Cognitive-costed agent model of the microblogging network

Mitsuhiro Nakamura and Hiroshi Deguchi

Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology, 4259 Nagatsuta-cho, Midori-ku Yokohama, 226-8502 Japan

Abstract. Microblogging is a new paradigm spreading in social web services that provides us a light-weight, speedy way of communication. In the microblogging system, users post short messages just as quickly as chatting. Users can easily communicate with each other. However, the microblogging brings users huge cognitive costs since they always need to follow up their friends' posts every second. Can such cognitive costs affect the structure of the microblogging network? Here we extracted data from the most major microblogging web service: *Twitter*. We find that the network structure in *Twitter* has two main features: link reciprocity and asymmetricity between distributions of the in-degree and out-degree. To explain such characteristics, we introduce a simple cognitive-costed agent model based on the Barabási-Albert model. With the mathematical analysis and computational experiments, we confirm that even such a simple model explains well the behavior of the observed data.

Keywords. complex network, social web service, microblogging, cognitive cost

1 Introduction

Microblogging is a new paradigm spreading in social web services that provides us a light-weight, speedy way of communication. Recently, more and more services have been implementing the microblogging systems, like *Twitter*[1], *Jaiku*[2], and so on. In these systems, users post short messages (usually less than 140 characters) as quickly as chatting via PC, mobile, or cell phone. Users can communicate and share information in the blink of an eye. Furthermore, they can add the other users as their friends therefore the microblogging system is a class of social networks.

Many private companies and public organizations plan to make use of the microblogging service as a marketing medium e.g. public relations channel and infrastructure for gathering latent customer information. For

[1] `http://twitter.com`
[2] `http://www.jaiku.com`

such purposes, it is important to understand the structure and function of the microblogging system. Despite that, there have been still few studies investigating microblogging systems [1, 2].

In this paper, we adopt a point of view of complex network science and attempt to understand properties of the microblogging network. We make our analysis based on the actual data obtained from *Twitter*, which is the first and most popular microblogging service. By studying some statistical features of the data, we introduce a simple stochastic agent model of the microblogging network. With mathematical analysis and agent-based simulation, we show that even such a simple model can explain macroscopic properties in the observation. Finally we discuss several issues contained in the model and point out directions for future research.

2 Data Analysis

Microblogging is an instance of the social network and *Twitter* shows ordinary characteristics already observed in the other networks [1] such as small-world and scale-free features [3–6]. Contrasting with the other social networking services such as *Mixi*[3], *Twitter* is quite unique because of its directedness when viewed as a network. In *Twitter*'s network, like the other social networks, vertices represent users and edges represent friendships between users. Edges have directions, "following" or "followed", and thus friendships can be either one-way or two-way (see Fig. 1).

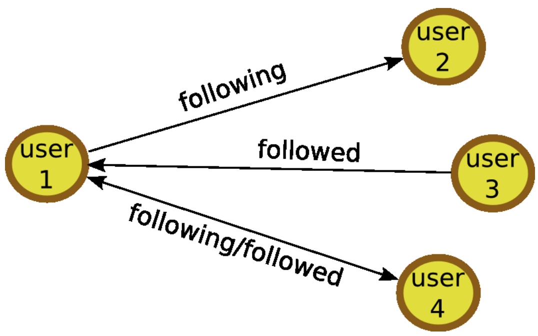
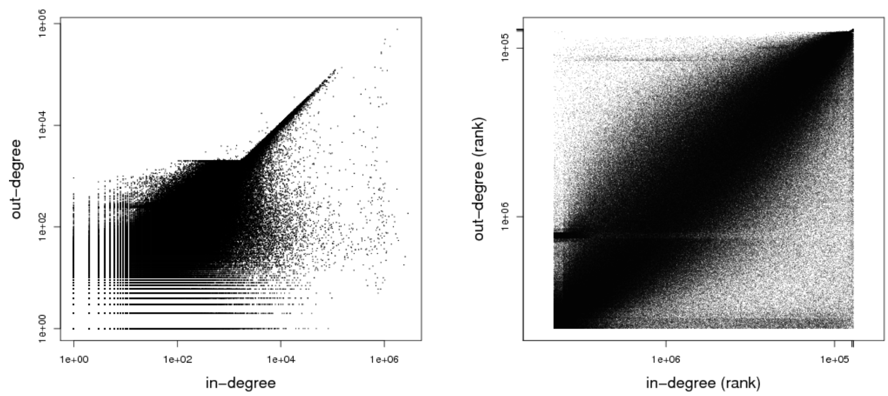


Fig. 1. Following/followed friendships in *Twitter*. To add a user as a friend is called "follow". User 1 "follows" user 2 and user 4, while "followed" by user 3 and user 4. Only user 1 and user 4 have a mutual friendship and the others are all one-sided. When a user follows another user, the former one follows up the latter one's updating of messages.

