# Power Law Modelling of Internet Topology

Shi Zhou

University College London, Adastral Park Campus
Ross Building, Adastral Park, Ipswich, IP5 3RE, United Kingdom
`s.zhou@adastral.ucl.ac.uk`

**Abstract.** In recent years there have been tremendous efforts to measure, characterise and model the internet topology. We discuss why the power law degree distribution is not an artifact but an integral property of the internet. On the other hand we argue that while it is one of the properties that fundamentally characterise the global internet structure, other properties should also be considered to obtain a full description of the network. We review the power law modelling of the internet topology and provide a critical look at the contribution of such research to the Internet engineering.

**Keywords:** Internet, topology, power-laws, network modelling, scale-free networks.

## 1 Introduction

In 1999 it was discovered that the global Internet structure is characterised by a power law [1]. That is, the probability distribution of a node's connectivity (measured for example by the number of BGP peering relations that an autonomous system has) follows a power law. This discovery invalidated previous internet models that were based on the classical random graphs. Since then there have been tremendous efforts to measure, characterise and model the internet topology [2,3,4,5,6,7,8,9,10,11].

There is an increasing recognition that effective engineering of the global internet should be based 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 [12].

This paper reviews the measurements and models of the internet topology, and comments upon whether the power law is in itself an adequate characterisation of the system. It questions whether models based on power laws provide an suitable platform for theoretical and simulation analysis of the internet's topological characteristics. Finally, it provides discussion of how such research could be of use in improving network performance.

## 2 Internet Topology at AS-Level

Topology is the connectivity graph of a network, upon which the network's physical and engineering properties are based. The internet contains millions

J. Zhou (Ed.): Complex 2009, Part II, LNICST 5, pp. 2090–2098, 2009.

of routers, which are grouped into about ten thousands of sub-networks, called Autonomous Systems (AS). The internet topology can be studied at the router level or the AS level. Connectivity between routers within an AS is a local structure, whereas the delivery of data traffic through the global internet depends on the complex interactions between ASs that exchange routing information using the border gateway protocol (BGP) [13]. Therefore the global internet topology is usually characterised at the AS level (AS graphs), on which a node is an AS which is typically owned by an Internet service provider (ISP), and a link represents a BGP-peering relation between two ASs (i.e. a commercial agreement between two ISPs), e.g. customer-provider or peer-peer relations.

## 3 Measuring Internet Topology

Measurements of the Internet topology first became available in late 1990s. There are two types of measurements using different data sources.

*Passive measurements* are constructed from BGP routing tables which contain the information of links from an AS to its immediate neighbours. The Routing Information Service of RIPE [14] is an importance source of BGP data. The widely used BGP AS graphs are produced by the National Laboratory for Applied Network Research [15] and the RouteViews Project at University of Oregon [16], which connect to a number of operational routers within the Internet for the purpose of collecting BGP tables. The Topology Project at University of Michigan [17] provided an extended version [18] of BGP AS graph by using additional data sources, such as the Internet Routing Registry (IRR) data and the Looking Glass (LG) data. A shortcoming of the BGP AS graphs is that they may contain links which do not actually exist in the real Internet.

*Active measurements* are based on traceroute which captures the sequence of IP hops along the forward path from the source to a given destination by sending either UDP or ICMP probe packets to the destination. CAIDA [19] has developed a tool, *skitter*, which probes about one million IPv4 addresses from 25 monitors placed in the global internet. By using the core BGP tables provided by RouteViews, CAIDA maps the IP addresses in the gathered IP paths to AS numbers [20] and constructs AS graphs on a daily basis. DIMES [21] is a more recent large-scale distributed measurement effort. It collects traceroute data by probing from $> 10,000$ software clients, installed by volunteers in $> 90$ countries, to destinations assigned by a central server at random from a set of five million destination addresses. To further improve the completeness, DIMES merges the resulted AS graph with that of RouteViews. So far DIMES has produced the most complete AS graph containing $\approx 20,000$ nodes and $\approx 70,000$ links. A criticism of the traceroute AS graphs is that the translation from IP addresses to AS numbers is not trivial and could introduce errors.

## 4 Power Law Degree Distribution

Power laws describe a wide range of phenomena in nature and a large body of ongoing research investigates their applicability in fields such as computer science, physics, biology, social sciences and economics [22,23]. Power law distributions are characterised by a slower than exponentially decaying probability tail, which loosely means that large values can occur with a non-negligible probability. They can be used to characterise a variety of relations such as for example the distribution of income, city population, citations of scientific papers, word frequencies, computer file sizes and the number of daily hits to a given website.

