Strategies for Generating and Evaluating Large-Scale Powerlaw-Distributed P2P Overlays

Ana-Maria Oprescu[1], Spyros Voulgaris[2], and Haralambie Leahu[2]

[1] Universiteit van Amsterdam
a.m.oprescu@uva.nl
[2] VU University
spyros@cs.vu.nl, h.leahu@vu.nl

Abstract. A very wide variety of physical, demographic, biological and man-made phenomena have been observed to exhibit powerlaw behavior, including the population of cities and villages, sizes of lakes, etc. The Internet is no exception to this. The connectivity of routers, the popularity of web sites, and the degrees of World Wide Web pages are only a few examples of measurements governed by powerlaw. The study of powerlaw networks has strong implications on the design and function of the Internet.

Nevertheless, it is still uncertain how to explicitly *generate* such topologies at a very large scale. In this paper, we investigate the generation of P2P overlays following a powerlaw degree distribution. We revisit and identify weaknesses of existing strategies. We propose a new methodology for generating powerlaw topologies with predictable characteristics, in a completely decentralized, emerging way. We provide analytical support of our methodology and we validate it by large-scale (simulated) experiments.

1 Introduction

Many real-world large-scale networks demonstrate a power-law degree distribution, that is, a very small fraction of the nodes has a high *degree* (i.e., number of neighbors), while the vast majority of nodes has a small degree. In nature, such networks typically emerge over time, rather than being instantiated on the spot based on a blueprint. Providing researchers from different disciplines with a framework that allows them to control the self-emerging process of power-law networks, could substantially help them in studying and better understanding such networks, as well as deploying them at will to serve new applications (e.g., bio-inspired algorithms for peer-to-peer systems).

There are several algorithms to generate power-law networks, however little has been done for a self-emerging method for building such networks [3,5,6,4]. In this work, we first investigate existing research with an emphasis on the decentralization properties of proposed algorithms. Next, we select one approach that looks promising for straightforward decentralization. We identify several

J. Dowling and F. Taïani (Eds.): DAIS 2013, LNCS 7891, pp. 71–84, 2013.

limitations within the existing approach and we present a novel algorithm that has been tailored specifically to the needs of a large P2P network. Starting from a given, static distribution of random values among the P2P network nodes, we control the emerging power-law overlay .

We summarize related research conducted on power-law generation in Section 2, where we assess the degree to which such approaches may be decentralized. In Section 3.1 we identify several limitations (both theoretical and empirical) with an existing sequential approach and proceed to present a novel algorithm to alleviate the respective issues. In Section 4 we show how the decentralized algorithm may be implemented in a P2P network and present our evaluation results. We summarize our findings in Section 5.

2 Related Work

There is a vast literature on properties and characteristics of scale-free and small-world networks. The research behind such literature is focused on the observation of aforementioned topologies and their behavior (like finding the λ value) rather than construction methodologies. However, there are several important generative mechanisms which produce specific models of power-law networks. It started with the *Erdös and Rényi* random-graph theory and continued with the *Watts and Strogatz* model, which was the first to generate a small-world topology from a regular graph, by random rewiring of edges. Drawbacks of this initial model are its degree homogeneity and static number of nodes. These limitations can be addressed by scale-free networks, but the clustering coefficient becomes an issue. In turn, the clustering coefficient can be controlled through the employed generative mechanism. However, generating a random scale-free network having a specific λ value is not trivial. Moreover, most existing algorithms to generate scale-free networks are centralized and their decentralization, again, far from trivial. We present several types of generative models.

Preferential Attachment. This model, also known as the "rich-get-richer" model, combines preferential attachment and growth. It assumes a small initial set of m_0 nodes, with $m_0 > 2$, forming a connected network. The remaining nodes are added one at a time. Each new node is attached to m existing nodes, chosen with probabilities proportional to their degrees. This model is referred to as the Barabási-Albert (BA) model [2], though it was proposed by Derek J. de Solla Price [7] in 1965 and Yule in 1925 [12]. The degree distribution is proven to be $P(k) \sim k^{-3}$. Dorogovtsev and Mendes [11] have extended the model to a linear preference function, i.e., instead of a preference function $f_{BA}(i) = k_i$ they use $f_{DM}(i) = k_i + D, D \geq 0$. Dangalchev [6] introduced the two-level network model, by considering the neighbor connectivity as a second "attractiveness" discriminator, $f_{Da}(i) = k_i + c \times \sum_j k_j$, where $c \in [0, 1]$. The global view required at each node attachment renders this algorithm difficult to decentralize.

Preferential Attachment with Accelerated Growth. This model [10] extends the previous model with a separate mechanism to add new links between

existing nodes, hence accelerating the growth of the number of links in the network (much like the Internet). This algorithm inherits the difficulties of the basic preferential attachment with respect to decentralization.

Non-linear Preferential Attachment. Krapivsky, Redner, and Leyvraz propose a model [14] that produces scale-free networks as long as $f_{KRL}(i) \sim k_i; k \to \infty$, where $f_{KRL}(i) = k_i^{\gamma}$. This algorithm inherits the difficulties of the basic preferential attachment with respect to decentralization.

