Guided Search and Distribution of Information Flow on Complex Graphs

Bosiljka Tadić

Jožef Stefan Institute, Box 3000, 1001 Ljubljana, Slovenia Bosiljka.Tadic ijs.si http://phobos.ijs.si/~tadic/

Abstract. Within the numerical model proposed in [1] we study the flow of information on a graph with power-law organization of links and an emergent superstructure formed by two highly interconnected hubs. The local search algorithms which navigate transport at each node can use information in the range of zero, one, and two-nearest neighbourhood of the node. We show how the flow carried by a particular link is distributed over graph when range of guided search and posting rate of packets are varied. The probability density function of the flow is given by a universal log-normal law, which is determined by the overall packet density in the stationary traffic. However, the distribution becomes unstable when the traffic experiences temporary or permanent jamming.

1 Introduction

The complexity of structure and function of evolving networks is often expressed in emergence of topological communities [2,3,4] and functionally connected units, which are based on the linking properties of the network. The inhomogeneity and sparseness of complex networks effect dynamic processes that these networks support. Recent study of traffic on networks of diverse emergent topologies [1, 5] revels that functional efficiency of a network, however, is not determined by the network's topology alone, but it also depends on the microscopic rules of the dynamics. In particular, in the transport processes on networks the rules which are adapted to the locally changing networks' structure lead to more efficient network performance. In the case of random walk dynamics the influence of the structural diversity can be incorporated into search algorithms that navigate the walkers through the graph [6].

Inhomogeneity on the level of local connectivity of nodes makes possible to design variety of navigations rules on complex networks. The requirement for locality of the search is essential for costs reasons. In [1] we have studied two different search algorithms that are guided by locally encountered network structure. It has been found that (a) better guided search leads to better performance of the network; and (b) the same guided search algorithm performs better on networks with higher organizational complexity [1]. An example which illustrates these conclusions is the traffic of information packets on the Web-type graph [7]

M. Bubak et al. (Eds.): ICCS 2004, LNCS 3038, pp. 1086-1093, 2004.

[©] Springer-Verlag Berlin Heidelberg 2004

with the next-near-neighbour (nnn-) search. The network's output was found to be 40 times higher than when the rules of random diffusion are applied.

These results suggest that in artificially grown networks, although the topological complexity of the network is essential, the functional units can not be identified with topological communities independently on the nature of the dynamic process that the network performs. This motivates study of details of the network's activity as it performs a particular type of function. A detailed knowledge about the distribution of activity throughout the network may be then used as a starting point for restructuring of the network in order to improve its function at local levels. Another approach of network's restructuring to optimize its performance under global condition was suggested in [8].

Here we use the numerical model, which was recently introduced in [1] for information packet transport on structured graphs, to study the intensity of traffic carried along a particular link (flow) on the cyclic scale-free Webgraph. In particular, we determine how the distribution of flow depends on the radius of the area around visited node which is explored by the guided search rules, and on the packet density.

2 Network Topology and Search Algorithms

The network which we use here is represented by a directed graph (Web graph) grown with the microscopic dynamic rules which are proposed in [7] to model the evolution of the world-wide Web. The rules consist of preferential attachment and rewiring while the graph grows, which leads to an emergent structure with inhomogeneous scale-free ordering in both in-coming and out-going links and a number of closed cycles, as shown in Fig. 1. The statistical properties of the graph topology are studied in detail [7,9], in particular, the local and global connectivity, a giant component and clustering (number of triangles per node). A substantial degree of correlations between in- and out-link connectivity at local level was found [10] which is characteristic for dis-assortative [2] graph topologies. Further topological properties—betweenness of nodes and links—are studied by use of random walk methods [11,2]. An important feature of the Web graph, which makes it different from common scale-free graphs [12] is the emergence of two types of hub nodes—a node with high in-link connectivity and a node with high out-link connectivity—which are closely interlinked by a number of short triangles. The superstructure of interlinked nodes which is associated with the two hubs appears to play an important role in the transport processes on the graph [1,13].

The diffusion rules for *dense* traffic of packets on the graph are summarized as follows (see [1] for detailed description and [10] for details of the implementation of the numerical code):

Packets are created with a finite rate R at a node and each packet is assigned an address (destination) where it should be delivered. Both the origin node and destination node are selected randomly within the giant component of the graph. The packets are navigated through the nodes by one of the search rules given



Fig. 1. (a) Webgraph with different widths of links representing four bins of flow: low, average, high, and very high flow in the case of *nnn*-search. (b) Only high flow links are displayed (solid lines), with the remaining graph connections shown by dotted lines.

below. The priority of move at a node is given according to the "last-in-firstout" (LIFO) queuing mechanism. When the packet arrives to its destination it is removed from the traffic.

