Methods And Problems Attempt in Scale-Free Models From Complex Networks

Bing Yao[1,b], Xiaomin Wang[1], Jing Su[1], Fei Ma[1], Xiyang Zhao[1], Ming Yao[2,a], Mingjun Zhang[3] and Guoxing Wang[3]

1. College of Mathematics and Statistics, Northwest Normal University, Lanzhou, 730070 CHINA
2.Department of Information Process and Control Engineering, Lanzhou Petrochemical College of Vocational Technology Lanzhou, 730060 CHINA
3. School of Information Engineering Lanzhou University of Finance and Economics Lanzhou, 730030 CHINA

[a]yybm918@163.com, [b]yybb918@163.com

Keywords:power law, scale-free, model, cumulative distribution, algorithm.

Abstract. The dynamical phenomena of complex networks are not easy to model and to characterize by current methods of mathematics. Newman, Barabási and Watts pointed out the direction of researching complex networks by graph theory, which was successfully applied in current investigation of complex networks, and Newman's network-based methods have been applied to a variety of fields. Some well-known and new methods of mathematics are provided in this article, associated with construction and problems on scale-free models from real networks.

Introduction

The emerging field of "network science" is an increasingly popular field. What in this article will go on according to the suggestion proposed by Newman, Barabási and Watts [3]: *Pure graph theory is elegant and deep, but it is not especially relevant to networks arising in the real world. Applied graph theory, as its name suggests, is more concerned with real-world network problems, but its approach is oriented toward design and engineering.* Doubtless, it is one of important directions and guidance of researching networks by means of graph theory. However, Alderson [12] pointed: *Recent efforts to develop a universal view of complex networks have created both excitement and confusion about the way in which knowledge of network structure can be used to understand, control, or design system behavior.* Let $P(k)$ be the *probability distribution* of a vertex joined with k vertices in $N(t)$.

As known, Barabási and Albert [1] have shown $P(k) = \Pr(x = k) \propto k^{-\alpha}$ to be one of standard properties of a complex scale-free network, where α falls in the range $2 < \alpha < 3$. This criterion has been shown in many real networks that have scale-free behavior. These models have *growth* and *preferential attachment*, two important criterions for having scale-free structure. But there is some exception, for example, Mitzenmacher [19] shown that the monkeys typing randomly would produce a power law associated with word frequency requires neither preferential attachment nor optimization.

Newman's network-based methods have been applied to a variety of fields, including psychology, sociology, economics and biology. With collaborators, Newman developed statistical methods for analyzing power-law distributions and applied them to the study of a wide range of systems, in various cases either confirming or denying the existence of previously claimed power-law behaviors (Ref. https://en.wikipedia.org/wiki/Mark-Newman). Using existing methods to depict network models is not a easy thing, even a horrible job. It is very important for us to focus on mathematical methods that were used or are using or will be born or will be discovered in researching networks. Some results on mathematical methods of researching scale-free networks will be shown in this article.

Mathematical methods and models

Networks are connected in the following argument if no special declaration. The notations $n_v(t)$ and $n_e(t)$ are the numbers of vertices and edges of a network $N(t)$. Let $d(u, v)$ stand for the distance between two vertices u, v of a network model $N(t)$, and let $\deg(x)$ be the degree of a vertex x in $N(t)$. Also, $\deg(x)$ is the number of neighbors of the vertex x. For each vertex u of $V(t)$, where $V(t)$ is the vertex set of the model $N(t)$, the set $X(u,t,\beta) = \{v: d(u,v) = \beta, v \in V(t)\}$ is called the β-neighbor set of the vertex u for a fixed integer $\beta \geq 1$. We define a β-distance set $Y_u(t,\beta,k')$ over $X(u,t,\beta)$ by $Y_u(t,\beta,k') = \{v: \deg(v) = k',\ v \in X(u,t,\beta)\} = \{v: d(u,v) = \beta, \deg(v) = k'\}$.

Probability $P(k)$. Barabási *et al.* [18] introduced that there are the continuum theory, the master equation and the rate equation for estimating the degree distribution $P(k)$. Clauset *et al.* [10] have shown that a continuous power-law distribution can be described by a probability density

$$P(k) = \frac{P(k \leq X \leq k + dk)}{dk} = Ck^{-\gamma}, \tag{1}$$

where X is the observed value and $C = (\gamma - 1)x_{\min}^{-\gamma}$ is a normalization constant, since the distribution (1) diverges at zero, so there must be a lower bound $x_{\min} > 0$. In the discrete case, $C^{-1} = \zeta(\gamma, x_{\min})$ from $\sum_{k=x_{\min}}^{\infty} Ck^{-\gamma} = 1$, where Hurwitz zeta function $\zeta(\gamma, x_{\min}) = \sum_{n=0}^{\infty}(n + x_{\min})^{-\gamma}$.