Deterministic Static Models. Dangalchev proposed two such networks, the k-control and the k-pyramid, where the latter can be extended to a growth model. Ravasz and Barabási [1] explored hierarchical (fractal-like) networks in an effort to meet both the power-law degree distribution of scale-free networks and the high clustering coefficient of many real networks. Their model starts with a complete q-node graph which is copied $q-1$ times ($q > 2$); the root of the initial graph (selected arbitrarily from the q nodes) is connected with all the leaves at the lowest level; these copy-connect steps can be repeated indefinitely. Such networks have degree distribution $P(k) \sim k^{\frac{\ln q}{\ln(q-1)}}$. Cohen and Havlin [5] use a very simple model which delivers an ultra-small world for $\lambda > 2$; it assumes an origin node (the highest degree site) and connects it to next highest degree sites until the expected number of links is reached. Since loops occur only in the last layer, the clustering coefficient is intuitively high for a large number of nodes. According to [9], some deterministic scale-free networks have a clustering coefficient distribution $C(q) \sim q^{-1}$, where q is the degree. This implies well-connected neighborhoods of small degree nodes. This algorithm seems promising with respect to decentralization, except for the initial phase of complete q-node connectedness.

Fitness-Driven Model. This was introduced by Caldarelli [4] and proves how scale-free networks can be constructed using a power-law fitness function and an attaching rule which is a probability function depending on the fitness of both vertices. Moreover, it shows that even non-scale-free fitness distributions can generate scale-free networks. Recently, the same type of model with infinite mean fitness-distribution was treated in [13]. This power-law network generative algorithm seems the most promising with respect to decentralization.

3 Decentralizable Algorithms for Building Scale-Free Networks

We are interested in analyzing approaches that are feasible to decentralize. We first look at an existing model, presented by Caldarelli in [4], for which we introduce an analytical and empirical verification. We then present an improved model to build scale-free networks, which we also analyze and verify empirically. Our model maintains the property of easy decentralization.

3.1 Caldarelli's Fitness-Driven Model

In this model, power-law networks are generated using a "recipe" that consists of two main ingredients: a fitness density, $\rho(x)$, and a vicinity function, $f(x_i, x_j)$. The fitness density is used to assign each node a fitness value, while the vicinity function is used to decide, based on the fitness values, whether a link should be placed between two nodes.

One instance of this model assumes each node to have a fitness value x_i drawn from a Pareto distribution with density $\rho(x) \sim x^{-\gamma}$. For each node i, a link to another node j is drawn with probability $f(x_i, x_j) = \frac{x_i x_j}{x_M^2}$, where x_M is the maximum fitness value currently in the network.

This model looks very appealing for a self-emerging approach to power-law network generation, since it requires very little information to be globally available. Using epidemic dissemination techniques [8], the maximum fitness value currently existing in the network, x_M, may be easily propagated throughout the network.

According to [4], this approach leads to a network with a power-law degree distribution, that should have the same exponent as the non-truncated Pareto distribution of the fitness values. However, our initial set of experiments show that Caldarelli's approach rarely converges for very large networks. Figure 1 presents the data collected from four different experiments. Each experiment corresponds to a different degree distribution exponent and was repeated for two network sizes: 10,000 nodes and 100,000 nodes. For each experiment we constructed 100 different graphs, each with the same fitness distribution and different random seeds for the neighbor selection. We remark that for a desired power-law degree distribution with exponent $\gamma \neq 3$ and larger values of N ($100K$), the obtained degree distribution exponent does not converge to its desired value. Also, for $\gamma = 4$ and a network of $10K$ nodes, the general approximation function used to determine the degree distribution exponent could not be applied here. We investigated the issue further, by constructing the histogram corresponding to the degree distribution. Figure 5 shows the histograms obtained for each experiment. We remark that the histograms do not coincide with a power-law distribution.

A second set of experiments evaluated how well the algorithm controlled the emerging degree distribution exponent, γ. We increased the control γ in steps of 0.1 and ran the algorithm ten times for each value on a network of 100,000 nodes. We collected the estimated value of the emerging degree distribution exponent and the percentage of isolated nodes (i.e., nodes of degree zero). Both types of results are plotted in Figures 3a and 3b. We note that in the Caldarelli model a large number of nodes remain isolated.

