One-Max Constant-Probability Models for Complex Networks

Mark Korenblit, Vadim Talis, and Ilya Levin

Abstract This paper presents a number of the tree-like networks that grow according to the following newly studied principles: i) each new vertex can be connected to at most one existing vertex; ii) any connection event is realized with the same probability p; iii) the probability Π that a new vertex will be connected to vertex i depends not directly on its degree d_i but on the place of d_i in the sorted list of vertex degrees. The paper proposes a number of models for such networks, which are called *one-max constant-probability models*. In the frame of these models, structure and behavior of the corresponding tree-like networks are studied both analytically, and by using computer simulations.

1 Introduction

According to the well-known Barabási-Albert model [1], scale-free networks are characterized by two main mechanisms: continuous growth and preferential attachment. That is, a) the networks expand continuously by addition of new vertices, and b) there is a higher probability that a new vertex will be linked to a vertex already having many connections (high-degree vertex). Most vertices have only a few connections while there are a few highly connected hubs. Vertices of a scale-free network are the elements of any system and its edges represent the interaction between them.

The Barabási-Albert random graph model is described as follows:

Mark Korenblit
Holon Institute of Technology, Israel, e-mail: korenblit@hit.ac.il

Vadim Talis
Jerusalem College of Technology, Israel, e-mail: talisv@yahoo.com

Ilya Levin
Tel-Aviv University, Israel, e-mail: ilia1@post.tau.ac.il

Starting with a small number m_0 of vertices, at every time step we add a new vertex with $m \leq m_0$ edges that link the new vertex to m different vertices already present in the system. To incorporate preferential attachment, we assume that the probability Π that a new vertex will be connected to vertex i depends on the degree d_i of that vertex.

The mechanism of preferential attachment is assumed to be linear in the model, i.e., $\Pi(d_i)$ is proportional to d_i [1]. However, as noted in the same work, in general relationship between $\Pi(d_i)$ and d_i could have an arbitrary form and, therefore, different types of preferential attachment may be considered.

It is of interest to consider a special case when in every step a new vertex is connected to only one of the old vertices ($m = 1$). In this case the resulting graph is a tree known as a *nonuniform random recursive tree*. The probability of linking to its vertex depends on its degree. The structure and properties of such trees are investigated in [2], [5], [6], and many other works. When the probability of linking to a vertex is proportional to its degree, this gives a *random plane-oriented recursive tree*.

Nonuniform random recursive trees have a number of applications. They may serve for modeling pyramidal structures based on the principle "success breeds success". In a pyramid scheme where each entrant competes with those already participating, the experience gained in successful recruiting enhances the prospects for further success as captured by the growth rule of these trees [6]. The example of simulation of stock markets with these trees is given in [4].

In our paper we introduce a number of new network models based on nonuniform random recursive trees, so called *one-max constant-probability models*. These models are characterized by the following features: i) each new vertex may be connected to at most one old vertex, i.e., in every time step at most one new edge appears in the network; ii) any connection event is realized with the same probability p due to external factors; iii) the probability Π that a new vertex will be connected to vertex i depends not directly on its degree d_i but on the place of d_i in the sorted list of vertex degrees.

The proposed network model is rather realistic because in real life the choice of an object may be determined not by an absolute characteristic of the object but by a relative status of this object among other objects. The status itself depends, in its turn, on the objects' characteristics. Besides, this model explicitly defines the order of priorities in the search of appropriate connection and, therefore, it allows not just to analyze the topology of networks, but also to examine the network dynamics step-by-step.

2 Constant-Probability Search Model

The first model (we call it *Constant-Probability Search Model* or *CPSM*) is based on a regular linear search of a vertex with a maximum degree realized by consecutive comparisons of a current maximum degree with a degree's value of a current checked vertex. If this value is greater than a current maximum, the maximum is updated. For vertices with equal degrees, an earlier arrived vertex is preferable. However, in contrast to the standard search, every comparison is performed not always but with probability p. A new vertex is connected to a vertex v with a found maximum degree which, correspondingly, is equal to a true maximum degree with probability p. The degree of vertex v is incremented by 1 and the new vertex's degree is assigned to 1 if it has been connected to any vertex.

