On the Connectivity and Giant Component Size of Random K-out Graphs Under Randomly Deleted Nodes

Eray Can Elumar Mansi Sood Osman Yağan

***Abstract*— Random K-out graphs, denoted $\mathbb{H}(n;K)$, are generated by each of the n nodes drawing K out-edges towards K distinct nodes selected uniformly at random, and then ignoring the orientation of the arcs. Recently, random K-out graphs have been used in applications as diverse as random (pairwise) key predistribution in ad-hoc networks, anonymous message routing in crypto-currency networks, and differentially-private federated averaging. In many applications, connectivity of the random K-out graph when some of its nodes are *dishonest*, have *failed*, or have been *captured* is of practical interest. We provide a comprehensive set of results on the connectivity and giant component size of $\mathbb{H}(n;K_n,\gamma_n)$, i.e., random K-out graph when γ_n of its nodes, selected uniformly at random, are deleted. First, we derive conditions for K_n and n that ensure, with high probability (whp), the connectivity of the remaining graph when the number of deleted nodes is $\gamma_n=\Omega(n)$ and $\gamma_n=o(n)$, respectively. Next, we derive conditions for $\mathbb{H}(n;K_n,\gamma_n)$ to have a *giant component*, i.e., a connected subgraph with $\Omega(n)$ nodes, whp. This is also done for different scalings of γ_n and *upper* bounds are provided for the number of nodes *outside* the giant component. Simulation results are presented to validate the usefulness of the results in the finite node regime.**

***Index Terms*— Connectivity, giant component, robustness, random graphs, random K-out graphs, security, privacy**

I. Introduction

Random graphs are widely used in modeling and analysis of diverse real-world networks including social networks [1], economic networks [2], and communication networks [3]. In recent years, a random graph model known as the *random K-out graph* has received interest in designing secure sensor networks [4], decentralized learning [5], and anonymity preserving crypto-currency networks [6]. Random K-out graphs, denoted $\mathbb{H}(n;K)$, are generated over a set of n nodes as follows. Each of the n nodes draws K out-edges towards K distinct nodes selected uniformly at random. The resulting *undirected* graph obtained by ignoring the orientation of the edges is referred to as a random K-out graph.

In the context of sensor networks, random K-out graphs have been used [4], [7], [8] to analyze the performance of the random *pairwise* key predistribution scheme [9] and its heterogeneous variants [10], [11]. The random *pairwise* scheme works as follows. Before deployment, each sensor chooses K others uniformly at random. A unique *pairwise* key is given to each node pair where at least one of them selected the other. After deployment, two sensors can securely communicate if they share a pairwise key. The topology of the sensor network can thus be represented by a random K-out graph; each edge of the random K-out represents a secure communication link between two sensors. Consequently, random K-out graphs have been analyzed to answer key questions on the values of the parameters n, K needed to achieve certain desired properties, including connectivity at the time of deployment [4], [12], connectivity under *link* removals [7], [8], and unassailability [13].

Despite many prior works on random K-out graphs, very little is known about its connectivity properties when some of its *nodes* are removed. This is an increasingly relevant problem since many deployments of sensor networks are expected to take place in *hostile* environments where nodes may be captured by an adversary, or fail due to harsh conditions. In addition, random K-out graphs have recently been used to construct the communication graph in a differentially-private federated averaging scheme called the GOPA (GOssip Noise for Private Averaging) protocol [5, Algorithm 1]. According to the GOPA protocol, a random K-out graph is constructed on a set of nodes of which an unknown subset is *dishonest*. It was shown in [5, Theorem 3] that the privacy-utility trade-offs achieved by the GOPA protocol is tightly dependent on the subgraph on *honest* nodes being *connected*. When the subgraph on honest nodes is not connected, it was shown that the performance of GOPA is tied to the *size* of the connected components of the honest nodes.

With these motivations in mind, this paper aims to fill a gap in the literature and provide a comprehensive set of results on the connectivity and size of the giant component of the random K-out graph when some of its nodes are *dishonest*, have *failed*, or have been *captured*. Let $\mathbb{H}(n;K_n,\gamma_n)$ denote the random K-out graph when γ_n of its nodes, selected uniformly at random, are deleted. First, we provide a set of conditions for K_n and n that ensure, *with high probability* (whp), the connectivity of the remaining graph when the number γ_n of deleted nodes is $\Omega(n)$ and $o(n)$, respectively. Our result for $\gamma_n=\Omega(n)$ (see Theorem 3.1) significantly improves a prior result [14] on the same problem and leads to a *sharp* zero-one law for the connectivity of $\mathbb{H}(n;K_n,\gamma_n)$. Our result for the case $\gamma_n=o(n)$ (see Theorem 3.2) expands the existing threshold of $K_n\geq 2$ required for connectivity by showing that the graph is still connected whp for $K_n\geq 2$ when $o(\sqrt{n})$ nodes are deleted. We then derive conditions on K_n that leads $\mathbb{H}(n;K_n,\gamma_n)$ to have a *giant component* with an upper bound on the number of nodes allowed outside the giant component. This is also done for both cases $\gamma_n=\Omega(n)$ and $\gamma_n=o(n)$. Finally, we present simulation results when the number of nodes is finite and compare the results with an Erdős-Rényi graph with same average node degree.

