Radial structure of the Internet

By Petter Holme[1,*], Josh Karlin[1] and Stephanie Forrest[1,2]

[1]*Department of Computer Science, University of New Mexico,
Albuquerque, NM 87131, USA*
[2]*Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA*

The structure of the Internet at the autonomous system (AS) level has been studied by the physics, mathematics and computer science communities. We extend this work to include features of the core and the periphery, taking a radial perspective on AS network structure. New methods for plotting AS data are described, and they are used to analyse datasets that have been extended to contain edges missing from earlier collections. The average distance from one vertex to the rest of the network is used as the baseline metric for investigating radial structure. Common vertex-specific quantities are plotted against this metric to reveal distinctive characteristics of central and peripheral vertices. Two datasets are analysed using these measures as well as two common generative models (Barabási–Albert and Inet). We find a clear distinction between the highly connected core and a sparse periphery. We also find that the periphery has a more complex structure than that predicted by degree distribution or the two generative models.

Keywords: Internet; autonomous systems graph; complex networks; radial structure; centrality; hierarchy

1. Introduction

Since the turn of the century, there has been increasing interest in the statistical study of networks (Albert & Barabási 2002; Dorogovtsev & Mendes 2003; Newman 2003), stimulated in large part by the availability of large-scale network datasets. One network of interest is the Internet (Pastor-Santorras & Vespignani 2004). The Internet is intriguing because its complexity and size preclude comprehensive study. It comprises millions of individual end nodes connected to tens of thousands of Internet service providers (ISPs) whose relationships are continually in flux and only partially observable. One way to cope with these complexities is by analysing a single scale of Internet data, for example, a local office network of computers and their interconnections; a network of email address book contacts; the network formed by URL links on the World Wide Web; or the interdomain AS level of the Internet. This paper is concerned with the last of these examples—the AS graph. The vertices in the graph are themselves computer networks; roughly speaking an AS is an independently operated network or set of networks owned by a single entity. Edges represent pairs of ASs that can directly communicate.

* Author for correspondence (holme@cs.unm.edu).

1

A major finding of earlier AS studies is that node degree (number of links to other ASs) has a power-law distribution (Faloutsos *et al.* 1999). The degree distribution is, however, not the only structure that affects Internet dynamics (Doyle *et al.* 2005). In this paper, we investigate higher-order (beyond the degree distribution) network structures that also impact network dynamics. We analyse the AS graph using methods that are appropriate for networks with a clear hierarchical organization (Pastor-Santorras & Vespignani 2004; Trusina *et al.* 2004). In particular, we study network quantities as a function of the average distance to other vertices. This approach separates vertices of different hierarchical levels, in a radial fashion, ranging from central (in the sense of the closeness centrality; Sabidussi 1966) to peripheral vertices. This is, furthermore, a way to measure how clearly separated the core and the periphery are. Most analysis methods developed by physicists (degree frequencies, correlations, etc.) are based on quantities averaged over the whole network and do not take a hierarchical partitioning into account (Pastor-Santorras & Vespignani 2004). Studies by computer scientists, on the other hand, assume a division of the AS-level Internet into hierarchical levels (Subramanian *et al.* 2002). We argue that observed AS-level networks do have pronounced core–periphery dichotomy but that the periphery has more structure than previously thought.

2. Networks

This section briefly reviews the organization of the AS-level Internet and describes how we obtained our datasets. We also describe the network models to which we compare our observed data. These models include one randomization scheme that samples random networks with the same set of degrees as the original networks, and the generative BA and Inet models. Technically, all three models are null models, but to contrast the randomized networks (having N degrees of freedom) with the generative models (having only a few degrees of freedom) we reserve the term null model to the former.

The data are represented as a network $G=(V, E)$, where V is a set of N vertices (ASs) and E is a set of M undirected edges (connections between ASs). The Internet is currently composed of roughly 22 000 individual networks known as ASs. Each of these systems peer with a (usually small) set of ASs to form a connected network. The protocol used to establish peering sessions and discover routes to distant ASs is called the Border Gateway Protocol (BGP). Two typical peering relationships are: customer–provider in which the provider provides connectivity to the rest of the Internet for the customer and peer–peer in which the peering ASs transfer traffic between their respective customers. The extreme core of the network, the Tier-1 ASs, has many peer–peer and customer links but no providers. Nodes closer to the periphery of the network have fewer customers and peers but more providers.

(a) Autonomous system networks

We analyse four real-world datasets (i.e. datasets collected using observed network data rather than simulated networks that are generated synthetically) of which two are original. The first two datasets are well known and well studied (Chang *et al.* 2002) dating from 2002 and the second two are recent, inferred from

2006 data. The first graph in each pair consists of edges learned solely from dumps of router state, known as routing information bases (RIBs) (http://www. routeviews.org/data.html). RIBs are a standard source of AS connectivity data. The second graph in each pair contains RIB information augmented with edges derived from other sources (such as routing registries, looking glass servers and update messages) which produces a more accurate representation of the real network. The additional sources are described below.

(i) *Obtaining routing information bases from Route Views*

BGP routers store the most recent AS path for each IP block (prefix) announced by its peers. These data are stored in the router's RIB, and periodic RIB dumps from a large number of voluntary sources are available from Route Views (http://www.routeviews.org). Each RIB represents a static snapshot of all routes available to the router from which it was obtained. Since BGP only disseminates each router's best path, and this value is dynamic as links go up and down, a sizable portion of the network is hidden from each router. In order to obtain a more complete topology, common practice is to take the union of the relationships found in a large number of RIB samples. From the samples, AS relationships are then inferred from the routing paths. A path is comprised of connected ASs, and therefore each pair of adjacent ASs in a path corresponds to an edge in the graph.

