

Fluctuation of the Download Network *

HAN Ding-Ding(韩定定)^{1**}, LIU Jin-Gao(刘锦高)¹, MA Yu-Gang(马余刚)²

¹Department of Electronic Engineering, School of Information Science and Technology, East China Normal University, Shanghai 200062 ²Shanghai Institute of Applied Physics, Chinese Academy of Sciences, Shanghai 201800

(Received 13 August 2007)

The scaling behaviour of fluctuation for a download network we investigated a few years ago based upon Zhang's Econophysics web page is presented. A power law scaling, namely $\sigma \sim \langle f \rangle^\alpha$ exists between the dispersion σ and average flux $\langle f \rangle$ of the download rates. The fluctuation exponent α is neither 1/2 nor 1, which were claimed as two universal fluctuation classes in previous publication, while it varies from 1/2 to 1 with the time window in which the download data are accumulated. The crossover behaviour of fluctuation exponents can be qualitatively understood by the external driving fluctuation model for a small-size system or a network traffic model which suggests congestion as the origin.

PACS: 89.75.Hc, 89.75.Da, 89.40.Dd

Many phenomenological and statistical analyses have been made for complex networks.^[1,2] Most studies focused on the long-time behaviour of a certain complex network. In this sense, the feature of the network corresponds to its static characteristic. However, time evolution of the network topology is also very important. During its evolution, the network nodes experience different traffic flux from time to time and the fluctuation is unavoidable. Actually, fluctuation is a universal phenomenon which exists in many different fields, such as nuclear fragmentation or hadron production,^[3-5] which can also be related to the critical behaviour or self-organized criticality. For instance, the dispersion (σ) of an order parameter, such as the charge of the largest fragments in nuclear fragmentation, shows a transition from $\sigma \propto \langle f \rangle^{1/2}$ (the ordered phase) to $\sigma \propto \langle f \rangle$ (the disordered phase) when the multifragmentation phase transition takes place in hot nuclear systems^[3,4] (here $\langle f \rangle$ is the average of the order parameter). For network dynamics, recently Menezes and Barabási investigated the fluctuation in a number of real world networks, including internet, river network, microchip, WWW and highway network dynamics, and presented a model to understand the origin of fluctuation in traffic process.^[6] They found that the fluctuation is dominantly driven by either internal or external dynamics of the complex system.^[6] In their studies, they found that there is a power-law scaling for the dispersion and the average flux, namely $\sigma \propto \langle f \rangle^\alpha$, and there are two classes of universality for real systems. In the Internet and the computer chip there is robust internal dynamics which leads to the fluctuation exponent $\alpha = 1/2$, while highway and Web traffic are driven by external demand which leads to the fluctuation exponent $\alpha = 1$. Researchers have used a stylized model of random walkers throughout network, they thought what is prob-

ably one of the most important factors in the traffic dynamics on networks is the limited capacity of nodes to handle packets simultaneously, which leads to pack-packet interaction and induce large fluctuations or even network congestion. However, a recent study on scaling of fluctuation in internet traffic shows that the fluctuation is different from 1/2 which was claimed in the above-mentioned papers. Researchers developed a model where the arrival and departure of 'packets' follow exponential distribution, and the processing capability of nodes is either unlimited or finite, proposed by Duch and Arenas.^[7] This model presents a wide variety of exponents between 1/2 and 1, revealing the dependence on the few parameters considered, and questioning the existence of universality classes. Hence it seems that the universal classes of fluctuation scaling for network dynamics are far from reaching consensus and therefore it is worth further investigating what about the fluctuation behaviour in other real networks. Nevertheless, so far there have been few analyses on the fluctuation behaviour of other specific networks other than the networks which have been investigated in Ref. [6,7]. In this Letter, we investigate the network evolution and fluctuation based on our previous study of the download network.

In our previous work in 2004,^[8] we reported, for the first time, that the download frequency of the papers in a web page is also a scale-free network. Its rank-ordered download distribution can be described by the Zipf law^[9,10] or Tsallis's non-extensive entropy.^[11] The data set of the download rates comes from a well constructed web page in the field of economical physics (so-called Econophysics) by Zhang since 1998.^[12] Furthermore, the mechanism of network growth was explained by the preferential attachment network model of Barabási and Albert. Since three years have passed after this network analysis, it is of

*Supported by the Shanghai Development Foundation for Science and Technology under Grant Nos 06JC14082 and 05XD14021.

