

March 2022

The Zagreb Index of Several Random Models

Panpan Zhang

University of Pennsylvania, Philadelphia, PA 19104, USA, panpan.zhang@pennmedicine.upenn.edu

Follow this and additional works at: <https://digitalcommons.lsu.edu/josa>



Part of the [Analysis Commons](#), and the [Other Mathematics Commons](#)

Recommended Citation

Zhang, Panpan (2022) "The Zagreb Index of Several Random Models," *Journal of Stochastic Analysis*: Vol. 3 : No. 1 , Article 1.

DOI: 10.31390/josa.3.1.01

Available at: <https://digitalcommons.lsu.edu/josa/vol3/iss1/1>

THE ZAGREB INDEX OF SEVERAL RANDOM MODELS

PANPAN ZHANG*

ABSTRACT. In this article, we investigate the Zagreb index, a kind of graph-based topological index, of several random networks, including a class of networks extended from random recursive trees, plane-oriented recursive trees, and random caterpillars growing in a preferential attachment manner. We calculate the mean and variance of the Zagreb index for each class. In addition, we prove that the asymptotic distribution of the Zagreb index for the first class is normal, and that the asymptotic distribution of the Zagreb index for the second class is skewed to the right, hence not normal.

1. Introduction

A topological index, in chemical graph theory, is a metric that quantifies the structure of the molecular graph of a chemical compound via a number. The *Zagreb index* [7] is a topological index that has found a plethora of applications in mathematical chemistry and chemoinformatics. It is best known for modeling quantitative structure-property relationship (QSPR) and quantitative structure-activity relationship (QSAR) between molecules [12]. We refer the readers to [9] for a concise review.

The Zagreb index of a graph, $G = (V, E)$, is the sum of the squared degrees of all the nodes in G . Mathematically, it is given by

$$\text{Zagreb}(G) = \sum_{v \in V} (\deg(v))^2, \quad (1.1)$$

where $\deg(v)$ is the degree of node v .

Recently, the Zagreb index of several random trees were investigated, such as random recursive trees (RRTs) [5], b -ary search trees [6], spider trees [11] and caterpillar trees (uniform attachment) [14]. In this article, we investigate the Zagreb index of three random structures: a class of networks extended from RRTs, plane-oriented recursive trees (PORTs), and a class of caterpillars growing in a preferential attachment manner.

2. Zagreb Index of Extended RRTs

Tree is a popular structure for data storage and sorting in computer science. A rooted tree is a tree in which there is one designated node called *root*. The root

Received 2021-4-15; Accepted 2021-12-22; Communicated by the editors.

2020 *Mathematics Subject Classification*. Primary 05C80, 05C09; Secondary 60G42, 62E20.

Key words and phrases. Combinatorial probability, martingale, moments, random networks, recurrence methods, Zagreb index.

* Corresponding author.

of a tree is usually thought of as the originator of the tree. A *random recursive tree* (RRT) is a non-planer rooted tree such that a node is uniformly chosen from all the nodes in the existing tree as a parent for a new child at each growth step. The children of any parent in a RRT are not ordered.

The Zagreb index of RRTs was studied by [5]. The exact mean and variance of the index were calculated. They both increase linearly with respect to time n . The asymptotic distribution of the Zagreb index (scaled by n) was proven to follow a Gaussian law. In this section, we look into the Zagreb index of a class of networks extended from RRTs. This class of networks evolve as follows. At initial time $n = 1$, there is an undirected *complete graph* on $m_0 \geq 1$ nodes (i.e., each pair of nodes is connected by an edge). If there is a single node ($m_0 = 1$) at the initial point, it exists as an isolated node, where no self loop is considered. At each subsequent time point, we randomly choose $m \leq m_0$ (distinct) nodes from the existing network and connect them with a newcomer by m edges. At time n , we denote the structure of this class of networks by U_n . Our goal is to study the Zagreb index of U_n . A RRT appears as a special case of this network under the setting of $m_0 = m = 1$.

We consider the following enumeration for the nodes in U_n . Label the initial m_0 nodes with distinct numbers in $\{1, 2, \dots, m_0\}$. Before recruiting any child, these m_0 nodes are structurally equivalent, so the order of labeling is arbitrary. The child that joins the network at time $n \geq 2$ is labeled with $(m_0 + n - 1)$. Thus, there is a total of $(m_0 + n - 1)$ nodes in the network at time n . For each $j = 1, 2, \dots, m_0 + n - 1$, let $D_{n,j}$ be the degree of the node labeled with j . There are a couple of additional instantaneous facts that will be useful for establishing the recurrence equations subsequently.

Remark 2.1. According to the model definition, there are $\binom{m_0+n-1}{m}$ equally likely sets of sampled nodes connecting the newcomer joining U_{n+1} .

Remark 2.2. According to the model definition, the total degree of the nodes in the initial network is $m_0(m_0 - 1)$, and the contribution to the total degree at each evolutionary step is deterministic, i.e., $2m$.

Let

$$Z_n = \text{Zagreb}(U_n) = \sum_{j=1}^{m_0+n-1} D_{n,j}^2 \quad (2.1)$$

be the Zagreb index of U_n at time n . Note that we will repeatedly use D and Z with proper subscripts as node degrees and the Zagreb index for all kinds of random graphs that are investigated through this manuscript. In the next proposition, we calculate the expectation of Z_n , and develop a weak law as well.

