Counting triangles, tunable clustering and the small-world property in random key graphs

Osman Yağan, *Member, IEEE* and Armand M. Makowski, *Fellow, IEEE*

Abstract—Random key graphs were introduced to study various properties of the Eschenauer-Gligor key predistribution scheme for wireless sensor networks (WSNs). Recently this class of random graphs has received much attention in contexts as diverse as recommender systems, social network modeling, and clustering and classification analysis. This paper is devoted to analyzing various properties of random key graphs. In particular, we establish a zero-one law for the the existence of triangles in random key graphs, and identify the corresponding critical scaling. This zero-one law exhibits significant differences with the corresponding result in Erdős-Rényi (ER) graphs. We also compute the clustering coefficient of random key graphs, and compare it to that of ER graphs in the many node regime when their expected average degrees are asymptotically equivalent. For the parameter range of practical relevance in both wireless sensor network and social network applications, random key graphs are shown to be much more clustered than the corresponding ER graphs. We also explore the suitability of random key graphs as small world models in the sense of Watts and Strogatz.

Index Terms—Random key graphs; existence of triangles; clustering coefficient; wireless sensor networks; social networks.

1 Introduction

Random key graphs are random graphs that belong to the class of random intersection graphs [26]; they are also called uniform random intersection graphs by some authors [3], [11], [12]. They have appeared recently in application areas as diverse as epidemics in social networks [2], clustering analysis [11], [12], collaborative filtering in recommender systems [18], and random key predistribution for wireless sensor networks (WSNs) [9]. In this last context, random key graphs naturally occur in the study of a random key predistribution scheme introduced by Eschenauer and Gligor [9]: Before deployment, each sensor in a WSN is independently assigned K distinct cryptographic keys which are selected at random from a large pool of P keys. These K keys constitute the key ring of the sensor node and are inserted into its memory module. Two sensor nodes can then establish a secure edge between them if they are within transmission range of each other and if their key rings have at least one key in common; see [9] for implementation details. If we assume *full visibility*, namely that nodes are all within communication range of each other, then secure communication between two nodes requires only that their key rings share at least one key. The resulting notion of adjacency defines the class of random key graphs; see Section 2 for precise definitions.

Much efforts have recently been devoted to developing zero-one laws for the property of connectivity in random key graphs. A key motivation can be found in the need to obtain conditions under which the scheme of Eschenauer and Gligor guarantees secure connectivity with high probability in large networks. An interesting feature of this work lies in the following fact: Although random key graphs are *not* stochastically equivalent to the classical Erdős-Rényi graphs [8], it is possible to *formally* transfer well-known zero-one laws for connectivity in Erdős-Rényi graphs to random key graphs by asymptotically matching their edge probabilities. This approach, which was initiated by Eschenauer and Gligor in their original analysis [9], has now been validated rigorously; see the papers [3], [7], [24], [28], [32], [36] for recent developments. Rybarczyk [24] has shown that this transfer from Erdős-Rényi graphs also works for a number of issues related to the giant component and its diameter.

In view of these developments, it is natural to wonder whether this (formal) transfer technique applies to other graph properties. In particular, in the literature on random graphs there is long standing interest [4], [8], [15], [16], [23], [26] in the containment of certain (small) subgraphs, the simplest one being the *triangle*. This particular case is also of some practical relevance: The number of triangles in a graph is closely related to its clustering coefficient, and for random key graphs this has implications on network resiliency under the EG scheme (e.g., see [7]) and also on its applicability and relevance in different domains including social networks – more on that later.

With these in mind, in the present paper we study the triangle containment problem in random key graphs. In particular, we establish a zero-one law for the existence of

This work was supported in part by NSF Grants CCF-0729093 and CCF-1617934. Part of the work was conducted during Fall 2014 while A.M. Makowski was a Visiting Professor with the Department of Statistics of the Hebrew University of Jerusalem with the support of a fellowship from the Lady Davis Trust. The authors also thank a colleague (who wishes to remain anonymous) for suggestions that lead to a shorter proof of the one law in Theorem 3.4.

Parts of the material were presented in the 47th Annual Allerton Conference on Communication, Control and Computing, Monticello (IL), September 2009, and in the First Workshop on Applications of Graph Theory in Wireless Ad hoc Networks and Sensor Networks (GRAPH-HOC 2009), Chennai (India), December 2009.

O. Yağan is with the Department of Electrical and Computer Engineering and CyLab, Carnegie Mellon University, Pittsburgh, PA 15213 USA (e-mail: oyagan@ece.cmu.edu).

A. M. Makowski is with the Department of Electrical and Computer Engineering, and the Institute for Systems Research, University of Maryland, College Park, MD 20742 USA (e-mail: armand@isr.umd.edu).

triangles and identify the corresponding *critical* scaling. By the help of this result (and its proof), we conclude that in the many node regime, the expected number of triangles in random key graphs is always at least as large as the corresponding quantity in asymptotically matched Erdős-Rényi graphs. For the parameter range of practical relevance in WSNs, we show that this expected number of triangles can be orders of magnitude larger in random key graphs than in Erdős-Rényi graphs, confirming the observations made earlier via simulations by Di Pietro et al. [7].

These results show that transferring results from Erdős-Rényi graphs to random key graphs by matching their edge probabilities is not a valid approach in general, and can be quite misleading in the context of WSNs. In particular, our results indicate that the asymptotic equivalence of random key graphs and Erdős-Rényi graphs (in the sense discussed in [26]) is possible only when the size of key rings is comparable to the network size, a case not very realistic in WSNs due to the severe constraints imposed on the memory and computational capabilities of sensors. This points to the inadequacy of Erdős-Rényi graphs to capture some key properties of the EG scheme in realistic WSN implementations, and reinforces the call for a direct investigation of random key graphs.

The number (and *fraction*) of triangles in a network is closely related to its *clustering coefficient*, a metric known to have a significant impact on the dynamics of many interesting processes that take place on the network; e.g., the diffusion of information and epidemic diseases [10], [20], [21], [37], the propagation of influence [13], [38], and cascading failures [14]. With this in mind, we also study the clustering coefficient of random key graphs and compare it with that of an Erdős-Rényi graph. We observe that the clustering coefficient of a random key graph is never smaller than the clustering coefficient of the corresponding Erdős-Rényi graph with *identical* expected average degree. For the parameter range that is relevant for large scale social networks (as well as WSNs), we show that random key graphs are in fact *much more clustered* than Erdős-Rényi graphs when expected average degrees are asymptotically equivalent. Recalling the fact that random key graphs also have a small *diameter* [24], [31], we then conclude that random key graphs are *small-worlds* in the sense introduced by Watts and Strogatz [27]. This reinforces the possibility of using random key graphs in a wide range of applications including social network modeling.

In line with results currently available for other classes of graphs, e.g., Erdős-Rényi graphs [15, Chap. 3] and random geometric graphs [23, Chap. 3], it would be interesting to consider the containment problem for small subgraphs other than triangles in the context of random key graphs. To the best of our knowledge, this issue has not been considered in the literature. Future work may also consider other properties of random key graphs that might be relevant in various applications; e.g., Hamiltonicity, spectral radius, percolation, etc.

The paper is organized as follows: We formally introduce the class of random key graphs in Section 2, with various definitions for the clustering coefficient presented in Section 2.2. In Section 2.3 we evaluate the first and second moments of the number of triangles in random key graphs. Our main results are presented in Section 3: A zero-one law concerning the containment of triangles in random key graphs is discussed in Section 3.2 while its clustering coefficient is computed in Section 3.3. Relevant definitions and facts concerning Erdős-Rényi graphs are given in Section 4. Section 5 and Section 6 are devoted to comparing random key graphs and Erdős-Rényi graphs in terms of their number of triangles and clustering coefficients, respectively. Section 7 and Section 8 discuss the implications of our results on utilizing random key graph in the context of WSN and social network applications, respectively. The proofs of the main results of the paper are available in Section 10, while some technical results are established in Section 9.

A word on the notation and conventions in use: Unless specified otherwise, all limiting statements, including asymptotic equivalences, are understood with n going to infinity. The random variables (rvs) under consideration are all defined on the same probability triple $(\Omega, \mathcal{F}, \mathbb{P})$; its construction is standard and omitted in the interest of brevity. Probabilistic statements are made with respect to this probability measure $\mathbb{P}$, and we denote the corresponding expectation operator by $\mathbb{E}$. We denote almost sure convergence (under $\mathbb{P}$) by a.s. The indicator function of an event E is denoted by $\mathbf{1}[E]$. For any discrete set S we write $|S|$ for its cardinality. We denote almost sure convergence by a.s.

2 Model and definitions

2.1 Random key graphs

Pick positive integers K and P such that $K \leq P$, and fix $n = 3, 4, \ldots$. We shall group the integers P and K into the ordered pair $\theta \equiv (K, P)$ in order to lighten the notation.

The model of interest here is parametrized by the number n of nodes, the size P of the key pool and the size K of each key ring. For each node $i = 1, \ldots, n$, let $K_i(\theta)$ denote the random set of K distinct keys assigned to node i. Thus, under the convention that the P keys are labeled $1, \ldots, P$, the random set $K_i(\theta)$ is a subset of $\{1, \ldots, P\}$ with $|K_i(\theta)| = K$. The rvs $K_1(\theta), \ldots, K_n(\theta)$ are assumed to be *i.i.d.*, each of which is *uniformly* distributed with

$$\mathbb{P}[K_i(\theta) = S] = \binom{P}{K}^{-1}, \qquad i = 1, \ldots, n \tag{1}$$

for any subset S of $\{1, \ldots, P\}$ with $|S| = K$. This corresponds to selecting keys randomly and *without* replacement from the key pool.

Distinct nodes $i, j = 1, \ldots, n$ are said to be adjacent if they share at least one key in their key rings, namely

$$K_i(\theta) \cap K_j(\theta) \neq \emptyset, \tag{2}$$

in which case an undirected edge is assigned between nodes i and j. The adjacency constraints (2) define an undirected random graph on the vertex set $\{1, \ldots, n\}$, hereafter denoted $\mathbb{K}(n; \theta)$. We refer to this random graph as the *random key graph*.

It is easy to check that

$$\mathbb{P}[K_i(\theta) \cap K_j(\theta) = \emptyset] = q(\theta) \tag{3}$$

with

$$q(\theta) = \begin{cases} 0 & \text{if } P < 2K \\ \frac{\binom{P-K}{K}}{\binom{P}{K}} & \text{if } 2K \leq P. \end{cases} \tag{4}$$

The probability $p(\theta)$ of edge occurrence between any two nodes is therefore given by

$$p(\theta) = 1 - q(\theta). \tag{5}$$

If $P < 2K$ there exists an edge between any pair of nodes, and $\mathbb{K}(n;\theta)$ coincides with the complete graph on the vertex set $\{1,\ldots,n\}$. While it is always the case that $0 \leq q(\theta) < 1$, it is plain from (4) that $q(\theta) > 0$ if and only if $2K \leq P$.

The expression (4) is a consequence of the general fact

$$\mathbb{P}\left[S \cap K_i(\theta) = \emptyset\right] = \frac{\binom{P-|S|}{K}}{\binom{P}{K}}, \quad i = 1,\ldots,n \tag{6}$$

valid for any subset S of $\{1,\ldots,P\}$ with $|S| \leq P - K$.

We close by introducing the events

$$E_{ij}(\theta) = \left[K_i(\theta) \cap K_j(\theta) \neq \emptyset\right], \quad i,j = 1,\ldots,n$$

whose indicator functions

$$\xi_{ij}(\theta) = \mathbf{1}\left[K_i(\theta) \cap K_j(\theta) \neq \emptyset\right], \quad i,j = 1,\ldots,n$$

are the edge rvs defining the random key graph $\mathbb{K}(n;\theta)$. For each $i = 1,\ldots,n$, it is a simple matter to check with the help of (6) that the events $\{E_{ij}(\theta),\ j \neq i, j = 1,\ldots,n\}$ are mutually independent, or equivalently, that the rvs $\{\xi_{ij}(\theta),\ j \neq i, j = 1,\ldots,n\}$ form a collection of i.i.d. rvs.

