Growing networks with communities: A distributive link model

Cite as: Chaos 30, 041101 (2020); doi: 10.1063/5.0007422
Submitted: 13 March 2020 · Accepted: 6 April 2020 · Published Online: 23 April 2020

Ke-ke Shang1,2,a Bin Yang1 Jack Murdoch Moore3 Qian Ji1 and Michael Small4,5,b

AFFILIATIONS
1 Computational Communication Collaboratory, Nanjing University, Nanjing 210093, People’s Republic of China
2 College of Information and Communication Engineering, Dalian Minzu University, Dalian 116600, People’s Republic of China
3 School of Physical Science and Engineering, Tongji University, Shanghai 200092, People’s Republic of China
4 Complex Systems Group, Department of Mathematics and Statistics, The University of Western Australia, Crawley, Western Australia 6009, Australia
5 Mineral Resources, CSIRO, Kensington, Western Australia 6151, Australia

a) Author to whom correspondence should be addressed: kekeshang@nju.edu.cn and keke.shang.1989@gmail.com
b) Electronic mail: michael.small@uwa.edu.au

ABSTRACT

Evolution and popularity are two keys of the Barabasi–Albert model, which generates a power law distribution of network degrees. Evolving network generation models are important as they offer an explanation of both how and why complex networks (and scale-free networks, in particular) are ubiquitous. We adopt the evolution principle and then propose a very simple and intuitive new model for network growth, which naturally evolves modular networks with multiple communities. The number and size of the communities evolve over time and are primarily subjected to a single free parameter. Surprisingly, under some circumstances, our framework can construct a tree-like network with clear community structures—branches and leaves of a tree. Results also show that new communities will absorb a link resource to weaken the degree growth of hub nodes. Our models have a common explanation for the community of regular and tree-like networks and also breaks the tyranny of the early adopter; unlike the standard popularity principle, newer nodes and communities will come to dominance over time. Importantly, our model can fit well with the construction of the SARS-Cov-2 haplotype evolutionary network.

Published under license by AIP Publishing. https://doi.org/10.1063/5.0007422

I. INTRODUCTION

Evolving networks are ubiquitous,1−6 from social networks2 to protein networks,3 from the Internet to contact networks,4 and so on. Hence, in all these instances, time is the basic element for a fitness network model.1,14 However, another basic element, the principle of preferential attachment—leading to the scale-free degree distribution—is increasingly controversial15 as an omnipresent...
model of complex systems. In one recent study, it was shown that the scale-free network is rare.\textsuperscript{11} In this report, we argue that time evolution is the key for any network model, but the birth time of a node should not entirely preordain its final state. While the first mover advantage is an advantage, it is not sufficient to guarantee final success—sometimes better ideas will appear later. In particular, the degree distribution of growing engineering networks is usually influenced by both geography and resource-driven demand; hence, links cannot be predicted well via an algorithm that is based solely on the principle of pure preferential attachment.\textsuperscript{15,16}

On the other hand, communities appear in many complex networks and are frequently associated with important functions of those networks.\textsuperscript{17–20} Research into community detection also has recently attracted increasing attention.\textsuperscript{21,22} Actually, the appearance of new communities will pillage the degree resource from old hub nodes, as the links within communities are denser than those between groups, and a new hub node is easily formed with the new community.

In this report, we adopt two keys—network evolution and new communities—to design a growing network model. With time evolution, upstarts (new nodes with high degrees) will appear in new groups, and our model provides a novel perspective on the “rich-get-richer” phenomenon. First, we construct a fully connected network, and every node is one initial community. Second, we evaluate the probability that generates a new community or choose a community by the same probability mass function. Third, we introduce a new node that connects nodes within the new or chosen community with a higher probability and external nodes of that with a lower probability. Fourth and finally, we repeat steps 3 and 4 as required.

In what follows, we show network structures that are constructed via our model and their degree distributions. In regular networks (networks with many crossing links\textsuperscript{23}), with the increasing probability of the appearance of new communities, the degree of the hub node decreases—consistent with the hypothesis of our model. Surprisingly, our model can also construct a tree-like network with natural increasing new branches (communities).