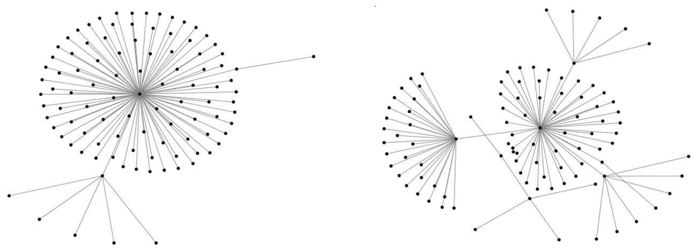
Therefore, the vertex with the 1-st largest degree will be chosen for connection by a new vertex with probability p, the vertex with the 2-nd largest degree – with probability $(1-p)p, \ldots,$ the vertex with the i-th largest degree – with probability $(1-p)^{i-1}p$ (for equal degrees, the degree of a vertex checked earlier is quasi larger). For n existing vertices, the probability that the new vertex will connect to no vertex is equal to $(1-p)^n$.

Proposition 1. *Given an n-vertex network which starts with a single vertex and is based on CPSM, the lower bound of the expected number M_n of the maximum degree in the network is equal to $p(n-1)$.*

Below, one can see that Proposition 1 holds not only for CPSM but also for all other one-max constant-probability models.

It is clear that the higher is p, the larger is degree of the first vertex in the network and the rather this degree is maximum. That is, older vertices increase their connectivity at the expense of the younger ones and a "rich-get-richer" phenomenon [1] is detected for high p.

Diagrams of two 100-vertex networks simulated for different values of p are presented in Fig. 1. Three the largest degrees in a network are indicated (degree of a vertex arrived in time step t is denoted by d_t).



(a) p=0.95, d_1=93, d_2=6, d_3=2 (b) p=0.5, d_2=51, d_5=28, d_8=7

Fig. 1 100-vertex networks based on CPSM

3 Constant-Probability Ordered Model

The second model, so called *Constant-Probability Ordered Model* (*CPOM*) is similar to CPSM. However, in contrast to CPSM, the list of existing vertices is kept sorted in decreasing order of their degrees so that the vertex with a maximum degree is in the top of the list. The list is scanned from the top and a new vertex is connected to the first vertex v which "is allowed to be connected by the probability p". The degree of vertex v is incremented by 1 and this vertex is moved toward the top of the list to find a proper new place for it. The new vertex's degree is assigned to 1 and this vertex is inserted into the list above vertices with degrees 0 (*isolated vertices*) if it has been connected to any vertex.

The running time of the search of appropriate connection in an n-vertex CPSM network is $O(n)$ for any p since always all existing vertices of the network have to be checked. At the same time, CPOM gives $O(n)$ running time in the average case only, while in the best case its running time is $O(1)$. Besides, CPOM exhibits a real network that has a mechanism which keeps most referred sites in the top of the list and makes them, correspondingly, more reachable than others.

Despite the different algorithms used by CPSM and CPOM, both models provide identical network topologies and diagrams illustrated in Fig. 1 are appropriate to CPOM as well.

CPOM (as CPSM) is characterized by the following phenomenon that becomes apparent for low p. Some vertices which come first may remain isolated since while a network is not large, a new vertex may rather connect to no existing vertices and find oneself at the bottom of the list. Next later vertices will find more vertices in the network and the probability of their connecting to one of existing vertices will be higher. At that, they will be linked with a higher probability to vertices with larger degrees and their degrees after connection will be 1. Therefore, as the size of the network increases, the chance of vertices with zero degrees "to be found" by new vertices decreases.

