MODELS OF NETWORKS WITH PREFERENTIAL ATTACHMENT

N. G. Shcherbakova

Institute of Computational Mathematics and Mathematical Geophysics SB RAS, 630090, Novosibirsk, Russia

Modeling is one of the methods of analyzing organizational principles of complex networks that define their topology and behavior. Traditionally networks with no apparent design principles were described as random graph that were first studied by Paul Erdös and Alfred Rényi [1, 2]. But it is known that the topology of real networks deviates from a random graph and many of them self-organize into scale-free (SF) state. The basic motivation of main theoretical models that come under review in this article is to explain the origin of this scale invariance.

The example of a network with in-degree and out-degree power-law distribution is the citation network observed by Derek de Solla Price in [4, 5]. He discovered that in order the network to have these properties the rate at which a paper gets new citations should be proportional to the number that it already has. He called the process *cumulative advantage*. Now it is known as *preferential attachment* due to [6]. Barabasi-Albert model (BA) incorporates two ingredients: growth and preferential attachment. The algorithm of network evolving starts with m_0 nodes and at each step one new node with $m < m_0$ is added. When choosing nodes to which the new node connects it is assumed that the probability Π that the new node will be connected to node i depends on the degree k_i of node i such that $\Pi\left(k_i\right) = \dfrac{k_i}{\sum_j k_j}$. Analytical studying predicts the emergence of power law degree distribution $P(k) \sim k^3$.

BA model lays the foundation of many extensions and modifications motivated by the reason that it can't explain a behavior of all real-world SF networks. Thus in [15] the model with edges rewire have been introduced. Krapvisky, Redner in [13] have examined the initial attractiveness $\Pi((k_i) = C + k_i)$ and the case where probability of attachment to a node is not linear. In [12, 17, 18] the effect of ageing have been investigated and the extended models have been suggested. In the BB-model [20, 21] parameter fitness of the node accounts for the competitive aspect of link obtaining. These added parameters have an effect on a network behavior and can vary the law of degree distribution.

Key words: complex network analysis, scale-free networks, networks models, preferential attachment.

References

1. Erdös P., Rényi A. On random graphs // Publicationes Mathematicae Debrecen. V. 6. 1959. P. 290–297.

2. Erdös P., Rényi A. On the evolution of random graphs // Bull. Inst. Internat. Statist. V. 38, № 4. 1961. P. 343–347.

3. Watts D. J., Strogatz S. H. Collective dynamics of “small-world” networks // Nature. 1998. V. 393. P. 440–442.

4. Price D. J. de Solla. Networks of Scientific Papers // Science. 1965. V. 149. P. 510–515.

5. Price D. J. de Solla. A general theory of bibliometric and other cumulative advantage processes // J. of the American Society for Information Science. 1976. V. 27(5-5). P. 292–306.

6. Barabási A.-L., Albert R. Emergence of scaling in random networks // Science. 1999. V. 286. P. 509–512.

7. Barabási A.-L., Albert R., Jeong H. Mean-field theory for scale-free random networks // Physica A 272. 1999. P. 173–187.

8. Albert R., Barabási A.-L. Statistical mechanics of complex networks // Rev. Mod. Phys. 2001.V. 47, N. 74. P. 47–97.

9. Dorogovtsev S. N., Mendes J. F. F. Evolution of networks // Advances in Phys. 2002. V. 51, 1079.

10. Dorogovtsev S. N., Mendes J. F. F., Samukhin A. N. Structure of growing network with preferential attachment // Phys. Rev. Lett. 2000. V. 85, 4633.

11. Kullmann L., Kertész J. Preferential growth: exact solution of the time dependent distributions // Phys. Rev. E. 63.051112. 2001.

12. Krapvisky P. L., Redner S., Leyvraz F. Connectivity of growing random networks // Phys. Rev. Lett. V. 85, 4629. 2000.

13. Krapvisky P. L., Redner S. Organization of growing random networks // Phys. Rev. E 63, 066123. 2001.

14. Albert R., Jeong H., Barabási A.-L. Error and attack tolerance of complex networks // Nature. 2000. V. 406. P. 378–482.

15. Albert R., Barabási A.-L. Topology of evolving networks: local events and universality // Phys. Review Letters. 2000. V. 85. P. 5234–5237.

16. Dorogovtsev S. N., Mendes J. F. F. Scaling behavior of developing and decaying networks // Europhes. Lettr. 2000. V. 52, N. 3. P. 33–39.

17. Dorogovtsev S. N., Mendes J. F. F. Evolution networks with aging of sites // Phys. Rev. E62, 1842. 2000.

18. Zhu H., Wang X., Zhu J-Y. The effect of aging on network structure // Phys. Rev. E 68, 056121. 2003.

19. Hajra K. B., Sen P. Phase transitions in an aging network // Phys. Rev. E 70, 056103. 2004.

20. Bianconi G., Barabási A.-L. Bose-Einstein condensation in complex networks // Phys. Rev. Lett. 86, 5632. 2001.

21. Bianconi G., Barabási A.-L. Competition and multiscaling in evolving networks // Europhysics Letters, 54: 436–442, 2001.

22. Ergün G, Rodgers G. J. Growing random network with fitness // arXiv: cond-mat/0103423.

23. Wang D., Song C., Barabási A.-L. Quantifying long-term scientific impact // Science. 2013. V. 342. P. 127–131.

24. Vázquez A. Knowing a network by walking on it: emergence of scaling // arXiv:cond-mat/0006132.

25. McGlohon M., Akoglu L., Faloutsos C. Weighted graphs and disconnected components // PAKDD'10 Proc. of the 14th Pacific-Asia conference on advances in knowledge discovery and data mining. 2010. Vol. P. 2. P. 410–421.

26. Kumar R., Raghavan P., Rajagopalan S., Sivakumar D., Tomkins A., Upfal E. Stochastic models for the Web graph // Proc. of the 41th IEEE symp. on foundations of computer science. 2000. P. 57–65.

27. Leskovec J., Kleinberg J. M., Faloutsos C. Graph evolution: Densification and shrinking diameters // ACM Transactions on Knowledge Discovery from Data (TKDD). 2007. V. 1, iss. 1. Art. 2.

