The Perturbation Method and Triangle-free Random Graphs

Nicholas C. Wormald*
Department of Mathematics
University of Melbourne
Parkville, VIC 3052, AUSTRALIA

Abstract

When a discrete random variable in a discrete space is asymptotically Poisson, there is often a powerful method of estimating its distribution, by calculating the ratio of the probabilities of adjacent values of the variable. The versatility of this method is demonstrated by finding asymptotically the probability that a random graph has no triangles, provided the edge density is not too large. In particular, the probability that $G \in \mathcal{G}(n,p)$ has no triangles is asymptotic to $\exp(-\frac{1}{6}p^3n^3 + \frac{1}{4}p^5n^4 - \frac{7}{12}p^7n^5)$ for $p = o(n^{-2/3})$, and for $G \in \mathcal{G}(n,m)$ it is asymptotic to $\exp(-\frac{1}{6}d^3n^3)$ for $d = \frac{2m}{n(n-1)} = o(n^{-2/3})$.

1 Introduction

From time to time the following idea has been used to estimate the probability of an event in a probability space Ω.
1. Partition the whole space into events $\mathcal{C}_0, \mathcal{C}_1, \ldots, \mathcal{C}_m$.
2. Estimate the ratio $\mathbf{P}(\mathcal{C}_{k+1})/\mathbf{P}(\mathcal{C}_k)$ for $k = 0, \ldots, m-1$.
3. Obtain an approximation for $\mathbf{P}(\mathcal{C}_i)$ from

$$\mathbf{P}(\mathcal{C}_i)^{-1} = \sum_{k=0}^{m} \frac{\mathbf{P}(\mathcal{C}_k)}{\mathbf{P}(\mathcal{C}_i)},$$

by breaking the ratio in the summation into a telescoping product of ratios estimated in 2.

The aim of this article is to act as a tutorial on how to apply this idea to probability spaces in a quite general way. This will be done in the setting of random graph spaces, and for didactic reasons is restricted to a computation of

*Research supported by the Australian Research Council

the probability that a random graph has no triangles. A much more extensive and general development is given in [13].

The probabilities $\mathbf{P}(\mathcal{C}_{k+1})$ and $\mathbf{P}(\mathcal{C}_k)$ in Step 2 are called here *adjacent probabilities.* The idea is that the ratio of these adjacent probabilities can be estimated by altering elements of $\mathcal{C}_{k+1}$ slightly so as to obtain elements of Ω predominantly in $\mathcal{C}_k$, and vice versa. Various modifications of the main idea have also been used, such as taking a partition of almost all of the space rather than all of it, or taking a partition into events indexed by two or more variables rather than one. In this article, this general approach will be called the *perturbation* method. In practice, the perturbation method is used to show that a variable X is approximately Poisson by letting $\mathcal{C}_k$ be the event that $X = k$. The simplest applications use the following lemma or a similar one for estimating $\mathbf{P}(\mathcal{C}_k)$.

Lemma 1. *If*

$$\frac{\mathbf{P}(\mathcal{C}_{k+1})}{\mathbf{P}(\mathcal{C}_k)} = \frac{\lambda}{(k+1)}(1 + o(1/m))$$

for $k = 0, \ldots, m-1$ *and* $\sum_{k=0}^{m} \mathbf{P}(\mathcal{C}_k) \sim 1$ *then*

$$\mathbf{P}(\mathcal{C}_0) \sim e^{-\lambda}.$$

Proof. This is elementary using the idea described above. There are at most m factors in the telescoping product form of $\mathbf{P}(\mathcal{C}_k)/\mathbf{P}(\mathcal{C}_i)$ in step 3, with a relative error of $o(1/m)$ each, giving error $o(1)$ overall. ∎

However, in more complex situations, the correction terms to the ratio of adjacent probabilities are significant and need to be included.

In many previous applications of the perturbation method, Ω is a uniform probability space of combinatorial structures and so the ratios in Step 2 were computed by direct counting. Moreover, the transformation of elements of $\mathcal{C}_{k+1}$ to elements of $\mathcal{C}_k$ was effected by deleting one or two bits of the structure and replacing them (in a more or less random manner). (An exception to this is Stein's use of "exchangeable pairs" [12].) For graphs, the natural operation of this type is a "switching". The setting of the present article is, however, somewhat different, and the transformation is from $\mathcal{C}_k$ to $\mathcal{C}_{k+1}$ (or, with lower probabilities, to $\mathcal{C}_{k+i}$ for some small values of i) by adding part of the structure.

An important innovation in the present article is the fact that the error bounds in the estimation of the ratios of adjacent probabilities are proved inductively. To be able to do this in an uncomplicated manner seems to be a key factor contributing to the ability to extend the method to a natural boundary of precision; in previous applications somewhat ad-hoc methods were used to increase precision or range of applicability. However, the name "perturbation" will be used to apply to the present as well as previous methods, since they all involve a small alteration to the structure and analysis of the corresponding (usually) small alteration to the value of a substructure count.

We will be concerned here mainly with the random graph model $\mathcal{G}(n,p)$, in which the edges of a random graph on n vertices are generated independently with probability p each. The results obtained in $\mathcal{G}(n,p)$ can also be translated to results in $\mathcal{G}(n, m)$. This does not depend on the general relationship that exists between these two models, by which an approximate result for one tranfers directly to the other. The type of information gained in this paper is far more delicate than that, and the results do not carry over without change between the models. However, the accuracy obtained is sufficient for the correct translation of the results.

As this paper is intended as an exposition of the perturbation method, for simplicity of definitions and notation it will be applied only to the distribution of the number X of triangles in $G \in \mathcal{G}(n, p)$ and $G \in \mathcal{G}(n, m)$, and then, only to the probability that $X = 0$. In Section 2 this is done for $p = o(n^{-4/5})$ by dividing the space into classes based on how many triangles they have. For higher values of p we need to consider clusters of two or more "edge-connected" triangles. By keeping count of clusters of two triangles it is possible to cover all $p = o(n^{-5/7})$; in Section 3 we go one step further, considering clusters of three triangles, to reach $p = o(n^{-2/3})$. In Section 4, a slight reworking of the technique in Section 3 gives the probability that $X = 0$ in $\mathcal{G}(n, m)$ for $m = o(n^{4/3})$.

A natural upper limit for the scope of this method in the case of triangles is not reached until $p = o(n^{-1/2})$ and $m = o(n^{3/2})$. This will be made explicit in [13] amidst a development of the method in $\mathcal{G}(n, p)$, and similar models, which extends the results of the present paper, adding more bells and whistles in several ways. For instance, the point probabilities in the distribution of the number of subgraphs can be found accurately over much of the range of the distribution, it applies to all strictly balanced subgraphs, the accuracy can be increased, the density of the graph model can be increased, other graph models and other types of subgraphs can be examined, and some other interesting consequences result. This general development however requires more careful treatment of various error terms than is attempted in the present paper.

