Planarity of Data Networks

Rhys Bowden, Hung X. Nguyen, Nickolas Falkner, Simon Knight, Matthew Roughan
University of Adelaide, Australia
{rhys.bowden,hung.nguyen,nickolas.falkner,simon.knight,matthew.roughan}@adelaide.edu.au
www.topology-zoo.org

***Abstract*—The study of network topology has attracted a great deal of attention in the last decade. However, this study has been hampered by a lack of accurate data. Existing methods for measuring topology have many flaws, and arguments about the importance of these flaws have overshadowed the more interesting questions about network structure. The Internet Topology Zoo is a store of network data created from the information that network operators make public. As such, it is the most accurate large-scale collection of network topologies available, and includes meta-data that we could never have derived from automated network measurements. With this data we can answer questions about network structure with more certainty than ever before, and in this paper we study a particular property of these networks – planarity. A surprising number of graphs in the Zoo are planar, many more than can be explained through the models for graph formation we tested. We speculate that planarity results from the requirement to build transparent networks that network operators can understand easily.**

I. Introduction

Graphs of communications networks have received a great deal of interest over the last few decades (for instances see [1]–[8]), both through purely scientific interest and for practical reasons. Network graphs determine many of the properties of the underlying communications network, for instance its reliability and performance. They are therefore valuable inputs into many network algorithms, and much effort has gone into their measurement and synthesis for use in testing algorithms.

More importantly, models of graph formation tell us something about how networks are designed. Any one engineer may be able to describe, at least loosely, their method for network design. However, the more interesting goal is to learn universal laws of network formation that may, for instance, still apply as technology evolves. Moreover, through understanding these laws we may learn how best to improve the underlying technology to fit the network design process, rather than putting the cart before the horse, as has been so often done in networking, by requiring engineers to work around technological limitations, or in providing them with features they do not need.

In this paper we note one feature – planarity – that is common in the networks we observe, but which is not explained well by the existing graph formation models. There are ongoing debates about what type of model best fits data networks: on the one hand lie the random graph models (starting with Erdos-Renyi and Gilbert [9] and going forward through Waxman [1], and more recently power-law graphs [2]–[5]). On the other hand lie "designed networks" such as the structured networks of G-ITM [6], [7] or HOT (Highly Optimized Tolerance) graphs [8]. Proponents of power-law and HOT graphs seem convincing, but both are hampered by lack of accurate data. In the few cases where a commercial network has been used the data have not been published.

There is ongoing research effort to improve the accuracy of measured networks, but we circumvent that issue entirely in this paper through a new source of network data, first described in [10]. The graphs in this dataset are derived from openly published network maps. Thus we circumvent the difficulties encountered by measurement based studies whose errors have confused the issue of topology modelling for years.

The results are interesting, in particular, we observe a high degree of planarity in our networks. A *planar graph* can be drawn in a plane without edges crossing. Planarity has interesting consequences (apart from the ability to simply draw the graph):

- A planar graph (without loops) is 4-colorable;
- For any planar graph we can define a dual (which is also planar), by taking one vertex in each *face* (including the outer face) and creating one edge in the dual between each face divided by an edge;
- The planar separator theorem states that every n-vertex planar graph can be partitioned into two subgraphs of size at most $2n/3$ by the removal of $O(\sqrt{n})$ vertices;

and so on. However, here we are less interested in the properties of planar graphs than we are in what this property tells us about graph formation, and hence about appropriate synthetic models for communications networks.

There are two broad approaches suggested for generating synthetic communications networks: variations on random graphs, and the structural methods most recently exemplified by optimization models [8]. The latter makes a great deal more sense when considering a computer network that will have typically been designed by a small group of engineers (rather than randomly grown). However, we use this approach to generate a large set of networks, and show that it does not generate networks with the observed degree of planarity.

That is not unexpected. The optimization approach seeks to improve some objective function usually related to the capital cost of building a network. It does not include any constraints, or cost associated with link crossings. However, in real network design there is a cost to *complexity*. A more complex network is harder to manage. It is harder for a network engineer to picture. Debugging is more difficult, as there are more possible sources of errors, and the relationship between error and observations of that error may be less direct. Therefore a more complex network has an associated *operations cost*.

Operations costs associated with network designs are very difficult to quantify and are, for this reason, often ignored in the operations research literature. However, these costs are real, and are qualitatively understood by many network engineers. The result is that they often avoid purely optimized networks in favor of simple designs. We speculate that this is the cause of the high degree of planarity we observe.

Rather than discarding the existing work on network synthesis through optimization, this simply implies that additional criteria should be included in such optimizations. Such criteria could crudely enforce planarity through constraints or additions to the objective function, or could be more subtle through including a "complexity" based cost, though we leave the choice of such a cost function for future research.

II. Data

Before we begin discussing the details of our topological data, let us first define our terminology precisely as topology is a woefully abused term. By topology we mean an undirected graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$, which abstracts the connectivity of a data communications network. In fact, we really mean a multigraph, as multiple edges are allowed between a single pair of nodes (formally, $\mathcal{E}$ is a multiset).

