Statistics and Probability Letters 195 (2023) 109779

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

On the time to identify the nodes in a random graph

Jonathan R. Stewart

Department of Statistics, Florida State University, 117 N Woodward Ave, Tallahassee, FL 32306-4330, United States of America

ARTICLE INFO

Article history:
Received 16 December 2021
Received in revised form 30 December 2022
Accepted 31 December 2022
Available online 11 January 2023

Keywords:
Sampling times
Random graphs
Egocentric network sampling
Erdős-Rényi random graph

ABSTRACT

The sampling of networks is an important problem at the frontier of statistical network analysis, and the identification of population members of a network is an important step in the sampling process. In this work, we study the random time τ to identify the nodes in an Erdős-Rényi random graph through egocentric sampling We derive the exact distribution of τ and give an exact formula for computing the mean time $\mathbb{E}\tau$ as a function of the size of the network. We explore how $\mathbb{E}\tau$ varies with the size of the network, the probability of edges, and network sparsity. We establish the scaling of τ with network size in both sparse and dense random graphs, highlighting special cases that demonstrate sub-linear scaling of τ with the size of the network. All theoretical results are non-asymptotic. Lastly, we discuss possible extensions to classes of random graphs beyond Erdős-Rényi random graphs.

1. Introduction

We consider simple, undirected random graphs denoted by $\boldsymbol{X}$ which are defined on a finite set of $N \geq 2$ nodes, which without loss is taken to be the set $\mathcal{N} = \{1, \ldots, N\}$. We set $X_{i,j} = 1$ if nodes $i \in \mathcal{N}$ and $j \in \mathcal{N}$ are connected in the graph and set $X_{i,j} = 0$ otherwise, making the standard conventions that $X_{i,j} = X_{j,i}$ and $X_{i,i} = 0$ $(1 \leq i < j \leq N)$.

An important problem at the frontier of statistical network analysis is the sampling of networks (Kolaczyk, 2017), where an observation process produces an observed subgraph of a larger population graph (Frank, 2005; Handcock and Gile, 2010). One of the most prevalent sampling designs for network data is egocentric sampling (Perry et al., 2018; Krivitsky and Morris, 2017), where individual nodes $i \in \mathcal{N}$ are sampled at random and the edges of each sampled node are observed, i.e., if node i is sampled then edges $X_{i,j}$ $(j \in \mathcal{N} \setminus \{i\})$ are observed. Concretely, an egocentric sample of the network $\boldsymbol{X}$ based on the random sample of nodes $\mathcal{S} \subseteq \mathcal{N}$ produces an observed subgraph $\boldsymbol{X}_{\mathcal{S}}$ given by

$$\boldsymbol{X}_{\mathcal{S}} = (X_{i,j} \,:\, \text{all } \{i,j\} \subset \mathcal{N} \text{ satisfying } \mathcal{S} \cap \{i,j\} \neq \emptyset). \tag{1}$$

In practice, egocentric sampling of networks is carried out through survey mechanisms which typically report only the connections of each sampled node $i \in \mathcal{S}$ in the network, that is the nodes $j \in \mathcal{N}$ satisfying $X_{i,j} = 1$ for each $i \in \mathcal{S}$. Due to the binary sample space of the edge variables, observing all edges connecting nodes $i \in \mathcal{S}$ to nodes $j \in \mathcal{N} \setminus \{i\}$ in the graph means we completely observe all edge variables $X_{i,j}$ between nodes $i \in \mathcal{S}$ and $j \in \mathcal{N} \setminus \{i\}$. Egocentric sampling designs for networks, such as link-tracing, have been used in respondent driven sampling designs as a means to sample hard-to-reach populations (Gile and Handcock, 2010). Understanding the time it takes to sample and identify nodes in a population graph can have important implications for studies that employ respondent driving methods for obtaining a sample from a population.