We verified the empirical results by revisiting the assumptions made in [4]. We localized a possible problem with the way the vicinity function is integrated. Intuitively, the problem is that while the X's (fitnesses) are independent r.v. (by assumption), their maximum (x_M) is dependent on all of them, hence can not be pulled out of the integral. To explain this formally, using the Law of Large Numbers we obtain the estimation

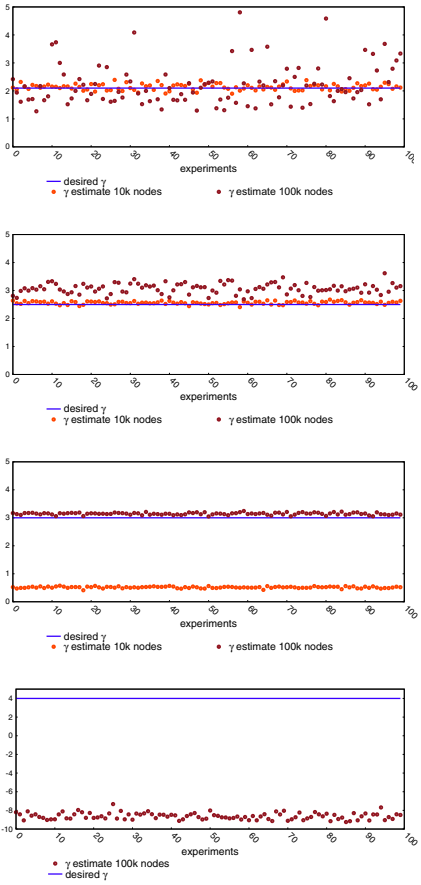


Fig. 1. Caldarelli's model

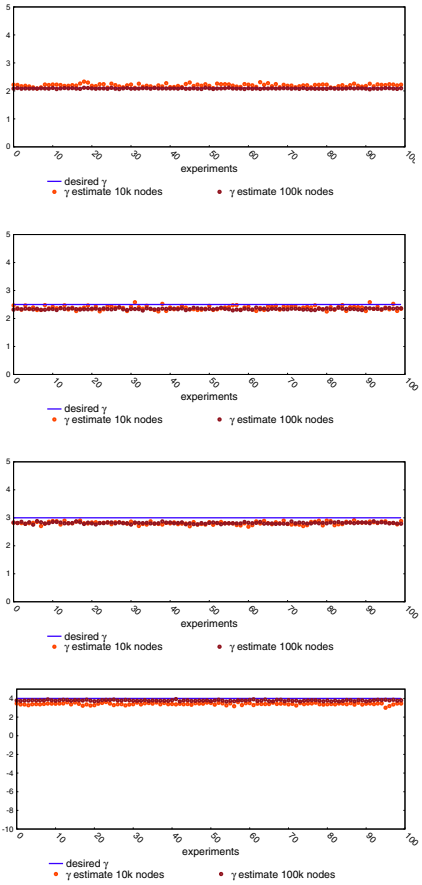
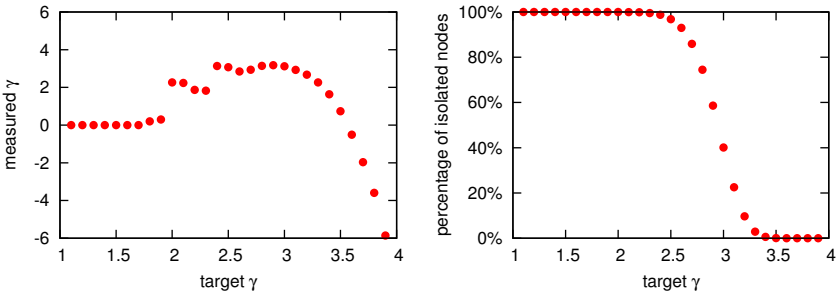


Fig. 2. Improved model

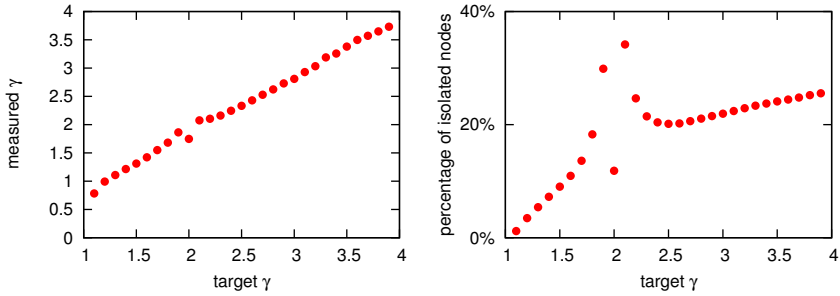
$$V_k = \frac{\text{Number of nodes of degree } k}{n} \approx \frac{1}{n} \frac{\rho\left(P^{-1}(k/n)\right)}{P'\left(P^{-1}(k/n)\right)},$$

where P^{-1} denotes the inverse of P, which is the probability that a node u with fitness x will be linked with any other node v ($P(x) := \mathbb{E}[p(X_u, X_v)|X_u = x]$). This approximation, in conjunction with the assumption $\rho(x) \sim x^{-\gamma}$ would provide the power-law behavior $V_k \sim k^{-\gamma}$, as claimed in [4]. However, this is not the case since x_M is a random variable dependent on all fitnesses (thus, also on X_u and X_v). Hence $P(x)$ is not linear, but is a rather intricate expression of x and an analytical expression for the inverse of P is infeasible. Even worse, if $\rho(x) \sim x^{-\gamma}$, the squared-maximum x_M^2 will grow to infinity at rate $n^{2/(\gamma-1)}$ (by Fisher-Tippet-Gnedenko Theorem), so that $D = (n-1)\mathbb{E}[p(X_1, X_2)]$, will tend to 0 when $\gamma < 3$ (the resulting graph will have a very large fraction of isolated



(a) γ estimates (b) percentage of isolated nodes

Fig. 3. Caldarelli's fitness-driven model



(a) γ estimates (b) percentage of isolated nodes

Fig. 4. Our model

nodes) and will grow to infinity for $n \to \infty$, when $\gamma > 3$; however, as explained before, this last fact is impossible in a power-law graph in which all nodes are linked with the same probability; see equation (3).