To investigate the community structure of our model more carefully, we measure the visibility of community structures by modularity.\textsuperscript{1} We find that networks that are constructed via our model with appropriate parameters have obvious community structures. Furthermore, we detect the predictability of communities and find that links within a community are more predictable than those between communities. All in all, our model can construct a regular or tree-like network with clear and predictable communities and controllable degree distributions. In the end, we successfully construct a tree-like network with a similar structure to the SARS-Cov-2\textsuperscript{24} haplotype evolutionary network and a regular network with a similar structure to the karate club network.

II. GROWING MODEL

(1) Given \(k (k \geq 1)\) nodes, construct a fully connected network, where each node is an independent community and each community has an independent label \(i\).

(2) Next, we introduce a free control parameter \(\beta\); the probability of the appearance of a new community is inversely proportional to

\[
P_{\text{inverse}} = \frac{n_1}{n + \beta} \times \frac{n_2}{n + \beta} \times \cdots \times \frac{n_i}{n + \beta} \times \frac{\beta}{n + \beta}.
\]

here, \(n\) is the number of nodes of community \(i\) and \(n\) is the number of nodes of the whole network.

(3) If the community is successfully formed, then we introduce a new node as the first member of the new community and randomly connect to \(m (m \leq k)\) existing nodes from other communities. Hence, the label \(i\) of the new community is \(c + 1\), where \(c\) is the number of old communities. Otherwise, introduce a new node as a member of an old community. The community to be chosen is decided by its label number \(i\); the bigger \(i\) means the later birth. In other words, the probability that a particular community is chosen is determined by its age, inversely proportional to \(P_{\text{inverse}}\). The later born community is easy to be chosen with the bigger \(\beta\) then forming more communities.

(4) Next, the new node has a probability \(p\) to connect its own community nodes and \(1 - p\) to randomly connect to other community nodes. Here, \(p\) can be 1; due to the community, the first node has connected the node from other communities at step 2. If the number of nodes within a community is not enough to be connected by its new node, let the new node first connect all nodes within the community randomly and then connect the external nodes randomly.

(5) Repeat (2), (3), and (4) \(n'\) times and then the number of nodes is \(n = k + n'\).

Time is the key power in the BA model, the birth time determines the popularity of nodes, and the popularity of a node determines accretion of links. Hence, the richer get richer and the link resource distribution becomes more unequal. However, our model introduces the control parameter \(\beta\) to create new communities to get the resource and also controls both new and old communities to share the resource, in contrast to a mechanism in which high degree nodes absorb more links over time.

For a large network, only a small number of nodes may connect to other communities. However, a small number of nodes will struggle to make an impact on the community structure—or be strongly wired with their intended community. To accelerate the computation, we can delete these nodes and then add new nodes. Actually, in a real-world growing network, that some nodes only connect other communities is natural. Such as in the SARS-Cov-2 haplotype evolutionary network,\textsuperscript{25} the haplotype h35 (group A) only connects other group B due to the relationship between them, and its community is uncovered. Hence, we can also ignore this phenomenon.

Furthermore, as depicted in Eq. (1), \(P_{\text{inverse}} \to 0\) when \(\beta \to \infty\), and the probability of the appearance of a new community tends to \(\infty\) and the number of communities \(N_c \to n\). On the contrary, \(P_{\text{inverse}} \to \infty\) when \(\beta \to 0\), and then the probability of the appearance of a new community tends to 0 and the number of communities \(N_c \to k\). We set the probability of the appearance of a new community equal to 0 when \(\beta = 0\). Hence, \(\beta \in [0, \infty)\). Here, we provide a fitting method for the relationship between the number of communities \(N_c\) and the \(\beta\),

\[
N_c \sim n^{\frac{\beta}{\ln(k)}}.
\]
\( \gamma \) is a free parameter for fitting. As shown in Fig. 1, we set \( k = 1 \) and then change the number of nodes \( n \). Results show that our fitting method can fit with the results of our model very well.

