Inferring and Visualizing Topological Structures of Large-scale Complex Network

Yuchun Guo
School of Electrical and Information Engineering
Beijing JiaoTong University, China
+86 10 51684759

ychguo@bjtu.edu.cn

Changjia Chen
School of Electrical and Information Engineering
Beijing JiaoTong University, China
+86 10 51684759

changjiachen@sina.com

Shi Zhou
Department of Computer Science
University College London, United Kingdom
+44 (0) 1473 663705

s.zhou@adastral.ucl.ac.uk

ABSTRACT

Many information and communications networks are very large and they exhibit extremely complex structures. Here we propose a topology visualisation tool, called the bitmap of sorted adjacency matrix (BOSAM), which illustrates the connectivity information of a network as a bitmap image. We show that by using carefully designed rules to sort network node in specific orders, the produced bitmaps are able to reveal interesting patterns associated with the network structures. This simple tool can facilitate the study on characterising and modelling large-scale complex networks such as the Internet and peer-to-peer (P2P) networks.

Categories and Subject Descriptors

C.2.1 [**Network Architecture and Design**]: Network topology;
I.6 [**Simulation and Modeling**]: Model Validation and Analysis

General Terms

Algorithms, Measurement, Verification.

Keywords

Topology inference, visualization, bitmap, index sorting.

1. INTRODUCTION

Structure, or topology, of large-scale complex networks [1] is often characterised by using a statistical physics approach, i.e., methods of statistics. For example a relevant property regards the degree of a node, that is the number of its connections to other nodes. The degree distribution $P(k)$ is the probability that a node in a graph having degree k. It is discovered that in many information and communications networks, such as the Internet [2] and the World Wide Web [3], the degree distribution exhibits a power-law tail with an exponent between 2 and 3 [4], [5], which significantly deviates from the Poisson distribution expected for a random graph [6].

Topology visualisation is a technique to illustrate the topology

information of a network as an image. This technique is useful because images can be more intuitive and intelligible than abstract mathematical descriptions, especially when a network is very large and complex. The existing visualization schemes display nodes and links with dots and lines respectively in a geometrical space with layouts depending on certain topological or geometrical or hybrid metric, e.g. k-coreness [7], outdegree and longitude [8]. Visualisation of some topological metric, such as the rich-clubness proposed in [9], is also used in research. In this paper we introduce a topology visualisation tool, called the bitmap of sorted adjacency matrix (BOSAM). BOSAM displays a network topology with a bitmap in which a link is represented as a dot in a grid position corresponding to the incident nodes. The bitmap layout can be transformed by sorting the node indexes in different ways to show the topology connectivity correlation from different perspectives. Some interesting patterns in topologies revealed with this tool are presented to show BOSAM is helpful for characterising and assessing large-scale complex topologies.

2. BITMAP OF SORTED ADJACENCY MATRIX (BOSAM)

If a network contains N nodes and each node is assigned a unique index $i(i=1,2,3,...,N)$, the network's connectivity information can be represented as an $N \times N$ matrix, called adjacency matrix. The matrix's entry a_{ij} is the number of links connecting between node i and node j. In this paper we consider simple undirected graphs (i.e., no self-loop and no multiple links between any pair of nodes), the adjacency matrix is then a symmetric (0, 1)-matrix with zeros on its diagonal. As shown in Figure 1, such an adjacency matrix can then be visualised as a black-and-white bitmap, on which a black pixel represents a connection between two nodes and the sum of black pixels on the ith column (or the jth row) equals to the degree of node i (or j).
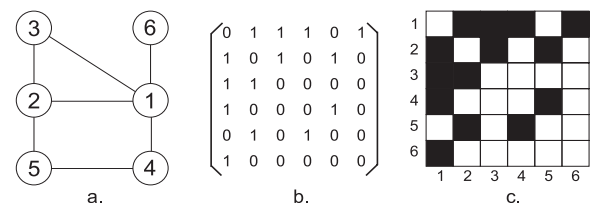

Figure 1. Visualisation of a network as a bitmap. (a) A network of 6 nodes. (b) Representation of the network's connectivity information as a 6×6 adjacency matrix. (c) Illustration of the matrix as a 6×6-pixel bitmap.