The 2002 graph taken from a single RIB (RIB'02) was inferred from Route Views on 15 May 2002. We constructed the 2006 RIB graph (RIB'06) from the Route Views RIB on 16 May 2006. The RIB'02 graph has $N=13\,233$ and $M=27\,724$, while RIB'06 has $N=22\,403$ and $M=46\,343$.

(ii) *Extending the routing information base dataset*

There are other sources of AS connectivity data besides Route Views. RIPE (http://www.ripe.net) has data collected from additional RIBs beyond those contained in the Route Views data. Peering information is directly available for a small number of ASs that are participating Looking Glass (http://www. traceroute.org) routers. Finally, some ASs register their peering relationships in regional registries such as RIPE. The extended 2002 AS graph (AS'02) was constructed using inferred topologies from all three of these sources, together with the original Route Views data.

RIB data represent a brief snapshot of routing state. There are many paths that a router sees only briefly, and the chances of capturing all of them from just a few RIB dumps are small. In the extended AS graph of 2006 (AS'06), we augmented the Route Views RIB data with all of the paths found in BGP update messages for the entire month of April 2006 from both Route Views and RIPE. This gives a more complete picture over time, although it is still biased by the limited number of routers from which the data were collected.

The extended 2002 AS graph (AS'02) has $N=13\,579$ and $M=37\,448$, and the corresponding 2006 network (AS'06) has $N=22\,688$ and $M=62\,637$. Thus, the extended datasets have 35% (2002) and 35% (2006) more edges than their RIB counterparts.

(b) Null-model networks

We are interested in network structure beyond degree distribution, so we compare our AS network data against a null model with the same degree distribution. Our null model is a random network constrained to have the same set of degrees as the original network. By comparing results for the observed networks with the same quantities for the null model, we can observe additional network structure if it exists. The standard way to sample such networks is by randomizing the original network with stochastic rewiring of the edges (see Gale 1957, for an early example). In our implementation, we create a new random network by enumerating the edges E of the original graph, and for each edge (i, j) we

(i) choose another edge (i', j') randomly and replace (i, j) and (i', j') with (i', j'). If this creates a multi- or self-edge, then revert to the original edges (i, j) and (i', j'), and repeat with a new (i', j') and

(ii) choose two edges (i_1, j_1) and (i_2, j_2) and replace them along with (i, j') by (i_1, j'), (i, j_2) and (i_2, j_1).

Step (ii) guarantees ergodicity of the sampling (Roberts 2000), i.e. that one can go between any pair of graphs with a given set of degrees by successive edge rewirings.

(c) Generative network models

We also study networks produced according to two previously proposed network-generation schemes (Barabási & Albert 1999; Winick & Jamin 2000). The first is the well-known Barabási–Albert (BA) preferential attachment model (Barabási & Albert 1999). The second, known as the Inet model, v. 3.0 (Winick & Jamin 2000), is more complex and designed specifically for creating networks with AS graph properties.

(i) Barabási–Albert model

The BA model is a general growth model for producing networks with power-law degree distributions (Barabási & Albert 1999). Vertices and edges are iteratively added to the network according to a preferential attachment rule that ensures that a power-law degree distribution emerges.

More precisely, the initial configuration consists of m isolated vertices. From this configuration, the network is iteratively grown. At each time-step, one vertex is added, together with m edges leading out from the new vertex. The edges are attached to vertices in the graph such that

(i) the probability of attaching to a vertex i is proportional to $k(i)$ and
(ii) no multiple edges, or self-edges, are formed.

This procedure produces a network which has, in the $N \to \infty$ limit, a degree distribution $P(k) \sim k^{-3}$ for $k \geq m$, and $P(k) = 0$ for $k < m$.

Since the BA model has only one integer parameter, it is not very flexible at fitting data. We use $m = 3$ to make the average degree as similar to the AS networks as possible. Other preferential attachment models (e.g. Pastor-Satorras et al. 2001)

can model the average degree and slope of the degree distribution more closely. Such improvements, we believe, are unlikely to change the conclusions drawn from the original BA model.

(ii) *Inet model*

The Inet model (Winick & Jamin 2000) is less general than the BA model. Its objective is to regenerate the AS graph as accurately as possible rather than focus on a single mechanism to create and explain scale-free networks. The scheme is rather detailed and we only sketch its strategy here. Starting with N vertices, Inet first generates random numbers that represent the final degree of the vertices such that the degree distribution matches the observed distribution of the AS graph as closely as possible. This means that the low-degree end of the distribution is more accurately modelled by Inet than the BA model, because the BA model does not produce a vertex with degree less than m. In the real AS graph, there are a considerable fraction of degree-one vertices. After the degrees are assigned to the vertices, edges are added in such a way that the degree–degree correlation properties of the original AS graph are matched as closely as possible.

A more detailed explanation of this procedure and its rationale are given in Winick & Jamin (2000). We use Inet's default parameter settings, except N, which we extracted from our datasets, producing an average degree that is approximately six.

3. Numerical results

This section presents the numerical results of our analysis. We first discuss the average distance metric used to display network properties from a radial perspective. Then, we define and present the results for each network structural measure as a function of the average distance to other vertices.

