

THE INSTITUTE FOR SYSTEMS RESEARCH

ISR TECHNICAL REPORT 2009-10

On the existence of triangles in random key graphs

Osman Yagan, Armand M. Makowski

The
Institute for
Systems
Research



A. JAMES CLARK
SCHOOL OF ENGINEERING

ISR develops, applies and teaches advanced methodologies of design and analysis to solve complex, hierarchical, heterogeneous and dynamic problems of engineering technology and systems for industry and government.

ISR is a permanent institute of the University of Maryland, within the A. James Clark School of Engineering. It is a graduated National Science Foundation Engineering Research Center.

www.isr.umd.edu

On the existence of triangles in random key graphs

Osman Yağın and Armand M. Makowski

oyagan@umd.edu, armand@isr.umd.edu

Department of Electrical and Computer Engineering
and Institute for Systems Research
University of Maryland, College Park, MD 20742.

July 3, 2009

Abstract

The random key graph, also known as the *uniform random intersection graph*, is a random graph induced by the random key predistribution scheme of Eschenauer and Gligor under the assumption of full visibility. We show the existence of a zero-one law for the appearance of triangles in random key graphs by applying the method of first and second moments to the number of triangles in the graph.

1 Introduction

Eschenauer and Gligor [4] have recently proposed the following random key predistribution scheme for wireless sensor networks (WSNs): Before network deployment, each sensor is independently assigned K distinct cryptographic keys which are selected at random from a pool of P keys. These K keys constitute the key ring of the node and are inserted into its memory. 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 [4] for implementation details.

Under the assumption of *full visibility*, namely that nodes are all within communication range of each other, two nodes can communicate securely if their key rings share at least one key. This notion of adjacency induces the *random key graph* $\mathbb{K}(n; (K, P))$ on the vertex set $\{1, \dots, n\}$ where n is the number of sensor nodes; see Section 2 for precise definitions.

In search of an indication for the feasibility of EG scheme in the context of WSNs, most of the research so far has focused on the connectivity properties of random key graphs. Setting aside the fact that random key graphs are not equivalent to Erdős-Renyi graphs [3], Eschenauer and Gligor [4] transferred the

well-known results for connectivity of Erdős-Rényi graphs to random key graphs by matching the corresponding graphs by identical link assignment probabilities. The validity of this transfer was later established at the cost of increased technicalities in [1], [2], [13]. In light of this success, it is natural to wonder whether this transfer from Erdős-Rényi graphs to random key graphs applies more generally to other graph properties.

Interestingly enough, random key graphs have appeared in application areas as diverse as clustering analysis [6], [7] and recommender systems [10], and their study is therefore of interest beyond the context of WSNs. The recent results in [1], [2], [13] settling the graph connectivity problem, we shift the focus in this paper towards other properties of random key graphs by considering the subgraph containment problem.

In the literature of random graphs, it has been of interest [3], [9] to look for thresholds for the containment of certain subgraphs, the most popular and simplest one being the *triangles*. Given the fact that the number of triangles in the graph is related to the clustering properties, this problem also bears practical importance.

We manage to show that random key graphs admit a *zero-one law* for the existence of triangles and we identify the corresponding critical thresholds. As we compare this threshold with the one obtained for Erdős-Rényi graphs, we see that the random key graphs evolve in a way such that the triangles start to appear earlier than in the case of Erdős-Rényi graphs. In fact, for the parameter range that is practically interesting in the context of WSNs, the threshold obtained for random key graphs turns out to be *much smaller* than the threshold for Erdős-Rényi graphs. Also, in that range it is easy to conclude from the results of this paper that the expected number of triangles is *much larger* in random key graphs compared to Erdős-Rényi graphs; a fact that was also observed in [2] via *simulations*. This prompts us to conclude that transferring the results for Erdős-Rényi graphs to random key graphs by matching them with identical link assignment probabilities can be quite *misleading* in some cases.

The paper is organized as follows: In Section 2 we formally introduce the class of random key graphs while in Section 3 we present the main results of the paper summarized as Theorem 3.1 and Theorem 3.2. These results are then used in Section 4 to compare random key graphs with Erdős-Rényi graphs in terms of their behavior for triangle appearance. In Section 5, we compute the expected value of the number of triangles in random key graphs and the asymptotic results that will be used in the proofs of the main results are collected in Section 6. In Section 7.1, we give a proof of the zero law (Theorem 3.1) while an outline for the proof of the one law (Theorem 3.2) is given in Section 7.2. Final parts of the paper (Sections 8 through 11) are devoted to completing the proof of Theorem 3.2.

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 triple $(\Omega, \mathcal{F}, \mathbb{P})$. 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 $\mathbf{1}[E]$. For any discrete set S we write $|S|$ for its cardinality.

2 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 with $K \leq P$. We often group the integers P and K into the ordered pair $\theta \equiv (P, K)$ in order to simplify the notation. Now, for each node $i = 1, \dots, n$, let $K_i(\theta)$ denote the random set of K distinct keys assigned to node i and let \mathcal{P} be the set of all keys. The rvs $K_1(\theta), \dots, K_n(\theta)$ are assumed to be *i.i.d.* rvs, each of which is *uniformly* distributed with

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

for any subset S of \mathcal{P} which contains exactly K elements. This corresponds to selecting keys randomly and *without* replacement from the key pool.

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

$$K_i(\theta) \cap K_j(\theta) \neq \emptyset, \quad (2)$$

in which case an undirected link is assigned between nodes i and j . The resulting random graph defines the *random key graph* on the vertex set $\{1, \dots, n\}$, hereafter denoted $\mathbb{K}(n; \theta)$.

For distinct $i, j = 1, \dots, n$, it is easy to check that

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

with

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

whence the probability of edge occurrence between any two nodes is equal to $1 - q(\theta)$. The expression given in (4) is a simple consequence of the often used fact that

$$\mathbb{P}[S \cap K_i(\theta) = \emptyset] = \frac{\binom{P-|S|}{K}}{\binom{P}{K}}, \quad i = 1, \dots, n \quad (5)$$

for every subset S of $\{1, \dots, P\}$ with $|S| \leq P - K$. Note that if $P < 2K$ there exists an edge between any pair of nodes, so that $\mathbb{K}(n; \theta)$ coincides with the complete graph K_n . Also, we always have $0 \leq q(\theta) < 1$ with $q(\theta) > 0$ if and only if $2K \leq P$.

For simplicity of exposition we refer to any pair of functions $P, K : \mathbb{N}_0 \rightarrow \mathbb{N}_0$ as a *scaling* provided the natural conditions

$$K_n \leq P_n, \quad n = 2, 3, \dots \quad (6)$$

are satisfied.

3 The main result

Pick positive integers K and P such that $K \leq P$. Fix $n = 3, 4, \dots$ and for distinct $i, j, k = 1, \dots, n$, define the indicator function

$$\chi_{n,ijk}(\theta) := \mathbf{1}[\text{Nodes } i, j \text{ and } k \text{ form a triangle in } \mathbb{K}(n; \theta)].$$

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

$$T_n(\theta) := \sum_{(ijk)} \chi_{n,ijk}(\theta) \tag{7}$$

where $\sum_{(ijk)}$ denotes summation over all distinct triples ijk with $1 \leq i < j < k \leq n$. The event $T(n, \theta)$ that there exists at least one triangle in $\mathbb{K}(n; \theta)$ is then characterized by

$$T(n, \theta) := [T_n(\theta) > 0] = [T_n(\theta) = 0]^c. \tag{8}$$

The main result of the paper is a zero-one law for the existence of triangles in random key graphs. To state the results we find it convenient to make use of the quantity

$$\tau(\theta) := \frac{K^3}{P^2} + \left(\frac{K^2}{P}\right)^3. \tag{9}$$

The zero law is given first.

Theorem 3.1 *For any scaling $P, K : \mathbb{N}_0 \rightarrow \mathbb{N}_0$, we have the zero law*

$$\lim_{n \rightarrow \infty} \mathbb{P}[T(n, \theta_n)] = 0 \tag{10}$$

under the condition

$$\lim_{n \rightarrow \infty} n^3 \tau(\theta_n) = 0. \tag{11}$$

The one law given next assumes a more involved form.

Theorem 3.2 *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, we have the one law*

$$\lim_{n \rightarrow \infty} \mathbb{P}[T(n, \theta_n)] = 1 \tag{12}$$

either if $0 \leq q^ < 1$ or if $q^* = 1$ under the additional condition*

$$\lim_{n \rightarrow \infty} n^3 \tau(\theta_n) = \infty. \tag{13}$$

Theorem 3.1 and Theorem 3.2 will be established by the method of first and second moments, respectively [8, p. 55]. The first step in doing so is taken in Section 5 by computing the first moment of the count variables defined at (7). In what follows we compare the random key graph to a related Erdős-Rényi graph [3] in terms of their behavior for the existence of triangles.

4 Comparison with Erdős-Rényi graphs

For each p in $[0, 1]$ let $\mathbb{G}(n; p)$ denote the Erdős-Rényi graph on the vertex set $\{1, \dots, n\}$ with link assignment probability p and let $T(n; p)$ denote the event that there is at least one triangle in $\mathbb{G}(n; p)$. As we set

$$\tau^*(p) = p^3, \quad (14)$$

the (well-known) zero-one law for the existence of a triangle in Erdős-Rényi graphs takes the following form:

Theorem 4.1 *For any scaling $p : \mathbb{N}_0 \rightarrow [0, 1]$, we have*

$$\lim_{n \rightarrow \infty} \mathbb{P}[T(n; p_n)] = \begin{cases} 0 & \text{if } \lim_{n \rightarrow \infty} n^3 \tau^*(p_n) = 0 \\ 1 & \text{if } \lim_{n \rightarrow \infty} n^3 \tau^*(p_n) = \infty. \end{cases} \quad (15)$$

It has been noted in [1], [2], [13] that p_n is analogous to $\frac{K_n^2}{P_n}$ under certain conditions as both expressions give the probability of link assignment; see Section 6 for details. Following this, we refer to an Erdős-Rényi graph $\mathbb{G}(n; p)$ as being *matched* to a random key graph $\mathbb{K}(n; \theta)$ if $p_n \sim \frac{K_n^2}{P_n}$. In that case, we have

$$\frac{\tau(\theta_n)}{\tau^*(p_n)} \sim 1 + \frac{P_n}{K_n^3}, \quad (16)$$

whence

$$\lim_{n \rightarrow \infty} \frac{\tau(\theta_n)}{\tau^*(p_n)} = \begin{cases} 1 & \text{if } \lim_{n \rightarrow \infty} \frac{P_n}{K_n^3} = 0 \\ \infty & \text{if } \lim_{n \rightarrow \infty} \frac{P_n}{K_n^3} = \infty. \end{cases} \quad (17)$$

This suggests that the threshold ensuring the existence of a triangle is always reached earlier in the evolution of a random key graph compared to the case of an Erdős-Rényi graph matched to it.

