Compact Routing on Internet-Like Graphs

Dmitri Krioukov
Email: dima@krioukov.net

Kevin Fall
Intel Research, Berkeley
Email: kfall@intel-research.net

Xiaowei Yang
Massachusetts Institute of Technology
Email: yxw@mit.edu

***Abstract*— The Thorup-Zwick (TZ) compact routing scheme is the first generic stretch-3 routing scheme delivering a nearly optimal per-node memory upper bound. Using both direct analysis and simulation, we derive the stretch distribution of this routing scheme on Internet-like inter-domain topologies. By investigating the TZ scheme on random graphs with power-law node degree distributions, $P_k \simeq k^{-\gamma}$, we find that the average TZ stretch is quite low and virtually independent of γ. In particular, for the Internet inter-domain graph with $\gamma \simeq 2.1$, the average TZ stretch is around 1.1, with up to 70% of all pairwise paths being stretch-1 (shortest possible). As the network grows, the average stretch slowly decreases. We find routing table sizes to be very small (around 50 records for 10^4-node networks), well below their theoretical upper bounds. Furthermore, we find that both the average shortest path length (i.e. distance) $\bar{d}$ and width of the distance distribution σ observed in the real Internet inter-AS graph have values that are very close to the minimums of the average stretch in the $\bar{d}$- and σ-directions. This leads us to the discovery of a unique critical point of the average TZ stretch as a function of $\bar{d}$ and σ. The Internet's distance distribution is located in a close neighborhood of this point. This is remarkable given the fact that the Internet inter-domain topology has evolved without any direct attention paid to properties of the stretch distribution. It suggests the average stretch function may be an indirect indicator of the optimization criteria influencing the Internet's inter-domain topology evolution.**

***Index Terms*— Routing, Internet Topology, Simulations, Graph Theory, Combinatorics, Statistics.**

I. INTRODUCTION

The question as to what drives the evolutionary process of the Internet's topology is of interest to many researchers. While various models of its topological structure appear to *describe* it reasonably well, most neither aid in understanding *why* the Internet's graph has evolved as it has, nor offer the "metric" which is effectively being optimized by its implementers as it grows.[1] In addition, as the network grows, its global routing scalability is being stressed [2], leading several groups to explore alternatives to the present Internet routing system. We believe that a better understanding of the Internet's topological growth process, coupled with knowledge of the theoretical underpinnings of the routing problem on graphs, could help in evaluating these proposals (or developing others). In particular, we are interested in the performance of the most scalable theoretical routing algorithms on realistic topology graphs.

To further investigate this relationship, we focus on the performance of *compact routing schemes* on *scale-free* graphs. Compact routing schemes comprise a set of algorithms that aim to make a good trade-off between *stretch* versus the amount of storage required at each vertex for routing tables. Stretch refers to the (usually worst-case) multiplicative factor increase of path length between a pair of vertices under a particular routing scheme versus the length of the shortest existing path between the same pair. The most efficient stretch-3 routing scheme for generic (arbitrary) graphs currently known is due to Thorup-Zwick [3], which we will simply refer to as "the TZ scheme." It is known to be optimal, up to a logarithmic factor, for per-node memory utilization.

We investigate the performance of the TZ scheme on scale-free graphs because these graphs represent our best current understanding of the Internet's inter-AS topological structure. It is worth mentioning that although we base our analysis on properties of the real-world Internet, we are not suggesting that the TZ scheme is ripe for use within the Internet to solve its scalability problems. Rather, we employ the TZ scheme as a tool to analyze the fundamental limits for average stretch and routing table sizes on realistic graphs. The scheme is *generic*, so that it can be directly applied to any graphs—to scale-free graphs, in particular. As we are unaware of any routing schemes developed specifically for scale-free graphs, we must turn to generic schemes to pursue our investigation.

One might expect that for scale-free graphs, the majority of known generic routing schemes would be very inefficient. Indeed, many routing schemes (including the TZ scheme) incorporate *locality* by carefully differentiating between close and remote nodes. This approach can make routing more efficient (in the stretch-versus-space trade-off sense, in particular) by keeping only approximate (non-shortest path) routing information for remote nodes, while full (shortest path) routing information is kept for local nodes. In scale-free graphs, with very low average distances and distance distribution widths, local nodes comprise huge portions of all the nodes in a network, so that one might suspect that locality-sensitive approaches might break for such networks. For a good example demonstrating that this might be quite plausible, see the Appendix, where the stretch factor is found to be very high for the Kleinrock-Kamoun (KK) routing scheme [4] applied to the scale-free networks. The choice of the KK scheme for such analysis is partially driven by the fact that many relatively recent "routing architecture" proposals [5], [6], aimed at resolving the Internet routing scalability issues, have been based on the ideas of [4].

[1] Several authors are currently pursuing such models, however. For one of the latest examples, see [1].

0-7803-8356-7/04/$20.00 (C) 2004 IEEE

In analyzing the TZ scheme's performance for the Internet graph, however, we find that the situation is quite opposite with respect to the KK schemes: the TZ scheme produces extremely low average stretch values and succinct routing tables that turn out to be well below their theoretical upper bounds.

A. Routing Background

The aim of compact routing schemes is to approach the optimal stretch-1 (shortest-path) routing but with significantly reduced memory requirements. It is shown in [7] that for *any* stretch-1 routing scheme, there exists a graph of size n and maximum node degree d, $3 \leqslant d < n$, such that $\Omega(n \log d)$ bits of memory are required at $\Theta(n)$ nodes. Since the trivial upper bound for shortest path full-table routing is also $O(n \log d)$, this result effectively demonstrates *incompressibility* of generic shortest path routing. That is to say, if we must accommodate both graphs of arbitrary topology and with all paths being stretch-1, we must be willing to have large routing tables. Although there are some results, [8], showing that the majority of graphs are "better," very little can be said conclusively regarding the practical implications of these results with respect to real-world graphs. Thus, in order to study compact routing on the Internet's graph, we must turn to the only existing tool we have for analyzing compact routing performance in all cases: *generic* routing schemes. Generic shortest path routing is incompressible, so if memory space is to be reduced, then the stretch must be increased.

The memory space lower bound dependence on stretch is not "continuous." As shown in [7], any generic routing scheme with stretch strictly less than 1.4 must use at least $\Omega(n \log n)$ bits of memory on some nodes of some graphs. In other words, the lower bound for generic schemes with stretch $s < 1.4$ is the same as in the incompressible case of shortest path routing discussed above (if one considers graphs with $d = \Theta(n)$). Furthermore, as shown in [9], the lower bound for schemes with stretch strictly less than 3 is nearly the same as for shortest path routing—$\Omega(n)$ bits of memory on some nodes of some graphs. The minimum stretch factor that we must be prepared to accept in order to significantly decrease memory requirements below the incompressible limits is therefore 3.

Cowen introduces a simple stretch-3 routing scheme with a local memory space upper bound of $O(n^{2/3} \log^{4/3} n)$ in [10]. Thorup and Zwick improve upon Cowen's result and deliver a per-node space upper bound of $O(n^{1/2} \log^{1/2} n)$ in [3].[2] We call these two schemes the *Cowen* and *TZ schemes*, respectively. The local memory space upper bound provided by the TZ scheme is nearly optimal (up to a logarithmic factor) because, as demonstrated in [11], any generic routing scheme with stretch strictly less than 5 must use at least $\Omega(n^{1/2})$ bits of memory on some nodes of some graphs. To the best of our knowledge, the TZ scheme is the only known generic stretch-3 routing scheme delivering a nearly optimal local memory upper bound. For this reason, we use it as the basis for our analytic work and simulations below.

