

Experiences with phonebook-centric social networks

Péter Ekler

Department of Automation and Applied Informatics
Budapest University of Technology and Economics
peter.ekler@aut.bme.hu

Tamás Lukovszki

Faculty of Informatics
Eötvös Loránd University
lukovszki@inf.elte.hu

Abstract- Social networks are becoming more and more popular nowadays. The increasing capabilities of mobile phones enable them to participate in such networks. The phonebooks in the mobile devices represent social relationships, that can be integrated in the social networks. Following we refer to this solution as phonebook-centric social networks. Such networks provide a synchronization mechanism between phonebooks of the users and the social network which allows detecting network members listed in the phonebooks (semi-) automatically. The detection of those members is based on the similarity of the personal data, e.g., similar name, same phone number, address, etc... Users can mark similarities between their phonebook contacts and members in the network. After this, if one of their contacts changes her or his personal detail, it will be propagated automatically into the phonebooks, after considering privacy settings. Synchronization time and energy consumption for synchronization on the mobiles and the scalability of the system strongly depend on the number of users and similarities. We have implemented a phonebook-centric social network, called *Phonebookmark* and investigated the structure of the network. We experienced that the distribution of similarities follows a power law, as well as the distribution of the in- and out-degrees in the social network. In this paper we propose a model for estimating the total number of similarities and we show that this estimation applies very well to the historical data of *Phonebookmark*.

Keywords: social networks, mobile phones, power law distribution

I. INTRODUCTION

The popularity of social networks is constantly increasing. Several social networks appeared and attracted thousands or even millions of users. The most popular online social networks Facebook [12] and Myspace [16] are among the top ten visited websites on the Internet [4]. The basic idea behind such networks was that users can manage personal relationships on these networks. It is an environment created by the people who are using it. The nodes of a social network are individuals or organizations and the edges represent social relations between them.

Mobile phones and mobile applications are another hot topic nowadays. The increasing capabilities of mobile devices allow them to participate in social network applications as well. Mobile phone support in general social networks are usually limited mainly to photo and video upload capabilities and access to the social network using the mobile web browser. Since the phonebook of the mobile device also describe the social relationships of its owner, discovering additional relations in social networks is beneficial for sharing personal data or other content. Given an implementation that allows us to upload as well as download

our contacts to and from the social networking application, we can completely keep our contacts synchronized so that we can see all of our contacts on the mobile phone as well as on the web interface. In addition to that if the system detects that some of my private contacts in the phonebook is similar to another registered members of the social network (i.e. may identify the same person), it can discover and suggest social relationships automatically. In the rest of this paper we refer to this solution as a *phonebook-centric social network*.

Discovering and handling similarities in phonebook-centric social networks is a key issue. If a member changes some of her or his detail, it should be propagated in every phonebook to which she or he is related after considering privacy settings. The algorithm for detecting and handling similarities consumes resources on server side and the number of similarities strongly influences the scalability of the system. From the mobile phone point of view it does not requires additional resources, however with the help of detected similarities the system can keep the phonebooks always up-to-date.

The rest of the paper is organized as follows. Section 2 describes related work in the field of social networks and power law distributions appearing in such networks. Section 3 discusses the structure of phonebook-centric social networks and introduces our implementation, called *Phonebookmark*. Section 4 shows measurements related to *Phonebookmark*. Section 5 proposes a model how to estimate the total number of accepted similarities in such networks. Finally, Section 6 summarizes the paper.

II. RELATED WORK

Huge amount of papers and popular books, such as Barabási's Linked [6] study the structure and principles of dynamically evolving large scale networks like the Internet and networks of social interactions. In [14] the authors discuss about that nowadays social networking on mobile phones is not only a buzz term for today's enthusiasts but also provides real possibilities to the users. Many features of social processes and the Internet are governed by power law distributions. Following the terminology in [13] a nonnegative random variable X is said to have a power law distribution if $\Pr[X \geq x] \sim cx^{-\alpha}$, for constant $c > 0$ and $\alpha > 0$. In a power law distribution asymptotically the tails fall according to the power α , which leads to much heavier tails than other common models.

