Distribution of subgraphs of random regular graphs

Zhicheng Gao*
Faculty of Business Administration
University of Macau
Macau
China
zcgao@umac.mo

N. C. Wormald†
Department of Combinatorics and Optimization
University of Waterloo
Waterloo ON
Canada N2L 3G1
nwormald@uwaterloo.ca

1 Introduction

The asymptotic distribution of small subgraphs of a random graph has been basically worked out (see Ruciński [5] for example). But for random regular graphs, the main techniques for proving, for instance, asymptotic normality, do not seem to be usable. One very recent result in this direction is to be found in [3], where switchings were applied to cycle counts. The aim of the present note is to show that another very recent method of proving asymptotic normality, given by the authors in [1], can easily be applied to this problem. In particular, it requires considerably less work than using switchings. The application is, however, not direct, in the sense that the result obtained is very weak if the random variable counting copies of a subgraph is examined directly. We obtain a much stronger result by considering isolated copies of a subgraph.

To be specific, we investigate the probability space $\mathcal{G}_{n,d}$ of uniformly distributed random d-regular graphs on n vertices (which we assume to be $\{1, 2, \dots, n\}$). Asymptotics are for $n \to \infty$, and here d is not fixed but may vary with n (though for all our results there is an upper bound on the growth of d, at least implicitly). As usual, we impose the restriction that for the asymptotics, the odd values of n are omitted in the case of odd d.

*Research supported by NSERC and University of Macau
†Research supported by the Canada Research Chairs program and NSERC.

We use $\mu(G)$ and $\nu(G)$ for the numbers of edges and vertices of a graph G respectively. A graph G is *strictly balanced* if

$$\frac{\mu(G)}{\nu(G)} > \frac{\mu(G_1)}{\nu(G_1)} \tag{1.1}$$

for all nontrivial proper subgraphs G_1 of G. A standard example: every connected regular graph is strictly balanced.

Throughout this paper, $[x]_m$ denoting the falling factorial: $x(x-1)\cdots(x-m+1)$. We will use the following from [1] to deduce asymptotic normality.

Theorem 1 *Let $s_n > -\mu_n^{-1}$ and*

$$\sigma_n = \sqrt{\mu_n + \mu_n^2 s_n}, \tag{1.2}$$

where $0 < \mu_n \to \infty$. Suppose that

$$\mu_n = o(\sigma_n^3), \tag{1.3}$$

and a sequence $\{X_n\}$ of nonnegative random variables satisfies

$$\mathbf{E}[X_n]_k \sim \mu_n^k \exp\left(\frac{k^2 s_n}{2}\right) \tag{1.4}$$

uniformly for all integers k in the range $c\mu_n/\sigma_n \leq k \leq c'\mu_n/\sigma_n$ for some constants $c' > c > 0$. Then $(X_n - \mu_n)/\sigma_n$ tends in distribution to the standard normal as $n \to \infty$.

We also use McKay [4, Theorem 2.10], in the form of the following simpler special case stated in [3]. Here, G denotes a random element of $\mathcal{G}_{n,d}$, E denotes the edge set, and K_n is the complete graph on n vertices (the same vertex set as G).

Theorem 2 *For any d and n such that $|\mathcal{G}_{n,d}| \neq 0$, let $J \subseteq E(K_n)$. Then, with j_i the number of edges in J incident with vertex i,*

(a) if $|J| + 2d^2 \leq nd/2$ then

$$\mathbf{P}(J \subseteq E(G)) \leq \frac{\prod_{k=1}^n [d]_{j_k}}{2^{|J|}[nd/2 - 2d^2]_{|J|}} \,;$$

(b) if $2|J| + 4d(d+1) \leq nd/2$ then

$$\mathbf{P}(J \subseteq E(G)) \geq \frac{\prod_{k=1}^n [d]_{j_k}}{2^{|J|}[nd/2 - 1]_{|J|}} \left(\frac{n-2d-2}{n+2d}\right)^{|J|}.$$

Corollary 1 *Provided $d|J| = o(n)$, the hypotheses of Theorem 2 imply that*

$$\mathbf{P}(J \subseteq E(G)) = \frac{\prod_{k=1}^n [d]_{j_k}}{(nd)^{|J|}} \left(1 + O\left((d|J|/n)^2\right)\right).$$

2 Distribution of number of copies of a graph

Throughout this section, G denotes a random graph in $\mathcal{G}_{n,d}$. Let H be a fixed strictly balanced graph with maximum vertex degree d. Let p and q be the number of vertices and edges of H. A *copy* of H in G is a subgraph of G which is isomorphic to H.