### III. Fitness of a Growing Model

In this report, we give three examples (Fig. 2) to qualitatively interpret our network growth model. Here, we set the number of initial nodes \( k \), the number of links per step \( m \), and the probability...
to connect to its own existing community ($p$) to be the same and only change the free control parameter $\beta$. We can see that with the increase in $\beta$, the number of communities increases, the size (number of nodes) of the biggest community decreases, and the degree of the largest hub node decreases (Table I). Furthermore, as shown in Table I, the increase in $\beta$ sharply increases the number of communities and will greatly inhibit the growth of the community size. Hence, this will influence the degree of the hub node in each community. The node with the lowest degree still has the same degree value. This phenomenon states that the new nodes are more likely to get richer (more degrees), and more degree resources can be allocated to previously not richer nodes under our control.

Next, we set $m = 1$ and $p = 1$ and then directly construct a tree-like network. As shown in Fig. 3, the surprising result is that our model is very similar to a naturally evolving tree. Initial nodes are tree roots; with the time evolution, each branch is the community and the branch of the branch is the new community; and the rest can be deduced by analogy. The number of communities will increase with the increase in $\beta$; hence, we set $\beta = \min(n_i)$ to avoid a large number of communities.

### A. Regular networks

To further discuss the fitness of our growing network model, we measure the community structure of our model by the default metric modularity. Here, we set $k$, $m$, and $n$ as in Fig. 2. As shown in Fig. 4(a), we can see that the modularity value ($Q$) for all networks will increase with the increase in $p$. This means that, with the increase in $p$, our model will have a clearer so-called community structure. In detail, compared to $\beta = \min(n_i)$ and $\beta = \max(n_i)$, we can see that $\beta = \max(n_i)$ has the worst performance under the metric of modularity. Moreover, $\beta = \text{mean}(n_i)$ has the best performance under the metric of modularity, and the network has the clear community structure ($Q$ is around 0.3) if $p > 0.7$. It should be noted that $\beta = \max(n_i)$ also has a clear community structure when $p > 0.8$.

Furthermore, many previous studies demonstrate that the community can improve the predictability of links. Hence, next, we measure the link predictability by the metric predictability ($Pr^*$),

$$S_{pr} = \frac{\sum_{i=1}^{n} \frac{S_{in} - S_{out}}{S_{in}}}{n}. \quad (3)$$

Here, $S_{in}$ is the algorithm prediction accuracy of internal links between communities and $S_{out}$ is that of all links. We adopt the default metric $AUC^{22}$ to measure our link prediction accuracy. In practice, it is not easy to improve the accuracy of a basis link

| $\beta$ | $N_c$ | $S_{max}$ | $D_{max}$ | $D_{min}$ |
|--------|-------|-----------|-----------|-----------|
| $\beta = \min(n_i)$ | 5     | 20        | 17        | 3         |
| $\beta = \text{mean}(n_i)$ | 8     | 17        | 14        | 3         |
| $\beta = \max(n_i)$ | 17    | 14        | 13        | 3         |

**FIG. 3.** The network structures and the corresponding degree distribution for a tree-like network of our model. In the network structure figure, each community has its unique node shape, the dashed link is the link between communities, and the solid link is the link within the community. We set $m = 1$, $p = 1$, $k = 3$, $n = 50$, and $\beta = \min(n_i)$ for it. Here, a new node randomly connects the old node within the same community.

**FIG. 4.** We set $\beta$ as $\min(n_i)$, $\text{mean}(n_i)$, and $\max(n_i)$, respectively. (a) is the $Q$ value of different regular networks with the increase in $p$. The ordinate is the $Q$ score of communities, and the abscissa is the $p$ value. (b) is the $Pr$ value of different regular networks with the increase in $p$. The ordinate is the $Pr$ score of networks, and the abscissa is the $p$ value.
As in particular, the virus haplotype evolutionary network has a high predictability of community links. Actually, not all networks have a high \( Q \) or \( Pr \) value. In other words, each network has its own particular community structure; actually, not all networks have a high \( Q \) or \( Pr \) value. In this report, we adopt the HEU (the algorithm for tree-like networks) as the basic link prediction algorithm. As depicted in Fig. 5, compared to \( \beta = \text{min}(n_i) \) and \( \beta = \text{mean}(n_i) \), the performance of modularity and predictability both can achieve the best and high-quality (\( Q > 0.3 \) and \( S_p > 0.1 \)) performance when we setting \( \beta = \text{mean}(n_i) \). Hence, for tree-like networks, we suggest that set \( \beta = \text{mean}(n_i) \) can achieve a clear community structure and a higher predictability of community links.