3.2 Improved Model

Here we present a novel model for a power-law graph with n nodes. It addresses the limitations found with the Caldarelli model by avoiding certain mathematical pitfalls. Our assumptions differ from the Caldarelli model in that we consider a *truncated* Pareto distribution, with density function $\rho(x) \sim x^{-2}$, for $x \in (l, b_n)$. We emphasize that, unlike Caldarelli, we start with a fixed distribution exponent.

Another considerable difference is the truncation and its upper bound $b_n \to \infty$. The upper bound will depend on the density $\rho(x)$ and on the desired outcome graph degree distribution exponent, denoted by γ in Caldarelli's model. The global variable x_M from Caldarelli's model will be replaced in our model by b_n.

We summarize the mathematical model below:

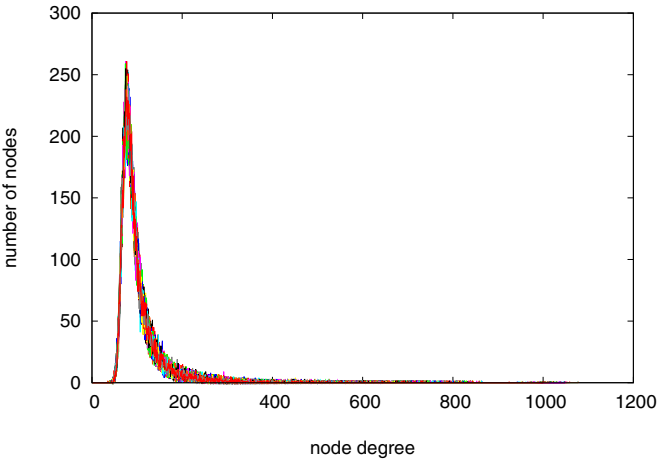


Fig. 5. Caldarelli degree histograms for $\gamma = 4$

(I) The fitnesses $X_1, \ldots, X_n$ are drawn from a truncated Pareto distribution, with lower bound $l = 1$, upper bound $b_n \to \infty$ (it will depend on the desired outcome) and density $\rho(x) \sim x^{-2}$, for $x \in (l, b_n)$.

(II) Every pair of nodes (u, v) will be linked with a probability given by $p(X_u, X_v)$, where we define

$$p(x_1, x_2) := \left(\frac{x_1 x_2}{b_n^2}\right)^{\eta}, \tag{1}$$

with $\eta > 0$ depending (again) on the desired outcome.

For appropriate choices of the upper-bound b_n (see details below), performing steps (I) and (II) will result in a power-law graph with index $\gamma := 1 + (1/\eta)$, satisfying (for large $k \leq n$)

$$V_k := \frac{\text{Number of nodes of degree } k}{n} \approx \frac{\gamma - 1}{k^{\gamma}};$$

in other words, if a power-law degree-distribution with exponent $\gamma > 1$ is desired, then one must choose $\eta = (\gamma - 1)^{-1}$ in step (II), while the upper-bound b_n must be chosen according to the following rules:

(i) For $\gamma \in (1, 2)$ we choose $b_n := \left[\left(\frac{\gamma-1}{2-\gamma}\right) n\right]^{\frac{\gamma-1}{\gamma}}$, which gives an expected degree

$$D \approx \left(\frac{\gamma - 1}{2 - \gamma}\right)^{\frac{2}{\gamma}} n^{\frac{2-\gamma}{\gamma}}.$$

(ii) For $\gamma = 2$ we choose $b_n := \sqrt{(n/2)\log(n)}$ and obtain for the expected degree

$$D \approx \frac{\log(n)}{2}.$$

(iii) For $\gamma > 2$ we choose $b_n := \left[\left(\frac{\gamma-1}{\gamma-2}\right) n\right]^{\frac{\gamma-1}{2}}$ which yields

$$D \approx \frac{\gamma - 1}{\gamma - 2}.$$

In the model described by steps (I) and (II), the probability that a node u, having fitness x, will be linked with any other node v is given by

$$P(x) := \mathbb{E}[p(X_u, X_v)|X_u = x] = \frac{x^\eta}{b_n^{2\eta}} \int_0^\infty z^\eta \rho(z)\, dz. \tag{2}$$

The expected degree of a node of fitness x is $(n-1)P(x)$. The (unconditional) probability of having the edge (u, v) is $\pi_n := \mathbb{E}[p(X_u, X_v)] = \mathbb{E}[P(X_u)]$ and the expected degree of a node is $D := (n-1)\pi_n$. For the choices (i)–(iii), the expected degree of a node of fitness x will be approximately x^η, for large enough n.

