

## К ВОПРОСУ О МОДЕЛИРОВАНИИ ТОПОЛОГИИ СОЦИАЛЬНОЙ СЕТИ

К. Г. Абрамов, Ю. М. Монахов, И. Ю. Бодров (Владимир)

В настоящее время можно с уверенностью утверждать, что тенденции развития технологий web 2.0 свидетельствуют об угрозе распространения нежелательной информации посредством социальных сетей. Для эффективного противодействия этой угрозе необходимо разрабатывать механизмы анализа и прогнозирования подобных процессов [1]. Для решения таких задач используется имитационное моделирование, но для того, чтобы его проводить, необходимо разобраться с первоочередным вопросом, касающимся природы социальных сетей. Чтобы начать моделирование, научиться генерировать подобные сети и решать другие задачи, следует определить структуру современных социальных сетей. В рамках общей задачи моделирования распространения нежелательной информации в социальных сетях встает научно-техническая подзадача определения топологии реальной социальной сети. Научная новизна работы заключается в том, что в ходе исследования был определен характер распределения степеней связности, отличающийся от результатов, полученных другими учеными. Обзор работ в данной области показывает, что на сегодняшний день нет единого мнения по этой проблеме. Рассмотрим одну из классификаций сетей, представленную на рис. 1.

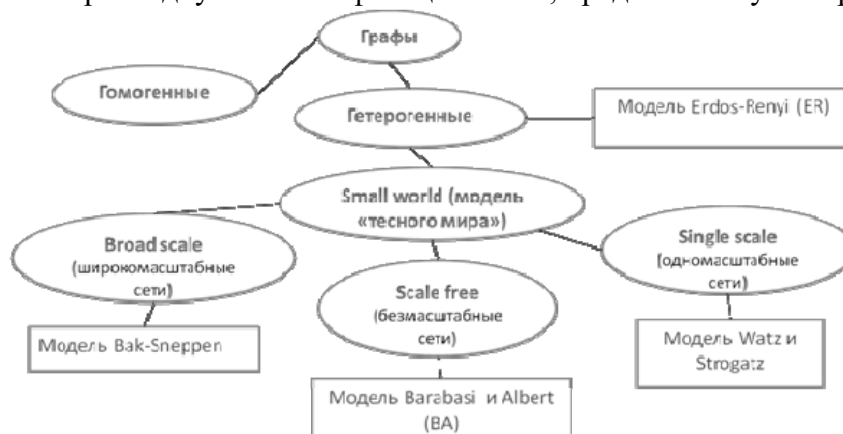


Рис. 1. Классификация сетей

Существуют две основные точки зрения на вопрос об отнесении топологии социальной сети к определенному типу. Pastor-Satorras и Vespignani в своих многочисленных работах [2–8] утверждают, что социальные сети представляют собой Scale-Free сети. Этой же точки зрения придерживается и Barabasi [9]. Newman в своей работе [10] и в работах в соавторстве [11, 12, 13] приводит доводы в пользу того, что социальные сети представляют собой small-world сети. Для классификации сетей в основном используют три характеристики: распределение степени связности, кластерный коэффициент и средняя длина пути (таблица).

При рассмотрении данного вопроса исследователи, как правило, опускают подробности об исходных данных сети. Однако в ряде последних исследований [15] этому уделяется большое внимание.

Необходимо понимать, что задача получения участка реальной социальной сети для исследования уже сама по себе является нетривиальной, что связано с рядом проблем. Исходные данные, доступные для исследования, не обладают необходимой полнотой, так как являются случайной выборкой с точки зрения топологии исходной сети. В выборке присутствуют узлы двух типов: открытые для исследования их положения в сетевом графе («открытые») и закрытые для исследования, сведения о местоположении в сети которых прямо получены быть не могут («закрытые»). Следовательно, встает

подзадача выбора из множества исходных данных репрезентативного подмножества, отражающего в большей мере топологию исходной сети. Сложность данной задачи состоит в невозможности выбрать подграф, полностью состоящий из открытых узлов, достаточно большого диаметра, чтобы его можно было считать репрезентативным. Причина проблемы состоит в низкой вероятности нахождения такого подграфа в выборке даже большего размера и является следствием малого отношения количества открытых узлов к закрытым, а также случайности самой выборки. Авторы считают, что именно из-за этих проблем исследователи получают некорректные данные, вследствие чего возникает ошибка с определением типа топологии социальной сети. Для решения этой проблемы авторы разработали собственную модель данных, а также алгоритмы вычисления степени связности и кластерного коэффициента.