Briefly, here are the known results on the probability that a random graph in $\mathcal{G}(n, p)$ or $\mathcal{G}(n, m)$ has no triangles. All of these results apply to more general subgraphs than just triangles. Erdős and Rényi [1] showed for $p \sim c/n$ with c fixed that X is asymptotically Poisson with expectation $c^3/6$. Thus $\mathbf{P}(X = 0) \sim e^{-c^3/6}$. Janson, Luczak and Ruciński [4] extended this to show that

$$\mathbf{P}(X = 0) \sim e^{-p^3 n^3/6} \tag{1.1}$$

for $p = o(n^{-4/5})$, using Janson's inequality. This inequality actually gives the following upper bound for all values of p, the lower bound coming from the correlation inequality:

$$M \leq \mathbf{P}(X = 0) \leq M \exp\left(\frac{\Delta}{2(1 - p^3)}\right) \tag{1.2}$$

where $M = (1 - p^3)^{n(n-1)(n-2)/6}$ and $\Delta = 6\binom{n}{4}p^5$. The upper and lower bounds do not match asymptotically unless $p = o(n^{-4/5})$.

Frieze [2], working with a type of switching in $\mathcal{G}(n, m)$ (see below) obtained the probability that a random graph has k triangles asymptotically, for a wide range of k, provided $m = o(n^{1+\delta})$. He did not attempt to state the best value of δ for which his argument applies, but it is certainly less than $1/5$. Prömel and

Steger [10] recently showed that the asymptotic formula for $\mathbf{P}(X = 0)$ in $\mathcal{G}(n, p)$ determined by (1.2) applies also to $\mathcal{G}(n, m)$; that is, in $\mathcal{G}(n, m)$, $\mathbf{P}(X = 0)$ is asymptotically $\exp(-p^3n^3/6)$ where $p = 2m/n^2$ provided $m = o(n^{6/5})$.

For more dense graphs, it is a recent result of Prömel and Steger [9] that almost all triangle-free graphs in $\mathcal{G}(n, m)$ are bipartite provided $m > cn^{7/4} \log n$. They also show that this is false for $c_1n < m < c_2n^{3/2}$. Of course, these results in $\mathcal{G}(n, m)$ imply similar results in $\mathcal{G}(n, p)$.

This section closes with brief mention of many random structures in which the switching version of the perturbation method has already been used. The next section considers the problem of estimating the probability that a random graph in $\mathcal{G}(n, p)$ has no triangles. The perturbation method is applied here and extended in the following section, where the translation to $\mathcal{G}(n, m)$ is also performed.

Stein [11] used a switching technique to estimate the asymptotic number of $k \times n$ Latin rectangles, for $k = o(n^{1/2})$. In this case, the switching is a permutation of a few of the entries of a rectangle with possible multiple occurrences of a number in the same column, so that the number of such occurrences is reduced. (Later Godsil and McKay [3] used switchings in conjuction with an integral formula for the number of 1-factors in bipartite graphs, to estimate the number of ways of extending a random $k \times n$ Latin rectangle by one row. The result was an asymptotic formula for the number of $k \times n$ Latin rectangles valid for $k = o(n^{6/7})$. But in this case the switchings were used to prove results about a random structure rather than to estimate the ratio of probabilities of adjacent classes of structures.)

Stein [11] uses switchings in the form of exchangeable pairs, whose use is further developed in [12], including applications to random allocations and the cycle lengths of random permutations. The use of exchangeable pairs is possibly almost as flexible as the perturbation method itself. In both cases, it is clear that the flexibility can be useful and so it may be detrimental to be too fixed in defining either method. Perhaps because of the fuzziness of definition, it is possible to argue that they are identical. What is clear is that they can achieve similar results, and that the power in each case depends on the treatment of error terms.

McKay [6] used switchings in the pairing model of random graphs with given degrees to estimate the probability that a random pairing induces no loops or multiple edges in the graph. This gave rise to an asymptotic formula for the number of graphs with given degrees, provided the degrees are $o(n^{1/3})$. In this case a switching is the replacement of two pairs of the pairing and by another two pairs using the same four points. In terms of graphs, this replaces two edges by another two edges on the same four points. This is a natural operation here because it preserves the degree sequence. The type of switching was modified in [8] to extend results to most degree sequences with maximum degree $o(n^{1/2})$. Here, a three-variate analogue of Lemma 1 was used. It was also shown in [7] that the new switchings can be used to generate a graph with vertex degrees $O(n^{1/3})$ uniformly at random in polynomial time. McKay [5] earlier used switchings to estimate probabilities in the space of random regular graphs.

More recently, Frieze [2] used a type of switching in $\mathcal{G}(n, m)$ to estimate the distribution of the number of strictly balanced subgraphs of a given type. Here the switching is the replacement of one edge by another one randomly chosen subject to some restrictions. Thus, the total number of edges remains unchanged. This is natural, but does not seem to give as simple an argument or as strong a result as the present paper.

Acknowledgments This research received a great boost from many conversations and communications with Charles Stein in 1990, when we discussed our two similar approaches to the problems considered here. I would also like to thank one of the referees for a careful reading and corrections.

2 $\mathcal{G}(n,p)$ with $p = o(n^{-4/5})$

This section introduces the perturbation method in a simple way by re-deriving the formula (1.1) for the probability of no triangles in $G \in \mathcal{G}(n,p)$ when $p = o(n^{-4/5})$. We regard a *triangle* of G as a subset of $E(G)$ inducing a 3-cycle of G. A *triangle-set* of G is a nonempty subset of $E(G)$ which is the union of a set of triangles of G. A *cluster* of G is a triangle-set which cannot be partitioned into disjoint triangle-sets. A *maximal cluster* K of G is cluster contained in no larger cluster. Note that every maximal cluster K of G is a nonempty triangle-set of G such that every triangle of G not contained by K has empty intersection with K. A maximal cluster of cardinality 3 is therefore a triangle sharing no edges with any other triangle.

Assume $p = o(n^{-4/5})$ and put $q = 1-p$. Let T_1 denote the set of triangles of the complete graph K_n on n vertices, and $X = X(G)$ the number of triangles in G. Define $\lambda = \mathbf{E}X$ $(= \binom{n}{3} p^3)$. In this section, a *large* cluster is any cluster which is not a triangle. Let $\mathcal{C}_k$ denote the set of graphs on n vertices with exactly k triangles and no large cluster. Since the expected number of clusters of cardinality 5 (the union of two triangles sharing an edge) is exactly $6\binom{n}{4} p^5 = o(1)$,

$$\sum_f \mathbf{P}(\mathcal{C}_k) = 1 - o(1). \tag{2.1}$$

As a substitute for switchings, associate each graph in $\mathcal{C}_k$ with the graphs obtained by adding a triangle in every possible way. This gives the equation

$$|T_1|\mathbf{P}(\mathcal{C}_k) = \sum_{(D,G): D\in T_1,\ G\in\mathcal{C}_k} \mathbf{P}(G). \tag{2.2}$$

Now evaluate this summation by grouping the terms according to $G + D$, which denotes the graph with the same vertices as G, and edge set $E(G) \cup D$.

Case 1: D is a maximal cluster of $G + D$.

