The Structure of Geographical Threshold Graphs

Milan Bradonjić, Aric Hagberg, and Allon G. Percus

Abstract. We analyze the structure of random graphs generated by the geographical threshold model. The model is a generalization of random geometric graphs. Nodes are distributed in space, and edges are assigned according to a threshold function involving the distance between nodes as well as randomly chosen node weights. We show how the degree distribution, percolation and connectivity transitions, clustering coefficient, and diameter relate to the threshold value and weight distribution. We give bounds on the threshold value guaranteeing the presence or absence of a giant component, connectivity and disconnectivity of the graph, and small diameter. Finally, we consider the clustering coefficient for nodes with a given degree l, finding that its scaling is very close to $1/l$ when the node weights are exponentially distributed.

1. Introduction

Large networks such as the Internet, World Wide Web, phone call graphs, infectious disease contacts, and financial transactions have provided new challenges for modeling and analysis [Bonato 05]. For example, web graphs may have billions of nodes and edges, which implies that the analysis on these graphs, i.e., processing and extracting information on these large sets of data, is "hard" [Abello et al. 02]. Extensive theoretical and experimental research has been done in web-graph modeling. Early measurements suggested that the Internet exhibits a power-law degree distribution [Faloutsos et al. 99] and that the web graph also follows a power-law distribution in in- and out-degree of links [Kleinberg et al. 99]. Modeling approaches using random graphs have attempted to capture both

the structure and dynamics of the web graph [Kumar et al. 00, Barabási and Albert 99, Aiello et al. 00, Bollobás et al. 01, Cooper and Frieze 01].

The study of random graphs began with the introduction of the uniform random graph model [Erdős and Rényi 59, Erdős and Rényi 60]. Since then, many other models have been proposed to better capture the structure seen in real-world networks [Bollobás 01, Durrett 06]. Some examples are random graph models with a given or expected degree sequence [Molloy and Reed 95, Chung and Lu 06], threshold graphs [Mahadev and Peled 95, Hagberg et al. 06] with edges created according to a function of node weights, and graphs with an underlying geometric structure, such as random geometric graphs [Penrose 03]. In this paper we study another recent addition to this collection of models: geographical threshold graphs (GTGs), a static model for networks that includes both geometric information and node weight information.

GTGs combine the geometric structure of random geometric graphs with node properties similar to threshold graphs. The properties of this graph ensemble have only recently begun to be studied [Masuda et al. 05, Bradonjić et al. 07, Bradonjić and Kong 07]. One motivation for analyzing this model is that many real networks need to be studied with a stochastic model that is "richer" than random geometric graphs. The GTG model has been applied, for instance, in the study of wireless ad hoc networks in systems in which the wireless nodes have different communication ranges or battery power [Bradonjić and Kong 07]. In that case, the weights represent available power or bandwidth of a wireless node. By varying the weights, properties such as the diameter or degree distribution can be tuned. Other possible applications of GTGs that are yet to be explored are epidemic modeling, where the weights could represent susceptibility or infectivity of an individual; and social networks, where the weights might be related to affinity or attractiveness.

2.   Geographical Threshold Graph Model

The GTG model is constructed from a set of n nodes placed independently and uniformly at random in a volume in $\mathbb{R}^d$. A nonnegative weight w_i, taken randomly and independently from a probability distribution function $f(w) : \mathbb{R}_0^+ \rightarrow \mathbb{R}_0^+$, is assigned to each node v_i for $i \in [n]$. Let $F(x) = \int_0^x f(w)\,dw$ be the cumulative density function. For two nodes v_i and v_j at distance r, the edge (i, j) exists if and only if the following connectivity relation is satisfied:

$$G(w_i, w_j)h(r) \geq \theta_n, \tag{2.1}$$

where θ_n is a given threshold parameter that depends on the size of the network. The function $h(r)$ specifies the connection probability as a function of distance and is assumed to be decreasing in r. In the following we take $h(r) = r^{-\beta}$, for some positive β, which is typical, for example, of the path-loss model in wireless networks [Bradonjić and Kong 07]. The interaction strength between nodes $G(w_i, w_j)$ is typically taken to be symmetric (to produce an undirected graph) and either multiplicatively or additively separable, i.e., in the form of $G(w_i, w_j) = g(w_i)g(w_j)$ or $G(w_i, w_j) = g(w_i) + g(w_j)$.

Some basic results have already been established. For the case of uniformly distributed nodes over a unit space it has been shown [Masuda et al. 05, Bradonjić and Kong 07] that the expected degree of a node with weight w is

$$\mathrm{E}[k(w)] = \frac{n\pi^{d/2}}{\Gamma(d/2+1)} \int_{w'} f(w')\big(h^{-1}(\theta_n/G(w, w'))\big)^d dw',$$

where h^{-1} is the inverse of h. The degree distribution has been studied for specific weight-distribution functions $f(w)$ [Masuda et al. 05]. In both the multiplicative and additive cases of $G(w, w')$, questions of diameter, connectivity, and topology control have been addressed [Bradonjić and Kong 07].

Here we restrict ourselves to the case of $g(w) = w$, $\beta = 2$, and nodes distributed uniformly over a two-dimensional space. For analytical simplicity we take the space to be a unit torus, and use the additive model for the connectivity relation

$$\frac{w_i + w_j}{r^2} \geq \theta_n. \tag{2.2}$$

Certain of our techniques may be generalized to other cases in a straightforward manner.

Finally, we impose the following relatively weak conditions on the weight distribution $f(w)$: (1) a finite mean $\mu = \mathrm{E}[w]$, (2) a finite variance $\sigma^2 = \mathrm{E}[w^2] - \mathrm{E}[w]^2$, and (3) nonvanishing support everywhere on $w \in (0, \infty)$.

Some examples of GTG instances with exponential weight distribution $f(w) = e^{-w}$ are shown in Figure 1.

The paper is organized as follows. We first state a basic property concerning the degree distribution of GTGs. In Section 4, Theorems 4.2 and 4.3 provide bounds on θ_n for the absence and the existence of a giant component. Similarly, in Section 5, Theorems 5.2 and 5.3 provide bounds on θ_n for the graph being disconnected and being connected. Section 6 gives upper bounds on the diameter, along with simulation results. Finally, in Section 7 we calculate the clustering coefficient, and discuss certain of its properties.
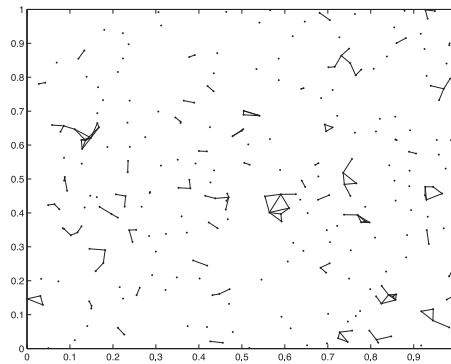
(a)
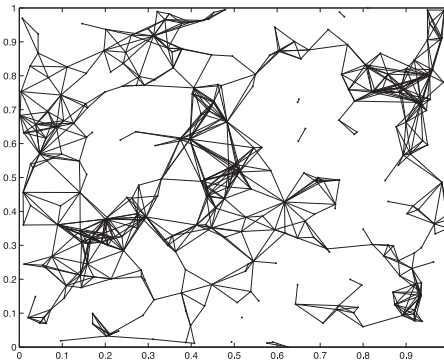
(b)
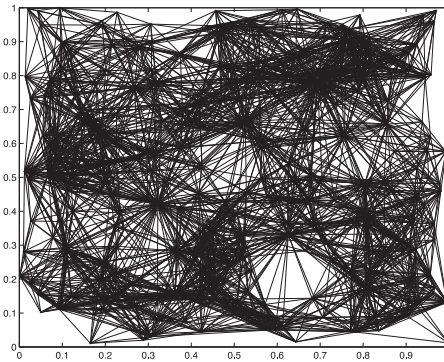
(c)

Figure 1. Instances of GTG with exponential weight distribution for $n = 300$ at decreasing threshold parameter values (increasing mean degree): (a) $\theta_n/n = 2\pi$, well below the percolation transition; (b) $\theta_n/n = 1$, above the percolation but below the connectivity transition; (c) $\theta_n/n = 1/2e$, well above connectivity.