2.2 Clustering coefficient

Many networks encountered in practice exhibit high clustering (or transitivity) in that the neighbors of a node are likely to be neighbors to each other [25] – Your friends are likely to be friends! Clustering properties are known to have a significant impact on the dynamics of many interesting processes that take place on a network, e.g., the diffusion of information and epidemic diseases [10], [20], [21], [37], the propagation of influence [13], [38], and cascading failures [14]. With this in mind we shall investigate clustering in random key graphs under various parameter regimes.

A formal definition of clustering is given next. Consider an undirected graph G with no self-loops on the vertex set V. For each i in V, let $T_i(G)$ denote the number of distinct *triangles* in G that contain vertex i. The *local* clustering coefficient of *node* i is given by

$$C_i(G) = \begin{cases} \frac{T_i(G)}{\frac{1}{2}d_i(d_i-1)} & \text{if } d_i \geq 2 \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

where d_i is the degree of node i in G.

There are, however, several possible definitions for a graph-*wide* notion of clustering [22]: Inspired by (7), it is natural to consider the *average* of the local clustering coefficient $C_{\text{Avg}}(G)$ over the graph G, i.e.,

$$C_{\text{Avg}}(G) = \frac{1}{|V'|}\sum_{i \in V'} C_i(G) \tag{8}$$

where $V' = \{i \in V :\ d_i \geq 2\}$. This last quantity, while natural, is often replaced by the *global* clustering coefficient defined as the "fraction of transitive triples" over the whole graph G, namely,

$$C^\star(G) = \frac{\sum_{i \in V} T_i(G)}{\frac{1}{2}\sum_{i \in V} d_i(d_i-1)} \tag{9}$$

provided $\sum_{i \in V} d_i(d_i-1) > 0$. It is convenient to set $C^\star(G) = 0$ otherwise.

In the context of random graphs, related (but simpler) definitions are possible when the edge assignment rvs are *exchangeable* (as is the case for the random graphs of interest here). Recall that an undirected random graph $\mathbb{G}$ defined over the set of nodes $\{1,\ldots,n\}$ is characterized by the $\{0,1\}$-valued edge rvs $\{\xi_{ij},\ i,j = 1,\ldots,n\}$ with the interpretation that $\xi_{ij} = 1$ (resp. $\xi_{ij} = 0$) if there is an edge (resp. no edge) between nodes i and j. As we consider graphs which are undirected with no self-loops, we impose the conditions

$$\xi_{ij} = \xi_{ji} \quad \text{and} \quad \xi_{ii} = 0, \quad i,j = 1,\ldots,n.$$

A case of great interest arises when the rvs $\{\xi_{ij},\ 1 \leq i < j \leq n\}$ form a family of *exchangeable* rvs [1]. In that setting, a popular approach (e.g., see [6]) is to define the clustering coefficient of the random graph $\mathbb{G}$ as the conditional probability

$$C(\mathbb{G}) = \mathbb{P}\left[E_{12} \mid E_{13} \cap E_{23}\right] \tag{10}$$

where we have used the notation

$$E_{ij} = \left[\xi_{ij} = 1\right], \quad i,j = 1,\ldots,n.$$

For the random graphs considered here, we show that the quantity (10) provides a good approximation to the global clustering coefficient defined at (9), when n is large; see Theorem 3.7 and Theorem 4.1. It is for this reason that we use the simpler definition (10) for studying clustering in the remainder of this paper.

2.3 Counting triangles

Pick positive integers K and P such that $K \leq P$, and fix $n = 3,4,\ldots$ For distinct $i,j,k = 1,\ldots,n$, we define the indicator function

$$\chi_{ijk}(\theta) = \mathbf{1}\left[\begin{array}{c}\text{Nodes } i,\ j \text{ and } k \text{ form} \\ \text{a triangle in } \mathbb{K}(n;\theta)\end{array}\right]. \tag{11}$$

The number of distinct triangles in $\mathbb{K}(n;\theta)$ is then simply given by

$$T_n(\theta) = \sum\nolimits_{1 \leq i < j < k \leq n}^{n} \chi_{ijk}(\theta). \tag{12}$$

Of particular interest is the event that there exists at least one triangle in $\mathbb{K}(n;\theta)$, namely $[T_n(\theta) > 0] = [T_n(\theta) = 0]^c$.

One of our main results is a zero-one law for the existence of triangles in random key graphs. These results will be established by the method of first and second moments applied to the count variables (12), e.g., see [4, p. 2], [15, p. 55]. They are stated in terms of the quantity

$$\tau(\theta) = \frac{K^3}{P^2} + \left(\frac{K^2}{P}\right)^3, \qquad \begin{array}{c}\theta = (K,P) \\ K,P = 1,2,\ldots\end{array} \tag{13}$$

As we shall see soon in Proposition 3.2, this quantity gives the *asymptotic* probability of a triangle in random key graphs, when the parameters K and P are suitably scaled.

Key to much of the discussion carried out in this paper are the first two moments of the count variables (12). The first moment, computed next, will be conveniently expressed with the help of the quantity $\beta(\theta)$ given by

$$\beta(\theta) = (1 - q(\theta))^3 + q(\theta)^3 - q(\theta)r(\theta) \tag{14}$$

with $r(\theta)$ defined by

$$r(\theta) = \begin{cases} 0 & \text{if } P < 3K \\ \frac{\binom{P-2K}{K}}{\binom{P}{K}} & \text{if } 3K \leq P. \end{cases} \tag{15}$$

Note that $r(\theta)$ corresponds to the probability (6) when $|S| = 2K$.

Proposition 2.1. *Fix $n = 3, 4, \ldots$. For positive integers K and P such that $K \leq P$, we have*

$$\mathbb{E}\left[\chi_{123}(\theta)\right] = \beta(\theta) \tag{16}$$

with $\beta(\theta)$ defined at (14), so that

$$\mathbb{E}\left[T_n(\theta)\right] = \binom{n}{3}\beta(\theta). \tag{17}$$

A proof of Proposition 2.1 is given in Section 9.1. We see from (16) that the quantity $\beta(\theta)$ gives the probability that three distinct vertices form a triangle in $\mathbb{K}(n;\theta)$. For future reference, we note that

$$r(\theta) \leq q(\theta)^2 \tag{18}$$

by direct inspection, whence

$$\beta(\theta) \geq (1 - q(\theta))^3 > 0. \tag{19}$$

The second moment of the count variables (12) is computed next; it will play a crucial role in the proofs of both Theorem 3.4 and Theorem 3.7 that are forthcoming.

Proposition 2.2. *For positive integers K and P such that $K \leq P$, we have*

$$\mathbb{E}\left[T_n(\theta)^2\right] = \mathbb{E}\left[T_n(\theta)\right] + \left(\frac{\binom{n-3}{3}}{\binom{n}{3}} + 3\frac{\binom{n-3}{2}}{\binom{n}{3}}\right)\left(\mathbb{E}\left[T_n(\theta)\right]\right)^2 + 3(n-3)\binom{n}{3} \cdot \mathbb{E}\left[\chi_{123}(\theta)\chi_{124}(\theta)\right] \tag{20}$$

for all $n = 3, 4, \ldots$

The proof of Proposition 2.2 is available in Section 9.2.

3 Main results

For simplicity of exposition we refer to any pair of functions $P, K : \mathbb{N}_0 \to \mathbb{N}_0$ as a *scaling* (for random key graphs) provided the natural condition $K_n \leq P_n$ holds for all $n = 3, 4, \ldots$.

3.1 Two asymptotic equivalences

The two asymptotic equivalence results (under such scalings) presented next will prove useful in a number of places. They provide easy asymptotic expressions for the edge probability and for the probability of a triangle, respectively, in large random key graphs. The first one, already obtained in [32], is given here for easy reference.

Lemma 3.1. *For any scaling $P, K : \mathbb{N}_0 \to \mathbb{N}_0$, we have*

$$\lim_{n\to\infty} q(\theta_n) = 1 \quad \textit{if and only if} \quad \lim_{n\to\infty} \frac{K_n^2}{P_n} = 0, \tag{21}$$

and under either condition at (21), the asymptotic equivalence

$$1 - q(\theta_n) \sim \frac{K_n^2}{P_n} \tag{22}$$

holds.

The next result shows that under certain conditions the quantity (13) behaves asymptotically like (14) (which gives the probability that three nodes form a triangle in random key graphs).

Proposition 3.2. *For any scaling $P, K : \mathbb{N}_0 \to \mathbb{N}_0$ satisfying (21), we have the asymptotic equivalence*

$$\beta(\theta_n) \sim \tau(\theta_n). \tag{23}$$

A proof of Proposition 3.2 is given in Section 9.3. In words, this result shows that under (21) the probability of three vertices forming a triangle in random key graphs is asymptotically equivalent to

$$\tau(\theta_n) = \frac{K_n^3}{P_n^2} + \left(\frac{K_n^2}{P_n}\right)^3.$$

3.2 Zero-one laws for the existence of triangles

The zero-law, which is given first, is established in Section 10.1.

Theorem 3.3. *For any scaling $P, K : \mathbb{N}_0 \to \mathbb{N}_0$, the zero-law*

$$\lim_{n\to\infty} \mathbb{P}\left[T_n(\theta_n) > 0\right] = 0$$

holds under the condition

$$\lim_{n\to\infty} n^3\tau(\theta_n) = 0. \tag{24}$$

The one-law given next assumes a more involved form; its proof is given in Section 10.2.

Theorem 3.4. *For any scaling $P, K : \mathbb{N}_0 \to \mathbb{N}_0$ for which the limit $\lim_{n\to\infty} q(\theta_n) = q^\star$ exists, the one-law*

$$\lim_{n\to\infty} \mathbb{P}\left[T_n(\theta_n) > 0\right] = 1$$

holds if either $0 \leq q^\star < 1$, or if $q^\star = 1$ and the additional condition

$$\lim_{n\to\infty} n^3\tau(\theta_n) = \infty \tag{25}$$

holds.

To facilitate an upcoming comparison with analogous results in ER graphs, we combine Theorem 3.3 and Theorem 3.4 into a single symmetric statement.

Theorem 3.5. For any scaling $P, K : \mathbb{N}_0 \to \mathbb{N}_0$ for which $\lim_{n\to\infty} q(\theta_n)$ exists, we have

$$\lim_{n\to\infty} \mathbb{P}\left[T_n(\theta_n) > 0\right] = \begin{cases} 0 & \text{if } \lim_{n\to\infty} n^3\tau(\theta_n) = 0 \\ \\ 1 & \text{if } \lim_{n\to\infty} n^3\tau(\theta_n) = \infty. \end{cases}$$

By Lemma 3.1 the condition $\lim_{n\to\infty} n^3\tau(\theta_n) = 0$ implies $\lim_{n\to\infty} q(\theta_n) = 1$, hence he limit $\lim_{n\to\infty} q(\theta_n)$ necessarily exists with $q^\star = 1$.

3.3 Clustering in random key graphs

In accordance with definition (10), the clustering coefficient of the random key graph $\mathbb{K}(n;\theta)$ is defined by

$$C_{\rm K}(\theta) = \mathbb{P}\left[E_{12}(\theta) \mid E_{13}(\theta) \cap E_{23}(\theta)\right]. \tag{26}$$

A closed form expression for this quantity is given next.

Proposition 3.6. For positive integers K, P such that $K \le P$, we have

$$C_{\rm K}(\theta) = \frac{\beta(\theta)}{(1-q(\theta))^2} \tag{27}$$

with $\beta(\theta)$ given by (14).