E-mail address: jrstewart@fsu.edu.

https://doi.org/10.1016/j.spl.2023.109779
0167-7152/

We set our problem up as follows. Let $\boldsymbol{\pi}(\mathcal{N}) = (\pi(1), \ldots, \pi(N))$ be a random permutation of the node indices $(1, \ldots, N)$. In what follows, we will consider iteratively sampling nodes at random from the set of nodes $\{1, \ldots, N\}$. The random permutation $\boldsymbol{\pi}(\mathcal{N})$ prescribes a randomized sampling order. Consider an egocentric sample obtained from a set of sampled nodes $\mathcal{S} \subseteq \mathcal{N}$, assuming that sampled nodes $i \in \mathcal{S}$ are identified. Define

$$\mathcal{E}_i = \left\{ j \in \mathcal{N} \setminus \{i\} \,:\, X_{i,j} = 1 \right\}, \qquad i = 1, \ldots, N, \tag{2}$$

to be the nodes in $\mathcal{N} \setminus \{i\}$ which are connected to node $i \in \mathcal{N}$ in $\boldsymbol{X}$. We assume that each node $j \in \mathcal{E}_i$ is identified. In other words, $\mathcal{E}_i$ represents the data obtained through the egocentric sample of node i. Note that $\mathcal{E}_i$ is a random set since $\boldsymbol{X}$ is a random graph. It is possible – and generally quite likely – that $\mathcal{E}_i \not\subseteq \mathcal{S}$ for $i \in \mathcal{S}$, as the set $\mathcal{E}_i$ may contain nodes which are not in $\mathcal{S}$ since we observe and identify all nodes $i \in \mathcal{S}$ and all nodes $j \in \mathcal{N} \setminus \mathcal{S}$ satisfying $X_{i,j} = 1$ for any $i \in \mathcal{S}$. Throughout, we assume the identification of all nodes is possible.

We consider iteratively sampling nodes in the node set $\{1, \ldots, N\}$ at random at discrete time steps $t \in \{1, \ldots, N\}$ by iterating through the random permutation $\boldsymbol{\pi}(\mathcal{N})$. For each such sampled node i, we produce a partial observation of the network $\mathcal{E}_i$ by producing an egocentric sample of the network based on node i. We define

$$\mathcal{N}(t) = \left[\bigcup_{k=1}^{t} \mathcal{E}_{\pi(k)} \right] \cup \{\pi(1), \ldots, \pi(t)\}, \qquad t = 1, \ldots, N, \tag{3}$$

to be the set of nodes identified up to the tth time step, and define

$$\tau = \min \{t^\star \,:\, \mathcal{N}(t^\star) = \mathcal{N}\} \tag{4}$$

to be the random time at which all nodes in the graph have been identified through egocentric sampling.

2. The time to identify all of the nodes in an Erdő s-Rényi random graph

We derive the distribution of τ when $\boldsymbol{X}$ is an Erdős-Rényi random graph (Erdős and Rényi, 1959; Gilbert, 1959; Frieze and Karoński, 2016) and calculate the expected time $\mathbb{E}\,\tau$ to identify all nodes in the graph through egocentric sampling in Theorem 1. An Erdős-Rényi random graph assumes edge variables $X_{i,j}$ ($\{i,j\} \subset \mathcal{N}$) are mutually independent and the probability of an edge is specified by a parameter $p \in (0,1)$, i.e., $\mathbb{P}(X_{i,j} = 1) = p$ for all $\{i,j\} \subset \mathcal{N}$. We will here consider homogeneous Erdős-Rényi random graphs for ease of presentation, but note that the methods we employ could be extended to inhomogeneous Erdős-Rényi random graphs and other classes of random graphs. Fig. 1 visualizes values of $\mathbb{E}\,\tau$ for various edge probabilities and over a range of network sizes. The effect of network sparsity on the distribution of τ is discussed in Section 2.1 and visualized in Fig. 2. Of note, both Figs. 1 and 2 suggest a sub-linear scaling of $\mathbb{E}\,\tau$ with N, in certain scenarios. This phenomena is explored rigorously in the next section.

Theorem 1. *Let $\boldsymbol{X}$ be an Erdős-Rényi random graph defined on the set of nodes $\mathcal{N} = \{1, \ldots, N\}$ with edge probability $p \in (0,1)$ and let τ be defined as in (4). Then the distribution of τ is given by*

$$\mathbb{P}(\tau \leq t) = \left[1 - (1-p)^t\right]^{N-t}, \qquad 1 \leq t \leq N.$$

