Giant Component and Connectivity in Geographical Threshold Graphs

Milan Bradonjić[1], Aric Hagberg[2], and Allon G. Percus[3,4]

[1] Department of Electrical Engineering, UCLA, Los Angeles, CA 90095
milan@ee.ucla.edu
[2] Mathematical Modeling and Analysis, Theoretical Division,
Los Alamos National Laboratory, Los Alamos, NM 87545
hagberg@lanl.gov
[3] Department of Mathematics, UCLA, Los Angeles, CA 90095
[4] Information Sciences Group, Los Alamos National Laboratory, Los Alamos, NM 87545
percus@ipam.ucla.edu

Abstract. The geographical threshold graph model is a random graph model with nodes distributed in a Euclidean space and edges assigned through a function of distance and node weights. We study this model and give conditions for the absence and existence of the giant component, as well as for connectivity.

Keywords: random graph, geographical threshold graph, giant component, connectivity.

1 Introduction

Large networks such as the Internet, World Wide Web, phone call graphs, infections disease contacts, and financial transactions have provided new challenges for modeling and analysis [1]. 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" [2]. Extensive theoretical and experimental research has been done in web-graph modeling. Early measurements suggested that the Internet exhibits a power-law degree distribution [3] and that the web graph also follows a power-law distribution in in- and out-degree of links [4]. Modeling approaches using random graphs have attempted to capture both the structure and dynamics of the web graph [5,6,7,8,9].

In this short paper we study geographical threshold graphs (GTGs), a static model for networks that includes both geometric information and node weight information. The motivation for analyzing this model is that many real networks need to be studied by using a "richer" stochastic model (which in this case includes both a distance between nodes and weights on the nodes). This model has already been applied in the study of wireless ad hoc networks for systems where the wireless nodes have different capabilities [10]. The weights, in this case, represent power or bandwidth resources available to wireless nodes in the network. By varying the weights in a GTG model, properties such as the diameter or degree distribution can be tuned. Other possible applications of GTGs are in epidemic modeling, where the weights might represent susceptibility to

A. Bonato and F.R.K. Chung (Eds.): WAW 2007, LNCS 4863, pp. 209–216, 2007.

infection, or other social networks where the weights may be related to attractiveness or other individual characteristic.

2 Geographical Threshold Graph Model

In addition to unstructured random graphs [11,12], recent research has focused on random geometric graphs (RGG) where edges are created according to a distance between nodes [13], and threshold graphs [14,15] with edges created according to a function of node weights. Geographical threshold graphs, which combine aspects of RGG and threshold graphs, have only recently begun to receive attention [16,10].

The GTG model is constructed from of a set of n nodes placed independently in $\mathbf{R}^d$ according to a Poisson point-wise process. A non-negative weight w_i, taken randomly and independently from a probability distribution function $f(w) : \mathbf{R}_0^+ \to \mathbf{R}_0^+$, 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 i and j at distance r, the edge (i, j) exists if and only if the following connectivity relation is satisfied:

$$G(w_i, w_j)h(r) \geq \theta_n\,, \tag{1}$$

where θ_n is a given threshold parameter that depends on the size of the network. The function $h(r)$ is assumed to be decreasing in r. We use $h(r) = r^{-\alpha}$, for some positive α, which is typical for e.g., the path-loss model in wireless networks [10]. The interaction strength between nodes $G(w_i, w_j)$ is usually taken to be symmetric 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 shown. For the case of uniformly distributed nodes over a unit space it has been shown [16,10] that the expected degree of a node with weight w is

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

where h^{-1} is the inverse of h. The degree distribution has been studied for specific weight distribution functions $f(w)$ [16]. In both the multiplicative and additive case of $G(w, w')$, questions of diameter, connectivity, and topology control have been addressed [10].

