

УДК: 519.1+538.911

Социальные сети

И. А. Евин^{1,а}, Т. Ф. Хабибуллин²

¹Учреждение Российской академии наук Институт машиноведения им. А. А. Благонравова РАН,
Россия, 101990, г. Москва, Малый Харитоньевский переулок, д. 4

²Московский физико-технический институт (государственный университет),
Россия, 141700, г. Долгопрудный, Институтский переулок, д. 9

E-mail: ^a yevin@list.ru

Получено 31 марта 2012 г.,
после доработки 2 мая 2012 г.

В статье дан обзор основных результатов изучения как реальных социальных сетей (сетей сотрудничества ученых и актеров, сетей цитирования научных публикаций, сетей друзей и знакомых и т. д.), так и современных онлайн-социальных сетей (Twitter, Facebook и т. д.) с точки зрения теории сложных сетей. На основе собственных исследований авторами выявлены особенности восприятия некоторых сложных сетей.

Ключевые слова: безмасштабные сети, ассортативность, онлайн-социальные сети

Social networks

I. A. Yevin¹, T. F. Khabibullin²

¹*Mechanical Engineering Institute, 4 Maliy Charitonjevskiy pereulok, Moscow, 101909, Russia*

²*Moscow Institute of Physics and Technology, 9 Institutskiy per, Dolgoprudny, 141700, Russia*

Abstract. — The paper reviews the main results of the study of real social networks (networks of collaboration between scientists and actors, networks of citation of scientific publications, networks of friends and acquaintances, etc.) and modern online social networks (Twitter, Facebook etc.) from the complex networks theory standpoint. Based on original research by the authors, it reveals peculiarities of perception of certain complex networks.

Keywords: scale-free networks, assortativity, online social networks

Citation: *Computer Research and Modeling*, 2012, vol. 4, no. 2, pp. 423–430 (Russian).

Введение

Многие сложные системы можно изучать, рассматривая их как сетевые структуры, образованные взаимодействием элементов этих систем между собой. Для большинства построенных таким образом сетей распределение вероятности, что данный узел имеет число связей, равное q , описывается формулой:

$$P(q) \propto (q + q_0)^{-\gamma} \exp(q/q_1), \quad (1)$$

где γ — показатель степени (обычно $2 \leq \gamma \leq 3$), q_0 и q_1 — константы, которые отражают соответственно явления насыщения при малых значениях и обрезания (cut-off) при больших значениях q . В случае ориентированных сетей распределения вида (1) описывают отдельно числа входных и выходных связи узлов. Сети с таким законом распределения стали называть безмасштабными или сложными сетями [Евин, 2010; Newman, 2010].

Сложные сети традиционно разделяются на социальные (сети сотрудничества ученых и актеров кино, сети знакомств, сети цитирования, сети телефонной связи, сети электронной почты и многие другие), биологические (сети белковых взаимодействий и биологических реакций, экологические сети) и технологические сети (Интернет, Всемирная паутина, сети электростанций и прочие). При этом социальные сети принципиально отличаются от биологических и технологических сетей характером корреляции узлов: социальным сетям свойственна тенденция связывания между собой хабов (узлов с большим числом связей), в то время как для биологических и технологических сетей свойственно связывание хабов с узлами, имеющими малое число связей. Для количественного описания характера корреляции узлов в теории сложных сетей используется понятие ассортативности.

Термин «ассортативное спаривание» (assortative mating), то есть подбор подобного к подобному, возник в биологических исследованиях, когда при селекции исследовалось, по каким параметрам и признакам происходит подбор пар при спаривании. Количественно понятие ассортативности описывается коэффициентом Пирсона:

$$r = \frac{L \sum_{i=1}^L j_i k_i - \left[\sum_{i=1}^L j_i \right]^2}{L \sum_{i=1}^L j_i^2 - \left[\sum_{i=1}^L j_i \right]^2}. \quad (2)$$