The bitmap of sorted adjacency matrix (BOSAM) is special in that it sorts a network's nodes in specific orders using carefully designed rules and assigns node indexes accordingly so that the produced bitmaps show meaningful patterns that associate with the network's topological properties. Thus the node-sorting rules are the key for the success of BOSAM. Firstly we sort nodes in a list of decreasing degrees, e.g. the best-connected node is at the top of the list and the second position if the second best-connected node. However there are always many groups of nodes having the same degree. Here we introduce two node-sorting rules which take into account not only a node's degree but also degrees of its neighbouring nodes, or neighbour degrees.

Rule 1: After sorting node degrees in a decreasing order, for each group of nodes with the same degree, we sort the largest neighbour degree of the nodes in a decreasing order and rearrange the sorting list accordingly. For those who have the same largest neighbour degree, we then sort their second largest neighbour degree in a decreasing order. For example in Figure 1, there are three 2-degree nodes, node 3, 4 and 5. The largest neighbour degree of node 3 and node 4 is larger than that of node 5, and the second largest neighbour degree of node 3 is larger than that of node 4. This process continues until all neighbour degrees have been considered. For nodes whose degree and all neighbour degrees are same, they are positioned in the list next to each other in arbitrary orders.

Rule 2: We sort nodes in a similar way as above, but we sort the smallest neighbour degree of nodes with the same degree in an increasing order. If necessary we do the second smallest neighbour degree in an increasing order and so on.

We assign each node an index according to its rank, or the node's position in the sorted list. The above two rules sort nodes in different orders and thus we can produce two BOSAMs for a given network. The rationale underlying the two rules is as follows. Suppose in a network high-degree nodes cluster into a core and low-degree ones scatter peripherally, then the neighbours largest degree is a measure of the 'cohesion' force a node receives from the network core, and the neighbours smallest degree is a measure of the 'radiation' force the node receives from the peripheral. Recently a number of studies have shown that the joint degree distribution $P(k, k')$, i.e., the probability that a link connects k- and k'-degree nodes, plays a significant role in defining a network's structure [10].

The above two rules are presented to show how BOSAM works whereas the sorting rule can be designed by the researchers as desired to visualize the degree-degree correlation in a network from different viewpoints.

3. APPLICATIONS OF BOSAM

The Internet contains millions of routers, which are grouped into thousands of subnetworks, called autonomous systems (AS). The topology of the Internet at the AS-level is relevant to the Internet engineering because the delivery of data traffic through the global Internet depends on the complex interactions between ASes that exchange routing information using the Border Gateway Protocol (BGP). Figure 2 shows BOSAMs of the three Internet AS graphs studied in [10]: (i) WHOIS graph, extracted from RIPE's WHOIS database [11]; (ii) BGP-table graph, using BGP tables collected by the Route Views project [12]; and (iii) skitter graph, using

traceroute data collected by CAIDA's active probing tool *skitter* [13]. Besides, Figure 2 shows two other Internet AS graphs: (iv) DIMES graph, provided by the active measurement project DIMES [14] and (v) BGP-update graph, extracted from BGP updates colleted by Route Views project [12] (at the same time when the above BGP-table graph was collected). The size and order of these graphs are given in Table 1.

Table 1. the number of nodes and links of the networks

Network	Nodes	Links
WHOIS Internet AS graph	7,485	56,949
BGP-table Internet AS graph	17,446	40,805
BGP-update Internet AS graph	17,417	42,484
DIMES Internet AS graph	13,386	24,670
skitter Internet AS graph	9,204	28,959
ER random model	9,204	28,859
BA scale-free model	9,204	27,597
PFP Internet model	9,204	27,612
Gnutella P2P network	317,592	7,396,948

