Extreme Events on Complex Networks

Vimal Kishore,[1] M. S. Santhanam,[2] and R. E. Amritkar[1]
[1]*Physical Research Laboratory, Navrangpura, Ahmedabad 380 009, India*
[2]*Indian Institute of Science Education and Research, Pashan Road, Pune 411 021, India*
(Received 12 February 2011; published 3 May 2011)

A wide spectrum of extreme events ranging from traffic jams to floods take place on networks. Motivated by these, we employ a random walk model for transport and obtain analytical and numerical results for the extreme events on networks. They reveal an unforeseen, and yet a robust, feature: small degree nodes of a network are more likely to encounter extreme events than the hubs. Further, we also study the recurrence time distribution and scaling of the probabilities for extreme events. These results suggest a revision of design principles and can be used as an input for designing the nodes of a network so as to smoothly handle extreme events.

DOI: 10.1103/PhysRevLett.106.188701 PACS numbers: 89.75.Hc, 05.40.−a, 05.40.Fb

Extreme events (EE) taking place on networks is a fairly commonplace experience. Traffic jams in roads and other transportation networks, web servers not responding due to heavy load of web requests, floods in the network of rivers, and power blackouts due to tripping of power grids are some of the common examples of extreme events on networks. Such events can be thought of as emergent phenomena due to transport on the networks. As EE lead to losses ranging from financial and productivity to even life and property [1], it is important to estimate probabilities for the occurrence of extreme events and, if possible, incorporate them to design networks that can handle such extreme events.

Transport phenomena on the networks have been studied vigorously in the last several years [2,3] though they were not focused on the analysis of EE. However, one kind of extreme event in the form of congestion has been widely investigated [4]. For instance, a typical approach is to define rules for (a) generation and transport of traffic on the network and (b) capacity of the nodes to service them. Thus, a node will experience congestion when its capacity to service the incoming "packets" has been exceeded [5]. In this framework, several results on the stability of networks, cascading failures to congestion transition have been obtained. An extreme event, on the other hand, is defined as exceedances above a prescribed quantile and is not necessarily related to the handling capacity of the node in question. It arises from natural fluctuations in the traffic passing through a node and not due to constraints imposed by capacity. Thus, in the rest of this Letter, we discuss transport on the networks and analyze the probabilities for the occurrence of EE arising in them without having to model the dynamical processes or prescribe the capacity at each of the nodes.

The transport model we adopt is the random walk on complex networks [3]. Random walk is of fundamental importance in statistical physics though in real network settings many variants of random walk could be at work [6]. For instance, in the case of road traffic, the flow typically follows a fixed, often shortest, path from node A to B and can be loosely termed deterministic. Thus, given the operational principle of network dynamics, i.e., deterministic or probabilistic or a combination of both, we obtain the probabilities for the occurrence of EE on the nodes. This reveals a significant and unexpected result: namely, that the EE are more prone to occur in a small degree node than in a hub. This feature is robust against fluctuating traffic and even upon the application of intelligent routing algorithms (e.g., shortest paths). This principal result implies that the design principles for networks should focus on small degree nodes which are prone to EE. Further, these probability estimates allow us to design nodes that can have sufficient capacity to smoothly handle EE of a certain magnitude. Currently, for univariate time series, there is a widespread interest on the extreme value statistics and their properties, in particular, in systems that display long memory [7]. Thus, we place our results in the context of both the random walks and EE in a network setting.

We consider a connected, undirected, finite network with N nodes with E edges. The links are described by an adjacency matrix $\mathbf{A}$ whose elements A_{ij} are either 1 or 0 depending on whether i and j are connected by a link or not, respectively. On this network, we have W noninteracting walkers performing the standard random walk. A random walker at time t sitting on the ith node with K_i links can choose to hop to any of the neighboring nodes with equal probability. Thus, transition probability for going from the ith to the jth node is A_{ij}/K_i. We can write down a master equation for the n-step transition probability of a walker starting from node i at time $n = 0$ to node j at time n as,

$$P_{ij}(n + 1) = \sum_k \frac{A_{kj}}{K_k} P_{ik}(n). \tag{1}$$

It can be shown that the n-step time-evolution operator corresponding to this transition, acting on an initial distribution, leads to stationary distribution with eigenvalue unity [3] and it turns out to be

$$\lim_{n\to\infty} P_{ij}(n) = p_j = \frac{K_j}{2E}. \quad (2)$$

The existence of stationary distribution is crucial for defining EE. Physically, the stationary probability in Eq. (2) implies that more walkers will visit a given node if it has more links.

