

## ИМИТАЦИОННОЕ МОДЕЛИРОВАНИЕ СОЦИАЛЬНОЙ СЕТИ НА ОСНОВЕ РЕТРОСПЕКТИВНЫХ ДАННЫХ

В. А. Бадрызлов (Омск)

### Введение

В современном обществе социальные сети, основанные на интернет-технологиях, стали одним из самых влиятельных каналов распространения информации. Через социальные сети происходит не только межличностное общение их участников, но и продвижение информации, идей, мнений. Иногда эти процессы носят деструктивный, разрушающий характер и хорошо бы было научиться управлять этими процессами. Но управление невозможно без понимания того, как организованы социальные сети, как они растут, как в них распространяется информация и как найти наиболее влиятельные узлы, обеспечивающие ее быстрое распространение. Новая отрасль науки «network science» ищет ответы на эти и многие другие вопросы. Основные трудности исследования сетевых структур состоят в том, что их непосредственное изучение осложнено очень большими размерами, а также тем, что происходят постоянные изменения в структуре сетей – отмирают старые связи, появляются новые связи и узлы. Эффективным методом исследования больших сетей является их моделирование с использованием случайных графов, которые воспроизводят основные, известные из эмпирических исследований, свойства сетей. Однако эти сведения носят ретроспективный характер, поскольку получены они задолго до момента исследования. Имитационное моделирование способно на основе ретроспективных данных дать оценки возможного будущего состояния сети.

### Случайные графы с нелинейным правилом предпочтительного связывания

Одним из эффективных методов моделирования больших сетевых структур является их представление в виде случайных графов с нелинейным правилом предпочтительного связывания (графов с НППС), теория которых представлена в работах [1–3]. Согласно этой теории, для выращивания графа с НППС используется граф-затравка из нескольких вершин, связанных ребрами. Любая вершина со степенью  $k$  имеет функцию предпочтения (вес)  $f(k)$ , где  $f(k) > 0$ , если  $g \leq k \leq M$ , иначе  $f(k) = 0$  (здесь  $g \geq 1$ ,  $M \leq \infty$ ). Генерация графа осуществляется путем добавления на каждом шаге вершины и случайного числа  $x$  инцидентных ей ребер, которые свободными концами соединяются с вершинами графа. Вероятность связывания ребра приращения с вершиной  $i$  графа, имеющего  $N$  вершин, определяется в виде

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

Количество ребер в приращении – случайная величина  $x \in \{g, g + 1, \dots, h\}$ , которая имеет дискретное распределение вероятностей  $\{r_k\}$ . Вероятность  $r_k = P(x = k) \geq 0$  при  $g \leq k \leq h$ ,  $\sum_{k=g}^h r_k = 1$ . При этом  $h \leq M$  и конечно.

Таким образом, алгоритм генерации графа с НППС задается функцией предпочтения  $f(k)$  и распределением вероятностей  $\{r_k\}$ . Предложенный класс графов с НППС позволяет реализовать самые разнообразные зависимости степени вершин от времени, в отличие от других случайных графов с предпочтительным связыванием.

### Калибровка генератора случайных графов

В рамках теории случайных графов с НППС разработаны методы синтеза случайных графов, которые по своим основным свойствам идентичны тем реальным сетям, которые они моделируют. В частности, в работе [1] предложен базовый метод калибровки случайных графов с НППС, далее уточненный в работах [4, 5]. Калибровка предполагает, что численными методами определяются два параметра генератора случайных графов:

- распределение вероятностей  $\{r_k\}$  числа ребер в приращении случайного графа;
- функция предпочтения  $f(k)$ , выраженная в табличной или аналитической форме.

Эти два параметра позволяют генератору строить случайный граф с распределением степеней связности (РСС) вершин таким же, как у реальной сети. Одновременно с расчетом параметров генератора метод калибровки дает точную оценку вероятностей  $Q_k$  того, что случайно выбранная вершина графа имеет степень связности  $k$ .

