Random key graphs can have many more triangles than Erdős-Rényi graphs

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

Abstract—Random key graphs are random graphs induced by the random key predistribution scheme of Eschenauer and Gligor (EG) under the assumption of full visibility. For this class of random graphs we establish a zero-one law for the existence of triangles, and identify the corresponding critical scaling. This zero-one law exhibits significant differences with the corresponding result in Erdős-Rényi graphs. We also compute the clustering coefficient of random key graphs, and compare them with that of Erdős-Rényi graphs in the many node regime when the expected average degrees are asymptotically equivalent. On the parameter range of practical relevance in wireless sensor networks (WSNs), random key graphs are found to be much more clustered than the corresponding Erdős-Rényi graphs. These results clearly show the inadequacy of random key graphs as small worlds in the sense of Watts and Strogatz is also discussed.

Keywords: Wireless sensor networks, Key predistribution, Random graphs, Existence of triangles, Clustering coefficient.

I. INTRODUCTION

Random key graphs are random graphs that belong to the class of random intersection graphs [19]; in fact they are also called uniform random intersection graphs by some authors [1], [8], [9]. They have appeared recently in application areas as diverse as clustering analysis [8], [9], collaborative filtering in recommender systems [13] and random key predistribution for wireless sensor networks (WSNs) [7]. In this last context, random key graphs naturally occur in the study of the following random key predistribution scheme introduced by Eschenauer and Gligor [7]: Before deployment, each sensor in a WSN is independently assigned $K$ distinct cryptographic keys which are selected at random from a very large pool of $P$ keys. These $K$ keys constitute the key ring of the node and are inserted into its memory module. Two sensor nodes can then establish a secure link between them if they are within transmission range of each other and if their key rings have at least one key in common; see [7] for implementation details. If we assume full visibility, namely 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 II 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 [6], 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 [7], has now been validated rigorously; see the papers [1], [5], [17], [21], [25] for recent developments. Furthermore, Rybarczyk [17] 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 [2], [6], [10], [11], [16], [19] in the containment of certain (small) subgraphs, the simplest one being the triangle. This last case is of some practical relevance: The number of triangles in a graph is closely related to its clustering properties, and for random key graphs this has implications for network resiliency under the EG scheme; see the paper by Di Pietro et al. [5]. With this in mind, in the present paper we study the zero-one law for the existence of triangles in random key graphs and identify the corresponding critical scaling.

On the way to this result, 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, this expected number of triangles can be orders of magnitude larger in random key graphs than in Erdős-Rényi graphs, a fact also observed earlier on via simulations by Di Pietro et al. [5].

This work was supported in part by NSF Grant CCF-0729093 and 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 was with the Department of Electrical and Computer Engineering, and the Institute for Systems Research, University of Maryland, College Park, MD 20742 USA. He is now with Department of Electrical and Computer Engineering and CyLab, Carnegie Mellon University, Moffett Field, CA 94035 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).
As a result, transferring results from Erdős-Rényi graphs by matching their edge probabilities is not a valid approach in general, and can be quite misleading in the context of WSNs.

Next, we compare the clustering coefficients of the two classes of random graphs. 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 practically relevant for large 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 [17], we then conclude that random key graphs are small-worlds in the sense introduced by Watts and Strogatz [20]. This reinforces the possibility of using random key graphs in a wide range of applications including modeling social networks.

Our results make it clear that the asymptotic equivalence of random key graphs and Erdős-Rényi graphs (in the sense discussed in [19]) is possible only when the size of key rings is comparable to the network size, a case not very realistic in WSNs where sensors have limited memory and computational capabilities. This points to the inadequacy of Erdős-Rényi graphs to capture some key properties of the EG scheme in WSNs where sensors have limited memory and computational capabilities. This reinforces the possibility of using random key graphs and Erdős-Rényi graphs in terms of their number of random key graphs are asymptotically equivalent. Recalling the fact that random key graphs in the sense introduced by Watts and Strogatz [20], this reinforces the possibility of using random key graphs and Erdős-Rényi graphs (in the sense introduced by Watts and Strogatz [20]) is possible only when the size of key rings is comparable to the network size, a case not very realistic in WSNs where sensors have limited memory and computational capabilities.

For each node \( i = 1, \ldots, n \), let \( K_i(\theta) \) denote the random set of keys assigned to node \( i \). Thus, under the convention that the \( P \) keys are labelled \( 1, \ldots, P \), 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. rvs, each of which is uniformly distributed with

\[
\mathbb{P}[K_i(\theta) = S] = \left( \frac{P}{K} \right)^{-1}, \quad i = 1, \ldots, n
\]

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

\[
K_i(\theta) \cap K_j(\theta) \neq \emptyset,
\]

in which case an undirected link is assigned between nodes \( i \) and \( j \). We find it convenient to introduce the event \( E_{ij}(\theta) \) where (2) takes place, i.e.,

\[
E_{ij}(\theta) = [K_i(\theta) \cap K_j(\theta) \neq \emptyset],
\]

with indicator function

\[
\xi_{ij}(\theta) = 1[E_{ij}(\theta)].
\]

The adjacency constraints (2) define a random graph on the vertex set \( \{1, \ldots, n\} \), hereafter denoted \( \mathbb{K}(n; \theta) \). We refer to it as the random key graph.

It is easy to check that

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

with

\[
q(\theta) = \begin{cases} 
0 & \text{if } P < 2K \\
\frac{(P-K)}{(K)} & \text{if } 2K \leq P.
\end{cases}
\]

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

\[
p(\theta) = 1 - q(\theta).
\]

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 \( 2K \leq P \).

A. Random key graphs

The model 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 \( i \). The results regarding the clustering coefficient of random key graphs were presented in the conference paper [23]. In line with results currently available for other classes of graphs, e.g., Erdős-Rényi graphs [10, Chap. 3] and random geometric graphs [16, 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.

The paper is organized as follows: In Section II we formally introduce the class of random key graphs and present some useful preliminaries. Our first step is to count triangles; first and second moments are computed in Section III (with proofs given in Sections IX and Section X). The main result of the paper is the zero-one law for the containment of triangles in random key graphs; it is presented in Section IV. The clustering coefficients are then discussed in Section V. Section VI contains well-known facts concerning Erdős-Rényi graphs; it also explains how to match random key graphs and Erdős-Rényi graphs, either exactly or asymptotically. Section VII and Section VIII are devoted to comparing matched random key graphs and Erdős-Rényi graphs in terms of their number of triangles and clustering coefficients, respectively. We prove the one-law (Theorem 4.4) in Section XI. The two appendices A and B contain proofs of various technical facts needed in the discussion.

II. MODEL AND PRELIMINARIES

First, a word on the notation and conventions in use: 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 space \( (\Omega, \mathcal{F}, \mathbb{P}) \). Probabilistic statements are made with respect to this probability measure \( \mathbb{P} \), and we denote the corresponding expectation operator by \( \mathbb{E} \). The indicator function of an event \( E \) is denoted by \( 1[E] \). For any discrete set \( S \) we write \( |S| \) for its cardinality.

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 \( \theta = (K, P) \) in order to lighten the notation.
B. Some useful facts

The expression (4) is a consequence of the fact
\[ P[S \cap K_i(\theta) = 0] = \frac{(P - |S|)}{(K_i(\theta))}, \quad i = 1, \ldots, n \] (6)
valid for every subset \( S \) of \{1, \ldots, P\} with \(|S| \leq P - K\). For each \( i = 1, \ldots, n \), it is a simple matter to check with the help of (6) that the events
\[ \{ E_{ij}(\theta), \quad j \neq i \} \]
are mutually independent, or equivalently, that the rvs \( i \) form a collection of i.i.d. rvs.