In the context of WSNs, it is desired to select the EG scheme parameters K_n and P_n such that the induced random key graph is *connected*. In fact, considering the *tradeoff* between connectivity and security [2], it is desirable to keep $\frac{K_n^2}{P_n}$ as close as possible to the connectivity threshold $\frac{\log n}{n}$; see [1], [2] and [13] for details. Therefore, in the case

$$\frac{K_n^2}{P_n} \sim \frac{\log n}{n} \quad (18)$$

we have

$$\frac{\tau(\theta_n)}{\tau^*(p_n)} \sim 1 \quad (19)$$

only if

$$K_n \gg \frac{n}{\log n}. \quad (20)$$

Given the limited memory and computational power of the sensor nodes, such key ring sizes are obviously *not* practical. Also they will lead to *high* node degrees which in turn will decrease the *resiliency* of the network against node capture attacks. Indeed, the criterion to ensure *security* in a WSN was suggested in [2] as selecting K_n and P_n such that

$$\frac{K_n}{P_n} \sim \frac{1}{n}$$

which then leads to

$$K_n \sim \log n \tag{21}$$

via (18). With (18) and (21) in effect, we find

$$\frac{\tau(\theta_n)}{\tau^*(p_n)} \sim 1 + \frac{n}{(\log n)^2} \rightarrow \infty. \tag{22}$$

Hence, for practical WSN scenarios the induced random key graph has a much smaller threshold for triangle appearance than the matched Erdős-Rényi graph. Therefore, transferring the well-known results for Erdős-Rényi graphs to random key graphs by matching them with identical edge assignment probabilities might be *misleading* for some graph properties. This shows that Erdős-Rényi graphs are not adequate in modelling the EG scheme, calling instead for a direct investigation of random key graphs in this context!

5 Computing the first moment

With positive integers K and P such that $K \leq P$, define

$$\beta(\theta) := (1 - q(\theta))^3 + q(\theta)^3 - q(\theta)r(\theta) \tag{23}$$

where we have set

$$r(\theta) := \begin{cases} 0 & \text{if } P < 3K \\ \frac{\binom{P-2K}{K}}{\binom{P}{K}} & \text{if } 3K \leq P. \end{cases} \tag{24}$$

Direct inspection shows that

$$r(\theta) \leq q(\theta)^2 \tag{25}$$

whence

$$\beta(\theta) \geq (1 - q(\theta))^3 > 0. \tag{26}$$

Lemma 5.1 *For positive integers K and P such that $K \leq P$, we have*

$$\mathbb{E}[T_n(\theta)] = \binom{n}{3} \beta(\theta), \quad n = 3, 4, \dots \tag{27}$$

To help deriving (27) we introduce the events

$$A(\theta) := [K_1(\theta) \cap K_2(\theta) \neq \emptyset] \cap [K_1(\theta) \cap K_3(\theta) \neq \emptyset] \quad (28)$$

and

$$\begin{aligned} B(\theta) &:= [K_1(\theta) \cap K_2(\theta) \neq \emptyset] \cap [K_1(\theta) \cap K_3(\theta) \neq \emptyset] \cap [K_2(\theta) \cap K_3(\theta) \neq \emptyset] \\ &= A(\theta) \cap [K_2(\theta) \cap K_3(\theta) \neq \emptyset]. \end{aligned} \quad (29)$$

The event $A(\theta)$ captures the existence of edges between node 1 and the pair of nodes 2 and 3, respectively, in $\mathbb{K}(n; \theta)$, while $B(\theta)$ is the event where the nodes 1, 2 and 3 form a triangle in $\mathbb{K}(n; \theta)$.

Lemma 5.2 *The probability of the event $A(\theta)$ is given by*

$$\mathbb{P}[A(\theta)] = (1 - q(\theta))^2. \quad (30)$$

In the proof of Lemma 5.2 (as well as in other proofs) we omit the explicit dependence on θ when no confusion arises from doing so.

Proof. Under the enforced independence assumptions we note that

$$\begin{aligned} \mathbb{P}[A(\theta)] &= \sum_{|S|=K} \mathbb{P}[K_1 = S, S \cap K_2 \neq \emptyset, S \cap K_3 \neq \emptyset] \\ &= \sum_{|S|=K} \mathbb{P}[K_1 = S] \mathbb{P}[S \cap K_2 \neq \emptyset] \mathbb{P}[S \cap K_3 \neq \emptyset] \\ &= (1 - q(\theta))^2 \end{aligned} \quad (31)$$

as we make use of (5) with $\sum_{|S|=K} \mathbb{P}[K_1 = S] = 1$. ■

In many of the forthcoming calculations we make repeated use of the fact that for any pair of events, say E and F , we have

$$\mathbb{P}[E \cap F] = \mathbb{P}[E] - \mathbb{P}[E \cap F^c]. \quad (32)$$

In particular, we can now conclude from Lemma 5.2 that

$$\begin{aligned} &\mathbb{P}[K_1(\theta) \cap K_2(\theta) = \emptyset, K_1(\theta) \cap K_3(\theta) \neq \emptyset] \\ &= \mathbb{P}[K_1(\theta) \cap K_2(\theta) \neq \emptyset, K_1(\theta) \cap K_3(\theta) = \emptyset] \\ &= q(\theta)(1 - q(\theta)) \end{aligned} \quad (33)$$

and

$$\mathbb{P}[K_1(\theta) \cap K_2(\theta) = \emptyset, K_1(\theta) \cap K_3(\theta) = \emptyset] = q(\theta)^2. \quad (34)$$

These facts will be used in computing the probability of

Lemma 5.3 With $\beta(\theta)$ given at (23) we have

$$\mathbb{P}[B(\theta)] = \beta(\theta). \quad (35)$$

Proof. Repeated use of (32) yields

$$\begin{aligned} \mathbb{P}[B(\theta)] &= \mathbb{P}[K_1 \cap K_2 \neq \emptyset, K_1 \cap K_3 \neq \emptyset] \\ &\quad - \mathbb{P}[K_1 \cap K_2 \neq \emptyset, K_1 \cap K_3 \neq \emptyset, K_2 \cap K_3 = \emptyset] \\ &= \mathbb{P}[A(\theta)] - \mathbb{P}[K_1 \cap K_2 \neq \emptyset, K_2 \cap K_3 = \emptyset] \\ &\quad + \mathbb{P}[K_1 \cap K_2 \neq \emptyset, K_1 \cap K_3 = \emptyset, K_2 \cap K_3 = \emptyset] \\ &= (1 - q(\theta))^2 - q(\theta)(1 - q(\theta)) + \mathbb{P}[K_1 \cap K_3 = \emptyset, K_2 \cap K_3 = \emptyset] \\ &\quad - \mathbb{P}[K_1 \cap K_2 = \emptyset, K_1 \cap K_3 = \emptyset, K_2 \cap K_3 = \emptyset] \\ &= (1 - q(\theta))^2 - q(\theta)(1 - q(\theta)) + q(\theta)^2 \\ &\quad - \mathbb{P}[K_1 \cap K_2 = \emptyset, K_1 \cap K_3 = \emptyset, K_2 \cap K_3 = \emptyset] \end{aligned} \quad (36)$$

as we recall (30), (33) and (34).

By independence we get

$$\begin{aligned} &\mathbb{P}[K_1 \cap K_2 = \emptyset, K_1 \cap K_3 = \emptyset, K_2 \cap K_3 = \emptyset] \\ &= \mathbb{P}[K_1 \cap K_2 = \emptyset, (K_1 \cup K_2) \cap K_3 = \emptyset] \\ &= \sum_{|S|=|T|=K, S \cap T = \emptyset} \mathbb{P}[K_1 = S, K_2 = T] \mathbb{P}[(S \cup T) \cap K_3 = \emptyset] \\ &= \sum_{|S|=|T|=K, S \cap T = \emptyset} \mathbb{P}[K_1 = S, K_2 = T] \cdot r(\theta) \\ &= \mathbb{P}[K_1 \cap K_2 = \emptyset] \cdot r(\theta) \end{aligned} \quad (37)$$

by invoking (5) (since $|S \cup T| = 2K$ under the constraints $|S| = |T| = K$ and $S \cap T = \emptyset$). Thus,

$$\mathbb{P}[B(\theta)] = (1 - q(\theta))^2 - q(\theta)(1 - q(\theta)) + q(\theta)^2 - q(\theta)r(\theta),$$

and the desired result follows upon noting the relation

$$(1 - q(\theta))^2 - q(\theta)(1 - q(\theta)) + q(\theta)^2 = (1 - q(\theta))^3 + q(\theta)^3.$$

■

The proof of Lemma 5.1 is now straightforward: Fix $n = 3, 4, \dots$. Exchangeability yields

$$\mathbb{E}[T_n(\theta)] = \binom{n}{3} \mathbb{E}[\chi_{n,123}(\theta)] \quad (38)$$

and the desired conclusion follows as we make use of Lemma 5.3.

6 Some useful asymptotics

In this section we collect a number of asymptotic results that prove useful in establishing some of the results derived in this paper. The first result, already obtained in [13], will be key to our approach.

Lemma 6.1 *For any scaling $P, K : \mathbb{N}_0 \rightarrow \mathbb{N}_0$, we have*

$$\lim_{n \rightarrow \infty} q(\theta_n) = 1 \quad (39)$$

if and only if

$$\lim_{n \rightarrow \infty} \frac{K_n^2}{P_n} = 0, \quad (40)$$

and under either condition the asymptotic equivalence

$$1 - q(\theta_n) \sim \frac{K_n^2}{P_n} \quad (41)$$

holds.

Since $1 \leq K_n \leq K_n^2$ for all $n = 1, 2, \dots$, the condition (40) implies

$$\lim_{n \rightarrow \infty} \frac{K_n}{P_n} = 0 \quad (42)$$

and

$$\lim_{n \rightarrow \infty} P_n = \infty. \quad (43)$$

so that for any $c > 0$, we have

$$cK_n \leq P_n \quad (44)$$

for all $n \in \mathbb{N}_0$ sufficiently large (dependent on c).

The proof of Lemma 6.1 is based on the following elementary bounds, whose proofs are available in [13],

Lemma 6.2 *For positive integers K, L and P such that $K + L \leq P$, we have*

$$\left(1 - \frac{L}{P - K}\right)^K \leq \frac{\binom{P-L}{K}}{\binom{P}{K}} \leq \left(1 - \frac{L}{P}\right)^K. \quad (45)$$

These bounds also form the basis for deriving the following asymptotic equivalence.

Proposition 6.3 *For any scaling $P, K : \mathbb{N}_0 \rightarrow \mathbb{N}_0$ satisfying (39)-(40), we have the asymptotic equivalence*

$$\beta(\theta_n) \sim \tau(\theta_n) \quad (46)$$

provided (13) holds.

Proof. From (23), we get

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

Under the enforced assumptions Lemma 6.1 already implies

$$(1 - q(\theta_n))^3 \sim \left(\frac{K_n^2}{P_n}\right)^3,$$

and

$$q(\theta_n)^3 \sim 1.$$

It is now plain that the equivalence (46) will hold if we show that

$$1 - \frac{r(\theta_n)}{q(\theta_n)^2} \sim \frac{K_n^3}{P_n^2}. \quad (47)$$

This key technical fact is established in Appendix A. ■

The final result of this section also relies on Lemma 6.1, and will prove useful in establishing the one law.

Proposition 6.4 *For any scaling $P, K : \mathbb{N}_0 \rightarrow \mathbb{N}_0$ satisfying (39)-(40), we have*

$$\lim_{n \rightarrow \infty} n^2(1 - q(\theta_n)) = \infty \quad (48)$$

provided the condition (13) holds.