В работах [4, 5] рассмотрена калибровка фрагмента социальной сети Твиттер [6] и выполнено сопоставление РСС узлов сети Твиттер с РСС, полученным расчетным путем. Это сопоставление представлено на рис. 1 и, как показали расчеты, эмпирическое и расчетное РСС очень близки друг к другу.

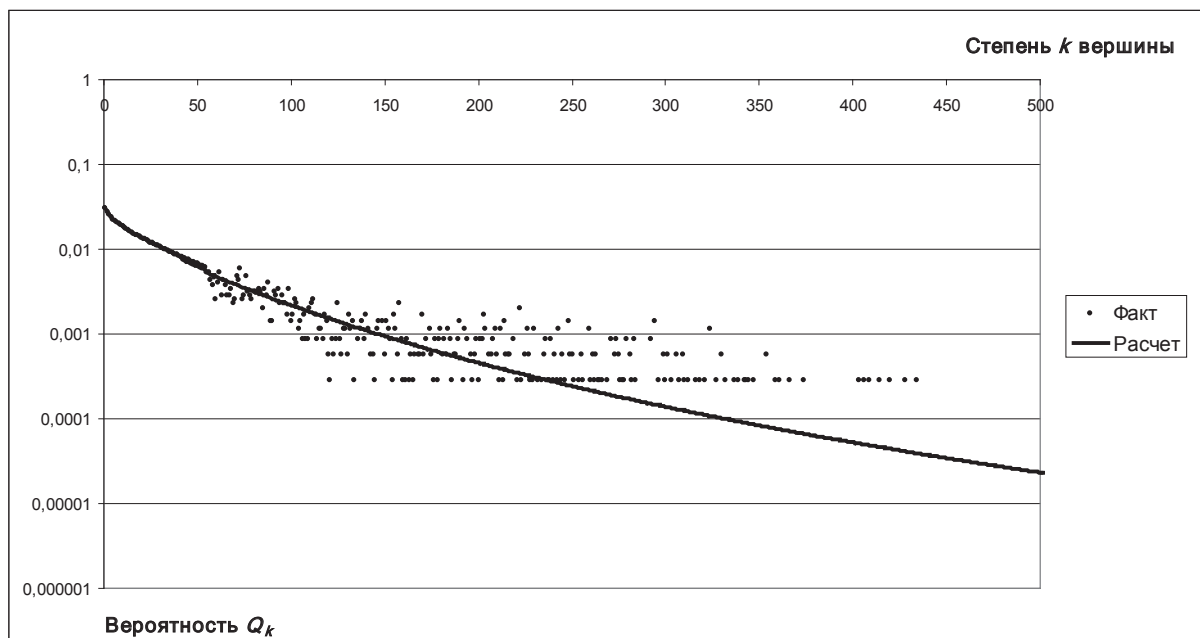


Рисунок 1 Распределение степени вершин (логарифмическая шкала вероятностей)

Сплошная линия – расчетные значения, маркеры – эмпирические значения

Метод калибровки параметров генератора случайных графов открывает новые перспективы для имитационного моделирования как социальных сетей, так и других сетевых структур. Получив ретроспективные данные о топологии сети, определяют параметры генератора. Выполняя далее многочисленные имитационные эксперименты по выращиванию случайного графа, можно получить оценки динамики роста как графа в целом, так и динамики роста его отдельных вершин.

## Программные средства имитационного моделирования

С целью исследования случайных графов в AnyLogic разработана модель «Генератор случайных графов предпочтительного связывания». Модель основана на теории случайных графов с НППС и позволяет генерировать широкий класс случайных графов с предпочтительным связыванием, начиная с графов Барабаши-Альберт [7]. Пример построения нескольких первых вершин графа, генерируемого в программе, представлен на рисунке 2. Далее случайный граф может разрастаться до 100000 и более вершин.

