

Why the Internet is so ‘small’?

Shi Zhou

Department of Computer Science, University College London
Malet Place, London, WC1E 6BT, United Kingdom
s.zhou@cs.ucl.ac.uk

Abstract. During the last three decades the Internet has experienced fascinating evolution, both exponential growth in traffic and rapid expansion in topology. The size of the Internet becomes enormous, yet the network is very ‘small’ in the sense that it is extremely efficient to route data packets across the global Internet. This paper provides a brief review on three fundamental properties of the Internet topology at the autonomous systems (AS) level. Firstly the Internet has a power-law degree distribution, which means the majority of nodes on the Internet AS graph have small numbers of links, whereas a few nodes have very large numbers of links. Secondly the Internet exhibits a property called disassortative mixing, which means poorly-connected nodes tend to link with well-connected nodes, and vice versa. Thirdly the best-connected nodes, or the rich nodes, are tightly interconnected with each other forming a rich-club. We explain that it is these structural properties that make the global Internet so ‘small’.

Key words: Internet, network, topology, autonomous systems, BGP, shortest path, power-law, scale-free, assortative mixing, rich-club

1 Introduction

The Internet is a network of autonomous systems (AS) which are collections of IP networks and routers under the control of one entity, typically an Internet service provider. The Border Gateway Protocol (BGP) is a critical component of the Internet’s infrastructure as it serves to connect these ASes together [1]. Performance of the BGP depends strongly on properties of the topological connectivity between these ASes. Surprisingly, it was only in 1999 that researchers reported that the Internet AS graph is a ‘scale-free’ network [2]. This discovery effectively invalidated all previous Internet models based on random graphs.

Since then researchers have reported many other topological properties of the Internet [3, 4]. In this paper we review three of them, namely the power-law degree distribution [5], the disassortative mixing [6, 7] and the rich-club phenomenon [8]. We explain that it is these structural properties that make the global Internet so ‘small’ in the sense that it is extremely efficient to route data packets across the global Internet.

So far much of the research on Internet topology occurs in the communities of physics, mathematics and complexity science. Interdisciplinary communica-

tion with other communities is much needed [9]. This paper is an effort towards this direction. The network under study is the Internet, but the statistical methods and the topological properties introduced here can be applied to any other network in nature and society.

2 Internet Topology at AS Level

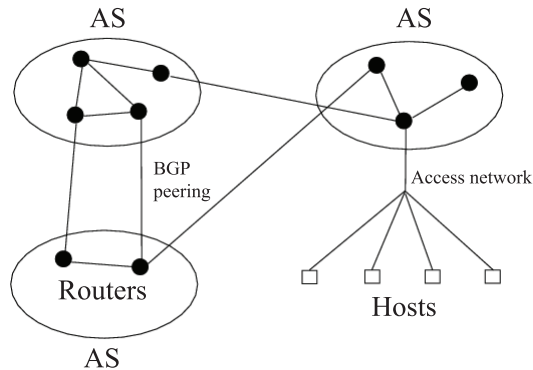


Fig. 1. Nodes of the Internet topology. (1) Hosts, that are the computers of users; (2) routers, that arrange data traffic across the Internet; and (3) autonomous systems (AS) that are subnetworks in the Internet.

The Internet is a capital example of complex networks. It is unlike any previous human invention in both scale and effect, and is now a global resource important to all of the people in the world. The Internet provides a low-level communication infrastructure upon which other communication mechanisms can be built, e.g. the ubiquitous email and Web protocols. Internet service providers (ISP) offer billions of users access to the Internet via telephone line, cable and wireless connections. Data packets generated by users are forwarded across the Internet toward their destinations by routers through a process known as routing.

The Internet contains millions of routers, which are grouped into thousands of subnetworks, called autonomous systems (AS) (see Fig. 1). The global Internet structure is characterised at the AS-level because the delivery of data traffic through the Internet depends on the complex interactions between ASes that exchange routing information using the Border Gateway Protocol (BGP). On the Internet AS-level topology, a node is an AS, which is usually controlled by an ISP; and a link represents a BGP peering relation between two ASes, which is a commercial agreement between two ISPs on forwarding data traffic.