B. Scale-free networks

Until fairly recently, most random graph analyses have been based on classical Erdős-Réni random n-node graphs [12], which have links between every pair of vertices with the uniform probability p. The ensemble of such graphs is called $\mathcal{G}_{n,p}$ and their average vertex degree is $\overline{k} \simeq np$. For large n, the vertex degree probability distribution for these graphs is the quickly-decaying Poisson distribution with an exponentially small number of high-degree nodes, $P_k \simeq \overline{k}^k \mathrm{e}^{-\overline{k}}/k!$, and average distance is $\overline{d} \simeq \log n / \log \overline{k}$ [13]. These graphs are uncorrelated,[3] and their entire statistical properties can be derived from this vertex degree distribution.

Almost all the networks observed in nature differ drastically from the $\mathcal{G}_{n,p}$ graphs. For our work, the most important difference is an inconsistency between the average distance and average vertex degree predicted by the Erdős-Réni model for the Internet. In the real Internet inter-domain 11000-node graph, $\overline{k} \simeq 5.7$ and $\overline{d} \simeq 3.6$ [14], while the $\mathcal{G}_{11000, 5.2\times10^{-4}}$ graphs have $\overline{d} \simeq 5.3$. The $\mathcal{G}_{n,p}$ graphs of the same size with the right average distance $\overline{d} \simeq 3.6$ would have to have the average degree $\overline{k} \simeq 14$. The simultaneously small values of the average distance and average vertex degree in the Internet necessarily imply a larger portion of its nodes are high-degree than in a comparably-sized $\mathcal{G}_{n,p}$ graph. In other words, the vertex degree distribution must be *fat-tailed*. The power-law distribution, $P_k \simeq k^{-\gamma}$, one of such fat-tailed distributions, is what has been observed in many real-world networks, with the exponent γ ranging between 2 and 3. For the Internet inter-domain graph, $\gamma \simeq 2.1$ [15], [14].

Both the $\mathcal{G}_{n,p}$ graphs and graphs with fat-tailed degree distributions are often said to possess the *small-world* property, [16], to emphasize that they have extremely low average distances (for networks of such size), even though average distances in $\mathcal{G}_{n,p}$ graphs are only slightly higher. Networks with power-law degree distributions are also called *scale-free* since their node degree distribution lacks any characteristic scale, [17], in contrast to the $\mathcal{G}_{n,p}$ graphs with the narrow Poisson degree distribution centered around the characteristic average value $\overline{k} \simeq np$.

The most popular model for growing scale-free networks uses *linear preferential attachment* by Barabási and Albert (BA), [17]. The BA model is very simple; it does not have external parameters, and in its "pure" form, it predicts $\gamma = 3$. The model can be easily modified to produce other values of γ, but its ability to help explain the evolutionary processes influencing the growth of the Internet has been questioned in [18], [19]. In particular, in [18], it is noted that the BA model and its derivatives are capable of reproducing what has been already measured but fail to predict correctly anything new about the Internet topology growth.[4] Construction of explanatory models capturing elements of the actual factors

[2] They also show how to implement routing decisions at *constant* time per node.

[3] That is, vertex-vertex degree correlations are absent.

[4] A good example is the power-law decay of the *clustering coefficient*. Models reproducing this effect were constructed only after it had been observed in real networks.

governing the Internet evolution is a *hot* topic of Internet-related research these days [1].

II. Compact Routing Schemes

In this section, we briefly review the Cowen and TZ schemes and establish the terminology and notation we will require. Both schemes are very simple. They involve four separate components: the landmark set (LS) construction procedure, routing table construction, labeling, and routing (message forwarding) function. The TZ scheme differs from the Cowen scheme by improving only the LS construction procedure.

Both schemes operate on any undirected connected graph $G = (V, E)$ with positive edge weights. Let $n = |V|$ be the graph size, $\delta(u, v)$ be the distance (in hops) between a pair of nodes $u, v \in V$, L be the LS, $L(v)$ be a landmark node closest to node $v \in V$, and $C(v)$ be v's *cluster*, defined for $\forall v \in V$ as the set of all nodes c that are closer to v than to their closest landmarks,

$$C(v) = \left\{ c \in V \;\middle|\; \delta(c, v) < \delta(c, L(c)) \right\}. \tag{1}$$

Clusters are similar to Voronoi diagrams but they can intersect. If $l \in L$, then $L(l) = l$ and $C(l) = \emptyset$ by definition. If L is empty, then for $\forall v \in V$, $L(v) = \emptyset$ and $C(v) = V$.

The TZ LS construction algorithm iteratively selects landmarks from the set of large-cluster nodes T. At the first iteration, $T = V$ and every node $t \in T$ is selected to be a landmark with a specific uniform probability q/n with $q = (n/\log n)^{1/2}$. The expected LS size after the first iteration is q. During subsequent iterations, T is redefined to be a set of nodes that have clusters of size greater than a specific threshold $\tilde{q} = 4n/q$,

$$T = \left\{ t \in V \;\middle|\; |C(t)| > \tilde{q} \right\}, \tag{2}$$

and additional portions of landmarks are selected from T with a uniform probability $q/|T|$. The iterations proceed until T is empty.

Every node $v \in V$ then calculates its outgoing port for the shortest path to every $l \in L$ and every $c \in C(v)$. This is the routing information that is stored locally at v. As one can see, the essence of the LS construction procedure is the right balance between the LS and cluster sizes (or, effectively, between q and $\tilde{q}$). The cluster sizes are upper-bounded by definition (2), and the involved part of the proof in [3] is to demonstrate that the algorithm terminates with a proper limit for the expected LS size, which turns out to be $2q \log n$. This guarantees the overall local memory upper bound of $O(n^{1/2} \log^{1/2} n)$.

The label of node v (used as its destination address in packet headers) is then a triple of its ID, the ID of its closest landmark $L(v)$, and the local ID of the port at $L(v)$ on the shortest path from $L(v)$ to v. With these labels, routing of a packet destined to v at some (intermediate) node u occurs as follows: if $v = u$, done; if $v \in L \cup C(u)$, the outgoing port can be found in the local routing table at u; if $u = L(v)$, the outgoing port is in the destination label in the packet; otherwise, the outgoing port for the packet is the outgoing port to $L(v)$—the $L(v)$ ID is in the label and the outgoing port for it can be found in the local routing table. The demonstrations of correctness of the algorithm and that the maximum stretch is 3 are straightforward ([10], [3]).

III. Analytical results

In this section, we provide analytical expressions for the TZ stretch distribution on a small-world graph with a given distance distribution.

To obtain our results we make a simplifying assumption: we consider only the first iteration of the TZ LS construction algorithm. There are two justifications making this assumption reasonable. First, as shown in [11], the first iteration guarantees that the *average* cluster size is n/q; the subsequent iterations guarantee that *all* cluster sizes are no larger than $4n/q$. Therefore, the error introduced by this assumption for the *average* stretch is small as we see in the next section. Secondly, we are concerned with small-world graphs which have very short average distances and narrow distance distributions. Indeed, if there are no long distances in a graph, then even after just the first iteration, the majority of clusters are small.

For the rest of this section, we let q denote the actual size of the LS ($q = |L|$) and D be the graph diameter (i.e. the maximum shortest path length in the graph). We also denote the distance p.m.f. and c.d.f. by $f(d)$ and $F(d)$, respectively. With D being the graph diameter, $d \in \{1 \dots D\}$.