Cumulative distributions. C1. **Cumulative degree distribution.** Dorogovtsev *et al.* [6], in order to obtain exact (analytical) and precise (numerical) answers for main structural and topological characteristics of scale-free graphs, defined the *cumulative degree distribution* as

$$P_{cum}(k) = \sum_{k' \geq k} \frac{N(k',t)}{n_v(t)} \sim k^{1-\gamma}, \tag{2}$$

where $N(k',t)$ represents the number of vertices which the degree greater than k at time step t, and $2 < \gamma = 1 + \frac{\ln 3}{\ln 2} < 3$. $P_{cum}(k)$ is useful in studying scale-free models (Ref. [5, 8]).

C2. **Edge-cumulative distribution.** The authors of the paper [7] motivated from (2) define the *edge-cumulative distribution* by $P_{ecum}(k) = \sum_{k' \geq k} \frac{E(k',t)}{n_e(t)} \sim k^{1-\delta}$, where $E(k',t)$ is the number of edges joined with vertices of degree k' greater than k at time step t. Wang *et al.* [9] conjecture that both $P_{cum}(k)$ and $P_{ecum}(k)$ are equivalent to each other after verifying their conjecture for some deterministic models. We define the *edge-cumulative distribution* $P_{ecum}^{d}(k) = \sum_{k' \geq k} \frac{k' N(k',t)}{n_e(t)} \sim k^{1-\varepsilon}$.

C3. **Mixed cumulative distributions.** Wang *et al.*, in [9], have conjectured that both $P_{cum}(k)$ and $P_{ecum}(k)$ are mutually equivalent in every deterministic scale-free network model. Again, Wang *et al.* define three mixed cumulative distributions with $0 < \tau < t$ as

$$P_{cum}^{1}(k) = \sum_{i \leq \tau} \frac{n_e(i) \cdot n_v(i)}{n_e(t) \cdot n_v(t)}, P_{cum}^{2}(k) = \sum_{i \leq \tau} \frac{n_e(i) - n_v(i)}{n_e(t) - n_v(t)}, P_{cum}^{3}(k) = \sum_{i \leq \tau} \frac{\sqrt{n_e(i) \cdot n_v(i)}}{\sqrt{n_e(t) \cdot n_v(t)}}. \tag{3}$$

and verified them by Comellas' recursive graphs, Sierpinski and Apollonian network models.

C4. **Another group of mixed distributions.**

$$P_{cum}^{\times}(k) = \sum_{k' \geq k} \frac{E(k',t) \cdot N(k',t)}{n_e(t) \cdot n_v(t)}, P_{cum}^{-}(k) = \sum_{k' \geq k} \left| \frac{E(k',t) - N(k',t)}{n_e(t) - n_v(t)} \right|, \tag{4}$$

$$P_{cum}^{\sqrt{\times}}(k) = \sum_{k' \geq k} \sqrt{\frac{E(k',t) \cdot N(k',t)}{n_e(t) \cdot n_v(t)}}, P_{cum}^{*}(k) = \sum_{k' \geq k} \left[\frac{E(k',t) - N(k',t)}{n_e(t) - n_v(t)} \right]^2. \tag{5}$$

Unfortunately, we do not know the physical meanings of the above four cumulative distributions in real networks, and do not estimate them in deterministic and random models.

C5. **Complementary cumulative distribution.** Clauset *et al.* [10] suggested to consider also the *complementary cumulative distribution* function $P_{cum}^{c}(x) = P_{cum}^{c}(x) = \int_{x}^{\infty} P(k)dk = \left(\frac{x}{x_{\min}}\right)^{-\alpha+1}$ of a power-law distributed variable X with $P_{cum}^{c}(x_{\min}) = 1$ and $P(k) = \Pr(X = k) = Ck^{-\alpha}$ with

$C = (\alpha - 1)x_{\min}^{\alpha-1}$, and which for both *continuous* and *discrete* cases is defined as $P^c_{cum}(x) = \Pr(X \geq x)$. Thereby, we have $\frac{\partial P^c_{cum}(k)}{\partial k} = -\frac{\partial}{\partial t}\int_\infty^x P(k)dk = -P(k)$. In many literature we can see

$$P(k) = -\frac{\partial P^c_{cum}(k)}{\partial k} = -\frac{\partial \Pr(X \geq x)}{\partial t} = \frac{\partial(1 - \Pr(X \geq x))}{\partial t} = \frac{\partial \Pr(X < x)}{\partial t} \tag{6}$$

