

УДК 519.175.4:004.421.5:004.7

СЛУЧАЙНЫЕ ГРАФЫ ПРЕДПОЧТИТЕЛЬНОГО СВЯЗЫВАНИЯ С ДОБАВЛЕНИЕМ ПОЛНЫХ ПОДГРАФОВ

Е. Б. Юдин

Институт математики им. С. Л. Соболева СО РАН, udinev@asoiu.com

Предлагается расширение модели случайных графов с нелинейным правилом предпочтительного связывания путем учета возможностей добавления в сеть целых сообществ. Эти сообщества моделируются полными графами. Такое расширение модели естественным образом обосновывается известными процессами в социальных сетях. В работе выводятся математические соотношения, которые позволяют рассчитать распределение степени связности вершин $\{Q_k\}$ сгенерированного графа. Также выводятся уравнения, позволяющие при заданном распределении степени связности узлов реальной сети подобрать параметры генерации, учитывающих процесс добавления сообществ. Полученные результаты проверяются путем генерации больших графов.

Ключевые слова: графы предпочтительного связывания, сетевые сообщества, число клик.

PREFERENTIAL ATTACHMENT RANDOM GRAPHS WITH COMPLETE SUBGRAPHS INCREMENT

E. B. Yudin

*Sobolev Institute of Mathematics, Siberian Branch, Russian Academy of Sciences,
udinev@asoiu.com*

The model extension of nonlinear preferential attachment random graphs considering possibilities of adding entire communities in the network is proposed. The added communities are represented by entire graphs. Such an extension of model is naturally based on existing processes on social networks. Mathematical relations are derived. These relations allow calculating the degree distribution of vertices $\{Q_k\}$ of a generated graph. Equations are also derived which allow, for a given node degree distribution of real network getting generation properties considering the process of adding communities. The results obtained are verified by generating large graphs.

Keywords: preferential attachment graphs, online communities, number of clicks.

Введение

Изучаемые Теорией сетей (Network Science) структуры имеют разную природу, но схожие свойства, среди которых исследователи выделяют следующие:

- распределение степеней связности узлов сети является степенным;
- сети имеют небольшой, часто уменьшающийся с увеличением размера сети, диаметр (правило «шести рукопожатий»);
- в сетях присутствуют сообщества (некоторые группы узлов, имеющих больше связей между собой, чем с узлами из других сообществ).

В качестве моделей больших сетей часто используют случайные графы предпочтительного связывания [1]. Эти модели растущих сетей позволяют генерировать графы, реализующие:

- заданное распределение степени связности вершин;
- заданное распределение степени связности вершин и заданный коэффициент кластеризации [2–5];
- заданные распределение степени связности вершин и совместное распределение концевых степеней ребер [6].

Схему генерации графов предпочтительного связывания удобно рассматривать на примере случайного графа с нелинейным правилом предпочтительного связывания (НППС). Генерация графа с НППС (рис. 1) представляет собой итерационный процесс. На каждом шаге этого процесса к существующему графу добавляется приращение – вершина со случайным числом ребер, свободные концы которых присоединяются к i -й вершине существующего графа с вероятностью, пропорциональной некоторой функции предпочтения $f(k_i)$, зависящей от степени связности k_i вершины графа:

$$p_i = \frac{f(k_i)}{\sum_{i=1}^n f(k_i)}.$$

Если число ребер в приращении является константой m , а функция предпочтения линейна: $f(k_i) = k_i$, то мы получим известный граф Барабаши – Альберт (БА) [1], в котором моделируется принцип «богатый становится богаче» (чем больше степень связности вершины, тем больше вероятность присоединения к ней). Если функция предпочтения логарифмическая: $f(k_i) = \log(k_i)$, то моделируется закон Вебера – Фихнера, в соответствии с которым интенсивность ощущения пропорциональна логарифму интенсивности раздражителя, в данном случае степени связности k_i вершины.

