"Winner takes it all": Strongest node rule for evolution of scale-free networks

H. Štefančić and V. Zlatić
Theoretical Physics Division, Rudjer Bošković Institute, P.O. Box 180, HR-10002 Zagreb, Croatia
(Received 29 September 2004; revised manuscript received 23 May 2005; published 7 September 2005)

We study a model for evolution of complex networks. We introduce information filtering for reduction of the number of available nodes to a randomly chosen sample, as a stochastic component of evolution. New nodes are attached to the nodes that have maximal degree in the sample. This is a deterministic component of network evolution process. This fact is unusual for evolution of scale-free networks and depicts a possible route for modeling network growth. We present both simulations and theoretical results for network evolution. The obtained degree distributions exhibit an obvious power-law behavior in the middle with the exponential cut off in the end. This highlights the essential characteristics of information filtering in the network growth mechanisms.

DOI: 10.1103/PhysRevE.72.036105 PACS number(s): 89.75.Hc, 02.50.Cw, 05.10.Gg, 89.75.Da

I. INTRODUCTION

Recently, there have been a number of extensive investigations in the field of complex networks. With such an extensive effort, a number of important theoretical and practical results have been reported [1–3]. Many real-world systems can be described as complex networks: the world wide web [4], internet routers [5–7], proteins [8], and scientific collaborations [9], among others. The main features that separate complex networks from "ordinary" networks are the famous small-world effect [10] and the scale-free degree distribution [11].

The first and simplest model for the scale-free distribution of degrees in a complex network was proposed by Albert and Barabási [12] (thereafter referred to as the BA model). This model is based on a simple principle of preferential attachment. The network grows in such a way that at each time step t a new node is introduced into the network and attaches itself to some of older nodes designated by the moment s when they entered the network. The probability that the node t will attach itself to a node s is linearly proportional to the degree k_s of the older node $P_{t\rightarrow s} \sim k_s$. Using this simple principle, a scale-free network of exponent 3 is easily reconstructed. Although very appealing because of its simplicity, the BA model cannot correctly reproduce all characteristics of real-world networks. First, it produces a temporally correlated network in the sense that older nodes tend to have more edges than the younger ones. This was not observed in the real data [13]. Second, it assumes that every new node has the complete information about the whole network, which is unrealistic for real network formations [14,15]. Third, in its original form, it reproduces only networks with degree distribution characterized by exponent $\gamma=3$. Nevertheless, the BA model has triggered a huge number of models that try to avoid these shortcomings, but are also a natural extension of the original. Among others, there are models with nonlinear preferentiality [16], with rewiring of edges at later times [17], with a fitness parameter as an intrinsic value of a node [18,19], etc. Novel and more complex approaches, describing a variety of degree distributions and with more support in the real data, have been studied recently [20,21]. We believe that it is also of fundamental importance to examine "as simple as possible" processes that capture essential behavior of real world networks.

In this paper, we present a model that exhibits power-law-like degree distribution of an undirected network or the in-degree power-law-like distribution of a directed network. The purpose of the model is to test information filtering as a stochastic component of the network evolution process, while using a simple deterministic rule for attachment of new nodes. The results we report in this paper clearly show that our model can reproduce power-law distributions with the cutoff, similar to some real data reported recently [22].

II. MODEL

Our model introduces two crucial features that make it different from the Albert-Barabási model. A new node is introduced into the network at each time step. For simulation purposes, we first generate a network of 1100 nodes that are completely randomly connected to each other. Each new node in this core is connected to one of the older ones with uniform probability, until a core is formed. The size of the core is taken to be 1100 because we investigated filtration subsets with maximal size of 1000 nodes. After the core is formed, the following procedure takes place. Each new node attaches itself to the network with ω links. To choose to which of the already present nodes in the network it will attach itself, the following rule is applied. (i) A sample of the already present nodes of fixed size m is randomly chosen from the network which contains t nodes. The probability of choosing any node in the sample equals m/t. (ii) Chosen nodes are sorted by their degree in decreasing order. For the nodes with the same degree, no additional rearrangement is applied. (iii) From such a sorted sample, a new node is attached to the first ω nodes with the highest degree. The third rule is a simple deterministic "winner takes it all" algorithm, which combined with the first two rules produces very interesting macroscopic effects, as will be presented in this paper.

