LOCA L-WORLD EVOLVING NETWORKS WITH TUNABLE CLUSTERING

ZHONG-ZHI ZHANG∗
Institute of System Engineering, Dalian University of Technology,
2 Ling Gong Rd., Dalian 116024, Liaoning, China
xinjizzz@sina.com

LI-LI RONG
Institute of System Engineering, Dalian University of Technology,
2 Ling Gong Rd., Dalian 116024, Liaoning, China
llrong@dlut.edu.cn

BING WANG
Department of Applied Mathematics, Dalian University of Technology,
2 Ling Gong Rd., Dalian 116024, Liaoning, China
bingbignmath@yahoo.com.cn

Received (Day Month Year)
Revised (Day Month Year)

We propose an extended local-world evolving network model including a triad formation step. In the process of network evolution, random fluctuation in the number of new edges is involved. Our model can unify the generic properties of real-life networks: scale-free degree distribution, high clustering and small average path length. We derive analytical expressions for degree distribution and clustering coefficient. Moreover, in our model, the clustering coefficient is tunable simply by changing the expected number of triad formation steps after a single local preferential attachment step.

Keywords: Local-world; Scale-free networks; Complex networks; Disordered systems.

1. Introduction

Complex networks[1,2,3] describe many systems in nature and society, such as Internet[4], World Wide Web[5], metabolic networks[6], protein networks in the cell[7], co-author networks[8] and sexual networks[9]. Most of complex networks share three apparent features. (a) The degree distribution of nodes is scale-free, i.e., it follows a power law. (b) The clustering coefficient is high. Two nodes having a common neighbor are far more likely to be linked to each other than are two nodes selected

∗Corresponding author.
randomly. (c) The average path length (APL) is small. That is, the expected number of edges needed to pass from one arbitrarily selected node to another is low.

How to model complex networks with these three properties? Traditionally the study of complex networks has been the scope of graph theory. While graph theory initially focused on regular graphs, since the 1950s large-scale networks with no apparent design principles have been depicted as random graphs, proposed as the simplest and most straightforward realization of a complex network. In the past few years, with the computerization of data acquisition process and the availability of high computing power, scientists have found that most real-life networks are neither completely regular nor completely random, but have three properties above. So they proposed some new models to depict real-life networks. Among them, two are the most well known. One is the small-world network model proposed by Watts and Strogatz (WS) in the year of 1998, which interpolates between regular and random graphs and has two properties of high clustering and short APL. The other is scale-free network model with features of power-law degree distribution and low APL addressed by Barabási and Albert (BA). Although these two pioneering models played an important role in network science and started an avalanche of research on complex networks, neither of them can completely describe the three characteristics of real-life networks. After that, a great number of attempts have been made to construct models with the three properties coinciding with real-life networks. Holme and Kim extended the BA model to include a triad formation step. Klemm and Eguíluz introduced a growing network model based on a finite memory of the nodes. Saramäki and Kaski presented an undirected scale-free network model generated by random walkers. In relation to the problem of Apollonian packing, Andrade et al. introduced Apollonian networks which were also proposed by Doye and Massen. Zhou et al. presented a simple rule generating random two-dimensional Apollonian networks. Zhang et al. offered a simple general algorithm producing high-dimensional (random) Apollonian networks. All these models may capture some mechanisms responsible for the above traits shared by real-life networks, but they have ignored some other significant factors. For example, in various real-life networks such as World Trade Web, when a new node enters the system, it doesn’t have the globe information of all nodes existed before, so preferential attachment mechanism only works on the local-world of the new node. Li and Chen (LC) proposed a local-world evolving network model which enable people to better understand and describe this real-life phenomenon. But LC model has a low clustering coefficient.

In order to portray real-life network more appropriately, we present a local-world evolving network model with changeable local-world size and tunable clustering, which capture both the mechanism of local preferential attachment and triad formation. Our model has the three common features of real-life networks. Moreover, it represents a transition between exponential scaling and power-law, so it may depict some real-networks such as scientific collaboration networks whose
degree distribution is neither power-law nor exponential. Additionally, in the existing model there is a random fluctuation in the number of new edges acquired by the network which is more realistic. We analyze the geometric characteristics of the model both analytically and numerically. The analytical expressions are in good agreement with the numerical simulation results.