In the BGP-table and skitter bitmaps, black pixels are densely concentrated along the top and the left borders, where $i < 0.5k$ or $j < 0.5k$. This indicates two topological properties: (1) power-law degree distribution, that a small number of nodes have very large numbers of links; and (2) negative degree-degree correlation [1][15], that low-degree nodes tend to link with high-degree nodes and vice versa. The rich-club phenomenon [9] describes the fact that high-degree nodes are tightly interconnected with each other. In the WHOIS bitmaps, black pixels are densely concentrated at the top-left corner where $i < 2k$ and $j < 2k$. This suggests that the WHOIS graph by comparison exhibits a stronger rich-club phenomenon than the BGP-table and skitter ones. On the other hand, WHOIS bitmaps have much less empty white area that corresponds to the 1-degree nodes. This indicates WHOIS graph has less 1-degree nodes and is much denser compared with the former two AS graphs. These observations are in agreement with the numerical results presented in [10]. Observe bitmaps of other two AS graphs. The BGP update bitmaps look quite similar with the BGP-table ones except that less 1-degree nodes appear in the BGP-update graph, which means they are similar in the degree-degree correlation properties. This phenomenon has been noticed by the authors of [10] as well. The DIMES graph has not been numerically compared with other AS graphs in literature. The DIMES bitmaps show roughly similar patterns with the BGP and skitter bitmaps but with quite larger empty areas. This indicates that DIMES graph is sparser than other AS graphs but has roughly similar degree-degree correlation properties with the BGP and skitter graphs.

Figure 2 also shows BOSAMs of three network models: (i) the Erdös-Rènyi (ER) model [6], which generates random graphs; (ii) the Barabási-Albert (BA) model [5], which grows scale-free networks having a power–law degree distribution $P(k) \sim k^{-3}$; and (iii) the positive-feedback preference (PFP) model [16], which is an Internet topology generator. As shown in Table 1, all models are grown to the same number of nodes as the skitter Internet AS graph. The BOSAMs show that the models generate distinctive

topologies. It is clear the PFP model well resembles the skitter Internet AS graph.

Gnutella is a popular P2P file-sharing network. A snapshot of the Gnutella network obtained by a crawl from October 13 to November 11 in 2005 [17] contains more than three hundred thousands of nodes (the so-called reachable ultrapeers [18]) and seven millions of links. Such a scale makes the statistical analysis of degree-degree correlation intractably burdensome in computation. Via BOSAM the visualization of such scale graph is feasible to be accomplished efficiently. Figure 3 shows that Gnutella is profoundly different from power-law networks such as the Internet, the BA model and the PFP model in Figures 2. Instead it shows similarity to the ER random model. Moreover, the top-left (bottom-right) 1: 4^L ($L = 1 \sim 7$) corners of rule-1

BOSAMs are presented with different resolutions in the subplots of Figure 3. They clearly exhibit some fractal patterns recurring at different scales. Such fractal scaling properties have not been reported and can be interesting for further study.

The interesting patterns in the BOSAMs revealed here actually represent the degree distribution and joint degree distribution of the original graphs. Further analysis of the implications of these patterns and their relationships with the statistical properties and metrics is ongoing and will be reported in the future.