Wang *et al.* [9] estimated $P(k) \sim \frac{\partial P_{cum}(k)}{\partial k}$ in discrete cases of deterministic scale-free network models. In [4], Newman shows the the cumulative distribution function $P(k) = \sum_{k'=k}^{\infty} \Pr(k')$, where $\Pr(k')$ is the probability that the degree is greater than or equal to k. We can compute $P(k_i(t) < k) = 1 - P(k_i(t) \geq k) = 1 - \int_k^{+\infty} P(x)dx = 1 - P_{cum}(k)$, and furthermore $P(k) = \frac{\partial P(k_i(t)<k)}{\partial k} = -\frac{\partial P_{cum}(k)}{\partial k}$, we then, by (1), can derive $P_{cum}(k) = -\int_0^k P(w)dw = -C\int_0^k w^{-\gamma}dw = \frac{C}{\gamma-1}k^{1-\gamma}$.

C6. **The d-cumulative degree distribution.** Su (Jing Su, Bing Yao and Ming Yao. Some Characteristics of A Class of Edge-iteration Network Model. submitted) define $P^d_{cum}(k) = \frac{1}{n_d(t)}\sum_{i=0}^{\tau} n_d(i)$, where $n_d(i)$ is the number of vertices of a fixed degree d in $N(i)$ at time step i, and $0 < \tau < t$.

C7. **Cumulative distribution of the clustering coefficient.** Let $C(u)$ be the clustering coefficient of a vertex u in $N(t)$. Then $N(t)$ has its own *average clustering coefficient* as $\langle c\rangle = \frac{1}{n_v(t)}\sum_{u\in V(t)} C(u)$. Barabási and Dorogovtsev *et al.*, in [18] and [6], defined independently a deterministic model to be *hierarchical* if the clustering coefficient of a vertex u having degree k holds the scaling law $C(k) = \frac{2|E(u)|}{k(k-1)} \propto \frac{1}{k}$ true, where $|E(u)|$ is the number of edges between the k neighbors of u. Dorogovtsev *et al.* [6] formulated the *cumulative distribution of the clustering coefficient* of $N(t)$ as $W_{cum}(\xi) = \frac{1}{n_v(t)}\sum_{\xi'\leq\xi} m_c(\xi', t) \propto \xi^{\gamma-1}$ with $\gamma = 1 + \ln 3/\ln 2$. where ξ and ξ' are points of the discrete spectrum, and $m_c(\xi', t)$ is the number of vertices with clustering coefficient ξ'. Dorogovtsev *et al.* claimed that their results in [6] can be reasonably applied to random growing networks.

Beta-distance parameters. B1. **Beta-distance average degree**

$$\langle k\rangle_\beta = \frac{1}{n_v(t)}\sum_{u\in V(t)} \langle k\rangle_{(u,\beta)} = \frac{1}{n_v(t)}\sum_{u\in V(t)}\sum_{x\in X(u,t,\beta)} \frac{\deg(x)}{|X(u,t,\beta)|}. \tag{7}$$

with $\langle k\rangle_{(u,\beta)} = \frac{1}{|X(u,t,\beta)|}\sum_{x\in X(u,t,\beta)} \deg(x)$. As $\beta = 1$, we have $\langle k\rangle_1 = \langle k\rangle = 2n_e(t)/n_v(t)$.

B2. **Beta-distance cumulative degree distributions.** The universal and local measurements are defined as

$$P^\beta_{cum}(k,u)_u = \frac{1}{n_v(t)}\sum_{k'>k}|Y_u(t,\beta,k')|, P^\beta_{cum}(k,u)_l = \frac{1}{|X(u,t,\beta)|}\sum_{k'>k}|Y_u(t,\beta,k')|. \tag{8}$$