**To whom correspondence should be addressed. Email: ddhan@ee.ecnu.edu.cn

© 2008 Chinese Physical Society and IOP Publishing Ltd

interesting to see how this network evolves and how about the fluctuation of download rate.

Firstly let us see some plots of rank distributions of the download numbers from the data on 2004/08/31 to 2007/07/29, which is shown in Fig. 1. Roughly, rank distribution is almost linear in double logarithm plots and they can be described by the Zipf law^[9]

$$N \simeq \text{rank}^{-\gamma}, \quad (1)$$

where γ is the Zipf law exponent. Zipf's law or scale free networks is different from the predictions of pure random networks introduced by Erdos and Renyi.^[13] Roughly, the shapes of these distributions keep similar in all times. However, quantitative analysis shows non-constant behaviour of the evolution of Zipf exponent (γ) which is shown in the inset of Fig. 1. Especially, there is a bump during 2006, i.e. the rank-ordered distributions tend to be steeper, which reflects higher download frequency for higher rank papers. However, the exponent decreases in 2007, i.e. more flatter distribution, which is obviously seen in Fig. 1 (diamonds). In this case, the web visitors prefer to download more papers listed in the web page, and they do not only focus on those top downloaded papers.

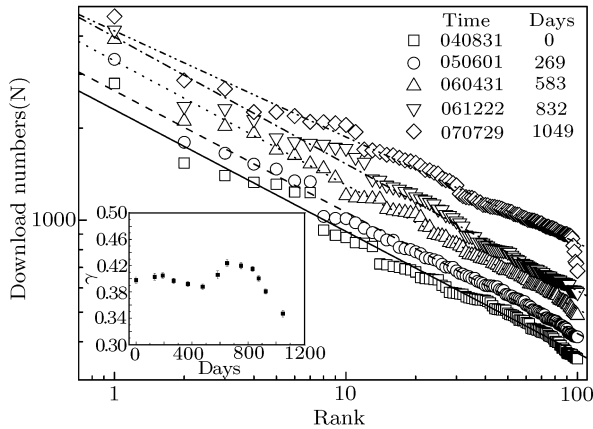


Fig. 1. Rank-ordered (Zipf-type) plot for the download numbers of the papers in <http://www.unifr.ch/econophysics> web page. Inset: the evolution of Zipf exponent as a function of the days starting from 2004/08/31.

To quantitatively see the increasing download numbers with time, we make a plot in Fig. 2 for the rank-sorted download numbers ΔN starting from the date 2004/08/31 (i.e. $\Delta N = 0$ on 2004/08/31) as a function of days starting from 2004/08/31. From the figure, all the curves do not follow exact linear increase. In other words, there exist fluctuations for the download rates day by day.

Quantitative fluctuation of the download rates can be described by the average download numbers per day, namely the average flux, $\langle f \rangle = \frac{\Delta N}{t_{i+1} - t_i}$, where

t_i is the time of the download day i (i.e., the abscissa of Fig. 2) and i from the starting date (2004/08/31) to the ending date which will be illustrated later. For each rank-ordered paper, these download rates change day by day, from which we can extract the average flux $\langle f \rangle$ and its dispersion σ (root of mean square of the download rate distribution) for each paper.

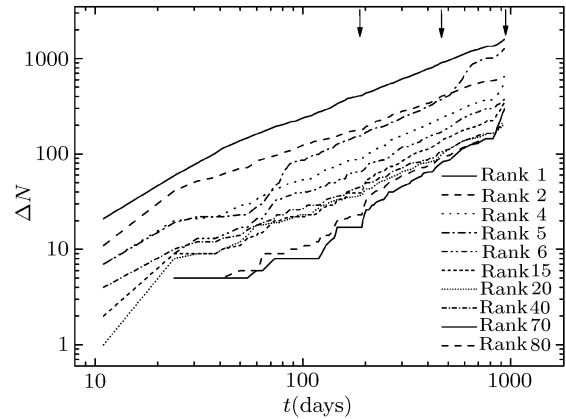


Fig. 2. Time evolution of download numbers which is sorted by the different ranks of the papers. The arrows in the upper axis illustrate three time points, namely 2005/03/13, 2005/12/26 and 2007/03/26 from left to right), which will be used to investigate the time window effects afterwards.