The expected time to identify all nodes in the random graph via egocentric sampling is given by

$$\mathbb{E}\,\tau = N + 1 - \sum_{t=1}^{N} [1 - (1-p)^t]^{N-t}.$$

Proof of Theorem 1. We first prove the distribution of τ. Consider iteratively sampling nodes in $\{1, \ldots, N\}$ at random. It suffices to construct a random permutation of the node indices $(1, \ldots, N)$ denoted by $\boldsymbol{\pi}(\mathcal{N}) = (\pi(1), \ldots, \pi(N))$ and iterate through $\boldsymbol{\pi}(\mathcal{N})$, observing the edges each node $\pi(t)$ is connected to at step $t \geq 1$, i.e., observing the random set $\mathcal{E}_{\pi(t)}$. At step t, we have sampled nodes $\pi(1), \ldots, \pi(t)$, and have therefore identified each such node. The remaining nodes $\mathcal{N} \setminus \{\pi(1), \ldots, \pi(t)\}$ are identified at step t if, for each node $v \in \{\pi(t+1), \ldots, \pi(N)\}$, at least one of the nodes $\pi(1), \ldots, \pi(t)$ is connected to v in $\boldsymbol{X}$. For each $v \in \{\pi(t+1), \ldots, \pi(N)\}$, let $\mathcal{A}_{t,v}$ be the event none of the nodes $\pi(1), \ldots, \pi(t)$ are connected to v. The probability of event $\mathcal{A}_{t,v}$ is then given by

$$\mathbb{P}(\mathcal{A}_{t,v}) = \mathbb{P}\left(X_{\pi(1),v} = 0, \ldots, X_{\pi(t),v} = 0\right) = \prod_{k=1}^{t} \mathbb{P}\left(X_{\pi(k),v} = 0\right) = (1-p)^t, \tag{5}$$

owing to the independence of edge variables in an Erdős-Rényi random graph and since $\mathbb{P}(X_{i,j} = 1) = p \in (0,1)$ for all pairs of nodes $\{i,j\} \subset \mathcal{N}$. The event $\tau \leq t$ occurs if and only if the event $\mathcal{A}^c_{t,\pi(t+1)} \cap \ldots \cap \mathcal{A}^c_{t,\pi(N)}$ occurs, i.e., every node in $\{\pi(t+1), \ldots, \pi(N)\}$ must be connected in $\boldsymbol{X}$ to at least one node in $\{\pi(1), \ldots, \pi(t)\}$ for the event $\tau \leq t$ to occur. Note

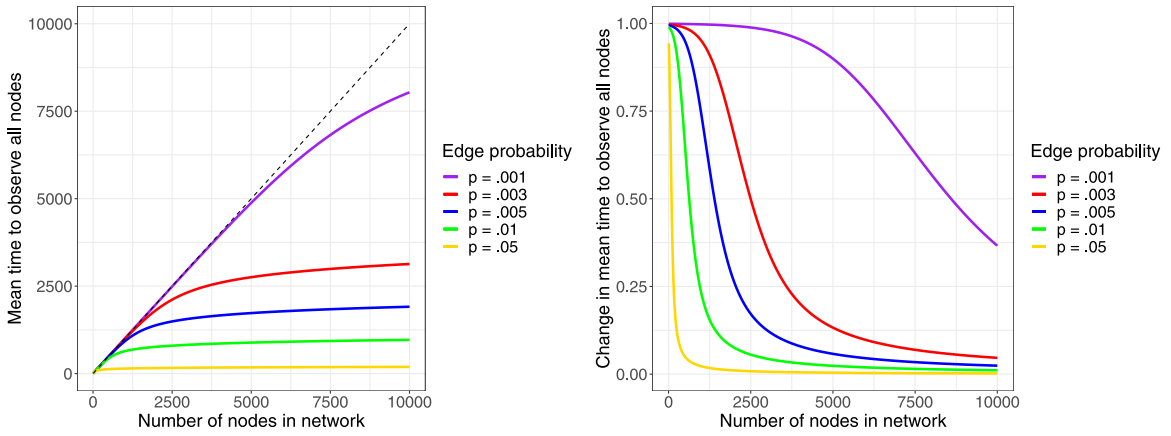

Fig. 1. Plot of the mean time to observe all nodes in the network as a function of the network size N and the change in mean times for various edge probabilities.

