Random walks on complex networks with inhomogeneous impact

Zoltán Eisler* and János Kertész†
Department of Theoretical Physics, Budapest University of Technology and Economics, Budapest, Hungary
(Received 17 January 2005; published 27 May 2005)

In many complex systems, for the activity f_i of the constituents or nodes i a power-law relationship was discovered between the standard deviation σ_i and the average strength of the activity: $\sigma_i \propto \langle f_i \rangle^\alpha$; universal values $\alpha=1/2$ or 1 were found, however, with exceptions. With the help of an impact variable we present a random walk model where the activity is the product of the number of visitors at a node and their impact. If the impact depends strongly on the node connectivity and the properties of the carrying network are broadly distributed (as in a scale-free network) we find both analytically and numerically nonuniversal α values. The exponent always crosses over to the universal value of 1 if the external drive dominates.

DOI: 10.1103/PhysRevE.71.057104 PACS number(s): 89.75.Da, 05.40.−a

Complex systems usually consist of many interacting units such that their scaffold is a network. The most puzzling questions are usually the ones regarding the dynamics of such systems. Examples range from traffic (vehicular or Internet) through the biochemistry of the cell to markets which are systems generated by human interactions. Due to its generality, the network aspect has turned out to be extremely fruitful in these studies, but the dynamics is usually rather individual, system dependent. However, recently an interesting unifying feature was found in systems with multichannel observations.

Processes taking place in many complex systems can be characterized by generalized activities $f_i(t) \geqslant 0$, defined at each constituent or node $i=1,\ldots,N$. For a wide range of systems, research [1–3] has revealed power-law scaling between the mean and the standard deviation of the activity of the nodes

$$\sigma_i \propto \langle f_i \rangle^\alpha, \tag{1}$$

where by definition

$$\sigma_i = \sqrt{\langle (f_i - \langle f_i \rangle)^2 \rangle}. \tag{2}$$

This is not unmotivated from equilibrium statistical physics. Many physical systems belong to the class $\alpha=1/2$; in most cases this is the fingerprint of equilibrium and the dominance of internal dynamics. Examples of such behavior are a computer chip or the hardware level Internet (the network of data transmission) [1].

One can show analytically the existence of a universality class with $\alpha=1$. This value always prevails in the presence of a strong driving force, when the dominant factor is such externally imposed dynamics. This limit is found for river networks, highway traffic and the World Wide Web [1].

It is instructive to recall that although only different aspects of the same system, the Internet and the Web fall into separate categories. The former has a robust internal activity even without outside (human) interaction, due to automatic queries, data transfer, etc. On the other hand, the latter consists of Web pages whose activity is generated by external demand (i.e., the clicks of users).

For such an analysis multichannel monitoring of a large number of elements is needed with a possibly broad range of $\langle f_i \rangle$. A previous work [2] introduced a method of decomposition to separate the effects of an external driving force from the system's internal dynamics originating from the constituents' individual behavior and interactions. Furthermore, when external driving was absent or subordinate, $\alpha=1/2$ seemed to hold for every investigated system, though universality could not be proven. In fact, more recent studies have shown that there are exceptions from this rule: Fluctuations of stock market trading activity (traded volume times price) are characterized by $\alpha \approx 0.72$ [3]. In this paper, we present a generalization of a random walk model by Menezes and Barabási [1], which accounts for such anomalous α values and clarifies their microscopic origins.

The model in Ref. [1] was motivated by the statistics of Web page visitations and we will also use this language. Let us take a scale-free Barabási-Albert network of N nodes (Web pages) [4,5]. First, we distribute W tokens (users, walkers) on the nodes randomly. In every time step every walker jumps from its present site to a random neighbor of the site along an edge (link). The new feature is that if a user steps to a site, it has to pay a certain value $V(i)$ (exerts a certain impact), which depends on the degree $k(i)$ of the visited node (more popular pages with more links cost more money) as a power law

$$V(i) \propto k(i)^\mu. \tag{3}$$

We must emphasize, that although this is a simplification, the assumption is not *ad hoc*. For example, the average activity for a given stock was found to scale with the exponent $\mu \approx 0.44$ as a function of the capitalization of the companies [8], which, in a sense, can be identified with the nodes' strength. This application will be discussed in detail.

