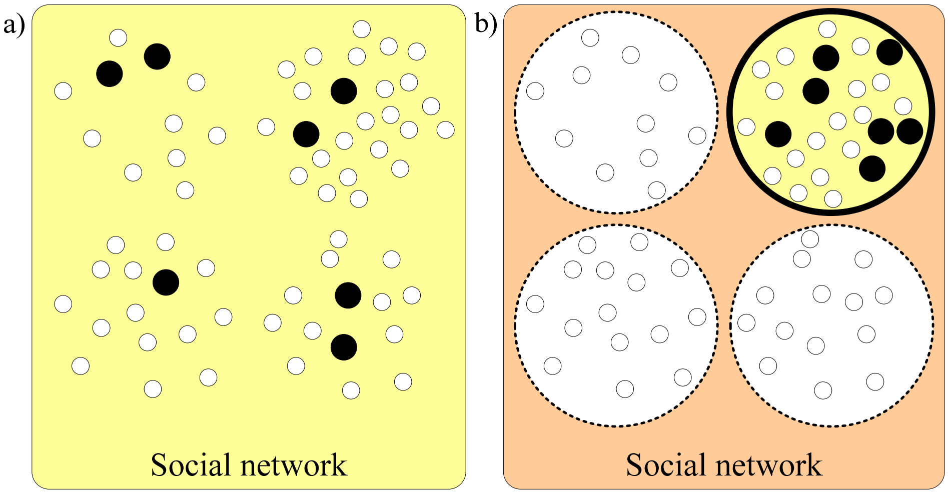

Figure 6 The outcome of human filtering process a) by single member; b) by social group. The black circles denote people who were selected in human filtering process

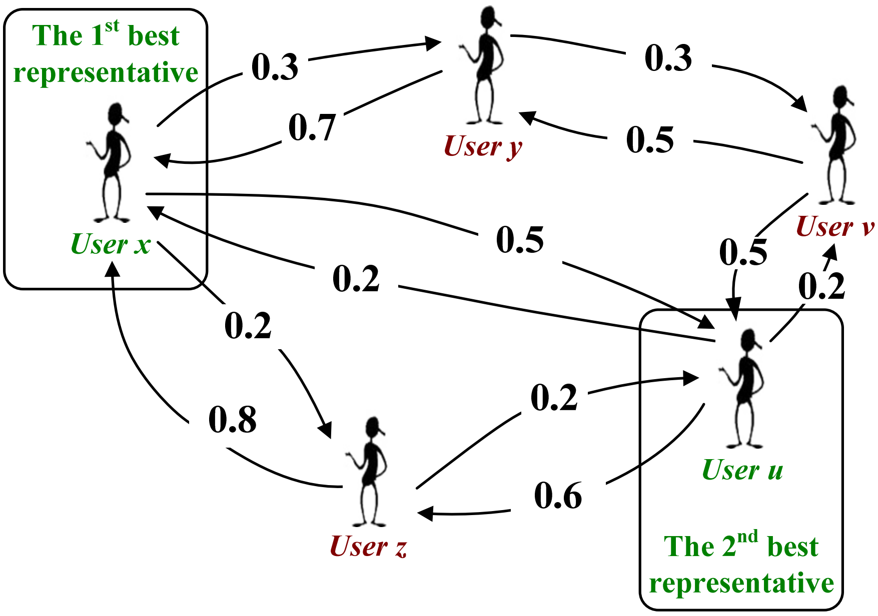

Figure 7 A human community and its two representatives

The tests consist of two main parts. Firstly, the influence of ε coefficient's value on the final social position of an individual is investigated. The aim of the second part is to assess the number of necessary iterations in relation to the initial social position values of the members of the human community.

5.1. The influence of ε

In the first step the initial social positions for all members (Figure 7) are established as follows: $SP(x) = 0.2$, $SP(y) = 0.2$, $SP(z) = 0.2$, $SP(u) = 0.2$, $SP(v) = 0.2$. The following values of ε have been taken into account: $\varepsilon = 0.1$, $\varepsilon = 0.5$, and $\varepsilon = 0.9$. For all cases the stop condition is: no difference in social position values with precision of 5 digits after the decimal point for all the members in two following iterations. The number of necessary iterations as well as the social position distribution has been studied in relation to ε (Table 1, Table 2, Table 3).