Some of the results will be conveniently stated in terms of the quantity
\[ \tau(\theta) = \frac{K^3}{P^2} + \left( \frac{K^2}{P} \right)^3, \quad \theta = (K, P) \] (7)

III. 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
\[ \chi_{ijk}(\theta) \]
(8)
The number of distinct triangles in \( \mathbb{K}(n; \theta) \) is then simply given by
\[ T_n(\theta) = \sum_{(ijk)} \chi_{ijk}(\theta) \] (9)
where \( \sum_{(ijk)} \) denotes summation over all distinct triples \( ijk \) with \( 1 \leq i < j < k \leq n \). Of particular interest to us will be the event that there exists at least one triangle in \( \mathbb{K}(n; \theta) \), namely
\[ [T_n(\theta) > 0] = [T_n(\theta) = 0]^c. \] (10)

Key to the analysis presented here is our ability to evaluate the first two moments of the count variables (9). The first moment, computed next, will be conveniently expressed with the help of the quantity \( \beta(\theta) \) given by
\[ \beta(\theta) = (1 - q(\theta))^3 + q(\theta)^3 - q(\theta)r(\theta) \] (11)
with \( r(\theta) \) defined by
\[ r(\theta) = \begin{cases} 
0 & \text{if } P < 3K \\
\frac{(P - 3K)}{(K)} & \text{if } 3K \leq P.
\end{cases} \] (12)

Note that \( r(\theta) \) corresponds to the probability (6) when \(|S| = 2K\). The next result is established in Section IX.

**Proposition 3.1:** Fix \( n = 3, 4, \ldots \). For positive integers \( K \) and \( P \) such that \( K \leq P \), we have
\[ E[\chi_{123}(\theta)] = \beta(\theta) \] (13)
with \( \beta(\theta) \) defined at (11), so that
\[ E[T_n(\theta)] = \left( \frac{n}{3} \right)^3 \beta(\theta). \] (14)

Thus, \( \beta(\theta) \) is simply the probability that three distinct vertices form a triangle in \( \mathbb{K}(n; \theta) \). The second moment of the count variables (9) can also be evaluated.

**Proposition 3.2:** For positive integers \( K \) and \( P \) such that \( K \leq P \), we have for all \( n = 3, 4, \ldots \)
\[ E[T_n(\theta)]^2 \]
(15)
\[ = E[T_n(\theta)] + \left( \frac{(n - 3)}{3} \right)^3 \cdot (E[T_n(\theta)])^2 + 3(n - 3) \left( \frac{n}{3} \right)^3 \cdot E[\chi_{123}(\theta)\chi_{124}(\theta)]. \]

These facts are discussed in Section X. For future reference, we note
\[ r(\theta) \leq q(\theta)^2 \] (16)
by direct inspection, whence
\[ \beta(\theta) \geq (1 - q(\theta))^3 > 0. \] (17)

IV. Zero-one laws

For simplicity of exposition we refer to any pair of functions \( P, K : \mathbb{N}_0 \rightarrow \mathbb{N}_0 \) as a scaling (for random key graph) provided the natural condition \( K_n \leq P_n \) holds for all \( n = 2, 3, \ldots \).

A. Asymptotic equivalences

The two asymptotic equivalence results presented next prove useful in a number of places. The first one was already obtained in [25], and is given here for easy reference.

**Lemma 4.1:** For any scaling \( P, K : \mathbb{N}_0 \rightarrow \mathbb{N}_0 \), we have
\[ \lim_{n \rightarrow \infty} q(\theta(n)) = 1 \quad \text{if and only if} \quad \lim_{n \rightarrow \infty} \frac{K_n^2}{P_n} = 0. \] (18)
Under either condition at (18) we have the asymptotic equivalence
\[ 1 - q(\theta(n)) \sim \frac{K_n^2}{P_n}. \] (19)

Because \( 1 \leq K_n \leq K_n^2 \) for all \( n = 1, 2, \ldots \), the condition (18) implies both
\[ \lim_{n \rightarrow \infty} \frac{1}{P_n} = 0 \quad \text{and} \quad \lim_{n \rightarrow \infty} \frac{K_n}{P_n} = 0. \] (20)
Therefore, \( \lim_{n \rightarrow \infty} P_n = \infty \), and for any \( c > 0 \), we have \( cK_n < P_n \) for all \( n \) sufficiently large in \( \mathbb{N}_0 \) (dependent on \( c \)).

The second asymptotic equivalence will allow us to state the results in a more explicit form; its proof is available in Appendix A.

**Proposition 4.2:** For any scaling \( P, K : \mathbb{N}_0 \rightarrow \mathbb{N}_0 \) satisfying (18), the asymptotic equivalence
\[ \beta(\theta(n)) \sim \tau(\theta(n)). \] (21)
B. Zero-one laws for the existence of triangles

A main result of the paper is a zero-one law for the existence of triangles in random key graphs. This result will be established by the method of first and second moments applied to the count variables \((9)\), e.g., see [2, p. 2], [10, p. 55]. The zero-law is given first.

Theorem 4.3: For any scaling \(P, K : \mathbb{N}_0 \rightarrow \mathbb{N}_0\), the zero-law

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

holds under the condition

\[
\lim_{n \rightarrow \infty} n^3 \tau(\theta_n) = 0.
\]

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

\[
P \left[ T_n(\theta_n) > 0 \right] \leq \left(\frac{n}{3}\right) \beta(\theta_n)
\]

by virtue of Proposition 3.1. Theorem 4.3 thus follows if under (22) we show that \(\lim_{n \rightarrow \infty} \left(\frac{n}{3}\right) \beta(\theta_n) = 0\). By Proposition 4.2 this convergence is equivalent to the assumed condition \(\lim_{n \rightarrow \infty} n^3 \tau(\theta_n) = 0\), and the proof of Theorem 4.3 is now complete. ■

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

Theorem 4.4: For any scaling \(P, K : \mathbb{N}_0 \rightarrow \mathbb{N}_0\) for which the limit \(\lim_{n \rightarrow \infty} q(\theta_n) = q^*\) exists, the one-law

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

holds if either \(0 \leq q^* < 1\) or \(q^* = 1\) with the additional condition

\[
\lim_{n \rightarrow \infty} n^3 \tau(\theta_n) = \infty.
\]

To facilitate an upcoming comparison with analogous results in Erdős-Rényi graphs, we combine Theorem 4.3 and Theorem 4.4 into the symmetric statement.

Theorem 4.5: For any scaling \(P, K : \mathbb{N}_0 \rightarrow \mathbb{N}_0\) for which \(\lim_{n \rightarrow \infty} q(\theta_n)\) exists, we have

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

By Lemma 4.1 we note that the condition \(\lim_{n \rightarrow \infty} n^3 \tau(\theta_n) = 0\) implies \(\lim_{n \rightarrow \infty} q(\theta_n) = 1\) (hence \(q^* = 1\)).

V. CLUSTERING

A. Defining clustering coefficients

Many real networks are known [18] to exhibit a phenomenon often called high clustering (or transitivity), informally defined as the propensity of a node’s neighbors to be neighbors as well. Clustering properties of a network is closely related to the existence and number of triangles. To that end, we find it of interest to also look at the clustering coefficient of random key graphs under various parameter regimes.