Fig. 2 illustrates the above phenomenon for $p = 0.1$. A network after 100 time steps (Fig. 2 (a)) and the same network after 1000 time steps (Fig. 2 (b)) have the same 6 isolated vertices with order numbers $1, 5, 11, 15, 23, 27$.

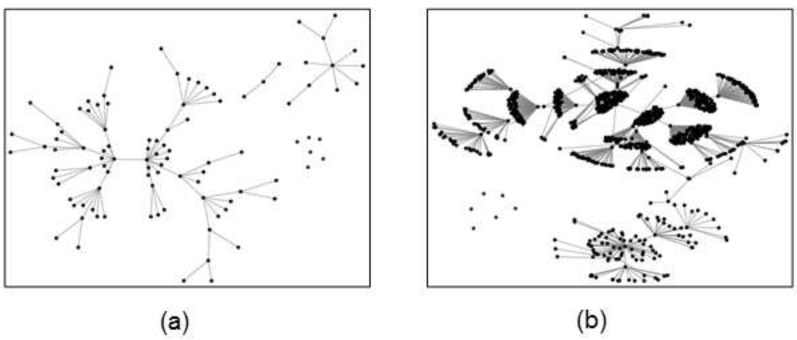


Fig. 2 The phenomenon of first isolated vertices for CPOM

Proposition 2. *Given an n-vertex network based on CPSM or CPOM, the expected number I_n of isolated vertices in the network is defined recursively as follows:* $I_1 = 1$; $I_{n+1} = I_n + 2(1-p)^n - (1-p)^{n-I_n}$.

The result is well-reasoned. For $p = 0$, $I_{n+1} = I_n + 1$ (the number of isolated vertices increases in every time step). For $p = 1$, $I_{n+1} = I_n$ (all new vertices are connected to the first one and the number of isolated vertices does not increase at all). For large n, I_{n+1} tends to I_n (probabilities of appearance of new isolated vertices and of connecting new vertices to old isolated vertices decrease).

Corresponding computational results for p from 0 to 1 are presented in Fig. 3. One can see that for $p < 0.5$, the higher is p, the smaller is n for which I_n reaches saturation and the smaller is I_n in saturation itself. For $p > 0.5$, the expected number of isolated vertices is less than 1.

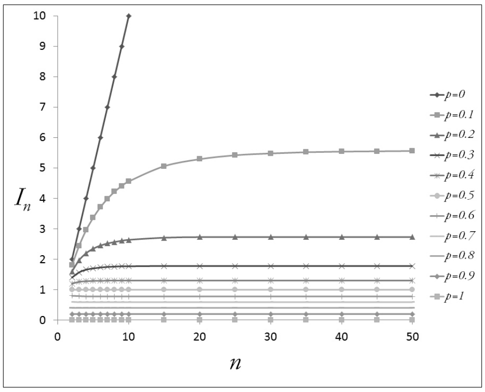


Fig. 3 Expected numbers of isolated vertices in CPOM and CPSM networks

4 Constant-Probability Ordered Non-0 Model (CPOM-N0)

In order to neutralize the negative effect described in the previous section, when some vertices which come first may remain isolated, we slightly modify CPOM. A new vertex connected to one of existing vertices is not inserted above isolated vertices and remains at the bottom of the list. Thus old vertices with zero degrees will not be at the bottom and the list will be sorted only concerning degrees exceeding 1. Such a model is appropriate to be called *Constant-Probability Ordered Non-0 Model* (*CPOM-N0*).

The example of this model's behavior for $p = 0.1$ is shown in Fig. 4. In Fig. 4 (a) one can see a network after 100 time steps. This network has 3 isolated vertices: 5, 12, and 17. The same network after 300 time steps is presented in Fig. 4 (b). It has the only isolated vertex 5. At last, after 1200 time steps, there are no isolated vertices in this network (Fig. 4 (c)).