Velocities of dynamic network models. We can compute $\frac{\partial|V(t)|}{\partial t}$ and $\frac{\partial|E(t)|}{\partial t}$ for the vertex set $V(t)$ and the edge set $E(t)$ of a network model $N(t)$. Thereby, we define $V_{el}(N(t)) = \frac{\partial|V(t)|}{\partial t}\cdot\frac{\partial|E(t)|}{\partial t}$ as the velocity of $N(t)$, since we regard that t is continuous. Clearly, $N(t)$ is growing when $V_{el}(N(t)) > 0$, otherwise $N(t)$ is decaying as $V_{el}(N(t)) < 0$. Also, we can compare network models by their velocities. We have found a phenomenon in the following deterministic models: For the recursive graph $= K(q,t)$ discussed in [5], $V_{el}(K(q,t)) = [(q+1)^t\ln(q+1)]^2/q(q+1)$. For the Sierpinski network model $N_S(t)$ investigated in [8], $V_{el}(N_S(t)) = 27(6^t\ln 6)^2/25$. For the High dimensional Apollonian network model $N_A(t)$ shown in [11], $V_{el}(N_A(t)) = [(d+1)^t\ln(d+1)]^2/d(d+1)$. Thereby, we guess: *Every deterministic scale-free network model $N(t)$ has a constant velocity, or has that $V_{el}(N(t)) \propto A\cdot B^t$ with constants $A, B \neq 0$.* On the other hands, we may consider $V^-_{el}(N(t)) = \frac{\partial|E(t)|}{\partial|V(t)|} = \frac{\partial|E(t)|}{\partial t}/\frac{\partial|V(t)|}{\partial t}$, as $V^-_{el}(K(q,t)) = 1$, as $V^-_{el}(N_S(t)) = 3$, as $V^-_{el}(N_A(t)) = (d+1)^{-1}$. Each one of three examples is almost equal to a constant, which shows that three models are *sparse*.

Construction of Models. No doubtless, one need more models to understand and control real networks in the world. Current models in literature can be partitioned into random models and deterministic models. Graph theory is one of useful and powerful tools applied successfully in network science. When constructing models having some properties of $P_{\rm rop}$ for understanding or simulating real networks, we use two generic mechanisms of *growth* and *preferential attachment* formulated by Barabási and Albert [1], where $P_{\text{rop}} = \{$SF,SW,HCC, HT,HM, SS$\}$ is a set with SF=*scale-free character*, SW=*small-wold structure*, HCC=*high clustering coefficient*, HT=*hierarchical topology*, HM=*high modularity*, SS=*self-similarity*. As known, some models exhibit two or more properties of P_{rop}. Some of techniques for constructing network models are listed in the following: **M1.** Some models can be made in the following ways: **M1.1.** The definition of scale-free graphs was introduced by Li *et al.* [2], so it needs more investigation of scale-free graphs, which is a new branch of graph theory. **M1.2.** We can define the *so-called power law (scale-free) graphs* as: A power law graph $G(t)$ for $t \geq 0$ holds: (i) $|G(t)| = 1 + |G(t-1)|$; (ii) a new vertex u is added to $G(t-1)$ and joins with k vertices of $G(t-1)$. **M2.** A growing network model $N(t)$ can be obtained by: **M2.1.** Doing a unique operation O to $N(t-1)$ having smaller numbers of vertices and edges such that each of numbers of vertices and edges of $N(t)$ is greater than that of $N(t-1)$. Bound growing network models were introduced in [7], most of them are nested. Genio *et al.* [13] have proven: *All scale-free networks are sparse.* We, by the above ways, have shown: *For any real number $M > 0$, there exist a scale-free graph $N(t)$ and a number $\beta \geq M$ such that $|E(t)| \propto \beta \cdot n_v(t)$.* **M2.2.** Doing two operations O_1 and O_2 to $N(t-1)$ such that each of numbers of vertices and edges of $N(t)$ is greater than that of $N(t-1)$. But, it may not have $N(t-1) \subset N(t)$. **M3.** No growth of vertices and edges by rewiring edges under probability p, or by some mechanisms of preferential attachments, or by adding a new vertex and then removing an old vertex simultaneously. Some such non-growing models are: the famous BW Models proposed first by Watts and Strogatz [15]; Yang-Chen-Chen Models [17]; Xie-Zhou-Wang Models [16]; Ghoshal-Newman Models [14]. **M4.** We are thinking of the following topics of constructing network models: **M4.1.** For each desired θ with respect to $2 < \theta < 3$, making a scale-free model $N(t)$ such that a vertex of $N(t)$ was joined with k vertices under the probability $P(k) \propto k^{-\theta}$. **M4.2.** For each desired η falling in the range $2 < \eta < 3$, making a scale-free model $N(t)$ having its own cumulative degree distribution $P_{cum}(k) \propto k^{1-\eta}$. **M4.3.** For each desired μ holding $2 < \mu < 3$, making a scale-free model $N(t)$ having its own edge cumulative distribution $P_{ecum}(k) \propto k^{1-\mu}$. **M5.** *Complementary models* of models are not mentioned more in literature in our memory. We consider such models for their structures and parameters in general. **M6.** No more network directed models appeared in literature, although WWW can be described by digraphs. **M7.** It is useful to apply hypergraph theory in construction of models, especially, society networks and WeChat etc.

