

Tracing Strength of Relationships in Social Networks

Ivan Srba and Mária Bieliková

Slovak University of Technology

Faculty of Informatics and Information Technologies

Ilkovičova 3, 842 16 Bratislava, Slovakia

srba07@student.fiit.stuba.sk, bielik@fiit.stuba.sk

Abstract

Current web is known as a space with constantly growing interactivity among its users. It is changing from the data storage into a social interaction place where people not only search interesting information, but also communicate and collaborate. Obviously, social networks are the most used places for common interaction among people. We present a method for analysis of the strength of relationships together with their evolution. This method is based on the various user activities in social networks. We evaluate our approach within the Facebook social network.

1. Introduction and related work

Nowadays we witness a rapid expansion of new generation of services known as the Web 2.0. One of the most popular is the social network where people express their relationships. The strength of these relationships among users differs and can rapidly change. Information about relationship intensity between two users is useful in many areas. For example, we can use it for recommendation based on an assumption that if one user is interested in an object, the other one can be also involved in the same object [1]. It is possible under the condition that the both users are somehow related. Another example is a web page adaptation based on preferences and behavior of the user who we are in a strong relationship with. For preserving of the relationships the user can also control their evolution towards to his important friends.

Recently, several works have been focused on analyzing of the structure and properties of a large-scale of social networks [2], [3], [4]. These studies cover social networks on a high level and provide only general information about nodes and links in the traced social network. They do not contain any information related to relationships between two concrete users. For example, the authors in [5] provide information about users of the social portal YouTube and divide

them into groups following their behavior. Although it is useful, it does not give us enough information about the relationship between two users.

There are only several researches within *activity network* aimed at particular users. The result of research of the activity network in Cyworld social portal was its similarity with the social network [6]. The authors in [8] discovered that the activity network in the social portal Facebook differs from the social network (considering graph properties).

The method of analysis relationships among users based on the Facebook's *wall posts* is described in [9]. It has several limitations - it uses only wall posts and does not involve any other relationship expressions (e.g., photo upload, messages, etc.). The extraction is limited to subset of users of one region (New Orleans Network). It does not use the access to all profiles, only to those publicly available. We consider as a serious limitation the fact that it is not universal and cannot be used in other social networks. The most related research to ours is presented in [10]. The authors proposed an unsupervised latent variable model for the estimation of relationship strength based on the interaction activity and users' similarity.

2. Analysis of relationship intensity

Our method primarily concerns the calculation of relationship intensity strength. However, it is important to guarantee an access to the social network's data and fulfill preconditions in terms of the input data.

2.1. Input data

Partial relationship intensity of the user can be expressed by elementary interaction between two users (e.g. sending a message) or static common information (e.g. common hobby). Such interaction or static information we denote as a *rate factor*. The rate factor represents an identifiable and countable expression of user relationships. It can influence the strength of

a relationship in a positive or negative way depending on the social aspect.

The rate factor can be shared with several *sources* (social networks). The rate factors of each source can have a different importance; numerically it is represented by *weight*. The weight of the rate factor expresses its relative influence on the final relationship intensity. The weight of the rate factor for each source is assigned experimentally (in this paper we provide weights estimated for the Facebook social network). Examples of the rate factors with the weights determined according to analysis and monitoring of user behavior are a common photography (weight 0.13), study at the same school (weight 0.012) and a boyfriend/girlfriend relationship (weight 0.95).

The final relationship strength is also influenced by the *count* of instance of the rate factor (not only in relationship of two users who are traced). This fact is important because the frequency of using of the social network has effect on the partial relationship intensity.

Time influences the partial relationship intensity too. To include effect of time, we differentiate the rate factors of a single activity, an interval activity or an unbounded activity.

The rate factor of a single activity represents a single-shot expression of a relationship. It is possible to identify the date and time of this activity. Furthermore, we can experimentally estimate the duration of influence on the common relationship. An example is sending the message where we know the date and time of its sending and duration of the influence estimated at two days.

The rate factor of an interval activity is used for the activity where we can identify the date and time of beginning and completion. Likewise the previous type we can estimate the duration of its influence. An example is studying at the same school from September 2006 to June 2010 and duration of the influence estimated at 365 days.