At this point, it should be noted that power-law graphs with index $\gamma > 1$ (regardless of how they are generated) in which every two nodes are linked with the same probability π_n, enjoy the following property: If $\gamma > 2$ then the expected degree D must remain bounded as the number of nodes n grows arbitrarily large. When $\gamma = 2$ the expected degree D may grow to infinity with the number of nodes, but no faster than $\log(n)$. Finally, when $\gamma \in (1, 2)$ the expected degree D may again grow to infinity with the number of nodes, but no faster than $n^{2-\gamma}$. To justify the above claims, one may express the total expected number N of edges in the graph in two ways: first, since any two nodes are linked with the same probability π_n, the expected number of edges $\mathbb{E}[N]$ is given by $n(n-1)\pi_n/2$. On the other hand, N is half of the sum of all degrees in the graph, hence

$$D = (n-1)\pi_n = \frac{2\mathbb{E}[N]}{n} = \sum_{k=1}^{n-1} k\mathbb{E}[V_k] \leq c \sum_{k=1}^{n-1} \frac{1}{k^{\gamma-1}}, \tag{3}$$

where V_k denotes the number of nodes of degree k and $c > 0$ is some finite constant. Since the r.h.s. in (3) is bounded for $\gamma > 2$ and using the estimates

$$\sum_{k=1}^{n-1} \frac{1}{k^{\gamma-1}} \sim \begin{cases} n^{2-\gamma}, & \gamma \in (1, 2), \\ \log(n), & \gamma = 2, \end{cases}$$

hence our claims are justified. The conclusion is that the power-law structure of a graph, in which every two nodes interact with the same probability, induces an upper-bound on the magnitude of the expected degree of the nodes. Comparing the expected degree estimates in (i)–(iii) with the maximal rates imposed by (3) reveals that our method maximizes the expected degree when $\gamma \geq 2$.

We also remark that the graph resulted at step (II) will have a certain fraction of isolated nodes which increases with γ. More precisely, for γ close to 1 this fraction will be very small (close to 0), while for very large γ it will approach $1/e \simeq 37\%$; when $\gamma \in (2, 3)$ this fraction will stay between $14 - 22\%$.

The existence of these isolated nodes in our model is a consequence of the upper-bound established by (3) since, in general, by Jensen's Inequality it holds that

$$\mathbb{E}[\deg(v) = 0] \approx \mathbb{E}[\exp(-(n-1)P(X))] \geq \exp[-(n-1)\pi_n] = \exp(-D),$$

so whenever the expected degree D is bounded (recall that this is necessarily the case when $\gamma > 2$) the expected fraction of isolated nodes will be strictly positive.

Similarly to the experiments conducted on the Caldarelli model, we also performed a set of extensive tests on our novel model. Results from the set of 100 experiments are collected in Figure 2. We remark that our model performs in a more stable fashion with respect to the emerging degree distribution exponent. We also notice that our model provides a better convergence with respect to the size of the network.

Next, we analyzed how well our model controlled the emerging degree distribution exponent, γ, by performing the same set of averaging experiments as in Caldarelli's case. All results are collected in Figures 4a and 4b. Our model outperforms Caldarelli's model both in terms of control over the emerging degree distribution exponent, γ, and in terms of the number of isolated nodes. Finally, we notice that the theoretically proven discontinuity at $\gamma = 2$ is illustrated by the experimental results.

In this section, we have presented and experimentally evaluated a novel method for generating connected power-law graphs with any index $\gamma > 1$. In our proposed model, we correct the issues in [4], by considering *t*runcated (bounded) fitness-values and use a deterministic bound b_n instead of a random one. While the lower bound ($l = 1$) is included in the model for technical purposes only, the upper bound b_n is crucial and plays the role of a tuning parameter which allows one to obtain the desired power-law index γ as well as the correct behavior for the expected degree. In fact, the upper-bound b_n is strongly related to the number of edges in the graph by means of the vicinity function defined in (1); namely, the larger the b_n the smaller the number of edges in the graph. In general, increasing the magnitude of b_n will damage the power-law behavior, while for $\gamma > 2$ decreasing b_n will result in an asymptotically empty (still power-law) graph. Therefore, the model is extremely sensitive to the choice of the upper-bound b_n when $\gamma > 2$.

4 Building Power-Law Overlays

4.1 Algorithm

Building power-law overlays in the real world is a nontrivial task. Following the standard methodology, that is, applying the vicinity function on all possible pairs of nodes to decide which edges to place is impractical: It assumes either centralized membership management, or complete membership knowledge by each node. Neither of these scales well with the size of the overlay.