Now we can obtain the distribution of random walkers on a given node. We ask for the probability $f(w)$ that there are w walkers on a given node having degree K. Since the random walkers are independent and noninteracting, the probability of encountering w walkers at a given node is p^w while the rest of the $W - w$ walkers are distributed on all the other nodes. This turns out to be binomial distribution given by

$$f(w) = \binom{W}{w} p^w (1-p)^{W-w}. \quad (3)$$

Now, the mean and variance for a given node can be explicitly written down as

$$\langle f \rangle = \frac{WK}{2E}, \qquad \sigma^2 = W\frac{K}{2E}\left(1 - \frac{K}{2E}\right). \quad (4)$$

As expected, the mean and variance depends on the degree of the node for fixed W and E. Note that $K/2E \ll 1$ and hence $\sigma \approx \langle f \rangle^{1/2}$. This reproduces the relation proposed in Ref. [8], later shown to have limited validity [9].

One natural extension of the result in Eq. (3) is to account for fluctuations in the number of walkers. We assume that the total number of walkers is a random variable uniformly distributed in the interval $[W - \Delta, W + \Delta]$. Then the probability of finding w walkers becomes

$$f^{\Delta}(w) = \sum_{j=0}^{2\Delta} \frac{1}{2\Delta + 1} \binom{\tilde{W} + j}{w} p^w (1-p)^{\tilde{W}+j-w}, \quad (5)$$

where $\tilde{W} = W - \Delta$. The mean and variance of this distribution can be obtained as

$$\langle f^{\Delta} \rangle = \langle f \rangle, \qquad \sigma_{\Delta}^2 = \langle f^{\Delta} \rangle \left[1 + \langle f^{\Delta} \rangle \left\{\frac{\Delta^2}{3W^2} + \frac{\Delta}{3W^2} - \frac{1}{W}\right\}\right]. \quad (6)$$

In the spirit of extreme value statistics, an extreme event is one whose probability of occurrence is small, typically associated with the tail of the probability distribution function. In the network setting, we will apply the same principle to each of the nodes. Based on Eqs. (3) and (4), we will designate an event to be extreme if more than q walkers traverse a given node at any time instant. The probability for the occurrence of an extreme event can be obtained as

$$F(K) = \sum_{j=0}^{2\Delta} \frac{1}{2\Delta + 1} \sum_{k=\lfloor q \rfloor + 1}^{\tilde{W}+j} \binom{\tilde{W} + j}{k} p^k (1-p)^{\tilde{W}+j-k}, \quad (8)$$

where $\lfloor u \rfloor$ is the floor function defined as the largest integer not greater than u. Notice that necessarily the cutoff q will have to depend on the node (or rather, the traffic flowing through the node) in question. Applying uniform threshold independent of the node ($q =$ const) will lead to some nodes always experiencing an extreme event while some others never encountering any extreme event at all. Hence we define the threshold for extreme event to be $q = \langle f \rangle + m\sigma$, where m is any real number.

It does not seem possible to write the summation in Eq. (8) in closed form. However, for the special case when $\Delta = 0$, Eq. (8) simplifies to

$$F(K) = \sum_{k=\lfloor q \rfloor + 1}^{W} f(k) = I_p(\lfloor q \rfloor + 1, w - \lfloor q \rfloor), \quad (9)$$

where $I_p(.,.)$ is the regularized incomplete beta function [10]. For a given choice of network parameter E and number of walkers W, the extreme event probability at any node depends only on its degree. In Fig. 1 we show $F(K)$ as a function of degree K superimposed on the results obtained from random walk simulations. The agreement between Eq. (8) and the simulated results is quite good. Further, each point in the figure represents an average over all the nodes with the same degree. We emphasize that the oscillations seen in Fig. 1 are inherent in the analytical and numerical results and not due to insufficient ensemble averaging.