Proof. Consider a scaling $P, K : \mathbb{N}_0 \rightarrow \mathbb{N}_0$ satisfying (39)-(40). By Lemma 6.1 the desired conclusion (48) will be established if we show

$$\lim_{n \rightarrow \infty} n^2 \frac{K_n^2}{P_n} = \infty. \quad (49)$$

As condition (13) reads

$$\lim_{n \rightarrow \infty} n^3 \left(\frac{K_n^3}{P_n^2} + \left(\frac{K_n^2}{P_n}\right)^3 \right) = \infty,$$

we immediately get (49) by virtue of the trivial bounds

$$n^3 \left(\frac{K_n^2}{P_n}\right)^3 = \left(\frac{nK_n^2}{P_n}\right)^3 \leq \left(\frac{n^2K_n^2}{P_n}\right)^3$$

and

$$n^3 \frac{K_n^3}{P_n^2} \leq n^4 \frac{K_n^4}{P_n^2} = \left(\frac{n^2K_n^2}{P_n}\right)^2$$

valid for all $n = 1, 2, \dots$ ■

Proposition 6.4 will be used as follows: Pick $a > 0$ and $b > 0$, and consider a scaling $P, K : \mathbb{N}_0 \rightarrow \mathbb{N}_0$ satisfying (39)-(40). For each $n = 2, 3, \dots$, we get

$$\begin{aligned} \frac{1}{n^2} \cdot \frac{(1 - q(\theta_n))^a}{\beta(\theta_n)^b} &\leq \frac{1}{n^2} \cdot \frac{(1 - q(\theta_n))^a}{(1 - q(\theta_n))^{3b}} \\ &= \frac{1}{n^2 (1 - q(\theta_n))} \cdot (1 - q(\theta_n))^{a-3b+1}. \end{aligned} \quad (50)$$

Therefore, under condition (13) Proposition 6.4 yields

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \cdot \frac{(1 - q(\theta_n))^a}{\beta(\theta_n)^b} = 0 \quad \text{if } a - 3b + 1 \geq 0 \quad (51)$$

as we make use of (39)-(40).

7 Proofs of Theorem 3.1 and Theorem 3.2

7.1 A proof of Theorem 3.1

Fix $n = 3, 4, \dots$. An elementary bound for \mathbb{N} -valued rvs yields

$$\mathbb{P}[T_n(\theta_n) > 0] \leq \mathbb{E}[T_n(\theta_n)], \quad (52)$$

so that

$$\mathbb{P}[T(n, \theta_n)] \leq \binom{n}{3} \beta(\theta_n). \quad (53)$$

The conclusion (10) follows if we show that

$$\lim_{n \rightarrow \infty} \binom{n}{3} \beta(\theta_n) = 0 \quad (54)$$

under (11).

The condition $\lim_{n \rightarrow \infty} n^3 \tau(\theta_n) = 0$ implies $\lim_{n \rightarrow \infty} \tau(\theta_n) = 0$ and (40) automatically holds. By Proposition 6.3 we conclude $\beta(\theta_n) \sim \tau(\theta_n)$, whence $n^3 \beta(\theta_n) \sim n^3 \tau(\theta_n)$, and condition (11) is indeed equivalent to (54) since $\binom{n}{3} \sim \frac{n^3}{6}$.

7.2 A proof of Theorem 3.2

Assume first that q^* satisfies $0 \leq q^* < 1$. Fix $n = 3, 4, \dots$ and partition the n nodes into the $k_n + 1$ non-overlapping groups $(1, 2, 3), (4, 5, 6), \dots, (3k_n +$

$1, 3k_n + 2, 3k_n + 3$ 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

$$\begin{aligned} & \mathbb{P}[T_n(\theta_n) = 0] \\ & \leq \mathbb{P} \left[\bigcap_{\ell=0}^{k_n} \left[\begin{array}{c} \text{Nodes } 3\ell + 1, 3\ell + 2, 3\ell + 3 \text{ do not form} \\ \text{a triangle in } \mathbb{K}(n; \theta_n) \end{array} \right] \right] \\ & = \prod_{\ell=0}^{k_n} \mathbb{P} \left[\begin{array}{c} \text{Nodes } 3\ell + 1, 3\ell + 2, 3\ell + 3 \text{ do not form} \\ \text{a triangle in } \mathbb{K}(n; \theta_n) \end{array} \right] \end{aligned} \quad (55)$$

$$\begin{aligned} & = (1 - \beta(\theta_n))^{k_n+1} \\ & \leq (1 - (1 - q(\theta_n))^3)^{k_n+1} \end{aligned} \quad (56)$$

$$\leq e^{-(k_n+1)(1-q(\theta_n))^3}. \quad (57)$$

Note that (55) follows from the fact that the events

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

are mutually independent due to the non-overlap condition, while the inequality (56) is justified with the help of (26). Let n go to infinity in the inequality (57). From the constraint $q^* < 1$ we conclude that $\lim_{n \rightarrow \infty} \mathbb{P}[T(n, \theta_n)^c] = 0$ since $k_n \sim \frac{n}{3}$ so that $\lim_{n \rightarrow \infty} (k_n + 1)(1 - q(\theta_n))^3 = \infty$. This establishes (12).

To handle the case $q^* = 1$, we use a standard bound which forms the basis of the method of second moment [8, remark 3.1, p. 55]. Here it takes the form

$$\frac{\mathbb{E}[T_n(\theta_n)]^2}{\mathbb{E}[T_n(\theta_n)^2]} \leq \mathbb{P}[T_n(\theta_n) > 0], \quad n = 2, 3, \dots \quad (58)$$

It is now plain that (12) will be established in the case $q^* = 1$ if we show the following result.

Proposition 7.1 *For any scaling $P, K : \mathbb{N}_0 \rightarrow \mathbb{N}_0$ satisfying (39)-(40), we have*

$$\lim_{n \rightarrow \infty} \frac{\mathbb{E}[T_n(\theta_n)^2]}{\mathbb{E}[T_n(\theta_n)]^2} = 1 \quad (59)$$

under the condition (13).

The remainder of the paper is devoted to establishing Proposition 7.1. As will soon become apparent this is a bit quite more involved than expected.

8 Computing the second moment

A natural step towards establishing Proposition 7.1 consists in computing the second moment of the count variables (7).

Proposition 8.1 For positive integers K and P such that $K \leq P$, we have

$$\begin{aligned} \mathbb{E} [T_n(\theta)^2] &= \mathbb{E} [T_n(\theta)] + \left(\frac{\binom{n-3}{3}}{\binom{n}{3}} + 3 \frac{\binom{n-3}{2}}{\binom{n}{3}} \right) \cdot \mathbb{E} [T_n(\theta)]^2 \\ &+ \binom{n}{3} \binom{3}{2} \binom{n-3}{1} \cdot \mathbb{E} [\chi_{n,123}(\theta) \chi_{n,124}(\theta)] \end{aligned} \quad (60)$$

for all $n = 3, 4, \dots$ with

$$\begin{aligned} \mathbb{E} [\chi_{n,123}(\theta) \chi_{n,124}(\theta)] &= -(1 - q(\theta))^5 + 2(1 - q(\theta))^2 \beta(\theta) \\ &- \frac{1}{q(\theta)} (\beta(\theta) - (1 - q(\theta))^3)^2 + \sum_{k=0}^K c_k(\theta) - q(\theta)^4 \end{aligned} \quad (61)$$

where we have set

$$c_k(\theta) := \frac{\binom{K}{k} \binom{P-K}{K-k}}{\binom{P}{K}} \cdot \left(\frac{\binom{P-2K+k}{K}}{\binom{P}{K}} \right)^2, \quad k = 0, 1, \dots, K. \quad (62)$$

As explained in Appendix B we have the probabilistic interpretation

$$c_k(\theta) = \mathbb{P} [|K_1(\theta) \cap K_2(\theta)| = k, (K_1(\theta) \cup K_2(\theta)) \cap K_i(\theta) = \emptyset, i = 3, 4] \quad (63)$$

for each $k = 0, 1, \dots, K$.

Proof. Consider positive integers K and P such that $K \leq P$ and fix $n = 3, 4, \dots$. By exchangeability and by the binary nature of the rvs involved we readily conclude that

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

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

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

so that

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

as we make use of the relation (27).

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

$$\mathbb{P} [\chi_{n,123}(\theta) = 1 | K_1(\theta) = S] = \mathbb{P} [\chi_{n,123}(\theta) = 1] = \beta(\theta), \quad S \in \mathcal{P}_K.$$

A similar statement applies to $\chi_{n,145}(\theta)$ and therefore the rvs $\chi_{n,123}(\theta)$ and $\chi_{n,145}(\theta)$ are independent and identically distributed so that

$$\mathbb{E} [\chi_{n,123}(\theta) \chi_{n,145}(\theta)] = \mathbb{E} [\chi_{n,123}(\theta)] \mathbb{E} [\chi_{n,145}(\theta)].$$

As before this last observation yields

$$\binom{n}{3} \binom{3}{1} \binom{n-3}{2} \mathbb{E} [\chi_{n,123}(\theta) \chi_{n,145}(\theta)] = 3 \frac{\binom{n-3}{2}}{\binom{n}{3}} \cdot \mathbb{E} [T_n(\theta)]^2 \quad (66)$$

bu virtue of (27).

The evaluation (61)–(62) of the moment $\mathbb{E} [\chi_{n,123}(\theta) \chi_{n,124}(\theta)]$ is rather lengthy, although quite straightforward; details are given in Appendix B. Reporting (61)–(62), (65) and (66) into (64) establishes Proposition 8.1. ■

In preparation of the proof of Proposition 7.1 we note that Proposition 8.1 readily implies

$$\begin{aligned} \frac{\mathbb{E} [T_n(\theta)^2]}{\mathbb{E} [T_n(\theta)]^2} &= \frac{1}{\mathbb{E} [T_n(\theta)]} + \left(\frac{\binom{n-3}{3}}{\binom{n}{3}} + 3 \frac{\binom{n-3}{2}}{\binom{n}{3}} \right) \\ &+ \frac{3(n-3)}{\binom{n}{3}} \cdot \frac{\mathbb{E} [\chi_{n,123}(\theta) \chi_{n,124}(\theta)]}{\mathbb{E} [\chi_{n,123}(\theta)]^2} \end{aligned} \quad (67)$$

for all $n = 2, 3, \dots$ as we make use of (38).

9 A proof of Proposition 7.1

Consider any scaling $P, K : \mathbb{N}_0 \rightarrow \mathbb{N}_0$ satisfying (39)–(40). By Proposition 6.3 we have $\lim_{n \rightarrow \infty} n^3 \beta(\theta_n) = \infty$ under the additional condition (13), whence

$$\lim_{n \rightarrow \infty} \mathbb{E} [T_n(\theta_n)] = \infty$$

by virtue of (38).

