

Modeling small-worlds with geographical factors: distance-bias & bounded-growth neighborhoods ^{*}

Van Nguyen and Chip Martel
{nguyenvk,martel}@cs.ucdavis.edu

Computer Science, UC Davis, CA 95616

Abstract. Studies of many real world networks show that geographical factors play a significant role. However, existing models for small-world properties and power-law degrees don't fully consider these geographical factors. We propose a *general model for small-world and power-law features which also considers geographical factors*, including the *distance-bias tendency* (links tend to favor closer distances) and the property of *bounded growth* in neighborhood expansion. Our refined model combines a growth bounded base graph with a distance-bias distribution of random links. We show when the small-world effect may occur and how the diameter changes *depending on the coordination between the distance-bias parameter and the two bounded growth parameters*. This helps explain why the Internet graph is considered as a small-world with low diameter, but is locally growth bounded. We develop analysis techniques for graphs with non-uniform random links, including a fractal-based analysis. We also discuss future work on applications to network design, where our models help augment networks for improving routing and other related issues.

1 Introduction

A large inter-disciplinary community is actively studying real-world complex networks. In 1967, Milgram confirmed folklore, that we live in a small-world where two strangers are linked by a short chain of acquaintances. In 1998, Watts and Strogatz [27] produced a graph model for small-world networks, which has small diameter (i.e. paths between any two nodes require only a few hops) and high clustering coefficient (two nodes with a shared neighbor are likely connected by a link). These small-world properties occur in many large-scale networks such as social networks, biological networks and the Internet. The power-law feature of vertex degrees has also been widely recognized as a common property. For example, the distribution of vertex degrees in the Internet topology (at both inter-domain & router levels) has a power-law shape [8]. An extensive review can be seen in, e.g., [21].

Good models can give insights into the mechanisms underlying those common phenomena and features and can also influence network design and help control the networks. We are developing new graph models for the Internet and related networks. These models can be used to augment existing networks for better performance, to design new computer networks, and to provide algorithmic support, i.e. efficient algorithms for basic routing or distributed search.

1.1 Our work in perspective

In this paper, we provide a general model for real-world random networks, where we focus on modeling two important geographical factors: the property of *bounded growth* in neighborhood expansion and the *distance-bias tendency*. These arise in many settings and particularly in Internet related settings. For example, Faloutsos et al. [8] observed in the Internet topology that a 'ball' of neighbors within distance R has size approximated by R^β when the radius distance R is small enough. Note that β may vary between locations due to different population density or physical characteristics. Thus, our model's ability to reflect variable neighborhood growth rates is an important feature. Researchers have also found similar observations in wireless networks and peer-to-peer networks; so, several papers use bounded growth when working with Internet-related problems, e.g. [25, 7, 13, 10, 1]. Recent studies of the Internet's topology evaluate the role of *distance-bias* where links are more likely to connect closer nodes [28, 19]. An approximate linearity of distance bias in the link distribution is observed in [28].

^{*} This work was supported by NSF grant ANI-04-35525 and NSF grant 0520190.

We consider a general graph model where an arbitrary base graph, representing local connections, is augmented with random (usually long) links. We study general conditions which result in short paths between almost all nodes. We then analyze a refined model, which simulates practical settings (detailed below), where the base graphs are growth-bounded and the random links are distance-biased. Moreover, we suggest an efficient routing strategy for this setting, using limited local information.

Previous models, e.g. [27, 16, 22], use simple grid or a grid-like base graphs that all have constant “growth degree” (e.g. 2 for a 2-dimension grid), so the rate of growth in neighborhood expansion, β , is the same for the entire graph. Instead, we allow the local growth rate to vary from place to place.

Example 1 (variable growth in wireless networks). Think of a wireless sensor network covering three nearby structures, a 3-D building, a 2-D lawn and a 1-D road: the network has growth degree 3 inside the building (so nodes in a neighborhood grow at a cubic rate), degree 2 on the lawn, and 1 on the road.

We also expect fluctuating growth rates in the physical layout of the Internet, due to population density patterns around the globe [28]. Thus, we model variable bounded growth by two parameters: a lower and upper bound on β . As an example application, our models help consider how to add a wired infrastructure into a wireless network to obtain faster routing while keeping a low total cost and reasonable congestion in these extra links (our preliminary result for the grid setting is in [24]).

Below we compare our approach with prior work, then summarize our results and discuss our new technical contributions.

1.2 Comparison with prior work

Recent mathematical models try to simulate small-world and power-law properties. There is a certain complexity and difficulty to contain them all in a single mathematical model, but there has been some effort to do that [6]. However, none of these mathematical models looks at all the geographical factors mentioned above. Below, we review two important models, which address complementary goals.

a) Chung-Lu model [4–6]. Chung and Lu start with a generalized random graph model [4, 5], which aims at modeling power-law networks where the number of nodes with degree k is proportional to $k^{-\beta}$ for some fixed exponent β . Their model can construct random graphs with specified expected degrees: they assign weights to the nodes and have an independent random link between any two nodes with a probability proportional to the product of the weights of these two nodes. Thus, the expected degree of each node is specified by its weight ¹.

However, their distribution of random links does not feature a distance bias and their graphs have exponential growth instead of locally bounded growth. Recently, Chung and Lu extended their approach to model the clustering effect by a hybrid model where they couple their power-law random graph - as a ‘global graph’ - with a specific ‘local graph’ rich in local links. They consider locally (k, l) -connected graphs where for any two nodes u and v , there are at least l edge-disjoint paths with length at most k between u and v . In contrast, our models allow a more general class of local graphs.

b) Kleinberg model [16]. Kleinberg extended Watts and Strogatz’s model to help explain another striking aspect of the small-world phenomenon: that short paths can be found using a simple greedy strategy with limited local information (reflecting Milgram’s search strategy based on a first-name basis). In this model, a grid graph is augmented by random links, where a link is more likely to go to closer nodes. Specifically, each node u has one random link which goes to a node v with probability inversely proportional to $d^\alpha(u, v)$, where $d(u, v)$ is the lattice distance between u and v - for brevity, we write $Pr[v \stackrel{R}{\leftarrow} u] \propto d^{-\alpha}(u, v)$. His main result shows that greedy routing, using local information, can find short routes only when $\alpha = 2$. This started the active study of searchable small-worlds where decentralized routing, e.g. by a greedy strategy, is possible [18].

Despite Kleinberg’s model’s use of distance-bias links, it does not satisfy our main goals above. The grid graphs or similar meshes have a uniform distribution of degrees and density, so, they are too simple

¹ The Chung-Lu model is classified as an off-line power-law model to opposite to on-line ones, i.e. the evolutionary models of growth with preferential attachment, in which nodes come one-by-one at each unit of time and ‘attach’ to the existing nodes with a stronger preference to ones with higher degrees (see e.g. [21] for a survey).

to simulate localities in real-world networks, where skewed distributions are typical. Also, Kleinberg’s and follow-up work only provide an efficient routing strategy except for the special setting when $\alpha = 2$.

We build on the nice elements in those two models (and earlier models): (i) a base graph with pre-determined local links is augmented by random links which are formed based on (ii) given node weights, and (iii) a distance-bias. It is natural to think of combining these together (and possibly additional features) in one model. However, building such a complex mathematical model can be challenging². Modeling local base graphs in an accurate flexible way is challenging: we want a general and versatile approach so our local graphs can feature arbitrary degrees and bounded-growth rates, and should be tractable for finding diameter analysis and routing algorithms.

This work also builds on [22], where we provided an extensive diameter analysis for Kleinberg’s models and a general framework to construct small-world graphs, but still restricted to settings with a constant growth rate.

1.3 Our contributions

We propose a more general model for analyzing real-world random networks. Our model couples an (arbitrarily) *weighted base graph* and a *node-based distribution* of random links where the number of random links generated by a node depends on the node weight. With the right conditions this structure has power-law degrees. In theorem 1 we show a *general sufficient condition* (§3) for this setting to have the small-world property that most paths are short³; specifically, we show that there exists a giant component of size $\theta(n)$ with small diameter. Our model is strong enough to show several interesting facts easily, e.g. an $O(\log n)$ upper-bound for the diameter of a power-law graph in corollary 2. In another illustration, for Kleinberg’s grid and tree-based settings [16, 17] without local links, if each node generates at least expected $1 + \epsilon$ random links (for any $\epsilon > 0$), there exists a giant $O(\log n)$ -diameter component.