Let $d(i, j)$ denote the graph distance between two vertices i and j—the number of edges in the shortest path between i and j. A simple measure for how peripheral a vertex is in the network is its *eccentricity*—the distance to the most distant vertex, $\max_{j \in V} d(i, j)$ (Buckley & Harary 1989). Eccentricity is thus an extremal property of the network and is determined by a small fraction of vertices. To better reflect the typical path length of a vertex, we rank vertices according to an average property. The average property corresponding to eccentricity is the average distance from one vertex to all of the others,

$$\bar{d}(i) = \frac{1}{N-1} \sum_{j} d(i, j), \tag{3.1}$$

where the sum is over all vertices, except i, in V. We note that the reciprocal value of $\bar{d}(i)$, the *closeness centrality*, is a common measure for centrality in social network studies (Sabidussi 1966; Buckley & Harary 1989). Average distance is a more intuitive measure in this context—$\bar{d}(i) \approx 2$ means that i is on average two hops away from other vertices, whereas the closeness value 0.5 does not have such a direct interpretation.

Another way to study eccentricity is by iteratively removing vertices of low degree to construct a sequence of k cores (subgraphs in which all vertices have degree k or more; Alvarez-Hamelin *et al.* 2006). We used the average distance metric

instead, because it measures separation of vertices, i.e. the values on the x-axis are not only integers as for the eccentricity. Further, because it is a global measure (in the sense that the entire network topology affects $\bar{d}(i)$ for every i), it is likely more robust to errors in the input data.

Peering policies do not always allow each router to pick the shortest topological path to every destination. For instance, a route learned through a customer might be longer than through a peer-to-peer link but because it would provide revenue, policy requires that the customer route be used. Therefore, the $\bar{d}(i)$ values shown in this paper do not always represent valid routing paths. More accurate measurements would require relationship information which is difficult to obtain, because companies have an incentive to conceal it and inferencing methods are inaccurate.

(a) Radial vertex density

We first plot the fraction of vertices as a function of $\bar{d}$. Figure 1 shows the distribution of $\bar{d}$ for our datasets and the model AS graphs. The observed networks produce graphs that are far from smooth unimodal distributions. Instead they have one peak close to $\bar{d} = 3$, a smaller peak around $\bar{d} = 4$, and for the 2006 data, a third peak near $\bar{d} = 5$. The difference between the RIB-only and the extended datasets is small, except around the second peak in figure 1b which is higher in the RIB-only data. The null-model curves are much more unimodal, although they do not follow a simple, smooth functional form. Such a unimodal form could be a result of the averaging of many null-model curves, but the observation holds even if single realizations of the randomization are plotted (data not shown). Thus, the observed AS graph is less homogeneous than we would predict by considering only vertex degree.

We interpret the two peaks as an effect of the hierarchical organization of the Internet. The core (Tier-1 providers and other large ISPs) is in the low-$\bar{d}$ tail, the $\bar{d} = 3$ peaks are vertices directly connected to the core, and the $\bar{d} = 4$ peaks are vertices whose closest neighbours are in the $\bar{d} = 3$ peaks. This explains the approximately integer distance between the peaks. As expected, the Tier-1 ASs (AS numbers 209, 701, 1239, 1668, 2914, 3356, 3549, 3561, 6461 and 7018 in our datasets) have an average $\bar{d} = 2.35 \pm 0.03$ in the AS'02 data and $\bar{d} = 2.41 \pm 0.03$ in the AS'06 data and appear in the centre of the network (left of the most central peak). Thus, the Tier-1 ASs are in the extreme low end of the $\bar{d}$-spectrum.

Results for the BA and Inet model networks are shown in figure 1c. The Inet model has a peak to the left of the middle of the range of distances, but no second or third peak. The BA model matches the observed network even less accurately—its peak is at a relatively high $\bar{d}$ value.

(b) Degree

Degree distribution is now a classical quantity in the study of the Internet topology. Faloutsos *et al.* (1999) report a highly skewed distribution of degree, fitting well to a power law with an exponent around 2.2. Since this finding, the degree distribution has become a core component in models of the AS graph—both the BA and Inet models as well as others (Medina *et al.* 2000; Fabrikant *et al.* 2002; Alvarez-Hamelin & Schabanel 2004) create networks with power-law
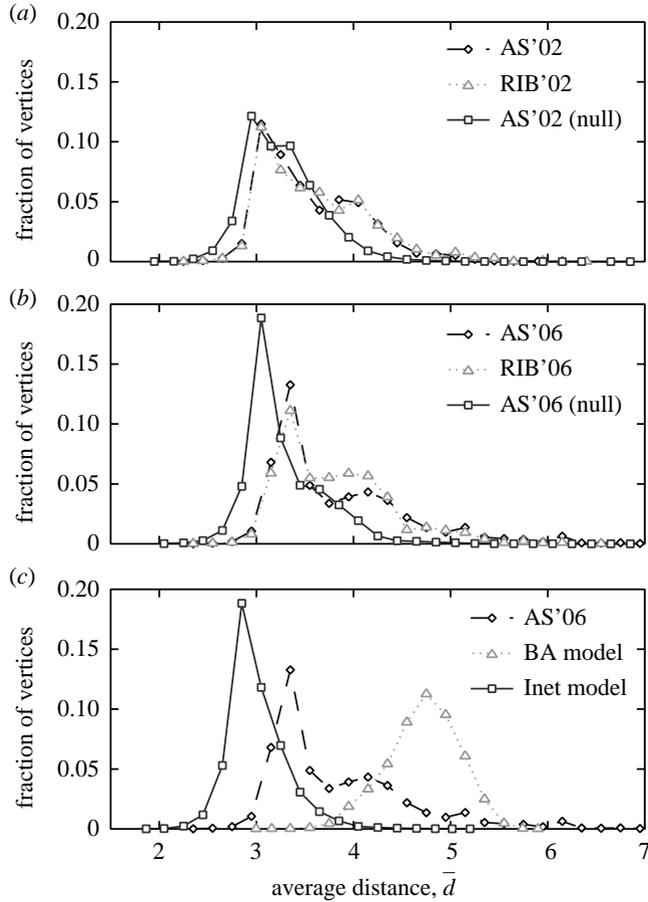
Figure 1. Normalized histograms of vertices with a specific average distance $\bar{d}$ to the rest of the vertices. (*a*) shows curves for the Oregon Route Views data (RIB'02), extended data (AS'02) and values for random networks with the same degree sequences as AS'02. (*b*) displays curves for the Oregon Route Views data (RIB'06), extended data (AS'06) as well as randomized networks with the degree sequence of AS'06. (*c*) shows the same AS'06 curve as (*b*) along with the BA and Inet model results for parameter values as close as possible to those of the AS'06 network. For the null-model curves in (*a*) and (*b*) as well as the model networks in (*c*), 100 averages were used. Lines are guides for the eyes. The error bars represent standard error (the point symbols are often larger than the error bars).

