International Journal of Modern Physics C
Vol. 25, No. 5 (2014) 1440012 (9 pages)

DOI: 10.1142/S0129183114400129

Non-Gaussian behavior of the internet topological fluctuations

Qu Chen*, Jiang-Hai Qian† and Ding-Ding Han*,‡

*School of Information Science and Technology
East China Normal University
Shanghai 200241, P. R. China

†Shanghai Institute of Applied Physics
Chinese Academy of Sciences
Shanghai 201800, P. R. China
‡ddhan@ee.ecnu.edu.cn

Received 30 June 2013
Accepted 30 September 2013
Published 22 January 2014

The traditional Gibrat's hypotheses were once used to model the topological fluctuations of Internet. Although it seems to reproduce the scaling relation of Internet's degree distribution, the detailed micro-dynamics have never been empirically validated. Here, we analyze the distribution of degree growth rates of the Internet for various time scales. We find that in contrast to the traditional Gibrat's assumptions, none of the degree growth rates are normally distributed, but behaves as an exponential decrease on its body and a power-law decay on its tail. Moreover, the observed growth rate distribution turns out independent of the initial degree when the time interval enlarges to a year. Our observations do not consist with the traditional Gibrat law model and suggest a more complex fluctuation mechanism underlying the evolution of Internet.

Keywords: Topological fluctuation; internet; non-Gaussian; micro-dynamics; Gibrat's law.

PACS Nos.: 89.20.Hh, 89.75.Fb, 05.40.−a, 02.50.−r.

1. Introduction

Internet is a typical complex system exhibiting simultaneously stationary statistical properties and strong fluctuations on both topological and dynamical aspects.[1–5] The preferential attachment (PA) process,[6] as mentioned in most scale-free network models, unveils the origin of the power-law behavior in the degree distribution. But things are not always that simple as edge connection or disconnection happens incessantly in evolving networks. Hence, the topological fluctuation can be considered as another possible candidate for the scale-free behavior. Compared with its dynamical flux,[7–12] however, it is less understood, and both theoretical and empirical basis remain insufficient.

From the fluctuation perspective, Goh *et al.*[13] proposed a fluctuation-driven model in which edge fluctuations are considered essential in the network growth. Following the classical Gibrat's law,[14] i.e. the temporal fluctuation of a variable is proportional to its present value, they predict a power-law degree distribution with the degree exponent $\gamma \approx 2.1$, which is in good agreement with the empirical measures $\gamma_{AS} \approx 2.2$.[1,2] Similar method has also been conducted by Huberman and Adamic (HA) in characterizing the growth dynamics of the Web.[15] However, the application of the classical Gibrat's hypothesis to Internet evolution has not been validated on the microscopic level. Indeed, the Central Limit Theorem, as the basis of most previous work for deriving the scaling exponent, requires only independent and identically distributed (i.i.d) random variables, but no other detailed information about the growth rate distribution. In real networks, however, heterogeneous behavior dominates, thus indicating a non-Gaussian distribution. In such context, existing understandings on the micro-dynamics of the Internet may be incomplete. Actually, the traditional Gibrat's law is not generally valid. No matter in the economic, social or even the traffic systems,[16–20] the increments of a specific large time interval are verified to be exponentially distributed. Hence, it is indispensable to discern the intrinsic dynamical fluctuations of the Internet.

As shown in this paper, topological fluctuation of the Internet cannot be well-described by the simplest random multiplicative process. We find that for any time scale, the degree growth rates are not normally distributed, but behave as an exponential decrease on its body with a power-law decay on its tail. In the next section, we will first revisit scaling properties on the macro-level. And in Sec. 3, we will briefly interpret the origin of power-law degree distribution from the fluctuation aspect. Later in Sec. 4, we will provide empirical evidence on Internet's degree growth pattern.

