Geodetic number of random graphs of diameter 2

Gab-Byung Chae, Edgar M. Palmer and Wai-Cheong Siu

Department of Mathematics
Michigan State University
East Lansing, MI, 48824
U.S.A.
chaegabb@pilot.msu.edu
palmer@math.msu.edu
siuwaich@pilot.msu.edu

Abstract

Buckley and Harary introduced several graphical invariants related to convexity theory, such as the geodetic number of a graph. These invariants have been the subject of much study and their determination has been shown to be *NP-hard*. We use the probabilistic method developed by Erdös to determine the asymptotic behavior of the geodetic number of random graphs with fixed edge probability. As a consequence we have a random greedy algorithm for a good approximation of a geodetic basis of a given graph G. Our technique can be applied to other random graphs of diameter 2 and to random digraphs.

1 Introduction

A graph G consists of a finite set $V = V(G)$ of n *vertices* and a set $E = E(G)$ of unordered pairs of vertices called *edges*. Let S be any subset of the vertex set V of a graph G. Then the *closure* of S, denoted $C(S)$, consists of all vertices w such that there exist vertices u and v in S such that w lies on a geodesic between u and v. Therefore $S \subseteq C(S)$ along with $C(\{u, v\})$ for all pairs of non-adjacent vertices u, v in S. The set S is *convex* if it is *closed*, i.e. $S = C(S)$. If $C(S) = V(G)$, then S is called a *geodetic cover* of G. Such a set of minimum cardinality is called a *geodetic basis* of G. The *geodetic number* of a graph, denoted $gn(G)$, is the cardinality of a geodetic basis. For example, the geodetic number of the Kuratowski graph $K_{3,3}$ is 3 while that of the Petersen graph is 4. This parameter was introduced by Harary and Buckley in their book [2] and has been the subject of much recent study. See [3], [4], [8], [9] and [10].

The determination of $gn(G)$ was found to be *NP-hard* by Harary, Loukakis and Tsouros [13]. We have found a random greedy algorithm that is very effective for approximating a geodetic basis of a random graph $G_{n,p}$ of order n with fixed edge

Australasian Journal of Combinatorics 26(2002), pp.11–20

probability p. Our proof is made by showing that $gn(G_{n,p})$ is asymptotic to $\log_b n$ where $b = 1/(1-p)$. The methods also apply to other random graphs of diameter 2 and to random digraphs.

There are several useful references for background material not provided here. We refer the reader to the books by Palmer [17] and Bollobás [1] for relevant theory of random graphs. For graph theory and graph algorithms see Chartrand and Oellermann [7]. Chartrand and Lesniak [6] and West [18] cover both graphs and random graphs.

2 Greedy geodetic cover algorithm

We begin with a greedy algorithm that always produces a geodetic cover and can be applied to any graph G. We assume that the input is in the form of an adjacency matrix. The algorithm recursively examines non-adjacent vertices to build the cover set.

GREEDY GEODETIC COVER ALGORITHM (GGCA)

Step 0. $H \leftarrow G$ and $S \leftarrow \phi$.

Step 1. IF H is a complete graph, $S \leftarrow S \cup V(H)$. EXIT.

Step 2. Choose two non-adjacent vertices u, v in $V(H)$
and $S \leftarrow S \cup \{u, v\}$.
IF $C(S) = V(G)$, EXIT.
ELSE let H be the subgraph of G induced by the vertex set
$V(G) - C(S)$ and GO to Step 1.

The algorithm always terminates with a geodetic cover in S. Observe that Step 2 is executed at most $n/2$ times. To find the closure $C(\{u, v\})$ of a pair of non-adjacent vertices u, v, one can build two breadth-first search (BFS) trees rooted at u and v. If the distance between u and v, denoted $d(u, v)$, is finite, then the closure consists of all vertices w such that $d(u, w) + d(w, v) = d(u, v)$. The BFS complexity is $O(n^2)$ for our input and since the two BFS trees supply the necessary distance information, the number of operations to determine $C(\{u, v\})$ is also $O(n^2)$. Step 2 may demand the determination of the closure of non-adjacent vetices as many as $2|S| + 1$ times. And so altogether Step 2 requires at most

$$\binom{n}{2} + O(n^2 |S|)$$

operations.