Instead, we explicitly designed a solution in which nodes are not required to traverse the whole network to determine their links. They form links by considering a small partial view of the network. The key point, however, in this

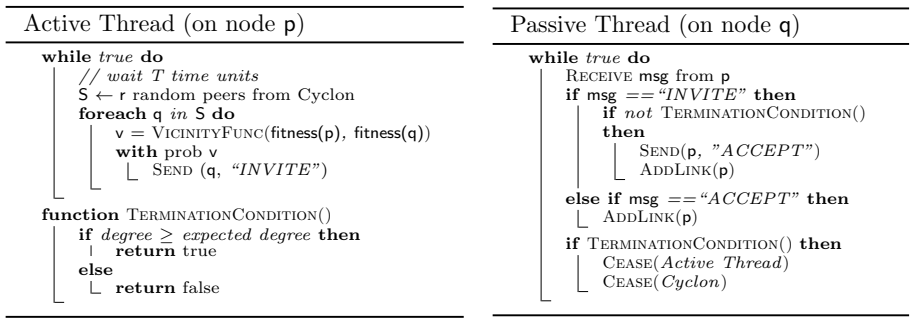


Fig. 6. The generic gossiping skeleton for building power-law overlays

approach is the termination condition, that is, a criterion that lets a node decide when to stop looking for additional links.

Our method exploits the analytic findings of the previous section. In a nutshell, each node periodically picks a few random other nodes, and feeds the two fitness values into the vicinity function to determine whether to set up a link or not. A node performs this repeatedly until it has satisfied its termination condition, that is, it has established a number of links equal to its expected degree, as computed by the respective formula.

In more detail, our protocol works as follows. Nodes run an instance of Cyclon [15], a peer sampling service that provides each node with a regularly refreshed list of pointers to random other peers, in a fully decentralized manner and at negligible bandwidth cost. Upon being handed a number of random other peers, a node applies the vicinity function and decides if it wants to set up a link with one or more of them. It sends an Invite message to the respective peers, and awaits their responses. Upon receiving an Invite, a node checks if its degree has already reached its expected degree value. If not, it sends back an Accept message as a response to the invitation, and the two nodes establish a link with each other on behalf of the power-law overlay.

When a node's termination condition is met, that is, the number of established links of that node has reached its expected degree, it refrains from further gossiping. That is, it stops considering new neighbors to send Invite messages to, and it responds to other nodes' invitations by a Reject message. Notably, a node also refrains from all Cyclon communication. This is particularly important for letting the network converge fast. By ceasing its Cyclon communication, a node is prompty and conveniently "forgotten" by the Cyclon overlay, letting the latter be populated exclusively by nodes that are still in search of additional links. Thus, Cyclon constitutes a good source of random other peers, as it picks random nodes out of a pool of peers that are willing to form additional links. Even in a network of hundreds of thousands of nodes, when a small number of nodes are left searching for additional links,they quickly discover each other and decide what links to establish.

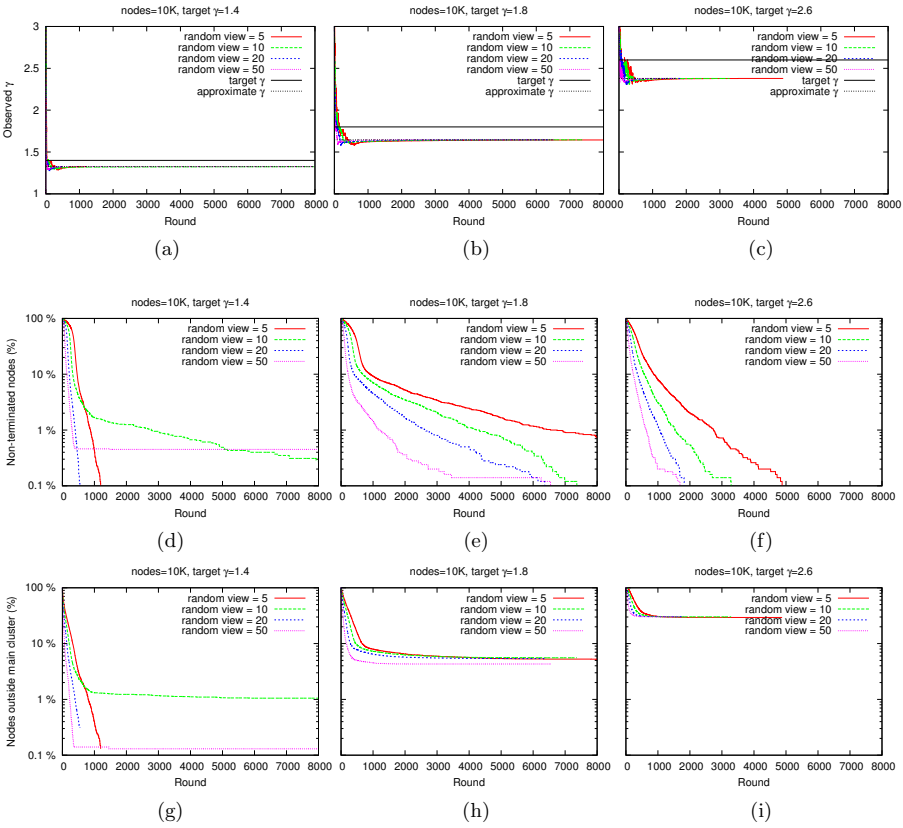


Fig. 7. Statistics collected for different γ values and different random views

Figure 6 shows the programming model of our protocol, without including Cyclon. As most gossiping protocols, it is modelled by two threads. An *active thread*, taking care of all periodic behavior and sending invitations, and a *passive thread* receiving messages (invitations or responses to invitations) and reacting accordingly.