In this case, if $i < 3$ edges of D are in G, then $G + D \in \mathcal{C}_{k+1}$, $\mathbf{P}(G + D) = p^{3-i}\mathbf{P}(G)/q^{3-i}$, and furthermore each graph in $\mathcal{C}_{k+1}$ arises in exactly $\binom{3}{i}(k+1)$ ways from this construction. Thus the contribution to (2.2) from these terms is

$$\sum_{i=0}^{2} \binom{3}{i} \left(\frac{q}{p}\right)^{3-i} (k+1)\mathbf{P}(\mathcal{C}_{k+1}) = \left(\frac{1}{p^3} - 1\right)(k+1)\mathbf{P}(\mathcal{C}_{k+1}).$$

On the other hand, the terms in which all three edges of D are in G contribute

$$k\mathbf{P}(\mathcal{C}_k)$$

to (2.2).
Case 2: D is contained in a large cluster of $G+D$.
In this case there must exist two edges x_1 and x_2 of G creating a second triangle with D. There are $0 \le i \le 2$ triangles of G containing x_1 or x_2. Deleting x_1 and x_2 and all other edges of these i triangles creates a graph G'. Then $G' \in \mathcal{C}_{k-i}$, and these terms contribute

$$O(p^2n^4)\mathbf{P}(\mathcal{C}_k) + O(p^3n^4 + p^4n^5)\mathbf{P}(\mathcal{C}_{k-1}) + O(p^6n^6)\mathbf{P}(\mathcal{C}_{k-2})$$

to (2.2).

Collecting the above contributions and noting that $|T_1| = \binom{n}{3}$, we obtain

$$\begin{aligned} |T_1|\mathbf{P}(\mathcal{C}_k) &= \left(\frac{1}{p^3} - 1\right)(k+1)\mathbf{P}(\mathcal{C}_{k+1}) + k\mathbf{P}(\mathcal{C}_k) \\ &\quad + O(p^2n^4)\mathbf{P}(\mathcal{C}_k) + O(p^3n^4 + p^4n^5)\mathbf{P}(\mathcal{C}_{k-1}) + O(p^6n^6)\mathbf{P}(\mathcal{C}_{k-2}). \end{aligned}$$

Hence

$$\frac{\mathbf{P}(\mathcal{C}_{k+1})}{\mathbf{P}(\mathcal{C}_k)} = \frac{\lambda}{k+1}(1 + O(\epsilon))$$

where

$$\epsilon = \frac{k}{n^3} + p^3 + p^2n + p^3n(1+pn)\frac{\mathbf{P}(\mathcal{C}_{k-1})}{\mathbf{P}(\mathcal{C}_k)} + p^6n^3\frac{\mathbf{P}(\mathcal{C}_{k-2})}{\mathbf{P}(\mathcal{C}_k)}.$$

Writing $\frac{\mathbf{P}(\mathcal{C}_{k-2})}{\mathbf{P}(\mathcal{C}_k)}$ as $\frac{\mathbf{P}(\mathcal{C}_{k-2})}{\mathbf{P}(\mathcal{C}_{k-1})}\frac{\mathbf{P}(\mathcal{C}_{k-1})}{\mathbf{P}(\mathcal{C}_k)}$, it follows by induction from this equation that for $k < C\lambda$ (C any constant),

$$\frac{\mathbf{P}(\mathcal{C}_{k+1})}{\mathbf{P}(\mathcal{C}_k)} = \frac{\lambda}{k+1}(1 + O(p^2n)). \tag{2.3}$$

Assume firstly that $pn \to \infty$. Then the expected number of sets of k disjoint triangles in $G \in \mathcal{G}(n,p)$ is at most

$$\frac{\binom{n}{3}^k}{k!}p^{3k} < \left(\frac{ep^3n^3}{6k}\right)^k. \tag{2.4}$$

Hence, $\sum_{k \ge 3\lambda} \mathbf{P}(\mathcal{C}_k) = o(1)$. Thus, we can choose $m = \lfloor 3\lambda \rfloor$ in Lemma 1, and its hypotheses are satisfied in view of (2.1) and (2.3). On the other hand, if $pn = O(1)$, then $m = \log n$ suffices by a similar argument. Hence

$$\mathbf{P}(X = 0) \sim e^{-\lambda}$$

for $p = o(n^{-4/5})$, as first shown in [4].

3 $\mathcal{G}(n,p)$ with several types of small clusters

The method of the previous section can be extended to treat $\mathcal{G}(n,p)$ with larger values of p by keeping track of the numbers of maximal clusters with a small number of triangles. In this section we do so for $p = o(n^{-2/3})$. Again put $q = 1-p$. Cluster definitions are as in the previous section, with the exception that a *small* cluster is redefined to be any cluster consisting of just one triangle (which we say is of type 1), or the union of two triangles (type 2) or three triangles on five vertices. In the last case there are two possibilities: all three triangles share one edge (type $3a$) or not (type $3b$). Any other clusters are called *large*. Then the probability that $G \in \mathcal{G}(n,p)$ has a large cluster is $o(1)$.

Theorem 1. *The probability that $G \in \mathcal{G}(n,p)$ has no triangles is asymptotic to*

$$\exp\left(-\frac{1}{6}p^3n^3 + \frac{1}{4}p^5n^4 - \frac{7}{12}p^7n^5\right)$$

for $p = o(n^{-2/3})$.

Proof. Let $\mathcal{S}$ denote the set $\{1, 2, 3a, 3b\}$ of small types. For each $t \in \mathcal{S}$, let T_t denote the set of subsets of K_n which can form a cluster of type t, let $X_t = X_t(G)$ denote the number of clusters of type t in a graph G, and define $\lambda_t = \mathbf{E}X_t$. Note that since X_t counts *all* clusters of type t , not just maximal ones,

$$\begin{aligned}
\lambda_1 &= \frac{p^3n^3}{6}(1 + O(\tfrac{1}{n})),\\
\lambda_2 &= \frac{p^5n^4}{4}(1 + O(\tfrac{1}{n})),\\
\lambda_{3a} &= \frac{p^7n^5}{12}(1 + O(\tfrac{1}{n})),\\
\lambda_{3b} &= \frac{p^7n^5}{2}(1 + O(\tfrac{1}{n})).
\end{aligned}$$

For any non-negative integer valued function f with domain $\mathcal{S}$, let $\mathcal{C}_f$ denote the set of graphs on n vertices with no large clusters and exactly $f(t)$ maximal clusters of type t for each $t \in \mathcal{S}$. For each of the four values of $t \in \mathcal{S}$, we will use the equation

$$|T_t|\mathbf{P}(\mathcal{C}_f) = \sum_{(D,G): D\in T_t,\ G\in\mathcal{C}_f} \mathbf{P}(G), \tag{3.1}$$

analogous to (2.2). In each case, u will denote the type of the maximal cluster K of $G + D$ which contains D. We define

$$\delta_t(u) = \begin{cases} 1 & u = t \\ 0 & \text{otherwise.} \end{cases}$$

(i) $t = 1$. First consider $u = 1$. This is exactly the same as Case 1 in the previous section, which gives a contribution to (3.1) of

$$\left(\frac{1}{p^3} - 1\right)(f(1) + 1)\mathbf{P}(\mathcal{C}_{f+\delta_1}) + f(1)\mathbf{P}(\mathcal{C}_f).$$