3.   Degree Distribution

We start by stating the degree distribution in our GTG model. Let the position vector of the nodes be $\mathbf{x}$ and the weight vector be $\mathbf{w}$. Without loss of generality, let us consider node v_1. It is straightforward to show that the probability of v_1 having degree k, given weights $\mathbf{w}$, is

$$\Pr[d_1 = k | \mathbf{w}] = \binom{n-1}{k} \prod_{i=2}^{k+1} \text{area}(B(x_i, r_{i1})) \prod_{j=k+2}^{n} (1 - \text{area}(B(x_j, r_{j1}))),$$

where $\text{area}(B(x_i, r_{i1}))$ is the area of the ball with center x_i and radius r_{i1}, and due to (2.2), the radii are given by

$$r_{i1} = \sqrt{\frac{w_1 + w_i}{\theta_n}}$$

for $i = 2, \ldots, n$. After marginalization, it follows that

$$
\begin{aligned}
\Pr[d_1 &= k | w_1] \\
&= \left(\prod_{i=2}^{n} \int_{w_i} f(w_i) dw_i \right) \Pr[d_1 = k | \mathbf{w}] \\
&= \binom{n-1}{k} \left(\int_{w} f(w) \frac{\pi(w_1 + w)}{\theta_n} dw \right)^k \left(1 - \int_{w} f(w) \frac{\pi(w_1 + w)}{\theta_n} dw \right)^{n-1-k} \\
&= \binom{n-1}{k} \left(\frac{\pi(w_1 + \mu)}{\theta_n} \right)^k \left(1 - \frac{\pi(w_1 + \mu)}{\theta_n} \right)^{n-1-k}.
\end{aligned}
\tag{3.1}
$$

That is, the degree distribution d_i of a node v_i with weight w_i follows the binomial distribution

$$d_i(\cdot | w_i) \sim \text{Bin}(n-1, p_i), \tag{3.2}$$

where

$$p_i = \frac{\pi}{\theta_n}(w_i + \mu). \tag{3.3}$$

4.   Giant Component

Definition 4.1. (Giant Component.)   The *giant component* is a connected component of size $\Theta(n)$.

In this section we analyze the conditions for the existence of the giant component, giving bounds on the threshold parameter value θ_n where it first appears.

For $\theta_n = gn$, we specify positive constants $g' > g''$ and prove that whp (with high probability, i.e., with probability $1 - o(1)$), if $g > g'$, the giant component does not exist, whereas if $g < g''$, the giant component exists.

We do not prove that there is a zero–one law governing the emergence of the giant component. However, given that the probability of a giant component is zero above g' and one below g'', it would be rather surprising if the transition were not a sharp one.

4.1.   Absence of Giant Component

Theorem 4.2. *Let $\theta_n = gn$ for $g > g'$, where $g' = 2\pi\mu$. Then whp there is no giant component in GTG.*

Proof. We start by introducing a slightly different GTG model from our usual one. Consider the space of possible node positions and weights, $S = \{(x, y, w) : x, y \in [0, 1], w \geq 0\}$. In the model we have already defined, we place n nodes in S, leading to a binomial degree distribution: call this the *binomial GTG*. Let us now instead place nodes in S according to a spatial Poisson process with rate $nf(w)$, so that the expected number of nodes is n: call this the *Poisson GTG*. We will prove that the Poisson GTG does not have a giant component. It is straightforward to see that if the binomial GTG had a giant component with nonvanishing probability, the analogous Poisson GTG would have one as well. Thus, the binomial GTG cannot have a giant component either.

The proof's approach is similar to one given in [Alon and Spencer 00]. Divide the nodes into three classes: alive, dead, and neutral. Denote the number of alive nodes by Y_i. Now apply the following algorithm. At time $t = 0$, designate one node (picked uniformly at random) as being alive and all others as neutral. At each subsequent time step t, pick a node v_t uniformly at random from among those that are alive, and then consider all neutral nodes connected to v_t. Denote the number of these nodes by Z_t. Change these nodes from neutral to alive, and change v_t itself from alive to dead. The random variables Y_i, Z_i satisfy the following recurrence relation: $Y_0 = 1$ and $Y_t = Y_{t-1} + Z_t - 1$, for $t \geq 1$. The number of alive nodes satisfies

$$Y_t - 1 = \sum_{i=1}^{t} Z_i - t.$$

Since neutral nodes are by definition those that have not yet been explored by the algorithm, the Z_i are independent random variables. We formalize this argument as follows. For a node $v_i = (x_i, y_i, w_i)$, define $S_i \subseteq S$ as the subspace

of all positions and weights of nodes that can be connected to v_i, namely $S_i = \{(x, y, w) : x, y \in [0, 1], w \geq 0, (x - x_i)^2 + (y - y_i)^2 \leq (w + w_i)/\theta_n\}$. At time $t = 0$, any node within S_0 is a neutral node connected to v_0. But at a subsequent time step t, nodes within any S_i for $i < t$ have already been designated alive, so only those in

$$B_t = S_t \setminus \bigcup_{i=0}^{t-1} S_i$$

can be neutral nodes connected to v_t. Thus, the nodes figuring within Z_t and $Z_{t'}$, for any two different time steps t and t', are drawn from *disjoint* subspaces B_t and $B_{t'}$.

The Z_i are simply restrictions of the Poisson process to the B_i. Due to the memoryless nature of the Poisson process, they are independent Poisson random variables. Given that $B_i \subseteq S_i$ and the expected population of S_i is np_i with p_i as defined in (3.3), Z_i satisfies the stochastic bound

$$\Pr[Z_i \geq k] \leq \Pr[\mathrm{Po}(np_i) \geq k].$$

Now consider nodes that are alive, and let T be the largest t such that $Y_t > 0$. Then T is the size of the component containing v_0, and the giant component exists if and only if $T = \Theta(n)$ with some nonvanishing probability. The variable T satisfies

$$\Pr[T \geq t] = \Pr[Y_t > 0] = \Pr[Y_t \geq 1] = \Pr\Big[\sum_{i=1}^{t} Z_i \geq t\Big] \leq \Pr\Big[\sum_{i=1}^{t} \mathrm{Po}(np_i) \geq t\Big].$$

We take the threshold to be $\theta_n = gn$. Since the sum of independent Poisson random variables is itself Poisson distributed, we need to prove that

$$\Pr\Big[\mathrm{Po}\Big(n\sum_{i=1}^{t} p_i\Big) \geq t\Big] \to 0$$

for $t = \Theta(n)$, for some $g > 0$. For any constant $\epsilon \in (0, 1)$, the following inequality holds:

$$\Pr\Big[\mathrm{Po}\Big(n\sum_{i=1}^{t} p_i\Big) \geq t\Big] \leq \Pr\Big[\mathrm{Po}\Big(n\sum_{i=1}^{t} p_i\Big) \geq t \Big| \sum_{i=1}^{t} w_i \in (1 \pm \epsilon)t\mu\Big]$$

$$+ \Pr\Big[\sum_{i=1}^{t} w_i \notin (1 \pm \epsilon)t\mu\Big]. \tag{4.1}$$

We will bound the first term on the right-hand side using the concentration of Poisson random variables [Penrose 03]. To maximize the conditional probability,

set $\sum w_i = (1 + \epsilon)t\mu$, and then let $\lambda = n\sum_{i=1}^{t} p_i = (2 + \epsilon)at\mu$, where $a = n\pi/\theta_n = \pi/g$. Now, given any constant $\gamma \in (0, 1)$, for $t \to \infty$, i.e., $\lambda \to +\infty$, it follows that

$$\Pr[\mathrm{Po}(\lambda) \notin (1 \pm \gamma)\lambda] \leq e^{-\lambda H(1-\gamma)} + e^{-\lambda H(1+\gamma)} \to 0,$$

where the function H is given by $H(x) = 1 - x + x\log x$, for $x > 0$. It is now sufficient to choose g large enough that $t > (1 + \gamma)\lambda$, which occurs when $g > (2+\epsilon)(1+\gamma)\pi\mu$. It follows that for any $g > 2\pi\mu$, ϵ and γ can be set so that the first term on the right-hand side in (4.1) goes to zero.

Now consider the second term on the right-hand side. By the central limit theorem, $(\sum w_i - t\mu)/(\sqrt{t}\sigma)$ tends to the normal distribution $N(0, 1)$ as $t \to \infty$, so

$$\Pr\Big[\sum_{i=1}^{t} w_i \notin (1 \pm \epsilon)t\mu\Big] = \Pr\Big[\frac{\sum w_i - t\mu}{\sqrt{t}\sigma} \notin (-\epsilon, \epsilon)\sqrt{t}\frac{\mu}{\sigma}\Big] \to 0$$

for any constant ϵ.