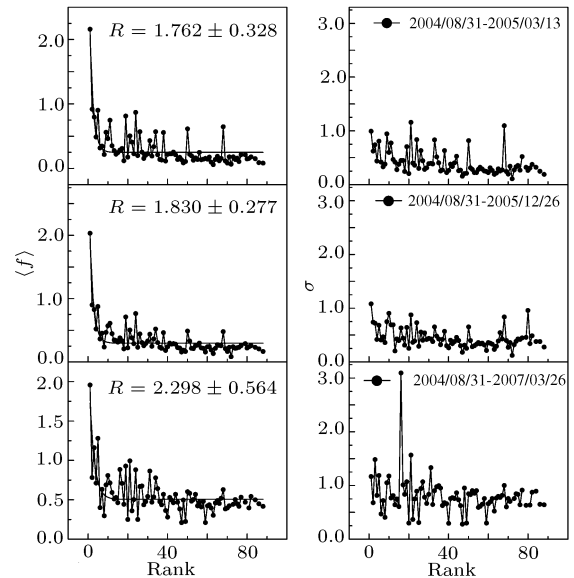


Fig. 3. Average flux $\langle f \rangle$ (left column) or the dispersion σ (left column) as a function of the rank for the accumulated data during 2004/08/31 to 2005/03/13 (top row), from 2004/08/31 to 2005/12/26 (middle row) and from 2004/08/31 to 2007/03/26 (bottom row). The lines in the left panels represent the exponential fits using Eq. (2).

Figure 3 shows the relationship between the average download rates $\langle f \rangle$ (left column) or the dispersion σ (left column) as a function of the rank for the accumulated data during 2004/08/31 to 2005/03/13

(i.e. ~ 6.5 months)(top row), from 2004/08/31 to 2005/12/26 (i.e. ~ 31 months)(i.e. ~ 16 months) (middle row) and from 2004/08/31 to 2007/03/26 (i.e. ~ 31 months) (bottom row). In left columns, the download rates show a fast decay with the increasing of rank for those most downloaded papers and the keep fluctuation for large rank values. It can be qualitatively understood that the web visitors prefer to download the top rank-ordered papers when he/she visits this page for the first time. This average day-by-day download flux can be roughly described by the exponential decay fits:

$$\langle f \rangle \propto \exp\left(-\frac{\text{rank}}{R}\right), \quad (2)$$

which is plotted in the figures and the half-lifetime decay exponent R is shown in the inset. R is small and seems to increase with the time period during which the data were accumulated. Right columns depict the dispersion as a function of the ranks which does not exhibit an obvious exponential decay as $\langle f \rangle$ versus rank shows, rather than frequent fluctuations.

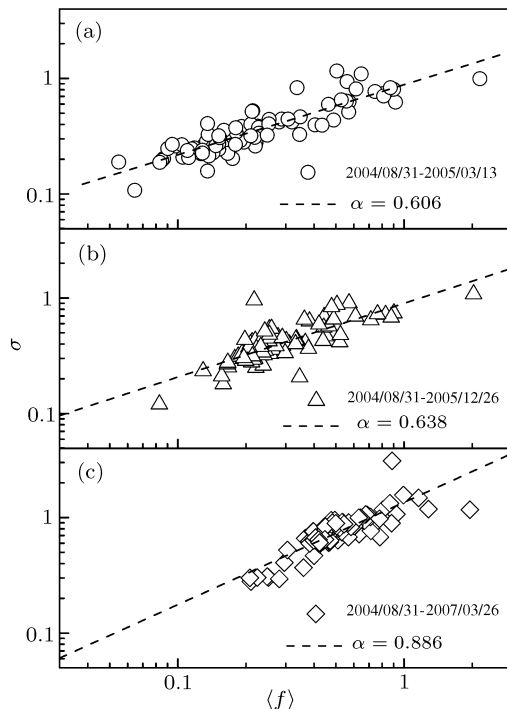


Fig. 4. Behaviour of σ as a function of $\langle f \rangle$ for all rank-ordered papers in three accumulated time windows.