We then analyze a refined model coupling a *growth-bounded* local graph with variable growth rates and a *distance-bias* distribution of random links ($Pr[v \xrightarrow{R} u] \propto d^{-\alpha}(u, v)$ for a distance measure $d(u, v)$ and some parameter $\alpha > 0$). We show in section 4 when the small-world effect occurs in theorem 2, and in theorem 3 we show how the diameter changes *depending on the relationship between the distance-bias parameter (α) and our two bounded-growth parameters*. Specifically, if α is less than twice the minimum growth degree (i.e. the lower bound on β), the setting has small diameter with high probability. This helps explain why the Internet graph is considered a small-world with low diameter, but is locally growth bounded .

Based on this analysis, we suggest a natural *decentralized routing strategy* which can perform efficiently in a broad class of the above distance-bias structures (where Kleinberg’s setting with $\alpha = 2$ is a very special case). However, due to space limitations, we only discuss this briefly in section 4. We also only discuss the main ideas of our theorems and leave most detailed proofs to the appendix.

Main technical contributions. Although, we follow Chung and Lu’s pattern in adding random links to the nodes based on their given weights, we face a different technical challenge. Since we add random links using a distance-bias distribution, we can’t directly exploit the classical study of random graphs, which focuses on uniform random links. A traditional technique to show the existence of a typical short $s - t$ path is to construct two sufficiently large ‘balls’ of neighbors centered around s and t , then show the existence of a random link connecting these two balls. For our general setting we prove a diameter result using a new technique based on our ‘*connecting lemma*’ in section 3. Instead of connecting just two pieces (s ’s and t ’s balls), we connect a collection of pieces together. We show how to partition the graph into pieces where each piece has small diameter, and then show that the super-graph (whose nodes are contractions of these pieces) has expected diameter $O(1)$.

We use the *connecting lemma* with another technique where *we view our graph model as a fractal*: the whole graph is a dense spatial mesh of pieces, each separates from another by at most a constant

² Simplistic models help provide insight into a certain aspect but do not reflect the full real-world network; more complex models can describe more practical settings but are usually harder to work out mathematically and may incur complex notions.

³ We do not look at the clustering property in this structure as it can be obtained by using a base graph rich in local links.

number of links; each piece then has a similar structure using smaller sub-pieces, and so on. This fractal technique not only helps in our diameter analysis but also in our design of routing strategy and algorithms.

2 Our general model: basic ideas, definitions and facts

2.1 Our general model.

For “adding random links”, we use the following random assignment operation based on a distribution rule τ . For a given node u in a graph G , an R_τ call on u creates a random link to another node v . We write this as $v = R_\tau(u)$ or $v \xleftarrow{R_\tau} u$. We only consider τ such that all random links, even with R_τ called on the same input node, are independent.

We now define our basic class of random graphs, where each node is assigned a weight which determines the expected number of random link ‘generated’ by this node. Let $\mathcal{H} = \{H(V, W, E)\}$ denote an infinite collection of undirected base graphs, where V is a set of vertices, E is a set of edges between nodes in V and W is a weight function: $V \rightarrow Z^+$ so each node $u \in V$ has weight W_u .

Each node u has $Be(W_u, p)$ random links generated from it, where $Be(k, p)$ denotes the number of successes by doing k Bernoulli’s random trials each with probability of success p . Intuitively, node u is assigned a fixed number of *seeds* W_u , where each seed gives birth to a random link with probability p ⁽⁴⁾.

Definition 1. For a set of undirected base graphs $\mathcal{H} = \{H(V, W, E)\}$, a distribution τ , and a number $p \in [0, 1]$, the family of undirected random graphs $\mathcal{P}\mathcal{L}\mathcal{U}\mathcal{S}(\mathcal{H}, \tau, p)$ consists of graphs, each of which is a base graph $H \in \mathcal{H}$ augmented by adding $Be(W_u, p)$ undirected random links for each node u , using distribution τ for each link. We also use $\mathcal{P}\mathcal{L}\mathcal{U}\mathcal{S}(\mathcal{H}, \tau)$ for the special case $p = 1$.

For simplicity, we use undirected random links here, but our results can be partially extended for directed random links (due to limited space, we defer this extension for directed links to the full version[23]). Intuitively, each node v generates and, so, “owns” certain random links, while some other random links also incident to v are not owned by v but by some other nodes (which generated these links).

Example 2 (Kleinberg’s grid setting). The undirected version of Kleinberg’s grid model ([16]) is a $\mathcal{P}\mathcal{L}\mathcal{U}\mathcal{S}(\mathcal{H}, \tau, p)$ family, where \mathcal{H} consists of all $n \times n$ grids ($n = 1, 2, 3, \dots$) and where all the weights are 1, $p = 1$, and τ is the inverse square distribution: $v \xleftarrow{R_\tau} u$ with $Pr[v \xleftarrow{R_\tau} u] \propto d^{-2}(u, v)$, where $d(u, v)$ is the lattice distance between u and v .

Note that we use a *general node-based* distribution τ which can be non-uniform while the Chung-Lu model uses a *specific pair-based* distribution which is uniform within a subset of nodes of the same weight. Although a bit more complicated sometimes, our node-based approach can help simulate dynamic graphs (on adding/removing nodes, weights or links) and preferential-attachment networks, and therefore is better for designing practical networks.

Example 3 (power-law). Consider $G = \mathcal{P}\mathcal{L}\mathcal{U}\mathcal{S}(H, \tau, p)$ where $H = (V, W, E)$ with $E = \emptyset$, i.e. a set of weighted nodes, and where $Pr[v \xleftarrow{R_\tau} u] \propto W_v$. It is not hard to see that the expected degree of a node u is $2pW_u$ (roughly, half generated by u itself and the other half from other nodes). Similar as with the Chung-Lu model, this setting can simulate a power-law graph by using power-law weights.

Notions on weight density. Define the weight of a vertex set S as $\sum_{u \in S} W_u$ for $\mathcal{P}\mathcal{L}\mathcal{U}\mathcal{S}(\mathcal{H}, \tau)$, and as $\sum_{u \in S} pW_u$ for $\mathcal{P}\mathcal{L}\mathcal{U}\mathcal{S}(\mathcal{H}, \tau, p)$. For $k > 0$, a node u is (k, t) -dense if it is in some connected component (of H) which has size at most k and weight at least t , i.e. a locality with a certain wealth of random links. A set of base graphs \mathcal{H} is (almost) t -dense if there exists a constant $c > 0$ such that for each $H \in \mathcal{H}$ large enough, (almost) all nodes in H are (ct, t) -dense.

⁴ Modeling with different values of p are particularly needed in the theorem 1’s proof in appendix ??

2.2 Basic ideas in our approach for diameter analysis.

As mentioned early in the introduction, we partition our graph into pieces where each piece has short paths between almost all node pairs. We then consider a super-graph, with a super-node for each such piece, and there is a link between two super-nodes if there is one connecting their counter-part pieces. This super-graph is dense enough such that it is favorable in term of short paths, compared to a random graph which has a constant diameter: these super-nodes are only $O(1)$ links from each other. Thus the length of a typical path in our original graph is at most a constant multiple of the typical path length in a single piece. We present this ‘connecting pieces’ technique in §3.

Other issues are on constructing these pieces, which has a small typical path but a large enough size. We can construct such a piece as a ball of neighbors, for a class of random link distributions with our expansion property, defined in §2.3 below. For settings without such expansion property, we propose a further decomposition on the graph where a fractal-based analysis is introduced in §4.3.

2.3 Fair random link distribution and exponential neighborhood expansion.

We now review important concepts that we use from [22].

Definition 2. 1) (*Expansion property*) Consider a family $\mathcal{F} = \mathcal{PLUS}(\mathcal{H}, \tau, p)$. For constants $\mu > 0$ and $\xi > 0$, distribution τ has ‘the (μ, ξ) expansion property’ if a random link leaves a vertex set of size n^μ with probability at least ξ . More formally, $\forall H = (V, E) \in \mathcal{H}$, $n = |V|$:

$$\forall u \in V, \forall \mathcal{C} \subset V, |\mathcal{C}| < n^\mu : \Pr[v \stackrel{R_\tau}{\leftarrow} u : v \notin \mathcal{C}] \geq \xi$$

We also say, distribution τ or family \mathcal{F} (with τ as a part of it) is (μ, ξ) -expansion.