A formal definition of clustering in a network 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 \(G\) with contain vertex \(i\). The clustering coefficient of node \(i\) is given by

\[
C_i(G) = \begin{cases} \frac{T_i(G)}{2 \sum_{j \neq i} d_i(j - 1)} & \text{if } d_i \geq 2 \\ 0 & \text{otherwise} \end{cases}
\]

where \(d_i\) is the degree of node \(i\) in \(G\).

There are, however, several possible definitions for a graph-wide notion of clustering [15]: Inspired by (24), it is natural to consider the average clustering coefficient \(C_{\text{Avg}}(G)\) of the graph \(G\) defined by

\[
C_{\text{Avg}}(G) = \frac{1}{|V'|} \sum_{i \in V'} C_i(G)
\]

where \(V' = \{i \in V : d_i \geq 2\}\). This last definition, while a natural one, is often replaced by the notion of clustering as “fraction of transitive triples,” namely

\[
C^*(G) = \frac{\sum_{i \in V} T_i(G)}{2 \sum_{i \in V} d_i(j - 1)}
\]

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

Related (but simpler) definitions are possible in the context of random graphs with exchangeable link assignments (as is the case for the random graphs of interest here), e.g., [4]. In particular, a possible approach is to define the clustering coefficient of the random key graph \(K(n; \theta)\) by

\[
C_K(\theta) = \mathbb{P} \left[ E_{12}(\theta) | E_{13}(\theta) \cap E_{23}(\theta) \right].
\]

Interest in this quantity stems from the fact that it is expected to provide a good approximation to (26) when \(n\) is large. The strong consistency result given next formalizes this idea, and is established in Appendix B.

Proposition 5.1: For positive integers \(K, P\) such that \(K \leq P\), we have

\[
\lim_{n \rightarrow \infty} C^*(K(n; \theta)) = C_K(\theta) \text{ a.s.}
\]

Simulation results given in Table I of Section V illustrate the convergence (28) (and an analog one for Erdős-Rényi graphs).

B. Evaluating \(C_K(\theta)\)

Throughout we shall use the simpler definition (27) for reasons of analytical tractability. This should already be plain from the next proposition.

Proposition 5.2: For positive integers \(K, P\) such that \(K \leq P\), we have

\[
C_K(\theta) = \frac{\beta(\theta)}{(1 - q(\theta))^2}
\]

with \(\beta(\theta)\) given by (11).
Proof. The definitions of \( C_K(\theta) \) and \( \chi_{123}(\theta) \) yield

\[
C_K(\theta) = \frac{\mathbb{P}[E_{12}(\theta) \cap E_{13}(\theta) \cap E_{23}(\theta)]}{\mathbb{P}[E_{13}(\theta) \cap E_{23}(\theta)]} = \mathbb{E}[\chi_{123}(\theta)] \left( 1 - q(\theta) \right)^2
\]

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

\[
\mathbb{P}[E_{13}(\theta) \cap E_{23}(\theta)] = \mathbb{P}[K_1(\theta) \cap K_3(\theta) \neq \emptyset, K_2(\theta) \cap K_3(\theta) \neq \emptyset] = (1 - q(\theta))^2
\]

by virtue of (6) (and comments following it). The conclusion (29) is immediate upon substituting (13) into (30). \( \blacksquare \)

VI. FACTS CONCERNING ERDÖS–RÉNYI GRAPHS

A little later in this paper, we shall compare random key graphs to related Erdős–Rényi graphs [6], but first some notation: For each \( n = 2, 3, \ldots \) and each \( p \) in \([0, 1]\), let \( \mathbb{G}(n; p) \) denote the Erdős–Rényi graph on the vertex set \( \{1, \ldots, n\} \) with link assignment probability \( p \). In analogy with earlier notation let \( E_{ij}(p) \) denote the event that there is an (undirected) link assigned between the distinct nodes \( i \) and \( j \) in \( \mathbb{G}(n; p) \). Thus, the random graph \( \mathbb{G}(n; p) \) is characterized by having the \( \frac{n(n-1)}{2} \) (undirected) links between the \( n \) nodes be independently assigned with probability \( p \), i.e., the events \( \{E_{ij}(p), 1 \leq i < j \leq n\} \) are mutually independent events, each of probability \( p \). For ease of exposition it will always be understood that \( E_{ij}(p) = E_{ji}(p) \) for distinct \( i, j = 1, \ldots, n \). We refer to any mapping \( p : \mathbb{N}_0 \to [0, 1] \) as a scaling for Erdős–Rényi graphs.

In analogy with (9) let \( T_{n}(p) \) denote the number of distinct triangles in \( \mathbb{G}(n; p) \). Under the enforced independence, we note that

\[
\mathbb{E}[T_{n}(p)] = \binom{n}{3} \tau^*(p), \quad 0 \leq p \leq 1
\]

with

\[
\tau^*(p) = p^3, \quad 0 \leq p \leq 1.
\]

Link assignments being exchangeable in Erdős–Rényi graphs, we again define the clustering coefficient in \( \mathbb{G}(n; p) \) by

\[
C_{ER}(p) = \mathbb{P}[E_{12}(p) \mid E_{13}(p) \cap E_{23}(p)]
\]

so that

\[
C_{ER}(p) = \frac{\mathbb{P}[E_{12}(p) \cap E_{13}(p) \cap E_{23}(p)]}{\mathbb{P}[E_{13}(p) \cap E_{23}(p)]} = p
\]

by mutual independence. Here as well, we expect

\[
\lim_{n \to \infty} C^*(\mathbb{G}(n; p)) = C_{ER}(p) \quad a.s.
\]

This can be established by the same arguments as the ones provided in the proof of Proposition 5.1.

Random key graphs are not equivalent to Erdős–Rényi 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 under the exact matching condition

\[
p = 1 - q(\theta) = p(\theta).
\]

See [22] for a discussion of (dis)similarities. However, in order to meaningfully compare the asymptotic regime of random key graphs with that of Erdős–Rényi graphs, we shall say that the scaling \( p : \mathbb{N}_0 \to [0, 1] \) (for Erdős–Rényi graphs) is asymptotically matched to the scaling \( P, K : \mathbb{N}_0 \to \mathbb{N}_0 \) (for random key graphs) if

\[
p_n \sim p(\theta_n).
\]

Under the condition (18), the asymptotic matching condition (37) amounts to

\[
p_n \sim \frac{K_n^2}{P_n}
\]

by virtue of Lemma 4.1.

Condition (36) (resp. (37)) is equivalent to requiring that the expected degrees in \( \mathbb{K}(n; \theta_n) \) and \( \mathbb{G}(n; p_n) \) coincide (resp. be asymptotically equivalent).

VII. COMPARING THE NUMBER OF TRIANGLES

Fix \( p \) in \([0, 1]\), and positive integers \( K \) and \( P \) such that \( K \leq P \). From (14) and (32) it is plain that

\[
\mathbb{E}[T_{n}(p)] = \frac{\beta(\theta)}{\tau^*(p)}, \quad n = 3, 4, \ldots
\]

A. Matching expected degrees

Under the exact matching condition (36), this last expression becomes

\[
\frac{\mathbb{E}[T_{n}(p)]}{\mathbb{E}[T_{n}(p(\theta))]} = \frac{\beta(\theta)}{\tau^*(p(\theta))} = 1 + \frac{q(\theta)}{(1 - q(\theta))^3} \cdot q(\theta)
\]

for each \( n = 3, 4, \ldots \), whence \( \mathbb{E}[T_{n}(p(\theta))] \leq \mathbb{E}[T_{n}(p)] \) by virtue of (16). Consequently, the expected number of triangles in a random key graph is always at least as large as the corresponding quantity in an Erdős–Rényi graph matched to it. This was already suggested by Di Pietro et al. [5] with the help of limited simulations.