As pointed out earlier the equivalent conditions (39)–(40) imply

$$3K_n \leq P_n \quad (68)$$

for all $n \in \mathbb{N}_0$ sufficiently large. On that range (67) is valid with θ replaced by θ_n . Letting n go to infinity in the resulting expression, we note that

$$\lim_{n \rightarrow \infty} \left(\frac{\binom{n-3}{3}}{\binom{n}{3}} + 3 \frac{\binom{n-3}{2}}{\binom{n}{3}} \right) = 1 \quad \text{and} \quad \frac{\binom{n}{3}}{3(n-3)} \sim \frac{n^2}{18}.$$

It is plain that the convergence (59) will hold if we show that

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \frac{\mathbb{E} [\chi_{n,123}(\theta_n) \chi_{n,124}(\theta_n)]}{\mathbb{E} [\chi_{n,123}(\theta_n)]^2} = 0. \quad (69)$$

In order to establish (69) under the assumptions of Proposition 7.1 we proceed as follows: Recall from Lemma 5.1 that

$$\mathbb{E} [\chi_{n,123}(\theta_n)]^2 = \beta(\theta_n)^2 \geq (1 - q(\theta_n))^6, \quad (70)$$

and from (61) observe that

$$\begin{aligned} & \frac{1}{n^2} \cdot \frac{\mathbb{E} [\chi_{n,123}(\theta_n) \chi_{n,124}(\theta_n)]}{(\mathbb{E} [\chi_{n,123}(\theta_n)])^2} \\ &= -\frac{1}{n^2} \cdot \frac{(1 - q(\theta_n))^5}{\beta(\theta_n)^2} + \frac{2}{n^2} \cdot \frac{(1 - q(\theta_n))^2}{\beta(\theta_n)} \\ & \quad - \frac{1}{n^2} \cdot \frac{1}{q(\theta_n)} \left(\frac{\beta(\theta_n) - (1 - q(\theta_n))^3}{\beta(\theta_n)} \right)^2 \\ & \quad + \frac{1}{n^2} \cdot \frac{\sum_{k=0}^{K_n} c_k(\theta_n) - q(\theta_n)^4}{\beta(\theta_n)^2} \end{aligned} \quad (71)$$

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

Let n go to infinity in (71). Using (51) (once with $a = 5$ and $b = 2$, then with $a = 2$ and $b = 1$), we get

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \cdot \frac{(1 - q(\theta_n))^5}{\beta(\theta_n)^2} = 0 \quad (72)$$

and

$$\lim_{n \rightarrow \infty} \frac{2}{n^2} \cdot \frac{(1 - q(\theta_n))^2}{\beta(\theta_n)} = 0. \quad (73)$$

The convergence

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \cdot \frac{1}{q(\theta_n)} \left(\frac{\beta(\theta_n) - (1 - q(\theta_n))^3}{\beta(\theta_n)} \right)^2 = 0 \quad (74)$$

is immediate since

$$\left| \frac{\beta(\theta_n) - (1 - q(\theta_n))^3}{\beta(\theta_n)} \right|^2 \leq 1, \quad n = 2, 3, \dots$$

and $\lim_{n \rightarrow \infty} q(\theta_n) = 1$. Consequently the proof of Proposition 7.1 will be completed if we show

Proposition 9.1 For any scaling $P, K : \mathbb{N}_0 \rightarrow \mathbb{N}_0$ satisfying (39)-(40), we have

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \cdot \frac{\sum_{k=0}^K c_k(\theta_n) - q(\theta_n)^4}{\beta(\theta_n)^2} = 0 \quad (75)$$

under the condition (13).

The proof of Proposition 9.1 will proceed in two steps:

10 A reduction step

We start with an easy bound.

Lemma 10.1 With positive integers K and P such that $2K \leq P$, we have

$$c_1(\theta) \leq 1 - q(\theta). \quad (76)$$

Proof. Specializing (63) with $k = 1$ we get

$$\begin{aligned} c_1(\theta) &= \mathbb{P}[|K_1(\theta) \cap K_2(\theta)| = 1, (K_1(\theta) \cup K_2(\theta)) \cap K_i(\theta) = \emptyset, i = 3, 4] \\ &\leq \mathbb{P}[|K_1(\theta) \cap K_2(\theta)| = 1] \\ &\leq \mathbb{P}[|K_1(\theta) \cap K_2(\theta)| \geq 1] \end{aligned}$$

and the conclusion is immediate as we identify

$$\mathbb{P}[|K_1(\theta) \cap K_2(\theta)| \geq 1] = \mathbb{P}[K_1(\theta) \cap K_1(\theta) \neq \emptyset] = 1 - q(\theta). \quad \blacksquare$$

Lemma 10.2 With positive integers K and P such that $3K \leq P$, the monotonicity property

$$\frac{c_1(\theta)}{c_0(\theta)} \geq \frac{c_2(\theta)}{c_1(\theta)} \geq \dots \geq \frac{c_K(\theta)}{c_{K-1}(\theta)} \quad (77)$$

holds.

Proof. Fix $k = 0, \dots, K - 1$. From the expression (62) we note that

$$\begin{aligned} \frac{c_{k+1}(\theta)}{c_k(\theta)} &= \frac{\binom{K}{k+1} \binom{P-K}{K-k-1} \binom{P-2K+k+1}{K}^2}{\binom{K}{k} \binom{P-K}{K-k} \binom{P-2K+k}{K}^2} \\ &= \frac{1}{k+1} \cdot \frac{(K-k)^2}{P-3K+k+1} \cdot \frac{P-2K+k+1}{P-3K+k+1} \end{aligned} \quad (78)$$

and by considering each factor in this last expression we readily conclude that the ratio $\frac{c_{k+1}(\theta)}{c_k(\theta)}$ decreases monotonically with k . \blacksquare

Lemma 10.3 For any scaling $P, K : \mathbb{N}_0 \rightarrow \mathbb{N}_0$ satisfying (39)-(40), we have

$$\frac{c_2(\theta_n)}{c_1(\theta_n)} \leq 1 - q(\theta_n) \quad (79)$$

for all $n \in \mathbb{N}_0$ sufficiently large.

Proof. Pick a scaling $P, K : \mathbb{N}_0 \rightarrow \mathbb{N}_0$ satisfying (39)-(40) so that (68) eventually holds. On that range replace θ by θ_n in (78) with $k = 1$ according to this scaling, yielding

$$\frac{c_2(\theta_n)}{c_1(\theta_n)} = \frac{1}{2} \cdot \frac{(K_n - 1)^2}{(P_n - 3K_n + 2)} \cdot \frac{P_n - 2K_n + 2}{P_n - 3K_n + 2}. \quad (80)$$

It is now plain that

$$\lim_{n \rightarrow \infty} \frac{P_n}{K_n^2} \cdot \frac{c_2(\theta_n)}{c_1(\theta_n)} = \frac{1}{2}$$

by making use of the consequences (42) and (43) of the assumption (39)-(40). From the equivalence (41) this last limiting fact can be rewritten as

$$\frac{c_2(\theta_n)}{c_1(\theta_n)} \sim \frac{1}{2} (1 - q(\theta_n))$$

and the desired conclusion follows. ■

Combining Lemma 10.1, Lemma 10.2 and Lemma 10.3 will lead to the following key bounds.

Lemma 10.4 For any scaling $P, K : \mathbb{N}_0 \rightarrow \mathbb{N}_0$ satisfying (39)-(40), we have

$$c_k(\theta_n) \leq (1 - q(\theta_n))^k, \quad k = 1, 2, \dots, K_n \quad (81)$$

for all $n \in \mathbb{N}_0$ sufficiently large.

Proof. Pick a scaling $P, K : \mathbb{N}_0 \rightarrow \mathbb{N}_0$ satisfying (39)-(40). For each $n = 2, 3, \dots$, we can use Lemma 10.1 and Lemma 10.2 to conclude that

$$\begin{aligned} c_k(\theta_n) &= \prod_{\ell=1}^{k-1} \frac{c_{\ell+1}(\theta_n)}{c_\ell(\theta_n)} \cdot c_1(\theta_n) \\ &\leq \left(\frac{c_2(\theta_n)}{c_1(\theta_n)} \right)^{k-1} \cdot c_1(\theta_n) \\ &\leq \left(\frac{c_2(\theta_n)}{c_1(\theta_n)} \right)^{k-1} \cdot (1 - q(\theta_n)) \end{aligned} \quad (82)$$

with $k = 1, \dots, K_n$. The desired conclusion is now a simple consequence of Lemma 10.3. \blacksquare

We are now in a position to take the first step towards the proof of Proposition 9.1.

Proposition 10.5 *For any scaling $P, K : \mathbb{N}_0 \rightarrow \mathbb{N}_0$ satisfying (39)-(40), we have*

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \cdot \frac{\sum_{k=5}^{K_n} c_k(\theta_n)}{\beta(\theta_n)^2} = 0 \quad (83)$$

under the condition (13).

Proof. Pick an admissible scaling $P, K : \mathbb{N}_0 \rightarrow \mathbb{N}_0$ satisfying (39)-(40). From Lemma 10.4 it follows that

$$\begin{aligned} \sum_{k=5}^{K_n} c_k(\theta_n) &\leq \sum_{k=5}^{K_n} (1 - q(\theta_n))^k \\ &\leq \sum_{k=5}^{\infty} (1 - q(\theta_n))^k \\ &= \frac{(1 - q(\theta_n))^5}{q(\theta_n)} \end{aligned} \quad (84)$$

for all $n \in \mathbb{N}_0$ sufficiently large. Letting n go to infinity in this last inequality we readily obtain (83) as an immediate consequence of Proposition 6.4, to wit (51) (with $a = 5$ and $b = 2$). \blacksquare

11 The second step

It is now plain from Proposition 10.5 that the proof of Proposition 9.1 will be completed if we show the following fact.

Proposition 11.1 *For any scaling $P, K : \mathbb{N}_0 \rightarrow \mathbb{N}_0$ satisfying (39)-(40), we have*

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \cdot \frac{\sum_{k=0}^4 c_k(\theta_n) - q(\theta_n)^4}{\beta(\theta_n)^2} = 0 \quad (85)$$

under the condition (13).

Fix positive integers K and P such that $3K \leq P$. We start by writing

$$\begin{aligned}
& \sum_{i=0}^4 c_k(\theta) - q(\theta)^4 \tag{86} \\
&= \frac{\binom{P-K}{K} \binom{P-2K}{K}^2 + K \binom{P-K}{K-1} \binom{P-2K+1}{K}^2 + \binom{K}{2} \binom{P-K}{K-2} \binom{P-2K+2}{K}^2}{\binom{P}{K}^3} \\
&\quad + \frac{\binom{K}{3} \binom{P-K}{K-3} \binom{P-2K+3}{K}^2 + \binom{K}{4} \binom{P-K}{K-4} \binom{P-2K+4}{K}^2}{\binom{P}{K}^3} - \frac{\binom{P-K}{K}^4}{\binom{P}{K}^4} \\
&:= \frac{F(\theta)}{G(\theta)}
\end{aligned}$$

where we have

$$\begin{aligned}
& F(\theta) \tag{87} \\
&= [P \dots (P-3K+1)] [(P-2K) \dots (P-3K+1)] \\
&\quad + K^2 [P \dots (P-3K+2)] [(P-2K+1) \dots (P-3K+2)] \\
&\quad + \frac{K^2(K-1)^2}{2} [P \dots (P-3K+3)] [(P-2K+2) \dots (P-3K+3)] \\
&\quad + \frac{K^2(K-1)^2(K-2)^2}{6} [P \dots (P-3K+4)] \times \\
&\quad \quad \times [(P-2K+3) \dots (P-3K+4)] \\
&\quad + \frac{K^2(K-1)^2(K-2)^2(K-3)^2}{24} [P \dots (P-3K+5)] \times \\
&\quad \quad \times [(P-2K+4) \dots (P-3K+5)] \\
&\quad - [(P-K)^4 \dots (P-2K+1)^4].
\end{aligned}$$

and

$$G(\theta) := \prod_{\ell=0}^{K-1} (P-\ell)^4. \tag{88}$$

Next, we write $F(\theta)$ as a polynomial in P (of order $4K$):

$$F(\theta) = a_0(\theta)P^{4K} + a_1(\theta)P^{4K-1} + \dots + a_{4K-1}(\theta)P + a_{4K}(\theta) \tag{89}$$

where the coefficients depend on θ only through K . We now compute the first six coefficients:

