Prob. Th. Rel. Fields 78, 1–10 (1988)

Probability Theory and Related Fields

When Are Small Subgraphs of a Random Graph Normally Distributed?

Andrzej Ruciński *

Institute of Mathematics, Adam Mickiewicz University, Poznań, Poland

Summary. Let G be a graph and let X_n count copies of G in a random graph $K(n, p)$. The random variable $(X_n - E(X_n))/\sqrt{\mathrm{Var}(X_n)}$ is asymptotically normally distributed if and only if $np^m \to \infty$ and $n^2(1-p) \to \infty$, where $m = \max\{e(H)/|H|: H \subset G\}$. In addition to, and in connection with this main result we investigate the formula for $\mathrm{Var}(X_n)$ and the Poisson convergence of X_n.

1. Introduction

A random graph $K(n, p)$ is a graph on the vertex set $\{1, \ldots, n\}$ whose edges appear independently from each other and with probability $p = p(n)$. One of the classical questions of the theory of random graphs concerns the probability of existence and distribution of the number of copies of a given graph G one can encounter in $K(n, p)$. The aim of this paper is to establish all instances of $p(n)$ for which the above random variable is asymptotically normal.

We assume the reader is familiar with elementary notions from graph theory. For a graph G we denote by $|G|$ and $e(G)$ the number of vertices and edges of G, respectively. For a random variable X, $E(X)$ and $V(X)$ stand for the expectation and variance of X, respectively.

Let X_n be the number of subgraphs of a random graph $K(n, p)$ isomorphic to a graph G. It is already known ([2, 4, 9]) that $P(X_n > 0) \to 1\,(0)$ if $np^m \to \infty\,(0)$, where $m = \max\{d(H): H \subset G\}$, $d(H) = e(H)/|H|$ and "$H \subset G$" means "H is a subgraph of G". On the threshold, i.e., when $np^m \to c > 0$ one can reduce the problem of limit distributions of X_n to the case of balanced G (see [11]). (G is balanced if $d(G) = m$). Then all moments of X_n converge to positive constants but, in general, there is no way to derive a limit distribution from that. However, if G is strictly balanced, i.e., for all $H \subsetneq G$, $d(H) < d(G)$, then X_n converges to a Poisson distribution ([2, 6]). In fact, the inverse of last implication is also true ([10]).

* Part of the research was done during the author's stay in Division of Mathematics and Science, St. John's University, Staten Island, New York

Using the well-known relationship between the Poisson and normal distribution some authors have established the asymptotic normality of X_n just above the threshold, i.e., when $np^m \to \infty$ sufficiently slowly ([1, 5, 6]). The method they had chosen, however, imposed two artificial restrictions. First, all those results were valid only for strictly balanced graphs G. Secondly, they were valid for a short range of $p(n)$. Until now the best result has been due to Karoński [5] who proved that, for $\varepsilon = \min\{(d(G)-d(H))/|H| : H \subsetneqq G\}$ and $\alpha = |G|/[e(G)+\varepsilon|G|/(|G|-2)]$ if $np^m \to \infty$ but $n^\alpha p \to 0$ then $\tilde{X}_n \overset{\mathscr{D}}{\rightsquigarrow} N(0, 1)$, where

$$\tilde{X}_n = \frac{X_n - E(X_n)}{\sqrt{V(X_n)}}.$$

The result followed from the fact that, in this range of p, X_n is Poisson convergent – a notion introduced by Barbour [1].

It was already noticed that $\tilde{X}_n$ may be normally distributed even if G is not strictly balanced (see [10]). Recently, Nowicki and Wierman [8] have established, using the projection method for U-statistics, the asymptotic normality of $\tilde{X}_n$ for an arbitrary graph G if $np^{e(G)-1} \to \infty$ but $n^2(1-p) \to \infty$. In this paper we "close the book" by proving that $\tilde{X}_n \overset{\mathscr{D}}{\rightsquigarrow} N(0,1)$ iff $np^m \to \infty$ and $n^2(1-p) \to \infty$. This has been accomplished by the use of method of moments. For the sake of completeness and unification we give the proof of all possible sequences $p = p(n)$. However, to avoid technical difficulties we assume that for every $\varepsilon \geqq 0$ the limit of $n^\varepsilon p$ exists or $n^\varepsilon p$ diverges to infinity and the same is true for $n^\varepsilon(1-p)$.