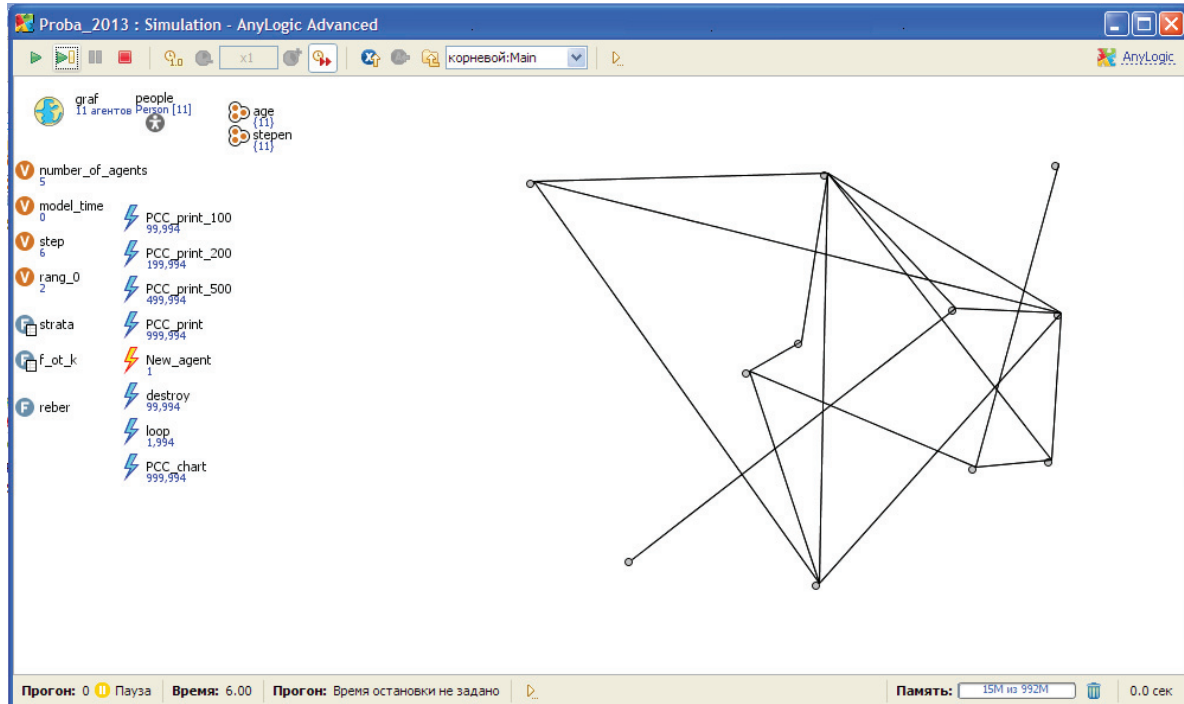


Рис. 2. Начальная фаза генерации случайного графа

В модели для формирования графа-затравки задаются две переменные:

**number\_of\_agents** – число вершин в графе затравке;

**rang\_0** – степень связности вершин графа-затравки.

Функция **reber** задает в модели распределение вероятностей числа ребер в приращении  $\{r_k\}$ . Табличная функция **f\_ot\_k** задает функцию предпочтения вершин  $f(k)$  в табличном виде. Для конкретных графов функции **reber** и **f\_ot\_k** могут быть переопределены так, как это необходимо для исследователя. Коллекции **age** и **stepen** хранят соответственно возраст и степень всех вершин.

Управление процессом моделирования осуществляется с помощью событий. Событие **New\_agent** запускается на каждом шаге модельного времени и состоит в том, что появляется новая вершина графа, определяется число ребер в приращении, устанавливаются связи новой вершины с уже существующими. То есть это событие выполняет всю совокупность действий по созданию приращения случайного графа в соответствии с теорией случайных графов с НППС. При его выполнении учитываются значение функции **reber** и функции предпочтения **f\_ot\_k**, заданной в табличном виде.

Событие **destroy** выполняется либо с определенной периодичностью, либо в случайные моменты времени, и состоит в том, что происходит удаление какой-либо вершины графа со всеми ее ребрами. Это позволяет моделировать сетевые структуры, в которых возможно исчезновение узлов.