Distributions with an inverse polynomial tail have been first observed in 1897 by Pareto [18] (see. [19]), while describing

the distribution of income in the population. Zipf observed similar statistical behavior in the distribution of inhabitants in cities [22].

In [7] the graph structure of the Web has been investigated and it was shown that the distribution of in- and out-degree of the web graph and the size of weekly and strongly connected components are well approximated by power law distributions. Nazir et al. [17] showed that the in- and out-degree distribution of the interaction graph of the studied Myspace applications also follow such distributions. Those distributions also approximate the degree distribution of the Gnutella network [19]. Crovella et al. [9] observed power law distributions in the sizes of files and transmission times in the Internet.

There has been a great deal of theoretical work on designing random graph models that result in a Web-like graph. Barabási and Albert [5] describe the preferential attachment model, where the graph grows continuously by inserting nodes, where new node establishes a link to an older node with a probability which is proportional to the current degree of the older node. Bollobás et al. [7] analyze this process rigorously and show that the degree distribution of the resulting graph follow a power law. Another model based on a local optimization process is described by Fabrikant et al. [13]. Mitzenmacher [15] gives an excellent survey on the history and generative models for power law distributions. Aiello et al. [3] studies random graphs with power law degree distribution and derives interesting structural properties in such graphs.

III. PHONEBOOK-CENTRIC SOCIAL NETWORKS

In a simple social network the nodes are the registered users and links represent social relationships between them. A phonebook-centric social network [10] is based on the idea that the phonebook of the mobile phones also represents social relationships of the owners. It allows for the registered members to synchronize their phonebooks to the social network. By synchronization, the phonebook of the mobile phone is automatically updated with the latest information provided by the friends of its owner.

In addition to that, the private contacts are also uploaded to the phonebook-centric social network. These contacts are not visible to others. However, having all of the contacts in the system has the following benefits:

- The contacts can be managed (list, view, edit, call, etc.) from a web browser.
- The service notices the user if duplicate contacts are detected in its phonebook and warns about it.
- The contacts are safely backed up in case the phone gets lost.
- The contacts can be easily transferred to a new phone if the user replaces the old one.
- The phonebook can be shared between multiple phones, if one happens to use more than one phone.
- It is not necessary to explicitly search for the friends in the service, because it notices if there are members similar to the contacts in the phonebooks and warns about it.

In order to analyze these networks we have implemented with Nokia Siemens Networks a phonebook-centric social network, called *Phonebookmark*. Before public introduction it was available for a group of general users from April to December of 2008. It had 420 registered members with more than 72000 private contacts. During this period we have collected different type of data related to the social network which was the base of the measurements and the proposed model in this paper. The structure of phonebook-centric social networks is illustrated on Figure 1.

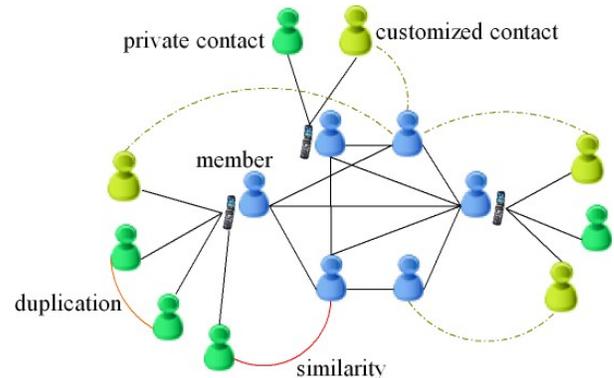


Figure 1. The structure of a phonebook-centric social network.

A *member* is a registered user of the social network. They can upload their contact list to the social network and maintain a backup phonebook there. *Private contacts* correspond to phonebook entries of the members. However, these private contacts are not shared between members. A private contact is transferred into the system when a member synchronizes her or his phonebook with the social network.