The rate factor of an unbounded activity represents a time unbounded activity or some static information for which we cannot define start and end date. An example is a family relationship.

2.2. Calculation of the relationship intensity

Partial relationship intensity depends on the weight, the count of instance of the rate factor and time:

$$I_f(k, j) = \frac{w_{kj} \sum_{i=1}^l f_i}{1 + \ln(1 + l_c)}$$

where I_f is the partial relationship intensity for one rate factor, w_{kj} is the weight of the rate factor j for source k , l is the count of instance of the rate factor in the

relationship of two traced users, l_c is the count of instance of the rate factor, and f_i is the function expressing time influence. The function expressing time influence depends on the type of the rate factor. We identified three time influence functions according to the type of the activity.

Function f_i for the *rate factor of a single activity* is:

$$f_i(t_{kj}, t_m, t_v) = \begin{cases} 1/(1 + \log_{\sqrt{t_{kj}}}(\max(1, t_m - t_v))) & t_m \geq t_v \\ 0 & t_m < t_v \end{cases}$$

where t_{kj} is the time of the influence duration in days, $\sqrt{t_{kj}}$ expresses the duration until the intensity logarithmically decreases to the half of the original value, t_m is the date and time when is the intensity calculated, t_v is the date and time of a single activity, and $t_m - t_v$ is the time difference in days between a single activity and the moment when the intensity is calculated.

Figure 1 illustrates how the time function influences the partial relationship intensity for a single activity. In this case the activity happened on the second day, the influence duration is set to 25 days, so we can see that on the 7th day the function has value 0.5.

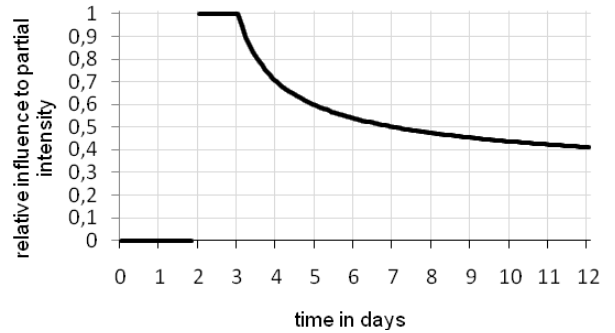


Figure 1. Time function for a single activity.

Function f_i for the *rate factor of an unbounded activity* is:

$$f_i = \frac{1}{2}$$

Function f_i for the *rate factor of an interval activity* is an adapted combination of previous two functions:

$$f_i(t_{kj}, t_m, t_z, t_k) = \begin{cases} 0 & t_m < t_z \\ 1 & t_m \in \langle t_z, t_k \rangle \\ 1/(1 + \log_{\sqrt{t_{kj}}}(\max(1, t_m - t_k))) & t_m > t_k \end{cases}$$

where t_z is the date and time of a beginning interval activity, and t_k is the date and time of an ending interval activity.

Figure 2 illustrates how the time function influences the partial relationship intensity of an interval activity (influence duration is set to 25 days, interval activity starts on the 2nd day and ends on the 5th day).

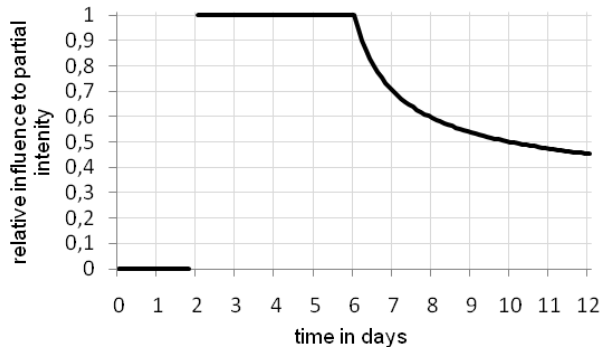


Figure 2. Time function for an interval activity.

The partial relationship intensity for one source is a summation of all values of the partial relationship intensity. The final relationship intensity is computed as arithmetic average of the partial relationship intensity of all sources.

3. Evaluation

We developed a web-based application for computing the strength of relationships in social networks. In our design we constrain neither the social network nor the rate factors.