E.C. Elumar, M. Sood and O. Yağan are with Department of Electrical and Computer Engineering and CyLab, Carnegie Mellon University, Pittsburgh, PA, 15213 USA. Email: {eelumar@andrew.cmu.edu, msood@andrew.cmu.edu, oyagan@ece.cmu.edu}

II. Notations and the Model

All random variables are defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and probabilistic statements are given with respect to the probability measure $\mathbb{P}$. The complement of an event A is denoted by A^{c}. The cardinality of a discrete set A is denoted by $|A|$. All limits are understood with n going to infinity. If the probability of an event tends to one as $n \to \infty$, we say that it occurs with high probability (whp). The statements $a_n = o(b_n)$, $a_n = \omega(b_n)$, $a_n = O(b_n)$, $a_n = \Theta(b_n)$, and $a_n = \Omega(b_n)$, used when comparing the asymptotic behavior of sequences $\{a_n\}, \{b_n\}$, have their meaning in standard Landau notation. The asymptotic equivalence $a_n \sim b_n$ is used to denote the fact that $\lim_{n\to\infty} \frac{a_n}{b_n} = 1$.

The random K-out graph is defined on the vertex set $V := \{v_1, \ldots, v_n\}$ as follows. Let $\mathcal{N} := \{1, 2, \ldots, n\}$ denote the set vertex labels. For each $i \in \mathcal{N}$, let $\Gamma_{n,i} \subseteq \mathcal{N} \setminus i$ denote the set of K_n labels corresponding to the nodes selected by v_i. It is assumed that $\Gamma_{n,1}, \ldots, \Gamma_{n,n}$ are mutually independent. Distinct nodes v_i and v_j are adjacent, denoted by $v_i \sim v_j$ if at least one of them picks the other. Namely,

$$v_i \sim v_j \quad \text{if} \quad [j \in \Gamma_{n,i}] \ \vee \ [i \in \Gamma_{n,j}]. \tag{1}$$

The random graph defined on the vertex set V through the adjacency condition (1) is called a random K-out graph [4], [15], [16] and denoted by $\mathbb{H}(n; K_n)$. It was previously established in [4], [12] that random K-out graph is connected whp when $K \geq 2$ and not connected when $K = 1$; i.e.,

$$\lim_{n\to\infty} \mathbb{P}\left[\mathbb{H}(n; K) \text{ is connected}\right] = \begin{cases} 1 & \text{if} \quad K \geq 2, \\ 0 & \text{if} \quad K = 1. \end{cases} \tag{2}$$

Next, we model random K-out graphs under random removal of nodes. As already mentioned, our motivation is to understand the properties of the underlying network when some nodes are *dishonest*, or have *failed*, or have been *captured*. We let γ_n denote the number of such nodes and assume, for simplicity, that they are selected uniformly at random among all nodes in V. The case where the set of dishonest/captured/failed nodes are selected carefully by an adversary might also be of interest, but is beyond of the scope of the current paper; see [13] for partial results in that case. A related model of interest is the random K-out graph under randomly deleted *edges*. The connectivity and k-connectivity under that case have been studied in [7], [17], [18].

Formally, let $D \subset V$, $|D| = \gamma_n$ denote the set of deleted nodes. We can then define $\mathbb{H}(n; K_n, \gamma_n)$ on the vertex set $R = V \setminus D$ and the corresponding set of labels $\mathcal{N}_R$, such that distinct vertices v_i and v_j (both in R) are adjacent if they were adjacent in $\mathbb{H}(n; K_n)$; i.e., if $[j \in \Gamma_{n,i}] \ \vee \ [i \in \Gamma_{n,j}]$. For each $i \in \mathcal{N}_R$, the set of labels adjacent to node v_i in $\mathbb{H}(n; K_n, \gamma_n)$ is denoted by $\Gamma_{n-\gamma_n,i} \subseteq \mathcal{N}_R \setminus i$.

Definition 2.1 (Connected Components): A pair of nodes in a graph $\mathbb{G}$ are said to be connected *if there exists a path of edges connecting them. A* component C_i *of $\mathbb{G}$ is a subgraph in which any two vertices are connected to each other, and no vertex is connected to a node outside of C_i.*

A graph with n nodes is said to have a *giant* component if its largest connected component is of size $\Omega(n)$.

III. Main Results and Discussion

Our main results are presented in Theorems $3.1 - 3.4$ below. Each Theorem addresses a design question as to how we should choose the parameter K_n such that when the given number γ_n of nodes are deleted, the remaining graph satisfies the given desired property (e.g., connectivity or a giant component with a specific size) whp.

A. Results on Connectivity

Let $P(n, K_n, \gamma_n) = \mathbb{P}\left[\mathbb{H}(n; K_n, \gamma_n) \text{ is connected}\right]$.

Theorem 3.1: *Let $\gamma_n = \alpha n$ with α in $(0, 1)$, and consider a scaling $K : \mathbb{N}_0 \to \mathbb{N}_0$ such that with $c > 0$ we have*

$$K_n \sim c \cdot r_1(\alpha, n), \quad \textit{where} \quad r_1(\alpha, n) = \frac{\log n}{1 - \alpha - \log \alpha} \tag{3}$$

is the threshold function. Then, we have

$$\lim_{n\to\infty} P(n, K_n, \gamma_n) = \begin{cases} 1, & \text{if} \quad c > 1 \\ 0, & \text{if} \quad 0 < c < 1. \end{cases} \tag{4}$$

The proof of the *one-law* in (4), i.e., that $\lim_{n\to\infty} P(n, K_n, \gamma_n) = 1$ if $c > 1$, is given in Section IV. The *zero-law* of (4), i.e., that $\lim_{n\to\infty} P(n, K_n, \gamma_n) = 0$ if $c < 1$, was established previously in [14, Corollary 3.3]. There, a one-law was also provided: under (3), it was shown that $\lim_{n\to\infty} P(n, K_n, \gamma_n)$ if $c > \frac{1}{1-\alpha}$, leaving a gap between the thresholds of the zero-law and the one-law. Theorem 3.1 presented here fills this gap by establishing a tighter one-law, and constitutes a *sharp* zero-one law; e.g., when $\alpha = 0.5$, the one-law in [14] is given with $c > 2$, while we show that it suffices to have $c > 1$.

