Probability in the Engineering and Informational Sciences, **27**, 2013, 247–260.
doi:10.1017/S0269964812000447

THE ZAGREB INDICES OF RANDOM GRAPHS

QUNQIANG FENG, ZHISHUI HU, AND CHUN SU
Department of Statistics and Finance, School of Management
University of Science and Technology of China
Hefei 230026, China
E-mail: fengqq@ustc.edu.cn; huzs@ustc.edu.cn; suchun@ustc.edu.cn

Several limit laws for the Zagreb indices of the classical Erdös–Rényi random graphs are investigated in this paper. We have obtained the necessary and sufficient condition for the asymptotic normality of the two Zagreb indices (suitably normalized), as well as the explicit values for the means and variances of both the indices. Besides, the limiting joint distribution of the numbers of paths of various lengths is also studied under several conditions.

1. INTRODUCTION

In chemistry, a topological index is a map from the set of chemical compounds represented by molecular graphs to the set of real numbers, where the graphs are generated from molecules by replacing atoms with vertices and bonds with edges, or represent only bare molecular skeletons, that is, molecular skeletons without hydrogen atoms. Many topological indices are closely correlated with some physico-chemical characteristics of the underlying compounds.

The well-known Zagreb indices were introduced by chemists Gutman and Trinajstić [4]. Let G be a simple graph with vertex set $V(G)$ and edge set $E(G)$. The first Zagreb index $Z(G)$ and the second Zagreb index $\widetilde{Z}(G)$ of G are defined, respectively, as

$$Z(G) = \sum_{v \in V(G)} [d(v)]^2, \quad \widetilde{Z}(G) = \sum_{uv \in E(G)} d(u)d(v),$$

where $d(v)$ denotes the degree of the vertex v in G. That is, the first Zagreb index of a graph is the sum of the squares of the degrees of all vertices, and the second Zagreb index is the sum of products of the degrees of all pairs of adjacent vertices.

For a survey of the application of the Zagreb indices in chemistry, we refer to Nikolić, Kovačević, and Miličević [8] and the references therein. Besides, the Zagreb indices also attracts attention in graph theory (see, e.g., Abdo, Dimitrov, and Gutman [1], Nikiforov [7], and Peled, Petreschi, and Sterbini [9]). Using the method of martingale limit theorem, the first Zagreb index of a random recursive tree is studied by Feng and Hu [3].

Our main concern here is to study the Zagreb indices of the classical Erdös–Rényi random graphs. This simple model is specified by two parameters: the number of vertices in the graph n, and the probability of an edge p. Given n and p, we choose a graph on n vertices by including an edge between each pair of vertices with probability p, independently

 0269-9648/13 $25.00

for each pair. For numerous results on Erdös–Rényi random graphs, we refer the reader to the books Bollobás [2] and Janson, Łuczak, and Ruciński [5].

The paper is organized as follows. We first show a simple relation between the first Zagreb index and the numbers of edges and paths of length two, from which the mean and variance of this index are given. Then as the size of the random graph goes to infinity, the necessary and sufficient condition for the asymptotic normality of this index is given. We also get the parallel results for the second Zagreb index, where the necessary and sufficient condition for the asymptotic normality is the same as that of the first one. Finally, we also study the asymptotic multivariate normality for the numbers of paths of various lengths under several conditions.

2. THE FIRST ZAGREB INDEX

Let $G(n,p)$ be an Erdös–Rényi random graph on the set of vertices $\{1,2,\ldots,n\}$. We shall usually consider the probability $p=p(n)$ as a function of n. For any pair $1\le i,j\le n$, we define indicators $I_{ij}=\mathbf{1}(i \text{ and } j \text{ is connected in } G(n,p))$. Note that $I_{ii}=0, I_{ij}=I_{ji}$, and $\{I_{ij}, 1\le i<j\le n\}$ is a family of i.i.d. Bernoulli random variables with success rate p. Then the degree of any vertex i in $G(n,p)$ is equal to $\sum_{j=1}^{n} I_{ij}$. It thus follows that the first Zagreb index of a random graph $G(n,p)$ can be expressed as

$$\begin{aligned} Z_n &= \sum_{i=1}^{n}\left(\sum_{j=1}^{n} I_{ij}\right)^2 \\ &= \sum_{i=1}^{n}\left(\sum_{j=1}^{n} I_{ij} + \sum_{j\neq k} I_{ij}I_{ik}\right) \\ &= 2(E_n + P_{2,n}), \end{aligned} \tag{1}$$

where

$$E_n = \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} I_{ij} \quad \text{and} \quad P_{2,n} = \sum_{i=1}^{n}\sum_{j<k} I_{ij}I_{ik}$$

denote the numbers of edges and paths of length two in $G(n,p)$, respectively.

2.1. The Mean and Variance

First, we shall compute the first two moments of Z_n. Note that E_n has a binomial distribution with parameters $\binom{n}{2}$ and p. Then

$$\mathbf{E}[E_n] = \frac{1}{2}n(n-1)p, \tag{2}$$

$$\mathbf{Var}[E_n] = \frac{1}{2}n(n-1)p(1-p). \tag{3}$$

For $P_{2,n}$, it is clear that

$$\mathbf{E}[P_{2,n}] = 3\binom{n}{3}p^2 = \frac{1}{2}n(n-1)(n-2)p^2. \tag{4}$$

For any path of length two T in K_n, the complete graph on vertices $\{1,2,\ldots,n\}$, define a indicator variable $\mathbf{1}_T = \mathbf{1}(T\subset G(n,p))$. Note that $\mathbf{1}_T$ and $\mathbf{1}_{T'}$ are independent if T and T'

have no common edges. In K_n there are $3\binom{n}{3}\cdot 2(2n-5)=n(n-1)(n-2)(2n-5)$ ordered pairs (T,T') such that T and T' have exactly one common edge. Then for the variance of $P_{2,n}$, we have

$$\begin{aligned}\mathbf{Var}[P_{2,n}] &= \sum_{T,T'\subset K_n}\mathrm{Cov}(\mathbf{1}_T,\mathbf{1}_{T'})\\ &= 3\binom{n}{3}\mathbf{Var}[I_{12}I_{13}] + n(n-1)(n-2)(2n-5)\mathrm{Cov}(I_{12}I_{13}, I_{12}I_{14})\\ &= \frac{1}{2}n(n-1)(n-2)p^2(1-p)\Big[1+(4n-9)p\Big]\\ &= \frac{1}{2}n^3p^2(1-p)(1+4np)(1+O(n^{-1})). \qquad (5)\end{aligned}$$

