

## ИССЛЕДОВАНИЕ ПРОЦЕССОВ РАСПРОСТРАНЕНИЯ ИНФОРМАЦИИ В СОЦИАЛЬНОЙ СЕТИ МЕТОДОМ ИМИТАЦИОННОГО МОДЕЛИРОВАНИЯ

В.А. Бадрызлов, М.Н. Юдина (Омск)

### Введение

Социальные сети, получившие в современном обществе необычайно широкое распространение, являются эффективным средством распространения информации разного рода, как важной, позитивной, так и вредоносной. Вопросы распространения информации по сетям недостаточно изучены, наука о сетях (Network Science), занимающаяся, в том числе, и этими вопросами, находится на этапе становления, поэтому выполненное методами имитационного моделирования исследование процессов распространения информации в реальной сети представляется актуальным и укладывается в рамки науки о сетях.

Следует отметить, что непосредственное изучение социальных сетей – сложная и трудоемкая задача. Сложность и трудоемкость связаны с тем, что сети могут состоять из сотен тысяч и миллионов узлов, что создает как проблемы сбора информации об узлах сети и их взаимосвязях, так и вычислительные проблемы, связанные с обработкой полученной информации. Кроме того, взаимосвязи между узлами социальной сети постоянно перестраиваются, узлы и связи в ней появляются и исчезают. Любой полученный граф реальной сети лишь «моментальный снимок» этой сети, а через мгновение социальная сеть хоть чуть-чуть меняется и ее граф уже отличается от ранее полученного. Поэтому исследования графов сетей носят ретроспективный характер, однако изучение прошлого дает материал для построения прогнозов на будущее.

Было бы полезно получить инструментарий для ответов, например, на такие вопросы: какой узел нужно выбрать, чтобы информация максимально быстро разошлась по сети; как быстро информация разойдется по сети, если для ее запуска выбрать произвольный узел; за какое время информацией может быть охвачена половина (или некоторая часть) социальной сети. Проведенное авторами исследование – первая часть работ в этом направлении.

### Постановка задачи

В ходе подготовки работы [1] был получен граф подсети пользователей социальной сети «ВКонтакте», указавших Омский государственный технический университет (ОмГТУ) своим местом обучения. Для анализа этой сети [2] было разработано специальное приложение, запрашивающее данные через API (интерфейс прикладного программирования). Узлами исследуемой подсети «ВКонтакте» являются пользователи, а связи выражают отношение «дружбы» между ними. Использование API существенно упрощает построение графа сети по сравнению с возможностями разбора веб-страниц. Будем далее называть его графом ОмГТУ.

В настоящей статье поставлена цель – выполнить исследование процессов распространения информации в этом графе с использованием методов имитационного моделирования. Исследование процессов распространения информации предваряет построение распределения степеней связности (РСС) вершин графа, поскольку РСС вершин является одной из самых важных характеристик, определяющих свойства графов.

Начиная исследование процессов распространения информации, авторам важно было убедиться, что выбранный граф имеет РСС, сходное с РСС растущих случайных графов с предпочтительным связыванием, предложенных известными зарубежными и отечественными исследователями [3-6]. Обзор основных правил построения наиболее известных графов сделан, например, в работах [7, 8].

Граф ОмГТУ на момент построения имел 9927 вершин и 35596 ребер. Построенное в ходе исследования РСС имеет вид, представленный на рис. 1. На графике в логарифмических шкалах по оси абсцисс показаны степени связности  $k$  вершин, а по оси ординат – соответствующие вероятности  $Q_k$  того, что случайно выбранная в графе вершина имеет степень связности  $k$ .