Proposition 2.3. *For $n \geq 1$, the mean of the Zagreb index of U_n is*

$$\mathbb{E}[Z_n] = \frac{(5m^2 + m)n^2 - 2mm_0(2m - m_0 + 1)n \log n + O(n)}{n + m_0 - 1}. \quad (2.2)$$

As $n \rightarrow \infty$, we have

$$\frac{Z_n}{n} \xrightarrow{L_1} 5m^2 + m. \quad (2.3)$$

This convergence takes place in probability as well.

Proof. Let \mathbb{F}_{n-1} denote the σ -field generated by the history of the first $(n-1)$ stages of the network, and let $\mathbf{1}_{n,S}$ denote the event indicating that the nodes labeled with the indices in set S are chosen as parents for the new child at time n . Upon the insertion of node n , an almost-sure relation between Z_n and Z_{n-1} , conditional on \mathbb{F}_{n-1} and $\mathbf{1}_{n,S}$ is given by

$$Z_n = Z_{n-1} + \sum_{j=s_1}^{s_m} (D_{n-1,j} + 1)^2 - \sum_{j=s_1}^{s_m} D_{n-1,j}^2 + m^2, \quad (2.4)$$

where $S := \{s_1, s_2, \dots, s_m\}$ is an m -long subset of $\{1, 2, \dots, m_0 + n - 2\}$. We simplify the almost-sure relation (c.f. Equation (2.4)) to get

$$Z_n = Z_{n-1} + 2 \sum_{j \in S} D_{n-1,j} + (m^2 + m). \quad (2.5)$$

Taking the average over all possible S 's, we obtain

$$\begin{aligned} \mathbb{E}[Z_n | \mathbb{F}_{n-1}] &= Z_{n-1} + \frac{2}{\binom{n+m_0-2}{m}} \sum_S \sum_{j \in S} D_{n-1,j} + (m^2 + m) \\ &= Z_{n-1} + 2 \times \frac{\binom{n+m_0-3}{m-1}}{\binom{n+m_0-2}{m}} \sum_{j=1}^{n+m_0-2} D_{n-1,j} + (m^2 + m) \\ &= Z_{n-1} + \frac{2m}{n+m_0-2} \sum_{j=1}^{n+m_0-2} D_{n-1,j} + (m^2 + m), \end{aligned} \quad (2.6)$$

where the sum $\sum_{j=1}^{n+m_0-2} D_{n-1,j}$ is not random; it is equal to

$$(m_0(m_0 - 1) + 2m(n - 2)).$$

We thus can take another expectation with respect to \mathbb{F}_{n-1} to get a recurrence for $\mathbb{E}[Z_n]$, which is given by

$$\begin{aligned} \mathbb{E}[Z_n] &= \mathbb{E}[Z_{n-1}] + \frac{m((5m+1)n + 2m_0^2 + (m-1)m_0 - 2(5m+1))}{n+m_0-2} \\ &= \mathbb{E}[Z_1] + \sum_{i=2}^n \frac{m((5m+1)i + 2m_0^2 + (m-1)m_0 - 2(5m+1))}{i+m_0-2} \\ &= \mathbb{E}[Z_1] + \sum_{i=2}^n \frac{m((5m+1)(2-m_0) + 2m_0^2 + (m-1)m_0 - 2(5m+1))}{i+m_0-2} \\ &\quad + \sum_{i=2}^n m(5m+1), \end{aligned} \quad (2.7)$$

where, in the last display, the second term is a harmonic number, and the summand of the third term does not contain i . We solve this recurrence with initial condition $\mathbb{E}[Z_1] = Z_1 = m_0(m_0 - 1)^2$, and get the result stated in the proposition.

In what follows, we have

$$|(n+m_0-1)Z_n - (5m^2+m)n^2| = O_{L_1}(n \log n), \quad (2.8)$$

where the big O notation $O_{L_1}(\cdot)$ is defined as follows: Given a real function $f(\cdot)$, we say a sequence of random variables $X_n = O_{L_1}(f(n))$ if and only if there exists a positive constant M and a positive integer n_0 such that $\mathbb{E}[|X_n|] \leq Mf(n)$ for all $n \geq n_0$. Divide by n^2 on both sides of Equation (2.8), and let n go to infinity to obtain an L_1 convergence for Z_n/n , which also implies an in-probability convergence. \square

The computation of the second moment of Z_n is based on squaring the almost-sure relation of Z_n presented in Equation (2.5). That is

$$\begin{aligned} Z_n^2 &= Z_{n-1}^2 + 4 \left(\sum_{j \in S} D_{n-1,j} \right)^2 + (m^2 + m)^2 + 4Z_{n-1} \sum_{j \in S} D_{n-1,j} \\ &\quad + 2(m^2 + m)Z_{n-1} + 4(m^2 + m) \sum_{j \in S} D_{n-1,j}. \end{aligned} \quad (2.9)$$