degree distributions. One interpretation of degree is that it is a local centrality measure (Buckley & Harary 1989). Further, different measures of centrality are known to be highly correlated (Nakao 1990; Holme *et al.* 2002; Lee 2006), so one can expect the average degree k to be a decreasing function of the average distance $\bar{d}$.

Figure 2 confirms this prediction for both the observed and model networks. In figure 2*a,b*, we observe that the $k(\bar{d})$ curves decrease dramatically until the approximate location of the first peak in the distribution plots figure 1*a,b*. Therefore, $\bar{d}$ identifies a natural border between the core vertices of high degree and low average distance, and the sparsely connected periphery. The observed graphs, however, have higher degree in the periphery compared with the
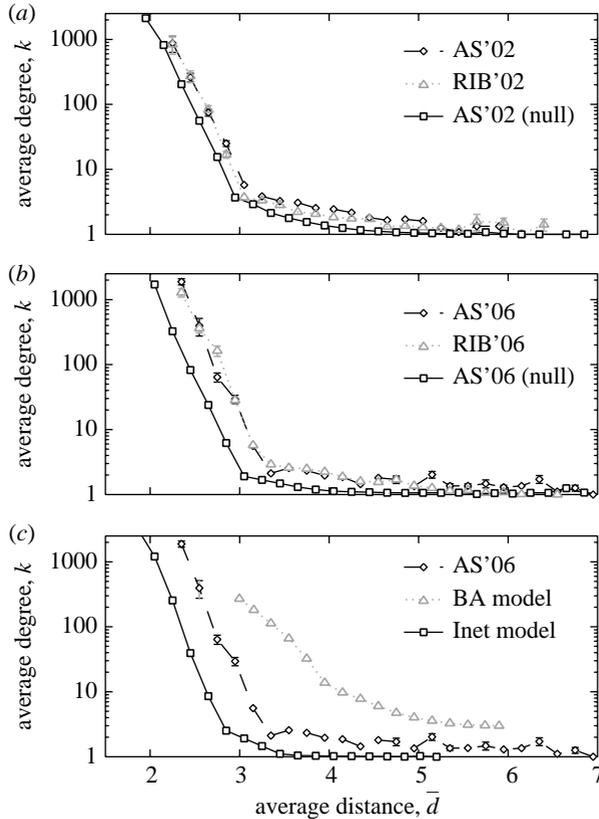
Figure 2. Degree k as a function of the average distance $\bar{d}$. The panels and symbols are the same as given in figure legend 1.

null-model curves. This suggests that the network periphery may have more complex wiring topology than that is predicted by degree distribution alone. This pattern occurs in our other network measurements as well.

The Inet model (figure 2c) fails to capture the higher degree (implying additional complexity) in the periphery. Since the BA model has a minimal degree of three, it is difficult to compare to the observed networks. However, the decrease of the $k(\bar{d})$ curves at the largest $\bar{d}$ peak is not conspicuous in the BA model curves. Thus, there is no clear core–periphery dichotomy in the BA model. This too is not surprising, because the BA model was designed to produce 'scale-free' networks in the sense of fractals (if one zooms in on any part of system, it looks similar to the whole).