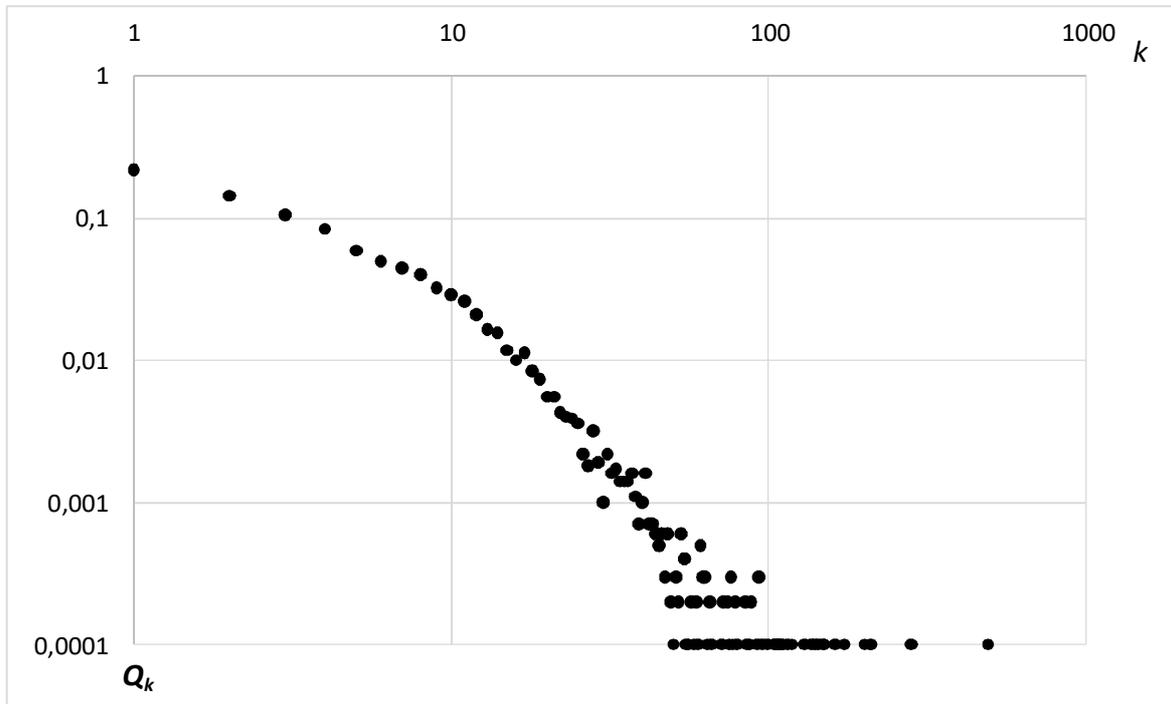


Рис. 1. РСС графа ОмГТУ

РСС вершин нескольких больших сетей представлено на рис. 2 [9]. На нем круглыми маркерами показано РСС сети, а прямой линией – РСС графа Барабаш-Альберт [3], являющегося широко известной моделью Интернета. На рис. 2 показана минимальная степень связности  $m$  соответствующего графа Барабаш-Альберт.

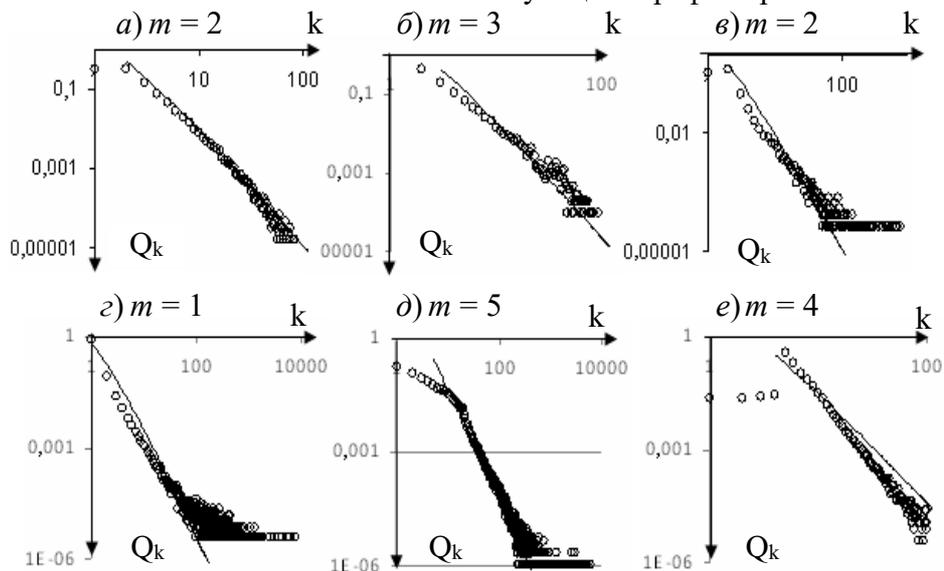


Рис. 2. РСС узлов реальных сетей

- а) сеть маршрутизаторов Интернет;
- б) сеть пользователей программы PGP;
- в) сеть автономных систем Интернет;

- г) сеть адресов электронной почты;
- д) сеть ссылок веб-страниц;
- е) сеть товаров-комплементов Интернет-магазина Amazon.