Since any path of length two contains exactly two edges, we can get that

$$\begin{aligned}\mathrm{Cov}(E_n,P_{2,n}) &= 2\cdot 3\binom{n}{3}\mathrm{Cov}(I_{12}, I_{12}I_{13})\\ &= n(n-1)(n-2)p^2(1-p). \qquad (6)\end{aligned}$$

Hence, by (1), (2) and (4), we have

$$\begin{aligned}\mathbf{E}[Z_n] &= 2(\mathbf{E}[E_n]+\mathbf{E}[P_{2,n}])\\ &= n(n-1)p + n(n-1)(n-2)p^2\\ &= n^2p(1+np)(1+O(n^{-1})), \qquad (7)\end{aligned}$$

and by (3), (5) and (6),

$$\begin{aligned}\mathbf{Var}[Z_n] &= 4(\mathbf{Var}[E_n]+\mathbf{Var}[P_{2,n}]+2\mathrm{Cov}(E_n,P_{2,n}))\\ &= 2n(n-1)p(1-p)[1+5(n-2)p+(n-2)(4n-9)p^2]\\ &= 2n^2p(1-p)(1+5np+4n^2p^2)(1+O(n^{-1})). \qquad (8)\end{aligned}$$

The first two moments lead to several weak laws for the first Zagreb index as follows.

Proposition 1: *Let Z_n be the first Zagreb index of a random graph $G(n,p)$. As $n\to\infty$, the following assertions hold:*

(i) If $n^2p\to 0$, then $Z_n \xrightarrow{\mathcal{P}} 0$;

(ii) If there exists a constant $c>0$ such that $n^2p\to c$, then $Z_n/2 \xrightarrow{\mathcal{D}} \mathrm{Poi}(c/2)$;

(iii) If $n^2p\to\infty$, then $Z_n/(n^2p+n^3p^2) \xrightarrow{\mathcal{P}} 1$.

Proof: If $n^2p\to 0$, then both $\mathbf{E}[Z_n]$ and $\mathbf{Var}[Z_n]$ tend to 0, which implies (i). If $n^2p\to\infty$, by (7) and (8) it is not hard to check that $\mathbf{Var}[Z_n]=o(\mathbf{E}[Z_n])^2$, which yields $Z_n/\mathbf{E}[Z_n]$ converges in probability to 1 by Chebyshev's inequality. Then (iii) follows by (7). We now let $n^2p\to c>0$. Note that both $\mathbf{E}[P_{2,n}]$ and $\mathbf{Var}[P_{2,n}]$ tend to 0, which implies that $P_{2,n}$ converges in probability to 0. Then by (1) and Slutsky's theorem, to prove (ii), it is sufficient to show that E_n converges in distribution to $\mathrm{Poi}(c/2)$, which in fact is known (see Theorem 3.19 of Janson et al. [5]). ∎

Symmetrically, if $n^2(1-p) \to c$ for some $0 \le c < \infty$, we have the following result.

Proposition 2: *Let Z_n be the first Zagreb index of a random graph $G(n,p)$. As $n \to \infty$, the following assertions hold:*

(i) If $n^2(1-p) \to 0$, then $\mathbf{P}(Z_n = n(n-1)^2) \to 1$;
(ii) If there exists a constant $c > 0$ such that $n^2(1-p) \to c$, then

$$\frac{n(n-1)^2 - Z_n}{4n} \xrightarrow{\mathcal{D}} \mathrm{Poi}(c/2);$$

(iii) If $n^2(1-p) \to \infty$, then

$$\frac{n(n-1)^2 - Z_n}{n^3(1-p^2)} \xrightarrow{\mathcal{P}} 1.$$

Proof: If $n^2(1-p) \to \infty$, it follows from (7) that

$$\begin{aligned} n(n-1)^2 - \mathbf{E}[Z_n] &= n(n-1)(1-p)[n-1+(n-2)p] \\ &= n^3(1-p^2)(1+O(n^{-1})). \end{aligned}$$

Then

$$\mathbf{Var}[n(n-1)^2 - Z_n] = \mathbf{Var}[Z_n] = o\Big(\mathbf{E}[n(n-1)^2 - Z_n]\Big)^2,$$

which implies (iii) by Chebyshev's inequality.

Consider now the complement graph $\overline{G}(n,p)$. Let $\overline{E}_n$ and $\overline{P}_{2,n}$ denote the numbers of edges and paths of length two in $\overline{G}(n,p)$, respectively. For any pair $1 \le i,j \le n$, we also define $\overline{I}_{ij} = \mathbf{1}(i$ and j is connected in $\overline{G}(n,p))$. It is easy to see that $I_{ij} + \overline{I}_{ij} = 1$ for any $1 \le i \ne j \le n$. Then

$$\begin{aligned} Z_n &= \sum_{i=1}^{n}\left(n-1-\sum_{j=1}^{n}\overline{I}_{ij}\right)^2 \\ &= n(n-1)^2 - 2(n-1)\sum_{i=1}^{n}\sum_{j=1}^{n}\overline{I}_{ij} + \sum_{i=1}^{n}\left(\sum_{j=1}^{n}\overline{I}_{ij}\right)^2 \\ &= n(n-1)^2 - (4n-6)\overline{E}_n + 2\overline{P}_{2,n}. \end{aligned}$$

If $n^2(1-p) \to c$, then from the proof of Proposition 1 and by symmetric $\overline{E}_n$ converges in distribution to $\mathrm{Poi}(c/2)$ and $\overline{P}_{2,n}$ converges in probability to 0, which implies (ii).