A similar result is available when the scalings are only asymptotically matched.

Corollary 7.1: Consider a scaling \( K, P : \mathbb{N}_0 \to \mathbb{N}_0 \) satisfying (18), and a scaling \( p : \mathbb{N}_0 \to [0, 1] \). Under the asymptotic matching condition (37), we have the equivalence

\[
\frac{\mathbb{E}[T_{n}(\theta_n)]}{\mathbb{E}[T_{n}(p_n)]} \sim 1 + \frac{P_n}{K_n^3}.
\]

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 Erdős–Rényi graphs.
Proof. Replacing $\theta$ by $\theta_n$ and $p$ by $p_n$ according to these scalings in the expression (39), we get

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

Under (18), Proposition 4.2 yields

$$
\frac{\mathbb{E} [T_n(\theta_n)]}{\mathbb{E} [T_n(p_n)]} \sim \frac{\tau(\theta_n)}{\tau^*(p_n)}
$$

with

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

With the help of (38), we now conclude

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

and the equivalence (41) follows from (42).

B. Dimensioning WSNs

In the context of a WSN with $n$ nodes, it is natural to select the parameters $K_n$ and $P_n$ such that the induced random key graph is connected. However, there is a tradeoff between connectivity and security [5], requiring $K_n^3$ to be kept as close as possible to the critical scaling $\log n$ for connectivity; see the papers [1], [5], [17], [21], [25]. The desired regime near the boundary can be achieved by taking

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

with $c > 1$ but close to one. On the other hand, it follows from (41) that

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

provided condition (18) holds. This obviously occurs when (44) holds, in which case the condition at (45) amounts to taking

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

Thus, combining we see that under (44) it holds that

$$
\frac{\mathbb{E} [T_n(\theta_n)]}{\mathbb{E} [T_n(p_n)]} \sim 1 \quad \text{if and only if} \quad K_n \gg \frac{n}{\log n}.
$$

In that case the expected number of triangles in random key graphs is of the same order as the corresponding quantity in asymptotically matched Erdős-Rényi graphs with $\mathbb{E} [T_n(\theta_n)] \sim \mathbb{E} [T_n(p_n)] \sim \frac{c}{\pi} (\log n)^3$. This is a direct consequence of (32), (38) and (44). This conclusion holds regardless of the value of $c$ in (44).

However, given the limited memory and computational power of the sensor nodes, the key ring sizes at (46) are not practical. In addition, they will lead to high node degrees and this in turn will decrease network resiliency against node capture attacks. Indeed, it was proposed by Di Pietro et al. [5, Thm. 5.3] that resiliency in WSNs against node capture attacks can be ensured by selecting $K_n$ and $P_n$ such that $K_n^3 \sim \frac{1}{n}$. Under (44) this additional requirement then leads to $K_n \sim c \cdot \log n$ so that $P_n \sim c \cdot n \cdot \log n$, and (41) now implies

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

This means that, for realistic WSN scenarios the expected number of triangles in the induced random key graphs will be orders of magnitude larger than in Erdős-Rényi graphs.

We conclude this section by comparing Theorem 4.5 with its analog for Erdős-Rényi graphs. Fix $n = 2, 3, \ldots$ 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 Erdős-Rényi graphs is well known [2, Chp. 4], [10, Thm. 3.4, p. 56].

Theorem 7.2: For any scaling $p : \mathbb{N}_0 \to [0, 1]$, we have

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

This result was also established by the method of first and second moments, and its form is easily understood once we recall (32).

As we compare Theorem 4.5 and Theorem 7.2, we see a direct analogy since the terms $\tau(\theta_n)$ and $\tau^*(p_n)$ both correspond to (asymptotic) probability that three arbitrary nodes form a triangle, in random key graphs and Erdős-Rényi graphs, respectively. More interestingly, we see that under the asymptotic matching condition (37) (and (18)), triangles will start appearing earlier in the evolution of a random key graph as compared to an Erdős-Rényi graph matched to it; this fact is evident from (43). Put differently, we can see triangles in a random key graph with smaller asymptotic edge probability (which corresponds to a smaller mean node degree) than in Erdős-Rényi graph. As shall be clear from the previous discussions and (43), the larger is the quantity $P_n/K_n^3$, the more emphasized these differences will be.

VIII. COMPARING CLUSTERING COEFFICIENTS

Fix $p$ in $(0, 1]$, and positive integers $K$ and $P$ such that $K \leq P$. Combining (29) and (34) we get

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

A. The exactly matched case

Under the exact matching condition (36) we find

$$
\frac{C_K(\theta)}{C_{ER}(p(\theta))} = 1 + \frac{q(\theta)^2 - r(\theta)}{(1 - q(\theta))^3} \cdot q(\theta)
$$

as we recall (5). Thus, $C_K(\theta) \geq C_{ER}(p(\theta))$ by virtue of (16).

The clustering coefficient of a random key graph is at least as large as that of the Erdős-Rényi graph exactly matched to it.
In fact, the lower bound mentioned above is achieved only when \( P < 2K \), i.e., from (4) we get
\[
\frac{C_K(\theta)}{C_{ER}(p)} = 1 \quad \text{whenever} \quad P < 2K.
\]
It is a simple matter to check that for \( K = 1 \), (51) yields
\[
\frac{C_K(1, P)}{C_{ER}(p)} = P
\]
for each \( P = 1, 2, \ldots, \) while for \( K = 2 \) we have
\[
\frac{C_K(2, P)}{C_{ER}(p)} = P - \frac{2P^3 - 4P^2 - P + 3}{2(2P - 3)^3} \geq P.
\]
It is also straightforward to show that
\[
1 \leq \frac{C_K(\theta)}{C_{ER}(p)} \leq P.
\]
Since \( P \) can be made very large, we understand that the parameters of the random key graph can be selected so that it has a much larger clustering coefficient than the Erdős-Rényi graph matched to it. This will especially be so for WSNs where the size of the key pool \( P \) is expected to be in the range \( 2^{17} - 2^{20} \) [7].

In Table I we compare the clustering coefficients of random key graphs and Erdős-Rényi graphs for several realistic parameter values. The numerical values of \( C_K(\theta) \) and \( C_{ER}(p) \) are obtained directly from the expressions (27) and (33), respectively. On the other hand, \( C_n^2(\theta) \) and \( C_n^3(p) \) stand for the clustering coefficient of \( \mathbb{G}(n; \theta) \) and \( \mathbb{G}(n; p) \), respectively, calculated through (26) and averaged over 1000 realizations; the number of nodes is set to \( n = 1000 \) in all simulations. The data support the claim that the definition (26) captures essentially the same feature as the quantities (27) and (33), i.e., the results given in Table I can be taken as an indication of the validity of (28) and (35).

