Degree and component size distributions in the generalized uniform recursive tree

2008 J. Phys. A: Math. Theor. 41 185101

(http://iopscience.iop.org/1751-8121/41/18/185101)

J. Phys. A: Math. Theor. **41** (2008) 185101 (8pp) doi:10.1088/1751-8113/41/18/185101

Degree and component size distributions in the generalized uniform recursive tree

Zhongzhi Zhang[1,2], **Shuigeng Zhou**[1,2], **Shanghong Zhao**[1,2] **and Jihong Guan**[3]

[1] Department of Computer Science and Engineering, Fudan University, Shanghai 200433, People's Republic of China
[2] Shanghai Key Lab of Intelligent Information Processing, Fudan University, Shanghai 200433, People's Republic of China
[3] Department of Computer Science and Technology, Tongji University, 4800 Cao'an Road, Shanghai 201804, People's Republic of China

E-mail: zhangzz@fudan.edu.cn, sgzhou@fudan.edu.cn and jhguan@mail.tongji.edu.cn

Received 7 January 2008, in final form 2 April 2008
Published 18 April 2008
Online at stacks.iop.org/JPhysA/41/185101

Abstract
We propose a generalized model for the famous uniform recursive tree (URT) by introducing an imperfect growth process, which may generate disconnected components (clusters). The model undergoes an interesting phase transition from a singly connected network to a graph consisting of fully isolated nodes. We investigate the distributions of degree and component sizes by both theoretical predictions and numerical simulations. For the nontrivial cases, we show that the network has an exponential degree distribution while its component size distribution follows a power law, both of which are related to the imperfect growth process. We also predict the growth dynamics of the individual components. All analytical solutions are successfully contrasted with computer simulations.

PACS numbers: 89.75.Hc, 36.40.Qv, 02.50.Ey

1. Introduction

Degree distribution and component (cluster) size distribution are two important properties [1] of complex networks, which have become a focus of attention for the scientific community [2–5]. It has been established that degree distribution has an important consequence on other structural characteristics of networks [1], such as average path length, clustering coefficient, betweenness centrality and others. Particularly, degree distribution also has profound effects on almost all aspects of dynamic processes taking place on networks, including robustness [6, 7], percolation [8–10], synchronization [11], games [12], epidemic spreading [13–16], etc.

On the other hand, as a fundamental structural feature of networks, component (especially the giant component) size is one of the most significant measures of the network robustness [6–10]; in particular, component size distribution is directly related to many significant practical issues such as the distribution of the sizes of disease outbreaks for disease propagation on contact networks [17–19].

A wide variety of network models have been proposed to describe real-life systems and study their structural properties [2–5], among which the uniform recursive tree (URT) is perhaps one of the earliest and most widely studied simplest models [20]. The URT is now acknowledged as one of the two principal models [21] of a random graph (the second one is the Erdös–Rényi (ER) model [22]). It is a growing tree constructed as follows: start with a single node, at each increment of time let a new node be added to the network linked to a randomly selected preexisting node. It has found applications in several areas. For example, it has been suggested as a model for the spread of epidemics [23], the family trees of preserved copies of ancient or medieval texts [24], chain letter and pyramid schemes [25], to name but a few.

Most previous models belong to idealized models [2–5], although they may provide valuable insight into reality. In fact, real-world systems may be viewed as imperfect realizations of idealized theoretical models [26], which have significant influences on their features and function. For instance, introducing defects (impurities) into magnetic materials may drastically change their properties [27]. Recently, many authors have focused on investigating the influence of deleting nodes (or edges) on the function of networks such as their integrity and communication ability [6, 10, 28].

In this paper, we present a generalized model of the URT by introducing an imperfect growth process which may generate disconnected components. The proposed model is governed by a tunable parameter q. We study both analytically and numerically the interesting characteristic aspects of the models, focusing on degree distribution, component size distribution, as well as the evolving dynamics of individual components. We show that both distributions can be computed accurately, although calculation of component size distribution seems difficult for previously studied models. The obtained results indicate that the imperfect growth process affects fundamentally the structure of the network.

2. The model

The proposed generalized model with an imperfection growth process is generated in the following way. We start from an initial state ($t = 0$) of one isolated node. Then, at each increment of time, a new node is added, whose behavior of growth dynamics depends on a parameter q. With probability q, the new node keeps isolated; and with complementary probability $1 - q$, the newly introduced node connects a randomly chosen old node. This growing process is repeated until the network reaches the desired size. One can easily see that at time t, the network consists of $t + 1$ nodes and expected $(1 - q)t$ edges. Thus, when t is large, the average node degree at time t is approximately equal to a constant value $2(1 - q)$.