In graph theory, degree $k$ is defined as the number of links or immediate neighbours of a node. Degree is the principal parameter when characterising network connectivity. The first step in describing and discriminating between different networks is to measure the degree distribution $P(k)$, the fraction of nodes in the network with degree $k$. Only in relatively recent years it has been discovered that the internet topology at the AS level (and the router level) exhibits a power law degree distribution $P(k) \sim Ck^{-\gamma}$ [1], where $C > 0$ is a constant and the exponent $\gamma \simeq 2.2 \pm 0.1$. This means a few nodes have very large numbers of connections, whereas the vast majority of nodes have only a few links. Although internet AS graphs produced from different data sources vary in the numbers of nodes and links, all the internet AS graphs are well characterised by a power law degree distribution [24].

The power law property is an evidence that the internet AS-level topology has evolved into a complex, heterogeneous structure that is profoundly different from a random graph which has been assumed in the original design of internet protocol suits.

## 5 Power Law or Sampling Bias?

A recent series of papers [25,26] reported that traceroute measurements based on data collected from a small number of observers to a large number of destinations could be biased in such a way that graphs which in fact have Poisson degree distributions appear to exhibit a power law. Indeed it has been shown [27,28] that there could be a considerable amount (35%) of the links in the AS level internet that are still to be unveiled.

Then a fundamental question is whether the power law degree distribution is indeed an integral property of the internet. The answer is yes. Firstly, the more recent DIMES project collects data from numerous observers distributed in thousands of AS networks around the world. Therefore the number of vantage points of DIMES is two orders of magnitude larger than that of previous measurements. The internet AS graph obtained by DIMES exhibits a clear power law degree distribution in agreement with previous observations (with slightly different power law exponent). Secondly, study [29] has shown that if the observed power law were due to sampling bias and the larger real graph had a Poisson degree distribution, then the real graph's average degree would be more than 100, which is obviously not realistic as it is known that the average number of BGP peering relations of an AS is around 6. Study [29] also showed that in traceroute-based studies, the more the underlying graph is heavy-tailed, the more it is clearly discriminated. And the heavy-tail is what is measured first and with better accuracy. Recent studies [27,28] also suggest that the power law seems to hold more accurately if we consider only the customer-provider links only.

## 6 Structures beyond the Power Law

Degree distribution is a first-order topological property which is based on the connectivity information of each individual node. However the internet has many other inherent structures associated with the connectivity of a pair, a triad or a set of nodes, which are characterised by higher order topological properties. For example the degree-degree correlation [30,31,32,33] indicates whether high-degree nodes tend to connect with high-degree nodes (assortative mixing) or low-degree nodes (disassortative mixing); the rich-club coefficient [34,35] quantifies how the best connected nodes connect with themselves; the clustering coefficient [36] measures the fraction of a node's neighbours which are neighbours to each other; the average shortest path gives the average hopping distance between any two nodes and the $k$-core decomposition [37] infers a network's underlying hierarchical structure. These high-order properties have clear physical meanings and have profound impacts on a network's overall functionality.

It is known that networks with exactly the same degree distribution can have completely different higher order properties [35]. When modelling the internet topology, it is vital to capture not only an accurate, but also a complete picture of the network structure. A power law-focused internet model without a proper examination of the higher order properties could produce misleading results. It is suggested [38] that the second order properties are sufficient for most practical purposes, while the third order properties essentially reconstructs the Internet AS- and router-level topologies exactly.

## 7 Modelling Internet Topology

Since the discovery of the power law degree distribution, a large number of models for internet topology have been proposed to generate and explain the power law [3,4,39,10,40,41]. Models from networking community, such as Tier, BRITE [42], GT-ITM (Transit-Stub) and Inet [43], often suffer from problems of no (or incorrect) power law, inaccurate large-scale hierarchy, parameter manipulations and lack of evolution mechanism; and models from physicists [44,45,46,47,48,9,49] also have problems as they often are too general and do not incorporate any real network specifics.

In general there are two main approaches for generating topologies of complex networks [50]. The equilibrium (top-down) approach is to construct an ensemble of static random graphs reproducing certain properties of observed networks and then to derive their other properties by the standard methods. The non-equilibrium approach (bottom-up) tries to mimic the actual dynamics of network growth: if this dynamics is accurately captured, then the modelling algorithm, when let to run to produce a network of the required size, will output the topology coinciding with the observations. The classic examples of this approach are the Barabási-Albert (BA) model [44] and the HOT model [51]. It is clear that the more ambitious non-equilibrium approach has the potential to hold the ultimate truth.