For experimenting we chose the popular social portal Facebook. We were researching (i) distribution of user interaction among friends, (ii) evolution of the calculated intensity of the relationships during a time period, (iii) comparison of calculated results with the expected ones (the first ten best friends). The results in this paper are based on the activity of 238 users with 25,435 relationships.

3.1. User interaction among friends

At first we investigated the distribution of wall posts and all activities (rate factors of a single activity) among friends. We calculated the cumulative distribution function (CDF) for this purpose (see Figure 3). About 82% of relationships have no wall activities. This result was expected and indicates that only a small subset of the users make majority of wall posts. About 50% of relationships have no occurrence of the rate factor for a single activity. This means that the graph representing an activity network differs from the graph representing a social network. This result also shows that distribution of wall posts is quite similar but not equal to the distribution of all activities.

Secondly, we summed up the intensity of the single activities of each user. Then we selected three values (70%, 90% and 100%) of the summed intensity and calculated the portion of the relationships involved in

the required part of the intensity. Figure 4 presents a cumulative distribution function for this percentage. This figure gives us a lot of useful information, e.g. information that for all users 90% of the intensity is created only by 54% of the relationships.

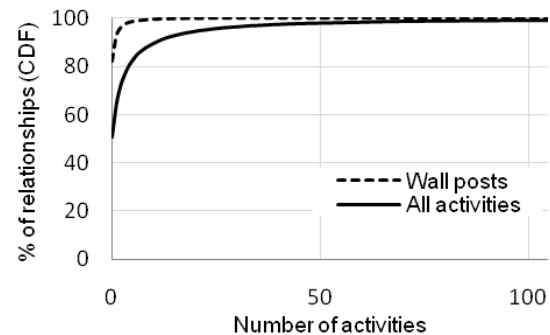


Figure 3. Distribution of user activities among friends.

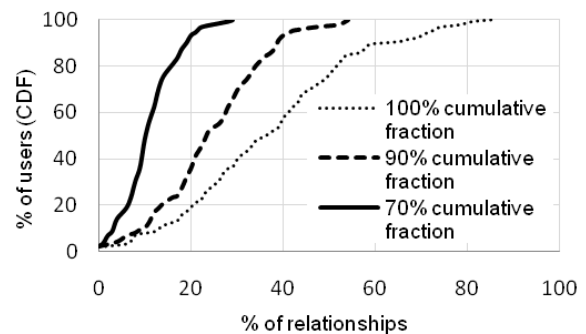


Figure 4. Distribution of relationship intensity among friends.

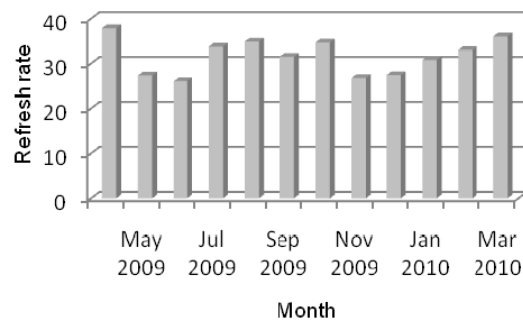


Figure 5. Refresh rate (one month interval).

3.2. User interaction during a time period

To evaluate distribution of user interaction during a time period we calculated *refresh rate* of the relationships within one year with a month interval (see Figure 5) in the same way as described in [9].

The result is that only 31% of user pairs continue interacting in the following month. This result is

almost the same as result 29% achieved in [9]. It proves that users regularly change friends involved in interaction activities.

To inspect dynamics of the relationships during a time period we calculated a relative change. We noticed big differences among more and less active relationships, so we divided them into five groups according to the calculated intensity. The results are plotted in Figure 6. Average relative change of all relationships is 17%. The results show that highly coupled users do not change their relationships so often and dramatically as the other do.