2. Macroscopic Topological Properties of the Internet

In this section, we will describe the topological properties of the Internet at the autonomous system (AS) level. Our data come from the Oregon Route Views project.[21] The original data are extracted from the Border Gateway Protocol (BGP) routing tables, which provide the detailed information about the routing path of two AS regions. Here, each AS is represented by a single node while each edge is the logical link between inter-connected ASs. In this way, the unweighted and undirected graph of the Internet can be established and its adjacency matrix $A(G)$ is represented as:

$$A(i,j) = \begin{cases} 1 & \text{if } \ i \text{ and } j \text{ are adjacent,} \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

First, we analyzed the increment of the total number of nodes and edges over the period of 10 years from 2003 to 2012. As shown in Fig. 1(a), the number of nodes $N(t)$ grows exponentially with time as $N(t) = N(0)\exp(\alpha t)$ with $\alpha \approx 0.11$, and the

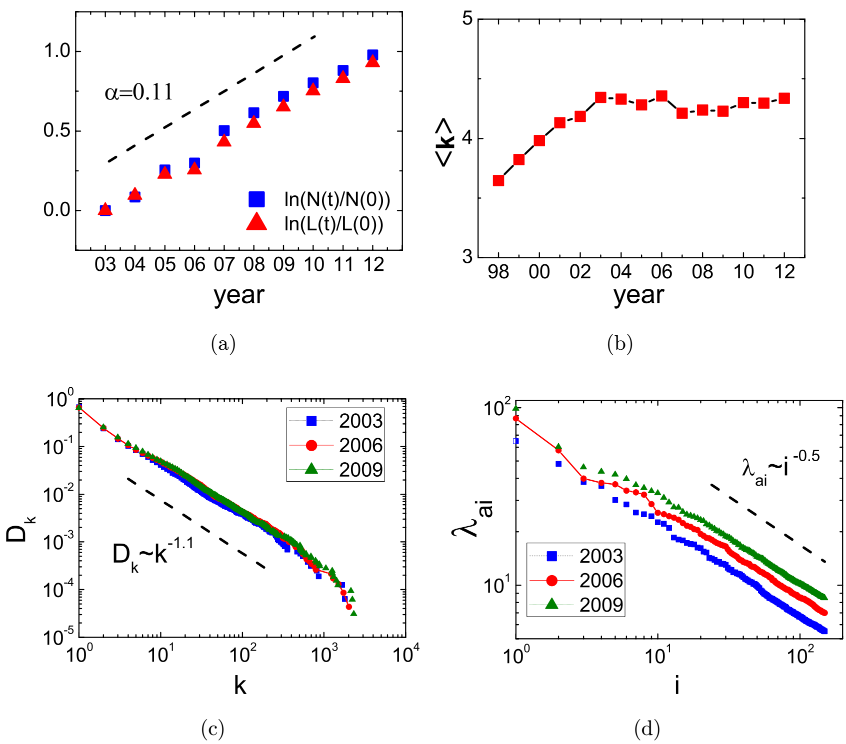


Fig. 1. (Color online) Macro topological properties of the Internet. (a) Growth of the number of vertices and edges. $L(t)$ increases exponentially with exponent 0.11 as the same order of $N(t)$. (b) Growth of the average degree $\langle k\rangle$. The average degree grows to $\langle k\rangle = 4.5$ in 2003 and remains stable since then. (c) Cumulative degree distribution. The cumulative distribution function of degree in the year 2003, 2006 and 2009 are in principle linearly overlapping with a slope of -1.1, which yields to a degree exponent $\gamma \approx -2.1$. (d) Descending sequence of the largest 150 eigenvalues. The descending eigenvalues λ_i are proportional to the power of its corresponding order, and practically equal exponents ϵ of different snapshots characterize the topological nature of the Internet.

number of edges $L(t)$ grows almost at the same rate. In the early development of the Internet (1997–2000), the growth of edges $L(t)$ was reported to be faster than that of nodes $N(t)$, leading to the relation $L(t) = N(t)^{1+\theta}$ with $\theta = 0.16$.[13] In contrast, our data indicates that $L(t)$ increases exponentially with exponent 0.11 as the same order of $N(t)$ after 2003, leading to a constant average degree $\langle k\rangle = 4.5$ (Fig. 1(b)). This result is consistent with the fact that scale-free network is sparse.[22] In this sense, the evolution of Internet is not stable until 2003, and it is necessary to re-examine the dynamics of the network.