B. Tree-like network

1. Growing model

Many real-world systems are naturally represented as tree-like networks with almost no closed triangular structure. Here, we set \( k, m, p, \) and \( n \) as shown in Fig. 3. We adopt the HEU (the algorithm for tree-like networks) as the basic link prediction algorithm. As depicted in Fig. 5, compared to \( \beta = \text{min}(n_i) \) and \( \beta = \text{mean}(n_i) \), the performance of modularity and predictability both can achieve the best and high-quality (\( Q > 0.3 \) and \( S_p > 0.1 \)) performance when we setting \( \beta = \text{mean}(n_i) \). Hence, for tree-like networks, we suggest that set \( \beta = \text{mean}(n_i) \) can achieve a clear community structure and a higher predictability of community links.

2. The haplotype evolutionary network of SARS-Cov-2

Actually, the tree-like networks are ubiquitous in nature and human society. In particular, the virus haplotype evolutionary network is a natural tree-like network with new communities. In this report, we adopt the SARS-Cov-2 haplotype evolutionary data as a real-world network, a node stands for one kind of a haplotype or a medium vector, and a link stands for an evolutionary relationship between two connected nodes. The SARS-Cov-2 haplotype evolutionary network with 5 groups (communities) has 59 nodes (58 haplotypes and one medium vector mv2) and 59 links before the statistical characteristic of the real network and then regulate \( \beta \) to construct a network with a similar \( Q \) value, a \( Pr \) value, and the number of communities to the real network.

### Table II

The number of communities \( N_c \), the maximum depth of the tree within a community \( L_{max} \), the size of the biggest community \( S_{max} \), the highest degree \( D_{max} \), the lowest degree \( D_{min} \), the modularity value \( Q \), and the predictability value \( Pr \) for networks in Fig. 6.

| Network          | \( N_c \) | \( L_{max} \) | \( S_{max} \) | \( D_{max} \) | \( D_{min} \) | \( Q \)  | \( Pr \)  |
|------------------|----------|---------------|---------------|---------------|---------------|-------|-------|
| SARS-Cov-2       | 5        | 2             | 30            | 27            | 1             | 0.5471| 0.0082|
| Growing model    | 5        | 2             | 34            | 31            | 1             | 0.5208| 0.0071|

(a) SARS-Cov-2   
(b) Growing Model   
(c) BA Model

![Fig. 6](image-url)  
(a) is the network structure of SARS-Cov-2 haplotypes. (b) is the network structure of our model, and we set \( m = 1 \), \( p = 1 \), \( k = 3 \), \( n = 59 \), and \( \beta = \text{min}(n_i) \) for it. Here, a new node connects the oldest node within the same community with a 90% probability and randomly connects other old nodes within the same community. (c) is the network structure of the BA model, and we set \( m = 1 \), \( k = 3 \), and \( n = 59 \).
4 February 2020. Here, in the SARS-Cov-2 network, we observe that the root node in each community is the hub node with the highest degrees; hence, we set that a new node connects the oldest node within the same community with a 90% probability. As depicted in Fig. 6, when we set $\beta = \min(n_i)$ to avoid a large number of communities, $n = 59$. Our model can construct a network that has the same number of communities, the lowest degree, and the maximum depth of the tree within a community as the real evolutionary network. Furthermore, as shown in Table II, our model network has the similar modularity value, the predictability value, and the highest degree as that of the real network. Results show that our model can fit well with the real-world evolving network that has new communities.