МОДЕЛИ СЕТЕЙ С ПРЕДПОЧТИТЕЛЬНЫМ ПРИСОЕДИНЕНИЕМ

Н. Г. Щербакова

Институт вычислительной математики и математической геофизики СО РАН,
630090, Новосибирск, Россия

УДК 519.177
DOI: 10.24411/2073-0667-2019-00011

В статье представлен краткий обзор механизмов генерации и роста, свойственных реальным комплексным сетям. Основное внимание уделено моделям, порождающим масштабно-инвариантные сети.

Ключевые слова: модели комплексных сетей, масштабно-инвариантные сети, механизм предпочтительного присоединения.

Введение. Моделирование является одним из методов анализа организационных принципов комплексных сетей, определяющих их топологию и функционирование. Модели призваны объяснить процессы, управляющие эволюцией реальных сетей. Базовые идеи были изложены в работах [1, 2] в виде двух *ER*-моделей порождения случайных графов (сетей). Модель $G(n,m)$ из всех графов с n вершинами и m ребрами равновероятно выбирается любой. Согласно *ER*-модели $G(n,p)$ граф порождается. Первоначально имеется n изолированных вершин. Ребро появляется с вероятностью p_{ER} независимо от других ребер. Вероятность того, что вершина будет иметь степень k, следует распределению Пуассона

$$P(k) = \mathrm{e}^{-\lambda}\lambda^{\mathrm{k}}/\mathrm{k}!,$$

где $\lambda = C_{n-1}^{k} p_{ER}^{k} (1-p_{ER})^{n-1-k}$. Обычно термин "случайный граф" без уточнения относится именно к графам или сетям, порожденным согласно одной из этих моделей.

Модель *WS*, представленная в работе [3], предполагает построение случайного графа, начинающееся с регулярной одномерной кольцевой решетки с n вершинами, путем переключения ребер. Первоначально каждая вершина имеет k (четное число) ребер с ближайшими соседями, $k/2$ с каждой стороны, число ребер $kn/2$. Переключаемое ребро выбирается с вероятностью p_{WS} (не допускаются петли и кратные ребра), вершина, к которой происходит переключение, выбирается случайным образом. При $p_{WS} \to 1$ распределение степеней стремится к распределению, свойственному модели *ER* с $p_{ER} = \dfrac{k}{n}$. Сети, построенные согласно *WS*-модели, обладают структурными свойствами "малого мира": небольшое среднее расстояние, с ростом n средняя длина пути между вершинами растет пропорционально $\log n$; большой коэффициент кластеризации.

Модели *ER* и *WS* предполагают, что число вершин зафиксировано изначально, а вероятности p_{ER} и p_{WS} не зависят от степени вершин. Распределение степеней вершин соответствующих графов имеет экспоненциальный хвост. При исследовании реальных сетей выясняется, что они, как правило, имеют другое распределение степеней узлов. Так, в работе [4] показано, что сеть научных публикаций обладает свойством: доля научных публикаций, получивших k цитирований (для больших k), уменьшается как $k^{-\gamma}$, $2 \le \gamma \le 3$. То есть вероятность того, что узел имеет входящую степень k, подчиняется степенному закону:

$$P(k) \sim k^{-\gamma}. \tag{1}$$

Сеть, для которой распределение степеней следует степенному закону (обычно подразумевается, что $2 \le \gamma \le 3$), называемую *безмасштабной* (*scale free*) или *масштабно-инвариантной* (*scale-invariantI*), обозначим как *SF*-сеть. Механизм, приводящий к тому, что распределение цитирований подчиняется степенному закону, назван в работе [5] "*cumulative advantage*" (кумулятивное преимущество). Идея Прайса заключается в том, что в результате роста сети скорость, с которой публикация получает новые цитирования, должна быть пропорциональна числу полученных ранее цитирований. Впоследствии механизм стал рассматриваться как возможное объяснение степенного распределения степеней узлов в различных реальных сетях. Он заложен в основу многих сетевых моделей. Модели случайных сетей, расширяющихся за счет добавления узлов с присущими им связями, в которых вероятность присоединения к узлу зависит только от степени узла k, будем называть *GN*-моделями.

1. Механизм предпочтительного присоединения. В работе [6] приводятся две основные причины, благодаря которым распределение степеней сети будет подчиняться степенному закону: а) сеть постоянно расширяется за счет добавления новых улов; б) новый узел присоединяется к узлам, уже имеющим высокую степень. Это свойство, называемое *preferential attachment* (предпочтительное присоединение), обозначим *PA*. Из работ [6, 7] следует, что правила (а) и (б) должны выполняться совместно. В общем случае вероятность того, что в результате *PA* узел i, имеющий степень k_i, приобретает новое ребро, выражается формулой

$$\Pi\left(k_i\right) = \frac{k_i^{\alpha}}{\sum\limits_j k_j^{\alpha}}. \tag{2}$$

1.1 *Модель Барабаши – Альберт* (*BA*). Предпочтительное присоединение является фундаментом модели, предложенной в работах [6, 8, 9]. Алгоритм построения сети таков. Рост сети начинается с небольшого числа узлов (m_0). На каждом шаге добавляется новый узел с m ($m < m_0$) ребрами, с помощью которых новый узел связывается с m различными уже присутствующими узлами, будем называть их узлами назначения. Предполагается, что присоединение к узлу назначения i, имеющему входящую степень k_i, происходит с вероятностью

$$\Pi\left(k_i\right) = \frac{k_j}{\sum\limits_j k_j}. \tag{3}$$

Зависимость вероятности присоединения от степени носит линейный характер. Заметим, что большинство реальных сетей ориентированные. Обычно интерес представляет входящая степень узла, но в модели *BA* ребра не имеют ориентации, поэтому речь идет о степени. Можно считать, что для модели *BA* исходящая степень любого узла – константа, равная m. После t временных шагов получится сеть с $n = t + m_0$ узлами и mt ребрами. Численное моделирование показывает, что такая сеть эволюционирует к масштабноинвариантному состоянию с вероятностью того, что узел будет иметь степень k, следующей степенному закону (1) со значением $\gamma_{SF} = 3$.