In Sect. 2 we discuss the Poisson convergence of X_n and examine the behaviour of $V(X_n)$. Our main result is proven in Sect. 3.

2. Poisson convergence

Let X_n be a sequence of nonnegative, integer-valued random variables, $\lambda_n = E(X_n)$. Barbour [1] has defined the Poisson convergence of X_n by

$$d(X_n, Y_n) = \sup_{A \subset \{0,1,\ldots\}} |P(X_n \in A) - P(Y_n \in A)| \to 0 \quad \text{as} \quad n \to \infty,$$

where Y_n is a Poisson random variables with $E(Y_n) = \lambda_n$. Set

$$\bar{X}_n = \frac{X_n - \lambda_n}{\sqrt{\lambda_n}}.$$

(It is an easy exercise to prove that when $\lambda_n \to \infty$, the Poisson convergence of X_n implies $\bar{X}_n \overset{\mathscr{D}}{\rightsquigarrow} N(0, 1)$.) Barbour [1] applied this approach to X_n being the number of copies of a given graph G in $K(n, p)$ and found the following bound for $d(X_n, Y_n)$:

Let $G_1, G_2, \ldots$ be all copies of G in the complete graph on $\{1, \ldots, n\}$ and let $l_i = 1$ if $G_i \subset K(n, p)$, $l_i = 0$ otherwise. Then

$$d(X_n, Y_n) \leqq 2p^e + 2\lambda_n^{-1} \sum_{\substack{i \neq j \\ e(G_i \cap G_j) > 0}}\sum E(l_i l_j). \tag{1}$$

Using (1) and Theorem 2 from the next section we are in position to determine almost precisely the range of $p = p(n)$ for which X_n is Poisson convergent. Let $v = |G|$, $e = e(G)$, $d = d(G)$, $q = 1 - p$, and

$$\beta = \max\left\{\frac{v - |H|}{e - e(H)} : H \subsetneqq G\right\}.$$

Theorem 1. *Let X_n be the number of subgraphs of $K(n,p)$ isomorphic to a given graph G. Then X_n is Poisson convergent if and only if $np^d \to 0$ or $n^\beta p \to 0$ as $n \to \infty$*

Comments. 1. Since $\beta \leqq v(v-2)/[e(v-2) + e|H| - e(H)v] < \alpha$ for all $H \subsetneqq G$, Theorem 1 extends the results of Karoński from [5] (see Introduction).

2. $\beta < d^{-1}$ iff G is strictly balanced.

Proof of Theorem 1. Assume first that $np^d \to 0$.
Since, for every $A \subset \{0, 1, \ldots\}$,

$$\begin{aligned}|P(X_n \in A) - P(Y_n \in A)| &\leqq P(X_n > 0) + P(Y_n > 0) + |P(X_n = 0) - P(Y_n = 0)| \\ &\leqq 2P(X_n > 0) + 2P(Y_n > 0),\end{aligned}$$

one has $d(X_n, Y_n) \leqq 4\lambda_n \to 0$, the last convergence following from the fact that

$$\lambda_n \asymp n^v p^e = (np^d)^v.$$

If $n^\beta p \to 0$ then

$$\sum_{e(G_i \cap G_j) > 0}\sum E(l_i l_j) \asymp \sum_{\substack{H \subsetneqq G \\ e(H) > 0}} n^{2v - |H|} p^{2e - e(H)} = o(\lambda_n) \tag{2}$$

and, again, $d(X_n, Y_n) \to 0$. Set $\sigma_n^2 = V(X_n)$. Assume now that $n^2 q \to \infty$. It is easy to check (see below) that if $n^\beta p \to \infty$ then $\lambda_n = 0(\sigma_n^2)$. Thus, if, in addition, $np^m \to \infty$ then X_n cannot be Poisson convergent. Indeed, $\bar{X}_n = a_n \tilde{X}_n$, $a_n \to 0$, $\tilde{X}_n \xrightarrow{\mathscr{D}} N(0,1)$, and by Slutsky's theorem [3, p. 249] $\bar{X}_n \xrightarrow{\mathscr{D}} 0$.