События `PCC_print_100`, `PCC_print_200`, `PCC_print_500`, `PCC_chart`, `PCC_print` – однотипные события, состоящие в том, что в заданные моменты времени рассчитывается PCC вершин графа и выводится для анализа на консоль или в файл в различном виде, что позволяет изучать динамику изменения PCC.

Рассмотренные элементы модели позволяют выполнять следующие действия:

- задание параметров графа-затравки число вершин и ребер;
- задание функции предпочтения для вершин графа;
- задание распределения вероятностей числа ребер в приращении;
- генерация случайного графа с заданными параметрами;
- построение PCC вершин графа в заданные моменты времени;
- отслеживание степени выделенной вершины графа;
- вывод в файл или на консоль списка ребер графа;
- загрузка из файла топологии графа для его последующего наращивания.

### Имитационное моделирование фрагмента социальной сети Твиттер

Параметры случайного графа, моделирующего фрагмент сети Твиттер, полученные в результате калибровки, таковы. Распределения вероятностей числа ребер в приращении  $\{r_k\}$  предполагает, что в приращении графа может быть от 1 до 43 с соответствующими вероятностями. Функция предпочтения вершин определяется следующим образом. Для  $1 \leq k \leq 41$  значение функции  $f(k) = 0$ , далее для  $42 \leq k \leq 54$  значение функции предпочтения определяется таблицей и для  $k > 54$  функция  $f(k) = 40,32754 + 0,222566k$ .

Значения функции предпочтения для вершин со степенью  $42 \leq k \leq 54$

$k$	$f(k)$	$k$	$f(k)$	$k$	$f(k)$	$k$	$f(k)$	$k$	$f(k)$
42	11,23015	454	48,059431	48	48,36598	51	48,664537	54	48,957582
43	47,849217	46	48,162671	49	48,466257	52	48,762718	55	Далее по формуле
44	47,954986	47	48,264818	50	48,565746	53	48,860373	56	

Поскольку в приращении графа возможно появление до 42 ребер, граф-затравка при каждом имитационном эксперименте создавался из 50 вершин, а у вершин равновероятно могло появляться от 1 до 5 ребер. Рост графа производился на протяжении 3450 шагов модельного времени, что позволяло получить 3500 вершин – именно столько вершин было во фрагменте сети Твиттер, выбранном для анализа. По результатам экспериментов построено усредненное PCC. Сравнение PCC реальной сети с двумя PCC, полученными разными способами, представлено на рис. 3.