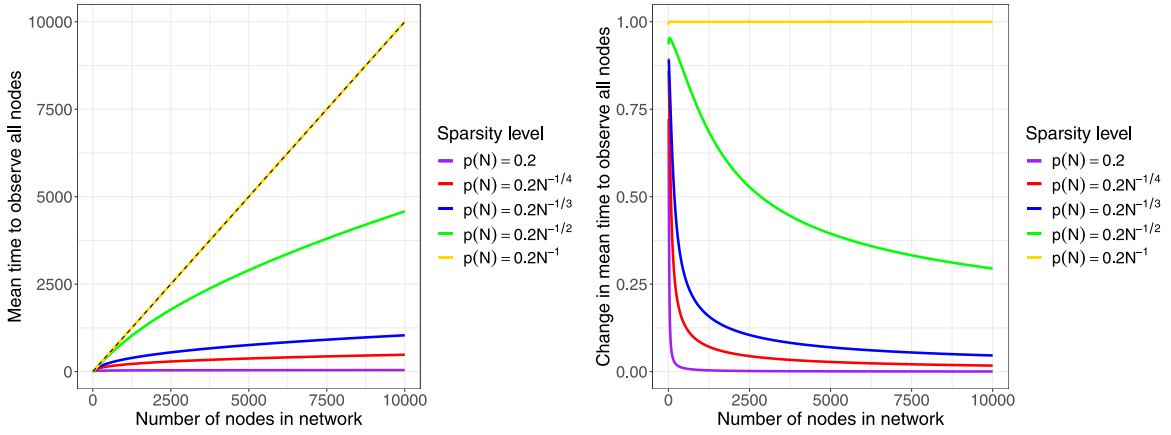

Fig. 2. Plot of the mean time to observe all nodes in the network as a function of the network size N and the change in mean times across various levels of network sparsity.

that the events $\mathcal{A}_{t,\pi(t+1)}, \ldots, \mathcal{A}_{t,\pi(N)}$ are independent events because each depends on non-overlapping subsets of edge variables in $\boldsymbol{X}$ and edge variables are independent in an Erdős-Rényi random graph. Thus,

$$\mathbb{P}\left(\mathcal{A}^c_{t,\pi(t+1)} \cap \ldots \cap \mathcal{A}^c_{t,\pi(N)}\right) = \prod_{k=1}^{N-t} \mathbb{P}\left(\mathcal{A}^c_{t,\pi(t+k)}\right) = \prod_{k=1}^{N-t} [1-(1-p)^t] = [1-(1-p)^t]^{N-t}.$$

The distribution of τ is derived to be

$$\mathbb{P}(\tau \leq t) = [1-(1-p)^t]^{N-t}, \qquad 1 \leq t \leq N.$$

Since τ is a non-negative stopping time, we compute directly

$$\mathbb{E}\,\tau = \sum_{t=0}^{N} \mathbb{P}(\tau > t) = \sum_{t=0}^{N} (1-\mathbb{P}(\tau \leq t)) = 1 + \sum_{t=1}^{N} \left(1-[1-(1-p)^t]^{N-t}\right) = N+1-\sum_{t=1}^{N}[1-(1-p)^t]^{N-t}. \quad \square$$

2.1. Effect of network sparsity on τ

Sparsity in networks refers to the scaling of edge probabilities with the network size, in which case the edge probabilities p are a function of N, which we denote by $p(N)$ throughout this section. The dense graph regime refers to the case when $p(N) \propto 1$, i.e., when the edge probability $p(N)$ is constant as a function of N, which implies the expected number of edges in the network is of order N^2. The sparse graph regime corresponds to scenarios where the edge probability $p(N)$ decreases as a function of N, typically monotonically. There are a number of special cases of interest. When $p(N) \propto N^{-1}$, Le Cam's theorem can be used to demonstrate the Poisson distribution as the limiting distribution for node degrees in an

Erdős-Rényi random graph, in which case the expected number of connections a given node has is bounded. The extreme case when $p(N) \propto N^{-2}$ corresponds to the case when the graph has a bounded expected number of edges. The evolution of sparse Erdős-Rényi random graphs has been extensively studied (Erdős and Rényi, 1960; Frieze and Karoński, 2016).