[3] `http://mixi.jp`

Here we analyze fetched data set (about 1.5 million users, obtained in July 2009) from *Twitter*'s data feed and study statistical properties of the network, especially degree distributions. We promise that a user's in-degree means the number of users he/she is "followed" by and out-degree indicates the number of users he/she is "following".

Fig. 2 shows strong correlation between the in-degree and out-degree. Most users have the same order of the in-degree and out-degree. This



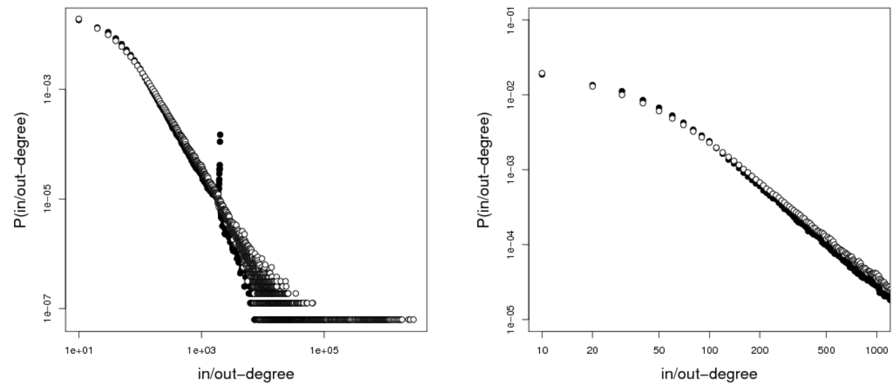
(a) Log-log scatter plot between in/out-degrees

(b) Rank scatter plot between in/out-degrees

Fig. 2. Log-log and rank scatter plots between the in-degree and out-degree. Kendall's $\tau = .64$. Vertices with huge degrees have high ranks.

feature is caused by link reciprocity: tendency of vertex pairs to form mutual connections between each other [7]. In many social networking services, creation of a link from one user to another tend to cause creation of the reverse link to be established [8, 9]. Such a behavior of users is called "reciprocation". Additionally, in Fig. 2(a), users in the right lower-triangular part present asymmetricity between their in-degree and out-degree, while few users are in the left upper-trianguler part.

Fig. 3 displays the in-degree and out-degree distributions that follow power-law (scale-free) statistics. Fig. 3(b) is a zoom-up of Fig. 3(a). We observed that the out-degree is huger than the in-degree in the small-degree range ($k < 10^2$), whereas the out-degree is smaller than the in-degree in the large-degree range ($k > 10^2$). What mechanism causes such phenomena?



(a) In/out-degree distributions (b) Zoom-up of the small-degree range

Fig. 3. Log-log plots of the in-degree and out-degree distributions. Blank points for in-degree and filled points for out-degree. Both of them exhibit scale-free behavior in the large-degree range, while showing a power-law decay in the small-degree range. A peculiar peak around $k = 2,000$ in the out-degree distribution is human-made [10].

3 Model

We now try to get further understanding about the phenomena reported above. To explain observed statistical behavior, basically we apply a rate equation approach to the extended Barabási-Albert model [11]. We first focus on a feature observed in the actual network: reciprocation. We also pay attension to the cognitive cost brought by the microblogging system.

3.1 Simple reciprocal model

A growth model of the directed network which considers vertices' reciprocation was studied by Ref. [12]. Here we briefly introduce such a stochastic agent model.

We start with one vertex in the network. In each time step, a new vertex i is added to the network and create an edge to an existing vertex j with probability $\Pi(k_j)$, which is proposal to j's in-degree k_j (preferential attachment). Then, the vertex j creates a return edge to the vertex i with probability ρ_j, where the constant ρ_j is determined with each vertex.