2. The LC Local-world Evolving Network

Two ingredients, i.e. growth and preferential attachment in local-world, inspired Li and Chen to introduce the LC model for dynamical evolving networks to capture the localization of real-life networks. The generation algorithm of the LC model is the following:

1. Initial condition: The network has a small number \( m_0 \) of nodes and small number \( e_0 \) of edges. And then we perform the following two steps:

2. Growth: At every time step, we add one node \( v \) with \( m(< m_0) \) edges to the existing networks.

3. Determining local-world: Randomly choose \( M \) nodes from the existing network, which are considered as the “Local-world” for the new node \( v \).

4. Local Preferential attachment (LPA): The node \( v \) connects to \( m \) different nodes in its local-world determined in step (3). We assume that the probability \( \Pi \) that node \( v \) will be connected to node \( i \) already present in the local-world of node \( v \) depends on the degree \( k_i \) of node \( i \), such that

\[
\Pi_{\text{Local}}(k_i) = \frac{M}{m_0 + t} \frac{k_i}{\sum_{\text{Local}} k_j}
\]

In the LC model, in the case of \( m < M < m_0 + t \), it represents a transition between power-law and exponential scaling networks. In particular, the original BA model is a special case of this local-world evolving network model.

3. Extended Local-world Evolving Networks with Changeable Local-world size and Tunable Clustering

The LC model captures a common characteristic of many real-life networks that nodes have local-world connectivity. However, clustering coefficient of the LC model approaches zero when the network is large, which will be addressed in the following section. To incorporate the high clustering, making use of the method introduced by Holme and Kim we modify the LC model by adding an additional triad formation (TF) step: In the previous LPA step if there is an edge connecting the new node \( v \) and an existing node \( w \), then we add one more edge from \( v \) to a randomly selected neighbor of \( w \) with a given probability (see Fig.1). If all neighbors of \( w \) have connected to \( v \), do a LPA step instead. The concrete expression is as following: at every time step, when a new node \( v \) enters the existing network, we perform a LPA step first, and then with the probability \( p \) we implement a TF step. In succession, we carry out a LPA step followed by a TF step with the probability \( p \). After this
process repeats \( m \) times we go to the next time step. It is worth noticing that in our model a LPA step is always followed by a TF step with probability \( p \), which we take as the control parameter in our proposed model. So in the existing model, at every time step there are \( m \) LPA steps and random fluctuated \( p \)-dependent TF steps between 0 and \( m \) whose expectation is \( mp \). That is to say, when new nodes are added to the network at different time steps, the number of new edges is generally not constant. \(^{34}\) After \( t \) time steps, the model develops to a network with \( m_0 + t \) nodes and expected \( (1 + p)mt + e_0 \) edges. Then we can compute the average node degree of the network at time \( t \), \( \langle k \rangle \), as \( t \to \infty \):

\[
\langle k \rangle_{t \to \infty} = \frac{2[(1 + p)mt + e_0]}{t + m_0} \approx 2m(1 + p) \tag{2}
\]

Fig. 1. Local preferential attachment and triad formation. In the LPA step (a) with a probability proportional to its degree a node \( u \) is selected to link to the new node \( v \). In the TF step (b) the new node \( v \) randomly selects one neighborhood node \( w \) of node \( u \) chosen by \( v \) to attach to in the previous LPA step. \( \times \) symbolizes “not allowed to attach to”. After [19].

Additionally, with the network growth more information is available for the new node, so the size \( M \) of local-world increases with time. Thus, different from the LC model, we allow for changes in \( M \), which is denoted here \( M_t \). We assume that

\[
M_t = a(m_0 + t) + b \quad \text{and limit} \quad m \leq M_t \leq m_0 + t.
\]

The interpretation for our model is that most real-life networks exhibit such evolving mechanism. For example, in the network of scientific citations, a new manuscript is more likely to cite well-known and thus much-cited publications than less-cited and consequently less-known papers in the same domain of the manuscript.
Moreover, with the lapse of time, there are more papers available in the domain for new node to refer, so the size of local-world increases with time. On the other hand, in the content of citations a not untypical scenario is that after citing a few famous references an author may simply cite secondary references from the famous ones.