4. CONCLUSIONS

We show that by sorting network nodes in specific orders, BOSAM is able to illustrate a number of network topological properties. This simple tool is indicated to be scalable and
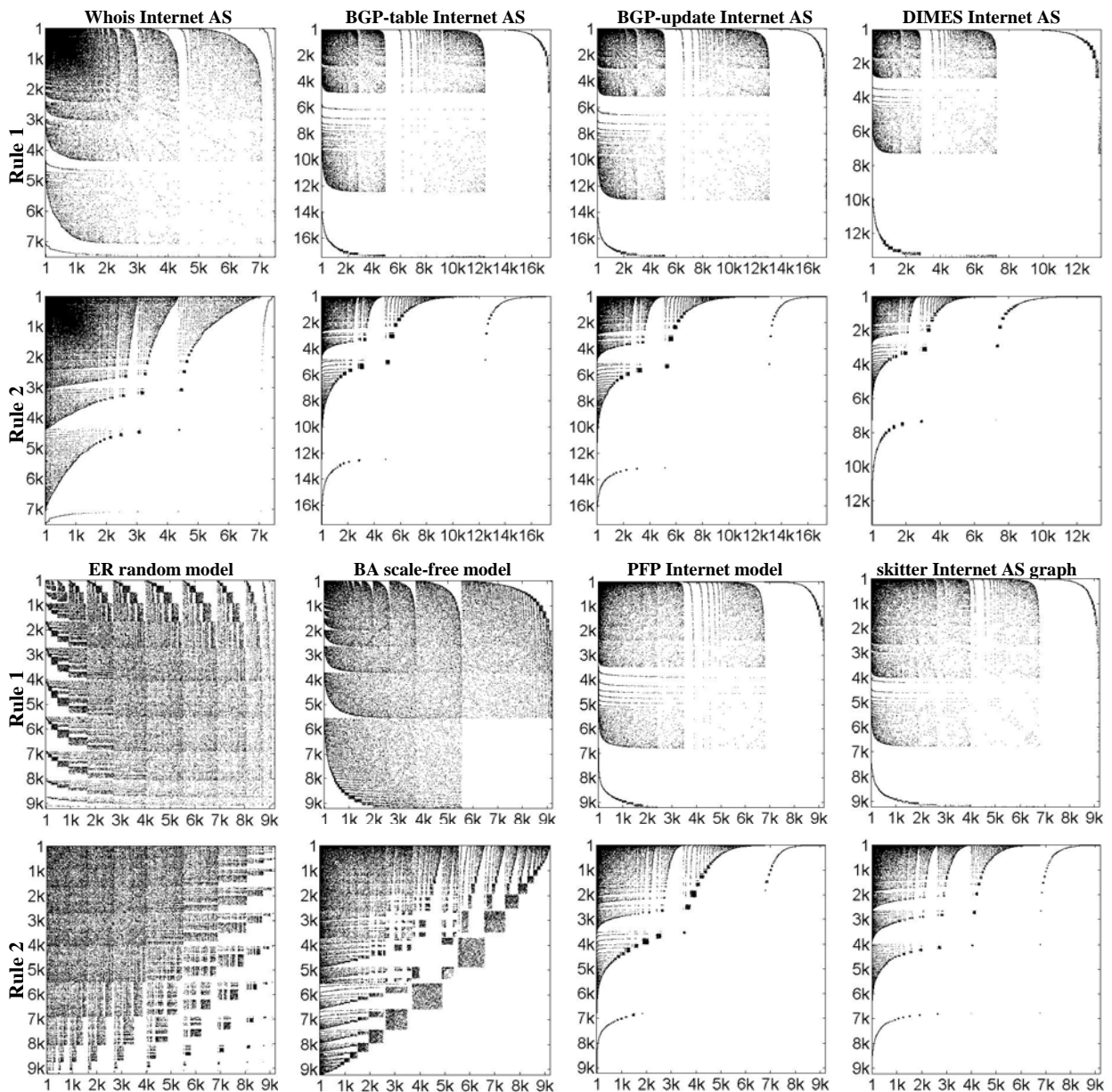

Figure 2. BOSAMs of 5 Internet AS graphs and 3 network models.

efficient for large-scale complex networks so that it can facilitate the study on large-scale complex networks. Some real networks and canonical topology models are visualized via BOSAM. The visual comparisons between these graphs are consistent with existing statistical researches and some more detailed correlation structures are exposed. The BOSAMs of a Gnutella snapshot with more than 300k nodes unveil the similarity between Gnutella network and the ER model instead of BA and PFP models. Furthermore, some interesting fractal patterns are also discovered in Gnutella topology. Though only two sorting rules are provided in this paper as examples to show how BOSAM works, users can add new ingredients to the node sorting rules to highlight other topological properties that they are interested.

5. REFERENCES

[1] M. Newman, A.-L. Barabási, and D. Watts, Eds., *The Structure and Dynamics of Networks*. Princeton University Press, 2006.

[2] R. Pastor-Satorras and A. Vespignani, *Evolution and Structure of the Internet - A Statistical Physics Approach*. Cambridge University Press, 2004.

[3] M. Levene and A. Poulovassilis, Eds., *Web Dynamics - Adapting to Changes in Content, Size, Topology and Use*. Springer, 2004.

[4] M. Faloutsos, P. Faloutsos, and C. Faloutsos, On power-law relationships of the Internet topology, *Comput. Commun. Rev.*, vol. 29, pp. 251–262, 1999.

[5] A. L. Barabási and R. Albert, *"Emergence of scaling in random networks,"* Science, vol. 286, pp. 509–512, 1999.

[6] P. Erdös and A. Rényi, *"On random graphs,"* Publ. Math., vol. 6, pp. 290–297, 1959.

[7] "Large Network visualization tool", http://xavier.informatics.indiana.edu/lanet-vi/.

[8] "Visualizing Internet Topology at a Macroscopic Scale", http://www.caida.org/analysis/topology/as_core_network/.

[9] S. Zhou and R. J. Mondragón, *"The rich-club phenomenon in the Internet topology,"* IEEE Comm. Lett., vol. 8, no. 3, pp. 180–182, 2004.