Thus the overall worst-case complexity of the algorithm is $O(n^4)$. This estimate can be reduced slightly with a bit more effort. Note that all BIG OH and little oh notation is used with respect to n, the number of vertices in our graphs.

Note that in Step 2 we could have enlarged set S one vertex at a time. The algorithm would have produced a geodetic cover with the same worst case complexity. But the analysis below is easier when we add two vertices at a time and furthermore the estimate of the size of the cover is the same.

Next we apply our greedy algorithm *GGCA* to a random graph $G_{n,p}$ of order

n with fixed edge probability p with $0 < p < 1$. It is well known (see [1] or [17]) that $G_{n,p}$ almost surely has diameter 2. Therefore after Step 2 is applied for the first time, almost surely S will consist of two non-adjacent vertices and there will be approximately $p^2 n$ vertices in the closure $C(S)$ of S. And there will be about $(1 - p^2)n$ vertices outside the closure of S. As the algorithm continues to execute Step 2, after j passes the cardinality of S is $2j$ and there are at most approximately $(1-p^2)^j n$ vertices remaining outside the closure of S. Suppose we halt the algorithm after j steps and simply set $S \leftarrow S \cup (V - C(S))$. Then S will be a geodetic cover with at most approximately $2j + (1 - p^2)^j n$ vertices. It will be seen at once that if $j = \log_a(n)$ with $a = 1/(1 - p^2)$ then $|S|$ is $O(\log(n))$. (Note that although we frequently deal with integer values such as j, we do not always round off expressions for these variables because the proofs remain valid without the distraction of the extra notation.)

Note that the algorithm is allowed to halt after Step 1 if H is a complete graph. A famous theorem of Matula (see [14] or [15]) shows that the clique number of $G_{n,p}$ is asymptotic to $2\log_b(n)$ where $b = 1/p$. Details are also available in both [1] and [17]. So if the number of vertices outside $C(S)$ is much more than $2\log_b(n)$, this exit will almost surely not be taken.

The outcome of these observations is summarized next.

Proposition 1 *Given a random graph $G_{n,p}$ of order n with fixed edge probability p, the random greedy geodetic covering algorithm almost surely finds a geodetic cover of order $O(\log(n))$, i.e. $gn(G_{n,p}) = O(\log(n))$.*

We could also apply the algorithm to other families of random graphs but we will see that Proposition 1 provides a target value for the geodetic number of a random graph which leads to better results in the next section.

3 The geodetic number of a random graph of diameter 2

In this investigation our goal was to determine the asymptotic behavior of the geodetic number of a random graph with fixed edge probability. Taking a cue from the Proposition in the previous section, we found a lower bound for $gn(G_{n,p})$ of the form $c_1 \log(n)$ with c_1 as large as possible. Then we turned our attention to the problem of establishing an upper bound of nearly the same value. The result was the following theorem in which the two bounds are asymptotically equal. The proof is straightforward and uses techniques well-known to experts in the field but we have filled in extra steps to reveal the method of discovery.

Theorem 1 *Let $G_{n,p}$ be a random graph of order n with fixed edge probability p and define $b = \frac{1}{1-p}$. Almost surely:*

$$gn(G_{n,p}) = (1 + o(1)) \log_b n.$$

Proof. First we determine a lower bound for $gn(G_{n,p})$. Let $k = (1-\varepsilon_n)\log_b n$ where $\varepsilon_n \to 0$ as $n \to \infty$ and ε_n will be defined later. Next, for any graph G with vertex set V of order n, we define the random variable $X = X(G)$ to be the number of k-subsets of $V = V(G)$ which are geodetic covers. Then the expected number $E[X]$ of geodetic covers of order k for $G_{n,p}$ is

