Social Capital in Online Social Networks

Przemysław Kazienko[1] and Katarzyna Musiał[1, 2]

[1] Wrocław University of Technology, Wyb..Wyspiańskiego 27, 50-370 Wrocław, Poland
[2] Blekinge Institute of Technology, S-372 25 Ronneby, Sweden
kazienko@pwr.wroc.pl, weaving@wp.pl

Abstract. The problem of social capital in context of the online social networks is presented in the paper. Not only the specific elements, which characterize the single person and influence the individual's social capital like static social capital, activity component, and social position, but also the ways of stimulation of the social capital are described.

1 Introduction

Social networks are the information systems available online that have recently been rapidly developed. A social network consists of the set of even millions of people worldwide, who are in mutual relationships, e.g. *Friendster* or *LinkedIn*. Each its participant can be characterized by the value called *social capital* that can be stimulated in order to either make the network more persistent or spread it. In most cases, both, the community as a whole and each individual that belongs to the social network will benefit if the social capital of the members grows [2]. However, definitions of an online social network and social capital are not well established. Some research in this area should be done not only to specify what these concepts are, but also to enable the appropriate evolution of the social network. Stanley Milgram conducted the small-world experiment, which conclusion was that people in USA form the social network and they are connected with "six degrees of separation" [15]. A social network is the set of actors, i.e. group of people or organizations, which are nodes of the network, and ties, called also relationships that link the nodes [1, 9]. Social networks indicate the ways in which actors are related. The nodes and ties are usually presented by graphs (sociograms) or matrices [10]. The evolution of the social network depends on the mutual experience, knowledge, relative interpersonal interests, and trust of human beings [8, 12]. Some measures can be defined to investigate the number and quality of the relationships within the network. The crucial methods, which are currently used to identify the structure of a social network, are: full network method, snowball method, and ego-centric methods [10]. Since there is no established classification of social networks, we propose our own taxonomy of social networks: dedicated (e.g. dating or business networks, networks of friends, graduates, fun clubs), indirect (online communicators, address books, e-mails), common activities (e.g. co-authors of scientific papers, co-organizers of events), local networks (e.g. people living in the neighbourhood), families, employees networks, hyperlink networks (links between home pages), etc. Additionally, the classification of social networks can be based not only

on the type of the relations that occur in the network, but also on the type of the communication between members i.e. they can be either online (virtual, via the computer network) or offline (tangible, with personal contact). Only the first ones will be considered in this paper. Online social networks (called also virtual communities) enable and support the communication between people who are in different places and on different schedules [18]. This makes the relationships not as tangible as those from the real world. The ways of communication within the online networks vary depending on the functionality of the network. The following ones can be distinguished: email, chat, forum, blog, comments, testimonials, photo/movie album, etc. (Fig.1). All social networks are defined by the static attributes of actors like their interests or demographic data as well as the description of the relationships between actors. All this data create the user profile [14], which can be analyzed in order to define and measure the social capital – one of the most important factors of social networks.

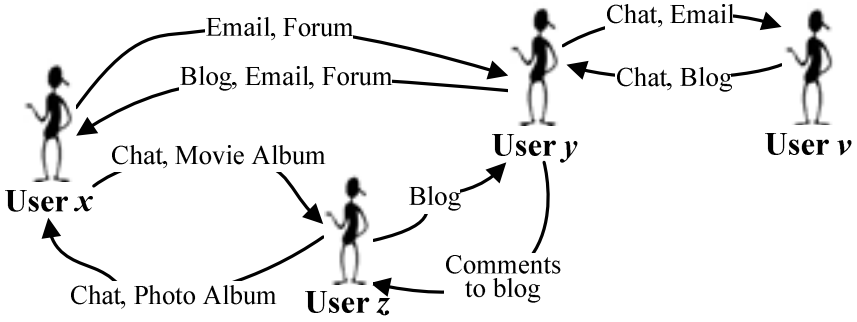

Fig. 1 An example of the online social network

2 Social Capital

There are many approaches to the concept of social capital and each of them presents the social capital in a bit different way [2, 7, 11]. However, all the definitions have something in common – they are implicitly [16, 17] or explicitly [6] functional i.e. social capital is described by its function rather than by its nature. Putnam defined social capital as "features of social organization such as networks, norms, and social trust that facilitate coordination and cooperation for mutual benefits" [17]. The scientist who defined the social capital as explicitly functional concept was Coleman. According to his theory "social capital is defined by its function. It is not a single entity, but a variety of different entities having two characteristics in common: they all consist of some aspect of social structure, and they facilitate certain actions of individuals who are within the structure" [6].