<table>
<thead>
<tr>
<th>( K )</th>
<th>( P )</th>
<th>( C_K(\theta) )</th>
<th>( C_n^2(\theta) )</th>
<th>( C_{ER}(p) )</th>
<th>( C_n^3(p) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>( 10^4 )</td>
<td>0.2590</td>
<td>0.2587</td>
<td>0.0160</td>
<td>0.0159</td>
</tr>
<tr>
<td>8</td>
<td>( 5 \times 10^4 )</td>
<td>0.1348</td>
<td>0.1349</td>
<td>0.0127</td>
<td>0.0128</td>
</tr>
<tr>
<td>16</td>
<td>( 2 \times 10^4 )</td>
<td>0.0737</td>
<td>0.0736</td>
<td>0.0127</td>
<td>0.0128</td>
</tr>
<tr>
<td>20</td>
<td>( 4 \times 10^4 )</td>
<td>0.0590</td>
<td>0.0590</td>
<td>0.0100</td>
<td>0.0100</td>
</tr>
<tr>
<td>24</td>
<td>( 10^5 )</td>
<td>0.0490</td>
<td>0.0498</td>
<td>0.0057</td>
<td>0.0057</td>
</tr>
<tr>
<td>32</td>
<td>( 10^5 )</td>
<td>0.0498</td>
<td>0.0498</td>
<td>0.0102</td>
<td>0.0102</td>
</tr>
<tr>
<td>40</td>
<td>( 5 \times 10^5 )</td>
<td>0.0280</td>
<td>0.0280</td>
<td>0.0032</td>
<td>0.0031</td>
</tr>
<tr>
<td>64</td>
<td>( 10^6 )</td>
<td>0.0196</td>
<td>0.0196</td>
<td>0.0041</td>
<td>0.0041</td>
</tr>
</tbody>
</table>

**Table I**

Clustering coefficients with fixed \( \theta \) and \( p = 1 - q(\theta) \)

**Proof.** As we replace \( \theta \) by \( \theta_n \) and \( p \) by \( p_n \) according to these scalings in the expression (48), we get
\[
\frac{C_K(\theta_n)}{C_{ER}(p_n)} = \frac{\beta(\theta_n)}{p_n(1 - q(\theta_n))^2}, \quad n = 1, 2, \ldots \quad (51)
\]
Note that
\[
\frac{C_K(\theta_n)}{C_{ER}(p_n)} \sim \frac{\beta(\theta_n)}{(1 - q(\theta_n))^3} \sim \frac{\tau(\theta_n)}{(1 - q(\theta_n))^3}. \quad (52)
\]
The first equivalence is a consequence of (37) and the second one is validated by Proposition 4.2 under (18). With (38) being still valid here, we easily conclude (50) by the same arguments as the ones used to obtain (41).

Thus, under (18) and (37), we conclude that
\[
\lim_{n \to \infty} \frac{C_K(\theta_n)}{C_{ER}(p_n)} = 1 \quad \text{if} \quad \lim_{n \to \infty} \frac{K^3_n}{P_n} = \infty, \quad (53)
\]
and
\[
\lim_{n \to \infty} \frac{C_K(\theta_n)}{C_{ER}(p_n)} = \infty \quad \text{if} \quad \lim_{n \to \infty} \frac{K^3_n}{P_n} = 0. \quad (54)
\]
It is now easy to see that asymptotically matched random key graphs and Erdős-Rényi graphs can have vastly different clustering coefficients: Under the condition (44), (53) can hold only if the key ring size \( K \) is much larger than \( n/\log n \). As already discussed, this condition cannot be satisfied in a practical WSN scenario due to limitations on the sensor nodes and security constraints. In fact, we see from (41) and (47) that, in a realistic WSN scenario, 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 Erdős-Rényi graph. This provides yet another property of the random key graphs that Erdős-Rényi graph models cannot adequately capture, further indicating the non-equivalence of the two models for practical WSN scenarios.

**C. Small worlds**

Since random key graphs can be highly clustered, a natural question arises as to their suitability to model the small world effect. This notion is linked to a well-known series of experiments conducted by Milgram [14] 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 a small world in the sense that the path lengths between pairs of individuals are short. As a way to capture Milgram’s experiments, Watts and Strogatz [20] defined small worlds as 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 effect if its average path length is of the same order as that of an Erdős-Rényi 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...
Thus, by repeated application of (55) we find
\[
E \left[ \chi_{123}(\theta) \right] = (1 - q(\theta))^2
\]
with high probability where \( \mathcal{K}(n; \theta_n) \) is the largest connected component of \( \mathcal{H}(n; \theta_n) \). This suggests that the diameter, hence the average path length, in random key graphs is small as was the case with Erdős-Rényi graphs [3]. We also note [24, Corollary 5.2] that random key graphs have very small \((\leq 2)\) diameter under certain parameter ranges. Therefore, random key graphs may indeed be considered as good candidate models for small worlds!

The fact that random key graphs can exhibit small world properties can be taken as a further motivation to pursue other application areas for them, particularly in the field of social networks. We provide one concrete possibility below, where random key graphs appear naturally in modeling common-interest relationships in a population: Suppose that there exists an object pool consisting of \( P \) objects and that each of the \( n \) users picks \( K \) distinct objects uniformly at random from this object pool; objects may represent hobbies, books, movies, etc. Two friends are said to have a common-interest relation if they have at least one common object in their object rings. This naturally suggests modeling the common interest relationship by a random key graph \( \mathcal{H}(n; K, P) \). Various properties of random key graphs would then be of interest in applications such as implementation of large-scale, distributed publish-subscribe services [26].

IX. A PROOF OF PROPOSITION 3.1

Fix positive integers \( K \) and \( P \) such that \( K \leq P \). As exchangeability yields (14), we need only show the validity of (13).

In the discussion that follows we omit the explicit dependence on \( \theta \) when no confusion arises from doing so. Also, we make repeated use of the fact that for any pair of events \( E \) and \( F \) in \( \mathcal{F} \), we have
\[
P[E \cap F] = P[E] - P[E \cap F^c]. \tag{55}
\]
Thus, by repeated application of (55) we find

\[
E \left[ \chi_{123}(\theta) \right] = (1 - q(\theta))^2
\]

Next, as we use (55) one more time, we get
\[
P[K_1 \cap K_2 \neq \emptyset, K_1 \cap K_3 \neq \emptyset, K_2 \cap K_3 \neq \emptyset] = P[K_1 \cap K_3 = \emptyset, K_2 \cap K_3 = \emptyset] - P[K_1 \cap K_2 = \emptyset, K_1 \cap K_3 = \emptyset, K_2 \cap K_3 = \emptyset].
\]

Again, by independence, with the help of (6) we conclude that
\[
P[K_1 \cap K_3 = \emptyset, K_2 \cap K_3 = \emptyset] = q(\theta)^2 \tag{57}
\]

and
\[
P[K_1 \cap K_2 = \emptyset, K_1 \cap K_3 = \emptyset, K_2 \cap K_3 = \emptyset] = q(\theta)r(\theta) \tag{58}
\]

since \( |K_1 \cup K_2| = 2K \) when \( K_1 \cap K_2 = \emptyset \). Collecting these facts we find
\[
E \left[ \chi_{123}(\theta) \right] = (1 - q(\theta))^2 - (1 - q(\theta)) q(\theta) + q(\theta)^2 - q(\theta)r(\theta)
\]

and the conclusion (13) follows by elementary algebra. \( \blacklozenge \)