Next we consider 3 subcases:

(a) G is not balanced. If $np^d \to c > 0$ then $P(X_n > 0) \to 0$ but $P(Y_n > 0) \to c_0 > 0$ which implies $\lim_{n \to \infty} d(X_n, Y_n) > 0$. If $np^d \to \infty$ but $np^m \to c \geqq 0$ then $\limsup_{n \to \infty} P(X_n > 0) < 1$ (see [10]) whereas $P(Y_n > 0) \to 1$ and again $d(X_n, Y_n) \not\to 0$.

(b) G is balanced but not strictly balanced. Now $\beta = d^{-1} = m^{-1}$. If $np^d \to c > 0$ then, as we already mentioned in the introduction, all moments of X_n converge to positive and finite limits different from those of $Po(\lambda)$, $\lambda = \lim_{n \to \infty} \lambda_n$ (see [10] for details). Thus, by [3, p. 254, Corollary 7] there exists a $k \in \{0, 1, \dots\}$ such that $\lim_{n\to\infty} P(X_n=k) \neq \lim_{n\to\infty} P(Y_n=k)$, and so $d(X_n, Y_n) \not\to 0$.

(c) G is strictly balanced. Let us find a precise asymptotic formula for σ_n^2. We have, provided $p \to 0$,

$$\sigma_n^2 = \sum_{i,j}^* Cov(I_i, I_j) \sim \sum_{i,j}^* E(I_i, I_j) = \sum_{\substack{H\subset G\\ e(H)>0}} \sum_{i,j}^{(H)} p^{2e-e(H)} \sim \sum_{\substack{H\subset G\\ e(H)>0}} c_H n^{2v-|H|} p^{2e-e(H)}, \quad (3)$$

where $\sum^{(H)}$ is taken over all pairs (i,j) such that $G_i \cap G_j = H$, $\sum^* = \sum\limits_{\substack{H\subset G\\ e(H)>0}}^{(H)}$, $c_H = \frac{\operatorname{aut}(H)}{\operatorname{aut}^2(G)} f(G, H)$, $\operatorname{aut}(K)$ is the number of automorphisms of the graph K and $f(G, H)$ is the number of copies of H in G.

It is visible now that if $n^\beta p \to c$ then $\lambda_n \sim c_0 \sigma_n^2$, $c_0 > 1$, and so $E(\bar{X}_n^2) \not\to 1$.

This excludes the Poisson convergence of X_n by [3, p. 254, Cor. 7], since $E(\bar{X}_n^4) = O(E(\tilde{X}_n^4)) = O(1)$ as shown in the proof of Theorem 2 below.

Finally observe that if $n^2 q \to c \in [0, \infty)$ then $\lim_{n\to\infty} p(X_n = t) = e^{-c/2}$, $t = \binom{n}{v} V!/\operatorname{aut}(G)$, whereas for all $k = 0, 1, \dots$ and all $\lambda > 0$, $e^{-\lambda}\lambda^k/k! < k^{-1/2}$ and so $\lim_{n\to\infty} P(Y_n = t) = 0$.