Applying continuous approximation, we get the following time evolution about the average in-degree.

$$\frac{\partial \bar{k}(s,t)}{\partial t} = \Pi(\bar{k}(s,t)) = \frac{\bar{k}(s,t)}{\displaystyle\int_0^t du\ \bar{k}(u,t)} = \frac{1}{1+\langle\rho\rangle}\frac{\bar{k}(s,t)}{t}\ ,\ \bar{k}(s,s) = \langle\rho\rangle\ , \tag{1}$$

where $\bar{k}(s,t)$ means the average in-degree of a vertex at any time t which is added at the time s and $\langle\rho\rangle$ denotes the expected value over time of ρ_i.

The solution of Eq. (1) is,

$$\bar{k}(s,t) = \langle\rho\rangle \left(\frac{s}{t}\right)^{-\frac{1}{1+\langle\rho\rangle}}\ . \tag{2}$$

Hence, at the equilibrium time, the cumulative in-degree distribution becomes,

$$P_<(k) = \Pr\left[\bar{k}(s,t) < k\right] = \Pr\left[\frac{s}{t} > \left(\frac{\langle\rho\rangle}{k}\right)^{1+\langle\rho\rangle}\right] = 1 - \left(\frac{k}{\langle\rho\rangle}\right)^{-(1+\langle\rho\rangle)}\ . \tag{3}$$

Then we obtain the in-degree distribution,

$$p(k) = \frac{\partial}{\partial k}P_<(k) \propto k^{-(2+\langle\rho\rangle)}\ . \tag{4}$$

We also get the out-degree distribution $p(q) \propto q^{-(2+\langle\rho\rangle)}$ from the boundary condition $\bar{q}(s,s) = 1$ and the following relation,

$$\frac{\partial \bar{q}(s,t)}{\partial t} = \langle\rho\rangle \frac{\partial \bar{k}(s,t)}{\partial t}\ . \tag{5}$$

The out-degree distribution is thiner than the in-degree distribution on account of the multiplicative factor $\langle\rho\rangle$ in Eq. (5). This result is statistically consistent with the observation in Fig. 3.

3.2 Reciprocal model with cognitive cost

We extend the reciprocal model described above with considering the effect of cognitive cost.

In *Twitter*, posted messages from one's "following" friends appear on his/her "timeline", which keeps updating every second (see Fig. 4). Consequently, the more the number of following users increases, the faster and the faster the timeline flows. Thus large amounts of following users

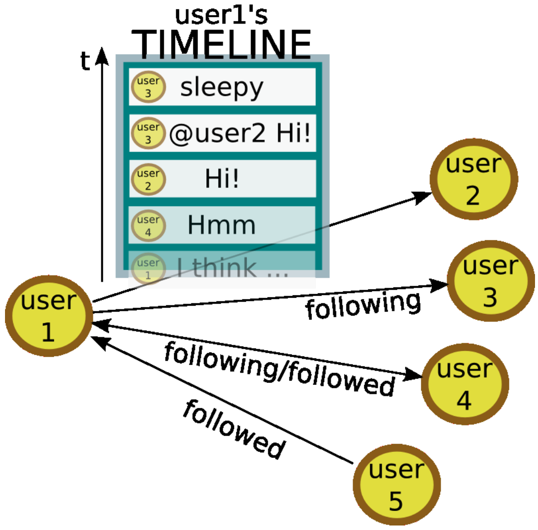


Fig. 4. The concept of "timeline". Posted messages from user 1's following users (user 2, user 3, user 4) keep real-time updating on a second scale. Updates from not-following users (user 5) will not appear on the timeline.

result in difficulty following up friends' posts. We regard it as that the cognitive cost increases with the out-degree.

We assume that the merginal cognitive cost is diminishing. The functional form of the cognitive cost should be,

$$C(q) \propto q^{\alpha} \ , \tag{6}$$

where q is the out-degree and $\alpha \in (0, 1)$ is a constant.

Consider the following situation. In each time step, a new vertex i is added to the network with probability p, then the vertex i creates a new edge to an old vertex j with probability $\Pi(k_j)$. The vertex j reciprocates the vertex i with probability ρ_j. On the other hand, with probability $\bar{p} \equiv 1 - p$, an old vertex i is chosen with probability Ψ_i, then the vertex i creates an edge to another old vertex j in accordance with the preferential attachment rule and the vertex j reciprocates in the same manner. Let Ψ_i depend on the cognitive cost $C(q_i)$. We simply assume that $\Psi_i \propto C(q_i)^{-\beta}$ where $\beta > 0$, since vertices with smaller cost should create more out-going edges. Therefore $\Psi_i = \Psi(q_i) \propto q_i^{-\kappa}$ where $\kappa \equiv \alpha\beta > 0$.