В работе [7] рассмотрены две ограниченные модели A и B, согласно которым сеть развивается либо за счет роста числа узлов, либо за счет механизма PA (3). Так, в модели A рост сети начинается с m_0 узлов и на каждом шаге добавляется новый узел с m ребрами, но узел назначения i выбирается из всех имеющихся с равной вероятностью, так что $\Pi\left(k_i\right) = \dfrac{1}{m_0 + t - 1}$. Показано, что $P(k)$ будет иметь экспоненциальный вид: $P(k) = B\exp(-\beta k)$, $B = e/m$, $\beta = 1/m$. В модели B построение сети начинается с n узлов, ребра отсутствуют. На каждом шаге выбирается случайным образом узел, который присоединяется к узлу i согласно (3). В результате $P(k)$ с течением времени будет иметь вид нормального распределения. То есть для того, чтобы распределение степеней узлов было степенным, важны оба момента: рост числа узлов и линейное PA.

1.2. *Методы исследования модели BA.*

1.2.1. При исследовании динамических свойств модели BA используются несколько теоретических подходов. В работах [6, 7] приведены аналитические выкладки, устанавливающие зависимость степени k_i узла i от времени, основанные на предположении, что k – непрерывная вещественная переменная и скорость ее изменения пропорциональна $\Pi(k)$. Такой подход назван *теорией непрерывности* (*continuum theory*). Скорость, с которой узел i приобретает новые ребра, определяется уравнением вида

$$\frac{dk_i}{dt} = m\Pi\left(k_i\right) = m\frac{k_i}{\sum\limits_{j=1}^{n-1} k_j}. \tag{4}$$

С учетом того, что на каждом шаге добавляется m ребер, знаменатель в (4) имеет вид

$$\sum_j k_j = 2mt - m.$$

Таким образом,

$$\frac{dk_i}{dt} = \frac{k_i}{2t.} \tag{5}$$

Каждый узел i при появлении в сети в момент t_i имеет степень $k_i(t_i) = m$, поэтому решение уравнения (5) имеет вид

$$k_i\left(t\right) = m\left(\frac{t}{t_i}\right)^{\beta}, \quad \beta = 1/2. \tag{6}$$

Равенство (6) указывает, что степени всех узлов изменяются одинаковым образом, следуя степенному закону. Дальнейшие выкладки показывают, что функция распределения степеней представляется равенством

$$P\left(k\right) = \frac{dP(k_i(t) < k)}{dk} = \frac{2m^{1/\beta}t}{m_0 + t} \times \frac{1}{k^{1/\beta + 1}}.$$

При $t \to \infty$ вероятность того, что узел будет иметь степень k, асимптотически равна:

$$P\left(k\right)\sim 2m^{1/\beta}k^{-\gamma},\quad \gamma=1/\beta+1=3. \tag{7}$$

Параметры β и γ не зависят от m.

Таким образом, несмотря на непрерывный рост, сеть достигает стационарного состояния, а коэффициент C в степенном распределении $P\left(k\right)\sim Ck^{-}\gamma$ пропорционален квадрату средней связности узлов сети, $C\sim m^2$.

1.2.2. В работе [10] предложен метод исследования GN-моделей на основе анализа вероятности $p(k,t_i,t)$ того, что узел i, появляющийся в сети в момент t_i, ко времени t будет иметь степень k. Так, для модели BA с учетом того, что при подключении нового узла с m ребрами степень уже имеющегося узла увеличивается на единицу с вероятностью $m\Pi(k)=k/2t$ (4, 5) или останется прежней, *основное уравнение* (*master equation*), определяющее $p(k,t_i,t)$, имеет вид

$$p\left(k,t_i,t+1\right)=\frac{k-1}{2t}p\left(k-1,t_i,t\right)+\left(1-\frac{k}{2t}\right)p\left(k,t_i,t\right). \tag{8}$$

Распределение степеней узлов сети:

$$P(k)=\lim_{t\,\to\,\infty}\left(\sum_{t_i}p(k,t_i,t)\right)\Big/t,$$

$$P\left(k\right)=\begin{cases}\dfrac{k-1}{k+2}P\left(k-1\right), & k\geq m+1,\\ \dfrac{2}{(m+2)}, & k=m.\end{cases} \tag{9}$$

В результате $P(k)$ оценивается как

$$P\left(k\right)=\frac{2m(m+1)}{k(k+1)(k+2)},$$

т. е. близко к (7).

Подобный метод используется также в работе [11], где рассматривается модель роста системы, состоящей из нескольких групп элементов. На каждом шаге новый элемент добавляется к некоторой группе. Вероятность p, с которой новый элемент добавляется к группе i, пропорциональна размеру N группы (k_i/N). С вероятностью $q=1-p$ новый элемент образует новую группу. Процесс описывается равенством

$$P_i(k,t)=p\frac{k-1}{t-1}P_i(k-1,\,t-1)+p\left(1-\frac{k}{t-1}\right)P_i(k,t-1)+$$
$$+(1-p)P_i(k,t-1)+(1-p)\Pi_{i-1}(t-1)\delta_{k,1}\left(1-\delta_{i1}\right),$$

где $P_i(k,t)$ – вероятность того, что группа i в момент t будет содержать k элементов, $\Pi_i(t)$ – вероятность того, что в момент t в системе будет i групп (здесь и далее δ – дельта Кронекера).