2) Let m_τ be the minimum value of $\Pr[R_\tau(u) = v]$ for all $u \neq v$, then distribution τ (or family \mathcal{F}) is μ -fair if m_τ is at least $\Omega(n^{-\mu})$.

If satisfied, these properties assure that the distribution of random links is fair and diverse enough (not too skewed, in other words): for any given node u , no small set of vertices takes a dominant role in attracting u ’s links and also no single node is ‘ignored’ (is connected to u with at least a probability $m_\tau = \Omega(n^{-\mu})$). Many existing networks feature these properties.

Example 4. Watts and Strogatz’s small-world networks are (μ, ξ) -expansion for any constants $\mu, \xi \in (0, 1)$, and 1-fair.

Example 5. Kleinberg’s grid-based [16] and tree-based [17] small-world models are $(\mu, 1 - \mu - o(1))$ -expansion for any constant $\mu \in (0, 1)$, and are $(1 + \epsilon)$ -fair for any constant $\epsilon > 0$

The expansion property assures an exponential growth in neighbor sets until a large enough size. Let $S_k(u)$ denotes the set of nodes within k links from u .

Lemma 1 (Expansion lemma). Consider $\mathcal{F} = \mathcal{PLUS}(\mathcal{H}, \tau, p)$. Suppose that \mathcal{F} has the (μ, ξ) -expansion property and the base graphs are almost $(\frac{1}{\xi} + \epsilon)$ -dense for any constant $\epsilon > 0$. There exists a constant $c > 0$ such that for a given node u in a connected component of dense enough weight, w.h.p., $|S_{c \log n}(u)| \geq n^\mu$.

Lemma 1 (proof in appendix) shows a general sufficient condition on when and how we can construct a neighbor ball with logarithmic diameter. The expansion property allows an exponential growth before reaching a threshold (n^μ), if we start with a local connected component with dense enough weight.

3 General conditions for short paths using ‘connecting pieces’

For brevity, in this and the next sections we often work with a random graph $G = \mathcal{PLUS}(H, \tau, p)$ with a specific base graph $H(V, W, E)$. For asymptotic bounding, we vary this base graph H (from a collection \mathcal{H}) with $n = |V|$ going to infinity.

Connecting pieces. As mentioned above, from each connected component of dense enough weight, w.h.p., we can grow a neighborhood of size at least n^μ with diameter $O(\log n)$. So, a giant component with $O(\log n)$ diameter will emerge in our random graph, if we can show that almost surely any two (disjoint) such pieces can be linked by $O(1)$ links. This ‘linking’ notion is formalized by the following ‘connecting lemma’.

Lemma 2 (Connecting lemma). *Let $G = (V, E)$ be a graph with $n = |V|$ and $m, D \in \mathbb{N}$ such that V can be partitioned into \hat{n} disjoint subsets $V_1, V_2, \dots, V_{\hat{n}}$, where $\hat{n} = \theta(\frac{n}{m})$, each subset with size $\theta(m)$, and the subgraph induced by V_i has diameter at most D . Now using a distribution τ , for any z such that $znm_\tau = \Omega(\hat{n}^\epsilon)$ for some $\epsilon > 0$, if we generate at least z independent random links from each set V_i , then the new graph has diameter $O(D)$ w.h.p.*

Proof. We collapse the subsets V_i 's into super-nodes and consider the new graph H of size \hat{n} where each node generate z random links, including possible loops. Given any two super-nodes, the probability that a given random link from one node fails to reach the other is $P = 1 - |V|m_\tau \leq e^{m_\tau\theta(m)}$. Thus, the probability of a new link between two given super-nodes is at least $p' = 1 - P^z \geq 1 - (e^{m_\tau\theta(m)})^z = \Omega(zmm_\tau)$.

We say the diameter of (random) graph G_1 is dominated by the diameter of (random) graph G_2 , denoted as $G_1 \leq_D G_2$, if the former is not greater than the latter w.h.p. We show that the diameter of H is dominated by that of a ((Erdős and Rényi's)) random graph $K = G(\hat{n}, \hat{p})$, where $\hat{p} = \theta(p')$. Note that, for a given constant d , K has diameter d w.h.p. if $(\hat{n}\hat{p})^{d-1}/\hat{n} \rightarrow 0$ and $(\hat{n}\hat{p})^d/\hat{n} \rightarrow \infty$ [15, 3], which is easily verified by $d = \lfloor 1 + \frac{1}{\epsilon} \rfloor$. Thus we just need to prove $H \leq_D K$ for a proper constant in $\hat{p} = \theta(p')$.

We show that by a few of more refined graph comparisons: $H \leq_D I_1 \leq_D I_2 \leq_D K$. First, clearly $H \leq_D I_1$ where I_1 is a modification of H where each node generates exactly z random links, which goes to another given node with probability exactly P , or becomes a loop with probability $1 - (\hat{n} - 1)P$. Now let I_2 be a (random) graph of \hat{n} nodes, where from each node u we generate $Z = c_1(znm_\tau)$ uniform random links, i.e. each go to any given node with equal probability $1/\hat{n}$, where $c_1 > 0$ is a constant to be chosen later. Let K be a random graph $K = G(\hat{n}, \hat{p})$, where $\hat{p} = c_1c_2(zmm_\tau) = \theta(p')$ and $c_2 > 0$ is a constant. Excluding the self-loop links, all the three I_1, I_2 and K have expected degree $\theta(znm_\tau)$, although I_2 has a constant degree Z while K has Poisson degrees and I_1 has degrees as sum of two different Poisson distributions (for 'out-' and 'in-' random links). Using Chernoff bound, by choosing c_1 and c_2 small enough, all the degrees of I_1 are greater than Z w.h.p. and all the degrees of K are smaller than Z (I_1 's constant degree) w.h.p. This results in that $I_1 \leq_D I_2 \leq_D K$, since all these I_1, I_2 and K use uniform random links ⁵. Thus, $H \leq_D K$ where $K = G(\hat{n}, \hat{p})$, where $\hat{p} = \Omega(zmm_\tau)$. \square

Our general setting, conditioned on a few reasonable criteria, can feature a broad class of random structures where paths of logarithmic length are typical.

Definition 3 (AlmostSW family). *A $\mathcal{PLUS}(\mathcal{H}, \tau, p)$ family is AlmostSW if there exist $\mu, \xi \in (0, 1)$ such that \mathcal{F} is (μ, ξ) -expansion, $(1 + \mu - \epsilon_1)$ -fair and each base graph $H \in \mathcal{H}$ is almost $(\frac{1}{\xi} + \epsilon_2)$ -dense for any constants $\epsilon_1, \epsilon_2 > 0$.*

The conditions on the expansion property and the weight density allow to construct large enough neighborhoods each has logarithmic diameter. The other fairness condition assures to 'connect pieces together'. All the settings in examples 2-5 are AlmostSW families for proper parameters; we show that for a few of these settings in §3.1.

Theorem 1. *An AlmostSW family almost surely has a giant connected component with size $n(1 - o(1))$ and diameter $O(\log n)$.*

The idea is to split the weights into two parts: the first part is used to construct pieces of size n^μ as balls of neighbors (lemma 1), then the second part can be used to connect these pieces, using paths of $O(1)$ links between any two pieces – now nodes in our super-graph (connecting lemma). However, a naive approach in constructing these neighbor balls may result in overlapping parts between these balls and violate the requirement for independent random links in our super-graph. We deal with this as below. When the base graph is connected, we use the shortest path distance d (using number of arcs) as a metric and create an L -net $X = \{u_1, u_2, \dots, u_k\}$ in the metric space (V, d) , for an appropriate L (see §4.1 below for detail). Now we choose the nodes in X to be centers of our neighbor balls: we partition V

⁵ For example, an instance (a graph after the links are fixed) in K can be included as a subgraph in an instance in I_2 : when we generate I_2 , we can, initially, create uniform links between pairs of nodes as with K . With high probability, we never use up the 'quota' that allows Z random link generated at each node in I_2 . In fact, we can go on to generate a few more random link at each node, but on the node-based mode instead.

into k disjoint subsets $\{A_i\}_{i=1}^k$, by putting any node $v \in V$ into A_i if u_i (among these k center nodes) is the closest to v (handle ties arbitrarily). We choose L so that our neighbor balls are large enough. A concrete proof for this connected case is below.