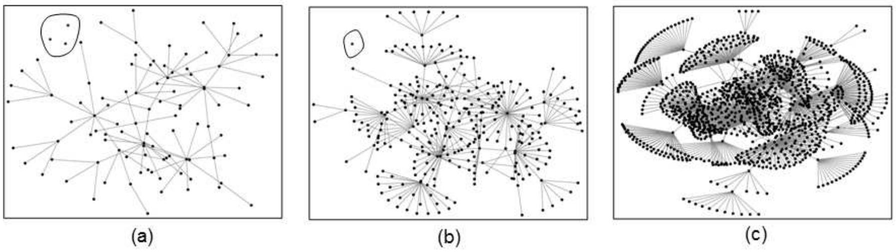


Fig. 4 A network based on CPOM-N0 ($p = 0.1$)

CPOM-N0 is evidence that the additional advantage of CPOM in contrast with CPSM is its flexibility. The list of existing vertices in CPOM is actually the priority list. While in CPSM a vertex's degree directly determines the vertex' priority, in CPOM the vertex's place in the list is this criterion. One can define this place not only as a function of a degree but as a function of additional parameters as well.

There are also other differences in behavior of CPOM and CPOM-N0. Isolated vertices not only disappear in networks based on CPOM-N0 for large n. For the same small n, the expected number of vertices with zero degree in a CPOM-N0 network is less than in a CPOM network. On the other hand, the expected number of *connected components* (collections of connected vertices which have no connections to one another) consisting of more than one vertex in a CPOM-N0 network is greater than in a CPOM network of the same size. The explanation of this phenomenon is the following. An isolated vertex of a CPOM network may rather remain isolated in the next time steps than in a CPOM-N0 network in which this vertex has a higher probability to become a start vertex of a new autonomous part of the network. In any case, both networks are characterized by the same expected number of connected components including isolated vertices that is equal to the number of vertices which were isolated some time, i.e., to the number of appearances of isolated vertices.

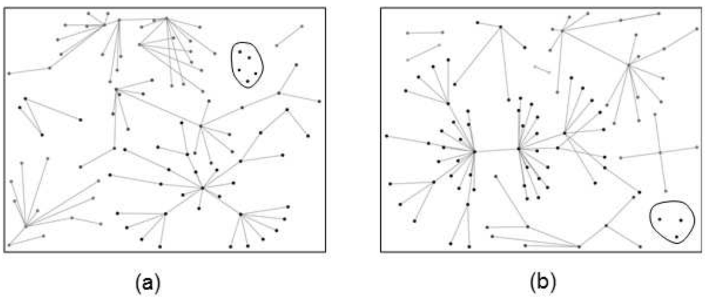


Fig. 5 100-vertex networks

Two corresponding examples are illustrated in Fig. 5. In Fig. 5 (a) one can see the CPOM network after 100 time steps. This network has 11 connected components, 5 of which are isolated vertices (5, 12, 14, 30, 57). The CPOM-N0 network after 100 time steps presented in Fig. 5 (b) has also 11 connected components and only 3 of them are isolated vertices (25, 33, 40). With increase of the network in Fig. 5 (b), new vertices will connect to these 3 vertices sooner or later, while the probability of connecting new vertices to 5 isolated vertices in Fig. 5 (a) will decrease in every time step. Herewith, both networks will consist of 11 connected components, and the probability of appearance of new connected components will decrease with increase of the networks.

The expected numbers of connected components including isolated vertices are equal in networks of the same size based on all one-max constant-probability models. This fact allows to formulate and to prove the following proposition:

Proposition 3. *Given an n-vertex network based on a one-max constant-probability model, the expected number C_n of connected components in the network is defined recursively as follows:* $C_1 = 1;\ C_{n+1} = C_n + (1-p)^n$.