Table 1 Social position calculation for the social network from Figure 7; $\varepsilon = 0.1$

$\varepsilon = 0.1$						
Iteration No.	1	2	…	7	8	**9**
***SP*(*x*)**	0.2	0.934	…	1.06758	1.06758	**1.06758**
***SP*(*y*)**	0.2	0.916	…	0.97951	0.97951	**0.97951**
***SP*(*z*)**	0.2	0.916	…	0.98258	0.98258	**0.98258**
***SP*(*u*)**	0.2	0.924	…	1.02051	1.02052	**1.02052**
***SP*(*v*)**	0.2	0.910	…	0.94979	0.94979	**0.94979**

Table 2 Social position calculation for the social network from Figure 7; $\varepsilon = 0.5$

$\varepsilon = 0.5$						
Iteration No.	1	2	…	19	20	**21**
***SP*(*x*)**	0.2	0.67	…	1.30447	1.30447	**1.30447**
***SP*(*y*)**	0.2	0.58	…	0.88142	0.88142	**0.88142**
***SP*(*z*)**	0.2	0.58	…	0.96289	0.96289	**0.96289**
***SP*(*u*)**	0.2	0.62	…	1.10816	1.10816	**1.10816**
***SP*(*v*)**	0.2	0.55	…	0.74302	0.74303	**0.74303**

Table 3 Social position calculation for the social network from Figure 7; $\varepsilon = 0.9$

$\varepsilon = 0.9$						
Iteration No.	1	2	…	120	121	**122**
***SP*(*x*)**	0.2	0.406	…	1.51933	1.51933	**1.51933**
***SP*(*y*)**	0.2	0.244	…	0.74265	0.74265	**0.74265**
***SP*(*z*)**	0.2	0.244	…	1.02148	1.02148	**1.02148**
***SP*(*u*)**	0.2	0.316	…	1.19999	1.20000	**1.20000**
***SP*(*v*)**	0.2	0.190	…	0.51651	0.51651	**0.51651**

The data from Table 1 to Table 3 provides the information about the influence of the coefficient ε value on the number of iterations that ought to be performed before the stop condition is fulfilled. It means that the greater ε is, the more iterations must be performed (Table 4).

Social positions for each user from the human community are separated in relation to the value of the coefficient ε, and are presented on the charts (Figure 8). Note that user *x*'s social position is always the highest while user *v* possesses the lowest social position. The highest or lowest values for all the members are reached when ε=0.9. Moreover, the social position of the individual is nearly linearly dependent on the value of ε (see regression lines in the Figure 8).