In a phonebook-centric social network it is possible that one of our private contacts in our phonebook is similar to a member of the network, i.e. they have the same phone number or similar names, etc... Following we will refer to this as a *similarity*. Similarities potentially identify the same person. A similarity detection algorithm allows us to detect and accept similarities in the network and recommend possible relationships for the members. In addition to that this algorithm enables also to recognize duplications in phonebooks.

Phonebookmark provides a semi-automatic similarity detecting and resolving mechanism. First it detects similarities and calculates a similarity weight for them, which indicates, how likely the entries identify the same person. (Figure 2).

After a detected similarity is being selected, *Phonebookmark* provides a user interface where the details of the two people can be merged. Here the user can choose whether to accept or ignore the similarity, which is the base of the semi-automatic behavior (Figure 3).



Figure 2. Handling multiple similarities



Figure 3. Semi-automatic similarity resolution

During the operational period of our *Phonebookmark*, the algorithm detected 1200 similarities and users have accepted more than 90% of these (1088), which is an encouraging number for analyzing the distribution of similarities and propose a model for it. Following we refer to this rate as $P_R \sim 0.9$. *Phonebookmark* was implemented within a project founded by Nokia Siemens Networks and it is the property of NSN.

If a member accepts a similarity between another member and her or his private contact a new *customized contact* appears. This way the owner can edit this contact in her or his phonebook but if the referred member changes her or his profile, the change will be propagated to the customized contact after considering the privacy settings. However this propagation will take effect only if the owner member has not edited that specific profile detail yet. We will refer to this mechanism as customization. This customization mechanism enables that members can edit their phonebook, but their phonebooks can also get updates from the network. The number of edges between members and customized contacts is the number of accepted similarities.

IV. MEASUREMENTS IN PHONEBOOKMARK

Distributions of in- and out-degrees

The first experiment we conducted was to analyze the distribution of in- and out-degree of the members in *Phonebookmark*. The in- (out-) degree of a member is the number of edges to (from) other members in the network. The distribution of in- and out-degrees are important to understand the basic structure of the network. We experienced, that the distribution of both in- and out-degrees follow a power law.

If X has a power law distribution, then on a log-log plot of $\Pr[X \geq x]$, also known as the *complementary cumulative distribution function*, asymptotically the behavior will be a

straight line. This provides a simple empirical test for whether a random variable has a power law given an appropriate sample. In this case the gradient of the log-log function is the α parameter of the given power law distribution:

$$\ln(\Pr[X \geq x]) = -\alpha \ln(x) + \ln(c) \quad (1)$$

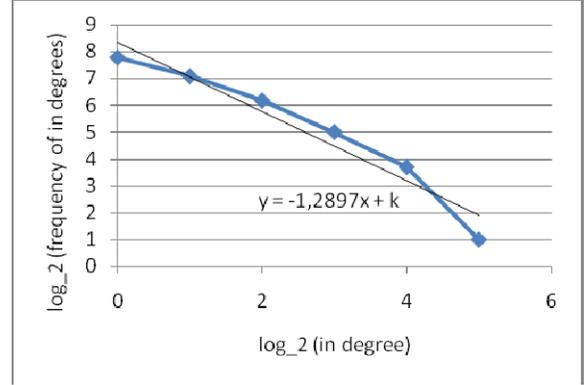


Figure 4. Distribution of in-degree

Figure 4 shows the complementary cumulative distribution function of in-degrees using logarithmically scaled x - and y -axis. The x -axis represents the in-degree and the y -axis the number of members with at least this in-degree. Figure 4 only shows the frequency of nodes with in-degree of at least one. There are also nodes with in-degree zero. Their frequency is not depicted, since the logarithm of zero is minus infinity. Figure 4 also shows the line obtained by the least squares fitting technique. The slope of this line corresponds to the exponent of the power law distribution. The exponent we obtain is $\alpha=1.2897$.

