

PHYSICAL REVIEW LETTERS, том 86, выпуск 14, 2 апреля 2001 г.

# Распространение эпидемий в безмасштабных сетях

Ромуальдо Пастор-Саторрас<sup>1</sup> и Алессандро Веспиньяни<sup>2</sup><sup>1</sup>Кафедра физики и ядерной инженерии, Политехнический университет Каталонии, Campus Nord, Mòdul B4, 08034 Барселона, Испания<sup>2</sup>Международный центр теоретической физики им. Абдуса Салама (ICTP), P.O. Box 586, 34100 Триест, Италия (Получено 20 октября 2000 г.)

Интернет обладает весьма сложной структурой связности, которую в последнее время описывают классом безмасштабных сетей. Эта особенность, по-видимому очень эффективная для коммуникационной сети, одновременно способствует распространению компьютерных вирусов. Мы анализируем реальные данные о заражениях компьютерными вирусами и определяем среднее время жизни и персистентность вирусных штаммов в Интернете. Предложена динамическая модель распространения инфекций в безмасштабных сетях; показано отсутствие эпидемического порога и связанного с ним критического поведения. Новая эпидемиологическая схема объясняет данные о компьютерных вирусах и может способствовать пониманию других процессов распространения в коммуникационных и социальных сетях.

DOI: 10.1103/PhysRevLett.86.3200

PACS: 89.75.Hc, 05.50.+q, 05.70.Ln

*Профессиональный перевод с английского. Формулы, нумерация рисунков и библиографические ссылки сохранены по оригиналу.*

Многие социальные, биологические и коммуникационные системы можно адекватно описывать сложными сетями, узлы которых представляют отдельных людей или организации, а связи моделируют взаимодействия между ними [1,2]. Особенно интересными примерами являются Интернет [3,4] и Всемирная паутина [5], которые широко исследовались ввиду их технологической и экономической значимости. Эти исследования, в частности, показали, что вероятность того, что узел таких сетей имеет  $k$  связей, подчиняется безмасштабному распределению  $P(k) \sim k^{-\gamma}$ , где показатель  $\gamma$  лежит в диапазоне от 2 до 3. Наличие узлов с очень большим числом связей (локальная кластеризация) действительно является ключевым элементом моделирования этих сетей в рамках недавно введённых безмасштабных (scale-free, SF) графов [6].

Поскольку сложные сети широко распространены в природе, большой интерес представляет влияние их особенностей на распространение эпидемий и заболеваний [7], а в более общем плане – на неравновесные фазовые переходы, характерные для этих явлений [8]. Исследование эпидемий в таких сетях имеет непосредственное практическое применение при анализе распространения компьютерных вирусов [9,10] и может быть актуально также для эпидемиологии [11] и контроля загрязнений [12].

В настоящей работе мы анализируем данные о реальных эпидемиях компьютерных вирусов и даём их статистическую характеристику, указывающую на необходимость учитывать специфическую топологию безмасштабных сетей в теоретическом описании таких заражений. С этой целью с помощью крупномасштабного моделирования и аналитических методов мы исследуем на SF-графах эпидемиологическую модель SIS (susceptible-infected-susceptible, «восприимчивый-инфицированный-восприимчивый») [11]. Мы обнаруживаем отсутствие эпидемического порога и связанного с ним критического поведения. Это означает, что безмасштабные сети предрасположены к распространению и длительному существованию инфекций при

любой, даже сколь угодно малой, скорости распространения возбудителя. Отсутствие эпидемического порога – стандартного элемента математической эпидемиологии [11] – радикально меняет многие обычные выводы эпидемического моделирования. Полученные результаты имеют значение и для теории фазовых переходов в поглощающее состояние, а также каталитических реакций [8].