Thus, for $g > 2\pi\mu$, the probability that T equals $\Theta(n)$ goes to zero, and so there is no giant component. $\square$

4.2.   Existence of Giant Component

Theorem 4.3. *Let $\theta_n = gn$ for $g < g'' = \sup_{\alpha \in (0,1)} \alpha F^{-1}(1 - \alpha)/\lambda_c$, where $\pi\lambda_c \approx 4.52$ is the mean degree at which the giant component first appears in random geometric graphs (RGG) [Penrose 03]. Then whp the giant component exists in GTG.*

Proof.   For any constant $\alpha \in (0, 1)$, we prove that whp there are αn "high-weighted" nodes, all with weights greater than or equal to some s_n; we state s_n later. Let X_i be the indicator of the event $w_i \geq s_n$. Then $\Pr[X_i = 1] = 1 - F(s_n) =: q$. Let $X = \sum_{i=1}^{n} X_i$ be the number of high-weighted nodes. Using the Chernoff bound $\Pr[X \leq (1 - \delta)\mathrm{E}[X]] \leq \exp(-\mathrm{E}[X]\delta^2/2)$, with $\delta = 1 - \alpha/q$, we obtain

$$\Pr[X \leq \alpha n] = \Pr[X \leq (1 - \delta)\mathrm{E}[X]] \leq \exp\big(-n(q - \alpha)^2/(2q)\big) = n^{-\gamma}$$

for some constant $\gamma > 1$ satisfying $(q - \alpha)^2 = 2q\gamma \log n/n$.

Solving this quadratic equation in q gives $q = \alpha + \Theta(\log n/n)$, so $F(s_n) = 1 - q = 1 - \alpha - \Theta(\log n/n)$. For any $\epsilon > 0$ and n sufficiently large, the following inequality is satisfied:

$$F^{-1}(1 - \alpha) \geq s_n \geq F^{-1}(1 - \alpha - \epsilon).$$

Thus, let s_n be a sequence such that its limit satisfies

$$s_n \to F^{-1}(1 - \alpha) = \Theta(1).$$

Now we consider the set of αn high-weighted nodes. For each such node v_i with weight w_i, define its characteristic radius to be

$$r_t^2(w_i) = \frac{w_i}{\theta_n}.$$

Then it follows that any other high-weighted node v_j within this radius is connected to v_i, since the connectivity relation is satisfied:

$$\frac{w_i + w_j}{r^2} \geq \frac{w_i}{r_t^2} = \theta_n.$$

Let $\theta_n = gn$, where $g < \alpha F^{-1}(1 - \alpha)/\lambda_c$. For the radius r_t, whp it follows that

$$r_t^2(w_i) = \frac{w_i}{\theta_n} \geq \frac{s_n}{\theta_n} > \frac{\lambda_c}{\alpha n}. \tag{4.2}$$

Let us therefore consider small circles with a fixed radius r_0 such that $\sqrt{s_n/\theta_n} > r_0 > \sqrt{\lambda_c/(\alpha n)}$ around each of these αn nodes. A subgraph of this must be an RGG with mean degree greater than $\pi \lambda_c$, which whp contains a giant component. Since its size is $\Theta(\alpha n) = \Theta(n)$, it is a giant component of the GTG too. We may optimize the bound by taking the supremum of g over $\alpha \in (0,1)$, and the theorem follows. $\square$

4.3.   Comparison of Upper and Lower Bounds

We again stress that we have not proven a zero–one law for the emergence of the giant component. If a sharp transition does indeed exist, then g' and g'' provide bounds on its location. Here we consider the size of the gap between the two bounds.

Claim 4.4. *For any weight distribution $f(w)$, $g'/g'' \geq 2\pi\lambda_c \approx 9.04$.*

Proof. First consider $g' = 2\pi\mu$. Using the telescope formula, μ satisfies

$$\mu = \int_0^\infty (1 - F(y))dy.$$

Now consider $g'' = \sup_{\alpha \in (0,1)} \alpha F^{-1}(1 - \alpha)/\lambda_c$. We have $F : [0, +\infty) \to [0, 1)$. Since F is a bijection, the inverse $F^{-1} : [0, 1] \to [0, +\infty)$ exists. Let $x = F^{-1}(1 - \alpha)$, and consequently $\alpha = 1 - F(x)$. Then

$$\sup_{\alpha \in (0,1)} \alpha F^{-1}(1 - \alpha) = \sup_{x \in (0,\infty)} x(1 - F(x)).$$

Define the function

$$h(x) = \int_0^x (1 - F(y))dy - x(1 - F(x)).$$

Since $h'(x) = xf(x) \geq 0$ and $h(0) = 0$, we know that $h(x) \geq 0$ for every $x \geq 0$. Let x_0 be the value at which $x(1 - F(x))$ has its supremum. Then,

$$\mu - \sup_{\alpha \in (0,1)} \alpha F^{-1}(1 - \alpha) = \int_0^\infty (1 - F(y))dy - x_0(1 - F(x_0)) \geq h(x_0) \geq 0,$$

from which the claim follows.   □

Remark 4.5. For the exponential distribution $f(w) = \gamma \exp(-\gamma w)$, we have $g' = 2\pi/\gamma$.

Remark 4.6. If $\alpha F^{-1}(1 - \alpha)$ has an extremum for $\alpha \in (0, 1)$, this occurs at

$$\alpha = F^{-1}(1 - \alpha)f(F^{-1}(1 - \alpha)).$$

For example, for the exponential distribution the maximum is at $\alpha = 1/e$, giving a bound of $g'' = 1/e\gamma\lambda_c$.

5.   Connectivity

Definition 5.1. (Connectivity.)   The graph on n vertices is *connected* if the largest component has size n.

In this section we analyze conditions for connectivity, giving bounds on the threshold parameter θ_n at which the entire graph first becomes connected. Similarly to our approach in the case of the giant component, for $\theta_n = \kappa n/\log n$, we specify positive constants $\kappa' > \kappa''$ and prove that whp, if $\kappa > \kappa'$, the graph is disconnected, whereas if $\kappa < \kappa''$, the graph is connected.

As in the case of the emergence of the giant component, it seems likely but has not been proven that there is a sharp phase transition at which GTGs become connected.

5.1.   Disconnected Graph

In this subsection we prove the conditions under which a graph is disconnected.

Theorem 5.2. *Let $\theta_n = \kappa n/\log n$ for $\kappa > \kappa'$, where $\kappa > \pi\mu$. Then the GTG is disconnected whp.*

Proof. For a node v_i, let Y_i be the indicator of the event that v_i is isolated. We will consider the sum

$$Y = \sum_{i=1}^{n} Y_i$$

and show that $\Pr[Y = 0] \to 0$. It will then follow that whp there is at least one isolated node, and so the graph is disconnected.

From the binomial degree distribution in (3.1), the probability that v_i is isolated, conditional on w_i, is

$$\Pr[Y_i = 1 | w_i] = \left(1 - \frac{w_i + \mu}{\theta_n}\pi\right)^{n-1}.$$

Now define

$$p \equiv \mathrm{E}[Y_i] = \Pr[Y_i = 1] = \int f(w_i)\left(1 - \frac{w_i + \mu}{\theta_n}\pi\right)^{n-1} dw_i.$$

Applying the second moment method yields

$$\Pr[Y = 0] \le \frac{\mathrm{Var}[Y]}{\mathrm{E}[Y]^2} = \frac{\sum_i \mathrm{Var}[Y_i] + \sum_{i \neq j} \mathrm{Cov}[Y_i, Y_j]}{(np)^2}.$$

The variance and covariance are given by

$$\mathrm{Var}[Y_i] = \mathrm{E}[Y_i^2] - \mathrm{E}[Y_i]^2 = p - p^2,$$
$$\mathrm{Cov}[Y_i, Y_j] = \mathrm{E}[Y_i, Y_j] - \mathrm{E}[Y_i]\mathrm{E}[Y_j] = \Pr[Y_i = 1, Y_j = 1] - p^2,$$

so

$$\Pr[Y = 0] \le \frac{n(p - p^2) + n(n - 1)(\Pr[Y_i = 1, Y_j = 1] - p^2)}{(np)^2}$$
$$< \frac{1}{np} + \frac{\Pr[Y_i = 1, Y_j = 1]}{p^2} - 1.$$

Let us first consider the $1/(np)$ term. Let $\theta_n = \kappa\frac{n}{\log n}$, where κ is a constant. We claim that if $\kappa > \pi\mu$, then $1/(np) \to 0$. To see this, let ζ be any positive constant that satisfies $\kappa > \pi(\mu + \zeta)$. We have

$$p = \int f(w)\left(1 - \frac{w + \mu}{\theta_n}\pi\right)^{n-1} dw$$
$$\ge \int_0^{\zeta} f(w)\left(1 - \frac{w + \mu}{\theta_n}\pi\right)^{n-1} dw$$
$$\ge F(\zeta)\left(1 - \frac{\mu + \zeta}{\theta_n}\pi\right)^{n-1}$$

$$= F(\zeta)\left(1 - \frac{\mu + \zeta}{\kappa n}\pi \log n\right)^{n-1}$$
$$= F(\zeta)n^{-(\mu+\zeta)\pi/\kappa}(1 + o(1)).$$

Therefore, if $\kappa > \pi(\mu + \zeta)$, then $pn \geq F(\zeta)\omega(n)$. Since $f(w)$ has nonvanishing support, $F(\zeta) > 0$, and so $1/(np) \to 0$ for $n \to \infty$.