Здесь L — число связей в сети, а j_i и k_i — число связей у узлов на обоих концах связи. Если узлы с большим числом связей (хабы) связаны друг с другом, то $r \approx 1$. Если узлы с большим числом связей связаны с узлами с небольшим числом связей, то $r \approx -1$. Для технологических и биологических сетей свойственна отрицательная ассортативность [Newman, 2003], в социальных сетях имеет место положительная ассортативность, что является следствием того, что в социальных системах действует принцип «подобное соединяется с подобным». Например, супруги в браке обычно приблизительно одного возраста, уровня образования, этноса и так далее. Это проявляется также в том, что формируются сообщества (кластеры) людей, связанных по увлечениям, интересам (коллекционеры, книголюбы) и профессиям. Знаменитые люди обычно знакомы друг с другом. Для богатых эта закономерность известна как «явление элитарного клуба» (rich club phenomenon). На языке теории сетей это как раз и означает наличие связей между узлами с наибольшим числом связей.

В таблице 1 приведены значения коэффициента Пирсона (коэффициента ассортативности) для реальных сетей различной природы.

Из всех типов сетей социальные сети имеют наиболее длительную историю изучения [Jackson, 2008; Watts, 2004]. Именно в сетях цитирования научных публикаций в 1965 году Д. Прайсом был впервые эмпирически обнаружен степенной закон распределения узлов по числу

Таблица 1. Ассортативность социальных, технологических и биологических сетей. Адаптировано из [Newman, 2003]

Тип сети	Сеть	Размер сети	Ассортативность, r
Социальные	соавторов по физике	52 909	0,363
	соавторов по биологии	1 520 251	0,127
	соавторов по математике	253 339	0,120
	сотрудничества актеров кино	449 913	0,208
	директоров компаний	7 673	0,276
	адресов электронной почты	16 881	0,092
Технологические	сеть электростанций	4 941	-0,003
	Интернет	10 697	-0,189
	«Всемирная паутина» (WWW)	269 504	-0,067
Биологические	взаимодействий белков	2 115	-0,156
	метаболическая сеть	765	-0,240
	нейронная сеть	307	-0,226
	морская пищевая сеть	134	-0,263
	пресноводная пищевая сеть	92	-0,326
	сеть сообщества дельфинов	62	-0,044

связей [Price, Derek, 1965]. Открытие в социальных сетях явления «тесного мира» С. Милграмом стало решающим фактором развития современной теории сложных сетей [Milgram, 1967].

Интересные и важные явления и закономерности были обнаружены и исследованы в социальных сетях друзей и знакомств. Оказалось, например, что в сетях друзей действует закон трех степеней влияния: наше влияние распространяется только на наших друзей и друзей наших друзей. На следующем шаге это влияние уже ничтожно мало [Christakis, Fowler, 2009]. Обратное также справедливо: наибольшее влияние на нас оказывают наши друзья и друзья наших друзей.

Еще на ранних этапах развития теории сложных сетей были детально исследованы законы распространения инфекционных заболеваний в социальных сетях, в том числе в сетях друзей и знакомых. Исследования последних лет показали, что аналогичным образом распространяются среди знакомых и друзей хорошее настроение и депрессия, алкоголизм и курение, ожирение и даже суицидальное поведение [Christakis, Fowler, 2009]. Остановимся подробнее на таком важном для политологии вопросе, как поведение избирателей в социальных сетях.

Политические предпочтения в социальных сетях

Первые исследования, показавшие важность сетевой структуры друзей и знакомых на поведение избирателей, были проведены в США в 1940-х годах [Lazarsfeld, Berelson, Gaudet, 1944]. В этих исследованиях было показано, что решения людей о том, за кого голосовать на выборах, определяются главным образом не средствами массовой информации, а так называемыми «лидерами общественного мнения». Именно эти лидеры перерабатывают имеющуюся информацию, в том числе поступающую из газет, радио, телевидения, и транслируют ее своим друзьям, знакомым и членам их семей, которые мало интересуются политикой.

