Gradient Networks

It has recently been recognized that a large number of systems are organized into structures best described by complex networks, or massive graphs. Many of these networks, also called *scale-free networks*, such as citation networks, the www, the internet (Fig. 1), cellular metabolic networks, the sex-web, the world-wide airport network possess power-law degree distribution, $P(k) \sim k^{-\gamma}$. Scale-free networks are very different from pure random graphs, which are well studied in the mathematical literature, and which have "bell curve" Poisson degree distributions. (Degree is the number of neighboring nodes a node is connected to via direct links.) Therefore, it is natural to ask: Why do scale-free networks emerge in nature? The diverse range of systems in which scale-

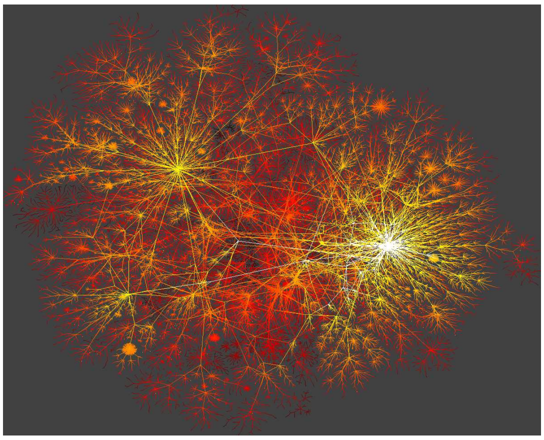

Figure 1: Skitter data depicting a macroscopic snapshot of Internet connectivity, with selected backbone ISPs colored separately.

free networks appear suggests that perhaps there is a simple common reason for their development.

Generally, real-world networks do not form or evolve simply by purely random processes. Instead, networks develop in order to fulfill a *main function*, which is to *transport* various entities such as information, cars, power/energy, water, forces, etc. All these large-scale networks mentioned above are non-globally designed. They evolve and grow through local changes, through a natural selection-like dynamics. For example, if a router on the internet is frequently congested and packets are lost or delayed due to that, it will get replaced by several interconnected new routers, and/or the connections rewired in that router's neighborhood. The question is whether such local "improvement" dynamics (local attempts for optimization) will constrain the global network structure. It is plausible that the structure that the network evolves into (scale-free in particular) will be one that ensures efficient and robust transport.

In order to investigate the connection between flow processing efficiency and network structure, we need to first define a flow dynamics on the network [1]. In particular, we consider the case when the flows are generated by gradients of a scalar field distributed on the nodes of a network. This approach is motivated by the idea that transport processes are often driven by local gradients of a scalar. Examples include electric current which is driven by a gradient of electric potential, and heat flow which is driven by a gradient of temperature. An example where gradient-induced transport on complex networks plays an important role is diffusive load balancing used in distributed computation (and also employed in packet routing on the internet). In this case, a computer (or a router) asks its neighbors on the network for their current job load (or packet load), and then the router balances its load with the neighbor that has the *minimum* number of jobs to run (or packets to route). Thus, the scalar at each node is the negative of the number of jobs at that node, and the *flow* occurs in the direction of the gradient of this scalar in the node's network neighborhood.

In order to construct a simple and general model of a transport process, assume that there are N nodes, and that the transport takes place on a fixed *substrate network* S which describes the interconnections of the nodes. Associated with each node i is a non-degenerate random number h_i which describes the "potential" of the node. Then a *gradient network* $\boldsymbol{G}$

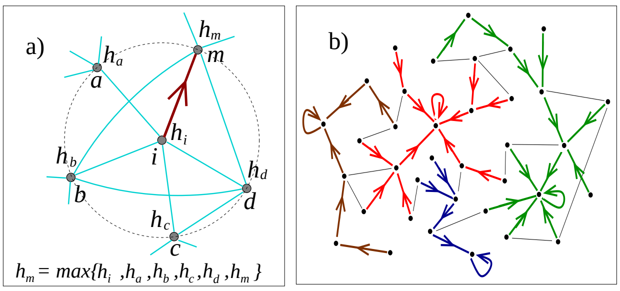

Figure 2: The gradient network. a) the gradient at node i is a directed edge pointing towards the largest increase of the scalar potential in the node's neighborhood. b) an example of a gradient network.