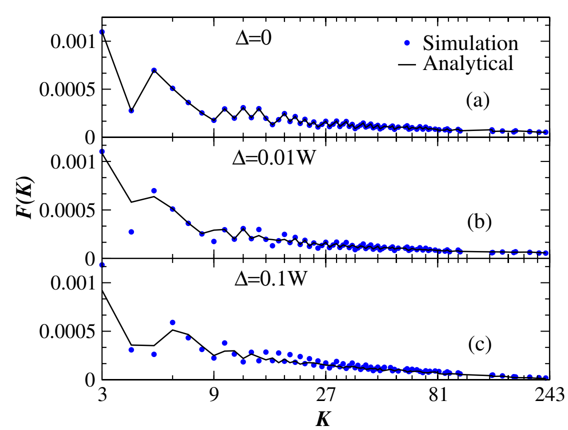

FIG. 1 (color online). Probability for the occurrence of extreme events as a function of degree K with fluctuations Δ in the total number of walkers on semilog plot. The threshold for EE is $q = \langle f \rangle + 4\sigma$. The solid lines are from the analytical result in Eq. (8). All the simulations shown in this Letter are obtained with a scale-free network (degree exponent $\gamma = 2.2$) with $N = 5000$ nodes, $E = 19\,815$ vertices, and $W = 2E$ walkers averaged over 100 realizations with randomly chosen initial conditions.

An important feature of this result is that the nodes with smaller degree ($K < 20$) reveal, on an average, a higher probability for the occurrence of EE as compared to the nodes with higher degree, say, $K > 100$. By careful choice of parameters, the probability $F(K)$ can differ by as much as an order of magnitude. This runs contrary to a naive expectation that higher degree nodes garner more traffic and hence are more prone to EE. While the former contention is still true in the random walk model we employ, the results here indicate that the latter one is not generally correct. As shown in Figs. 1(b) and 1(c), this feature is robust even when the number of walkers becomes a fluctuating quantity. We note that Eqs. (8) and (9) for the extreme event probability do not depend on the topology of the network. Even though the simulation results are shown for scale-free graphs, it holds good for other types of graphs (not shown here) with random and small world topologies. However, the difference in probability for EE between hubs and smaller degree nodes is not pronounced in the case of random graphs.

The threshold q that defines an event to be extreme depends on the traffic flowing through a given node. The choice $q = \langle f \rangle + m\sigma$ is arbitrary. Now, we show that the extreme event probability in Eq. (9) scales with the choice of threshold q or, equivalently, m. In the Fig. 2(a) we show $F_m(K)$ for various choices of m in log-log scale. Clearly, as m decreases, ignoring the local fluctuations, the curves tend to become horizontal. Physically, this can be understood as follows: $q \to 0$ implies that the threshold for EE decreases and this leads to larger number of EE and hence a higher probability of occurrence. In the limiting case of $q = 0$, $F(K) = 1$ for all nodes and all the events would be extreme. The graph in Fig. 2(a) suggests that it might be scaling with respect to q or m. Starting from Eq. (9), we were not able to determine the scaling analytically. Hence, we empirically show that the following scaling relation holds for the probability of EE,

$$\frac{F_m(K)}{K^{1-S_m}} = \text{constant}, \tag{10}$$

where $F_m(K)$ represents extreme event probability for threshold value q with parameter m. In this, S_m is the slope of the curves $F_m(K)$ in the Fig. 2(a). Using Eq. (10) on the simulated data for $\Delta = 0$, we find that all the curves for the probability of EE, shown in Fig. 2(b), collapse into one curve to a good approximation.

In the study of EE, distribution of their return intervals is an important quantity of interest. This carries the signature of the temporal correlations among the EE and is useful for hazard estimation in many areas. We focus on the return intervals for a given node of the network. Since the random walkers are noninteracting, the events on the nodes are uncorrelated. Then, the recurrence time distribution is given by $P(\tau) = e^{-\tau/\langle\tau\rangle}$, where the mean recurrence time is $\langle\tau\rangle = 1/F(K)$. In the inset of Fig. 3, we show $P(\tau)$ obtained from simulations for three nodes with different degrees. In semilog plot, they reveal an excellent agreement with the analytical distribution $P(\tau)$ (shown as a solid line). The main graph of Fig. 3 shows the mean recurrence time $\langle\tau\rangle$, the only parameter that characterizes the recurrence distribution, as a function of K and it agrees with the analytical result.