Table 4 The number of iterations in relation to the value of ε

ε	**Number of iterations**
0.1	9
0.5	21
0.9	122

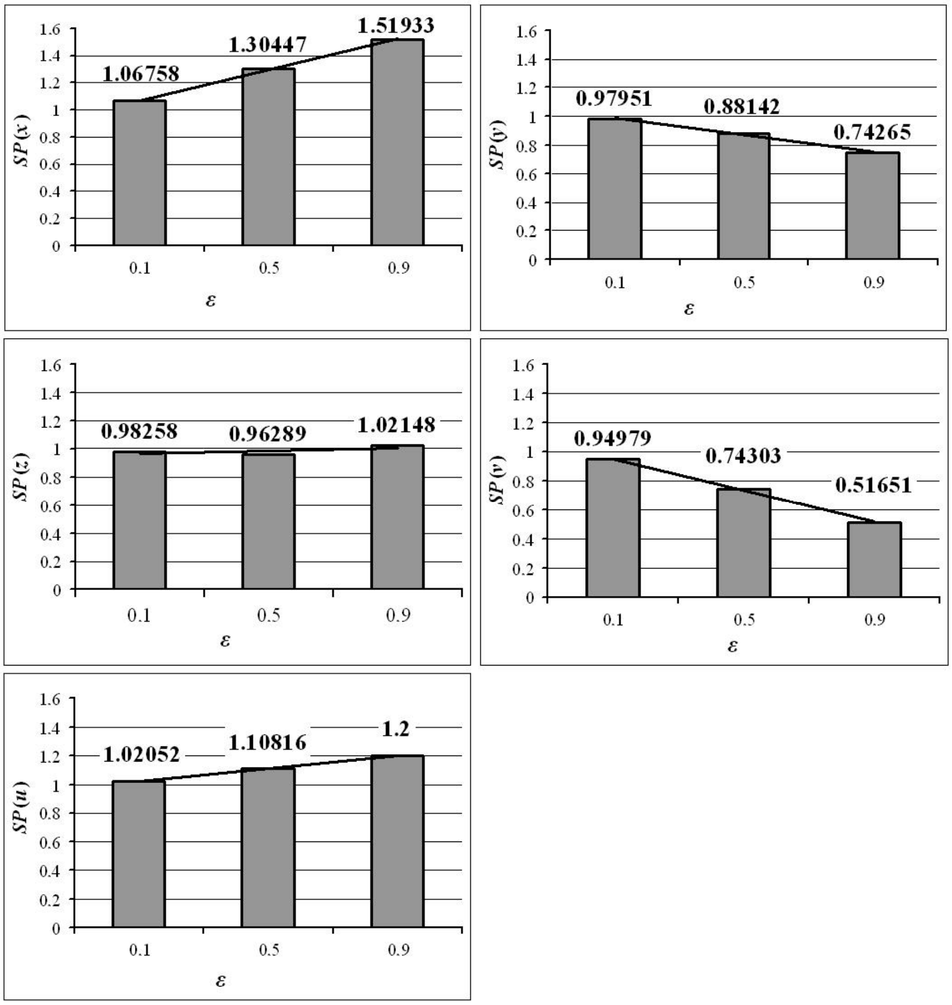

Figure 8 The values of members' social positions in relation to ε value

All the calculations of the social positions are gathered together in Figure 9. If we would need to extract the best representative of the community from Figure 7, then we always would select member *x* and next in order – member *u*. Note that they both have the greatest number of acquaintances – three, as compared to all the others who possess only two (Figure 7).

Some additional information about the influence of the coefficient ε onto the members' social positions provides the average social position within the human community and the standard deviation of the social position's value (Figure 10). If ε is greater, the distance between the minimum and maximum social position within community increases. The next conclusion is that the average social position does not depend on the value of ε. In all cases, it equals around 1 (Figure 10, Figure 11). However, the standard deviation differs depending on the coefficient ε value. The greater ε is, the bigger standard deviation is. Furthermore, the dependence between the value of the coefficient ε and standard deviation is linear (Figure 11).