$$E[X] = \sum_{\substack{S \subseteq V \\ |S|=k}} P(C(S) = V). \tag{1}$$

But since these random graphs have diameter 2 almost surely, when $C(S) = V$, each vertex in $V - S$ must have at least 2 neighbors in S. Hence $P(C(S) = V)$ is bounded above by the probability that each vertex in $V - S$ has at least 2 neighbors in S. Therefore we have almost surely

$$\begin{aligned} E[X] &\leq \binom{n}{k}(1-(1-p)^k - kp(1-p)^{k-1})^{n-k} \\ &\leq \binom{n}{k}(1-kp(1-p)^{k-1})^{n-k} \\ &\leq \frac{n^k}{k!} exp\{-(n-k)kp(1-p)^{k-1}\} \\ &= \frac{f(n)^k}{k!}, \end{aligned} \tag{2}$$

where

$$f(n) = n\, exp\{-(n-k)p(1-p)^{k-1}\}. \tag{3}$$

We emphasize that the equations (2) rely on the random graph having diameter 2, and so are conditional. They can be made unconditional by adding on the right side

$$\binom{n}{k} P(d(G_{n,p}) \neq 2) \leq \binom{n}{k}\binom{n}{2}(1-p)(1-p^2)^{n-2} = o(1)$$

where $d(G)$ is the diameter of G. The last equality holds for $k = O(\log_b n)$ and so this omission posses no difficulty.

Now we can finish this part of the proof by defining ε_n so that $f(n)$ is bounded above by a constant independent of n. First note that in our notation

$$(1-p)^k = n^{-1+\varepsilon_n} \tag{4}$$

and so if $\varepsilon_n \to 0$ as $n \to \infty$, then

$$kp(1-p)^{k-1} = o(1). \tag{5}$$

Therefore

$$\log f(n) = \log n - \frac{p}{1-p} n^{\varepsilon_n} + o(1). \tag{6}$$

Observe that we could now define ε_n by

$$n^{\varepsilon_n} = \frac{1-p}{p} \log n, \tag{7}$$

but it is convenient to pick a small $\varepsilon > 0$ and let

$$\varepsilon_n = (1+\varepsilon)\frac{\log \log n}{\log n} = (1+\varepsilon)\frac{\log_b \log n}{\log_b n}. \tag{8}$$

Now it can be seen that with ε_n and hence k so defined, we have the expectation $E[X] \to 0$ as $n \to \infty$. Therefore $G_{n,p}$ almost surely has no geodetic covers of order k. That is, almost surely

$$gn(G_{n,p}) \geq (1-\varepsilon_n)\log_b n = \log_b n - (1+\varepsilon)\log_b \log n. \tag{9}$$

Now we turn to the upper bound. Set $k = (1+\varepsilon_n)\log_b n$ where $\varepsilon_n \to 0$ as $n \to \infty$ with ε_n to be defined later. Next let S be a subset of $V = V(G_{n,p})$ of order k chosen uniformly at random and define the random variable X_S by

$$X_S(G) = |\{v : v \notin C(S)\}|. \tag{10}$$

Thus $E[X_S]$ is the expected number of vertices that lie outside the closure of S. We will show that $E[X_S] \to 0$ as $n \to \infty$. It then follows that almost surely no vertices lie outside the closure of S. We begin with a formula for $P(v \notin C(S))$, i.e. the probability that a vertex v is not in $C(S)$. For convenience and to emphasize that it depends on k, we denote this probability by α_k. We find that

$$\alpha_k = \sum_{i=0}^{k} \binom{k}{i} p^{i+\binom{i}{2}}(1-p)^{k-i}. \tag{11}$$

To verify the formula, consider a vertex $v \in V - S$ with exactly i neighbors in S. The neighbors can be chosen in $\binom{k}{i}$ ways. The probability that the vertex has exactly the i neighbors chosen is $p^i(1-p)^{k-i}$. But the neighbors must be mutually adjacent and that probability is $p^{\binom{i}{2}}$. Since these probabilities are independent, we multiply them before summing over all $i = 0$ to k to obtain the result above.