Let w and v be independently selected random vertices from the random graph $G = (V, E)$ corresponding to a source and destination node, respectively. We introduce the following three random variables X, Y, and Z:

$$X = \delta(w, L(v)), \qquad \text{p.m.f.} \equiv p_X(x), \tag{3}$$

$$Y = \delta(v, L(v)), \qquad \text{p.m.f.} \equiv p_{Y_1}(y), \tag{4}$$

$$Z = \delta(w, v), \qquad \text{p.m.f.} \equiv p_Z(z) = f(z). \tag{5}$$

These random variables correspond to the distances from some random node w to the landmark nearest another random node v, the distance from that landmark to v, and the actual shortest path between the two random nodes. From these, we may construct another random variable, S^*, defined when $w \neq v$ to describe the stretch value

$$S^* = \frac{X + Y}{Z}. \tag{6}$$

This expression for stretch is approximate for two reasons. First, it does not account for stretch-1 paths to destinations in the local cluster of W. Second, it does not incorporate the *shortcut effect*. Recall that the Cowen routing algorithm is such that if destination $v \notin L$ and if a message on its way to $L(v)$ passes some node $u \mid v \in C(u)$, then the message never reaches $L(v)$ but instead goes along the shortest path from u to v. To refine our approximation, we correct the above definition of S^* to form S as follows:

$$S = S(X, Y, Z) = \begin{cases} 1 & \text{if} \quad Z < Y, \\ 1 & \text{if} \quad Z < X, \\ 1 & \text{if} \quad Z = 0, \\ \frac{X+Y}{Z} & \text{otherwise.} \end{cases} \tag{7}$$

Note that the first case on the r.h.s. of (7) accounts exactly for the stretch-1 paths to the destinations in the local cluster (cf. definition (1)), while the second case accounts approximately for the shortcut effect as shown in [21].

With the above notations, the p.m.f. for the distance Y_i between a random node and its i'th closest landmark is very similar to the p.d.f. for *order statistics* (see, for example, [20]). One can show (cf. [21]) that

$$p_{Y_i}(d) = i\binom{q}{i}F(d)^{i-1}f(d)(1-F(d))^{q-i}. \tag{8}$$

The p.m.f. for the average distance to all landmarks from a randomly-selected node is

$$p_X(d) = \frac{1}{q}\sum_{i=1}^{q} p_{Y_i}(d). \tag{9}$$

Since landmarks are just q random nodes, $p_X(d)$ is equivalent to $f(d)$,

$$\begin{aligned} p_X(d) &= f(d)\frac{1}{q}\sum_{i=1}^{q} i\binom{q}{i}F(d)^{i-1}(1-F(d))^{q-i} \\ &= f(d). \end{aligned} \tag{10}$$

Our problem now is to find the p.m.f. $p_S(s)$ for the random variable S. If X, Y, and Z were independent and unconstrained then $p_S(s)$ would be given by a simple sum over the joint distribution where $p_{XYZ}(x,y,z) = p_X(x)p_{Y_1}(y)p_Z(z)$. They are not independent, however, because they are defined in equations (3)-(5) on the same random events. Furthermore, the form of their definition results in the triangle inequality,

$$|X-Y| \leqslant Z \leqslant X+Y, \tag{11}$$

which causes some portions of the joint p.m.f to be zero. As such, the complete joint p.m.f. we require is derived in [21] to be

$$p_{XYZ}(x,y,z) = \frac{p_X(x)p_{Y_1}(y)p_Z(z)I_t(x,y,z)}{F(x+y)-F(|x-y|)}, \tag{12}$$

where the denominator is positive and the "triangle" indicator function is defined as follows

$$I_t(x,y,z) = \begin{cases} 1 & \text{if } \quad |x-y| \leqslant z \leqslant x+y, \\ 0 & \text{otherwise.} \end{cases} \tag{13}$$

The intuition behind equation (12) is as follows. We are selecting three random nodes on a graph. These nodes form a triangle. The distance distribution of one of the triangle's sides $p_{Y_1}(y)$ is special (given by the LS construction procedure). The probability mass is therefore concentrated only in the feasible combinations of x, y and z (i.e. those for which the triangle inequality holds). The denominator normalizes the product to form a proper p.m.f.

Using equation (12), the stretch distribution $p_S(s)$ and the average stretch $\overline{s}$ are computed as follows:

$$p_S(s) = \sum_{x=1}^{D}\sum_{y=1}^{D}\sum_{z=1}^{D} p_{XYZ}(x,y,z)|S(x,y,z)=s \tag{14}$$

$$\overline{s} = \sum_{s} s\cdot p_S(s) \tag{15}$$

For the average stretch, the summation in equation (15) is over the finite set of rational stretch values s bounded above by 2, as follows from equation (7).

Equations (14) and (15) are our final analytical results that we require for the numerical evaluations of the next section. Of particular note is that the stretch distribution and average depend only on $f(d)$ and q.

IV. Simulation and Numeric Results

We are now ready to substitute the Internet inter-AS distance distribution into the analytical expressions of the previous section. Because the Internet is an evolving network, it contains vertex-vertex correlations [22], and so to fully achieve our immediate goal, we would need to know an analytic result for the distance distribution in *correlated* scale-free networks. Unfortunately, this problem has not yet been solved.[5] Surprisingly, the *deterministic* scale-free graph model by Dorogovtsev, Goltsev, and Mendes (the DGM model) [24] analytically produces the Gaussian distance distribution, which is very close to the distance distribution observed in the real Internet inter-domain graph [22]. Given this observation, we choose to parameterize distance distributions in small-world graphs we consider in this section by Gaussian distributions. Using a discrete form of the Gaussian distribution as the distance p.m.f. $f(d)$ from the previous section transforms equations (14) and (15) into expressions that cannot be evaluated analytically, so that we retreat to numeric evaluations with $f(d)$ taken to be an explicitly normalized Gaussian,

$$f(d) = c\,e^{-\frac{1}{2}\left(\frac{d-\overline{d}}{\sigma}\right)^2}, \tag{16}$$

where c is s.t. $\sum_{d=1}^{D} f(d) = 1$, and $\overline{d}$ and σ are respectively the average distance and the width (standard deviation) of the distance distribution. Distributions $p_X(x)$ and $p_{Y_1}(y)$ are also explicitly normalized. Variables X, Y, and Z, defined in (3)-(5), take integer values within the following ranges:

$$\begin{aligned} D &= \left[\overline{d}\right] + \left\lceil 10\sigma\sqrt{2}\right\rceil, & (17)\\ X,Y &= 1\ldots D, & (18)\\ Z &= \max\left(1,|X-Y|\right)\ldots\min\left(D,X+Y\right), & (19) \end{aligned}$$

where $\left[\overline{d}\right] \equiv \text{round}\left(\overline{d}\right)$ and diameter D becomes a distance distribution cutoff parameter, $f(d) \ll 1$, $\forall d > D$ since $f(D)/f(\overline{d}) \simeq e^{-100}$. The TZ LS size q (cf. Section II) is rounded, $q = \left[(n/\log_2 n)^{1/2}\right]$.

[5] Although, there are some recent analytical results on distance distributions in *uncorrelated* scale-free graphs, [23].

For the following simulation results, we developed our own TZ scheme simulator and used it on graphs produced by the PLRG generator [25]. For a given parameter set, all the data is averaged over 10 random graphs. All average graph sizes n are between $10,000$ and $11,000$ unless mentioned otherwise.

A. Distance distribution

While the PLRG generator has been found to produce topologies largely consistent with those observed in the Internet inter-AS graph [28], it outputs uncorrelated graphs, and, hence, there are some concerns regarding its capability of reproducing *all* the features of strongly correlated nets, such as the Internet. However, since the stretch distribution is a function of the distance distribution and the graph size only (Section III), all we need from a graph generator for our purposes is that distance distributions in graphs produced by it be close to distance distributions observed in real-world graphs. Based on the experiments performed in [28], one can expect that the distance distribution in PLRG-generated graphs with the node degree distribution exponent $\gamma = 2.1$ should be close to the one in the Internet. We find that it is indeed so. See Fig. 1(a) for details.