Figure 4 demonstrates the relationship of $\langle f \rangle$ and σ for all rank-ordered papers. To investigate the possible effect of time window in which the download data were accumulated, we use the data ensembles which correspond to the period from 2004/08/31 to 2005/03/13 (a), from 2004/08/31 to 2005/12/26 (b) and from 2004/08/31 to 2007/03/26 (c), respectively. From each double logarithm plot, all points basically show linear increases between the average flux and its

dispersion. In this context, we fit the data points using the power law

$$\sigma \propto \langle f \rangle^\alpha \quad (3)$$

to extract the scaling parameter α which are shown in the inset of each panel. There are two points which we can learn from the figure: (1) the scaling parameter α is neither $1/2$ nor 1 . In the work of Menezes and Barabási, they thought there are two universal fluctuation classes: $\alpha \simeq 1/2$ or 1 systems. The typical example of the former is the Internet network, which was claimed to be dominantly driven by the internal dynamics. The typical example of the latter is the WWW URL links, which was claimed to be dominantly driven by the external dynamics. The exponents of our download network are between $1/2$ and 1 . (2) The scaling exponents seem to depend on the time windows during which the data samples are collected. In the other words, the longer the time windows, the larger the fluctuation exponent. Hence, in this viewpoint, we cannot exclude the possibility that fluctuation exponent could reach to 1 if we take very long time windows from the present work.

Qualitatively, our network fluctuation could be explained by the interplay of different fluctuation types, namely internal fluctuation and external fluctuation, which was proposed by Menezes and Barabási^[6]. In their model, they consider the random diffusion of W walkers (here they represent the visitors who download the papers) on the network, such that each walker that reaches a node i (here it represents the rank-ordered paper) departs in the next time step along one of other nodes. Originally each walker is placed on the network at a randomly chosen location and removed after it performs M steps, mimicking in a highly simplified fashion a human browser surfing the Web for information.^[6] In this way, finally the relationship between the average flux and the fluctuations follow a fluctuation scaling with $\alpha = 1/2$, corresponding to internal fluctuation driven behaviour. However, in real systems the fluctuation on a given node is determined not only by the system's internal dynamics, but also by changes in the external condition. To incorporate externally induced fluctuation, Menezes and Barabási allow W (the number of walkers in the web page), to vary from one day to the other. This is of course true, especially in the case that peoples visit un-congested web page, such as the Econophysics web page. Assuming that the day-to-day variations of $W(t)$ define a dynamic variable chosen from an uniform distribution in the interval $[W - \Delta W, W + \Delta W]$, for $\Delta W = 0$ one recovers $\alpha = 1/2$. However, when ΔW exceeds a certain threshold, in both the models the dynamical exponent changes to $\alpha = 1$.^[6] In this case, the external fluctuation can overshadow the internal fluctuation so that $\alpha = 1$. However, our network fluctuation behaviour is not exactly the above extreme cases, in-

stead it seems to be located in the transition region between the two extreme cases, namely the internal fluctuation type ($\alpha = 1/2$) and external fluctuation type ($\alpha = 1$). This could be explained by a smaller ΔW which does not exceed the certain threshold corresponding to a transition condition from $\alpha = 1/2$ to 1 in our download network. We think this is reasonable since the Econophysics web page is not a popular web page, such as Yahoo or Sina web pages, instead it is a small-circle scientific web page and no many people browse it often. In this case, ΔW could be small due to a few people browse this web page day-by-day so that the day variations ΔW cannot exceed the certain threshold. Actually, there exists a wide transition region where α is between 1/2 to 1 for a finite network system in Menezes and Barabási's model. Therefore, our interpretation for the origin of the download network fluctuation is not contradicted to their model. Using the above scenario, we can qualitatively learn what the fluctuation originates from for our download network.