Then we have for the expectation:

$$E[X_S] = (n-k)\alpha_k \leq n\alpha_k \tag{12}$$

and so we need a nice upper bound for α_k. To do this we split the sum in α_k in two parts and focus first on the lower portion where $0 \leq i \leq s = c\log k$, and c is a constant that depends on p and will be chosen shortly. Then for the lower portion of the sum we have the following upper bound, which holds for all n sufficiently large:

$$\sum_{i=0}^{s} \binom{k}{i} p^{i+\binom{i}{2}}(1-p)^{k-i} \leq \sum_{i=0}^{s} \binom{k}{i} p^{i}(1-p)^{k-i} \leq \frac{3}{2}\binom{k}{s} p^{s}(1-p)^{k-s}. \tag{13}$$

The second (crude) inequality of (13) follows from well known estimates of the tail of the binomial distribution (see [17], p. 133).

Now we look at the upper portion of the sum in (11) and show that it is dominated by the right side of (13). Here are some of the steps:

$$\sum_{i=s+1}^{k} \binom{k}{i} p^{i+\binom{i}{2}}(1-p)^{k-i} \leq (1-p)^k \sum_{i=s+1}^{k} \frac{k^i}{i!}\left(\frac{p^{\frac{i+1}{2}}}{1-p}\right)^i$$
$$= (1-p)^k \sum_{i=s+1}^{k} \frac{\left(\frac{kp^{\frac{i+1}{2}}}{1-p}\right)^i}{i!} \qquad (14)$$
$$= O(1)(1-p)^k \frac{\left(\frac{kp^{\frac{s+2}{2}}}{1-p}\right)^{s+1}}{(s+1)!},$$

where the last equality holds if the constant c in the definition of s is large enough so that

$$\frac{kp^{\frac{s+2}{2}}}{1-p} < 1. \qquad (15)$$

To show that the right side of (13) dominates, just divide the right side of (14) by the right side of (13) and note that the quotient is negligible provided that

$$kp^{\frac{s}{2}} < 1. \qquad (16)$$

Now it can be seen that if $c = \frac{3}{\log \frac{1}{p}}$ and $s = c \log k$ then,

$$\alpha_k \leq 2\binom{k}{s} p^s (1-p)^{k-s} \qquad (17)$$

for sufficiently large n.

Here is another interesting way to establish this bound on α_k that we have heard about. It makes use of the important estimate mentioned earlier of the clique size of a random graph first found by Matula (see [14], [15] and also the texts [1] or [17]). In a random graph of order k with edge probability p the order of a largest clique is asymptotic to $2\frac{\log k}{\log \frac{1}{p}}$. Therefore the clique number of the random set S of order k is at most $3\frac{\log k}{\log \frac{1}{p}} = s$. Now suppose that a vertex v in $V - S$ is not covered by S, i.e. $v \notin C(S)$. Then v has at most s neighbors in S. Therefore

$$\alpha_k \leq \sum_{i=0}^{s} \binom{k}{i} p^i (1-p)^{k-i} \qquad (18)$$

which implies the bound in the previous inequality.

From our definitions of s and k we have for sufficiently large n

$$\begin{aligned} E[X_S] &\leq n\alpha_k \\ &\leq 2\,n\frac{(kp)^s}{s!}(1-p)^{k-s} \\ &= 2\,\frac{g(n)^s}{s!}, \end{aligned} \tag{19}$$

where

$$g(n) = \frac{kp}{1-p}n^{-\epsilon_n/s}. \tag{20}$$

Now we can finish by defining ϵ_n so that $g(n)$ is bounded by a constant independent of n. Since p is fixed we only need

$$kn^{-\epsilon_n/s} = O(1). \tag{21}$$

On taking the log of both sides and using the definitions of k and s it can be seen that ϵ_n can be defined by

$$\epsilon_n = \frac{(3+\varepsilon)}{\log\frac{1}{p}}\frac{(\log\log_b n)^2}{\log n}. \tag{22}$$

where $\epsilon > 0$ is arbitrary. Then almost surely

$$\begin{aligned} gn(G_{n,p}) &\leq (1+\varepsilon_n)\log_b n \\ &= \log_b n + \frac{3+\varepsilon}{\log\frac{1}{p}\log\frac{1}{1-p}}(\log\log_b n)^2. \quad \square \end{aligned} \tag{23}$$