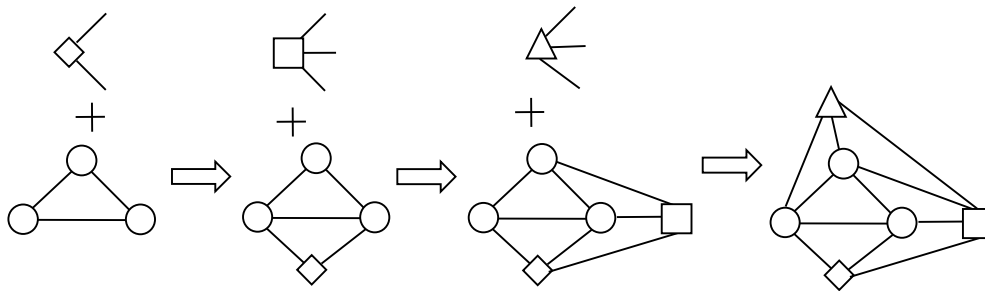


Рис. 1. Генерация графов с НППС

Примечание: к графу-затравке добавляются вершины со случайным числом ребер. На рисунке добавляется вершина с двумя ребрами, затем дважды с тремя

Однако в графах предпочтительного связывания (особенно это касается наиболее исследованного графа предпочтительного связывания – графа БА) сообщества, как правило, отсутствуют, в некотором смысле такие графы являются однородными. Даже для небольшого размера полных подграфов (клик), содержащих всего лишь три вершины, в графах БА количество таких «треугольных» клик существенно меньше по сравнению с их количеством в реальных сетях с соизмеримым числом узлов и ребер [3]. Тем более это справедливо для числа клик большего размера. В данной работе предлагаются графы предпочтительного связывания, в котором эта проблема решается. В этих графах предполагается, что сообщество добавляется в сеть целиком, тем самым моделируется этап «добавления сообщества» [7]. Добавляемое сообщество представляет собой заданное число n узлов, любые два узла из которых связаны между собой.

Графы с добавлением полных подграфов

Графы с добавлением полных подграфов представляет собой модифицированную версию графов с НППС. Пусть на каждом шаге выращивания графа с вероятностью γ поступает полный граф, содержащий n связанных между собой вершин (n -ада), с вероятностью $(1 - \gamma)$ – обычное приращение графа – монада (рис. 2). Число добавляемых ребер у вершины монады определяется в соответствии с дискретным распределением $\{r_k^{(1)}\}$, у каждой из вершин n -ады

часть ребер являются частью полного подграфа, а часть – разыгрывается в соответствии с распределением $\{r_k^{(n)}\}$, у всех вершин n -ады число добавляемых ребер одинаково. Свободные концы ребер монады присоединяются к графу независимо, в соответствии с вероятностью:

$$p_i = \frac{f(k_i)}{\sum_{j=1}^N f(k_j)}, \quad i = 1, \dots, N, \quad (1)$$

где N – число вершин графа.

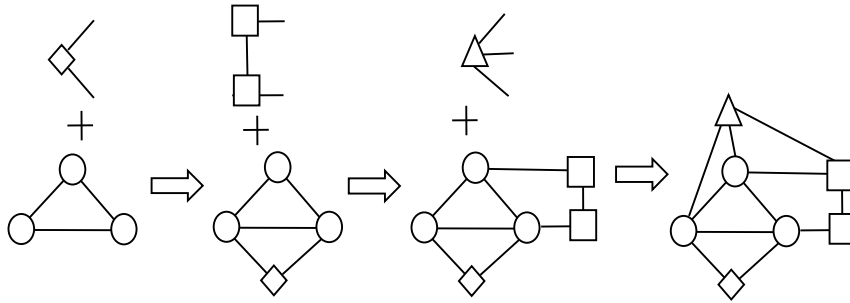


Рис. 2 Генерация графа с НППС с добавлением сообществ, на каждом шаге генерации графа добавляется двухвершинное приращение (диада) с вероятностью γ или одновершинное приращение (монада) с вероятностью $(1-\gamma)$

У каждой вершины n -ады μ ребер присоединяются к одной и той же вершине, разыгрываемой по формуле (1). Остальные свободные концы ребер n -ады присоединяются к существующим вершинам графа с вероятностью (1) независимо.

Можно записать для предлагаемого графа следующие свойства.

1. В среднем за один шаг поступает $(1 + (n-1)\gamma)$ вершин.

2. Обозначим среднее число ребер монады $m^{(1)}$, тогда $m^{(1)} = \sum_{k=1}^t i \cdot r_i^1$, где t – максимальное значение степени связности монады.

3. Среднее число ребер n -ады $(n \cdot m^{(n)} + n(n-1)/2)$, свободных (не входящих в клику размером n) ребер $n \cdot m^{(n)}$, где $m^{(n)}$ – среднее число свободных ребер каждой из вершин n -ады.

4. За один шаг к графу добавляется в среднем $\gamma \cdot (n \cdot m^{(n)} + m^{(n)} \cdot (n-1)/2) + (1-\gamma) \cdot m^{(1)}$ ребер.