[10] P. Mahadevan, D. Krioukov, M. Fomenkov, B. Huffaker, X. Dimitropoulos,K. Claffy, and A. Vahdat, *"The Internet AS-level topology: Three data sources and one definitive metric,"* Comput. Commun. Rev., vol. 36, no. 1, pp. 17–26, 2006.

[11] "Internet Routing Registries," http://www.irr.net/.

[12] "University of Oregon RouteViews Project," http://www.routeviews.org/.

[13] kc claffy, T. E., Monk, and D. McRobb, *"Internet tomography,"* Nature, January 1999. http://www.nature.com/nature/webmatters/tomog/tomog.html.

[14] Yuval Shavitt and Eran Shir, DIMES - Letting the Internet Measure Itself, http://www.arxiv.org/abs/cs.NI/0506099.

[15] A. Vazquez, M. Boguñá, Y. Moreno, R. Pastor-Satorras, and A. Vespignani, *"Topology and correlations in structured scale-free networks,"* Phys. Rev. E, vol. 67, no. 046111, 2003.

[16] S. Zhou and R. J. Mondragón, *"Accurately modelling the Internet topology,"* Phys. Rev. E, vol. 70, no. 066108, 2004.

[17] "University of Oregon Ion P2P Project," http://mirage.cs.uoregon.edu/ P2P/.

[18] D. Stutzbach, R. Rejaie, and S. Sen, *"Characterizing unstructured overlay topologies in modern p2p file-sharing systems,"* in Proc. of the ACM Internet Measurement Conference (IMC), 2005.
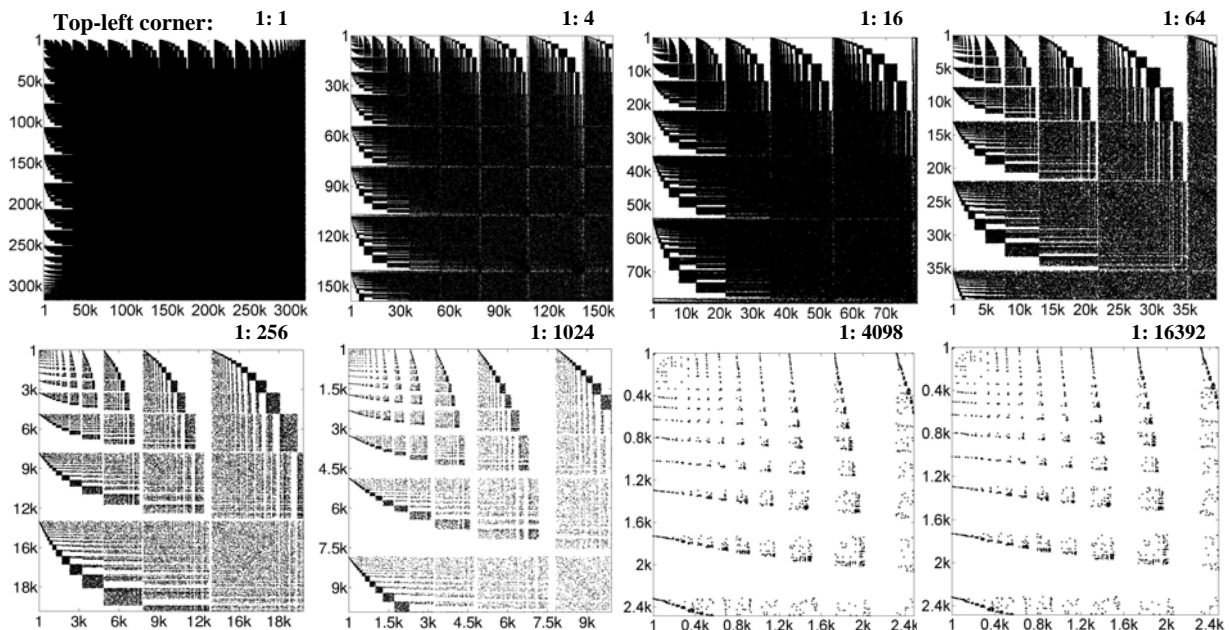
Figure 3. Rule-1 BOSAMs of the Gnutella P2P network and their enlargements at increasing scales.