COMBINATORICA 1(4) (1981) 369—376

MATCHING BEHAVIOUR IS ASYMPTOTICALLY NORMAL

C. D. GODSIL

Institut für Mathematik
Montanuniversität Leoben
A-8700 Leoben, Austria

and

Simon Fraser University
Burnaby, B. C. V5A 1S6, Canada

Received 2 July 1981

A k-matching in a graph G is a set of k edges, no two of which have a vertex in common. The number of these in G is written $p(G, k)$. Using an idea due to L. H. Harper, we establish a condition under which these numbers are approximately normally distributed. We show that our condition is satisfied if $n=|V(G)|$ is large compared to the maximum degree Δ of a vertex in G (i.e. $\Delta=o(n)$) or G is a large complete graph. One corollary of these results is that the number of points fixed by a randomly chosen involution in the symmetric group S is asymptotically normally distributed.

1. Introduction

A *matching* in a graph G is a set of edges, no two of which have a vertex in common. A matching consisting of exactly k edges is a *k-matching*. We denote the number of k-matchings in G by $p(G, k)$, with the convention that $p(G, 0)=1$.

We will show here that for many graphs the number of edges in a randomly chosen matching (assuming all matchings are equally likely to be chosen) is asymptotically a normal random variable. Another way of expressing this is that the graph of the numbers $p(G, k)$ versus k can be well fitted by a function of the form $A\exp\left((x-B)^2/2\right)$.

The accuracy of this approximation increases with the variance of the number of edges in a randomly chosen matching. We note one corollary of this result.

1.1. Theorem. *Let $\{G_n\}_{n=1}^{\infty}$ be a sequence of graphs, each regular of degree d, such that $|V(G_n)|$ increases with n. Then when n is large enough, the numbers $p(G_n, k)$ are normally distributed.*

We will prove this following Lemma 3.2. Note that if G_n is taken to be the disjoint union of n copies of K_2 then $p(G_n, k)=\binom{n}{k}$ and so Theorem 1.1 then yields the case $p=1/2$ of the DeMoivre—Laplace central limit theorem.

The constraint on the degree of the graphs G_n in Theorem 1.1 is sufficient, but not all necessary. In fact this theorem also holds when $G_n=K_n$, the complete graph on n vertices.

AMS subject classification (1980): 05 A 15, 05 C 50; 60 F 05

It is worth noting that many number sequences of combinatorial interest arise as coefficients $p(G, k)$, for a suitable graph G. For example the Stirling numbers of second kind can be so represented (see Section 4). Also the Hermite polynomial of degree n is the matchings polynomial of K_n. The Laguerre and Chebyshev polynomials also arise as matchings polynomials. For more details, see [10], [5].

2. A normality criterion

Our results will follow from the properties of the matchings polynomial, which we now introduce.

2.1. Definitions. Let G be a graph. Denote the largest number of edges in a matching of G by $\nu(G)$. The *matchings polynomial* of G is

$$\mu(G, x) = \sum_{k=0}^{\nu(G)} (-1)^k p(G, k) x^{n-2k}.$$

We set

$$p(G) = \sum_{k=0}^{\nu(G)} p(G, k)$$

and define the matchings *generating function* of G to be

$$\beta(G, x) = \sum_{k=0}^{\nu(G)} p(G, k) x^k.$$

Note that the coefficient of $\beta(G, x)/\beta(G, 1)$ can be interpreted as the probability that a random chosen matching in G has k edges. (Thus it is the probability generating function for a random variable taking values $0, 1, \ldots, \nu(G)$.)

The matchings polynomial has many interesting properties (see [5], [6]). One of the more remarkable is

2.2. Theorem. (Heilmann and Lieb [10: Lemma 4.1]). *Let G be a graph. The zeros of $\mu(G, x)$ are real.*

Actually the paper of Heilmann and Lieb gives three different proofs of the above result. A fourth is given in [7]. We note two consequences of it.

2.3. Corollary. *Let G be a graph and let $m = \nu(G)$. Then*

(*a*) *the zeros of $\beta(G, x)$ are real and negative and*

(*b*) *the sequence $p(G, k)\Big/\binom{m}{k}$ is log-concave i.e. the numbers*

$$\frac{p(G, k)\Big/\binom{m}{k}}{p(G, k-1)\Big/\binom{m}{k-1}}$$

for $k = 1, \ldots, m$ form a non-increasing sequence.

Proof. (a) This follows at once from the theorem together with the observation that if $\mu(G,\theta)=0$ then $-1/\theta^2$ is a zero of $\beta(G,x)$.