In the unconnected case ($d(u, v) = \infty$ for some u, v), the L -net argument can not be used, however a similar construction can still be used to construct the A_i in a greedy fashion. Note that the giant component will include most nodes, except some isolated small component - ‘islands’ (with size $o(\log n)$). That is, almost all shortest paths have length only $O(\log n)$.

Proof (for the connected case). Consider a random graph $G = \mathcal{PLUS}(H, \tau, p)$ for a connected base graph $H(V, W, E)$. We now use the intuition that node u is assigned a fixed number of seeds W_u , where each seed gives birth to a random link with probability p . Choose $\hat{p} \in (0, 1)$ such that $\hat{p}(\frac{1}{\xi} + \epsilon_2) > \frac{1}{\xi} + \frac{\epsilon_2}{2}$. As mentioned above, our idea is to split the weights and hence, we generate G in two stages accordingly, where only the first part of weights is in the early stage. In order to do split the weights, in the first stage we do this random pre-processing: each seed can be used in this stage with probability \hat{p} , otherwise it needs to wait for the second stage (but in both cases, it may also fail to generate a random link with probability p). Let G_1 be the graph obtained after the first state; clearly $G_1 = \mathcal{PLUS}(H, \tau, \hat{p}p)$ ⁽⁶⁾.

We now can assume $p = 1$ without loss of generality, and consider $G_1 = \mathcal{PLUS}(H, \tau, \hat{p})$. We partition V into disjoint subsets $\{V_i\}_{i=1}^m$ (where $|V| = \cup_{i=1}^m |V_i|$) as follows. From lemma 1, there exists a constant c such that for any node u which is $(O(t), t)$ -dense, where $t = \frac{1}{\xi} + \frac{\epsilon_2}{2}$, with high probability, $|N_{c \log n}(u)| \geq n^\mu$. Use metric function d as the shortest path distance (using number of arcs) and let $X = \{u_1, u_2, \dots, u_k\}$ be a $(2c \log n)$ -net in metric space (V, d) . We partition V into disjoint subsets $\{A_i\}_{i=1}^k$, $|V| = \sum_{i=1}^k |A_i|$, by putting any node $v \in V$ into A_i if u_i (among these k center nodes in X) is the closest to v (handle ties arbitrarily). Note that, for any A_i , almost surely $|A_i| \geq n^\mu$, and any two nodes in A_i are at most distance $4c \log n$ apart. For any A_i with $|A_i| > n^\mu$, we chop it into pieces of size n^μ and possibly one of size $< n^\mu$. Now our partition $\{V_i\}_{i=1}^m$ is the collection of all the full or chopped-up pieces from the A_i ’s.

Consider the V_i ’s which have size n^μ . Clearly, (with high probability) the union, V' , of these is at least a constant fraction of V . We now go to the second stage and release all the remaining seeds which we hold off in the first stage (these occupy an expected $(1 - \hat{p})$ fraction of the whole). By doing so, clearly, each V_i' has $\Omega(n^\mu)$ fresh weight, which are now used for our ‘connecting pieces’. Apply the connecting lemma to $V' = \cup V_i'$, by having: $m = \theta(n^\mu)$, $\hat{n} = \theta(n^{1-\mu})$, $m_\tau = \Omega(n^{-(1+\mu-\epsilon_1)})$, and hence, $znm_\tau = \Omega(n^{\epsilon_1}) = \Omega(\hat{n}^{\frac{\epsilon_1}{1-\mu}})$. Thus, with high probability, there exists a (giant) connected component with node set V' and diameter $O(\log n)$. Note that the other pieces from $V - V'$ are also linked to this giant component by $O(\log n)$ links; so, if it is connected the whole graph has expected diameter $O(\log n)$. \square

3.1 Examples using our general model.

Kleinberg’s tree-based setting [16]. In this setting, nodes are the leaves of a complete b -ary tree T ⁽⁷⁾, where b is a constant. Let $h(u, v)$ denote the height of the least common ancestor of u and v in tree T . There are no local links in this setting (so tree T is only for computing $h(u, v)$ and not included in the structure) but there are a number of undirected random links, following a distribution τ such that: $Pr[v \xleftrightarrow{R_\tau} u] \propto b^{-h(u, v)}$.

Corollary 1. *In Kleinberg’s undirected tree-based setting [17], if each node generates at least expected $1 + \epsilon$ random links (for any $\epsilon > 0$) then there exists a giant component which contains most of the nodes.*

Corollary 1 improves on a prior result in [22], where at least 3 random links per node were needed. The same result also applies to Kleinberg’s grid model in [16] without local links. We need to show that these settings are *AlmostSW*, then by theorem 1, a giant component emerges with high probability

⁶ Note to distinguish this two-stage generation of G from the superposition of these two random graphs: $G_1 = \mathcal{PLUS}(H, \tau, \hat{p}p)$ and $G_2 = \mathcal{PLUS}(H, \tau, (1 - \hat{p})p)$

⁷ One motivation for this is for a hierarchical semantic model of searched subjects in the Internet.

even if there is only expected $(1 + \epsilon)$ random links generated by each node, for any constant $\epsilon > 0$ ⁸. A similar result is well known in percolation graphs (e.g. see [2] for a review) where the random links are pair-based.

Proof. We show the setting is an *AlmostSW* then use theorem 1. As mentioned in example 5, this setting is (μ, ξ) -expansion for any $\mu \in (0, 1)$ and $\xi = 1 - \mu - o(1)$. Also, it is easy to see that $m_\tau = \theta(\frac{1}{n \log n})$, i.e. the setting is $(1 + \epsilon_3)$ -fair for any $\epsilon_3 > 0$. Thus, for any $\epsilon > 0$, we can choose μ close enough to 0, hence any $\xi \in (0, \mu)$, such that $\frac{1}{\xi} < 1 + \epsilon$. It is not hard to show that most nodes are $(1 + \epsilon)$ -dense (unless a node is in an isolated component of size $o(\log n)$). Then, the setting satisfies all the 3 conditions of an *AlmostSW* family for appropriate $\epsilon_1, \epsilon_2 > 0$. \square

Arbitrary expected degrees with degree-bias random links.

Corollary 2. *The power-law graph in example 3 has diameter $O(\log n)$ w.h.p.*

Proof. The expected degree of each node u is $2pW_u$ (approximately, half generated by u itself and the other half from other nodes). For a predetermined power-law weight function which has exponent $\in (2, 3)$ (this range is observed in different Internet networks), it is not hard to see that a constant fraction of the vertices have expected degree at least 3, and therefore, most of these nodes will join a connected component with diameter $O(\log n)$ ⁹. Most of the low-degree nodes can also join this component directly or by a short chain. \square

A result similar to corollary 2 is in [5]¹⁰.

4 Bounded-growth, distance-bias, and a fractal-based analysis

We review basic concepts we use on metrics and bounded growth (see e.g. [26, 11] for more background), then present our results on a refined model where a bounded-growth graph is augmented by distance-bias random links. We finish by discussing our future work on applications to network design, suggested by our fractal-based analysis. For brevity, most results in this sections are stated for $\mathcal{PLUS}(H, \tau)$ although they can be naturally extended for $\mathcal{PLUS}(H, \tau, p)$.

4.1 Metrics and growth bounded graphs.

The pair (V, d) , including a set of nodes V and a distance function $d : V^2 \rightarrow \mathbf{R}^+$ is a *metric space*, and d is a *metric*, if for any $u, v, w \in V$, $d(u, v) = 0 \Leftrightarrow u = v$, $d(u, v) = d(v, u)$ (symmetry) and $d(u, v) \leq d(u, w) + d(w, v)$ (triangle inequality). We define $N_r(u) = \{v \in V | d(u, v) \leq r\}$, the subset of nodes within ‘distance’ r from u .