При этом

$$\Pi_i(t) = C_{t-1}^{i-1} p^{t-1-(i-1)} (1-p)^{i-1}.$$

Для всей системы распределение вычисляется как среднее распределений размеров групп:

$$P(k,t) = \frac{1}{t} \sum_{i=1}^{t} P_i(k,t).$$

При достаточно большом t вероятность стремится к стационарной величине

$$P(k) = \lim_{t\to\infty} P(k,\, t).$$

Зависимость этой величины от p при достаточно больших k и t оценивается как

$$P(k) \sim k^{-1-(1/p)},$$

а вероятность $P_i(k,t)$ при росте числа групп оценивается как

$$\lim_{t,\ i\ \to\infty} P_i(k,t) = \left(\frac{i}{t}\right)^p.$$

1.2.3. В работе [12] авторы исследуют функцию $N_k(t)$ распределения среднего числа узлов, имеющих на момент времени t степень k. В рамках модели BA при появлении в сети нового узла $N_k(t)$ изменяется как

$$\frac{dN_k}{dt} = m\frac{(k-1)N_{k-1}(t) - kN_k(t)}{\sum\limits_{k\geq 1} kN_k(t)} + \delta_{km}. \tag{10}$$

Это так называемое *уравнение скорости* (*rate equation*). В (10) учитывается, что новый узел может присоединиться к узлу со степенью $k-1$, в результате чего степень станет равной k, или присоединиться к узлу со степенью k, уменьшив число узлов со степенью k, кроме того, добавляемый узел может иметь $m = k$. Асимптотические пределы $N_k(t) \sim tP(k)$, $\sum\limits_k kN_k(t) \sim 2mt$, что приводит к рекурсивному равенству (9).

Рассмотренные методы предсказывают одни и те же асимптотические результаты, и любой из них может использоваться для исследования динамики распределения степеней узлов сети в рамках модели.

1.3. *Свойства модели BA. SF*-сети, порождаемые согласно модели BA, обладают рядом свойств, отличающих их от случайных графов с той же средней степенью вершин и с тем же размером.

а) Признаки “малого мира” (см. [8]): с ростом n средняя длина пути SF-сети растет приблизительно как $\log n$; коэффициент кластеризации с ростом сети уменьшается приблизительно следуя степенному закону, а у случайных графов как $<k>\ /n$ ($<k>$ – среднее значение степени).

б) Корреляция степеней соседних узлов. В работе [13] показано, что если n_k – число узлов степени k, n_{kl} – число непосредственно связных узлов со степенями k и l, то $n_{kl} \neq n_k n_l$. Более вероятно, что узлы с подобными степенями спонтанно соединяются между собой. Если ребра подключаются случайным образом, то число связных узлов степени k и l, $n_{kl} = n_k n_l$.

в) Специфика изменения размеров компонент с ростом сети. *Входящая компонента* узла x – это множество всех узлов, из которых существует ориентированный путь в узел x; *исходящая компонента* – это множество всех узлов, которые достижимы из x. В работе [13] проведено исследование изменения размеров компонент с ростом сети. Для модели BA с $m = 1$ исходящая компонента – это узлы единственного ориентированного пути, начинающегося в x. Число $I_s(t)$ входящих компонент с s узлами удовлетворяет уравнению скорости:

$$\frac{dI_s}{dt} = \frac{(s-1)I_{s-1} - sI_s}{t} + \delta_{s1}$$

Показано, что при линейной скорости присоединения I_s подчиняется степенному закону: $I_s \propto s^{-2}$. При изучении исходящих компонент исследовался размер “поколений”. Так, поколение 0 содержит единственный начальный узел. Поколение 1 – это присоединенные к нему узлы вне зависимости от того, когда это произошло, и т. д. Изучается $L_g(t)$ – размер поколения g в момент t, показано что для некоторого фиксированного (большого) значения t имеет место $L_{\max} \simeq \dfrac{t}{\sqrt{2\pi \ln t}}$. Число “поколений” растет логарифмически с ростом числа узлов при условии линейной зависимости предпочтительного присоединения от степени узлов.

г) Устойчивость к разрушению. В работе [14] показано, что SF-сети более устойчивы к выходу из строя случайных узлов (диаметр практически не меняется), чем сети с экспоненциальным хвостом распределения степеней узлов, в то же время они более уязвимы к атакам, заключающимся в выходе из строя важных узлов, плотно связных с остальными.

2. Обобщения модели *BA*. Исследования реальных сетей показывают, что схема развития даже SF-сетей не исчерпывается моделью BA. Так, модель прогнозирует экспоненту $\gamma = 3$ (см. (1)), в то время как для реальных сетей она варьируется от 3 до 5. Кроме того, реальные сети имеют отклонение от степенного закона. Сеть может иметь локальные свойства, от которых зависит тенденция роста, например возникновение локальных связей (ребер между ранее добавленными узлами); переключение ребер; исчезновение ребер. Поэтому появилось значительное число работ, посвященных расширению модели BA и другим моделям развития сетей.

2.1. *Нелинейное предпочтительное присоединение.* Для нелинейного предпочтительного присоединения вероятность того, что узел i, имеющий степень k_i, получит присоединение со стороны нового узла, определяется формулой (2).

Такие модели (для случая $m = 1$) рассматриваются в работе [12]. Анализ свойств моделей производится на основе уравнения скорости, подобного (10), но уравнение имеет вид

$$\frac{dN_k}{dt} = \frac{1}{M_\alpha}\big[(k-1)^\alpha N_{k-1}(t) - k^\alpha N_k(t)\big] + \delta_{k1}, \tag{11}$$

где $M_\alpha = \sum k^\alpha N_k(t)$. Показано, что от значения α существенно зависит вид функции распределения степеней узлов: для $\alpha < 1$ распределение степеней относится к растянутому экспоненциальному (степенной закон в экспоненте $P(k) \sim e^{k^{-\gamma}}$), в то время как для $\alpha > 1$ ожидается феномен “застывания”, когда один узел присоединяется почти ко всем другим. Если $\alpha = 1$, то это модель BA.

2.2. *Асимптотическая зависимость от степени.* В работе [13] рассмотрено несколько моделей GN-сетей. Вероятность того, что новый узел присоединится к узлу назначения, имеющему степень k, при условии, что вероятность зависит исключительно от k, обозначена A_k. В дополнение к строгому равенству $A_k = k^\alpha$ рассмотрена модель, для которой характерна асимптотическая зависимость: $A_k \sim k^\alpha$, $k \to \infty$. Показано, что если $\alpha = 1$, то $N_k \sim k^{-\gamma}$, но γ может принимать значения в диапазоне $2 < \gamma < \infty$.