It is evident that at every time step, the parameters in the above-modified model always meet the following conditions: \( m \leq M_t \leq m_0 + t \) and \( 0 \leq p \leq 1 \). So there are three limiting cases of our model, now we consider them below.

Case A: When \( p = 0 \), and \( M_t = m \), it is a growing network with uniform attachment which is the same case of model A in the Barabási-Albert scale-free model.\(^{14}\)

Case B: When \( p = 0 \), and \( M_t = t + m_0 \), the local-world of the new node is the whole network. Our model reduces to the original BA model.\(^{13,14}\)

Case C: When \( p = 0 \), \( M_t = \text{const} \) and \( m \leq M_t \leq m_0 + t \), the model reduces to the LC model.\(^{31}\)

As discussed in Section 2, when \( m \leq M_t \leq m_0 + t \), the LC model represents a transition between power-law and exponential scaling networks. It is the same with our model, because the LC model is a special case of ours. So our considered model may depict some real-life networks whose degree distribution is neither power-law nor exponential.

### 4. Analytical Calculation of Relevant Network Parameters

#### 4.1. Degree distribution

We focus on investigating the case of \( M_t \gg m \). If the local-world scale has \( M_t \gg m \), our model described above has a power-law degree distribution, similar to the BA network.\(^{13,14}\) We can interpret this by calculating analytically based on the mean-field approach in Ref. [14], [20] and [31]. We assume that the degree \( k_i \) of node \( i \) is continuous, and thus the probability given by Eq. (1) can be interpreted as a continuous rate of change of \( k_i \). In a LPA step, node \( i \) increases its degree with the rate

\[
\frac{\partial k_i}{\partial t} = \frac{M_t}{m_0 + t} \sum_{\text{Local}} k_j\]  

(3)

For a TF step, we can gain the average increase of \( k_i \) via the probability given by the following equation

\[
\frac{\partial k_i}{\partial t} = \frac{M_t}{m_0 + t} \sum_{n \in \Omega} k_n \sum_{\text{Local}} k_j k_n = \frac{M_t}{m_0 + t} \sum_{\text{Local}} k_j\]  

(4)

where \( \Omega \) is the set of neighbors of node \( i \), and \( k_j \) is the number of nodes in \( \Omega \).

Similar to the fluctuation in Ref. [28] and [29], here the fluctuation of triad formation steps also has little impact on the growth dynamics \( k_i(t) \) of node \( i \) and degree distribution \( P(k) \) when network is large enough. So we can suppose that in
one time step we perform $m$ LPA steps and $mp$ TF steps on average. From Eq. (3) and (4) the total rate per time step is expressed as

$$\frac{\partial k_i}{\partial t} = m\frac{M_i}{m_0 + t \sum_{\text{Local}} k_j} + mp\frac{M_i}{m_0 + t \sum_{\text{Local}} k_j} = \frac{m(1 + p)M_i}{m_0 + t \sum_{\text{Local}} k_j}$$

(5)

We assume that the cumulative degree of the local-world meets the following expression

$$\sum_j k_j = \langle k \rangle M_t$$

(6)

where $\langle k \rangle$ is the average degree of all the nodes in the networks given by Eq. (2). Substituting Eq. (6) into Eq. (5) we get

$$\frac{\partial k_i}{\partial t} = \frac{m(1 + p)M_i}{m_0 + t} \approx \frac{k_i}{2t}$$

(7)

which has the same form as the original BA model. The solution of this equation, with the initial condition that node $i$ was added to the system at time $t_i$ with the expected value of connectivity $k_i(t_i) = m(1 + p)$, is

$$k_i(t) = m(1 + p) \left( \frac{t}{t_i} \right)^{0.5}$$

(8)

and this results in the power-law degree distribution of form

$$P(k) = \frac{2m(1 + p)(m(1 + p) + 1)}{k(k + 1)(k + 2)}$$

(9)

which at the limit $m(1 + p) \gg 1, k \gg 1$ can be written in the common form $P(k) = 2(1 + p)^2m^2k^{-3}$. In Fig. 2, the degree distributions at various values of $p$ and $m$ are shown. In the process of simulation, the local-world size $M_t$ scales as $M_t = 0.1(m_0 + t) + m$. From Fig. 2, one can easily see that the simulated degree distributions match very well with the theoretical ones. Comparing (b) and (c) with (a) and (d), we observe that for small values of $k$, there is a deviation from power-law behavior in (b) and (c), which originates from the fluctuation in the number of new links acquired by the system (see Ref. [28] and [29]). It should be noted that many real-life networks such as the World Wide Web, the actor collaboration graph and scientific collaboration network indeed exhibit this phenomenon of deviation from power-law to some degree for small $k$ values.