For $r > 0$, a subset $X \subset V$ is an *r-cover* of V if any node u in V is within distance r from at least one node $v \neq u$ in X . A subset $X \subset V$ is an *r-packing* if any two nodes in X are at least distance $2r$ apart. A subset $X \subset V$ is an *r-net* if X is both an *r-cover* and an $\frac{r}{2}$ -packing. It is well known that there exists an *r-net* for any metric space, which can be constructed greedily [26]. For any connected undirected graph, the shortest path distance (using the number of arcs) is clearly a metric, and there exists an *r-net* with respect to this distance.

For a node u in G , define $R_u = \min\{r : N_r(u) = V\}$, the maximum distance of any node from u . For a constant $\Delta > 0$, a graph G with metric d is *growth bounded (GB) below* by growth degree Δ , or is GB_Δ , if

$$\forall u \in V, \forall r \leq R_u/2, \Delta |N_u(r)| \leq |N_u(2r)|$$

⁸ E.g., if each node generates at least one but $\theta(n)$ nodes generate two random links each; also if each node generates one random link and there are $\theta(n)$ local links in the base graph.

⁹ For any vertex sets of size at most $o(n)$ (even with only top degree nodes) there is a probability at least some constant $\xi > 0$ that a random link escapes from this set (e.g. comes to one of the low degrees).

¹⁰ There, Chung and Lu gave a thorough analysis to show that their power law graphs have diameter $O(\log n)$ w.h.p., and the average distance between any two nodes is shorter: $O(\log \log n)$.

Also, for $\bar{\Delta} > 0$, a graph G with metric d is *GB above* by growth $\bar{\Delta}$, or is $GB^{\bar{\Delta}}$, if

$$\forall u \in V, \forall r, |N_u(2r)| \leq \bar{\Delta} |N_u(r)|$$

Intuitively, nodes from V “come into view” at a constant rate (especially, when $\Delta \approx \bar{\Delta}$) when we expand a ball around any node $u \in V$. Similar notions were used in e.g. [25]. For some constant $\beta > 0$, it is easy to see that $GB_{2\beta}$ implies $|N_u(r)| = \Omega(r^\beta)$ and also $GB^{2\beta}$ implies $|N_u(r)| = O(r^\beta)$ ($\forall u \in V$). Note that the reverse direction is not true, i.e., $GB_{2\beta}$ is stronger than having $|N_u(r)| = \Omega(r^\beta)$. As before, a graph has *growth degree* β if for r large enough, $\forall u \in V, |N_u(r)| = \theta(r^\beta)$.

4.2 Expansion from a bounded-growth neighborhood.

Definition 4 (*DistBias family*). For $\alpha, \beta > 0$, a $DistBias(\alpha, \beta)$ family is a $\mathcal{PLUS}(\mathcal{H}, \tau, p)$ family where each base graph $H \in \mathcal{H}$ has an associated metric function d s.t. H is $GB_{2\beta}$ w.r.t d , and $Pr[R_\tau(u) = v] \propto d^{-\alpha}(u, v)$. We use $DistBias(\alpha)$ for $\alpha = \beta$.

Similarly, $DistBias(\alpha, \beta_1, \beta_2)$ is also defined with H being both $GB_{2\beta_1}$ and $GB^{2\beta_2}$.

Example 6. Kleinberg’s two-dimensional grid model (example 2) is $DistBias(2)$; Kleinberg’s tree-based setting (appendix 3.1) is $DistBias(1)$ with $d(u, v)$ as the size of the minimum subtree containing both u and v .

Theorem 2. For each α and β , $0 < \alpha \leq \beta$ there exists a constant $q > 0$ such that a $DistBias(\alpha, \beta)$ family using q -dense base graphs is *AlmostSW*.

Informally, for exponent α small enough ($\leq \beta$), many random links are to distant nodes; so, the expansion property occurs. If instead $\alpha > \beta$, most random links are concentrated within a small neighborhood and there is no expansion property. For example, consider the 1-D grid setting ($\beta = 1$). For $\alpha > 1$, for any constant $\mu > 0$, the probability to get from a source node u to distance at least $m = n^\mu$ from u , is $\theta(\sum_{k=m}^n 1/k^\alpha) = O(\frac{1}{m^{\alpha-1}})$, which tends to 0 when n goes to infinity, violating the expansion property.

See appendix §A.2 for theorem 2’s proof. A key part is showing the graphs have the expansion property. This depends on bounding $Pr[v \stackrel{R_\tau}{\leftarrow} u : m \leq d(u, v) \leq M]$, the probability a random link from a node u to a distance between given distance values m, M where $1 \leq m \leq M \leq R_u = \min\{r : N_r(u) = V\}$. The nodes between distances m and M can be partitioned into several layers divided by distances 2^k for appropriate $k \in N$ (using balls with center u and these exponential radiuses 2^k). We then bound the probability from u to the nodes in each layer, using our definitions on bounded-growth.

Corollary 3. For the Internet topology (e.g. at AS-level), if $\alpha \approx 1$ (as observed in [28]) and β is between 1 and 2 (as in [8, 28]), then the Internet topology has poly-log diameter.

4.3 On a strong distance-bias regime: a fractal perspective.

We now extend our small-world settings when the expansion property doesn’t apply: a random link is unlikely to escape from a vertex set of moderate size (the probability of escaping a set of size n^μ tends to zero when n goes to infinity, for any fixed constant $\mu \in (0, 1)$). Because links stay in a local neighborhood, we can’t partition the graph into large neighborhoods each with small diameter and then link them together by paths of $O(1)$ links. Instead this process of ‘partition-then-connect’ is multiply repeated, recursively, from the very local to global levels. With the parameter scheme introduced below in theorem 3, we can create a fractal picture instead: the whole graph is an $O(1)$ -diameter super graph of nodes representing pieces, and each piece itself is a smaller-scale $O(1)$ -diameter super graph, and so on.

Intuitively, one can think of an online map with different zoom levels (country, state, county, district, etc). For example, a typical short $s - t$ route in this setting consists of $O(\log n)$ super-links between super ‘state’ node at the highest zoom-out level. However, when we point and further zoom-in a super ‘state’ node, we unfold a sub-path of $O(\log n)$ links between super ‘county’ nodes. Figure 1 traces such a route, unfolding to the highest zoom-in at both the source and destination sites (s and t).

Theorem 3. For a $\text{DistBias}(\alpha, \beta_1, \beta_2)$ family where $0 < \alpha < 2\beta_1, \beta_1 < \beta_2$, there exists a poly-log function $L = L(n)$ such that if the base graphs are L -dense, then w.h.p., the graphs have expected diameter $O(\log^\gamma n)$ for $\gamma = \gamma(\alpha, \beta_1, \beta_2)$. In fact, γ increases if α or β_2 increases, or β_1 decreases¹¹.

Proof (outline). Consider random graph $G = \mathcal{P}\mathcal{L}\mathcal{U}\mathcal{S}(H, \tau)$ for a base graph $H(V, E)$ and an associated metric d . We first partition the graph into sets which (after contraction) form an $O(1)$ -diameter super-graph, w.h.p. For any subset $B \subset V$ define $\text{MetricDiam}(B)$ as the maximum metric distance between any two nodes in B ¹². Let $\Delta = \text{MetricDiam}(V)$ and construct a δ -net, $\delta = \Delta^\xi$, of H , for any $\xi \in (\frac{\alpha}{2\beta_1}, 1)$. As in theorem 1's proof, we can partition V into a set of disjoint neighborhoods V_i , by using the nodes in this δ -net as centers, where each piece V_i has $\delta < \text{MetricDiam}(V_i) < 2\delta$. Using the ‘connecting pieces’ technique, it is not hard to see that the super-graph formed by contracting the pieces has expected diameter $O(C)$ where $C = \frac{(1-\xi)\beta_2}{\xi\beta_1}$: there are at most $O(\Delta^{(1-\xi)\beta_2})$ super nodes¹³ each has weight at least $\theta(\Delta^{\beta_1})$, and the probability of a new link connecting two given nodes is almost 1¹⁴.