If $n^2(1-p) \to 0$, also by symmetric both $\mathbf{E}[\overline{E}_n]$ and $\mathbf{Var}[\overline{E}_n]$ tend to 0. It thus follows that the probability that $\overline{G}(n,p)$ is empty tends to 1. Then

$$\mathbf{P}(Z_n = n(n-1)^2) = \mathbf{P}(G(n,p) \text{ is a complete graph}) \to 1,$$

and (i) holds. ■

2.2. Asymptotic Normality

Let $\{X_i\}_{i\in\mathcal{I}}$ be a family of random variables defined on a common probability space. A dependency graph for $\{X_i\}$ is any graph L with vertex set $V(L)=\mathcal{I}$ such that if A and B are two disjoint subsets of $\mathcal{I}$ with no edges between A and B, then the families $\{X_i\}_{i\in A}$ and $\{X_i\}_{i\in B}$ are mutually independent. For any integer $r\geq 1$ and $i_1,i_2,\ldots,i_r\in\mathcal{I}$, we also let

$$\overline{N}_L(i_1,i_2,\ldots,i_r)=\bigcup_{k=1}^{r}\{j\in\mathcal{I}:j=i_k \text{ or } j \text{ is adjacent to } i_k \text{ in } L\}$$

denote the closed neighborhood of $\{i_1,i_2,\ldots,i_r\}$ in L.

The following lemma plays an important role in our proofs of the asymptotic normality for the Zagreb indices of $G(n,p)$ as $n\to\infty$. It first appeared in Mikhailov [6], and was also stated as Theorem 6.33 in Janson et al. [5].

Lemma 1: *Suppose that $\{S_n\}_{n=1}^{\infty}$ is a sequence of random variables such that $S_n=\sum_{\alpha\in A_n}X_{n\alpha}$, where for each n, $\{X_{n\alpha},\alpha\in A_n\}$ is a family of random variables with dependency graph L_n. Suppose further that there exist numbers M_n and Q_n such that $\sum_{\alpha\in A_n}\mathbf{E}[|X_{n\alpha}|]\leq M_n$ and, for every $\alpha_1,\alpha_2\in A_n$,*

$$\sum_{\alpha\in\overline{N}_{L_n}(\alpha_1,\alpha_2)}\mathbf{E}[|X_{n\alpha}||X_{n\alpha_1},X_{n\alpha_2}]\leq Q_n.$$

As $n\to\infty$, if $M_nQ_n^2/(\mathbf{Var}[S_n])^{3/2}\to 0$, then

$$\frac{S_n-\mathbf{E}[S_n]}{\sqrt{\mathbf{Var}[S_n]}}\xrightarrow{\mathcal{D}}N(0,1).$$

We now state the necessary and sufficient condition for the asymptotic normality of the first Zagreb index of $G(n,p)$ as $n\to\infty$.

Theorem 1: *Let Z_n be the first Zagreb index of a random graph $G(n,p)$. As $n\to\infty$, if $n^2p(1-p)\to\infty$, then*

$$\frac{Z_n-n(n-1)p-n(n-1)(n-2)p^2}{n\sqrt{2p(1-p)(1+5np+4n^2p^2)}}\xrightarrow{\mathcal{D}}N(0,1).$$

Conversely, if $(Z_n-a_n)/b_n\xrightarrow{\mathcal{D}}N(0,1)$ for some constants a_n and b_n, then $n^2p(1-p)\to\infty$.

Proof: The second statement of the theorem is clear from Propositions 1 and 2. If $n^2p(1-p)$ does not go to ∞, then there exists a sequence $\{n_k,k=1,2,\ldots\}$ such that $n_k^2p(1-p)\to c_1$ and $p(n_k)\to c_2$ with $0\leq c_1<\infty$ and $0\leq c_2\leq 1$. Clearly, $c_2=0$ or 1. Thus $n_k^2p\to c_1$ or $n_k^2(1-p)\to c_1$. By Propositions 1 and 2, either $n_k^2p\to c$ or $n_k^2(1-p)\to c$, the subsequence $\{Z_{n_k}\}$ rules out asymptotic normality for any normalization.

Now we turn to the first statement. By (7) and (8), the desired result is equivalent to

$$\frac{Z_n-\mathbf{E}[Z_n]}{\sqrt{\mathbf{Var}[Z_n]}}\xrightarrow{\mathcal{D}}N(0,1). \tag{9}$$

We shall prove (9) by three steps.

https://doi.org/10.1017/S0269964812000447 Published online by Cambridge University Press

Step 1: we show that (9) holds under a stronger condition $n^2p(1-p)^3 \to \infty$.

To proceed, Lemma 1 will be applied. Let us denote the set of all edges and paths of length two in K_n by $\{T_\alpha\}_{\alpha\in A_n}$, where A_n is an index set and its cardinality is equal to $\binom{n}{2}+3\binom{n}{3}=\frac{1}{2}n(n-1)^2$. For any $\alpha \in A_n$, we let $X_\alpha = \mathbf{1}(T_\alpha \subset G(n,p))$, for simplicity omitting subscripts n. Construct a graph L_n with vertex set A_n by connecting every pair of indices α and β such that the corresponding graphs T_α and T_β have a common edge. Then it is evidently a dependency graph for $\{X_\alpha\}$. Note that

$$Z_n = 2(E_n + P_{2,n}) = 2\sum_{\alpha\in A_n} X_\alpha.$$

We now verify the conditions of Lemma 1 for $Z_n/2 = \sum_{\alpha\in A_n} X_\alpha$. The value of M_n can be simply set as $n^2p(1+np)$. Suppose that $\alpha_1, \alpha_2 \in A_n$ are given. Consider the union $T_{\alpha_1} \cup T_{\alpha_2}$, which has the vertex set V. Let K_V be the complete graph on V. Note that K_V has at most 6 vertices, and the cardinality of $\overline{N}_{L_n}(\alpha_1,\alpha_2) \cap \{\alpha : T_\alpha \subset K_V\}$ is not more than 36. Moreover, each T_β for $\beta \in \overline{N}_{L_n}(\alpha_1,\alpha_2) \setminus \{\alpha : T_\alpha \subset K_V\}$ is such a path that one edge is in K_V and another one is not, hence the number of such β is less than $8n$. We thus have

$$\sum_{\alpha\in\overline{N}_{L_n}(\alpha_1,\alpha_2)} \mathbf{E}[|X_\alpha| | X_{\alpha_1}, X_{\alpha_2}] < 36 + 8np < 36(1+np).$$

Therefore, we can set $Q_n = 36(1+np)$. It follows that

$$\begin{aligned}\frac{M_n Q_n^2}{(\mathbf{Var}[E_n+P_{2,n}])^{3/2}} &= \frac{36^2 n^2 p(1+np)^3}{\left[\frac{1}{2}n^2p(1-p)(1+5np+4n^2p^2)\right]^{3/2}}(1+O(n^{-1})) \\ &= O([n^2p(1-p)^3]^{-1/2}) \to 0,\end{aligned}$$

which implies by Lemma 1 that (9) holds if $n^2p(1-p)^3 \to \infty$.