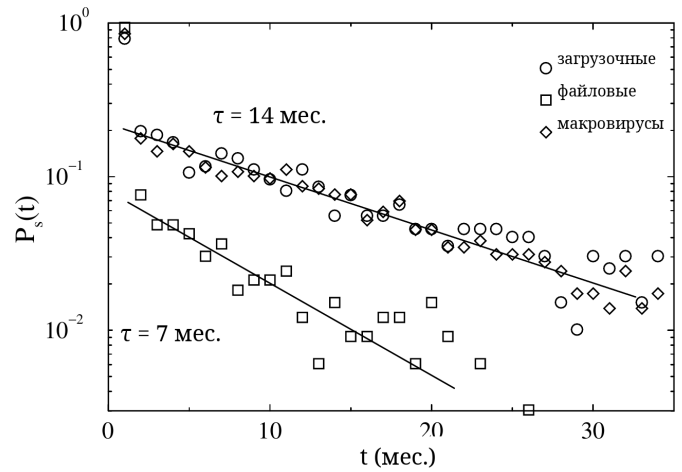
Исследование компьютерных вирусов давно привлекает постоянное внимание специалистов по информатике [10,13–15]; при этом в основном используются подходы, заимствованные из биологической эпидемиологии [11]. Стандартной моделью компьютерной вирусной инфекции служит эпидемиологическая модель SIS. Каждый узел сети представляет отдельную систему, а каждая связь – канал, по которому инфекция может перейти к другим системам. Узел может находиться лишь в одном из двух дискретных состояний: «здоров» или «инфицирован». На каждом временном шаге любой восприимчивый (здоровый) узел заражается со скоростью  $\nu$ , если он связан с одним или несколькими инфицированными узлами. Одновременно инфицированные узлы излечиваются и вновь становятся восприимчивыми со скоростью  $\delta$ ; тем самым определяется эффективная скорость распространения  $\lambda = \nu/\delta$  [16]. Без ограничения общности можно положить  $\delta = 1$ . Модель неявно учитывает наличие антивирусного программного обеспечения, поскольку все инфицированные узлы в конечном счёте возвращаются в восприимчивое состояние. Она соответствует ситуации, в которой пользователи после очистки компьютера не становятся более настороженными по отношению к вирусной инфекции, так что их компьютеры могут быть заражены вновь [15]. Обновление состояния узлов можно выполнять как параллельно, так и последовательно [8].

В моделях с локальной связностью (евклидовы решётки) и на случайных графах наиболее существенным результатом является общий прогноз ненулевого эпидемического порога  $\lambda_c$  [8,11]. Если  $\lambda$  выше порога,  $\lambda \geq \lambda_c$ , инфекция распространяется и ста-

новится устойчивой. Ниже порога, при  $\lambda < \lambda_c$ , инфекция экспоненциально быстро угасает. Эпидемический порог фактически эквивалентен критической точке неравновесного фазового перехода. В данном случае критическая точка отделяет активную фазу со стационарной плотностью инфицированных узлов от фазы, содержащей только здоровые узлы и характеризующейся нулевой активностью. В частности, нетрудно видеть, что модель SIS является обобщением модели контактного процесса, подробно исследованной в контексте фазовых переходов в поглощающее состояние [8]. С другой стороны, статистические наблюдения за вирусными инцидентами в реальных условиях показывают, что все сохраняющиеся вирусы выходят на очень низкий уровень распространённости, затрагивая лишь ничтожную долю общего числа компьютеров [10]. Это резко противоречит теоретическим прогнозам, если не предположить крайне маловероятное: эффективная скорость распространения каждого компьютерного вируса настроена лишь бесконечно мало выше порога. Следовательно, выработанное к настоящему времени представление, хотя и весьма поучительно, не вполне адекватно описывает реальное явление.

Чтобы лучше понять свойства распространения вирусов в реальных условиях, мы проанализировали данные о распространённости, опубликованные *Virus Bulletin* [17] за период с февраля 1996 по март 2000 г., то есть за временное окно в 50 месяцев. В частности, исследовалась вероятность выживания однородных групп вирусов, классифицированных по механизму заражения [9]. Мы учитывали общее число вирусов определённого типа, которые возникли и исчезли внутри выбранного окна наблюдения. Вероятность выживания штамма  $P_s(t)$  определялась как доля вирусов, всё ещё существующих через время  $t$  после своего появления. На рис. 1 видно, что в первые два месяца жизни вируса вероятность выживания резко падает. Это хорошо известная особенность [10,13], означающая, что статистически лишь малая доля вирусов вызывает заметную вспышку в компьютерном сообществе. Вместе с тем при больших временах на рис. 1 отчётливо наблюдается экспоненциальный хвост,  $P_s(t) \sim \exp(-t/\tau)$ , где  $\tau$  – характерное время жизни вирусного штамма [18]. Численная аппроксимация данных даёт  $\tau \simeq 14$  месяцев для загрузочных вирусов и макровирусов и  $\tau \simeq 6-9$  месяцев для файловых вирусов. Значения  $\tau$  сравнительно слабо зависят от выбранного окна наблюдения: анализ вирусов, возникших и исчезнувших за интервал менее 50 месяцев, даёт результаты, совместимые с полной выборкой, хотя из-за меньшей статистики флуктуации возрастают.