Moreover, we analyzed the scale invariant property of the Internet in the year 2003, 2006 and 2009. The cumulative distribution function (CDF) of degree is defined as the percentage of nodes having degree more than or equal to k, i.e. $D_k = P\{x \geq k\}$. The cumulative degree distributions, plotted on the double logarithmic scale, are all well-approximated by a linear form (Fig. 1(c)), indicating that

D_k is proportional to the power of degree k:

$$D_k \propto k^D. \tag{2}$$

CDFs of different year are in principle overlapping, and linear fitting result gives the slope $D \approx -1.1$, which yields to a degree exponent $\gamma = D - 1 \approx -2.1$.

Adjacency matrix is closely related to network dynamics and its eigenvalues intimately unveil some basic topological properties such as the diameter, the number of spanning trees and connected components.[23] Here we calculated the largest 150 eigenvalues $\lambda_i (i = 1, 2, \ldots, 150)$, where i is the order of λ_i in the descending sequence. Figure 1(d) depicts that whenever in 2003, 2006 or 2009, a power-law relation between eigenvalues and its corresponding order exists:

$$\lambda_i \propto i^{\varepsilon}. \tag{3}$$

Intriguingly, the power-law exponents ε obtained with least-square errors method in log–log scale are practically equal: -0.51, -0.49 and -0.48. Therefore the appearance of this power-law relation is not a coincidence, and the historical invariant exponent, despite the rapid growth of the network size, suggests that the eigen exponent ε characterizes the topological nature of the Internet. For instance, the eigen exponent at the router level is merely -0.177,[1] significantly different from that of the AS level.

3. Degree Distribution Induced by Fluctuation

The good consistence between our empirical analysis and other previous studies[1,2,4] based on different data sources and time reveals a stationary macroscopic topology of Internet. On the other hand, violent fluctuations can still occur at the microscopic level as connections between routers appear and disappear constantly. The most appealing fact is that the two seemingly inconsistent phenomenon not only coexist, but correlate with each other. This scenario is well-established by the classical Gibrat model[14] which provides a possible explanation of the power-law degree induced by the system fluctuation.

Classical Gibrat model provides another scenario in describing Internet fluctuations. Gibrat model is based on a stochastic multiplicative process, that is, the temporal fluctuation of a variable is proportional to its present value. In other words, the higher connectivity of a node, the more new links it can capture. Although this model does not specify how nodes are connected with each other, the implicit PA mechanism in nature helps in reproducing the skewed degree distribution. Hence it has been widely used in explaining the growth property in fields like firms, gross domestic product (GDP) and city population.[16–20,24–27] Besides, Huberman interpreted the scale-free property of the WWW with an extension of the model.[15]

Generally the degree growth can be viewed as a random multiplicative process,

$$k_i(t + 1) = k_i(t)[1 + \xi_i(t + 1)], \tag{4}$$

where $k_i(t)$ is the degree of a vertex i at time t and $\xi_i(t)$ is the corresponding growth rate. $\{\xi_i(t)\}$ is a set of independent random variables with the same mean g_0 and standard deviation σ_0. We define the logarithmic growth rate as $r_i(t) = \ln(1 + \xi_i(t+1))$, therefore after a sufficiently large time T, the degree growth rate of node i follows:

$$\ln\left(\frac{k_i(t_0+T)}{k_i(t_0)}\right) = \sum_{t=t_0+1}^{t_0+T} r_i(t). \tag{5}$$

There are two basic assumptions in Gibrat model:

Assumption 1. $\xi_i(t)$ is independent random variable, i.e. the degree growth rate is independent of the previous node degree and also uncorrelated in time.
Assumption 2. $\xi_i(t)$ is identically distributed, i.e. the growth factor of each AS belongs to the same distribution.

