

一种基于距离偏好连接的网络演化模型

应君*, 严传魁

温州大学数理学院, 浙江 温州

收稿日期: 2024年4月29日; 录用日期: 2024年5月22日; 发布日期: 2024年5月31日

摘要

考虑到现实网络的实际情况以及BA无标度网络模型不能展现真实网络的高聚类特性, 本文提出了一个带调节参数的基于距离偏好连接的网络演化模型。通过仿真实验显示, 模型的度分布不再遵循无标度分布, 而是满足负指数分布, 并且通过调节距离因素依赖程度参数的数值, 可以调整网络模型的聚类系数, 从而获得具有更高聚类性质的网络模型, 更贴近真实网络的特征。

关键词

复杂网络, 演化机制, 度分布, 聚类系数

A Network Evolution Model Based on Distance-Preference Connections

Jun Ying*, Chuankui Yan

School of Mathematics and Physics, Wenzhou University, Wenzhou Zhejiang

Received: Apr. 29th, 2024; accepted: May 22nd, 2024; published: May 31st, 2024

Abstract

Taking into account the actual situation of real networks and the inability of the BA scale-free network model to capture the high clustering characteristics of real networks, this paper proposes a distance-preference based network evolution model with a tunable parameter. Through simulation experiments, it is shown that the degree distribution of the model no longer follows a scale-free distribution, but instead follows a negative exponential distribution. By adjusting the value of the distance-preference dependence parameter, the clustering coefficient of the network model can be adjusted, thus obtaining a network model with higher clustering properties that is more closely aligned with the characteristics of real networks.

Keywords

Complex Networks, Evolutionary Mechanisms, Degree Distribution, Clustering Coefficient

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着互联网、社交网络和物联网的普及，各事物之间的联系变得更加复杂和密切，研究复杂网络结构有助于揭示这些网络的规律性。复杂网络理论在生物医学领域有着广泛的应用，例如新陈代谢网络[1]、蛋白质相互作用网络[2]，有助于研究疾病传播和药物靶点等重要问题。此外，研究人类社会网络也对理解信息传播和社会动态等现象至关重要，在社会学和人类行为学领域具有重要意义[3] [4]。总的来说，复杂网络研究能够帮助人们更好地理解现实世界中的复杂系统，揭示系统内部的规律和结构，为解决各种实际问题提供理论和方法支持[5]。近年来，依据现实世界的真实网络选择合适的网络演化机制构建网络演化模型，并研究其相关拓扑性质已经成为复杂网络科学的重要研究方向之一。

随着对复杂网络的深入研究，人们逐渐认识到许多真实网络的特性并不简单地符合随机或规则结构，而是展现出高度复杂的拓扑结构。大多数研究指出，现实世界的真实网络具有幂律度分布，较小的平均最短路径长度和高聚类这三个显著特征。然而，一些经典的网络模型如 Watts 和 Strogatz 的小世界网络模型(WS 模型) [6]和 Barabasi 和 Albert 的无标度网络模型(BA 模型) [7]并未同时具备这三个特征。WS 模型缺乏幂律度分布，而 BA 模型则呈现低聚类特性。通过 BA 网络聚类系数的平均场理论[8]可知：BA 无标度网络模型的聚类系数正比于 $(\ln N)^2/N$ 。可见，当节点数趋近于无穷大时，BA 模型的聚类系数将趋近于零，这与现实网络的高聚类特性并不相符。

现代复杂网络科学的核心理论之一是大部分甚至所有真实世界的网络都呈现无标度特性[9] [10]，并且通常认为度为 k 的节点遵循幂律分布 $k^{-\alpha}$ ，其中 $\alpha > 1$ 。然而，关于无标度网络是否普遍存在这一问题一直备受争议。近年来的研究表明，并非所有网络都表现出无标度分布[11] [12]。Tanaka 的研究指出无标度网络解释无法准确描述代谢网络的特性[13]，Clore 的研究则发现 RNA 二级结构网络并非无标度的[14]。这些研究结果表明，对于不同类型的网络，无标度特性并非普遍存在，这促使人们进一步深入探讨网络拓扑结构的多样性和复杂性，以更好地理解真实世界网络的特征和演化规律。

2. 距离偏好网络模型演化算法

考虑到真实网络中的节点之间的连接会受到空间地理因素或特定距离权重的影响，例如在电力网络中，电站之间的地理距离越近，连接的可能性越大；而在社交网络中，人们之间的关系距离越近，结交的可能性也越高等等。为了更好地模拟这一现象，本文提出了一种名为距离偏好网络模型(距离模型)的新型复杂网络演化模型。模型旨在模拟节点之间的距离偏好连接机制，以更准确反映实际网络的特性。