The social capital can be seen as "metaphor about advantage" [5]. In other words, people have to make an effort in order to receive higher returns [5, 11]. In general, social capital usually grows with the use of all its component resources. Although social capital generally brings benefits, it happens sometimes the investment in its building is bigger than the outcomes on the efforts [2]. The social capital of the group that form a social network is aggregation of social capital of all individuals.

The definition that will serve as the basis for the further consideration was formulated by Bourdieu and Wacquant: the social capital is "the sum of the resources, actual or virtual, that accrue to an individual or a group by virtue of possessing a durable network of more or less institutionalized relationships of mutual acquaintance and recognition" [3]. Although this definition does not concern explicitly the online social network, we applied its general idea in our further research.

3 Components of Social Capital in Online Social Networks

3.1. User Social Capital

The user social capital can be defined as the set of features that describe the ability of cooperation between the people The social capital consists of two main parts: static and dynamic (Fig.2). The users themselves deliver the information about the former while the latter is monitored by the system.

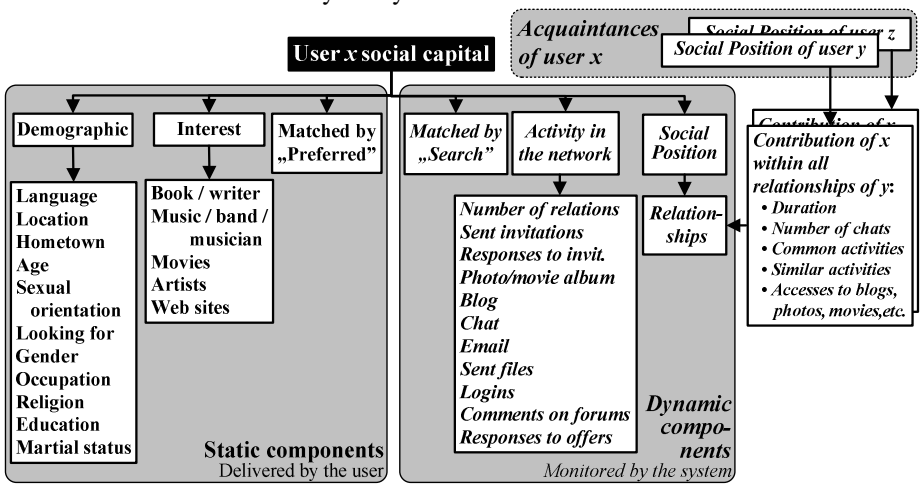

Fig. 2 Social capital of user *x* grouped in components. *Location* is the place of residence, *Hometown* is the place of origin, *Looking for* denotes the purpose of inviting new friends.

The static social capital $S(x)$ provides information about *Interest* and *Demographic* features of user *x*. Additionally, each user can deliver the data about people sought by them and the system counts the number of participants who match these constraints – *Matched by "Preferred"* component. The more overall these conditions are, the more open-minded and flexible the user is, and that is why their social capital should be greater. The users themselves maintain all this data and neither their behaviour within the social network nor the system itself can change these static components.

The dynamic part of social capital depends on the activity and position of the user within social network. Three components of the dynamic social capital for user *x* can be distinguished: *Matched by "Search"* – $MS(x)$, *Activity* – $A(x)$, and *Social Position* – $SP(x)$. The social capital $SK(x)$ for user *x* is defined as follows:

$$SK(x) = \alpha \cdot S(x) + \beta \cdot MS(x) + \gamma \cdot A(x) + \delta \cdot SP(x) \quad (1)$$

where α, β, γ, δ – coefficients of importance for particular components, which are constant at the beginning. Latter on, their values can be recalculated based on the user feedback [13]. The analysis of user choices within the network will be done to establish their most appropriate values. They can be either the same for all the network members or separate for each individual.

The *Matched by "Search" MS*(x) component is derived from all searches made by user x within the network. The set of participants that match search criteria is used to estimate the range of the network that covers user's interest. The more network members fit to performed searches, the more sociable and open-minded the user for new acquaintances is, and in consequence $MS(x)$ should also be higher.

The *Activity* part characterizes the relative activity of the user compared to all other community members. It consists of many components that describe user activity within certain domain like frequency of login, sent emails, invitations, files, chats, as well as total number of relationships, or published multimedia. Some of these elements are calculated for fixed periods (1 month, half a year) – frequencies, e.g. sent emails, while some others are absolute numbers e.g. sent invitations, number of relationships. Nevertheless, all of them are normalized according to the highest values among all network members.