The three local search algorithms which we consider here differ in the range of area where the search for the destination node is performed in the vicinity of the visited node. If the node just visited by the packet is not its destination node, the search performed around that node can be within

- range zero, when the packet proceeds by randomly selecting a neighbour node on the network;
- range one, where the search for the destination is asked among first neighbours of the visited node and, if found, packet is delivered to it, else it proceeds to a randomly selected neighbour;
- range two, where through all first neighbours the signaling to next-nearneighbours (nnn) of the visited node is done; if the message is returned by a neighbour of the destination node, then the packet is delivered to that (common) neighbour, else it proceeds to a randomly selected neighbour node.

The packet transport on the same network appears to be markedly different when different search algorithm is used. The efficiency of the transport measured by the transit times of packets increases with the range of search. The distributions of the transit times of packets in Fig. 2 show power-law dependences



Fig. 2. Distribution of transit times for three algorithms of search in ranges 0, 1, and 2, and for fixed posting rate R = 0.005. The slopes of solid lines are: 1.5, 1.2, and 0.25. All data are averaged over many network realizations and log-binned.

with different slopes and cut-offs for different search algorithms. The maximum transit time increases in the case of less effective search, leading to increased accumulation of packets in the traffic (small output rate) and the network enters jamming regime at much lower posting rate, as compared to the advanced search algorithms. For shortest transit times T = 1, resulting in the case when selected pairs of nodes are at unit distance, the search rules with the range one and range two coincide.

3 Information Flow Distribution

The topological importance of a particular node on the graph is measured by betweenness, which is defined as the number of shortest paths between pairs on nodes on the graph that pass through that node. An analogous property, betweenness-centrality, is defined for links in the graph theory [14]. Here we consider a dynamically generalized betweenness-centrality called *flow* [14], by which we measure the importance of a particular link for the packet traffic on the graph. The actual paths of packets differ from topologically shortest paths, being subjected to navigation and queuing processes.

The flow of the traffic which is carried by a particular link is given by the number of packets transferred along that link within given transport process. In



Fig. 3. Distribution of flow for three search algorithms of information range (right to left) zero, one, and two, and for fixed posting rate (R = 0.005) and time interval.

the packet transport between pairs of nodes on the graph we identify the exact path of a packet and update the flow on each link along the path. Depending on local network structure and the search algorithm, in the same time interval and fixed posting rate certain links carry much higher flow compared to the others. An example is shown in Fig. 1a: the intensity of flow on the links of the graph is expressed by the increased width of a link (within four bins). The results for the probability density function of flow along a link on the graph for different situations studied in this work are shown in Figs. 3 and 4 (averaged over network realizations and log-binned). The distribution appears to be close to a log-normal law (cf. Figs. 3 and 4).

Fig. 3 demonstrates the influence that the three search algorithms with different ranges of information have on the probability density of flow carried by a link for identical posting rates and time intervals. The efficient search algorithm (i.e., with the information range two) removes packets from the traffic quite quickly after their creation. Therefore, although certain links appear to carry quite large flow, the overall distribution is centered at low mean value. The location of links with heavy flow on the graph are shown in Fig. 1b for the case of the most effective search. On the other hand, in the case of less effective searches, packets remain on the network for much longer time, increasing the overall packet density. Consequently, the probability of large flow along a link will increase. The mean value of the distribution is shifted towards increasingly larger values when the range of the search algorithm is reduced (cf. Fig. 3).



Fig. 4. Distribution of flow in the case of nnn-search (a) for different values of posting rate R = 0.005, 0.01, 0.05, 0.1, and 0.3 (left to right) for fixed time (40000 timesteps), and (b) for fixed number of packets (2000 packets) in different time intervals, corresponding to stationary flow rates R = 0.005, 0.01, 0.05, 0.1 (collapsed curves), and temporary and permanent jamming rates R = 0.3, and 0.4 (shifted curves).