**C. Karate club network**

Dissimilar to the haplotype evolutionary network of SARS-Cov-2, previous famous traditional networks have no natural communities. However, researchers proposed many detection algorithms to uncover their defined communities. In this report, we adopt the karate club network as the real regular network: A node represents one club user, and a link indicates that there is a relationship between a pair of nodes. The karate club network is the common used data and has been widely applied for community detection. Moreover, we adopt the foundational community detection algorithm, the Girvan–Newman (GN) algorithm, to detect the karate club network. As depicted in Fig. 7, when we set $\beta = \min(n_i)$ to avoid a large number of communities, $k = 3$, $m = 2$, and $n = 34$. Our model can construct a network that has the same number of communities and the size of the biggest community as the real karate club network. Surprising, our model and the real network each have a community with only one node. Furthermore, as shown in Table III, our model network has a similar modularity value, a predictability value, and the lowest degree as that of the real network. Results show that our model can also fit with the traditional network, which has been widely used for community detection.

**IV. CONCLUSION AND DISCUSSION**

Growing network models or static network models with communities have been proposed many times, but a fitness growth model that incorporates the birth of communities and can achieve tree-like networks is very rare. Our model exhibits a clear community structure, the more predictable links of communities, and the varying degree distribution at the same time. Particularly, in the construction of tree-like networks, our growing model can grow the branches and leafs for the tree-like network, and the almost perfect new branches or leafs that are precisely the new communities. More importantly, compared to the SARS-Cov-2 haplotype evolutionary network, our model can construct a network with the same number of communities and a similar network structure. In addition, our model also can fit with the common used network of community detection.

“Rich get richer” is the common sense for network science; however, new communities often expand rapidly in the real world. Our work helps us to utilize the new community phenomenon and to contribute a novel network model. Another possible way is that...
using the increase in new communities helps us improve the link prediction accuracy of tree-like networks.

**ACKNOWLEDGMENTS**

Ke-ke Shang is supported by the National Natural Science Foundation of China (NNSFC) (No. 61803047) and the Major Project of the National Social Science Foundation of China (No. 19ZDA324). Michael Small is supported by the ARC Discovery Project (No. DP180100718).

**DATA AVAILABILITY**

The genomic data used in this study are available from the article W.-B. Yu, G.-D. Tang, L. Zhang, and R. T. Corlett, “Decoding evolution and transmissions of novel pneumonia coronavirus using the whole genomic data,” ChinaXiv:202002.00033 (2020) and the associated supplementary material. The karate club data used in this study are available from the article M. Girvan and M. E. Newman, “Community structure in social and biological networks,” Proc. Natl. Acad. Sci. U.S.A. 99, 7821–7826 (2002) and the associated supplementary material.