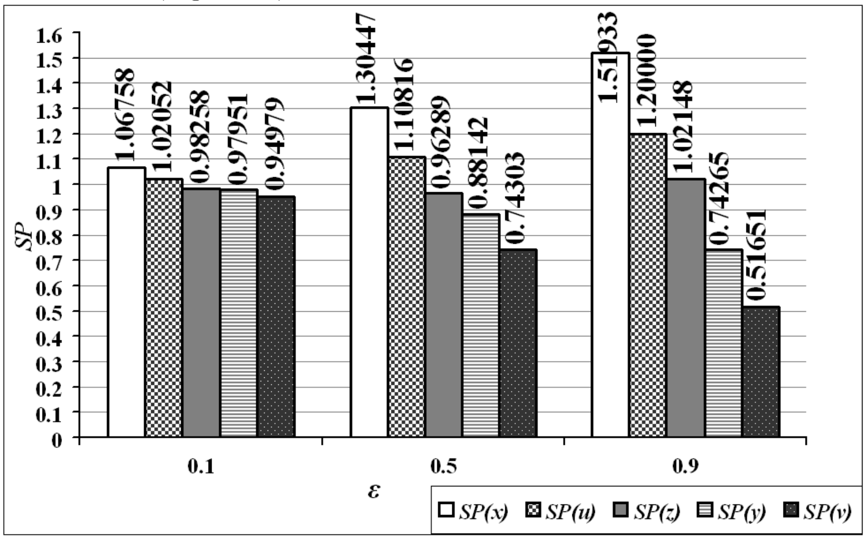

Figure 9 The value of social position in relation to ε

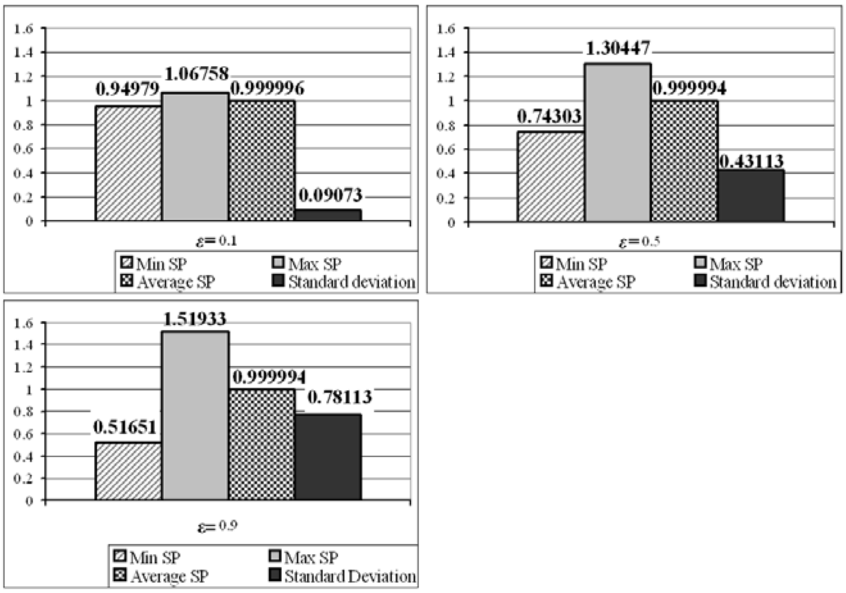

Figure 10 The minimum, maximum, average, and standard deviation of social position calculated for the same community but for different values of ε

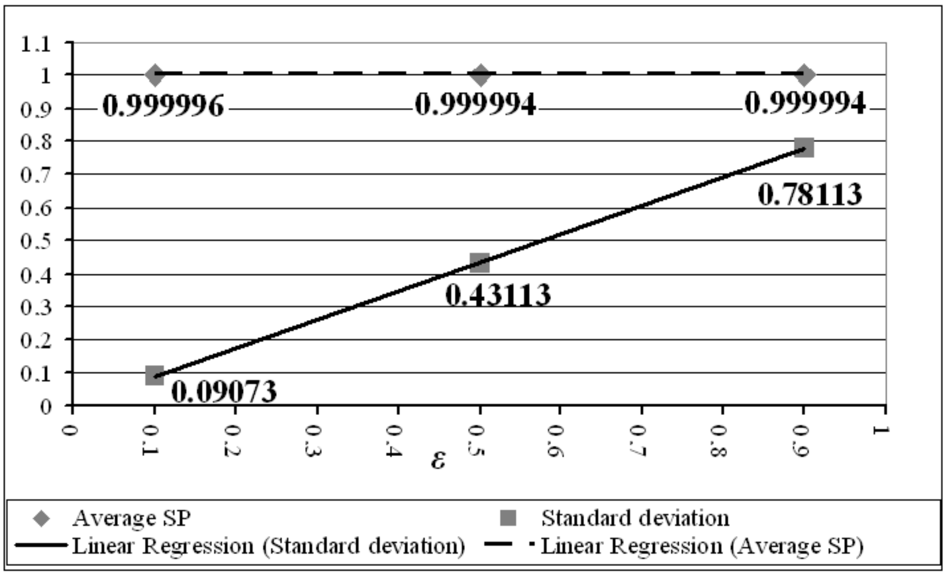

Figure 11 The linear regression for average social position and standard deviation

As it was presented, ε influences the number of iterations and the value of the distance between the members' social positions. The coefficient should be picked out very carefully. On one hand the large number of calculations would slow down the process due to a big number of iterations. On the other hand too few iterations may cause the values of all social positions to be too close to each other.