Proof. The definitions of $C_{\rm K}(\theta)$ and $\chi_{123}(\theta)$ yield

$$C_{\rm K}(\theta) = \frac{\mathbb{P}\left[E_{12}(\theta) \cap E_{13}(\theta) \cap E_{23}(\theta)\right]}{\mathbb{P}\left[E_{13}(\theta) \cap E_{23}(\theta)\right]} = \frac{\mathbb{E}\left[\chi_{123}(\theta)\right]}{(1-q(\theta))^2} \tag{28}$$

since the events $E_{13}(\theta)$ and $E_{23}(\theta)$ are independent, with

$$\begin{aligned} &\mathbb{P}\left[E_{13}(\theta) \cap E_{23}(\theta)\right] \\ &= \mathbb{P}\left[K_1(\theta) \cap K_3(\theta) \neq \emptyset, K_2(\theta) \cap K_3(\theta) \neq \emptyset\right] \\ &= \left(1-q(\theta)\right)^2 \end{aligned} \tag{29}$$

by virtue of (6) (and comments following it). The conclusion (27) is immediate upon substituting (16) into (28). ∎

For random key graphs there is strong consistency between the definitions (9) and (26) of clustering coefficient.

Theorem 3.7. For positive integers K, P such that $K \le P$, we have

$$\lim_{n\to\infty} C^\star(\mathbb{K}(n;\theta)) = C_{\rm K}(\theta) \quad a.s. \tag{30}$$

A proof of Theorem 3.7 is given in Section 10.3. To the best of our knowledge, Theorem 3.7 is the first rigorous result in the literature that shows that the conditional probability definition (10) of clustering coefficient converges asymptotically almost surely to the empirical clustering coefficient measure of (9). For instance, Deijfen and Kets indicated [6], for another class of random graphs, that the two definitions should be closely related, but that a rigorous proof would need significant additional work.

Simulation results given in Table I illustrate the convergence (30) for several realistic parameter values. The numerical values of $C_{\rm K}(\theta)$ are obtained directly from the expressions (26). The quantity $\widehat{C}_n^\star(\theta)$ stands for the clustering coefficient of $\mathbb{K}(n;\theta)$, calculated through (9) and averaged over 1000 realizations; the number of nodes is set to $n = 1000$ in all simulations. The data support the validity of (30), and confirm the claim that for large networks the quantity (9) captures essentially the same structural information as (26).

K	P	$1-q(\theta)$	$C_K(\theta)$	$\widehat{C}_n^\star(\theta)$
4	10^3	0.0159	0.2590	0.2587
8	5×10^3	0.0127	0.1348	0.1349
16	2×10^4	0.0127	0.0737	0.0736
20	4×10^4	0.0100	0.0590	0.0590
24	10^5	0.0057	0.0469	0.0468
32	10^5	0.0102	0.0408	0.0408
40	5×10^5	0.0032	0.0280	0.0280
64	10^6	0.0041	0.0196	0.0196

TABLE 1
Clustering coefficients with fixed θ for random key graphs

4 Facts concerning Erdős-Rényi graphs

A little later in this paper, we shall compare random key graphs to related Erdős-Rényi (ER) graphs [8], but first some notation: For each $n = 2, 3, \ldots$ and each p in $[0,1]$, let $\mathbb{G}(n;p)$ denote the ER graph on the vertex set $\{1,\ldots,n\}$ with edge probability p. The ER graph $\mathbb{G}(n;p)$ is characterized by the fact that the $\frac{n(n-1)}{2}$ possible undirected edges between the n nodes are independently assigned with probability p. Thus, if in analogy with earlier notation, with distinct $i, j = 1,\ldots,n$, we denote by $E_{ij}(p)$ the event that there is an (undirected) edge between nodes i and j in $\mathbb{G}(n;p)$, then the events $\{E_{ij}(p),\ 1 \le i < j \le n\}$ are mutually independent, each of probability p. For ease of exposition it will always be understood that $E_{ij}(p) = E_{ji}(p)$ for distinct $i, j = 1,\ldots,n$.

Random key graphs are *not* stochastically equivalent to ER graphs even when their edge probabilities are matched exactly: As graph-valued rvs, the random graphs $\mathbb{G}(n;p)$ and $\mathbb{K}(n;\theta)$ have different distributions even under the *exact matching* condition

$$p = 1 - q(\theta) = p(\theta). \tag{31}$$

See [29] for a discussion of (dis)similarities. Under (31) the random graphs $\mathbb{G}(n;p)$ and $\mathbb{K}(n;\theta)$ are said to be exactly matched.

In analogy with (12) let $T_n(p)$ denote the number of distinct triangles in $\mathbb{G}(n;p)$. Under the enforced independence, we note that

$$\mathbb{E}\left[T_n(p)\right] = \binom{n}{3}\tau^\star(p), \quad n = 3, 4, \ldots \tag{32}$$

with

$$\tau^\star(p) = p^3, \quad 0 \le p \le 1.$$

The edge assignment rvs being exchangeable in ER graphs, we can again define the clustering coefficient in $\mathbb{G}(n;p)$ according to (10) by setting

$$C_{\rm ER}(p) = \mathbb{P}\left[E_{12}(p) \mid E_{13}(p) \cap E_{23}(p)\right]. \tag{33}$$

By mutual independence of the edge rvs it follows that

$$C_{\rm ER}(p) = \frac{\mathbb{P}\left[E_{12}(p) \cap E_{13}(p) \cap E_{23}(p)\right]}{\mathbb{P}\left[E_{13}(p) \cap E_{23}(p)\right]} = p. \tag{34}$$

Here as well, strong consistency holds between the two notions of clustering (9) and (26).

Theorem 4.1. For every p in $(0,1)$, we have

$$\lim_{n\to\infty} C^\star(\mathbb{G}(n;p)) = C_{\rm ER}(p) \quad a.s. \tag{35}$$

K	P	$C_K(\theta)$	$\widehat{C}^\star_n(\theta)$	$C_{ER}(p)$	$\widehat{C}^\star_n(p)$
4	10^3	0.2590	0.2587	0.0159	0.0159
8	5×10^3	0.1348	0.1349	0.0127	0.0128
16	2×10^4	0.0737	0.0736	0.0127	0.0128
20	4×10^4	0.0590	0.0590	0.0100	0.0100
24	10^5	0.0469	0.0468	0.0057	0.0057
32	10^5	0.0408	0.0408	0.0102	0.0102
40	5×10^5	0.0280	0.0280	0.0032	0.0031
64	10^6	0.0196	0.0196	0.0041	0.0041

TABLE 2
Clustering coefficients with fixed θ and $p = 1 - q(\theta)$ computed via (4)

This result can be established by arguments similar to the ones provided in the proof of Proposition 3.7; due to space limitations we give it in [33]. Table II expands on Table I given earlier in that we now compare the clustering coefficients of exactly matched random key graphs and ER graphs for the parameter values used in Table I. The quantities for random key graphs are as before. The numerical values of $C_{\rm ER}(p)$ are obtained directly from the expressions (33). Here $\widehat{C}^\star_n(p)$ stands for the clustering coefficient of $\mathbb{G}(n;p)$. It is calculated through (9) and averaged over 1000 realizations. The number of nodes is still set to $n = 1000$ in all simulations. Again the data support the claim that for large networks the definition (9) captures essentially the same information as the quantity (33).

Any mapping $p : \mathbb{N}_0 \to [0,1]$ will be called a scaling for ER graphs. In order to meaningfully compare the asymptotic regime of random key graphs with that of ER graphs under their respective scalings, we shall say that the scaling $p : \mathbb{N}_0 \to [0,1]$ (for ER graphs) is *asymptotically matched* to the scaling $P, K : \mathbb{N}_0 \to \mathbb{N}_0$ (for random key graphs) if

$$p_n \sim p(\theta_n) = 1 - q(\theta_n). \tag{36}$$

Sometimes, when (36) holds, we shall also say that the random graphs $\mathbb{G}(n;p_n)$ and $\mathbb{K}(n;\theta_n)$ are asymptotically matched. Under condition (21), by Lemma 3.1 the asymptotic matching condition (36) amounts to

$$p_n \sim \frac{K_n^2}{P_n}. \tag{37}$$

Condition (31) (resp. (36)) is equivalent to requiring that the expected degrees in $\mathbb{K}(n;\theta)$ and $\mathbb{G}(n;p)$ (resp. $\mathbb{K}(n;\theta_n)$ and $\mathbb{G}(n;p_n)$) coincide (resp. are asymptotically equivalent).

5 Comparing the number of triangles in random key graphs and ER graphs

Fix p in $(0,1]$, and positive integers K and P such that $K \leq P$. From (17) and (32) it is plain that

$$\frac{\mathbb{E}\left[T_n(\theta)\right]}{\mathbb{E}\left[T_n(p)\right]} = \frac{\beta(\theta)}{\tau^\star(p)}, \quad n = 3,4,\ldots \tag{38}$$

Under the *exact* matching condition (31), with $p(\theta)$ given by (5), this last expression yields

$$\frac{\mathbb{E}\left[T_n(\theta)\right]}{\mathbb{E}\left[T_n(p(\theta))\right]} = \frac{\beta(\theta)}{\tau^\star(p(\theta))} = 1 + \frac{q(\theta)^2 - r(\theta)}{(1-q(\theta))^3} \cdot q(\theta)$$

for each $n = 3,4,\ldots$, whence

$$\mathbb{E}\left[T_n(p(\theta))\right] \leq \mathbb{E}\left[T_n(\theta)\right], \quad n = 3,4,\ldots$$

by virtue of (18). Consequently, the expected number of triangles in a random key graph is always at least as large as the corresponding quantity in an ER graph exactly matched to it. This was already suggested by Di Pietro et al. [7] with the help of limited simulations.

An analogous result is available when the scalings are only *asymptotically* matched.

Corollary 5.1. *Consider a scaling $K, P : \mathbb{N}_0 \to \mathbb{N}_0$ satisfying (21), and a scaling $p : \mathbb{N}_0 \to [0,1]$. Under the asymptotic matching condition (36), we have the equivalence*

$$\frac{\mathbb{E}\left[T_n(\theta_n)\right]}{\mathbb{E}\left[T_n(p_n)\right]} \sim 1 + \frac{P_n}{K_n^3}. \tag{39}$$

In other words, for large n the expected number of triangles in random key graphs is always at least as large as the corresponding quantity in asymptotically matched ER graphs – In fact, if the ratio P_n/K_n^3 is large, the number of triangles in random key graphs can be several orders of magnitude larger than that of ER graphs. In Sections 7 and 8 this issue is explored in the context of wireless sensor networks and social networks, respectively.

Proof. Replacing θ by θ_n and p by p_n according to the given scalings in the expression (38), we get

$$\frac{\mathbb{E}\left[T_n(\theta_n)\right]}{\mathbb{E}\left[T_n(p_n)\right]} = \frac{\beta(\theta_n)}{\tau^\star(p_n)}, \quad n = 3,4,\ldots$$

Under (21), Proposition 3.2 yields

$$\frac{\mathbb{E}\left[T_n(\theta_n)\right]}{\mathbb{E}\left[T_n(p_n)\right]} \sim \frac{\tau(\theta_n)}{\tau^\star(p_n)} \tag{40}$$

with

$$\frac{\tau(\theta_n)}{\tau^\star(p_n)} = \frac{1}{p_n^3} \cdot \left(\frac{K_n^3}{P_n^2}\right) + \frac{1}{p_n^3} \cdot \left(\frac{K_n^2}{P_n}\right)^3, \quad n = 3,4,\ldots$$

With the help of (37), we conclude

$$\frac{\tau(\theta_n)}{\tau^\star(p_n)} \sim 1 + \frac{P_n}{K_n^3} \tag{41}$$

and the equivalence (39) follows from (40). ■

From (41) it follows that under the asymptotic matching condition (36) (together with (21)), triangles will start appearing *earlier* in the evolution of a random key graph as compared to an ER graph (asymptotically) matched to it. It should also be clear from (41) that the larger the quantity P_n/K_n^3, the more pronounced will such difference be.

We close this section by comparing Theorem 3.5 with its analog for ER graphs. Fix $n = 3,4,\ldots$ and p in $[0,1]$. Consider the event that there exists at least one triangle in $\mathbb{G}(n;p)$, i.e., $[T_n(p) > 0]$. The following zero-one law for triangle containment in ER graphs is well known [4, Chap. 4], [15, Thm. 3.4, p. 56].