Как видно из рисунков, граф ОмГТУ имеет РСС вершин, близкое к РСС других сетевых структур, можно предполагать, что результаты, полученные для графа социальной сети ВКонтакте ОмГТУ можно обобщить и на другие большие сетевые структуры.

#### **Организация экспериментов**

Исследование процессов распространения информации в графе ОмГТУ проводилось методом имитационного моделирования по следующей схеме.

1. В среду имитационного моделирования в момент модельного времени  $t = 0$  загружается файл графа, полученный в формате электронных таблиц – на листе электронной таблицы по строкам в два столбца для каждого ребра графа указаны номера вершин, которые соединяет это ребро.

2. В момент времени  $t = 1$  генерируется новая вершина графа с единственным ребром, которое своим свободным концом присоединяется к указанной в ходе эксперимента вершине графа.

3. Имитируется передача информации от новой вершины к выбранной.

4. В течение 10 шагов модельного времени каждая вершина графа, получившая информацию, воспринимает ее и с вероятностью 0,2 передает всем вершинам, с которыми она связана.

5. По окончании 10 шагов времени определяется, сколько вершин получили распространяемую информацию, и рассчитывается доля вершин графа ОмГТУ, получивших запущенную в сеть информацию.

По предложенной схеме работает механизм вирусного маркетинга, когда в сеть запускается нетривиальное сообщение (вирус), которое получивший его человек может переслать своим друзьям. В результате происходит распространение информации по сети, охватывается широкий круг заранее неопределенных лиц, а скрытая реклама, содержащаяся в сообщении, доходит до очень большой аудитории. Система вирусного маркетинга рассмотрена, например, в работе [10].

#### **Результаты моделирования**

В графе ОмГТУ определены степени связности всех вершин на момент построения графа. В частности, было установлено, что из 9927 вершин наиболее высокосвязными являются следующие вершины, представленные в таблице 1, остальные имеют меньшую степень связности.

Таблица 1. Степени связности вершин графа ОмГТУ в порядке убывания

№ п/п	Номер вершины в графе	Степень связности	№ п/п	Номер вершины в графе	Степень связности
1	3165	491	12	6879	136
2	231	281	13	4703	129
3	4882	210	14	165	118
4	2863	208	15	5738	115
5	3902	200	16	8963	111
6	216	173	17	6956	109
7	1115	162	18	5404	108
8	8883	149	19	6559	107
9	9508	148	20	9355	105
10	1212	143	21	2523	104
11	4611	139	22	3652	99

Для участия в эксперименте были выбраны четыре вершины, выделенные в таблице серым цветом – две самые высокосвязные (степени 491 и 281), со степенью связности 200 и со степенью 99 (вершины со степенью связности 100 в графе нет).

Через каждую из этих выбранных вершин в граф запускалось сообщение по предложенной выше схеме, и этот имитационный эксперимент выполнялся несколько сотен раз. Поскольку вероятность передачи полученного сообщения была задана как 20%, то практически в 80% случаев распространение информации от первой вершины не происходило. В оставшихся 20% случаев происходил первый шаг передачи информации и далее она распространялась по графу. Понятно, что вероятностный характер факта передачи информации в каждом эксперименте приводит к различному результату. Примером того, как в отдельных экспериментах происходил охват вершин графа распространяемой информацией, является таблица 2.

Таблица 2. Некоторые результаты имитационных экспериментов

Номер эксперимента	Количество вершин графа ОмГТУ, не получивших информации	Количество вершин графа ОмГТУ, получивших информации
1	5186	4742
2	1732	8196
3	1748	8180
4	5634	4294
5	1202	8726
6	2695	7233
7	1988	7940
8	1609	8319
9	1570	8358
10	8151	1777

После обработки информации, полученной в ходе имитационных экспериментов, фрагмент которой представлен выше, получаем окончательные результаты распространения информации в графе ОмГТУ, представленные в таблице 3.