**REFERENCES**

1. G. Bianconi and A.-L. Barabási, “Competition and multiscaling in evolving networks,” Europhys. Lett. 54, 436 (2001).
2. A. Barrat, M. Barthélemy, and A. Vespignani, *Dynamical Processes on Complex Networks* (Cambridge University Press, 2008).
3. J. Cendra, M. C. Gonzalez, P. Wang, T. Schoenharl, G. Madye, and A.-L. Barabási, “Uncovering individual and collective human dynamics from mobile phone records,” J. Phys. A: Math. Theor. 41, 224015 (2008).
4. T. Gross and B. Blasius, “Adaptive coevolutionary networks: A review,” J. R. Soc. Interface 5, 259–271 (2007).
5. P. Holme and J. Saramäki, “Temporal networks,” Phys. Rep. 519, 97–125 (2012).
6. K. K. Shang, M. Small, X. K. Xu, and W. S. Yan, “The role of direct links for link prediction in evolving networks,” Europhys. Lett. 117, 28002 (2017).
7. P. Holme, “Network dynamics of ongoing social relationships,” Europhys. Lett. 64, 427 (2003).
8. J.-D. J. Han, N. Bertin, T. Hao, D. S. Goldberg, G. F. Berriz, L. V. Zhang, D. Dupuy, A. J. Walhout, M. E. Cusick, F. P. Roth et al., “Evidence for dynamically organized modularity in the yeast protein–protein interaction network,” Nature 430, 88 (2004).
9. R. Pastor-Satorras and A. Vespignani, *Evolution and Structure of the Internet: A Statistical Physics Approach* (Cambridge University Press, 2007).
10. J. Stehlé, N. Voirin, A. Barrat, C. Cattuto, V. Colizza, L. Isella, C. Régis, J.-F. Pinton, N. Khanafer, W. Van den Broeck et al., “Simulation of an SEIR infectious disease model on the dynamic contact network of conference attendees,” BMC Med. 9, 87 (2011).
11. A.-L. Barabási and R. Albert, “Emergence of scaling in random networks,” Science 286, 509–512 (1999).
12. E. Papadopoulos, M. Kitsak, M. Á. Serrano, M. Boguñá, and D. Krioukov, “Popularity versus similarity in growing networks,” Nature 489, 337 (2012).
13. A. D. Broio and A. Clauset, “Scale-free networks are rare,” Nat. Commun. 10, 1017 (2019).
14. V. Voitalov, P. van der Hoorn, R. van der Hofstad, and D. Krioukov, “Scale-free networks well done,” Phys. Rev. Res. 1, 033034 (2019).
15. L. Lü and T. Zhou, “Link prediction in complex networks: A survey,” Phys. A: Stat. Mech. Appl. 390, 1150–1170 (2011).
16. K.-K. Shang, T.-C. Li, M. Small, D. Burton, and Y. Wang, “Link prediction for tree-like networks,” Chaos 29, 061103 (2019).
17. M. Girvan and M. E. Newman, “Community structure in social and biological networks,” Proc. Natl. Acad. Sci. U.S.A. 99, 7821–7826 (2002).
18. V. Spirin and I. A. Mirny, “Protein complexes and functional modules in molecular networks,” Proc. Natl. Acad. Sci. U.S.A. 100, 12123–12128 (2003).
19. A. E. Krause, K. A. Frank, D. M. Mason, R. E. Ulanowicz, and W. T. Taylor, “Compartments revealed in food-web structure,” Nature 426, 282 (2003).
20. A. Clauset, M. E. Newman, and C. Moore, “Finding community structure in very large networks,” Phys. Rev. E 70, 066111 (2004).
21. A. Mahmood and M. Small, “Subspace based network community detection using sparse linear coding,” IEEE Trans. Knowl. Data Eng. 28, 801–812 (2015).
22. F. Fortunato and D. Hric, “Community detection in networks: A user guide,” Phys. Rep. 659, 1–44 (2016).
23. M. E. Newman, “Modularity and community structure in networks,” Proc. Natl. Acad. Sci. U.S.A. 103, 8577–8582 (2006).
24. The virus of coronavirus disease 2019 (COVID-19) has been named as “severe acute respiratory syndrome coronavirus 2” (SARS-CoV-2) by the International Committee on Taxonomy of Viruses.
25. W.-B. Yu, G.-D. Tang, L. Zhang, and R. T. Corlett, “Decoding evolution and transmissions of novel pneumonia coronavirus using the whole genomic data,” ChinaXiv:202002.00033 (2020).
26. Y. Yan and S. Gregory, “Finding missing edges in networks based on their community structure,” Phys. Rev. E 85, 056112 (2012).
27. C. V. Cannastraci, G. Alanis-Lobato, and T. Ravasi, “From link-prediction in brain connectome and protein interactome to the local-community-paradigm in complex networks,” Sci. Rep. 3, 1613 (2013).
28. K.-K. Shang, M. Small, Y. Wang, D. Yin, and S. Li, “A novel metric for community detection,” EPL (Europhys. Lett.) 129, 68002 (2020).
29. J. A. Hanley and B. J. McNeil, “The meaning and use of the area under a receiver operating characteristic (ROC) curve,” Radiology 143, 29 (1982).
30. L. Lü and T. Zhou, “Link prediction in weighted networks: The role of weak ties,” EPL (Europhys. Lett.) 89, 18001 (2010).
31. T. Ibrahim and J. Kleinberg, “The link-prediction problem for social networks,” J. Am. Soc. Inf. Sci. Technol. 58, 1019–1031 (2007).
32. E. A. Leicht, P. Holme, and M. E. Newman, “Vertex similarity in networks,” Phys. Rev. E 73, 026120 (2006).
33. In this report, our model can construct many networks with similar structures. Here, we choose the applicable one to show.
34. W. W. Zachary, “An information flow model for conflict and fission in small groups,” J. Anthropol. Res. 33, 452–473 (1977).