Theorem 5.2. *For any scaling $p : \mathbb{N}_0 \to [0,1]$, we have*

$$\lim_{n\to\infty} \mathbb{P}\left[T_n(p_n) > 0\right] = \begin{cases} 0 & \text{if } \lim_{n\to\infty} n^3 \tau^\star(p_n) = 0 \\ \\ 1 & \text{if } \lim_{n\to\infty} n^3 \tau^\star(p_n) = \infty. \end{cases}$$

This result, which is also established by the method of first and second moments, is easily understood once we recall (32). As we compare Theorem 3.5 with Theorem 5.2, we note a direct analogy since the terms $\tau(\theta_n)$ and $\tau^\star(p_n)$ correspond to the (asymptotic) probability that three arbitrary nodes form a triangle in random key graphs and ER graphs, respectively.

6 Comparing the clustering coefficients of random key graphs and ER graphs

Fix p in $(0,1]$, and positive integers K and P such that $K \leq P$. Combining (27) and (34) we get

$$\frac{C_K(\theta)}{C_{\rm ER}(p)} = \frac{\beta(\theta)}{p(1-q(\theta))^2}. \tag{42}$$

Under the exact matching condition (31) we find

$$\begin{aligned} \frac{C_K(\theta)}{C_{\rm ER}(p(\theta))} &= \frac{\beta(\theta)}{(1-q(\theta)))^3} \\ &= 1 + \frac{q(\theta)^2 - r(\theta)}{(1-q(\theta))^3} \cdot q(\theta) \end{aligned} \tag{43}$$

as we recall (5). Thus,

$$C_{\rm ER}(p(\theta)) \leq C_{\rm K}(\theta) \tag{44}$$

by virtue of (18) – The clustering coefficient of a random key graph is at least as large as that of the ER graph exactly matched to it.

Several conclusions can be extracted from these expressions: Equality in (44) holds only when $P < 2K$, i.e., from (4) we get

$$\frac{C_{\rm K}(\theta)}{C_{\rm ER}(p(\theta))} = 1 \text{ if } K \leq P < 2K$$

since then $q(\theta) = r(\theta) = 0$. If $2K \leq P < 3K$, then $0 < q(\theta) < 1$ but $r(\theta) = 0$, whence

$$\frac{C_K(\theta)}{C_{\rm ER}(p(\theta))} = 1 + \left(\frac{q(\theta)}{1-q(\theta)}\right)^3 > 1.$$

Understanding the case $3K \leq P$ is more challenging due to a lack of simple expressions. Therefore, before dealing with the case of an arbitrary positive integer K, we first consider a couple of special cases as a way to explore the relative ranges possibly exhibited by the clustering coefficients. For $K = 1$ it is a simple matter to check from (43) that

$$\frac{C_{\rm K}(1,P)}{C_{\rm ER}(p(\theta))} = P \tag{45}$$

for each $P = 2, 3, \ldots$. For $K = 2$ uninteresting calculations show that

$$\frac{C_{\rm K}(2,P)}{C_{\rm ER}(p(\theta))} = \frac{P}{2} \cdot \frac{2P^3 - 4P^2 - P + 3}{(2P-3)^3}$$

for each $P = 6, 7, \ldots$, whence

$$\frac{P}{8} < \frac{C_{\rm K}(2,P)}{C_{\rm ER}(p(\theta))} < P \tag{46}$$

on that range. This upper bound is seen to hold by noting that $4(2P^3 - 4P^2 - P + 3) = (2P-3)^3 + (P-1)(20P - 38) + 1 > (2P-3)^3$ for all $P = 2, 3, \ldots$. The lower bound follows from the easily checked fact that $2P^3 - 4P^2 - P + 3 < 2(2P-3)^3$ for all $P = 2, 3, \ldots$.

The cases $K = 1$ and $K = 2$ may not be interesting from the perspective of envisioned modeling applications of random key graphs. However, the discussion already shows that the parameters of the corresponding random key graph can be selected (e.g., by taking P very large in these two cases) so that it has a much larger clustering coefficient than the ER graph exactly matched to it. Additional limited numerical evidence along these lines is also available in Table II discussed earlier. In fact, for any *given* K we see that the linear behavior found in (45) and (46) holds asymptotically for large P.

Corollary 6.1. *For each positive integer K, it holds that*

$$\frac{C_{\rm K}(\theta)}{C_{\rm ER}(p(\theta))} \sim 1 + \frac{P}{K^3} \quad (P \to \infty). \tag{47}$$

Thus, exactly matched random key graphs and ER graphs will have vastly different clustering coefficients when P is large. This will be especially so for WSNs where the size of the key pool P in the Eschenauer-Gligor scheme is expected to be in the range $2^{17} - 2^{20}$ (with K much smaller) [9].

Proof. Fix positive integers K and P such that $2K \leq P$. We can rewrite (43) as

$$\frac{C_K(\theta)}{C_{\rm ER}(p(\theta))} = 1 + \left(\frac{q(\theta)}{1-q(\theta)}\right)^3 \cdot \left(1 - \frac{r(\theta)}{q(\theta)^2}\right). \tag{48}$$

With K fixed and P getting large, we see from Lemma 3.1 that $1 - q(\theta) \sim \frac{K^2}{P}$ and $q(\theta) \sim 1$ $(P \to \infty)$, so that

$$\left(\frac{q(\theta)}{1-q(\theta)}\right)^3 \sim \left(\frac{P}{K^2}\right)^3 \quad (P \to \infty).$$

The arguments given in the proof of Proposition 3.2 to establish (69) can also be used to establish

$$1 - \frac{r(\theta)}{q(\theta)^2} \sim \frac{K^3}{P^2} \quad (P \to \infty). \tag{49}$$

Collecting we conclude to the validity of (47). ■

Next we compare the clustering coefficients of asymptotically matched random key graphs and ER graphs when the parameters θ and p are scaled with n.

Corollary 6.2. *Consider a scaling $K, P : \mathbb{N}_0 \to \mathbb{N}_0$ satisfying (21) and a scaling $p : \mathbb{N}_0 \to [0,1]$. Under the asymptotic matching condition (36), we have the equivalence*

$$\frac{C_{\rm K}(\theta_n)}{C_{\rm ER}(p_n)} \sim 1 + \frac{P_n}{K_n^3}. \tag{50}$$

Proof. As we replace θ by θ_n and p by p_n according to these scalings in the expression (42), we get

$$\frac{C_{\rm K}(\theta_n)}{C_{\rm ER}(p_n)} = \frac{\beta(\theta_n)}{p_n(1-q(\theta_n))^2}, \quad n = 3, 4, \ldots \tag{51}$$

Note that

$$\frac{C_{\rm K}(\theta_n)}{C_{\rm ER}(p_n)} \sim \frac{\beta(\theta_n)}{(1-q(\theta_n))^3} \sim \frac{\tau(\theta_n)}{(1-q(\theta_n))^3}. \tag{52}$$

The first equivalence is a consequence of (36) while the second equivalence follows by Proposition 3.2 under (21). With (37) being still valid here, we easily conclude (50) by the same arguments as the ones used to obtain (39). ■

Under (21) and (36), we conclude that

$$\lim_{n\to\infty} \frac{C_{\mathrm{K}}(\theta_n)}{C_{\mathrm{ER}}(p_n)} = 1 \quad \text{if} \quad \lim_{n\to\infty} \frac{K_n^3}{P_n} = \infty, \tag{53}$$

and

$$\lim_{n\to\infty} \frac{C_{\mathrm{K}}(\theta_n)}{C_{\mathrm{ER}}(p_n)} = \infty \quad \text{if} \quad \lim_{n\to\infty} \frac{K_n^3}{P_n} = 0. \tag{54}$$

Thus, asymptotically matched random key graphs and ER graphs can in principle have vastly different clustering coefficients. We explore this possibility in the next two sections where the implications of the main results are discussed in the context of wireless sensor networks and of social networks based on common interest relationships.

7 Wireless Sensor Networks

Random key graphs were originally introduced to model the random key pre-distribution scheme proposed by Eschenauer and Gligor [9] in the context of WSNs. When the WSN comprises n nodes, it is natural to select the parameters K_n and P_n in order for the induced random key graph to be *connected*. However, there is a tradeoff between connectivity and security [7], requiring that $\frac{K_n^2}{P_n}$ be kept as close as possible to the critical scaling $\frac{\log n}{n}$ for connectivity (but above it); see the papers [3], [7], [24], [28], [32]. The desired regime near the boundary can be achieved by taking

$$\frac{K_n^2}{P_n} \sim c \cdot \frac{\log n}{n} \tag{55}$$

with $c > 1$ but close to one.

Now, consider the situation where the random key graph $\mathbb{K}(n;\theta_n)$ is matched asymptotically to the ER random graph $\mathbb{G}(n;p_n)$ under the asymptotic matching condition (36). It follows from (39) that

$$\frac{\mathbb{E}\left[T_n(\theta_n)\right]}{\mathbb{E}\left[T_n(p_n)\right]} \sim 1 \quad \text{if and only if} \quad \frac{P_n}{K_n^3} = o(1) \tag{56}$$

under the condition (21). This last condition obviously occurs when (55) holds, in which case the condition at (56) amounts to taking

$$\frac{1}{K_n} = o(1)\left(c \cdot \frac{\log n}{n}\right).$$

Thus, under the connectivity condition (55) it holds that

$$\frac{\mathbb{E}\left[T_n(\theta_n)\right]}{\mathbb{E}\left[T_n(p_n)\right]} \sim 1 \quad \text{if and only if} \quad \lim_{n\to\infty} \frac{K_n}{n/\log n} = \infty. \tag{57}$$

The expected number of triangles in random key graphs is then of the same order as the corresponding quantity in asymptotically matched ER graphs with $\mathbb{E}\left[T_n(\theta_n)\right] \sim \mathbb{E}\left[T_n(p_n)\right] \sim \frac{c^3}{6}\left(\log n\right)^3$ – This is a direct consequence of (32), (37) and (55). This conclusion holds regardless of the value of c in (55).

However, given the limited memory and computational power of the sensor nodes, key ring sizes satisfying (57) are not practical since requiring $K_n \gg \frac{n}{\log n}$. Furthermore, they will also result in *high* node degrees, and this in turn will decrease network *resiliency* against node capture attacks. It was proposed by Di Pietro et al. [7, Thm. 5.3] that resiliency in large WSNs against node capture attacks can be ensured by selecting K_n and P_n such that $\frac{K_n}{P_n} \sim \frac{1}{n}$. Under (55) this additional requirement then leads to $K_n \sim c \cdot \log n$, whence $P_n \sim c \cdot n \log n$, and (39) now implies

$$\lim_{n\to\infty} \frac{\mathbb{E}\left[T_n(\theta_n)\right]}{\mathbb{E}\left[T_n(p_n)\right]} = \lim_{n\to\infty} \left(1 + \frac{n}{(c\cdot\log n)^2}\right) = \infty. \tag{58}$$

Therefore, for such realistic WSN implementations the expected number of triangles in the induced random key graphs will be orders of magnitude larger than in ER graphs.

Concerning the clustering coefficients, we see that under the condition (55), (53) can hold only if the key ring size K is much larger than $n/\log n$. As already discussed, this condition can not be satisfied in a practical WSN scenario due to storage limitations at the sensor nodes and security constraints. In fact, we see from (39) and (58) that, in a realistic WSN, the condition (54) is always in effect and the clustering coefficient of the random key graph is much larger than that of the asymptotically matched ER graph.