Measurement data of the Internet AS-level topology became available in late 1990's. Since then there have been a number of projects to collect the topol-

ogy data of the Internet [4]. These measurements have greatly improved our understanding of the structure and evolution of the Internet [3]. The Internet macroscopic structure has become relatively stable in recent years. This suggests that the Internet has entered a fairly mature stage in terms of network evolution. Today the Internet AS graph is very large. It contains about twenty thousand AS nodes and fifty thousands of links. It is important to realise that the Internet is actually very much sparsely connected. The actual number of links present in the Internet is only about 0.04% of all possible links that the network can have. This is relevant because the fewer links the Internet has the cheaper and easier to invest, manage and maintain the network.

A remarkable property of the Internet is that it is extremely efficient in routing data packets across the global network. On the AS graph the average of shortest path between any two nodes is around 3 hops. The length of routing path observed on real BGP data is around 4 hops. This is in spite of the fact that the Internet contains tens of thousands of nodes and the network is very sparsely connected. Why is the Internet so ‘small’? The key lies in the topological structure of the Internet.

3 Node Connectivity

In graph theory, degree k is defined as the number of links or immediate neighbours a node has. The average node degree is $\langle k \rangle = \frac{1}{N} \sum_i k_i = 2L/N$, where k_i is the degree of node i , N is the number of nodes and L is the number of links. The average node degree in the Internet is around 6. The principal statistical property to describe and discriminate between different networks is to measure the degree distribution $P(k)$, i.e. the fraction of nodes in the network with degree k .

It had been assumed that the Internet topology resembles a random graph where nodes are connected with each other randomly (with a uniform probability). A random network features a Poisson degree distribution where most node degrees are close to the average degree and only a few nodes have very large or very small degrees. It is known from graph theory that a random graph is small in the sense that one node is never too far away from other nodes. Is this the reason that the Internet is so small?

The answer is no, because in 1999 it was discovered that the internet topology at the AS level (and at the router level) exhibits a power law degree distribution $P(k) \sim Ck^\gamma$ [5], where $C > 0$ is a constant and the exponent $\gamma \simeq -2.2 \pm 0.1$ (see Fig. 2). This means a few nodes have very large numbers of connections, whereas the vast majority of nodes have only a few links. The power law property is an evidence that the Internet AS-level topology has evolved into a very heterogeneous structure, which is completely different from the random graph model. A power-law network is called a ‘scale-free’ network [2] because it is not the average degree, but the exponent of power-law distribution that fundamentally characterises the network’s connectivity and the value of the exponent is not related to the size (scale) of the network.

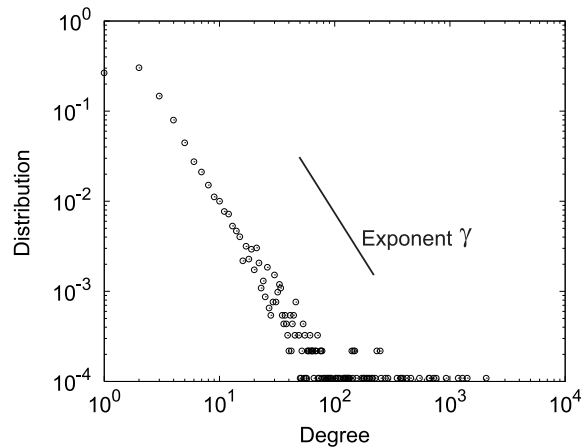


Fig. 2. Internet power-law degree distribution on log-log scale.

Recently the small-world theory [10] has attracted a lot of attention. A small-world network resembles a typical social network which is highly clustered where one's friends are likely to be friends to each other. At the same time a small-world contains random connections which serve as shortcuts and make the network small. Most scale-free networks can be regarded as small-world networks in the sense that the average shortest paths of scale-free networks are always small. But some scale-free networks are smaller than others. For example the average shortest path of many social networks is around 6 (which is comparable to that of random graphs); whereas the Internet's average shortest path is 3 which is actually substantially smaller than predicted by the random graph and small-world theories.