Исследования, проведенные в шестидесятых–восьмидесятых годах прошлого века, выявили кластеризацию в социальных сетях на основе политических предпочтений [Huckfeldt, Sprague, 1995]. Например, в США демократы в качестве друзей чаще всего выбирают себе демократов, а республиканцы — республиканцев. Те люди, которые ходят на выборы, обсуждают политические вопросы также с теми, кто ходит на избирательные участки, и так далее.

Д. Фаулер с помощью построенной им компьютерной модели исследовал возможность появления лавинообразных процессов в социальных сетях во время голосования [Fowler, 2005]. Оказалось, что в некоторых случаях один избиратель может инициировать до сотни других избирателей пойти проголосовать в день выборов, хотя обычно в социальных сетях один человек имеет в среднем три–четыре связи с ближайшими соседями.

В 2006 году Д. Никерсоном было показано, что многие предсказанные на компьютерных моделях эффекты — это реальные явления в современной политической жизни некоторых стран. Например, исследование, проведенное в ходе избирательной компании в городе Денвер (штат Миннеаполис, США), показало, что решение принять участие в голосовании одного из избирателей вызвало приход на избирательные участки до тридцати дополнительных избирателей [Nickerson, 2008].

Некоторые тенденции в исследованиях онлайн-социальных сетей

В последние годы, наряду с онлайн-социальными сетями первого поколения (сети электронной почты или сети Web 1.0), получили очень широкое распространение социальные сети второго поколения (сети Web 2.0), такие как МайСпейс (MySpace), Фейсбук (Facebook), Твиттер (Twitter), Живой Журнал (Live-Journal), «ВКонтакте», «Одноклассники», «МойКруг» и другие.

Данные о пользователях таких сетей представляют уникальные возможности для разнообразных исследований поведения общества в широких временных и географических масштабах. Особенно это свойственно социальной сети Твиттер, в которой 91% пользователей делает свой профиль и историю своих коммуникационных связей публично открытыми. Исследователи стали изучать содержание твитов (коротких сообщений, состоящих не более чем из 140 знаков) для количественного анализа и предсказания некоторых процессов. Например, стали предсказывать кассовые сборы от проката кинокартин [Asur, Huberman, 2010], курсы акций на фондовом рынке [Bollen, Mao, Zeng, 2011], результаты региональных выборов [O'Connor et al., 2010], влияние погоды на настроение [Hannak et al., 2012] и так далее.

Остановимся подробнее на исследовании, целью которого было выяснение того, представляют ли пользователи Твиттера репрезентативную выборку всего общества [Mislove, 2011]. В этом исследовании было проанализировано более 1,7 миллиарда сообщений, посланных почти 5,5 миллионами пользователей с марта 2009 г. по август 2009 г. Среди этих пользователей было отобрано около 3 миллионов жителей США, что составляет примерно 1% всего населения этой страны.

Была разработана методика исследования отобранных пользователей Твиттера по трем основным параметрам.

1. Было проведено сравнение географического распределения пользователей со всем населением США. Для этого использовалось поле «местоположение» («location») в профилях пользователей. Показано, что пользователи Твиттера в основном живут в областях с высокой плотностью населения (больших городах и их пригородах), а малонаселенные регионы представлены недостаточно полно.
2. По имени пользователей определялась их гендерная принадлежность и было выявлено значительное доминирование мужчин, которое, впрочем, уменьшалось во времени.
3. По фамилии была исследована расово-этническая принадлежность пользователей и была выявлена сильная корреляция этого признака с географическим местоположением.

На основе этих данных были сделаны выводы, как должны быть скорректированы самые разнообразные исследования пользователей Твиттера, чтобы они были репрезентативными, то есть отражали структуру всего населения США.

Многие эксперты объясняют победу Барака Обамы на президентских выборах широким использованием интернет-технологий, например YouTube, а также технологий, связанных с мобильными телефонами. В последние годы в политическую жизнь многих стран активно вовлекаются онлайн-социальные сети Facebook, Twitter и другие.