Here we restrict ourselves to the case of $g(w) = w$, $\alpha = 2$, and nodes distributed uniformly over a two-dimensional space. For analytical simplicity we take the space to be a unit torus. We concentrate on the analysis of the additive model, i.e., when the connectivity relation is given by

$$\frac{w_i + w_j}{r^2} \geq \theta_n. \tag{3}$$

Our techniques may be generalized to other cases in a straightforward manner. Our contribution in this short paper is to provide the first bounds on θ_n for the emergence of the giant component, and for connectivity.

3 Giant Component in GTG

Definition 1 (Giant Component). *The giant component is a connected component with 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 $\theta_n = cn$, we specify positive constants $c' > c''$ and prove that whp, if $c > c'$ the giant component does not exist whereas if $c < c''$ the giant component exists.

3.1 Absence of Giant Component

Lemma 1. *Let* $\theta_n = cn$ *for* $c > c'$, *where* $c' = 2\pi E[w]$. *Then whp there is no giant component in GTG.*

Proof. We use an approach similar to one given in [17]. Divide the nodes into three classes: alive, dead and neutral. Denote the number of alive nodes as Y_i. The algorithm works as follows. At time $t = 0$, designate one node (picked u.a.r.) as being alive and all others as neutral. Now, at each subsequent time step t, pick a node v_t u.a.r. from among those that are alive, and then consider all neutral nodes connected to v_t. Denote the number of these nodes as 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 recursion relation: $Y_0 = 1$ and $Y_t = Y_{t-1} + Z_t - 1$, for $t \geq 1$. The number of alive nodes satisfies

$$Y_t - 1 = \sum_{k=1}^{t} Z_k - t. \tag{4}$$

At a time step k, let d_k be the degree of node v_k. Since Z_k only includes the neutral nodes connected to v_k,

$$Z_k \leq d_k. \tag{5}$$

Now 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 the following relation

$$\Pr[T \geq t] = \Pr[Y_t > 0] = \Pr[Y_t \geq 1] = \Pr[\sum_{k=1}^{t} Z_k \geq t] \leq \Pr[\sum_{k=1}^{t} d_k \geq t]. \tag{6}$$

Consider the threshold $\theta_n = cn$ for some $c > 0$. It is shown in the Appendix that for a node v_k with random weight w_k, the vertex degree distribution is Poisson: $d(w_k) \sim Po(a(w_k + \mu))$, where $a = n\pi/\theta_n$ and $\mu = \mathrm{E}[w]$. Since the sum of independent random Poisson variables is a Poisson random variable,

$$\Pr\Big[\sum_{k=1}^{t} d_k \geq t\Big] = \Pr\Big[Po(a\sum_{k=1}^{t}(w_k + \mu)) \geq t\Big]. \tag{7}$$

We now use the following inequality. For any $\varepsilon \in (0,1)$,

$$\Pr\Big[Po(a\sum(w_k+\mu))\geq t\Big] \leq \Pr\Big[Po(a\sum(w_k+\mu))\geq t \,|\, \sum w_k\in(1\pm\varepsilon)t\mu\Big] + \Pr\Big[\sum w_k\notin(1\pm\varepsilon)t\mu\Big].$$

By the central limit theorem, for $t\to\infty$, the sum $(\sum w_k - t\mu)/(\sqrt{t}\sigma)$ tends to the normal distribution $N(0,1)$. That is,

$$\Pr\Big[\sum w_k\notin(1\pm\varepsilon)t\mu\Big] = \Pr\Big[\frac{\sum w_k - t\mu}{\sqrt{t}\sigma}\notin(-\varepsilon,\varepsilon)\sqrt{t}\frac{\mu}{\sigma}\Big]\to 0. \tag{8}$$

Finally, we use the concentration on the Poisson random variable [13]. Define $\lambda = \mathrm{E}[a\sum(w_k+\mu)] = 2at\mu$. Given any $\varepsilon_0\in(0,1)$, for $t\to\infty$, i.e., $\lambda\to+\infty$, it follows that

$$\Pr[Po(\lambda)\notin(1\pm\varepsilon_0)\lambda]\leq e^{-\lambda H(1-\varepsilon_0)}+e^{-\lambda H(1+\varepsilon_0)}\to 0, \tag{9}$$

where the function $H(x) = 1-x+x\ln x$, for $x>0$. It is now sufficient to choose a small enough that $t>2at\mu(1+\varepsilon_0)$ for some positive constant ε_0. This is equivalent to $1>2a\mu$, i.e., $c>2\pi\mu$. It follows that $\Pr[Po(a\sum(w_i+\mu))\geq t] = o(1)$ for $t=\Theta(n)$, which completes the proof.

