Biology helps to construct weighted scale-free networks

A. Ramezanpour

Department of Physics, Sharif University of Technology
P.O. Box 11365-9161, Tehran, Iran and
Institute for Advanced Studies in Basic Sciences - Zanjan 45195-159, Iran

received 14 June 2004; accepted in final form 25 August 2004

PACS. 89.75.-k – Complex systems.
PACS. 89.75.Fb – Structures and organization in complex systems.

Abstract. – In this work we study a simple evolutionary model of bipartite networks whose evolution is based on the duplication of nodes. Using analytical results along with numerical simulation of the model, we show that the above evolutionary model results in weighted scale-free networks. Indeed we find that in the one-mode picture we have weighted networks with scale-free distributions for interesting quantities like the weights, the degrees and the weighted degrees of the nodes and the weights of the edges.

Introduction. – Most of interacting systems can be regarded as complex networks [1–3]. Certainly, seeking the structural and universal properties of these networks is a main goal in studying the behavior of these systems [4–8]. Among these one can refer to the small-world phenomenon [4] and the scale-free behavior of degree distribution [1], where the degree denotes the number of nearest neighbors of a node. Clearly, finding the basic ingredients to produce such behaviors helps us in a better understanding of the real networks.

An interesting feature of the real networks is that they are complex weighted networks [9, 10]. For example, we can associate a weight to each node of a network which might represent the size or power of that node to create connections with the other nodes. In a protein complex network this weight is the number of proteins attributed to a protein complex [11,12]. We could also assign a weight to each edge of a network which might be a measure of interaction between the end point nodes of the edge in the network. In the example of protein complex network, this weight shows the number of proteins that two protein complexes have in common. The weight of an edge in this case would be an appropriate measure to quantify the functional correlation of two protein complexes connected by that edge. In the same way, one could consider social collaboration networks, e.g. scientific coauthorship networks [13–15], as weighted networks.

As the above paragraph reveals, there are a large number of weighted networks which can be exhibited as a one-mode picture of a bipartite network [15,16]. For example, in the case of the protein complex network we could make a bipartite network of proteins and protein complexes. An edge in this bipartite network only connects a protein (a node of type I) to a protein complex (a node of type II) and represents a membership relation. Thus, one can

(*) E-mail: ramzanpour@iasbs.ac.ir

© EDP Sciences
ask if there is a simple rule for the evolution of bipartite networks which reproduces the basic features of the above complex weighted networks.

In this paper we study a simple model for the evolution of bipartite networks. To this end we apply a well-known rule of biology in the context of protein evolution, that is duplication of proteins, to the evolution of bipartite networks. It has been shown that this mechanism can well reproduce the structural properties of the protein interaction networks [17,18]. In ref. [12] the same procedure has been applied to simulate the evolution of a protein complex network. Let us illustrate the duplication mechanism in the example of a scientific coauthorship network. A new article in this network could well be assumed as a result of an old article (it has been duplicated) with some changes probably in the list of the authors (its connections have undergone mutation). This new article may also introduce a new author to the list of the present authors (it creates a new node of the other type). Note that an author with a higher number of articles has a larger probability to produce a new article. This feature automatically enters the model if we select randomly an article for duplication. This is an important property of the duplication mechanism which results in the emergence of scale-free distributions. It turns out that the simple model introduced in this paper can generate complex weighted networks with scale-free distributions for both the weight of the nodes and the edges and also for the degree and the weighted degree of the nodes in the one-mode pictures. We define the weighted degree of a node as the sum of the weights of the edges emanating from that node.

The paper is organized as follows. In the next section we give the model definition in detail. The third section is devoted to the analytic study of the model along with the results of the numerical simulations. In the fourth section we briefly study the effect of limited ages of the nodes on the structure of the network. The fifth section includes the conclusion remarks of the paper.

The model definition. – Consider a bipartite network with $n$ nodes of the first type and $N$ nodes of the second type. We will label nodes of type I by small letters like $a, b, c, \ldots$ and nodes of the other type by capital letters, that is $A, B, C, \ldots$. In the same way each quantity will be represented by small or capital letter according to its relation with the type of nodes. For example, by $m_a(t)$ and $m_A(t)$ we mean, respectively, the weight of a first- and a second-type node at time $t$. Here the weight of a node is the number of its connections in the bipartite network. To evolve the network we go through the following rules:

i) In each step we choose one type of nodes for duplication. With probability $\lambda$ a node of the first type and with probability $1 - \lambda$ a node of the second type will be chosen for duplication. Suppose that we have decided to duplicate a node of the first type. Now a node of this type is randomly chosen and fully duplicated.

ii) Finally, the new node creates a node of the other type and connects to it.