Care must be taken to define the nature of the nodes and edges of the graph. Internet topologies have been given for each of the seven OSI layers: e.g., edges may refer to physical cables, virtual network layer connections, or even the HTML links between WWW pages. Other types of topology are also possible, such as those reflecting hierarchical approximations: e.g., groupings of routers into Autonomous Systems (ASs) or Points-of-Presence (PoPs). The datasets we are using contain various levels of detail, from physical fiber, through to virtual/-logical connectivity between ASs. We admit various Internet communications networks to the Zoo, but we ensure that in each case the type of nodes and links are precisely specified.

A. The Collection Process

There are various strategies available for measuring network topology. The most direct way is to ask the network itself. IP routers are managed through *configuration files* describing the current operation of the router, and which can be used to measure a network [11][1]. However, because of the quality and detail of information contained in these files, they are considered sensitive and are rarely allowed outside an organization. Thus such data may be used to construct the type of map we use here, but is otherwise rarely available to researchers.

The second class of techniques involve IP-level hacks that ideally return the path between two points. The IP header option field "record route" [13], [14] returns the route of a packet as it traverses the network, but is reputedly not enabled on many routers due to security and performance concerns. The more common approach is `traceroute` [15], [16]. Despite being commonly used, `traceroute` has many well-known deficiencies (summarized in [17]).

There are nevertheless many studies of network topology using traceroutes (for examples see [18]–[23]), but the resulting network topologies are sometimes *very* inaccurate. Moreover, none of these studies perform large-scale verification against ground-truth data. One of the potential uses of the Zoo data is to establish ground-truth data to use in testing and improving measurement-based approaches, which have, in principle, the potential to survey a much wider range of networks than our process.

We performed comparisons between our dataset and one of the most recent and advanced traceroute based methods [23] and found large differences. For instance, one network (Cogent) was found in [23] to have 35 PoPs. The network operator publicly advertises 182 PoPs.

There is a third group of strategies for topology inference based on the ideas of network tomography. The statistical nature of these approaches again leads to errors.

Instead of the existing automated methods we adopt here a simple, manual approach. Many companies present public material about their network, primarily for promotional purposes. They wish to sell their network.

Some care goes into such maps because they are a form of advertisement and therefore have legal requirements for accuracy; they are highly visible to potential customers; and finally, network engineers are often proud of their work, and many would very much like to display it at its best.

The most important form of published information, from our point of view, is a network map, though other supplementary data can often be very useful as well. Such maps often only show PoPs and their interconnects, but sometimes they provide much more detail. We have collected over 200 such maps and associated data, and make no claim that we have an exhaustive list. In fact it is likely that many more such maps exist, and will exist in the future. Our collection and transcription process is described in detail in [10]. In brief, we use a group of tools to aid manual transcription of the network maps. The manual process, along with the quality checks we implement, insures a highly accurate representation of the published network map is transcribed into our database. The graphs are stored in a flexible and easy to read data format — GML (the Graph Markup Language) — which allows us to include meta-data about the graph (e.g., its link capacities and node locations) and the data collection (e.g., the date of collection). GML is easily read using the Python based graph library NetworkX [24], and easily converted into other formats such as the XML derivative GraphML [25], or the dot format used by GraphViz [26].

The data is stored at **www.topology-zoo.org**. It is viewable through a table containing meta-data about the networks, or as a large batch file. Scripts are provided for easy access and translation of the data.

We ask that any researchers who make use of this data take care to first understand the limitations of the data as documented in [10], and second acknowledge the data appropriately and report any resulting publications to us via email.

[1]A related approach is to use a routing monitor (e.g., [12]), which observes routing protocols and uses this information to construct a network topology, but this also requires privileged access to the network in question.

B. Accuracy

How accurate is the Zoo's data? The maps are created by network companies themselves, so they are based directly on ground truth. However, some network operators clearly produce these maps manually, potentially leading to inaccuracies in their depiction of their own network. There are two reasons that these errors must be much less significant than those in, for instance, traceroute studies.

- The network maps we use are all public documents, and so must satisfy standard due diligence requirements for an advertisement or official corporate publication. That is not to say that all corporations are perfect – it is easy to make mistakes in drawing the map – but a network operator is unlikely to publish a worse map than the one they use in their own network operations. Thus, at worst, we possess the best available map of a network.
- Some network maps may idealize the network. However, we argue that in these cases, we are seeing what was in the mind of the network engineer when the network was designed. In this sense, the idealized view of the network is actually more interesting than its implementation.

A second question of accuracy is "How accurate are our transcriptions?" We have transcribed a large number of maps so it is inevitable that some errors occur. However, we have tried to minimize errors by (i) using a graphical tool so that the transcription process is closely matched to the maps, and (ii) making sure that each network is transcribed by one person, and then checked by at least one other person.

C. Classification

As noted above, it is important to be precise about exactly what topology is being considered. One of the advantages of the manual transcription of the data from public information is that we can provide a number of additional classification tags for the data.

At the most basic level we classify our networks as Commercial (COM) or Research and Education Networks (REN). Our secondary type classification is related to the role the network plays: *backbone, testbed, customer, transit, access* and *internet exchange*. These are not exclusive groupings, but tags we attach to each network as appropriate. In analysing planarity we found some differences between customer and non-customer networks, and so only define that precisely here (see [10] for details of the other secondary classifications).