(*c*) *Neighbour degree*

Degree is a property of individual vertices, with no information about how they are interconnected. In this sense, degree is a measure of local network structure. A common way to broaden the perspective to understand the network's non-local organization (Mahadevan *et al.* 2005) is to measure the correlations of degrees between neighbours in the network. There are three common approaches. The first, known as *assortative mixing coefficient* (Newman 2003), measures the
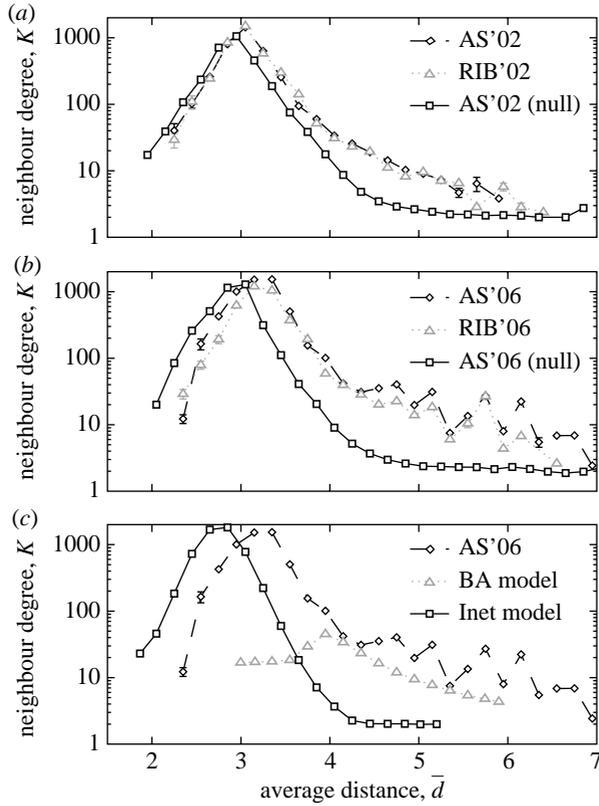
Figure 3. Neighbour degree K as a function of the average distance $\bar{d}$. The panels and symbols are the same as given in figure legend 1.

Pearson correlation coefficient for each edge. This provides one number for the entire network and is thus appropriate for comparisons between networks. The second approach makes a density plot that displays the fraction of edges with degree (k_1, k_2). This kind of two-dimensional plot is called a *correlation profile* (Maslov *et al.* 2004; Mahadevan *et al.* 2006). Correlation profiles provide more detailed information than the assortative mixing coefficient, but they are less concise and more sensitive to noisy data. The third approach measures average neighbour degree

$$K(i) = \frac{1}{k(i)} \sum_{j \in \Gamma_i} k(j),\tag{3.2}$$

(where Γ_i is the neighbourhood of i) as a function of degree $k(i)$ (Pastor-Satorras *et al.* 2001). All approaches must be compared to null models because skewed degree distributions are known to induce anticorrelations (Maslov *et al.* 2004). The third approach produces a one-dimensional plot and thus forms a middle ground between the assortative mixing coefficient and the correlation profile. It is also a method that can be adapted to our radial-plot framework—by plotting K against $\bar{d}$, we can monitor the correlation between centrality and neighbour degree. For the AS-level Internet, it has been observed that the $K(k)$ curves

decay (Pastor-Satorras *et al.* 2001). In other words, high-degree vertices are, on average, connected to vertices of low degree and vice versa. Then, since degree decreases with $\bar{d}$, one would then expect K to be an increasing function of $\bar{d}$.

As seen in figure 3, vertices at intermediate distances have neighbours of highest degree. The peak in $K(\bar{d})$ coincides with the largest peak in the histograms found in figure 1 and the change of slope in figure 2. This suggests that the periphery is composed of two levels: the intermediate majority that is primarily connected to the core and the extreme periphery that is connected to other periphery vertices.

It is also apparent in figure 3*a,b* that the null model qualitatively has the same shape as the observed network; but, just as for k; K are larger in the observed networks than the null model. Also, the Inet model underestimates the average neighbour degree in the periphery. Finally, the BA model exhibits less correlation between K and $\bar{d}$.

(*d*) Deletion impact

If a vertex is not actively routing packets due to fault or attack, other vertices might be affected. We are interested in knowing how susceptible a given network structure is to random node failures. Assuming that the network is connected, let S_i be the number of vertices in the largest connected subgraph after the deletion of i. We define the *deletion impact* as

$$\phi(i) = \frac{N-1-S_i}{N-2}. \tag{3.3}$$

This measure can take values in the interval $[0, 1]$. A value of 0 means that the entire network, except i, is still connected after the deletion. A value of 1 means that all of the network's edges are attached to i and that all of the vertices are isolated after the deletion.

Figure 4 plots deletion impact as a function of the average distance for the same datasets as the previous figures. All curves are roughly decreasing. This means that the network is more sensitive to the deletion of central, than peripheral, vertices. This observation is anticipated from earlier studies showing that the Internet is vulnerable to targeted attacks at the vertices of highest degree (Albert *et al.* 2000) but robust to random failures. This is because the majority of vertices have low ϕ-values. However, the deletion impact measure can detect more subtle effects in the periphery.