8 Social networks – Can random key graphs be small worlds?

With an obvious change in terminology, random key graphs can be used to model certain types of social networks, e.g., see [2], [36]: Instead of viewing $\{1,\dots,P\}$ as a collection of cryptographic keys randomly assigned to the nodes of a WSN according to the Eschenauer-Gligor scheme, we can think of it as a list of "interests," e.g., hobbies, books, movies, sports, etc., which are pursued by the members of a social group. In that reformulation, the i.i.d. random sets $K_1(\theta),\dots,K_n(\theta)$ appearing in the definition of the random key graph $\mathbb{K}(n;\theta)$ can now be interpreted as the interests assigned to the individual members of that group.[1] The random key graph $\mathbb{K}(n;\theta)$ then naturally describes a *common-interest* relationship between community members since two individuals are now adjacent in $\mathbb{K}(n;\theta)$ when they have at least one interest in common.

In parallel to the discussion given in Section 7 for WSNs, we explore the parameter ranges likely to appear in practice for random key graphs modeling social networks. We do so with an eye towards understanding the behavior of the expression appearing at (39).

We begin with the observation that most real-world social networks are known to be *sparse* in the sense that the expected number of edges per node appears to remain (nearly) constant as the size of the network increases. In the case of random key graphs, the expected degree of a node is given by $(n-1)(1-q(\theta_n))$ and sparsity amounts to $1-q(\theta_n) \sim \frac{c}{n}$ for some $c > 0$, or equivalently, to

$$\frac{K_n^2}{P_n} \sim \frac{c}{n} \tag{59}$$

1. Here we assume that each individual has exactly K interests drawn from the list $\{1,\dots,P\}$. More realistic models can be obtained through more complex randomization mechanisms as in the work of Godehardt et al. on general random intersection graphs [11], [12] and as in the work of Yağan on *inhomogeneous* random key graphs [34].

by virtue of Lemma 3.1, whence

$$\frac{P_n}{K_n^3} \sim \frac{n}{cK_n}. \tag{60}$$

In view of Corollary 5.1 and Corollary 6.2, in the sparse regime, random key graphs will have many more triangles and will be much more clustered (by orders of magnitude) than the asymptotically matched ER graphs unless

$$\limsup_{n\to\infty} \frac{n}{K_n} < \infty. \tag{61}$$

This condition is equivalent to

$$K_n = \Omega(n), \tag{62}$$

and is even more stringent than the corresponding condition (57) derived for WSN applications. More importantly, under the condition (59) we have $P_n \sim c^{-1} n K_n^2$, requiring

$$P_n = \Omega(n^3)$$

if (62) is also enforced. Thus, under (59) and (62) generating the random key graph will require each of the n nodes to choose $K_n = \Omega(n)$ objects from a universe of size $P_n = \Omega(n^3)$. The computational complexity of this task quickly becomes prohibitively high as the number of individuals in the social network becomes large. Yet, we would expect the realistic values for the number K_n of interests of a single individual to be much smaller than the network size n, in sharp contrast with (62). In other words, the condition (62) is naturally eliminated in realistic applications of random key graphs to such social networks – The resulting random key graphs will naturally have very high clustering and contain very large number of triangles when used for social network modeling.

Since random key graphs can be highly clustered, a natural question arises as to their suitability for modeling the *small world effect*. This notion is linked to a well-known series of experiments conducted by Milgram [19] in the late sixties. The results, commonly known as six degrees of separation, suggest that the social network of people in the United States is *small* in the sense that path lengths between pairs of individuals are short. As a way to capture Milgram's experiments, Watts and Strogatz [27] introduced *small world* network models that are highly clustered and yet have a small average path length. More precisely, a random graph is considered to be a small world if its average path length is of the same order as that of an ER graph with the same expected average degree, but with a much larger clustering coefficient.

The results of this paper already show that random key graphs can satisfy the high clustering coefficient requirement of a small world. Under (55), Rybarczyk [24] has shown that

$$\text{diam}[\mathcal{K}(n;\theta_n)] \sim \frac{\log n}{\log\log n}$$

with high probability where $\mathcal{K}(n;\theta_n)$ is the largest connected component of $\mathbb{K}(n;\theta_n)$. This suggests that the diameter, hence the average path length, in random key graphs is *small* as was the case with ER graphs [5]. We also note [31, Corollary 5.2] that random key graphs have *very small* (e.g., ≤ 2) diameter under certain parameter ranges (e.g., with $P_n = O(n^\delta)$ with $0 < \delta < \frac{1}{2}$). Thus, random key graphs may indeed be considered good candidate models for small worlds!

9 Proofs of the preliminary results

In Sections 9.1 and 9.2, we fix positive integers K and P such that $K \leq P$, and $n = 3, 4, \ldots$.

9.1 A proof of Proposition 2.1

As exchangeability yields (17), we need only show the validity of (16). We make repeated use of the fact that for any pair of events E and F in $\mathcal{F}$, we have

$$\mathbb{P}[E \cap F] = \mathbb{P}[E] - \mathbb{P}[E \cap F^c]. \tag{63}$$

Thus, by repeated application of (63) we find

$$\begin{aligned}
&\mathbb{E}[\chi_{123}(\theta)] \\
&= \mathbb{P}\left[\begin{array}{c} K_1(\theta)\cap K_2(\theta)\neq\emptyset, K_1(\theta)\cap K_3(\theta)\neq\emptyset, \\ K_2(\theta)\cap K_3(\theta)\neq\emptyset \end{array}\right] \\
&= \mathbb{P}[K_1(\theta)\cap K_2(\theta)\neq\emptyset, K_1(\theta)\cap K_3(\theta)\neq\emptyset] \\
&\quad - \mathbb{P}\left[\begin{array}{c} K_1(\theta)\cap K_2(\theta)\neq\emptyset, K_1(\theta)\cap K_3(\theta)\neq\emptyset, \\ K_2(\theta)\cap K_3(\theta)=\emptyset \end{array}\right] \\
&= \mathbb{P}[K_1(\theta)\cap K_2(\theta)\neq\emptyset, K_1(\theta)\cap K_3(\theta)\neq\emptyset] \\
&\quad - \mathbb{P}[K_1(\theta)\cap K_2(\theta)\neq\emptyset, K_2(\theta)\cap K_3(\theta)=\emptyset] \\
&\quad + \mathbb{P}\left[\begin{array}{c} K_1(\theta)\cap K_2(\theta)\neq\emptyset, K_1(\theta)\cap K_3(\theta)=\emptyset, \\ K_2(\theta)\cap K_3(\theta)=\emptyset \end{array}\right].
\end{aligned}$$

By independence, with the help of (6), we readily obtain the expressions

$$\mathbb{P}[K_1(\theta)\cap K_2(\theta)\neq\emptyset, K_1(\theta)\cap K_3(\theta)\neq\emptyset] = (1-q(\theta))^2$$

and

$$\begin{aligned}
&\mathbb{P}[K_1(\theta)\cap K_2(\theta)\neq\emptyset, K_2(\theta)\cap K_3(\theta)=\emptyset] \\
&= (1-q(\theta))\,q(\theta).
\end{aligned}$$

Next, as we use (63) one more time, we get

$$\begin{aligned}
&\mathbb{P}\left[\begin{array}{c} K_1(\theta)\cap K_2(\theta)\neq\emptyset, K_1(\theta)\cap K_3(\theta)=\emptyset, \\ K_2(\theta)\cap K_3(\theta)=\emptyset \end{array}\right] \\
&= \mathbb{P}[K_1(\theta)\cap K_3(\theta)=\emptyset, K_2(\theta)\cap K_3(\theta)=\emptyset] \\
&\quad - \mathbb{P}\left[\begin{array}{c} K_1(\theta)\cap K_2(\theta)=\emptyset, K_1(\theta)\cap K_3(\theta)=\emptyset \\ K_2(\theta)\cap K_3(\theta)=\emptyset \end{array}\right].
\end{aligned}$$

Again, by independence, with the help of (6) we conclude that

$$\mathbb{P}[K_1(\theta)\cap K_3(\theta)=\emptyset, K_2(\theta)\cap K_3(\theta)=\emptyset] = q(\theta)^2$$

and

$$\begin{aligned}
&\mathbb{P}\left[\begin{array}{c} K_1(\theta)\cap K_2(\theta)=\emptyset, K_1(\theta)\cap K_3(\theta)=\emptyset \\ K_2(\theta)\cap K_3(\theta)=\emptyset \end{array}\right] \\
&= \mathbb{P}\left[\begin{array}{c} K_1(\theta)\cap K_2(\theta)=\emptyset, \\ K_3(\theta)\cap(K_1(\theta)\cup K_2(\theta))=\emptyset \end{array}\right] \\
&= q(\theta)r(\theta)
\end{aligned}$$

since $|K_1(\theta)\cup K_2(\theta)| = 2K$ when $K_1(\theta)\cap K_2(\theta) = \emptyset$. Collecting these facts we find

$$\begin{aligned}
&\mathbb{E}[\chi_{123}(\theta)] \\
&= (1-q(\theta))^2 - (1-q(\theta))\,q(\theta) + q(\theta)^2 - q(\theta)r(\theta)
\end{aligned}$$

and the conclusion (16) follows by elementary algebra. ■

9.2 A proof of Proposition 2.2

By exchangeability and the binary nature of the rvs involved we readily obtain

$$\begin{aligned}
&\mathbb{E}\left[T_n(\theta)^2\right] \\
&= \mathbb{E}\left[T_n(\theta)\right] + \binom{n}{3}\binom{3}{2}\binom{n-3}{1}\mathbb{E}\left[\chi_{123}(\theta)\chi_{124}(\theta)\right] \\
&\quad + \binom{n}{3}\binom{3}{1}\binom{n-3}{2}\mathbb{E}\left[\chi_{123}(\theta)\chi_{145}(\theta)\right] \\
&\quad + \binom{n}{3}\binom{n-3}{3}\mathbb{E}\left[\chi_{123}(\theta)\chi_{456}(\theta)\right]. \qquad (64)
\end{aligned}$$

Under the enforced independence assumptions the rvs $\chi_{123}(\theta)$ and $\chi_{456}(\theta)$ are independent and identically distributed. As a result,

$$\mathbb{E}\left[\chi_{123}(\theta)\chi_{456}(\theta)\right] = \mathbb{E}\left[\chi_{123}(\theta)\right]\mathbb{E}\left[\chi_{456}(\theta)\right] = \beta(\theta)^2,$$

and using the relation (17) yields

$$\binom{n}{3}\binom{n-3}{3}\mathbb{E}\left[\chi_{123}(\theta)\chi_{456}(\theta)\right] = \frac{\binom{n-3}{3}}{\binom{n}{3}}\left(\mathbb{E}\left[T_n(\theta)\right]\right)^2. \quad (65)$$

On the other hand, with the help of (6) we readily check that the indicator rvs $\chi_{123}(\theta)$ and $\chi_{145}(\theta)$ are independent and identically distributed *conditionally* on $K_1(\theta)$ with

$$\mathbb{P}\left[\chi_{123}(\theta) = 1 | K_1(\theta)\right] = \mathbb{P}\left[\chi_{123}(\theta) = 1\right] = \beta(\theta). \qquad (66)$$

As a similar statement applies to $\chi_{145}(\theta)$, we conclude that the rvs $\chi_{123}(\theta)$ and $\chi_{145}(\theta)$ are (unconditionally) independent and identically distributed with

$$\mathbb{E}\left[\chi_{123}(\theta)\chi_{145}(\theta)\right] = \mathbb{E}\left[\chi_{123}(\theta)\right]\mathbb{E}\left[\chi_{145}(\theta)\right] = \beta(\theta)^2.$$

Again by virtue of (17), this last observation yields

$$\begin{aligned}
&\binom{n}{3}\binom{3}{1}\binom{n-3}{2}\mathbb{E}\left[\chi_{123}(\theta)\chi_{145}(\theta)\right] \\
&= 3\frac{\binom{n-3}{2}}{\binom{n}{3}} \cdot \left(\mathbb{E}\left[T_n(\theta)\right]\right)^2. \qquad (67)
\end{aligned}$$

Substituting (65) and (67) into (64) establishes Proposition 2.2. ∎

9.3 A proof of Proposition 3.2

Since $1 \le K_n \le {K_n}^2$ for all $n = 1, 2, \ldots$, the condition (21) implies both

$$\lim_{n\to\infty} \frac{1}{P_n} = 0 \quad \text{and} \quad \lim_{n\to\infty} \frac{K_n}{P_n} = 0. \qquad (68)$$

Therefore, $\lim_{n\to\infty} P_n = \infty$, and for any $c > 0$, we have $cK_n < P_n$ for all n sufficiently large in $\mathbb{N}_0$ (dependent on c). Thus, we have $3K_n < P_n$ for all n sufficiently large in $\mathbb{N}_0$. On that range we can use the expression (14) to write

$$\beta(\theta_n) = \left(1 - q(\theta_n)\right)^3 + q(\theta_n)^3\left(1 - \frac{r(\theta_n)}{q(\theta_n)^2}\right).$$

As Lemma 3.1 already implies $q(\theta_n)^3 \sim 1$ and $\left(1 - q(\theta_n)\right)^3 \sim \left(\frac{K_n^2}{P_n}\right)^3$, the asymptotic equivalence $\beta(\theta_n) \sim \tau(\theta_n)$ will be established if we show that

$$1 - \frac{r(\theta_n)}{q(\theta_n)^2} \sim \frac{K_n^3}{P_n^2}. \qquad (69)$$

This is an easy consequence of the fact that all terms involved are non-negative.