However, the above scenario assumes that any external driving force is not unique to interpret the observed crossover fluctuation exponents between 1/2 and 1. A simple traffic model in complex networks, which suggests congestion as the origin of the increase of α and captures the essential parameters governing the dynamical process,^[7] is also possible to explain the download network fluctuation. In that model, traffic process in a complex network of N nodes as N queue systems, and a random walk simulation for the movement of packets on the network. The arrival process of packets to the network is controlled by a Poisson distribution with parameter λ , each packet enters the network at a random selected node. Once the packet arrives at the node, then it enters a queue. The delivery of the packets in the queue is controlled by an exponential distribution of service times with parameter μ . In that model, the packets will perform S random steps in the network before disappearing. This dynamics is performed in continuous time, assuming that the time expended by packets travelling through a link is negligible. The model can finally account for different scaling exponents α depending on the parameters λ , μ , S , and the time period P . Especially, we are interesting to see that α is a function of the time window length P in which the average was taken, which changes from 1/2 to 1: α increases with the time window length in the transition region. In the present study, even though our time window means the whole statistical one in which the download data were accumulated, which is different from the above mentioned time window in which the average were taken in the above-mentioned model, the effect could be analogous: the larger accumulated time windows can be somewhat equivalent to the larger time window length P in which the average were taken. The same trend which α increases with time was observed. Actually, the fluctuation exponents between 1/2 and 1 have been observed in the stock market transaction and other human dynamics such as emails from a particular company and data on the printing activity, etc., and their exponent shows the dependences on the time window size. A detailed review can be found in Ref. [14].

In summary, we have investigated the evolution of the download network for the rank-ordered papers which were listed in Zhang's Econophysics web page. In recent three years, the download distribution shows the change of the exponents even though the rank-ordered distribution still keeps scale-free feature, reflecting the change of traffic on nodes which represent the given downloaded papers. Further, we give a quantitative analysis for the average download rates $\langle f \rangle$ per day, which shows day-by-day fluctuation. The average flux shows a fast exponential decay as a function of the rank, while the dispersion does not show an obvious dependence of the rank. Interestingly, the dispersion of the download rate distributions shows power-law scaling behaviour with its average flux, namely $\sigma \propto \langle f \rangle^\alpha$. In different time windows ranging from about 6.5 months to 31 months in which the download distributions are accumulated, the scaling parameter α changes with the time windows, namely from 0.60 to 0.89. The origins are qualitatively interpreted by two models. Future work on quantitative model simulation and a possible Δ -scaling of network fluctuation is in progress.

References

- [1] Albert R, Barabási A-L 2002 *Rev. Mod. Phys.* **74** 47
- [2] Dorogovtsev S N and Mendes J F 2002 *Adv. Phys.* **51** 1079
- [3] Botet R et al 2001 *Phys. Rev. Lett.* **86** 3514
- [4] Ma Y G 2001 *J. Phys. G* **27** 2455
Ma Y G et al 2004 *Phys. Rev. C* **69** 031604(R)
Ma Y G et al 2005 *Phys. Rev. C* **71** 054606
Ma Y G et al 2005 *J. Phys. G* **31** S1179
Ma G L et al 2003 *Chin. Phys. Lett.* **20** 1013
- [5] Gulminelli F et al 2006 *Eur. Phys. J. A* **30** 253
Gulminelli F and D Agostino M 2006 *Dynamics and Thermodynamics with Nuclear Degree of Freedom* ed Chomaz P, Gulminelli F, Trautmann W and Yennello S J (Berlin: Springer) chap 5(6)
- [6] Algotto de Menezes M and Barabási A L 2004 *Phys. Rev. Lett.* **92** 028701
Algotto de Menezes M and Barabási A L 2004 *Phys. Rev. Lett.* **93** 068701
- [7] Duch J and Arenas A 2006 *Phys. Rev. Lett.* **96** 218702
- [8] Han D D et al 2004 *Chin. Phys. Lett.* **21** 1855
- [9] Zipf G K 1949 *Human behavior and the Principle of Least Effort* (Cambridge: Addison-Wesley)
- [10] Ma Y G 1999 *Phys. Rev. Lett.* **83** 3617
Ma Y G 2000 *Chin. Phys. Lett.* **17** 340
Ma Y G 2006 *Eur. Phys. J. A* **30** 227
Ma Y G 2006 *Dynamics and Thermodynamics with Nuclear Degree of Freedom* ed Gulminelli F, Trautmann W and Yennello S J (Berlin: Springer) chap 5(4)
- [11] Tsallis C et al 2000 *Eur. Phys. J. B* **13** 777
- [12] <http://www.unifr.ch/econophysics/>
- [13] Erdos P and Renyi A 1960 *Publ. Math. Inst. Hung. Acad. Sci.* **5** 17
- [14] Eisler Z, Bartos I and Kertész 2007 arXiv:0708.2053