Summary

Our results are focusing on the following problems: 1. The sizes of real networks around us are too big in general, and they are complex and random for formulating models. 2. Lack of recognized and fixed standards, precise definitions of objects of networks. 3. Need many experts coming from two or more subjects in cooperative research as we are facing various disciplines, large amount of literature. 4. Lack of appropriate approaches in network science, since current methods are directly taken from existing mathematics. Moreover, the new mathematical methods produced from network research are far more from to be an effective theoretical system. 5. Technology and equipments are updated quickly every day. New technology and new equipments change greatly structures of real networks, even set up new networks. The time period of producing new technology and new equipments is shorter than that of researching. 6. The gap between network development and theoretical study is too large. It is very difficult to make synchronous development of investigation and real network technology.

Acknowledgment. We are indebted to Zhongzhi Zhang (School of Computer Science, Fudan University, Shanghai 200433, China) for his valuable and thoughtful suggestions. The author, *Bing Yao*, was supported by the National Natural Science Foundation of China under grants No. 61163054 (*Studying Topologies And Behaviors Of Scale Free Networks By Applying Graphical Structure Analysis Based*

On Spanning Trees) No. 61363060 (*Research Of The Function And Topological Structures Of The Internet Of Things Based On Graph Theory*), No. 61662066 (*Research On The Equivalent Relations Of Complex Network Models And Construction Of High Quality Networks*). The author, *Ming Yao*, was supported by the Special Funds of Finance Department of Gansu Province of China under Grant No. 2014-63.

References

[1] Albert-László Barabási and Reka Albert. Science **286** (1999) 509-512.

[2] Li, L., D. L. Alderson, J. C. Doyle, W. Willinger. 2006. Internet Math. **2**(4) 431-23.

[3] M. E. J. Newman, A.-L. Barabási, and D. J.Watts. Princeton University Press, Princeton (2006).

[4] M. E. J. Newman. SIAM Review, **45** (2003), 167-256.

[5] Francesc Comellas, Guillaume Fertin, Andre Raspaud. Physical Review E **69**, 03710-1, 037104-4 4 (2004)

[6] S.N.Dorogovtsev, A.V.Goltsev, J.F.F.Mendes. Physical reviewer, 2002, **65**, 066122-066125.

[7] Xia Liu, Bing Yao, Wanjia Zhang, Xiang-en Chen, Xinsheng Liu, Ming Yao. The 2014 International Conference on Information and Communication Technologies, 2014, 35-38.

[8] Zhongzhi zhang, Shuigeng Zhou, Lujun Fang, Jihong Guan, Yichao zhang. physical A, 2007, **380**,621.

[9] Xiaomin Wang, Bing Yao and Jin Xu. arXiv:1601.06357v1 [cs.SI] 24 Jan 2016. Xiaomin Wang, Bing Yao, Jin Xu. ArXiv 1512.08657v1 (2015).

[10] Aaron Clauset, Cosma Rohilla Shalizi, M. E. J. Newman. Power-law distributions in empirical data. SIAM Review **51**, 661-703 (2009)

[11] Zhang Zhongzhi, Rong Lili, Comellas Francesc. Physica A, 2006, **364**: 610-618.

[12] David L. Alderson. Operations Research Vol. **56**, No. 5, 2008, 1047-065.

[13] Charo I. Del Genio, Thilo Gross, and Kevin E. Bassler. Physical Review Letters **107**, 178701 (2011)

[14] Gourab Ghoshal, M. E. J. Newman. Eur. Phys. J. B **59**, 75-83 (2007).

[15] D. J. Watts and S. H. Strogatz. Nature, Vol.**393**, 440-442, June 1998.

[16] Yan-Bo Xie, Tao Zhou, Bing-Hong Wang. Physica A **387** (2008) 1683-1688.

[17] Xu-Hua Yang, Guang Chen, Sheng-Yong Chen. Physica A **392** (2013) 2547-2554.

[18] A.-L. Barabási, Z. Dezső, E. Ravasz, S.-H. Yook, and Z. Oltvai. Spain (2002) (AIP, Melville New York, 2003).

[19] Michael Mitzenmacher. Internet Mathematics Vol. 1, No. 2: 226-251, 2003.