在距离模型中，节点的增长会受到距离偏好连接的影响，即新加入网络中的节点更倾向于连接到距离较近的节点，这一连接方式能更好地模拟真实网络中节点间的关联性。通过考虑距离因素，使得节点的连接方式更贴近实际情况，从而使得网络的拓扑结构更加真实。具体的演化算法如下：在初始网络中，假设有 m_0 个节点随机分布在单位圆上，并且节点之间进行完全连接。随后网络会进行演化过程，在每一

时间步中会引入一个新的节点 v , 该节点随机分布在单位圆上, 并且连接到 m 个已存在的节点上, 这里 $m \leq m_0$ 。新引入的节点 v 与已存在的节点 i 连接的概率 Π_i 受到节点属性之间的距离影响

$$\Pi_i = \frac{1/d_{iv}^\alpha}{\sum_{u \neq v} 1/d_{uv}^\alpha} \quad (1)$$

其中, d_{iv} 取节点间的欧氏距离平方 $d_{iv} = (x_i - x_v)^2 + (y_i - y_v)^2$, x_i 和 y_i 为节点 i 在单位圆上的坐标, α 为距离因素依赖程度的调节参数, 表示对距离的偏好程度, 即 α 的数值越大该网络中的节点越倾向于和自己距离较近的节点相连接。考虑到真实网络的现实情况, 调节参数的取值不会太大, 本文只讨论参数取值 $1 \leq \alpha \leq 4$ 的情况。

3. 数值仿真实验

通过进行数值仿真实验, 研究了网络模型的拓扑特性。在仿真过程中, 设定仿真参数为 $N = 10000$, $m = m_0 = 5$, 并在后续仿真实验中保持这一参数组合不变, 除非另有特殊说明。为减小随机因素对仿真结果的影响, 采用了 50 个网络的平均值进行分析, 以确保结果的稳定性和可靠性。

3.1. 度分布

在复杂网络理论中, 度分布是指描述网络中节点度数(即节点的连接数量)分布情况的概率分布。简单来说, 度分布表示了网络中不同节点的度数出现的频率或概率。具体而言, 复杂网络度分布的定义如下: 假设一个网络包含 N 个节点, 每个节点的度数可以用 k_i 表示, 其中 i 表示节点的编号, k_i 表示第 i 个节点的度数。那么度分布可以定义为: $P(k)$ 表示网络中一个随机选取的节点的度数为 k 的概率。通过度分布, 可以了解网络中节点度数的分布情况, 从而揭示网络的整体结构特征。在实际应用中, 通常会绘制度分布的直方图或概率分布曲线, 以直观地展示网络中节点度数的分布情况。

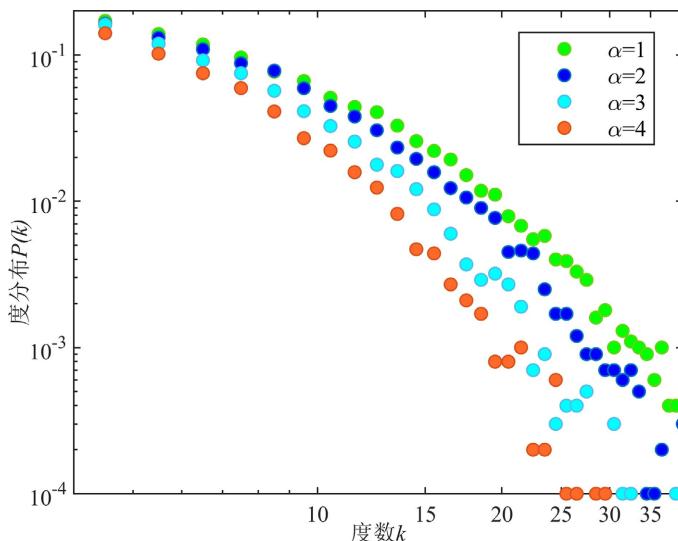


Figure 1. Node degree distribution of Distance-Preference Network
Model in double logarithmic coordinates

图 1. 双对数坐标下距离偏好网络模型的节点度分布

本文利用 MATLAB 对距离偏好网络模型和 BA 无标度网络模型进行了仿真分析。如图 1 所示为双对数坐标下距离偏好网络模型的节点度分布, 图中不同 α 值的度分布函数具有明显弯曲的形状, 与负指数