Theorem 3.2: *Consider a scaling $K : \mathbb{N}_0 \to \mathbb{N}_0$.*
a) If $\gamma_n = o(\sqrt{n})$, then we have

$$\lim_{n\to\infty} P(n, K_n, \gamma_n) = 1, \quad \text{if} \quad K_n \geq 2 \ \forall n \tag{5}$$

b) If $\gamma_n = \Omega(\sqrt{n})$ and $\gamma_n = o(n)$, and if for some sequence w_n, it holds that

$$K_n = r_2(\gamma_n) + \omega_n, \quad \textit{where} \quad r_2(\gamma_n) = \frac{\log(\gamma_n)}{\log 2 + 1/2}$$

is the threshold function, then we have

$$\lim_{n\to\infty} P(n, K_n, \gamma_n) = 1, \quad \text{if} \quad \lim_{n\to\infty} \omega_n = \infty \tag{6}$$

Random K-out graph was known [4], [12] to be connected whp when $K_n \geq 2$ (viz. (2)). Theorem 3.2 extends this result by showing that $K_n \geq 2$ is sufficient to have the random K-out graph remain connected whp when $o(\sqrt{n})$ of its nodes (selected randomly) are deleted.

B. Results on the Size of the Giant Component

Let $C_{max}(n, K_n, \gamma_n)$ denote the set of nodes in the *largest* connected component of $\mathbb{H}(n; K_n, \gamma_n)$ and let $P_G(n, K_n, \gamma_n, \lambda_n) := \mathbb{P}[|C_{max}(n, K_n, \gamma_n)| > n - \gamma_n - \lambda_n]$. Namely, $P_G(n, K_n, \gamma_n, \lambda_n)$ is the probability that less than λ_n nodes are *outside* the largest component of $\mathbb{H}(n; K_n, \gamma_n)$.

Theorem 3.3: *Let $\gamma_n = o(n)$, $\lambda_n = \Omega(\sqrt{n})$ and $\lambda_n \leq \lfloor (n-\gamma_n)/3 \rfloor$. Consider a scaling $K : \mathbb{N}_0 \to \mathbb{N}_0$ and let*

$$r_3(\gamma_n, \lambda_n) = 1 + \frac{\log(1+\gamma_n/\lambda_n)}{\log 2 + 1/2}$$

be the threshold function. Then, we have

$$\lim_{n\to\infty} P_G(n, K_n, \gamma_n, \lambda_n) = 1, \quad \text{if} \quad K_n > r_3(\gamma_n, \lambda_n), \ \ \forall n.$$

We remark that if $\lambda_n = \beta n$ with $0 < \beta < 1/3$, then $r_3(\gamma_n, \lambda_n) = 1 + o(1)$. This shows that when $\gamma_n = o(n)$, it suffices to have $K_n \geq 2$ for $\mathbb{H}(n; K_n, \gamma_n)$ to have a giant component containing $\Omega((1-\beta)n)$ nodes for arbitrary $0 < \beta < 1/3$.

Theorem 3.4: *Let $\gamma_n = \alpha n$ with α in $(0,1)$, $\lambda_n = o(n)$, and $\lambda_n = \omega(1)$. Consider a scaling $K : \mathbb{N}_0 \to \mathbb{N}_0$ and let*

$$r_4(\alpha, \lambda_n) = 1 + \frac{\log(1+\frac{n\alpha}{\lambda_n}) + \alpha + \log(1-\alpha)}{\frac{1-\alpha}{2} - \log\left(\frac{1+\alpha}{2}\right)}$$

be the threshold function. Then, we have

$$\lim_{n\to\infty} P_G(n, K_n, \alpha, \lambda) = 1, \quad \text{if} \quad K_n > r_4(\alpha, x_n), \ \ \forall n.$$

Due to space limitations, we only provide a proof of Theorem 3.1 here. Proofs of all results are available in [19].

C. Simulation Results

To check the usefulness of our results when the number n of nodes is finite, we examine the probability of connectivity and the number of nodes outside the giant component (i.e., $n - \gamma_n - |C_{max}(n, K_n, \gamma_n)|$) in two different experimental setups. The first setup is to obtain the results for the case where $\gamma_n = \alpha n$, with α in $(0,1)$. We generate instantiations of the random graph $\mathbb{H}(n; K_n, \gamma_n)$ with $n = 5000$, varying K_n in the interval $[1, 25]$ and several α values in the interval $[0.1, 0.8]$. Then, we record the empirical probability of connectivity and λ_n from 1000 independent experiments for each (K_n, α) pair. The results of this experiment are shown in Fig. 1 (Left) and Fig. 2.

Fig. 1 (Left) depicts the empirical probability of connectivity of $\mathbb{H}(n; K_n, \gamma_n)$. The vertical lines stand for the critical threshold of connectivity asserted by Theorem 3.1. In each curve, $P(n, K_n, \gamma_n)$ exhibits a threshold behaviour as K_n increases, and the transition from $P(n, K_n, \gamma_n) = 0$ to $P(n, K_n, \gamma_n) = 1$ takes place around $K_n = \frac{\log n}{1-\alpha-\log\alpha}$, validating the claims of Theorem 3.1.