The proof suggests an improved heuristic greedy geodetic cover algorithm that can be applied to a random graph $G_{n,p}$.

IMPROVED GREEDY GEODETIC COVER ALGORITHM (IGGCA)

Step 0. Let $\varepsilon > 0$ be given.

Step 1. Choose $k = (1+\varepsilon)\log_b n$ vertices at random.
and put them in set S.

Step 2. WHILE there is a vertex v in $V - S$ which does not have two
non-adjacent neighbors in S, DO $S \leftarrow S \cup \{v\}$.

The theorem shows that when applied to $G_{n,p}$ the algorithm IGGCA will almost surely terminate with a geodetic cover without putting *any* new vertices in S. It also shows that the cardinality of the cover will be the same order of magnitude as a geodetic basis. The computational effort is $O(n)$ for Step 1. To implement Step 2, consider each of the $n-k$ vertices outside S and determine if they have a pair of non-adjacent common neighbors in S. This requires about $O(n(\log n)^2)$ computations. So the overall worst-case complexity is $O(n(\log n)^2)$.

The random graph threshold for diameter 2 is well-known and easy to establish. It is a fact that if the edge probability p is given by

$$p^2 = \frac{2\log n + \omega_n}{n}, \tag{24}$$

where $\omega_n \to \infty$, but also

$$n^2(1-p) \to \infty, \tag{25}$$

then almost surely $G_{n,p}$ has diameter 2. The next theorem is a result that covers a range of edge probability at least as large as this threshold. But we redefine the edge probability as follows:

$$p^2 = \frac{2\log n + \omega_n}{n^{1-c}} \tag{26}$$

where $\omega_n = o(\log(n))$ and c is a constant between 0 and 1.

Theorem 2 *Let $G_{n,p}$ be a random graph of order n and let c be a positive constant less than 1. Then with edge probability p given by formula (26), almost surely*

$$(\frac{1+c}{2} - o(1))\frac{\log n}{p} \leq gn(G_{n,p}) \leq (1+o(1))\frac{\log n}{p}.$$

The proof requires a bit more effort than the previous theorem but since the method is similar, it is omitted. The corresponding problem for random graphs of diameter greater than 2 is considerably more complicated and remains unsolved.

4 Conclusion

Buckley and Harary discussed several other graph invariants whose values can be determined by our theorems for some families of random graphs. For example, the *hull number* of a graph was introduced by Everett and Seidman [12] and studied further in [5], [11] and [16]. Let S be a subset of the vertex set $V = V(G)$ of the graph G. Following [2] we denote the iteration of the closure operation in the usual way so that $C(C(S)) = C^2(S)$ and $C(C(C(S))) = C^3(S)$, etc. Then the *geodetic iteration number* of S, denoted $gin(S)$, is the smallest integer i such that $C^i(S) = C^{i+1}(S)$. The *geodetic iteration number* of G, denoted by $gin(G)$ is the maximum value of $gin(S)$ taken over all subsets S of V. For example, it is easy to verify that $gin(K_{2,3}) = 2$.

The *convex hull* of S, denoted by $H(S)$, is $C^i(S)$ where $i = gin(G)$. Then the *hull number* of G, denoted by $h(G)$ is the smallest cardinality of a set S whose hull is V. For example, the hull number of the Petersen graph is 3.