2.3. *Модели с переключением ребер.* В работе [15] рассмотрена модель, в которой развитие сети, начинающейся с m_0 изолированных узлов, допускает три возможности:

1) С вероятностью p добавляется m новых ребер между уже существующими узлами. При этом в качестве начальной точки ребра произвольно выбирается узел, а второй конец – согласно предпочтительному присоединению с вероятностью

$$\Pi(k_i) = \frac{k_i + 1}{\sum\limits_j (k_j + 1)}. \tag{12}$$

Процесс повторяется m раз.

2) С вероятностью q переключаются m ребер. Сначала произвольно выбирается узел i и ребро (i, j), оно удаляется и образуется новое ребро (i, l). При этом l выбирается с вероятностью $\Pi(k_l)$, согласно (12). Процесс повторяется m раз.

3) С вероятностью $1 - p - q$ добавляется новый узел, имеющий m новых ребер и подключающийся к существующему узлу i с вероятностью $\Pi(k_i)$.

Для этой модели справедлива обобщенная форма степенного закона

$$P(k) \propto (k + \kappa(p, q, m))^{-\gamma(p,q,m)},$$

где $\kappa(p, q, m) = (p - q)\left(\dfrac{2m(1 - q)}{1 - p - q} + 1\right) + 1$; $\gamma(p, q, m) = \dfrac{2m(1 - q) + 1 - p - q}{m} + 1$.

Данная модель предполагает существование двух режимов: безмасштабного и экспоненциального, в зависимости от параметров p, q, m. Подробности см. в работе [15].

Модели с переключением ребер, названные *GNR*-моделями, представлены также в работе [13]. Так, рассматривается модель развития сети GNR_1, когда в каждый момент времени добавляется новый узел l и единообразно из существующих выбирается узел назначения x $(m = 1)$. С вероятностью $1 - r$ создается ребро от l до x, а с вероятностью r ребро направляется от узла l на узел y, являющийся предшественником $x (x \to y)$. Здесь уравнение скорости (ср. (10), (11)) имеет вид:

$$\frac{dN_k}{dt} = \delta_{k1} + \frac{1 - r}{M_0}[N_{k-1} - N_k] + \frac{r}{M_0}\Big[(k - 2)N_{k-1} - (k - 1)N_k\Big]. \tag{13}$$

Если $r = 0$, то уравнение скорости (13) для GNR_1 имеет тот же вид, что и для *GN*.

В работе [13] показано, что рассматриваемый случай *GNR*-модели можно свести к *GN*. Замечено, что предпочтительное присоединение модели *BA* можно рассматривать как линейную зависимость со сдвигом $A_k = k + w$ и перейти к рассмотрению ориентированной сети. Степень k в ориентированном графе – это сумма входящей степени g и исходящей степени h. В данном случае исходящая степень $h = 1$, $k = g + 1$, зависимость от степени определяется как

$$A_k = a(k - 1) + b = k + w, \quad \text{где} \quad w = -1 + b/a; \quad -1 < w < \infty.$$

Степенной закон для распределения $P(k)$ сохраняется, так как при $k \to \infty$ влияние константы сдвига уменьшается. GNR_1 можно свести к GN со сдвигом:

$$A_k = (k-1)r + 1 - r = k + w, \quad w = (1-r)/r - 1 = \frac{1}{r} - 2.$$

2.4. *Добавление и удаление ребер.* Класс моделей развития неориентированных сетей с добавлением и удалением ребер между старыми узлами рассмотрен в работе [16]. Предполагается, что сеть развивается по следующим правилам: в каждый момент времени добавляется новый узел с одной связью, вероятность его присоединения к имеющемуся узлу следует из (3); вместе с этим может добавиться c ребер ($c \geq 0$), вероятность появления ребра между несвязными узлами i и j пропорциональна произведению их степеней $k_i k_j$ (развивающаяся сеть); или могут исчезнуть равновероятно $|c|$ связей ($c \leq 0$, распадающаяся сеть). Независимо от знака параметра c в пределе степень узла соответствует (6) с динамическим параметром $\beta = \dfrac{1+2c}{2(1+c)}$, распределение степеней $P(k)$ соответствует (7) с параметром $\gamma = 2 + \dfrac{1}{1+2c}$. Для развивающейся сети: если $c \to \infty$, то $\beta \to 1$, $\gamma \to 2$, обе величины $<k_i(t)>$ и $P(k)$ демонстрируют следование степенному закону. Для распадающейся сети только $<k_i(t)>$ соответствует степенному закону на всем диапазоне $-1 < c < 0$, а распределение $P(k)$ соответствует степенному закону только для значения c близкого к нулю. То есть постоянное удаление ребер приводит к более значительным изменениям структуры сети, чем добавление.

2.5. *Эффект возраста.* При изучении реальных сетей цитирования выясняется, что интерес к выдающимся работам со временем не снижается, в то время как в основном значимость работ постепенно утрачивается, так как представленные в них идеи уже внедрены и расширены в других работах. В моделях с линейным предпочтительным присоединением, учитывающих возраст узлов, вероятность присоединения к узлу i можно представить формулой

$$\Pi(k_i,\, t_i) \propto \; k_i f(t_i), \tag{14}$$

где t_i – время появления в сети узла i. Определение функции от времени $f(t_i)$ зависит от модели: если $f(t_i) = 1$ для любого i, то это модель BA. В результате вид функции $f(t_i)$ влияет на распределение степеней $P(k)$ в сети, развивающейся соответственно модели.

В работе [17] аналитические выкладки относятся к модели, в которой в каждый момент времени присоединяется один узел согласно (14), возрастное затухание определяется как $f(t_i) = (t - t_i)$. Показано, что в случае, когда параметр $\beta < 1$ (см. (6)), сохраняется степенной закон (1). Причем если β стремится от $-\infty$ к 0, то значение γ стремится от 2 к 3, т. е., приближается к распределению, свойственному модели без учета возраста. Дальнейшее возрастание β от 0 к 1 сопровождается стремлением γ к бесконечности. При $\beta > 1$ распределение приближается к экспоненциальному.