By invoking the Central Limit Theorem, the degree increment $k_i(t)/k_i(t_0)$ of each vertex for every time interval t is log-normally distributed with mean $g_0 t$ and variance $\sigma_0^2 t$. Together with Assumption 2, the probability that a vertex with initial degree k_0 at time t_0 grows to k after τ can be formulized as:

$$P(k,\tau|k_0) = \frac{1}{k\sqrt{2\pi\sigma_0^2\tau}}\exp\left\{-\frac{[\ln(k/k_0) - g_0\tau]^2}{2\sigma_0^2\tau}\right\}. \tag{6}$$

Since the degree distribution is actually a mixture of such log-normal distributions over different τ, then it is possible to approximate the degree distribution as:

$$P(k) \sim \int d\tau\rho(\tau)P(k,\tau|k_0) \sim k^{-\gamma}, \tag{7}$$

where $\rho(\tau)$ is the density of ASs with age τ. From analytical calculation, the exponent γ is determined by:

$$\gamma = 1 - \frac{g_0}{\sigma_0^2} + \frac{\sqrt{g_0^2 + 2\alpha\sigma_0^2}}{\sigma_0^2}. \tag{8}$$

With empirical values of α, g_0 and σ_0, Goh obtained the degree exponent $\gamma \approx 2.1$,[13] which is well-consistent with directly measured value.[1,2] Above interpretation and empirical results indicate that power-law can emerge from the seemingly disordered fluctuations. And the Gibrat model is the bond which properly integrates fluctuation and the power-law degree distribution.

Despite the simplicity of the model, its basic assumptions have not been examined by the empirical data. An unevadable fact is that various studies have demonstrated that traditional Gibrat's law is not always sufficient to account for many nontrivial observations. For example empirical results ranging from firm size, GDP of countries to city populations all suggest an exponential distribution of the growth rate, which is inconsistence with the normal distribution result predicted by Gibrat's law.[16–20]

In the next section, we will focus on the growth rate of degree and provide evidence that the fluctuation indeed originated from a more complex mechanism.

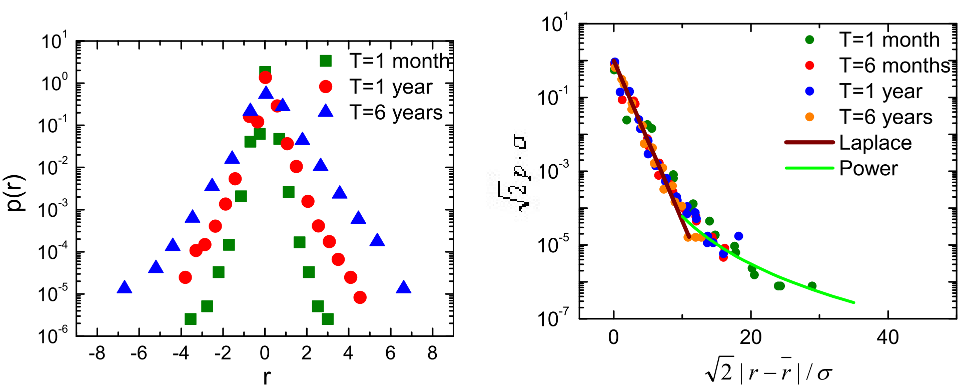
4. Detecting Fluctuation Mechanism: Degree Growth Pattern

To verify the Gibrat assumptions, we concentrate on the degree growth pattern of the Internet. The (logarithmic) degree growth rate of node i is defined as:

$$r_i(t) = \log R_i(t) = \log\left(\frac{k_i(t+T)}{k_i(t)}\right), \tag{9}$$

where $k_i(t)$ is the degree of node i at t and T is the time interval for observation. Here we select several time scales for analysis, for example, $T = 1$ month, $T = 1$ year and $T = 6$ years. Monthly data are collected within the period of 36 months from January 2005 to December 2007 and yearly data cover 15 years from 1998 to 2012. For simplicity, we will denote degree growth rate as $r = \log R = \log(\frac{k_1}{k_0})$ below.