5.2. The influence of initial social positions

Another issue that is investigated in this paper is the influence of the initial social position values on the number of iterations that must be performed before the stop condition is reached. Two groups of tests have been carried out for the social network in Figure 7 and for two different values of ε, i.e. ε=0.5 and ε=0.9. For each group, five different sets of initial values have been studied: $SP(x) = SP(y) = SP(z) = SP(u) = SP(v) = 0$ (case 1), $SP(x) = SP(y) = SP(z) = SP(u) = SP(v) = 0.2$ (case 3), $SP(x) = SP(y) = SP(z) = SP(u) = SP(v) = 1$ (case 4), $SP(x) = SP(y) = SP(z) = SP(u) = SP(v) = 3$ (case 5). In case 2, initial values have been assigned relatively close to the final ones (Figure 8): $SP(x)$=1.4, $SP(y)$=0.8, $SP(z)$=1, $SP(u)$=1.1, $SP(v)$=0.6. The stop conditions are the same as in the previous calculations. This means that there is no difference in the social position values with precision to the 5-th decimal place for all the members in two following iterations. The results of experiments are presented in Table 5 and Figure 12, for ε=0.5 and in Table 6 and Figure 13, for ε=0.9, respectively.

Table 5 Social position calculation for different sets of their initial values (ε=0.5)

Case	*SP*(*x*)	*SP*(*y*)	*SP*(*z*)	*SP*(*u*)	*SP*(*v*)	Number of iterations
Case 1	0	0	0	0	0	**21**
Case 2	1.4	0.8	1	1.1	0.6	**16**
Case 3	0.2	0.2	0.2	0.2	0.2	**21**
Case 4	1	1	1	1	1	**12**
Case 5	3	3	3	3	3	**20**

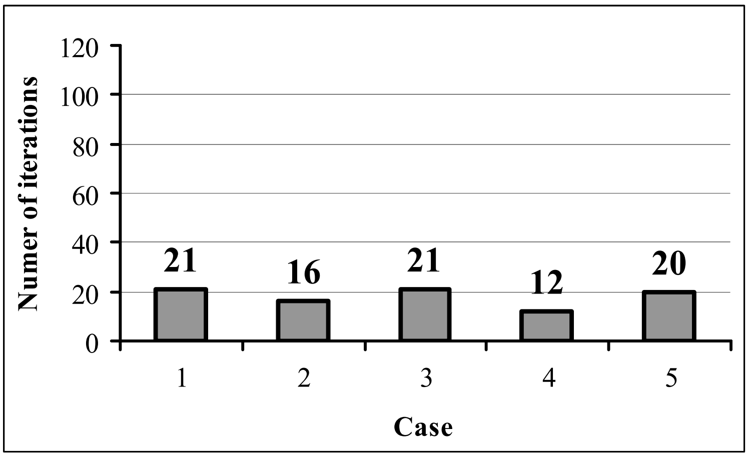

Figure 12 The number of iterations in relation to the initial values of social position in the human community (ε=0.5)

The first thing that should be emphasized is that the final value of the social position is not influenced by the initial values of social positions assignment. In all cases the final values of social positions are very similar. They are exactly the same with a precision to the 4-th decimal place. In consequence, the representatives of the community from Figure 7, who will be selected based on their social position, would always be the same: member *x* and next member *u*; regardless of the initial values.

Table 6 Social position calculation for different sets of their initial values (ε=0.9)

Case	*SP*(*x*)	*SP*(*y*)	*SP*(*z*)	*SP*(*u*)	*SP*(*v*)	Number of iterations
Case 1	0	0	0	0	0	**110**
Case 2	1.4	0.8	1	1.1	0.6	**65**
Case 3	0.2	0.2	0.2	0.2	0.2	**113**
Case 4	1	1	1	1	1	**21**
Case 5	3	3	3	3	3	**118**

The smallest number of iterations is reached when the initial values of social positions for every member equal 1 (case 4). Among all other cases the best result has been achieved when initial values are close to their final values (case 2). In other cases, the number of iterations is significantly larger: between 20 and 21, for ε=0.5 (Figure 12, Table 5) or at least 110, for ε=0.9 (Figure 13, Table 6).