As we have already seen in the above proof, the main term of Barbour's estimate (1) is strongly related to the variance of X_n provided $p \to 0$ (compare (2) and (3) above). This is not an accident, what we will try to explain now. We have

$$V(X_n) \sim \sum_{\substack{H\subset G\\ e(H)>0}} c_H n^{2v-|H|} p^{2e-e(H)} (1 - p^{e(H)}).$$

We call a subgraph H of G a leading overlap of G if

$$V(X_n) = O(n^{2v-|H|} p^{2e-e(H)} (1 - p^{e(H)})).$$

Now, condition (2) is equivalent to the fact that the only leading overlap of a strictly balanced graph G is G itself. On the other hand, if a proper subgraph of G is a leading overlap then $E(X_n) \nsim V(X_n)$ and the Poisson convergence is unlikely. It is also interesting to know how the leading overlaps of G change as the order of p increases. If $p \not\to 0$ then clearly K_2 is the only leading overlap of G. In fact, K_2 becomes such as soon as $np^\gamma \to \infty$, where

$$\gamma = \max\left\{\frac{e(H)-1}{|H|-2} : H \subset G, e(H) > 1\right\}.$$

At the other end, when $np^m \to \infty$ arbitrarily slowly, the smallest subgraph G_1 which maximizes $d(H)$ is a leading overlap of G. For G strictly balanced, $G_1 = G$ and G remains the only leading overlap of itself as long as $n^\beta p \to 0$ (i.e., exactly as long as X_n in Poisson convergent). In between other subgraphs take their turns unless G is s-balanced, in which case the change from G to K_2 is very sudden. A graph G is called s-balanced if for every $H \subset G$, $e(H) > 1$,

$$\frac{e(H)-1}{|H|-2} \leqq \frac{e-1}{v-2}.$$

G is called strictly s-balanced if the above inequality is strict for all $H \neq G$. Notice that every s-balanced graph is strictly balanced unless it is a union of disjoint edges. For an s-balanced graph G, $\frac{1}{\gamma} = \beta = (v-2)/(e-1)$ and, assuming G is strictly balanced, the only leading overlap of G is G itself when $pn^\beta \to 0$ and K_2 when $pn^\beta \to \infty$.

If $pn^\beta \to c > 0$ then both G and K_2 are leading overlaps of G (the only ones if G is strictly s-balanced). For instance, every tree T is s-balanced but not strictly s-balanced. Therefore, when $np \to c > 0$, all connected subgraphs of T are leading overlaps. An example of a strictly balanced graph which is not s-balanced is the graph G with vertex-set $\{1, \ldots, 5\}$ such that the vertices 1, 2, 3, 4 form K_4 and the vertex 5 is joined to 1 and 2. Consequently, if $np^2 \to \infty$ but $n^2 p^5 \to 0$, K_4 is the leading overlap of G.

For further references we summarize here our knowledge about the asymptotic behaviour of the variance of X_n. Let $p = p(n) \to c$.

$$V(X_n) \sim \begin{cases} c_{K_2} n^{2v-2} c^{2e-1}(1-c) & \text{if } 0<c<1, \\ c_{K_2} n^{2v-2} q & \text{if } c=1, \\ \left(\sum_{i=1}^{u} a_i\right) n^{2v-|H|} p^{2e-e(H)} & \text{if } c=0, \end{cases}$$

where $a_i = c_{H_i} n^{|H|-|H_i|} p^{e(H)-e(H_i)}$ and $H_1, \ldots, H_u$ are all pairwise nonisomorphic leading overlaps of G.

3. *Asymptotic Normality*

As we have seen in Sect. 2, if G is strictly balanced, $np^m \to \infty$, and $n^\beta p \to 0$ then $\tilde{X}_n \overset{\mathscr{D}}{\rightsquigarrow} N(0, 1)$. There is no hope, however, to extend the normal phase of X_n any further using the technique of Poisson convergence. Surprisingly enough, the problem of asymptotic normality of X_n, the number of copies of G one can find in $K(n, p)$, can be solved once and for ever by the standard method of moments. This approach was inspired by the way Maehara applied the method of moments in [7].