Исследование сетевой структуры блогосферы США [Adamic, Glance, 2005] с использованием современных алгоритмов нахождения сообществ выявило сильную кластеризацию таких сетей, отражающую, прежде всего, политические предпочтения. Аналогичные исследования были проведены и для некоторых других стран. Первой такой страной стал Иран, для которого ежедневно в течение семи месяцев собирали информацию с нескольких миллионов блогов [Kelly, Etlig, 2008]. Ожидалось, что политический контроль в этой стране должен сделать структуру блогосферы более однородной, чем в демократических странах. Однако полученные данные показали, что эта структура также гетерогенна. В блогосфере Ирана, помимо политических кластеров, связанных с нынешним президентом Ахмадинижадом и экс-президентом Хатами, были выявлены неполитические сообщества, например любителей персидской литературы и поэзии, поклонников различных знаменитостей и тому подобные [Kelly, Etlig, 2008]. Недавние события в Тунисе, Египте и других странах показали растущее влияние онлайн-социальных сетей на динамику политических процессов.

Несомненно, изучение структурных и динамических свойств онлайн-социальных сетей чрезвычайно актуально, а обзор текущего состояния этих исследований заслуживает отдельной публикации [Mislove, 2009].

Особенности восприятия некоторых сложных сетей

К какому из перечисленных выше типов сетей (технологические, биологические, социальные) следует отнести сети общественного транспорта: автобусов, троллейбусов, метро? Из множества опрошенных нами людей подавляющее большинство относило их к технологическим сетям. Между тем проведенные нами исследования показывают, что сети общественного транспорта имеют положительную ассортативность, следовательно, их структура ближе к структуре социальных сетей.

На рис. 1 показана сеть троллейбусов г. Москвы. Узлами этой сети являются маршруты троллейбусов. Два узла соединены связью, если соответствующие им маршруты имеют хотя бы одну общую остановку.

Была также построена сеть автобусных маршрутов г. Москвы. Число узлов сети равно 645, ассортативность — 0.314. Нет никакого сомнения, что сеть московского метро также имеет положительную ассортативность, поскольку многие станции с наибольшими количествами пересадок на другие линии связаны напрямую между собой.

Персонажи литературных романов образуют сети друзей и знакомых. Казалось бы, естественно предположить, что такие сетевые структуры являются социальными сетями. В таблице 2 представлены данные по числу персонажей и показателю ассортативности пяти известных литературных произведений. Как мы видим, ни одну из сетевых структур взаимосвязи персонажей этих произведений по характеру корреляции узлов нельзя отнести к социальным сетям.

Эти сети принадлежат к особому типу, который мы назвали «когнитивные сети» [Евин и др., 2011]. К ним относятся также сетевые структуры музыкальных произведений, живописи и тому подобные. Эти структуры придуманы авторами соответствующих произведений, и они полностью или частично хранятся в памяти людей, прочитавших роман, прослушавших музыку и так далее. Более того, именно такого рода сети, отражающие связи слов, понятий, фонем, музыкальных звуков, участвуют в формировании той области сознания человека, которая отличает его от животных.