We continue the diffusion for T steps and finally calculate the total profit of every Web site, which equals the $N(i)$ number of visitations multiplied by $V(i)$; these values will be the

*Electronic address: eisler@maxwell.phy.bme.hu
†Also at Laboratory of Computational Engineering, Helsinki University of Technology, Espoo, Finland.

activities $f_i(t=1)$ of the nodes on day 1. We iterate this procedure for the same network for $t=1,\ldots,D$ days to generate the whole time series $f_i(t)$. The original variant presented in Ref. [1] corresponds to the special case $\mu=0$: it only counts $N(i)$ for each node. Later we will address the case when the $V(i)$ values are allowed to fluctuate.

We expect a power-law scaling relationship such as

$$\langle N(i)\rangle \propto k(i)^{\nu}, \tag{4}$$

i.e., the mean number of the visitations to a node is proportional to its degree. Higher connected vertices should have higher traffic. Stationary solution of the master equation yields $\nu=1$ analytically, in line with Ref. [6]. Also, if initially tokens are distributed uniformly, after the first step the expectation value of the number of tokens at any node will be directly proportional to its degree. Therefore the model reaches the stationary state in one step. For the sake of generality we will keep the notation for a general value of ν, which could be generated by different dynamics.

We can write the mean profit of node i as

$$\langle f_i\rangle = \langle N(i)V(i)\rangle \propto k(i)^{\mu+\nu}. \tag{5}$$

Fluctuations can be expressed by the application of the central limit theorem. As walkers do not interact, their visits are independent, and thus for large enough $\langle N(i)\rangle$ and finite $\langle N(i)^2\rangle$ the distribution converges to a Gaussian. Moreover, the variance of the visits at node i is $\sigma_{Ni}^2=\langle[N(i)-\langle N(i)\rangle]^2\rangle \propto\langle N(i)\rangle\propto k(i)^{\nu}$. The variance of the signal detected on node i can then be written as

$$\sigma_i^2 = \sigma_{Ni}^2 V(i)^2 \propto k(i)^{2\mu+\nu}, \tag{6}$$

where the proportionality comes from (3). Finally, one can combine (5) and (6) to get

$$\alpha = \frac{1}{2}\left(1 + \frac{\mu/\nu}{\mu/\nu+1}\right). \tag{7}$$

α is the internal dynamical exponent defined by (1) in the absence of external forces.

We performed simulations of such a process and found perfect agreement with the above calculation. We fixed a Barabási-Albert network of $N=2000$ nodes, $W=200$ tokens, $T=100$ steps per day, and averaged over $D=10^5$ days. We also varied $\mu=-0.5, \ldots, 5.0$. Examples for the scaling relation (1) are shown in Fig. 1 [7]. There is a clear dependence of the slope on the value of μ. The measured exponents α, compared with the analytical formula are shown in Fig. 2.

The right-hand side of (7) is governed by the single parameter μ/ν. By setting $\mu=0$, we recover the original, non-independent impacts and $\alpha=1/2$. If $\mu/\nu>0$, the scaling exponent of fluctuations increases, $\alpha>1/2$. As $\mu/\nu\to\infty$, $\alpha\to 1$, which is the same exponent but due to a different mechanism as $\alpha=1$ arising from strong driving. Note that by choosing $\mu/\nu<0$, $\alpha<1/2$ values are also accessible.

This result is robust against fluctuations in V, i.e., if different users spend different amounts of money while visiting the same Web page, provided

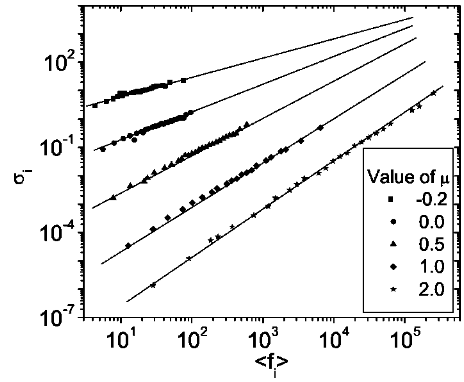

FIG. 1. Scaling of the standard deviation of node activity with the mean signal of the same node. A single point in the graph represents the average standard deviation of all nodes with approximately the same flux. Slopes on the log-log scale give the internal dynamical exponent α. Varying the impact distribution by changing μ causes a continuous change in α, as expected from (7).

$$V(i) = \langle V(i)\rangle X \propto k(i)^{\mu} X. \tag{8}$$

X is drawn independently from a fixed distribution for every visitation of node i. It is also assumed to have a finite second moment.