One of the most successful non-equilibrium model for the Internet is the Positive-Feedback Preference (PFP) model proposed in 2004 [52]. It is able to reproduce many of the known characteristics of the AS topology [53,54] and also reproduce the characteristics of the topology of smaller sub-graphs, for example the Chinese internet AS graph [55]. The PFP model achieves this by using two simple growth mechanisms (a variation of the BA model's preferential attachment approach), which are in accordance with the dynamics observed on internet history data [30,56,57]. Firstly the model grows from a small random graph by two interactive actions, ie the attachment of new nodes to old nodes in the existing system and the addition of new links between these old nodes to other old nodes. This resembles the dynamics that a large ISP extends its connections to peering ISPs as a reaction to the increasing number of new customers. Secondly the preference probability of a new link attaching to node $i$ with $k_i$ connections is given as

$$\Pi(i) = \frac{k_i^{1+\delta \log_{10} k_i}}{\sum_j k_j^{1+\delta \log_{10} k_j}}, \tag{1}$$

where the parameter $\delta = 0.048$. This means a node's ability of competing for new links increases more and more rapidly with its growing number of links, like a positive-feedback loop, such that 'rich not just get richer, they get disproportionately richer'. This resembles the 'winner-takes-all' trend in the social, economic, technological realities of the internet.

## 8 Practical Responses to Power Law Topology Models

Studies have suggested that the Internet's power law structure may be relevant to a number of problems in the internet [58], such as the network's robustness and vulnerability [59], the severely biased distribution of traffic flow, the slow convergence of BGP routing tables [60] and the large-scale cascading failure caused by incidents or deliberate attacks [61]. As such, the power law property also provides novel insights into the solutions of these problems. For example study [62] showed that the power law property makes it possible to mitigate the Distributed DoS attacks by implementing the route-based filtering on less than 20% of Internet AS sites; [63] showed that a compact routing scheme based on the power law property requires a significantly smaller routing table size; [64] showed that the power law property is relevant to the epidemic threshold for a network; and [65] presented a more realistic simulation of the Internet which combines power law topologies with Internet power law traffic models.

## 9 Criticisms and Commentary

In summary the power law degree distribution is an integral property of the internet AS-level topology. It is vital, however, for researchers to look beyond this first-order property in order to obtain a fuller description of the global Internet structure.

So far much of the research in this area studies the internet topology as a pure graph. However the reachability between two AS nodes on the internet is not only affected by the underlying connectivity graph, but also constrained by many other factors such as routing policies, capacities, demo-geographic distributions etc. Future internet models should more closely reflect the internet reality in order to produce practically useful results.

In addition, as pointed out in [66], an interdisciplinary communication among networking, physics, mathematics, and engineering communities is much needed to facilitate the interdisciplinary flow of knowledge. This will facilitate the network research community to convert theoretical results into more practical solutions that matter for real networks, e.g. performance, revenue and engineering.

As can be seen in the previous section, the power law models of topology have begun to stimulate research which takes advantage of this network structure. In this way, the power law models of topology can be used to inform and also improve network engineering.

**Acknowledgments.** S. Zhou is supported by the Royal Academy of Engineering and EPSRC under grant no. 10216/70.

## References

1. Faloutsos, M., Faloutsos, P., Faloutsos, C.: On power–law relationships of the Internet topology. Computer Comm. Rev. 29, 251–262 (1999)
2. Strogatz, S.H.: Exploring complex networks. Nature (London) 410, 268 (2001)
3. Krapivsky, P.L., Redner, S.: Organization of growing random networks. Phys. Rev. E 63, 066123 (2001)
4. Albert, R., Barabási, A.L.: Statistical mechanics of complex networks. Rev. Mod. Phys. 74, 47–97 (2002)
5. Bornholdt, S., Schuster, H.G.: Handbook of Graphs and Networks - From the Genome to the Internet. Wiley-VCH, Weinheim (2002)
6. Subramanian, L., Agarwal, S., Rexford, J., Katz, R.H.: Characterizing the internet hierarchy from multiple vantage points. In: Proc. IEEE INFOCOM, pp. 618–627 (2002)
7. Chen, Q., Chang, H., Govindan, R., Jamin, S., Shenker, S., Willinger, W.: The origin of power laws in internet topologies revisited. In: Proc. IEEE INFOCOM, pp. 608–617 (2002)
8. Dorogovtsev, S.N., Mendes, J.F.F.: Evolution of Networks - From Biological Nets to the Internet and WWW. Oxford University Press, Oxford (2003)
9. Caldarelli, G., Rios, P.D.L., Pietronero, L.: Generalized network growth: from microscopic strategies to the real internet properties (2004), arXiv:cond-mat/0307610 v1
10. Pastor-Satorras, R., Vespignani, A.: Evolution and Structure of the Internet - A Statistical Physics Approach. Cambridge University Press, Cambridge (2004)
11. Chang, H., Jamin, S., Willinger, W.: To peer or not to peer: Modeling the evolution of the internet's AS-level topology. In: Proc. IEEE INFOCOM, pp. 1–12 (2006)