Step 2: We show that (9) also holds if $n^2(1-p) \to \infty$ and $np \to \infty$.

At this time, we have $\mathbf{Var}[E_n] = o(\mathbf{Var}[E_n + P_{2,n}])$ and $\mathbf{Var}[P_{2,n}] \sim \mathbf{Var}[E_n + P_{2,n}]$. By Theorem 6.5 of Janson et al. [5],

$$\frac{E_n - \mathbf{E}[E_n]}{\sqrt{\mathbf{Var}[E_n]}} \xrightarrow{\mathcal{D}} N(0,1), \quad \frac{P_{2,n} - \mathbf{E}[P_{2,n}]}{\sqrt{\mathbf{Var}[P_{2,n}]}} \xrightarrow{\mathcal{D}} N(0,1).$$

Then

$$\begin{aligned}\frac{Z_n - \mathbf{E}[Z_n]}{\sqrt{\mathbf{Var}[Z_n]}} &= \frac{(E_n + P_{2,n}) - \mathbf{E}[E_n + P_{2,n}]}{\sqrt{\mathbf{Var}[E_n + P_{2,n}]}} \\ &= \frac{E_n - \mathbf{E}[E_n]}{\sqrt{\mathbf{Var}[E_n]}}\frac{\sqrt{\mathbf{Var}[E_n]}}{\sqrt{\mathbf{Var}[E_n + P_{2,n}]}} + \frac{P_{2,n} - \mathbf{E}[P_{2,n}]}{\sqrt{\mathbf{Var}[P_{2,n}]}}\frac{\sqrt{\mathbf{Var}[P_{2,n}]}}{\sqrt{\mathbf{Var}[E_n + P_{2,n}]}} \\ &\xrightarrow{\mathcal{D}} N(0,1).\end{aligned}$$

Step 3: we shall summarize the above two steps to get (9) under the desired condition $n^2p(1-p) \to \infty$.

It is enough to show that for any sequence $\{n_k, k = 1, 2, \ldots\}$, there exists a subsequence $\{n'_k, k = 1, 2, \ldots\}$ such that

$$\frac{Z_{n'_k} - \mathbf{E}[Z_{n'_k}]}{\sqrt{\mathbf{Var}[Z_{n'_k}]}} \xrightarrow{\mathcal{D}} N(0,1). \tag{10}$$

For any sequence $\{n_k, k = 1, 2, \ldots\}$, if $n_k^2 p(1-p)^3 \to \infty$ as $k \to \infty$, then by Step 1, we have $(Z_{n_k} - \mathbf{E}[Z_{n_k}])/(\mathbf{Var}[Z_{n_k}])^{1/2}$ converges in distribution to $N(0,1)$. If $n_k^2 p(1-p)^3 \to \infty$ does not hold, then there exists a subsequence $\{n'_k, k = 1, 2, \ldots\}$ and a constant $c < \infty$ such that $(n'_k)^2 p(1-p)^3 \to c$ as $k \to \infty$. Since $(n'_k)^2 p(1-p) \to \infty$, we have

$$(1-p)^2 = \frac{(n'_k)^2 p(1-p)^3}{(n'_k)^2 p(1-p)} \to 0.$$

Thus $p \to 1$ and

$$n'_k p \to \infty, (n'_k)^2(1-p) \to \infty.$$

Now by Step 2, it follows that (10) holds. The proof of Theorem 1 is complete. ■

Remark: The condition $n^2 p(1-p) \to \infty$ is equivalent to that $n^2 p \to \infty$ and $n^2(1-p) \to \infty$. Then the above theorem implies that Z_n is asymptotic normal (suitably normalized) if and only if both the numbers of edges in $G(n,p)$ and its complement $\overline{G}(n,p)$ goes to infinity with probability 1.

3. THE SECOND ZAGREB INDEX

Let $\widetilde{Z}_n$ be the second Zagreb index of a random graph $G(n,p)$. By the definition,

$$\begin{aligned}\widetilde{Z}_n &= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} I_{ij}(1 + \sum_{k\neq j} I_{ik})(1 + \sum_{l \neq i} I_{jl}) \\ &= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} I_{ij} + \sum_{i=1}^{n-1} \sum_{k \neq j} I_{ij} I_{ik} + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \sum_{k\neq j} \sum_{l \neq i} I_{ik} I_{ij} I_{jl} \\ &= E_n + 2P_{2,n} + P_{3,n} + 3\triangle_n, \end{aligned} \tag{11}$$

where $P_{3,n}$ and $\triangle_n$ denote the numbers of paths of length three and triangles in $G(n,p)$, respectively.

It is easy to see that

$$\mathbf{E}[P_{3,n}] = 12\binom{n}{4}p^3, \quad \mathbf{E}[\triangle_n] = \binom{n}{3}p^3.$$

By (2), (4), and (11), then we have

$$\begin{aligned}\mathbf{E}[\widetilde{Z}_n] &= \binom{n}{2}p + 6\binom{n}{3}p^2 + 12\binom{n}{4}p^3 + 3\binom{n}{3}p^3 \\ &= \frac{1}{2}n(n-1)p[1 + (n-2)p]^2 \\ &= \frac{1}{2}n^2 p(1+np)^2(1 + O(n^{-1})).\end{aligned}$$

https://doi.org/10.1017/S0269964812000447 Published online by Cambridge University Press

We can further obtain the variance of $\widetilde{Z}_n$ (see the details for the calculation in the Appendix):

$$\begin{aligned}\mathbf{Var}[\widetilde{Z}_n] &= \frac{1}{2}n(n-1)p(1-p)[1+(n-2)p(12+(31n-60)p \\ &\quad + (28n^2-114n+105)p^2 + (9n^3-58n^2+117n-69)p^3)] \\ &= \frac{1}{2}n^2p(1-p)[1+12np+31n^2p^2+28n^3p^3+9n^4p^4](1+O(n^{-1})).\end{aligned}$$

Analogous to Propositions 1 and 2, and Theorem 1, we state the corresponding results for the second Zagreb index in the following. They can be proved in a way very similar to that used for the first Zagreb index, and hence we omit their proofs.