Next, we will show that $\Pr[Y_i = 1, Y_j = 1]/p^2 = o(1)$. Consider the joint probability conditional on a set of weights $\mathbf{w}$. Denoting the neighborhood relation by $v_i \sim v_j$, we have

$$\Pr[Y_i = 1, Y_j = 1|\mathbf{w}] = \Pr\left[v_i \nsim v_j, \bigcap_{k \neq i,j} v_i \nsim v_k, v_j \nsim v_k|\mathbf{w}\right]$$

$$= \Pr[v_i \nsim v_j|w_i, w_j]\Pr\left[\bigcap_{k \neq i,j} v_i \nsim v_k, v_j \nsim v_k|v_i \nsim v_j, \mathbf{w}\right]$$

$$= \Pr[v_i \nsim v_j|w_i, w_j]\prod_{k \neq i,j}\Pr[v_i \nsim v_k, v_j \nsim v_k|v_i \nsim v_j, \mathbf{w}].$$

We now use the easily verified property that given events Q, R, and S that depend on $\mathbf{w}$,

$$\Pr[R^c, S^c|Q^c, \mathbf{w}] = 1 - \Pr[R|\mathbf{w}] - \Pr[S|\mathbf{w}] + (1 - \Pr[Q|R, S, \mathbf{w}])\frac{\Pr[R, S|\mathbf{w}]}{\Pr[Q^c|\mathbf{w}]}.$$

Let $a = \Pr[v_i \sim v_j|w_i, w_j]$, $b = \Pr[v_i \sim v_k|w_i, w_k]$, $c = \Pr[v_j \sim v_k|w_j, w_k]$, and define the *clustering coefficient*

$$C = \Pr[v_i \sim v_j|v_i \sim v_k, v_j \sim v_k, w_i, w_j, w_k].$$

Then

$$\Pr[Y_i = 1, Y_j = 1|\mathbf{w}] = (1 - a)\prod_{k \neq i,j}\left[1 - b - c + (1 - C)\frac{bc}{1 - a}\right]. \qquad (5.1)$$

Note that $a = (w_i + w_j)\pi/\theta_n$, and similarly for b and c.

In Section 7.1, we show (Lemma 7.1) that if $w_i, w_j, w_k \leq \hat{w} = (1-3\sqrt{3}/4\pi)\theta_n/2\pi$, then $C \geq a$. Thus, under those conditions,

$$1 - b - c + (1 - C)\frac{bc}{1 - a} \leq 1 - b - c + bc = (1 - b)(1 - c).$$

Now average (5.1) over all weights. It follows from the finite variance of $f(w)$ that for any constant M, we have $F(M\theta_n) = 1 - o(1/n)$, and so

$$\Pr[Y_i = 1, Y_j = 1]$$

$$= \int f(w_i)dw_i \int f(w_j)dw_j(1-a)$$

$$\times \left(\int f(w_k)dw_k \left[1-b-c+(1-C)\frac{bc}{1-a} \right] \right)^{n-2}$$

$$= \int_0^{\hat{w}} f(w_i)dw_i \int_0^{\hat{w}} f(w_j)dw_j(1-a)$$

$$\times \left(\int_0^{\hat{w}} f(w_k)dw_k \left[1-b-c+(1-C)\frac{bc}{1-a} \right] \right)^{n-2}(1+o(1))$$

$$\leq \int_0^{\hat{w}} f(w_i)dw_i \int_0^{\hat{w}} f(w_j)dw_j(1-a)\left(\int_0^{\hat{w}} f(w_k)dw_k(1-b)(1-c) \right)^{n-2}$$

$$\times (1+o(1))$$

$$= \int_0^{\hat{w}} f(w_i)dw_i \int_0^{\hat{w}} f(w_j)dw_j$$

$$\times \left\{ \left(1 - \frac{\pi}{\theta_n}(w_i + w_j) \right) \right.$$

$$\left. \times \left(\int_0^{\hat{w}} f(w_k)dw_k(1-\frac{\pi}{\theta_n}(w_i+w_k))(1-\frac{\pi}{\theta_n}(w_j+w_k)) \right)^{n-2} \right\}$$

$$= \int_0^{\hat{w}} f(w_i)dw_i \int_0^{\hat{w}} f(w_j)dw_j$$

$$\times \left\{ (1 - \frac{\pi}{\theta_n}(w_i + w_j)) \right.$$

$$\left. \times \left(1 - \frac{\pi}{\theta_n}(w_i + w_j + 2\mu) + \frac{\pi^2}{\theta_n^2}(w_i w_j + \mu(w_i + w_j) + \mu^2 + \sigma^2) \right)^{n-2} \right\}.$$

Now consider p^2. Using the fact that μ, σ, and $\frac{1}{2} - (\hat{w} + \mu)\pi/\theta_n$ are all $\Theta(1)$, we have

$$p^2 = \int f(w_i)dw_i \int f(w_j)dw_j \left(1 - \frac{\pi}{\theta_n}(w_i + \mu) \right)^{n-1} \left(1 - \frac{\pi}{\theta_n}(w_j + \mu) \right)^{n-1}$$

$$\geq \int_0^{\hat{w}} f(w_i)dw_i \int_0^{\hat{w}} f(w_j)dw_j \left(1 - \frac{\pi}{\theta_n}(w_i + \mu) \right)^{n-1} \left(1 - \frac{\pi}{\theta_n}(w_j + \mu) \right)^{n-1}$$

$$= \int_0^{\hat{w}} f(w_i)dw_i \int_0^{\hat{w}} f(w_j)dw_j$$

$$\times \left(1 - \frac{\pi}{\theta_n}(w_i + w_j + 2\mu) + \frac{\pi^2}{\theta_n^2}(w_i w_j + \mu(w_i + w_j) + \mu^2) \right)^{n-1}$$

$$
\begin{aligned}
&= \int_0^{\hat{w}} f(w_i)dw_i \int_0^{\hat{w}} f(w_j)dw_j \\
&\quad \times \left(1 - \frac{\pi}{\theta_n}(w_i + w_j + 2\mu) + \frac{\pi^2}{\theta_n^2}(w_i w_j + \mu(w_i + w_j) + \mu^2 + \sigma^2)\right)^{n-1}(1 - o(1)) \\
&> \int_0^{\hat{w}} f(w_i)dw_i \int_0^{\hat{w}} f(w_j)dw_j \left(1 - \frac{\pi}{\theta_n}(w_i + w_j + 2\mu)\right) \\
&\quad \times \left(1 - \frac{\pi}{\theta_n}(w_i + w_j + 2\mu) + \frac{\pi^2}{\theta_n^2}(w_i w_j + \mu(w_i + w_j) + \mu^2 + \sigma^2)\right)^{n-2}(1 - o(1)) \\
&= \int_0^{\hat{w}} f(w_i)dw_i \int_0^{\hat{w}} f(w_j)dw_j \left(1 - \frac{\pi}{\theta_n}(w_i + w_j)\right) \\
&\quad \times \left(1 - \frac{\pi}{\theta_n}(w_i + w_j + 2\mu) + \frac{\pi^2}{\theta_n^2}(w_i w_j + \mu(w_i + w_j) + \mu^2 + \sigma^2)\right)^{n-2}(1 - o(1)).
\end{aligned}
$$

Finally, this gives the desired ratio

$$
\frac{\Pr[Y_i = 1, Y_j = 1]}{p^2} < 1 + o(1).
$$

By the second moment method it then follows that $\Pr[Y = 0] < o(1)$. $\square$

5.2. Connected Graphs

Theorem 5.3. *Let $\theta_n = \kappa n / \log n$ for $\kappa < \sup_{\alpha \in (0,1)} \alpha F^{-1}(1 - \alpha)/4$. Then the GTG is connected whp.*

Proof. The proof is divided into two parts. In the first part, we prove that a constant fraction of nodes αn are connected. In the second part we prove that the rest of the $(1 - \alpha)n$ nodes are connected to the first set of αn nodes.

First part: Invoking the proof of the appearance of the giant component, there are αn nodes all with weights greater than or equal to $s_n \to F^{-1}(1 - \alpha) = \Theta(1)$.

Let $\theta_n = \kappa n / \log n$, where $\kappa < \alpha F^{-1}(1 - \alpha)\pi$. Analogously to r_t, define the connectivity radius r_c by

$$
r_c^2(w_i) = \frac{w_i}{\theta_n} \geq \frac{s_n}{\theta_n} > \frac{\log n}{\alpha \pi n} \, .
$$

Similarly to Theorem 4.3, let us consider small circles around each of these αn nodes, and consider these nodes as an RGG. It is known that $r_n = \sqrt{\log n / (\pi n)}$ is the connectivity threshold in RGG [Gupta and Kumar 98]. The connectivity of RGG implies the connectivity of these αn nodes in our GTG.