This process of partitioning then ‘connecting pieces together’ can be repeated recursively at smaller scales (until reaching neighborhoods of size $\theta(L)$). It is not hard to see that the number of partitioning levels is $m = O(\log \log n)$. A probabilistic recurrence approach [14] can be used to show that the graph diameter is LC^m , a poly-log function of n , with high probability (assured by a large enough L). A concrete treatment on a similar probabilistic recurrence can be found in [22]. \square

Example 7. Consider a wireless network where growth rate varies from 1 to 3 (e.g. in example 1). If we add one random link to each node, using distribution τ where $\Pr[v \xrightarrow{R\tau} u] \propto d^{-\alpha}(u, v)$ for $\alpha < 2$, then the new graph has poly-log expected diameter.

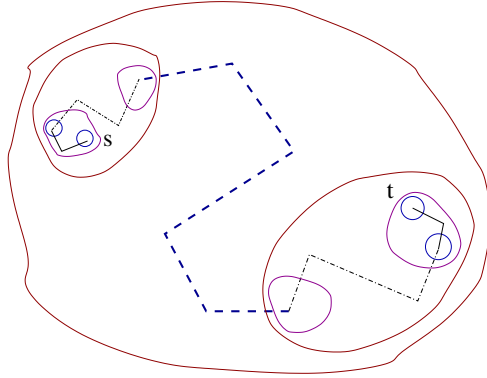


Fig. 1. Tracing a short route in our fractal-based analysis.

Here, we use links of different levels (unfolding super-nodes at different depth): super-links (dashed) of a super graph, or physical (solid) of the actual network. The (super-) links within a sub-graph (super-node at the next zoom-out level) create a path of $O(\log n)$ length. Note that each super-node (collapsed sub-graph) hides within it a chunk of actual physical path, which may have significant length – but still poly-logarithmic.

¹¹ Note some other, easier cases: polynomial diameter (i.e. ‘large-world’ state) for $\alpha > 2\beta_2$, either poly-log or polynomial for $2\beta_1 < \alpha \leq 2\beta_2$ or $2\beta_1 \leq \alpha < 2\beta_2$. The special case that $\frac{\alpha}{2} = \beta_1 = \beta_2$ seems hard and still open. More details are in the full version.

¹² Note we use the metric d to construct the subsets, but our goal is to analyze diameter with respect to number of hops, not the metric d .

¹³ Consider the smallest piece V_j with center u , then the number of pieces $\leq |V|/|V_j| \leq N_\Delta(u)/N_{\delta/2}(u) = O(\Delta^{(1-\xi)\beta_2})$.

¹⁴ If instead $\alpha \geq 2\beta_1$, the probability of having a connecting arc tends to zero and the diameter of the super graph will be at least a poly-log function of n .

4.4 Routing and network design: Initial results and future work.

We observe that our models and random graph constructions provide a natural platform for designing efficient routing networks. We can tailor our routing algorithms to best exploit the topological structures (especially of long-distance links using our fractal-based analysis). Using our refined models (with growth-bounded base graphs and distance-bias random links) we propose to develop: (i) an efficient routing strategy and (ii) a mechanism to improve network performance by changing construction parameters. Based on these, we consider two new applications in (iii) hybrid wireless networks and (iv) optical networks.

A new general routing strategy. Kleinberg's work in [16] and follow-up work e.g. [17, 20, 9] show that greedy routing and similar strategies (using limited local information) only work efficiently in a specific case of our general model when $\alpha = \beta$ ¹⁵. However, our fractal-based analysis and theorem 3 naturally suggest *a new general routing strategy based on a new partition scheme*. In our partition scheme, a graph is recursively partitioned such that a parent block (sub-graph) of metric diameter Δ is partitioned into child blocks of metric diameter Δ^μ , for a fixed constant exponent $\mu \in (\frac{\alpha}{2\beta_1}, 1)$. This hierarchy has only a poly-log number of levels and yet, is far less steep than a b -arc tree (for constant $b \geq 2$) where nodes near the root can be overloaded and highly congested. Thus, we suggest a *hierarchical routing strategy* which is very efficient for a large class of our *distance-bias* structures: a short $s-t$ route within a parent block can be formed by combining a few sub-routes within a few sub-blocks. In our initial results, this routing scheme uses a *small distributed routing database and finds routes of poly-log expected length*.

Hybrid wireless ad-hoc networks. We consider network constructions for building hybrid ad-hoc networks by adding a wired infrastructure to an ad-hoc wireless network [12]. Our initial results [24] consider *adding long links to* a grid-like network with uniform traffic demand between any two nodes, where the cost of a link is proportional to its weight (length). Given a budget to buy long links, we choose links so the typical route has a poly-log length, while the congestion ratio (between the most congested and the average links) is as small as possible: by adding $O(1)$ long links to each node we can maintain a *near optimal trade-off* between congestion ratio and weight.

Routing with dynamic capabilities. We consider routing with Quality of Service provisioning in a dynamic network where the links' bandwidth capabilities may change arbitrarily. Our working example is in *optical wavelength division multiplexing (WDM) networks* where the capability of fiber-optic cable is enormous while communication is mainly restricted by limited resources at the hosts (switching devices and other expensive hardware). We propose a new approach with the assumption that links tend to favor closer distances, and improve *adaptivity* which measures the maximum number of sites to be updated upon a single capability change. Our initial result improves the result in [7] to ideal adaptivity $O(1)$ compared to their polynomial scheme.

References

1. I. Abraham, D. Malkhi, and O. Dobzinski, "Land: Stretch $(1 + \epsilon)$ locality-aware networks for dhds," in *Proc. of ACM Symp. on Discrete Algorithms (SODA)*, 2004.
2. I. Benjamini and N. Berger, "The diameter of a long-range percolation clusters on finite cycles," *Random Structures and Algorithms*, vol. 19, no. 2, pp. 102–111, 2001.
3. F. Chung and L. Lu, "The diameter of random sparse graphs," *Advances in Applied Math.*, vol. 26, pp. 257–279, 2001.
4. —, "Connected components in random graphs with given degree sequences," *Annals of Combinatorics*, vol. 6, pp. 125–145, 2002.
5. —, "The average distances in random graphs with given expected degrees," *Internet Mathematics*, vol. 1, pp. 91–114, 2003.
6. —, "The small world phenomenon in hybrid power law graphs," in *Complex Networks*, E. B.-N. et. al., Ed. Springer-Verlag, 2004, pp. 91–106.
7. S. Dolev, E. Kranakis, D. Krizanc, and D. Peleg, "Bubbles: Adaptive routing scheme for high-speed dynamic networks," *SIAM J. Comput.*, vol. 29, no. 3, pp. 804–833, 1999.
8. C. Faloutsos, P. Faloutsos, and M. Faloutsos, "On power-law relationships of the internet topology," 1999.

¹⁵ These routing strategies look for a long link near the current node which can make a great jump toward the destination node, but for $\alpha \neq \beta$ there is often no long nearby link