The *Social Position SP*(x) component describes the general position of user x in the network and it can be derived from their relationships. $SP(x)$ increases with the number of acquaintances who are in relationship with x. Moreover, if acquaintance y of user x possesses the high *Social Position SP*(y), then also $SP(x)$ is high. In such case we can say that member x has the "significant" friend y, i.e. x gets on with real authorities. If user y allocates most contribution of their activity directly to user x, then user x is the best friend of y. As a result, the greater part of $SP(y)$ is inherited by user x. This contribution i.e. the significance of x for y consists of many partial activities of y like duration of their acquaintance, number of direct communications e.g. chats, common and similar activities as well as number of accesses of y to the content published by x (blogs, photos, movies). Common activities are for example co-authorships in projects or studies, co-organization of events, etc. Similar activities are derived from usually unintentional meetings of x by y in the network, e.g. comments on the same forums or to the same news, auctions or purchases similar products, trips to the same places, etc.

The values of all the components are summed in (1) and social capital is additive because each of its elements have positive influence on its final value. We also considered multiplication but in this case each of the components would have crucial influence on social capital.

3.2. Static Components

The static component $S(x)$ of the social capital define the user x characteristic, which does not change over time. Note that the conventional definition does not contain such elements since they do not "accrue by virtue of possessing a durable network" [3]. Nevertheless, all of them are the resources which enable to improve individuals' social networks. Based on the *interests* and *demographic* features of the network

member x, the system can find and recommend another person y who will fit their expectations. Obviously, this is mutually beneficial for social capital of both participants x and y. Moreover, some static features can be utilized at direct calculation of social capital, e.g. location. A person from a bigger city is a member of larger local community so they are able to invite more new network members or co-operate with larger number of people also out of the network. People supposed to be more mobile, for whom the country of origin and residence are different, have greater opportunity to benefit from this fact. In online communities the accessibility and quality of the access to the Internet play an important role in assessing the social capital and people from highly developed regions have usually better access. Furthermore, a user who speaks more languages and especially speaks world-wide English has more possibilities to create new relationships. Well-educated people usually do better in the sense of receiving higher return on investment and additionally they are well-connected [5].

The search preferences characterize network members as well. If the participant provides very precise description of the people they want to be in contact, then there may not be enough members who match these expectations.

3.3. Activity of the User

The *Activity* component $A(x)$ of social capital for user x respects the relative frequency of all possible activities performed by user x within the social network, separately for all activity types, like updates of blogs or multimedia albums, number of comments on forums, sent invitations, etc (Fig. 2). $A(x)$ describes how active is user x compared to the most active members in the particular activity type, e.g. in updating their blogs:

$$A(x) = \frac{1}{N} \sum_{i=1}^{N} \frac{A_i(x)}{A_i^{\max}} \qquad (2)$$

where $A_i(x)$ – measure of activity type i for user x; $A_i^{\max}$ – maximum value of activity i; N – number of all activity types.

In addition, the user can be more active in one period and less in another one and the older periods should have less influence on the final measure of activity $A_i(x)$. The precise formula for $A_i(x)$ was presented in [14]. Since all elements $A_i(x)$ and $A_i^{\max}$ are dynamic and change over time, the calculation of $A(x)$ should be periodically repeated for all network members.

3.4. Social Position

The position and value of the user in the network tightly depends on the strength of the relationships that this user maintains. In order to assess the general strength of the relationships of user x *Social Position* function $SP(x)$ has been introduced. Its inspiration was the PageRank algorithm, which is the basic method used by Google to determine the page's relevance or importance [4]. *Social Position* $SP(x)$ of user x respects both $SP(y)$ value of user x acquaintances as well as the activity of y with relation to x:

$$SP(x) = (1 - \varepsilon) + \varepsilon \cdot (SP(y_1) \cdot C(y_1 \rightarrow x) + \ldots + SP(y_m) \cdot C(y_m \rightarrow x)) \qquad (3)$$

where: ε – constant coefficient from the range [0,1]; $y_1, \ldots y_m$ – acquaintances of x, i.e. members who are in the direct relation to x; m – the number of acquaintances of user x; $C(y_1 \rightarrow x), \ldots, C(y_m \rightarrow x)$ – the function that describes the contribution which has user x in the set of the relationships of user $y_1, \ldots, y_m$, respectively.