PROPOSITION 3: *Let $\widetilde{Z}_n$ be the second Zagreb index of a random graph $G(n,p)$. As $n \to \infty$, the following assertions hold:*

(i) If $n^2p \to 0$, then $\widetilde{Z}_n \xrightarrow{\mathcal{P}} 0$;
(ii) If there exists a constant $c > 0$ such that $n^2p \to c$, then $\widetilde{Z}_n \xrightarrow{\mathcal{D}} \mathrm{Poi}(c/2)$;
(iii) If $n^2p \to \infty$, then $2\widetilde{Z}_n/(n^2p(1+np)^2) \xrightarrow{\mathcal{P}} 1$;
(iv) If $n^2(1-p) \to 0$, then $\mathbf{P}(\widetilde{Z}_n = n(n-1)^3/2) \to 1$;
(v) If there exists a constant $c > 0$ such that $n^2(1-p) \to c$, then

$$\frac{n(n-1)^3/2 - \widetilde{Z}_n}{3n^2} \xrightarrow{\mathcal{D}} \mathrm{Poi}(c/2);$$

(vi) If $n^2(1-p) \to \infty$, then

$$\frac{n(n-1)^3 - 2\widetilde{Z}_n}{n^4(1-p^3)} \xrightarrow{\mathcal{P}} 1.$$

THEOREM 2: *Let $\widetilde{Z}_n$ be the second Zagreb index of a random graph $G(n,p)$. As $n \to \infty$, if $n^2p(1-p) \to \infty$, then*

$$\frac{\widetilde{Z}_n - \frac{1}{2}n(n-1)p[1+(n-2)p]^2}{n\sqrt{\frac{1}{2}p(1-p)(1+12np+31n^2p^2+28n^3p^3+9n^4p^4)}} \xrightarrow{\mathcal{D}} N(0,1).$$

Conversely, if $(\widetilde{Z}_n - a_n)/b_n \xrightarrow{\mathcal{D}} N(0,1)$ for some constants a_n and b_n, then $n^2p(1-p) \to \infty$.

From the fact $\mathbf{Var}[\triangle_n] = o(\mathbf{Var}[\widetilde{Z}_n])$, one can see that $(\triangle_n - \mathbf{E}[\triangle_n])/(\mathbf{Var}[\widetilde{Z}_n])^{1/2}$ converges in probability to 0, that is, the contribution of $\triangle_n$ to the second Zagreb index can be always negligible. Moreover, the influence of the coefficients in (11) on our proof can be also negligible. In fact, using the method in the proof of Theorem 1 one can obtain that for any real numbers $a > 0, b \geq 0, c \geq 0$, the sequence $\{aE_n + bP_{2,n} + cP_{3,n}\}$ is also asymptotically normal distributed (suitably normalized), as $n \to \infty$. We shall generalize this result for the numbers of paths of various lengths in the next section.

4. PATHS

For any integer $k \geq 1$, let $P_{k,n}$ be the numbers of the paths of length k in a random graph $G(n, p)$ with $P_{1,n} = E_n$. As a consequence of Theorem 6.5 of Janson et al. [5], if $n^{k+1}p^k \to \infty$ and $n^2(1-p) \to \infty$, then $(P_{k,n} - \mathbf{E}[P_{k,n}])/(\mathbf{Var}[P_{k,n}])^{1/2}$ converges in distribution to $N(0, 1)$ as $n \to \infty$ for any $k \geq 1$. In this section, we shall study the limiting joint distribution of the random (column) vector $(P_{1,n}, P_{2,n}, \ldots, P_{m,n})'$ under suitable normalization for any fixed positive integer m as $n \to \infty$.

The first two moments of $P_{k,n}$ are given first. One can easily get the expectation of $P_{k,n}$ for $k \geq 1$:

$$\mathbf{E}[P_{k,n}] = \frac{(k+1)!}{2}\binom{n}{k+1}p^k = \frac{1}{2}n^{k+1}p^k(1 + O(n^{-1})). \tag{12}$$

For the variance of $P_{k,n}$, it is manifestly difficult to compute the exact expression when both k and n are large. However, we have the asymptotic expansion of $\mathrm{Cov}(P_{k,n}, P_{l,n})$ for $1 \leq k \leq l \leq n-1$ as follows (also see the details for the calculation in the Appendix):

$$\mathrm{Cov}(P_{k,n}, P_{l,n}) = \frac{n(1-p)}{2}\sum_{j=1}^{k} j(l-k+j)(np)^{l+j-1}(1 + O(n^{-1})). \tag{13}$$

We now state the main result in this section as follows.

THEOREM 3: *For any fixed integer $m \geq 1$, as $n \to \infty$, the following assertions hold:*

(i) If $np \to c$ for some constant $c > 0$, then

$$\frac{1}{\sqrt{n}}\left(P_{1,n} - \frac{1}{2}cn, P_{2,n} - \frac{1}{2}c^2 n, \ldots, P_{m,n} - \frac{1}{2}c^m n\right)' \xrightarrow{\mathcal{D}} N(\mathbf{0}, \mathbf{\Sigma}_m),$$

where $\mathbf{0}$ is an m-dimensional vector of zeros and $\mathbf{\Sigma}_m = (\sigma_{kl})_{m\times m}$ with

$$\sigma_{kl} = \frac{1}{2}\sum_{j=1}^{\min(k,l)} j(|l-k| + j)c^{\max(k,l)+j-1}, \quad 1 \leq k, l \leq m.$$

(ii) If $np \to 0$ and $n^{m+1}p^m \to \infty$, then

$$\left(\frac{P_{1,n} - \frac{1}{2}n^2 p}{\sqrt{\frac{1}{2}n^2 p}}, \frac{P_{2,n} - \frac{1}{2}n^3 p^2}{\sqrt{\frac{1}{2}n^3 p^2}}, \ldots, \frac{P_{m,n} - \frac{1}{2}n^{m+1}p^m}{\sqrt{\frac{1}{2}n^{m+1}p^m}}\right)' \xrightarrow{\mathcal{D}} N(\mathbf{0}, \mathbf{I}_m),$$

where $\mathbf{I}_m$ is an identity matrix of size m.