Similar to the proof of Proposition 2.3, we tend to take the expectation with respect to $\mathbf{1}_{n,S}$, then to take another expectation with respect to \mathbb{F}_{n-1} , and ultimately to get a recurrence for the second moment of Z_n . Before implementing this strategy, we take the most complex term in Equation (2.9) out and simplify it separately as follows:

$$\sum_S \left(\sum_{j \in S} D_{n-1,j} \right)^2 = \sum_S \left(\sum_{j \in S} D_{n-1,j}^2 + 2 \sum_{j \neq k \in S} D_{n-1,j} D_{n-1,k} \right). \quad (2.10)$$

The first part is simple. It is

$$\sum_S \sum_{j \in S} D_{n-1,j}^2 = \binom{n+m_0-3}{m-1} Z_{n-1}. \quad (2.11)$$

The second part is

$$\begin{aligned} &2 \sum_S \sum_{j \neq k \in S} D_{n-1,j} D_{n-1,k} \\ &= 2 \binom{n+m_0-4}{m-2} \sum_{j \neq k} D_{n-1,j} D_{n-1,k} \\ &= \binom{n+m_0-4}{m-2} \sum_{j=1}^{n+m_0-2} D_{n-1,j} \left(\sum_{j=1}^{n+m_0-2} D_{n-1,j} - D_{n-1,j} \right) \\ &= \binom{n+m_0-4}{m-2} \sum_{j=1}^{n+m_0-2} D_{n-1,j} (m_0(m_0-1) + 2m(n-2) - D_{n-1,j}) \\ &= \binom{n+m_0-4}{m-2} \left((m_0(m_0-1) + 2m(n-2))^2 - Z_{n-1} \right). \end{aligned} \quad (2.12)$$

We are now ready to derive the second moment of Z_n , the result of which is presented in the next proposition.

Proposition 2.4. *For $n \geq 1$, the second moment of the Zagreb index of U_n is*

$$\mathbb{E}[Z_n^2] = \frac{(5m^2 + m)^2 n^5 - 4m^2 m_0 (5m + 1)(2m - m_0 + 1)n^4 \log n + O(n^4)}{(n + m_0 - 2)(n + m_0 - 1)^2}.$$

Proof. Recall the squared almost-sure relation in Equation (2.9). Take the expectation with respect to $\mathbf{1}_{n,S}$, and plug in the expressions of $\sum_S (\sum_{j \in S} D_{n-1,j})^2$ and $\sum_S \sum_{j \in S} D_{n-1,j}$ (in the proof of Proposition 2.3) to get

$$\mathbb{E}[Z_n^2 | \mathbb{F}_{n-1}] = Z_{n-1}^2 + C_1(n, m, m_0)Z_{n-1} + C_2(n, m, m_0), \quad (2.13)$$

where C_1 and C_2 are two constant functions (free of Z_n) depending only on n , m and m_0 . We have the exact expressions of C_1 and C_2 , but they are too lengthy to report in the manuscript. To give the readers some general idea about C_1 and C_2 , we present the leading term for each:

$$C_1(n, m, m_0) = 5m + 1 + O\left(\frac{1}{n}\right), \quad (2.14)$$

$$C_2(n, m, m_0) = m^2(25m^2 - 6m + 1) + O\left(\frac{1}{n}\right), \quad (2.15)$$

Taking another expectation on both sides of Equation (2.13) with respect to \mathbb{F}_{n-1} and plugging in $\mathbb{E}[Z_n]$ derived in Proposition 2.3, we obtain a recurrence for the second moment of Z_n . Solving the recurrence with initial condition $\mathbb{E}[Z_1^2] = Z_1^2 = m_0^2(m_0 - 1)^4$, we get the result stated in the proposition. \square

Although we only present the top two highest-order terms of $\mathbb{E}[Z_n^2]$ in Proposition 2.4, we obtain the exact expressions of a few more terms in the derivation. These terms are needed to determine the order of the leading term of the variance of Z_n . These terms are available upon request by the readers. In the next corollary, we give the variance of Z_n , computed by taking the difference between the second moment and the square of the first moment of Z_n .

Corollary 2.5. *For $n \geq 1$, the variance of the Zagreb index of U_n is*

$$\text{Var}[Z_n] = 4m^2(m + 1)n + 4m^2 m_0(m_0 - 2m - 1) \log^2 n + O(\log n).$$

In [5], the authors proved that the variance of the Zagreb index of RRT is asymptotically equal to $8n$, which is the special case of Corollary 2.5 ($m = 1$). According to Corollary 2.5, we find that the variance of the Zagreb index of U_n is linear in n , and its asymptotic value does not depend on m_0 . In the next corollary, we show that Z_n/n converges to $5m^2 + m$ in L_2 -space (stronger than the L_1 convergence and the in-probability convergence presented in Proposition 2.3).

Corollary 2.6. *As $n \rightarrow \infty$, we have*

$$\frac{Z_n}{n} \xrightarrow{L_2} 5m^2 + m.$$

Proof. According to the asymptotic mean and variance of Z_n , we have

$$\begin{aligned} \mathbb{E} \left[|Z_n - (5m^2 + m)n|^2 \right] &= \mathbb{E} \left[(Z_n - \mathbb{E}[Z_n] + \mathbb{E}[Z_n] - (5m^2 + m)n)^2 \right] \\ &\leq \text{Var}[Z_n] + O(\log^2 n) \\ &= O(n), \end{aligned} \tag{2.16}$$

which completes the proof. \square

The next corollary follows Corollary 2.6 immediately by applying the *Cauchy-Schwarz inequality*, and it is helpful in deriving the limit law.