The use of Theorem 1 calls for computing high moments of a random variable. It turns out that the random variable counting copies of H has badly behaved moments and consequently does not produce a very useful result. Instead we consider a related random variable whose behaviour is more easily analysed. We say that a subgraph of a graph G isomorphic to H is an *isolated copy* of H if it shares no edges with any other subgraph of G isomorphic to H. Let X_H be the random variable which is the number of isolated copies of H in a random d-regular graph. Let a denote the order of the automorphism group of H. Set

$$\mu = \mathbf{P}(H_1 \subseteq G)[n]_p/a \tag{2.1}$$

where H_1 is a fixed copy of H on the vertex set $\{1, 2, \ldots, V(H)\}$. The probability that *any* given copy of H in K_n occurs in G is equal to $\mathbf{P}(H_1 \subseteq G)$, and there are $[n]_p/a$ such copies. Hence, μ is the expected number of copies of H in G, isolated or not. By Corollary 1, for $d = o(n)$ (noting that q is fixed),

$$\mu = \Theta(n^{p-q}d^q) \tag{2.2}$$

where $f = \Theta(g)$ if $f = O(g)$ and $g = O(f)$. Also, let

$$r = r(n, d, H) = \mathbf{P}(H_1 \text{ is not isolated} \mid H_1 \subseteq G). \tag{2.3}$$

Fix a proper subgraph F of H_1 containing at least one edge, and consider the probability that G contains not only H_1 but also the edges of a subgraph $H_2 \cong H$ with $H_1 \cap H_2 = F$. Again using Corollary 1, this (unconditional) probability is $\Theta(n^{p-p(F)}\,(d/n)^{2q-q(F)})$ for $d = o(n)$. Since there is a bounded number of such subgraphs F, we have for $d = o(n)$,

$$r = \Theta(r_H(n, d)) \quad \text{where} \quad r_H(n, d) = n^{p-p(F)}\left(\frac{d}{n}\right)^{q-q(F)}, \tag{2.4}$$

$p(F) = |V(F)|$, $q(F) = |E(F)|$, and F is a subgraph of H which maximises $n^{q(F)-p(F)}d^{-q(F)}$ subject to $1 \le q(F) < q$. (See [2, Section 3.2] for a related discussion in the setting of random graphs without the regularity condition.)

Theorem 3 *Define μ and r as in (2.1) and (2.3). Suppose that $\mu \to \infty$, $\mu = o(n)$, $\mu = o(n^2/d^2)$ and $r = O(1/\sqrt{\mu})$. Then $(X_H - \mu e^{-r})/\sigma$ tends in distribution to the standard normal as $n \to \infty$, where $\sigma^2 = \mu e^{-r}$.*

Note 1 If $r = o(1/\sqrt{\mu})$ then the mean and variance of the asymptotic distribution can both be taken as μ. Moreover, the proof of the theorem then simplifies considerably. However, by including the case $r \approx 1/\sqrt{\mu}$ we highlight why the method does not easily extend.

Note 2 The distribution result in [3], which is only for cycles, does not extend to the full range of d covered by Theorem 3. (It does however apply to non-fixed subgraphs, a modification which could also be done easily using the techniques of the present paper.) One could presumably extend the methods used in [3] to obtain distribution results for all strictly balanced subgraphs, but this is not as economical as the method in the present paper, and we believe that the range of d obtained would not be any greater than that in Theorem 3.

Note 3 Distribution results for subgraph counts in the other common models of random graphs apply for wider ranges of density of the parent graph than expressed in Theorem 3. Given the much greater accessability of those models due to edge independence, this is not very surprising.