To establish (69) we proceed as follows: With positive integers K, P such that $3K \le P$, we note that

$$\begin{aligned}
&\frac{r(\theta)}{q(\theta)^2} \\
&= \left(\frac{(P-2K)!}{(P-K)!}\right)^2 \cdot \frac{(P-2K)!}{(P-3K)!} \cdot \frac{P!}{(P-K)!} \\
&= \frac{(P-2K)!(P-2K)!}{(P-K)!(P-3K)!} \cdot \frac{P!(P-2K)!}{(P-K)!(P-K)!} \\
&= \prod_{\ell=0}^{K-1}\left(\frac{P-2K-\ell}{P-K-\ell}\right) \cdot \prod_{\ell=0}^{K-1}\left(\frac{P-\ell}{P-K-\ell}\right) \\
&= \prod_{\ell=0}^{K-1}\left(1 - \frac{K}{P-K-\ell}\right) \cdot \prod_{\ell=0}^{K-1}\left(1 + \frac{K}{P-K-\ell}\right) \\
&= \prod_{\ell=0}^{K-1}\left(1 - \left(\frac{K}{P-K-\ell}\right)^2\right)
\end{aligned}$$

upon grouping factors appropriately. Elementary bounding arguments now yield the two bounds

$$1 - \left(1 - \left(\frac{K}{P-K}\right)^2\right)^K \le 1 - \frac{r(\theta)}{q(\theta)^2}$$

and

$$1 - \frac{r(\theta)}{q(\theta)^2} \le 1 - \left(1 - \left(\frac{K}{P-2K}\right)^2\right)^K.$$

Pick a scaling $P, K : \mathbb{N}_0 \to \mathbb{N}_0$ satisfying the equivalent conditions (21) and consider n sufficiently large in $\mathbb{N}_0$ so that $3K_n < P_n$. On that range, we replace θ by θ_n in the last chain of inequalities according to this scaling. A standard sandwich argument will yield the desired equivalence (69) if we show that

$$1 - \left(1 - \left(\frac{K_n}{P_n - cK_n}\right)^2\right)^{K_n} \sim \frac{K_n^3}{P_n^2}, \quad c = 1, 2. \qquad (70)$$

To do so we proceed as follows: Fix $c = 1, 2$. With

$$A_n(c) = \left(\frac{K_n}{P_n - cK_n}\right), \quad n = 1, 2, \ldots$$

standard calculus yields

$$\begin{aligned}
&1 - \left(1 - \left(\frac{K_n}{P_n - cK_n}\right)^2\right)^{K_n} \\
&= 1 - \left(1 - A_n(c)^2\right)^{K_n} \\
&= K_n A_n(c)^2 \int_0^1 \left(1 - A_n(c)^2 t\right)^{K_n - 1} dt \qquad (71)
\end{aligned}$$

on the appropriate range. The asymptotic equivalences

$$A_n(c)^2 = \left(\frac{K_n}{P_n - cK_n}\right)^2 \sim \left(\frac{K_n}{P_n}\right)^2 \qquad (72)$$

and

$$K_n A_n(c)^2 \sim \frac{K_n^3}{P_n^2} \tag{73}$$

follow from (68), so that (70) will hold if we show that

$$\lim_{n\to\infty} \int_0^1 \left(1 - A_n(c)^2 t\right)^{K_n - 1} dt = 1. \tag{74}$$

In view of (72) we conclude from (68) that for all n sufficiently large in $\mathbb{N}_0$ we have $\sup_{0\le t\le 1} \left|1 - A_n(c)^2 t\right| \le 1$. Therefore, the Bounded Convergence Theorem will yield (74) as soon as we establish

$$\lim_{n\to\infty} \left(1 - A_n(c)^2 t\right)^{K_n - 1} = 1, \quad 0 \le t \le 1. \tag{75}$$

To that end, recall the decomposition

$$\log(1-x) = -\int_0^x \frac{1}{1-t} dt = -x - \Psi(x) \tag{76}$$

where

$$\Psi(x) = \int_0^x \frac{t}{1-t} dt, \quad 0 \le x < 1.$$

It is easy to check that

$$\lim_{x\downarrow 0} \frac{\Psi(x)}{x} = 0. \tag{77}$$

Fix n sufficiently large in $\mathbb{N}_0$ as required above. For each t in the interval $(0, 1]$, with the help of (76) we can write

$$\begin{aligned} &\left(1 - A_n(c)^2 t\right)^{K_n - 1} \\ &\quad = e^{(K_n - 1)\log\left(1 - A_n(c)^2 t\right)} \\ &\quad = e^{-(K_n-1)A_n(c)^2 t - (K_n - 1)\Psi(A_n(c)^2 t)}. \end{aligned} \tag{78}$$

Returning to (73), we use (21) and (68) to find

$$\lim_{n\to\infty} K_n A_n(c)^2 = \lim_{n\to\infty} \left(\frac{K_n^2}{P_n} \cdot \frac{K_n}{P_n}\right) = 0.$$

It is then plain that $\lim_{n\to\infty}(K_n - 1)A_n(c)^2 = 0$, whence

$$\begin{aligned} &\lim_{n\to\infty} (K_n - 1)\Psi(A_n(c)^2 t) \\ &\quad = \lim_{n\to\infty} (K_n - 1) A_n(c)^2 t \cdot \frac{\Psi(A_n(c)^2 t)}{A_n(c)^2 t} = 0 \end{aligned}$$

with the help of (77) in the last step. Finally, letting n go to infinity in (78), we readily get (75) as desired. ∎

10 Proofs of the main results

10.1 A proof of Theorem 3.3

Consider a scaling $P, K : \mathbb{N}_0 \to \mathbb{N}_0$. For each $n = 3, 4, \ldots$, the elementary bound $\mathbb{P}[T_n(\theta_n) > 0] \le \mathbb{E}[T_n(\theta_n)]$ implies

$$\mathbb{P}[T_n(\theta_n) > 0] \le \binom{n}{3} \beta(\theta_n)$$

by virtue of Proposition 2.1. Theorem 3.3 thus follows if under (24) we show that $\lim_{n\to\infty} \binom{n}{3}\beta(\theta_n) = 0$. By Proposition 3.2 this convergence is equivalent to the assumed condition $\lim_{n\to\infty} n^3 \tau(\theta_n) = 0$, and the proof of Theorem 3.3 is now complete. ∎

10.2 A proof of Theorem 3.4

Assume first that $q^\star$ satisfies $0 \le q^\star < 1$. Fix $n = 3, 4, \ldots$ and partition the n nodes into the $k_n + 1$ non-overlapping groups $(1, 2, 3)$, $(4, 5, 6)$, …, $(3k_n + 1, 3k_n + 2, 3k_n + 3)$ with $k_n = \lfloor \frac{n-3}{3} \rfloor$. If $\mathbb{K}(n; \theta_n)$ contains no triangle, then *none* of these $k_n + 1$ groups of nodes forms a triangle. With this in mind we get

$$\begin{aligned} &\mathbb{P}[T_n(\theta_n) = 0] \\ &\quad \le \mathbb{P}\left[\bigcap_{\ell=0}^{k_n} \left[\begin{array}{c}\text{Nodes } 3\ell+1, 3\ell+2, 3\ell+3 \text{ do not} \\ \text{form a triangle in } \mathbb{K}(n;\theta_n)\end{array}\right]\right] \\ &\quad = \prod_{\ell=0}^{k_n} \mathbb{P}\left[\begin{array}{c}\text{Nodes } 3\ell+1, 3\ell+2, 3\ell+3 \text{ do not} \\ \text{form a triangle in } \mathbb{K}(n;\theta_n)\end{array}\right] \qquad (79) \\ &\quad = \left(1 - \beta(\theta_n)\right)^{k_n+1} \\ &\quad \le \left(1 - (1 - q(\theta_n))^3\right)^{k_n+1} \qquad (80) \\ &\quad \le e^{-(k_n+1)(1-q(\theta_n))^3}. \qquad (81) \end{aligned}$$

Note that (79) follows from the fact that the events

$$\left[\begin{array}{c}\text{Nodes } 3\ell+1, 3\ell+2, 3\ell+3 \text{ do not} \\ \text{form a triangle in } \mathbb{K}(n;\theta_n)\end{array}\right], \ \ell = 0, \ldots, k_n$$

are mutually independent due to the non-overlap condition, while the inequality (80) is justified with the help of (19). Let n go to infinity in the inequality (81). From the constraint $q^\star < 1$ we conclude that $\lim_{n\to\infty} \mathbb{P}[T_n(\theta_n) = 0] = 0$ since $k_n \sim \frac{n}{3}$ so that $\lim_{n\to\infty}(k_n + 1)(1 - q(\theta_n))^3 = \infty$. This establishes the one law in the case $q^\star < 1$.

To handle the case $q^\star = 1$, we use a standard bound which forms the basis of the method of second moment [15, Remark 3.1, p. 55]. Here this bound takes the form

$$\frac{\left(\mathbb{E}[T_n(\theta_n)]\right)^2}{\mathbb{E}[T_n(\theta_n)^2]} \le \mathbb{P}[T_n(\theta_n) > 0], \quad n = 3, 4, \ldots \tag{82}$$

Theorem 3.4 then will be established in the case $q^\star = 1$ if we show under (21) that the condition (25) implies

$$\lim_{n\to\infty} \frac{\mathbb{E}[T_n(\theta_n)^2]}{\left(\mathbb{E}[T_n(\theta_n)]\right)^2} = 1. \tag{83}$$

As pointed earlier, the conditions (21) imply $3K_n < P_n$ for all n sufficiently large in $\mathbb{N}_0$. On that range, with θ replaced by θ_n, Proposition 2.2 yields

$$\begin{aligned} \frac{\mathbb{E}[T_n(\theta_n)^2]}{\left(\mathbb{E}[T_n(\theta_n)]\right)^2} &= \frac{1}{\mathbb{E}[T_n(\theta_n)]} + \left(\frac{\binom{n-3}{3}}{\binom{n}{3}} + 3\frac{\binom{n-3}{2}}{\binom{n}{3}}\right) \\ &\quad + \frac{3(n-3)}{\binom{n}{3}} \cdot \frac{\mathbb{E}[\chi_{123}(\theta_n)\chi_{124}(\theta_n)]}{\left(\mathbb{E}[\chi_{123}(\theta_n)]\right)^2} \end{aligned}$$

as we make use of (17) in the last term.