4.2. Clustering coefficient

By definition, clustering coefficient $C_i$ of a vertex $i$ is the ratio of the total number $e_i$ of existing edges between all $k_i$ its nearest neighbors and the number $k_i(k_i - 1)/2$ of all possible edges between them, i.e. $C_i = 2e_i/[k_i(k_i - 1)]$. The clustering coefficient of the whole network is the average of all individual $C_i$'s.
Fig. 2. Degree distribution $P(k)$ versus $k$ on a logarithmic scale. The size of networks is 10000. The open circles, stars, and squares denote the cases of $m = 1, 3, 5$, respectively. The slope of the dashed lines is -3. (a),(b),(c),(d) exhibit cases of $p = 0, 0.1, 0.5, 1$, respectively.

Using the mean-field rate-equation theory, we can calculate $C_i$ analytically. Here we also place our emphasis on the case of $M_t \gg m$. Fig. 3 illustrates the main microscopic mechanisms increasing $e_i$: (a) node $i$ is connected to the new node in a LPA step, which is potentially followed by one TF step; (b) in a LPA step the new node attaches to one of the neighbors of $i$, and then in one of the subsequent TF steps the new node conversely gets linked to $i$; (c) in a LPA step node $i$ is connected to the new node and in another LPA step a neighbor of $i$ is also selected for connection to the new node; (d) in a TF step node $i$ is connected to the new node and in another TF step a neighbor of $i$ is also selected for connection to the new node. (c) node $i$ is connected to the new node in a LPA step, and in the potential TF steps which follow LPA steps when the new node connects to the neighbor nodes of $i$, the new node gets linked to $i$. Here we exclude secondary triangle formation that takes place if two TF steps from the new node form a triangle composed of two of $i$’s neighbors and the new node, which has little effect on the clustering of node $i$. So the rate equation for $e_i$ reads

$$
\frac{\partial e_i}{\partial t} = m \frac{k_i}{2m(1+p)t} p + m \sum_{n \in \Omega} \frac{k_n}{2m(1+p)t} \frac{1}{k_n} p \\
+ m \frac{k_i}{2m(1+p)t} (m-1) \sum_{n \in \Omega} \frac{k_n}{2m(1+p)t} \\
+ mp \frac{k_i}{2m(1+p)t} (m-1) p \sum_{n \in \Omega} \frac{k_n}{2m(1+p)t} \\
+ m \frac{k_i}{2m(1+p)t} (m-1) p \sum_{n \in \Omega} \frac{k_n}{2m(1+p)t}
$$

(10)
Fig. 3. The microscopic mechanisms increasing $e_i$. The dashed edges increase $e_i$.

The five terms in the right hand of Eq. (10) give the increase in $e_i$ by mechanism from (a) to (e) in turn. It should be noted that the third term describes mechanism (c) and it is the only one that would remain if we consider the LC model. In Eq. (10), $k_i/[2m(1+p)t]$ is the local preferential attachment probability and triad formation probability to node $i$; $k_n$ denote the degrees of the neighbors of node $i$, and $1/k_n$ come from the fact that the neighboring node where a TF step links is chosen uniformly from the neighbors; $\sum_{n\in\Omega} k_n$ is the sum of the degrees of neighbors of $i$.

After some simplifications to Eq. (10), we get

$$\frac{\partial e_i}{\partial t} = 2m \frac{k_i}{2m(1+p)t} p + (1 + p + p^2)m \frac{k_i}{2m(1+p)t} (m - 1) \sum_{n\in\Omega} \frac{k_n}{2m(1+p)t}$$  \hspace{1cm} (11)

In addition, for uncorrelated random networks we have

$$\sum_{n\in\Omega} k_n = k_i \frac{\langle k \rangle}{4} \ln t = k_i \frac{(1+p)m}{2} \ln t$$  \hspace{1cm} (12)