The distributions of $N(i)$ and $V(i)$ are independent and so they factorize in (5), whose formula hence remains unchanged. In order to prove that scaling suggested by (6) also persists, let us write

$$\sigma_i^2 = \left\langle \left(\sum_{n=1}^{N} V_i(n) - \left\langle \sum_{n=1}^{N} V_i(n) \right\rangle \right)^2 \right\rangle, \tag{9}$$

where n runs for all the N visits to site i during the day and $V_i(n)$ is the profit from the nth visit. By denoting $\sum_{n=1}^{N} V_i(n)$ as V_N, its density function as $\mathbb{P}(V_N)$, and that of $N(i)$ by $\mathbb{P}(N)$, it is possible to rewrite (9) as

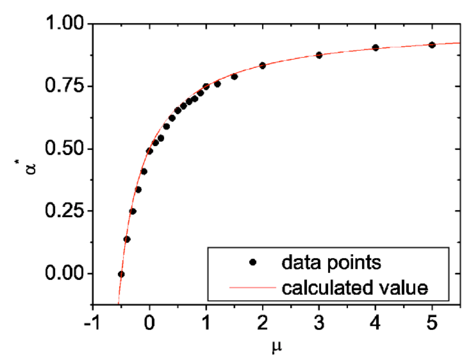

FIG. 2. (Color online) The values of the internal dynamical exponent α as a function of μ governing the distribution of node-dependent impacts. Circles show simulation results ($N=2000$ nodes, $W=200$ walkers, $T=100$ steps/day, averaged for $D=10^5$ days). The solid line represents the analytical formula (7) for $\nu=1$. By setting $\mu=0$, we recover $\alpha=1/2$, which is observed for several equilibrium systems. If one allows for node-dependent impacts ($\mu\neq 0$), nonuniversal behavior emerges and α can change continuously between 0 and 1.

$$\sigma_i^2 = \sum_{N=0}^{\infty} \mathbb{P}(N)\int_0^{\infty} dV_N \mathbb{P}(V_N)V_N^2 - \left(\sum_{N=0}^{\infty} \mathbb{P}(N) \int dV_N \mathbb{P}(V_N)V_N\right)^2. \quad (10)$$

By both adding and subtracting the term $\Sigma_{N=0}^{\infty}\mathbb{P}(N)[\int_0^{\infty}dV_N\mathbb{P}(V_N)V_N]^2$, then applying the equality $\int_0^{\infty}dV_N\mathbb{P}(V_N)V_N = N\langle V(i)\rangle$ and that for any fixed N the variance of V_N is $\sigma_{V_N}^2 = N\sigma_V^2$, one finally finds

$$\sigma_i^2 = \sigma_{Ni}^2\langle V(i)\rangle^2 + \sigma_{Vi}^2\langle N(i)\rangle \propto k(i)^{2\mu+\nu}. \quad (11)$$

The final proportionality comes from similar arguments as in the case of (6). In particular, we know that $\sigma_{Ni}^2 \propto \langle N(i)\rangle \propto k(i)^{\nu}$. On the other hand, with respect to scaling with the node degree $k(i)$, we defined $V(i)^2 \propto k(i)^{2\mu}$, while also $\sigma_V^2 \propto k(i)^{2\mu}$.

One can see explicitly the new source that contributes to fluctuations. The first term, basically the same as before, comes from diffusive dynamics. The second, additional term, is the one that describes the effect of visit to visit variations present in impacts $V(i)$. Regardless of this more complicated structure, the scaling of σ_i^2 with the vertex degree $k(i)$ is preserved, similarly to $\langle f_i\rangle$. Thus the dynamical exponent α is unaffected. Simulations based on various distributions of X confirmed this calculation.

Next, in order to analyze the behavior of the system under the influence of an external drive, we allowed day to day changes in the number of walkers W. Following Ref. [1], we introduced $W(t) = \langle W\rangle + \Delta W(t)$, where $\Delta W(t)$ is Gaussian white noise with standard deviation ΔW [9]. This acts as an external driving force and contributes to fluctuations. It is known, that increasing ΔW toward the strongly driven limit $(\Delta W/\langle W\rangle \gg 1)$, any system displays a crossover to $\alpha=1$ as a sign of the growing dominance of exogenous behavior. We used the above set of parameters and varied $\Delta W/\langle W\rangle = 5\times 10^{-3},\ldots, 15$. For all values of the internal exponent (7) we recovered this expected tendency, as shown in Fig. 3. Note that the intermediate values above that given by (7) but below 1 are effective exponents, actual scaling breaks down due to the crossover between them.

