Deciding on the type of the degree distribution of a graph from traceroute-like measurements

Xiaomin Wang[1], Matthieu Latapy, Michèle Soria

firstname. lastname@ lip6. fr
LIP6 – CNRS and UPMC, 4 place Jussieu, 75005 Paris, France

Abstract

The degree distribution of the Internet topology is considered as one of its main properties. However, it is only known through a measurement procedure which gives a biased estimate. This measurement may in first approximation be modeled by a BFS (Breadth-First Search) tree. We explore here our ability to infer the type (Poisson or power-law) of the degree distribution from such a limited knowledge. We design procedures which estimate the degree distribution of a graph from a BFS of it, and show experimentally (on models and real-world data) that this approach succeeds in making the difference between Poisson and power-law degree distributions.

Keywords: Internet, graph, topology, measurement, degree distribution, traceroute, simulations.

1. Introduction

The Internet may be seen as a graph at several levels: autonomous systems, routers and links between them, or IP hops between interfaces for instance. For a decade, these graphs have been at the core of an intense research activity [10, 15, 23, 27] aimed at a better understanding and management of the Internet, which plays a crucial role in our society.

A property of high interest in these graphs is the *degree distribution*, *i.e.* the fraction P_k of nodes with k links, for all k: it may have a strong influence on the robustness of the network [2, 20], on protocol design [21], and on spreading of information [6, 23]. Moreover, it is often claimed that these degree distributions may deviate significantly from what classical models assume [2, 13, 23, 27], which leads to an intense activity on modeling issues [12, 3].

However, the degree distribution of the Internet topology (at any of the levels cited above) is not readily available: one only has access to *samples* of these graphs, obtained through measurement procedures which are intricate, time and resource consuming, and far from complete. Even more important,

these samples are *biased* by the measurement procedure, which may have a strong influence on the observed degree distribution [9, 1, 17, 27].

As a consequence, the current situation regarding the degree distribution of the Internet is unclear [27, 19, 16, 18]. The relevance of obtained samples regarding the degree distribution observed from them is far from being established. In particular, there is a controversy on whether the Internet topology may have a homogeneous (typically Poisson), or heterogeneous (typically power-law) degree distribution [16, 27]. In order to obtain an answer to this question, the most widely used approach currently is to conduct larger and larger measurements, in the expectation that these will lead to accurate observations [18, 24, 8]. However, this may be a dead end: the degree distribution may be intrinsically biased by the measurement process [1, 17] and in practice it may depend much on the sample size [18].

We explore here a completely different approach: we consider a simple model of Internet topology measurements and try to derive the type of the degree distribution of the underlying graph from this limited observation. Our basic goal therefore is to answer the following question: given the limited information obtained from measurement, does the underlying topology more likely have a power-law or a Poisson degree distribution?

In many cases (`traceroute` measurements, BGP tables, and AS-level `traceroute`, typically), the measurement process may be approximated by a BFS (Breadth First Search) tree from a given node of the network. Indeed, Internet measurements mostly consist in sets of *routes* (i.e. paths in the considered topology) going from a monitor to a set of targets, collected from as many monitors as possible. Since each route is modeled as a shortest path and since one may expect routes to have long common prefixes, the view from each monitor may be approximated by a BFS. Although this is a rough approximation, in the lack of a widely accepted and better solution, it has been used in many occasions [17, 22, 1, 25, 9].

Deciding on the degree distribution from such a view is a challenging task, and we will make a few assumptions in order to make a first step towards this ambitious goal. We first assume that the *size* of the graph, *i.e.* its number of nodes n, is given. In the case of the Internet, this is a reasonable assumption [25, 14]. In addition, we will assume that the underlying graph is a random graph with either a Poisson or power-law degree distribution. And finally, we assume that we have a *complete* BFS of the considered graph: all nodes (but not all links) in the graph are reached by the exploration. Finally, we assume that the number of links m of the graph is known. It is clear that the last two assumptions are very strong, and are not attainable in practice. We however consider them as reasonable for a first approximation, and give hints of how to get rid of the knowledge of the number of links in the last section of the paper.

This paper contains six main sections. We describe our methodology in Section 2. It relies on several strategies to infer a degree distribution from a BFS, which we detail in Section 3. In Sections 4 and 5 we experimentally evaluate the validity of our approach on model graphs and real-world graphs respectively. In Section 6, we propose a method to reduce the initial information
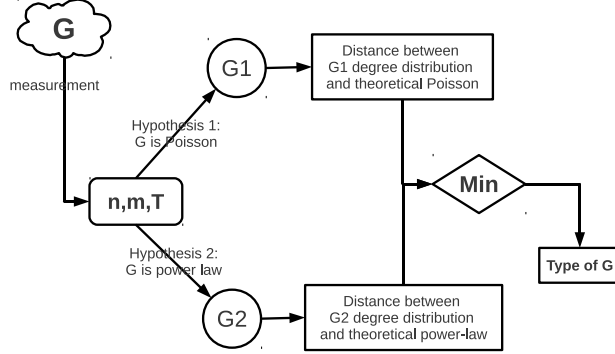
Figure 1: Scheme of our method. G is an unknown graph on which we perform a measurement which gives its number of nodes n, its number of links m and a BFS T. We then consider two different hypotheses: G has a Poisson degree distribution with average degree λ or it has a power-law degree distribution with exponent α. We build two graphs $G1$ and $G2$ each with a strategy in accordance with the corresponding hypothesis. We then compare the degree distribution of $G1$ to the expected one of G if hypothesis 1 is true, and the one of $G2$ to the expected one of G if hypothesis 2 is true. The hypothesis which leads to the most similar degree distributions is expected to be correct.

needed by our approach. We discuss limitations and future work in Section 7.