The first peak in the $\bar{d}$ distribution is, as mentioned above, around $\bar{d}=3$. At this distance, ϕ has decreased 1000 times from the core where $\phi \sim 10^{-2}$. In this quantity, we see a substantial difference from the null model; the peripheral vertices of the inferred networks have significantly lower deletion impact than the peripheral vertices of the null-model networks. This, we believe, is another effect of the high degree of peripheral vertices. The fact that the periphery is relatively highly connected suggests that there are alternate routes which could be used if a regular path is obstructed by a vertex failure. In the case of the Inet model, which has very few vertices of high $\bar{d}$, the peripheral ϕ values are low because the periphery is well connected to the core. As expected, $\phi=0$ for all vertices in the BA model since all vertices have degree of at least three. The BA model thus produces networks that are more robust to vertex deletion than the observed networks are.
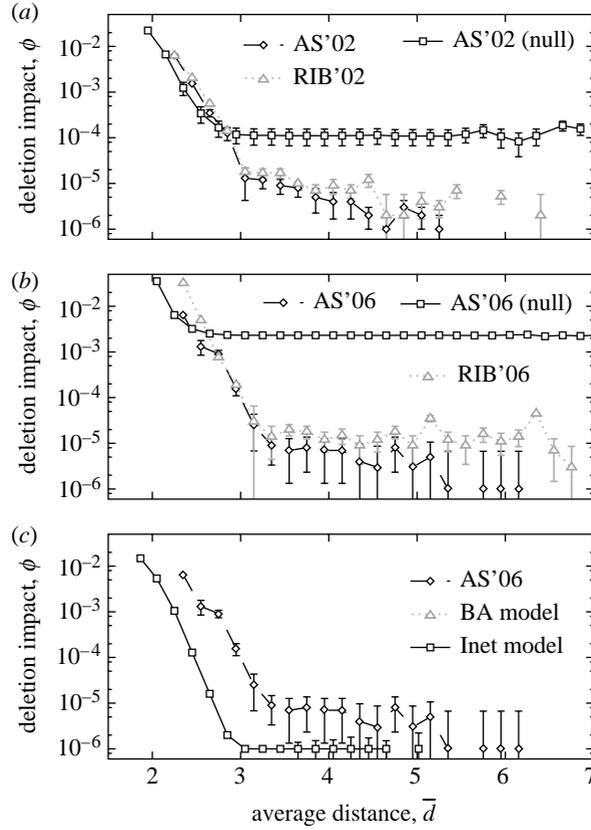
Figure 4. Deletion impact ϕ as a function of the average distance $\bar{d}$. The panels and symbols are the same as given in figure legend 1.

(e) Clustering coefficient

The *clustering coefficient* $C(i)$ (Watts & Strogatz 1998) is another frequently studied network property:

$$C(i) = \frac{M(\Gamma_i)}{\binom{k(i)}{2}}. \tag{3.4}$$

$M(X)$ denotes the number of edges in a subgraph X, and Γ_i is the neighbourhood of i. The clustering coefficient measures how interconnected the neighbourhood of a vertex is. One interpretation is that $C(i)$ is the number of connected neighbour pairs rescaled by the theoretical maximum. $C(i)$ can also be seen as the fraction of triangles that i is a member of, normalized to the interval $[0, 1]$.

In figure 5, we display the clustering coefficient as a function of the average distance. The curves for the observed graph, null-model and Inet model networks show a peak around the same point as the peak in the $\bar{d}$ distribution. However, the clustering in the periphery is higher in the inferred networks than in the null models. In other words, there are more triangles in the periphery than can be expected from only the network's degree distribution. In fact, for 100 null-model
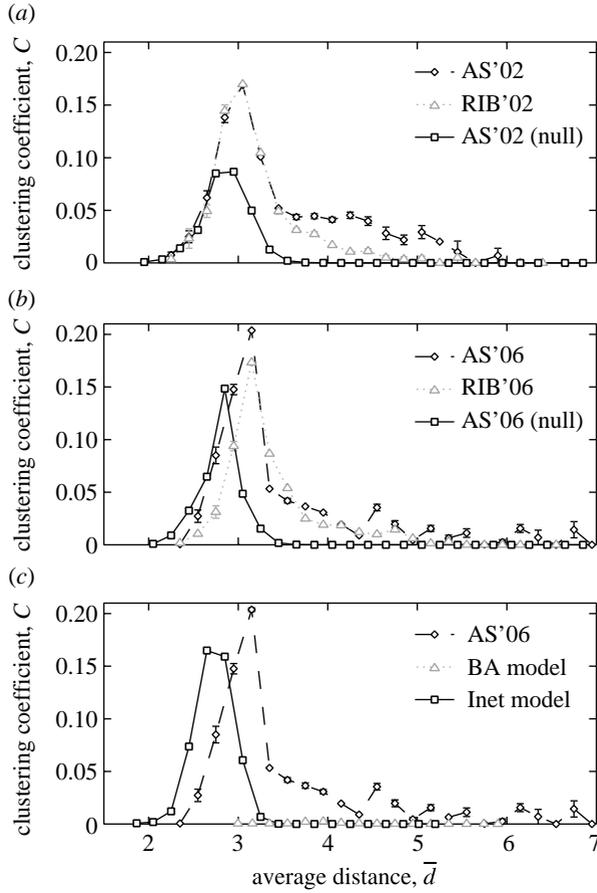
(*a*)

(*b*)

(*c*)

Figure 5. Clustering coefficient C as a function of the average distance $\bar{d}$. The panels and symbols are the same as given in figure legend 1.

networks based on the AS'06 network, no triangles existed for $\bar{d} > 3.8$ with any vertex having $\bar{d} > 3.8$. This should be compared with 1124 triangles for the AS'06 network itself (there are even 83 triangles where all vertices have $\bar{d} > 3.8$). This provides additional evidence that the periphery of the observed AS graphs is complex. As triangles represent redundancy (the three vertices will still be connected if any one of the edges are cut); this could help explain the increased robustness to deletion seen in figure 4. As seen in figure 5*b*, neither the Inet nor the BA model predicts a significant number of peripheral triangles. The low deletion impact values for peripheral vertices in these models may be attributed to the presence of longer cycles.