The *customer* tag is used when a network provided a higher level of services to its customers[2] than simple transit. We classified a network with this tag if the services provided required per-customer state: for instance, web hosting or electronic mail. With the introduction of per-customer state, the provider must have a customer service model that is not driven purely by the technical requirements of maintaining connectivity and core services (DNS, routing, etc.). This tag is applied when a provider clearly advertises a web-hosting, e-mail or co-location facility, or similar per-customer state service, for their connected organizations.

[2]In many cases customers might not be individual users, they may be businesses or research organizations.

The other major aspect of network type is the layer of the network. We provide tags indicating the layer (1-3) and perhaps some more information about the type of technology being used, for instance *IP*.

We may more accurately compare networks if we focus on their area of influence [27]. The tags for this categorization are taken from the set *metro, region, country, country+, continent, continent+* and *global*.

A *metro* network is one that spans a city, or a city-sized area possibly including a small number of adjacent townships. Likewise the *country* and *continent* designations. In each of these cases we add a tag describing the range, e.g., in the case of continent, it designates the "continent": North America, Europe, Asia-Pacific, Latin America, and Africa.

A *region* network is approximately the size of a province, state or a small number of states, where the number of states involved is not a substantial part of the containing country. The difference between the size of states in Australia, which can be as big as countries in Europe, demonstrates the need for a flexible definition of the "greater-than-metro but smaller-than-country" range classification.

The *country+* (and *continent+*) classifications are used when the network is mostly located within one country (or continent) but has routers in another that do not correspond to a significant number of the total. The label is needed because there are many networks that are easily identified as belonging to a country (or continental) region, but for expedience have one or more routers outside the country.

Where a network has significant presence in at least two continents, it is labelled a *global* network.

III. Results

A. Planarity Analysis of Zoo Maps

We currently have 147 networks transcribed into the Zoo, which we analyze here. Some of the maps are disconnected. In these cases, we take the largest connected components. If the graphs has multi-edges, we convert them (for the purpose of analyzing planarity) to single edge before performing our analysis. This step does not change the planar property of the graphs, but can change the average node degree.

Planarity was determined using the Boyer-Myrvold planarity test algorithm [28] implemented in the Matlab BGL (Boost Graph Library) [29]. Among the 147 networks analyzed, 21 (14%) were non-planar.

It would be natural to expect that larger, more complicated graphs are likely to be non-planar, and so we investigate the relationship between planarity, network size, and the average node degree. Figure 1 shows a scatter plot of the planar $(+)$ and non-planar (o) graphs in the Zoo. It is quite clear that the non-planar graphs tend have higher average node degree, and that the average node degree at which the networks become commonly non-planar decreases with network size.

Both conclusions are to be expected. For instance, there are well known upper bounds for the number of edges for a planar graph as a function of the network size. Euler's results [32] give two upper bounds: $3n-6$ for arbitrary graphs, and $2n-4$ for a graph where there are no triangles (cycles of

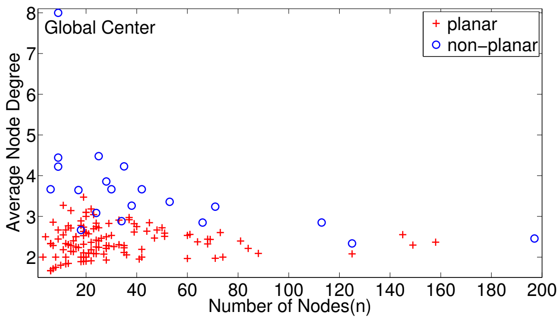


Fig. 1. Planarity of the network maps as a function of network size, and average node degree. We omit a network with 754 nodes and 895 edges whose graph is non-planar because it would not allow us to focus on the majority of smaller graphs.

length 3). Figure 2 shows the size of the maps (the number of nodes n) versus the number of edges for both planar and non-planar graph. We also plot in the figure the upper-bound on the number of edges that a planar graph can have, given the number of nodes n. As shown in Figure 2, all of the measured planar graphs lie below the second upper bound. We observe that many of the non-planar graphs also lie below the upper bound, though this is of little significance.

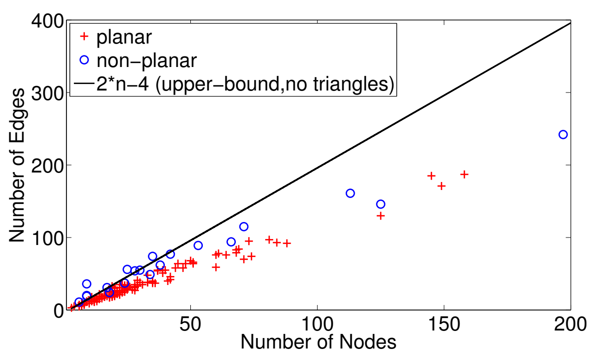


Fig. 2. Planarity of the network maps against the Euler upper bound.

The 21 non-planar networks are given in table III-A, where the names are attributed from the source of the data.