Proposition 11.2 *With positive integers K and P such that $3K \leq P$, we have*

$$a_0(\theta) = a_1(\theta) = a_2(\theta) = 0 \tag{90}$$

and

$$a_3(\theta) = K^4 \tag{91}$$

whereas

$$a_4(\theta) = -6K^6 + 6K^5 - K^4 \quad (92)$$

and

$$a_5(\theta) = -\frac{1}{120}K^{10} + \frac{1}{6}K^9 + \frac{199}{12}K^8 - 34K^7 + \frac{1207}{120}K^6 + \frac{161}{6}K^5 - \frac{209}{6}K^4 + 20K^3 - \frac{24}{5}K^2. \quad (93)$$

The proof of Proposition 11.2 is tedious and is given in Appendix C. For the remaining coefficients, we rely on the bounds obtained next:

Proposition 11.3 *With positive integers K and P such that $3K \leq P$, we have*

$$|a_i(\theta)| \leq 2 \cdot (12K^2)^i, \quad i = 6, 7, \dots, 4K. \quad (94)$$

Proof. Pick positive integers K, P such that $3K \leq P$. The first term in (87) is a polynomial in P with order $4K$ and leading coefficient 1. Therefore, it contributes to all of the coefficients $a_1(\theta), \dots, a_{4K}(\theta)$ in the expression (89) and for each $i = 1, 2, \dots, 4K$, its contribution to $a_i(\theta)$ can be determined by the products of i of its roots summed over all $\binom{4K}{i}$ possible i -uple of roots. In fact, we have

$$a_{i,1}(\theta) = (-1)^i \sum_{l=1}^{\binom{4K}{i}} \prod_{j=1}^i r_{1,l,j}, \quad i = 1, 2, \dots, 4K$$

where for each $l = 1, \dots, \binom{4K}{i}$, $\{r_{1,l,j}\}_{j=1}^i$ is a sequence of i distinct roots of the first term in (87) and $a_{i,1}(\theta)$ stands for the contribution of that term to $a_i(\theta)$. As all roots of the first term is bounded in absolute value by $3K$, it is now a simple matter to conclude that

$$|a_{i,1}(\theta)| \leq \binom{4K}{i} (3K)^i, \quad i = 0, 1, \dots, 4K.$$

Similarly, with $a_{i,2}(\theta)$ representing the contribution of the second term of (87) to $a_i(\theta)$, we have

$$|a_{i,2}(\theta)| \leq K^2 \binom{4K-1}{i-1} (3K)^{i-1}, \quad i = 1, 2, \dots, 4K$$

since the second term in (87) defines a polynomial in P with order $4K - 1$. Proceeding in a similar manner for the 3rd, 4th, 5th and the 6th terms of (87),

we readily obtain

$$\begin{aligned}
& |a_i(\theta)| \\
&= |a_{i,1}(\theta) + a_{i,2}(\theta) + a_{i,3}(\theta) + a_{i,4}(\theta) + a_{i,5}(\theta) + a_{i,6}(\theta)| \\
&\leq |a_{i,1}(\theta)| + |a_{i,2}(\theta)| + |a_{i,3}(\theta)| + |a_{i,4}(\theta)| + |a_{i,5}(\theta)| + |a_{i,6}(\theta)| \\
&\leq \binom{4K}{i} (3K)^i + K^2 \binom{4K-1}{i-1} (3K)^{i-1} + \frac{K^2(K-1)^2}{2} \binom{4K-2}{i-2} (3K)^{i-2} \\
&\quad + \frac{K^2(K-1)^2(K-2)^2}{6} \binom{4K-3}{i-3} (3K)^{i-3} \\
&\quad + \frac{K^2(K-1)^2(K-2)^2(K-3)^2}{24} \binom{4K-4}{i-4} (3K)^{i-4} + \binom{4K}{i} (2K)^i \\
&\leq (12K^2)^i + K^2(12K^2)^{i-1} + K^4(12K^2)^{i-2} + K^6(12K^2)^{i-3} \\
&\quad + K^8(12K^2)^{i-4} + (8K^2)^i \\
&\leq 2 \cdot (12K^2)^i.
\end{aligned}$$

■

We now obtain a bound for $F(\theta_n)$ by the help of Proposition 11.2 and Proposition 11.3:

Proposition 11.4 *Consider scaling $P, K : \mathbb{N}_0 \rightarrow \mathbb{N}_0$ satisfying (39)-(40). We have*

$$F(\theta_n) \leq K_n^4 P_n^{4K_n-3} \quad (95)$$

for all $n \in \mathbb{N}_0$ sufficiently large.

Proof. Recalling (92), we find

$$a_4(\theta) = -6K^6 + 6K^5 - K^4 \leq -K^5, \quad K = 1, 2, \dots, \quad (96)$$

whereas from (93), we get

$$a_5(\theta) \leq -\frac{1}{120}K^{10} + \frac{1}{6}K^9 + \frac{199}{12}K^8 \leq -\frac{1}{120}K^{10} + 17K^9, \quad K = 1, 2, \dots \quad (97)$$

by crude bounding arguments. Now, pick an admissible scaling $P, K : \mathbb{N}_0 \rightarrow \mathbb{N}_0$ satisfying (39)-(40) and replace θ by θ_n in (89) according to this scaling. It is clear by Proposition 11.2 that (95) will follow if we show that

$$a_4(\theta_n)P_n^{4K_n-4} + a_5(\theta_n)P_n^{4K_n-5} + \dots + a_{4K_n-1}(\theta_n)P_n + a_{4K_n}(\theta_n) \leq 0 \quad (98)$$

for all n sufficiently large. With the help of (94), (96) and (97), we find

$$\begin{aligned}
& a_4(\theta_n)P_n^{4K_n-4} + a_5(\theta_n)P_n^{4K_n-5} + \dots + a_{4K_n-1}(\theta_n)P_n + a_{4K_n}(\theta_n) \\
&\leq P_n^{4K_n-4} \left(-K_n^5 + \frac{-\frac{1}{120}K_n^{10} + 17K_n^9}{P_n} + \frac{|a_6(\theta_n)|}{P_n^2} + \dots + \frac{|a_{4K_n}(\theta_n)|}{P_n^{4K_n-4}} \right)
\end{aligned}$$

$$\begin{aligned}
&\leq P_n^{4K_n-4} \left(-K_n^5 + \frac{-\frac{1}{120}K_n^{10} + 17K_n^9}{P_n} + 2 \cdot \frac{(12K_n^2)^6}{P_n^2} \sum_{i=0}^{4K_n-6} \left(\frac{12K_n^2}{P_n} \right)^i \right) \\
&\leq K_n^5 P_n^{4K_n-4} \left(-1 + \frac{-\frac{1}{120}K_n^5 + 17K_n^4}{P_n} + 2 \cdot (12)^6 \frac{K_n^7}{P_n^2} \sum_{i=0}^{\infty} \left(\frac{12K_n^2}{P_n} \right)^i \right) \\
&= K_n^5 P_n^{4K_n-4} \left(-1 + \frac{-\frac{1}{120}K_n^5 + 17K_n^4}{P_n} + 2 \cdot (12)^6 \cdot \frac{K_n^7}{P_n^2} \cdot \frac{1}{1 - \frac{12K_n^2}{P_n}} \right) \\
&= K_n^5 P_n^{4K_n-5} \left(-P_n - \frac{1}{120}K_n^5 + 17K_n^4 + 2 \cdot (12)^6 \cdot \frac{K_n^7}{P_n} \cdot \frac{1}{1 - \frac{12K_n^2}{P_n}} \right) \quad (99)
\end{aligned}$$

for all n large enough to ensure that $\frac{K_n^2}{P_n} < 1$. By virtue of (40), there exists a positive integer n_1^* such that

$$-\frac{1}{240}K_n^5 + 2 \cdot (12)^6 \frac{K_n^7}{P_n} \cdot \frac{1}{1 - \frac{12K_n^2}{P_n}} = K_n^5 \left(-\frac{1}{240} + 2 \cdot (12)^6 \frac{\frac{K_n^2}{P_n}}{1 - \frac{12K_n^2}{P_n}} \right) \leq 0$$

for all $n \geq n_1^*$. Therefore, $n \geq n_1^*$ implies that

$$\begin{aligned}
&-P_n - \frac{1}{120}K_n^5 + 17K_n^4 + 2 \cdot (12)^6 \cdot \frac{K_n^7}{P_n} \cdot \frac{1}{1 - \frac{12K_n^2}{P_n}} \\
&\leq -P_n - \frac{1}{240}K_n^5 + 17K_n^4. \quad (100)
\end{aligned}$$

Also, in view of the consequence (43) of (39)-(40), there exist a positive integer n_2^* such that

$$P_n \geq 17 \cdot (17 \cdot 240)^4 \quad (101)$$

for all $n \geq n_2^*$. Now pick $n \geq n_2^*$ and assume that $K_n \leq 17 \cdot 240$. We obtain

$$-P_n - \frac{1}{240}K_n^5 + 17K_n^4 \leq -P_n + 17K_n^4 \leq -P_n + 17 \cdot (17 \cdot 240)^4 \leq 0 \quad (102)$$

via (101). If on the other hand $K_n > 17 \cdot 240$, we have

$$-P_n - \frac{1}{240}K_n^5 + 17K_n^4 \leq -\frac{1}{240}K_n^5 + 17K_n^4 \leq 0. \quad (103)$$

Consequently, we have for all $n \geq n_2^*$

$$-P_n - \frac{1}{240}K_n^5 + 17K_n^4 \leq 0. \quad (104)$$

Invoking (100), we now have

$$-P_n - \frac{1}{120}K_n^5 + 17K_n^4 + 2 \cdot (12)^6 \cdot \frac{K_n^7}{P_n} \cdot \frac{1}{1 - \frac{12K_n^2}{P_n}} \leq 0$$

for all $n \geq \max\{n_1^*, n_2^*\}$. This last fact readily implies (98) via (99). \blacksquare

The proof of Proposition 11.1 can now be completed:

Proposition 11.5 *For any scaling $P, K : \mathbb{N}_0 \rightarrow \mathbb{N}_0$ satisfying (39)-(40), we have*

$$\lim_{n \rightarrow \infty} \frac{1}{n^2 \beta(\theta_n)^2} \cdot \frac{F(\theta_n)}{G(\theta_n)} = 0 \quad (105)$$

under the condition (13).