The nodes are numbered from 0, and the network is grown to the size n_{max}. We averaged over multiple simulations (30–100) for every investigated ω, m, and n_{max} in order to get a statistically relevant ensemble of network realiza-

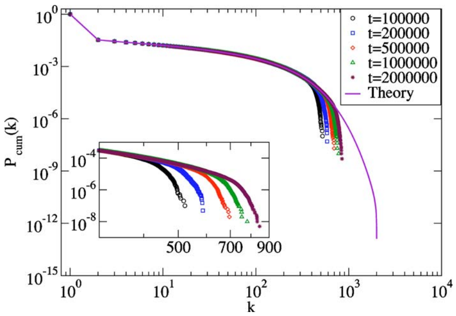

FIG. 1. (Color online) Simulated degree cumulative probability functions with $m=100$ and $\omega=1$ for different final network sizes n_{max} are compared to the theoretically obtained one. The figure clearly depicts the asymptotic approach of simulation curves to the theoretical result. This implies that analytical results are precise and that they sufficiently well describe the behavior of the system when $n_{max}\rightarrow\infty$. The inset gives the enlarged section with the tails of the simulated cumulative probability distributions to better illustrate the effects of the finite network size.

tions. We also performed a scaling investigation presented in Fig. 1 to see how a simulated distribution behaves for different network sizes.

To show the stability of the obtained distributions for different parameters, we also give figures of simulated distributions behavior for different starting core sizes in Fig. 2 and for different parameters m, i.e., sample sizes, in Fig. 3.

III. THEORY

In the theoretical treatment of the node degree distribution we decide to limit ourselves to the description of network with $\omega=1$. The reason for such an approach is a cumbersome analytical study for the case of $\omega>1$, which would include many more summation terms that are analytically almost unsolvable. We use the master equation approach of Dorogovtsev *et al.* [23]. In this approach, a new node enters the network at every moment s and is therefore denoted by s. It connects with one edge to the node with maximum degree in

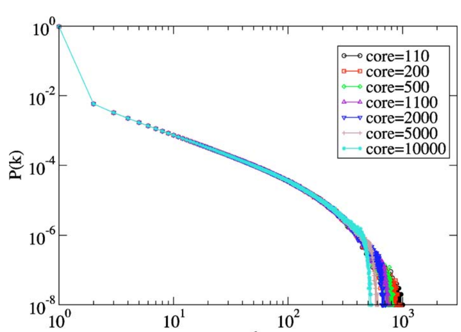

FIG. 2. (Color online) Simulated degree probability distributions with $m=100$, $\omega=1$ and $n_{max}=10^6$ for different starting core sizes. The figure distinctly demonstrates the stability of probability distribution over many decades. The differences that the functions exhibit in their maximal degrees are the consequence of finite size (FS) effects.

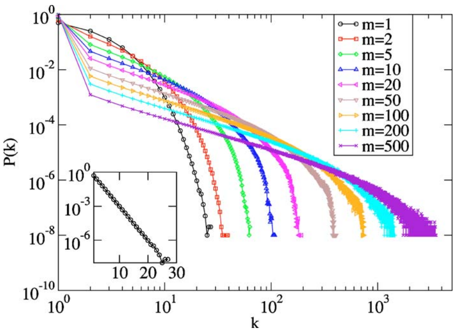

FIG. 3. (Color online) Simulated degree probability distributions for different m, but constant $\omega=1$, and $n_{max}=10^6$. When $m=1$ the network is the exponential one, as can be easily seen from the straight line in the lin-log plot in the inset. With the increase of m, the power-law behavior emerges and becomes more pronounced as m increases.

the randomly selected sample of size m. Nodes in sample are selected from t nodes that are already present in the network, so that every existing node has the probability m/t of entering the sample.