Влияние возраста исследуется также в работе [13]. Узел имеет возраст a означает, что он добавлен в момент $t_i = t - a$. Среднее число узлов возраста a, имеющих входящую степень $k-1$ в момент t, обозначим как $c_k(t,a)$. Тогда $N_k(t) = \int\limits_0^t da\, c_k(t,a)$. Совместное распределение эволюционирует согласно формуле

$$\frac{A_{k-1}c_{k-1} - A_k c_k}{A(t)} + \delta_{k1}\delta(a),$$

где $A(t) = \sum\limits_{j\ \geq 1} A_j N_j(t)$. Для $A_k = k$ показано, что распределение степеней узлов фиксированного возраста имеет экспоненциальное затухание и при $a \to t$ средняя степень $<k> \sim (1 - a/t)^{-1/2}$. "Молодые" узлы ($a/t \to 0$) имеют небольшую степень, а "старые" – большую. Слабое затухание распределения степеней старых узлов приводит к степенному закону.

В работе [18] рассматривается модель, в которой вероятность присоединения определяется формулой (14), а функция $f(t) = \exp(\alpha t)$. Показано, что сеть не будет SF-сетью, если $\alpha < 0$.

Модель развития сети с добавлением в каждый момент времени одного узла ($m = 1$) рассмотрена в работе [19]. Вероятность присоединения определяется формулой $\Pi(k_i, t_i) \propto k_i^\beta t_i^\alpha$. Влияние параметров α и β на структуру сети представлено в виде диаграммы на α–β плоскости. Пространство делится на две фазы: малый мир и регулярные сети. Безмасштабное поведение наблюдается вдоль прямой линии для $\beta \geq 1$ в рамках фазы малого мира. В фазе малого мира при $\beta \gg 1$ наблюдается тенденция узлов подключаться к одному и тому же узлу; для $\beta < 1$, $\alpha \leq -1$ распределение степеней приближается к растянутому экспоненциальному.

3. Модель Бианкони – Барабаши (*BB*). Модель Бианкони – Барабаши (*BB*), представленная в работах [20, 21], является расширением модели *BA*. Предполагаются рост и предпочтительное присоединение, но предпочтительное присоединение помимо степени k зависит еще от одного параметра – *качества* (*fitness*) η, характеризующего скрытые свойства узла, обеспечивающие получение дополнительных связей. Например, это может быть оценка узла со стороны сообщества. В каждый момент времени к системе добавляется новый узел j с соответствующим значением η_j, выбираемым из распределения $\rho(\eta)$. Вероятность того, что новый узел присоединится к узлу i, определяется формулой

$$\Pi_i = \frac{\eta_i k_i}{\sum\limits_j \eta_j k_j}. \tag{15}$$

Параметр η_i, участвующий в формуле (15), называют *мультипликативным.* Согласно теории непрерывности (см. (4)), степень узла изменяется как

$$\frac{dk_i}{dt} = m\frac{\eta_i k_i}{\sum\limits_j \eta_j k_j}. \tag{16}$$

Если все узлы будут иметь одинаковое качество, то (16) сводится к (4) для модели *BA*, т. е. $k_i(t) \sim t^{1/2}$. Допустим, временная эволюция степени узла k_i также следует степенному закону, но с экспонентой $\beta(\eta_i)$:

$$k(t, t_i, \eta_i) = m\left(\frac{t}{t_i}\right)^{\beta(\eta_i)}. \tag{17}$$

Согласно (16), (17), динамическую экспоненту можно представить в виде

$$\beta(\eta) = \frac{\eta}{C}, \tag{18}$$

где

$$C = \int \rho(\eta) \frac{\eta}{1-\beta(\eta)} d\eta.$$

Таким образом, более молодые узлы могут приобретать ребра несмотря на небольшое значение степени, а старые узлы с большим значением параметра качество могут играть центральную роль. В модели BA степень со временем росла как квадратный корень от времени, экспонента $\beta = 1/2$ (см. (6)). В модели BB, в предположении, что эволюция k_i следует степенному закону, экспонента, согласно (18), пропорциональна η_i, т. е. уникальна для каждого узла.

С применением теории непрерывности (см. (3)–(6)) получено, что распределение степеней имеет вид

$$P(k) \sim \int d\eta \rho(\eta) \frac{C}{\eta} \left(\frac{m}{k}\right)^{\frac{C}{\eta}+1}, \tag{19}$$

т. е. является взвешенной суммой степенных законов. Согласно (19), $P(k)$ зависит от вида распределения параметра качество $\rho(\eta)$.

Модель получила дальнейшее развитие. Так, в работе [22] рассматриваются модели с предпочтительным присоединением, модифицированным с помощью *аддитивного* параметра качества η:

$$\Pi_i = (k_i - 1) + \eta_i$$

и *аддитивно-мультипликативного* качества:

$$\Pi_i = \eta_i (k_i - 1) + \zeta_i.$$

Показано, что в первом случае экспонента γ в равенстве (1) зависит от среднего значения параметра качество, а во втором случае – от всего множества значений параметра для узлов сети. Кроме того, рассмотрена модель ориентированной сети, в которой входящие и исходящие ребра имеют разные значения мультипликативного качества.

4. Комбинированная модель роста сети. В работе [23] представлена комбинированная модель, учитывающая одновременно предпочтительное присоединение, эффект возраста и качество применительно к сети цитирования публикаций.

Вероятность того, что публикация i в момент времени t будет процитирована, определяется как

$$\Pi_i(t) \sim \eta_i c_i^t \, P_i(t). \tag{20}$$

Здесь c_i^t – общее число цитирований к моменту t (предпочтительное присоединение), η_i – качество. Затухание влияния, связанного с возрастом, $Pi(t)$, определено как

$$P_i(t) = \frac{1}{\sqrt{2\pi} t \sigma_i} e^{-\frac{(\ln t - \mu_i)^2}{2\sigma_i^2}},$$

μ_i – среднее, отражает актуальность, σ_i – стандартное отклонение, отражает продолжительность. То есть затухание имеет логнормальный вид. Используя уравнение (8), на основе определения (20), можно прогнозировать рост цитирований в зависимости от времени:

$$c_i^t = m\left(e^{\frac{\beta\eta_i}{A}\Phi\left(\frac{\ln t-\mu_i}{\sigma_i}\right)}-1\right), \tag{21}$$

где $\Phi(x)=\dfrac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{x}e^{-y^2/2}\,dy$; m – среднее число ссылок в публикациях; β отражает рост общего числа публикаций; A – нормализующая константа. Параметры m, β, A являются глобальными, их значение одинаково для всех публикаций. Величина $\eta_i' \equiv \eta_i\beta/A$ – относительное качество публикации i.