The breakdowns of non-planar networks according to their classifications are given in table II. Note that some networks could not be classified. Therefore, the table elements don't necessary sum to 147. Noteworthy points are that

- layer 1 networks are slightly more likely to be planar;
- all of the research networks are planar;
- customer networks are more likely to be non-planar;
- larger networks (in geographic extent) are more likely to be non-planar.

However, the above results may all be conflated with network size, for instance, country-wide networks (and larger) tend to be larger than regional or metropolitan networks. As we expand the size of the Zoo it should be possible to do a more formal statistical analysis, controlling for the network size and degree, to understand the true factors that influence planarity.

Network Name	Size(n)	Av. Node Degree
Airtel	9	4.2
AT&T(IP-MPLS)	25	4.48
BT(North America)	35	4.23
Chinanet	38	3.26
Cogentco	197	2.46
Dataexchange	6	3.66
Deltacom	113	2.84
Deutsche Telekom (IPTransit Pops Only)	30	3.67
Globalcenter	9	8
Globenet	66	2.85
Goodnet	17	3.65
Gridnet	9	4.44
HurricaneElectric	24	4.6
IBM	18	2.67
IIJ	28	3.86
Ion	125	2.34
Kdl	754	2.37
Telcove	53	3.36
TW	71	3.24
UUNET	42	3.67
Xspedius	34	2.88

TABLE I
THE NON-PLANAR NETWORKS.

	Classification	Planar	Non-planar
Layers	Layer1	24	3
	Layer3	90	18
Types	REN	53	0
	Commercial	73	21
Classes	Customer	37	12
	Non-Customer	36	1
Geoextent	Regional	16	1
	Country-Wide	81	13
	Country+	24	7

TABLE II
PLANARITY OF NETWORKS FOR DIFFERENT CLASSIFICATIONS.

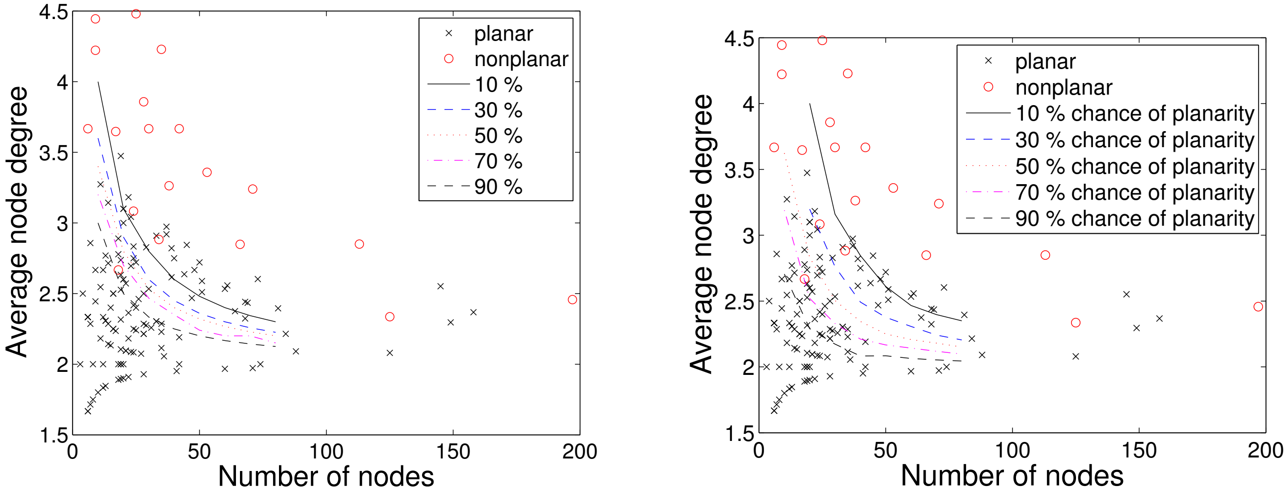
B. Planarity of Random Graphs

An obvious question remains. Is the degree of planarity we observe unusual, or should we expect to see these results?

There are a small set of results discussing planarity of random graphs, but these are asymptotic results for large graphs. Instead, we approach this question by generating a set of synthetic topologies of comparable size and node degree to those in the Zoo. From these we can gain some insight into the likelihood of planarity.

Our first approach is to generate a set of Erdos-Renyi (or Gilbert) random graphs [9], and examine their planarity. An Erdos-Renyi random graph is generated by randomly choosing the node-pairs connected by a fixed number m of links (chosen to set the average node degree, which is given by $2m/n$). However, such a network is not guaranteed to be connected, so we modify it as follows. Assume that you want to generate a random connected graph with n nodes and m edges. Start by picking a node at random and calling that the *connected tree*, then pick another node at random and join it to the tree at a random position, and so on. Continue this process until all n nodes are connected. This forms the connected core of the network. After that, pick the remaining $m + 1 - n$ links randomly from the $(n^2 - 3n + 2)/2$ possible node pairs that haven't been chosen yet.

For each combination of average node degree and network



(a) Random topologies.

(b) Optimized topologies.