The probability that the node s with degree k will enter the sample of size m at time t and will be connected to the new node is

$$v(k,m,t-1)=\sum_{l=0}^{m-1}\binom{\hat{B}(k,t-1)}{l}\frac{\binom{N(k,t-1)-1}{m-l-1}}{(m-l)\binom{t}{m}}. \quad (1)$$

Here, the first binomial coefficient in the numerator represents number of possible ways to chose l nodes with degree smaller than k into the sample, and

$$\hat{B}(k,t-1)=\sum_{q=1}^{k-1}N(q,t-1), \quad (2)$$

where $N(k,t)$ is the number of nodes with degree k at time t. The second binomial coefficient counts the number of possible ways to chose $m-1-l$ nodes with the same degree as node s into the sample. This part of expression (1) accounts for the possibility that in the selected sample there exist other nodes with the same maximal degree as s. Using the fact that $N(q,t)=P(q,t)\cdot t$, together with an approximation that for large t one can approximate

$$\binom{t}{m}$$

with $t^m/m!$, we reduce the expression (1) to the following form:

$$v(k,m,t-1)\simeq\frac{1}{t}\sum_{l=0}^{m-1}\binom{m}{l}\Pi(k,t-1)^{l}P(k,t-1)^{m-l-1}, \quad (3)$$

where

$$\Pi(k,t-1) \equiv \sum_{q=1}^{k-1} P(q,t-1). \quad (4)$$

Using the well-established Dorogovtsev-Mendes master equation approach for calculating the node-degree distribution, for $k \geq 2$ we write

$$p(k,s,t) = v(k-1,m,t-1)p(k-1,s,t-1) + [1 - v(k,m,t-1)]p(k,s,t-1). \quad (5)$$

To calculate the probability distribution $P(k,t)$ such that a randomly chosen node has k edges at time t, we average the probability distribution of all nodes s; i.e.,

$$P(k,t) = \frac{1}{t+1}\sum_{s=0}^{t} p(k,s,t). \quad (6)$$

Thus, we obtain

$$P(k,t) = \frac{\zeta(k-1,t-1)}{t+1}P(k-1,t-1) + \left(\frac{t}{t+1} - \frac{\zeta(k,t-1)}{t+1}\right)P(k,t-1), \quad (7)$$

where

$$\zeta(k,t-1) \equiv \sum_{l=0}^{m-1} \binom{m}{l} \Pi(k,t-1)^{l} P(k,t-1)^{m-l-1}. \quad (8)$$

Assuming that Eq. (7) has a stable asymptotic solution for $t \gg 1$, thus changing the time-dependent probability distribution into time-independent $P(k,t)=P(k)$, we obtain the following closed form:

$$P(k) = \zeta(k-1)P(k-1) - \zeta(k)P(k). \quad (9)$$

Equations (9) are polynomials of order m and hold for all $k \geq 2$. Written as polynomials, they adopt the following form:

$$a(0)P(k)^m + a(1)P(k)^{m-1} + \cdots + a(l)P(k)^{m-l} + \cdots + [1 + a(m-1)]P(k) - \sum_{l'=0}^{m-1}\binom{m}{l'}\left(\sum_{q=1}^{k-2}P(q)\right)^{l'} \times P(k-1)^{m-l'} = 0, \quad (10)$$

where the coefficients $a(l)$ are

$$a(l) = \binom{m}{l}\left(\sum_{q=1}^{k-1}P(q)\right)^{l}. \quad (11)$$

For theoretical treatment of $P(1)$ as our boundary condition, the following equation holds:

$$p(1,s,t) = \delta_{s,t} + (1-\delta_{s,t})[1 - v(1,m,t-1)]p(1,s,t-1), \quad (12)$$

with an obvious relation for probability that a node with one edge at time $t-1$ will adopt a new edge at time t:

$$v(1,s,t-1) = \frac{P(1,t-1)^{m-1}}{t}. \quad (13)$$

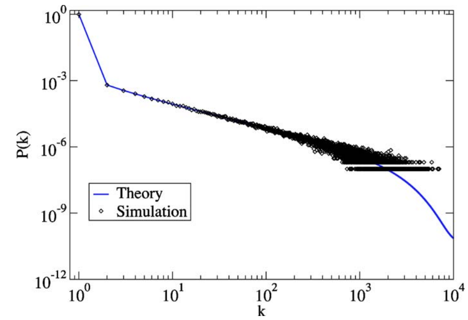

FIG. 4. (Color online) Theoretical probability distribution (solid line) nicely follows simulation data (black diamonds) for $m=1000$. Scattering in the tail is a consequence of low probability fluctuations induced by finite size effects. The reader should also note a big jump of probability for $P(k=1)$.