(iii) If $np \to \infty$ and $n^2(1-p) \to \infty$, then

$$\sqrt{\frac{2p}{1-p}}\left(\frac{P_{1,n} - \mathbf{E}[P_{1,n}]}{np}, \frac{P_{2,n} - \mathbf{E}[P_{2,n}]}{2(np)^2}, \ldots, \frac{P_{m,n} - \mathbf{E}[P_{m,n}]}{m(np)^m}\right)' \xrightarrow{\mathcal{D}} X\mathbf{1}_m,$$

where X is a standard normal random variable and $\mathbf{1}_m$ is the vector with all entries of 1.

PROOF: We shall prove (i), and only give a sketch of proofs of (ii) and (iii).

https://doi.org/10.1017/S0269964812000447 Published online by Cambridge University Press

If $np \to c > 0$, we have $\mathbf{E}[P_{k,n}] = c^k n/2 + O(1)$ by (12), and for $1 \le k \le l$,

$$\lim_{n\to\infty} \operatorname{Cov}\left(\frac{P_{k,n}}{\sqrt{n}}, \frac{P_{l,n}}{\sqrt{n}}\right) = \frac{1}{2}\sum_{j=1}^{k} j(l-k+j)c^{l+j-1}, \tag{14}$$

by (13). For any integer $m \ge 1$, let $\boldsymbol{\Sigma}_m(n)$ be the covariance matrix of random vector $\mathbf{P}_m(n)/\sqrt{n}$, where

$$\mathbf{P}_m(n) := (P_{1,n}, P_{2,n}, \ldots, P_{m,n})'.$$

By (14), one can get that the matrix $\boldsymbol{\Sigma}_m$, the limit of $\boldsymbol{\Sigma}_m(n)$ as $n \to \infty$, is

$$\begin{pmatrix} \frac{1}{2}c & c^2 & \frac{3}{2}c^3 & \cdots & \frac{m}{2}c^m \\ c^2 & \frac{1}{2}c^2(1+4c) & c^3(1+3c) & \cdots & \frac{1}{2}c^m[(m-1)+2mc] \\ \frac{3}{2}c^3 & c^3(1+3c) & \frac{1}{2}c^3(1+4c+9c^2) & \cdots & \frac{1}{2}c^m\sum_{j=1}^{3} j(m-3+j)c^{j-1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{m}{2}c^m & \frac{1}{2}c^m[(m-1)+2mc] & \frac{1}{2}c^m\sum_{j=1}^{3} j(m-3+j)c^{j-1} & \cdots & \frac{1}{2}c^m\sum_{j=1}^{m} j^2c^{j-1} \end{pmatrix}.$$

By adding $-kc^{k-1}$ multiplies of the first column to the kth column for $k = 2, 3, \ldots, m$, the determinant of $\boldsymbol{\Sigma}_m$ satisfies that

$$\det(\boldsymbol{\Sigma}_m) = \frac{1}{2}c^m \det(\boldsymbol{\Sigma}_{m-1}), \quad m \ge 2.$$

With the initial value $\det(\boldsymbol{\Sigma}_1) = c/2$, the above recurrence implies that

$$\det(\boldsymbol{\Sigma}_m) = \lim_{n\to\infty} \det(\boldsymbol{\Sigma}_m(n)) = \frac{1}{2^m}c^{\frac{m(m+1)}{2}}, \quad m \ge 1. \tag{15}$$

Hence the limiting covariance matrix $\boldsymbol{\Sigma}_m$ is positive definite.

Let $\{a_1, a_2, \ldots, a_m\}$ be a sequence of real numbers, and

$$S_{m,n} := \sum_{j=1}^{m} a_j P_{j,n}.$$

It is now sufficient to prove that for any real numbers $a_1, a_2, \ldots, a_m$ with $\sum_{j=1}^{m} a_j^2 = 1$,

$$\frac{S_{m,n} - \mathbf{E}[S_{m,n}]}{\sqrt{\mathbf{Var}[S_{m,n}]}} \xrightarrow{\mathcal{D}} N(0,1). \tag{16}$$

We shall also apply Lemma 1. Denote by $\{T_\alpha\}_{\alpha\in B_n}$ the set of all paths of length not more than m in K_n, where B_n is an index set and its cardinality $|B_n|$ is the number of all

https://doi.org/10.1017/S0269964812000447 Published online by Cambridge University Press

such paths, that is,

$$|B_n| = \frac{1}{2}\sum_{i=1}^{m}(i+1)!\binom{n}{i+1}.$$

For any $\alpha \in B_n$, let t_α be the length of T_α and $X_\alpha = a_{t_\alpha}\mathbf{1}(T_\alpha \subset G(n,p))$. Then

$$S_{m,n} = \sum_{\alpha \in B_n} X_\alpha.$$

We can also define a dependency graph L_n^* with vertex set B_n for $\{X_\alpha : \alpha \in B_n\}$ by connecting every pair of indices α and β such that the corresponding graphs T_α and T_β have at least one common edge.

The procedure to verify the conditions of Lemma 1 is analogous to that in the proof of Theorem 1. Write $a_{\max} := \max\{|a_1|, |a_2|, \ldots, |a_m|\}$, and

$$M_n := a_{\max}\sum_{j=1}^{m}\mathbf{E}[P_{j,n}].$$

Then

$$\sum_{\alpha \in B_n}\mathbf{E}[|X_\alpha|] \le M_n,$$

and $M_n = O(n)$ by (12). We now determine the number Q_n for sufficient large n. For any given $\alpha_1, \alpha_2 \in B_n$, let V^* be the vertex set of the subgraph $T_{\alpha_1} \cup T_{\alpha_2}$, in which the number of vertices in V^* is not more than $2(m+1)$. For any $\alpha \in \overline{N}_{L_n^*}(\alpha_1, \alpha_2)$, consider the number of vertices of T_α which are not in V^*. Denote

$$\overline{N}_i := \{\alpha : \alpha \in \overline{N}_{L_n^*}(\alpha_1, \alpha_2) \text{ and } T_\alpha \text{ has exactly } i \text{ vertices out of } V^*\},$$

for $i = 0, 1, \ldots m-1$. Then $\cup_{i=0}^{m-1}\overline{N}_i = \overline{N}_{L_n^*}(\alpha_1, \alpha_2)$, and for any $\alpha \in \overline{N}_i$,

$$\mathbf{E}[|X_\alpha||X_{\alpha_1}, X_{\alpha_2}] \le a_{\max}p^i,$$

which implies that for any $0 \le i \le m-1$,

$$\sum_{\alpha \in \overline{N}_i}\mathbf{E}[|X_\alpha||X_{\alpha_1}, X_{\alpha_2}] = O(n^i p^i) = O(1).$$