Let n go to infinity in the resulting expression: Under condition (25), we have $\lim_{n\to\infty} n^3\beta(\theta_n) = \infty$ by Proposition 3.2, whence $\lim_{n\to\infty} \mathbb{E}[T_n(\theta_n)] = \infty$ by virtue of (17). Since

$$\lim_{n\to\infty} \left(\frac{\binom{n-3}{3}}{\binom{n}{3}} + 3\frac{\binom{n-3}{2}}{\binom{n}{3}}\right) = 1 \tag{84}$$

and

$$\frac{\binom{n}{3}}{3(n-3)} \sim \frac{n^2}{18}, \tag{85}$$

the convergence (83) will hold if we show that

$$\lim_{n\to\infty} \frac{1}{n^2} \frac{\mathbb{E}\left[\chi_{123}(\theta_n)\chi_{124}(\theta_n)\right]}{\left(\mathbb{E}\left[\chi_{123}(\theta_n)\right]\right)^2} = 0 \tag{86}$$

under the foregoing conditions on the scaling.

This is shown as follows: Given positive integers K and P such that $K \leq P$, fix $n = 3, 4, \ldots$. It is immediate that

$$\begin{aligned} &\mathbb{E}\left[\chi_{123}(\theta)\chi_{124}(\theta)\right] \\ &\quad \leq \quad \mathbb{E}\left[\chi_{123}(\theta)\mathbf{1}\left[K_1(\theta) \cap K_4(\theta) \neq \emptyset\right]\right]. \end{aligned} \tag{87}$$

From (6) it follows that the rvs $\chi_{123}(\theta)$ and $\mathbf{1}\left[K_1(\theta) \cap K_4(\theta) \neq \emptyset\right]$ are independent conditionally on $K_1(\theta)$, and an easy conditioning argument yields

$$\mathbb{E}\left[\chi_{123}(\theta)\mathbf{1}\left[K_1(\theta) \cap K_4(\theta) \neq \emptyset\right]\right] = \beta(\theta)(1 - q(\theta)) \tag{88}$$

as we recall (4) and (16). Using (87) together with (16) and (88) we readily obtain the inequalities

$$\frac{\mathbb{E}\left[\chi_{123}(\theta)\chi_{124}(\theta)\right]}{\left(\mathbb{E}\left[\chi_{123}(\theta)\right]\right)^2} \leq \frac{\beta(\theta)(1 - q(\theta))}{\beta(\theta)^2} \leq \beta(\theta)^{-2/3} \tag{89}$$

where in the last step we noted that $1 - q(\theta) \leq \beta(\theta)^{1/3}$ by appealing to (19).

Returning to the convergence (86) we see from (89) that we need only show

$$\lim_{n\to\infty} n^2 \beta(\theta_n)^{2/3} = \infty. \tag{90}$$

As Proposition 3.2 yields $n^2\beta(\theta_n)^{2/3} \sim n^2\tau(\theta_n)^{2/3} = \left(n^3\tau(\theta_n)\right)^{2/3}$, the desired conclusion (90) follows under the condition (25). ∎

10.3 A proof of Theorem 3.7

Throughout P and K are positive integers such that $K \leq P$, and fix $n = 3, 4, \ldots$. For each $i = 1, \ldots, n$, we introduce the index set

$$\mathcal{P}_{n,i} = \{(j,k) : \ 1 \leq j < k \leq n, \ j \neq i, \ k \neq i\}. \tag{91}$$

Next, define the count rvs $T_{n,i}(\theta)$ and $T^\star_{n,i}(\theta)$ by

$$T_{n,i}(\theta) = \sum_{(j,k)\in\mathcal{P}_{n,i}} \xi_{ij}(\theta)\xi_{ik}(\theta)\xi_{jk}(\theta)$$

and

$$T^\star_{n,i}(\theta) = \sum_{(j,k)\in\mathcal{P}_{n,i}} \xi_{ij}(\theta)\xi_{ik}(\theta).$$

The rv $T_{n,i}(\theta)$ counts the number of distinct triangles in $\mathbb{K}(n;\theta)$ which have node i as a vertex, while $T^\star_{n,i}(\theta)$ counts the number of (unordered) distinct pairs of nodes which are both connected to node i in $\mathbb{K}(n;\theta)$. The rv $D_{n,i}(\theta)$ is the degree of node i in $\mathbb{K}(n;\theta)$ and is given by

$$D_{n,i}(\theta) = \sum_{k=1, \ k\neq i}^{n} \xi_{ik}(\theta).$$

We have

$$\sum_{i=1}^{n} T_{n,i}(\theta) = 3T_n(\theta)$$

while

$$D_{n,i}(\theta)\left(D_{n,i}(\theta) - 1\right) = 2T^\star_{n,i}(\theta).$$

Under the condition

$$\sum_{i=1}^{n} D_{n,i}(\theta)\left(D_{n,i}(\theta) - 1\right) > 0,$$

the definition of $C^\star(\mathbb{K}(n;\theta))$ yields

$$\begin{aligned} C^\star(\mathbb{K}(n;\theta)) &= \frac{\sum_{i=1}^n T_{n,i}(\theta)}{\frac{1}{2}\sum_{i=1}^n D_{n,i}(\theta)\left(D_{n,i}(\theta) - 1\right)} \\ &= \frac{\sum_{i=1}^n T_{n,i}(\theta)}{\sum_{i=1}^n T^\star_{n,i}(\theta)} \end{aligned}$$

so that

$$C^\star(\mathbb{K}(n;\theta)) = \frac{3T_n(\theta)}{\sum_{i=1}^n T^\star_{n,i}(\theta)} \mathbf{1}\left[\sum_{i=1}^{n} T^\star_{n,i}(\theta) > 0\right]. \tag{92}$$

The desired conclusion (30) is now immediate from Lemma 10.1 and Lemma 10.2 established below. They deal with the a.s. convergence of the numerator and denominator (properly normalized) appearing in the ratio (92), respectively.

Lemma 10.1. *For positive integers P and K such that $K \leq P$, we have*

$$\lim_{n\to\infty} \frac{T_n(\theta)}{\binom{n}{3}} = \beta(\theta) \quad a.s. \tag{93}$$

Proof. Fix $n = 3, 4, \ldots$ and $\varepsilon > 0$. Markov's inequality already gives

$$\mathbb{P}\left[\left|\frac{T_n(\theta)}{\binom{n}{3}} - \beta(\theta)\right| > \varepsilon\right] \leq \varepsilon^{-2}\mathrm{Var}\left[\frac{T_n(\theta)}{\binom{n}{3}}\right]$$

as we recall (17). It is now plain from (20) that

$$\begin{aligned} &\mathrm{Var}\left[\frac{T_n(\theta)}{\binom{n}{3}}\right] \\ &= \mathbb{E}\left[\left(\frac{T_n(\theta)}{\binom{n}{3}}\right)^2\right] - \left(\frac{\mathbb{E}\left[T_n(\theta)\right]}{\binom{n}{3}}\right)^2 \\ &= \frac{\mathbb{E}\left[T_n(\theta)\right]}{\binom{n}{3}^2} + \left(\frac{\binom{n-3}{3}}{\binom{n}{3}} + 3\frac{\binom{n-3}{2}}{\binom{n}{3}} - 1\right) \cdot \left(\frac{\mathbb{E}\left[T_n(\theta)\right]}{\binom{n}{3}}\right)^2 \\ &\quad + 3(n-3)\binom{n}{3} \cdot \frac{\mathbb{E}\left[\chi_{123}(\theta)\chi_{124}(\theta)\right]}{\binom{n}{3}^2} \\ &= \frac{\beta(\theta)}{\binom{n}{3}} + \left(\frac{\binom{n-3}{3}}{\binom{n}{3}} + 3\frac{\binom{n-3}{2}}{\binom{n}{3}} - 1\right) \cdot \beta(\theta)^2 \\ &\quad + \frac{3(n-3)}{\binom{n}{3}} \cdot \mathbb{E}\left[\chi_{123}(\theta)\chi_{124}(\theta)\right] \end{aligned} \tag{94}$$

as we again make use of the expression (17).

With the help of (84) and (85), it is easy to see that

$$\lim_{n\to\infty} \mathrm{Var}\left[\frac{T_n(\theta)}{\binom{n}{3}}\right] = 0, \tag{95}$$

a fact which would readily imply a weaker form of (93) with a.s. convergence replaced by convergence in probability.

However, elementary algebra on (94) shows that (95) takes place according to

$$\lim_{n\to\infty} n^2 \mathrm{Var}\left[\frac{T_n(\theta)}{\binom{n}{3}}\right] = C$$

with

$$C = 18\left(\mathbb{E}\left[\chi_{123}(\theta)\chi_{124}(\theta)\right] - \beta(\theta)^2\right) > 0.$$

As a result, for every $\varepsilon > 0$, we have

$$\sum_{n=3}^{\infty} \mathbb{P}\left[\left|\frac{T_n(\theta)}{\binom{n}{3}} - \beta(\theta)\right| > \varepsilon\right] \leq \frac{C'}{\varepsilon^2}\sum_{n=3}^{\infty} n^{-2} < \infty$$

for some $C' > C$, and the conclusion (93) follows by the Borel-Cantelli Lemma. ∎

Lemma 10.2. *For positive integers P and K such that $K \leq P$, we have*

$$\lim_{n\to\infty} \frac{\sum_{i=1}^n T^\star_{n,i}(\theta)}{\binom{n}{3}} = 3p(\theta)^2 \quad a.s. \tag{96}$$

Proof. Fix $n = 3, 4, \ldots$. Note that

$$\begin{aligned} T^\star_{n,1}(\theta) &= \sum_{j=2}^{n-1}\sum_{k=j+1}^{n} \xi_{1j}(\theta)\xi_{1k}(\theta) \\ &= \Phi_n(\xi_{12}(\theta), \ldots, \xi_{1n}(\theta)) \end{aligned} \tag{97}$$

where the mapping $\Phi_n : [0,1]^{n-1} \to \mathbb{R}_+$ is given by

$$\begin{aligned} \Phi_n(x_2, \ldots, x_n) &= \sum_{\ell=2}^{n-1}\sum_{k=\ell+1}^{n} x_\ell x_k \\ &= \sum_{\ell=2}^{n-1} x_\ell\left(\sum_{k=\ell+1}^{n} x_k\right) \end{aligned} \tag{98}$$

with $(x_2, \ldots, x_n)$ arbitrary in $[0,1]^{n-1}$. For each $j = 2, \ldots, n$, consider pairs of elements $(x_2, \ldots, x_n)$ and $(y_2, \ldots, y_n)$ in $[0,1]^{n-1}$ which differ only in the j^{th} component, i.e.,

$$x_\ell = y_\ell, \quad \ell \neq j, \quad \ell = 2, \ldots, n.$$

Under such conditions, it is easy to check that

$$\begin{aligned} &|\Phi_n(x_2, \ldots, x_n) - \Phi_n(y_2, \ldots, y_n)| \\ &\leq |x_j - y_j| \cdot \sum_{\ell=2,\ \ell\neq j}^{n-1} x_\ell \\ &\leq n - 1. \end{aligned} \tag{99}$$

Recall that the $(n-1)$ rvs $\{\xi_{1j}(\theta),\ j = 2, \ldots, n\}$ are i.i.d. Bernoulli rvs. In view of the constraints (99) we can now apply McDiarmid's inequality [17] (with $c_j = (n-1)$ for all $j = 2, \ldots, n-1$); see also Corollary 2.17 and Remark 2.28 in the monograph [15, p. 38]. Thus, for every $t > 0$ we find

$$\mathbb{P}\left[\left|T^\star_{n,1}(\theta) - \mathbb{E}\left[T^\star_{n,1}(\theta)\right]\right| > t\right] \leq 2e^{-\frac{2t^2}{(n-1)^3}} \tag{100}$$

with

$$\begin{aligned} \mathbb{E}\left[T^\star_{n,1}(\theta)\right] &= \sum_{j=2}^{n-1}\sum_{k=j+1}^{n} \mathbb{E}\left[\xi_{1j}(\theta)\xi_{1k}(\theta)\right] \\ &= \sum_{j=2}^{n-1}(n-j)p(\theta)^2 \\ &= \frac{(n-1)(n-2)}{2}\cdot p(\theta)^2 \end{aligned} \tag{101}$$

under the independence noted earlier.