2. Methodology

Our methodology is sketched in Figure 1. It aims at deciding the type of the degree distribution of an unknown graph G from one of its BFS tree T, its number of nodes n and its number of links m obtained through a measurement. To do so, we consider the two following hypotheses:

- (H1) G has a Poisson degree distribution $P^{(1)}$: $P_k^{(1)} = \frac{\lambda^k e^{-\lambda}}{k!}$, in which $\lambda = \frac{2m}{n}$ is the average degree.

- (H2) G has a power-law degree distribution $P^{(2)}$: $P_k^{(2)} = Ck^{-\alpha}$,where C is a normalizing coefficient and the average degree is $\frac{\sum k^{1-\alpha}}{\sum k^{-\alpha}}$.

For each hypothesis, we build a graph according to a strategy which we detail in Section 3, thus obtaining $G1$ and $G2$, respectively. Our expectation is that if hypothesis $H1$ is true (G is Poisson) then the degree distribution P^{G1} of $G1$ will be closer to the theoretical distribution $P^{(1)}$ than P^{G2} to $P^{(2)}$, and conversely if $H2$ is true (G is power-law) then the degree distribution P^{G2} of $G2$ will be closer to $P^{(2)}$ than P^{G1} to $P^{(1)}$. To compare two distributions, we will use different distances D (we will define several below) and compare $D(P^{(1)}, P^{G1})$ to $D(P^{(2)}, P^{G2})$.

3

We experimentally assess the validity of this approach by applying it to cases where we know the original graph G (we obtain such graphs using models in Section 4 and using real-world data in Section 5). We then compare the expected theoretical degree distribution to the ones of the graphs obtained from each strategy and check conformance of results with expectations. Comparing distributions is a challenge itself, for which no general automatic procedure is commonly accepted. To perform this, we will use two complementary approaches: visual inspection of plots of the inverse cumulative degree distributions (ICDD), together with two classical statistical tests, the Kolmogorov-Smirnov test [11] and Statistical Distance [5] distances, defined as:

Definition 1. Kolmogorov-Smirnov test:
The KS test gives the maximum difference of the cumulative values of two degree distributions P and Q: $KS(P,Q) = max_i \sum_{k=1}^{i} (P_k - Q_k)$.

Definition 2. Statisitical Distance:
The SD gives the sum of the absolute difference between two degree distributions P and Q: $SD(P,Q) = \sum_{k=1} |P_k - Q_k|$.

In Section 6, we extend our approach to get information on the degree distribution of a graph when the number of links is not known, which is a more realistic assumption. In that case we use our rebuilding strategies for a wide range of possible values of m and infer the most probable type of degree distribution as previously, see Figure 2.

3. Building strategies

Starting from a BFS T of a graph G with n nodes and m links and an hypothesis (H1 or H2) on the degree distribution of G (type Poisson or power-law), our objective here is to iteratively add $m - n + 1$ links to T in order to build a graph G' with n nodes, m links, and degree distribution similar to the one of G. We define different link addition strategies according to the supposed type of G, Poisson or power-law. And in each case, we show how to compute the degree distribution of the resulting graph.

3.1. H1: G is Poisson (RR strategy)

Suppose G is a Poisson random graph. It may therefore be seen as the result of an Erdös Rényi (ER) construction [7]: starting with n nodes and no link, m links are uniformly chosen among the $\frac{n(n-1)}{2}$ possible pairs of nodes. The expected node degree distribution obtained by this way follows a Poisson law: $P_k = \frac{\lambda^k e^{-\lambda}}{k!}$ where $\lambda = \frac{2m}{n}$ is the average degree.

We may think of building a graph G' similar to G by using a variant of the ER construction: starting with the n nodes and $n - 1$ links of a BFS tree T, the $m - n + 1$ missing links are randomly added as in the ER model. But then T may not be a possible BFS of G': any link in G which is not in T is necessarily between two nodes in consecutive levels of T, or in the same level
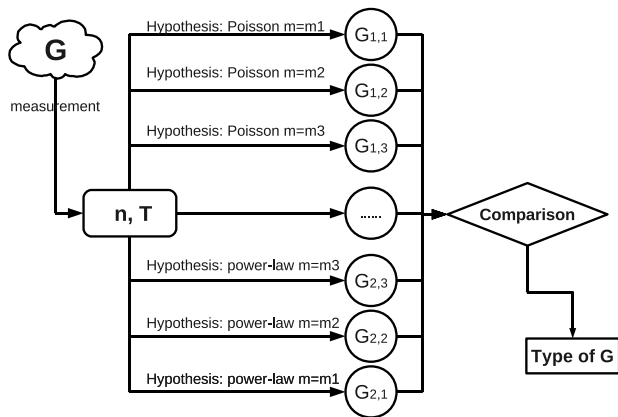
4