Second part: Color the αn high-weighted nodes blue, and the remaining $(1 - \alpha)n$ nodes red. Now let us tile our space into $n/(c_0 \log n)$ squares of size

$c_0 \log n / n$. We state c_0 later. Consider any square S_i, and let B_i be the number of blue nodes in S_i. In expectation there are $\mathrm{E}[B_i] = \alpha c_0 \log n$ blue nodes in each square. From the Chernoff bound it follows that

$$\Pr[B_i \geq (1 - \delta)\alpha c_0 \log n] \geq 1 - n^{-\alpha c_0 \delta^2 / 2}. \tag{5.2}$$

Let us consider one red node r. The node r belongs to some square S_i. Let M_r be the event that the red node r is connected to some blue node $b \in S_i$. Let the weights of r, b be w_r, w_b, respectively. The probability of the complement of M_r, conditioned on there being at least one blue node in S_i, is given by

$$\Pr[M_r^c | B_i \geq 1] = \Pr[w_r + w_b \leq r^2 \theta_n] \leq \Pr\left[w_r + w_b \leq 2c_0 \frac{\log n}{n} \kappa \frac{n}{\log n}\right]$$

$$= \Pr[w_r + w_b \leq 2c_0 \kappa].$$

As long as $F^{-1}(1 - \alpha) > 2c_0\kappa$, then $w_b > 2c_0\kappa$, and hence $\Pr[M_r^c | B_i \geq 1] = 0$. For large enough n it must hold that $(1 - \delta)\alpha c_0 \log n > 1$, and so from (5.2), we have

$$\Pr[M_r^c] \leq \Pr[M_r^c | B_i \geq (1 - \delta)\alpha c_0 \log n] + \Pr[B_i < (1 - \delta)\alpha c_0 \log n]$$

$$\leq 0 + n^{-\alpha c_0 \delta^2 / 2}.$$

If $\alpha c_0 \delta^2 / 2 \geq 1 + \epsilon$ for some $\epsilon > 0$, then by the union bound,

$$\Pr[\bigcup_r M_r^c] \leq \sum_r \Pr[M_r^c] \leq (1 - \alpha)n n^{-(1+\epsilon)} = (1 - \alpha)n^{-\epsilon}.$$

Finally, the probability that all red nodes are connected to the set of blue nodes is given by the following relation:

$$\Pr[\bigcap_r M_r] = 1 - \Pr[\bigcup_r M_r^c] \geq 1 - (1 - \alpha)n^{-\epsilon} \to 1. \tag{5.3}$$

The requirements we have imposed on constants so far are $\kappa < \alpha F^{-1}(1 - \alpha)\pi$, $\kappa < F^{-1}(1 - \alpha)/(2c_0)$, and $\alpha c_0 \geq 2(1 + \epsilon)/\delta^2$. These conditions combine to give

$$\kappa < \alpha F^{-1}(1 - \alpha) \min\left(\pi, \frac{\delta^2}{4(1 + \epsilon)}\right).$$

Since $\alpha \in (0, 1)$, $\delta \in (0, 1)$, and $\epsilon > 0$ are arbitrary, we obtain

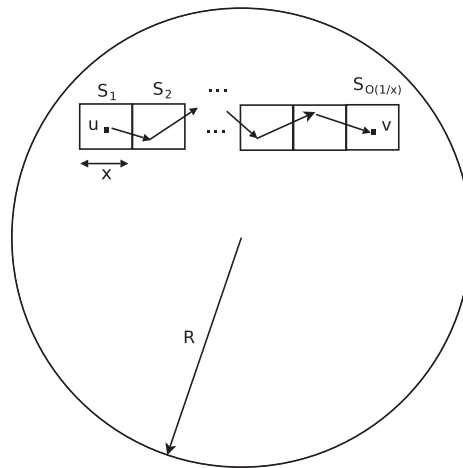$$\kappa < \sup_{\alpha \in (0,1)} \alpha F^{-1} \frac{1 - \alpha}{4}. \qquad \square$$

Figure 2. Illustration of our diameter proof technique: a sequence of adjacent squares of size $x \times x$ links an arbitrary pair of nodes u and v in a unit-area disk.

6. Diameter

In this section we analyze the diameter of GTG, and provide an upper bound on it. In the design of large networks, such as the Internet and wireless networks, it is desirable to achieve low latency in the graph (i.e., the hop count between any pair of nodes in the network is small). In other words, a graph with a small diameter is desired.

We give conditions on the threshold function θ_n such that the graph has a desired diameter in general. Furthermore, we derive conditions on θ_n, in terms of the cumulative distribution function on weights $F(w)$, such that diam belongs to the specific classes diam $= O(1)$, diam $= O(\log^q n)$, and diam $= O(\sqrt{n/\log n})$. These correspond respectively to ultralow-, low-, and high-latency networks. For these three classes, we give the exact expressions on θ_n in the case of the exponentially distributed weights. Note that all of these classes correspond to denser graphs than those we have considered so far, i.e., with θ_n scaling as $o(\log n/n)$ as opposed to the $\Theta(\log n/n)$ scaling for connectivity.

Let u and v be two arbitrary nodes. Construct a sequence of adjacent squares $S_1, S_2, \ldots, S_{O(1/x)}$, of size $x \times x$, linking u and v, such that u and v are the centers of the first and last squares respectively (see Figure 2).[1] The geometric distance between any two nodes is $r \leq \Theta(1)$. Thus, there are $O(1/x)$ squares on the straight path $u - v$ in total.

---

[1] The centers of the squares lie on the straight line $u - v$.

Let V_i be the number of nodes that for $i = 1, 2, \ldots, O(1/x)$ lie within the square S_i. We have $\mathrm{E}[V_i] = nx^2$. Using the Chernoff bound, the following is satisfied:

$$\Pr[V_i \leq (1 - \delta)\mathrm{E}[V_i]] \leq e^{-\mathrm{E}[V_i]\delta^2/2}.$$

Taking $\delta = 1/2$, we get $\Pr[V_i \leq nx^2/2] \leq e^{-nx^2/8}$, i.e., in each square S_i, there are at least $nx^2/2$ nodes whp.

Let M_i be the event that in a square S_i, there is at least one node with weight $w \geq s_n$. We will specify s_n later. We now derive a lower bound on the probability $\Pr[M_i]$. This probability is greater than the probability conditioned on the event that there are at least $nx^2/2$ nodes in S_i, i.e.,

$$\begin{aligned}
\Pr[M_i] &\geq \Pr[M_i | V_i \geq nx^2/2]\Pr[V_i \geq nx^2/2] \\
&\geq (1 - \Pr[W \leq s_n]^{nx^2/2})(1 - e^{-nx^2/8}) \\
&= (1 - F(s_n)^{nx^2/2})(1 - e^{-nx^2/8}).
\end{aligned}$$

We now explain how we choose s_n such that any two neighboring squares S_j and S_{j+1} are connected by an edge (i.e., there are two connected nodes $a \in S_j$ and $b \in S_{j+1}$). Let weights of a and b be w and w', respectively. We showed that in any square S_i there is at least one node with weight greater than or equal to s_n, whp. We want the connectivity relation (2.1) for nodes a and b to be satisfied. The maximal distance $\|a - b\|$ between a pair of nodes is $\|a - b\| \leq x\sqrt{5}$. Conditioned on the events that weights w, w' are greater then s_n, we have the following relation for the connectivity of nodes a and b:

$$\Pr[a \sim b | w, w' \geq s_n] \geq \Pr[2s_n/r^2 \geq \theta_n].$$

Let us choose $s_n = \Theta(x^2\theta_n)$. If an arbitrary pair of nodes (u, v) is connected by a path of nodes belonging to the squares $S_1, S_2, \ldots, S_{O(1/x)}$, then the following relation on diam is satisfied:

$$\Pr[\mathrm{diam} = O(1/x)] \geq \Pr[\cap_{i=1}^{O(1/x)} M_i] = \left((1 - e^{-nx^2/8})(1 - F(s_n)^{nx^2/2}\right)^{O(1/x)},$$

since the nodes, as well as weights, are distributed independently.

We now present a lemma on the diameter.

Lemma 6.1. *Let the cumulative weight distribution function be $F(w)$ in the GTG model. Let x and a sequence $s_n = \Theta(x^2\theta_n)$ be such that*

$$\lim_{n \to \infty} \left(1 - F(s_n)^{nx^2/2}\right)^{1/x} = 1. \tag{6.1}$$

Then whp, $\mathrm{diam} = O(1/x)$.