In Fig. 2, we plot the *maximum* number of nodes outside the giant component observed in 1000 experiments for each parameter pair, and compare these with our result, namely the upper bound on $n - \gamma_n - |C_{max}|$ obtained from Theorem 3.4 by taking the maximum γ_n value that gives a threshold less than or equal to the K_n value tested in the simulation. As can be seen, for any K_n and γ_n value, the experimental maximum number of nodes outside the giant component is smaller than the upper bound obtained from Theorem 3.4, reinforcing the usefulness of our results in practical settings.

We ran a second set of experiments for the case where $\gamma_n = o(n)$. As before, we generate instantiations of the random graph $\mathbb{H}(n; K_n, \gamma_n)$, with $n = 50,000$, varying K_n in $[2, 5]$ and varying λ_n in $[10, 2000]$. For each (K_n, γ_n) pair, we generate 1000 experiments and record the maximum number of nodes seen outside the giant component; in some case no nodes are seen outside the giant component indicating that the graph is connected. The results of this experiment are shown in Fig. 1 (Right) and Fig. 3.

In Fig. 1 (Right), the maximum number of nodes seen outside the giant component in 1000 experiments is depicted as a function of K_n. The plots for $\gamma_n = 10$ and $\gamma_n = 100$ correspond to the $\gamma_n = o(\sqrt{n})$ case in Theorem 3.2a. As can be seen, there is only one node outside the giant component in the worst case for $\gamma_n = 10$ and $\gamma_n = 100$ when $K_n = 2$,

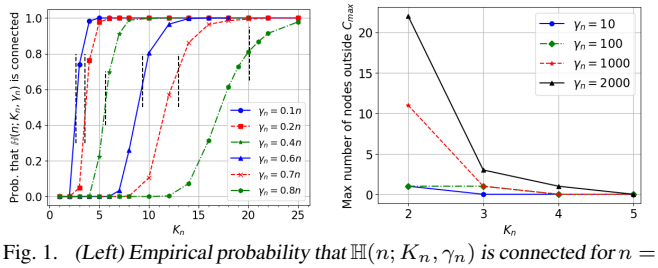

Fig. 1. *(Left) Empirical probability that $\mathbb{H}(n; K_n, \gamma_n)$ is connected for $n = 5000$ calculated from 1000 experiments. The vertical lines are the theoretical thresholds given by Theorem 3.1. (Right) Maximum number of nodes outside the giant component of $\mathbb{H}(n; K_n, \gamma_n)$ for $n = 50,000$ in 1000 experiments.*

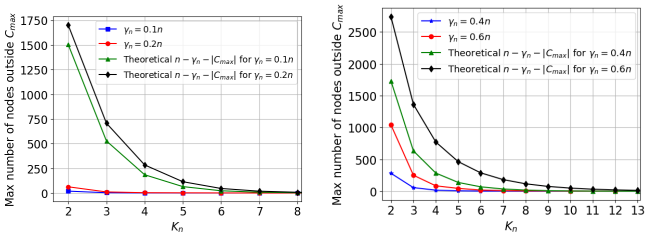

Fig. 2. *Maximum number of nodes outside the giant component of $\mathbb{H}(n; K_n, \gamma_n)$ for $n = 5000$ and $\gamma_n = 0.1n$, $\gamma_n = 0.2n$ cases (Left); and for $n = 5000$ and $\gamma_n = 0.4n$, $\gamma_n = 0.6n$ cases (Right), obtained through 1000 experiments along with respective plot of theoretical $n - \gamma_n - |C_{max}|$.*

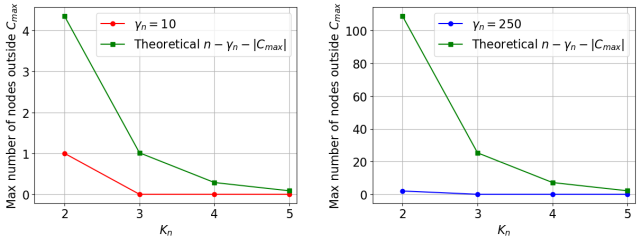

Fig. 3. *Maximum number of nodes outside the giant component of $\mathbb{H}(n; K_n, \gamma_n)$ for $n = 50,000$ and $\gamma_n = 10$ cases (Left); and for for $n = 50,000$ and $\gamma_n = 250$ cases (Right), obtained through 1000 experiments along with the plot of theoretical $n - \gamma_n - |C_{max}|$.*

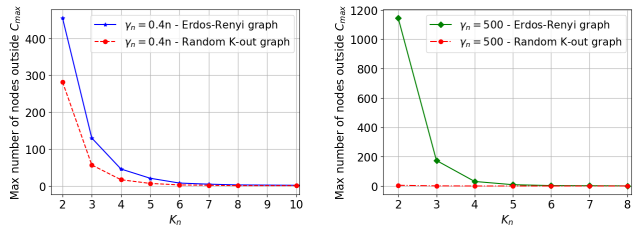

Fig. 4. *Comparison of maximum number of nodes outside the giant component of a random K-out graph $\mathbb{H}(n; K_n, \gamma_n)$ and an Erdős-Rényi graph with same mean node degree when $n = 5000$, $\gamma_n = 0.4n$ (Left); and when $n = 50,000$ and $\gamma_n = 500$ (Right). Each data-point is obtained through 1000 experiments.*