Proof of Theorem 3 We compute the k'th factorial moment $\mathbf{E}[X_H]_k$, for $k = O(\sqrt{\mu})$, $k \to \infty$. Note that

$$\mathbf{E}[X_H]_k = \sum_{J_1,\dots,J_k \subseteq E(K_n)} \mathbf{P}\left(\bigwedge_i A_{J_i}\right) \tag{2.5}$$

where A_J denotes the event that $J \subseteq E(G)$ and forms an isolated copy of H. (For $k = 1$, this differs from (2.1) because the copies here are isolated.) To find the number of nonzero summands contributing in (2.5), for which a prerequisite is that the J_i are pairwise disjoint, consider placing k ordered copies of H on the vertices of the complete graph. Since $k = O(\sqrt{\mu}) = o(\sqrt{n})$, the number of ways of doing this, where each copy is placed independently (ignoring possible overlaps) is asymptotic to

$$\frac{n^{kp}}{a^k}, \tag{2.6}$$

and we get the same expression if we insist that the copies have disjoint vertex sets (so after j copies have been placed there are $n - pj$ vertices to choose from). Thus, by sandwiching, this is also asymptotically the number of ways of choosing edge-disjoint copies, as required for isolated copies, and almost all these placements are pairwise vertex-disjoint. Clearly

$$\mathbf{P}\left(\bigwedge_i A_{J_i}\right) \le \mathbf{P}(J_1 \cup \cdots \cup J_k \subseteq E(G)) \le (\mathbf{P}(J_1 \subseteq G))^k (1 + o(1))$$

using Corollary 1 and noting that the assumption $\mu = o(n^2/d^2))$ implies the required bound on $d|J|$. So we have from (2.5) that

$$\begin{aligned}
\mathbf{E}[X_H]_k &= o\left(\frac{n^{kp}}{a^k}\right)(\mathbf{P}(J_1 \subseteq G))^k + \sum_{\substack{J_1,\dots,J_k \subseteq E(K_n) \\ J_i \text{ vertex-disjoint}}} \mathbf{P}\left(\bigwedge_i A_{J_i}\right) \\
&= o\left(\frac{n^{kp}}{a^k}\right)(\mathbf{P}(J_1 \subseteq G))^k + \frac{n^{kp}}{a^k}\mathbf{P}\left(\bigwedge_i A_{J_i}\right) \qquad (2.7)
\end{aligned}$$

for any particular choice of $J_1, \dots, J_k \subseteq E(K_n)$ which induce vertex-disjoint copies of H in K_n. Letting B denote the conditional probability that these sets induce *isolated* copies in G, given that they are subsets of the edge set of G, we have

$$\begin{aligned}\mathbf{P}\left(\bigwedge_i A_{J_i}\right) &= B\,\mathbf{P}(J_1,\dots,J_k \subseteq E(G)) \\ &\sim B(\mathbf{P}(J_1 \subseteq G))^k \end{aligned} \tag{2.8}$$

using Corollary 1 again.

For one of the copies to be nonisolated, it must share an edge with some other copy of H, and we may use the same machinery to compute the factorial moments of the number Y of i for which J_i induces a nonisolated copy in G, conditional upon $J_1, \dots, J_k \subseteq E(G)$. We have

$$kr = O(1) \tag{2.9}$$

since $r\sqrt{\mu} = O(1)$, and hence also, since $\mu \to \infty$,

$$r = o(1). \tag{2.10}$$

Thus

$$r = O(r^2) + \sum_{H_2} \frac{\mathbf{P}(H_1 \cup H_2 \subseteq G)}{\mathbf{P}(H_1 \subseteq G)}$$

where the sum is over copies H_2 of H in K_n which share at least one edge with H_1, and the r^2 term bounds the overcounting in inclusion-exclusion. With this observation it is easy to compute the jth factorial moment $\mathbf{E}[Y]_j$ of Y. There are $[k]_j$ ways to choose a j-subset of $J_1, \dots, J_k$ each of which induce a copy of H to be nonisolated, and the probability that all the required edges are present in G is, again using Corollary 1, asymptotic to

$$r^j \mathbf{P}(J_1,\dots,J_k \subseteq E(G)) \sim r^j (\mathbf{P}(J_1 \subseteq G))^k.$$

The probability that all these edges are in G, conditional upon the event that $J_1, \dots, J_k \subseteq E(G)$, is thus r^j, and so $\mathbf{E}[Y]_j \sim (kr)^j$. Thus by (2.9) and the method of moments (the usual one, that is), $\mathbf{P}(Y = 0) \sim e^{-kr}$. Thus $B \sim e^{-kr}$, and we have from (2.7) and (2.8) that

$$\mathbf{E}[X_H]_k \sim \frac{n^{kp}}{a^k} e^{-kr} (\mathbf{P}(J_1 \subseteq G))^k.$$