Using a procedure similar to that already mentioned above, we obtain the asymptotic value for $P(1)$:

$$P(1) = 1 - P(1)^m. \quad (14)$$

Unfortunately, the set of Eqs. (10) and (14) is analytically unsolvable and is therefore solved numerically. The solutions of these polynomial equations show excellent agreement with numerical simulations as can be seen in Figs. 4, 5, 6, and 7. These findings further vindicate the master equation approach followed in this paper.

IV. DISCUSSION

An interesting aspect of the model is its behavior when the model parameters acquire some limiting values. Let us first consider the limit $m=1$. In this case the new node attaches completely randomly to one of the existing nodes. The network obtained by this sort of growth is *exponential*; i.e., its node-degree probability distribution is exponential [2]. This feature is clearly demonstrated in the inset of Fig. 3. The opposite limit $m \to t$ would be another interesting limit of our growth mechanism. Strictly speaking, this growth rule falls out of scope of this paper, since our model implies *fixed* m. However, it is easy to see that this limiting growth rule

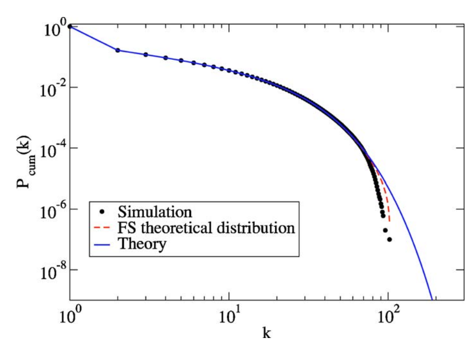

FIG. 5. (Color online) For $m=10$ the theoretical distribution (solid line) nicely follows simulation data (black dots). The disagreement in the very tail is explained by finite size effects of simulated data. However, a FS theoretical distribution obtained by transformation (16) shows even better agreement with simulation.

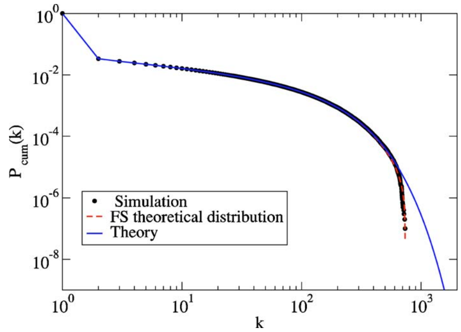

FIG. 6. (Color online) For $m=100$ it is easy to see that the theoretical distribution follows simulation data very well, and FS theoretical distribution even better.

results in a network consisting of one superconnected hub to which practically all other nodes are connected.

As we have mentioned in the preceding section, a master equation approach yields a chain of the polynomial Eqs. (10) and (14). Note the fact that $P(k^*)$ representing the probability that a randomly chosen node will have a degree k^* depends only on degree probabilities that are equal to or less than k^* (10). We have calculated the roots of the system to get a degree probability distribution.

All simulated data and analytical roots of polynomial equations exhibit a big jump from $P(k=1)$ to $P(k=2)$ of order of a magnitude or more. The difference $P(k=1)-P(k=2)$ depends strongly on the size of a chosen sample m. If the size of the sample is larger, then there is higher probability that a node of degree larger than 1 will enter the sample, and collect the new link. The smaller the sample the greater the probability that only nodes of degree one will be chosen in the sample, thus lowering the overall amount of nodes of degree one. The obtained analytical solutions from Eq. (14) are in excellent agreement with simulation results regarding to this jump. The average relative error for $m \in \{10,100,1000\}$ simulation and theory is 4.3×10^{-5}, and gets smaller as the sample size m grows larger for $n_{max}=10^6$.