Эти характерные времена поразительно велики по сравнению с интервалом, через который антивирусное программное обеспечение появляется на рынке (обычно через несколько дней или недель после первого сообщения об инциденте), и соответствуют возникновению метастабильных эндемических состояний. Столь длительное время жизни в масштабе типичных скоростей распространения и восстановления должно было бы указывать на эффективную скорость распространения, значительно превышающую эпидемический порог. Однако это вновь не согласо-

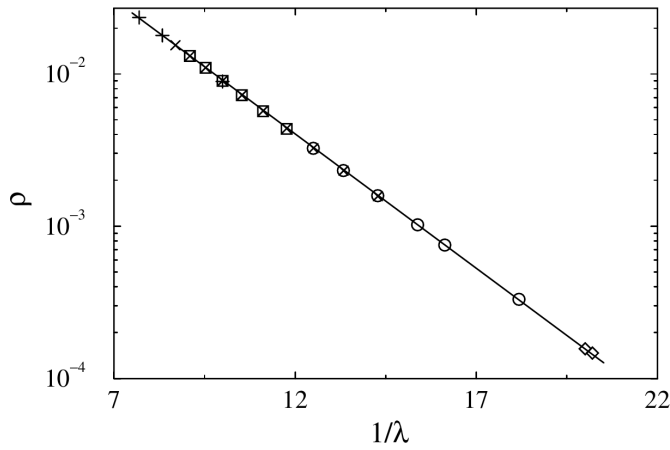


**Рис. 1.** Вероятность выживания вирусов, циркулирующих в реальных условиях. 814 проанализированных вирусов объединены в три основные группы [9]. Файловые вирусы заражают компьютер при запуске инфицированного приложения. Загрузочные вирусы также распространяются через инфицированные приложения, но копируют себя в загрузочный сектор жёсткого диска и поэтому не удаляются перезагрузкой компьютера. Макровирусы заражают файлы данных и потому не зависят от платформы. На графике отчётливо виден экспоненциальный спад с характерным временем  $\tau$ .

ется с неизменно низкими уровнями распространённости компьютерных вирусов.

Ключ к пониманию этих парадоксальных свойств компьютерных вирусов заключается в способности многих из них распространяться посредством обмена данными через коммуникационные протоколы (FTP, электронная почта и т. п.) [10]. Вирусы преимущественно попадают на компьютеры, интенсивно связанные с внешним миром и поэтому обменивающиеся пропорционально большим объёмом данных и информации. Следовательно, естественно считать топологию Интернета той эффективной топологией, на которой происходит распространение. Безмасштабная связность Интернета означает, что каждый узел имеет статистически значимую вероятность обладать очень большим числом связей по сравнению со средней степенью сети  $\langle k \rangle$ . Это противоположно обычным случайным сетям (как локальным, так и нелокальным), в которых каждый узел имеет приблизительно одно и то же число связей,  $k \simeq \langle k \rangle$  [19]. Поэтому безмасштабные свойства необходимо включить в теорию эпидемического распространения компьютерных вирусов.

Чтобы выяснить влияние безмасштабной связности на распространение эпидемий, мы исследуем модель SIS на SF-сетях. В качестве прототипа рассматривается граф, создаваемый алгоритмом из работы [6]. Сначала берётся небольшое число  $m_0$  несвязанных узлов. На каждом временном шаге добавляется новый узел с  $m$  связями; каждая связь присоединяется к уже существующему узлу  $i$  степени  $k_i$  с вероятностью  $k_i / \sum_j k_j$ . После достаточно большого числа итераций получается сеть из  $N$  узлов с распределением степеней  $P(k) \sim k^{-3}$  и средней степенью  $\langle k \rangle = 2m$ . В настоящей работе принято  $m = 3$ . Мы выполнили численное моделирование на графах с числом узлов от  $N = 10^3$  до  $N = 8,5 \times 10^6$  и изучили временную эволюцию и стационарные свойства плотности инфицированных узлов  $\rho$  в сохраняющихся заражени-