Fig. 3. Planarity as a function of average node degree and network size compared to contours of constant probability of planarity derived from two different network models. The percentage for each contour gives the probability of planarity along that curve. Note that we focus on the most important region of the graph which omits a few networks that appear in Figure 1.

size we can generate a set of such networks, and measure their planarity, and from these samples derive an estimate of the probability that a network with those parameters will be planar. However, in order to compare these results with those of the Zoo, it is more useful to have contours of constant probability for particular network size and average node degree. We construct these contours as follow: for each network size n we perform a binary search on the number of links m until we find a value that has given planarity probability p in a 95% confidence interval for 10000 trials. This procedure appears to work well within the limits of the problem (for instance, we cannot always achieve an exact match between a given probability and a particular average node degree because the number of links in a network is a discrete variable).

The contours are plotted against a scatter plot of the Zoo networks in Figure 3 (a). The contours are truncated at a maximum network size of 80 to reduce the computational cost of deriving the contours, focusing on the most interesting region.

The contours do not show a distribution function. They should be interpreted as follows. If a network lies on one of the contours, say for instance the 10% contour (near the top), then there is a one in ten probability that the network will be planar. Most networks do not sit exactly on a displayed contour but we can still estimate their probabilities from the closest contours.

The most obvious feature of Figure 3 (a) is that the number of planar graphs far exceeds what the contours predict. For example, along the 10% contour we would expect approximately 90% of the networks to be non-planar, but in fact the reverse is true – less than 10% of these graphs are non-planar. Thus, based on this simple random model, the degree of planarity observed is highly unexpected.

C. Planarity of Optimized Networks

No-one would seriously argue that Internet-like networks, such as we examine in the Zoo, are well modelled by a simple random graph of the type used above. In this section we use a recent approach for generating realistic Internet-like networks. To generate these topologies we use the idea that network topologies can be modelled as the result of an optimization problem, see for example [33]. This optimization problem is intended to model the fact that networks are designed by network engineers to fulfil a purpose, not purely as the result of a simple random process. Consequently, it is reasonable to expect that networks are optimized with respect to monetary cost and service provided, within the bounds of technological constraints. However, our approach is slightly different from [33] because we aim to be able to control the average node degree in order to match the results to the Zoo networks.

Our approach starts with a set of node locations randomly chosen, independently and uniformly distributed over a rectangle representing the area for the network. To determine traffic demands between the nodes, each node is assigned a population independently at random from a chosen distribution (in this case exponential, although Pareto distributed populations give similar topologies). Traffic demand between a pair of nodes is then proportional to the product of the populations for those nodes, as per a standard gravity model [34].

Once the node positions and traffic have been randomly generated we derive a (near) optimal network using a genetic algorithm. The optimization has three tunable costs to allow us to obtain different types of networks:

0) A link existence cost, $C_0 = k_0$. This is a fixed cost for a link to be in the network.
1) A link length cost, $C_1 = k_1\ell$, where ℓ is the link length, determined by the positions of the two nodes at either end of the link.
2) A bandwidth-length cost, $C_2 = k_2\ell W$, where W is the

link bandwidth. Required bandwidth for a link is equal to a constant multiple of the sum of the traffic demands for the paths traversing that link (we find paths using shortest path routing based on link distances ℓ).

The total cost of a link is given by

$$C = C_0 + C_1 + C_2 = k_0 + k_1\ell + k_2\ell W,$$

We can understand how to choose the parameters k_i if we consider their impact on the final network. The network is required to be connected, so if we were to optimize only with respect to C_0 or C_1, then we would produce a minimum spanning tree. If we were to optimize only with respect to C_2, then a fully-connected network would be the optimum. We can tune between these extremes by modifying the values of the three parameters so there is a balance between the costs. In the cases presented here we do so by fixing k_0 and k_1 and changing the relative value of k_2, which allows us to control the average node degree in the resulting optimized networks. Figure 4 shows three examples of tuning the k_2 parameter for a network with 40 nodes.

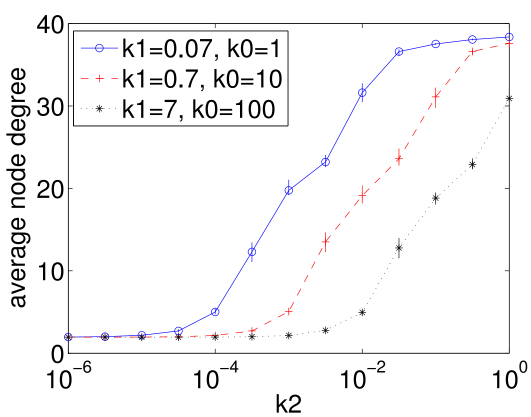


Fig. 4. Average Node Degree vs the parameter k_2. We see that varying k_2 while keeping the other parameters constant allows us fine grained control of node degree. As k_2 tends to infinity the optimal networks become fully connected, and as k_2 goes to zero, the networks tend to a tree.

For each choice of n and k_2 we generate 500 network (from different starting node/population distributions). We measure the proportion of these that are planar as with the random graphs. We also use the bootstrap method to generate 95% confidence intervals for the estimates of the proportion that are planar. We plot the resulting estimates of the expected proportion of planar graphs in Figure 5. We perform the same analysis for networks of sizes 10, 20, 40, ..., 80 though we only plot 30 and 60 in Figure 5 for clarity.