These processes have been shown in fig. 1. Note that all the above processes occur in one time step that is from $t$ to $t + 1$ and the same events could happen for a node of the second type. Moreover, all the processes except the selection of type of the duplicated node, are deterministic.

Note that in any stage of the evolution we can construct two simple networks which are the two one-mode pictures of the above bipartite network. For instance, the one-mode picture of nodes of type I at time $t$ is constructed as follows: Take all the nodes of type I. Then two nodes $a$ and $b$ in this network are connected by an edge if they have at least one node of type II in common in the bipartite picture. We also assign weight $w_{ab}(t)$ to this edge which gives the number of common neighbors of type II. In fig. 1 we have showed the one-mode picture of nodes of type I for the bipartite network represented in this figure. Similarly, one can construct the other one-mode picture which consists of nodes of type II.
Analytic study of the model. – Note that due to the symmetry of the evolution if we compute the behavior of nodes of type II we could get the behavior of the other type only by replacing $\lambda$ with $1 - \lambda$.

As the initial condition we start at time $t = 1$ with one node of type I which has been connected to one node of type II. Thus, according to the deterministic creation of the nodes, the number of nodes at time $t$ will be given by $n(t) = N(t) = t$.

Let us start by writing the average weight of a typical node of type II at time $t + 1$ if it entered the network at time $t_A \leq t$. Note that the probability that a member of node $A$ be selected for duplication is $\lambda m_A(t)/n(t)$. With this probability the weight of node $A$ will increase by one due to the connection with the new node. Thus,

$$m_A(t + 1) = m_A(t) + \lambda m_A(t)/n(t). \quad (1)$$

But the average weight of this node at the time of its birth $t_A$ is $1$ or $1 + m_B(t_A - 1)$, respectively, for the case of first-type duplication or duplication of node $B$. These events occur with probability $\lambda$ and $(1 - \lambda)/N(t_A - 1)$, respectively. Then, averaging over different candidates of node $B$, we find

$$m_A(t_A) = 1 + (1 - \lambda) \sum_{B=1}^{N(t_A - 1)} m_B(t_A - 1)/N(t_A - 1). \quad (2)$$

In the continuum approximation where we take $t$ as a continuum variable, eq. (1) has the following solution:

$$m_A(t) = m_A(t_A)(t/t_A)^{\lambda} = [1 + (1 - \lambda)(1 + \ln t_A)](t/t_A)^{\lambda}, \quad (3)$$

where we have used the fact that $n(t) = N(t) = t$. For $\lambda = 1$ this equation gives

$$m_A(t) = t/t_A. \quad (4)$$

Note that in each step a new node of type II has been introduced to the network. Thus, using the conservation of probabilities $S(m)dm \sim dt_A$ and the above relation between $m_A$ and $t_A$ we find that $S(m)$, the weight distribution of nodes of type II, behaves like

$$S(m) \sim m^{-2}. \quad (5)$$

On the other hand, for $\lambda = 0$ we have from eq. (3)

$$m_A(t) = 2 + \ln t_A, \quad (6)$$

that means the weight of node $A$ is fixed for ever at $t_A$. Utilizing the same procedure as for the determination of $S(m)$ in the case of $\lambda = 1$, we find

$$S(m) \sim e^m. \quad (7)$$

But the network is finite and in this case we expect to see the finite-size effects for the smaller values of $m$. Therefore, for large values of $m$ we will have an exponential decay for $S(m)$ due to the finite size of the network. These arguments have been confirmed in fig. 2 which shows the results of the numerical simulations in these cases.

Now let us focus on the evolution of the degree of a node in the one-mode picture of nodes of type II. In fact the number of neighbors of node $A$ increases by one when the duplicated
A. Ramezanpour: Biology helps to construct weighted scale-free networks

Fig. 1 – A step of the evolution of a bipartite network in which node c of type I has been duplicated. The results of this duplication are node e and E. One can also see the one-mode picture of this bipartite network constructed by nodes of type I. The number of lines between two nodes gives the number of common neighbors of second type that is the weight of the edge connecting them.