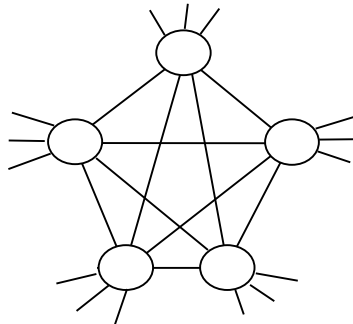


Рис. 3. Полный граф с пятью вершинами в качестве приращения (пентада)

Примечание: разыгранных свободных ребер три, входящих в клику – десять

Поставим следующие задачи:

1. В рамках предложенной модификации модели графов с НППС вывести уравнения, которые позволят рассчитать распределение степени связности (РСС) вершин $\{Q_k\}$, сгенерированного графа.

2. В рамках предложенной модификации модели графов с НППС вывести уравнения, позволяющие при заданном РСС узлов реальной сети подобрать параметры генерации, учитывающих процесс добавления сообществ.

Вывод распределения степени связности вершин

Разобьем множество всех вершин имеющегося графа на подмножества A_k (слои) вершин, содержащие вершины с одинаковой связностью k ($k = m, m+1, \dots$). Вероятность Q_k слоя A_k определим как вероятность выбора вершины именно из слоя A_k при случайном (равновероятном) выборе любой из N вершин имеющегося графа: $Q_k = |A_k| / N$, где $|A_k|$ – число вершин со связностью k .

Финальное значение P_k вероятности $\sum_{i \in A_k} p_i$ того, что для связывания с ребром новой вершины будет выбрана вершина слоя A_k , можно получить из формулы (1):

$$\sum_{i \in A_k} p_i = \frac{\sum_{i \in A_k} f(k_i)}{\sum_{j=1}^N f(k_j)} = \frac{f(k)|A_k|}{\sum_{l \geq g} f(l)|A_l|} = \frac{f_k |A_k| / N}{\sum_{l \geq g} f_l |A_l| / N} \sim \frac{Q_k f_k}{\sum_{l \geq g} Q_l f_l} = \frac{Q_k f_k}{\langle f \rangle} = P_k.$$

Таким образом, вероятность P_k выбора слоя A_k можно вычислить как

$$P_k = \frac{Q_k f_k}{\langle f \rangle}, \quad (2)$$

где $\langle f \rangle = \sum_{l \geq g} Q_l f_l$ – среднее значение функции предпочтения.

Рассмотрим, как изменяется $|A_k|$ при поступлении монады:

1. Если поступает монада, то $|A_k|$ с вероятностью r_k увеличится на 1. В среднем $|A_k|$ за счет этого увеличивается на $1 \cdot r_k^{(1)} + 0 \cdot (1 - r_k^{(1)}) = r_k^{(1)}$.

2. Кроме того, в среднем $m^{(1)}P_k$ ребер монады связываются с вершинами слоя A_k и эти вершины уходят в слой A_{k+1} , т. е. A_k уменьшается в среднем на $m^{(1)}P_k$.

3. В среднем $m^{(1)}P_{k-1}$ ребер монады связываются с вершинами слоя A_{k-1} и эти вершины уходят в слой A_k , т. е. A_k увеличивается в среднем на $m^{(1)}P_{k-1}$.

Итого при добавлении монад:

$$\Delta_1 = r_k^{(1)} + m^{(1)}P_{k-1} - m^{(1)}P_k. \quad (3)$$

Рассмотрим, как изменяется $|A_k|$ при поступлении n -ады (рис. 4).

1. Вершины n -ады с вероятностью $r_{k+1-n}^{(n)}$ попадают в слой A_k , это дает увеличение $|A_k|$ на $n \cdot r_{k+1-n}^{(n)}$.

2. Для каждой из вершин n -ады справедливо, что μ исходящих из нее ребер связываются с вероятностью P_{k-n} с вершинами слоя A_{k-n} и эти вершины переходят в слой A_k . Это дает увеличение $|A_k|$ на μP_{k-n} .

3. В среднем μ сопряженных n -ок ребер связываются с вершинами слоя A_k и эти вершины переходят в слой A_{k+n} . Это дает уменьшение $|A_k|$ на μP_k .

4. Свободные ребра n -ады связываются с вершинами слоя A_{k-1} с вероятностью P_{k-1} . Таких одиночных ребер $n \cdot m^{(n)} - n \cdot \mu$. Увеличение $|A_k|$ составляет $(n \cdot m^{(n)} - n \cdot \mu) P_{k-1}$.