Here we derives degree distributions in the same way in Sec. 3.1. The time evolution of the average in-degree consists of the following three cases:

1. A new vertex is added with probability p, then one gets a new edge from the new vertex with probability $\Pi(\bar{k})$.
2. One creates a new edge with probability $\bar{p}\Psi(\bar{q})$, then receives reciprocation with probability $\langle\rho\rangle$.
3. One gets a new edge from an old vertex with probability $\bar{p}\left(1-\Psi(\bar{q})\right)\Pi(\bar{k})$.

Thus,

$$\begin{aligned}\frac{\partial \bar{k}}{\partial t} &= p\Pi(\bar{k}) + \bar{p}\left[\Psi(\bar{q})\langle\rho\rangle + \left(1-\Psi(\bar{q})\right)\Pi(\bar{k})\right] \\ &= \Pi(\bar{k}) + \bar{p}\langle\rho\rangle\Psi(\bar{q}) - \bar{p}\Pi(\bar{k})\Psi(\bar{q}) \ . \end{aligned} \tag{7}$$

In just the same way, the time evolution of the average out-degree is the summation of the following three cases:

1. A new vertex is added with probability p, then one gets a new edge from the new vertex with probability $\Pi(\bar{k})$, after that the one reciprocates with probability $\langle\rho\rangle$.
2. One creates a new edge with probability $\bar{p}\Psi(\bar{q})$.
3. One gets a new edge from an old vertex with probability $\bar{p}\left(1-\Psi(\bar{q})\right)\Pi(\bar{k})$, then reciprocates with probability $\langle\rho\rangle$.

Thus,

$$\begin{aligned}\frac{\partial \bar{q}}{\partial t} &= p\Pi(\bar{k})\langle\rho\rangle + \bar{p}\left[\Psi(\bar{q}) + \left(1-\Psi(\bar{q})\right)\Pi(\bar{k})\langle\rho\rangle\right] \\ &= \langle\rho\rangle\Pi(\bar{k}) + \bar{p}\Psi(\bar{q}) - \bar{p}\langle\rho\rangle\Pi(\bar{k})\Psi(\bar{q}) \ . \end{aligned} \tag{8}$$

For Eqs. (7) (8), obtaining,

$$\frac{\partial \bar{q}}{\partial t} = \langle\rho\rangle\frac{\partial \bar{k}}{\partial t} + \bar{p}\left(1-\langle\rho\rangle^2\right)\Psi(\bar{q}) \ . \tag{9}$$

As a consequence of Eq. (9), large enough $\langle\rho\rangle$ results in that the average in-degree and out-degree have strong correlation. Assuming such a condition, we can ignore $\Psi(\bar{q})$ in the large-degree range and thus Eqs (7) (8) become,

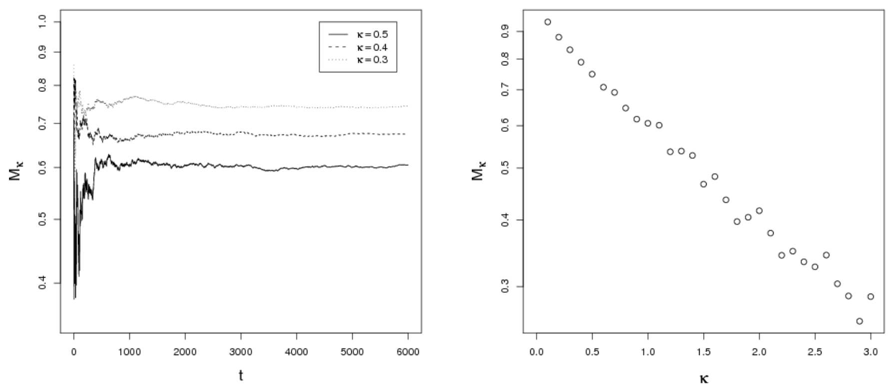
$$\left\{\begin{array}{rcll}\dfrac{\partial \bar{k}}{\partial t} &=& \Pi(\bar{k}) \ , & \bar{k}(s,s) = \langle\rho\rangle \\ \dfrac{\partial \bar{q}}{\partial t} &=& \langle\rho\rangle\Pi(\bar{k}) \ , & \bar{q}(s,s) = 1\end{array}\right. . \tag{10}$$

As the same as Sec. 3.1, we get $p(k) \propto k^{-(2+\langle\rho\rangle)}$ and $p(q) \propto q^{-(2+\langle\rho\rangle)}$.