There are two limiting cases of the present model with the URT as one of its particular case. Therefore, we call the presented model the 'generalized uniform recursive tree'. For $q = 0$, the model coincides with the uniform recursive tree which is a singly connected network. When $q = 1$, the network is reduced to a fully disconnected network. Thus, varying q in the interval (0,1) allows the formation of disconnected clusters as part of the growing dynamics.

3. Distributions of degree and component sizes

Below we will show that some interesting characteristics (i.e., degree distribution, component size distribution and evolutionary dynamics of individual components) may be investigated analytically, which depend on the parameter q, i.e., the imperfect growth process.

3.1. Degree distribution

First we focus on the degree distribution. For $q = 1$, all nodes have the same number of connections 0, the network exhibits a completely homogeneous degree distribution. For the case of $0 \leqslant q < 1$, we can address the degree distribution using the master-equation approach [29]. For simplicity, we label nodes by their time of birth, so that node s refers to the node introduced at time s, and use $p(k, s, t)$ to denote the probability that at time t the node s has a degree k. Then we may analyze the dynamics through a set of master equations [29–31] governing the evolution of the degree distribution of an individual node, which are of the following form:

$$p(k,s,t+1) = \frac{1-q}{t+1}p(k-1,s,t) + \left(1 - \frac{1-q}{t+1}\right)p(k,s,t) + q\delta_{k,0} + (1-q)\delta_{k,1} \tag{1}$$

with the initial condition, $p(k, s = 0, t = 0) = \delta_{k,0}$. This accounts for two possibilities for a node: first, with probability $\frac{1-q}{t+1}$, it may get an extra edge from the new node, and thus increase its own degree by 1 and, second, with the complimentary probability $1 - \frac{1-q}{t+1}$, the node may remain in the former state with the former degree.

Based on equation (1), one can obtain the total degree distribution $P_t(k)$ specifying the probability that a randomly chosen node has degree k at time t:

$$P_t(k) = \frac{1}{t+1}\sum_{s=0}^{t} p(k,s,t). \tag{2}$$

Summing up both sides of equation (1) over s, we get the following master equation for the degree distribution:

$$(t+2)P_{t+1}(k) - (t+1)P_t(k) = (1-q)P_t(k-1) - (1-q)P_t(k) + q\delta_{k,0} + (1-q)\delta_{k,1}. \tag{3}$$

In the infinite t limit, each $P_t(k)$ converges to some limit $P(k)$. Then, the corresponding stationary equation takes the form

$$(2-q)P(k) - (1-q)P(k-1) = q\delta_{k,0} + (1-q)\delta_{k,1}. \tag{4}$$

It has the solution of an exponential form

$$P(k) = \begin{cases} \dfrac{q}{2-q}, & k = 0, \\ \dfrac{2}{2-q}\left(\dfrac{1-q}{2-q}\right)^k, & k \geqslant 1. \end{cases} \tag{5}$$

For $q = 0$, $P(k) = 2^{-k}$ which is exactly the same degree distribution of the uniform recursive tree [3]. Note that this degree distribution $P(k) = 2^{-k}$ is also shared by the Yule tree [32]. There is one major difference between the URT and Yule tree: the former is small-world [33], while the latter lacks this property [32].

In order to confirm the validity of the obtained analytical prediction of degree distribution, we have performed extensive numerical simulations of the networks. In figure 1, we report the simulation results of the degree distribution for several values of q, from which we can see that the degree distribution decays exponentially with the degree, in agreement with the analytical results.

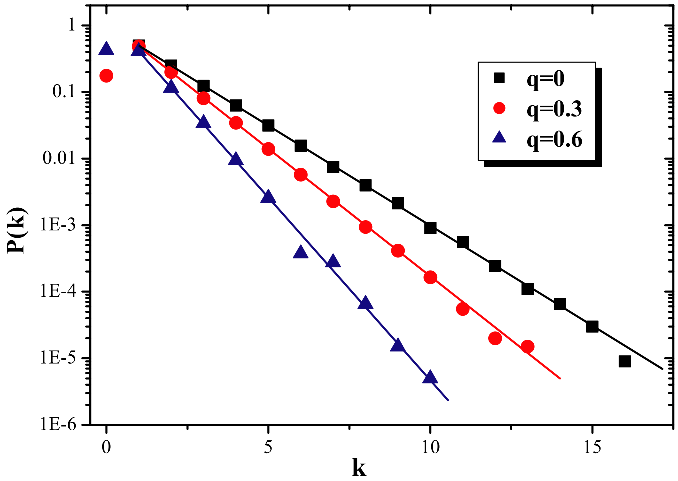

Figure 1. Semilogarithmic graph of degree distribution of the networks with order $N = 200\,000$. Each point is an average over 10 independent simulations. The solid lines are the analytic calculation values given by equation (5).