4.2 Evaluation

We implemented our algorithm in PeerNet, a branch of the popular PeerSim simulator written in Java.

We consider a network consisting of a fixed set of N nodes. We assume that communication is reliable. Links are persistent and bidirectional (i.e., when x establishes a link to y, y gets a message to establish a link to x). A node's active thread operates in a periodic manner, and all nodes' periods are the same, yet they are triggered completely asynchronously with respect to each other.

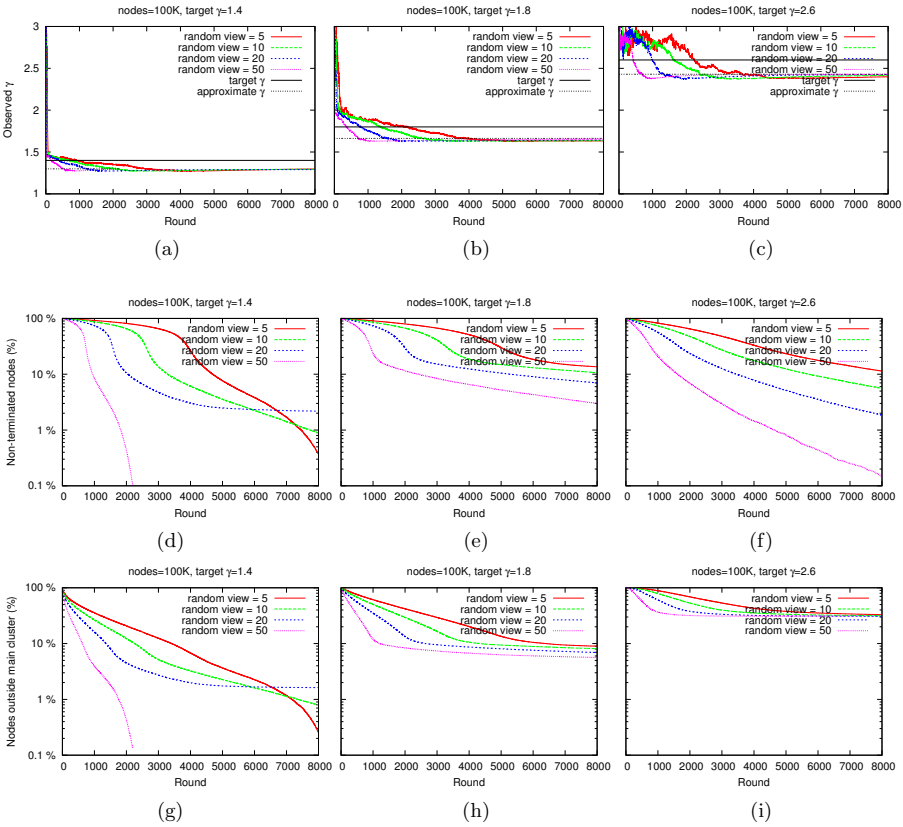


Fig. 8. Statistics collected for different γ values and different random views

The behavior of the protocol depends on three main parameters. First, the *target* γ; second, the *number of nodes* in the network; and third, the *random view size*, that is, the number of random links a node is handed by CYCLON in each round.

Figures 7 and 8 show the results of our experiments for 10,000 and 100,000 nodes, respectively. The first row in each figure (i.e., Figure 7(a-c) and Figure 8(a-c)) shows the observed γ of the emerged overlay, as a function of the number of rounds elapsed since the beginning of the experiment, for three sample values of γ, namely, 1.4, 1.8, and 2.6. The four different lines in each plot correspond to four different random view sizes. In the case of 10K nodes, all four lines converge equally fast to the (approximate) target γ. For the larger network of 100K nodes, checking out more random nodes per round provides some advantage with respect to convergence time.

Note that each graph shows a different target value of γ and the corresponding approximate value. Our formula for a node's expected degree is derived from the mathematical model presented in Section 3.2. However, it is based on the

assumption of a large enough number of nodes and therefore we evaluate the error introduced by this approximation. We construct the histogram of all expected degrees (i.e., the expected degree distribution) and we use it to compute an *a*pproximate γ. In each Figure 7a–7c and 8a–8c we compare the target γ, the approximate γ and the γ values of the self-emerging overlays.