In the last few years many studies have shown [4, 11] that the degree distribution alone does not uniquely characterise a network topology. Networks having exactly identical degree distribution can exhibit vastly different other topological properties. The fact that a network is scale-free only tells us limited information. In order to obtain a full picture of a network's structure, we need to look beyond the question of 'how many links does a node have?' and ask 'To whom the node is connected to?'

4 Who Connects with Whom?

4.1 Disassortative Mixing

Recently a number of studies have shown that the degree correlation plays a significant role in defining a network's structure [4, 7, 12, 13]. The degree correlation defines the mixing pattern in a network. Social networks are assortative mixing where nodes tend to attach to alike nodes, i.e. high-degree nodes to high-degree nodes and low-degree nodes to low-degree nodes. On contrast technological and

biological networks, including the Internet, are disassortative mixing where high-degree nodes tend to connect with low-degree nodes, and visa versa. We can not use a network's degree distribution to predict its mixing pattern. Networks with the same degree distribution can have completely opposite mixing patterns [11].

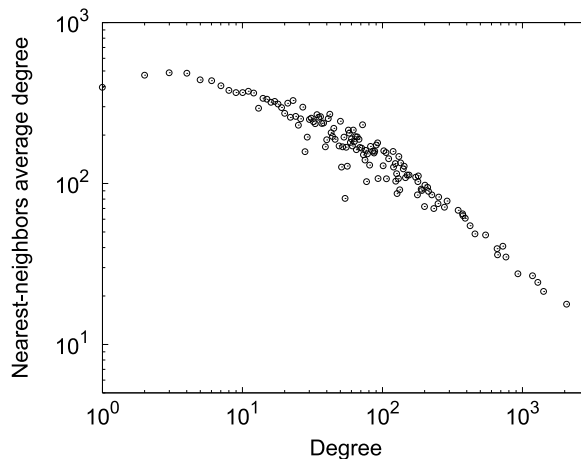


Fig. 3. Nearest-neighbours average degree of k -degree nodes.

A network's mixing pattern can be inferred by the correlation between a node's degree k and its nearest-neighbours average degree k_{nn} [6]. Fig. 3 shows the Internet exhibits a negative correlation between the two quantities suggesting the network is disassortative mixing. For example for a node with degree 1000, the average degree of its 1000 neighbours is around 20; whereas for a node with degree 2, the average degree of its two neighbours is around 500.

The mixing pattern can also be inferred by computing the so-called assortative coefficient [7], $-1 < \alpha < 1$, which is defined as

$$\alpha = \frac{L^{-1} \sum_m j_m k_m - [L^{-1} \sum_m \frac{1}{2}(j_m + k_m)]^2}{L^{-1} \sum_m \frac{1}{2}(j_m^2 + k_m^2) - [L^{-1} \sum_m \frac{1}{2}(j_m + k_m)]^2},$$

where L is the number of links a network has, and j_m, k_m are the degrees of the nodes at the ends of the m th link, with $m = 1, \dots, L$. If $0 < \alpha < 1$, a network is assortative mixing; and if $-1 < \alpha < 0$, a network is disassortative mixing, e.g. $\alpha = -0.19 \sim -0.24$ for the Internet.

The disassortative mixing property of the Internet is relevant because it means the peripheral, low-degree nodes are almost always connect directly with some high-degree nodes. On the other hand the chance of finding a chain of low-degree nodes is low. Then, how about the high-degree nodes themselves? Are they interconnected with each other? Does the disassortative mixing property mean that the high-degree nodes only connect with low-degree nodes?

4.2 Rich-Club Phenomenon

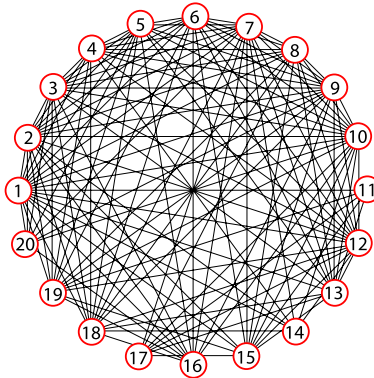


Fig. 4. Internet connections among the top 20 richest nodes themselves. Other connections to the nodes are omitted for clarity.