Уравнение (21) охватывает все три механизма, влияющие на историю цитирования публикаций. Если определить отмасштабированные переменные как

$$\tilde t \equiv (\ln t-\mu_i)/\sigma_i;\quad \tilde c \equiv \ln\left(1+c_i^t/m\right)/\eta_i,$$

получаем

$$\tilde c \;=\; \Phi\left(\tilde t\right). \tag{22}$$

Равенство (22) – основной результат работы [23], прогнозирующий, что история цитирования каждой публикации следует одной и той же универсальной формуле, только отмасштабированной согласно индивидуальным c_i', μ_i, σ_i.

5. Модели с неявным *PA*. Свойство хорошо связных узлов успешно приобретать новые связи наблюдается во многих реальных системах. В модели BA этот принцип заложен в алгоритм роста сети. Однако существуют модели сетей, в которых ребра между узлами организуются согласно иным законам, но в результате роста сети становятся SF-сетями. Так, в работе [24] предложена модель ориентированной сети, алгоритм построения которой основан на случайном блуждании. Начальное состояние – один узел без ребер. Эволюция состоит из двух этапов: добавление нового узла, который присоединяется к одному случайно выбранному узлу; блуждание по узлам, которые достижимы из первого выбранного, и присоединение к ним с вероятностью p, пока это возможно. Затем повторяется добавление узла. То есть этап добавления узла отделен от этапа построения ребер. Для этой модели, независимо от значения p, половина узлов имеет входящую степень, равную нулю. Показано, что в случае $p=0$ распределение входящих степеней имеет экспоненциальный вид: $P(k)=2^{-(k+1)}$. Для $p=1$ входящая степень рассматриваемого узла увеличивается при добавлении нового, если новый присоединится к нему сразу или присоединится к другому узлу, который имеет ребро с рассматриваемым. Таким образом, узлы с большой входящей степенью имеют преимущество в приобретении ребер, т. е., имеет место предпочтительное присоединение вида $\Pi(k)=(k+1)/n$. Хвост распределения входящих степеней имеет вид степенного закона $P(k)\sim k^{-2}$. Параметр $\gamma\simeq 2$ в отличие от значения $\gamma\simeq 3$ для модели BA. Так как значения 0 и 1 для вероятности p определяют качественно разное поведение сети, предполагается, что должно быть пороговое значение p_C, такое что для $p> p_C$ хвост распределения начинает подчиняться степенному закону. Экспериментально установлено, что $p_C\simeq 0{,}4$. Мотивацией являлось моделирование сети цитирования, однако для реальных сетей цитирования параметр $\gamma\simeq 3$.

Модель порождения сети на основе случайного блуждания (“*butterfly*”) представлена также в работе [25]. Модель имеет три параметра: *phost* – вероятность, с которой новый узел выбирает имеющийся узел (“*host*”); *plink* – вероятность образования ребра между двумя узлами; p_{step} – определяет, переходить ли к соседнему узлу, который будет выступать в качестве нового "*host*". В представленной модели в каждый момент времени добавляется один узел. Параметры p_{host} и p_{link} одинаковы для всех новых узлов, а p_{step} – индивидуальный.

Модели эволюции сети с копированием ссылок, предложенные в работе [26], мотивированы на выявление принципов роста веб-графов. Для модели с линейным ростом в каждый момент времени t добавляется один узел, имеющий исходящую степень, равную константе. Узел назначения для ориентированных ребер выбирается следующим образом. Из числа имеющихся узлов выбирается "прототип". Затем i-е ребро с вероятностью p присоединяется к любому узлу и с вероятностью $1-p$ присоединяется к узлу назначения i-го ребра "прототипа". В результате вероятность того, что узел с входящей степенью k получит новое ребро, пропорциональна $(1-p)k$, т. е. наблюдается линейное предпочтительное присоединение.

Частично метод копирования ссылок прототипа используется в модели *FF* (*forest fire*), представленной в работе [27]. Определяются два параметра: p и r. Процесс добавления нового узла v состоит из трех этапов:

а) Случайным унифицированным образом выбирается "прототип" w и организуется исходящая дуга к w.

б) Генерируются два случайных числа x и y геометрически распределенные со средними значениями $p/(1-p)$ и $rp/(1-rp)$ соответственно. Новый узел v выбирает x исходящих дуг и y входящих дуг узла w. Пусть $w_1,w_2,\ldots,w_{x+y}$ – концы этих дуг.

в) Узел v образует исходящие дуги к $w_1,w_2,\ldots,w_{x+y}$.

Этап (б) повторяется для каждого узла $w_1,w_2,\ldots,w_{x+y}$, при этом ни один узел не выбирается дважды. В сети, развивающейся согласно модели *FF*, при присоединении нового узла предпочтение имеют хорошо связные узлы, а хвосты распределений входящих и исходящих степеней подчиняются степенному закону.

Заключение. Для изучения реальных сетей в области науки и социальных отношений разрабатываются модели, способные демонстрировать основные свойства этих сетей. Здесь рассмотрены модели сетей, в процессе роста которых вероятность присоединения нового узла к присутствующему в основном зависит от степени k узла назначения. В модели *BA* зависимость от степени узла назначения носит линейный характер. Рост и линейное предпочтительное присоединение являются основными механизмами, ответственными за то, что хвост распределения степеней узлов сети, развивающихся согласно модели, для достаточно больших k подчиняется степенному закону. Модель *BA* получила развитие за счет введения дополнительных мотиваций, влияющих на предпочтительное присоединение, таких как "возраст" узла назначения, первоначальная "привлекательность" или "качество". От вида функций, определяющих эти дополнения, зависит, останется ли сеть безмасштабной или распределение степеней будет иметь экспоненциальный вид. Масштабная инвариантность при определенных условиях свойственна также рассмотренным здесь моделям, допускающим добавление, переключение и удаление ребер между существующими узлами.