5. Свободные ребра n -ады связываются с вершинами слоя A_k с вероятностью P_k . Уменьшение $|A_k|$ составляет $(n \cdot m^{(n)} - n \cdot \mu) P_k$.

Итого получаем:

$$\Delta_2 = n \cdot r_{k+1-n}^{(n)} + \mu P_{k-n} - \mu P_k + (n \cdot m^{(n)} - n \cdot \mu) P_{k-1} - (n \cdot m^{(n)} - n \cdot \mu) P_k. \quad (4)$$

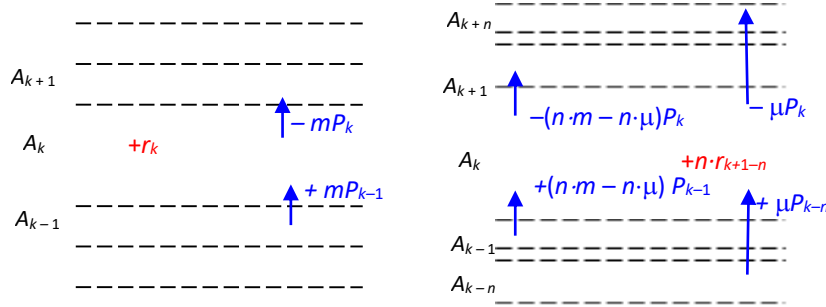


Рис. 4. Схема изменения числа вершин в слое A_k при добавлении монады (слева) и n -ады (справа)

На каждом шаге выращивания графа модуль поступает с вероятностью γ , обычное приращение – с вероятностью $(1 - \gamma)$. С учетом этого:

$$\begin{aligned} \Delta|A_k| &= (1 - \gamma)\Delta_1 + \gamma \cdot \Delta_2 = (1 - \gamma)(r_k^{(1)} + m^{(1)}P_{k-1} - m^{(1)}P_k) + \\ &+ \gamma(n \cdot r_{k+1-n}^{(n)} + \mu P_{k-n} - \mu P_k + (n \cdot m^{(n)} - n \cdot \mu)P_{k-1} - (n \cdot m^{(n)} - n \cdot \mu)P_k). \end{aligned}$$

Приравняв долю вершин в слое A_k до выполнения шага выращивания графа к соответствующей доле A_{k+1} после выполнения этого шага, получаем следующее уравнение баланса:

$$\frac{|A_k|}{N} = \frac{|A_k| + \Delta|A_k|}{N + ((1 - \gamma) + \gamma \cdot n)}.$$

Что можно переписать как

$$\begin{aligned} \frac{|A_k|}{N} &= \frac{|A_k| + r_k^{(1)} + m^{(1)}P_{k-1} - m^{(1)}P_k - \gamma \cdot r_k^{(1)} - \gamma \cdot m^{(1)}P_{k-1} + \gamma \cdot m^{(1)}P_k}{N + ((1 - \gamma) + \gamma \cdot n)} + \\ &+ \frac{\gamma \cdot n \cdot r_{k+1-n}^{(n)} + \gamma \cdot \mu \cdot P_{k-n} - \gamma \cdot \mu \cdot P_k + \gamma \cdot (n \cdot m^{(n)} - n \cdot \mu)P_{k-1} - \gamma \cdot (n \cdot m^{(n)} - n \cdot \mu)P_k}{N + ((1 - \gamma) + \gamma \cdot n)}. \end{aligned}$$

Полученное уравнение перепишем следующим образом:

$$\begin{aligned} |A_k| N + |A_k| \cdot ((1 - \gamma) + \gamma \cdot n) &= N |A_k| + N (r_k^{(1)} + m^{(1)}P_{k-1} - m^{(1)}P_k - \gamma \cdot r_k^{(1)} - \gamma \cdot m^{(1)}P_{k-1} + \\ &+ \gamma \cdot m^{(1)}P_k + \gamma \cdot n \cdot r_{k+1-n}^{(n)} + \gamma \cdot \mu \cdot P_{k-n} - \gamma \cdot \mu \cdot P_k + \gamma (n \cdot m^{(n)} - n \cdot \mu)P_{k-1} - \gamma (n \cdot m^{(n)} - n \cdot \mu)P_k) = \\ &= N \cdot |A_k| + N \cdot (r_k^{(1)}(1 - \gamma) + \gamma \cdot n \cdot r_{k+1-n}^{(n)} + (m^{(1)}(1 - \gamma) + \gamma \cdot n \cdot m^{(n)} - \gamma \cdot n \cdot \mu) \cdot P_{k-1} + \\ &+ \gamma \cdot \mu \cdot P_{k-n} + (m^{(1)} \cdot (\gamma - 1) - \gamma \cdot n \cdot m^{(n)} + \gamma(n - 1) \cdot \mu) P_k). \end{aligned}$$