We call nodes with large numbers of links the ‘rich’ nodes. Fig.4 shows that in the Internet the richest nodes are tightly interconnected with each other forming the core of the network. This is called the rich-club phenomenon [8]. This phenomenon does not conflict with the Internet’s disassortative mixing property. Each rich node shown in the figure has a large number of links. The majority the links are connect to poorly connected nodes (disassortative mixing) whereas only a few of them are enough to provide the interconnectivity with other rich nodes, whose number is anyway small.

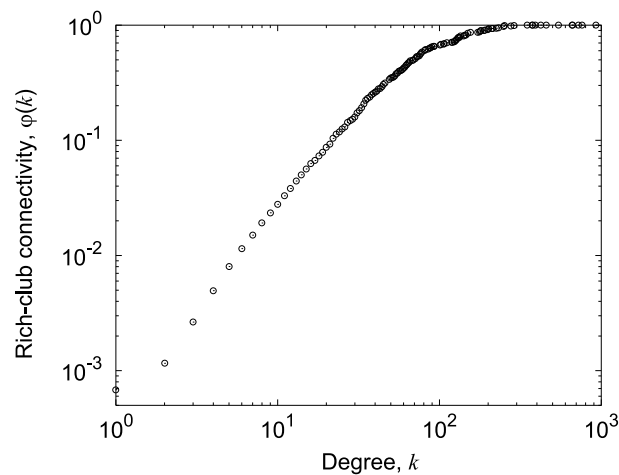


Fig. 5. Rich-club coefficient as a function of node degree.

The rich-club phenomenon can be quantitatively assessed by measuring the rich-club coefficient,

$$\phi(k) = \frac{2E_{\geq k}}{N_{\geq k}(N_{\geq k} - 1)}, \quad (1)$$

where $N_{\geq k}$ is the number of nodes with degrees greater or equal to k and $E_{\geq k}$ is the number of links among the $N_{\geq k}$ nodes themselves. The rich-club coefficient measures the ratio of the actual number of links to the maximum possible number of links among nodes with degrees no less than k . Fig. 5 shows the rich-club coefficient as a function of node degree k . It shows that the rich nodes are becoming more tightly interconnected when the degree increases, in other words, when the club becomes more ‘exclusive’. A small number of nodes with degrees larger than 300 have $\phi = 1$ which means they are fully interconnected.

5 Discussion

The internet is ultra small because it exhibits both disassortative mixing and rich-club phenomenon. These two structural properties together contribute to the routing efficiency of the network. The rich club consists of a small number of highly connected nodes. The club members are tightly interconnected between each other. If two club members do not have a direct connection, they are very likely to share a neighbouring club member. Thus the hop distance between the members is very small (either 1 or 2 hops). The rich-club functions as a ‘super’ traffic hub of the Internet by providing a large selection of shortcuts for routing. The disassortative mixing property ensures that the majority of network nodes, which are peripheral low-degree nodes, are always near the rich-club. Thus a typical shortest path between two peripheral nodes consists of three hops, the first hop is from the source node to a member of the rich club, the second hop is between two rich-club members, and the final hop is to the destination node.

As we know a fully connected graph is the most efficient topology because the shortest path is 1 between any two nodes. Whereas for sparsely connected networks, the star-like topology is very efficient because the average shortest path is less than 2. If we abstract the rich-club as a single node then the Internet topology will look like a star topology, where all nodes connect with the centre of the star. A star topology has a big problem that the centre of the star is very easy to be congested and becomes the single-point-of-failure. The Internet ‘solves’ these problems by replacing the centre with a club of rich nodes. The rich-club contains plenty of redundant links. The network’s performance and integrity will not be affected if some of the links between the club members are removed. In addition, incoming traffic to the ‘centre’ is handled in a distributed manner by a handful of rich-club members such that the chance of getting congested is reduced. In case a few club members are congested (or removed), the peripheral nodes served by these rich nodes may suffer, but the rest of the network can still work as normal. Of course if the rich-club core is broken by removing too many rich nodes or inter-rich links, the network will fall apart.