Hence we can set Q_n as a sufficient large constant, which is independent of n. Let $\lambda_{\min} = \lambda_{\min}(m)$ be the smallest eigenvalue of $\mathbf{\Sigma}_m$. Then $\lambda_{\min} > 0$ by (15), and for any real vector $\mathbf{a} = (a_1, a_2, \ldots, a_m)'$ with $\mathbf{a}'\mathbf{a} = 1$,

$$\mathbf{Var}\Big(\frac{1}{\sqrt{n}}\mathbf{a}'\mathbf{P}_m(n)\Big) = \mathbf{a}'\mathbf{\Sigma}_m(n)\mathbf{a} \ge \lambda_{\min}(1+o(1)).$$

It thus follows that for any real numbers $a_1, a_2, \ldots, a_m$ with $\sum_{j=1}^m a_j^2 = 1$, if n is sufficiently large, then

$$\mathbf{Var}\left[\sum_{j=1}^{m} a_j P_{j,n}\right] \ge \frac{n}{2}\lambda_{\min}.$$

We thus have

$$\frac{M_n Q_n^2}{(\mathbf{Var}[S_{m,n}])^{3/2}} = O(n^{-1/2}),$$

which implies (16) by Lemma 1.

Applying Lemma 1 to prove (ii), one can verify that (16) is valid for the sum

$$S'_{m,n} = \sum_{j=1}^{m} a_j (np)^{(m-j)/2} P_{j,n},$$

where $a_1, a_2, \ldots, a_m$ are also arbitrary fixed real numbers with $\sum_{j=1}^{m} a_j^2 = 1$. Note that, by (13), in such case $\mathrm{Cov}(P_{k,n}, P_{l,n}) \to 0$ for any $k \neq l$.

If $np \to \infty$, then (13) also implies that $\mathrm{Corr}(P_{k,n}, P_{l,n}) \to 1$ for any $k, l \geq 1$. To prove (iii), one can verify that the sum

$$S''_{m,n} = \sqrt{\frac{2p}{1-p}} \sum_{j=1}^{m} \frac{a_j P_{j,n}}{j(np)^{j-1}}$$

also satisfies (16) for any such real numbers $a_1, a_2, \ldots, a_m$. ∎

APPENDIX: THE CALCULATIONS

Exact Variance of the Second Zagreb Index. To compute $\mathbf{E}[\widetilde{Z}_n]$, by (11) we need to get the covariance matrix of the random vector $(E_n, P_{2,n}, P_{3,n}, \triangle_n)$. The technique in (5) will be repeatedly used in the following calculations.

Since there are $12\binom{n}{4}$ paths of length three in K_n and each such path contains exactly three edges,

$$\mathrm{Cov}(E_n, P_{3,n}) = 3 \cdot 12 \binom{n}{4} (p^3 - p^4) = \frac{3}{2} n(n-1)(n-2)(n-3)p^3(1-p).$$

Similarly,

$$\mathrm{Cov}(E_n, \triangle_n) = 3 \binom{n}{3} (p^3 - p^4) = \frac{1}{2} n(n-1)(n-2)p^3(1-p).$$

For any path of length three in K_n, count such paths of length two that have exactly one or two common edges with it. It is not hard to get the quantities are $6n - 16$ and 2, respectively. Then

$$\begin{aligned}\mathrm{Cov}(P_{2,n}, P_{3,n}) &= 12 \binom{n}{4} [(6n-16)p^4(1-p) + 2p^3(1-p^2)] \\ &= n(n-1)(n-2)(n-3)p^3(1-p)[1 + (3n-7)p].\end{aligned}$$

Also in a similar way, we have

$$\begin{aligned}\mathrm{Cov}(P_{2,n}, \triangle_n) &= \binom{n}{3} [6(n-3)p^4(1-p) + 3p^3(1-p^2)] \\ &= \frac{1}{2} n(n-1)(n-2)p^3(1-p)[1 + (2n-5)p].\end{aligned}$$

Note that any two distinct triangles in K_n have at most one common edge. It follows that

$$\begin{aligned}\mathbf{Var}[\triangle_n] &= \binom{n}{3} [p^3(1-p^3) + 3(n-3)p^5(1-p)] \\ &= \frac{1}{6} n(n-1)(n-2)p^3(1-p)[1 + p + (3n-8)p^2].\end{aligned}$$

In K_n, fix a path of length three and count the triangles which have common edges with it. Then there are $2(n-3) + n - 4 = 3n - 10$ and 2 such triangles that have one and two common edges

https://doi.org/10.1017/S0269964812000447 Published online by Cambridge University Press

with the fixed path, respectively. We thus have

$$\begin{aligned}\mathrm{Cov}(P_{3,n},\triangle_n) &= 12\binom{n}{4}[(3n-10)p^5(1-p)+2p^4(1-p^2)]\\ &= \frac{1}{2}n(n-1)(n-2)(n-3)p^4(1-p)[2+(3n-8)p].\end{aligned}$$

We now treat the variance of $P_{3,n}$. Count the paths of length three which have common edges with the specified one $1-2-3-4$. For the case of only one common edge, consider the edge $1-2$ and $2-3$ separately. If a path of length three only have the common edge $1-2$ with the specified one, then it is one of the following three types:

$$x-y-1-2, \quad x-1-2-y, \quad 1-2-x-y,$$

where x,y are suitably chosen in $\{3,4,\ldots,n\}$. It is straightforward to see the numbers of desired paths in the above three types are $(n-1)(n-4)$, $(n-3)^2$ and $(n-2)(n-4)$, respectively. Then there are $3n^2-17n+21$ paths just having the common edge $1-2$ with the specified one. For the paths that have only common edge $2-3$, by symmetry we may just consider two types of paths: $x-y-2-3$ and $x-2-3-y$. One can see the total number of such paths is $2(n-3)^2+[(n-4)^2+(n-3)]=3n^2-19n+31$. Hence, the number of paths of length three in K_n which have only one common edge with the path $1-2-3-4$ is

$$2(3n^2-17n+21)+3n^2-19n+31=9n^2-53n+73.$$

With the fact that in K_n there are $4n-11$ paths of length three containing just two edges of the path $1-2-3-4$, we have

$$\begin{aligned}\mathbf{Var}[P_{3,n}] &= 12\binom{n}{4}[p^3(1-p^3)+(4n-11)p^4(1-p^2)+(9n^2-53n+73)p^5(1-p)]\\ &= \frac{1}{2}n(n-1)(n-2)(n-3)p^3(1-p)[1+(4n-10)p+(9n^2-49n+63)p^2].\end{aligned}$$