All simulated data exhibit a strong scattering in the tail. The scattering is a consequence of low probability fluctuations and makes the comparison between theory and simulation more difficult (Fig. 4). In order to eliminate these fluctuations in the data and compare theory and simulation, it is possible to use exponential binning or to transform probability distribution into the cumulative probability distribution. We implemented the second approach and produced a cumulative degree probability distribution P_{cum}.

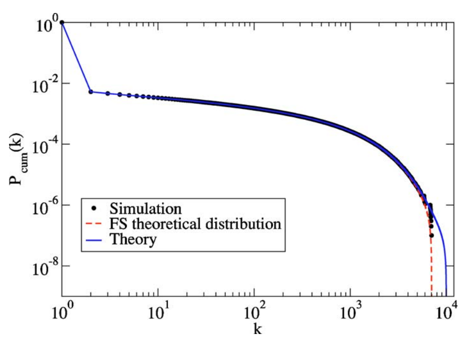

FIG. 7. (Color online) $m=1000$ is the largest monitored sample size but is still small enough compared to simulation number of nodes $n_{max}=10^6$. Theory is in excellent agreement with simulation.

TABLE I. Fitted cumulative degree distribution parameters for different sample sizes. Correlation coefficients show excellent agreement between the theoretical distribution data and the presented fits.

	$P_{cum}\sim k^{-\gamma}e^{-\alpha k}$			$P_{cum}\sim e^{-\eta k^{\beta}}$		
	γ	α	Corr	η	β	Corr
$m=10$	0.5501	0.0765	0.9980	0.9829	0.4718	0.9976
$m=100$	0.4026	0.0092	0.9995	0.3894	0.4385	0.9987
$m=1000$	0.3067	0.0011	0.9994	0.2230	0.3823	0.9978

$$P_{cum}(k)=\sum_{q=k}^{\infty}P(q). \quad (15)$$

This distribution contains the same system information as the degree distribution, but is much smoother in the tail. We compared our theoretical curve with the simulated one and found an excellent match between theory and simulations. The results of the comparison between simulation and theory are presented in Figs. 5, 6, and 7. The relative disagreement observed in the tails is a consequence of finite size effects (Fig. 1). Since our theoretical curve falls relatively slowly, as can be seen in Table I, the summation of probabilities for $k>k_{max}$ in Eq. (15) contributes strongly to the cumulative degree probability in the tail. To get an even better match, we calculated "renormalized" cumulative probability distribution

$$\tilde{P}_{cum}(k)=\frac{\sum_{q=k}^{k_{max}}P(q)}{1-\sum_{q=k_{max}+1}^{\infty}P(q)}. \quad (16)$$

This finite size cumulative probability distribution is even better in describing finite size effects, as shown in the Figs. 5, 6, and 7.

To obtain a description of the degree distribution in the thermodynamical limit, we fitted theoretical cumulative degree distribution (theoretical and not simulation distribution was also used since it does not suffer from finite size effects) with the stretched exponential (17) and power-law distribution with the exponential cutoff (18) [15]:

$$P_{cum}\sim e^{-\eta k^{\beta}}, \quad (17)$$

$$P_{cum}\sim k^{-\gamma}e^{-\alpha k}. \quad (18)$$

For fitting purposes we used all theoretical $P_{cum}(k)$ values, except $P_{cum}(1)$, because its value is clearly not determined by the scale-free-like behavior as opposed to all other k values. Both distributions fit our overall results very well, as presented in Table I, and Figs. 8, 9, and 10. The correlation

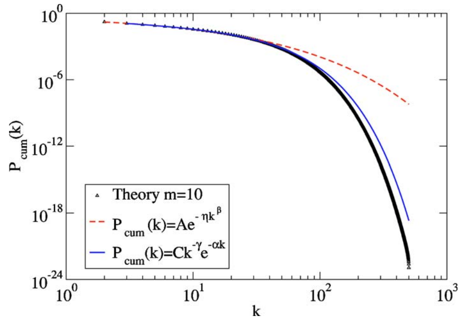

FIG. 8. (Color online) Two different functions: (i) stretched exponential and (ii) power law with the cutoff are fitted on theoretical data for sample size $m=10$. This figure clearly shows that the power law with the exponential cutoff better describes the tail of the theoretical distribution.

coefficients of the fitted distributions are all above the 0.99 margin, proving that both fitting models are capable of describing theoretically obtained curves very well. The power law with the exponential cutoff always has just a slightly higher correlation coefficients than stretched exponential for the same sample size m. The figures clearly show that the power law with exponential cutoff describes the tail of the simulated distribution very accurately. A stretched exponential is clearly not suitable for the description of the tail properties.