In the small-degree range, we can ignore $\Pi(\bar{k})$. Therefore,

$$\begin{cases} \dfrac{\partial \bar{k}}{\partial t} = & \bar{p}\langle\rho\rangle\Psi(\bar{q}) \ , \quad \bar{k}(s,s) = \langle\rho\rangle \\ \dfrac{\partial \bar{q}}{\partial t} = & \bar{p}\Psi(\bar{q}) \ , \qquad \bar{q}(s,s) = 1 \end{cases} . \tag{11}$$

$\Psi(\bar{q})$ has $\sum_j q_j^{-\kappa}$ at its denominator, whose behavior is unknown. We adopt a mean-field approximation and rewrite it as $M_\kappa t$, where $M_\kappa \equiv \left\langle q_j^{-\kappa} \right\rangle_j$. This is supported by a numerical evaluation (see Fig. 5).



(a) Evaluation of M_κ (b) κ-dependency of M_κ

Fig. 5. (a) Evaluation of M_κ. At large enough time, M_κ becomes stable. (b) κ-dependency of M_κ. Only the vertical axis is plotted as logarithm. M_κ is inferred to be a form of exponential function.

Applying this approximation to Eqs. 11, we obtain,

$$p(q) = \frac{M_\kappa}{\bar{p}} q^\kappa \exp\left[-\frac{M_\kappa}{\bar{p}} \frac{q^{\kappa+1}-1}{\kappa+1}\right] \ , \tag{12}$$

$$p(k) = \frac{M_\kappa}{\bar{p}\langle\rho\rangle^{\kappa+1}} k^\kappa \exp\left[-\frac{M_\kappa}{\bar{p}\langle\rho\rangle^{\kappa+1}} \frac{k^{\kappa+1}-1}{\kappa+1}\right] \ . \tag{13}$$

Eqs. (12) (13) do not follow the power law. Replacing $M_\kappa/\bar{p}$ in Eq. (12) and $M_\kappa/\bar{p}\langle\rho\rangle^{\kappa+1}$ in Eq. (13) as θ, these functions behave as shown in Fig. 6. Small θ (namely, small p and large κ) makes those distributions have

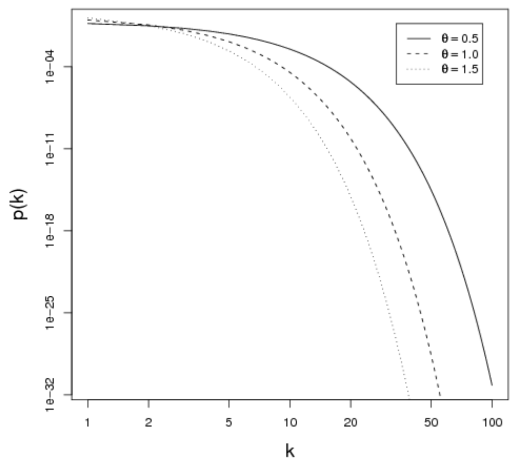


Fig. 6. Log-log plots of the functional form of Eqs. (12) (13).

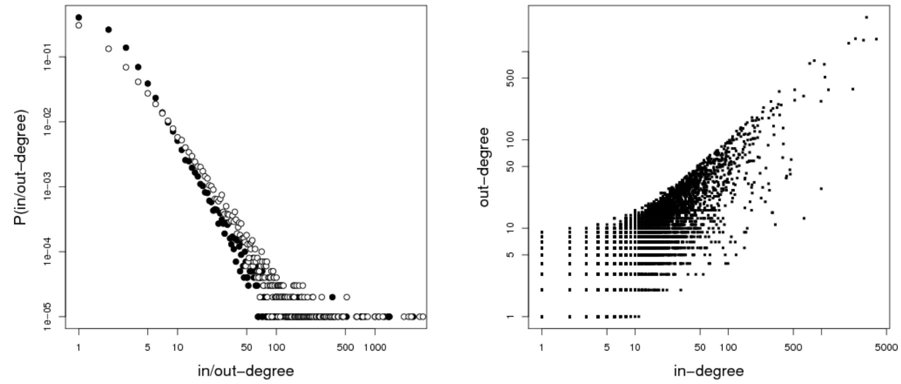
gentle slopes. Moreover, the factor $\langle\rho\rangle^{\kappa+1}$ in Eq. (13) affects the asymmetricity between the in-degree and out-degree distributions. Smaller $\langle\rho\rangle$ makes the in-degree distribution sharper.