Theorem 2. *Let G be an arbitrary graph with at least one edge. Then*

$$\tilde{X}_n \overset{\mathscr{D}}{\rightsquigarrow} N(0,1) \quad \text{if and only if } np^m \to \infty \quad \text{and} \quad n^2(1-p)\to\infty.$$

Moreover, if $n^2(1-p)\to c>0$ then $-\tilde{X}_n \overset{\mathscr{D}}{\rightsquigarrow} \tilde{P}o\left(\frac{c}{2}\right)$.

Proof I. Sufficiency. Set μ_k for the kth central moment of X_n. It is enough to prove

$$\mu_{2k} \sim \frac{(2k)!}{k!\,2^k}\mu_2^k \quad \text{and} \quad \mu_{2k+1} = o(\mu_2^{k+\frac{1}{2}}), \qquad k=1,2,\ldots. \tag{5}$$

Indeed, then

$$E(\tilde{X}_n^k) \to \begin{cases} 0 & \text{if } k \text{ is odd, (since } \mu_2\to\infty) \\ \dfrac{k!}{\left(\frac{k}{2}\right)!\,2^{k/2}} & \text{if } k \text{ is even} \end{cases}$$

and the thesis follows from the fact that the distribution of $N(0,1)$ is uniquely determined by its moments.

We split the proof of sufficiency into 3 cases according to the value of $c = \lim_{n\to\infty} p(n)$: $0<c<1$, $c=1$, $c=0$.

In each case we will make use of the expression

$$\mu_k = \sum\nolimits^{(*)} E[(l_{i_1}-p^e)\ldots(l_{i_k}-p^e)] = \sum\nolimits^{(*)} a(i_1,\ldots,i_k),$$

where the sum $\sum^{(*)}$ is taken over all sequences $(G_{i_1},\ldots,G_{i_k})$ of not necessarily distinct copies of G one can find in the complete graph with vertex set $\{1,\ldots,n\}$ which satisfy

$$\forall h=1,\ldots,k:\ e(G_{i_h}\cap \bigcup_{j\neq h} G_{i_j})>0. \tag{*}$$

(Let us recall that l_i is the indicator of the event "$G_i \subset K(n,p)$".)

Also we say that $(G_{i_1},\ldots,G_{i_k})$ satisfies (**) [(***)] if $\forall h=1,\ldots,k\ \exists$ *unique* $j\neq h$: $e(G_{i_h}\cap G_{i_j})>0$ [and, moreover, $e(G_{i_h}\cap G_{i_j})=1$].

We begin with the easiest case which may serve as the essence of the method applied in all three cases.

Case 1. $p\to c$, $0<c<1$. In this case μ_k is a polynomial in n of degree $\max|\bigcup_j G_{i_j}|$.

We have

$$\mu_k = \sum_l \sum_{|\cup G_{i_j}|=l} a(i_1,\ldots,i_k) = \sum\nolimits^{(***)} a(i_1,\ldots,i_k) + O(n^{kv-k-1}).$$

Thus

$$\mu_{2k} \sim \binom{2k}{2, \ldots, 2} \frac{1}{k!} c_{K_2}^k n^{2k(v-1)} c^{2e-1} (1-c)^k \sim \frac{(2k)!}{k!\,2^k} \mu_2^k.$$

On the other hand, if k is odd then no k-tuple of copies of G satisfies (***) and so $\mu_k = o(\mu_2^{k/2})$.

Case 2. $p \to 1$. Set $\bar{l}_j = 1 - l_j$, $E\bar{l}_j = 1 - p^e = v \sim eq$, $q = 1 - p$. Then

$$a(i_1, \ldots, i_k) = (-1)^k E[(\bar{l}_{i_1} - v) \ldots (\bar{l}_{i_k} - v)].$$