Collecting above all variances and covariances, by (3), (5), (6) and (11), we have

$$\begin{aligned}\mathbf{Var}[\widetilde{Z}_n] &= \mathbf{Var}[E_n]+4\mathbf{Var}[P_{2,n}]+\mathbf{Var}[P_{3,n}]+9\mathbf{Var}[\triangle_n]\\ &\quad +4\mathrm{Cov}(E_n,P_{2,n})+2\mathrm{Cov}(E_n,P_{3,n})+6\mathrm{Cov}(E_n,\triangle_n)\\ &\quad +4\mathrm{Cov}(P_{2,n},P_{3,n})+12\mathrm{Cov}(P_{2,n},\triangle_n)+6\mathrm{Cov}(P_{3,n},\triangle_n)\\ &= \frac{1}{2}n(n-1)p(1-p)\Big[1+4(n-2)p[1+(4n-9)p]\\ &\quad +(n-2)(n-3)p^2[1+(4n-10)p+(9n^2-49n+63)p^2]\\ &\quad +3(n-2)p^2[1+p+(3n-8)p^2]+8(n-2)p+6(n-2)(n-3)p^2\\ &\quad +6(n-2)p^2+8(n-2)(n-3)p^2[1+(3n-7)p]\\ &\quad +12(n-2)p^2[1+(2n-5)p]+6(n-2)(n-3)p^3[2+(3n-8)p]\Big]\\ &= \frac{1}{2}n(n-1)p(1-p)\Big[1+(n-2)p\Big(12+(31n-60)p\\ &\quad +(28n^2-114n+105)p^2+(9n^3-58n^2+117n-69)p^3\Big)\Big]\\ &= \frac{1}{2}n^2p(1-p)[1+12np+31n^2p^2+28n^3p^3+9n^4p^4](1+O(n^{-1})).\end{aligned}$$

Covariances of Various Paths. It is now the goal to give the expressions of the covariances between the numbers of paths of various lengths in general. In K_n, let $C_n(k,l;i)$ denote the number

https://doi.org/10.1017/S0269964812000447 Published online by Cambridge University Press

of such paths of length k that have exactly i common edges with any given path of length l. It is easy to see that $C_n(k,l;i)$ is well defined. We claim that for $1 \le i \le \min(k,l)$ and $1 \le k, l \le n-1$,

$$C_n(k,l;i) = (k-i+1)(l-i+1)n^{k-i}(1+O(n^{-1})). \tag{17}$$

In fact, the main contribution to this quantity is the number of paths of length k, in which the i common edges (with a given path of length l) are connected, that is, the common subgraph itself is a path of length i. Note that each path of length $m \ge i$ contains exactly $m-i+1$ paths of length i. Consider to construct a path of length k which have exactly connected i common edges with a given path of length l. Then we should choose suitably $k-i$ more vertices from $n-(i+1)$ vertices (the common component has already $i+1$ vertices). Hence, there are $(k-i+1)(l-i+1)n^{k-i}(1+O(n^{-1}))$ ways to construct such paths of length k. If the i common edges are not connected, to construct such a path of length k we should choose at most $k-i-1$ more vertices. It is clear that there are at most $O(n^{k-i-1})$ ways.

Analogous to (5), for $1 \le k \le l \le n-1$, by (17) we have

$$\begin{aligned}\mathrm{Cov}(P_{k,n}, P_{l,n}) &= \frac{(l+1)!}{2}\binom{n}{l+1}\sum_{i=1}^{k} C_n(k,l;i)p^{k+l-i}(1-p^i)\\ &= \frac{n(1-p)}{2}\sum_{i=1}^{k}(k-i+1)(l-i+1)(np)^{k+l-i}(1+O(n^{-1}))\\ &= \frac{n(1-p)}{2}\sum_{j=1}^{k} j(l-k+j)(np)^{l+j-1}(1+O(n^{-1})).\end{aligned}$$

Acknowledgment

This work was partially supported by NSFC (No. 10801122 and No. 11101395), RFDP (No. 20093402120014) and the Fundamental Research Funds for the Central Universities.

References

1. Abdo, H., Dimitrov, D., & Gutman, I. (2012). On the Zagreb indices equality. *Discrete Applied Mathematics* 160: 1–8.
2. Bollobás, B. (2001). *Random graphs, 2nd Ed.* Cambridge: Cambridge University Press.
3. Feng, Q. & Hu, Z. (2011). On the Zagreb index of random recursive trees. *Journal of Applied Probability* 48: 1189–1196.
4. Gutman, I. & Trinajstić, N. (1972). Graph theory and molecular orbitals. Total φ-electron energy of alternant hydrocarbons. *Chemical Physics Letters* 17: 535–538.
5. Janson, S., Łuczak, T., & Ruciński, A. (2000). *Random graphs*. New York: Wiley-Interscience.
6. Mikhailov, V.G. (1991). On a theorem of Janson. *Teoriya Veroyatnostei i ee Primeneniya* 36, 168–170 (In Russian); an English translation appears in *Theory of Probability Application* 36: 173–176.
7. Nikiforov, V. (2007). The sum of the squares of degrees: sharp asymptotics. *Discrete Mathematics* 307, 3187–3193.
8. Nikolić, S., Kovačević, G., Miličević, A., & Trinajstić, N. (2003). The Zagreb indices 30 years after. *Croatica Chemica ACTA* 76, 113–124.
9. Peled, U.N., Petreschi, R., & Sterbini, A. (1999). (n,e)-graphs with maximum sum of squares of degrees. *Journal of Graph Theory* 31, 283–295.