The effect of sparsity on the time to observe all nodes in a graph through egocentric sampling can be studied through Theorem 1 by replacing p with $p(N)$ and computing the mean times $\mathbb{E}\,\tau$ as a function of N. This is done in Fig. 2 for the cases when $p(N) = .2\,N^{-r}$, varying $r \in \{0,\ 1/4,\ 1/3,\ 1/2,\ 1\}$. As mentioned above, the case when $r = 1$ corresponds to the scaling $p(N) \propto N^{-1}$, for which the expected node degrees are bounded (from above and below). This implies that the expected number of nodes reported and identified in the egocentric sampling of any given node would likewise be bounded in expectation. The resulting mean time $\mathbb{E}\,\tau$ scales approximately with N in this case (shown in Fig. 2), with exact scaling confirmed rigorously in Theorem 2 in the next section. When $r < 1$, the mean degrees of nodes tend to infinity with N. However, larger values of r correspond to slower rates of growth for the expected degrees, the impact of which is shown in the ordering of the mean times in Fig. 2 with respect to the level of network sparsity.

3. The scaling of τ as a function of N

The scaling of τ (under a probability guarantee) is established in Theorem 2, confirming the sub-linear growth of τ with N suggested in Figs. 1 and 2 in certain scenarios. Of particular interest, we demonstrate that in the dense graph regime where p is fixed, τ scales logarithmically in N. In the sparse graph regime, we have related the scaling of τ to the ratio $(\log N)\,/\,p$. When $p \sim N^{-r}$ $(0 < r < 1)$, we have $(\log N)\,/\,p \sim N^{r}\,\log N$, showing the faster scaling due to network sparsity as visualized in Fig. 2. When $r = 1$, the scaling of $N \log N$ is too fast, noting that $\mathbb{P}(\tau \le N) = 1$. Theorem 2 confirms the linear scaling of τ with N, implied in Fig. 2, in this case.

Theorem 2. *Consider an Erdős-Rényi random graph with edge probability $p \in (0,\,1)$ bounded away from 1. Then there exists, for each $\delta \in (0,\,1)$, constants $N_0 \ge 2$ and $C_1, C_2 > 0$ such that, for all $N \ge N_0$,*

$$\mathbb{P}\left(C_2\,\min\left\{N,\ \frac{\log N}{p}\right\} \ \le\ \tau\ \le\ C_1\,\min\left\{N,\ \frac{\log N}{p}\right\}\right) \ \ge\ \ \delta.$$

Proof of Theorem 2. From Theorem 1, the distribution of τ is given by

$$\mathbb{P}(\tau \le t) = [1-(1-p)^t]^{N-t}, \qquad t = 1,\ldots,N.$$

For numbers $a, b \in \mathbb{R}$,

$$\mathbb{P}(a \le \tau \le b) \ \ \ge\ \ \mathbb{P}(\tau \le b) - \mathbb{P}(\tau \le a),$$

noting that $\tau \in \{1,\ldots,N\}$ is a discrete valued stopping time. Let $\delta \in (0,\,1)$. It suffices to find $\delta_1, \delta_2 \in (0,\,1)$ such that $\delta = \delta_1 - \delta_2 \in (0,\,1)$ satisfying $\mathbb{P}(\tau \le b) \ge \delta_1$ and $\mathbb{P}(\tau \le a) \le \delta_2$ in order to show $\mathbb{P}(a \le \tau \le b) \ge \delta$. We demonstrate a sufficient condition for $t > 0$ in order to ensure $\mathbb{P}(\tau \le t) \ge \delta_1$ for a given $\delta_1 \in (0,\,1)$. Write

$$\mathbb{P}(\tau \le t) = \left[1-(1-p)^t\right]^{N-t} = \left[1-\left(1-\frac{p\,t}{t}\right)^t\right]^{N-t} \ \ \ge\ \ (1-e^{-p\,t})^{N-t} \ \ \ge\ \ 1-(N-t)\,e^{-pt},$$

first using the inequality

$$1-\left(1-\frac{x}{n}\right)^n \ge 1-e^{-x} \qquad \text{implied from} \qquad e^{-x} \ge \left(1-\frac{x}{n}\right)^n \qquad (\text{valid for } x \in (0,\,n)),$$

and then using Bernoulli's inequality $(1-x)^k \ge 1-kx$ (valid for $x \in (0,\,1)$ and k a positive integer). It suffices for $1-(N-t)\,e^{-pt} \ge \delta_1$ to ensure $\mathbb{P}(\tau \le t) \ge \delta_1$. Re-arranging, we obtain $1-\delta_1 \ge (N-t)\,e^{-p\,t}$. Noting $N\,e^{-p\,t} \ge (N-t)\,e^{-p\,t}$, it suffices for $1-\delta_1 \ge N\,e^{-p\,t}$. Taking the logarithm of both sides and re-arranging, we obtain $p\,t \ge \log N + |\log(1-\delta_1)|$, noting $\delta_1 \in (0,\,1)$ so that $\log(1-\delta_1) < 0$. This yields the equation

$$t \ \ \ge\ \ \frac{\log N}{p}\left(1+\frac{|\log(1-\delta_1)|}{\log N}\right).$$