By (2.1) and the fact that $k = o(\sqrt{n})$, this implies

$$\mathbf{E}[X_H]_k \sim (\mu e^{-r})^k$$

and thus by Theorem 1 with $s_n = 0$ the distribution of X_H is asymptotically normal with mean and variance μe^{-r} by (2.10). ∎

Example: Cycles

Consider the graph $H = C_t$, the cycle of length $t \geq 3$. Here by (2.2) $\mu = \Theta(d^t)$, so we require $d \to \infty$. Also it is easy to check that $r_H(n,d) = \Theta(\mu/(nd))$, the maximum in (2.4) occurring for $F \cong K_2$. Thus the range of μ is bounded at the maximum end by $\mu^{3/2} = O(nd)$, $\mu = o(n)$ and $\mu = o(n^2/d^2)$. By considering the implied upper bounds on d, we see that the first is strictest, and thus the number of isolated copies of H is asymptotically normally distributed provided $d \to \infty$ and $d = O(n^{2/(3t-2)})$. This can be compared with the result in [3], for which the bound is $d = o(n^{1/(2t-1)})$.

Example: Complete graphs

Consider the graph $H = K_t$, where $t \geq 3$. Here by (2.2)

$$\mu = \Theta(d^{t(t-1)/2} n^{-t(t-3)/2}). \tag{2.11}$$

Also $r_H(n,d) = \max_{2\leq s<t} d^{t(t-1)/2-s(s-1)/2} n^{-t(t-3)/2+s(s-3)/2}$, and considering $d = n^\alpha$, the maximum occurs (as with cycles) at $s = 2$, and so $r = \Theta(\mu/(nd))$. Thus (as with cycles), the range of μ is bounded at the maximum end by $\mu^{3/2} = O(nd)$, $\mu = o(n)$ and $\mu = o(n^2/d^2)$. By considering the implied upper bounds on d, it is straightforward to verify that the first gives the strictest bound for $t = 3$, and the last does for $t \geq 4$. These imply the upper bounds $d = O(n^{2/7})$ in the case $t = 3$, and $d = o(n^{(t(t-3)/2+2)/(t(t-1)/2+2)})$, i.e. $d = o(n^{1-2t/(t^2-t+4)})$, for $t \geq 4$. Therefore, the number of isolated copies of H is asymptotically normally distributed provided this upper bound on d holds and the expression in (2.11) tends to ∞.

Finally, we may conclude something about the distribution of the total number of copies of H, isolated or not. Denote this number by $\hat{X}_H$.

Corollary 2 *Suppose that $\mu \to \infty$, $\mu = o(n)$, $\mu = o(n^2/d^2)$ and $r = o(1/\sqrt{\mu})$. Then $(\hat{X}_H - \mu)/\sqrt{\mu}$ tends in distribution to the standard normal as $n \to \infty$.*

Proof: The expected number of nonisolated copies is $O(\mu r)$. So we may conclude that the total number of copies of H is asymptotically normal provided $\mu r = o(\sqrt{\mu})$, i.e. $r\sqrt{\mu} = o(1)$. This is an assumption of the corollary which is stronger than the corresponding one in the theorem. ∎

3 Concluding remarks

For the distribution results obtained in Theorem 3, the mean and variance are asymptotically equal. This means that it could equivalently be stated as giving asymptotically Poisson distribution. It would be interesting to know the range of the degree d for which the subgraph count remains asymptotically Poisson. Theorem 1 can potentially be used to deduce asymptotic normality outside the Poisson range (as for instance the previous applications in [1]); one challenge is to find a way to apply it for such d in the present context. Another challenge is to find a way to apply any of the other methods of deducing asymptotic normality to significantly higher values of d than we do here.

One possibility is to use switchings rather than standard inclusion-exclusion to extend the range of d for which the nonisolated copies may be treated in the proof of Theorem 3. However, the extra effort may not pay very big dividends.

References

[1] Z. Gao and N.C. Wormald, Asymptotic normality determined by high moments, and submap counts of random maps, *Probability Theory and Related Fields* (to appear).

[2] S. Janson, T. Łuczak and A. Ruciński, **Random graphs**, Wiley, New York, 2000.

[3] B.D. McKay, N.C. Wormald and B. Wysocka, Short cycles in random regular graphs, *Electronic Journal of Combinatorics* (submitted).

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

[5] A. Ruciński, When are small subgraphs of a random graph normally distributed? *Probability Theory and Related Fields* 78 (1988), 1–10.