We then proceed as follows. Paying special attention to the value of the node degree distribution exponent γ equal to 2.1, which is observed in the Internet, we generate a series of graphs with γ ranging from 2 to 3, and calculate their distance distributions. We fit these distributions by explicitly normalized Gaussians (16) yielding values of $\overline{d}$ and σ that we use in numerical evaluations of our analytical results. For fitting, we use the standard non-linear least squares method. All fits are very good: the maximum SSE we observe in our fits is 0.003 and the minimum R-square is 0.9905.

The values of $\overline{d}$ and σ in fitted Gaussians are slightly off from the means and standard deviations of distance distributions in generated graphs as depicted in Fig. 1(b). In fact, Fig. 1(b) is a parametric plot of $\sigma(\overline{d})$ with γ being a parameter. We observe an almost linear relationship between $\overline{d}$ and σ with such parametrization. Note that the almost linear relationship between the distance c.d.f. center and width parameterized by γ is analytically obtained in [23] as well. We further discuss this subject in Section V. In Fig. 1(c,d), we show fitted $\overline{d}$ and σ as functions of γ (cf. with the results in [23], [29]).

Average graph sizes for different values of γ are slightly different, but the dependence of $\overline{d}$ and σ on n (not shown) is negligible compared to their dependence on γ. This is in agreement with [23], [29].

B. Stretch distribution

We obtain a very close match between the simulations and analysis of the average TZ stretch and stretch distribution. The average stretch as a function of γ is shown in Fig. 2(a). For the Internet-like graphs, $\gamma = 2.1$, the average stretch we observe in simulations is 1.09 and the average stretch given by (15) with $f(d)$ in (16), with $\overline{d} = 3.4$ and $\sigma = 0.9$, is 1.14. Thus, we find that the average stretch is *very low*.

TABLE I

THE TOP TEN STRETCH VALUES AND PERCENTAGE OF PATHS ASSOCIATED WITH THEM.

Stretch	Analysis (%)	Simulations (%)
1	58.7	70.8
$4/3$	16.0	13.1
$5/4$	14.8	9.71
$3/2$	4.95	2.33
$5/3$	2.88	0.731
$6/5$	2.10	2.54
2	0.434	0.210
$7/5$	0.173	6.77×10^{-2}
$7/6$	5.20×10^{-2}	0.460
$8/7$	3.01×10^{-4}	7.42×10^{-2}

Furthermore, while both the average distance and distance distribution width in power-law graphs do depend on γ (cf. Fig. 1(c,d)), the average stretch *does not*. We delay the discussion of this topic until Section V.

The stretch distributions obtained both analytically, (14), and in simulations are shown in Fig. 2(b). The sets of significant stretch values (that is, stretch values having noticeable probabilities) match between the analysis and simulations. The top ten stretch values corresponding to virtually 100% of paths are presented in Table I.

We notice that a majority of paths (up to $\sim 71\%$ according to the simulations) are *shortest*. There are only a very few significant stretch values for the rest of paths. All the significant stretch values are below 2.

The small amount of stretch values with noticeable probabilities is due to the narrow width of the distance distribution. Indeed, in $\sim 86\%$ cases, two random nodes are either 3 or 4 hops away from each other. That is, the probability for X or Z to be either 3 or 4 is ~ 0.86, see Fig. 1(a). In $\sim 82\%$ cases, a random node is just one hop away from its closest landmark, $p_{Y_1}(1) \sim 0.82$. This explains why stretch-$4/3$ ($X = 3$, $Y = 1$, and $Z = 3$) and stretch-$5/4$ ($X = 4$, $Y = 1$, and $Z = 4$) paths are most probable among stretch $s > 1$ paths in Table I.

In Fig. 2(c), the analytical results for the average stretch as a function of the graph size are shown. Note that dependence on n in (15) is only via the LS size q. We present data for the case when $\overline{d}$ and σ are fixed at their values observed in the Internet, and the case when they are allowed to scale as in the DGM model. In both cases, the average stretch slowly *decreases* as the network grows, although this decrease is spread over multiple orders of magnitude of n and the stretch change is confined to a narrow region between 1.3 and 1.1. We also notice that after a certain point, the stretch stops decreasing. Although it becomes very small, it does not reach its minimal value 1.

Finally, in Fig. 2(d), we report the simulation data on the average cluster and LS sizes. (Recall that the sum of the cluster and LS sizes in the TZ scheme is the number of records in the local routing tables.) We notice that they are well below their theoretical bounds. Indeed, for the Internet-like graphs we studied, $n \sim 10^4$, $\gamma \sim 2.1$, this sum is ~ 52, while the theoretical upper bound, $6(n \log n)^{1/2}$, is ~ 2200.