This approach can be reversed. Driving hides the microscopic dynamics, because all systems display the universal value $\alpha=1$. However, if it is possible to measure the internal exponent $\alpha(\Delta W/\langle W\rangle \ll 1)$, one can decide about the presence of impact inhomogeneity. It has been found [1], that for the hardware level Internet $\alpha = 1/2$. In this case $f_i(t)$ is the data flow through node i at time t, $\langle V(i)\rangle$ is the mean size of the passing data packets and $\langle N(i)\rangle$ is their mean number per unit time. This shows, that across the nodes of this system (routers) only the number of packets varies, but their size does not $(\mu=0)$. This homogeneous dynamics can be expected: As the same packets pass many computers, the mean of their sizes can well be independent of node degree.

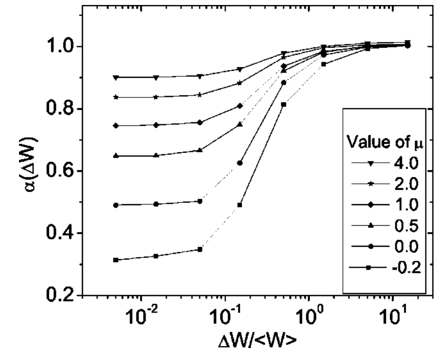

FIG. 3. Measured values of α as a function of the relative strength of driving force $(\Delta W/\langle W\rangle)$ for several fixed values of μ ($N=2000$ nodes, $W=200$ walkers, $T=100$ steps/day, averaged for $D=10^6$ days). Although the internal value $(\Delta W/\langle W\rangle \ll 1)$ varies, all systems display the universal behavior $\alpha=1$ in the exogenous limit $(\Delta W/\langle W\rangle \gg 1)$.

Although not readily represented as a network, a similar analysis can be carried out on stock market data [10]. Here, $f_i(t)$ is the value of stocks of the company i bought and/or sold at time t, $\langle N(i)\rangle$ is the mean number of transactions per unit time, and $V(i)$ is the value of stocks exchanged in a single trade. The role of degree (node size) is taken by company capitalization. In this case, it has been found that $\nu \approx 0.39$ and $\mu \approx 0.44$ [8]. This is direct evidence for the existence of scaling proposed in (3). Accordingly, α has the nontrivial value 0.72 [3], due to the presence of inhomogeneous impacts.

If the dynamics that generates the activities $f_i(t)$ is much slower than the method used to record them, one can observe the single events at each node. This happens, e.g., if we track each walker in our model. In this case, given the node size distribution $k(i)$, it is straightforward to measure μ and ν directly from $f_i(t)$ by their definitions. Again, this has been possible for stock markets [8], because all individual trades are documented as so called tick-by-tick data. Other possibilities could be distributed computing or telephone networks, where events take a longer time, while logs of activity can be written instantly. It would be interesting to check the validity of our assumptions in these networks too.

The partial support of the Center for Applied Mathematics and Computational Physics of the BUTE is acknowledged.

[1] M. A. de Menezes and A.-L. Barabási, Phys. Rev. Lett. **92**, 028701 (2004).
[2] M. A. de Menezes and A.-L. Barabási, Phys. Rev. Lett. **93**, 068701 (2004).
[3] Z. Eisler, J. Kertész, S.-H. Yook, and A.-L. Barabási, Europhys. Lett. **69**, 664 (2005).
[4] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[5] The theory itself is valid for any type of network. While our assumptions are independent of scale freedom, it ensures vertices with a broad range of degrees and so of $\langle f \rangle$. This is necessary for the observation of scaling.
[6] H. Rieger and J. Noh, Phys. Rev. Lett. **92**, 118701 (2004).
[7] To correct for bias coming from that there are many low activity nodes, whereas only a few highly active nodes, it is straightforward to take equal size bins in log $\langle f_i \rangle$ and average log σ_i for points in the same bin. This way, equal weight is given to all regimes in $\langle f_i \rangle$. This method was used throughout the paper.
[8] G. Zumbach, e-print cond-mat/0407769, Quant. Fin. (to be published).
[9] Of course the number of walkers $W(t)$ is always non-negative. Thus when $\langle W \rangle + \Delta W(t) < 0$, we define $W(t) = 0$.
[10] Z. Eisler and J. Kertész (unpublished).