分布的图像特征相符。而并非为图 2 所示的 BA 无标度网络表现的直线特征, 即幂律度分布。图中结果表明距离偏好网络模型的度分布遵循负指数分布, 文献[15]讨论了 1000 个真实网络数据集的度分布, 结果表明严格无标度网络占比不足 4%, 采用负指数分布拟合得到的结果并不逊色于幂律分布, 甚至可能更为优越。因此, 距离偏好网络模型呈现的负指数度分布在刻画实际网络度分布方面具有可行性, 并具有实际意义。

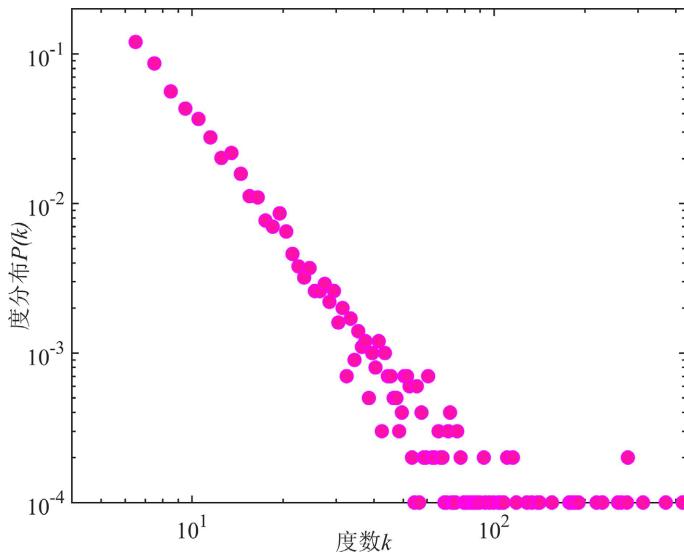


Figure 2. Node degree distribution of BA Model in double logarithmic coordinates

图 2. 双对数坐标下 BA 模型的节点度分布

文献[16]统计分析了 2011 年 5 月的全球 Facebook 朋友关系网络的节点度分布, 其中节点数为 7.21 亿。结果显示全球 Facebook 朋友关系网络的节点度分布在双对数坐标下有很大的曲率, 表现为弯曲的形状, 并非幂律度分布所对应的直线。文献[17]对世界机场网络和电影演员合作网络的度分布进行了统计分析, 从统计结果可以看到这两个网络的度分布曲线在双对数坐标平面上也是弯曲的, 结果表明世界机场网络的度分布衰减呈指数级, 电影演员合作网络的合作次数在 30 到 300 之间的值, 数据符合指数衰减。文献[18]研究了随着站点老化而增长的网络, 从模拟和分析中都发现, 该网络只在 $\alpha < 1$ 时表现出无标度行为, 对于 $\alpha > 1$, 度分布 $P(k)$ 是指数分布。

网络度分布对于网络的演化和特征具有重要影响。对比图 1 和图 2 可以知道, 具有负指数度分布的网络的度数增长速度逐渐减缓, 而幂律度分布的网络则未表现出度数增长速度持续减弱的特征。本文提出距离偏好网络模型的度分布相较于 BA 无标度网络模型的来说更为均匀, 文献[19]研究表明, 由于度分布更均匀, 距离偏好网络模型在应对随机故障和恶意攻击时不仅能够保持 BA 模型的同步鲁棒性, 还能改善同步脆弱性。因此, 距离偏好网络模型在实际背景中具有更广泛的应用前景。

3.2. 聚类系数

复杂网络的聚类系数是指一个节点的邻居节点之间实际存在的连接数与所有可能存在的连接数之比。具体定义如下: 对于一个复杂网络中的某个节点 i , 设其邻居节点数为 k_i , 那么该节点的聚类系数 C_i 定义为该节点的邻居节点之间实际存在的连接数 E_i 与所有可能存在的连接数 $k_i(k_i - 1)/2$ 之比, 即:

$$C_i = \frac{2E_i}{k_i(k_i-1)} \quad (2)$$

其中, E_i 表示节点的邻居节点之间实际存在的连接数。通过计算所有节点的聚类系数并求平均值, 可以得到整个网络的平均聚类系数 $\langle C \rangle$ 为