roughly in line with Theorem 3.2a which expects the graph to be connected. The plots for $\gamma_n = 1000$ and $\gamma_n = 2000$ correspond to the $\gamma_n = w(\sqrt{n})$ and $\gamma_n = o(n)$ case in Theorem 3.2b. The thresholds on K_n for these γ_n values, obtained using Theorem 3.2b are $r_2(1000) = 6.79$ and $r_2(2000) = 7.37$, rounded to two digits after decimal when the $\omega(1)$ term in Theorem 3.2b is ignored due to n having a finite value in the simulations. As can be seen, the graph becomes connected for $\gamma_n = 1000$ when $K_n \geq 4$, and for $\gamma_n = 2000$ when $K_n \geq 5$. Hence, we can see that graphs for $\gamma_n = 1000$ and $\gamma_n = 2000$ are connected when K_n is selected above the theoretical threshold obtained from 3.2b, supporting Theorem 3.2b.

In Fig. 3, the maximum number of nodes seen outside the giant component in 1000 experiments is plotted as a function of K_n. The corresponding theoretical plots are obtained by the upper bound on $n - \gamma_n - |C_{max}|$ asserted by Theorem 3.3. For any K_n and γ_n pair, the experimental values are smaller than the theoretical values, supporting the usefulness of Theorem 3.3 in the finite node regime.

D. Discussion

In Theorem 3.1, we improve the results given in [14] by using tighter upper bounds on the probability of not being connected. With this, we close the gap between the zero law and the one law, and hence establish a sharp zero-one law for connectivity when $\gamma_n = \Omega(n)$ nodes are deleted from $\mathbb{H}(n; K_n, \gamma_n)$.

In Theorem 3.2, we establish that the graph $\mathbb{H}(n; K_n, \gamma_n)$ with $\gamma_n = o(n)$ is connected whp when $K_n \sim \log(\gamma_n)$; and when $\gamma_n = o(\sqrt{n})$, $K_n \geq 2$ is sufficient for connectivity. The latter result is especially important, since $K_n \geq 2$ is the previously established threshold for connectivity [12], we improve this result by showing that the graph is still connected with $K_n \geq 2$ even after $o(\sqrt{n})$ nodes (selected randomly) are deleted. We also note that Theorem 3.3 and 3.4 constitute the first results concerning the giant component size of random K-out graphs under randomly deleted nodes.

To put these results in perspective, we compare them with an Erdős-Rényi graph $G(n, p)$, which is connected whp if $p > \log n/n$. This translates to having an average node degree of $< k > \sim \log n$ [20]. The $< k >$ required for the random K-out graph to be connected whp is much lower, with $< k > = O(1)$ when $o(\sqrt{n})$ nodes are removed, and $< k > \sim \log(\gamma_n)$ when $\gamma_n = \Omega(\sqrt{n})$ nodes are removed.

For a better comparison, we examine the experimental maximum number of nodes outside the giant component out of 1000 experiments of a random K-out graph $\mathbb{H}(n; K_n, \gamma_n)$ and an Erdős-Rényi graph $G(n, p)$ with same mean node degree when γ_n random nodes are removed from the graph. Since these graphs are defined using different parameters (p for Erdős-Rényi and K for random K-out), to achieve the same node degree, p in the Erdős-Rényi graph is selected as $p = 2K_n/n$. The results are given in Fig. 4 for $n = 5000$, $\gamma_n = 0.4n$ on (Left), and $n = 50,000$, $\gamma_n = 500$ on (Right). As can be seen, the random K-out graph has less maximum number of nodes outside the giant component than the Erdős-Rényi graph and this difference is more pronounced when γ_n is smaller. From these plots for the finite node regime, and also comparing the Theorems given in our paper with the relevant literature on ER graphs, we can conclude that random K-out graphs are more robust to random node removals than Erdős-Rényi graphs in the sense of probability of connectivity and size of the giant component being larger. This reinforces the efficiency of the K-out construction in various distributed network applications including federated averaging [5], [21] where it is desirable to maintain connectivity in the event of node failures or adversarial capture of nodes.

IV. A Proof of Theorem 3.1

We start by defining a *cut*.

Definition 4.1 (Cut): [22, Definition 6.3] For a graph $\mathcal{G}$ defined on the node set V, a *cut* is a non-empty subset $S \subset V$ of nodes *isolated* from the rest of the graph. Namely, $S \subset V$ is a cut if there is no edge between S and $S^{\mathrm{c}} = V \setminus S$.

Definition 4.1 implies that if S is a cut, then so is S^{c}. Recall from Section II that we defined $\mathbb{H}(n; K_n, \gamma_n)$ as the graph when the set D of nodes is removed from the graph $\mathbb{H}(n; K_n)$. Namely, the vertex set of $\mathbb{H}(n; K_n, \gamma_n)$ is given by $R = V \setminus D$. Let $\mathcal{E}_n(K_n, \gamma_n; S)$ denote the event that $S \subset R$ is a cut in $\mathbb{H}(n; K_n, \gamma_n)$ as per Definition 4.1. With $S^{\mathrm{c}} = R/S$, the event $\mathcal{E}_n(K_n, \gamma_n; S)$ occurs if no nodes in S pick neighbors in S^{c}, and no nodes in S^{c} pick neighbors in S. Note that nodes in S or S^{c} can still pick neighbors in the set D. Thus, with $\mathcal{N}_S$, $\mathcal{N}_{S^{\mathrm{c}}}$ denoting the set of labels of the vertices in S and S^{c}, respectively, we have

$$\mathcal{E}_n(K_n, \gamma_n; S) = \bigcap_{i \in \mathcal{N}_S} \bigcap_{j \in \mathcal{N}_{S^{\mathrm{c}}}} (\{i \notin \Gamma_{n-\gamma_n, j}\} \cap \{j \notin \Gamma_{n-\gamma_n, i}\})$$