The general concept of $SP(x)$ is the inheritance of *Social Position* from all acquaintances of user x, especially from those for whom x is the really good friend. This goodness of the friendship is derived from the contribution of acquaintances' activities directed to x. User x possesses the high $SP(x)$ if x would have the real authorities as friends and these authorities would really be in contact with x. Moreover, $SP(x) \in [1-\varepsilon, 1]$ and it equals $1-\varepsilon$ when user x does not have any relationships within the network. $SP(x)$ for all users is calculated iteratively and this process should be repeated periodically. The number of iterations can be fixed to constant value l or the calculation stops when the differences in values of $SP(x)$ between the following iterations are below the given threshold.

Note that the same *Social Position* can be achieved by x if user x would have many relationships with people who have medium $SP(y)$ or if x would have only few relationships but with participants with high $SP(y)$ (Fig.3).

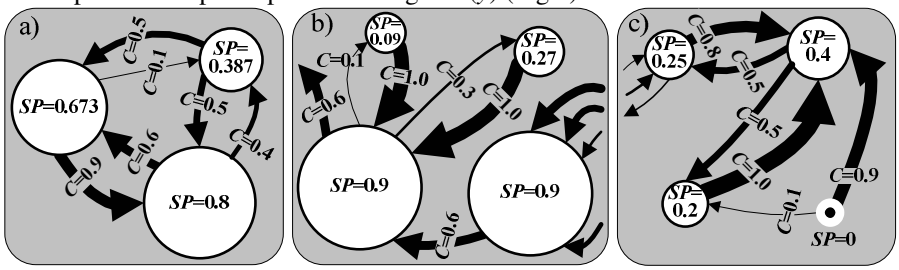

Fig. 3 Examples of social networks with calculated $SP(x)$ and $C(y \rightarrow x)$ for each user, ε=1

The contribution $C(y \rightarrow x)$ of user x within activity of their acquaintance y is the sum of all contacts, cooperation, and communications from y to x in relation to all activities of y:

$$C(y \rightarrow x) = \frac{\mu_1 \cdot \frac{a_1(y \rightarrow x)}{a_1^{sum}(y)} + \mu_2 \cdot \frac{a_2(y \rightarrow x)}{a_2^{sum}(y)} + \ldots + \mu_k \cdot \frac{a_k(y \rightarrow x)}{a_k^{sum}(y)}}{\mu_1 + \mu_2 + \ldots + \mu_k} \quad (4)$$

where: $\mu_1, \ldots, \mu_k$ – fixed coefficients of importance for the particular activity type; k – the number of criteria (activity types) that describe strength of relationship; $a_1(y \rightarrow x)$, $\ldots$, $a_k(y \rightarrow x)$ – the number of activities common for both user x and y of the first, k-th type, respectively, e.g. the number of common projects; $a_1^{sum}(y)$, $\ldots$, $a_k^{sum}(y)$ – the total number of the activities of type 1, k that performed user y.

There is only one requirement for activity functions: $\sum_{j=1}^{m} a_i(y \rightarrow x_j) = a_i^{sum}(y)$.

One of the activity types is communication via chat. In this case, $a_i(y \rightarrow x)$ is the number of chats that are common for x and y; and $a_i^{sum}(y)$ is the number of all chats in which y took part in. If user x has many common chats with y in comparison to the number of all chats with y, then x has greater contribution within activities of y i.e. $C(y \rightarrow x)$ will have greater value and in consequence *Social Position* of user x will grow. Note that $C(y \rightarrow x)$ will have value 1 when user x is the only friend of user y.

However, not all of the elements can be calculated in such simple way. Much more complex are similar activities, e.g. comments on forums. Each forum consists of many threads where people can submit their comments. In this case, $a_i(y \rightarrow x)$ is the number of user's x comments in the threads in which y has also commented, whereas the function $a_i^{sum}(y)$ is the number of comments that have been made by all friends of y on these threads.

4 Social Capital in Data Mining and Recommendation Systems

People behave in the network community similarly to their attitude in the real life. Thus, we can say that the analysis of online communities will deliver the interesting knowledge about human beings. Social capital is the combination of single features that characterize participants within the online society. The research on social capital and its dynamic over the course of time for the entire online social as well as for its users will gain in better understanding of human behaviours and limits.