Сократим одинаковые члены $|A_k|N$ и разделим равенство на N . В результате выражение $|A_k|((\gamma - 1) + \gamma \cdot n)$ в левой части превращается в $Q_k((\gamma - 1) + \gamma \cdot n)$, а множители N в правой

части сокращаются. Заменяя все P_k в правой части равенства их выражением $P_k = \frac{Q_k f_k}{\langle f \rangle}$, получаем уравнение баланса для Q_k :

$$Q_k \cdot \left(1 + \gamma(n-1) - \frac{f_k(m^{(1)}(\gamma-1) - \gamma \cdot n \cdot m^{(n)} + \gamma(n-1)\mu)}{\langle f \rangle} \right) = \\ = \frac{(r_k^{(1)}(1-\gamma) + \gamma \cdot n \cdot r_{k+1-n}^{(n)})\langle f \rangle + (m^{(1)}(1-\gamma) + \gamma \cdot n \cdot m^{(n)} - \gamma \cdot n \cdot \mu - \gamma \cdot m^{(1)})Q_{k-1} + \gamma \cdot \mu \cdot Q_{k-n}}{\langle f \rangle}.$$

В явной форме для Q_k можно записать выражение:

$$Q_k = \frac{(r_k^{(1)}(1-\gamma) + \gamma \cdot n \cdot r_{k+1-n}^{(n)})\langle f \rangle + (m^{(1)}(1-\gamma) + \gamma \cdot n \cdot m^{(n)} - \gamma \cdot n \cdot \mu) f_{k-1} Q_{k-1} + \gamma \cdot \mu \cdot f_{k-n} Q_{k-n}}{\langle f \rangle + \langle f \rangle \gamma(n-1) + f_k [m^{(1)} \cdot (1-\gamma) + \gamma \cdot n \cdot m^{(n)} - \gamma(n-1) \cdot \mu]}, \quad (5)$$

где $k = g, g+1, \dots, M$, $\langle f \rangle = \sum_{k \geq g} f_k Q_k$.

Если нет n -ад, т. е. $\gamma = 0$, то получим известную формулу [8]:

$$Q_k = \frac{r_k^{(1)} \langle f \rangle + m^{(1)} \cdot f_{k-1} Q_{k-1}}{\langle f \rangle + m^{(1)} \cdot f_k}.$$

Если нет монад, т. е. $\gamma = 1$, при $n = 2$ получим:

$$Q_k = \frac{2 \cdot r_{k-1}^{(2)} \langle f \rangle + (2m^{(n)} - 2\mu) \cdot f_{k-1} Q_{k-1} + \mu \cdot f_{k-2} Q_{k-2}}{2\langle f \rangle + 2f_k \cdot m^{(2)} - f_k \mu}.$$

Калибровка графа по распределению степеней вершин

В задаче калибровки графа требуется найти параметры $\{r_k^{(1)}\}, \{r_k^{(n)}\}, \{f_k\}, \gamma, \mu$ при которых граф будет иметь заданное РСС вершин $\{Q_k\}$. Эту задачу можно решить, используя уравнение (5).

Выражая из (5) f_k , получаем рекуррентное соотношение:

$$f_k = \frac{[r_k^{(1)}(1-\gamma) + \gamma \cdot n \cdot r_{k+1-n}^{(n)}]\langle f \rangle}{a \cdot Q_k} - \frac{[\langle f \rangle + \langle f \rangle \gamma(n-1)]Q_k}{a \cdot Q_k} + \\ + \frac{[(m^{(1)}(1-\gamma) + \gamma \cdot n \cdot m^{(n)} - \gamma \cdot n \cdot \mu) f_{k-1} Q_{k-1} + \gamma \mu f_{k-n} Q_{k-n}]\langle f \rangle}{\langle f \rangle \cdot a \cdot Q_k}, \quad (6)$$

где $a = m^{(1)} \cdot (1-\gamma) + \gamma \cdot n \cdot m^{(n)} - \gamma(n-1) \cdot \mu$.

Соотношением (6) определяется определяющая функцию предпочтения последовательность $\{f_k\}$, каждый член которой имеет общий мультипликативный множитель $\frac{\langle f \rangle}{a}$.