$$\langle C \rangle = \frac{1}{N} \sum_i C_i \quad (3)$$

其中 N 为网络的总节点数。聚类系数可以用来衡量网络中节点之间的紧密程度, 反映了网络的聚集性和社团结构。

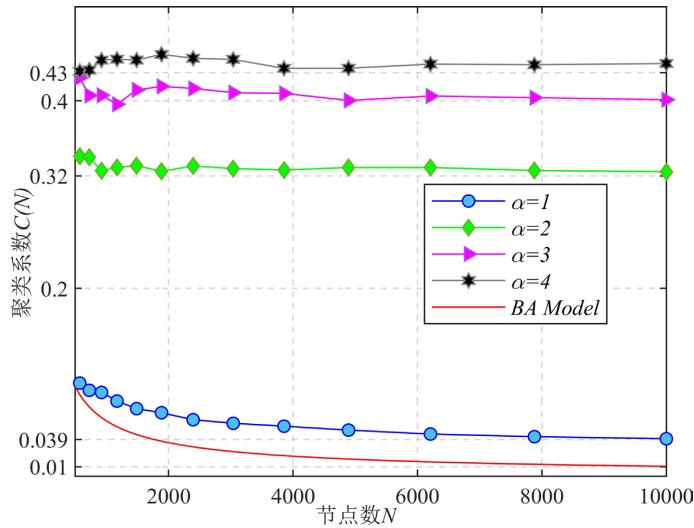


Figure 3. Clustering coefficient of the distance-preference network model and the BA model

图 3. 距离偏好网络模型和 BA 模型的聚类系数

根据图 3 的结果显示, 随着调节参数 α 的增大, 即距离因素的影响程度增加, 网络的聚类系数也相应增大, 这表明距离偏好连接机制对网络的聚类特性具有重要影响。这种现象可以解释为距离偏好机制促进了网络中邻居节点之间的连接, 从而增加了聚类系数。当参数 $\alpha > 1$ 时, 模拟仿真所得网络的聚类系数显著增加, 远远大于 BA 模型的聚类系数, 并且随着参数的增大, 距离偏好网络模型的聚类系数呈现增长的趋势。

值得注意的是, 距离模型的聚类系数增长的幅度随着参数的增大而减小, 在 $1 < \alpha < 2$ 时大幅增加; 在 $2 < \alpha < 4$ 时增加的幅度减小, 这表明网络的聚类系数并非随着参数 α 线性增加。这种非线性关系反映了网络结构中的复杂动态调整过程, 即随着距离因素影响程度的增加, 网络的聚类特性受到多方面影响, 导致聚类系数的增长呈现出一定的饱和效应。这一发现深化了对复杂网络中聚类现象的理解, 并为探索网络演化规律和设计高效网络算法提供了重要参考。

网络的聚类系数是衡量网络中节点形成紧密群集的程度的重要指标。当网络的聚类系数较大时, 表明网络中的节点更倾向于形成紧密的群集或群组, 这意味着节点之间的连接更加密切, 形成了更多的三角形结构。本文提出的速度模型相较于传统的 BA 模型具有更大的聚类系数。这意味着距离偏好网络模型更具有鲁棒性, 同时在距离偏好网络模型中信息传播速度更快, 也更有助于节点之间的合作和协作。因此, 相较于 BA 模型, 距离偏好网络模型在模拟和描述现实世界中众多复杂网络的演化和特征方面更为准确和广泛。这种模型具有潜在的应用前景, 可以为实际问题的解决提供更有效的方法和策略。

4. 结论

本文提出了一个带调节参数的网络演化模型, 其基本机制为距离偏好连接。通过对该网络模型的度分布和聚类系数进行仿真实验, 结果显示网络模型的度分布呈现负指数分布特征, 而且模型得出的聚类系数明显高于 BA 无标度网络模型, 从而改善了 BA 无标度网络模型低聚类的缺陷。而通过对距离依赖程度参数的调节, 能够获得更大的聚类系数, 这表明距离偏好连接机制对网络模型的聚类系数具有关键影响, 这一发现提示距离偏好连接机制可能是现实世界真实网络高聚类现象涌现的基本机制之一, 突出了该模型在解释网络聚类特性方面的重要性。

本文提出的带调节参数的网络演化模型结合距离偏好连接机制在仿真实验中展现出度分布呈负指数分布的特征, 与无标度网络模型不同, 这为无标度网络模型是否普遍的争论提供了新的视角和证据, 表明并非所有网络的度都符合幂律分布, 不同的网络演化机制可能导致不同的度分布形式。未来的研究可以进一步探讨不同网络模型的度分布特征, 以及这些特征背后的演化机制, 从而更全面地理解复杂网络的结构和特性。

此外距离模型采取简单的演化机制在仿真实验中展现出了较好的性能, 能够更好地描述现实世界真实网络的聚类特性。未来的研究可以进一步探索该模型在复杂网络演化和结构形成方面的应用, 同时也可以考虑将该模型应用于实际网络问题中, 以验证其有效性和实用性。这一研究成果为理解和分析复杂网络中的聚类现象提供了新的思路和方法, 具有重要的理论和实践意义。