Figure 2: Scheme of our method, in the case when m is unknown. G is an unknown graph on which we perform a measurement which gives its number of nodes n and a BFS tree T. We then consider two different type of hypotheses: (H1) G has a Poisson degree distribution with average degree λ or (H2) it has a power-law degree distribution with exponent α. Both parameters λ and α are equivalent to the number of links m. Then we build two families of graphs $G_{1,m}$ and $G_{2,m}$ in accordance with these two hypothesis and different m. We then compare the degree distribution of $G_{1,m}$ to the expected one of G if (H1) is true, and the one of $G_{2,m}$ to the expected one of G if (H2) is true. The hypothesis which leads to the most similar degree distributions is expected to be correct.
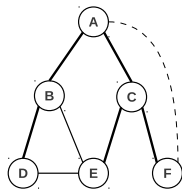

Figure 3: Let us consider a part of a BFS of G, composed of a set of links A-B, A-c, B-D, C-E and C-F. Links B-E and D-E may also be present in G, but not link A-F.

of T (otherwise T would not be a shortest path tree and thus not a BFS, see Figure 3). In order to ensure that T is also a possible BFS of G' we therefore add links only between nodes in consecutive levels or in the same level. Since both extremities of links are randomly chosen, we call this construction the RR (Random-Random) strategy.

We now show that the expected node degree distribution of G' obtained with the RR strategy can be directly computed from n, m and T without explicitly constructing G'.

Let n_j denote the number of nodes at level j in the BFS tree T. For each node v at level j with degree k (one father node and $k-1$ son nodes), three kinds of links may be added: (1) to upper level $j-1$, there are $n_{j-1}-1$ possible positions, all nodes in the upper level except the father of v; (2) to the same level of v, there are $n_j - 1$ possible positions, all nodes in this level except v itself; (3) to lower level $j+1$, there are $n_{j+1}-k+1$ possible positions, all nodes at lower level except the $k-1$ sons of v.

Let S_{jk} be the number of possible positions linked to a node at level j having k neighbors: $S_{jk} = n_{j-1}+n_j+n_{j+1}-k-1$. On the other hand, let S be the total number of possible positions in a BFS tree T: $S = \sum_j \left(n_j n_{j+1} - n_{j+1} + \frac{n_j(n_j-1)}{2} \right)$. Therefore, if we make a random choice, the probability that this position links to a given node v (at level j with degree k) is $\frac{S_{jk}}{S}$. We finally obtain the following theorem.

Theorem 1. *Given a tree T with n nodes, if we construct a graph G' using the RR strategy, then the expectation of a node v with degree l in G' is:*

$$E\left(d_{G'}(v) = l\right) = \frac{1}{n} \sum_{j>0} \sum_{k=1}^{l} n_{jk} P\left(k \rightarrow l, j\right) \tag{1}$$

where n_{jk} is the number of nodes with degree k at level j in T and $P\left(k \rightarrow l, j\right) = \binom{m-n+1}{l-k} \left(\frac{S_{jk}}{S}\right)^{l-k} \left(1 - \frac{S_{jk}}{S}\right)^{m-n+1-(l-k)}$ is the probability that a node v with degree k at level j becomes a node of degree l in G' after $m-n+1$ links have been added.

As each newly added link impacts two nodes, the values of $P(k \rightarrow l, j)$ are independent from the other $P(k' \rightarrow l', j')$. We will consider that this impact can be ignored, when n and m is large enough. From this result, one may estimate the expectation of the degree distribution of G' from n, m and T, without constructing it explicitly. This is of high interest in practice, since it allows to compute the expectation of degree distribution and compare it with degree distributions obtained by the constructing strategies. Going further would however need precise results on the expectation of degrees in T, which is a difficult problem [1].

3.2. H2: G is power-law (PP strategies)

Suppose now that G is a power-law graph. We therefore aim at designing a process which builds from a BFS T of G, a graph G' with power-law node degree distribution. To do this, as before, we add $m - n + 1$ links between nodes in appropriate levels of T. However, these pairs of nodes are no longer chosen uniformly at random. Instead, we use a selection scheme inspired from the preferential attachment of the classical Barabási-Albert model [4]: we choose (in the appropriate levels) nodes randomly with probability proportional to their degree in T. As we choose both extremities of added links according to preferential attachment, we call this procedure the PP (Preferential-Preferential) strategies.

We now show how to compute, for this strategy, the expected obtained degree distribution.

Theorem 2. *Given a tree T with n nodes, if we construct a graph G' using the PP strategy, then the expectation of a node v with degree l in G' is:*

$$E\left(d_{G'}(v) = l\right) = \frac{1}{n} \sum_{j>0} \sum_{k=1}^{l} n_{jk} P(k \to l, j, m') \tag{2}$$

where n_{jk} is the number of nodes with degree k at level j in T and $P\left(k \to l, j, m'\right)$ is the probability that a node v with degree k at level j in T is constructed as a node with degree l in G' after m' links have been added into T.

Computation process.. The term $P(k \to l, j, m')$ may be obtained recursively:

$$
\begin{aligned}
P\left(k \to l, j, m'\right) = &\left(1 - \theta\right) P\left(k \to l, j, m' - 1\right) \\
&+ \theta\, P\left(k + 1 \to l, j, m' - 1\right)
\end{aligned}
\tag{3}
$$

We split $P(k \to l, j, m')$ into two parts which correspond to two cases (linked to this node or not) when a new link is added. The probability that a newly added link changes the node from degree k to $k + 1$ is denoted by θ.