Secondly, consider $u = 2$. In this case K contains D together with two edges x_1 and x_2 of G completing another triangle with D. If the third edge of the triangle with x_1 and x_2 is not in $E(G)$, then $G + D \in \mathcal{C}_{f+\delta_2}$, and G can be reconstructed from $G + D$ by selecting a maximal cluster of type 2 and deleting some edges. Therefore, the contribution here is

$$\left(\frac{2}{p^3} + O(p^{-2})\right)(f(2)+1)\mathbf{P}(\mathcal{C}_{f+\delta_2}),$$

where the factor 2 occurs because there are two ways to select the triangle D in the selected cluster of type 2. On the other hand, if x_1 and x_2 are in a triangle of G, then this triangle is not a maximal cluster of $G + D$. In this case, either $D \subseteq G$ and $G + D \in \mathcal{C}_f$, contributing

$$2f(2)\mathbf{P}(\mathcal{C}_f),$$

or $G + D \in \mathcal{C}_{f+\delta_2-\delta_1}$, contributing

$$O(p^{-2})(f(2)+1)\mathbf{P}(\mathcal{C}_{f+\delta_2-\delta_1}).$$

Thirdly, consider $u = 3a$. Then $|K - D| \geq 4$. If $K \cap E(G)$ contains no triangles, then $G + D \in \mathcal{C}_{f+\delta_{3a}}$, and the contribution here is

$$\left(\frac{3}{p^3} + O(p^{-2})\right)(f(3a)+1)\mathbf{P}(\mathcal{C}_{f+\delta_{3a}}).$$

On the other hand, if $K \cap E(G)$ contains a triangle, then either $G + D \in \mathcal{C}_{f+\delta_{3a}-\delta_2}$ or $G + D \in \mathcal{C}_f$, contributing

$$O(p^{-2})(f(3a)+1)\mathbf{P}(\mathcal{C}_{f+\delta_{3a}-\delta_2}) + 3f(3a)\mathbf{P}(\mathcal{C}_f).$$

Fourthly, consider $u = 3b$. This is a little different from $u = 3a$. If $K \cap E(G)$ contains no triangles, the contribution is

$$\left(\frac{1}{p^3} + O(p^{-2})\right)(f(3b)+1)\mathbf{P}(\mathcal{C}_{f+\delta_{3b}}).$$

It is also possible that $K \cap E(G)$ contains just one triangle, in which case $G + D \in \mathcal{C}_{f+\delta_{3b}-\delta_1}$, and the contribution is

$$\left(\frac{2}{p^3} + O(p^{-2})\right)(f(3b)+1)\mathbf{P}(\mathcal{C}_{f+\delta_{3b}-\delta_1}).$$

Similarly, the other possibilities here contribute

$$O(p^{-2})(f(3b)+1)(\mathbf{P}(\mathcal{C}_{f+\delta_{3b}-\delta_2}) + \mathbf{P}(\mathcal{C}_{f+\delta_{3b}-2\delta_1})) + 3f(3b)\mathbf{P}(\mathcal{C}_f).$$

Finally, suppose that K is large. The situation here is much more complicated than the analogous Case 2 in the previous section. Choose a large subcluster L of K such that $D \subseteq L$ and L has no large proper subclusters. Since all small clusters have cardinality at most 7, it follows that $|L| \leq 8$. Otherwise, some edge or pair of edges in a triangle in L could be deleted to leave a large proper subcluster of L. At most $|L|$ distinct maximal clusters of G can have non-trivial intersection with L. Since G has no clusters of cardinality greater than 7, the union of L with all these maximal clusters is a large cluster J of $G + D$ containing at most 56 edges. Deleting J from $G+D$ gives a graph G' such that $\mathbf{P}(G) = O(p^{|J|-3})\mathbf{P}(G')$ and such that the maximal clusters of G' are the same as those of G except for the omission of the small maximal clusters of $G \cap J$. For $t \in \mathcal{S}$ let $h(t)$ be the number of maximal clusters of $G \cap J$ of type t. Then the function h with domain $\mathcal{S}$ belongs to a finite set H_1 of bounded non-negative integer functions. The number of possible positions of J gives an upper bound on the number of G corresponding to G'. Since this number of positions multiplied by $p^{|J|}$ is the expected number of clusters of the same type as J, which is $o(1)$ as J is large, the contribution to (3.1) for K large is

$$o(p^{-3}) \sum_{h \in H_1} \mathbf{P}(\mathcal{C}_{f-h}).$$

For an integer-valued function h defined on $\mathcal{S}$, write

$$\rho(h) = \rho(f,h) = \frac{\mathbf{P}(\mathcal{C}_{f+h})}{\mathbf{P}(\mathcal{C}_f)}.$$

Then the above evaluation of (3.1) can be written as

$$\begin{aligned}\rho(f,\delta_1) \quad &= \quad \frac{\lambda_1}{f(1)+1}\Bigg(1 - \frac{2(f(2)+1)\rho(\delta_2) + 3(f(3a)+1)\rho(\delta_{3a})}{\lambda_1} \\ &\qquad - \frac{(f(3b)+1)(\rho(\delta_{3b}) + 2\rho(\delta_{3b}-\delta_1))}{\lambda_1} + O(\epsilon_1)\Bigg) \qquad (3.2)\end{aligned}$$

where

$$\begin{aligned}\epsilon_1 \quad &= \quad (f(2)+1)\frac{\rho(\delta_2)+\rho(\delta_2-\delta_1)}{p^2n^3} + (f(3a)+1)\frac{\rho(\delta_{3a})+\rho(\delta_{3a}-\delta_2)}{p^2n^3} \\ &\quad +(f(3b)+1)\frac{\rho(\delta_{3b})+\rho(\delta_{3b}-\delta_1)+\rho(\delta_{3b}-\delta_2)+\rho(\delta_{3b}-2\delta_1)}{p^2n^3} \\ &\quad +p^3 + \frac{f(1)+f(2)+f(3a)+f(3b)}{n^3} + \sum_{h\in H_1}\frac{o(\rho(-h))}{\lambda_1}.\end{aligned}$$

(ii) $t = 2$. Arguments similar to those above give a contribution to (3.1) of

$$\left(\frac{1}{p^5} + O(p^{-4})\right)(f(2)+1)\mathbf{P}(\mathcal{C}_{f+\delta_2}) + O(p^{-2})(f(2)+1)\mathbf{P}(\mathcal{C}_{f+\delta_2-\delta_1}) + O(f(2))\mathbf{P}(\mathcal{C}_f)$$

for $u = 2$,

$$\begin{aligned}&\left(\frac{3}{p^5} + O(p^{-4}))\right)(f(3a)+1)\mathbf{P}(\mathcal{C}_{f+\delta_{3a}}) + O(f(3a))\mathbf{P}(\mathcal{C}_f) \\ &\qquad +O(p^{-4})(f(3a)+1)(\mathbf{P}(\mathcal{C}_{f+\delta_{3a}-\delta_1}) + \mathbf{P}(\mathcal{C}_{f+\delta_{3a}-\delta_2}))\end{aligned}$$