3.2 Existence of Giant Component

Lemma 2. *Let* $\theta_n = cn$ *for* $c<c'' = \sup_{\alpha\in(0,1)}\alpha F^{-1}(1-\alpha)/\lambda_c$, *where* $\pi\lambda_c\approx 4.52$ *is the mean degree at which the giant component first appears in Random Geometric Graphs (RGG) [13]. 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 $\Pr[X_i=1] = 1-F(s) =: q$. 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 $\delta = 1-\alpha/q$,

$$\Pr[X\leq\alpha n] = \Pr[X\leq(1-\delta)\mathrm{E}[X]]\leq\exp\big(-n(q-\alpha)^2/(2q)\big) = n^{-\beta} \tag{10}$$

for some constant $\beta>1$ satisfying $(q-\alpha)^2 = 2q\beta\ln n/n$. Solving that quadratic equation in q gives $q=\alpha+\Theta(\ln n/n)$, so $F(s_n) = 1-q = 1-\alpha-\Theta(\ln n/n)$. For any $\varepsilon>0$ and n sufficiently large the following is satisfied

$$F^{-1}(1-\alpha)\geq s_n\geq F^{-1}(1-\alpha-\varepsilon). \tag{11}$$

Thus, let us define the sequence s_n by its limit

$$s_n\to F^{-1}(1-\alpha) = \Theta(1). \tag{12}$$

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

$$r_t^2(w_i) = w_i/\theta_n. \tag{13}$$

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:

$$(w_i + w_j)/r^2 \geq w_i/r_t^2 = \theta_n. \tag{14}$$

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

$$r_t^2(w_i) = \frac{w_i}{\theta_n} \geq \frac{s_n}{\theta_n} > \frac{\lambda_c}{\alpha n}. \tag{15}$$

Let us therefore consider small circles, with a fixed radius r_0 s.t. $\sqrt{s_n/\theta_n} > r_0 > \sqrt{\lambda_c/(\alpha n)}$, around each of these αn nodes. A subgraph of this must be a RGG with mean degree $> \lambda_c$, 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 c over $\alpha \in (0,1)$, and the lemma follows.

4 Connectivity in GTG

Definition 2 (Connectivity). *The graph on n vertices is connected if the largest component has size n.*

In this section we analyze sufficient conditions for the entire graph to be connected. We consider the connectivity threshold $\theta_n = cn/\ln n$ and specify a bound on c.

Lemma 3. *Let* $\theta_n = cn/\ln n$ *for* $c < \sup_{\alpha\in(0,1)} \alpha F^{-1}(1-\alpha)/4$. *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 $\geq s_n \to F^{-1}(1-\alpha) = \Theta(1)$.

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

$$r_c^2(w_i) = \frac{w_i}{\theta_n} \geq \frac{s_n}{\theta_n} > \frac{\ln n}{\alpha\pi n}. \tag{16}$$

Similarly to Lemma 2 let us consider small circles around each of these αn nodes, and consider these nodes as a RGG. It is known that $r_n = \sqrt{\ln n/(\pi n)}$ is the connectivity threshold in RGG [18]. 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 $n/(c_0 \ln n)$ squares of size $c_0 \ln n/n$. 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 $\mathrm{E}[B_i] = \alpha c_0 \ln n$ blue nodes in each square. From the Chernoff bound it follows

$$\Pr[B_i \geq (1-\delta)\alpha c_0 \ln n] \geq 1 - n^{-\alpha c_0 \delta^2/2}. \tag{17}$$