Let $\mathcal{Z}(x_n; K_n, \gamma_n)$ denote the event that $\mathbb{H}(n; K_n, \gamma_n)$ has no cut $S \subset R$ with size $x_n \leq |S| \leq n - \gamma_n - x_n$ where $x : \mathbb{N}_0 \to \mathbb{N}_0$ is a sequence such that $x_n \leq (n - \gamma_n)/2 \ \forall n$. Namely, $\mathcal{Z}(x_n; K_n, \gamma_n)$ is the event that there are no cuts in $\mathbb{H}(n; K_n, \gamma_n)$ whose size falls in the range $[x_n, n - \gamma_n - x_n]$.

Lemma 4.2: [11, Lemma 4.3] For any sequence $x : \mathbb{N}_0 \to \mathbb{N}_0$ such that $x_n \leq \lfloor (n - \gamma_n)/3 \rfloor$ for all n, we have

$$\mathcal{Z}(x_n; K_n, \gamma_n) \Rightarrow |C_{max}(n, K_n, \gamma_n)| > n - \gamma_n - x_n. \quad (7)$$

Lemma 4.2 states that if the event $\mathcal{Z}(x_n; K_n, \gamma_n)$ holds, then the size of the largest connected component of $\mathbb{H}(n; K_n, \gamma_n)$ is greater than $n - \gamma_n - x_n$; i.e., there are less than x_n nodes outside of the giant component of $\mathbb{H}(n; K_n, \gamma_n)$. Hence, we can see that $\mathbb{H}(n; K_n, \gamma_n)$ is connected if $\mathcal{Z}(x_n; K_n, \gamma_n)$ takes place with $x_n = 1$. Thus, the one-law will be established if we show that $\lim_{n \to \infty} \mathbb{P}[\mathcal{Z}(x_n; K_n, \gamma_n)^{\mathrm{c}}] = 0$ with $x_n = 1$. From the definition of $\mathcal{Z}(x_n; K_n, \gamma_n)$, we have

$$\mathcal{Z}(x_n; K_n, \gamma_n) = \bigcap_{S \in \mathcal{P}_n : \ x_n \leq |S| \leq \lfloor \frac{n-\gamma}{2} \rfloor} (\mathcal{E}_n(K_n, \gamma_n; S))^{\mathrm{c}},$$

where $\mathcal{P}_n$ is the collection of all non-empty subsets of R. Complementing both sides and using union bound, we get

$$\mathbb{P}\left[(\mathcal{Z}(x_n; K_n, \gamma_n))^{\mathrm{c}}\right] \leq \sum_{S \in \mathcal{P}_n : x_n \leq |S| \leq \lfloor \frac{n-\gamma}{2} \rfloor} \mathbb{P}[\mathcal{E}_n(K_n, \gamma_n; S)]$$

$$= \sum_{r=x_n}^{\lfloor \frac{n-\gamma}{2} \rfloor} \sum_{S \in \mathcal{P}_{n,r}} \mathbb{P}[\mathcal{E}_n(K_n, \gamma_n; S)], \quad (8)$$

where $\mathcal{P}_{n,r}$ denotes the collection of all subsets of R with exactly r elements. For each $r = 1, \ldots, \lfloor (n-\gamma)/2 \rfloor$, we can simplify the notation by denoting $\mathcal{E}_{n,r}(K_n, \gamma_n) = \mathcal{E}_n(K_n, \gamma_n; \{v_1, \ldots, v_r\})$. From the exchangeability of the node labels and associated random variables, we have

$$\mathbb{P}[\mathcal{E}_n(K_n, \gamma_n; S)] = \mathbb{P}[\mathcal{E}_{n,r}(K_n, \gamma_n)], \quad S \in \mathcal{P}_{n,r}.$$

$|\mathcal{P}_{n,r}| = \binom{n-\gamma_n}{r}$, since there are $\binom{n-\gamma_n}{r}$ subsets of R with r elements. Thus, we have

$$\sum_{S \in \mathcal{P}_{n,r}} \mathbb{P}[\mathcal{E}_n(K_n, \gamma_n; S)] = \binom{n-\gamma_n}{r} \mathbb{P}[\mathcal{E}_{n,r}(K_n, \gamma_n)].$$

Substituting this into (8), we obtain

$$\mathbb{P}\left[(\mathcal{Z}(x_n; K_n, \gamma_n))^{\mathrm{c}}\right] \leq \sum_{r=x_n}^{\lfloor \frac{n-\gamma}{2} \rfloor} \binom{n-\gamma_n}{r} \mathbb{P}[\mathcal{E}_{n,r}(K_n, \gamma_n)] \quad (9)$$

Remember that $\mathcal{E}_{n,r}(K_n, \gamma_n)$ is the event that the $n-\gamma_n-r$ nodes in S and r nodes in S^{c} do not pick each other; but they can pick from the γ_n nodes from D. Thus, we have

$$\begin{aligned}\mathbb{P}[\mathcal{E}_{n,r}(K_n, \gamma_n)] &= \left(\frac{\binom{\gamma_n+r-1}{K_n}}{\binom{n-1}{K_n}}\right)^r \left(\frac{\binom{n-r-1}{K_n}}{\binom{n-1}{K_n}}\right)^{n-\gamma_n-r} \\ &\leq \left(\frac{\gamma_n + r}{n}\right)^{rK_n} \left(\frac{n-r}{n}\right)^{K_n(n-\gamma_n-r)}\end{aligned}$$