As pointed out before, many types of flow on the network, such as the information packets flowing through the network of routers and traffic on roads, use more intelligent routing algorithms [11] rather than a random walk. To check the robustness of results in Eqs. (8) and (9), we implemented the random walk simulation with the constraint that the traffic from node i to j takes the shortest path (SP) on the network. If multiple shortest paths are available to go from node i to j, the algorithm chooses any one of them with equal probability. Thus, in this setting,

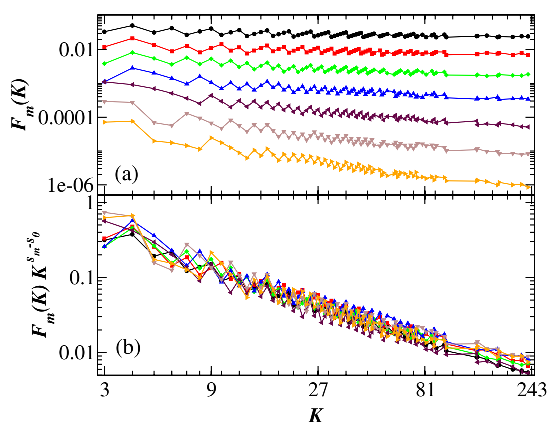

FIG. 2 (color online). Probability for occurrence of extreme events for several values of threshold $q = \langle f \rangle + m\sigma$. (a) shows the extreme event probabilities in log-log plot obtained from simulations with $\Delta = 0$. (b) shows scaling EE probabilities. S_0 represents the reference slope with $m = 2$. The threshold applied for curves from top to bottom are $m = 2.0$, 2.5, 3.0, 3.5, 4.0, 4.5, and 5.0.

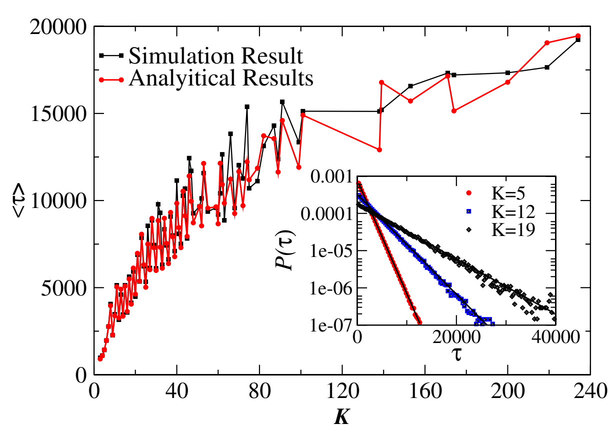

FIG. 3 (color online). The inset shows the recurrence time distribution for extreme events from simulations (symbols) with $\Delta = 0$ for nodes with 5, 12, and 19 links. The solid line is the analytical distribution. The main figure shows the mean recurrence time as a function of degree K.

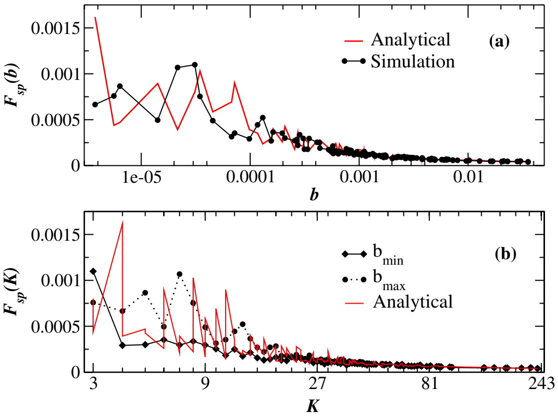

FIG. 4 (color online). Extreme event probability F_{sp} for $\Delta = 0$ with shortest path algorithm implemented for random walkers. The data are plotted in two different ways. (a) $F_{sp}(b)$ as a function of betweenness centrality, (b) $F_{sp}(K)$ as a function of degree K of the node. Nodes with same value of K can have different betweenness centrality. In (b), in order to reduce the clutter, for every value of K, the extreme event probability for the node with largest (b_{max}, solid circles) and least value (b_{min}, solid square) of b is plotted.

for every random choice of source-destination pair the paths are laid out by the algorithm and randomness arises only when multiplicity of SPs are available. Thus, this can be thought of as a walk with a large deterministic component. The simulation results with the SP algorithm [12] shown in Fig. 4 are qualitatively similar to the trend displayed in Fig. 1. In this scenario of predominantly deterministic dynamics, it is conceivable that the degree of a node does not determine the flux passing through it. This role is played by the centrality of the node with respect to the SPs in the network, quantified by the betweenness centrality b of a given node [13]. Based on this qualitative argument, the results in Fig. 4 can be understood if we replace Eq. (2) with $p = \beta b/B$ where B is the normalization factor that depends on the sum of betweenness centrality of all the nodes on the network. From the numerical simulations, we obtain $\beta \approx 0.94$. Using this p in Eq. (2), we can go through the same arguments as before and analytically obtain $\langle f \rangle$, σ^2, q, and the probability $F_{sp}(b)$ for occurrence of EE. In Fig. 4(a), $F_{sp}(b)$ is shown as solid curve. In Fig. 4(b), the same data for $F_{sp}(b)$ are shown as a function of K for easier comparison with Fig. 1. Thus, even with the SP algorithm thrown in, the EE probabilities are higher for the nodes with smaller degree ($K < 20$) than for the ones with larger degree ($K > 100$).