We approximate $e_i$ by integrating both sides in Eq. (11). The integral for the first term in the right hand of Eq. (11) is simply

$$\int_1^N 2m \frac{k_i}{2m(1+p)t} p dt = \frac{2p}{1+p} \int_1^N \frac{dk_i}{dt} dt = \frac{2p}{1+p} [k_i(N) - m(1+p)]$$  \hspace{1cm} (13)
where we made use of Eq. (7). Using Eq. (8) and (12), we can integrate the second term in the right hand of eq. (11)

$$\int_1^N (1 + p + p^2) m_{\frac{k_i}{m(1+p)^t}} (m-1) \sum_{n \in \Omega} \frac{k_n}{m(1+p)} \ln t dt$$

$$= (1 + p + p^2) \frac{m-1}{m(1+p)} \int_1^N \frac{k^2}{m(1+p)} \ln t dt$$

$$= (1 + p + p^2) \frac{m^2(m-1)(1+p)}{8(1+p)} \frac{(\ln N)^2}{2} N^2 k_i^2(N)$$

Combining this with Eq. (13) yields

$$e_i = e_{i,0} + \frac{2p}{1 + p} k_i(N) + (1 + p + p^2) \frac{m-1}{8(1+p)} \frac{(\ln N)^2}{N} k_i^2(N)$$

The clustering coefficient for nodes with large degree $k$ becomes

$$C(k) = \frac{e}{k(k-1)/2} \approx \frac{2p/(1 + p)}{k} + (1 + p + p^2) \frac{m-1}{8(1+p)} \frac{(\ln N)^2}{N}$$

after neglecting $n_{i,0}$ (see Fig. 4).
The parameter $p$ in our model introduces the clustering effect into the system by allowing the formation of triads. By setting $p$ to a value between 0 and 1 the average clustering can be adjusted continuously and grows monotonically with an increasing $p$. In the representation of $C(k)$, the first term can be ascribed to the triad formation induced clustering, and shows the $k^{-1}$ behavior that has been observed in several real-life networks. Furthermore, this clustering property is similar to other models such as pseudofractal scale-free network model and the highly clustered model and (random) Apollonian networks. Note that $C(k)$ consists of a power law and a constant, so perfect power-law behavior follows only when the former one dominates. Moreover, since $n_{i,0}$ has been neglected, Eq.(16) and the inverse proportionality hold true for the nodes with $k_i$ large enough, only. It is easy to know that for $p = 0$, we get the clustering coefficient $C(k)$ of nodes in the LC model

$$C(k) = \frac{m - 1}{8} \left( \ln N \right)^2 N$$

which goes to zero as $N$ becomes large enough.

From the above discussions, we find that the existing model shows both the scale-free nature and the high clustering at the same time. Moreover, our model exhibits small-word property. In Fig. 5 we present average path length $L(N)$ versus network size $N$ with $m = 1$ at various values of $p$. One can see that $L(N)$ increases logarithmically with $N$. In fact, for $m > 1$, $L$ grows slower than the case of $m = 1$ with the same $p$, because in the latter case the network is sparser.

![Fig. 5. Average path length $L(N)$ versus network size $N$ on a semilogarithmic scale with $m = 1$. The open circles, diamonds, and squares denote the cases of $p=0, 0.5, 1$, respectively. The straight dashed lines are fits to the data.](image-url)
Local-world Evolving Networks with Tunable Clustering

5. Conclusion

In summary, by mimicking principles in network formation, we have presented an extended local-world evolving network model for creation of growing networks with power-law degree distribution, a finite clustering and small average path length. And we have obtained the analytic solutions for relevant network parameters of the considered model. By changing the expected value of triad formation steps after a single LPA, one can tune the clustering coefficient in a systematic way. In addition, in the evolution of the network, random fluctuation in the number of new edges is involved which can be adjusted via tuning $p$ and have little effect on the characteristics of the networks. Although local-world exists in many real-life networks, it should be pointed out that the choice of a local-world in real-life world networks is more intricate and flexible. We use here the most generic case, i.e. random selection, in our proposed model as in the LC model. Future work should include studying in detail the real formation mechanisms of local-worlds in different real-life networks as well as their impacts on network topology and dynamics.

Acknowledgments

This research was supported by the National Natural Science Foundation of China under Grant No. 70431001.