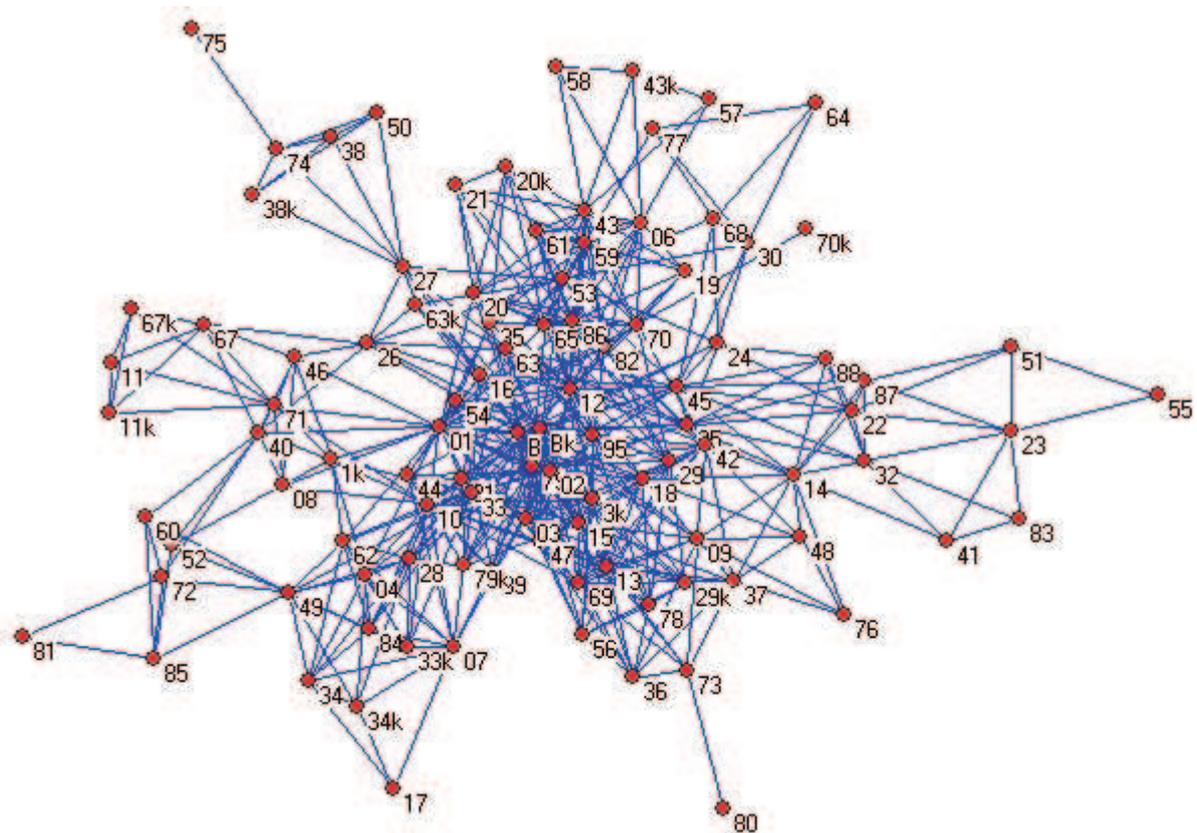


Рис. 1. Сеть троллейбусных маршрутов г. Москвы. У этой сети число узлов — 102, ассортативность — 0.287

Таблица 2. Сравнительные характеристики сетевых структур персонажей некоторых произведений мировой литературы

Автор и название произведения	Число персонажей	Ассортативность, r
Л. Н. Толстой «Анна Каренина»	138	-0.350
М. Твен «Гекельберри Финн»	74	-0.173
Ч. Диккенс «Дэвид Копперфильд»	87	-0.258
Гомер «Илиада»	561	-0.118
В. Гюго «Отверженные»	80	-0.165

Хотя взаимодействие некоторых персонажей Нового Завета основано на реальных событиях, их сетевая структура не образует социальную сеть. Число узлов этой сети равно 80, ассортативность имеет отрицательное значение: -0.133 . Эту сеть также следует отнести к когнитивным сетям, поскольку ее структура (или ее отдельные фрагменты) содержится в памяти миллионов людей.

Заключение

Более чем десятилетний опыт исследования социальных и других сложных систем методами теории сложных сетей показывает, что получаемые результаты зачастую оказываются

более полными и содержательными, чем при использовании многих других традиционных подходов, поскольку свойства сети дают информацию о всей соответствующей сложной системе в целом. Хотя некоторые исследователи называют теорию сложных сетей основной научной парадигмой XXI века [Newman, 2008], всегда следует иметь в виду, что само представление сложной системы в виде сетевой структуры есть неизбежная дань традиционной парадигме редукционизма. Только разумное сочетание этих двух подходов может выдвинуть изучение сложных процессов в экономике, политике, культуре и тому подобных на качественно новый уровень.