Proof. Pick an admissible scaling $P, K : \mathbb{N}_0 \rightarrow \mathbb{N}_0$ satisfying (39)-(40) and assume that (13) holds. Invoking Proposition 11.4 and recalling (88), we find

$$\frac{1}{n^2 \beta^2(\theta_n)} \cdot \frac{F(\theta_n)}{G(\theta_n)} \leq \frac{K_n^4}{n^2 \beta(\theta_n)^2 P_n^3} \prod_{\ell=1}^{K_n-1} \left(\frac{P_n}{P_n - \ell} \right)^4. \quad (106)$$

Also, we have

$$\left(\prod_{\ell=1}^{K_n-1} \left(\frac{P_n}{P_n - \ell} \right)^4 \right)^{-1} = \prod_{\ell=1}^{K_n-1} \left(1 - \frac{\ell}{P_n} \right)^4 \quad (107)$$

and an easy bounding argument yields

$$\left(1 - \frac{K_n - 1}{P_n} \right)^{4(K_n-1)} \leq \prod_{\ell=1}^{K_n-1} \left(1 - \frac{\ell}{P_n} \right)^4 \leq 1. \quad (108)$$

Noting that

$$1 - \left(1 - \frac{K_n - 1}{P_n} \right)^{4(K_n-1)} = \int_{1 - \frac{K_n-1}{P_n}}^1 4(K_n - 1)t^{4K_n-5} dt \leq \frac{4(K_n - 1)^2}{P_n},$$

we find

$$1 - \frac{4(K_n - 1)^2}{P_n} \leq \prod_{\ell=1}^{K_n-1} \left(1 - \frac{\ell}{P_n} \right)^4 \leq 1 \quad (109)$$

via (108). Now, let n go to infinity in this last expression: We have

$$\lim_{n \rightarrow \infty} \prod_{\ell=1}^{K_n-1} \left(1 - \frac{\ell}{P_n} \right)^4 = 1 \quad (110)$$

by virtue of (40) and this readily implies

$$\lim_{n \rightarrow \infty} \prod_{\ell=1}^{K_n-1} \left(\frac{P_n}{P_n - \ell} \right)^4 = 1. \quad (111)$$

Substituting (46) and (111) into (106), we obtain

$$\begin{aligned}
\frac{1}{n^2 \beta(\theta_n)^2} \cdot \frac{K_n^4}{P_n^3} \prod_{\ell=1}^{K_n-1} \left(\frac{P_n}{P_n - \ell} \right)^4 &\sim \frac{K_n^4}{n^2 P_n^3 \left(\frac{K_n^3}{P_n^2} + \left(\frac{K_n^2}{P_n} \right)^3 \right)^2} \quad (112) \\
&\leq \frac{K_n^4}{n^2 P_n^3 \left(\frac{K_n^3}{P_n^2} \right)^2} \\
&= \frac{1}{n^2 \frac{K_n^2}{P_n}}
\end{aligned}$$

and (105) follows from the consequence (49) of (13). ■

A A proof of Proposition 6.3 (Continued)

With positive integers K, P such that $3K \leq P$, we write

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

Thus, an elementary bounding argument yields

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

whence

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

Now, pick a scaling $P, K : \mathbb{N}_0 \rightarrow \mathbb{N}_0$ satisfying the equivalent conditions (39)-(40), and replace $\theta_n = (K_n, P_n)$ by $\theta = (P, K)$ in the last expression according to this scaling. We find

$$1 - \left(1 - \left(\frac{K_n}{P_n - K_n} \right)^2 \right)^{K_n} \leq 1 - \frac{r(\theta_n)}{q(\theta_n)^2} \leq 1 - \left(1 - \left(\frac{K_n}{P_n - 2K_n} \right)^2 \right)^{K_n}. \quad (\text{A.1})$$

Invoking (42), we now obtain

$$\left(\frac{K_n}{P_n - cK_n}\right)^2 = \left(\frac{K_n}{P_n}(1 + o(1))\right)^2 = \frac{K_n^2}{P_n^2}(1 + o(1)) \quad (\text{A.2})$$

for each $c = 1, 2$, whence

$$\lim_{n \rightarrow \infty} K_n \left(\frac{K_n}{P_n - cK_n}\right)^2 = 0, \quad c = 1, 2$$

by virtue of (40) and (42). Finally, let n go to infinity in (A.1) and use the elementary convergence relation

$$(1 - a)^b \sim 1 - ab \quad \text{if } ab \rightarrow 0$$

with

$$a = \left(\frac{K_n}{P_n - cK_n}\right)^2, \quad c = 1, 2 \quad \text{and} \quad b = K_n.$$

Noting that (A.2) also implies

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

we readily get (47) by a sandwich argument. ■

B Evaluating (61)–(62)

For notational convenience, we define

$$K_{ij} := [K_i(\theta) \cap K_j(\theta) \neq \emptyset].$$

for distinct $i, j = 1, 2, \dots, n$. Moreover, for any non-empty subset S of $\{1, \dots, P\}$, we write

$$K_{Si} := [S \cap K_i(\theta) \neq \emptyset], \quad i = 1, \dots, n.$$

In what follows we make repeated use of the decomposition (32). Beginning with the observation

$$\begin{aligned} & \mathbb{E}[\chi_{n,123}(\theta)\chi_{n,124}(\theta)] \\ &= \mathbb{P}[K_{12}, K_{13}, K_{23}, K_{14}, K_{24}] \\ &= \mathbb{P}[K_{13}, K_{23}, K_{14}, K_{24}] - \mathbb{P}[K_{12}^c, K_{13}, K_{23}, K_{14}, K_{24}]. \end{aligned} \quad (\text{B.3})$$

we shall compute each term in turn.

To compute the second term in (B.3) we condition on the sets K_1 and K_2 such that $K_1 \cap K_2 = \emptyset$. Thus,

$$\begin{aligned}
& \mathbb{P}[K_{12}^c, K_{13}, K_{23}, K_{14}, K_{24}] \\
&= \sum_{|S|=|T|=K, S \cap T = \emptyset} \mathbb{P}[K_1 = S, K_2 = T, K_{S3}, K_{T3}, K_{S4}, K_{T4}] \\
&= \sum_{|S|=|T|=K, S \cap T = \emptyset} \mathbb{P}[K_1 = S, K_2 = T] \mathbb{P}[K_{S3}, K_{T3}, K_{S4}, K_{T4}] \\
&= \sum_{|S|=|T|=K, S \cap T = \emptyset} \binom{P}{K}^{-2} \mathbb{P}[K_{S3}, K_{T3}] \cdot \mathbb{P}[K_{S4}, K_{T4}] \\
&= \sum_{|S|=|T|=K, S \cap T = \emptyset} \binom{P}{K}^{-2} (\mathbb{P}[K_{S3}, K_{T3}])^2 \\
&= \binom{P}{K}^{-2} \sum_{|S|=|T|=K, S \cap T = \emptyset} (\mathbb{P}[K_{S3}] - \mathbb{P}[K_{T3}^c] + \mathbb{P}[K_{S3}^c, K_{T3}^c])^2 \\
&= \binom{P}{K}^{-2} \sum_{|S|=|T|=K, S \cap T = \emptyset} (1 - \mathbb{P}[K_{S3}^c] - \mathbb{P}[K_{T3}^c] + \mathbb{P}[K_{S3}^c, K_{T3}^c])^2 \\
&= \binom{P}{K}^{-2} \sum_{|S|=|T|=K, S \cap T = \emptyset} (1 - 2q(\theta) + r(\theta))^2 \\
&= \binom{P}{K}^{-2} \binom{P}{K} \binom{P-K}{K} (1 - 2q(\theta) + r(\theta))^2 \\
&= q(\theta) (1 - 2q(\theta) + r(\theta))^2 \tag{B.4}
\end{aligned}$$

as we note from (5) that $\mathbb{P}[K_{S3}^c] = \mathbb{P}[K_{T3}^c] = q(\theta)$ for S and T in \mathcal{P}_K with $\mathbb{P}[K_{S3}^c, K_{T3}^c] = r(\theta)$ whenever $S \cap T = \emptyset$.

We now turn to the first term in (B.3). Again, upon making repeated use of (32) we find

$$\begin{aligned}
& \mathbb{P}[K_{13}, K_{23}, K_{14}, K_{24}] \\
&= \mathbb{P}[K_{23}, K_{14}, K_{24}] - \mathbb{P}[K_{13}^c, K_{23}, K_{14}, K_{24}] \\
&= \mathbb{P}[K_{14}, K_{24}] - \mathbb{P}[K_{23}^c, K_{14}, K_{24}] - \mathbb{P}[K_{13}^c, K_{14}, K_{24}] + \mathbb{P}[K_{13}^c, K_{23}^c, K_{14}, K_{24}] \\
&= (1 - q(\theta))^2 - 2\mathbb{P}[K_{23}^c, K_{14}, K_{24}] + \mathbb{P}[K_{13}^c, K_{23}^c, K_{24}] - \mathbb{P}[K_{13}^c, K_{23}^c, K_{14}, K_{24}] \\
&= (1 - q(\theta))^2 - 2\mathbb{P}[K_{23}^c, K_{14}, K_{24}] + \mathbb{P}[K_{13}^c, K_{23}^c, K_{24}] \\
&\quad - \mathbb{P}[K_{13}^c, K_{23}^c, K_{14}] + \mathbb{P}[K_{13}^c, K_{23}^c, K_{14}, K_{24}] \tag{B.5}
\end{aligned}$$

as we note that $\mathbb{P}[K_{23}^c, K_{14}, K_{24}] = \mathbb{P}[K_{13}^c, K_{14}, K_{24}]$. Next, we find

$$\begin{aligned}
\mathbb{P}[K_{23}^c, K_{14}, K_{24}] &= \sum_{|S|=K} \mathbb{P}[K_4 = S, K_{23}^c, K_{S1}, K_{S2}] \\
&= \sum_{|S|=K} \mathbb{P}[K_4 = S] \mathbb{P}[K_{23}^c, K_{S1}, K_{S2}]
\end{aligned}$$

$$\begin{aligned}
&= \sum_{|S|=K} \binom{P}{K}^{-1} \mathbb{P}[K_{S1}] \cdot \mathbb{P}[K_{23}^c, K_{S2}] \\
&= \sum_{|S|=K} \binom{P}{K}^{-1} (1 - q(\theta)) \cdot q(\theta)(1 - q(\theta)) \quad (\text{B.6}) \\
&= q(\theta)(1 - q(\theta))^2. \quad (\text{B.7})
\end{aligned}$$

upon using (33) in (B.6).

In a similar manner, we obtain

$$\begin{aligned}
\mathbb{P}[K_{13}^c, K_{23}^c, K_{24}] &= \sum_{|S|=K} \mathbb{P}[K_2 = S, K_{13}^c, K_{S3}^c, K_{S4}] \\
&= \sum_{|S|=K} \mathbb{P}[K_2 = S] \mathbb{P}[K_{13}^c, K_{S3}^c, K_{S4}] \\
&= \sum_{|S|=K} \binom{P}{K}^{-1} \mathbb{P}[K_{S4}] \cdot \mathbb{P}[K_{13}^c, K_{S3}^c] \\
&= \sum_{|S|=K} \binom{P}{K}^{-1} (1 - q(\theta)) \cdot q(\theta)^2 \quad (\text{B.8}) \\
&= q(\theta)^2(1 - q(\theta)) \quad (\text{B.9})
\end{aligned}$$

where (B.8) follows from (34).