(b) This result holds for the coefficients of any polynomial with negative real roots, by virtue of Newton's inequalities (see Theorem 144 in [8]). ∎

The fact that (b) holds for the numbers $p(G,k)$ was first noted by Heilmann and Lieb (Theorem 7.1 in [10]). Note also that it is trivial to show using (b) that the numbers $p(G,k)$ themselves form a log-concave sequence.

The next set of definitions will enable us to state and prove our main results.

2.4. Definitions. Let G be a graph. We use $S=S(G)$ to denote the random variable whose value is the number of edges in a randomly chosen matching in G. The mean and variance of S will be denoted $\hat{\beta}=\hat{\beta}(G)$ and $\sigma=\sigma(G)$ respectively.

2.5. Theorem. *Let G be a graph. Then there exist positive constants K and L such that if $\sigma>K$ then*

$$|\sigma p(G,k)/p(G)-e^{-x^2/2}(2\pi)^{-1/2}|<L\sigma^{-1/2} \tag{1}$$

where $x=(k-\hat{\beta})/\sigma$.

Proof. We will show that S can be represented as a sum of independent random variables. Our conclusion can then be derived from a version of the central limit theorem. This approach is essentially due to L. H. Harper [9]; we will comment further on this at the end of the proof.

From our definitions we see that the probability generating function of S is just $\beta(G,x)/\beta(G,1)$. By Corollary 2.3 (a) we see that this polynomial can be expressed in the form

$$\prod_{i=1}^{m}\frac{(x+\lambda_i)}{1+\lambda_i},$$

where the λ_i are just the zeros of $\beta(G,x)$ and $m=v(G)$.

We now introduce random variables X_i, $i=1,\dots,m$ such that

$$\operatorname{Prob}(X_i=0)=1/(1+\lambda_i)$$

$$\operatorname{Prob}(X_i=1)=\lambda_i/(1+\lambda_i).$$

Then X_i has pgf (probability generating function) $(x+\lambda_i)/(1+\lambda_i)$ and so $X_1+\dots+X_m$ has pgf $\beta(G,x)/\beta(G,1)$, since the generating function for a sum of independent random variables equals the product of the generating functions of the terms in the sum. It follows that we may identify this sum with S.

We claim now that, if $x=(k-\hat{\beta})/\sigma$ and $K=33/4$

$$\left|\sum_{l=0}^{k}\frac{p(G,l)}{p(G)}-\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-u^2/2}\,du\right|<\frac{K}{\sigma}. \tag{2}$$

Given this, our theorem follows immediately from Theorem (B) in [2]. (This theorem asserts that if a log-concave sequence $p(G,k)$ satisfies (2) for some $K>7$ and $\sigma>9.46\times10^8$ then (1) holds with $L=124.5$.)

It remains to establish (1). We will derive this from Theorem 1 on p. 521 of [4]. This theorem is a form of the so-called Berry—Esséen theorem. If we define

x_i to be the mean of X_i, this result implies that, if $\mathrm{E}(|X_i-x_i|^3)<\mathrm{E}((X_i-x_i)^2)$ then (2) holds with $K=33/4$.

However X_i only takes values 0 and 1, both with positive probability. Therefore $0<x_i<1$ and so $0<|X_i-x_i|<1$. Hence $|X_i-x_i|^3<|X_i-x_i|^2$, no matter what value X_i takes, and so our theorem is proved. ∎

To make any use of this theorem we must establish the existence of graphs G such that σ is arbitrarily large. We will do this in the next section.

We have already remarked that the method used above is basically due to L. H. Harper [9]. He used it to show that the Stirling numbers of the second kind are asymptotically normally distributed.

Harper's arguments alone would, in our situation, yield the conclusion that if $\{G_n\}_{n=1}^{\infty}$ was a sequence of graphs such that $\sigma(G_n)$ went to infinity with n then the sum in (2) would converge to the integral in (2).

Subsequently E. A. Bender extended Harper's method in [1]. In our case his extension would enable us to conclude additionally that

$$\frac{\sigma(G_n)\,p(G_n,k)}{p(G_n)} \to (2\pi)^{-1/2}e^{-x^2/2}$$

as $n\to\infty$. The explicit bound we quote was extracted from Bender's arguments by E. Rodney Canfield in [2].

3. Bounds on $\sigma(G)$

We require some information on the distribution of the zeros of $\beta(G,x)$.