for $u = 3a$,

$$\left(\frac{2}{p^5} + O(p^{-4})\right)(f(3b)+1)\mathbf{P}(\mathcal{C}_{f+\delta_{3b}}) + O(f(3b))\mathbf{P}(\mathcal{C}_f)$$
$$+O(p^{-4})(f(3b)+1)(\mathbf{P}(\mathcal{C}_{f+\delta_{3b}-\delta_1}) + \mathbf{P}(\mathcal{C}_{f+\delta_{3b}-\delta_2}) + \mathbf{P}(\mathcal{C}_{f+\delta_{3b}-2\delta_1}))$$

for $u = 3b$, and

$$o(p^{-5}) \sum_{h \in H_2} \mathbf{P}(\mathcal{C}_{f-h})$$

for K large, where H_2 is a bounded set of bounded functions like H_1. Hence, since $|T_2| = 6\binom{n}{4}$ and $\lambda_2 = p^5|T_2|$,

$$\rho(f,\delta_2) = \frac{\lambda_2}{f(2)+1}\left(1 - \frac{3(f(3a)+1)\rho(\delta_{3a}) + 2(f(3b)+1)\rho(\delta_{3b})}{\lambda_2} + O(\epsilon_2)\right) \tag{3.3}$$

where

$$\begin{aligned}
\epsilon_2 \;\; = \;\; & (f(2)+1)\frac{\rho(\delta_2-\delta_1)}{p^4n^4} + (f(3a)+1)\frac{\rho(\delta_{3a}) + \rho(\delta_{3a}-\delta_1) + \rho(\delta_{3a}-\delta_2)}{p^4n^4} \\
& +(f(3b)+1)\frac{\rho(\delta_{3b}) + \rho(\delta_{3b}-\delta_1) + \rho(\delta_{3b}-\delta_2) + \rho(\delta_{3b}-2\delta_1)}{p^4n^4} \\
& +p + \frac{f(2)+f(3a)+f(3b)}{n^4} + \sum_{h\in H_2}\frac{o(\rho(-h))}{\lambda_2}.
\end{aligned}$$

(iii) $t = 3a$. Here the contributions to (3.1) are

$$\left(\frac{1}{p^7} + O(p^{-6})\right)(f(3a)+1)\mathbf{P}(\mathcal{C}_{f+\delta_{3a}}) + O(f(3a))\mathbf{P}(\mathcal{C}_f)$$
$$+O(p^{-6})(f(3a)+1)(\mathbf{P}(\mathcal{C}_{f+\delta_{3a}-\delta_1}) + \mathbf{P}(\mathcal{C}_{f+\delta_{3a}-\delta_2}))$$

for $u = 3a$, nothing for $u = 3b$, and

$$o(p^{-7}) \sum_{h \in H_{3a}} \mathbf{P}(\mathcal{C}_{f-h})$$

for K large, where H_{3a} is similar to H_1 and H_2. Hence, since $|T_{3a}| = 10\binom{n}{5}$ and $\lambda_{3a} = p^7|T_{3a}|$,

$$\rho(f,\delta_{3a}) = \frac{\lambda_{3a}}{f(3a)+1}(1 + O(\epsilon_{3a})) \tag{3.4}$$

where

$$\epsilon_{3a} = (f(3a)+1)\frac{\rho(\delta_{3a}-\delta_1) + \rho(\delta_{3a}-\delta_2)}{p^6n^5} + p + \frac{f(3a)}{n^4} + \sum_{h\in H_{3a}}\frac{o(\rho(-h))}{\lambda_{3a}}.$$

(iv) $t = 3b$. Here the contributions to (3.1) are

$$\left(\frac{1}{p^7} + O(p^{-6})\right)(f(3b)+1)\mathbf{P}(\mathcal{C}_{f+\delta_{3b}}) + O(f(3b))\mathbf{P}(\mathcal{C}_f)$$
$$+O(p^{-6})(f(3b)+1)(\mathbf{P}(\mathcal{C}_{f+\delta_{3b}-\delta_1}) + \mathbf{P}(\mathcal{C}_{f+\delta_{3b}-\delta_2}) + \mathbf{P}(\mathcal{C}_{f+\delta_{3b}-2\delta_1}))$$

for $u = 3b$, and

$$o(p^{-7}) \sum_{h \in H_4} \mathbf{P}(\mathcal{C}_{f-h})$$

for K large, where H_4 is similar to H_1, H_2 and H_3.

Hence, since $|T_{3b}| = 60 \binom{n}{5}$ and $\lambda_{3b} = p^7 |T_{3b}|$,

$$\rho(f, \delta_{3b}) = \frac{\lambda_{3b}}{f(3b)+1}(1 + O(\epsilon_{3b})) \tag{3.5}$$

where

$$\begin{aligned} \epsilon_{3b} &= (f(3b)+1)\frac{\rho(\delta_{3b}-\delta_1) + \rho(\delta_{3b}-\delta_2) + \rho(\delta_{3b}-2\delta_1)}{p^6 n^5} \\ &\quad + p + \frac{f(3b)}{n^4} + \sum_{h \in H_{3b}} \frac{o(\rho(-h))}{\lambda_{3b}}. \end{aligned}$$

Define for $t \in \mathcal{S}$

$$\gamma(t) = \gamma(f,t) = \frac{\rho(f,\delta_t)(f(t)+1)}{\lambda_t} = \frac{(f(t)+1)\mathbf{P}(\mathcal{C}_{f+\delta_t})}{\lambda_t \mathbf{P}(\mathcal{C}_f)}.$$

Then, for example,

$$(f(3b)+1)\rho(\delta_{3b}-\delta_1) = \lambda_{3b}\gamma(f-\delta_1, 3b)\rho(-\delta_1) = \frac{\lambda_{3b}\gamma(f-\delta_1,3b)(f(1)+1)}{\gamma(1)\lambda_1},$$

and (3.2)–(3.5) become

$$\begin{aligned} \gamma(f,1) &= 1 - \frac{2\lambda_2}{\lambda_1}\gamma(2) - \frac{3\lambda_{3a}}{\lambda_1}\gamma(3a) - \frac{\lambda_{3b}}{\lambda_1}\gamma(3b) \\ &\qquad - \frac{2\lambda_{3b}\gamma(f-\delta_1,3b)(f(1)+1)}{\gamma(1)\lambda_1^2} + O(\epsilon_1') && (3.6) \\ \gamma(f,2) &= 1 - \frac{3\lambda_{3a}}{\lambda_2}\gamma(3a) - \frac{2\lambda_{3b}}{\lambda_2}\gamma(3b) + O(\epsilon_2') && (3.7) \\ \gamma(f,3a) &= 1 + O(\epsilon_{3a}') && (3.8) \\ \gamma(f,3b) &= 1 + O(\epsilon_{3b}') && (3.9) \end{aligned}$$

where

$$\begin{aligned} \epsilon_1' &= \frac{\lambda_2\hat\gamma(2) + \lambda_{3a}\hat\gamma(3a) + \lambda_{3b}\hat\gamma(3b)}{p^2 n^3}(1 + \rho(-\delta_1) + \rho(-\delta_2) + \rho(-2\delta_1)) \\ &\quad + p^3 + \frac{f(1)+f(2)+f(3a)+f(3b)}{n^3} + \sum_{h \in H_1} \frac{o(\rho(-h))}{\lambda_1} \\ \epsilon_2' &= \frac{\lambda_2\hat\gamma(2) + \lambda_{3a}\hat\gamma(3a) + \lambda_{3b}\hat\gamma(3b)}{p^4 n^4}(1 + \rho(-\delta_1) + \rho(-\delta_2) + \rho(-2\delta_1)) \\ &\quad + p + \frac{f(2)+f(3a)+f(3b)}{n^4} + \sum_{h \in H_2} \frac{o(\rho(-h))}{\lambda_2} \end{aligned}$$