In the following, we compute θ using four terms:

- $\theta 1$: the probability that the node v is selected as the first endpoint of the newly added link.

- $\theta 2$: the probability that a node at upper level $j - 1$ is selected as the first endpoint, and v is selected as the second endpoint.

- $\theta 3$: the probability that a node (except v) at level j is selected as the first endpoint, and v is selected as the second endpoint.

- $\theta 4$: the probability that a node at lower level $j + 1$ is selected as the first endpoint, and v is selected as the second endpoint.

The sum of degrees of all nodes at the level i after t links have been added is $d_{i,t}$. In the following, we give the details of how to calculate for the case of PP strategy.

$$\theta 1 = \frac{k}{2\left(n-1+t\right)} \tag{4}$$

$$\theta 2 = \frac{d_{j-1,t}}{2\left(n-1+t\right)} \frac{k}{d_{j-2,t} + d_{j-1,t} + d_{j,t}} \tag{5}$$

$$\theta 3 = \frac{d_{j,t} - k}{2\left(n-1+t\right)} \frac{k}{d_{j+1,t} + d_{j,t} + d_{j-1,t}} \tag{6}$$

$$\theta 4 = \frac{d_{j+1,t}}{2\left(n-1+t\right)} \frac{k}{d_{j+2,t} + d_{j+1,t} + d_{j,t}} \tag{7}$$

The next problem is how to compute $d_{j,t}$. We denote by $K_{j \to j'}$ the probability that a directed link is added from level j and j', where j' must be j, $j-1$, $j+1$.

$$d_{j,t} = K_{j \to j}\left(d_{j,t-1} + 2\right) + \left(K_{j\pm1 \to j} + K_{j \to j\pm1}\right)\left(d_{j,t-1} + 1\right)$$
$$+ \left(1 - K_{j \to j} - K_{j \to j\pm1} - K_{j\pm1 \to j}\right) d_{j,t-1} \tag{8}$$

$$K_{j \to j'} = \frac{d_{j,t-1}}{2\left(n-1+t\right)} \frac{d_{j',t-1}}{d_{j-1,t-1} + d_{j,t-1} + d_{j+1,t-1}} \tag{9}$$

Using equation (9) in the expressions (8) shows that $d_{j,t}$ is a function of $d_{j',t-1}$, which can be calculated by dynamic programming techniques.

Our computation process is not exact, since we have neglected to take into account possible collision, i.e. positions to be selected several times. However, since we deal with sparse graphs, in the case of model graphs as well as real graphs, the number of links to be added to the BFS tree is much smaller than the number of possible positions, and so there are very few collisions. From a practical viewpoint, in the building process, we just ignore multiple links.

4. Validation using model graphs

Our expectation is that the strategies described in previous section succeed in building a graph G' similar (regarding degree distribution) to G when the appropriate strategy is used with an appropriate graph (RR if G is Poisson, PP if G is power-law). In addition, we expect that the degree distribution of G' will differ significantly from that of G if a wrong strategy is applied (RR if G is power-law, PP if G is Poisson). In this section we conduct experiments on *model graphs*, *i.e.* random graphs in the classes of Poisson graphs or power-law graphs with given parameters (average degree and exponent respectively). To ensure that the BFS covers all nodes of the graph, we use a software which generates *random simple connected graphs* according to a given degree sequence (sampled from the given degree distribution) [26].
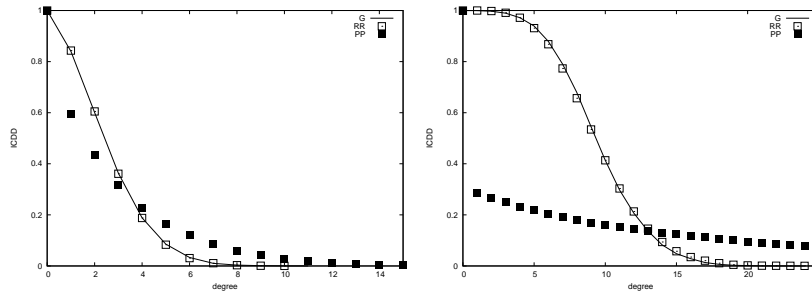
Figure 4: Reconstruction of a graph Poisson 3 and Poisson 10. We draw the ICDD (Inverse Cumulative Degree Distribution) for three graphs: the original graph G and the ones obtained with the RR and the PP strategies. Horizontal axis: degree k; vertical axis: fraction of the number of nodes with degree lower than or equal to k.

For each model graph G, we first extract a BFS tree T from a randomly chosen node, then build G'_{RR} (using RR strategy) and G'_{PP} (using PP strategy), and we denote by P^{RR} and P^{PP}, the corresponding degree distributions. According to the method described in Figure 1, we suppose that we know n and m. Therefore, the corresponding parameters λ (for RR strategy), α (for PP strategy) and the corresponding theoretical degree distributions P^{λ} and P^{α} can be derived from n and m. By comparing the values $KS(P^{RR}, P^{\lambda})$ and $KS(P^{PP}, P^{\alpha})$ (or $SD(P^{RR}, P^{\lambda})$ and $SD(P^{PP}, P^{\alpha})$), we may then decide on the type of the underlying graph. If the concluded type is the same as the type that we have used to generate the underlying graph, and so is the same as the actual one, then our method has succeed.