Proof. The proof follows from the previous discussion.   □

6.1. Some Classes of Diameter

We now analyze and state conditions on θ_n such that diam $= O(1)$, diam $= O(\log^q n)$, and diam $= O(\sqrt{n/\log n})$. We work out the case in which the weight distribution is exponential, $f(w) = e^{-w}$, $w \geq 0$ (i.e., $F(w) = 1 - e^{-w}$, $w \geq 0$), and derive an upper bound on the threshold function θ_n in this particular case. For some other weight distribution, the analysis would be similar. Our results are these:

- *Ultralow Latency:* diam $= O(1)$. Let $x < 1$ be a constant and $s_n = \theta_n$. If $F(\theta_n)^n \to 0$, then diam $= O(1)$ whp. For the exponential weight distribution it follows that $\theta_n = o(\log n)$.

- *Low Latency:* diam $= O(\log^q n)$. Let $x = 1/\log^q n$ and $s_n = \theta_n/\log^{2q} n$. If $F(\theta_n/\log^{2q} n)^{n/(2\log^{2q} n)} \log^q n \to 0$, then diam $= O(\log^q n)$ whp. For the exponential weight distribution it follows that

$$\theta_n = o\Big((\log n)^{2q(1-(\log^{2q} n)/n)}\Big).$$

- *High Latency:* diam $= O(\sqrt{n/\log n})$. Let $x = \sqrt{\log n/n}$ and $s_n = \theta_n \log n/n$. If $\sqrt{n/\log n} F(\theta_n \log n/n)^{\log n} \to 0$, then diam $= O(\sqrt{n/\log n})$ whp. For the exponential weight distribution it follows that

$$\theta_n = o\Big((n/\log n)^{1-1/(2\log n)}\Big).$$

We now prove these claims.

6.1.1. Ultralow Latency: diam $= O(1)$. For the diameter to be a constant, let $x < 1$ be a constant. Invoking Lemma 6.1, it follows that diam $= O(1)$ whp if and only if $1 - F(s_n)^{nx^2/2} \to 1$, i.e., if and only if $F(s_n)^n \to 0$. The condition on the size of diam is given by the following claim, and we can derive θ_n such that diam $= O(1)$ whp.

Claim 6.2. *If $F(\theta_n)^n \to 0$, then* diam $= O(1)$ *whp.*

For the exponential weight distribution it follows that $F(\theta_n)^n = (1 - e^{-\theta_n})^n \to e^{-n/e^{\theta_n}}$. The last equation tends to zero if and only if $n/e^{\theta_n} \to \infty$. That is, we make the following claim.

Claim 6.3. *For the exponential weight distribution $f(w) = e^{-w}$, the diameter in GTG satisfies* diam $= O(1)$ *if $\theta_n = o(\log n)$.*

6.1.2. Low Latency: diam $= O(\log^q n)$. Let us choose $x = 1/\log^q n$. Invoking Lemma 6.1, we obtain

$$(1 - F(s_n)^{nx^2/2})^{1/x} = \left(1 - F(s_n)^{n/(2\log^{2q} n)}\right)^{\log^q n}. \tag{6.2}$$

For $s_n \to 0$, the last expression tends to 1 if and only if

$$F(s_n)^{n/(2\log^{2q} n)} \log^q n \to 0,$$

where we use $\lim_{t \to +\infty}(1 - 1/t)^t = 1/e$. The condition on the size of diam is given by the following claim.

Claim 6.4. *if $F(\theta_n/\log^{2q} n)^{n/(2\log^{2q} n)} \log^q n \to 0$, then* diam $= O(\log^q n)$ *whp.*

For the exponential weight distribution, the following is to be satisfied:

$$F(s_n)^{n/(2\log^{2q} n)} \log^q n = \log^q n(1 - e^{-s_n})^{n/(2\log^{2q} n)} \to s_n^{n/(2\log^{2q} n)} \log^q n \to 0,$$

or equivalently,

$$s_n = o\left((\log n)^{-\frac{2q}{n}\log^{2q} n)}\right).$$

Claim 6.5. *For the exponential weight distribution $f(w) = e^{-w}$, the diameter in GTG satisfies* diam $= O(\log^q n)$ *if $\theta_n = o\left((\log n)^{2q(1-(\log^{2q} n)/n)}\right)$.*

6.1.3. High Latency: diam $= O(\sqrt{n/\log n})$. Let us choose $x = \sqrt{\log n/n}$. Invoking Lemma 6.1, we obtain

$$(1 - F(s_n)^{nx^2/2})^{1/x} = (1 - F(s_n)^{\log n})^{\sqrt{\frac{n}{\log n}}}.$$

It can be shown, using $\lim_{t \to +\infty}(1 - 1/t)^t = 1/e$, that the last expression tends to 1 if and only if $\sqrt{n/\log n} F(s_n)^{\log n} \to 0$. The condition on the size of diam is given by the following claim.

Claim 6.6. *If $\sqrt{n/\log n} F(\theta_n \log n/n)^{\log n} \to 0$, then* diam $= O(\sqrt{n/\log n})$ *whp.*

For the exponential weight distribution, the following is to be satisfied:

$$\sqrt{n/\log n} F(s_n)^{\log n} = \sqrt{n/\log n}(1 - e^{-s_n})^{\log n} \to \sqrt{n/\log n} s_n^{\log n} \to 0,$$

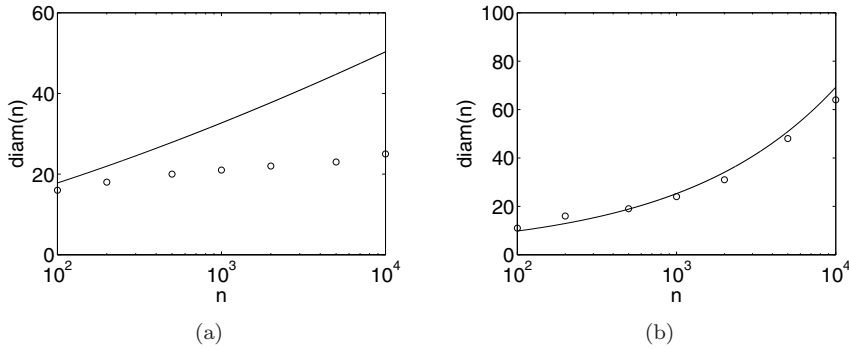or equivalently,

$$s_n = o\left((\log n/n)^{1/(2\log n)}\right).$$

(a)   (b)

Figure 3. (a) For the case of diam $= O(\log^q n)$, with $q = 1.5$, the analytical solid curve is the upper bound on diam(n). Simulation results match with theoretical predictions, since the simulation points all lie below the analytical curve. (b) For the case of diam $= O(\sqrt{n/\log n})$, the solid curve plots the upper bound on diam(n), and this bound closely matches the experimental values.

Claim 6.7. *For the exponential weight distribution $f(w) = e^{-w}$, the diameter in GTG satisfies*

$$\mathrm{diam} = O(\sqrt{n/\log n}) \quad \text{if} \quad \theta_n = o\Big((n/\log n)^{1-1/(2\log n)}\Big).$$

Simulation results are shown for the GTG with path-loss exponent $\beta = 3$ (not $\beta = 2$) for the case of diam $= O(\log^{1.5} n)$ in Figure 3(a) and diam $= O(\sqrt{n/\log n})$ in Figure 3(b). Exponentially distributed weights with mean 1 are used. The network sizes simulated are $n = \{100, 200, 500, 1000, 2000, 10000\}$. The threshold values θ_n for the two cases are obtained by invoking previous claims.