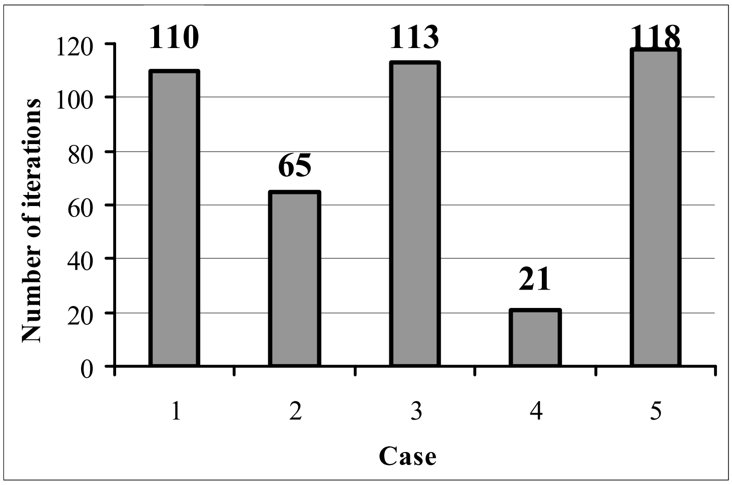

Figure 13 The number of iterations in relation to the initial values of social position in the human community (ε=0.9)

6. Conclusions

The human filtering process can differ depending on the goals of target group selection. Two variants of the process were presented: human filtering by single member and human filtering by social group. In the social group approach we obtain the list of members only from one, previously identified social group, i.e. those who are in relationships with one another. On the other hand, the single member method provides a group of people that are not necessarily related to one another.

The presented solution gives the opportunity to improve the selection of groups regardless the goal of such selection. This framework can be a powerful tool, which can be used to choose people for project teams, find new potential employees, search the consumers' group for advertising campaigns, and finally for use in target marketing. The social position enables to analyze the social network with respect to social aspects of peoples' behaviours.

Acknowledgements

This work was partly supported by the Polish Ministry of Science and Higher Education, grant no. N516 037 31/3708.

References

(Adamic and Adar, 2003) – Adamic, L.A. and Adar, E., 2003, "Friends and Neighbors on the Web", Social Networks, 25 (3), 211-230.

(Adomavicius and Tuzhilin, 2005) – Adomavicius, G. and Tuzhilin, A., 2005, "Toward the Next Generation of Recommender Systems: A Survey of the State–of–the–Art and Possible Extensions", IEEE Transactions on Knowledge and Data Engineering, 17 (6), 734-749.

(Botafogo, Rivlin, and Shneiderman, 1992) – Botafogo, R.A., Rivlin, E., and Shneiderman, B., 1992, "Structural analysis of hypertexts: identifying hierarchies and useful metrics", ACM Transaction on Information Systems, 10(2),, 142-180.

(Boyd, 2004) – Boyd, D.M., 2004, "Friendster and Publicly Articulated Social Networking", CHI 2004, ACM Press, 1279-1282.

(Burt, 2000) – Burt, R.S., 2000, "The Network Structure of Social Capital", Research in Organizational Behavior, 22.

(Carrington, Scott, and Wasserman, 2005) – Carrington, P.J., Scott, J., and Wasserman, S., 2005, "Models and Methods in Social Network Analysis", Cambridge University Press, New York.

(Coleman, 1990) – Coleman, J.S., 1990, "Foundations of Social Theory", Harvard University Press, Cambridge, MA.

(Culotta, Bekkerman, and McCallum, 2004) – Culotta, A., Bekkerman, R., McCallum, A., 2004, "Extracting social networks and contact information from email and the Web", CEAS 2004, First Conference on Email and Anti-Spam.

(Flake, Lawrence, and Giles, 2000) – Flake, G., Lawrence, S., Giles, C.L., 2000, "Efficient identification of web communities", The Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM Press, 150-160.

(Frank, 1995) – Frank, K.A., 1995, "Identifying cohesive subgroups", Social Networks, 17 (1), 27-56.

(Freeman, 1977) – Freeman, L.C., 1977, "A set of measures of centrality based on betweenness", Sociometry 40, 35-41.

(Freeman, 1979) – Freeman, L.C., 1979, "Centrality in social networks: Conceptual clarification", Social Networks, 1(3), 215-239.