3.1. Lemma. (Heilman and Lieb [10, Thm. 4.3]). *Let Δ be the maximum degree of a vertex in the graph G and let θ be a zero of $\mu(G,x)$. Then if $\Delta>1$, $\theta<2(\Delta-1)^{1/2}$.* ∎

3.2. Lemma. *Let G be a graph with maximum degree $\Delta>0$ and M edges. Then*

$$\sigma(G) < M^{1/2}/4\Delta.$$

Proof. Let S, X_i, λ_i be as in the previous section. Let $m=v(G)$. We have

$$\mathrm{Var}(X_i) = \lambda_i/(1+\lambda_i)^2.$$

Since $S=X_1+\ldots+X_m$ and the X_i are independent, this implies

$$\mathrm{Var}(S) = \sum_{i=1}^{m} \lambda_i/(1+\lambda_i)^2. \tag{3}$$

In the proof of Corollary 2.3 (a) we saw that if λ_i is a zero of $\beta(G,x)$ then $\lambda_i=1/\theta_i^2$, where θ_i is a zero of $\mu(G,x)$. Therefore if $\Delta>1$,

$$\lambda_i/(1+\lambda_i)^2 = (1/\lambda_i)/(1+(1/\lambda_i))^2 > (1/\lambda_i)/(4\Delta-3)^2,$$

since by Lemma 3.1, $\lambda_i^{-1}=\theta_i^2<4\Delta-4$. Accordingly for $\Delta>1$ our lower bound will follow if we show that $\sum\lambda_i^{-1}=M$.

Since $\beta(G, x)$ has constant term equal to 1, the coefficient of x equals $\sum \lambda_i^{-1}$. But from the definition of $\beta(G, x)$ this coefficient is just $p(G, 1)=M$. To complete the proof we note that if $\Delta \leqq 1$ then G is a Δ disjoint union of isolated vertices and edges. In this case it is not difficult to show that $\beta(G, x)=(1+x)^m$. Hence $\lambda_i=1$ for all i and so, by equation (3) above, $\sigma(G)=m/2=M/2$. This is clearly greater than $M^{1/2}/4\Delta$. ∎

If G has n vertices and is regular of degree d then $M=nd/2$. Hence Lemma 3.2 gives $\sigma(G)>(n/16d)^{1/2}$. Taking this with Theorem 2.5 yields the result stated as Theorem 1.1 in the introduction.

We remark that (3) can be rewritten as

$$\mathrm{Var}(S) = \sum_{i=1}^{m}\left(\frac{1}{1+\lambda_i}-\frac{1}{(1+\lambda_i)^2}\right).$$

Since we also have

$$\mathrm{E}(S) = \sum_{i=1}^{m}\frac{1}{1+\lambda_i},$$

it follows that $\mathrm{Var}(S)<\mathrm{E}(S)$, no matter what graph G we take. This is essentially a corollary of the fact that S can be expressed as a sum of independent $0-1$ random variables. Nevertheless it is of interest since it shows that $\mathrm{E}(S)$ and $\mathrm{V}(S)$ cannot be prescribed arbitrarily.

We will now show that $\sigma(K_n)$ tends to infinity with n. This implies that the numbers $p(K_n, k)$, $(k=0, 1, \ldots)$ are asymptotically normally distributed. Of course $p(K_n, k)$ is the number of involutions in the symmetric group S_n with $n-2k$ fixed points.

We will use the polynomial

$$\varphi_n(x) = \sum_{k=0}^{[n/2]} p(K_n, k)x^{n-2k}.$$

Let T_n be the random variable $n-2S(K_n)$. Then $\varphi_n(x)/\varphi_n(1)$ is the probability generating function for T_n.

3.3. Lemma. *We have*

(a) $\mathrm{E}(T_n) = n\varphi_{n-1}(1)/\varphi_n(1)$,

(b) $\mathrm{E}(T_n^2) = (n^2-n)\varphi_{n-2}(1)/\varphi_n(1)$.

Proof. We know that

$$\mathrm{E}(T_n) = \frac{d}{dx}(\varphi_n(x)/\varphi_n(1))|_{x=1} \tag{4a}$$

$$\mathrm{E}(T_n^2-T_n) = \frac{d^2}{dx^2}(\varphi_n(x)/\varphi_n(1))|_{x=1}. \tag{4b}$$

(This is a basic property of probability generating functions and can be found, for example, in Chapter XI of [3].) It is known and easily proved that

$$p(K_n, k) = \frac{n!}{(n-2k)!\,k!\,2^k}.$$

Hence it follows that $\varphi_n'(x)=n\varphi_{n-1}(x)$, $\varphi_n''(x)=n(n-1)\varphi_{n-2}(x)$. Substituting these identities in (4a) and (4b) yields the lemma. ▮

3.4. Lemma. *Let $\hat{\beta}_n$ and σ_n denote respectively the mean and standard deviation of the number of edges in a randomly chosen matching from K_n. Then we have*

(a) $\hat{\beta}_n = (n-n^{1/2}+1/2)/2+O(n^{-1/2})$,
(b) $\sigma_n^2 = (2n^{1/2}-3)/8+O(n^{-1/2})$.