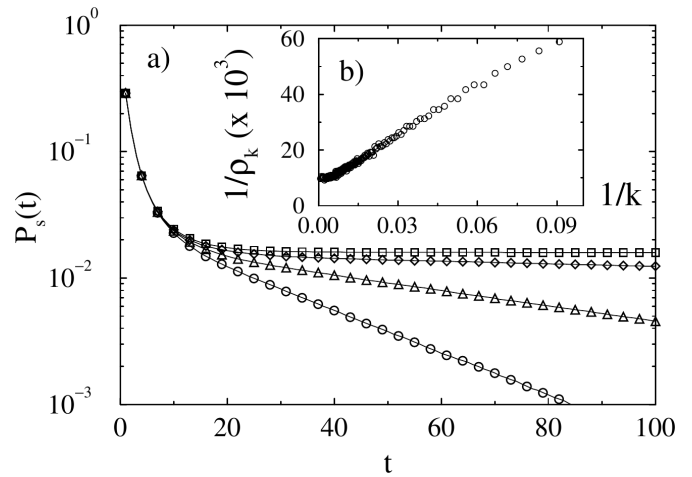


**Рис. 2.** Персистентность  $\rho$  как функция  $1/\lambda$  для различных размеров сети:  $N = 10^5$  (знаки плюс),  $N = 5 \times 10^5$  (квадраты),  $N = 10^6$  (кресты),  $N = 5 \times 10^6$  (кружки) и  $N = 8,5 \times 10^6$  (ромбы). Линейность в полулогарифмическом масштабе подтверждает предсказанную для  $\rho$  экспоненциальную зависимость. Сплошная линия – аппроксимация выражением  $\rho \sim \exp(-C/\lambda)$ .

ях, то есть распространённость вируса. В начальный момент заражалась половина узлов сети, после чего правила SIS-модели итеративно применялись с параллельным обновлением. По окончании начального переходного режима система стабилизировалась в стационарном состоянии с постоянной средней плотностью инфицированных узлов. В этом состоянии узлы заражались повторно без заметной периодичности. Распространённость определялась усреднением как минимум по 100 различным начальным конфигурациям, рассчитанным не менее чем на 10 различных реализациях случайной сети.

Первый поразительный результат моделирования – отсутствие эпидемического порога, то есть  $\lambda_c = 0$ . На рис. 2 показана стационарная распространённость вируса, которая при уменьшении  $\lambda$  убывает как  $\rho \sim \exp(-C/\lambda)$ , где  $C$  – константа. Следовательно, при любом конечном значении  $\lambda$  вирус может охватить достаточно большую сеть, сохраняя конечную распространённость. Во всех сетях с ограниченной степенью узлов стационарная распространённость ниже эпидемического порога всегда равна нулю: все заражения угасают. Дополнительным подтверждением нашего результата является полное отсутствие масштабной зависимости  $\rho$  от числа узлов, которая, напротив, характерна для эпидемических переходов вблизи конечного порога [8]. Это позволяет исключить артефакты, обусловленные конечным размером сети. Результат можно понять интуитивно: на обычных решётках чем выше степень узла, тем ниже эпидемический порог. В SF-сети неограниченные флуктуации степени,  $\langle k^2 \rangle = \infty$ , играют роль бесконечной связности и тем самым обнуляют порог.

Наконец, мы анализируем распространение инфекции из локализованного источника вируса. Рост заражения со временем имеет степенной характер, что согласуется с реальными данными, в которых экспоненциального роста численности циркулирующего вируса не наблюдалось. Примечательно, что, применяя использованное выше определение вероятности выживания  $P_s(t)$ , в нашей модели мы получаем ту же экспоненциальную временную зависи-



**Рис. 3.** (а) Вероятность выживания  $P_s(t)$  при скорости распространения  $\lambda = 0,065$  в безмасштабных сетях размеров  $N = 5 \times 10^5$  (квадраты),  $N = 2,5 \times 10^4$  (ромбы),  $N = 1,25 \times 10^4$  (треугольники) и  $N = 6,25 \times 10^3$  (кружки). Экспоненциальное поведение после резкого начального спада согласуется с анализом данных на рис. 1. (б) Относительная плотность  $\rho_k$  как функция  $k^{-1}$  в SF-сети размера  $N = 5 \times 10^5$  при  $\lambda = 0,1$ . График воспроизводит зависимость, предсказанную уравнением (2).

мость (рис. 3а). Характерное время жизни зависит от скорости распространения и размера сети, что позволяет связать среднее время жизни вирусного штамма с эффективной скоростью распространения и размером Интернета [20]. Одновременно расхождение времён жизни при увеличении размера сети указывает, что по мере расширения Интернета вирусы живут дольше.