### Основные топологические характеристики сетей

Параметр	Small-world	Scale-Free	Случайные графы
Средняя длина пути	$\sim \frac{\ln N}{\ln k}$	$m=1: l \sim \ln N;$ $m \geq 2: l_{BA}^{\alpha > 3} \approx \ln N;$ $l_{BA}^{\alpha = 3} \approx \ln N / \ln \ln N;$ $l_{BA}^{2 < \alpha < 3} \approx \ln \ln N.$	$\sim \frac{\ln N}{\ln \langle k \rangle}$
Кластерный коэффициент	$C_{p \rightarrow 1} \sim k/N,$ $C_p > 0 \gg C_{p \rightarrow 1}$	$C \sim 5 \frac{\langle k \rangle}{N}$	$C \sim \frac{k}{N}$
Распределение степеней вершин	Закон Пуассона: $p(k) \sim e^{-pN} \frac{(pN)^k}{k!}$	Степенной закон: $p(k) \sim k^{-\alpha}$	Закон Пуассона: $p(k) \sim e^{-pN} \frac{(pN)^k}{k!}$

Разработанные механизмы были применены к участку социальной сети «Facebook» размером в 1 млн узлов. Из этих данных была получена репрезентативная выборка из 7200 узлов. В результате ее анализа получены следующие результаты. Из рис. 2 можно увидеть, что полученные данные нельзя аппроксимировать ни пуассоновским, ни степенным распределением. Хорошо подходит логнормальное распределение, описываемое функцией:

$$p(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}. \quad (1)$$

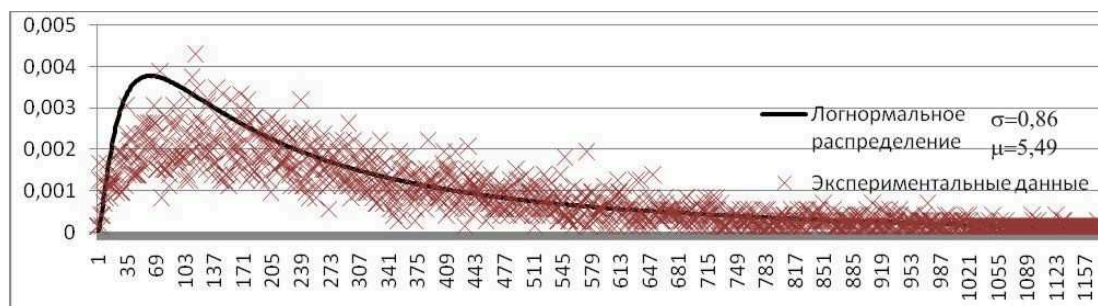


Рис. 2. Гистограмма распределения степеней связности и логнормальное распределение

Для полученных экспериментальных данных с помощью программы Microsoft Excel были подсчитаны математическое ожидание и мода. Далее, используя формулы математического ожидания и моды (формулы (2) и (3) соответственно) для логнормального распределения, были высчитаны значения параметров  $\mu$  и  $\sigma$ . После подстановки этих значения в формулу (1), был получен график логнормального распределения, представленный на рис. 1. Эксперименты проводились на разных выборках различных размеров, картина распределения при этом не изменилась. Средняя степень связности составила 375. Следует отметить, что к выводу о логнормальном распределении приходили и другие исследователи [16]:

$$M = e^{\mu + \sigma^2/2}; \quad (2)$$

$$Mo = e^{\mu - \sigma^2}. \quad (3)$$

В ходе обработки было получено частично репрезентативное подмножество из 50 тысяч открытых узлов, для которого был рассчитан экстраполированный кластерный коэффициент, составивший 0,166139. Это значение практически остается неизменным для участков различных размеров. Исходя из этого, можно говорить о том, что кластерный коэффициент не зависит от размера сети ( $C \sim \text{const}$ ). При построении распределения значений кластерного коэффициента была получена картина, представленная на рис. 3. Результат говорит о том, что связи разрежены и наблюдается низкая кластеризация.