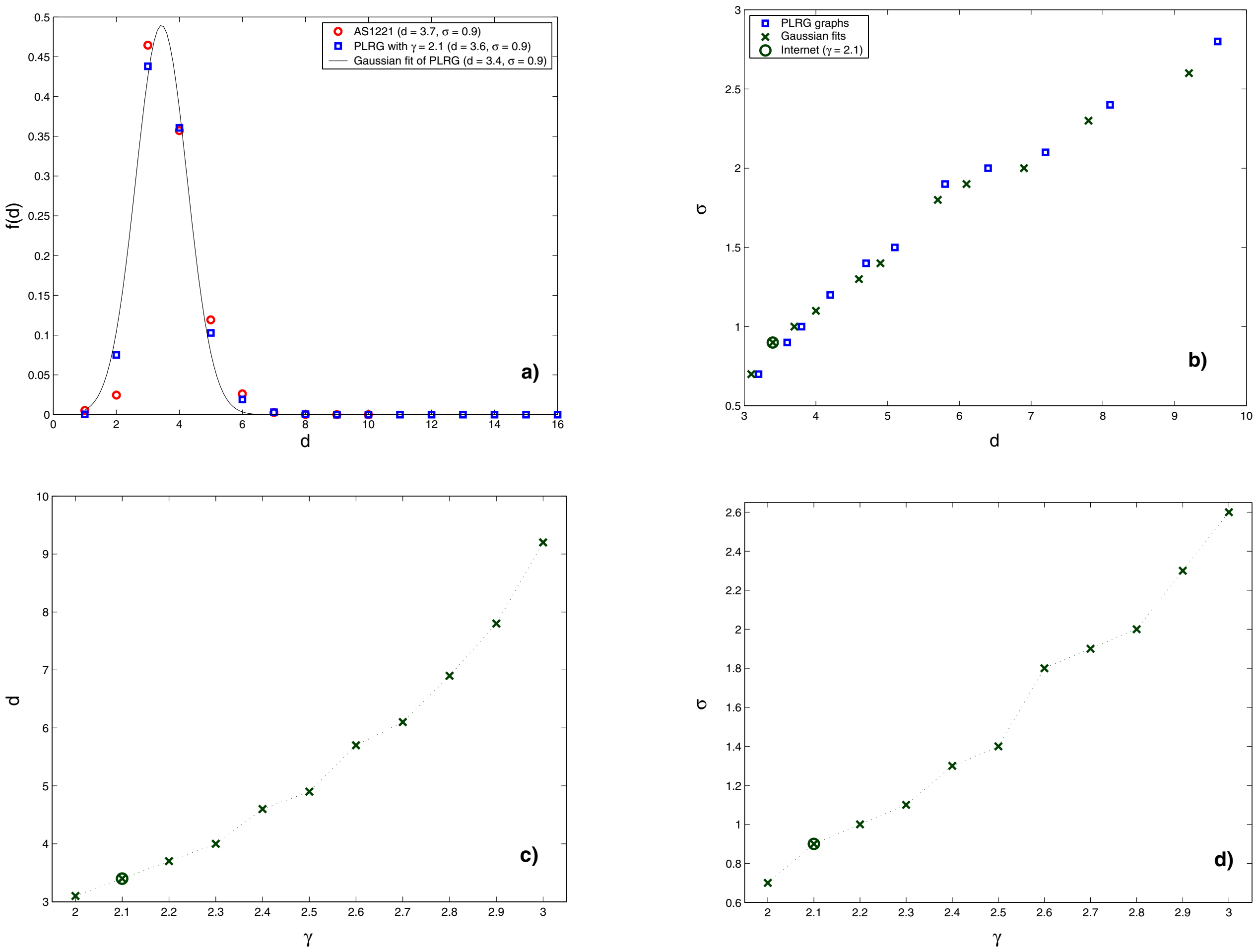

Fig. 1. **a)** The distance distributions. The circles represent the distance distribution from a typical AS (AS #1221) averaged over a period of approximately March-May, 2003. (The source of data is [26]; for other measurements, see [14], [27].) The mean and standard deviation is 3.7 and 0.9 respectively. The distance distribution in PLRG-generated graphs with $\gamma = 2.1$ is shown by squares. The standard deviation is the same as before, the mean is 3.6. The solid line is the Gaussian fit of the PLRG distribution, $\overline{d} = 3.4$ and $\sigma = 0.9$. **b)** The means and standard deviations (squares) of distance distributions in PLRG-generated graphs with $\gamma = 2.0, 2.1, \ldots, 3.0$ (from left to right), and the corresponding values of $\overline{d}$ and σ (crosses) in their Gaussian fits. The fitted values of $\overline{d}$ and σ as functions of γ are shown in **(c)** and **(d)** respectively. The Internet value of $\gamma = 2.1$ is circled in (b)-(d).

C. $\mathcal{G}_{n,p}$ *graphs*

Looking at Figs. 2(a,c), one may be tempted to assume that the average stretch just moderately depends on n and does not depend on either $\overline{d}$ or σ for a wide class of random graphs.

To demonstrate that this is incorrect, we consider the most common class of random graphs, $\mathcal{G}_{n,p}$. We take $n \sim 10^4$ and choose p to match approximately the Internet average distance ($p \sim 1.3 \times 10^{-3}$) and average node degree ($p \sim 5.7 \times 10^{-4}$). The analytical and simulation results for the average stretch in these two cases are presented in Table II. We find that the average stretch is substantially higher than in the case of random graphs with power-law node degree distributions.

V. Minimum Stretch and the Internet Graph

Our investigation so far suggests that the average TZ stretch depends strongly on the characteristics of the graph distance distribution—its average distance and width, in particular. Recall that now we are taking the distance distribution in a graph to be *Gaussian*, (16), and, hence, the average TZ stretch $\overline{s}$ in (15) is a function of the average distance $\overline{d}$ and the width of the distance distribution σ, $\overline{s} \equiv \overline{s}(\overline{d}, \sigma)$. At this point, we wish to explore the analytical structure of $\overline{s}(\overline{d}, \sigma)$ in more detail.

The natural starting point is to fix either $\overline{d}$ or σ to their observed values in the Internet, (3.4 and 0.9 respectively), and vary the other. This results of this exercise are illustrated in Figures 3(a,b). The left graph shows the stretch values when σ is fixed at 0.9 and $\overline{d}$ is allowed to vary between 0 and 7. The right graph shows the stretch values when $\overline{d}$ is fixed at 3.4 and the width σ is allowed to vary between 0 and 7. To our great surprise, we discover that these two functions have *unique minimums* and that the point corresponding to the Internet distance distribution (the large dots, which we will call the “Internet point”) are *very close* to them. In other words, one may get an impression that the Internet topology has been carefully “crafted” to have a distance distribution that would

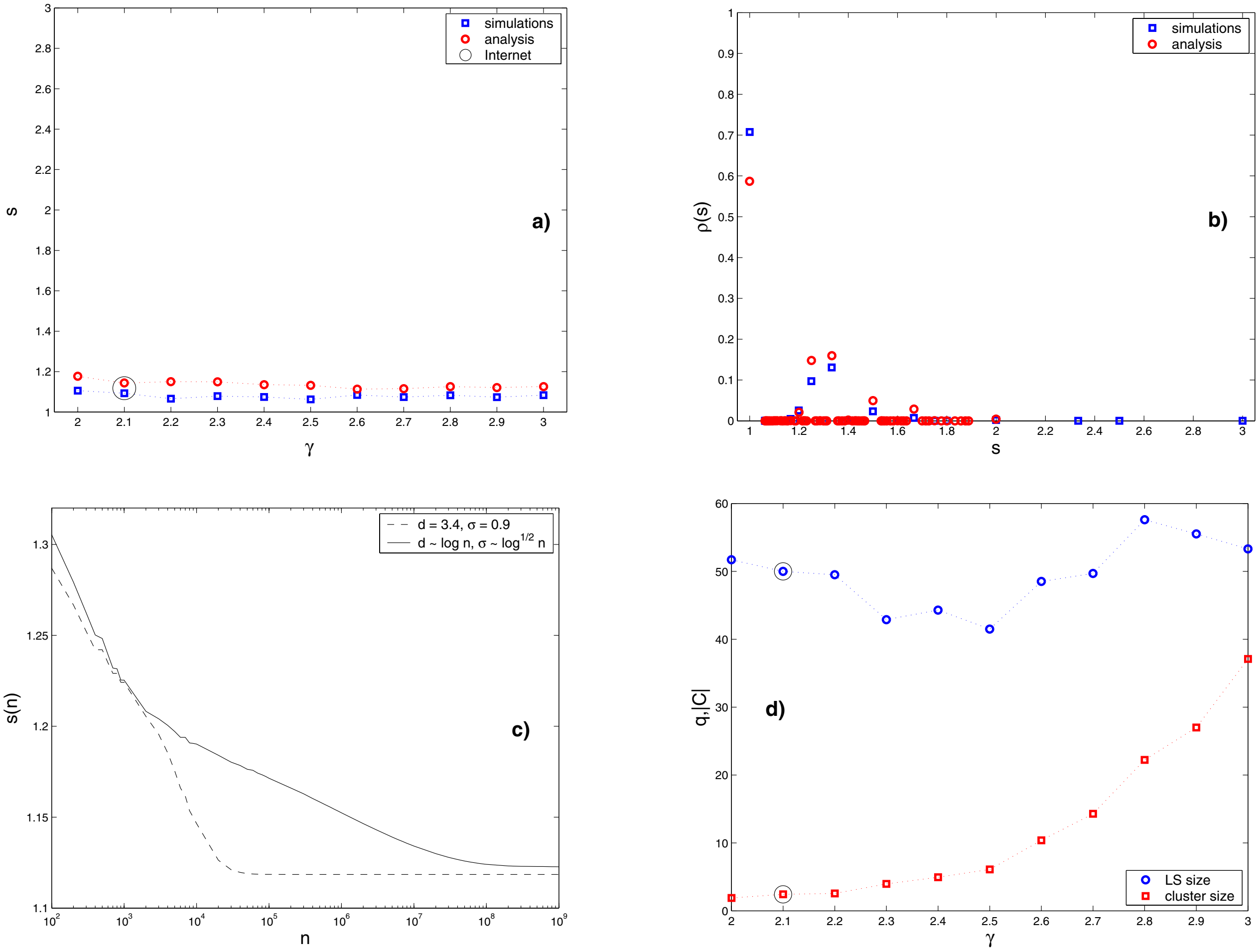

Fig. 2. **a)** The analytical results (circles) and simulation data (squares) for the average TZ stretch as a function of γ. **b)** The same for the TZ stretch distribution with $\gamma = 2.1$. **c)** The analytical data for the average stretch as a function of the graph size. The dashed line corresponds to the case when the distance distribution parameters $\overline{d}$ and σ are fixed to the values observed in the Internet. The solid line presents the data when $\overline{d}$ and σ scale according to the DGM model. **d)** The simulation data for the LS (circles) and cluster (squares) sizes. In the Internet case, $\gamma = 2.1$, the average graph size in simulations is 10,687, the average LS size is 50.0, and the average cluster size is 2.43.