X. A PROOF OF PROPOSITION 3.2

Fix positive integers \( K \) and \( P \) such that \( K \leq P \), and \( n = 3, 4, \ldots \). By exchangeability and the binary nature of the rvs involved we readily obtain
\[
E \left[ T_n(\theta)^2 \right] = E \left[ T_n(\theta) \right] + \binom{n}{3} \binom{3}{2} \binom{n - 3}{1} E \left[ \chi_{123}(\theta) \chi_{124}(\theta) \right] + \binom{n}{3} \binom{n - 3}{2} E \left[ \chi_{123}(\theta) \chi_{145}(\theta) \right] + \binom{n}{3} \binom{n - 3}{3} E \left[ \chi_{123}(\theta) \chi_{456}(\theta) \right]. \tag{59}
\]

Under the enforced independence assumptions the rvs \( \chi_{123}(\theta) \) and \( \chi_{456}(\theta) \) are independent and identically distributed. As a result,
\[
E \left[ \chi_{123}(\theta) \chi_{456}(\theta) \right] = E \left[ \chi_{123}(\theta) \right] E \left[ \chi_{456}(\theta) \right] = \beta(\theta)^2
\]

so that using the relation (14) we obtain
\[
\binom{n}{3} \binom{n - 3}{3} E \left[ \chi_{123}(\theta) \chi_{456}(\theta) \right] = \frac{(n - 3)(n - 3)}{3} E \left[ T_n(\theta) \right]^2. \tag{60}
\]

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
\[
P[\chi_{123}(\theta) = 1 | K_1(\theta)] = P[\chi_{123}(\theta) = 1] = \beta(\theta). \tag{61}
\]

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
\[
E \left[ \chi_{123}(\theta) \chi_{145}(\theta) \right] = E \left[ \chi_{123}(\theta) \right] E \left[ \chi_{145}(\theta) \right] = \beta(\theta)^2.
\]
Again by virtue of (14), this last observation yields
\[
\binom{n}{3} \binom{3}{1} \binom{n-3}{2} E \left[ \chi_{123}(\theta) \chi_{145}(\theta) \right] \\
= 3 \binom{n-3}{2} \binom{3}{3} \cdot \left( E \left[ T_n(\theta) \right] \right)^2.
\] (62)
Substituting (60) and (62) into (59) establishes Proposition 3.2.

\section{XI. Proving Theorem 4.4}

Assume first that \( q^* \) satisfies \( 0 \leq q^* < 1 \). Fix \( n = 3, 4, \ldots \) and partition the \( n \) nodes into the \( k_n + 1 \) non-overlapping groups \( (1, 2, 3), (4, 5, 6), \ldots, (3k_n + 1, 3k_n + 2, 3k_n + 3) \) with \( k_n = \left\lfloor \frac{n-1}{3} \right\rfloor \). If \( K(n; \theta_n) \) contains no triangle, then \( q^* \) of these \( k_n + 1 \) groups of nodes forms a triangle. With this in mind we get
\[
P \left[ T_n(\theta_n) = 0 \right] \\
\leq \prod_{\ell=0}^{k_n-1} P \left[ \text{Nodes } 3\ell + 1, 3\ell + 2, 3\ell + 3 \text{ do not form a triangle in } K(n; \theta_n) \right] \\
= \prod_{\ell=0}^{k_n-1} P \left[ \text{Nodes } 3\ell + 1, 3\ell + 2, 3\ell + 3 \text{ do not form a triangle in } K(n; \theta_n) \right] \\
= (1 - \beta(\theta_n))^{k_n+1} \\
\leq (1 - (1 - q(\theta_n))^3)^{k_n+1} \\
\leq e^{-k_n(1-q(\theta_n))^3}. 
\] (63) 
(64) 
(65)
Note that (63) follows from the fact that the events
\[
\left[ \text{Nodes } 3\ell + 1, 3\ell + 2, 3\ell + 3 \text{ do not form a triangle in } K(n; \theta_n) \right], \quad \ell = 0, \ldots, \ell_n
\] are mutually independent due to the non-overlap condition, while the inequality (64) is justified with the help of (17). Let \( n \) go to infinity in the inequality (65). From the constraint \( q^* < 1 \) we conclude that \( \lim_{n \to \infty} P \left[ T_n(\theta_n) \right] = 0 \) 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 \( q^* < 1 \).

To handle the case \( q^* = 1 \), we use a standard bound which forms the basis of the method of second moment [10, Remark 3.1, p. 55]. Here this bound takes the form
\[
\left( \frac{E \left[ T_n(\theta_n) \right]}{E \left[ T_n(\theta_n) \right]^2} \right)^2 \leq P \left[ T_n(\theta_n) > 0 \right], \quad n = 3, 4, \ldots
\] (66)
Theorem 4.4 then will be established in the case \( q^* = 1 \) if we show under (18) that the condition (23) implies
\[
\lim_{n \to \infty} \frac{E \left[ T_n(\theta_n) \right]^2}{E \left[ T_n(\theta_n) \right]^2} = 1.
\] (67)
As pointed earlier, the conditions (18) imply \( 3K_n < P_n \) for all \( n \) sufficiently large in \( \mathbb{N}_0 \). On that range, with \( \theta \) replaced by \( \theta_n \), Proposition 3.2 yields
\[
\frac{E \left[ T_n(\theta_n) \right]^2}{E \left[ T_n(\theta_n) \right]^2} = 1 + \frac{\binom{n-3}{3}}{\binom{n}{3}} + \frac{\binom{n-3}{2}}{\binom{n}{3}} E \left[ \chi_{123}(\theta_n) \chi_{145}(\theta_n) \right] \\
\geq \frac{1}{\binom{n}{3}} \cdot \left( E \left[ \chi_{123}(\theta_n) \right] \right)^2.
\]

as we make use of (14) in the last term.

Let \( n \) go to infinity in the resulting expression: Under condition (23), we have \( \lim_{n \to \infty} n^3 \beta(\theta_n) = \infty \) by Proposition 4.2, whence \( \lim_{n \to \infty} E \left[ T_n(\theta_n) \right] = \infty \) by virtue of (14). Since
\[
\lim_{n \to \infty} \left( \frac{n^3}{3(n-3)} + \frac{n^3}{3} \right) = 1
\] (68)
and
\[
\frac{n^3}{3(n-3)} \sim \frac{n^2}{18},
\] (69)
the convergence (67) will hold if we show that
\[
\lim_{n \to \infty} \frac{E \left[ \chi_{123}(\theta_n) \chi_{145}(\theta_n) \right]}{E \left[ \chi_{123}(\theta_n) \right]^2} = 0
\] (70)
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
\[
\frac{E \left[ \chi_{123}(\theta_n) \chi_{145}(\theta_n) \right]}{E \left[ \chi_{123}(\theta_n) \right]^2} \leq \frac{E \left[ \chi_{123}(\theta_n) \right] \mathbf{1} \left[ K_1(\theta) \cap K_4(\theta) \neq \emptyset \right]}{E \left[ \chi_{123}(\theta_n) \right]^2},
\] (71)
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
\[
\frac{E \left[ \chi_{123}(\theta) \mathbf{1} \left[ K_1(\theta) \cap K_4(\theta) \neq \emptyset \right] \right]}{E \left[ \chi_{123}(\theta) \right]^2} = \beta(\theta)(1 - q(\theta))
\] (72)
as we recall (4) and (61). Using (71) together with (13) and (72) we readily obtain the inequalities
\[
\frac{E \left[ \chi_{123}(\theta) \chi_{145}(\theta) \right]}{E \left[ \chi_{123}(\theta) \right]^2} \leq \frac{\beta(\theta)(1 - q(\theta))}{E \left[ \chi_{123}(\theta) \right]^2} \leq \beta(\theta)^{-2/3}
\] (73)
where in the last step we noted that \( 1 - q(\theta) \leq \beta(\theta)^{1/3} \) by appealing to (17).