References

1. R. Albert, A.-L. and Barabási, Rev. Mod. Phys. 74 (2002) 47.
2. S. N. Dorogovtsev and J. F. F. Mendes, Adv. Phys. 51 (2002) 1079.
3. M. E. J. Newman, SIAM Review 45 (2003) 167.
4. M. Faloutsos, P. Faloutsos and C. Faloutsos, Comput. Commun. Rev. 29 (1999) 251.
5. R. Albert, H. Jeong and A.-L. Barabási, Nature 401 (1999) 130.
6. H. Jeong, B. Tombor, R. Albert, Z.N. Oltvai and A.-L. Barabási, Nature 407 (2000) 651.
7. H. Jeong, S. Mason, A.-L. Barabási and Z. N. Oltvai, Nature 411 (2001) 41.
8. M. E. J. Newman, Proc. Natl. Acad. Sci. U.S.A. 98 (2001) 404.
9. F. Liljeros, C.R. Edling, L.A.N. Amaral, H.E. Stanley, Y. Åberg, Nature 411 (2001) 907.
10. P. Erdős and A. Rényi, Pub. Math 6 (1959) 290.
11. P. Erdős and A. Rényi, Pub. Math. Insti. Hung. Acad. Sci. 5 (1960) 17.
12. D. J. Watts and S. H. Strogatz, Nature 393 (1998) 440.
13. A.-L. Barabási and R. Albert, Science 286 (1999) 509.
14. A.-L. Barabási, R. Albert and H. Jeong, Physica A272 (1999) 173.
15. M. E. J. Newman and D. J. Watts, Phys. Lett. A263 (1999) 341.
16. M. Barthélemy and L.A.N. Amaral, Phys. Rev. Lett. 82 (1999) 3180.
17. A. Barrat and M. Weigt, Eur. Phys. J. B13 (2000) 547.
18. S. N. Dorogovtsev, J. F. F. Mendes and A. N. Samukhin, Phys. Rev. Lett. 85 (2000) 4633.
19. P. L. Krapivsky, S. Redner and F. Leyvraz, Phys. Rev. Lett. 85 (2000) 4629.
20. P. Holme and B. J. Kim, Phys. Rev. E65 (2002) 026107.
21. K. Klemm and V. M. Eguíluz, Phys. Rev. E65 (2002) 036123.
22. K. Klemm and V. M. Eguíluz, *Phys. Rev.* **E65** (2002) 057102.
23. J. Saramäki and K. Kaski, *Physica* **A341** (2004) 80.
24. J. S. Andrade Jr., H. J. Herrmann, R. F. S. Andrade and L. R. da Silva, *Phys. Rev. Lett.* **94** (2005) 018702.
25. J. P. K. Doye and C. P. Massen, *Phys. Rev.* **E71** (2005) 016128.
26. T. Zhou, G. Yan, B. H Wang, *Phys. Rev.* **E71** (2005) 046141.
27. Z.Z. Zhang, F. Comellas, G. Fertin and L.L. Rong, arXiv: cond-mat/0503316
28. Z.Z. Zhang, L.L. Rong and F. Comellas, arXiv: cond-mat/0502591
29. M. A. Serrano and M. Boguñá, *Phys. Rev.* **B68** (2003) 015101.
30. X. Li, Y. Y. and G. R. Chen, *Physica* **A328** (2003) 287.
31. X. Li, and G. R. Chen, *Physica* **A328** (2003) 274.
32. M. E. J. Newman, *Phys. Rev.* **E64** (2001) 016132.
33. Z. H., Liu, Y. C., Lai, N. Ye and P. Dasgupta *Phys. Lett.* **A303** (2002) 337.
34. Z. H., Liu, Y. C., Lai and N. Ye, *Phys. Rev.* **E66** (2002) 036112.
35. G. Szabó, M. Alava and J. Kertész, *Phys. Rev.* **E67** (2003) 056102.
36. V. M. Eguíluz and K. Klemm, *Phys. Rev. Lett.* **89** (2002) 108701.
37. E. Ravasz and A.-L. Barabási, *Phys. Rev.* **E67** (2003) 026112.
38. S. N. Dorogovtsev, A. V. Goltsev and J. F. F. Mendes, *Phys. Rev.* **E65** (2002) 066122.