Corollary 2.7. *We have*

$$Z_n = (5m^2 + m)n + O_{L_1}(\sqrt{n}).$$

As both the mean and the variance of Z_n are linear in n , we suspect that the limit distribution of Z_n scaled by n is normal for general U_n , not just for the class of RRTs [5]. To prove the conjecture, our strategy is to apply a *Martingale Central Limit Theorem* (MCLT). According to Equation (2.7), $\{Z_n\}_n$ is not a martingale. We consider the following transformation such that the transformed array is a martingale.

Lemma 2.8. *For $n \geq 1$, the array*

$$M_n = Z_n - C_3(n, m, m_0),$$

where $C_3(n, m, m_0)$ is a function of n , m and m_0 , is a martingale.

Proof. Given a sequence $\{\beta_n\}_n$, consider $M_n := Z_n + \beta_n$ such that $\{M_n\}_n$ is a martingale. We retrieve β_n based off the fundamental martingale property, i.e.,

$$\begin{aligned} \mathbb{E}[Z_n + \beta_n | \mathbb{F}_{n-1}] &= Z_{n-1} + \frac{(5m+1)n + 2m_0^2 + (m-1)m_0 - 2(5m+1)}{n + m_0 - 2} \\ &\quad + \beta_n \\ &= Z_{n-1} + \beta_{n-1} \end{aligned} \tag{2.17}$$

We thus obtain a recurrence for β_n . We solve the recurrence with an arbitrary choice of initial value of β_n , e.g., $\beta_1 = 0$, to get the exact expression of $C_3(n, m, m_0)$. The solution is

$$C_3(n, m, m_0) = \frac{(5m+1)n^2 + O(n \log n)}{n + m_0 - 1}. \tag{2.18}$$

\square

Though we only give the leading term of $C_3(n, m, m_0)$ in Lemma 2.8, we use the exact solution of $C_3(n, m, m_0)$ in the follow-up calculation. There are different forms of MCLTs listed in [8], based on different sets of conditions. We choose a MCLT [8, Corollary 3.2] based on martingale differences. Consider a backward difference operator ∇ such that $\nabla M_j := M_j - M_{j-1}$. The MCLT that we exploit for the proof requires a *conditional Lindeberg's condition* and a *conditional variance condition*.

Lemma 2.9. *The conditional Lindeberg's condition is given by*

$$W_n := \sum_{j=1}^n \mathbb{E} \left[\left(\frac{\nabla M_j}{\sqrt{n}} \right)^2 \mathbf{1}_{\{|\nabla M_j/\sqrt{n}| > \varepsilon\}} \middle| \mathbb{F}_{j-1} \right] \xrightarrow{P} 0.$$

Proof. By the construction of the martingale, we have

$$\begin{aligned} |\nabla M_j| &= |Z_j + \beta_j - (Z_{j-1} + \beta_{j-1})| \\ &\leq |Z_j - Z_{j-1}| + \frac{(5m+1)j + 2m_0^2 + (m-1)m_0 - 2(5m+1)}{j + m_0 - 2} \\ &\leq 2 \max_S \sum_{i \in S} D_{j-1,i} + m^2 + m + (5m+1) + O\left(\frac{1}{j}\right) \\ &= O(\log_2 j) + (5m+1) + O\left(\frac{1}{j}\right) \end{aligned} \quad (2.19)$$

for large j , where the bound for the maximal degree of a node is obtained according to [2, Theorem 1]. Therefore, $|\nabla M_j/\sqrt{n}|$ is uniformly bounded for all j . In other words, for any $\varepsilon > 0$, there exists $n_0(\varepsilon) > 0$ such that the sets $\{|\nabla M_j/\sqrt{n}| > \varepsilon\}$ are empty for all $n > n_0(\varepsilon)$. In what follows, we conclude that W_n converges to 0 almost surely, which is stronger than the in-probability convergence required for the Lindeberg's condition. \square

Lemma 2.10. *The conditional variance condition is given by*

$$V_n := \sum_{j=1}^n \mathbb{E} \left[\left(\frac{\nabla M_j}{\sqrt{n}} \right)^2 \middle| \mathbb{F}_{j-1} \right] \xrightarrow{P} \eta^2,$$

where η^2 is a random variable that is either finite or converges almost surely. Particularly for our case, η^2 is equal to $4m^3 + 4m^2$.