$$
\begin{aligned}
\epsilon'_{3a} &= \frac{\lambda_{3a}\hat{\gamma}(3a)(\rho(-\delta_1)+\rho(-\delta_2))}{p^6n^5} + p + \frac{f(3a)}{n^4} + \sum_{h\in H_{3a}} \frac{o(\rho(-h))}{\lambda_{3a}} \\
\epsilon'_{3b} &= \frac{\lambda_{3b}\hat{\gamma}(3b)(\rho(-\delta_1)+\rho(-\delta_2)+\rho(-2\delta_1))}{p^6n^5} + p + \frac{f(3b)}{n^4} + \sum_{h\in H_{3b}} \frac{o(\rho(-h))}{\lambda_{3b}},
\end{aligned}
$$

and

$$
\hat{\gamma}(t) = \gamma(f,t) + \gamma(f-\delta_1,t) + \gamma(f-\delta_2,t) + \gamma(f-2\delta_1,t).
$$

We now derive asymptotic estimates of $\gamma(f,t)$, $t \in \mathcal{S}$, by induction on f. We can identify f with the vector $(f(1), f(2), f(3a), f(3b))$. The induction proceeds on the lexicographic ordering of these vectors; that is, $f < g$ iff $f \neq g$ and f has a smaller value than g at the first component where they differ. The induction requires no special initial step; we need only to keep in mind that $\rho(-\delta_t) = 0$ if $f(t) = 0$.

As in the previous section, the upper bound on the value of $f(t)$ which needs to be considered depends on whether $\lambda_t \to \infty$ for each $t \in \mathcal{S}$. First, assume that $p^7n^5 \to \infty$. In this case, define

$$
S = \{f : 0 \le f(t) < \lfloor 3\lambda_t \rfloor \text{ for each } t \in \mathcal{S}\}.
$$

We prove by induction on f in S that

$$
\begin{aligned}
\gamma(f,1) &= 1 - 3p^2n + \frac{21}{2}p^4n^2 - 6p^4n^2\frac{f(1)}{\lambda_1} + o(1/\lambda_1) && (3.10)\\
\gamma(f,2) &= 1 - 5p^2n + o(1/\lambda_2) && (3.11)\\
\gamma(f,3a) &= 1 + o(1/\lambda_{3a}) && (3.12)\\
\gamma(f,3b) &= 1 + o(1/\lambda_{3b}). && (3.13)
\end{aligned}
$$

Note that since $p = o(n^{-2/3})$ and $p^7n^5 \to \infty$, (3.10)–(3.13) imply $\gamma(f,t) \sim 1$ for $t \in \mathcal{S}$.

Now assume (3.10)–(3.13) for all $g < f \in S$. Since $f - \delta_t < f$ and $f(t)/\lambda_t < 3$ for any $t \in \mathcal{S}$,

$$
\rho(f,-\delta_t) = \frac{1}{\rho(f-\delta_t,\delta_t)} = \frac{f(t)}{\lambda_t\gamma(f-\delta_t,t)} = O(1)
$$

Similarly, for any non-negative function h defined on $\mathcal{S}$, $\rho(-h) = O(1)$. Hence, using $p^3n^2 = o(1)$, we obtain

$$
\epsilon'_{3b} = O(p\gamma(f,3b) + p) + o(1/\lambda_{3b})
$$

and then (3.13) from (3.9). A similar argument gives (3.12). Using these,

$$
\epsilon'_2 = O(p\gamma(2) + p) + o(1/\lambda_2)
$$

and so (3.11) comes from (3.7), (3.12) and (3.13). From here we obtain (3.10) in a similar fashion.

Since $p^7n^5 \to \infty$, arguing as at (2.4) gives

$$\sum_{f(t)\geq 3\lambda_t} \mathbf{P}(\mathcal{C}_f) = o(1)$$

for each $t \in \mathcal{S}$. Therefore, following the line of the proof of Lemma 1, we see that with $m_t = \lfloor 3\lambda_t \rfloor$,

$$\mathbf{P}(X_1 = 0)^{-1} \sim \sum_{0\leq f(t)\leq m_t \ \ \forall t\in\mathcal{S}} \frac{\mathbf{P}(\mathcal{C}_f)}{\mathbf{P}(\mathcal{C}_0)}. \tag{3.14}$$

Choose an ordering, say $1 < 2 < 3a < 3b$, of the elements of $\mathcal{S}$. For f as in (3.14) and $t \in \mathcal{S}$, define $f^*_{t,i}$ to be the function agreeing with f on each $t' < t$, with $f^*_{t,i}(t) = i$, and with value 0 on each $t' > t$. Rewriting (3.14),

$$\begin{aligned}
\mathbf{P}(X_1 = 0)^{-1} \quad &\sim \quad \sum_{0\leq f(t)\leq m_t \ \ \forall t\in\mathcal{S}} \ \prod_{\substack{t\in\mathcal{S} \\ i=0,\ldots,f(t)-1}} \rho(f^*_{t,i}, \delta_t) \\
&= \quad \sum_{0\leq f(t)\leq m_t \ \ \forall t\in\mathcal{S}} \ \prod_{\substack{t\in\mathcal{S} \\ i=0,\ldots,f(t)-1}} \gamma(f^*_{t,i}, t)\frac{\lambda_t}{f^*_{t,i}(t)+1} \\
&\sim \quad \sum_{0\leq f(t)\leq m_t \ \ \forall t\in\mathcal{S}} \ \prod_{t\in\mathcal{S}} \prod_{i=0}^{f(t)-1} \frac{\gamma(i\delta_t, t)\lambda_t}{i+1} \\
&= \quad \prod_{t\in\mathcal{S}} \sum_{0\leq j\leq m_t} \prod_{i=0}^{j-1} \frac{\gamma(i\delta_t, t)\lambda_t}{i+1}.
\end{aligned} \tag{3.15}$$

For $t = 1$, the summation here is, using (3.10),

$$\begin{aligned}
&\sum_{0\leq j\leq m_t} \prod_{i=0}^{j-1} \left(1 - 3p^2n + \frac{21}{2}p^4n^2 - 36pn^{-1}i + o(\lambda_1^{-1})\right) \frac{\lambda_1}{i+1} \\
&= \sum_{0\leq j\leq m_t} \left(\prod_{i=0}^{j-1} \left(1 - 3p^2n + \frac{21}{2}p^4n^2 + o(\lambda_1^{-1})\right) \frac{\lambda_1}{i+1} \right) \prod_{i=0}^{j-1}(1 - 36pn^{-1}i) \\
&\sim \sum_{0\leq j\leq m_t} \left(1 - 3p^2n + \frac{21}{2}p^4n^2\right)^j \frac{\lambda_1^j}{j!} \exp\left(\sum_{i=0}^{j-1}(-36pn^{-1}i + O(p^8n^4)) \right).
\end{aligned}$$