К системе можно подойти и аналитически, записав одноузельное уравнение, описывающее временную эволюцию  $\rho(t)$ . Для учёта флуктуаций степени введём относительную плотность  $\rho_k(t)$  инфицированных узлов заданной степени  $k$ , то есть вероятность того, что узел с  $k$  связями инфицирован. Динамические уравнения скоростей реакций в приближении среднего поля имеют вид [8,21]

$$\partial_t \rho_k(t) = -\rho_k(t) + \lambda k [1 - \rho_k(t)] \Theta(\lambda). \quad (1)$$

Член рождения инфекции учитывает вероятность того, что узел с  $k$  связями здоров,  $[1 - \rho_k(t)]$ , и заражётся через соседний узел. Вероятность этого события пропорциональна скорости заражения, числу связей и вероятности  $\Theta(\lambda)$  того, что произвольно выбранная связь ведёт к инфицированному узлу. Среднеполевой характер уравнения обусловлен тем, что корреляции плотностей между различными узлами не учитываются. Вместе с тем мы отказались от предположения об однородности степеней узлов, обычно используемого для регулярных сетей. Налагая условие стационарности,  $\partial_t \rho_k(t) = 0$ , получаем стационарные плотности

$$\rho_k = \frac{k\lambda\Theta(\lambda)}{1 + k\lambda\Theta(\lambda)}, \quad (2)$$

то есть чем выше степень узла, тем больше вероятность его заражения. Эту неоднородность необходимо учитывать при самосогласованном вычислении  $\Theta(\lambda)$ . Действительно, вероятность того, что связь ведёт к узлу с  $s$  связями, пропорциональна  $sP(s)$ . Иными словами, случайно выбранная связь с большей ве-

роятностью соединена с узлом высокой степени. Поэтому

$$\Theta(\lambda) = \sum_k \frac{kP(k)\rho_k}{\sum_s sP(s)}. \quad (3)$$

Поскольку  $\rho_k$ , в свою очередь, зависит от  $\Theta(\lambda)$ , мы получаем уравнение самосогласования, позволяющее определить  $\Theta(\lambda)$  и  $\rho_k$ . Наконец, параметр порядка можно вычислить из соотношения  $\rho = \sum_k P(k)\rho_k$ , выражающего среднюю плотность инфицированных узлов в системе. В рассматриваемой SF-модели распределение степеней имеет вид  $P(k) = 2m^2/k^3$ , причём  $k$  аппроксимируется непрерывной переменной [6]. Интегрирование уравнения (3) в низшем порядке по  $\lambda$  даёт

$$\Theta(\lambda) = \frac{e^{-1/(m\lambda)}}{\lambda m}.$$

После усреднения по  $\rho_k$  окончательно получаем

$$\rho \simeq 2e^{-1/(m\lambda)}. \quad (4)$$

Этот наглядный расчёт воспроизводит численные результаты и подтверждает неожиданное отсутствие в модели какого-либо эпидемического порога или критической точки, то есть  $\lambda_c = 0$ . В качестве дополнительной проверки аналитических результатов мы численно вычислили относительные плотности  $\rho_k$  в модели и восстановили предсказанную уравнением (2) зависимость от  $k$  (рис. 3б). Следует также отметить, что предложенную схему можно обобщить на сети с  $2 < \gamma \leq 3$ , получив качественно те же результаты. Лишь при  $\gamma > 4$  эпидемии в SF-сетях имеют те же свойства, что и в случайных сетях. Подробный анализ различных случаев будет представлен отдельно [22].

Возникающая картина эпидемического распространения в сложных сетях подчёркивает роль топологии в эпидемическом моделировании. В частности, отсутствие эпидемического порога и критического поведения в широком классе безмасштабных сетей является неожиданным результатом, радикально меняющим многие стандартные выводы о распространении эпидемий. Инфекции могут размножаться в таких сетях при сколь угодно малой скорости распространения. Однако эта крайне неблагоприятная особенность компенсируется экспоненциально малой распространённостью в широком диапазоне скоростей,  $\lambda \ll 1$ . Данный вывод особенно важен для технологических сетей, таких как Интернет [4], чья безмасштабная связность характеризуется показателем  $\gamma \simeq 2.5$ . Предложенная картина, например, качественно согласуется с реальными данными о распространении компьютерных вирусов и может разрешить давнюю проблему повсеместно низкой распространённости таких вирусов без предположения о какой-либо глобальной настройке скоростей распространения.