Fig. 7 shows results of a numerical simulation. As the actual observation, the out-degree distribution is thicker than the in-degree distribution in the small-degree range ($k < 10$), while the out-degree distribution is thinner than the in-degree distribution in the large-degree range ($k > 10$). Additionally, the in-degree and out-degree are asymmetric as shown in Fig. 7(b).

4 Discussion and Conclusions

In summary, we have taken a look at the actual microblogging network and introduced a simple stochastic agent model containing two features: users' reciprocation and the effect of cognitive cost. Despite its simplicity, the model succeeds to reproduce structural properties observed in the experimental data and thus it should capture some key characteristics of the behavior of users in the real microblogging systems.

On the other hand, some assumptions in our model failed to catch the reality in the actual system. For example, we assumed that the probability of reciprocation ρ_i is constant, while it is suggested that ρ_i depends on the



(a) In/out-degree distributions (b) Scatter plot between in/out-degrees

Fig. 7. Results of a numerical simulation, where $p = 0.5$, $\kappa = 0.5$, and ρ_i obeys a uniform distribution in $[0, 1]$. The number of vertices is $100,000$. (a) The in-degree and out-degree distributions. Blank points for the in-degree and filled points for the out-degree. (b) Scatter plot between the in-degree and out-degree.

out-degree in the real microblogging service. In future, We will enhance our model for such inconsistencies.

Microblogging is a new social web service which will be having significant importance in the next generation. Toward application in marketing, we will continue further research for understanding of such a social web system.

5 Acknowledgements

We thank *Twitter Inc.* for providing an API and allowing us to fetch data sets from their amazing service. We also thank *Leave a nest Co.,Ltd* for a research grant, and Oscar Alonso for reading the manuscript.

References

1. Java, A., Song, X., Finin, T., Tseng, B.: Why we twitter: understanding microblogging usage and communities. In: Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis, San Jose, California, ACM (2007) 56–65
2. Cheng, A., Evans, M.: Inside twitter. `http://www.sysomos.com/insidetwitter/` (June 2009)
3. Albert, R., Barabási, A.: Statistical mechanics of complex networks. Reviews of Modern Physics **74**(1) (2002) 47

4. Dorogovtsev, S.N., Mendes, J.F.F.: Evolution of networks. Advances In Physics **51**(4) (2002) 1079–1187
5. Newman, M.E.J.: The structure and function of complex networks. Arxiv preprint cond-mat/0303516 (2003)
6. Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., Hwang, D.: Complex networks: Structure and dynamics. Physics Reports **424**(4-5) (February 2006) 175–308
7. Garlaschelli, D., Loffredo, M.I.: Patterns of link reciprocity in directed networks. Physical Review Letters **93**(26) (December 2004) 268701
8. Kumar, R., Novak, J., Tomkins, A.: Structure and evolution of online social networks. In: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, Philadelphia, PA, USA, ACM (2006) 611–617
9. Mislove, A., Koppula, H.S., Gummadi, K.P., Druschel, P., Bhattacharjee, B.: Growth of the flickr social network. In: Proceedings of the first workshop on Online social networks, Seattle, WA, USA, ACM (2008) 25–30
10. Schonfeld, E.: Twitter's 2000-follow limit raises a ruckus. but how many people can you seriously watch anyway? `http://www.techcrunch.com/2008/08/12/twitters-2000-follow-limit-raises-a-ruckus-but-how-many-people-can-you-seriously-keep-track-of-anyway/` (August 2008)
11. Barabási, A., Albert, R.: Emergence of scaling in random networks. Science **286**(5439) (October 1999) 509–512
12. Zlatić, V., Štefančić, H.: Influence of reciprocal arcs on the degree distribution and degree correlations. 0902.3542 (February 2009)