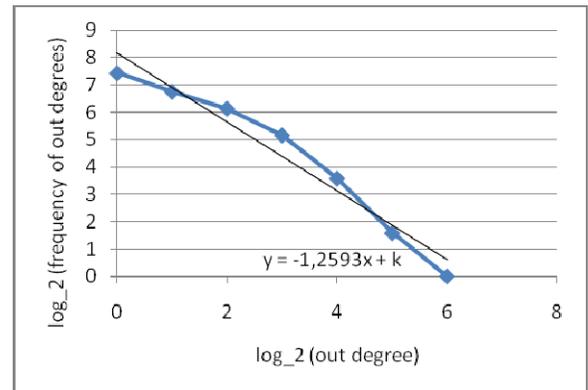


Figure 5. Distribution of out-degree

Figure 5 shows the complementary cumulative distribution function of out-degrees. This distribution also follows a power law with an exponent $\alpha=1.2593$.

Distribution of phonebook sizes

Figure 6 shows the tail distribution of the phonebook-sizes such that the x -axis has linear scale and the y -axis logarithmic scale. The points on this figure fit very well to a line, which

means that the tail of the phonebook sizes decreases exponentially. This is a big contrast to the inverse polynomial tail of in- and out-degrees.

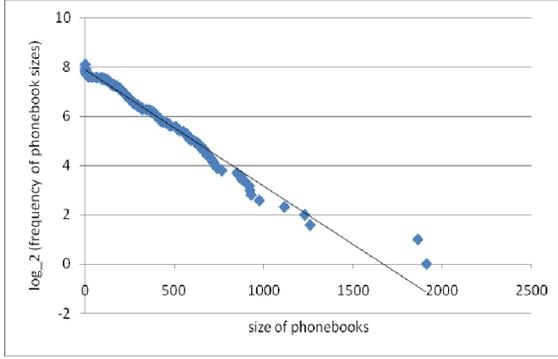


Figure 6. Size of phonebooks in Phonebookmark

Distribution of similarities

Based on the database and database logs of *Phonebookmark* we managed to measure the distribution of similarities raised by a member during registration and first phonebook synchronization.

Figure 7 shows the complementary cumulative distribution function of the number of similarities, where the x -axis is the number of similarities and the y -axis means how many people arises at least that amount of similarities when registers and synchronizes.

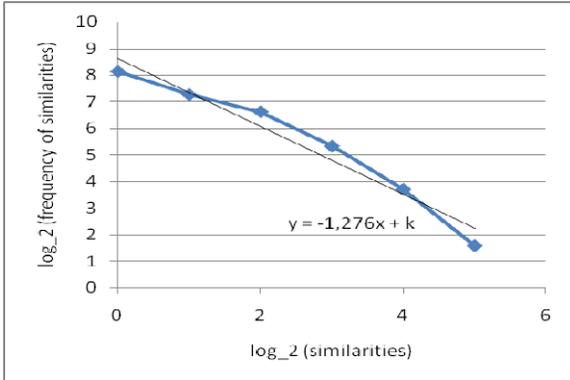


Figure 7. Distribution of similarities

Again we use logarithmically scaled x - and y -axis. We can see on Figure 6 that the points are close to a line, thus the distribution of similarities can be well approximated by a power law. The exponent of the power law distribution is $\alpha=1.276$.

According to this measurement the distribution of similarities in our case can be well approximated as follows:

$$\Pr[X \geq x] \sim x^{-1.276}. \quad (2)$$

The evidence that the distribution of the similarities follows a power law has practical consequences. The expected number of users involving at least a certain number of similarities x

can be estimated by $N_M \Pr[X \geq x] \sim N_M x^{-1.276}$, where N_M is the number of members in the network.

It worth to note that the tail of the distribution of phonebooks decreases exponentially and very large phonebooks are very unlikely. But a new member can cause a much higher number of similarities than the size of his or her phonebook, since he or she can appear in the phonebook of many other members.

V. ESTIMATION OF THE TOTAL NUMBER OF SIMILARITIES

According to the empirical results of the previous section we model the number of similarities generated during a member registration by a probability variable X . More precisely, X models the number of similarities proposed by the automatic similarity detection algorithm.