It is worth mentioning that the power-law distribution with the exponential cutoff has already been obtained in a similar model [15], which has shown that exponential parameter α is trivially connected with the sample size m by the relation $\alpha=1/m$. Although one cannot expect this relation to be valid for this model also, the parameter α is very close to $1/m$, and this coincidence is better for larger m, as can be seen in Table I. In our opinion, it would be interesting to measure α in some observed network distributions of a similar shape and compare it with the expected sizes of samples on which the new node has the possibility of creating a link.

One of the most interesting features of the growth mechanism introduced in this paper is the small value of the power-law exponents γ of the node degree probability distribution. The values of these exponents, displayed in Table I, are among the smallest obtained in the complex network models so far. This fact further emphasizes the importance of our model as a possible framework for the study of the power laws with small exponents.

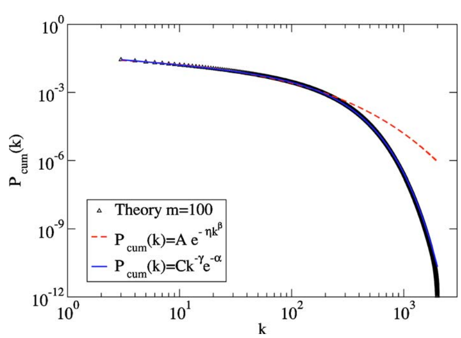

FIG. 9. (Color online) Fitted curves for $m=100$. Power law with the exponential cutoff represents the theoretical distribution very well.

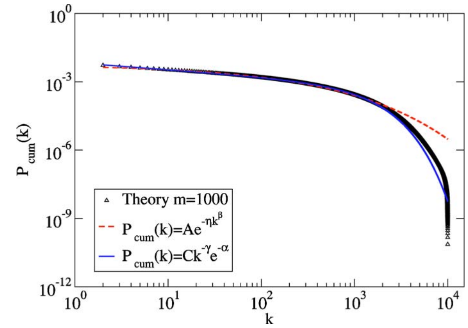

FIG. 10. (Color online) Excellent agreement of fitted and theoretical distributions for $m=1000$.

Finally, let us briefly discuss simulation results for $\omega>1$. Simulation results for the cumulative probability distribution [without $P(\omega)$] are displayed in Fig. 11. The typical characteristics of the distribution are equivalent to the $\omega=1$ case. The degree $k=\omega$ has a substantially larger probability compared to all other degrees. The cumulative probability distribution for $k>\omega$ obtained in the simulations displays the scale-free-like properties. These simulation distributions can be well fitted to the power law with the exponential cutoff (18), as shown in Fig. 11.

An important characteristic of the model is the temporal correlation of the networks produced by the model of growth studied in this paper. The investigation of this characteristic shows that, similar to the BA model, there exists temporal correlation. Therefore, additional modifications of the two-step rule are needed to overcome this shortcoming.

In the study of the two-step process of the attachment of new nodes to the existing network, the steps can be modified in order to describe other types of the network growth. For instance, in the first step it is possible to preferentially select nodes into the sample and then randomly select ω nodes from the sample. Such a two-step attachment process results in a power-law degree probability distribution with a number of outliers with a high degree [24]. The other possibility is the preferential selection in the first step and the preferential selection in the second step. Such a model of network growth results in a strong condensation of links at a very small number of nodes [24]. These models seem more realistic for cer-

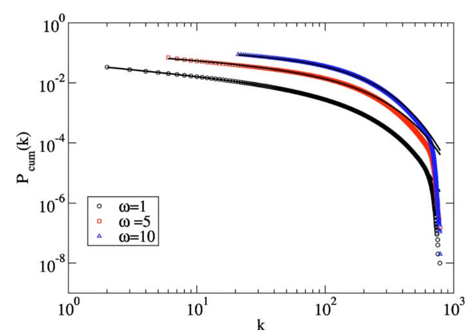

FIG. 11. (Color online) Evidence that the distributions for $\omega>1$ fall in the same class as the distributions studied analytically. The situation with $m=100$ is presented.

tain types of network growth, such as, e.g., the network growth based on the use of search engines.