7. Clustering Coefficient

7.1. Weights Given

Let us consider in more detail the clustering coefficient defined in Section 5, namely the neighbor probability

$$C(w_i, w_j, w_k) = \Pr[v_i \sim v_j | v_i \sim v_k, v_j \sim v_k, w_i, w_j, w_k].$$

Let $x = \sqrt{(w_i + w_j)/\theta_n}$, $y = \sqrt{(w_j + w_k)/\theta_n}$, and $z = \sqrt{(w_i + w_k)/\theta_n}$. Then if d_{ij} represents the distance between points i and j, we have

$$C(w_i, w_j, w_k) = \Pr[d_{ij} \leq x | d_{jk} \leq y, d_{ik} \leq z]$$

$$= \frac{1}{\pi z^2} \int_0^z \Pr[d_{ij} \leq x | d_{jk} \leq y, d_{ik} = r]\, 2\pi r\, dr$$

$$= \frac{1}{\pi^2 y^2 z^2} \int_0^z A(r) 2\pi r\, dr,$$

where $A(r)$ is the overlap area of a disk of radius x centered at i and a disk of radius y centered at k. Now consider a triangle $\triangle IJK$, with sides $IJ = r$, $IK = x$, $JK = y$, and $\angle JKI = \eta$, $\angle IJK = \xi$. Following arguments similar to those in [Dall and Christensen 02], there are three possible cases for $A(r)$:

$$A(r) = \begin{cases} \pi[\min(x, y)]^2 & \text{if } r \leq |x - y|, \\ x^2(\eta - \sin\eta\cos\eta) + y^2(\xi - \sin\xi\cos\xi) & \text{if } |x - y| < r < x + y, \\ 0 & \text{if } r \geq x + y, \end{cases}$$

where

$$\eta = \cos^{-1}\left(\frac{r^2 + x^2 - y^2}{2xr}\right),$$

$$\xi = \cos^{-1}\left(\frac{r^2 - x^2 + y^2}{2yr}\right).$$

From the definitions of x, y, and z, we have $|x - y| < z < x + y$. After some algebraic manipulation, one then obtains

$$C = \frac{x^2}{\pi}\left[\frac{1}{y^2}\cos^{-1}\left(\frac{z^2 + x^2 - y^2}{2xz}\right) + \frac{1}{x^2}\cos^{-1}\left(\frac{z^2 - x^2 + y^2}{2yz}\right)\right.$$

$$+ \frac{1}{z^2}\cos^{-1}\left(\frac{x^2 + y^2 - z^2}{2xy}\right)$$

$$\left. - \frac{x^2 + y^2 + z^2}{4x^2y^2z^2}\sqrt{2x^2y^2 + 2x^2z^2 + 2y^2z^2 - x^4 - y^4 - z^4}\right]$$

$$= \frac{1}{\pi(w_i + w_k)(w_j + w_k)}$$

$$\times \left\{(w_i + w_j)(w_i + w_k)\cos^{-1}\left(\frac{w_i}{\sqrt{w_i + w_j}\sqrt{w_i + w_k}}\right)\right.$$

$$+ (w_i + w_j)(w_j + w_k)\cos^{-1}\left(\frac{w_j}{\sqrt{w_i + w_j}\sqrt{w_j + w_k}}\right)$$

$$+ (w_i + w_k)(w_j + w_k)\cos^{-1}\left(\frac{w_k}{\sqrt{w_i + w_k}\sqrt{w_j + w_k}}\right)$$

$$\left. - (w_i + w_j + w_k)\sqrt{w_i w_j + w_j w_k + w_k w_i}\right\}.$$

Note that while C is a function of the weights w_i, w_j, and w_k, it is independent of θ_n. This reflects a similar property in random geometric graphs [Dall and Christensen 02], where the clustering coefficient is independent of the graph's mean degree.

In terms of the connection probabilities a, b, and c defined in Theorem 5.2, $x = \sqrt{a/\pi}$, $y = \sqrt{b/\pi}$, and $z = \sqrt{c/\pi}$. Then, C is given by

$$
C\pi bc = ab \cos^{-1} \frac{a+b-c}{2\sqrt{ab}} + bc \cos^{-1} \frac{-a+b+c}{2\sqrt{bc}} + ca \cos^{-1} \frac{a-b+c}{2\sqrt{ac}}
$$
$$
- \frac{a+b+c}{4} \sqrt{2ab + 2ac + 2ca - a^2 - b^2 - c^2}.
$$

We now prove the bound on the clustering coefficient that we needed for Theorem 5.2.

Lemma 7.1. *If $w_i, w_j, w_k \leq \hat{w} = (1 - 3\sqrt{3}/4\pi)\theta_n/2\pi$, then $C \geq a$.*

Proof. Define

$$
S(a, b, c) = \frac{C\pi}{a}
$$
$$
= \frac{1}{a} \cos^{-1} \frac{-a+b+c}{2\sqrt{bc}} + \frac{1}{b} \cos^{-1} \frac{a-b+c}{2\sqrt{ac}} + \frac{1}{c} \cos^{-1} \frac{a+b-c}{2\sqrt{ab}} - \gamma \frac{a+b+c}{4abc},
$$

where

$$
\gamma = \sqrt{2ab + 2ac + 2ca - a^2 - b^2 - c^2} = 2\sqrt{bc}\sqrt{1 - \left(\frac{-a+b+c}{2\sqrt{bc}}\right)^2}.
$$

It is easy to verify that $S = \pi$ when $w_i = w_j = w_k = \hat{w}$. We will now show that S is nonincreasing over the weights, and thus $S \geq \pi$ for all smaller values of w_i, w_j, w_k.

Consider the sign of the derivative

$$
\frac{dS}{dw_i} = \frac{\partial S}{\partial a} \frac{\partial a}{\partial w_i} + \frac{\partial S}{\partial b} \frac{\partial b}{\partial w_i} + \frac{\partial S}{\partial c} \frac{\partial c}{\partial w_i} = \frac{\pi}{\theta_n} \left(\frac{\partial S}{\partial a} + \frac{\partial S}{\partial b} \right).
$$

Since S is symmetric in a and b, it is sufficient to consider the sign of $\partial S/\partial a$:

$$
\frac{\partial S}{\partial a} = \frac{\gamma}{4a^2bc}(-a+b+c) - \frac{1}{a^2} \cos^{-1} \frac{-a+b+c}{2\sqrt{bc}}.
$$

Now let

$$
t = \frac{-a+b+c}{2\sqrt{bc}} \in [0, 1].
$$

Then

$$\frac{\partial S}{\partial a} = \frac{1}{a^2}\left(t\sqrt{1-t^2} - \cos^{-1} t\right).$$

Given the function $\varphi(t) = t\sqrt{1-t^2} - \cos^{-1} t$ on $[0, 1]$, we have $\varphi'(t) = 2\sqrt{1-t^2} \geq 0$, and $\varphi(1) = 0$. It follows that $\partial S/\partial a \leq 0$, and so $dS/dw_i \leq 0$. Finally, S is symmetric in (w_i, w_j, w_k), so it must be nonincreasing over each of the weights and bounded below by the value at $w_i = w_j = w_k = \hat{w}$. $\qquad\square$

7.2. Degree Given

Define C_l to be the neighbor probability for a node with a given degree:

$$C_l = \Pr[v_i \sim v_j | v_i \sim v_k, v_j \sim v_k, d(v_k) = l]$$

$$= \frac{\Pr[v_i \sim v_j, v_i \sim v_k, v_j \sim v_k, d(v_k) = l]}{\Pr[v_i \sim v_k, v_j \sim v_k, d(v_k) = l]}$$

$$= \frac{\int f(\mathbf{w})\Pr[v_i \sim v_j, v_i \sim v_k, v_j \sim v_k, d(v_k) = l|\mathbf{w}]d\mathbf{w}}{\int f(\mathbf{w})\Pr[v_i \sim v_k, v_j \sim v_k, d(v_k) = l|\mathbf{w}]d\mathbf{w}}.$$

A straightforward calculation shows that C_l is given by the ratio I_n/I_d of two integrals, where

$$I_d = \int f(w_k)\left(\frac{\pi}{\theta_n}(\mu + w_k)\right)^l \left(1 - \frac{\pi}{\theta_n}(\mu + w_k)\right)^{n-l} dw_k$$

and

$$I_n = \iiint f(w_i)f(w_j)f(w_k)\pi y^2 \pi z^2 C(w_i, w_j, w_k)dw_i\, dw_j$$

$$\times \left(\frac{\pi}{\theta_n}(\mu + w_k)\right)^{l-2} \left(1 - \frac{\pi}{\theta_n}(\mu + w_k)\right)^{n-l} dw_k$$

$$= \iiint f(w_i)f(w_j)f(w_k)(w_i + w_k)(w_j + w_k)C(w_i, w_j, w_k)dw_i\, dw_j$$

$$\times \frac{1}{(\mu + w_k)^2}\left(\frac{\pi}{\theta_n}(\mu + w_k)\right)^l \left(1 - \frac{\pi}{\theta_n}(\mu + w_k)\right)^{n-l} dw_k$$

For a specific weight distribution, these integrals may be evaluated numerically. It is intuitive that when l is very large ($l = \Theta(n)$), C_l should scale as $1/n$: nodes that connect to very many neighbors presumably do so because of their high weights, and their neighbors are no more likely to be connected than any two random nodes are. Interestingly, in the case shown in Figure 4, C_l scales almost perfectly as l^{-1} even at relatively small values of l. For an exponential weight distribution with mean 1 and parameters $n = 1000$ and $\theta_n = 1000$, the slope on the log-log plot already appears very close to -1 at $l = 8$.
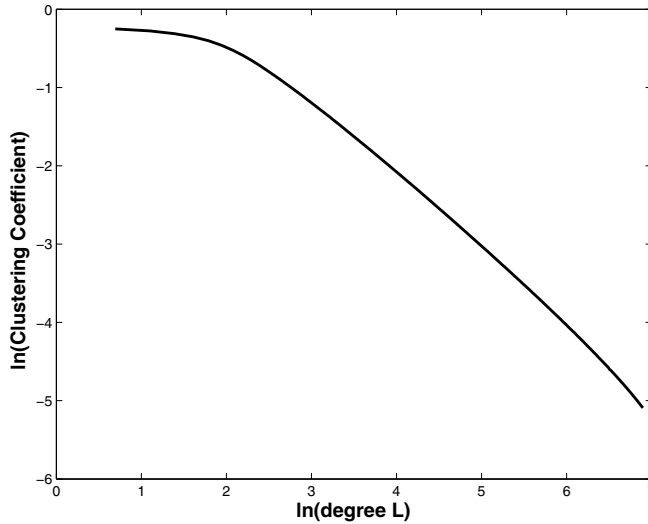
Figure 4. Clustering coefficient vs. degree L.