基金项目

本文工作由国家自然科学基金(No.11502062)支持。

参考文献

- [1] Proulx, S.R., Promislow, D. and Phillips, P.C. (2005) Network Thinking in Ecology and Evolution. *Trends in Ecology & Evolution*, **20**, 345-353. <https://doi.org/10.1016/j.tree.2005.04.004>
- [2] Berahmand, K., Nasiri, E., Mohammadiani, R. and Li, Y. (2021) Spectral Clustering on Protein-Protein Interaction Networks via Constructing Affinity Matrix Using Attributed Graph Embedding. *Computers in Biology and Medicine*, **138**, 104933. <https://doi.org/10.1016/j.combiomed.2021.104933>
- [3] Azaouzi, M. and Romdhane, L.B. (2018) An Efficient Two-Phase Model for Computing Influential Nodes in Social Networks Using Social Actions. *Journal of Computer Science & Technology*, **33**, 286-304. <https://doi.org/10.1007/s11390-018-1820-9>
- [4] Ilany, A., Holekamp, K.E. and Akçay, E. (2021) Rank-Dependent Social Inheritance Determines Social Network Structure in Spotted Hyenas. *Science*, **373**, 348-352. <https://doi.org/10.1126/science.abc1966>
- [5] Newman, M.E.J. (2003) The Structure and Function of Complex Networks. *SIAM Review*, **45**, 167-256. <https://doi.org/10.1137/S003614450342480>
- [6] Watts, D.J. and Strogatz, S.H. (1998) Collective Dynamics of “Small-World” Networks. *Nature*, **393**, 440-442. <https://doi.org/10.1038/30918>
- [7] Barabasi, A.L. and Albert, R. (1999) Emergence of Scaling in Random Networks. *Science*, **286**, 509-512. <https://doi.org/10.1126/science.286.5439.509>
- [8] Fronczak, A., Fronczak, P. and Holyst, J.A. (2003) Mean-Field Theory for Clustering Coefficients in Barabasi-Albert Networks. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, **68**, 046126. <https://doi.org/10.1103/PhysRevE.68.046126>
- [9] Ichinose, G. and Sayama, H. (2017) Invasion of Cooperation in Scale-Free Networks: Accumulated versus Average Payoffs. *Artificial Life*, **23**, 25-33. https://doi.org/10.1162/ARTL_a_00220
- [10] Liu, L. Shen, M. and Tan, C. (2021) Scale Free Is Not Rare in International Trade Networks. *Scientific Reports*, **11**, 13359. <https://doi.org/10.1038/s41598-021-92764-1>
- [11] Khanin, R. and Wit, E. (2006) How Scale-Free Are Biological Networks. *Journal of Computational Biology*, **13**, 810-818. <https://doi.org/10.1089/cmb.2006.13.810>

-
- [12] Stumpf, M.P.H. and Porter, M.A. (2012) Critical Truths about Power Laws. *Science*, **335**, 665-666.
<https://doi.org/10.1126/science.1216142>
 - [13] Tanaka, R. (2005) Scale-Rich Metabolic Networks. *Physical Review Letters*, **94**, 168101.
<https://doi.org/10.1103/PhysRevLett.94.168101>
 - [14] Clite, P. (2020) Are RNA Networks Scale-Free? *Journal of Mathematical Biology*, **80**, 1291-1321.
<https://doi.org/10.1007/s00285-019-01463-z>
 - [15] Broido, A.D. and Clauset, A. (2019) Scale-Free Networks Are Rare. *Nature Communications*, **10**, 1017.
<https://doi.org/10.1038/s41467-019-08746-5>
 - [16] Ugander, J., Karrer, B.L., Backstrom, L. and Marlow, C. (2011) The Anatomy of the Facebook Social Graph. arXiv: 1111.4503v1.
 - [17] Amaral, L.A., Scala, A., Barthelemy, M. and Stanley, H.E. (2000) Classes of Small-World Networks. *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 11149-11152.
<https://doi.org/10.1073/pnas.200327197>
 - [18] Dorogovtsev, S.N. and Mendes, J.F. (2000) Evolution of Networks with Aging of Sites. *Physical Review E-Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, **62**, 1842-1845.
<https://doi.org/10.1103/PhysRevE.62.1842>
 - [19] Li, Z. and Chen, G.R. (2006) Global Synchronization and Asymptotic Stability of Complex Dynamical Networks. *IEEE Transactions on Circuits and Systems II: Express Briefs*, **53**, 28-33. <https://doi.org/10.1109/TCSII.2005.854315>