Proof. We rewrite V_n as follows:

$$\begin{aligned} V_n &= \frac{1}{n} \sum_{j=1}^n \mathbb{E} \left[(Z_j + \beta_j - (Z_{j-1} + \beta_{j-1}))^2 \middle| \mathbb{F}_{j-1} \right] \\ &= \frac{1}{n} \sum_{j=1}^n \left(\mathbb{E} [(Z_j - Z_{j-1})^2 \middle| \mathbb{F}_{j-1}] + 2 \mathbb{E} [(Z_j - Z_{j-1})(\beta_j - \beta_{j-1}) \middle| \mathbb{F}_{j-1}] \right. \\ &\quad \left. + \mathbb{E} [(\beta_j - \beta_{j-1})^2 \middle| \mathbb{F}_{j-1}] \right). \end{aligned} \quad (2.20)$$

We calculate the three expectations in the summand one after another. The first part is given by

$$\begin{aligned} \mathbb{E} [(Z_j - Z_{j-1})^2 \middle| \mathbb{F}_{j-1}] &= \mathbb{E} [Z_j^2 \middle| \mathbb{F}_{j-1}] + Z_{j-1}^2 - 2Z_{j-1} \mathbb{E}[Z_j \middle| \mathbb{F}_{j-1}] \\ &= Z_{j-1}^2 + C_1(j, m, m_0)Z_{j-1} + C_2(j, m, m_0) + Z_{j-1}^2 \\ &\quad - 2Z_{j-1}(Z_{j-1} + C_4(j, m, m_0)), \end{aligned} \quad (2.21)$$

where $C_1(j, m, m_0)$ and $C_2(j, m, m_0)$ are identical to those given in Equation (2.13) and $C_4(j, m, m_0) = m((5m+1)j + 2m_0^2 + (m-1)m_0 - 2(5m+1))/(j + m_0 - 2)$.

The second part is

$$\begin{aligned} 2 \mathbb{E}[(Z_j - Z_{j-1})(\beta_j - \beta_{j-1}) \mid \mathbb{F}_{j-1}] &= (\beta_j - \beta_{j-1})(\mathbb{E}[Z_j \mid \mathbb{F}_{j-1}] - Z_{j-1}) \\ &= -2 C_4(j, m, m_0)^2. \end{aligned} \quad (2.22)$$

The third part is

$$\mathbb{E}[(\beta_j - \beta_{j-1})^2 \mid \mathbb{F}_{j-1}] = C_4(j, m, m_0)^2. \quad (2.23)$$

By Corollary 2.7, we have

$$\begin{aligned} V_n &= \frac{1}{n} \sum_{j=1}^n ((25m^4 + 14m^3 + 5m^2) - 2m^2(5m+1)^2 + m^2(5m+1)^2 \\ &\quad + O(\log j/j)) \\ &= 4m^3 + 4m^2 + O_{L_1}(\log^2 n/n) \\ &\xrightarrow{L_1} 4m^3 + 4m^2, \end{aligned} \quad (2.24)$$

which is stronger than in-probability convergence required for the conditional variance condition. \square

Theorem 2.11. *As $n \rightarrow \infty$, we have*

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

Proof. Upon the verifications of the two conditions in Lemmata 2.9 and 2.10, we have

$$\frac{Z_n + \beta_n - (Z_1 + \beta_1)}{\sqrt{n}} \sim \frac{Z_n - (5m^2 + m)n}{\sqrt{n}} \xrightarrow{D} \mathcal{N}(0, 4m^3 + 4m^2), \quad (2.25)$$

by the MCLT [8, Corollary 3.2]. This is equivalent to the stated result in the theorem. \square

3. Zagreb Index of PORTs

In contrast to a RRT, a *plane-oriented recursive tree* (PORT) accounts for orders in tree evolution. One simple way to interpret this process is that the probability that a node is chosen as a parent for a new child is proportional to its *degree* in the current tree. Mathematically, it is given by

$$\mathbb{P}(\mathbf{1}_v) = \frac{\deg(v)}{\sum_{u \in V} \deg(u)},$$

where $\mathbf{1}_v$ indicates the event that node v is chosen as a parent for the newcomer, and V is the set of all nodes in the current tree. Therefore, PORTs are a class of nonuniform trees. As its evolutionary process coincides with an attractive network characteristic—*preferential attachment* [1], PORTs are of substantial interest in the community.

The Zagreb index of PORTs was investigated in a recent article [13], where the exact mean and variance were determined. The author claimed that the Zagreb index of PORTs does not follow a Gaussian law as time goes to infinity via numeric

experiments. In this paper, we provide a more rigorous proof in support of the conjecture.

Let T_n be a PORT at time n . As one node joins the tree at each step, there is a total of n nodes in T_n . We label these n nodes with $\{1, 2, \dots, n\}$ according to the time points of their appearance in the tree. Let $D_{n,j}$ be the degree of the node labeled with j at time n . The Zagreb index of T_n is given by

$$Z_n = \text{Zagreb}(T_n) = \sum_{j=1}^n D_{n,j}^2. \quad (3.1)$$

Proposition 3.1 ([13]). *For $n \geq 1$, we have*

$$\begin{aligned} \mathbb{E}[Z_n] &= 2(n-1)(\Psi(n) + \gamma) \\ \mathbb{E}[Y_n] &= \frac{32\Gamma(n+1/2)}{\sqrt{\pi}\Gamma(n-1)} - 6(n-1) \left(\Psi(n) + \gamma + \frac{8}{3} \right), \\ \mathbb{E}[Z_n^2] &= 4(n \log n)^2 + 8\gamma(n^2 \log n) + \left(16 + 4\gamma^2 - \frac{2\pi^2}{3} \right) n^2 + O(n^{3/2}), \end{aligned}$$

where $\Psi(\cdot)$ is the standard digamma function, γ is the Euler's constant, and $Y_n := \sum_{j=1}^n D_{n,j}^3$ is a topological index summing the cubic degrees of all the nodes.