6 Summary

In a way the Internet has an ideal structure. It is very large in size and very sparse in connectivity. Yet the Internet has a sophisticated structure which makes the network very ‘small’ in terms of routing. At the same time the Internet is quite robust because the function of the network’s core is shared by a group of rich nodes which are tightly interconnected with each other and therefore have plenty of redundant links. It is remarkable that such a structure is the consequence of the Internet evolution process which is driven by distributed, local decisions without any central plan. In fact until ten years ago there was not even any map of the global Internet .

In the last decade there is a increasing recognition that effective engineering of the Internet is predicated on a detailed understanding of issues such as the large-scale structure of its underlying physical topology, the manner in which it evolves over time, and the way in which its constituent components contribute to its overall function [14].

In recent years there have been a lot of research activities on characterising and modelling the Internet topology. Various metrics and models are proposed [15, 16]. Researchers are interested in understanding not only the Internet’s structure but also the dynamics and the evolution [17]. One important goal is to provide better models for more realistic simulations which differentiates the different natures of links between AS nodes and incorporates suitable traffic model as well as other relevant factors such as capacity and demo-geographic distributions of network elements.

Acknowledgments. S. Zhou is supported by The Royal Academy of Engineering/EPSRC Research Fellowship under grant no. 10216/70.

References

1. Quoitin, B., Pelsser, C., Swinnen, L.: Interdomain traffic engineering with BGP. *IEEE Communications Magazine* **41** (2003) 122–128
2. Barabási, A., Albert, R.: Emergence of scaling in random networks. *Science* **286** (1999) 509
3. Pastor-Satorras, R., Vespignani, A.: *Evolution and Structure of the Internet - A Statistical Physics Approach*. Cambridge University Press, Cambridge (2004)
4. Mahadevan, P., Krioukov, D., Fomenkov, M., Huffaker, B., Dimitropoulos, X., Claffy, K., Vahdat, A.: The internet AS-level topology: Three data sources and one definitive metric. *Computer Comm. Rev.* **36** (2006) 17–26
5. Faloutsos, M., Faloutsos, P., Faloutsos, C.: On power-law relationships of the Internet topology. *Computer Comm. Rev.* **29** (1999) 251–262
6. Pastor-Satorras, R., Vázquez, A., Vespignani, A.: Dynamical and correlation properties of the internet. *Phys. Rev. Lett.* **87** (2001)
7. Newman, M.E.J.: Mixing patterns in networks. *Phys. Rev. E* **67** (2003) 026126
8. Zhou, S., Mondragón, R.J.: The rich-club phenomenon in the Internet topology. *IEEE Comm. Lett.* **8** (2004) 180–182

9. Krioukov, D., Chung, F., k. claffy, Fomenkov, M., Vespignani, A., Willinger, W.: The workshop on internet topology (WIT) report. *Computer Comm. Rev.* **37** (2007) 69–73
10. Watts, D.J., Strogatz, S.H.: Collective dynamics of ‘small-world’ networks. *Nature* **393** (1998) 440
11. Zhou, S., Mondragón, R.: Structural constraints in complex networks. *New Journal of Physics* **9** (2007) 1–11
12. Vázquez, A., Boguñá, M., Moreno, Y., Pastor-Satorras, R., Vespignani, A.: Topology and correlations in structured scale-free networks. *Phys. Rev. E* **67** (2003)
13. Maslov, S., Sneppenb, K., Zaliznyaka, A.: Detection of topological patterns in complex networks: correlation profile of the internet. *Physica A* **333** (2004) 529–540
14. Floyd, S., Kohler, E.: Internet research needs better models. *Computer Comm. Rev.* **33** (2003) 29–34
15. Tangmunarunkit, H., Govindan, R., Jamin, S., Shenker, S., Willinger, W.: Network topology generators: Degree-based vs. structural. *Proc. ACM SIGCOMM* (2002) 147–159
16. Zhou, S., Mondragón, R.J.: Accurately modelling the Internet topology. *Phys. Rev. E* **70** (2004) 066108
17. Dorogovtsev, S.N., Mendes, J.F.F.: *Evolution of Networks - From Biological Nets to the Internet and WWW*. Oxford University Press, Oxford (2003)