Figure 2(a) depicts the probability density $p(r)$ of degree growth rate of the Internet. Intuitively, for any time scale, none of the degree growth rates are normally distributed as Gibrat's law suggests, but behaves as an exponential decrease on its body with a slower decay on its tail. To determine the functional form of the distribution, we plot the rescaled probability density of four typical time intervals, i.e. 1 month, 6 months, 1 year and 6 years. Meaningfully, curves in Fig. 2(b) are well-collapsed, indicating a consistent distribution form in both mid-term and long-term evolution. The fitting analysis for rescaled distribution shows a Laplacian form $p_{\text{scal}} = \exp(-|r_{\text{scal}}|)$ on its body (Brown line in Fig. 2(b)), while the tail decays approximately as a power-law with exponent -4.3 (Green line in Fig. 2(b)). Hence, the



(a) (b)

Fig. 2. (Color online) Non-Gaussian distribution of degree growth rates. (a) Distribution of degree growth rates for $T = 1$ month, 1 year and 6 years, respectively. For any time scale, none of the degree growth rates are normally distributed as Gibrat's law suggests, but in a "tent-shaped" form. (b) Rescaled distribution. The distribution behaves as a Laplacian decrease on its body (Brown line) and the fitting result suggests a power-law decay with the exponent -4.3 on its tail (Green line).

probability density can be approximately described by a Laplacian (double exponential) function:

$$p(r|k_0) = \frac{1}{\sqrt{2}\sigma(k_0)} \exp\left(-\frac{\sqrt{2}|r - \bar{r}(k_0)|}{\sigma(k_0)}\right), \tag{10}$$

with a power-law decay on its tail.

Similar distribution forms for different time intervals proved in another aspect that the non-Gaussian behavior has no coincidence. Contradicted with the conventional Gibrat assumptions, degrees of consecutive time are empirically correlated. Actually, similar phenomenon have been reported in other field. For instance, Fu *et al.*[28] observed that the empirical distribution of business-firm growth rates can be characterized by a Laplacian cusp in its central part and a power-law tail with an