Работа частично поддержана контрактом European Network No. ERBFMRXCT980183. Р. Пастор-Саторрас также благодарит за поддержку по гранту CICYT

PB97-0693. Авторы признательны S. Franz, M.-C. Miguel, R. V. Solé, M. Vergassola, S. Visintin, S. Zapperi и R. Zecchina за полезные замечания и обсуждения.

## Список литературы

- [1] См. специальный раздел, посвящённый сложным системам: *Science* **284**, 79 (1999); S. Wasserman and K. Faust, *Social Network Analysis* (Cambridge University Press, Cambridge, 1994).
- [2] L. A. N. Amaral, A. Scala, M. Barthélemy, and H. E. Stanley, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 11149 (2000).
- [3] M. Faloutsos, P. Faloutsos, and C. Faloutsos, *ACM SIGCOMM '99, Comput. Commun. Rev.* **29**, 251 (1999).
- [4] A. Medina, I. Matt, and J. Byers, *Comput. Commun. Rev.* **30**, 18 (2000); G. Caldarelli, R. Marchetti, and L. Pietronero, *Europhys. Lett.* **52**, 386 (2000).
- [5] R. Albert, H. Jeong, and A.-L. Barabási, *Nature (London)* **401**, 130 (1999).
- [6] A.-L. Barabási and R. Albert, *Science* **286**, 509 (1999); A.-L. Barabási, R. Albert, and H. Jeong, *Physica (Amsterdam)* **272A**, 173 (1999).
- [7] C. Moore and M. E. J. Newman, *Phys. Rev. E* **61**, 5678 (2000).
- [8] J. Marro and R. Dickman, *Nonequilibrium Phase Transitions in Lattice Models* (Cambridge University Press, Cambridge, 1999).
- [9] F. B. Cohen, *A Short Course on Computer Viruses* (John Wiley & Sons, New York, 1994).
- [10] J. O. Kephart, G. B. Sorkin, D. M. Chess, and S. R. White, *Sci. Am.* **277**, No. 5, 56 (1997); S. R. White, in *Proceedings of the Virus Bulletin Conference*, Munich, 1998. Доступно в сети: <http://www.research.ibm.com/antivirus/SciPapers.htm>.
- [11] N. T. J. Bailey, *The Mathematical Theory of Infectious Diseases* (Griffin, London, 1975), 2nd ed.; J. D. Murray, *Mathematical Biology* (Springer-Verlag, Berlin, 1993).
- [12] M. K. Hill, *Understanding Environmental Pollution* (Cambridge University Press, Cambridge, 1997).
- [13] J. O. Kephart, S. R. White, and D. M. Chess, *IEEE Spectr.* **30**, 20 (1993).
- [14] W. H. Murray, *Comput. Sec.* **7**, 130 (1988).
- [15] J. O. Kephart and S. R. White, in *Proceedings of the 1991 IEEE Computer Society Symposium on Research in Security and Privacy (SSP '91)* (IEEE, Washington, 1991), p. 343.
- [16] Можно также определить модели, в которых скорость заражения пропорциональна числу инфицированных ближайших соседей [8]. В интересующем нас режиме малой распространённости оба определения дают в точности одинаковое поведение.
- [17] Данные о распространённости вирусов доступны на веб-сайте <http://www.virusbtn.com/Prevalence/>.
- [18] Именно так вероятность выживания обычно определяется в численном моделировании процессов распространения; см. [8].
- [19] P. Erdős and P. Rényi, *Publ. Math. Inst. Hung. Acad. Sci.* **5**, 17 (1960); D. J. Watts and S. H. Strogatz, *Nature (London)* **393**, 440 (1998); A. Barrat and M. Weigt, *Eur. Phys. J. B* **13**, 547 (2000).
- [20] Такое характеристическое масштабирование часто встречается при фазовых переходах в поглощающее состояние в системах конечного размера [8]. В общем случае ненулевая величина  $P_s(\infty)$  возможна только для сетей бесконечного размера.
- [21] G. Szabó, *Phys. Rev. E* **62**, 7474 (2000).
- [22] R. Pastor-Satorras and A. Vespignani (не опубликовано).

Источник: R. Pastor-Satorras and A. Vespignani, "Epidemic Spreading in Scale-Free Networks," *Physical Review Letters* **86**, 3200–3203 (2001). Русский перевод подготовлен по предоставленному пользователем оригиналу.