Список литературы

1. Erdős P., Rényi A. On random graphs // Publicationes Mathematicae Debrecen. V. 6. 1959. P. 290–297.

2. Erdős P., Rényi A. On the evolution of random graphs // Bull. Inst. Internat. Statist. V. 38, № 4. 1961. P. 343–347.

3. Watts D. J., Strogatz S. H. Collective dynamics of “small-world” networks // Nature. 1998. V. 393. P. 440–442.

4. Price D. J. de Solla. Networks of Scientific Papers // Science. 1965. V. 149. P. 510–515.

5. Price D. J. de Solla. A general theory of bibliometric and other cumulative advantage processes // J. of the American Society for Information Science. 1976. V. 27(5-5). P. 292–306.

6. Barabási A.-L., Albert R. Emergence of scaling in random networks // Science. 1999. V. 286. P. 509–512.

7. Barabási A.-L., Albert R., Jeong H. Mean-field theory for scale-free random networks // Physica A 272. 1999. P. 173–187.

8. Albert R., Barabási A.-L. Statistical mechanics of complex networks // Rev. Mod. Phys. 2001.V. 47, N. 74. P. 47–97.

9. Dorogovtsev S. N., Mendes J. F. F. Evolution of networks // Advances in Phys. 2002. V. 51, 1079.

10. Dorogovtsev S. N., Mendes J. F. F., Samukhin A. N. Structure of growing network with preferential attachment // Phys. Rev. Lett. 2000. V. 85, 4633.

11. Kullmann L., Kertész J. Preferential growth: exact solution of the time dependent distributions // Phys. Rev. E. 63.051112. 2001.

12. Krapvisky P. L., Redner S., Leyvraz F. Connectivity of growing random networks // Phys. Rev. Lett. V. 85, 4629. 2000.

13. Krapvisky P. L., Redner S. Organization of growing random networks // Phys. Rev. E 63, 066123. 2001.

14. Albert R., Jeong H., Barabási A.-L. Error and attack tolerance of complex networks // Nature. 2000. V. 406. P. 378–482.

15. Albert R., Barabási A.-L. Topology of evolving networks: local events and universality // Phys. Review Letters. 2000. V. 85. P. 5234–5237.

16. Dorogovtsev S. N., Mendes J. F. F. Scaling behavior of developing and decaying networks // Europhes. Lettr. 2000. V. 52, N. 3. P. 33–39.

17. Dorogovtsev S. N., Mendes J. F. F. Evolution networks with aging of sites // Phys. Rev. E62, 1842. 2000.

18. Zhu H., Wang X., Zhu J-Y. The effect of aging on network structure // Phys. Rev. E 68, 056121. 2003.

19. Hajra K. B., Sen P. Phase transitions in an aging network // Phys. Rev. E 70, 056103. 2004.

20. Bianconi G., Barabási A.-L. Bose-Einstein condensation in complex networks // Phys. Rev. Lett. 86, 5632. 2001.

21. Bianconi G., Barabási A.-L. Competition and multiscaling in evolving networks // Europhysics Letters, 54: 436–442, 2001.

22. Ergün G, Rodgers G. J. Growing random network with fitness // arXiv: cond-mat/0103423.

23. Wang D., Song C., Barabási A.-L. Quantifying long-term scientific impact // Science. 2013. V. 342. P. 127–131.

24. Vázquez A. Knowing a network by walking on it: emergence of scaling // arXiv:cond-mat/0006132.

25. McGlohon M., Akoglu L., Faloutsos C. Weighted graphs and disconnected components // PAKDD'10 Proc. of the 14th Pacific-Asia conference on advances in knowledge discovery and data mining. 2010. Vol. P. 2. P. 410–421.

26. Kumar R., Raghavan P., Rajagopalan S., Sivakumar D., Tomkins A., Upfal E. Stochastic models for the Web graph // Proc. of the 41th IEEE symp. on foundations of computer science. 2000. P. 57–65.

27. Leskovec J., Kleinberg J. M., Faloutsos C. Graph evolution: Densification and shrinking diameters // ACM Transactions on Knowledge Discovery from Data (TKDD). 2007. V. 1, iss. 1. Art. 2.

Щербакова Наталья Григорьевна — ст. науч. сотр. Ин-та вычислительной математики и математической геофизики СО РАН; e-mail: nata@nsc.ru.

Наталия Щербакова окончила Новосибирский государственный университет по специальности „Математическая лингвистика“ в 1967 году. С 1967 г. работала в Институте математики СО РАН, затем в Институте автоматики и электрометрии СО РАН в области создания программного обеспечения систем передачи данных. С 2000 года — сотрудник Института вычислительной математики и математической геофизики СО РАН, где с 2002 занимает должность старшего научного сотрудника. Являлась участником проекта „Сеть Интернет Новосибирского Научного Центра“, занималась вопросами мониторинга и анализа IP-сетей. Автор и соавтор более 40 работ, соавтор монографии „Анализ цитирования в библиометрии“. Текущие интересы лежат в области исследования методов оценки научной деятельности на основе анализа цитирования научной литературы.

Natalia Shcherbakova graduated from Novosibirsk State University in 1967 (mathematical linguistics). Since 1967 she worked at Institute of Mathematics SB RAS, then at Institute of Automation and Electrometry SB RAS in the field of software design for data transmission systems. In 2000 — the employee of Institute of Computational Mathematics and Mathematical Geophysics SB RAS, since 2002 works as senior researcher. She is a member of „Akademgorodok Internet Project“, dealt with software of monitoring and the analysis of IP networks. She is the author and co-author of more than 40 works, the co-author of the monograph „Ansliz tsitirovaniya v bibliometrii“. The current research interests lie in the field of bibliometrics: methods of measuring of scientific.

Дата поступления — 20.06.2019