The dominant terms in this summation are for $j = \lambda_1 + O(\sqrt{\lambda_1})$, and for such terms the exponential factor is

$$\exp(-18pn^{-1}\lambda_1^2 + O(p^{11}n^7 + pn^{-1}\lambda_1^{3/2})) \sim \exp\left(-\frac{1}{2}p^7n^5\right).$$

Hence, the factor in (3.15) due to $t = 1$ is asymptotic to

$$\exp\left(\left(1 - 3p^2n + \frac{21}{2}p^4n^2\right)\lambda_1 - \frac{1}{2}p^7n^5 \right) \sim \exp\left(\frac{1}{6}p^3n^3 - \frac{1}{2}p^5n^4 + \frac{5}{4}p^7n^5 \right).$$

Similar but much simpler computations using (3.11)–(3.13) show that the factor due to $t = 2$ is asymptotic to $\exp\left(\frac{1}{4}p^5n^4 - \frac{5}{4}p^7n^5\right)$, and for $t = 3a$ and $t = 3b$ the factors are $\exp\left(\frac{1}{2}p^7n^5\right)$ and $\exp\left(\frac{1}{12}p^7n^5\right)$ respectively. Hence, (3.15) is asymptotic to

$$\exp\left(\frac{1}{6}p^3n^3 - \frac{1}{4}p^5n^4 + \frac{7}{12}p^7n^5\right),$$

which gives the theorem when $p^7n^5 \to \infty$. Smaller values of p are easily dealt with using a simplified version of this argument. For instance, if p^7n^5 is roughly constant, we can choose $m_{3a} = m_{3b} = \log n$, analogous to $pn = O(1)$ in the previous section. The remaining details for such p are left to the reader. ∎

Theorem 1 shows that the upper bound in (1.2) (where X is the same as X_1) is the true value of the first correction to the Poisson approximation $e^{-\lambda_1}$ of $\mathbf{P}(X = 0)$. One possible explanation that the Poisson approximation is not asymptotically correct is that the numbers of maximal clusters of the various small types could be behaving, at least to this level of accuracy, as independent Poisson variables, making X approximately compound Poisson. For $p = o(n^{-2/3})$, the expected numbers of maximal clusters of the various small types are asymptotically λ_{3a} and λ_{3b} for types $3a$ and $3b$, and hence $\lambda_2 - 3\lambda_{3a} - 2\lambda_{3b}$ for type 2, and $\lambda_1 - 2(\lambda_2 - 3\lambda_{3a} - 2\lambda_{3b}) - 3\lambda_{3a} - 3\lambda_{3b}$, that is $\lambda_1 - 2\lambda_2 + 3\lambda_{3a} + \lambda_{3b}$, for type 1. This suggests using $\exp(-\lambda_1 + \lambda_2 - \lambda_{3a})$ as an approximation for $\mathbf{P}(X = 0)$, which is asymptotically correct for $p = o(n^{-5/7})$, but no further.

4 $\mathcal{G}(n, m)$

By considering edges not in any triangles as clusters of their own right, it is possible to investigate the distribution of the number of edges in the class of graphs in $\mathcal{G}(n, p)$ with a given number of triangles. This gives access to probabilities in $\mathcal{G}(n, m)$. The following theorem is an example.

Theorem 2. *The probability that $G \in \mathcal{G}(n, m)$ has no triangles is asymptotic to*

$$\exp\left(-\frac{1}{6}d^3n^3\right)$$

for $d = \frac{2m}{n(n-1)} = o(n^{-2/3})$.

Proof. The definitions of the previous section need only slight modifications. For the proof we still work in $\mathcal{G}(n, p)$ with $p = o(n^{-2/3})$. We introduce a new small cluster: an edge, which we say is of type 0. Thus $\mathcal{S} = \{0, 1, 2, 3a, 3b\}$, and

$$\lambda_0 = \mathbf{E}X_0 = \frac{pn^2}{2}(1 + O(\tfrac{1}{n})).$$

No new maximal clusters are introduced other than a single edge not in any triangle, so the probability that $G \in \mathcal{G}(n, p)$ has a large cluster is still $o(1)$. Along with the new equations corresponding to (3.1) for the expanded set $\mathcal{S}$, we will use the corresponding equation for $t = t_0$, but only for functions $f = k\delta_0$, $k \geq 0$:

$$\frac{n(n-1)}{2}\mathbf{P}(\mathcal{C}_{k\delta_0}) = \sum_{\substack{x\in T_0\\ G\in\mathcal{C}_{k\delta_0}}} \mathbf{P}(G). \tag{4.1}$$

Consider this equation first. Again let u denote the type of the maximal cluster K of $G+x$ which contains x. If $u=0$, there are two cases: $x\in E(G)$ and $x\neq E(G)$. The contribution to (4.1) is thus

$$\frac{q}{p}(k+1)\mathbf{P}(\mathcal{C}_{(k+1)\delta_0}) + k\mathbf{P}(\mathcal{C}_{k\delta_0}).$$

Similarly, $u=1$ gives

$$3\frac{q}{p}\mathbf{P}(\mathcal{C}_{(k-2)\delta_0+\delta_1}),$$

$u=2$ gives

$$\frac{q}{p}\mathbf{P}(\mathcal{C}_{(k-4)\delta_0+\delta_2})$$

and $u=3a$ gives

$$\frac{q}{p}\mathbf{P}(\mathcal{C}_{(k-6)\delta_0+\delta_{3a}}).$$

Note that there are no terms corresponding $x\in K$ because $G\in\mathcal{C}_{k\delta_0}$ and so G contains no triangles. There is no contribution from $u=3b$ because there is no way to add a single edge to a graph with no triangles to form a cluster of type $3b$. Finally, suppose that K is large. Then there is a cluster J of $G+x$ of type $3a$. Deleting J from $G+x$ gives a graph G' which has exactly $k-6$ edges and no triangles; that is, $G'\in\mathcal{C}_{(k-6)\delta_0}$. This contributes $o(p^6n^5)\mathbf{P}(\mathcal{C}_{(k-6)\delta_0})$ to (4.1). Hence this equation becomes

$$\rho(k\delta_0,\delta_0) = \frac{\lambda_0}{k+1}\left(1-\frac{3\rho(k\delta_0,-2\delta_0+\delta_1)}{\lambda_0}+O(\epsilon_0)\right) \tag{4.2}$$

where

$$\epsilon_0 = p + \frac{k+o(p^6n^5)\rho(k\delta_0,-6\delta_0)}{n^2} + \frac{\rho(k\delta_0,-4\delta_0+\delta_2)+\rho(k\delta_0,-6\delta_0+\delta_{3a})}{pn^2},$$

and so

$$\gamma(k\delta_0,0) = 1-\frac{3\gamma((k-2)\delta_0,1)\lambda_1(k-1)k}{\gamma((k-2)\delta_0,0)\gamma((k-1)\delta_0,0)\lambda_0^3}+O(\epsilon_0). \tag{4.3}$$

It is now possible to reconsider $\rho(k\delta_0,\delta_1)$ and so on, along the lines of (3.2)–(3.5), and derive equations like (3.10)–(3.13). However, much less is required than this, since now we only need to estimate $\gamma(k\delta_0,0)$ accurately. For this, we only need the approximate values of γ's and ρ's in (4.3).