Next we consider the flow distribution in the case of the efficient search algorithm (i.e., for the searched range in next-near-neighbourhood) for different density of packets. When the number of posted packets increases by increasing posting rate R within same time interval, the mean of the flow distribution moves towards higher values, as shown in Fig. 4a. In the conditions shown in Fig. 4a the traffic remains stationary, which can be checked by monitoring the work-load timeseries [13]. In order to further investigate this effect, we determine the flow distribution for the conditions when the time interval and posting rate vary so that the overall number of created packets remains fixed. The results are shown in Fig. 4b. For the posting rates where the traffic is stationary, the distributions for different posting rates collapse. Whereas, the flow distribution deviates from the universal curve for larger rates when the network experiences temporary "crisis" [13] which the system manages to clear out within long time intervals, or when permanent jamming occurs.

4 Conclusions

The topology of underlying graph plays an important role in the transport processes on networks. For example, the graph with high complexity, as the Web graph described in Section 2, can carry huge amount of traffic before getting jammed. The detailed analysis of traffic in terms of the transit time and flow statistics reveals the importance of search algorithm in the transport of information packets to specified destinations on the graph. In particular, the global efficiency of the traffic is directly correlated with the radius of the locally searched area around a node in the navigation process. The maximum transit time of packets decreases by a factor of two to ten when the range of search is increased from zero- to one- or two-near neighbours of a node. The distribution of flow along the links changes accordingly with the efficiency of the local search. Under the identical other conditions in the less effective local search the packets accumulation on the network, which leads to generally increased flow along each link. The mean of the flow distribution moves towards larger values.

On the global level, the probability distribution of flow appears to be close to a log-normal law. It is unstable with respect to random addition processes, which in the case of traffic is realized in the conditions of increased rate of packet creation. We find that for given (effective) search rule in the stationary traffic the distribution of flow along links of graph is universal when the overall number of created packets is fixed, independently on the time interval in which the packets are created. The observed deviations from the universal curve are correlated with the volatile fluctuations of the packet density (temporary jamming) or permanent jamming in the network. These results are in agreement with other findings based on queuing and network-load timeseries analysis [13], which suggest the occurrence of three distinct phases in the transport on complex graphs: stationary free flow, stationary flow with temporary jamming, and jammed traffic. In the artificially grown networks the detailed knowledge of the flow distribution of packet traffic for a given search rule can be used to rebuild the network structure so that it performs that transport process with higher efficiency.

References

- Tadić, B., Thurner, S.: Information Super-Diffusion on Structured Networks. Physica A 332 (2004) 566-584; cond-mat/0307670
- Newman, M. E. J.: The structure and function of complex networks. SIAM Rev. 46 (2003) 167-225
- 3. Zhou, H.: Network landscape from a Brownian particle's perspective. Phys. Rev. E 67 (2003) 041908
- 4. Arenas, A., Danon, L., Dıaz-Guilera, A., Gleiser, P.M., Guimerà, R.: Community analysis in social networks. Europ. Phys. J. B (in press).
- Tadić, B., Rodgers, G.J.: Packet transport on scale-free networks. Advances in Complex Systems 5 (2002) 445-456
- 6. Tadić, B.: Adaptive random walks on the class of Web graphs. European Physical Journal B, **23** (2001) 221-228
- 7. Tadić, B.: Dynamics of directed graphs: the world-wide Web. Physica A **293** (2001) 273-284
- Guimera, R., Diaz-Guilera, A., Vega-Redondo, F., Cabrales, A., Arenas, A.: Optimal network topologies for local search with congestion. Phys. Rev. Lett. 86 (2001) 3196-3199
- 9. Tadić, B.: Temporal fractal structures: origin of power laws in the world-wide Web. Physica A **314** (2002) 278-283
- Tadić, B.: Modeling Traffic of Information Packets on Graphs with Complex Topology. Proceedings ICCS 2003, P. Sloot *et al.* Eds., Lecture Notes in Computer Science, Part I, **2567** (2003) 136-142, Springer (Berlin)
- Tadić, B.: Exploring Complex Graphs by Random Walks. Proceedings of The Granada Seminar on Computational Physics "Modeling Complex Systems", P.L. Garrido and J. Marro (Eds.), AIP Conference Proceedings 661 (2002) 24-27
- Albert, R., Barabasi A.-L.: Statistical Mechanics of Complex Networks. Reviews Mod. Phys. 74 (2002) 47-101
- 13. Tadić, B., Thurner, T., Rodgers, G. J.: Traffic on complex networks: Towards understanding global statistical properties from microscopic density fluctuations. Phys. Rev. E (in press).
- 14. Bollobás, B.: Modern Graph Theory. Springer (New York) 1998.