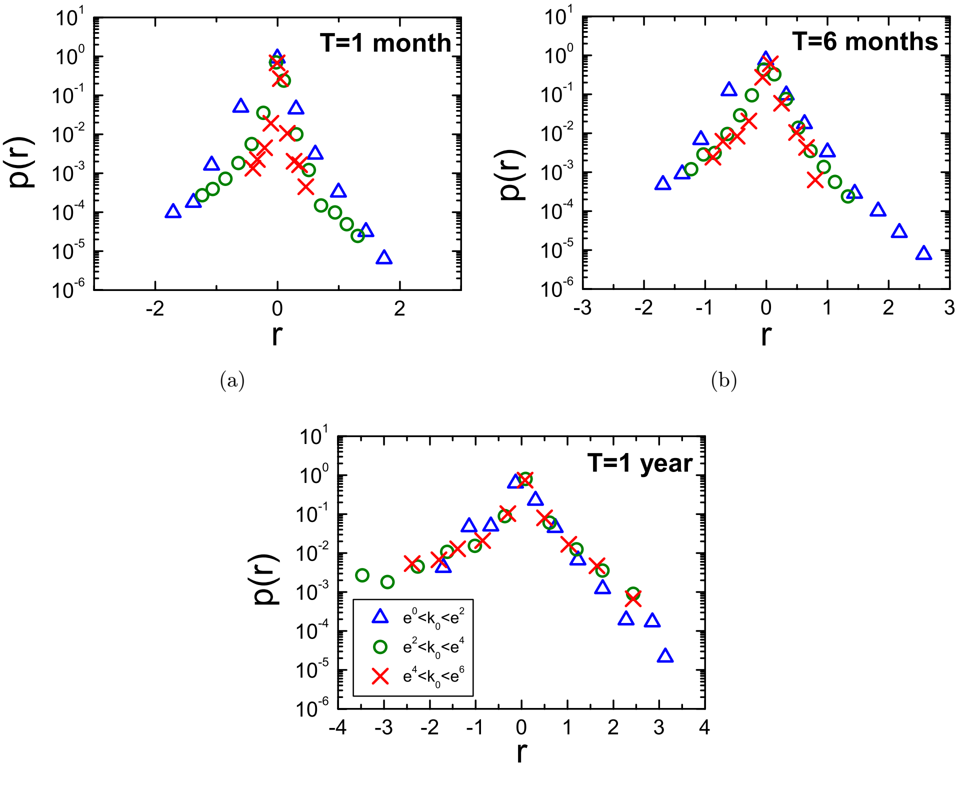


Fig. 3. (Color online) Dependence on the degree growth distribution of the initial value. (a) $T = 1$ month. The distribution shrinks obviously with increasing k_0, namely the volatility of degree growth correlates with the initial degree k_0 within a mid-term. (b) $T = 6$ months. Dependence on the initial degree appears as well, though the shrinking trend has greatly faded. (c) $T = 1$ year. The approximately collapsed curves indicate that the yearly growth rate distribution is independent of initial degree. Generally, the correlation between volatility and initial degree gradually weakens during the long-term evolution of the network.

exponent of 3. Moreover, from firm organization aspect, they proposed a reasonable model. Compared with firm growth, however, the tail in the Internet case decays faster so that their model is not suitable for interpreting the inner fluctuation mechanism of the Internet. Growth patterns of firm size and Internet node degree are similar, but also possess essential differences.

A step further, we investigate how the degree growth distribution depends on the initial degree k_0. ASs are classified into groups by selecting three equal intervals of $\log k_0$, and then calculate the conditional probability density $p(r|k_0)$ for each group. On monthly scale, the distribution shrinks obviously with increasing k_0, namely the volatility of degree growth correlates with the initial degree k_0. Therefore, less active ASs are much more volatile and easier to get or lose links within a mid-term, while the relative change of "hub" nodes are trivial. Notably, the shrinking trend is greatly eliminated with the increasing time interval, just as Fig. 3(b) indicates. And the dependence almost disappear, when the time window enlarges to a year as curves of each group collapse together in Fig. 3(c). This intriguing observation suggests that in mid-term evolution, the degree growth rates of the Internet are not identically distributed, but closely associated with initial values, and this correlation gradually weakens during the long-term evolution of the network.

5. Conclusion

Here, we have analyzed the topological fluctuation of the Internet. Empirical results illustrate that the degree growth rates are not normally distributed, but rather behave as a Laplacian distribution with a power-law decay on its tail. On the other hand, standard deviation of the degree growth rates correlates with the initial degree k_0. Nodes with larger connectivity are usually less volatile. More significantly, we make a comparison of the degree growth distribution for different time intervals. Our observations not only confirm the non-Gaussian behavior, but also indicate that the increasing time interval would weaken the above-mentioned correlation. Specifically, the yearly observed growth rate distribution turns out independent of the initial degree.

Compared with other studies in fields like economic, traffic and human interaction activity, we have specially analyzed the topological fluctuations of the Internet for various temporal resolutions. We find that the influence of the time interval should not be ignored. And neither the fluctuation properties of the network nor the independent identical distributed variables in Gibrat's assumptions are universal. Indeed, there is a more complex fluctuation mechanism underlying the evolution of the Internet.

Acknowledgment

The authors acknowledge the support of the National Natural Science Foundation of China (Grant Nos. 11075057).