Fig. 2 – Weight distribution of the nodes of type II. The parameters are $t = 1000$, $\lambda = 1$ (squares), $\lambda = 0.5$ (circles) and $\lambda = 0$ (triangles). The data are the result of averaging over 50 runs of the evolution of the model. This number is the same in all the numerical simulations of the model represented in this paper. The guideline shows a power law behavior of exponent 2.

node is one of its members in the case of first-type duplication. The probability for this to happen is $\lambda m_A(t)/n(t)$. In the case of second-type duplication, the number of neighbors increases only when the duplicated node is a neighbor of node A or the node A itself. This probability is given by $(1 - \lambda) (k_A(t) + 1)/N(t)$. Thus for node A with $t_A \leq t$, we have

$$k_A(t + 1) = k_A(t) + \lambda m_A(t)/n(t) + (1 - \lambda) (k_A(t) + 1)/N(t).$$

(8)

Here we restrict ourselves to the simple case of $\lambda = 1$. Then, the average degree of node A at the time of its birth is

$$k_A(t_A) = \frac{n(t_A - 1)}{\sum b=1 m_b(t_A - 1)/n(t_A - 1)}.$$

(9)

We can solve eq. (8) for $\lambda = 1$ and in the continuum approximation we find

$$k_A(t) = t/t_A + \ln(t_A - 1),$$

(10)

which for $t \to \infty$ predicts a power law degree distribution of exponent 2 for the large degrees. In fig. 3 we have shown the degree distribution for some values of $\lambda$. Note that for $\lambda = 0$ we have a fully connected network in which $k_A(t) = t - 1$ for all the nodes.

But each edge of the above network has a weight $w$ and one can speak of weighted degree of a node $Z_A(t) = \sum_{B \neq A} w_{AB}(t)$. Let us define $m(a|A; t)$ as the average weight of a member of node A at time $t$. We also take $w(B|A; t)$ as the average weight of an edge emanating from node A in the one-mode picture. Then one can easily find the following equation for $Z_A(t+1)$:

$$Z_A(t+1) = Z_A(t) + \lambda m_A(t)m(t|A)/n(t) + (1 - \lambda) [m_A(t)/N(t) + k_A(t)w(t|A)/N(t)].$$

(11)
Again we take $\lambda = 1$. In this situation the weighted degree of node $A$, when it enters the network is

$$Z_A(t_A) = \sum_a m_a(t_A - 1)/n(t_A - 1).$$

(12)

Then, solving eq. (11) for $\lambda = 1$ we find

$$Z_A(t) = [(\ln t + 1 + \ln[(t_A - 1)/t_A])] t/t_A,$$

(13)

where we have used the following relation:

$$Z_A(t) = m_A(t)[m(t|A) - 1].$$

(14)

As is seen, for $t \to \infty$ we expect a power law behavior of exponent 2 for $P(Z)$. Figure 4 shows this distribution for some values of $\lambda$. Note that for $\lambda = 0$ we have a complete network of nodes of type II in the one-mode picture. Thus as the figure shows, in this case $Z_A(t) \geq N(t)$ for all nodes. Indeed solving eq. (11) for $\lambda = 0$, we find

$$Z_A(t) = [4(t_A - 1) - 1 - \ln[(t_A - 1)/t_A]] t/t_A - 2 - \ln t_A,$$

(15)

that for $t, t_A \gg 1$ is nearly given by $4t - \ln t_A$. Thus, we expect that weighted degrees of nodes distribute around $4t$ as fig. 4 displays.

Finally, we study the average weight of an edge in the one-mode picture. The average weight of the edge between two nodes $A$ and $B$ with $t_A < t_B \leq t$, increases by one only when the duplicated node is of type I and moreover is a member of both the nodes. Thus, we find

$$w_{AB}(t + 1) = w_{AB}(t) + \lambda w_{AB}(t)/n(t).$$

(16)

Let us take $\lambda = 1$. Then, when node $B$ enters the network at time $t_B$, the average weight of edge $AB$ with $t_A < t_B$ is given by

$$w_{AB}(t_B) = m_A(t_B - 1)/n(t_B - 1).$$

(17)
Using \( n(t) = t \) and solving eq. (16) for \( \lambda = 1 \) we find

\[
 w_{AB}(t) = \frac{t}{(t_A t_B)}.
\]

One can easily check that in this case the weight distribution of the edges behaves like \( E(w) \sim w^{-2} \) for large times. On the other hand, for \( \lambda = 0 \) we expect to have an exponential behavior. These statements are confirmed by virtue of the numerical simulations in fig. 5.