By the *FKG* inequality the term $E(\bar{l}_{i_1} \ldots \bar{l}_{i_k})$ dominates among all terms obtained by multiplying the product under expectation. Let $r = r(i_1, \ldots, i_k)$ be the minimum number of edges whose removal destroys all G_i's. Then $E(\bar{l}_{i_1} \ldots \bar{l}_{i_k}) \asymp q^r$ and there are $O(n^{k(v-2)+2r})$ such sequences $(i_1, \ldots, i_k)$. Thus, given r, the terms which dominate in μ_k correspond to k-tuples $(G_{i_1}, \ldots, G_{i_k})$ of copies of G clustered into r disjoint "star-shaped" bunches, i.e., all mutual intersections within a bunch are the very same single edge. We call such a k-tuple a "Milky Way". Note that for a "Milky Way" all terms of the form $E[(\prod_{j \in J} \bar{l}_j) v^{k-|J|}]$, $J \subsetneqq \{i_1, \ldots, i_k\}$, are $o(E(\bar{l}_{i_1} \ldots \bar{l}_{i_k}))$. Note also that if $(G_{i_1}, \ldots, G_{i_k})$ is not a "Milky Way" then $a(i_1, \ldots, i_k) = o(n^{k(v-2)+2r})$, where r has the above meaning. Hence

$$\mu_k \sim (-1)^k \sum_{r=1}^{\lfloor k/2 \rfloor} S_2(k, r)\, c_{K_2}^{k/2}\, 2^{k/2-r}\, n^{2r+k(v-2)}\, q^r, \tag{6}$$

where $S_2(k, r)$ is the number of unordered partitions of a k-element set into r classes of size at least 2. However, $n^2 q \to \infty$ and so

$$\mu_k \sim (-1)^k S_2(k, \lfloor k/2 \rfloor)\, n^{kv - 2\lceil k/2 \rceil} q^{\lfloor k/2 \rfloor} c_{K_2}^{k/2}\, 2^{k/2 - \lfloor k/2 \rfloor}.$$

Thus (5) is fulfilled.

Case 3. $p \to 0$. We will prove (5) by induction on k. For $k = 1, 2$ there is nothing to do. For $k \geqq 3$ let us assume that (5) holds for $t \leqq k-1$, which, in particular, implies that $\mu_t = O(\mu_2^{t/2})$. We split $\mu_k = A_k + B_k$, where $A_k = \sum^{(**)} a(i_1, \ldots, i_k)$. Recall that $A_k = 0$ for k odd. If $(G_{i_1}, \ldots, G_{i_k})$ satisfies (*) but not (**) there exists j such that $(G_{i_1}, \ldots, G_{i_{j-1}}, G_{i_{j+1}}, \ldots, G_{i_k})$ satisfies (*) too. The best way to see this is to imagine the hypergraph whose vertices are the edges of $\bigcup_s G_{i_s}$ and edges are the edge-sets $E(G_{i_s})$, $s = 1, \ldots, k$. There must be a connected component with at least 3 edges and one of them is just $E(G_{i_j})$. Since there may be more than one index j with the above property, we always choose the smallest one. Furthermore we denote $K = G_{i_j} \cap \bigcup_{s \neq j} G_{i_s}$. Of course, both j and K are functions of $(i_1, \ldots, i_k)$. This way we have defined a mapping between k-tuples satisfying

(*) but not (**) and $(k-1)$-tuples satisfying (*) in which every $(k-1)$-tuple is the image of $O(\sum_{\substack{K\subset G\\ e(K)>0}} r^{|G|-|K|})$ k-tuples. Hence

$$\begin{aligned}B_k &\sim \sum_{\sim(**)}^{(*)} E(l_{i_1}\dots l_{i_k}) = \sum_{\sim(**)}^{(*)} E(l_{i_1}\dots l_{i_{j-1}} l_{i_{j+1}}\dots l_{i_k})\, p^{e(G)-e(K)}\\ &= O\Big(\sum_{\substack{K\subset G\\ e(K)>0}} n^{|G|-|K|} p^{e(G)-e(K)} \sum{}^{(*)} E(l_{i_1}\dots l_{i_{k-1}})\Big)\\ &= O\Big(\sum_K n^{|G|-|K|} p^{e(G)-e(K)} \mu_{k-1}\Big) = o(\mu_2^{k/2}),\end{aligned}$$