Letting $P_Z = \mathbb{P}\left[\mathcal{Z}(1; K_n, \gamma_n)^{\mathrm{c}}\right]$, $x_n = 1$ and $\gamma_n = \alpha n$ with $0 < \alpha < 1$ in (9), and using the standard bounds $\binom{n}{k} \leq (\frac{ne}{k})^k$ and $1 - x \leq e^{-x}$, we get

$$P_Z \leq \sum_{r=1}^{\lfloor \frac{n-\alpha n}{2} \rfloor} \left(\frac{n-\alpha n}{r}\right)^r e^r \left(\alpha + \frac{r}{n}\right)^{rK_n} e^{\frac{-rK_n(n-\alpha n-r)}{n}}$$

We will show that the right side of the above expression goes to zero as n goes to infinity. Let

$$A_{n,r,\alpha} := \left(\frac{n-\alpha n}{r}\right)^r e^r \left(\alpha + \frac{r}{n}\right)^{rK_n} e^{\frac{-rK_n(n-\alpha n-r)}{n}}.$$

We write

$$P_Z \leq \sum_{r=1}^{\lfloor n/\log n \rfloor} A_{n,r,\alpha} + \sum_{r=\lfloor n/\log n \rfloor}^{\lfloor \frac{n-\alpha n}{2} \rfloor} A_{n,r,\alpha} := Q_1 + Q_2,$$

and show that both Q_1 and Q_2 go to zero as $n \to \infty$. We start with the first summation Q_1.

$$Q_1 \leq \sum_{r=1}^{\lfloor \frac{n}{\log n} \rfloor} \left((1-\alpha)en \cdot e^{K_n \log(\alpha + \frac{1}{\log n}) - K_n(1-\alpha-\frac{1}{\log n})}\right)^r$$

Next, assume as in the statement of Theorem 3.1 that

$$K_n = \frac{c_n \log n}{1 - \alpha - \log \alpha}, \quad n = 1, 2, \ldots \quad (10)$$

for some sequence $c : \mathbb{N}_0 \to \mathbb{R}_+$ such that $\lim_{n\to\infty} c_n = c$ with $c > 1$. Also define

$$a_n := (1-\alpha)en \cdot e^{K_n \log(\alpha + \frac{1}{\log n}) - K_n(1-\alpha-\frac{1}{\log n})}$$

$$\begin{aligned}&= (1-\alpha)en \cdot e^{-c_n \log n \frac{1-\alpha-\log(\alpha+\frac{1}{\log n}) - \frac{1}{\log n}}{1-\alpha-\log\alpha}} \\ &= (1-\alpha)en^{1-c_n} e^{c_n \frac{\log n \log(1+\frac{1}{\alpha \log n}) + 1}{1-\alpha-\log\alpha}} = O(1)n^{1-c_n}\end{aligned}$$

where we substituted K_n via (10) and used the fact that $\log n \cdot \log(1 + \frac{1}{\alpha \log n}) \leq \frac{1}{\alpha}$. Taking the limit as $n \to \infty$ and recalling that $\lim_{n\to\infty} c_n = c > 1$, we see that $\lim_{n\to\infty} a_n = 0$. Hence, for large n, we have

$$Q_1 \leq \sum_{r=1}^{\lfloor n/\log n \rfloor} (a_n)^r \leq \sum_{r=1}^{\infty} (a_n)^r = \frac{a_n}{1-a_n} \quad (11)$$

where the geometric sum converges since $\lim_{n\to\infty} a_n = 0$. Using this once again, it is clear from the last expression that $\lim_{n\to\infty} Q_1 = 0$. Now, similarly consider Q_2.

$$Q_2 \leq \sum_{r=\lfloor n/\log n \rfloor}^{\lfloor (n-\alpha n)/2 \rfloor} \left((1-\alpha)e \log n \cdot e^{K_n \log(\frac{1+\alpha}{2}) - K_n \frac{1-\alpha}{2}}\right)^r$$

Next, we define

$$b_n := (1-\alpha)e \log n \cdot e^{-K_n\left(\frac{1-\alpha}{2} - \log(\frac{1+\alpha}{2})\right)} \quad (12)$$

Substituting for K_n via (10) and taking the limit as $n \to \infty$ it can be seen that $\lim_{n\to\infty} b_n = 0$ upon noting that $\frac{1-\alpha}{2} - \log(\frac{1+\alpha}{2}) > 0$ and $\lim_{n\to\infty} c_n = c > 1$. With arguments similar to those used in the case of Q_1, we can show that when n is large $Q_2 \leq b_n/(1-b_n)$, leading to Q_2 converging to zero as n gets large. With $P_Z \leq Q_1 + Q_2$, and both Q_1 and Q_2 converging to zero when n is large, we establish that P_Z converges to zero as n goes to infinity. This result also yields the desired conclusion $\lim_{n\to\infty} P(n, K_n, \gamma_n) = 1$ in Theorem 3.1 since $P_Z = 1 - P(n, K_n, \gamma_n)$. We direct readers to [19] for proof of other Theorems presented in Section III.