(*f*) *Distance balance*

In the context of scientific collaboration networks, it has been shown (Newman 2001) that the number of shortest paths leaving a vertex via a specific neighbour is skew distributed. In other words, most of the shortest paths from a vertex i to the rest of the network traverse a single neighbour of i.
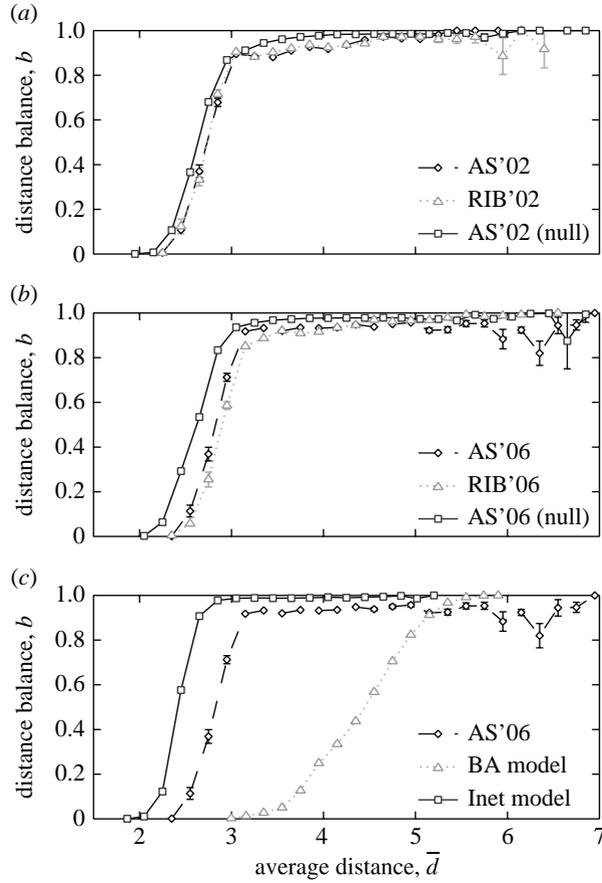
Figure 6. Distance balance b as a function of the average distance $\bar{d}$. The panels and symbols are the same as given in figure legend 1.

Rephrasing this in terms of the average distance, central vertices are likely to have few neighbours with smaller $\bar{d}$ values. This leads us to another view of centrality. Let the *distance balance* of $b(i)$ be the fraction of i-neighbours j with $\bar{d}(j) < \bar{d}(i)$. Clearly, one can expect this to be an increasing function of $\bar{d}$, but is it a linear increase?

In figure 6, we plot the distance balance as a function of $\bar{d}$. As expected, all of the curves generally increase but not linearly. Almost all the increase from 0 to 1 takes place around the highest peak in figure 1, which gives another characterization of the core and periphery: in the core, the typical vertex has relatively few neighbours of higher centrality than itself (and vice versa in the periphery). The $b(i)$ values in the peripheral region of all curves approach values close to 1. In figure 6b, the curves of the observed data are somewhat lower. This supports the previous observation that—as seen in quantities such as degree, neighbour degree and the clustering coefficient—the periphery is structurally less different from the core than what can be expected from random networks constrained to the degree sequence of the observed networks. As seen in figure 1c, the Inet model behaves like the null model—the same observation holds for the

average neighbour degree (figure 3) and clustering coefficient (figure 5). Unlike the Inet model, the BA model's curve increases more smoothly which suggests (in accordance with what has been observed above) a less pronounced core–periphery structure than the observed networks.

4. Summary and conclusions

This paper investigated how vertex-specific network measures of the AS-level Internet vary with the average distance from a vertex to the other vertices of the graph. This projection of vertices to the space of average distances gives a picture of how the network structure changes from the most central to the most peripheral vertices. Using the distance separation measure, we find that there is a well-defined core–periphery dichotomy in the inferred networks. To some extent, this can be explained as an effect of the set of degrees of the network—we note that the average degree as a function of the average distance has the same qualitative form for the observed networks as our null-model networks. However, the periphery is more complex than what is predicted by degree alone. This is manifested in higher average degree, higher average neighbour degree, lower deletion impact, higher clustering coefficient and lower distance balance than the observed networks. To summarize, the AS graph has a more clear split into a core and a periphery than can be anticipated by its degree distribution and simple models of scale-free networks. At the same time, the split is less dramatic and more nuanced than expected from a strict hierarchy. The additional network structure in the periphery may have consequences for spread of attacks and methods to defend against attack. Further, the two topology generators (Inet and BA model) that we tested could be extended to model the periphery more accurately.

We used two kinds of observed AS data—easily accessible router RIBs and more complete datasets where edges missing from the RIBs are added. The effect of the missing edges is clearly visible: the peripheries of the RIB networks (with missing edges) have lower average degree, lower number of triangles and other traits. On the other hand, the missing links do not change the network structure qualitatively. Our conclusions would be unchanged if we used only the RIB data. This suggests that though our datasets are incomplete, the addition of the edges yet missing would not significantly affect the network structure.

Future modelling and measuring research needs to be undertaken to elucidate the detailed structure of the core and periphery of the AS graph. Furthermore, the structures should be related to the strategies of AS management (Winick & Jamin 2000; Chang *et al.* 2002, in press; Daubechies *et al.* 2006).

References

Albert, R. & Barabási, A.-L. 2002 Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**, 47–98. (doi:10.1103/RevModPhys.74.47)

Albert, R., Jeong, H. & Barabási, A.-L. 2000 Attack and error tolerance of complex networks. *Nature* **406**, 378–382. (doi:10.1038/35019019)

Alvarez-Hamelin, J. I. & Schabanel, N. 2004 An Internet graph model based on trade-off optimization. *Eur. Phys. J. B* **38**, 231–237. (doi:10.1140/epjb/e2004-00116-y)

Alvarez-Hamelin, J. I., Dall'Asta, L., Barrat, A. & Vespignani, A. 2006 Large scale networks fingerprinting and visualization using the *k*-core *decomposition*. Advances in neural information processing systems, vol. 18, pp. 41–50. Cambridge, MA: MIT Press.