Список литературы

- Евин И. А.* Введение с теорию сложных сетей // Компьютерные исследования и моделирование. — 2010. — Т. 2, № 2. — С. 121–141.
- Евин И. А., Кобляков А. А., Савриков Д. В., Шувалов Н. Д.* Когнитивные сети // Компьютерные исследования и моделирование. — 2011. — Т. 3, № 3. — С. 231–239.
- Adamic L. A., Glance N.* The political Blogosphere and the 2004 U.S. Election: Divided They Blog // Proceedings of the 3rd International Workshop on Link Discovery — New York: Association for Computing Machinery, 2005. — P. 36–43.
- Asur S., Huberman B.* Predicting the future with social media // arXiv:1003.5699v1 — 2010.
- Bollen J., Mao H., Zeng X.-J.* Twitter mood predicts the stock market // In Proceedings of the 6th International AAAI Conference on Weblogs and Social Media (ICWSM'12) — Dublin — 2012.
- Christakis N. A., Fowler J. H.* Connected: The Surprising Power of Our Social Networks and How They Shape Our Lives. — New York: Back Bay Books, 2009.
- Corbett P.* Facebook demographics and statistics report 2010 // <http://www.istrategylabs.com/2010/01/facebook-demographics-and-statistics-report-2010-145-growth-in-1-year/>
- Fowler J. H.* Turnout in a Small World, in: The Social Logic of Politics: Personal Networks as Contexts for Political Behaviour. Edited by A. Zuckerman — Philadelphia: Temple University Press, 2005 — P. 269–287.
- Hannak A., Anderson E., Feldman-Barrett L., Lehmann S., Mislove A., Riedewald M.* Tweetin' in the Rain: Exploring societal-scale effects of weather on mood // In Proceedings of the 6th International AAAI Conference on Weblogs and Social Media (ICWSM'12). — Dublin, 2012.
- Huckfeldt R., Sprague J.* Citizens, Parties, and Social Communication. — New York: Cambridge University Press, 1995.
- Jackson M. O.* Social and Economic Networks. — Princeton University Press, 2008.
- Kelly J., Etlig B.* Mapping Iran's Online Public: Politics and Culture in the Persian Blogosphere // Berkman Center Research Publication 2008-01. — 2008. — P. 1–36.
- Lazarfeld P. F., Berelson B., Gaudet H.* The People's Choice. — New York: Columbia University, 1944.
- Milgram S.* The small world problem // Psychology Today. — 1967. — Т. 2, № 2. — P. 60–67.

- Mislove A., Lehmann S., Ahn Y., Onnela J., and Rosenquist J. N.* Understanding the Demographics of Twitter Users // In Proceedings of the 5th International AAAI Conference on Weblogs and Social Media (ICWSM'11). — Barcelona, 2011.
- Mislove A.* Online Social Networks: Measurement, Analysis, and Applications to Distributed Information Systems // A Thesis for the Degree Doctor of Philosophy. — Houston: Rice University, 2009.
- Newman M. E. J.* The Physics of Networks // *Physics Today*. — November, 2008 — P. 33–38.
- Newman M. E. J.* Networks. An Introduction. — Oxford University Press, 2008.
- Newman M. E. J.* Mixing patterns in networks // *Phys. Rev. E* 67, 026126; [arXiv:cond-mat/0209450v2](https://arxiv.org/abs/cond-mat/0209450v2) — 2003.
- Nickerson D. W.* Is Voting Contagious? Evidence from Two Field Experiments // *American Political Science Review*. — 2008. — V. 102. — P. 49–57.
- O'Connor B., Balasubramanyan R., Routledge B., Smith N.* From tweets to polls: Linking text sentiment to public opinion time series // In Proceedings of the 6th International AAAI Conference on Weblogs and Social Media (ICWSM'10). — Washington, DC, 2010.
- de Solla Price, Derek J.* Networks of Scientific Papers // *Science*. — 1965. — V. 149, № 3683. — P. 510–515.
- Strayhorn T.* Sex Differences in Use of Facebook and MySpace among First-Year College Students // *Student Affairs*. — 2009. — V. 10, № 2.
- Watts D.* Six degrees: The Science of a Connected Age. — New York: W. W. Norton & Company, 2004.