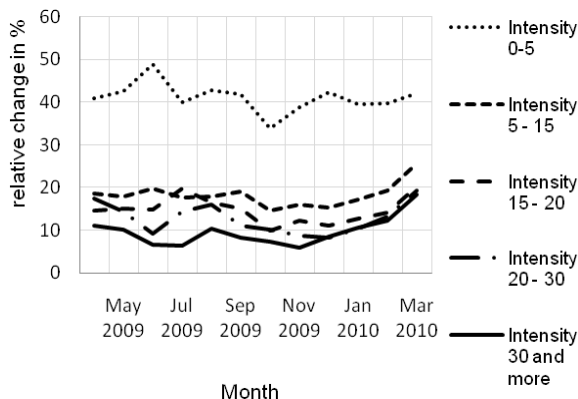


Figure 6. Relative intensity change for different groups of relationships.

3.3. User reception of calculated results

We also got an explicit feedback from users. We compared the list of top ten best friends chosen by users with the list of 10 friends who had the highest calculated relationship intensity. Evaluating 95 acquired lists of the friends chosen by the users we found their 88% overlap in both lists. The result is also influenced by the fact that default weights and time of duration of some rate factors are not appropriate for all traced users.

4. Conclusion

In this paper we propose and evaluate the method for analysis of relationship intensity in social networks. Approximation of user relationships presents valued information about users who share the same interests, the same ways of spending free time, etc. If we want to gain more accurate information for individual users, we can use the estimated relationship strengths in many ways, for example for the adaptation of the web page content in collaborative filtering methods either independently or in combination with social based recommendation [7].

Acknowledgements. This work was supported by grants No. VG1/0508/09, No. 028-025STU-4/2010, and it is a partial result of the Research & Development Operational Program for the project Support of Center of Excellence for Smart Technologies, Systems and Services II, ITMS 25240120029, co-funded by ERDF.

5. References

- [1] Herlocker, J.L., Konstan, J.A., Riedl, J.: Explaining collaborative filtering recommendations. In: *Proc. of the 2000 ACM Conf. on Computer Supported Cooperative Work*, Philadelphia (USA): ACM New York, 2000, pp. 241-250.
- [2] Hanneman, R., Riddle, M.: *Introduction to social network methods* [online], Riverside (California): Univ. of California, 2005. <http://faculty.ucr.edu/~hanneman/nettext> (09.02. 2010)
- [3] Mislove, A., Marcon, M., Gummadi, K. P., Druschel, P., Bhattacharjee, B.: Measurement and Analysis of Online Social Networks. In: *Proc. of the 7th ACM SIGCOMM Conf.*, San Diego (California): ACM New York, 2007, pp. 29 - 42.
- [4] Ahn, Y., Han, S., Kwak, H., Moon, S., Jeong, H.: Analysis of topological characteristics of huge online social networking services. In: *Proc. of the Int. Conf. on WWW*, Alberta (Canada): ACM New York, 2007, pp. 835 - 844.
- [5] Maia, M., Alemida, J., Almeida, V.: Identifying user behavior in online social networks. In: *Proc. of the 1st Workshop on Social Network Systems*, Glasgow (Scotland): ACM New York, 2008, pp. 1-6.
- [6] Chun, H., Ahn, Y., Eom, Y., Kwak, H., Moon, S., Jeong, H.: Comparison of online social relations in volume vs interaction: a case study of cyworld. In: *Proc. of the 8th ACM SIGCOMM Conf. on Internet Measurement*, Vouliagmeni (Greece): ACM New York, 2008, pp. 57-70.
- [7] Suchal, J., Návrát, P.: Full text search engine as scalable k-nearest neighbor recommendation system. In *Proc. of the Artificial Intelligence in Theory and Practice*, WCC 2010, Brisbane (Australia): Springer Boston, 2010.
- [8] Wilson, C., Boe, B., Sala, A.: User interactions in social networks and their implications. In: *Proc. of the 4th ACM European Conf. on Computer Systems*, Nuremberg (Germany): ACM New York, 2009, pp. 205-218.
- [9] Viswanath, B., Mislove, A., Cha, M., Gummadi, K.P.: On the evolution of user interaction in Facebook, In: *Proc. of the 2nd ACM Workshop on Online Social Networks*, Barcelona (Spain): ACM New York, 2009, pp. 37-42.
- [10] Xiang, R., Neville, M., Rogati, M.: Modeling relationship strength in online social networks, In: *Proc. of the 19th Int. Conf. on WWW*, Raleigh (North Carolina): ACM New York, 2010, pp. 981-990.