Barabási, A.-L. & Albert, R. 1999 Emergence of scaling in random networks. *Science* **286**, 509–512. (doi:10.1126/science.286.5439.509)

Buckley, F. & Harary, F. 1989 *Distance in graphs*. Reading, MA: Addison-Wesley.

Chang, H., Govindan, R., Jamin, S., Shenker, S. J. & Willinger, W. 2002 Towards capturing representative AS-level Internet topologies. Technical Report UM-CSE-TR-454-02, Electrical Engineering and Computer Science Department, University of Michigan.

Chang, H., Jamin, S. & Willinger, W. In press. To peer or not to peer: modeling the evolution of the Internet's AS-level topology. In *Proc. of IEEE INFOCOM 2006*.

Daubechies, I., Drakakis, K. & Khovanova, T. 2006 A detailed study of the attachment strategies of new autonomous systems in the AS connectivity graph. *Internet Math.* **2**, 185–246.

Dorogovtsev, S. N. & Mendes, J. F. F. 2003 *Evolution of networks: from biological nets to the Internet and WWW*. Oxford, UK: Oxford University Press.

Doyle, J. C., Alderson, D. L., Li, L., Low, S., Roughan, M., Shalunov, S., Tanaka, R. & Willinger, W. 2005 The "robust yet fragile" nature of the Internet. *Proc. Natl Acad. Sci. USA* **102**, 14 497–14 502. (doi:10.1073/pnas.0501426102)

Fabrikant, A., Koutsoupias, E. & Papadimitriou, C. H. 2002 *Heuristically optimized trade-offs: a new paradigm for power laws in the Internet*. Lecture notes in computer science, vol. 2380, pp. 110–122. Heidelberg, Germany: Springer.

Faloutsos, M., Faloutsos, P. & Faloutsos, C. 1999 On power-law relationships of the Internet topology. *Comput. Commun. Rev.* **29**, 251–262. (doi:10.1145/316194.316229)

Gale, D. 1957 A theorem of flows in networks. *Pac. J. Math.* **7**, 1073–1082.

Holme, P., Kim, B. J., Yoon, C. N. & Han, S. K. 2002 Attack vulnerability of complex networks. *Phys. Rev. E* **65**, 056 109. (doi:10.1103/PhysRevE.65.056109)

Lee, C.-Y. 2006 Correlations among centrality measures in complex networks. e-print physics/0605220.

Mahadevan, P., Krioukov, D., Fomenkov, M., Huffaker, B., Dimitropoulos, X., Claffy, K. C. & Vahdat, A. 2005 Lessons from three views of the Internet topology. Cooperative Association for Internet Data Analysis, tr-2005-02.

Mahadevan, P., Krioukov, D., Fomenkov, M., Huffaker, B., Dimitropoulos, X., Claffy, K. C. & Vahdat, A. 2006 The Internet AS-level topology: three data sources and one definitive metric. *ACM SIGCOMM Comput. Commun. Rev.* **36**, 17–26. (doi:10.1145/1111322.1111328)

Maslov, S., Sneppen, K. & Zaliznyak, A. 2004 Detection of topological patterns in complex networks: correlation profile of the Internet. *Phys. A* **333**, 529–540. (doi:10.1016/j.physa.2003.06.002)

Medina, A., Matta, I. & Byers, J. 2000 On the origin of power laws in Internet topologies. *ACM Comput. Commun. Rev.* **30**, 18–28. (doi:10.1145/505680.505683)

Nakao, K. 1990 Distribution of measures of centrality: enumerated distributions of Freeman's graph centrality measures. *Connections* **13**, 10–22.

Newman, M. E. J. 2001 Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Phys. Rev. E* **64**, 016 132. (doi:10.1103/PhysRevE.64.016132)

Newman, M. E. J. 2003 The structure and function of complex networks. *SIAM Rev.* **45**, 167–256. (doi:10.1137/S003614450342480)

Pastor-Santorras, R. & Vespignani, A. 2004 *Evolution and structure of the Internet: a statistical physics approach*. Cambridge, UK: Cambridge Univeristy Press.

Pastor-Satorras, R., Vázquez, A. & Vespignani, A. 2001 Dynamical and correlation properties of the Internet. *Phys. Rev. Lett.* **87**, 258 701. (doi:10.1103/PhysRevLett.87.258701)

Roberts Jr, J. M. 2000 Simple methods for simulating sociomatrices with given marginal totals. *Soc. Networks* **22**, 273–283. (doi:10.1016/S0378-8733(00)00026-5)

Sabidussi, G. 1966 The centrality index of a graph. *Psychometrika* **31**, 581–603. (doi:10.1007/BF02289527)

Subramanian, L., Agarwal, S., Rexford, J. & Katz, R. H. 2002 Characterizing the Internet hierarchy from multiple vantage point, pp. 618–627. New York, NY: IEEE.

Trusina, A., Maslov, S., Minnhagen, P. & Sneppen, K. 2004 Hierarchy measures in complex networks. *Phys. Rev. Lett.* **92**, 178 702. (doi:10.1103/PhysRevLett.92.178702)

Watts, D. J. & Strogatz, S. H. 1998 Collective dynamics of "small-world" networks. *Nature* **393**, 440–442. (doi:10.1038/30918)

Winick, J. & Jamin, S. 2000 Inet-3.0: Internet topology generator. Technical Report UM-CSE-TR-456-02, Electrical Engineering and Computer Science Department, University of Michigan.