(Garton, Haythorntwaite, and Wellman, 1997) – Garton, L., Haythorntwaite, C., and Wellman, B., 1997, "Studying Online Social Networks", Journal of Computer-Mediated Communication, 3 (1).

(Gibson, Kleinberg, and Raghavan, 1998) – Gibson, K.D., Kleinberg, J., Raghavan, P., 1998, "Inferring Web communities from link topology", Hypertext 1998, The Ninth ACM Conference on Hypertext and Hypermedia, ACM Press 1998, 225-234.

(Hanneman and Riddle, 2005) – Hanneman, R., Riddle, M., 2005, "Introduction to social network methods", online textbook, available from Internet <http://faculty.ucr.edu/~hanneman/nettext/> (01.04.2006).

(Hatala, 2006) – Hatala, J.P., 2006, "Social Network Analysis in Human Resources Development: A New Methodology", Human Resource Development Review, 5 (1), 45-71.

(Katz, 1953) – Katz, L. 1953, “A new status derived from sociometrics analysis”, Psychometrica 18, 39-43.

(Kazienko and Musiał, 2006a) – Kazienko, P. and Musiał, K., 2006, “Recommendation Framework for Online Social Networks”, The 4th Atlantic Web Intelligence Conference, AWIC 2006, Studies in Computational Intelligence, Springer Verlag, 111-120.

(Kazienko and Musiał, 2006b) – Kazienko, P. and Musiał, K., 2006, “Social Capital in Online Social Networks”, 10th International Conference on Knowledge-Based Intelligent Information & Engineering Systems, KES 2006, Lecture Notes in Artificial Intelligence LNAI 4252, Springer Verlag, 417-424.

(Kazienko and Musiał, 2007) – Kazienko P., Musiał K., 2007, “Social Position of Individuals in Virtual Social Networks”, submitted to AI Communications, The European Journal on Artificial Intelligence, Special Issue on "Network Analysis in Natural Sciences and Engineering".

(Kumar, *et al.*, 2002) – Kumar, R., Raghavan, P., Rajagopalan, S, and Tomkins, A., 2002, “The Web and Social Networks”, IEEE Computer, 35(11), 32-36.

(Montaner, Lopez, and De la Rosa, 2003) – Montaner, M., Lopez, B., and De la Rosa J.L., 2003, “A Taxonomy of Recommender Agents on the Internet”, Artificial Intelligence Review, Kluwer Academic Publisher, 19, 285-330.

(Putnam, 2000) – Putnam, R. D., 2000. “Bowling Alone: The Collapse and Revival of American Community”, Simon & Schuster, New York.

(Sabidussi, 1966) – Sabidussi, G., 1966, “The centrality index of a graph”, Psychmetrica, 31(4), 581-603.

(Shaw, 1954) – Shaw, M.E., 1954, “Group structure and the behavior of individuals in small groups”, Journal of Psychology, 38, 139-149.

(Shetty and Adibi, 2005) – Shetty, J., Adibi, J., 2005, “Discovering Important Nodes through Graph Entropy The Case of Enron Email Database”, LinkKDD '05, 3rd International Workshop on Link Discovery, KDD 2005, ACM Press 2005, 74-81.

(Tan, Steinbach, and Kumar, 2006) – Tan P.-N., Steinbach M., Kumar V., 2006, “Cluster Analysis: Basic Concepts and Algorithms”, Chapter 8 in “Introduction to Data Mining”, Addison-Wesley, 487-567.

(Wasserman and Faust, 1994) – Wasserman, S., Faust, K., 1994, “Social network analysis: Methods and applications”, Cambridge University Press, New York.

(Yang, *et al.*, 2006) – Yang, W.S., Dia, J.B., Cheng, H.Ch., Lin, H.T., 2006, “Mining Social Networks for Targeted Advertising”, 39th Hawaii International Conference on Systems Science (HICSS-39 2006), IEEE Computer Society 2006.

(Zhu, Chen, and Allen, 2006) – Zhu, W., Chen, C., and Allen, R.B., 2006, “Visualizing an enterprise social network from email”, 6th ACM/IEEE-CS joint Conference on Digital Libraries, ACM Press, 2006, 383.