We conducted experiments on Poisson model graphs with average degree from 3 to 10 (step is 0.1), and power-law model graphs with exponents from 2.1 to 2.5 (step is 0.01), which are typical values used in Internet modeling. In each case, we consider graphs with 10000 and 100000 nodes, and all results in tables and figures are averaged over ten samples. We display below the results for average degrees 3 and 10 and for exponents 2.1 and 2.3, which are representative of our observations.

The general conclusion of our experiments is that for both cases of Poisson and power-law, we succeed in deciding the type of the underlying graphs: for Poisson model graphs, the type decided by our method is actually Poisson and for power-law graphs, the type decided is actually power-law.

4.1. Poisson graphs

In Figure 4 we present the results for Poisson graphs with average degrees 3 and 10.

The degree distribution obtained with RR strategy is closer to the original one, as expected. This is confirmed by the KS and SD statistics (Table 1): the smallest values are obtained with RR strategy.

Notice that a Poisson graph with a higher degree gives better results. This is probably due to the fact that we add more links in this case, and so strategies

9

Table 1: KS and SD for Poisson model graphs. The smallest KS or SD values mean the smallest difference on degree distribution.

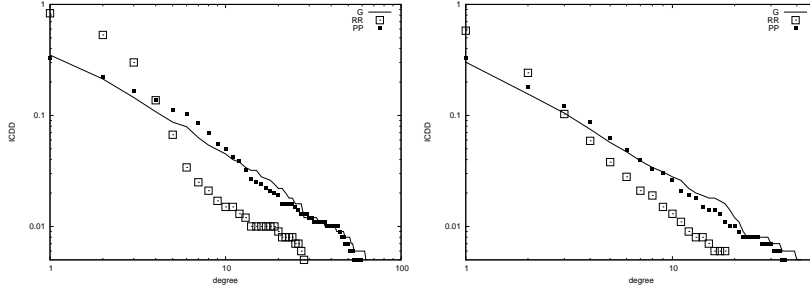|	Poisson 3					Poisson 10			
	$n = 10000$		$n = 100000$			$n = 10000$		$n = 100000$	
	KS	SD	KS	SD		KS	SD	KS	SD
RR	0.043	0.091	0.037	0.075	RR	0.025	0.116	0.014	0.048
PP	0.284	0.610	0.268	0.571	PP	0.119	0.444	0.111	0.446

Figure 5: Reconstruction of a graph power-law 2.1 and 2.3. We draw the ICDD for three graphs: the original graph G, and the ones obtained with the RR and the PP strategies.

for doing this make much more difference.

Finally, we conclude that our method succeeds in recognizing random Poisson graphs. This is true for all average degrees, but performs best on graphs with a relatively high average degree.

4.2. Power-law graphs

Similar to Poisson model graphs, we conduct our experiments with power-law model graphs.

In Figure 5 we present obtained results for power-law graphs with exponent 2.1 and 2.3. To better show the characteristic of the power-law, all plots are in log-log scale. Both the ICDD plot and the statistic test (Table 2) support our conclusion for all exponents and all sizes.

Table 2: KS and SD for power-law model graphs

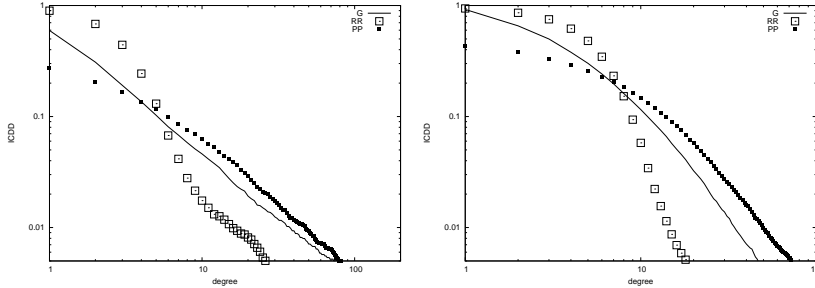|	Power-law 2.1					Power-law 2.3			
	$n = 10000$		$n = 100000$			$n = 10000$		$n = 100000$	
	KS	SD	KS	SD		KS	SD	KS	SD
RR	0.201	0.432	0.194	0.405	RR	0.278	0.591	0.274	0.553
PP	0.038	0.138	0.049	0.180	PP	0.030	0.086	0.024	0.095

10

Figure 6: ICDD of the Skitter and Radar graphs G and the ones obtained with RR and PP strategies.

Table 3: KS and SD for Skitter graph.

	KS	SD
theory-RR	0.166	0.359
theory-PP	0.082	0.235

5. Experiments on real-world data

Previous section shows that our method succeeds in making the difference between Poisson and power-law *random graphs*. It is clear however that, in practice, considered graphs have neither perfect Poisson nor power-law degree distribution, and are not random.

We consider in this section several real-world datasets among the current largest measurements of the Internet topology. Although obtained graphs are still partial views and probably are strongly biased, they constitute current state-of-the-art of available data and we use them as benchmarks.