12. Floyd, S., Kohler, E.: Internet research needs better models. Computer Comm. Rev. 33, 29–34 (2003)
13. Quoitin, B., Pelsser, C., Swinnen, L.: Interdomain traffic engineering with BGP. IEEE Communications Magazine 41, 122–128 (2003)
14. Routing information service, `http://www.ripe.net/`
15. The National Laboratory for Applied Network Research (NLANR), `http://moat.nlanr.net/`
16. The Route Views project, `http://www.routeviews.org/`
17. The topology project, `http://topology.eecs.umich.edu/`
18. Chang, H., Govindan, R., Jamin, S., Shenker, S., Willinger, W.: Towards capturing representative as-level internet topologies. Computer Networks Journal 44, 737–755 (2004)
19. The Cooperative Association for Internet Data Analysis (CAIDA), `http://www.caida.org/`
20. Murray, M., Claffy, K.: Measuring the immeasurable: global internet measurement infrastructure. In: Proc. of PAM, pp. 159–167 (2001)
21. The DIMES project, `http://www.netdimes.org/`
22. Mitzenmacher, M.: A brief history of generative models for power law and lognormal distributions. Internet Mathematics 1, 226–251 (2004)
23. Newman, M.: Power laws, Pareto distributions and Zipf's law. Contemporary Physics 46, 323–351 (2005)
24. 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, 17–26 (2006)
25. Lakhina, A., Byers, J.W., Crovella, M., Xie, P.: Sampling biases in ip topology measurements. In: Proc. IEEE INFOCOM, vol. 1, pp. 332–341 (2003)
26. Clauset, A., Moore, C.: Accuracy and scaling phenomena in internet mapping. Phys. Rev. Lett. 94, 018701 (2005)
27. Cohen, R., Raz, D.: The internet dark matter - on the missing links in the AS connectivity map. In: Proc. IEEE INFOCOM, pp. 1–12 (2006)
28. He, Y., Siganos, G., Faloutsos, M., Krishnamurthy, S.: A systematic framework for unearthing the missing links: Measurements and impact. In: Proc. of 4th USENIX Symposium on Networked Systems Design & Implementation, pp. 187–200 (2007)
29. Dall-Asta, L., Alvarez-Hamelin, I., Barrat, A., Vazquez, A., Vespignani, A.: Exploring networks with traceroute-like probes: Theory and simulations. Theoretical Computer Science 355, 6–24 (2006)
30. Pastor-Satorras, R., Vázquez, A., Vespignani, A.: Dynamical and correlation properties of the internet. Phys. Rev. Lett. 87 (2001)
31. Newman, M.E.J.: Assortative mixing in networks. Phys. Rev. Lett. 89, 208701 (2002)
32. Newman, M.E.J.: Mixing patterns in networks. Phys. Rev. E 67, 026126 (2003)
33. Maslov, S., Sneppenb, K., Zaliznyaka, A.: Detection of topological patterns in complex networks: correlation profile of the internet. Physica A 333, 529–540 (2004)
34. Zhou, S., Mondragón, R.J.: The rich-club phenomenon in the Internet topology. IEEE Comm. Lett. 8, 180–182 (2004)
35. Zhou, S., Mondragón, R.: Structural constraints in complex networks. New Journal of Physics 9, 1–11 (2007)
36. Watts, D.J., Strogatz, S.H.: Collective dynamics of 'small-world' networks. Nature 393, 440 (1998)
37. Carmi, S., Havlin, S., Kirkpatrick, S., Shavitt, Y., Shir, E.: A model of internet topology using k-shell decomposition. PNAS 104, 11150–11154 (2007)