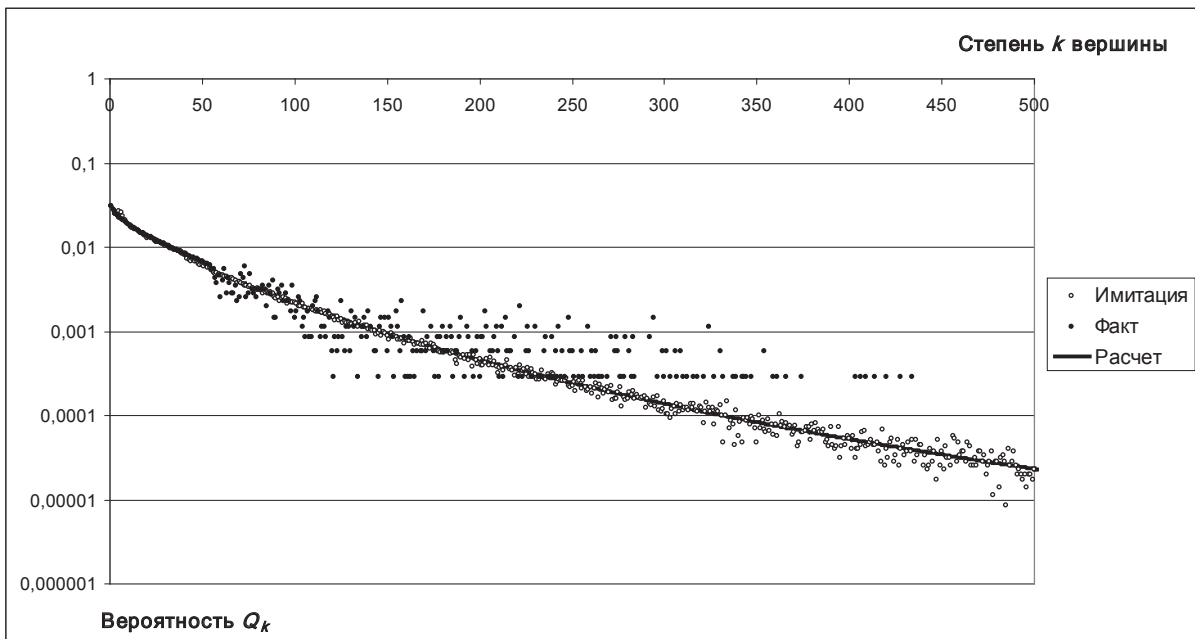


Рис. 3. Сравнение РСС, полученного разными методами

По рисунку можно отметить близость фактического, расчетного и экспериментального РСС, что подтверждает правильность результатов калибровки генератора случайных графов на основе ретроспективных данных о топологии реальной социальной сети.

### Выводы

Имея ретроспективные данные о топологии социальной сети или зная методы их получения, предложенные, например, в работах [8–10], а также используя методы имитационного моделирования, можно исследовать динамику роста конкретной социальной сети и процессы распространения в ней разнообразной информации. Эта возможность может быть актуальна для ряда экономических приложений, таких как определение стоимости, эффективности и скорости распространения рекламной информации в социальных сетях, а также поиска наиболее «выгодных» узлов сети в целях продвижения новых идей и распространения вредоносной или пропагандистской информации в информационной борьбе.

## Литература

1. **Задорожный В.Н.** Случайные графы с нелинейным правилом предпочтительного связывания // Проблемы управления. 2011. № 6. С. 2–11.
2. **Zadorozhnyi V., Yudin E.** Growing Network: Nonlinear Extension of the Barabasi-Albert Model // Communications in Computer and Information Science. 2014. Vol. 487. P. 432–439.
3. **Юдин Е. Б.** Генерация случайных графов предпочтительного связывания // Омский научный вестник. 2010. № 2 (90). С. 188–192.
4. **Бадрызлов В.А., Задорожный В.Н., Юдин Е.Б.** Особенности калибровки графов высокосвязных сетей //Россия молодая: передовые технологии – в промышленность! : матер. V Всерос. науч.-техн. конф. с междунар. участием (Омск, 12-14 ноября 2013 г.) : в 3 кн. Кн.2. Омск: Изд-во ОмГТУ, 2013. С.3–6.
5. **Бадрызлов В.А.** Идентификация и калибровка графов многосвязных социальных сетей // Омский научный вестник. 2014. №1 (127). С.176–180.
6. Граф фрагмента сети Твиттер [Электронный ресурс] URL: <http://wiki. Gephi.org/index.php/Datasets> (дата обращения 18.04.2013).
7. Barabási A.-L. Albert R. Emergence of scaling in random networks. Science 286. P. 509–512 (1999).
8. **Бондаренко А.С.** Алгоритм получения подсетей социальной сети «Вконтакте» с использованием фильтрующего критерия/Е.С. Бондаренко, Е.Б. Юдин// Информационные технологии и автоматизация управления: материалы VIII Всероссийской научно-практической конференции студентов, аспирантов, работников образования и промышленности. Омск, 2016. С. 196–199.
9. **Ниткин Д.А., Юдин Е.Б.** Исследование социальной сети «Вконтакте» // Информационные технологии и автоматизация управления: материалы VI Всероссийской научно-практической конференции студентов, аспирантов, работников образования и промышленности. Омск. 2015. С. 144–150.
10. **Пахомов И.А., Юдин Е.Б.** Получение структурной информации о сети WEB / И.А. Пахомов, // Информационные технологии и автоматизация управления: материалы VIII Всероссийской научно-практической конференции студентов, аспирантов, работников образования и промышленности. Омск. 2016. С. 241–246.