**Role of limited lifetime.** – In this section we briefly address the effect of limited age of the nodes on the behavior of the distributions studied in the previous section. To this end, we assign a lifetime to each type of the nodes. That is, a node will be active only during its life which is \( t^* \) or \( T^* \), according to the type of the node. Then only the active nodes of each type will have the opportunity to be selected for duplication. Moreover, the new node can only establish connections with the active nodes of the other type. Obviously, the number of active nodes of each type during the evolution will be always less than or equal to the assigned lifetimes. Nevertheless, the total number of nodes of each type is as before equal to \( t \). To see the role of the limited ages, we consider the case of \( \lambda = 1 \) with \( t^* = \infty \). Since the qualitative behavior of studied distributions is nearly the same, we shall only focus on \( E(w) \), the weight distribution of the edges in the one-mode picture of the nodes of type II. In fig. 6 we show \( E(w) \) for some values of \( T^* \). As the figure shows, by decreasing \( T^* \) the general behavior of distribution does not change and even its exponent remains constant. We found that this picture does not change very much even for the case of \( t^* = T^* \).

**Conclusion.** – In summary, we have shown how weighted scale-free networks could be generated by a simple model for the evolution of bipartite networks. The evolution of this model is based on the duplication of the nodes. We showed that by tuning \( \lambda \) which controls the rate of duplication of the nodes of different types, one can go from a power law regime to an exponential one where the tail of the distributions fall off exponentially. We also studied the effect of limited age for the nodes and showed that the qualitative behavior of the studied distributions does not change.
Certainly, the above-studied model is far from the evolution of real networks. For example, the scale-free behavior of most of biological networks does not only stem from the duplication mechanism. Here mutations play also an important role in the evolution of the network. However, we expect that the qualitative behavior of distributions studied in this paper does not significantly change by introducing mutations. This, for instance, has been shown for a variant of the above model in ref. [12]. Here the aim was to show how a simple mechanism for the evolution of un-weighted bipartite networks can result in the generation of weighted scale-free networks in the one-mode picture.

***

The author is grateful to V. Karimipour for helpful suggestions in writing the manuscript.

REFERENCES

[1] Albert R. and Barabási A.-L., Rev. Mod. Phys., 74 (2002) 47.
[2] Dorogovtsev S. N. and Mendes J. F. F., Evolution of Networks: From Biological Nets to the Internet and WWW (Oxford University Press) 2003.
[3] Newman M. E. J., SIAM Rev., 45 (2003) 167.
[4] Watts D. J. and Strogatz S. H., Nature, 393 (1998) 440.
[5] Barabási A.-L. and Albert R., Science, 286 (1999) 509.
[6] Amaral L. A. N., Scala A., Barthélémy M. and Stanley H. E., Proc. Natl. Acad. Sci. USA, 97 (2000) 11149.
[7] Albert R., Jeong H. and Barabási A.-L., Nature, 406 (2000) 378.
[8] Newman M. E. J., Phys. Rev. Lett., 89 (2002) 208701.
[9] Barrat A., Barthélemy M., Pastor-Satorras R. and Vespignani A., Proc. Natl. Acad. Sci. USA, 101 (2004) 3747.
[10] Barrat A., Barthélemy M. and Vespignani A., Phys. Rev. Lett., 92 (2004) 228701.
[11] Gavin A.-C. et al., Nature, 415 (2002) 141.
[12] Mashaghi A., Ramezanpour A. and Karimipour V., to be published in Eur. Phys. J. B, cond-mat/0304207.
[13] Newman M. E. J., Proc. Natl. Acad. Sci. USA, 98 (2001) 404.
[14] Barabási A. L., Jeong H., Neda Z., Ravasz E., Schubert A. and Vicsek T., Physica A, 311 (2002) 590.
[15] Ramasco J. J., Dorogovtsev S. N. and Pastor-Satorras R., cond-mat/0403438.
[16] Newman M. E. J., Strogatz S. H. and Watts D. J., Phys. Rev. E, 64 (2001) 026118.
[17] Wagner A., Mol. Biol. Evol., 18 (2001) 1283.
[18] Sole R. V., Pastor-Satorras R., Smith E. and Kepler T. B., Adv. Complex Syst., 5 (2002) 43.