Corollary 1. *Given a network discussed in Proposition 3, the expected number C_n of connected components in the network is expressed explicitly as follows:* $C_n = 1 + (1-p)\frac{1-(1-p)^{n-1}}{p}$. *With increase of n, C_n tends to* $\frac{1}{p}$.

5 Constant-Probability Ordered Directed Model

Previous models assume that connecting a new vertex to an old one leads to increase of a number of connections both of the old and the new vertices. However, not always a subject that initiates a connection is considered as acquiring this connection. At the same time, a referred object is regarded as a possessor of this connection in any case. Thus while most networks (from social to biological ones) are undirected, there are systems that should be simulated by directed networks. For example, Web pages are connected by directed links [3], [7], software modules are taken as vertices of a directed graph with links according to their interaction [3].

We slightly modify CPOM and introduce a *Constant-Probability Ordered Directed Model* (*CPODM*). An edge corresponding to a new connection leaves the new vertex and enters the old one. The list of vertices is sorted by their in-degrees. It is clear that the in-degree of a new vertex is 0 even if it has been connected to any existing vertex and, therefore, a new vertex is always in the bottom of the list.

Out-degree of any vertex in a network based on CPODM is 1 (if the vertex has been connected to any vertex when arriving) or 0 (if the vertex has been connected to no vertex when arriving). As follows from the model's description, the list of vertices does not distinguish between vertices with zero and non-zero out-degrees. For two vertices with zero in-degrees, the older vertex will be nearer to the top. Thus old *isolated vertices* (with zero in-degrees and out-degrees) will not be at the bottom of the list.

One can see that CPODM is similar to CPOM-N0. Although CPOM-N0 describes an undirected network, it distinguishes in special cases between a vertex that is connected to another one and a vertex to which another vertex is connected. In fact, both CPOM-N0 and CPODM identically process new vertices. For this reason, the same characteristic features inherent in both models. Like in CPOM-N0 networks, isolated vertices disappear in networks based on CPODM for large n. For small n, expected numbers of isolated vertices and of connected components consisting of more than one vertex for CPODM are the same as for CPOM-N0.

6 Conclusion

In this paper we proposed a number of new models of tree-like networks and studied genesis and evolution of these networks' topology. Some remarkable network effects were observed. We provided the interpretation of the network behavior on the base of analysis of simulation results.

Specifically, we have discovered the phenomenon of the existence of isolated vertices when subjects that were at the origins of a complex network creation may ultimately find oneself out of the network. We have interpreted the cause of this phenomenon and have shown how it can be prevented. The absence of isolated vertices in a large network, in turn, does not prevent it from splitting on unlinked autonomous parts (connected components) whose number tends to $\frac{1}{p}$ with increase of the network.

References

1. Barabási, A.-L., Albert, R.: Emergence of scaling in random networks, Science **286**, 509–512 (1999).
2. Devroye, L., Fawzi, O., Fraiman, N.: The height of scaled attachment random recursive trees. In: DMTCS Proc. **AM**, AofA'10, pp. 129-142 (2010).
3. Hein, O., Schwind, M., König, W.: Scale-free networks – the impact of fat tailed degree distribution on diffusion and communication processes, Wirtschaftsinformatik **48** (4), 267–275 (2006).
4. Hoffmann, A. O. I., Jager, W., Von Eije, J. H. (2007). Social simulation of stock markets: taking it to the next level. Journal of Artificial Societies and Social Simulation **10** (2), http://jasss.soc.surrey.ac.uk/10/2/7.html
5. Katona, Z.: Levels of a scale-free tree, Random Structures and Algorithms **29** (2), 194–207 (2006).
6. Mahmoud, H. M., Smythe, R. T., Szymański, J.: On the structure of random plane-oriented recursive trees and their branches, Random Structures and Algorithms **4** (2), 151–176 (1993).
7. Raigorodski, A. M.: Models of random graphs and their use (in Russian), Trudy MFTI **2** (4), 130–140 (2010).