V. CONCLUSIONS

We have shown that using a simple "winner takes it all" algorithm, together with the fact that nodes do not possess complete information on network structure, a macroscopic node-degree power law is created. We have shown that realistic assumption of incomplete knowledge can have a substantial effect on the network growth. Although the field of complex networks has made great progress during the last few years, there is still much open space for research of microscopic models that describe the formation of complex networks with certain expected features. Our results clearly show that stochastic-deterministic processes even as simple as that described in this paper can be used to reproduce some macroscopic effects of complex networks. Moreover, in this paper as well as in [15], we have demonstrated that the power law with the exponential cutoff can be a significant distribution for types of networks in which information filtering is performed. Recent findings in social contact networks [22] lead us to believe that the power law with the exponential cutoff and stretched exponentials should be studied more intensively in the future.

ACKNOWLEDGMENT

This work was supported by the Ministry of Science and Technology of the Republic of Croatia under the Contract Numbers 0098002 and 0098004.

[1] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[2] S. N. Dorogovtsev and J. F. F. Mendes, Adv. Phys. **51**, 1079 (2002).
[3] M. E. J. Newman, SIAM Rev. **45**, 167 (2003).
[4] B. A. Huberman and L. A. Adamic, Nature (London) **401**, 131 (1999).
[5] M. Faloutsos, P. Faloutsos, and C. Faloutsos, Comput. Commun. Rev. **29**, 251 (1999).
[6] G. Caldarelli, R. Marchetti, and L. Pietronero, Europhys. Lett. **52**, 386 (2000).
[7] R. Pastor-Satorras, A. Vázques, and A. Vespignani, Phys. Rev. Lett. **87**, 258701 (2001).
[8] H. Jeong, S. Mason, A.-L. Barabási, and Z. N. Oltvai, Nature (London) **411**, 41 (2001).
[9] J. W. Grossman and P. D. F. Ion, Congr. Numer. **108**, 129 (1995).
[10] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).
[11] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[12] A.-L. Barabási, R. Albert, and H. Jeong, Physica A **272**, 173 (1999).
[13] L. Adamic and B. Huberman, e-print arXiv: cond-mat/0001459.
[14] S. Mossa, M. Barthélémy, H. E. Stanley, and L. A. Nunes Amaral, Phys. Rev. Lett. **88**, 138701 (2002).
[15] H. Štefančić and V. Zlatić, Physica A **350**, 657 (2005).
[16] P. L. Krapivsky, S. Redner, and F. Leyvraz, Phys. Rev. Lett. **85**, 4629 (2000).
[17] R. Albert and A.-L. Barabási, Phys. Rev. Lett. **85**, 5234 (2000).
[18] G. Bianconi and A.-L. Barabási, Phys. Rev. Lett. **86**, 5632 (2001).
[19] G. Bianconi and A.-L. Barabási, Europhys. Lett. **54**, 436 (2001).
[20] M. A. Serrano, M. Boguna, and A. Díaz-Guilera, Phys. Rev. Lett. **94**(3), 038701 (2005).
[21] V. Colizza, J. R. Banavar, A. Maritan, and A. Rinaldo, Phys. Rev. Lett. **92**, 198701 (2004).
[22] D. Batorski, *The Structures of Relationships and Communication in Large Instant Messaging Network*, International Social Networks Conference, Portoroz, Slovenia, May, 2004 (unpublished).
[23] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, Phys. Rev. Lett. **85**, 4633 (2000).
[24] V. Zlatić and H. Štefančić, (unpublished).