V. Conclusions

In this paper, we provide a comprehensive set of results on the connectivity and giant component size of $\mathbb{H}(n; K_n, \gamma_n)$, i.e., random K-out graph with randomly selected γ_n nodes deleted. Computer simulations are used to validate the results in the finite node regime. Using our results, we compare random K-out graphs with Erdős-Rényi graphs with the same mean node degree and same number of deleted nodes, and show that random K-out graphs are either connected with higher probability or have a larger giant component. This reinforces the usefulness of random K-out graphs in various distributed network applications including federated averaging [5], [21]. Our results can help design networks with desired levels of robustness or tolerance to nodes failing, being captured, or being dishonest in many applications including wireless sensor networks and distributed averaging.

Acknowledgements

This work was supported by the National Science Foundation through Grant # CCF-1617934, and by CyLab through the Secure and Private IoT Initiative. In addition, Mansi Sood was supported by the CyLab Presidential Fellowship and the David H. Barakat and LaVerne Owen-Barakat CIT Dean's Fellowship.

REFERENCES

[1] M. E. Newman, D. J. Watts, and S. H. Strogatz, "Random graph models of social networks," *Proceedings of the National Academy of Sciences*, vol. 99, no. suppl 1, pp. 2566–2572, 2002.

[2] S. M. Kakade, M. Kearns, L. E. Ortiz, R. Pemantle, and S. Suri, "Economic properties of social networks," in *Advances in Neural Information Processing Systems*, 2005, pp. 633–640.

[3] A. Goldenberg, A. X. Zheng, S. E. Fienberg, E. M. Airoldi *et al.*, "A survey of statistical network models," *Foundations and Trends in Machine Learning*, vol. 2, no. 2, pp. 129–233, 2010.

[4] O. Yağan and A. M. Makowski, "On the connectivity of sensor networks under random pairwise key predistribution," *IEEE Transactions on Information Theory*, vol. 59, no. 9, pp. 5754–5762, Sept 2013.

[5] C. Sabater, A. Bellet, and J. Ramon, "Distributed differentially private averaging with improved utility and robustness to malicious parties," 2020.

[6] G. Fanti, S. B. Venkatakrishnan, S. Bakshi, B. Denby, S. Bhargava, A. Miller, and P. Viswanath, "Dandelion++: Lightweight cryptocurrency networking with formal anonymity guarantees," *Proc. ACM Meas. Anal. Comput. Syst.*, vol. 2, no. 2, pp. 29:1–29:35, Jun. 2018.

[7] O. Yağan and A. M. Makowski, "Modeling the pairwise key predistribution scheme in the presence of unreliable links," *IEEE Transactions on Information Theory*, vol. 59, no. 3, pp. 1740–1760, March 2013.

[8] F. Yavuz, J. Zhao, O. Yağan, and V. Gligor, "Designing secure and reliable wireless sensor networks under a pairwise key predistribution scheme," in *2015 IEEE International Conference on Communications (ICC)*. IEEE, 2015, pp. 6277–6283.

[9] H. Chan, A. Perrig, and D. Song, "Random key predistribution schemes for sensor networks," in *Proc. of IEEE S&P 2003*, 2003.

[10] R. Eletreby and O. Yağan, "Connectivity of inhomogeneous random K-out graphs," *IEEE Transactions on Information Theory*, vol. 66, no. 11, pp. 7067–7080, 2020.

[11] M. Sood and O. Yağan, "On the size of the giant component in inhomogeneous random k-out graphs," *arXiv preprint arXiv:2009.01610*, 2020, to appear in 59th IEEE Conference on Decision and Control (CDC 2020), December 2020.

[12] T. I. Fenner and A. M. Frieze, "On the connectivity of random m-orientable graphs and digraphs," *Combinatorica*, vol. 2, no. 4, pp. 347–359, Dec 1982.

[13] O. Yağan and A. M. Makowski, "Wireless sensor networks under the random pairwise key predistribution scheme: Can resiliency be achieved with small key rings?" *IEEE/ACM Transactions on Networking*, vol. 24, no. 6, pp. 3383–3396, 2016.

[14] O. Yağan and A. M. Makowski, "On the scalability of the random pairwise key predistribution scheme: Gradual deployment and key ring sizes," *Performance Evaluation*, vol. 70, no. 7, pp. 493 – 512, 2013.

[15] A. Frieze and M. Karoński, *Introduction to random graphs*. Cambridge University Press, 2016.

[16] B. Bollobás, *Random graphs*. Cambridge university press, 2001.

[17] F. Yavuz, J. Zhao, O. Yağan, and V. Gligor, "Toward k-connectivity of the random graph induced by a pairwise key predistribution scheme with unreliable links," *IEEE Transactions on Information Theory*, vol. 61, no. 11, pp. 6251–6271, 2015.

[18] ——, "k-connectivity in random K-out graphs intersecting Erdős-Rényi graphs," *IEEE Transactions on Information Theory*, vol. 63, no. 3, pp. 1677–1692, 2017.

[19] E. C. Elumar, M. Sood, and O. Yağan, "On the connectivity and giant component size of random k-out graphs under randomly deleted nodes," 2020. [Online]. Available: http://www.andrew.cmu.edu/user/oyagan/Conferences/ISIT2021b.pdf

[20] P. Erdős and A. Rényi, "On the evolution of random graphs," *Publ. Math. Inst. Hung. Acad. Sci*, vol. 5, no. 1, pp. 17–60, 1960.

[21] P. Dellenbach, A. Bellet, and J. Ramon, "Hiding in the crowd: A massively distributed algorithm for private averaging with malicious adversaries," 2018.

[22] A. Mei, A. Panconesi, and J. Radhakrishnan, "Unassailable sensor networks," in *Proc. of the 4th International Conference on Security and Privacy in Communication Netowrks*, ser. SecureComm '08. New York, NY, USA: ACM, 2008.