TABLE II

THE AVERAGE TZ STRETCH ON THE $\mathcal{G}_{n,p}$ GRAPHS.

n	p	Avg. degree $\overline{k}$	$(\overline{d}, \sigma)$ in graphs	$(\overline{d}, \sigma)$ in Gaussian fits	$\overline{s}$ (analysis)	$\overline{s}$ (simulations)
10^4	1.3×10^{-3}	13	$(3.9, 0.6)$	$(3.9, 0.5)$	1.51	1.60
10^4	5.7×10^{-4}	5.7	$(5.5, 0.9)$	$(5.6, 0.8)$	1.37	1.50

(nearly) minimize the average TZ stretch. Of course, this can be only an impression and not an explanation since the Internet evolution, as we know it today, has had nothing to do with stretch.

The next question we have to ask is if the minimums we observe in Fig. 3(a,b) correspond to a true local minimum of the stretch function. Our analytical results allow us to construct Fig. 3(c), where the stretch function is plotted against both $\overline{d}$ and σ. Note that not all combinations of $(\overline{d}, \sigma)$ correspond to Gaussian-like distance distributions. Indeed, when $\sigma > \overline{d}$, $f(d)$ from (16) looks more like an exponential decay since it is cut off from the left by condition $d \geqslant 1$. Also, when σ is very small, (corresponding to highly regular graphs like complete graphs, stars, etc.), the peculiar peak formation in the $\sigma \sim 0$ area in the picture occurs. In this region, accurate computation of the stretch requires detailed knowledge of the particular graph topology.[6]

The region of primary interest to us, and which corresponds to real-world networks where our Gaussian model is likely to be most accurate, is when σ is somewhat greater than 0 and somewhat less then $\overline{d}$. Here, we observe a concave area in a form of a channel between the other regions described above. This area, which we shall term the *minimal stretch*

[6] Note, however, that for the complete network case, $\overline{d} = 1$, $\sigma = 0$, we obtain the correct answer for the average stretch, 2. For detailed explanation of the peak structure, see [21].

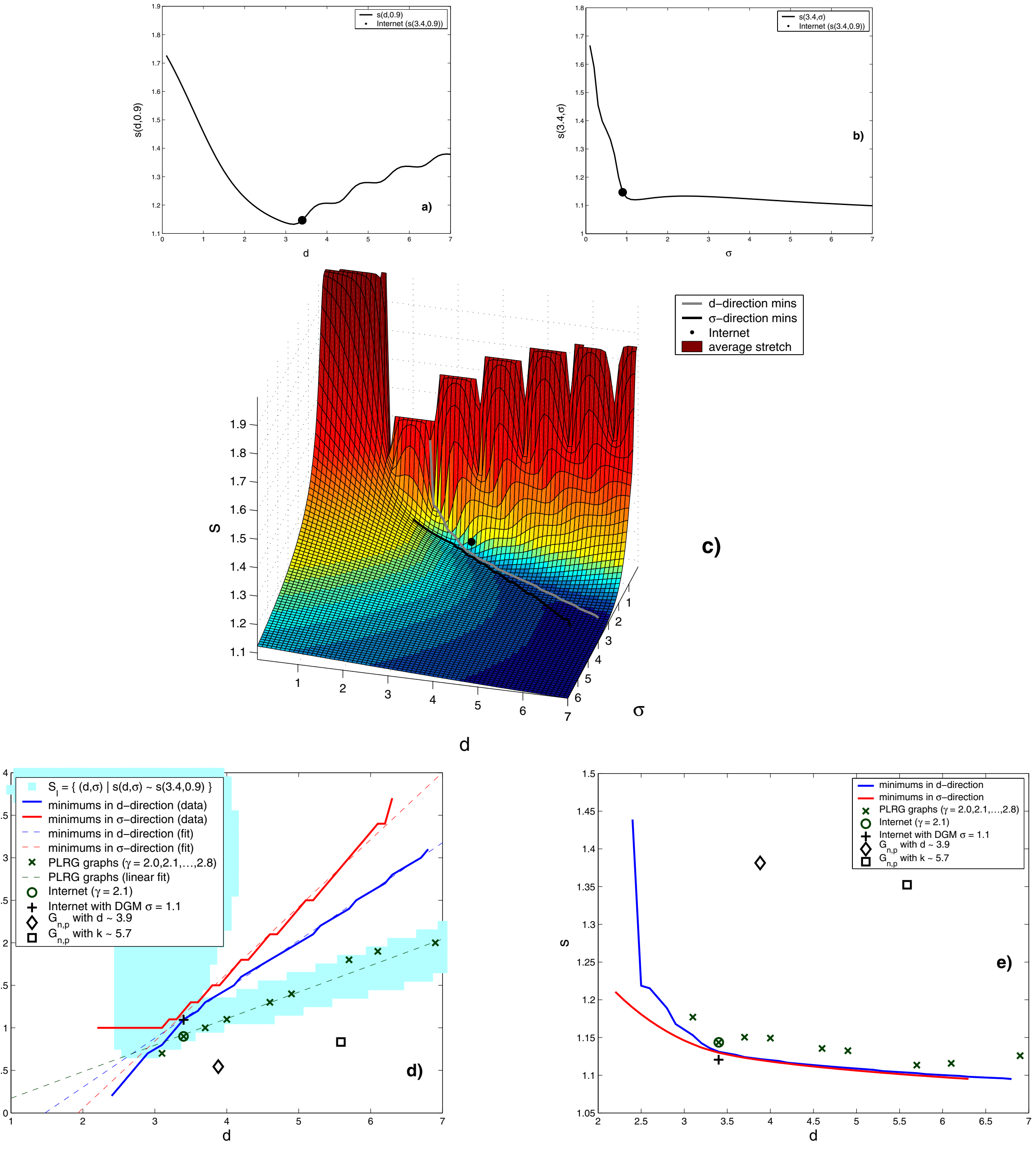

Fig. 3. **a),b)** The average stretch as functions of $\overline{d}$ with $\sigma = 0.9$ and of σ with $\overline{d} = 3.4$ respectively. The Internet is represented by the dots. **c)** The average stretch as a function of $\overline{d}$ and σ. The Internet is represented by the dot. The stretch minimums along the $\overline{d}$- and σ-axes, $M_{\overline{d}}$ and M_σ, are the light-grey and black lines respectively. **d)** The projection of (c) onto the $\overline{d}$-σ plane. The solid bottom and top lines represent respectively $M_{\overline{d}}$ and M_σ (the light-grey and black lines from (c)). The two dashed lines are their linear fits in the *MSR*. The crosses are the same as in Fig. 1(b), the bottom-most dashed line being their linear fit. The Internet, $\gamma = 2.1$, is circled. The shaded area is M_I from the text. The plus is the point with the average distance observed in the Internet and the Gaussian width predicted by the DGM model, $\overline{d} = 3.4$, $\sigma = 1.1$. The diamond and square are the distance distributions of the $\mathcal{G}_{n,p}$ graphs from Table II matching the Internet average distance and node degree. **e)** The projection of (c) onto the $\overline{d}$-$\overline{s}$ plane. The notations are the same as in (d). The graph sizes $n \sim 10^4$ everywhere.

region (MSR), is characterized by particularly low stretch values. The width and depth of the MSR slowly *increase* as $(\overline{d}, \sigma)$ grow. For small $(\overline{d}, \sigma)$, the MSR has a unique critical point, which we call the *MSR apex*. The Internet point is located very close to the apex, which is characterized by the shortest distance between the sets of minimums of the average stretch function $\overline{s}(\overline{d}, \sigma)$—along the $\overline{d}$- and σ-axes. We denote these two sets by $M_{\overline{d}}$ and M_σ respectively. We find that $M_{\overline{d}}$ and M_σ *almost touch* each other at the apex.