The total number of accepted similarities N_S in a phonebook-centric social network can be estimated with the following formula:

$$N_S = N_M E[X] P_R, \quad (3)$$

where N_M is the number of registered members and P_R is the rate of the similarities accepted by the users, as described in Section 3. In order to calculate $E[X]$, we need the probabilities $\Pr[X=x]$, which can be obtained from the complementary cumulative distribution function $\Pr[X \geq x] \sim cx^{-\alpha}$ by derivation (see, e.g. [1, 2]):

$$\Pr[X = x] \sim c' x^{-(\alpha+1)}. \quad (4)$$

In order to be a probability distribution, $\sum_{x=0}^{\infty} c' x^{-(\alpha+1)} = 1$. Thus, $c' = 1/\zeta(\alpha+1)$, where $\zeta(\cdot)$ denotes the Riemann Zeta function. Then the expected value is:

$$\begin{aligned} E[X] &= \sum_{x=0}^{\infty} x \Pr[X = x] \\ &= \sum_{x=0}^{\infty} x \frac{1}{\zeta(\alpha+1)} x^{-(\alpha+1)} \\ &= \frac{1}{\zeta(\alpha+1)} \sum_{x=0}^{\infty} x^{-\alpha} = \frac{\zeta(\alpha)}{\zeta(\alpha+1)}. \end{aligned} \quad (5)$$

The expected total number of accepted similarities N_S in a phonebook-centric social network can be estimated with the following formula:

$$N_S = N_M \frac{\zeta(\alpha)}{\zeta(\alpha+1)} P_R. \quad (6)$$

For $\alpha > 1$, $\zeta(\alpha)/\zeta(\alpha+1)$ is a finite constant. In our case, for $\alpha=1.276$, we obtain that the expected total number of similarities is

$$N_s = 2.9196 * 420 * 0.9 = 1103.$$

Measured similarities

During the operation of *Phonebookmark* the trends of users and similarities was measured. According to these measurements we can verify the previous model. Figure 7 and 8 illustrates that although the number of users increases, the average number of accepted similarities per user remains on the same level.

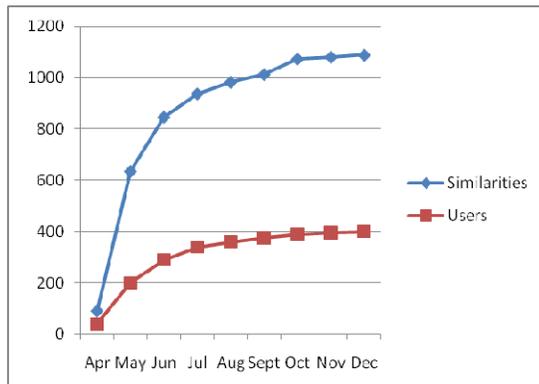


Figure 7. Trend of similarities and users in Phonebookmark

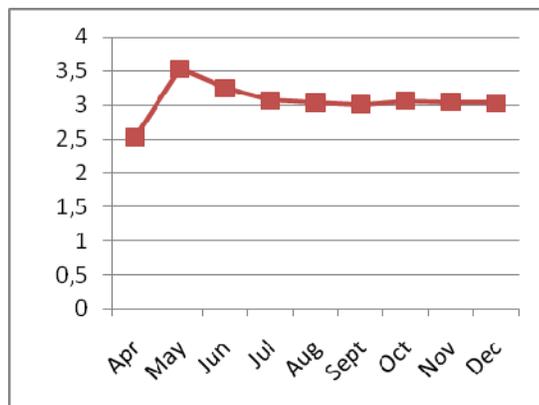


Figure 8. Average similarity per user in Phonebookmark

According to the previous model and the measurements, the total number of similarities increases linearly with the number of members.

VI. SUMMARY

Phonebook-centric social networks belong to the new generation of social networks. They also involve the social relationships defined by the phonebook entries into the network.

In this paper we investigated the structure of phonebook-centric social networks and we have introduced our implementation, called *Phonebookmark*. The system was available for a group of general users from April to December of 2008. It had 420 registered members with more than 72000 private contacts. Based on measurements from this period we have experienced that the distribution of in- and out-degrees follow a power law. In contrast to this, we found that the tail

of the distribution of phonebook sizes decreases exponentially. The distribution of similarities also follows a power law.