8.   Summary

Geographical threshold graphs are a rich model with the possibility of control-
ling structural properties by choosing specific weight distributions and tuning
threshold values. The model is a versatile one and can be used for the genera-
tion and analysis not only of web graphs or large complex networks, but more
generally of relation graphs in a large data set. If the data have a metric and can
be mapped to nodes in Euclidean space, much of the foregoing analysis applies:
one may hope to control structural properties of the data set by studying it as
a GTG.

In this paper we have analyzed some of the structural properties of a GTG.
Given a node weight distribution $f(w)$ and threshold θ_n, the degree distribution
can be easily calculated.

We have given bounds on the threshold value θ_n guaranteeing the absence or
existence of the giant component. We have also given bounds on θ_n guaranteeing
a disconnected or connected graph, and provided upper bounds on the diameter
for sufficiently dense graphs.

Finally, we have derived a formula for the clustering coefficient in terms of
the weight distribution and threshold, as well as discussed the general clus-
tering coefficient for nodes with a given degree, and evaluated it numerically.

Our analysis has used the additive threshold function $(w_i + w_j)/r^2 \geq \theta_n$ for the connectivity relation, but not all of our techniques require it. For this reason, many of the results may be generalized to other threshold functions, other path-loss exponents, and other spatial dimensionalities in a straightforward manner.

References

[Abello et al. 02] James Abello, Panos M. Pardalos, and Mauricio G. C. Resende, editors. *Handbook of Massive Data Sets*. Norwell, MA: Kluwer Academic Publishers, 2002.

[Aiello et al. 00] William Aiello, Fan Chung, and Linyuan Lu. "A Random Graph Model for Massive Graphs." In *Proceedings of the Thirty-Second Annual ACM Symposium on Theory of Computing*, pp. 171–180. New York: ACM, 2000.

[Alon and Spencer 00] Noga Alon and Joel H. Spencer. *The Probabilistic Method*, Second Edition. New York: John Wiley & Sons, 2000.

[Barabási and Albert 99] A. L. Barabási and R. Albert. "Emergence of Scaling in Random Networks." *Science* 286:5439 (1999), 509–512.

[Bollobás 01] Béla Bollobás. *Random Graphs*, Second Edition. Cambridge, UK: Cambridge University Press, 2001.

[Bollobás et al. 01] Béla Bollobás, Oliver Riordan, Joel Spencer, and Gábor Tusnády. "The Degree Sequence of a Scale-Free Random Graph Process." *Random Struct. Algorithms* 18:3 (2001), 279–290.

[Bonato 05] Anthony Bonato. "A Survey of Models of the Web Graph." In *Combinatorial and Algorithmic Aspects of Networking: First Workshop on Combinatorial and Algorithmic Aspects of Networking, CAAN 2004, Banff, Alberta, Canada, August 5–7, 2004, Revised Selected Papers*, Lecture Notes in Computer Science 3405, pp. 159–172. Heidelberg: Springer, 2005.

[Bradonjić and Kong 07] Milan Bradonjić and Joseph Kong. "Wireless Ad Hoc Networks with Tunable Topology." Paper presented at the 45th Annual Allerton Conference on Communication, Control and Computing, Urbana-Champaign, IL, September 26–28, 2007. Available at http://www.csl.uiuc.edu/allerton/archives/allerton07/PDFs/papers/0051.pdf.

[Bradonjić et al. 07] Milan Bradonjić, Aric A. Hagberg, and Allon G. Percus. "Giant Component and Connectivity in Geographical Threshold Graphs." In *Algorithms and Models for the Web-Graph: 5th International Workshop, WAW 2007, San Diego, CA, USA, December 11–12, 2007, Proceedings*, Lecture Notes in Computer Science

4863, edited by Anthony Bonato and Fan R. K. Chung, pp. 209–216. New York: Springer, 2007.

[Chung and Lu 06] Fan Chung and Linyuan Lu. "The Volume of the Giant Component of a Random Graph with Given Expected Degrees." *SIAM J. Discret. Math.* 20:2 (2006), 395–411.

[Cooper and Frieze 01] Colin Cooper and Alan M. Frieze. "A General Model of Undirected Web Graphs." In *Algorithms—ESA 2001: 9th Annual European Symposium, Aarhus, Denmark, August 28–31, 2001, Proceedings*, Lecture Notes in Computer Science 2161, pp. 500–511. London: Springer, 2001.

[Dall and Christensen 02] Jesper Dall and Michael Christensen. "Random Geometric Graphs." *Physical Review E* 66:1 (2002), 016121.

[Durrett 06] Rick Durrett. *Random Graph Dynamics*. Cambridge, UK: Cambridge University Press, 2006.

[Erdős and Rényi 59] P. Erdős and A. Rényi. "On Random Graphs." *Publicationes Mathematicae Debrecen* 6 (1959), 290-297.

[Erdős and Rényi 60] P. Erdős and A. Rényi. "On the Evolution of Random Graphs." *Publ. Math. Inst. Hungar. Acad. Sci.*, 5 (1960), 17–61.

[Faloutsos et al. 99] Michalis Faloutsos, Petros Faloutsos, and Christos Faloutsos. "On Power-Law Relationships of the Internet Topology." In *Proceedings of the Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication*, pp. 251–262. New York: ACM, 1999.

[Gupta and Kumar 98] P. Gupta and P. R. Kumar. "Critical Power for Asymptotic Connectivity." In *Proceedings of the 37th IEEE Conference on Decision and Control*, Vol. 1, pp. 1106–1110. Los Alamitos, CA: IEEE Press, 1998.

[Hagberg et al. 06] Aric Hagberg, Pieter J. Swart, and Daniel A. Schult. "Designing Threshold Networks with Given Structural and Dynamical Properties. *Phys. Rev. E* 74 (2006), 056116.

[Kleinberg et al. 99] Jon M. Kleinberg, Ravi Kumar, Prabhakar Raghavan, Sridhar Rajagopalan, and Andrew S. Tomkins. "The Web as a Graph: Measurements, Models, and Methods." In *Computing and Combinatorics: 5th Annual International Conference, COCOON'99, Tokyo, Japan, July 26–28, 1999, Proceedings*, Lecture Notes in Computer Science 1627, pp. 1–17. Heidelberg: Springer, 1999.

[Kumar et al. 00] R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins, and E. Upfal. "Stochastic Models for the Web Graph." In *Proceedings of the 41st Annual Symposium on Foundations of Computer Science*, p. 57–65. Washington, DC: IEEE, 2000.

[Mahadev and Peled 95] N. V. R. Mahadev and U. N. Peled. *Threshold Graphs and Related Topics*, Annals of Discrete Mathematics 56. New York: Elsevier, 1995.

[Masuda et al. 05] Naoki Masuda, Hiroyoshi Miwa, and Norio Konno. "Geographical Threshold Graphs with Small-World and Scale-Free Properties." *Physical Review E* 71:3 (2005), 036108.

[Molloy and Reed 95] M. Molloy and B. Reed. "A Critical Point for Random Graphs with a Given Degree Sequence." In *Proceedings of the Sixth International Seminar on Random Graphs and Probabilistic Methods in Combinatorics and Computer Science*, pp. 161–179. New York: John Wiley & Sons, Inc., 1995.

[Penrose 03] Mathew D. Penrose. *Random Geometric Graphs.* Oxford: Oxford University Press, 2003.

Milan Bradonjić, Department of Electrical Engineering, UCLA, Los Angeles, CA 90095 and Mathematical Modeling and Analysis Group, Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545 (milan@lanl.gov)

Aric Hagberg, Mathematical Modeling and Analysis Group, Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545 (hagberg@lanl.gov)

Allon G. Percus, Department of Mathematics, UCLA, Los Angeles, CA 90095 and School of Mathematical Sciences, Claremont Graduate University, Claremont, CA 91711 (allon.percus@cgu.edu)