can be constructed as the collection of directed links that point from each node to the nearest neighbor of that node on the substrate network S which has the highest potential, see Fig. 2. Of course, in general, the potential for each point can evolve in time, and as a result, the gradient network $\boldsymbol{G}$ will be time-dependent. If we furthermore assume that all links have the same 'conductance', or transport properties, the gradient network will describe the *instantaneous* substructure carrying the maximum flow. It can be shown that all non-degenerate (the probability that two neighboring nodes having exactly the same scalar value is zero) gradient networks are forests, i.e. they have no loops (except for self-loops), and consist only of trees. Furthermore, if S is a simple random graph, in which each pair of nodes is linked with probability p, and the scalars h_i are i.i.d. random variables, then the distribution of the number of links pointing to each node (the in-degree distribution) on the gradient network becomes the power law [2]:

$$R(l) \simeq \frac{c}{l}\ , \quad 0 < l < z\ , \tag{1}$$

where c is a constant. This is valid in the the limit $N \to \infty$ and $p \to 0$, such that $Np = z = const. \gg 1$, see Fig. 3a). Therefore, in this "scaling" limit, *gradient networks are scale-free networks* (being characterized by a power-law $\sim l^{-\gamma}$, $\gamma = 1$). This behavior is rather surprising, since the substrate network S is *not* scale-free, and in the same limit has a "bell curve" Poisson degree distribution.

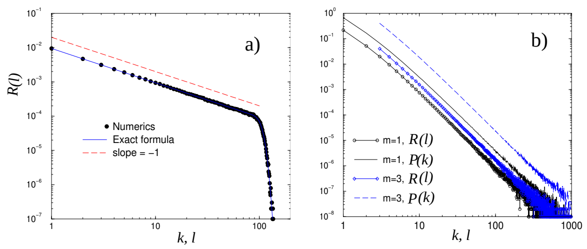

Figure 3: a) comparison between the exact formula (not shown here, see [1]) and numerics. b) the degree distributions of the gradient network and the substrate, when the substrate is a Barabási-Albert scale-free graph.

Alternatively, if the substrate network S is scale-free, for example, a Barabási-Albert (BA) network, then the associated gradient network *is also a scale-free network*, see Fig. 3b). In this case, the gradient network has the same type of structure as the substrate network, i.e., it is a scale-free (power law) graph characterized by the same exponent.

This observation has an important consequence on the efficiency of flow processing. If the flow is processed in non-zero time at the nodes, such as in the case of the internet, where packets have to be read and redirected, which is a non-instantaneous, physical process, *queues* may be generated at nodes. It is easy to see that the total length of queues at an instant is proportional to the fraction of nodes having at least one incoming link on the corresponding gradient network. Thus, the notion of gradient networks allows a transport characteristic related to congestion or jamming in the substrate network to be defined:

$$J = 1 - \left\langle \left\langle \frac{N_{receive}}{N_{send}} \right\rangle_h \right\rangle_{netw} = R(0) \tag{2}$$

where $N_{receive}$ is the number of nodes that receive gradient flow, and N_{send} is the number of nodes that send it. The value of J is always between 0 and 1, with $J = 1$ corresponding to maximal congestion (vanishing number of receivers/processors), and $J = 0$ corresponding to no congestion. Note that J is rather a *congestion pressure* characteristic generated by gradients, than an actual throughput characteristic. For a random graph substrate network the expression of J can be calculated to give the asymptotic behavior in the large network scaling limit, $p = const.$ and $N \to \infty$ as $J \simeq 1 - \frac{\ln N}{N \ln\left(\frac{1}{1-p}\right)}$. Therefore, in that limit *random networks become maximally congested*. However, for scale-free networks the conclusion in the same limit is drastically different. In that case, J tends to a positive constant bounded away from unity, i.e., scale-free networks are not prone to jamming. Figure 4 shows as comparison the congestion factors as function of network size both for random and scale-free substrate networks. Although we

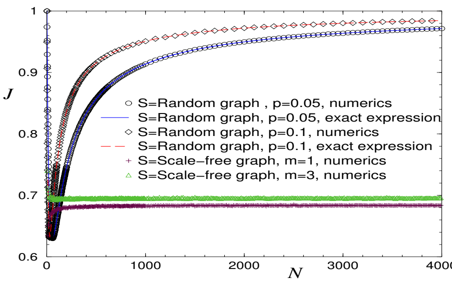

Figure 4: The congestion coefficient for random graphs) and scale-free networks.

have not specified a network evolution mechanism, what we have shown is that from the point of view of efficient flow processing, scale-free structures will more likely be selected for global structure (consistent with observation), than non-scale-free structures, like random graphs.

References

[1] Z. Toroczkai, K.E. Bassler, *Nature*, **428**, 716 (2004).

[2] Z. Toroczkai, B. Kozma, K.E. Bassler, N.W. Hengartner and G. Korniss. Gradient Networks, *submitted* (2005); cond-mat/0408262;

Contact Information: Zoltán Toroczkai – Center for Nonlinear Studies and Complex systems Group, Theoretical Division, Los Alamos National Laboratory, MS-B258, Los Alamos, NM, 87545, Phone: (505) 667-3218, email: toro@lanl.gov.