With $\varepsilon > 0$ we now substitute

$$t = \frac{(n-1)(n-2)}{2}\varepsilon \tag{102}$$

into (100). Since

$$\frac{2t^2}{(n-1)^3} = \frac{(n-2)^2}{2(n-1)}\cdot\varepsilon^2 \sim \frac{n}{2}\cdot\varepsilon^2, \tag{103}$$

we obtain from (100) and (101) that

$$\mathbb{P}\left[\left|\frac{T^\star_{n,1}(\theta)}{\frac{(n-1)(n-2)}{2}} - p(\theta)^2\right| > \varepsilon\right] \leq 2e^{-\frac{n}{2}(1+o(1))\varepsilon^2}. \tag{104}$$

Since

$$\begin{aligned} \left|\frac{\sum_{i=1}^n T^\star_{n,i}(\theta)}{\frac{n(n-1)(n-2)}{2}} - p(\theta)^2\right| &= \left|\frac{1}{n}\sum_{i=1}^{n}\left(\frac{T^\star_{n,i}(\theta)}{\frac{(n-1)(n-2)}{2}} - p(\theta)^2\right)\right| \\ &\leq \frac{1}{n}\sum_{i=1}^{n}\left|\frac{T^\star_{n,i}(\theta)}{\frac{(n-1)(n-2)}{2}} - p(\theta)^2\right|, \end{aligned}$$

it is plain that

$$\begin{aligned} &\mathbb{P}\left[\left|\frac{\sum_{i=1}^n T^\star_{n,i}(\theta)}{\frac{n(n-1)(n-2)}{2}} - p(\theta)^2\right| > \varepsilon\right] \\ &\leq \mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n}\left|\frac{T^\star_{n,i}(\theta)}{\frac{(n-1)(n-2)}{2}} - p(\theta)^2\right| > \varepsilon\right] \\ &\leq \mathbb{P}\left[\bigcup_{i=1}^{n}\left[\left|\frac{T^\star_{n,i}(\theta)}{\frac{(n-1)(n-2)}{2}} - p(\theta)^2\right| > \varepsilon\right]\right] \\ &\leq \sum_{i=1}^{n}\mathbb{P}\left[\left|\frac{T^\star_{n,i}(\theta)}{\frac{(n-1)(n-2)}{2}} - p(\theta)^2\right| > \varepsilon\right] \\ &= n\mathbb{P}\left[\left|\frac{T^\star_{n,1}(\theta)}{\frac{(n-1)(n-2)}{2}} - p(\theta)^2\right| > \varepsilon\right] \end{aligned} \tag{105}$$

where the last inequality follows by a union bound argument and (105) is a consequence of exchangeability.

Invoking (104) (with $\frac{\varepsilon}{3}$ instead of ε) we get

$$\mathbb{P}\left[\left|\frac{\sum_{i=1}^n T^\star_{n,i}(\theta)}{\binom{n}{3}} - 3p(\theta)^2\right| > \varepsilon\right] \leq 2ne^{-\frac{n}{18}(1+o(1))\varepsilon^2}$$

with

$$\sum_{n=3}^{\infty} ne^{-\frac{n}{18}(1+o(1))\varepsilon^2} < \infty$$

for every $\varepsilon > 0$. The a.s. convergence (96) now follows by the Borel-Cantelli Lemma. ∎

REFERENCES

[1] D. J. Aldous, "Exchangeability and related topics," *École d'Été de Probabilités de Saint-Flour XIII 1983, Lecture Notes in Math.* **1117**, pp. 1198, Springer, Berlin, 1985, doi:10.1007/BFb0099421

[2] F. G. Ball, D. J. Sirl and P. Trapman, "Epidemics on random intersection graphs," *The Annals of Applied Probability* **24** (2014), pp. 1081-1128.

[3] S.R. Blackburn and S. Gerke, "Connectivity of the uniform random intersection graph," *Discrete Mathematics* **309** (2009), pp. 5130-5140.

[4] B. Bollobás, *Random Graphs*, Second Edition, Cambridge Studies in Advanced Mathematics, Cambridge University Press, Cambridge (UK), 2001.

[5] F. Chung and L. Lu, "The diameter of sparse random graphs," *Advances in Applied Mathematics* **26** (2001), pp. 257-279.

[6] M. Deijfen and W. Kets, "Random intersection graphs with tunable degree distribution and clustering," *Probability in the Engineering and Informational Sciences* **23** (2009), pp. 661-674.

[7] R. Di Pietro, L.V. Mancini, A. Mei, A. Panconesi and J. Radhakrishnan, "Redoubtable sensor networks," *ACM Transactions on Information Systems Security* **TISSEC 11** (2008), pp. 1-22.

[8] P. Erdős and A. Rényi, "On the evolution of random graphs," *Publ. Math. Inst. Hung. Acad. Sci.* **5** (1960), pp. 17-61.

[9] L. Eschenauer and V.D. Gligor, "A key-management scheme for distributed sensor networks," in Proceedings of the ACM Conference on Computer and Communications Security (2002), Washington (DC), November 2002.

[10] J. P. Gleeson, S. Melnik and A. Hackett, A, "How clustering affects the bond percolation threshold in complex networks," *Physical Review E*, 066114, 2010.

[11] E. Godehardt and J. Jaworski, "Two models of random intersection graphs for classification," in *Studies in Classification, Data Analysis and Knowledge Organization* **22**, Eds. O. Optiz and M. Schwaiger, Springer, Berlin (2003), pp. 67-82.

[12] E. Godehardt, J. Jaworski and K. Rybarczyk, "Random intersection graphs and classification," in *Studies in Classification, Data Analysis and Knowledge Organization* **33**, Eds. H.J. Lens and R,. Decker, Springer, Berlin (2007), pp. 67-74.

[13] A. Hackett, S. Melnik and J. P. Gleeson, "Cascades on a class of clustered random networks," *Physical Review E*, 056107, 2011.

[14] X. Huang, S. Shao, H. Wang, S. V. Buldyrev, H. E. Stanley and S. Havlin, "The robustness of interdependent clustered networks," *Europhysics Letters* **101** (2013), 18002.

[15] S. Janson, T. Łuczak and A. Ruciński, *Random Graphs*, Wiley-Interscience Series in Discrete Mathematics and Optimization, John Wiley & Sons, 2000.

[16] M.K. Karoński, E.R. Scheinerman and K.B. Singer-Cohen, "On random intersection graphs: The subgraph problem," *Combinatorics, Probability and Computing* **8** (1999), pp. 131-159.

[17] C. McDiarmid, "On the method of bounded differences," in *Surveys in Combinatorics*, Cambridge University Press, Cambridge (UK), 1989. pp. 148-188.

[18] P. Marbach, "A lower-bound on the number of rankings required in recommender systems using collaborative filtering," in Proceedings of the 42nd Annual Conference on Information Sciences and Systems (CISS 2008), Princeton University, Princeton (NJ), March 2008.

[19] S. Milgram, "The small world problem," *Psychology Today* **2** (1967), pp. 60-67.

[20] J. C. Miller, "Percolation and epidemics in random clustered networks," *Physical Review E*, 020901, 2009.

[21] M. E. J. Newman, "Random Graphs with Clustering," *Phys. Rev. Lett.*, 058701, 2009.

[22] M. E. J. Newman, "The structure and function of complex networks," *SIAM Review* **45** (2003), pp. 167-256.

[23] M.D. Penrose, *Random Geometric Graphs*, Oxford Studies in Probability **5**, Oxford University Press, New York (NY), 2003.

[24] K. Rybarczyk, "Diameter, connectivity, and phase transition of the uniform random intersection graph," *Discrete Mathematics* **311** (2011), pp. 1998-2019.

[25] M. A. Serrano and M. Boguna, "Clustering in complex networks. i. General formalism, *Physical Review E* **74** (2006), p. 056114.

[26] K.B. Singer, *Random Intersection Graphs*, Ph.D. Thesis, The Johns Hopkins University, Baltimore (MD), 1995.

[27] D. Watts and S. Strogatz, "Collective dynamics of "small-world" networks," *Nature* **393** (1998), pp. 440-442.

[28] O. Yağan and A. M. Makowski, "Connectivity results for random key graphs," in Proceedings of the IEEE International Symposium on Information Theory (ISIT 2009), Seoul (Korea), June 2009.

[29] O. Yağan and A. M. Makowski, "On the existence of triangles in random key graphs," in Proceedings of the 47th Annual Allerton Conference on Communication, Control and Computing, Monticello (IL), September 2009.

[30] O. Yağan and A. M. Makowski, "Random key graphs – Can they be small worlds?," in Proceedings of the First Workshop on Applications of Graph Theory in Wireless Ad hoc Networks and Sensor Networks (GRAPH-HOC 2009), Chennai (India), December 2009.

[31] O. Yağan and A. M. Makowski, "Connectivity in random graphs induced by a key predistribution scheme: Small key pools," in Proceedings of the 44th Annual Conference on Information Sciences and Systems (CISS 2010), March 2010.

[32] O. Yağan and A.M. Makowski, "Zero-one laws for connectivity in random key graphs," *IEEE Transactions on Information Theory* **IT-58** (2012), pp. 2983-2999.

[33] O. Yağan and A.M. Makowski, "Counting triangles, tunable clustering and the small-world property in random key graphs (Extended version)," available in Arxiv.

[34] O. Yağan, "Zero-one laws for connectivity in inhomogeneous random key graphs," *IEEE Transactions on Information Theory*, **IT-62** (2016), pp. 4559-4574, August 2016.

[35] J. Zhao, O. Yağan and V. Gligor, "On the strengths of connectivity and robustness in general random intersection graphs," in Proceedings of the 53rd IEEE Conference on Decision and Control (CDC 2014), Los Angeles (CA), December 2014, pp. 3661-3668.

[36] J. Zhao, O. Yağan and V. Gligor, "k-connectivity in random key graphs with unreliable edges," *IEEE Transactions on Information Theory*, **IT-61** (2015), pp. 3810-3836.

[37] Y. Zhuang and O. Yağan, "Information Propagation in Clustered Multilayer Networks," in *IEEE Transactions on Network Science and Engineering*, vol. 3, no. 4, pp. 211-224, Oct.-Dec. 1 2016, doi: 10.1109/TNSE.2016.2600059.

[38] Y. Zhuang, A. Arenas, and O. Yağan, "Clustering determines the dynamics of complex contagions in multiplex networks," arXiv preprint arXiv:1608.08237 (2016).

Osman Yağan (S'07-M'12) received the B.S. degree in Electrical and Electronics Engineering from the Middle East Technical University, Ankara (Turkey) in 2007, and the Ph.D. degree in Electrical and Computer Engineering from the University of Maryland, College Park, MD in 2011. He is an Assistant Research Professor of Electrical and Computer Engineering (ECE) at Carnegie Mellon University (CMU). Prior to joining the faculty of the ECE department in August 2013, he was a Postdoctoral Research Fellow in CyLab at CMU. He has also held a visiting Postdoctoral Scholar position at Arizona State University during Fall 2011. His research interests include wireless communication networks, security, social and information networks, and cyber-physical systems.

Armand M. Makowski (M'83-SM'94-F'06) received the Licence en Sciences Mathématiques from the Université Libre de Bruxelles in 1975, the M.S. degree in Engineering-Systems Science from U.C.L.A. in 1976 and the Ph.D. degree in Applied Mathematics from the University of Kentucky in 1981. In August 1981, he joined the faculty of the Electrical Engineering Department at the University of Maryland College Park, where he is Professor of Electrical and Computer Engineering. He has held a joint appointment with the Institute for Systems Research since its establishment in 1985. Armand Makowski was a C.R.B. Fellow of the Belgian-American Educational Foundation (BAEF) for the academic year 1975-76; he is also a 1984 recipient of the NSF Presidential Young Investigator Award and became an IEEE Fellow in 2006. His research interests lie in applying advanced methods from the theory of stochastic processes to the modeling, design and performance evaluation of engineering systems, with particular emphasis on communication systems and networks.