Finally, we comment on how these results can be applied as a basis to design nodes of a network. The central result in this Letter in Eq. (8) allows us to *a priori* estimate the EE probabilities. These depend on whether operating principle of dynamics is deterministic or probabilistic. If the idea is to avoid congestion or other problems arising due to EE of certain magnitude, then these estimates can be used as an input to the design principles for the nodes. For instance, for the road traffic that operates broadly on the shortest path principle the probabilities can be used as a basis to provision for higher capacity to nodes that will avoid bottlenecks arising from EE of a given magnitude.

In scale-free networks, small degree nodes form the bulk and are more prone to encounter EE. But network design principles and practice generally focus on the hubs. Such evolved practices might work best most of the time. Our work suggests that they might fail in the context of extreme events and hence a revised approach is necessary. A careful design for the capacity of small degree nodes is important as well. It must be emphasized that incorporating such EE estimates in design principles will only help in better preparedness to meet the expected EE. The EE discussed here being due to inherent fluctuations will nevertheless take place and cannot be avoided.

The simulations were carried out on computer clusters at PRL, Ahmedabad and IISER, Pune.

[1] *Extreme Events in Nature and Society*, edited by S. Albeverio, V. Jentsch, and Holger Kantz (Springer, New York, 2005).
[2] S. Boccaletti *et al.*, Phys. Rep. **424**, 175 (2006); C. Nicolaides, L. Cueto-Felgueroso, and R. Juanes, Phys. Rev. E **82**, 055101(R) (2010); V. Tejedor, O. Bénichou, and R. Voituriez, Phys. Rev. E **80**, 065104(R) (2009).
[3] J. D. Noh and H. Rieger, Phys. Rev. Lett. **92**, 118701 (2004).
[4] D. De Martino *et al.*, Phys. Rev. E **79**, 015101(R) (2009); P. Echenique *et al.*, Europhys. Lett. **71**, 325 (2005); K. Kim *et al.*, Europhys. Lett. **86**, 58002 (2009); B. Tadic, G. J. Rogers, and S. Thurner, Int. J. Bifurcation Chaos Appl. Sci. Eng. **17**, 2363 (2007).D. J. Ashton, T. C. Jarrett, and N. F. Johnson, Phys. Rev. Lett. **94**, 058701 (2005); L. Zhao *et al.*, Phys. Rev. E **71**, 026125 (2005); R. Germano and A. P. S. de Moura, Phys. Rev. E **74**, 036117 (2006).
[5] Wen-Xu Wang *et al.*, Chaos **19**, 033106 (2009).
[6] F. Reif, *Statistical Physics* (McGraw-Hill, New York, 1965); C. Song *et al.*, Nature Phys. **6**, 818 (2010); D. Brockmann *et al.*, Nature (London) **439**, 462 (2006).
[7] J. F. Eichner *et al.*, Phys. Rev. E **75**, 011128 (2007); M. S. Santhanam and Holger Kantz, Phys. Rev. E **78**, 051113 (2008).
[8] M. Argollo de Menezes and A.-L. Barabási, Phys. Rev. Lett. **92**, 028701 (2004).
[9] S. Meloni *et al.*, Phys. Rev. Lett. **100**, 208701 (2008).
[10] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1964).
[11] D. Methi and K. Ramasamy, *Network Routing: Algorithms, Protocols and Architectures* (Morgan Kaufmann, San Francisco, 2007).
[12] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and S. Clifford, *Introduction to Algorithms* (MIT Press, Cambridge, 2001).
[13] M. Barthelemy, Eur. Phys. J. B **38**, 163 (2004); H. Wang, J. M. Hernandez, and P. V. Mieghem, Phys. Rev. E **77**, 046105 (2008); S. Dolev *et al.*, J. ACM **57**, 1 (2010).