Причем следует заметить, что последовательности $\{f_k\}$ и $\{Cf_k\}$, которые отличаются множителем $C > 0$, эквивалентны, учитывая формулу (1). Следовательно, мы можем приравнять

множитель $\frac{\langle f \rangle}{a}$ в правой части соотношения (6) к единице (определяя тем самым равенство $\langle f \rangle = a$) и записать это соотношение в более компактном виде:

$$f_k = \frac{r_k^{(1)}(1-\gamma) + \gamma n r_{k+1-n}^{(n)} - (1+\gamma(n-1))Q_k}{Q_k} + \frac{(m^{(1)}(1-\gamma) + \gamma n m^{(n)} - \gamma n \mu) f_{k-1} Q_{k-1} + \gamma \mu f_{k-n} Q_{k-n}}{\langle f \rangle \cdot Q_k}. \quad (7)$$

Учитывая, что в результате сделанного выбора мы имеем равенство $\langle f \rangle = a$, параметра $a = m^{(1)} \cdot (1-\gamma) + \gamma \cdot n \cdot m^{(n)} - \gamma(n-1) \cdot \mu$, получаем следующее рекуррентное соотношение:

$$f_k = \frac{r_k^{(1)}(1-\gamma) + \gamma \cdot n \cdot r_{k+1-n}^{(n)} - (1+\gamma(n-1))Q_k}{Q_k} + \frac{(m^{(1)}(1-\gamma) + \gamma \cdot n \cdot m^{(n)} - \gamma \cdot n \cdot \mu) f_{k-1} Q_{k-1} + \gamma \mu f_{k-n} Q_{k-n}}{[m^{(1)} \cdot (1-\gamma) + \gamma \cdot n \cdot m^{(n)} - \gamma(n-1) \cdot \mu] \cdot Q_k}, \quad (8)$$

где $k = g, g+1, \dots, M$.

Формула (8) является решением задачи калибровки графа. Она позволяет для реализации заданного РСС $\{Q_k\}$ получить бесконечно много последовательностей весов $\{f_k\}$, фактически задающих функцию предпочтения $f(k)$, в зависимости от того, какие параметры $\{r_k\}$, γ , μ и m будут выбраны.

Численные эксперименты

Полученные формулы можно проверить экспериментально путем использования электронных таблиц.

На рис. 5 в среде электронных таблиц представлен результат калибровки графа, когда в качестве приращения использовались диады (связанные пары вершин). При этом требовалось реализовать следующее распределение степени связности вершин в генерируемом графе:

$$Q_k = \frac{12}{k(k+1)(k+2)},$$

что соответствует распределению степени связности вершин графа БА с параметром $m = 2$ (при использовании графу предпочтительного связывания с одновершинным приращением).

В процессе калибровки по формуле (8) получены значения функции предпочтения $f(k)$. Результаты были проверены пересчетом по формуле (5) с параметрами $\gamma = 1$, $n = 2$, $\mu = 1$.

Как можно видеть, результаты расчетов в электронной таблице приводят к получению искомого распределения.

На рис. 4 приведены результаты генерации графа с заданными параметрами после добавления 50 тыс. приращений, в качестве графа затравки использовался полный граф, содержащий 4 вершины. Отклонение статистических оценок первых десяти значений распределения степени связности, полученных путем генерации графа, находятся в пределах статистической погрешности.

На рис. 6 представлена калибровка графа с использованием пентад (полных графов в приращении, содержащих 5 вершин) и монад. Вероятность появления пентад $\gamma = 0.01$. В качестве базового распределения степеней связности вершин взято сглаженное распределение, полученное по данным о сети автономных систем Интернет (доступного по электронному адресу <http://www-personal.umich.edu/~mejn/netdata/as-22july06.zip>).