since $\mu_{k-1}=O(\mu_2^{(k-1)/2})$ by the induction assumption and

$$n^{|G|-|K|} p^{e(G)-e(K)} = o(\mu_2^{1/2}),$$

the last following from the fact that

$$\infty \leftarrow n^{|H|} p^{e(H)} = O(n^{|K|} p^{e(K)})$$

for each leading overlap H of G. Thus $\mu_k = o(\mu_2^{k/2})$ for k odd. In order to prove the other part of (5) we partition $A_{2k}=C_{2k}+D_{2k}$, where the sum C_{2k} is taken over those $(G_{i_1},\dots,G_{i_{2k}})$ whose intersections are leading overlaps of G. Recall that $H\subset G$ is a leading overlap of G if $\mu_2=O(n^{2v-|H|}p^{2e-e(H)})$. To each $2k$-tuple of D_{2k} we associate a $(2k-2)$-tuple by removing the lexicographically first pair of copies of G which intersect on a nonleading subgraph of G. Since every $(2k-2)$-tuple is the image of $O(\sum_K n^{2v-|K|})$ $2k$-tuples

$$D_{2k} = O\Big(\sum_K n^{2v-|K|} p^{2e-e(K)} \mu_{2k-2}\Big) = o(\mu_2^k)$$

by the induction assumption and the fact that the sum is taken over all nonleading $K\subset G$, $e(K)>0$.

Finally, let $H=H_1, H_2, \dots, H_u$ be all, pairwise nonisomorphic leading overlaps of G. Then

$$\begin{aligned}C_{2k} &\sim \binom{2k}{2,\dots,2}\frac{1}{k!}\sum_{l_1+\dots+l_u=k}\binom{k}{l_1,\dots,l_u}\prod_{i=1}^{u} n^{2v-|H_i|} p^{2e-e(H_i)} c_{H_i}^{l_i}\\ &= \binom{2k}{2,\dots,2}\frac{1}{k!}\Big(\sum_{i=1}^{u} a_i\Big)^k (n^{2v-H} p^{2e-e(H)})^k = \frac{(2k)!}{2^k k!}\mu_2^k.\end{aligned}$$

II. Necessity. Set $\lambda_n=E(X_n)$, $\sigma_n^2=V(X_n)$. If $np^m\to 0$ then $\lambda_n=o(\sigma_n)$ and so, for every $\varepsilon>0$,

$$\begin{aligned}P(|\tilde{X}_n|>\varepsilon) &= P(X_n>\varepsilon\sigma_n+\lambda_n)+P(X_n/\sigma_n+\varepsilon<\lambda_n/\sigma_n)\\ &< P(X_n>0)+P(\varepsilon<\lambda_n/\sigma_n)\to 0.\end{aligned}$$

Hence $\tilde{X}_n \overset{\mathcal{D}}{\rightsquigarrow} 0$. If $np^m \to c > 0$ and G is strictly balanced then X_n converges to a Poisson random variable (see Introduction for the references). Assume now that G is not strictly balanced, i.e., there is $H \subsetneqq G$ with $d(H) = m$. It is easy to check that for $np^m \to c > 0$ H is a leading overlap of G if and only if $d(H) = m$. Thus B_k is no longer $o(\mu_2^{k/2})$. In particular, B_4 is at least equal to the sum of those terms $a(i_1, \ldots, i_4)$ which correspond to four copies of G mutually intersecting at H. So, $B_4 \geqq c_0 n^{4v - 3|H|} p^{4e - 3e(H)} \sim c_1 \mu_2^2$, $c_0, c_1 > 0$, and $\lim_{n \to \infty} E(\tilde{X}_n^4) \geqq 3 + c_1$. Moreover, it follows that $E(\tilde{X}_n^6) = O(1)$ which implies that $\tilde{X}_n \overset{\mathcal{D}}{\rightsquigarrow} N(0, 1)$ by [3, p. 254, Corollary 7]. Finally, if $n^2 q \to 0$ then we divide (4) by σ_n and after applying Markov's inequality we conclude that $\tilde{X}_n \overset{\mathcal{D}}{\rightsquigarrow} 0$.