9. P. Fraigniaud, C. Gavoille, and C. Paul, “Eclecticism shrinks the world,” in *Proc. of ACM Symp. on Princ. of Dist. Comp. (PODC)*, 2004, pp. 169–178.
10. J. Gao and L. Zhang, “Tradeoffs between stretch factor and load balancing ratio in routing on growth restrict graphs,” in *Proc. of ACM Symp. on Princ. of Dist. Comp. (PODC)*, 2004.
11. J. Heinonen, *Lectures on Analysis on Metric Spaces*. Springer Verlag, Universitext 2001.
12. A. Helmy, “Small worlds in wireless networks,” *IEEE Commun. Lett.*, vol. 7, no. 10, pp. 490–492, Oct. 2003.
13. D. Karger and M. Ruhl, “Finding nearest neighbors in growth-restricted metrics,” in *Proc. of ACM Symp. on Theory of Computing (STOC)*, 2002.
14. R. Karp, “Probabilistic recurrence relations,” *Journal of the ACM (JACM)*, vol. 41, no. 6, pp. 1136–1150, 1994.
15. V. Klee and D. Larman, “Diameters of random graphs,” *Canad J. Math.*, vol. 33, pp. 618–640, 1981.
16. J. Kleinberg, “The small-world phenomenon: An algorithmic perspective,” in *Proc. of ACM Symp. on Theory of Computing (STOC)*, 2000.
17. —, “Small-world phenomena and the dynamics of information,” in *Neural Info. Proce. Sys. (NIPS)*, 2001.
18. —, “Complex networks and decentralized search algorithms,” in *Proceedings of the International Congress of Mathematicians (ICM)*, 2006 (to appear).
19. A. Lakhina, J. W. Byers, M. Crovella, and I. Matta, “On the geographic location of internet resources,” *IEEE Journal of Selected Areas in Communications*, vol. 21, no. 6, 2003.
20. C. Martel and V. Nguyen, “Analyzing kleinberg’s (and other) smallworld models,” in *Proc. of ACM Symp. on Princ. of Dist. Comp. (PODC)*, 2004.
21. M. E. J. Newman, “The structure and function of complex networks,” *SIAM Review*, vol. 45, pp. 167–256, 2003.
22. V. Nguyen and C. Martel, “Analyzing and characterizing small-world graphs,” in *Proc. of ACM Symp. on Discrete Algorithms (SODA)*, 2005.
23. —, “Modeling small-worlds with geographical factors: distance-bias & bounded-growth neighborhoods,” Tech. Rep., 2006. [Online]. Available: <http://wwwcsif.cs.ucdavis.edu/nguyenvk//>
24. —, “Designing low cost networks with short routes and low congestion,” to appear in Infocom’06. [Online]. Available: <http://wwwcsif.cs.ucdavis.edu/nguyenvk//>
25. C. G. Plaxton, R. Rajaraman, and A. W. Richa, “Accessing nearby copies of replicated objects in a distributed environment,” 1997, pp. 311–320.
26. K. Talwar, “Bypassing the embedding: algorithms for low dimensional metrics,” in *Proc. of ACM Symp. on Theory of Computing (STOC)*, 2004, pp. 281–290.
27. D. Watts and S. Strogatz, “Collective dynamics of small-world networks,” *Nature*, vol. 393, pp. 440–32, 1998.
28. S. Yook, H. Jeong, and A. Barabasi, “Modeling the internets large-scale topology,” in *Proceedings of the National Academy of Sciences*, vol. 99, 2002, pp. 13 382–13 386.

A More on our general model and analysis results

A.1 Proof (outline) of lemma 1

Lemma 1. Consider $\mathcal{F} = \mathcal{PLUS}(\mathcal{H}, \tau, p)$. Suppose that \mathcal{F} has the (μ, ξ) -expansion property and the base graphs are almost $(\frac{1}{\xi} + \epsilon)$ -dense for any constant $\epsilon > 0$. There exists a constant $c > 0$ such that for a given node u in a connected component of dense enough weight, w.h.p., $|S_{c \log n}(u)| \geq n^\mu$.

Proof. To show lemma 1, the basic idea is to selectively construct a chain of neighbor sets $\{S_k\}$. For any node u , let $core_t(u)$ be a minimum connected component which contains u and has at least weight t . Starting with $S_0 = core_{\frac{1}{\xi} + \epsilon}(u)$, we aim at showing a (probabilistic) lower bound $\gamma > 1$ on the growth rate of this set chain in each early step. That is, the process can be compared to a branching process where the growth rate is (with high probability) at least some constant $\gamma > 1$. Such a chain of neighbor set can be defined as follows using an ‘expansion function’ χ . Given any node $u \in V$, this operation will call operation $R_\tau W_u$ times to obtain u ’s random link contacts then $\chi(u)$ denotes the union of the cores of these W_u random contacts: $\chi(u) = \cup core_j(v)$, where v is a random contact of u and $j = \frac{1}{\xi} + \epsilon$. Also, $\chi(S) = \bigcup_{u \in S} \chi(u)$ for any set S . From an arbitrary initial set $S_0 \subset V$, we construct a *neighbor subset chain* $\{S_k\}$ such that S_k is the nodes can be reached from S_0 using k random links (and $O(k)$ local links).

$$S_{k+1} = \chi(S_k) - \cup_{i=0}^k S_i; k = 1, 2, 3, \dots \text{ (16)}$$

¹⁶ In fact, we can think the random links are all generated at the birth of the graph and our construction only probes necessary nodes and explores the random links ‘owned’ by them.

See [22] for bounding of a similar chain of neighbor sets (which is quite simpler than the one here¹⁷; however, the same basic idea applies for both). \square

A.2 Proof of theorem 2

Theorem 2. For each α and β , $0 < \alpha \leq \beta$ there exists a constant $q > 0$ such that a $\text{DistBias}(\alpha, \beta)$ family using q -dense base graphs is AlmostSW .

Proof. Showing this theorem is essentially showing the expansion property, which heavily based on bounding $\Pr[v \xrightarrow{R_r} u : m \leq d(u, v) \leq M]$, the probability of having a random link from a node u to a distance within two given distance values m, M where $1 \leq m \leq M \leq R_u = \min\{r : N_r(u) = V\}$. For convenience, let $R = R_u$ and $N_k = |N_k(u)|$. Consider this summation $A(m, M) = \sum_{k=m}^M n_k \times \frac{1}{k^\alpha}$, where $n_k = N_k - N_{k-1}$ is the number of nodes at distance k from u . Clearly, $1/A(1, R)$ is the normalized coefficient C_u (of the distribution of the random links at u). So $\Pr[v \xrightarrow{R_r} u : m \leq d(u, v) \leq M] = A(m, M)/A(1, R)$. Intuitively, we see $A(m, M)$ as a measure of attractiveness (to draw a random link from u) of the set of nodes between distances m and M from u .

We work on bounding $A(m+1, M)$ so that it can be bounded above and below by values depending on α, β, m, M (and N_m, N_M). We lower bound $A(m+1, M)$ by relating to $\frac{N_m}{m^\alpha}$ and upper bound it by relating to $\frac{N_M}{M^\alpha}$. The basic technique is by grouping terms in $A(m+1, M)$, each group consists of terms with index between $2^i m$ and $2^{i+1} m - 1$ ($0 \leq i \leq \lfloor \log(\frac{M}{m}) \rfloor$); thus, we can approximate $A(m+1, M)$ using terms with index $2^i m$ only, which later can be bounded based on $\frac{N_m}{m^\alpha}$ or $\frac{N_M}{M^\alpha}$.

We first show that $A(m+1, M)$ can be bounded above and below by values depending on α, β, m, M (and N_m, N_M). We lower bound $A(m+1, M)$ by relating to $\frac{N_m}{m^\alpha}$ and upper bound it by relating to $\frac{N_M}{M^\alpha}$. For $l = \lfloor \log(\frac{M}{m}) \rfloor$, by grouping terms in $A(m+1, M)$,

$$\begin{aligned} A(m+1, M) &= \sum_{i=0}^{l-1} \sum_{j=2^i m+1}^{2^{i+1} m} \frac{n_j}{j^\alpha} + \sum_{j=2^l m+1}^M \frac{n_j}{j^\alpha} \geq \sum_{i=0}^{l-1} \frac{N_{2^{i+1} m} - N_{2^i m}}{(2^{i+1} m)^\alpha} + \frac{N_M - N_{2^l m}}{(2^l m)^\alpha} \\ &\geq \frac{2^\beta - 1}{2^\alpha} \sum_{i=0}^{l-1} \frac{N_{2^i m}}{(2^i m)^\alpha} + \frac{N_M - N_{2^l m}}{(2^l m)^\alpha} \end{aligned}$$

Call the last expression A . Note that we have used $N_{2^{i+1} m} - N_{2^i m} \geq (2^\beta - 1)N_{2^i m}$ in the last inequality. Also note that

$$N_m \leq \frac{N_{2m}}{2^\beta} \leq \frac{N_{2^2 m}}{2^{2\beta}} \leq \dots \leq \frac{N_{2^{l-1} m}}{2^{(l-1)\beta}} \leq \frac{N_{2^l m}}{2^{l\beta}}$$

Therefore, if N_m and N_M are fixed, A gets minimized if $N_{2^i m} = 2^{i\beta} N_m$, $i = 1..l$, and $A(m+1, M) \geq \frac{2^\beta - 1}{2^\alpha} \sum_{i=0}^{l-1} \frac{2^{i\beta} N_m}{(2^i m)^\alpha} + \frac{N_M - 2^{l\beta} N_m}{(2^l m)^\alpha}$. So,

$$A(m+1, M) \geq \frac{N_m}{m^\alpha} \times \left(\frac{2^\beta - 1}{2^\alpha} L + \frac{N_M - 2^{l\beta} N_m}{2^{l\alpha} N_m} \right) \quad (1)$$

where $L = \sum_{i=0}^{l-1} (2^{\beta-\alpha})^i$, which equals $2^{(\beta-\alpha)l} - 1$ if $\beta > \alpha$, and $L = l$ if $\alpha = \beta$.