Based off the simulation results in [13], we suspect that the asymptotic distribution of Z_n is skewed to the right, violating the property of symmetry of normal distribution. Therefore, it suffices to show that the skewness of Z_n is not zero; in fact, it is always positive.

In probability theory, the *skewness* of a random variable refers to its standardized third central moment, i.e.,

$$\mathcal{S}(Z_n) = \mathbb{E} \left[\left(\frac{Z_n - \mu_{Z_n}}{\sigma_{Z_n}} \right)^3 \right] = \frac{\mathbb{E}[Z_n^3] - 3\mathbb{E}[Z_n]\mathbb{E}[Z_n^2] + 2(\mathbb{E}[Z_n])^3}{(\text{Var}[Z_n])^{3/2}}, \quad (3.2)$$

in which most of the elements have already been determined in Proposition 3.1, except for the third moment of Z_n . We resort to a recurrence method to calculate $\mathbb{E}[Z_n^3]$ exactly. To construct a recurrence for $\mathbb{E}[Z_n^3]$, we need the results presented in the next two lemmata. The first lemma is based on a new topological index, X_n , the sum of the degrees to the fourth power of nodes in T_n , i.e., $X_n := \sum_{j=1}^n D_{n,j}^4$.

Lemma 3.2. *For $n \geq 2$, we have*

$$\mathbb{E}[X_n] = 36(n-1) \left(n + \frac{7\Psi(n)}{18} + \frac{7\gamma}{18} + \frac{5}{3} \right) - \frac{192\Gamma(n+1/2)}{\sqrt{\pi}\Gamma(n-1)}.$$

As $n \rightarrow \infty$, we have

$$\frac{X_n}{n^2} \xrightarrow{L_1} 36.$$

The convergence takes place in probability as well.

Proof. Upon time n (node n is not yet inserted), there is a total of $(n-1)$ nodes in the current tree. In addition, the total of node degrees is $2(n-2)$. By the definition of X_n , we have the following almost-sure relation from X_{n-1} (right before the insertion of node n) to X_n (right after the insertion of node n), conditional on

\mathbb{F}_{n-1} and $\mathbf{1}_{n,j}$, the event indicating that the node labeled with j is chosen as a parent for the newcomer:

$$\begin{aligned} X_n &= X_{n-1} + (D_{n-1,j} + 1)^4 - D_{n-1,j}^4 + 1 \\ &= X_{n-1} + 4D_{n-1,j}^3 + 6D_{n-1,j}^2 + 4D_{n-1,j} + 2. \end{aligned} \quad (3.3)$$

We average Equation (3.3) out over j to get

$$\begin{aligned} \mathbb{E}[X_n | \mathbb{F}_{n-1}] &= X_{n-1} + 4 \sum_{j=1}^{n-1} \frac{D_{n-1}^4}{2(n-2)} + 6 \sum_{j=1}^{n-1} \frac{D_{n-1}^3}{2(n-2)} \\ &\quad + 4 \sum_{j=1}^{n-1} \frac{D_{n-1}^2}{2(n-2)} + 2 \\ &= \frac{n}{n-2} X_{n-1} + \frac{3}{n-2} Y_{n-1} + \frac{2}{n-2} Z_n + 2. \end{aligned} \quad (3.4)$$

Taking another expectation (with respect to \mathbb{F}_{n-1}) and plugging in the results of $\mathbb{E}[Y_n]$ and $\mathbb{E}[Z_n]$, we obtain a recurrence for $\mathbb{E}[X_n]$. Solving the recurrence with initial condition $\mathbb{E}[X_2] = X_2 = 2$, we obtain the result stated in the lemma.

Note that the digamma function $\Psi(n)$ is asymptotically equivalent to $\log n$. Meanwhile, the second fraction is of order $n^{3/2}$ according to the *Stirling's approximation*. Thus, we have X_n/n^2 converges to 36 in L_1 -space, which also suggests a weak law for X_n . \square

In the second lemma, we derive the mixed moment of Z_n and Y_n ; namely, $\mathbb{E}[Z_n Y_n]$. Apparently, the variables Z_n and Y_n are not independent. Our strategy is to establish a recurrence on the expectation of the product of $Z_n Y_n$.