The second row of the figures (i.e, Figure 7(d-f) and Figure 8(d-f)) shows the percentage of nodes that have not yet established as many links as their expected degree mandates, and are, therefore, still gossiping in search of new connections. We see that, particularly for the 10K network, the most of the nodes meet their termination criterion within the first few hundred rounds, which means they do not spend any network resources thereafter.

Our formula for a node's expected degree is derived from the mathematical model presented in Section 3.2. However, it is based on the assumption of a large enough number of nodes and therefore we evaluate the error introduced by this approximation. We construct the histogram of all expected degrees, which corresponds to the expected degree distribution and use it to compute an *a*pproximate γ. In each Figure 7a–7c and 8a–8c we compare the target γ, the approximate γ and the γ values of the self-emerging overlays.

Most importantly, though, the graphs of the second row show that the vast majority of the nodes reach their exact expected degree, contributing to the excellent γ approximation observed in the first row graphs.

Finally, the third row graphs (i.e, Figure 7(g-i) and Figure 8(g-i)) show the number of nodes *not* contained in the largest cluster. For low values of γ the largest cluster is massive, containing virtually the whole set of nodes. This is expected, as nodes tend to have high degrees. For higher values of γ, though, which experience long tails of nodes with very low degrees, we see that the resulting overlay is split in many disconnected components. This does not mean that nodes are isolated at an individual level (as confirmed by the graphs of the second rows), but that nodes are connected according to their expected degrees in smaller components. Making sure of connecting all these components in a single connected overlay is the subject of future work.

5 Conclusions

Self-emerging power-law networks are an important area of research. However, algorithms that generate such topologies in a controlled manner are still scarce.

In this work, we investigated existing approaches to sequential power-law graphs generation and selected a model that allowed for straightforward decentralization. We then experimentally identified limitations with the selected model which have been supported by our theoretical findings. We presented a novel model, built on a thorough mathematical support, that addressed the limitations found with previous models. Under the same experimental settings, our results show that our proposed model significantly outperforms the initial one in different convergence aspects.

Next, we implemented a prototype self-emerging power-law network based on our model and gossiping protocols. We show that the theoretical and

sequential implementations of the novel model are closely followed in performance by the decentralized prototype. Furthermore, the theoretical bounds are observed throughout an extensive set of experiments. Such a result encourages us to consider the theoretical model already robust with respect to implementation approximations and to continue our research efforts having this model as a foundation. One interesting future research question, identified by our decentralized prototype evaluation, is how to alleviate the the problem of (many) disconnected components.

References

1. Albert, R., Barabási, A.-L.: Emergence of scaling in random networks. Science (1999)
2. Barabasi, A.L., Albert, R.: Emergence of scaling in random networks. Science 286, 509–512 (1999)
3. Batagelj, V., Brandes, U.: Efficient generation of large random networks. Phys. Rev. E 71(3), 036113 (2005)
4. Caldarelli, G., Capocci, A., De Los Rios, P., Muñoz, M.A.: Scale-free networks from varying vertex intrinsic fitness. Phys. Rev. Lett. 89(25), 258702 (2002)
5. Cohen, R., Havlin, S.: Scale-free networks are ultrasmall. Phys. Rev. Lett. 90(5), 058701 (2003)
6. Dangalchev, C.: Generation models for scale-free networks. Physica A: Statistical Mechanics and its Applications 338(3-4), 659–671 (2004)
7. de Solla Price, D.J.: A general theory of bibliometric and other cumulative advantage processes. Journal of the American Society for Information Science 27(5), 292–306 (1976)
8. Demers, A., Greene, D., Hauser, C., Irish, W., Larson, J., Shenker, S., Sturgis, H., Swinehart, D., Terry, D.: Epidemic Algorithms for Replicated Database Maintenance, pp. 1–12. ACM Press, New York (1987)
9. Dorogovtsev, S.N., Goltsev, A.V., Mendes, J.F.F.: Pseudofractal scale-free web. Phys. Rev. E 65, 66122 (2002)
10. Dorogovtsev, S.N., Mendes, J.F.F.: Effect of the accelerating growth of communications networks on their structure. Phys. Rev. E 63(2), 025101 (2001)
11. Dorogovtsev, S.N., Mendes, J.F.F., Samukhin, A.N.: Structure of growing networks with preferential linking. Phys. Rev. Lett. 85(21), 4633–4636 (2000)
12. Yule, G.U.: A Mathematical Theory of Evolution Based on the Conclusions of Dr. J. C. Willis, F.R.S. Journal of the Royal Statistical Society 88(3), 433–436 (1925)
13. Flegel, F., Sokolov, I.M.: Canonical fitness model for simple scale-free graphs. Phys. Rev. E 87, 022806 (2013)
14. Krapivsky, P.L., Redner, S., Leyvraz, F.: Connectivity of growing random networks. Phys. Rev. Lett. 85(21), 4629–4632 (2000)
15. Voulgaris, S., Gavidia, D., van Steen, M.: Cyclon: Inexpensive membership management for unstructured p2p overlays. 13(2), 197–217 (June 2005)