Note that we also looked at the variation in the average node degree for the generated networks. Although the expected average node degree is set by the parameter k_2 there is some variation for particular networks as shown by the confidence interval in Figure 4. However, this variation was found to be small enough that we have not tried to show it in Figure 5.

The procedure for generating contours is slightly more complex here, as we cannot (within a reasonable time) generate as many samples as for the random graph. Instead, we generate the contours by interpolating an inverted version of the curves shown in Figure 5. The resulting contours are shown in Figure 3 (b).

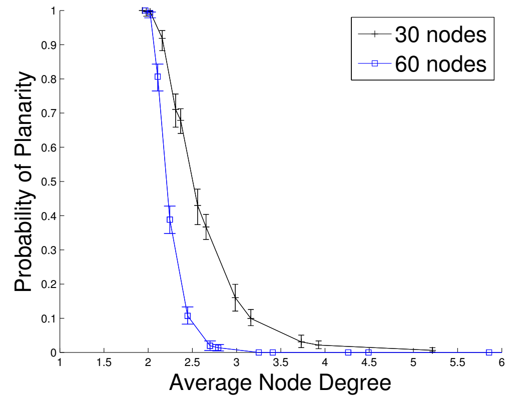


Fig. 5. Probability of planarity as a function of average node degree for optimized networks.

We can see in the figure that the contours for optimized networks are more widely separated than those for random graphs, and that, for instance, the 10% curve is higher indicating that networks can be larger and more connected (higher node degree) before non-planarity sets in. That gives a closer match to the Zoo data, but it is still very far away from explaining the high degree of planarity observed in many of the larger networks.

D. Subgraphs

We now investigate the reasons why some networks are non-planar. Kuratowski's theorem states that a graph is planar if and only if it does not contain a subgraph that is a homeomorphic[3] to K_5 (the complete graph with 5 vertices) or $K_{3,3}$ (complete bipartite graph) [32]. These subgraphs are shown in Figure 6. We identify the Kuratowski subgraph, i.e., the subgraph homeomorphic K_5 or $K_{3,3}$ of our non-planar networks using the the Boyer-Myrvold algorithm [28] implemented in Matlab [29]. Note that a non-planar graph can have more than one Kuratowski subgraph, but that planar graphs cannot contain any.

Table III shows the size of the Kuratowski subgraphs. In comparison to the original graph size, we see that these are 2-3 times smaller, as illustrated in Figure 7. This suggest that perhaps in these networks only a small sub-set of nodes are densely connected and lead to the non-planarity result.

We investigate this further by noting that tree-like structures are always planar, and that such a component cannot be the cause of non-planarity. We removed portions of the Kuratowski subgraphs that were tree-like by taking the 2-core component of the Kuratowski subgraph. We then contract the resulting graphs by removing all degree 2 nodes, effectively reversing subdivision operation. The results are also shown in

[3] Two graphs are homeomorphic if there is an isomorphism from a subdivision of one to a subdivision of the other [35], where a subdivision of a graph results from inserting vertices into edges. Note that subdivision preserves planarity.

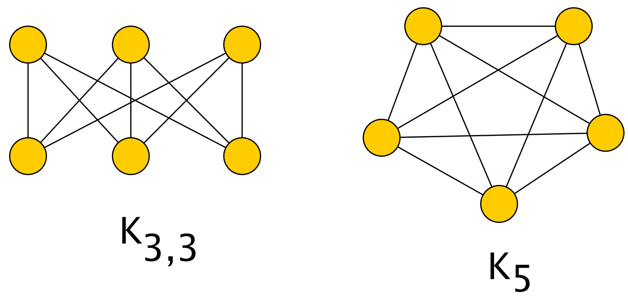


Fig. 6. Example of $K_{3,3}$ and K_5 .

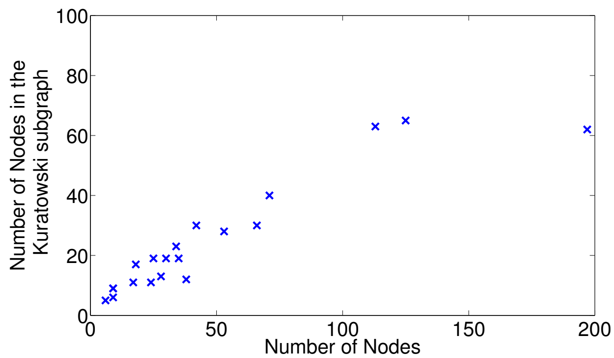


Fig. 7. Size of the Kuratowski subgraph versus the network size.

Table III. These simple operations reduced all of the networks down to less than a dozen nodes, and only a few more edges. Again, this suggests that only a very small subset of nodes are really responsible for non-planarity in these networks. This may be misleading, as at present we only consider one Kuratowski subgraph of each network, but it certainly is suggestive that non-planarity is unusual even in the larger, more complicated networks.