38. Mahadevan, P., Krioukov, D., Fall, K., Vahdat, A.: Systematic topology analysis and generation using degree correlations. In: Proc. ACM SIGCOMM, pp. 135–146 (2006)
39. Dorogovtsev, S.N.: Networks with given correlations, `http://arxiv.org/abs/cond-mat/0308336v1`
40. Mitzenmacher, M.: The future of power law research. Internet Mathematics 2, 525–534 (2005)
41. Leskovec, J., Kleinberg, J., Faloutsos, C.: Graph evolution: Densification and shrinking diameters. ACM Transactions on Knowledge Discovery from Data 1, 1–40 (2007)
42. Medina, A., Matta, I.: Brite: A flexible generator of internet topologies. Technical Report BU-CS-TR-2000-005, Boston University (2000)
43. Winick, J., Jamin, S.: Inet-3.0 internet topology generator. Technical Report UM-CSE-TR-456-02, University of Michigan (2002)
44. Barabási, A., Albert, R.: Emergence of scaling in random networks. Science 286, 509 (1999)
45. Dorogovtsev, S.N., Mendes, J.F.F.: Scaling behaviour of developing and decaying networks. EuroPhys. Lett. 52, 33 (2000)
46. Tauro, S.L., Palmer, C., Siganos, G., Faloutsos, M.: A simple conceptual model for the internet topology. In: Proc. of Global Internet, pp. 1667–1671 (2001)
47. Bu, T., Towsley, D.: On distinguishing between internet power law topology generators. In: Proc. IEEE INFOCOM, p. 638 (2002)
48. Bianconi, G., Caldarelli, G., Capocci, A.: Number of h-cycles in the internet at the autonomous system level (2003), ArXiv:cond-mat/0310339
49. Krapivsky, P.L., Redner, S., Leyvraz, F.: Connectivity of growing random networks. Phys. Rev. Lett. 85, 4629 (2000)
50. Status report of IRTF routing research group's (RRG) future domain routing (FDR) scalability research subgroup (RR-FS) (2005), `http://rr-fs.caida.org/`
51. Li, L., Alderson, D., Willinger, W., Doyle, J.: A first-principles approach to understanding the internet's router-level topology. In: The 2004 conference on Applications, technologies, architectures, and protocols for computer communications, pp. 3–14 (2004)
52. Zhou, S., Mondragón, R.J.: Accurately modelling the Internet topology. Phys. Rev. E 70, 066108 (2004)
53. Krapivsky, P., Krioukov, D.: Scale-free networks as pre-asymptotic regimes of super-linear preferential attachment. Phys. Rev. E 72, 026114 (2008)
54. Haddadi, H., Fay, D., Jamakovic, A., Maennel, O., Moore, A.W., Mortier, R., Rio, M., Uhlig, S.: Beyond node degree: evaluating as topology models. Technical Report UCAM-CL-TR-725, University of Cambridge (2008)
55. Zhou, S., Zhang, G.Q.: Chinese internet AS-level topology. IET Commun. 1, 209–214 (2007)
56. Vázquez, A., Pastor-Satorras, R., Vespignani, A.: Large-scale topological and dynamical properties of internet. Phys. Rev. E 65, 066130 (2002)
57. Park, S.T., Pennock, D., Giles, C.L.: Comparing static and dynamic measurements and models of the internet's AS topology. In: Proc. IEEE INFOCOM, pp. 1616–1627 (2004)
58. Handley, M.: Why the internet only just works. BT Technology Journal 24, 119–129 (2006)
59. Albert, R., Jeong, H., Barabasi, A.: Error and attack tolerance of complex networks. Nature 406, 378 (2000)

60. Labovitz, C., Ahuja, A., Wattenhofer, R., Venkatachary, S.: The impact of internet policy and topology on delayed routing convergence. In: Proc. IEEE INFOCOM, pp. 537–546 (2001)
61. Park, S.T., Khrabrov, A., Pennock, D.M., Lawrence, S., Giles, C.L., Ungar, L.H.: Static and dynamic analysis of the internet's susceptibility to faults and attacks. In: Proc. IEEE INFOCOM, pp. 2144–2154 (2003)
62. Park, K., Lee, H.: On the effectiveness of route-based packet filtering for distributed DoS attack prevention in power-law internets. Computer Comm. Rev. 31, 15–26 (2001)
63. Krioukov, D., Fall, K., Yang, X.: Compact routing on internet-like graphs. In: PAS. vol. 1, p. 219 (2004)
64. Wang, Y., Chakrabarti, D., Wang, C., Faloutsos, C.: Epidemic spreading in real networks: an eigenvalue viewpoint. In: Proc. of 22nd International Symposium on Reliable Distributed Systems (SRDS), pp. 25–34 (2003)
65. Arrowsmith, D., Woolf, M.: Modelling of TCP packet traffic in a large interactive growth network. In: Proc. IEEE Systems and Circuits, Vancouver, pp. 477–480 (2004)
66. Krioukov, D., Chung, F., Claffy, K., Fomenkov, M., Vespignani, A., Willinger, W.: The workshop on internet topology (WIT) report. Computer Comm. Rev. 37, 69–73 (2007)