Moreover, we also get

$$\begin{aligned}
\mathbb{P}[K_{13}^c, K_{23}^c, K_{14}^c] &= \sum_{|S|=K} \mathbb{P}[K_1 = S, K_{S3}^c, K_{23}^c, K_{S4}^c] \\
&= \sum_{|S|=K} \mathbb{P}[K_1 = S] \mathbb{P}[K_{S3}^c, K_{23}^c, K_{S4}^c] \\
&= \sum_{|S|=K} \binom{P}{K}^{-1} \mathbb{P}[K_{S4}^c] \cdot \mathbb{P}[K_{S3}^c, K_{23}^c] \\
&= \sum_{|S|=K} \binom{P}{K}^{-1} q(\theta) \cdot q(\theta)^2 \\
&= q(\theta)^3. \quad (\text{B.10})
\end{aligned}$$

Finally consider the term $\mathbb{P}[K_{13}^c, K_{23}^c, K_{14}^c, K_{24}^c]$: By conditioning on the cardinality of the intersection $K_1 \cap K_2$, we obtain

$$\begin{aligned}
&\mathbb{P}[K_{13}^c, K_{23}^c, K_{14}^c, K_{24}^c] \\
&= \sum_{|S|=|T|=K} \mathbb{P}[K_1 = S, K_2 = T, K_{S3}^c, K_{T3}^c, K_{S4}^c, K_{T4}^c] \\
&= \sum_{|S|=K} \sum_{k=0}^K \sum_{|T|=K, |T \cap S|=k} \mathbb{P}[K_1 = S, K_2 = T, K_{S3}^c, K_{T3}^c, K_{S4}^c, K_{T4}^c]
\end{aligned}$$

$$\begin{aligned}
&= \sum_{|S|=K} \mathbb{P}[K_1 = S] \sum_{k=0}^K \sum_{|T|=K, |T \cap S|=k} \mathbb{P}[K_2 = T] \mathbb{P}[K_{S_3}^c, K_{T_3}^c] \cdot \mathbb{P}[K_{S_4}^c, K_{T_4}^c] \\
&= \sum_{|S|=K} \binom{P}{K}^{-1} \sum_{k=0}^K \sum_{|T|=K, |T \cap S|=k} \mathbb{P}[K_2 = T] \cdot (\mathbb{P}[K_{S_3}^c, K_{T_3}^c])^2 \\
&= \sum_{|S|=K} \binom{P}{K}^{-1} \sum_{k=0}^K \frac{\binom{K}{k} \binom{P-K}{K-k}}{\binom{P}{K}} \cdot \left(\frac{\binom{P-2K+k}{K}}{\binom{P}{K}} \right)^2 \\
&= \sum_{k=0}^K \frac{\binom{K}{k} \binom{P-K}{K-k}}{\binom{P}{K}} \cdot \left(\frac{\binom{P-2K+k}{K}}{\binom{P}{K}} \right)^2 \\
&= \sum_{k=0}^K c_k(\theta) \tag{B.11}
\end{aligned}$$

as we make use of (62).

Substituting (B.4) and (B.5) (by the help of (B.7), (B.9), (B.10), and (B.11)) into (B.3), we find

$$\begin{aligned}
&\mathbb{E}[\chi_{n,123}(\theta) \chi_{n,124}(\theta)] \\
&= (1 - q(\theta))^2 - 2q(\theta)(1 - q(\theta))^2 + q(\theta)^2(1 - q(\theta)) - q(\theta)^3 \\
&\quad - q(\theta)(1 - 2q(\theta) + r(\theta))^2 + \sum_{k=0}^K c_k(\theta) \tag{B.12}
\end{aligned}$$

where we have used the notation (62).

As we seek to simplify this last expression, we note that

$$\begin{aligned}
&(1 - q(\theta))^2 - 2q(\theta)(1 - q(\theta))^2 + q(\theta)^2(1 - q(\theta)) - q(\theta)^3 \\
&= (1 - q(\theta))^2(1 - 2q(\theta)) + q(\theta)^2(1 - q(\theta)) - q(\theta)^3 \\
&= (1 - q(\theta))^2(1 - 2q(\theta) + q(\theta)^2) - q(\theta)^2(1 - q(\theta))^2 \\
&\quad + q(\theta)^2(1 - q(\theta)) - q(\theta)^3 \\
&= (1 - q(\theta))^4 + q(\theta)^2((1 - q(\theta)) - (1 - q(\theta))^2) - q(\theta)^3 \\
&= (1 - q(\theta))^4 + q(\theta)^2(1 - q(\theta))(1 - (1 - q(\theta))) - q(\theta)^3 \\
&= (1 - q(\theta))^4 + q(\theta)^3(1 - q(\theta)) - q(\theta)^3 \\
&= (1 - q(\theta))^4 - q(\theta)^4. \tag{B.13}
\end{aligned}$$

Next, we observe that

$$\begin{aligned}
&q(\theta)(1 - 2q(\theta) + r(\theta))^2 \\
&= q(\theta)(1 - 2q(\theta) + q(\theta)^2 - q(\theta)^2 + r(\theta))^2 \\
&= q(\theta)\left((1 - q(\theta))^2 - (q(\theta)^2 - r(\theta))\right)^2 \\
&= q(\theta)\left((1 - q(\theta))^4 - 2(1 - q(\theta))^2(q(\theta)^2 - r(\theta)) + (q(\theta)^2 - r(\theta))^2\right)
\end{aligned}$$

$$\begin{aligned}
&= q(\theta) (1 - q(\theta))^4 - 2q(\theta) (1 - q(\theta))^2 (q(\theta)^2 - r(\theta)) \\
&\quad + q(\theta) (q(\theta)^2 - r(\theta))^2.
\end{aligned} \tag{B.14}$$

Subtracting (B.14) from (B.13) gives

$$\begin{aligned}
&(1 - q(\theta))^4 - q(\theta)^4 - q(\theta) (1 - 2q(\theta) + r(\theta))^2 \\
&= (1 - q(\theta))^4 - q(\theta)^4 - q(\theta) (1 - q(\theta))^4 + 2q(\theta) (1 - q(\theta))^2 (q(\theta)^2 - r(\theta)) \\
&\quad - q(\theta) (q(\theta)^2 - r(\theta))^2 \\
&= (1 - q(\theta))^4 (1 - q(\theta)) - q(\theta)^4 + 2q(\theta) (1 - q(\theta))^2 (q(\theta)^2 - r(\theta)) \\
&\quad - q(\theta) (q(\theta)^2 - r(\theta))^2 \\
&= (1 - q(\theta))^5 - q(\theta)^4 + 2q(\theta) (1 - q(\theta))^2 (q(\theta)^2 - r(\theta)) \\
&\quad - q(\theta) (q(\theta)^2 - r(\theta))^2
\end{aligned} \tag{B.15}$$

Reporting the outcome of this last calculation into (B.12) we then get

$$\begin{aligned}
&\mathbb{E} [\chi_{n,123}(\theta) \chi_{n,124}(\theta)] \\
&= (1 - q(\theta))^5 + 2q(\theta) (1 - q(\theta))^2 (q(\theta)^2 - r(\theta)) \\
&\quad - q(\theta) (q(\theta)^2 - r(\theta))^2 + \sum_{k=0}^K c_k(\theta) - q(\theta)^4
\end{aligned} \tag{B.16}$$

and the conclusion (61) follows as we make use of the expression (23) for $\beta(\theta)$.

C A proof of Proposition 11.2

Pick positive integers K and P such that $3K \leq P$. First we note that in (87) the first and last terms are of order $4K$ whereas the second, third, fourth, and the fifth terms are of order $4K - 1$, $4K - 2$, $4K - 3$ and $4K - 4$, respectively. Therefore $a_0(\theta)$ is determined only by the first and last terms and it is immediate that $a_0(\theta) = 0$. First, second and the last terms in (87) determines $a_1(\theta)$ and a careful inspection gives

$$a_1(\theta) = - \left(\sum_{i=0}^{3K-1} i + \sum_{i=2K}^{3K-1} i \right) + K^2 - \left(- \binom{4}{1} \sum_{i=K}^{2K-1} i \right) = 0. \tag{C.17}$$

The third coefficient $a_2(\theta)$ depends only on the first three terms and the last term of (87). Again, it is straightforward to check that

$$\begin{aligned}
&a_2(\theta) \\
&= \sum_{i=1}^{3K-2} i \sum_{j=i+1}^{3K-1} j + \sum_{i=2K}^{3K-2} i \sum_{j=i+1}^{3K-1} j + \left(\sum_{i=2K}^{3K-1} i \right) \left(\sum_{i=1}^{3K-1} i \right)
\end{aligned} \tag{C.18}$$

$$\begin{aligned}
& -K^2 \left(\sum_{i=1}^{3K-2} i + \sum_{i=2K-1}^{3K-2} i \right) + \frac{K^2(K-1)^2}{2} \\
& - \left(\binom{4}{1}^2 \cdot \sum_{i=K}^{2K-2} i \sum_{j=i+1}^{2K-1} j + \binom{4}{2} \sum_{i=K}^{2K-1} i^2 \right) \\
& = 0.
\end{aligned}$$

We now compute the fourth coefficient $a_3(\theta)$ and start by noting that it depends on all but the fifth term in (87). Straightforward computation gives

$$\begin{aligned}
a_3(\theta) & \tag{C.19} \\
& = - \left(\sum_{v=1}^{3K-3} v \sum_{i=v+1}^{3K-2} i \sum_{j=i+1}^{3K-1} j + \sum_{v=2K}^{3K-3} v \sum_{i=v+1}^{3K-2} i \sum_{j=i+1}^{3K-1} j \right) \\
& - \left(\sum_{i=1}^{3K-1} i \cdot \sum_{i=2K}^{3K-2} i \sum_{j=i+1}^{3K-1} j + \sum_{i=2K}^{3K-1} i \cdot \sum_{i=1}^{3K-2} i \sum_{j=i+1}^{3K-1} j \right) \\
& + K^2 \left(\sum_{i=1}^{3K-3} i \sum_{j=i+1}^{3K-2} j + \sum_{i=2K-1}^{3K-3} i \sum_{j=i+1}^{3K-2} j + \binom{3K-2}{i=2K-1} \binom{3K-2}{i=1} \right) \\
& - \frac{K^2(K-1)^2}{2} \left(\sum_{i=1}^{3K-3} i + \sum_{i=2K-2}^{3K-3} i \right) + \frac{K^2(K-1)^2(K-2)^2}{6} \\
& + \binom{4}{1}^3 \cdot \sum_{v=K}^{2K-3} v \sum_{i=v+1}^{2K-2} i \sum_{j=i+1}^{2K-1} j + \binom{4}{2} \binom{4}{1} \sum_{j=K}^{2K-1} j^2 \left(\sum_{i=K}^{2K-1} i - j \right) \\
& + \binom{4}{3} \sum_{i=K}^{2K-1} i^3 \\
& = K^4
\end{aligned}$$

It is clear that $a_4(\theta)$ and $a_5(\theta)$ depends on all of the terms in (87). For $a_4(\theta)$, we proceed in a similar manner to get ¹

$$\begin{aligned}
a_4(\theta) & \tag{C.20} \\
& = \sum_{l=1}^{3K-4} l \sum_{v=l+1}^{3K-3} v \sum_{i=v+1}^{3K-2} i \sum_{j=i+1}^{3K-1} j + \sum_{l=2K}^{3K-4} l \sum_{v=l+1}^{3K-3} v \sum_{i=v+1}^{3K-2} i \sum_{j=i+1}^{3K-1} j \\
& + \sum_{i=1}^{3K-1} i \cdot \sum_{v=2K}^{3K-3} v \sum_{i=v+1}^{3K-2} i \sum_{j=i+1}^{3K-1} j + \sum_{i=2K}^{3K-1} i \cdot \sum_{v=1}^{3K-3} v \sum_{i=v+1}^{3K-2} i \sum_{j=i+1}^{3K-1} j
\end{aligned}$$