Proof. Our result will follow from 2.12 in [12]. In our notation this asserts that

$$\varphi_n(1)/\varphi_{n-1}(1) = n^{1/2}+\frac{1}{2}-\frac{1}{8}n^{-1/2}+O(n^{-1}). \tag{5}$$

We need an expression for $n\varphi_{n-1}(1)/\varphi_n$. To get this we note that $\varphi_{n+1}(1)=\varphi_n(1)+ +n\varphi_{n-1}(1)$, whence

$$n\varphi_{n-1}(1)/\varphi_n(1) = -1+\varphi_{n+1}(1)/\varphi_n(1). \tag{6}$$

(The identity we have just used follows on observing that $\varphi_{n+1}(1)$ is the number of involutions in S_{n+1} and that $\varphi_n(1)$ can be viewed as the number of involutions in S_{n+1} fixing some specified point.)

From (5), (6) and Lemma 3.3a we obtain

$$\mathrm{E}(T_n) = n^{1/2}-\frac{1}{2}+\frac{3}{8}n^{-1/2}+O(n^{-1}). \tag{7}$$

Since $\varphi_{n-2}(1)\varphi_n(1)=(\varphi_{n-2}(1)/\varphi_{n-1}(1))(\varphi_{n-1}(1)/\varphi_n(1))$ we also find that

$$\begin{aligned} \mathrm{E}(T_n^2-T_n) &= n-n^{1/2}+O(n^{-1/2}) \\ \mathrm{E}(T_n)^2 &= n-n^{1/2}+1+O(n^{-1/2}). \end{aligned} \tag{8}$$

As $T_n=n-2S(K_n)$ it follows that $\hat{\beta}_n=(n-\mathrm{E}(T_n))/2$ and $\sigma_n^2=\mathrm{Var}\,(T_n)/4$. Thus the first claim of the lemma follows from (7). Noting that $\mathrm{Var}\,(T_n)=\mathrm{E}(T_n^2)-\mathrm{E}(T_n)^2$, we deduce the second claim from (7) and (8). ▮

3.5. Corollary. *The numbers $p(K_n, k)$ are asymptotically normally distributed, with mean and variance as given in Lemma* 3.4.

Proof. This is just a combination of Lemma 3.4 and Theorem 2.5. ▮

Corollary 3.5 is of course equivalent to the assertion that the number of points fixed by a random involution in S_n is asymptotically normally distributed. In this case the mean and variance are $n^{1/2}-(1/2)$ and $n^{1/2}-(3/2)$ respectively.

4. Weaker sufficient conditions for normality?

The difficulty in applying 2.5 is that we need an estimate for $\sigma(S)$. Lemma 3.2 provides an estimate which covers a number of interesting cases. However it is clear from Corollary 3.5 that it fails to cover some possibilities of equal interest.

We hope to show in a subsequent paper that Corollary 3.5 holds when K_n is replaced by a graph G_n in which the minimum degree is $n-o(n)$, $n=|V(G)|$. However even this will not be enough.

For consider the graph B with vertex set

$$V(B) = \{b_1, \ldots, b_n\} \cup \{w_1, \ldots, w_n\},$$

where vertex b_i is adjacent to w_j iff $i>j$. Then by Theorem 4 on page 213 of [13] or by Exercise 4.31 of [11], $p(B, k)=S(n, n-k)$. That is, the numbers $p(B, k)$ coincide with the Stirling numbers of the second kind, written down in reverse order. We noted following Theorem 2.5 that L. H. Harper has proved these numbers are asymptotically normally distributed.

By now the reader may wonder if there does exist a non-trivial sequence G_n of graphs for which the matchings numbers $p(G_n, k)$ are not normally distributed. However if we take G_n to be n copies of $K_{1,n}$ then $p(G_n, k)$ is $n^k \binom{n}{k}$. A trivial computation shows that, for fixed n^k, $p(G_n, k)$ increases with n. Therefore we cannot possibly achieve normality in the limit as n tends to infinity.