Like in previous section, for each case of real-world graph G, we consider two hypotheses: (H1) G has a degree distribution close to a Poisson law; (H2) G has a degree distribution close to a power-law. Using strategies RR and PP, we build graphs G'_{RR} and G'_{PP}, respectively. Our expectation is that if (H1) is true then G'_{RR} degree distribution is close to the theoretical Poisson law; if (H2) is true then G'_{PP} degree distribution is close to the theoretical power-law; and the converse is not true.

5.1. Skitter graph

We first try our method on an AS-level map collected by the Skitter project of CAIDA [15]. The obtained graph has 5776 nodes and 12025 links.

Figure 6 (left) shows the ICDD obtained with our strategies and shows that the entire Skitter graph follows a degree distribution of type power-law (a perfect power-law in a log-log scale is a straight line). The distribution obtained with RR strategy is clearly far from a power-law. The one obtained with PP strategy is much closer. Table 3 confirms this, even though the difference between the

Table 4: KS and SD for Radar graphs.

	KS	SD		KS	SD		KS	SD
	Cm			Japon			Ortolan	
RR	0.064	0.151	RR	0.074	0.202	RR	0.062	0.156
PP	0.241	0.521	PP	0.163	0.363	PP	0.213	0.476

RR strategy and the PP strategy is as strong as for model graphs (see Table 1 and 2).

Finally, our method succeeds in deciding that the Skitter graph has a degree distribution close to a power-law.

5.2. Radar graphs

A radar graph is a part of the Internet topology observed by periodic running traceroute-like measurements from one monitor to a set of targets during several weeks, See [22] for details and the original data. We use here three instances of radar graphs, from three different monitors: Cm (21430 nodes and 37938 links), Japon (27066 nodes and 77545 links) and Ortolan (24725 nodes and 48516 links).

In Figure 6 (right), we plot the ICDD of Radar-Japon graph and those obtained by the RR and the PP strategies. The shape of the original ICDD indicates that the degree distribution of the underlying graph is likely to be a mixture of both Poisson and power-law.

Table 4 shows numerical results for the three radar graphs. All results show that the difference between the RR strategy and its corresponding theoretical distribution is smaller than that of the PP strategy. Therefore our method decide that the type of the radar graphs is more likely Poisson even though their distributions have a long-tail. Note that the difference between RR and PP are much smaller than for previous cases, thus indicating that the confidence in the conclusion is poor.

Finally, our method decides that Radar graphs are more likely to have Poisson degree distribution, but with poor confidence (which is in accordance with their actual type).

6. Deciding without m

The goal of this section is to decide the type of degree distribution and estimate m for an unknown graph G, using only n and one of its complete BFS tree T. We proceed as follows: for each building strategy, we compute the KS distance between the obtained distribution and the theoretical one for a wide range of realistic values of m: for each value of m we compare with KS the distance between the theoretical distribution corresponding to m and the distribution of the graph resulting from the building strategy. We then plot this statistics and select the value m' which gives the minimum KS value. If

Table 5: Results for Poisson model graphs

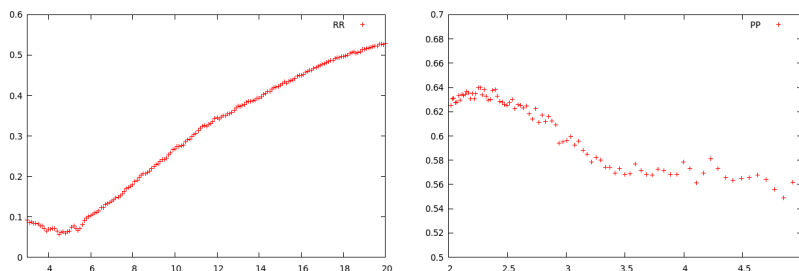| n=1000	KS	m	m'	n=10000	KS	m	m'
Poisson 3	0.068	1500	2085	Poisson 3	0.027	15000	23750
Poisson 5	0.074	2500	2835	Poisson 5	0.025	25000	36250
Poisson 10	0.096	5000	5650	Poisson 10	0.041	50000	68750

Figure 7: KS obtained for a Poisson model graph with average degree 3. In both figures, we plot KS as a function of the estimated degree. In the left figure we use the RR strategy, which shows minimum KS of 0.05 in the region of $\lambda = 4$. In the right figure we use the PP strategy , which shows much bigger KS's. Thus we conclude to a Poisson graph.

the minimum value is given by the RR strategy, we conclude that the degree distribution of the original graph is Poisson, whereas if m' is given by the PP strategy, we conclude that the degree distribution is power-law (Figure 2). The RR strategy is conducted with average degree from 2 to 50, with step 0.1. The PP strategy is conducted with exponent from 2.05 to 2.40, with step 0.05.

In the following we only mention the results of KS tests, but the result with SD are similar.

6.1. Poisson graphs

We show how the obtained KS test evolves as a function of the value of $\lambda = \frac{2m}{n}$ in Figure 7. The RR strategy performs much better than the PP one, and the KS reaches a minimal for a value m' close to the actual value of m.

Table 5 shows the results in a different way, for Poisson model graphs of size 1000 and 10000. First since the minimum KS is always observed in the case of RR (and not PP) strategy, we conclude that the type of graph is Poisson. Second, according to the minimum value of KS we can estimate the value of λ and thus the value of the (unknown) number of links of the original graph.