Corollary 4.1 *Let $G_{n,p}$ be a random graph of order n with edge probability p fixed and $0 < p < 1$. Almost surely*

$$h(G_{n,p}) = gin(G_{n,p}) = 2.$$

Proof. Consider two non-adjacent vertices u and v of $G_{n,p}$. The following statements all hold almost surely. First the cardinality of their common neighborhood, i.e.

$|C(\{u,v\})|$, is approximately p^2n. But since p is fixed, $p^2n >> \log n$ and so it follows from the proof of the upper bound in our first theorem that $C^2(\{u,v\}) = V$. Thus the hull number is 2 because $|S| = |\{u,v\}| = 2$ and the geodetic iteration number is 2 because the closure operation must be performed twice. $\square$

The methods used above can also be applied to random digraphs. But a number of problems involving convexity in random graphs remain. For example, can the result of Theorem 2 be sharpened? And what is the behavior of the geodetic number for random graphs of diameter greater than 2? Even the determination of other invariants like the iteration number or the hull number remains unsolved for random graphs of diameter 2.

Acknowledgment: We have had many interesting and fruitful discussions about this problem area with a number of colleagues and friends. We would especially like to thank Gary Chartrand and Ping Zhang for calling our attention to the geodetic invariants, Abdol Esfahanian for his helpful suggestions and the referee of this paper for recommendations that helped to improve the clarity and accuracy of the exposition.

References

[1] B. Bollobás, *Random Graphs*, Academic, London (1985).

[2] F. Buckley and F. Harary, *Distance in Graphs*, Addison-Wesley, Reading (1990).

[3] G. Chartrand, F. Harary and P. Zhang, On the geodetic number of a graph, *Networks*, **39** (2002) 1–6.

[4] G. Chartrand, F. Harary and P. Zhang, Extremal problems in geodetic graph theory, *Congress. Numer.*, **130** (1998) 157–168.

[5] G. Chartrand, F. Harary and P. Zhang, On the hull number of a graph, *Ars Combin.*, **57** (2000) 129–138.

[6] G. Chartrand and L. Lesniak, *Graphs and Digraphs*, 3rd edition, Chapman and Hall, London (1996).

[7] G. Chartrand and O. Oellermann, *Applied and Algorithmic Graph Theory*, McGraw-Hill, New York (1993).

[8] G. Chartrand and P. Zhang, The forcing geodetic number of a graph, *Discuss. Math. Graph Theory*, **19** (1999) 45–58.

[9] G. Chartrand and P. Zhang, Realizable ratios in graph theory: geodesic parameters, *Bull. Inst. Combin. Appl.*, **27** (1999) 69–80.

[10] G. Chartrand and P. Zhang, The geodetic number of an oriented graph, *Europ. J. Combinatorics*, **21** (2000) 181–189.

[11] G. Chartrand and P. Zhang, The forcing hull number of a graph, *J. Combin. Math. Combin. Comput.*, **36** (2001) 81–94.

[12] M.G. Everett and S.B. Seidman, The hull number of a graph, *Discrete Math.* **57** (1985) 217–223.

[13] F. Harary, E. Loukakis and C. Tsouros, The geodetic number of a graph, *Math. Comput. Modeling* **17** (1993) 89–95.

[14] D.W. Matula, On the complete subgraphs of a random graph,*Proc. 2nd Chapel Hill Conf. Combinatorial Math. and its Applications* (R.C. Bose, et al., eds.) Univ. North Carolina, Chapel Hill (1970) 356–369.

[15] D.W. Matula, The employee party problem, *Notices Amer. Math. Soc.*, **19** (1972) A-382.

[16] H.M. Mulder, The expansion procedure for graphs, *Contemporary Methods in Graph Theory* (R. Bodendiek, ed.) Wissenschaftsverlag, Mannheim (1990) 459–477.

[17] E.M. Palmer, *Graphical Evolution: an introduction to the theory of random graphs*, Wiley-Interscience, New York (1985).

[18] D. West, *Introduction to Graph Theory*, Prentice Hall, Englewood Cliffs (1996).

(Received 26/2/2001)