Returning to the convergence (70) we see from (73) that we need only show
\[
\lim_{n \to \infty} \frac{n^2 \beta(\theta_n)^{2/3}}{\beta(\theta)^{2/3}} = \infty.
\] (74)
As Proposition 4.2 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 (74) follows under the condition (23).
REFERENCES


APPENDIX

A. A proof of Proposition 4.2

Under the enforced assumptions, by the comments following (20), we have $3K_n < P_n$ for all $n$ sufficiently large in $\mathbb{N}_0$. On that range we can use the expression (11) to write

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

As Lemma 4.1 already implies $q(\theta_n)^3 \sim 1$ and $(1 - q(\theta_n))^3 \sim \left(\frac{K^2}{P_n}\right)^3$, the asymptotic equivalence $\beta(\theta_n) \sim \tau(\theta_n)$ will be established if we show that

$$1 - r(\theta) \frac{K^3}{P_n} \sim \frac{K^3}{P_n},$$

(A.1)

We proceed as follows: With positive integers $K, P$ such that $3K \leq P$, we note that

$$\frac{r(\theta)}{q(\theta)^2} = \left(\frac{(P - 3K)!}{(P - K)!}\right)^2 \cdot \left(\frac{(P - 2K)!}{(P - K)!}\right)^2 \frac{P!}{(P - 3K)!} = \frac{P - 3K}{P - K} \cdot \frac{P - 2K}{P - K} \cdots \frac{P - K}{P - K}.$$

and elementary bounding arguments yield

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

Pick a scaling $P, K : \mathbb{N}_0 \to \mathbb{N}_0$ satisfying the equivalent conditions (18) and consider $n$ sufficiently large in $\mathbb{N}_0$ so that $3K_n < P_n$. On that range, we replace $\theta$ by $\theta_n$ in the last chain of inequalities according to this scaling. A standard sandwich argument will yield the desired equivalence (A.1) if we show that

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

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

$$1 - \left(1 - \left(\frac{K_n}{P_n - cK_n}\right)^2\right)^{K_n} = K_n A_n(c)^2 \int_0^1 \left(1 - A_n(c)^2 t\right)^{K_n-1} dt.$$


on the appropriate range. The asymptotics
\[ A_n(c)^2 = \left( \frac{K_n}{P_n - cK_n} \right)^2 \sim \left( \frac{K_n}{P_n} \right)^2 \]  
and
\[ K_n A_n(c)^2 \sim \frac{K_n^2}{P_n^2} \]
follow from (20), so that (A.2) will hold if we show that
\[ \lim_{n \to \infty} \int_0^1 (1 - A_n(c)^2) t^{K_n - 1} dt = 1. \]  
(A.5)

In view of (A.3) we conclude from (20) that for all \( n \) sufficiently large in \( \mathbb{N}_0 \) we have \( \sup_{0 \leq t \leq 1} |1 - A_n(c)^2 t| \leq 1 \). Therefore, the Bounded Convergence Theorem will yield (A.5) as soon as we establish
\[ \lim_{n \to \infty} \left( 1 - A_n(c)^2 t \right)^{K_n - 1} = 1, \quad 0 \leq t \leq 1. \]  
(A.6)

To that end, recall the decomposition
\[ \log(1 - x) = - \int_0^x \frac{1}{1 - t} dt = -x - \Psi(x) \]  
(A.7)
where
\[ \Psi(x) = \int_0^x \frac{t}{1 - t} dt, \quad 0 \leq x < 1. \]

It is easy to check that
\[ \lim_{x \to 0} \frac{\Psi(x)}{x} = 0. \]  
(A.8)

Fix \( n \) sufficiently large in \( \mathbb{N}_0 \) as required above. For each \( t \) in the interval \( (0, 1] \), with the help of (A.7) we can write
\[ (1 - A_n(c)^2 t)^{K_n - 1} = e^{(K_n - 1) \log(1 - A_n(c)^2 t)} \]
\[ = e^{-(K_n - 1) A_n(c)^2 t - (K_n - 1) \Psi(A_n(c)^2 t)}. \]  
(A.9)

Returning to (A.4), we use (18) and (20) 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 \( \lim_{n \to \infty} (K_n - 1) A_n(c)^2 t = 0 \), whence
\[ \lim_{n \to \infty} (K_n - 1) \Psi(A_n(c)^2 t) = \lim_{n \to \infty} (K_n - 1) A_n(c)^2 t \cdot \frac{\Psi(A_n(c)^2 t)}{A_n(c)^2 t} = 0 \]
with the help of (A.8) in the last step. Finally, letting \( n \) go to infinity in (A.9), we readily get (A.6) as desired.

\begin{center}
\section*{B. A proof of Proposition 5.1}
\end{center}

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
\[ P_{n,i} = \{ (j,k) : 1 \leq j < k \leq n, \ j \neq i, \ k \neq i \}. \]

Next, define the count rvs \( T_{n,i}(\theta) \) and \( T_{n,i}^*(\theta) \) by
\[ T_{n,i}(\theta) = \sum_{(j,k) \in P_{n,i}} \xi_{ij}(\theta) \xi_{ik}(\theta) \xi_{jk}(\theta), \]
and
\[ T_{n,i}^*(\theta) = \sum_{(j,k) \in P_{n,i}} \xi_{ij}(\theta) \xi_{ik}(\theta). \]

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 \( T_{n,i}^*(\theta) \) counts the number of distinct pairs of nodes which are both connected to node \( i \) in \( \mathbb{K}(n;\theta) \). We also introduce the rv \( D_i(\theta) \) as the degree of node \( i \) in \( \mathbb{K}(n;\theta) \) given by
\[ D_i(\theta) = \sum_{k=1, k \neq i}^n \xi_{ik}(\theta). \]

Observe that we have
\[ \sum_{i=1}^n T_{n,i}(\theta) = 3T_n(\theta) \]
while
\[ D_i(\theta) (D_i(\theta) - 1) = 2T_{n,i}^*(\theta). \]

Under the condition
\[ \sum_{i=1}^n D_i(\theta) (D_i(\theta) - 1) > 0, \]
the definition of \( C^*(\mathbb{K}(n;\theta)) \) yields
\[ C^*(\mathbb{K}(n;\theta)) = \frac{\sum_{i=1}^n T_{n,i}(\theta)}{\frac{1}{2} \sum_{i=1}^n D_i(\theta) (D_i(\theta) - 1)} = \frac{\sum_{i=1}^n T_{n,i}(\theta)}{\sum_{i=1}^n T_{n,i}^*(\theta)} \]
so that
\[ C^*(\mathbb{K}(n;\theta)) = \frac{3T_n(\theta)}{\sum_{i=1}^n T_{n,i}^*(\theta)} \left[ \sum_{i=1}^n T_{n,i}^*(\theta) > 0 \right]. \]  
(A.10)

The desired conclusion (28) is now immediate from Lemma A.1 and Lemma A.2 established below. They deal with the a.s. convergence of the numerator and denominator (properly normalized) appearing in the ratio (A.10), respectively.