There are two simple conditions which are necessary for $\sigma(G_n)$ to increase with n. The first is that $\hat{\beta}(G_n)$ must increase. This is a consequence of our remark following Lemma 3.2, that $\sigma^2(G) < \hat{\beta}(G)$ for all graphs G. The second condition is that $\nu(G_n)$, the size of the largest matching in G_n, be unbounded. If this does not hold then $S(G_n)$ is the sum of a bounded number of discrete random variables. Hence it cannot possibly converge to any continuous distribution.

Interestingly enough the second of these conditions implies the first. This was pointed out privately to me by L. Babai.

4.1. Lemma. (L. Babai, unpublished). *For any graph G, we have* $E(S(G)) \geqq \nu(G)/3$.

Proof. We call a matching *non-maximal* if it is a proper subset of another matching. We note that

$$p'(G) = \sum_{k=0}^{\nu(G)} kp(G, k)$$

counts the number of ordered pairs (A, b), where A is a k-matching and b is an edge in A. Hence it also counts the number of ordered pairs consisting of a non-maximal $(k-1)$-matching together with an edge disjoint from it.

Let $m=\nu(G)$. If $k<m/2$ then a k-matching must be non-maximal, in fact given any m-matching there must be at least $m-2k$ edges in this disjoint from the k-matching. Hence

$$p'(G) \geqq \sum_{k=0}^{\lfloor m/2 \rfloor} p(G, k)(m-2k) \geqq \sum_{k=0}^{m} p(G, k)(m-2k).$$

But the last sum equals $mp(G)-2p'(G)$. Accordingly $p'(G)/p(G) \geqq m/3$. Since $p'(G)/p(G)=E(S)$, we are finished. ∎

In view of the above discussion we make the following:

4.2. Conjecture. *Let $\{G_n\}_{n=1}^{\infty}$ be a sequence of graphs such that $\nu(G_n)/|V(G_n)|$ is bounded below by a constant. Then asymptotically the numbers $p(G_n, k)$ $(k=0, 1, \ldots)$ are normally distributed.*

This conjecture is offered more in the hope of provoking an answer than from a confident belief in its truth. The condition given does at least exclude the known counterexamples and include all cases we know where normality occurs.

If the above conjecture proves intractable, or false, we offer as a replacement:

4.3. Conjecture. *Let* $\{G_n\}_{n=1}^{\infty}$ *be a sequence of graphs, such that the average vertex degree* $|E(G_n)|/|V(G_n)|$ *is proportional to* $|V(G_n)|$. *Then asymptotically the numbers* $p(G_n, k)$ *are normally distributed.*

The truth of the first conjecture would imply the truth of the second. The proof of this is left as an exercise.

References

[1] E. A. Bender, Central and local limit theorems applied to asymptotic enumeration, *J. Combinatorial Theory A*, **15** (1973), 91—111.
[2] E. Rodney Canfield, Application of the Berry—Esséen inequality to combinatorial estimates, *J. Combinatorial Theory A*, **28** (1980), 17—25.
[3] W. Feller, *An introduction to Probability Theory and its Applications*, Vol. 1, Wiley, New York, 1968.
[4] W. Feller, *An introduction to Probability Theory and its Applications*, Vol. II, Wiley, New York, 1966.
[5] C. D. Godsil and I. Gutman, On the theory of the matching polynomial, *J. Graph Theory*, **5** (1981), 137—144.
[6] C. D. Godsil, Hermite polynomials and a duality relation for matchings polynomials, *Combinatorica*, **1** (1981), 257—262.
[7] C. D. Godsil, Matchings and walks in graphs, *J. Graph Theory*, to appear.
[8] G. H. Hardy, J. E. Littlewood and G. Pólya, *Inequalities*, Cambridge University Press, Cambridge (1934).
[9] L. H. Harper, Stirling behaviour is asymptotically normal, *Ann. Math. Statist.* **38** (1967), 410—414.
[10] O. J. Heilmann and E. H. Lieb, Theory of monomer-dimer systems, *Comm. Math. Phys.* **25** (1972), 190—232.
[11] L. Lovász, *Combinatorial Problems and Exercises*, North-Holland, Amsterdam (1979).
[12] L. Moser and M. Wyman, On solutions of $x^d=1$ in symmetric groups, *Canad. J. Math.* **7** (1955), 159—168.
[13] J. Riordan, *An Introduction to Combinatorial Analysis*, Wiley, New York (1958).