For example for $n = 1000$ in Table 5 shows that the minimal value for KS is obtained for $m' = 2085$ (whereas the real value is $m = 1500$, corresponding to a Poisson model graph with $\lambda = 3$).

Notice that the estimate m' of m is always a little overestimated and this becomes worse when the size of the graph grows. We have no explanation for this observation currently.

Table 6: Results for power-law model graphs

| n=1000	KS	m	m'	n=10000	KS	m	m'
power-law 2.10	0.07	4233	3002	power-law 2.10	0.06	42326	28129
power-law 2.15	0.09	3662	2936	power-law 2.15	0.04	36622	25043
power-law 2.20	0.07	3220	2550	power-law 2.20	0.03	32198	25043
power-law 2.25	0.07	2873	2418	power-law 2.25	0.03	28730	21085
power-law 2.30	0.06	2598	2168	power-law 2.30	0.03	25982	21377
power-law 2.35	0.05	2378	2200	power-law 2.35	0.02	23780	20534
power-law 2.40	0.06	2200	2109	power-law 2.40	0.02	21996	20274
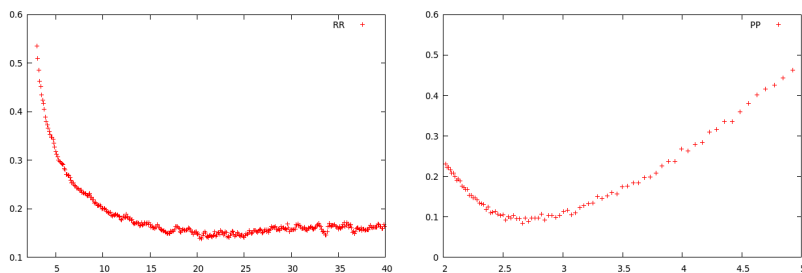
Figure 8: KS for a power-law model graph with exponent 2.2. The left figure is the estimation with the RR strategy with a minimum greater than 0.1. The right figure is the estimation with the PP strategy with a minimum less than 0.1

6.2. Power-law graphs

In Table 6 and Figure 8, we give the result of our experiments for power-law graphs with exponents between 2.10 and 2.40.

Again, the method succeeds in deciding the appropriate type of the graph: with PP, the value of the KS test is smaller than with RR. The estimate m' is reasonable, but lower than the actual value.

6.3. Real-world graphs

Finally, we apply our method on some real world graph, see Table 7 and Figure 9.

Table 7: Results for real-world graphs

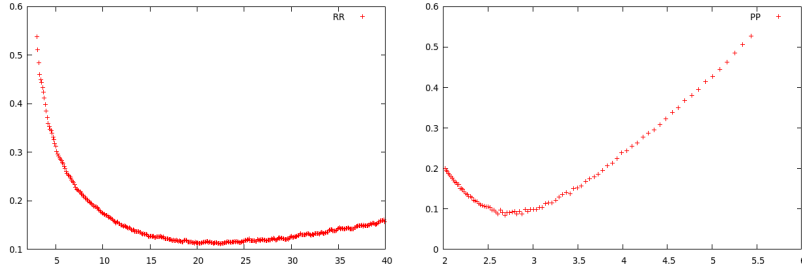|	RR	PP	m	m'
Skitter	0.109	0.091	12822	7683
Radar-Orlanto	0.150	0.435	48516	49450
Radar-Japon	0.067	0.162	77545	79041
Radar-Cm	0.051	0.225	37938	49289

Figure 9: KS for Skitter-AS graph.

As before, our method succeeds in deciding that the Skitter graph is very close to a power-law graph, while radar graphs are closer to a Poisson graph, but are actually in between, the both cases, we obtain reasonable estimates m' of the number of links m.

7. Conclusion

In this paper, we presented a new approach to decide on the type of the degree distribution of a graph when only a BFS and its size are known. We use different strategies to reconstruct a graph from a BFS, according to the presupposed type of the graph. We then show how these strategies allow to decide between Poisson and power-law degree distributed graphs, with experiments on random graphs as well as real-world graphs.

We worked on two sets of hypotheses: we first supposed that, in addition to a BFS T on the n vertices of the unknown original graph G, we know its number of edges m: in this case we formally compute the expected degree distribution of the graphs obtained with the different reconstructing strategies; moreover experiments on random graphs confirm that the knowledge of a BFS and m allows to recognize the type of the graph. Second, we use the same reconstruction strategies without knowing the number of edges m, and we can still reasonably predict the type of the original graph.

The main limitation of our contribution is that we suppose the knowledge of a complete BFS of the unknown graph. This hypothesis is not realistic, though, since practical measurements rather provide only paths to a subset of the real graph vertices. Our main perspective therefore is to reduce requirements on data, and design strategies needing truncated BFS (BFS until a certain level) only, and partial BFS (that do not contain all vertices of the graph). Assuming that such measurements are available is much more realistic[22].

In future work, we also want to improve our method by investigating various refinements: (i) design reconstruction strategies taking into account more subtle properties of the BFS, such as node distribution on each level, and also other local properties such as the clustering coefficient; (ii) extend the reconstruction strategies in the direction of other types of distributions: a first step would be

15

to mix RR and PP strategies for mixed Poisson power-law graphs, but we also aim at investigating power-laws with exponential cut-off, and other laws.