Note that $1+|\log(1-\delta_1)|\,/\,\log N$ is monotonically decreasing in N. Let $U = 1+|\log(1-\delta_1)|\,/\,\log 2 > 0$. Then

$$1+\frac{|\log(1-\delta_1)|}{\log N} \ \ \le\ \ 1+\frac{|\log(1-\delta_1)|}{\log 2} = U, \qquad \text{for all } N \ge 2.$$

It therefore suffices for $t \ge (U \log N)\,/\,p$. As $\mathbb{P}(\tau \le N) = 1$, there exists $C_1 > 0$ independent of N (but dependent on choice of δ_1) such that

$$\mathbb{P}\left(\tau \le C_1\,\min\left\{N,\ \frac{\log N}{p}\right\}\right) \ \ \ge\ \ \delta_1.$$

We next demonstrate a sufficient condition for $t > 0$ in order to ensure $\mathbb{P}(\tau \leq t) \leq \delta_2$ for a given $\delta_2 \in (0, 1)$. We want

$$\mathbb{P}(\tau \leq t) = \left[1 - (1-p)^t\right]^{N-t} \quad \leq \quad \delta_2.$$

Taking the logarithm of both sides, we get $(N-t)\log(1-(1-p)^t) \leq \log \delta_2$. Using the inequality $\log(1+x) \leq x$,

$$(N-t)\log(1-(1-p)^t) \quad \leq \quad -(N-t)(1-p)^t.$$

It therefore suffices for $-(N-t)(1-p)^t \leq \log \delta_2$ to ensure $\mathbb{P}(\tau \leq t) \leq \delta_2$. Multiplying by -1 and dividing by N,

$$(1-p)^t \left(1 - \frac{t}{N}\right) \quad \geq \quad \frac{|\log \delta_2|}{N}.$$

It must be that $t < N$ for the left-hand side to be positive, in line with the absolute bound on τ implied by $\mathbb{P}(\tau \leq N) = 1$. Taking logarithm of both sides again,

$$t \log(1-p) + \log\left(1 - \frac{t}{N}\right) \quad \geq \quad \log|\log \delta_2| - \log N.$$

Take $t \leq (1-e^{-1})N$. As $\log(1-t/N)$ is monotonically decreasing in t, we have the bound $\log(1-t/N) \geq -1$ provided $t \leq (1-e^{-1})N$, in which case it suffices for $t\log(1-p) - 1 \geq \log|\log \delta_2| - \log N$. Multiplying by -1 and re-arranging,

$$t \quad \leq \quad \frac{\log N - 1 - \log|\log \delta_2|}{|\log(1-p)|}. \tag{6}$$

We next lower bound the right-hand side of (6):

$$\frac{\log N - 1 - \log|\log \delta_2|}{|\log(1-p)|} = \frac{\log N}{|\log(1-p)|}\left(1 - \frac{1 + \log|\log \delta_2|}{\log N}\right) \quad \geq \quad \frac{(1-p)\log N}{p}\left(1 - \frac{1 + \log|\log \delta_2|}{\log N}\right),$$

where the inequality follows from the inequality $-x/(1-x) \leq \log(1-x)$ $(x \in (0,1))$ so that $|\log(1-p)| \leq p/(1-p)$. Since $(1 + \log|\log \delta_2|)/\log N \downarrow 0$ as $N \to \infty$, there exists $N_0 \geq 2$ and $L > 0$ such that, for all $N \geq N_0$,

$$(1-p)\left(1 - \frac{1 + \log|\log \delta_2|}{\log N}\right) \quad \geq \quad L,$$

owing to the fact that p is assumed bounded away from 1. It therefore suffices for $t \leq \min\{(L\log N)/p, (1-e^{-1})N\}$, showing there exists $N_0 \geq 2$ and $C_2 > 0$ independent of N (but dependent on choice of δ_2) such that, for all $N \geq N_0$,

$$\mathbb{P}\left(\tau \leq C_2 \min\left\{N, \frac{\log N}{p}\right\}\right) \quad \leq \quad \delta_2.$$