СУММПРОИЗВ X ✓ ≠ =((B21*(1-\$C\$2)+\$C\$2*2*B20)*\$G\$12+(\$E\$10+\$C\$2*(2-1)*\$E\$10-\$C\$2*2*\$C\$1)*G20*F20+\$C\$2*\$C\$1*G19*F19)/(\$G\$12+\$G\$12*\$C\$2*(2-1)+G21*\$K\$14)														
	A	B	C	D	E	F	G	H	O	P	Q	R	S	T
13	в приращении				k	Qk(k)	f(k)	Qk(fk)						
14	i	r(i)				теор		по нов формуле						
18	0	0			0	0	0	0		Результаты генерации графа мю = 1 гамма=1				
19	1	0			1	0	1	0						
20	2	1			2	0	2	0						
21	3	0			3	0,526316	3	\$2*(2-1)*\$E\$10-\$						
22					4	0,143541	4	0,14354067		Вершин	100004			
23					5	0,109091	5	0,109090909		Ребер	250006			
24					6	0,059467	6	0,059466849		m	2,49996			
25					7	0,040615	7	0,040614733		k	n(k)	Qk(k)		
26					8	0,027218	8	0,027217864		0	0	0		
27					9	0,019454	9	0,019453755		1	0	0		
28					10	0,014198	10	0,014197763		2	0	0		
29					11	0,010675	11	0,010675327		3	52580	0,525779		
30					12	0,008192	12	0,008192061		4	14376	0,143754		
31					13	0,006409	13	0,00640894		5	11001	0,110006		
32					14	0,005095	14	0,005094946		6	6051	0,060508		
33					15	0,004109	15	0,004108631		7	3984	0,039838		
34					16	0,003355	16	0,003354968		8	2706	0,027059		
35					17	0,00277	17	0,002770303		9	1946	0,019459		
36					18	0,00231	18	0,002310465		10	1427	0,014269		
37					19	0,001944	19	0,001944357		11	1025	0,01025		
38					20	0,00165	20	0,001649628		12	769	0,00769		
39					21	0,00141	21	0,001409971		13	614	0,00614		
40					22	0,001213	22	0,001213307		14	536	0,00536		
41					23	0,001051	23	0,001050568		15	405	0,00405		
										16	317	0,00317		

Рис. 5. Проверка полученных формул для графа БА с параметром $m = 4$

СУММПРОИЗВ																=(L23*(1-SC\$2)+SC\$2*5*B19-F23*(1+SC\$2*4))/F23+((SE\$11*(1-SC\$2)+SC\$2*5*SE\$10-SC\$2*5*SC\$1)*F22*G22+SC\$2*SC\$1*F18*G18)/F23/513															
A		B		C		D		E		F		G		H		J		O		P		Q		R		S		T		U	
		Мю		1																											
1		гамма		0,010000																											
2																															
3																															
4																															
7																															
8																															
9																															
10																															
11																															
12		Вер. числ.реб.																													
13		в приращении																													
14		i		r(i)																											
15																															
16																															
17																															
18		0		0																											
19		1		0,39091																											
20		2		0,04																											
21		3		0,08																											
22		4		0,12																											
23		5		0,16																											
24		6		0,2																											
25		7		0,00909																											
26						1																									
27																															
28																															
29																															
30																															
31																															
32																															
33																															
34																															
35																															
36																															
37																															
38																															
39																															
40																															
41																															
42																															
43																															
44																															
45																															
46																															

Рис. 6. Калибровка случайного графа с использованием электронных таблиц

На рис. 6 представлено сглаженное распределение для степени связности вершин исследуемой сети автономных систем и распределение для степени связности вершин сгенерированного графа. Как можно видеть на рис. 7, при малых значениях локальной степени связности распределения совпадают, при больших значениях выборка отсутствует (отсюда

большие отклонения), поскольку генерируется граф небольшого размера. Заметим также, что теоретическое распределение для генерируемых графов соответствует сглаженному распределению.

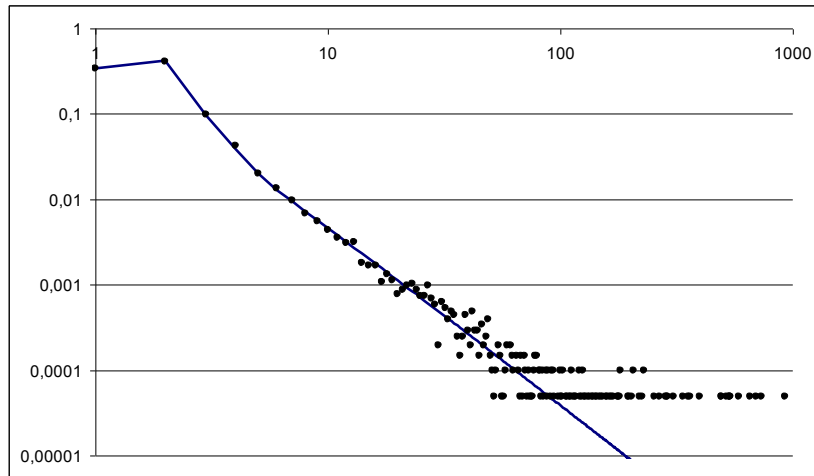


Рис. 7. Сглаженное распределение степени связности узлов в сети автономных систем (линия) и в реализации случайного графа того же размера с использованием пентад (маркеры)