Take $G\in\mathcal{C}_{k\delta_0}$. Since $O(kn)$ elements of T_1 can contain an edge of G, a simple version of the argument leading to (3.2) establishes $\rho(k\delta_0,\delta_1)=\lambda_1-O(p^3kn)$. Similarly,

$$\rho(k\delta_0,\delta_2)=O(\lambda_2) \qquad \text{and} \qquad \rho(k\delta_0,\delta_{3a})=O(\lambda_{3a}).$$

Substituting into (4.3) and restricting to $0 \le k \le Cpn^2$ for any C gives firstly that $\gamma(k\delta_0, 0) \sim 1$ for all such k, and then from this that $\epsilon_0 = O(p)$ and

$$\gamma(k\delta_0, 0) = 1 - \frac{4k^2}{n^3} + O(p). \tag{4.4}$$

However, the probability that $G \in \mathcal{G}(n,p)$ has more than Cpn^2 edges for $C > 1/2$ is $o({C_1}^{pn^2}) = o(\exp(-p^3n^3)) = o(\mathbf{P}(X_1 = 0))$ (assuming $pn^2 \to \infty$; we have no trouble in any case for smaller p). Hence, since $\mathbf{P}(X_1 = 0) = \sum_{k \ge 0} \mathbf{P}(\mathcal{C}_{k\delta_0})$,

$$\frac{\mathbf{P}(X_1 = 0)}{\mathbf{P}(\mathcal{C}_{k\delta_0})} \sim \sum_{j=0}^{pn^2} \frac{\mathbf{P}(\mathcal{C}_{j\delta_0})}{\mathbf{P}(\mathcal{C}_{k\delta_0})}, \tag{4.5}$$

and (4.4) now implies asymptotic normality of X_0 restricted to the graphs in $\mathcal{G}(n,p)$ with no triangles, with mean and variance approximately

$$\mu = \frac{1}{2}pn^2(1 - p^2n), \qquad \sigma^2 = \mu$$

respectively. In fact, for $k = \mu + x\sigma$ (x bounded), we have

$$\gamma(k\delta_0, 0) = 1 - p^2n + O(p) \quad \text{and} \quad \rho(k\delta_0, \delta_0) = \frac{\mu}{(k+1)}(1 + O(p + n^{-1})).$$

Hence from (4.5), if $\omega(n) \to \infty$ arbitrarily slowly,

$$\begin{aligned}
\frac{\mathbf{P}(X_1 = 0)}{\mathbf{P}(\mathcal{C}_{k\delta_0})} &\sim \sum_{j=\lfloor \mu - \omega(n)\sigma \rfloor}^{\lfloor \mu + \omega(n)\sigma \rfloor} \rho(k\delta_0, (j-k)\delta_0) \\
&\sim \frac{k!}{\mu^k} \sum_{j=\lfloor \mu - \omega(n)\sigma \rfloor}^{\lfloor \mu + \omega(n)\sigma \rfloor} \frac{\mu^j}{j!} \\
&\sim \frac{k! e^\mu}{\mu^k} \\
&\sim \sigma\sqrt{2\pi}\, e^{x^2/2},
\end{aligned} \tag{4.6}$$

by the normal approximation to the Poisson probability.

It follows that for $k = \lfloor \mu \rfloor$,

$$\mathbf{P}(X_0 = k | X_1 = 0) \sim \frac{1}{n\sqrt{\pi p}}.$$

On the other hand, for $N = \binom{n}{2}$ and $\epsilon = O(p^2n)$,

$$\binom{N}{\lfloor Np(1+\epsilon) \rfloor} p^{Np(1+\epsilon)}(1-p)^{N(1-p(1+\epsilon))} \sim \frac{1}{\sqrt{2\pi pN}} \exp\left(N\left(-\frac{1}{2}p\epsilon^2 + \frac{1}{6}p\epsilon^3\right)\right).$$

Thus, for $k = \lfloor \mu \rfloor$, putting $\epsilon = -p^2n$ shows

$$\mathbf{P}(X_0 = k) \sim \frac{\exp(-\frac{1}{4}p^5n^4 - \frac{1}{12}p^7n^5)}{n\sqrt{\pi p}},$$

and Theorem 1 gives $\mathbf{P}(X_1 = 0)$. Dividing (4.6) by the first of these probabilities and multiplying by the second gives

$$\mathbf{P}(X_1 = 0 | X_0 = k) \sim \exp(-\frac{1}{6}p^3n^3 + \frac{1}{2}p^5n^4 - \frac{1}{2}p^7n^5). \qquad (4.7)$$

But the subspace of $\mathcal{G}(n,p)$ restricted to $X_0 = k$ is equivalent to $\mathcal{G}(n,m)$ with $m = k$. Thus (4.7) gives the probability that $G \in \mathcal{G}(n,m)$ has no triangles for $m = k$; that is, for edge density

$$d = 2k/n^2 + O(k/n^3) = p - p^3n + O(pn^{-1}).$$

The theorem follows. ∎

References

[1] P. Erdős and A. Rényi, On the evolution of random graphs, *Publ. Math. Inst. Hungar. Acad. Sci.* **5** (1960) 17–61.

[2] A. Frieze, On small subgraphs of random graphs, in *Random Graphs Vol. 2* (A. Frieze and T. Luczak, eds). Wiley (New York) 1992, pp 67–90.

[3] C.D. Godsil and B.D. McKay, Asymptotic enumeration of Latin rectangles, *J. Combinatorial Theory, Ser. B* **48** (1990) 19–44.

[4] S. Janson, T. Luczak and A. Ruciński, An exponential bound for the probability of nonexistence of a specified subgraph in a random graph, in *Random Graphs '87* (M. Karonski, J. Jaworski, A. Rucinski eds). Wiley (New York) 1990, pp 73-87.

[5] B. D. McKay, Subgraphs of random graphs with specified degrees, *Congressus Numerantium* **33** (1981) 213–223.

[6] B. D. McKay, Asymptotics for symmetric 0-1 matrices with prescribed row sums, *Ars Combinatoria* **19A** (1985) 15–25.

[7] B.D. McKay and N.C. Wormald, Uniform generation of random regular graphs of moderate degree, *J. Algorithms* **11** (1990) 52–67.

[8] B.D. McKay and N.C. Wormald, Asymptotic enumeration by degree sequence of graphs with degrees $o(n^{1/2})$, *Combinatorica* **11** (1991) 369–382.

[9] H.J. Prömel and A. Steger, On the asymptotic structure of sparse triangle free graphs, *J. Graph Theory*, to appear.

[10] H.J. Prömel and A. Steger, Counting H-free graphs, *Discrete Math.*, to appear.

[11] C. Stein, Asymptotic evaluation of the number of Latin rectangles, *J. Combinatorial Theory, Ser. A.* **25** (1978) 38–49.

[12] C. Stein, *Approximate Computation of Expectations. Lecture Notes* **7**. IMS (Hayward, CA) 1988.

[13] N.C. Wormald, The perturbation method for small subgraphs of random graphs, manuscript.