The MSR apex can be more easily observed in Fig. 3(d) showing a projection of Fig. 3(c) on the $\overline{d}$-σ plane. The solid lines represent the above two sets of minimums forming the MSR, and the Internet point is very near their closest segment.

An opportunity to look at the apex from yet another angle is presented in Fig. 3(e) showing a projection of Fig. 3(c) on the $\overline{d}$-$\overline{s}$ plane. We see that starting from the apex, as $\overline{d}$ increases, the minimum stretch values along the $\overline{d}$- and σ-directions become virtually equal and slowly *decrease* as $\overline{d}$ grows. We also note that $\mathcal{G}_{n,p}$ graphs are far away from the apex and that they have average stretch values that are far from minimal.

We can see now that the apex is indeed a critical or "phase transition" point since it is located at the boundary of the two regions of the average stretch function. The first region, the MSR, is characterized by lowest possible stretch values corresponding to distance distributions observed in real-world graphs. The second region, with substantially higher average stretch values, corresponds to distance distributions in more regular graphs.

To illustrate this point in more detail, we turn our attention back to Fig. 3(d). We observe that the two sets of minimums, $M_{\overline{d}}$ and M_σ, are linear when $\sigma > 1$. The dashed lines represent the linear fits of $M_{\overline{d}}$ and M_σ in the area with $\sigma > 1$. The exact location of the intersection of these fits is $(\overline{d}^\star, \sigma^\star) = (3.16, 0.97)$. If the linear form of $M_{\overline{d}}$ and M_σ sustained for $\sigma < 1$ as well, then $M_{\overline{d}}$ and M_σ would intersect at $(\overline{d}^\star, \sigma^\star)$, where we would observe a stationary[7] point of $\overline{s}(\overline{d}, \sigma)$, which we could then test for the presence of an extremum of the stretch function. This does not happen, however. Instead, as $\overline{d}$ and σ become small, the linear behavior breaks near the apex due to increasingly "more discrete" structure of the distance distribution ([21]).

In the extended version of this paper [21], we show that linearity of $M_{\overline{d}}$ and M_σ can be analytically derived from the fact that the distance distribution is taken to be Gaussian. Of course, this does not explain why the *Internet* is so close either to the MSR or to its apex.

The linear form of $M_{\overline{d}}$ and M_σ in the MSR sheds some light on a closely related issue of why the average stretch is virtually independent of γ. In Fig. 3(d), the shaded area represents a set of $(\overline{d}, \sigma)$, for which the average stretch is approximately the same as for the Internet, $M_I = \left\{ (\overline{d}, \sigma) \mid \overline{s}(\overline{d}, \sigma) \sim \overline{s}(3.4, 0.9) \right\}$. We see that in the MSR, the M_I boundaries are almost parallel straight lines. Therefore, if the average stretch is to be independent of γ, which is observed in Section IV-B, then the points representing distance distributions in power-law graphs, $(\overline{d}_\gamma, \sigma_\gamma)$ from Fig. 1(b), should lie along the M_I boundaries, and this is what indeed happens. Yet again, the linear relation between $\overline{d}_\gamma$ and σ_γ in the power-law graphs, and the fact that this relation is just as required for the average TZ stretch being virtually independent of γ, come from two seemingly disjoint domains.

To finish the list of various "coincidences," we construct a linear fit of $(\overline{d}_\gamma, \sigma_\gamma)$ (the bottom-most dashed line in Fig. 3(d)). This line is located exactly at the MSR edge. Indeed, in the area of higher values of σ (that is, in the MSR), the stretch function is completely concave, while for smaller σ, we observe multiple convex and concave regions caused by "more discrete" structure of the distance distribution. In other words, the line corresponding to scale-free graphs is right at the boundary between "more random" and "more regular" graphs.

Furthermore, the Internet point, $\gamma = 2.1$, lies on this line, and our numeric analysis shows that the Internet value of $\gamma = 2.1$ minimizes the distance between the linear fit of $(\overline{d}_\gamma, \sigma_\gamma)$ and $(\overline{d}^\star, \sigma^\star)$, which is the intersection of the linear fits of $M_{\overline{d}}$ and M_σ. In other words, the Internet distance distribution is the point that is *closest* to the MSR apex, compared to distance distributions in all other scale-free graphs with power-law node degree distributions.

VI. Conclusions

We find that the TZ routing scheme applied to the Internet inter-AS graph results in a very low average stretch and succinct routing tables that are well below their upper bounds. The primary reason why the average stretch is of a great concern is that the TZ scheme is not a stretch-1 scheme, while Internet inter-domain routing is essentially shortest path routing.[8] Thus, any stretch $s > 1$ routing scheme applied to the Internet would involve augmentation, in one form or another, of the routing information provided by the scheme with the shortest path routing information for some (or all) non-shortest paths.

Our principal finding, that the TZ stretch on the Internet graph is reasonably low, opens a well-defined path for future work in the area of applying relevant theoretical results obtained for routing to realistic scale-free networks. One of the immediate next problems on this path is the performance analysis of *dynamic* low-stretch routing schemes on scale-free graphs. The TZ scheme is not optimal for the dynamic case since it labels nodes with topology-sensitive information. In other words, it is not name-independent. As soon as the topology changes, nodes need to be relabeled. Significant progress in construction of name-independent low-stretch routing schemes has been recently made by Arias, Cowen, et al. in [30].

[7] Recall that a function has a stationary point where all its first-order partial derivatives are zero. Thus, we can call the apex a *quasi- stationary* point emphasizing that $\partial\overline{s}/\partial\overline{d}$ and $\partial\overline{s}/\partial\sigma$ are both *nearly* zero at the apex.

[8] A routing scheme that would prevent, for example, a pair of ASs from utilizing a peering link between them is not realistic, of course.

More importantly, however, we find that the Internet shortest path length distribution is at the minimal distance from the unique critical point of the average stretch function. At present we lack sufficient information to show cause for this effect, but we do believe it strongly suggests the average stretch function may be an indirect (or even direct) indicator of some yet-to-be discovered process that has influenced the Internet's topological evolution. In other words, a rigorous explanation of this phenomenon would probably require much deeper understanding of the Internet evolution principles and demonstration of a link between them and the TZ scheme. This question is of great interest, as the fundamental laws governing the Internet evolution remain unclear. Therefore, a proper explanation of this effect may help us in our intent to move, perhaps along the lines of [1], from purely *descriptive* Internet evolution models to more *explanatory* ones, in the terminology of the program outlined in [18].

References

[1] H. Chang, S. Jamin, and W. Willinger, "Internet connectivity at the AS level: An optimization driven modeling approach," in *Proc. of MoMeTools*, 2003.