As a main contribution of this paper we have proposed a model that enables to estimate the total number of accepted similarities in phonebook-centric social networks depending on the number of registered members N_M . We have showed that the proposed model fits well to our measurements.

Acknowledgement: The authors thank to Balázs Bakos, Zoltán Ivánfi for their support from Nokia Siemens Networks

REFERENCES

- [1] L. A. Adamic. Zipf, power-law, Pareto -- a ranking tutorial. <http://www.hpl.hp.com/research/idl/papers/ranking>, 2000.
- [2] L. A. Adamic, B.A. Huberman. Zipf's law and the Internet. *Glottometrics*, Vol. 3, 143-150, 2002.
- [3] W. Aiello, F. R. K. Chung, L. Lu. A random graph model for massive graphs. In: *Proc. 32nd Symposium on Theory of Computing STOC*, 171-180, 2000.
- [4] Alexa. http://www.alexa.com/site/ds/top_sites. August 2009.
- [5] A.-L. Barabási, R. Albert. Emergence and scaling in random networks. *Science*, Vol. 286, 509-512, 1999.
- [6] A.-L. Barabási. *Linked: How Everything Is Connected to Everything Else*, Perseus Publishing, 2002.
- [7] B. Bollobás, O. Riordan, J. Spencer, G. Tusnady. The degree sequence of a scale-free random graph process. *Random Structures and Algorithms*, Vol. 18(3), 279-290, 2001.
- [8] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener. Graph structure in the Web. In: *Proc. 9th International World Wide Web Conference on Computer Networks*, 2000.
- [9] M. E. Crovella, M. S. Taqqu and A. Bestavros. Heavy-Tailed Probability Distributions in the World Wide Web. In: R. J. Adler, R. E. Feldman, M. S. Taqqu (eds.), *A Practical Guide To Heavy Tails*. 1, 3-26. *Chapman and Hall*, New York, 1998.
- [10] P. Ekler, T. Lukovszki. Similarity Distribution in Phonebook-centric Social Networks. In: *5th International Conference on Wireless and Mobile Communications (ICWMC)*. 2009.
- [11] P. Ekler, Z. Ivánfi, K. Aczél. Similarity Management in Phonebook-Centric Social Networks. In: *International Conference on Internet and Web Applications and Services (ICIW)*, 2009.
- [12] Facebook. <http://www.facebook.com>. August 2009.
- [13] A. Fabrikant, E. Koutsoupias, and C. H. Papadimitriou. Heuristically Optimized Trade-offs: A New Paradigm for Power Laws in the Internet. In: *Proc. 29th International Colloquium on Automata, Languages and Programming (ICALP)*, 110-122, 2002.
- [14] B. Forstner, I. Kelényi. Mobile Social networking – Beyond the Hype. In: *Mobile Peer to Peer: A Tutorial Guide*. Wiley, 161-190. 2009.
- [15] M. Mitzenmacher. A brief history of generative models for power law and lognormal distributions. *Internet Mathematics*, Vol. 1, 225-251, 2001.
- [16] Myspace. <http://www.myspace.com>. August 2009.
- [17] Nazir, S. Raza and C.-N. Chuah. Unveiling Facebook: A Measurement Study of Social Network Based Applications. In: *Proc. ACM Internet Measurement Conference (IMC)*, 43-56, 2008.
- [18] V. Pareto. *Course d'economie politique professé à l'université de Lausanne*, 3 volumes, 1896-7.
- [19] M. Ripeanu, I. Foster, and A. Iamnitch. Mapping the Gnutella Network: Properties of Large-Scale Peer-to-Peer Systems and Implications for System Design. *IEEE Internet Computing Journal*, Vol. 6(1), 50-57, 2002.
- [20] G. K. Zipf. Human behavior and the principle of least effort. *Addison-Wesley*, 1949.