A closer look at the contraction results reveal that only 3 networks (Globalcenter, Gridnet and Dataexchange) have a Kuratowski subgraph homeomorphic to K_5 (the clique on five nodes). In the other 18 networks, their Kuratowski subgraphs contain $K_{3,3}$ subdivisions. We don't know the significance of this finding as yet, but it suggests some interesting possibilities for future research.

IV. Speculation

The results above are not definitive by any means, but they are very suggestive. The data at our disposal reveal an interesting, and unexpected feature of Internet networks, that is, a high likelihood of being planar. The likelihood is much higher than for the two models of network generation that we tested (it will be interesting in the future to compare other models of network generation to see if any are so likely to generate planar networks). Even in the non-planar cases, it seems that only a small subset of nodes contribute to non-planarity. We speculate that planarity does not occur by accident. It is much more likely that it is being deliberately imposed on network designs.

Why might a network engineer require planarity? Simply speaking, network engineers have to understand their network. They need to work with it to debug problems, and to manage devices. A planar graph, by virtue of being easy to draw, can be more easily understood, and therefore manage. A more complex, harder to understand network is certainly possible, but will have some cost (in terms of the engineer's time at least). These types of cost are hard to quantify and hence have often been ignored in the literature on network design and optimization. In fact, most networks are not designed using formal mathematical algorithms. They are designed "by eye". Perhaps this is in part because of the fact that formal optimization networks generate complicated, hard to understand networks.

We do not suggest that network engineers deliberately design planar networks, nor would they design for other obscure mathematical properties, but rather that planarity is one signature of the desire for simple designs.

V. Conclusion and future work

This paper has shown a surprising degree of planarity in observed networks. We speculated that this is caused by network designers who prefer network designs that are simple to understand. There is another potential explanation, namely that the Zoo is preferentially biased towards networks that can be drawn, because we use network maps to populate the Zoo. However, remember that non-planarity does not mean a network cannot be drawn. It only means there will be link crossings, and we observe such in many of the graphs. Even maps of planar networks often contain crossings simply because maps are usually a geographic approximation to the network, and therefore the nodes and links aren't be placed arbitrarily. In any case, it is an ongoing task to better measure this property and confirm the results.

Our ideas about the cause of planarity seem intuitive, but are speculation none-the-less. We also aim to test these ideas more thoroughly through mathematical methods, e.g., by testing other types of network synthesis models to see if it arises naturally through other means, or through looking more carefully for the sources of non-planarity in graphs. We can also look for other signatures of "simplicity" in network design, to see if the underlying hypothesis about how network engineers work is justified.

As a network evolves, and grows, engineers may not mathematically optimize the network, but they certainly do try to improve it. So the optimization paradigm for network synthesis should not be discarded as a result of this work. We need, however, to consider the affect of more complicated optimization objectives, which penalize "complex" designs. However, it is not at all clear how to create such an objective, or optimize a network against it, and so this is yet another topic for future research.

References

[1] B.M.Waxman, "Routing of multipoint connections," *IEEE J. Select. Areas Commun.*, vol. 6, no. 9, pp. 1617–1622, 1988.

[2] M. Faloutsos, P. Faloutsos, and C. Faloutsos, "On power-law relationships of the Internet topology," in *ACM SIGCOMM*, pp. 251–262, 1999.

Network Name	Kuratowski subgraph		2-core		Contracting degree 2 nodes	
	# of nodes	#of edges	# of nodes	# of edges	# of nodes	# of edges
Airtel	6	9	6	9	6	9
AT&T(IP-MPLS)	19	22	13	16	8	11
BT(North America)	19	22	16	19	7	10
Chinanet	12	15	10	13	8	11
Cogentco	62	65	25	28	8	11
Dataexchange	5	10	5	10	5	10
Deltacom	63	66	31	34	8	11
Deutsche Telekom (IPTransit Pops Only)	19	22	14	17	9	12
Globalcenter	9	14	6	11	6	11
Globenet	30	33	21	24	8	11
Goodnet	11	14	11	14	7	10
Gridnet	9	14	9	14	5	10
HurricaneElectric	11	14	10	13	7	10
IBM	17	20	14	17	8	11
IIJ	13	16	11	14	8	11
Ion	65	68	27	30	8	11
Kdl	263	266	202	205	11	14
Telcove	28	31	27	30	8	11
TW	40	43	29	32	9	12
UUNET	30	33	22	25	7	10
Xspedius	23	26	20	23	7	10

TABLE III
KURATOWSKI SUBGRAPHS OF NON-PLANAR GRAPHS.

[3] A. Barabási and R. Albert, “Emergence of scaling in random networks,” *Science*, vol. 286, 1999.

[4] R. Albert, H.Jeong, and A.-L. Barabási, “Error and attack tolerance of complex networks,” *Nature*, vol. 406, pp. 378–382, 2000.

[5] S.-H. Yook, H.Jeong, and A.-L. Barabási, “Modeling the Internet’s large-scale topology,” *PNAS*, no. 99, pp. 13382–13386, 2002.

[6] E. W. Zegura, K. Calvert, and S. Bhattacharjee, “How to model an internetwork,” in *Proceedings of IEEE Infocom ’96*, (San Francisco, CA), 1996.