As the choice of $\delta_1 \in (0,1)$ and $\delta_2 \in (0,1)$ in the above was arbitrary, we have shown there exists, for each $\delta \in (0,1)$, constants $N_0 \geq 2$ and $C_1, C_2 > 0$ such that, for all $N \geq N_0$,

$$\mathbb{P}\left(C_2 \min\left\{N, \frac{\log N}{p}\right\} \leq \tau \leq C_1 \min\left\{N, \frac{\log N}{p}\right\}\right) \quad \geq \quad \delta,$$

provided $p \in (0,1)$ is bounded away from 1. □

4. Discussion

We have characterized the time it takes to identify the nodes in an Erdős-Rényi random graph through egocentric sampling in both dense and sparse graph settings. Networks we encounter in our world are generally not Erdős-Rényi random graphs. However, our results can provide conservative estimates in certain settings. For example, social networks often exhibit positive transitivity (i.e., a stochastic proclivity towards triangle motifs) (Holland and Leinhardt, 1972; Stewart et al., 2019), a phenomena also found in other networks (Schweinberger et al., 2020). In these cases, it is possible to find lower and upper bounding probabilities for the conditional probabilities of edges (Butts, 2011), thus providing a method to bound densities and probabilities of random graphs using Erdős-Rényi random graphs. Such techniques provide a means of establishing conservative estimates of $\mathbb{E}\tau$ and its distribution in certain settings, offering the ability to extend insights developed in this work to other classes of random graphs beyond Erdős-Rényi random graphs.

Data availability

No data was used for the research described in the article.

Acknowledgments

This research was supported by the Office of the Secretary of Defense, Directorate of Operational Test and Evaluation, and the Test Resource Management Center under the Science of Test research program. Contract #FA8075-18-D-0002.

References

Butts, C.T., 2011. Bernoulli graph bounds for general random graph models. Sociol. Methodol. 41, 299–345.
Erdős, P., Rényi, A., 1959. On random graphs. Publ. Math. 6, 290–297.
Erdős, P., Rényi, A., 1960. On the evolution of random graphs. Publ. Math. Inst. Hung. Acad. Sci. 5, 17–61.
Frank, O., 2005. Network sampling and model fitting. In: Carrington, P.J., Scott, J., Wasserman, S. (Eds.), Models and Methods in Social Network Analysis. Cambridge University Press, New York, pp. 31–56.
Frieze, A., Karoński, M., 2016. Introduction to Random Graphs. Cambridge University Press.
Gilbert, E.N., 1959. Random graphs. Ann. Math. Stat. 30, 1141–1144.
Gile, K., Handcock, M.H., 2010. Respondent-driven sampling: An assessment of current methodology. Sociol. Methodol. 40, 285–327.
Handcock, M.S., Gile, K., 2010. Modeling social networks from sampled data. Ann. Appl. Stat. 4, 5–25.
Holland, P.W., Leinhardt, S., 1972. Some evidence on the transitivity of positive interpersonal sentiment. Am. J. Sociol. 77, 1205–1209.
Kolaczyk, E.D., 2017. Topics at the Frontier of Statistics and Network Analysis:(re) visiting the Foundations. Cambridge University Press.
Krivitsky, P.N., Morris, M., 2017. Inference for social network models from egocentrically-sampled data, with application to understanding persistent racial disparities in HIV prevalence in the US. Ann. Appl. Stat. 11, 427–455.
Perry, P.L., Pescosolido, B.A., Borgatti, S.P., 2018. Egocentric Network Analysis: Foundations, Methods, and Models. Vol. 44, Cambridge University Press.
Schweinberger, M., Krivitsky, P.N., Butts, C.T., Stewart, J.R., 2020. Exponential-family models of random graphs: Inference in finite-, super-, and infinite-population scenarios. Statist. Sci. 35 (4), 627–662.
Stewart, J., Schweinberger, M., Bojanowski, M., Morris, M., 2019. Multilevel network data facilitate statistical inference for curved ERGMs with geometrically weighted terms. Social Networks 59, 98–119.