Lemma 3.3. *For $n \geq 2$, we have*

$$\mathbb{E}[Z_n Y_n] = \frac{64(\log n + O(1))\Gamma(n + 3/2) + 32/15}{\sqrt{\pi}\Gamma(n-1)}.$$

As $n \rightarrow \infty$, we have

$$\frac{Z_n Y_n}{n^{5/2} \log n} \xrightarrow{L_1} \frac{64}{\sqrt{\pi}}.$$

Proof. By the definition of Z_n (Y_n), we have the following almost-sure relations from Z_{n-1} (Y_{n-1}) to Z_n (Y_n), conditional on \mathbb{F}_{n-1} and $\mathbf{1}_{n,j}$:

$$Z_n = Z_{n-1} + (D_{n-1,j} + 1)^2 - D_{n-1,j}^2 + 1 = Z_{n-1} + 2D_{n-1,j} + 2. \quad (3.5)$$

$$\begin{aligned} Y_n &= Y_{n-1} + (D_{n-1,j} + 1)^3 - D_{n-1,j}^3 + 1 \\ &= Y_{n-1} + 3D_{n-1,j}^2 + 3D_{n-1,j} + 2. \end{aligned} \quad (3.6)$$

Taking the product of Equations (3.5) and (3.6), we get

$$\begin{aligned} Z_n Y_n &= Z_{n-1} Y_{n-1} + 3Z_{n-1} D_{n-1,j}^2 + 3Z_{n-1} D_{n-1,j} + 2Z_{n-1} \\ &\quad + 2Y_{n-1} D_{n-1,j} + 6D_{n-1,j}^3 + 6D_{n-1,j}^2 + 4D_{n-1,j} \\ &\quad + 2Y_{n-1} + 6D_{n-1,j}^2 + 6D_{n-1,j} + 4. \end{aligned} \quad (3.7)$$

Averaging it out over j , we then have

$$\begin{aligned}
\mathbb{E}[Z_n Y_n | \mathbb{F}_{n-1}] &= Z_{n-1} Y_{n-1} + 3Z_{n-1} \sum_{j=1}^{n-1} \frac{D_{n-1,j}^3}{2(n-2)} + 3Z_{n-1} \sum_{j=1}^{n-1} \frac{D_{n-1,j}^2}{2(n-2)} \\
&\quad + 2Z_{n-1} + 2Y_{n-1} \sum_{j=1}^{n-1} \frac{D_{n-1,j}^2}{2(n-2)} + 6 \sum_{j=1}^{n-1} \frac{D_{n-1,j}^4}{2(n-2)} \\
&\quad + 6 \sum_{j=1}^{n-1} \frac{D_{n-1,j}^3}{2(n-2)} + 4 \sum_{j=1}^{n-1} \frac{D_{n-1,j}^2}{2(n-2)} + 2Y_{n-1} \\
&\quad + 6 \sum_{j=1}^{n-1} \frac{D_{n-1,j}^3}{2(n-2)} + 6 \sum_{j=1}^{n-1} \frac{D_{n-1,j}^2}{2(n-2)} + 4 \\
&= \frac{2n+1}{2(n-2)} Z_{n-1} Y_{n-1} + \frac{3}{2(n-2)} Z_{n-1}^2 + \frac{3}{n-2} X_{n-1} \\
&\quad + \frac{2(n+1)}{n-2} Y_{n-1} + \frac{2n+1}{n-2} Z_{n-1} + 4. \tag{3.8}
\end{aligned}$$

The recurrence for $\mathbb{E}[Z_n Y_n]$ is then obtained by taking another expectation with respect to \mathbb{F}_{n-1} and plugging in the results of $\mathbb{E}[Z_{n-1}^2]$, $\mathbb{E}[X_{n-1}]$, $\mathbb{E}[Y_{n-1}]$, and $\mathbb{E}[Z_{n-1}]$. Solving the recurrence with initial condition $\mathbb{E}[Z_2 Y_2] = Z_2 Y_2 = 2 \times 2 = 4$, we obtain the solution of $\mathbb{E}[Z_n Y_n]$. The L_1 convergence follows by applying the Stirling's approximation. \square

The exact solution for $\mathbb{E}[Z_n Y_n]$ is derived. However, it cannot be written in a closed form. Instead, it is the sum of four fraction terms involving gamma functions, digamma functions, and first order polygamma functions. We thus only present the leading term of the solution in Lemma 3.3 for better readability. The exact solution is available upon request.

We are now ready to derive the third moment of Z_n by utilizing the results from Lemmata 3.2 and 3.3 and Proposition 3.1.

Proposition 3.4. *For $n \geq 2$, we have*

$$\mathbb{E}[Z_n^3] = (n+1)n(n-1) (8 \log^3 n + 24 \log^2 n + O(\log n)).$$

As $n \rightarrow \infty$, we have

$$\frac{Z_n^3}{(n \log n)^3} \xrightarrow{L_1} 8.$$

Proof. Recall the almost-sure relation for Z_n . Raising Equation (3.5) to the third power on both sides, we have

$$\begin{aligned}
Z_n^3 &= Z_{n-1}^3 + 8D_{n-1,j}^3 + 8 + 6Z_{n-1}^2 D_{n-1,j} + 6Z_{n-1}^2 + 12D_{n-1,j}^2 Z_{n-1} \\
&\quad + 24D_{n-1,j}^2 + 12Z_{n-1} + 24D_{n-1} + 24Z_{n-1} D_{n-1,j}. \tag{3.9}
\end{aligned}$$