3.2. Component size distribution

Next, we show that the distribution of component sizes follows a power law with exponent determined by parameter q. For the case of $q = 0$, there is only one component in the network, and the component size distribution is trivial. For $q = 1$, there are $t + 1$ components, every node belongs to a component of size 1 (the node itself). In the range $0 < q < 1$, the imperfect growth process allows the formation of multiple disconnected components. For this case, we can derive an analytical expression for the component size distribution using the rate-equation method [34].

Initially ($t = 0$), there is only one component in the network. At each subsequent step, a new cluster is created with probability q. Thus, after t step evolution, there are expected $N(c, t) = tq + 1$ components in the network, and the average value of component sizes is asymptotically equal to $\frac{1}{q}$ for large t. Let $N_c(t)$ denote the average number of components with size c at time t. By the very construction of the generalized uniform recursive tree, when a new node enters the network, the rate equations [34–37] that account for the evolution of $N_c(t)$ with time t are

$$\frac{\mathrm{d}N_c(t)}{\mathrm{d}t} = (1 - q)\frac{(c - 1)N_{c-1}(t) - cN_c(t)}{\sum_{c'} c' N_{c'}(t)} + q\delta_{c,1}. \tag{6}$$

Here the first term on the right-hand side of equation (6) accounts for the process in which the new node is connected to one node in a cluster with size $c - 1$, leading to a gain in the number of components with size c. Since there are $N_{c-1}(t)$ components of size $c - 1$ and the new node creates a new edge with probability $1 - q$, such processes occur at a rate proportional to $(1 - q)(c - 1)N_{c-1}(t)$, while the factor $\sum_{c'} c' N_{c'}(t)$ converts this rate into a normalized probability. The second (loss) term describes the new edge connecting to one of the nodes in a cluster with size c turning it into a component with size $c + 1$. The last term on the right-hand side of equation (6) accounts for the continuous introduction of a new component with size 1 (the isolated new node itself).

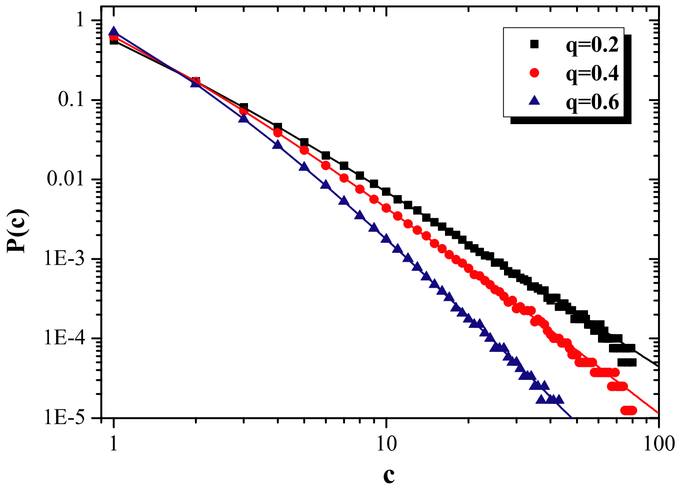

Figure 2. Double-logarithmic plot of the component size distribution of the networks with 200 000 nodes each. The points are the results of computer simulations, and each point is obtained by 20 independent network realizations. The solid lines correspond to the analytic solution provided by equation (8).

Let $P(c)$ be the component size distribution that is the probability of a randomly chosen cluster having size c. In the asymptotic limit $N_c(t) = N(c,t)P(c) \simeq tqP(c)$ and $\sum_{c'} c'N_{c'}(t) = t + 1$. Inserting these into equation (6), we have the following recursive equation:

$$P(c) = \begin{cases} \dfrac{c-1}{c+\frac{1}{1-q}} P(c-1), & c > 1, \\ \dfrac{1}{2-q}, & c = 1, \end{cases} \tag{7}$$

giving

$$P(c) = \frac{1}{2-q} \prod_{m=2}^{c} \frac{m-1}{m+\frac{1}{1-q}} = \frac{1}{2-q} \frac{\Gamma(c)\Gamma\left(2+\frac{1}{1-q}\right)}{\Gamma\left(c+1+\frac{1}{1-q}\right)} \tag{8}$$

for $c > 1$. Thus, in the infinite c limit

$$P(c) \sim c^{-(1+\frac{1}{1-q})}, \tag{9}$$

which shows that component size distribution follows a power-law form with an exponent $\gamma_c = 1 + \frac{1}{1-q}$ dependent on parameter q.