Acknowledgements

References

[1] D. Achlioptas, D. Kempe, A. Clauset, and C. Moore. On the bias of traceroute sampling: or, power-law degree distributions in regular graphs. *ACM Symposium on Theory of Computing*, pages 694–703, 2005.

[2] R. Albert, H. Jeong, and A.-L. Barabasi. Error and attack tolerance of complex networks. *Nature*, 406(6794):378–382, July 2000.

[3] D. Alderson, L. Li, W. Willinger, and J. C. Doyle. Understanding internet topology: principles, models, and validation. *IEEE/ACM Transactions on Networking*, 13(6):1205–1218, December 2005.

[4] A. L. Barabási, R. Albert, and H. Jeong. Mean-field theory for scale-free random networks. *Physica A: Statistical Mechanics and its Applications*, 272(1-2):173–187, October 1999.

[5] M. Basseville. Distance measures for signal processing and pattern recognition. *Signal Process.*, 18:349–369, December 1989.

[6] N. Berger, C. Borgs, J. T. Chayes, and A. Saberi. On the spread of viruses on the internet. *Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 301–310, 2005.

[7] B. Bollobas. *Random Graphs*. Cambridge University Press, January 2001.

[8] K. Claffy, Y. Hyun, K. Keys, M. Fomenkov, and D. Krioukov. Internet Mapping: From Art to Science. *Cybersecurity Applications and Technology Conference for Homeland Security*, pages 205–211, March 2009.

[9] L. Dall'asta, I. Alvarez-Hamelin, A. Barrat, A. Vazquez, and A. Vespignani. Exploring networks with traceroute-like probes: Theory and simulations. *Theoretical Computer Science*, 355(1):6–24, April 2006.

[10] B. Donnet and T. Friedman. Internet topology discovery: A survey. *Communications Surveys & Tutorials, IEEE*, 9(4):56–69, 2007.

[11] W. T. Eadie, D. Drijard, F. E. James, M. Roos, and B. Sadoulet. *Statistical methods in experimental physics*. American Elsevier Publishing Co, June 2008.

[12] A. Fabrikant, E. Koutsoupias, and C. Papadimitriou. Heuristically Optimized Trade-Offs: A New Paradigm for Power Laws in the Internet. *Automata, Languages and Programming*, 2380:781, January 2002.

[13] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the Internet topology. *SIGCOMM: Conference on Applications, technologies, architectures, and protocols for computer communication*, 29(4):251–262, October 1999.

[14] J. Heidemann, Y. Pradkin, R. Govindan, C. Papadopoulos, G. Bartlett, and J. Bannister. Census and Survey of the Visible Internet. *8th ACM SIGCOMM conference on Internet measurement*, pages 169–182, 2008.

[15] D. Krioukov, F. Chung, K. Claffy, M. Fomenkov, A. Vespignani, and W. Willinger. The Workshop on Internet Topology (WIT) Report. *SIGCOMM Computer Communication Review*, pages 69–73, Dec 2007.

[16] B. Krishnamurthy and W. Willinger. What are our standards for validation of measurement-based networking research? *ACM SIGMETRICS Performance Evaluation Review*, 36:64–69, August 2008.

[17] A. Lakhina, J. W. Byers, M. Crovella, and P. Xie. Sampling biases in IP topology measurements. *INFOCOM: 22th Conference on Computer and Communications Societies*, 1:332–341, 2003.

[18] M. Latapy and C. Magnien. Complex Network Measurements: Estimating the Relevance of Observed Properties. *INFOCOM: 27th Conference on Computer Communications*, pages 1660–1668, May 2008.

[19] L. Li, D. Alderson, W. Willinger, and J. Doyle. A first-principles approach to understanding the internet's router-level topology. *SIGCOMM: Conference on Applications, technologies, architectures, and protocols for computer communications*, 34(4):3–14, October 2004.

[20] C. Magnien, M. Latapy, and J.-L. Guillaume. Impact of Random Failures and Attacks on Poisson and Power-Law Random Networks. *ACM Computing Surveys*, Aug 2009.

[21] D. Magoni and J.-J. Pansiot. Influence of Network Topology on Protocol Simulation. *International Conference on Networking*, 2093:762–770, June 2001.

[22] F. Ouédraogo, C. Magnien, and M. Latapy. A radar for the internet. *1st International Workshop on Analysis of Dynamic Networks*, 2008.

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

[24] Y. Shavitt and E. Shir. DIMES: let the internet measure itself. *SIGCOMM Computer Communication Review*, 35(5):71–74, 2005.

[25] F. Viger, A. Barrat, L. Dall'Asta, C. H. Zhang, and E. D. Kolaczyk. What is the real size of a sampled network? The case of the Internet. *Physical Review E*, 75(5):056111+, May 2007.

[26] F. Viger and M. Latapy. Efficient and simple generation of random simple connected graphs with prescribed degree sequence. *11th International Computing and Combinatorics Conference*, 3595:440–449, 2005.

[27] W. Willinger, D. Alderson, and J. C. Doyle. Mathematics and the Internet: A Source of Enormous Confusion and Great Potential. *Notices of the American Mathematical Society*, 56, May 2009.