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

$$\Pr[M_r^c | B_i \geq 1] = \Pr[w_r + w_b \leq r^2 \theta_n] \leq \Pr[w_r + w_b \leq 2c_0 \frac{\ln n}{n} c \frac{n}{\ln n}] = \Pr[w_r + w_b \leq 2c_0 c]. \tag{18}$$

As long as $F^{-1}(1-\alpha) > 2c_0 c$, $w_b > 2c_0 c$ and hence $\Pr[M_r^c | B_i \geq 1] = 0$. For large enough n it must hold that $(1-\delta)\alpha c_0 \ln n > 1$, and so from Eq. (17),

$$\Pr[M_r^c] \leq \Pr[M_r^c | B_i \geq (1-\delta)\alpha c_0 \ln n] + \Pr[B_i < (1-\delta)\alpha c_0 \ln n] \leq 0 + n^{-\alpha c_0 \delta^2/2}. \tag{19}$$

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

$$\Pr[\bigcup_r M_r^c] \leq \sum_r \Pr[M_r^c] \leq (1-\alpha) n n^{-(1+\varepsilon)} = (1-\alpha) n^{-\varepsilon}. \tag{20}$$

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

$$\Pr[\bigcap_r M_r] = 1 - \Pr[\bigcup_r M_r^c] \geq 1 - (1-\alpha) n^{-\varepsilon} \to 1. \tag{21}$$

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

$$c < \alpha F^{-1}(1-\alpha) \min(\pi, \frac{\delta^2}{4(1+\varepsilon)}). \tag{22}$$

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

$$c < \sup_{\alpha \in (0,1)} \alpha F^{-1}(1-\alpha)/4. \tag{23}$$

5 Discussion

The GTG model is a versatile one and can be used not only for the generation and analysis of web-graphs or large complex networks, but more generally for 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.

Furthermore, while we considered the GTG model as a static structure, the set of weights in the model could vary in time. This would introduce dynamics, as might be appropriate for particular applications such as wireless networking.

Acknowledgements

Part of this work was funded by the Department of Energy at Los Alamos National Laboratory under contract DE-AC52-06NA25396 through the Laboratory Directed Research and Development Program.