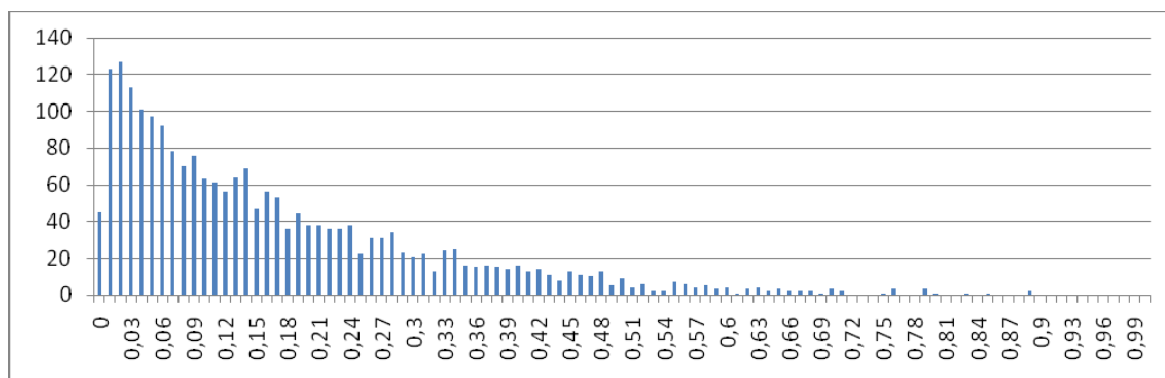


Рис. 3. Распределение кластерного коэффициента

По полученным данным можно говорить о том, что социальные сети нельзя отнести к какому-либо существующему типу сети. Авторы полагают, что в данном случае можно говорить о новом типе сетей, обладающих представленными выше характеристиками. Практическая значимость полученных результатов заключается в том, что они позволяют нам получать более точную картину при имитационном моделировании топологии социальных сетей. В дальнейшем планируется рассчитать среднюю длину пути в социальной сети с целью уточнения параметров ее топологии. Также планируется разработать алгоритмы генерации подобных сетей для того, чтобы проанализировать корреляцию между средней степенью связности и кластерным коэффициентом.

### Литература

1. **Абрамов К. Г., Монахов Ю. М., Никиташенко А. В.** К вопросу об уточнении моделей распространения нежелательной информации в социальных сетях Интернета // Информационные системы и технологии ИСТ-2011: Материалы XVII Международной научно-технической конференции (Н.Новгород, 22 апреля 2011 года).

- Н. Новгород: Электронное издание. 2011.; Гарнитура Times New Roman. Усл. печ. л. 28,0. Уч.-изд. л. 47,6. С.149. ISBN 978-5-9902087-2-8
2. **Pastor-Satorras R., Vespignani A.** Absence of epidemic threshold in scale-free networks with connectivity correlations // *Physical Review Letters*. 2002. Vol. 90, Issue: 2, Publisher: American Physical Society. P. 1–4.
  3. **Pastor-Satorras R., Vespignani A.** Critical load, congestion instabilities in scale-free networks // *Europhys. Lett.* 2002. Vol. 62. P. 292.
  4. **Pastor-Satorras R., Vespignani A.** Emergence of clustering, correlations, communities in a social network model, 2003.
  5. **Pastor-Satorras R., Vespignani A.** Epidemic dynamics, endemic states in complex networks // *Phys. Rev. E*. 2001.
  6. **Pastor-Satorras R., Vespignani A.** Epidemic dynamics in finite size scale-free networks // *Phys. Rev. E*. 2002.
  7. **Pastor-Satorras R., Vespignani A.** Epidemic spreading in correlated complex networks // *Phys. Rev. E. Stat Nonlin Soft Matter Phys.* 2002.
  8. **Pastor-Satorras R., Vespignani A.** Velocity, hierarchical spread of epidemic outbreaks in scale-free networks // *Phys. Rev. Lett.* 2004. Vol. 92. P. 178701.
  9. **Barabasi A-L, Albert R., Jeong H.** Mean-field theory for scale-free random networks // *Physica A*. 1999; 272: 173–189.
  10. **Newman M. E. J.** The structure and function of complex networks // *SIAM Review* (2003) 45. P. 167–256.
  11. **Newman M. E. J., Moore Christopher.** Disease spreading and percolation in small-world networks, Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, New Mexico, 1999.
  12. **Moore C., Newman M. E. J.** Epidemics, percolation in small-world networks // *Phys. Rev. E*. 2000. 61.
  13. **Newman M. E. J., Moore C., Watts D. J.** Mean-field solution of the small-world network model // *Phys. Rev. Lett.*, in press, 2000.
  15. **Жаринов И. В., Крылов В. В.** Конструирование графов с минимальной средней длиной пути // *Вестник Ижевского государственного технического университета*. 2008. № 4. С. 164–169.
  16. **Minas Gjoka, Carter T. Butts, Maciej Kurant, Athina Markopoulou.** Multigraph Sampling of Online Social Networks CoRR abs/1008.2565, 2010.
  17. **Alessandra Sala, Haitao Zheng, Ben Y. Zhao, Sabrina Gaito, Gian Paolo Rossi.** Brief announcement: revisiting the power-law degree distribution for social graph analysis. *PODC 2010*: 400–401