References

1. M. Faloutsos, P. Faloutsos and C. Faloutsos, *ACM SIGCOMM Comput. Commun. Rev.* **29**, 251 (1999).
2. G. Siganos, M. Faloutsos, P. Faloutsos and C. Faloutsos, *IEEE/ACM Trans. Netw.* **11**, 514 (2003).
3. R. Pastor-Satorras, A. Vázquez and A. Vespignani, *Phys. Rev. Lett.* **87**, 258701 (2001).
4. A. Vázquez, R. Pastor-Satorras and A. Vespignani, *Phys. Rev. E* **65**, 066130 (2002).
5. S.-H. Yook, H. Jeong and A.-L. Barabási, *Proc. Natl. Acad. Sci. USA* **99**, 13382 (2002).
6. A.-L. Barabási and R. Albert, *Science* **286**, 509 (1999).
7. D.-D. Han, J.-G. Liu and Y.-G. Ma, *Chin. Phys. Lett.* **25**, 765 (2008).
8. Z. Eisler, I. Bartos and J. Kertész, *Adv. Phys.* **57**, 89 (2008).
9. M. A. De Menezes and A.-L. Barabási, *Phys. Rev. Lett.* **92**, 28701 (2004).
10. J. Duch and A. Arenas, *Eur. Phys. J.-Special Topics* **143**, 253 (2007).
11. J. Duch and A. Arenas, *Phys. Rev. Lett.* **96**, 218702 (2006).
12. M. A. De Menezes and A.-L. Barabási, *Phys. Rev. Lett.* **93**, 68701 (2004).
13. K.-I. Goh, B. Kahng and D. Kim, *Phys. Rev. Lett.* **88**, 108701 (2002).
14. R. Gibrat, *Les Inégalités Économiques* (Recueil Sirey, Paris, 1931).
15. B. A. Huberman and L. A. Adamic, *Nature* **401**, 131 (1999).
16. H. D. Rozenfeld, D. Rybski, J. S. Andrade, M. Batty, H. E. Stanley and H. A. Makse, *Proc. Natl. Acad. Sci. USA* **105**, 18702 (2008).
17. A. Gautreau, A. Barrat and M. Barthélemy, *Proc. Natl. Acad. Sci. USA* **106**, 8847 (2009).
18. R. N. Mantegna and H. E. Stanley, *Nature* **376**, 46 (1995).
19. M. H. Stanley, L. A. Amaral, S. V. Buldyrev, S. Havlin, H. Leschhorn, P. Maass, M. A. Salinger and H. E. Stanley, *Nature* **379**, 804 (1996).
20. D. Canning, L. A. N. Amaral, Y. Lee, M. Meyer and H. E. Stanley, *Econ. Lett.* **60**, 335 (1998).
21. Route views project, https://www.routeviews.org.
22. C. I. Del Genio, T. Gross and K. E. Bassler, *Phys. Rev. Lett.* **107**, 178701 (2011).
23. D. M. Cvetkovic, M. Doob and H. Sachs, *Spectra of graphs: Theory and application* (Academic Press New York, 1980).
24. D. Rybski, S. V. Buldyrev, S. Havlin, F. Liljeros and H. A. Makse, *Proc. Natl. Acad. Sci. USA* **106**, 12640 (2009).
25. K. Matia, L. A. Nunes Amaral, M. Luwel, H. F. Moed and H. E. Stanley, *J. Am. Soc. Inf. Sci. Technol.* **56**, 893 (2005).
26. V. Plerou, L. A. N. Amaral, P. Gopikrishnan, M. Meyer and H. E. Stanley, *Nature* **400**, 433 (1999).
27. Y. Lee, L. A. Nunes Amaral, D. Canning, M. Meyer and H. E. Stanley, *Phys. Rev. Lett.* **81**, 3275 (1998).
28. D. Fu, F. Pammolli, S. V. Buldyrev, M. Riccaboni, K. Matia, K. Yamasaki and H. E. Stanley, *Proc. Natl. Acad. Sci. USA* **102**, 18801 (2005).