[2] G. Huston, *Commentary on Inter-Domain Routing in the Internet*, IETF, RFC 3221, 2001.

[3] M. Thorup and U. Zwick, "Compact routing schemes," in *Proc. of the 13th SPAA*. ACM, 2001.

[4] L. Kleinrock and F. Kamoun, "Hierarchical routing for large networks: Performance evaluation and optimization," *Computer Networks*, vol. 1, 1977.

[5] I. Castineyra, N. Chiappa, and M. Steenstrup, *The Nimrod Routing Architecture*, IETF, RFC 1992, 1996.

[6] F. Kastenholz, *ISLAY: A New Routing and Addressing Architecture*, IETF, Internet Draft, 2002.

[7] C. Gavoille and S. Pérennès, "Memory requirement for routing in distributed networks," in *Proc. of the 15th PODC*. ACM, 1996.

[8] H. Buhrman, J.-H. Hoepman, and P. Vitány, "Space-efficient routing tables for almost all networks and the incompressibility method," *SIAM J. on Comp.*, vol. 28, no. 4, 1999.

[9] C. Gavoille and M. Genegler, "Space-efficiency for routing schemes of stretch factor three," *J. of Parallel and Distributed Comp.*, vol. 61, no. 5, 2001.

[10] L. Cowen, "Compact routing with minimum stretch," *J. of Algorithms*, vol. 38, no. 1, 2001.

[11] M. Thorup and U. Zwick, "Approximate distance oracles," in *Proc. of the 33rd STOC*. ACM, 2001.

[12] P. Erdős and A. Réni, "On random graphs," *Publicationes Mathematicae*, vol. 6, 1959.

[13] B. Bollobás, *Random Graphs*, Academic Press, New York, 1985.

[14] A. Vázquez, R. Pastor-Satorras, and A. Vespignani, "Internet topology at the router and Autonomous System level," cond-mat/0206084.

[15] M. Faloutsos, P. Faloutsos, and C. Faloutsos, "On power-law relationships of the Internet topology," in *Proc. of the ACM SIGCOMM*, 1999.

[16] S. Milgram, "The small world problem," *Psychology Today*, vol. 61, 1967.

[17] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, 1999.

[18] W. Willinger, R. Govindan, S. Jamin, V. Paxson, and S. Shenker, "Scaling phenomena in the Internet: Critically examining criticality," *PNAS*, vol. 99, 2002.

[19] Q. Chen, H. Chang, R. Govindan, S. Jamin, S. J. Shenker, and W. Willinger, "The origin of power laws in Internet topologies revisited," in *Proc. of the IEEE INFOCOM*, 2002.

[20] S. Ross, *Introduction to Probabilty Models (7th ed)*, Academic Press, San Diego, 2000.

[21] D. Krioukov, K. Fall, and X. Yang, "Compact routing on Internet-like graphs," Tech. Report IRB-TR-03-010, Intel Research, 2003.

[22] S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of Networks: From Biological Nets to the Internet and WWW*, Oxford University Press, Oxford, 2003.

[23] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, "Metric structure of random networks," *Nucl. Phys. B*, vol. 653, no. 3, 2003.

[24] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, "Pseudofractal scale-free web," *Phys. Rev. E*, vol. 65, 066122, 2002.

[25] W. Aiello, F. Chung, and L. Lu, "A random graph model for massive graphs," in *Proc. of the 32nd STOC*. ACM, 2000.

[26] "BGP Table Data, http://bgp.potaroo.net/".

[27] A. Broido, E. Nemeth, and k claffy, "Internet expansion, refinement, and churn," *European Trans. on Telecommunications*, vol. 13, no. 1, 2002.

[28] H. Tangmunarunkit, R. Govindan, S. Jamin, S. Shenker, and W. Willinger, "Network topology generators: Degree-based vs. structural," in *Proc. of the ACM SIGCOMM*, 2002.

[29] F. Chung and L. Lu, "The average distance in a random graph with given expected degrees," *Internet Mathematics*, vol. 1, no. 1, 2003.

[30] M. Arias, L. Cowen, K. A. Laing, R. Rajaraman, and O. Taka, "Compact routing with name independence," in *Proc. of the 15th SPAA*. ACM, 2003.

Appendix

In this appendix, we calculate a rough estimate of the stretch factor of the Kleinrock-Kamoun (KK) hierarchical routing scheme, [4], applied to the observed Internet interdomain topology. We find that the stretch is very high, which is consistent with the observation made in [4] that the approach used there works reasonably well only for sparsely connected networks. The scale-free networks, on the contrary, are extremely densely connected.

Recall that [4] assumes the existence of a hierarchical partitioning of a network of size n into m levels of clusters. Each k-level cluster consists of $n^{1/m}$ $(k-1)$-level clusters, $k = 1 \ldots m$, 0-level clusters being nodes. The optimal clustering is achieved when $m \sim \log n$. There are a few other fairly strong assumptions about the properties of the required partitioning. Neither an algorithm for its construction nor proof of its existence are delivered, but if it does exist then the stretch factor is shown to be

$$s = 1 + \frac{1}{\overline{d}} \sum_{k=1}^{m-1} \left[1 - \frac{n^{\frac{k}{m}} - 1}{n - 1} \right] d_k, \qquad (20)$$

where $\overline{d}$ is the network average distance and d_k is the diameter of a k-level cluster.

It is further assumed in [4] that both the network diameter and average distance are power-law functions of the network size. This is certainly not true for scale-free networks with power-law *node degree* distributions. For recent results on the average distance in such networks, see [29], [23]. In the numerical evaluations in this appendix, we use the value of $\overline{d} \sim 3.6$ observed in the Internet, [26].

As shown in [29], the *diameter* of networks with power-law node degree distribution with exponent γ lying between 2 and 3 scales almost surely as $\Theta(\log n)$. For the Internet, $\gamma \sim 2.1$, and since the Internet size $n \sim 1.5 \times 10^4$ is relatively large, we may write the Internet diameter D as $D \sim c \log n$ with some multiplicative coefficient c. The observed value of D, $D \sim 13$ ([26]), defines c.

The size of a k-level cluster is obviously $n^{k/m}$ but nothing rigorous can be said about its degree distribution since there is no procedure for its construction. Thus, it is natural to assume that its degree distribution is also power-law with $2 < \gamma < 3$, which gives an estimate of the k-level cluster diameter as $d_k \sim c \log n^{k/m} \sim Dk/m$. Substituting this in (20) and performing summation gives

$$s \sim 1 + \frac{D}{2\overline{d}} \left[m \frac{n}{n-1} - \frac{n(n^{\frac{2}{m}} - 1)}{(n-1)(n^{\frac{1}{m}} - 1)^2} + \frac{2}{m} \frac{n^{\frac{1}{m}}}{(n^{\frac{1}{m}} - 1)^2} \right]. \tag{21}$$

Using the numerical values for n, $\overline{d}$, D, and optimal $m = 10$, we can see that the KK stretch factor on the Internet inter-domain topology is

$$s \sim 15. \tag{22}$$

Note that a 15-times path length increase in the Internet would lead to AS path lengths of ~ 55 and IP hop path lengths of ~ 150.

The stretch factor is a nearly linear function of the number of hierarchical levels m, which follows directly from equation (21) since it can be rewritten for large n as

$$s \sim 1 + \frac{D}{2\overline{d}}(m-1). \tag{23}$$

Using $D \sim \log n$, $\overline{d} \sim \log\log n$ ([29]), and optimal $m \sim \log n$, we obtain the following estimate of the stretch factor as a function of the network size:

$$s \sim \frac{\log^2 n}{\log\log n}. \tag{24}$$