III. The case $n^2 q \to c > 0$. Let us focus on formula (6). By inclusion – exclusion $S_2(k, r) = \sum_{l=0}^{r} (-1)^l \binom{k}{l} S(k-1, r-1)$, where $S(\,,\,)$ are the Stirling numbers of the second kind. After substituting and dividing by $\mu_2^{k/2}$ we get

$$E(\tilde{X}_n^k) \sim (-1)^k \lambda^{-k/2} \sum_{l=0}^{k} (-1)^l \sum_{r=0}^{k-l} S(k-1, r) \lambda^{r+l}, \qquad \lambda = c/2.$$

On the other hand, if Y is a Poisson random variable with expectation λ then

$$E(\tilde{Y}^k) = \lambda^{-k/2} \sum_{i=0}^{\infty} P(Y = i)(i - \lambda)^k = \lambda^{-k/2} \sum_{l=0}^{k} (-1)^l \binom{k}{l} \lambda^l \sum_{i=0}^{\infty} P(Y = i)\, i^{k-l}$$

$$= \lambda^{-k/2} \sum_{l=0}^{k} (-1)^l \binom{k}{l} \lambda^l E(Y^{k-l}).$$

But $E(Y^{k-l}) = \sum_{r=0}^{k-1} S(k-l, r) \lambda^r$ and so, for every $k = 1, 2, \ldots,$

$$\lim_{n \to \infty} E(\tilde{X}_n^k) = E((-\tilde{Y})^k),$$

which completes the proof, since $-\tilde{Y}$ is uniquely determined by its moments. □

Remark. Let Z_n be the number of nonedges in $K(n, p)$. Then $Z_n \overset{\mathcal{D}}{\rightsquigarrow} Po\left(\frac{c}{2}\right)$ provided $n^2 q \to c$.

Acknowledgements. I wish to thank Michał Karoński for drawing my attension to the paper of Maehara. I am also grateful to Tomasz Łuczak and the referee for valuable remarks leading to an improvement of the text and to simplification of the proof of Theorem 2.

References

1. Barbour, A.: Poisson convergence and random graphs. Math. Proc. Camb. Philos. Soc. **92**, 349–359 (1982)
2. Bollobás, B.: Threshold functions for small subgraphs. Math. Proc. Camb. Philos. Soc. **90**, 187–206 (1981)
3. Chow, Y.S., Teicher, H.: Probability theory. Springer Verlag, Berlin Heidelberg New York: 1978
4. Györi, E., Rothschild, B., Ruciński, A.: Every graph is contained in a sparsest possible balanced graph. Math. Proc. Camb. Philos. Soc. **98**, 397–401 (1985)
5. Karoński, M.: Balanced subgraphs of large random graphs. Poznań: AMU Press 1984
6. Karoński, M., Ruciński, A.: On the number of strictly balanced subgraphs of a random graph. In: Borowiecki, M., Kennedy, J.W., Syslo, M.M. (eds.). Proceedings of conference held in Lagów, Poland, 1981. Lect. Notes Math., vol. 1018, pp. 79–83. Berlin Heidelberg New York: Springer 1983
7. Maehara, H.: On random simplices in product distributions. J. Appl. Probab. **17**, 553–558 (1980)
8. Nowicki, K., Wierman, J.: Subgraph counts in random graphs using incomplete U-statistics method. Technical report. Lund 1986
9. Ruciński, A., Vince, A.: Strongly balanced graphs and random graphs. J. Graph Theory **10**, 251–264 (1986)
10. Ruciński, A., Vince, A.: Balanced graphs and the problem of subgraphs of random graphs, Congr. Numerantium **49**, 181–190 (1985)
11. Ruciński, A.: Small subgraphs of a random graph (a survey), Proceedings of conference "Random graphs '87" held in Poznań, 1987, submitted