Таблица 3. Итоги исследования распространения информации в графе ОмГТУ

Информация запущена через вершину	Степень связности вершины	Процент вершин графа ОмГТУ, не получивших информации	Процент вершин графа ОмГТУ, получивших информации
3165	491	26,4	73,6
231	281	28,9	71,1
3902	200	30,0	70,0
3652	99	32,5	67,5

Заключительная часть исследования – сравнение полученных результатов с процессом распространения информации в широко известном случайном графе Барабаши-Альберт. Для этого методом имитационного моделирования с использованием разработанных авторами средств [11], сгенерирован граф Барабаши-Альберт с тем же количеством вершин, что и у графа ОмГТУ. В полученной версии графа у вершины номер 6 максимальная степень связности равная 270. Именно через

эту вершину в граф Барабаши-Альберт запускалась информация. Имитационные эксперименты дали результат – в графе получили информацию 44,2% вершин, а не получили 55,8%.

#### **Выводы и направления дальнейших исследований**

В результате исследования процессов распространения информации в графе реальной социальной сети ОмГТУ получены ожидаемые результаты – чем больше степень связности вершины, через которую в граф запускается информация, тем шире охват вершин, до которых по сети доходит запускаемая информация. Однако в графе имеются «особые» вершины, обладающие свойством центральности в различном смысле. Различные варианты понятия центральности рассмотрены, например, в источнике [12]. Запуск информации в граф через такие вершины может оказаться более эффективным с точки зрения распространения информации. Поэтому первым этапом предстоящих исследований является задача поиска «особых» вершин с последующим запуском информации в граф через них.

Вторым этапом является этап проведения имитационных экспериментов с запуском информации в граф через «особые» вершины.

Третий этап – разработка аналитических методов оценки процессов распространения информации и проверка их результатов с использованием имитационного моделирования.

Вторым направлением исследований является исследование того, как процессы распространения информации зависят от внутренних свойств графа. Как показали выполненные исследования, у графа Барабаши-Альберт процент охвата вершин графа значительно меньше, чем в графе ОмГТУ при одинаковом числе вершин. Возможно, что на охват вершин влияет коэффициент кластеризации или PCC вершин графа.

#### **Литература**

1. **Юдина М. Н.** Анализ выборки пользователей из социальной сети «ВКонтакте» / Ниткин Д. А., Юдина М. Н. // Информационные технологии и автоматизация управления: материалы VIII Всероссийской научно-практической конференции студентов, аспирантов, работников образования и промышленности / ОмГТУ. Омск, 2016. С. 112-116.
2. **Юдина, М. Н.** Исследование подсети обучавшихся (обучающихся) в ОмГТУ лиц в социальной сети «ВКонтакте» / М. Н. Юдина, А. С. Бондаренко // Информационные технологии и автоматизация управления: материалы VIII Всероссийской научно-практической конференции студентов, аспирантов, работников образования и промышленности / ОмГТУ. Омск, 2016. С. 173-178.
3. **Barabasi A. L., Albert R.** Emergence of scaling in random networks // Science. 1999. V.286. P. 509-512.
4. **Krapivsky P. L., Redner S.** Organization of growing random networks // Phys. Reviews E. 2001. V.63. P.066123.
5. **Задорожный В. Н.** Случайные графы с нелинейным правилом предпочтительного связывания / В.Н. Задорожный // Проблемы управления. 2010. № 6. С. 2-11.
6. **Bollobas B., Borgs C., Chayes J., Riordan O.** Directed scale-free graphs. // In Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms (Baltimore, MD, 2003), P. 132-139, New York, (2003). ACM.
7. **Бадрызлов В.А.** Классификация случайных графов с предпочтительным связыванием / В. А. Бадрызлов // Омский научный вестник. 2017. №4 (154). С. 124-128.
8. **Щербакова Н. Г.** Модели сетей с предпочтительным присоединением / Н. Г. Щербакова // Проблемы информатики. 2019. №3 (44). С.46-61.

9. **Юдин Е. Б.** Генерация случайных графов предпочтительного связывания // Омский научный вестник, 2010. № 2 (90). С. 188-192.
10. **Годин С.** Идея-вирус? Эпидемия! Заставьте клиентов работать на ваш сбыт. – СПб.: Питер. 2005.
11. **Бадрызлов В.А.** Генерация случайного графа с предпочтительным связыванием v.1» / В.А. Бадрызлов // Хроники объединенного фонда электронных ресурсов «Наука и образование». 2017. №10 (101). С.12.
12. Центральность [Электронный ресурс]: Википедия. Свободная энциклопедия. – URL: <https://ru.wikipedia.org/wiki/центральность> (дата обращения: 06.10.2021).