Based on the historical data related to social capital of individuals in the given network we can build a model for its prediction. It can be also used for prediction of new desirable relationships between network members. The social capital is depleted by lack of communication rather than by activities within relationships. Building new or strengthening the existing relationships causes that both individuals and group will benefit and the social capital will increase. The number of relationships can be increased by tailoring and utilizing the concepts known as bonding and bridging social capital [17]. The goal of the former is the interconnection of two or more homogeneous and similar but separate groups whereas in the latter the different heterogeneous groups are linked. Usually, the joined groups are internally very close. Both bonding and bridging enable to achieve the larger community in which the associations between human beings are permanent. Another approach is to stimulate new relationships within groups. In this way the social network contains many disconnected but very close groups of members that know one another very well. Practically, the creation of new relationships and in consequence the growth of the social capital can be stimulated by various types of recommendation systems that process social capital values using diverse data mining techniques [12, 14].

Association rules and clustering, typical data mining methods can be utilized to identify interesting groups of people in the network, especially those that share the same level of social capital or some of its components: static components, *Activity* or *Social Position*. On the other hand, by means of sequential patterns and time sequences we can analyse general or unusual human behaviours with respect of time. Next, this knowledge could be harnessed at developing more "human sensitive" information systems.

5 Conclusions and Future Work

Social capital is an important indicator of significance and position of particular members within the online social network. The above presented method for estima-

tion of social capital consists of several components that cover wide range of static and dynamic features of network users. It also includes *Social Position* of the participant in the context of the number and strength of direct relationships with other participants as well as *Social Position* of these acquaintances.

The future work will focus on the application of the social capital concept to artificial environments like multi-agent systems and on the development of specialized recommendation systems that would stimulate the expansion of the network [14].

References

1. Adamic L.A., Adar E.: Friends and Neighbors on the Web. Social Networks 25(3) (2003) 211-230.
2. Adler P., Know S.: Social Capital: Prospects for a new concept. Academy of Management Review 27(1) (2002) 17-40.
3. Bourdieu P., Wacquant L. J. D.: An invitation to reflexive sociology. University of Chicago Press (1992).
4. Brin B., Page L.: The Anatomy of a Large-Scale Hypertextual Web Search Engine. WWW7, Computer Networks 30 (1-7) (1998) 107-117.
5. Burt R.S.: The Network Structure of Social Capital. Research in Organizational Behavior 22, JAI Press (2000).
6. Coleman J.S.: Foundations of Social Theory. Harvard University Press, Cambridge, MA, (1990).
7. Fukuyama F.: Social Capital and Development: The Coming Agenda. SAIS Review 22 (1) (2002) 23-37.
8. Golbeck J.: Computing and Applying Trust in Web-Based Social Networks. Ph.D. Thesis. University of Maryland (2005) http://www.cafepress.com/trustnet.20473616.
9. Golbeck J., Hendler J.A.: Accuracy of Metrics for Inferring Trust and Reputation in Semantic Web-Based Social Networks. EKAW 2004, Springer Verlag, LNCS 3257 (2004) 116-131.
10. Hanneman R., Riddle M.: Introduction to social network methods. Online textbook. (2005), http://faculty.ucr.edu/~hanneman/nettext/.
11. Kadushin C.: Too Much Investment in Social Capital? Social Networks 26(1)(2004)75-90.
12. Kazienko P., Kiewra M.: Personalized Recommendation of Web Pages. Chapter 10 in: Nguyen T. (ed.): Intelligent Technologies for Inconsistent Knowledge Processing. Advanced Knowledge International, Adelaide, South Australia (2004) 163-183.
13. Kazienko P., Kołodziejski P.: Personalized Integration of Recommendation Methods for E-commerce. Int. Journal of Computer Science and Applications 3(3) (2006) to appear.
14. Kazienko P., Musiał K.: Recommendation Framework for Online Social Networks. AWIC 2006, Springer Verlag, (2006).
15. Kumar R., Raghavan P., Rajagopalan S., Tomkins A.: Social Networks: From the Web to Knowledge Management. Chapter 17 in Zhong N. Liu J., Yao, Y. (eds): Web Intelligence, Springer-Verlag (2003) 367-379.
16. Lin N.: Social Capital: A Theory of Social Structure and Action. Cambridge University Press, New York (2001).
17. Putnam, R. D.: Bowling Alone: The Collapse and Revival of American Community. Simon & Schuster, New York (2000).
18. Wellman B., Salaff J., Dimitrova D., Garton L., Gulia M., Haythorntwaite C.: Computer Networks As Social Networks: Collaboratie Work, Telework, and Virtual Community. Annual Review of Sociology 22 (1996) 213-238.