[7] H. Tangmunarunkit, R. Govindan, S. Jamin, S. Shenker, and W. Willinger, “Network topology generators: degree-based vs. structural,” *SIGCOMM Comput. Commun. Rev.*, vol. 32, pp. 147–159, August 2002.

[8] J. C. Doyle, D. L. Alderson, L. Li, S. Low, M. Roughan, S. Shalunov, R. Tanaka, and W. Willinger, “The “robust yet fragile” nature of the Internet,” *Proceedings of the National Academy of Sciences of the USA (PNAS)*, vol. 102, pp. 14497–502, October 2005.

[9] E. Gilbert, “Random graphs,” *Annals of Mathematical Statistics*, vol. 30, pp. 1441–1144, 1959.

[10] S. Knight, H. X. Nguyen, N. Falkner, R. Bowden, and M. Roughan, “The internet topology zoo,” tech. rep., University of Adelaide, Australia, www.topology-zoo.org, 2011.

[11] A. Feldmann, A. Greenberg, C. Lund, N. Reingold, and J. Rexford, “Netscope: Traffic engineering for IP networks,” *IEEE Network Magazine*, pp. 11–19, March/April 2000.

[12] A. Shaikh and A. Greenberg, “OSPF Monitoring: Architecture, Design and Deployment Experience,” in *Proc. USENIX Symposium on Networked System Design and Implementation (NSDI)*, March 2004.

[13] “Internet Protocol.” IETF RFC 791, September 1981.

[14] F. Baker, “Requirements for IP version 4 routers.” IETF, Network Working Group, Request for Comments: 1812, July 1993.

[15] C. Smith, “Traceroute - whitepaper.” http://www.informatik.uni-trier.de/~smith/networks/tspec.html.

[16] V. Jacobson, “Traceroute.” ftp://ftp.ee.lbl.gov/traceroute.tar.gz, 1989-04.

[17] W. Willinger, “The science of complex networks and the Internet: Lies, damned lies, and statistics,” Feb 2010. http://www.maths.adelaide.edu.au/matthew.roughan/workshops.html.

[18] CAIDA, “Skitter.” http://www.caida.org/tools/measurement/skitter/.

[19] N. Spring, R. Mahajan, and D. Wetherall, “Measuring ISP topologies with Rocketfuel,” in *ACM SIGCOMM*, (Pittsburg, USA), August 2002.

[20] H. V. Madhyastha, T. Isdal, M. Piatek, C. Dixon, T. Anderson, A. Krishnamurthy, and A. Venkataramani, “iPlane: An information plane for distributed services,” in *OSDI*, November 2006.

[21] D. Feldman and Y. Shavitt, “Automatic large scale generation of Internet PoP level maps,” in *IEEE Globecom 2008*, pp. 2426–2431, 2008.

[22] K. Yoshida, Y. Kikuchi, M. Yamamoto, Y. Fujii, K. Nagami, I. Nakagawa, and H. Esaki, “Inferring PoP-level ISP topology through end-to-end delay measurement,” in *PAM 2009*, pp. 35–44, 2009.

[23] Y. Shavitt and N. Zilberman, “A structural approach for PoP geo-location,” in *IEEE Infocom 2010*, 2010.

[24] A. A. Hagberg, D. A. Schult, and P. J. Swart, “Exploring network structure, dynamics, and function using NetworkX,” in *Proceedings of the 7th Python in Science Conference (SciPy2008)*, (Pasadena, CA USA), pp. 11–15, August 2008.

[25] U. Brandes, M. Eiglsperger, I. Herman, M. Himsolt, and M. Marshall, “GraphML Progress Report Structural Layer Proposal,” in *Graph Drawing*, pp. 109–112, Springer, 2002.

[26] E. Gansner and S. North, “An open graph visualization system and its applications to software engineering,” *Software: practice and experience*, vol. 30, no. 11, pp. 1203–1233, 2000.

[27] A. H. Rasti, N. Magharei, R. Rejaie, and W. Willinger, “Eyeball ASes: From geography to connectivity,” in *ACM Sigcomm IMC*, (Melbourne, Australia), 2010.

[28] J. M. Boyer and W. J. Myrvold, “On the cutting edge: simplified O(n) planarity by edge addition,” *Journal of Graph Algorithms and Applications*, vol. 8, no. 3, p. 241273, 2004.

[29] www.stanford.edu/ dgleich/programs/matlab_bgl/.

[30] “Abilene network.” http://www.internet2.edu/pubs/200502-IS-AN.pdf, 2005.

[31] “Uunet.” http://www.nthelp.com/images/uunet.jpg.

[32] http://en.wikipedia.org/wiki/Planar_graph.

[33] L. Li, D. Alderson, W. Willinger, and J. Doyle, “A first-principles approach to understanding the Internet’s router-level topology,” in *Proceedings of the 2004 conference on Applications, technologies, architectures, and protocols for computer communications*, SIGCOMM ’04, (New York, NY, USA), pp. 3–14, ACM, 2004.

[34] M. Roughan, “Simplifying the synthesis of Internet traffic matrices,” *ACM SIGCOMM Computer Communications Review*, vol. 35, pp. 93–96, October 2005.

[35] http://en.wikipedia.org/wiki/Graph_homeomorphism.