\begin{center}
\textbf{Lemma A.1:} For positive integers \( P \) and \( K \) such that \( K \leq P \), we have
\end{center}
\[ \lim_{n \to \infty} \frac{T_n(\theta)}{\binom{n}{3}} = \beta(\theta) \quad \text{a.s.} \]  
(A.11)

\begin{center}
\textbf{Proof.} Fix \( n = 3, 4, \ldots \) and \( \varepsilon > 0 \). Markov’s inequality already gives
\end{center}
\[ \mathbb{P} \left[ \left| \frac{T_n(\theta)}{\binom{n}{3}} - \beta(\theta) \right| > \varepsilon \right] \leq \varepsilon^{-2} \text{Var} \left[ \frac{T_n(\theta)}{\binom{n}{3}} \right]. \]
as we recall (14). It is now plain from (15) that

\[
\text{Var} \left[ T_n(\theta) \right] = \mathbb{E} \left[ \left( T_n(\theta) - \mathbb{E}[T_n(\theta)] \right)^2 \right] = \mathbb{E}[T_n(\theta)]^2 + 3(n-3) \left( \frac{n}{3} \right) \mathbb{E}[\chi_{123}(\theta)\chi_{124}(\theta)]
\]

with a.s. convergence replaced by convergence in probability.

\[
\beta(\theta) + \left( \frac{n-3}{3} \right) + 3 \left( \frac{n-3}{3} \right)\mathbb{E}[\chi_{123}(\theta)\chi_{124}(\theta)]
\]

+ \frac{3(n-3)}{n} \mathbb{E}[\chi_{123}(\theta)\chi_{124}(\theta)]
\]

(A.12)

as we again make use of the expression (14).

With the help of (68) and (69), it is easy to see that

\[
\lim_{n \to \infty} \text{Var} \left[ T_n(\theta) \right] = 0,
\]

(A.13)

a fact which would readily imply a weaker form of (A.11) with a.s. convergence replaced by convergence in probability. However, elementary algebra on (A.12) shows that (A.13) takes place according to

\[
\lim_{n \to \infty} n^2 \text{Var} \left[ T_n(\theta) \right] = C
\]

with

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

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

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

for some \( C' > C \), and the conclusion (A.11) follows by the Borel-Cantelli Lemma.

Lemma A.2: For positive integers \( P \) and \( K \) such that \( K \leq P \), we have

\[
\lim_{n \to \infty} \sum_{i=1}^{n} T^*_{ni}(\theta) = 3\beta(\theta)^2 \ a.s.
\]

(A.14)

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

\[
T_{n,1}(\theta) = \sum_{j=2}^{n-1} \sum_{k=j+1}^{n} \xi_{ij}(\theta)\xi_{jk}(\theta)
\]

\[
= \Phi_n(\xi_{12}(\theta), \ldots, \xi_{1n}(\theta))
\]

(A.15)

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

\[
\Phi_n(x_2, \ldots, x_n) = \sum_{\ell=2}^{n-1} \sum_{k=\ell+1}^{n} x_\ell x_k
\]

(A.16)

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.,

\[
x_\ell = y_\ell, \quad \ell \neq j.
\]

Under such conditions, it is easy to check that

\[
\left| \Phi_n(x_2, \ldots, x_n) - \Phi_n(y_2, \ldots, y_n) \right| \leq \sum_{\ell=2, \ell \neq j}^{n-1} x_\ell
\]

\[
\leq n - 1.
\]

(A.17)

Recall that the \( (n-1) \) rvs \( \{\xi_{ij}(\theta), j = 2, \ldots, n\} \) are i.i.d. Bernoulli rvs. In view of the constraints (A.17) we can now apply McDiarmid’s inequality [12] (with \( c_j = (n-1) \) for all \( j = 2, \ldots, n-1 \); see also Corollary 2.17 and Remark 2.28 in the monograph [10, p. 38]). Thus, for every \( t > 0 \) we find

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

(A.18)

with

\[
\mathbb{E}[T^*_{n,1}(\theta)] = \sum_{j=2}^{n-1} \sum_{k=j+1}^{n} \mathbb{E}[\xi_{ij}(\theta)\xi_{jk}(\theta)]
\]

\[
= \sum_{j=2}^{n-1} (n-j)p(\theta)^2
\]

\[
= \frac{(n-1)(n-2)}{2} \cdot p(\theta)^2
\]

(A.19)

under the independence noted earlier.

With \( \varepsilon > 0 \) we now substitute

\[
t = \frac{(n-1)(n-2)}{2} \cdot \varepsilon
\]

into (A.18). Since

\[
\frac{2t^2}{(n-1)^3} = \frac{(n-2)^2}{2(n-1)} \cdot \varepsilon^2 \sim \frac{n}{2} \cdot \varepsilon^2,
\]

we obtain from (A.18) and (A.19) that

\[
\mathbb{P} \left[ \left| T^*_{n,1}(\theta) - \mathbb{E}[T^*_{n,1}(\theta)] \right| > \varepsilon \right] \leq 2e^{-\frac{2\varepsilon^2}{(n-1)^3}}.
\]

(A.20)

Since

\[
\left| \sum_{i=1}^{n} T^*_{ni}(\theta) - p(\theta)^2 \right|
\]

\[
= \frac{1}{n(n-1)(n-2)} \left( \sum_{i=1}^{n} T^*_{ni}(\theta) - p(\theta)^2 \right)
\]

\[
\leq \frac{1}{n(n-1)(n-2)} \left| \sum_{i=1}^{n} T^*_{ni}(\theta) - p(\theta)^2 \right|
\]

\[
= \frac{1}{n(n-1)(n-2)} \left| \sum_{i=1}^{n} T^*_{ni}(\theta) - p(\theta)^2 \right|
\]
it is plain that

\[
\Pr \left[ \left| \frac{1}{n(n-1)(n-2)} \sum_{i=1}^{n} T_{n,i}^* - p(\theta)^2 \right| > \varepsilon \right] \\
\leq \Pr \left[ \left| \frac{1}{n \sum_{i=1}^{n} \frac{T_{n,i}^*}{(n-1)(n-2)}} - p(\theta)^2 \right| > \varepsilon \right] \\
\leq \Pr \left[ \bigcup_{i=1}^{n} \left[ \left| \frac{T_{n,i}^*}{(n-1)(n-2)} - p(\theta)^2 \right| > \varepsilon \right] \right] \\
= \sum_{i=1}^{n} \Pr \left[ \left| \frac{T_{n,i}^*}{(n-1)(n-2)} - p(\theta)^2 \right| > \varepsilon \right] \\
= n \Pr \left[ \left| \frac{T_{n,1}^*}{(n-1)(n-2)} - p(\theta)^2 \right| > \varepsilon \right]
\]

where the equality before last follows by a union bound argument and the last equality is a consequence of exchangeability.

Invoking (A.20) (with \( \varepsilon \) instead of \( \varepsilon \)) we get

\[
\Pr \left[ \left| \frac{1}{n \binom{n}{3}} \sum_{i=1}^{n} T_{n,i}^* - 3p(\theta)^2 \right| > \varepsilon \right] \leq 2ne^{-n \frac{\theta^2}{4}(1+o(1))}\varepsilon^2
\]

with

\[
\sum_{n=3}^{\infty} ne^{-n \frac{\theta^2}{4}(1+o(1))}\varepsilon^2 < \infty
\]

for every \( \varepsilon > 0 \). The a.s. convergence (A.14) now follows by the Borel-Cantelli Lemma. \( \blacksquare \)