We have performed extensive numerical simulations for the full range of q between 0 and 1. In figure 2, we plot the component size distribution $P(c)$ as a function of q, which agrees well with the analytic result.

3.3. Evolving dynamics of component sizes

In addition to component size distribution, using the continuum approach [38] we can calculate the time dependence of component size c_i of a given component i, $i = 1, 2, \ldots, N(c,t)$. The size will increase by 1 every time a new node is added to the system and linked to one of the nodes belonging to this component. Assuming that c_i is a continuous real variable, according

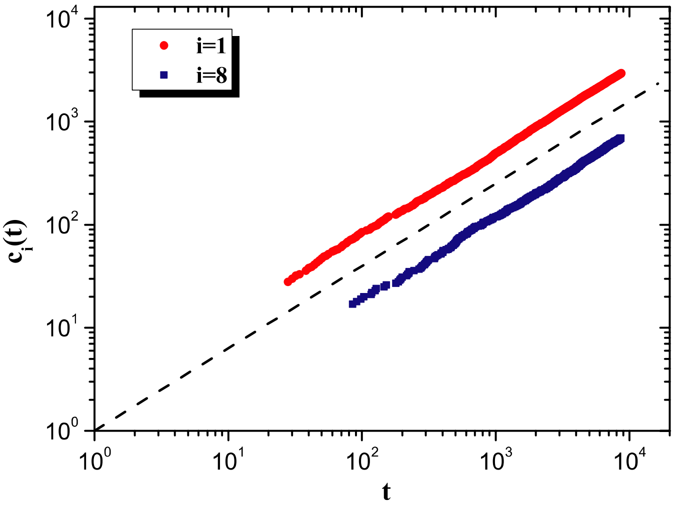

Figure 3. Double-logarithmic graph showing the time-evolution for the sizes of two components. Here $q = 0.2$, and the dashed line has slope 0.8, as predicted by equation (11).

to the generating algorithm of the model, the rate at which c_i changes is clearly proportional to c_i itself. Consequently, c_i satisfies the dynamical equation [38, 39]

$$\frac{\partial c_i(t)}{\partial t} = (1-q)\frac{c_i(t)}{t}. \tag{10}$$

The solution of this equation, with the initial condition that component i was born at time t_i with size $c_i(t_i) = 1$, is

$$c_i(t) = \left(\frac{t}{t_i}\right)^{1-q}. \tag{11}$$

Equation (11) shows that the size of all components evolves the same way, following a power law, the only difference being the intercept of this power law, see figure 3.

4. Conclusion and discussion

To conclude, by introducing an imperfect network growth process we have proposed an extended model for the uniform recursive tree. The presented model interpolates between a network with fully disconnected clusters and the uniform recursive tree with only one single component, which allows us to explore the crossover between the two limiting cases. We have provided both analytically and numerically the solutions for degree and component size distributions of our model. We found that in the case of $0 < q < 1$ the model exhibits an exponential degree distribution and a power-law component size distribution. We also presented that the size of all components evolves as a power-law function of time t.

Our model has a remarkable character that some nodes may become disconnected from the rest of the network. This property has been less reported in earlier studied models and thus has not yet been paid enough attention. Actually, our model may be further extended to include initial attractiveness [29] and preferential attachment [40]. That is, all nodes are born with some initial attractiveness $A > 0$. In the growth progress, the probability that an old node will receive a link from the new node is proportional to the sum of the initial attractiveness and its degree. In this more general case, the degree distribution is power law with exponent dependent on A, while the component size distribution is the same as that of the present model addressed here and is independent of parameter A.

In fact, in the case of a particular q, the proposed generalized uniform recursive tree can also be obtained from the URT by deleting all its edges with probability q, which is exactly the bond percolation problem on the URT. It is known to us all, the bond percolation model can be mapped to a dynamic susceptible-infected-removed (SIR) model [15, 16] for disease propagation, and a percolation transition in the bond percolation represents the onset of an epidemic [17, 18]. Thus, our model could stimulate and find some obvious implications for disease propagation and some other related researches in future. Finally, it should be mentioned that other models with loops are known to be locally tree-like near the percolation point observed upon the bond percolation, hence the results found here may also be relevant to such studies [26].