References

1. Bonato, A.: A survey of models of the web graph. In: López-Ortiz, A., Hamel, A.M. (eds.) CAAN 2004. LNCS, vol. 3405, pp. 159–172. Springer, Heidelberg (2005)
2. Abello, J., Pardalos, P.M., Resende, M.G.C. (eds.): Handbook of massive data sets. Kluwer Academic Publishers, Norwell, MA (2002)
3. Faloutsos, M., Faloutsos, P., Faloutsos, C.: On power-law relationships of the internet topology. In: SIGCOMM 1999: Proceedings of the conference on Applications, technologies, architectures, and protocols for computer communication, pp. 251–262. ACM Press, New York (1999)
4. Kleinberg, J.M., Kumar, R., Raghavan, P., Rajagopalan, S., Tomkins, A.S.: The web as a graph: Measurements, models, and methods. In: Asano, T., Imai, H., Lee, D.T., Nakano, S.-i., Tokuyama, T. (eds.) COCOON 1999. LNCS, vol. 1627, pp. 1–17. Springer, Heidelberg (1999)
5. Kumar, R., Raghavan, P., Rajagopalan, S., Sivakumar, D., Tomkins, A., Upfal, E.: Stochastic models for the web graph. In: FOCS 2000: Proceedings of the 41st Annual Symposium on Foundations of Computer Science, Washington, DC, USA, p. 57. IEEE Computer Society, Los Alamitos (2000)
6. Barabási, A.L., Albert, R.: Emergence of scaling in random networks. Science 286, 509–512 (1999)
7. Aiello, W., Chung, F., Lu, L.: A random graph model for massive graphs. In: STOC 2000: Proceedings of the thirty-second annual ACM symposium on Theory of computing, pp. 171–180. ACM Press, New York (2000)
8. Bollobás, B., Riordan, O., Spencer, J., Tusnády, G.: The degree sequence of a scale-free random graph process. Random Struct. Algorithms 18, 279–290 (2001)
9. Cooper, C., Frieze, A.M.: A general model of undirected Web graphs. In: Meyer auf der Heide, F. (ed.) ESA 2001. LNCS, vol. 2161, pp. 500–511. Springer, Heidelberg (2001)
10. Bradonjić, M., Kong, J.: Wireless ad hoc networks with tunable topology. In: Proceedings of the 45th Annual Allerton Conference on Communication, Control and Computing (to appear, 2007)
11. Erdős, P., Rényi, A.: On random graphs. Publ. Math. Inst. Hungar. Acad. Sci. (1959)
12. Erdős, P., Rényi, A.: On the evolution of random graphs. Publ. Math. Inst. Hungar. Acad. Sci. (1960)
13. Penrose, M.D.: Random Geometric Graphs. Oxford University Press, Oxford (2003)
14. Mahadev, N.V.R., Peled, U.N.: Threshold graphs and related topics. Annals of discrete mathematics, vol. 56. Elsevier, New York (1995)
15. Hagberg, A., Swart, P.J., Schult, D.A.: Designing threshold networks with given structural and dynamical properties. Phys. Rev. E 74, 056116 (2006)
16. Masuda, N., Miwa, H., Konno, N.: Geographical threshold graphs with small-world and scale-free properties. Physical Review E 71, 036108 (2005)
17. Alon, N., Spencer, J.H.: The probabilistic method, 2nd edn. John Wiley & Sons, Inc., New York (2000)
18. Gupta, P., Kumar, P.R.: Critical power for asymptotic connectivity. In: Proceedings of the 37th IEEE Conference on Decision and Control, vol. 1, pp. 1106–1110 (1998)

Appendix

Degree Distribution

The nodes are placed into the unit torus. W.l.o.g. let us consider the degree of the node v_1. Let the weight vector be $\mathbf{w}$. Let the position vector of the nodes be $\mathbf{x}$. It is straightforward to show that the probability of v_1 having degree k, given weights $\mathbf{w}$, is

$$\Pr[d_1 = k|\mathbf{w}] = \binom{n-1}{k} \prod_{i=2}^{k+1} \text{Area}(B(x_i, r_{i1})) \prod_{j=k+2}^{n} (1 - \text{Area}(B(x_j, r_{j1}))), \quad (24)$$

where $\text{Area}(B(x_i, r_{i1}))$ is the area of the ball at center x_i with radius r_{i1}, and due to (3) the radii are given by

$$r_{i1} = \sqrt{\frac{w_1 + w_i}{\theta_n}} \quad (25)$$

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

$$\begin{aligned} \Pr[d_1 = k|w_1] &= \Big(\prod_{i=2}^{n} \int_{w_i} dw_i f(w_i)\Big) \Pr[d_1 = k|\mathbf{w}] \\ &= \binom{n-1}{k} \Big(\int_w dw f(w) \frac{(w_1 + w)\pi}{\theta_n}\Big)^k \Big(1 - \int_w dw f(w) \frac{(w_1 + w)\pi}{\theta_n}\Big)^{n-1-k} \\ &= \binom{n-1}{k} \Big(\frac{(w_1 + \mu)\pi}{\theta_n}\Big)^k \Big(1 - \frac{(w_1 + \mu)\pi}{\theta_n}\Big)^{n-1-k} \\ &\to e^{-\lambda} \frac{\lambda^k}{k!}, \end{aligned}$$

where

$$\lambda = (w_1 + \mu) n\pi / \theta_n. \quad (26)$$

That is, the degree distribution of a node with weight w, in the limit follows the Poisson distribution

$$d(k|w) \sim Po((w + \mu) n\pi / \theta_n). \quad (27)$$