Заключение

Несмотря на многочисленные публикации по анализу сообществ в реальных сетях, число работ, решающих более фундаментальные вопросы о моделировании механизмов возникновения сообществ, немного. Так, в работах [9, 10] предлагаются геометрические графы предпочтительного связывания (где геометрия задает степень подобия, влияющую на предпочтение присоединения, и определяющую форму сообществ), в работах [11, 12], предлагается «генерировать» графы как отдельные слабосвязанные между собой компоненты (которые тоже можно рассматривать как сообщества). В данной работе предложен случайный граф, с помощью которого моделируется добавление сообществ как единого целого. Такому процессу можно найти интерпретацию в социальных сетях. Так, при появлении очередной учебной группы, в подсеть добавляются новые узлы (представляющие собой студентов), которые сразу же становятся связанными между собой (об этом свидетельствуют исследования, в том числе [13]: студенты, как правило, связываются в подсети с другими студентами из своей учебной группы).

Дальнейшее развитие предложенной модели может состоять в учете и других процессов, происходящих в сетях. Так в работах [14, 15] путем анализа динамики изменений структуры сети выявляются такие процессы как разрушения сообществ (разбиение на несколько меньших или полное уничтожение сообщества), связь возраста сообщества и его размера.

Исследование выполнено при финансовой поддержке Российского фонда фундаментальных исследований в рамках научного проекта № 16-31-60023 мол_а_дк.

Литература

1. Barabási A. L., Albert R. Emergence of scaling in random networks // Science. 1999. Vol. 286. P. 509–512.
2. Dorogovtsev S. N., Mendes J. F. F., Samukhin A. N. Generic scale of the “scale-free” growing networks // Phys. Rev. Vol. 63. 2001.
3. Klemm K., Eguíluz V. Highly clustered scale-free networks // Phys. Rev. Vol. 65. 2002.
4. Jackson M. O., Rogers B. W. Meeting strangers and friends of friends: How random are social networks? // Am. Econ. Rev. 2007. № 97 (3). P. 890–915.

5. Zadorozhnyi V. N., Yudin E. B. Structural properties of the scale-free Barabasi – Albert graph // Automation and Remote Control. 2012. № 73 (4). P. 702–716. doi: 10.1134/S0005117912040091.
6. Zadorozhnyi V. N., Yudin E. B. Dynamic equations of node degrees in growing networks with connection losses // Dynamics of Systems, Mechanisms and Machines (Dynamics). 2016. art. no. 7819111. P. 1–5. doi: 10.1109/Dynamics.2016.7819111.
7. Palla G., Barabási A.-L., Vicsek T. Quantifying social group evolution // Nature. 2007. № 446. P. 664–667.
8. Задорожный В. Н. Случайные графы с нелинейным правилом предпочтительного связывания // Проблемы управления. 2011. № 6. С. 2–11.
9. Zuev K., Boguñá M., Bianconi G., Krioukov D. Emergence of Soft Communities from Geometric Preferential Attachment // Scientific Reports. 2015. № 9421. doi:10.1038/srep09421.
10. Dorodnykh A., Ostroumova Prokhorenkova L., Samosvat E. Preferential placement for community structure formation // In International Workshop on Algorithms and Models for the Web-Graph. Springer, 2017. P. 75–89.
11. Klemm K., Eguiluz V. M. Growing scale-free network with small world behavior // Phys Rev E Stat Nonlin Soft Matter Phys. 2002. № 65.
12. Zadorozhnyi V. N., Yudin E. B., Yudina M. N. Analytical and numerical methods of calibration for preferential attachment random graphs // International Siberian Conference on Control and Communications, SIBCON ; S. Seifullin Kazakh Agrotechnical University Astana; Kazakhstan; 29–30 June 2017. Astana, 2017. doi: 10.1109/SIBCON.2017.7998461.
13. Юдина М. Н. Узлы в социальных сетях: меры центральности и роль в сетевых процессах // Омский научный вестник. 2016. № 4 (148). С. 161–165.
14. Chi Y., Song X., Zhou D., Hino K., and Tseng B. L. Evolutionary spectral clustering by incorporating temporal smoothness // KDD '07: Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM. N.Y., USA, 2007. P. 153–162.
15. Lin Y.-R., Chi Y., Zhu S., Sundaram H., and Tseng B. L. Facetnet: a framework for analyzing communities and their evolutions in dynamic networks // WWW '08: Proceedings of the 17th International Conference on the World Wide Web, ACM. N.Y., USA, 2008. P. 685–694.