¹Evaluating the expression (C.20) (as well as (C.21) given next) by hand is quite cumbersome. To avoid this, one can make use of a computer software (such as Mathematica, MATLAB, etc.) that can perform calculations symbolically.

$$\begin{aligned}
& + \binom{3K-2}{i=1} \binom{3K-1}{j=i+1} \binom{3K-2}{i=2K} \binom{3K-1}{j=i+1} - K^2 \sum_{v=1}^{3K-4} v \sum_{i=v+1}^{3K-3} i \sum_{j=i+1}^{3K-2} j \\
& - K^2 \left(\sum_{v=2K-1}^{3K-4} v \sum_{i=v+1}^{3K-3} i \sum_{j=i+1}^{3K-2} j + \sum_{i=1}^{3K-2} i \cdot \sum_{i=2K-1}^{3K-3} i \sum_{j=i+1}^{3K-2} j \right) \\
& - K^2 \sum_{i=2K-1}^{3K-2} i \cdot \sum_{i=1}^{3K-3} i \sum_{j=i+1}^{3K-2} j + \frac{K^2(K-1)^2}{2} \sum_{i=1}^{3K-4} i \sum_{j=i+1}^{3K-3} j \\
& + \frac{K^2(K-1)^2}{2} \left(\sum_{i=2K-2}^{3K-4} i \sum_{j=i+1}^{3K-3} j + \binom{3K-3}{i=2K-2} \binom{3K-3}{j=1} \right) \\
& - \frac{K^2(K-1)^2(K-2)^2}{6} \left(\sum_{i=1}^{3K-4} i + \sum_{i=2K-3}^{3K-4} i \right) \\
& + \frac{K^2(K-1)^2(K-2)^2(K-3)^2}{24} - \binom{4}{1}^4 \cdot \sum_{l=K}^{2K-4} l \sum_{v=K+1}^{2K-3} v \sum_{i=v+1}^{2K-2} i \sum_{j=i+1}^{2K-1} j \\
& - \binom{4}{2} \binom{4}{1}^2 \cdot \sum_{v=K}^{2K-1} v^2 \left(\sum_{i=K}^{2K-2} i \sum_{j=i+1}^{2K-1} j - v \sum_{i=K}^{2K-1} i + v^2 \right) \\
& - \binom{4}{3} \binom{4}{1} \sum_{j=K}^{2K-1} j^3 \left(\sum_{i=K}^{2K-1} i - j \right) - \binom{4}{2}^2 \cdot \sum_{i=K}^{2K-2} i^2 \sum_{j=i+1}^{2K-1} j^2 - \sum_{i=K}^{2K-1} i^4 \\
& = -6K^6 + 6K^5 - K^4.
\end{aligned}$$

Finally, $a_5(\theta)$ is given by

$$\begin{aligned}
& a_5(\theta) \tag{C.21} \\
& = - \sum_{u=1}^{3K-5} u \sum_{l=u+1}^{3K-4} l \sum_{v=l+1}^{3K-3} v \sum_{i=v+1}^{3K-2} i \sum_{j=i+1}^{3K-1} j \\
& - \sum_{u=2K}^{3K-5} u \sum_{l=u+1}^{3K-4} l \sum_{v=l+1}^{3K-3} v \sum_{i=v+1}^{3K-2} i \sum_{j=i+1}^{3K-1} j \\
& - \sum_{i=1}^{3K-1} i \cdot \sum_{l=2K}^{3K-4} l \sum_{v=l+1}^{3K-3} v \sum_{i=v+1}^{3K-2} i \sum_{j=i+1}^{3K-1} j \\
& - \sum_{i=2K}^{3K-1} i \cdot \sum_{l=1}^{3K-4} l \sum_{v=l+1}^{3K-3} v \sum_{i=v+1}^{3K-2} i \sum_{j=i+1}^{3K-1} j \\
& - \left(\sum_{v=1}^{3K-3} v \sum_{i=v+1}^{3K-2} i \sum_{j=i+1}^{3K-1} j \right) \left(\sum_{i=2K}^{3K-2} i \sum_{j=i+1}^{3K-1} j \right)
\end{aligned}$$

$$\begin{aligned}
& - \binom{3K-3}{v=2K} \binom{3K-2}{i=v+1} \binom{3K-1}{j=i+1} \binom{3K-2}{i=1} \binom{3K-1}{j=i+1} \\
& + K^2 \left(\sum_{l=1}^{3K-5} l \sum_{v=l+1}^{3K-4} v \sum_{i=v+1}^{3K-3} i \sum_{j=i+1}^{3K-2} j + \sum_{l=2K-1}^{3K-5} l \sum_{v=l+1}^{3K-4} v \sum_{i=v+1}^{3K-3} i \sum_{j=i+1}^{3K-2} j \right) \\
& + K^2 \sum_{i=1}^{3K-2} i \cdot \sum_{v=2K-1}^{3K-4} v \sum_{i=v+1}^{3K-3} i \sum_{j=i+1}^{3K-2} j \\
& + K^2 \sum_{i=2K-1}^{3K-2} i \cdot \sum_{v=1}^{3K-4} v \sum_{i=v+1}^{3K-3} i \sum_{j=i+1}^{3K-2} j \\
& + K^2 \binom{3K-3}{i=1} \binom{3K-2}{j=i+1} \binom{3K-3}{i=2K-1} \binom{3K-2}{j=i+1} \\
& - \frac{K^2(K-1)^2}{2} \left(\sum_{v=1}^{3K-5} v \sum_{i=v+1}^{3K-4} i \sum_{j=i+1}^{3K-3} j + \sum_{v=2K-2}^{3K-5} v \sum_{i=v+1}^{3K-4} i \sum_{j=i+1}^{3K-3} j \right) \\
& - \frac{K^2(K-1)^2}{2} \left(\sum_{i=1}^{3K-3} i \cdot \sum_{i=2K-2}^{3K-4} i \sum_{j=i+1}^{3K-3} j + \sum_{i=2K-2}^{3K-3} i \cdot \sum_{i=1}^{3K-4} i \sum_{j=i+1}^{3K-3} j \right) \\
& + \frac{K^2(K-1)^2(K-2)^2}{6} \left(\sum_{i=1}^{3K-5} i \sum_{j=i+1}^{3K-4} j + \sum_{i=2K-3}^{3K-5} i \sum_{j=i+1}^{3K-4} j \right) \\
& + \frac{K^2(K-1)^2(K-2)^2}{6} \sum_{i=1}^{3K-4} i \cdot \sum_{i=2K-3}^{3K-4} j \\
& - \frac{K^2(K-1)^2(K-2)^2(K-3)^2}{24} \left(\sum_{i=1}^{3K-5} i + \sum_{i=2K-4}^{3K-5} i \right) \\
& + \binom{4}{1}^5 \cdot \sum_{u=K}^{2K-5} u \sum_{l=u+1}^{2K-4} l \sum_{v=K+1}^{2K-3} v \sum_{i=v+1}^{2K-2} i \sum_{j=i+1}^{2K-1} j \\
& + \binom{4}{2} \binom{4}{1}^3 \\
& \times \sum_{l=K}^{2K-1} l^2 \left(\sum_{v=K}^{2K-3} v \sum_{i=m+1}^{2K-2} i \sum_{j=i+1}^{2K-1} j - l \sum_{i=K}^{2K-2} i \sum_{j=i+1}^{2K-1} j + l^2 \sum_{i=K}^{2K-1} -l^3 \right) \\
& + \binom{4}{2}^2 \binom{4}{1} \cdot \sum_{v=K}^{2K-1} v \left(\sum_{i=K}^{2K-2} i^2 \sum_{j=i+1}^{2K-1} j^2 - v^2 \sum_{i=K}^{2K-1} i^2 + v^4 \right)
\end{aligned}$$

$$\begin{aligned}
& + \binom{4}{3} \binom{4}{2} \sum_{i=K}^{2K-1} i^3 \left(\sum_{j=K}^{2K-1} j^2 - i^2 \right) \\
& + \binom{4}{3} \binom{4}{1}^2 \sum_{v=K}^{2K-1} v^3 \left(\sum_{i=K}^{2K-2} i \sum_{j=i+1}^{2K-1} j - v \sum_{i=K}^{2K-1} + v^2 \right) \\
& + \binom{4}{4} \binom{4}{1} \sum_{i=K}^{2K-1} i^4 \left(\sum_{j=K}^{2K-1} j - i \right) \\
= & -\frac{1}{120}K^{10} + \frac{1}{6}K^9 + \frac{199}{12}K^8 - 34K^7 + \frac{1207}{120}K^6 + \frac{161}{6}K^5 \\
& - \frac{209}{6}K^4 + 20K^3 - \frac{24}{5}K^2.
\end{aligned}$$

■

References

- [1] S.R. Blackburn and S. Gerke, “Connectivity of the uniform random intersection graph,” May 2008. arXiv:0805.2814v2 [math.CO]
- [2] R. Di Pietro, L.V. Mancini, A. Mei, A. Panconesi and J. Radhakrishnan, “Redoubtable sensor networks,” in *ACM Transactions on Information Systems Security TISSEC* **11** (2008), pp. 1-22.
- [3] P. Erdős and A. Rényi, “On the evolution of random graphs,” *Publ. Math. Inst. Hung. Acad. Sci.* **5** (1960), pp. 17-61.
- [4] 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.
- [5] J. Fill, E.R. Schneierman and K.B. Cohen-Singer, “Random intersection graphs when $m = \omega(n)$: An equivalence theorem relating the evolution of the $G(n, m, p)$ and $G(n, p)$ models,” *Random Structures and Algorithms* **16** (2000), pp. 249-258.
- [6] 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.
- [7] 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, Eds., Springer, Berlin (2007), pp. 67-74.

- [8] S. Janson, T. Łuczak and A. Ruciński, *Random Graphs*, Wiley-Interscience Series in Discrete Mathematics and Optimization, John Wiley & Sons, 2000.
- [9] M.K. Karoński, E.R. Schneinerman, and K.B. Singer-Cohen, “On random intersection graphs: The subgraph problem,” *Combinatorics, Probability and Computing* **8** (1999), pp. 131-159.
- [10] P. Marbach, “A lower-bound on the number of rankings required in recommender systems using collaborativ filtering,” CISS 2008, Princeton University, Princeton (NJ), March 2008.
- [11] K.B. Singer, *Random Intersection Graphs*, Ph.D. Thesis, The Johns Hopkins University, Baltimore (MD), 1995.
- [12] O. Yağan and A.M. Makowski, “On the random graph induced by a random key predistribution scheme under full visibility,” In Proceedings of the IEEE International Symposium on Information Theory (ISIT 2008), Toronto (ON), June 2008.
- [13] O. Yağan and A.M. Makowski, “Zero-one laws for connectivity in random key graphs,” Available online at <http://www.lib.umd.edu/drum/>, January 2009.