Similarly,

$$A(m+1, M) = \sum_{i=0}^{l-1} \sum_{j=2^i m+1}^{2^{i+1} m} \frac{n_j}{j^\alpha} + \sum_{j=2^l m+1}^M \frac{n_j}{j^\alpha} \leq \sum_{i=0}^{l-1} \frac{N_{2^{i+1} m} - N_{2^i m}}{(2^i m)^\alpha} + \frac{N_M - N_{2^l m}}{(2^l m)^\alpha}$$

However, we have $N_{2^{i+1} m} - N_{2^i m} \leq N_{2^{i+1} m} \leq 2^{-(l-i-1)\beta} N_M$ and $2^{l+1} m \geq N_M$ so

$$\begin{aligned} A(m+1, M) &\leq \sum_{i=0}^{l-1} \frac{2^{-(l-i-1)\beta} N_M}{(2^i m)^\alpha} + \frac{N_M}{(2^l m)^\alpha} = \sum_{i=0}^{l-1} \frac{2^{-(l-i-1)\beta-\alpha} N_M}{m^\alpha} + \frac{N_M}{(2^l m)^\alpha} \\ &= \sum_{i=0}^{l-1} 2^{i(\beta-\alpha)} \frac{2^{-(l-1)\beta+(l+1)\alpha} N_M}{(2^{l+1} m)^\alpha} + \frac{2^\alpha N_M}{(2^{l+1} m)^\alpha} \leq \frac{N_M}{M^\alpha} \left(2^{-(l-1)\beta+(l+1)\alpha} \sum_{i=0}^{l-1} 2^{i(\beta-\alpha)} + 2^\alpha \right) \end{aligned}$$

¹⁷ There, each node generates q random links for a given fixed integer q , each of which has probability at least ξ to leave the current neighbor set while here, each node generates $Be(W_u, p)$ such random links. However, since we only care about the total of random links which escape the neighbor set, similar Chernoff-bound argument work for both cases.

Thus,

$$A(m+1, M) \leq \frac{N_M}{M^\alpha} \left(2^{-l(\beta-\alpha)+\beta+\alpha} L + 2^\alpha \right) \quad (2)$$

Consider a $\text{DistBias}(\alpha, \beta)$ family using q -dense base graphs. We now need to show that by choosing q big enough, there exist $\mu, \xi \in (0, 1), \xi > \frac{1}{q}$ such that this family is (μ, ξ) -expansion while its distribution τ is $(1 + \mu + \Omega(1))$ -fair.

We now justify the expansion property. For any $\mathcal{C} \in V$ with size n^μ and $0 < \mu < 1$, we consider $\Pr[R_\tau(u) \notin \mathcal{C}]$. Define $m \in \mathbf{N}$ such that $N_{m-1} < |\mathcal{C}| = n^\mu \leq N_m$. For simplicity, we assume $N_m = |\mathcal{C}| = n^\mu$ and, using the observation that a ball is the best shape for \mathcal{C} to minimize the probability $\Pr[R_\tau(u) \notin \mathcal{C}]$, it is easy to see that

$$\Pr[R_\tau(u) \notin \mathcal{C}] \geq \Pr[R_\tau(u) \notin N_m(u)] = \Pr[v \stackrel{R_\tau}{\leftarrow} u : m+1 \leq d(u, v) \leq R_u] = A(m+1, R)/A(1, R).$$

Here we denote $R = R_u = \min\{r : N_r(u) = V\}$. We now claim that $\Pr[R_\tau(u) \notin \mathcal{C}]$ can be lower bounded by a constant (depending on μ and α only). We show that by proving so for the ratio $A(m+1, R)/A(1, m)$ ⁽¹⁸⁾. Using (1),(2), we only need to consider the case $\alpha = \beta$ which is actually the worst case (for $\alpha < \beta$, it is easy to see that $A(m+1, R) \gg A(1, m)$ when n is large enough).

Define $0 < \nu < 1$, such that $m = R^\nu$; clearly, $\log(\frac{R}{m}) = (1 - \nu) \log R$. From (1), we have

$$\begin{aligned} A(m+1, R) &\geq \frac{N_m}{m^\alpha} \times \left((1 - \frac{1}{2^\alpha}) [(1 - \nu) \log R] + \frac{n}{2^\alpha N_m} - 1 \right) \\ &= \frac{N_m}{m^\alpha} \log R \times (C + D) \end{aligned}$$

where $C = (1 - \frac{1}{2^\alpha})(1 - \nu)$ and $D = \frac{1}{\log R} (\frac{n}{2^\alpha N_m} - 2)$.

From (2),

$$A(1, m) \leq \frac{N_m}{m^\alpha} \times (2^{2\alpha} \nu \log R + 2^\alpha) = \frac{N_m}{m^\alpha} \log R \times (E + O(\log^{-1} R))$$

where $E = 2^{2\alpha} \nu$.

Consider the two following cases. If $\nu \leq \eta = \frac{1+\mu}{2}$ then we lower bound $A(m+1, R)/A(1, m)$ by

$$\frac{C}{2E} = 2^{-2\alpha-1} (1 - \frac{1}{2^\alpha}) \frac{1-\eta}{\eta} \quad (3)$$

when n (and then $\log R$) is large enough. If $\nu > \frac{1+\mu}{2}$ then we lower bound $A(m+1, R)/A(1, m)$ by $\frac{D}{2E}$. Here,

$$D = \frac{1}{\log R} \left(\frac{n}{2^\alpha N_m} - 2 \right) \geq \frac{1}{\log R} \left(\frac{n^{1-\mu}}{R^{(1-\nu)\alpha}} - 2 \right) = \frac{1}{\log R} \left(\frac{\Omega(R^{(1-\mu)\beta})}{R^{(1-\nu)\alpha}} - 2 \right) = \Omega \left(\frac{R^{\frac{1-\mu}{2}\alpha}}{\log R} \right)$$

which goes to infinity when R tends to infinity, while E is upper bounded by $2^{2\alpha}$. So $D/2E$ can even be lower bounded by any positive constant for R large enough.

Thus, $\Pr[R_\tau(u) \notin \mathcal{C}]$ is lower bounded by a value $\xi > 0$ depending on μ and α , i.e. the family meets (μ, ξ) -expansion. Thus we need to choose $q > 1/\xi$.

Now we only need to assure that distribution τ is $(1 + \mu + \epsilon)$ -fair, i.e. $\frac{1}{m_\tau} = O(n^{1+\mu+\epsilon})$, for some $\epsilon > 0$ (recall that m_τ is the minimum value of $\Pr[R_\tau(u) = v]$ for all $u \neq v$). Let u and v are the most far apart nodes and $R = d(u, v)$. From (2), $A(1, R) \leq \frac{n}{R^\alpha} (\log R + 2^\alpha)$. So, $m_\tau = \Pr[R_\tau(u) = v] = \frac{1}{A(1, R)} \times \frac{1}{R^\alpha} \geq \frac{1}{n(2^{2\alpha} \log R + 2^\alpha)}$ and $\frac{1}{m_\tau} = O(n \log n)$. Thus τ is even $(1 + \epsilon)$ -fair for any $\epsilon > 0$ (and μ is free to choose for this issue).

Thus we can choose μ close to zero to minimize q , and by (3), can choose any $q > \frac{2^{2\alpha+1}}{1 - \frac{1}{2^\alpha}}$. \square

¹⁸ Note that $A(1, R) = A(1, m) + A(m+1, R)$, so bounding $x/(x+y)$ is equivalent to bounding x/y . If $N_m > n^\mu$, the ratio to be considered is $\frac{A(m+1, R)+W}{A(1, m)-W}$, where $W = \frac{N_m - n^\mu}{m^\alpha}$. It is not hard to see that our proof can be extended for this case also.