Acknowledgments

This research was supported by the National Basic Research Program of China under grant no. 2007CB310806, the National Natural Science Foundation of China under grant nos. 60496327, 60573183, 90612007, 60773123 and 60704044, the Shanghai Natural Science Foundation under grant no. 06ZR14013, the China Postdoctoral Science Foundation funded project under grant no. 20060400162, the Program for New Century Excellent Talents in University of China (NCET-06-0376), and the Huawei Foundation of Science and Technology (YJCB2007031IN).

References

[1] Costa L da F, Rodrigues F A, Travieso G and Boas P R V 2007 *Adv. Phys.* **56** 167
[2] Albert R and Barabási A-L 2002 *Rev. Mod. Phys.* **74** 47
[3] Dorogvtsev S N and Mendes J F F 2002 *Adv. Phys.* **51** 1079
[4] Newman M E J 2003 *SIAM Rev.* **45** 167
[5] Boccaletti S, Latora V, Moreno Y, Chavez M and Hwanga D-U 2006 *Phys. Rep.* **424** 175
[6] Albert R, Jeong H and Barabási A-L 2000 *Nature* **406** 378
[7] Albert R, Jeong H and Barabási A-L 2001 *Nature* **409** 542
[8] Callaway D S, Newman M E J, Strogatz S H and Watts D J 2000 *Phys. Rev. Lett.* **85** 5468
[9] Cohen R, Erez K, ben-Avraham D and Havlin S 2000 *Phys. Rev. Lett.* **85** 4626
[10] Cohen R, Erez K, ben-Avraham D and Havlin S 2001 *Phys. Rev. Lett.* **86** 3682
[11] Barahona M and Pecora L M 2002 *Phys. Rev. Lett.* **89** 054101
[12] Szabó G and Fáth G 2007 *Phys. Rep.* **446** 97
[13] Pastor-Satorras R and Vespignani A 2001 *Phys. Rev. Lett.* **86** 3200
[14] Pastor-Satorras R and Vespignani A 2001 *Phys. Rev.* E **63** 066117
[15] Moreno Y, Pastor-Satorras R and Vespignani A 2002 *Eur. Phys. J.* B **26** 521
[16] Colizza V and Vespignani A 2007 *Phys. Rev. Lett.* **99** 148701
[17] Grassberger P 1982 *Math. Biosci.* **63** 157
[18] Newman M E J 2002 *Phys. Rev.* E **66** 016128
[19] Newman M E J 2007 *Phys. Rev.* E **76** 045101
[20] Smythe R T and Mahmoud H 1995 *Theor. Probab. Math. Stat.* **51** 1
[21] Dorogovtsev S N, Krapivsky P L and Mendes J F F 2008 *Europhys. Lett.* **81** 30004
[22] Erdös P and Rényi A 1960 *Pub. Math. Inst. Hung. Acad. Sci.* **5** 17
[23] Moon J W 1974 *London Math. Soc. Lecture Note* **13** 125
[24] Najock D and Heyde C 1982 *J. Appl. Probab.* **19** 675
[25] Gastwirth J 1977 *Am. Stat.* **31** 79
[26] Zen A, Kabakçioğlu A and Stella A L 2007 *Phys. Rev.* E **76** 026110
[27] Hui K and Berker A N 1989 *Phys. Rev. Lett.* **62** 2507

[28] Crucitti P, Latora V, Marchiori M and Rapisarda A 2003 *Physica* A **320** 622
[29] Dorogovtsev S N, Mendes J F F and Samukhin A N 2000 *Phys. Rev. Lett.* **85** 4633
[30] Dorogovtsev S N, Mendes J F F and Samukhin A N 2001 *Phys. Rev.* E **63** 062101
[31] Zhang Z Z, Zhou S G, Shen Z and Guan J H 2007 *Physica* A **385** 765
[32] Chan D Y C, Hughes B D, Leong A S and Reed W J 2003 *Phys. Rev.* E **68** 066124
[33] Dorogovtsev S N, Mendes J F F and Oliveira J G 2006 *Phys. Rev.* E **73** 056122
[34] Krapivsky P L, Redner S and Leyvraz F 2000 *Phys. Rev. Lett.* **85** 4629
[35] Krapivsky P L and Redner S 2001 *Phys. Rev.* E **63** 066123
[36] Wang L, Du F, Dai H P and Sun Y X 2006 *Eur. Phys. J.* B **53** 361
[37] Zhang Z Z and Zhou S G 2007 *Physica* A **380** 621
[38] Barabási A-L, Albert R and Jeong H 1999 *Physica* A **272** 173
[39] Zhang Z Z, Rong L L and Comellas F 2006 *Physica* A **364** 610
[40] Barabási A-L and Albert R 1999 *Science* **286** 509