Average it out over j to get

$$\begin{aligned}
\mathbb{E}[Z_n^3 | \mathbb{F}_{n-1}] &= Z_{n-1}^3 + 8 \sum_{j=1}^{n-1} \frac{D_{n-1,j}^4}{2(n-2)} + 8 + 6Z_{n-1}^2 \sum_{j=1}^{n-1} \frac{D_{n-1,j}^2}{2(n-2)} + 6Z_{n-1}^2 \\
&\quad + 12Z_{n-1} \sum_{j=1}^{n-1} \frac{D_{n-1,j}^3}{2(n-2)} + 24 \sum_{j=1}^{n-1} \frac{D_{n-1,j}^3}{2(n-2)} + 12Z_{n-1} \\
&\quad + 24 \sum_{j=1}^{n-1} \frac{D_{n-1,j}^2}{2(n-2)} + 24Z_{n-1} \sum_{j=1}^{n-1} \frac{D_{n-1,j}^2}{2(n-2)} \\
&= \frac{n+1}{n-2} Z_{n-1}^3 + \frac{6n}{n-2} Z_{n-1}^2 + \frac{6}{n-2} Z_{n-1} Y_{n-1} \\
&\quad + \frac{12(n-1)}{n-2} Z_{n-1} + \frac{4}{n-2} X_n + \frac{12}{n-2} Y_{n-1} + 8. \tag{3.10}
\end{aligned}$$

Taking expectation on both sides, and plugging in all the results of lower moments, we obtain a recurrence for the third moment of Z_n . We solve the recurrence with initial condition $\mathbb{E}[Z_2^3] = Z_2^3 = 8$ to get the stated result in the proposition, and the L_1 convergence of Z_n^3 after it is properly scaled. \square

Similar to $\mathbb{E}[Z_n Y_n]$, we get the exact solution for $\mathbb{E}[Z_n^3]$. However, the solution is even more complex than that for $\mathbb{E}[Z_n Y_n]$, involving digamma functions, *Meijer G functions* and nested sums which can not be simplified to closed forms. However, the leading terms that we have developed are sufficient to characterize the asymptotic behavior of the skewness of Z_n .

Theorem 3.5. *As $n \rightarrow \infty$, the distribution of Z_n is skewed to the right. Hence, it is not normal.*

Proof. Recall the definition formula for $\mathcal{S}(Z_n)$:

$$\mathcal{S}(Z_n) = \frac{\mathbb{E}[Z_n^3] - 3\mathbb{E}[Z_n]\mathbb{E}[Z_n^2] + 2(\mathbb{E}[Z_n])^3}{(\text{Var}[Z_n])^{3/2}}. \tag{3.11}$$

Plugging in $\mathbb{E}[Z^3]$, $\mathbb{E}[Z^2]$ and $\mathbb{E}[Z]$, we find that the top three leading terms (of order $n^3 \log^3 n$, $n^3 \log^2 n$ and $n^3 \log n$) in the numerator are exactly canceled out, left with the highest nonzero term of order n^3 . This is the same as the order of the leading term in the denominator. We thus arrive at

$$\mathcal{S}(Z_n) \sim \frac{-72\gamma^2 + 164\gamma - 41/3 + 4\pi^2\gamma - 8\gamma^3}{(16 - 2\pi^2/3)^{3/2}} \approx 2.7, \tag{3.12}$$

as $n \rightarrow \infty$, and conclude that Z_n does not converge to normal asymptotically. \square

4. Zagreb Index of Caterpillars

In mathematical chemistry, *caterpillar* is a popular model for representing the structure of benzoid hydrocarbon molecules [3, 4]. In this section, we look into a class of random caterpillars by incorporating caterpillars with randomness. The class of random caterpillars considered grow in a preferential attachment manner, as described in Section 3. More precisely, at time 0, there is a *spine* consisting

of $m \geq 2$ (fixed) nodes, which are labeled with distinct numbers in $\{1, 2, \dots, m\}$ from one end to the other. At each subsequent point, a *leaf* is connected to one of the spine nodes by an edge, the probability being equal to its degree over the total degree of all spine nodes.

At time n , we denote the structure of a random caterpillar by C_n . We first give some graph invariants of C_n . The total number of nodes in C_n is $(n + m)$, and the total degree (of all spine nodes) is $(n + 2m - 2)$. Let $X_{n,j}$ be the number of leaves attached to spine node j , and let $D_{n,j}$ be the degree of spine node j . There is an instantaneous relation between $X_{n,j}$ and $D_{n,j}$. That is $D_{n,j} = X_{n,j} + 1$ for $j = 1, m$; $D_{n,j} = X_{n,j} + 2$ for $j = 2, 3, \dots, m - 1$. According to the evolution of C_n , the probability that the spine node j is selected for recruiting a leaf at time n is

$$\frac{D_{n-1,j}}{\sum_{j=1}^m D_{n-1,j}}.$$

The Zagreb index of C_n is given by

$$Z_n = \text{Zagreb}(C_n) = \sum_{j=1}^m D_{n,j}^2 + n. \quad (4.1)$$

Note that we only account for caterpillars of $m \geq 2$ in this section, as the Zagreb index of a caterpillar of $m = 1$ (*star*) is deterministic; It is $(n^2 + n)$. In the next proposition, we derive the expectation of Z_n , and develop a weak law as well.

Proposition 4.1. *For $n \geq 1$, the mean of the Zagreb index of caterpillars is*

$$\mathbb{E}[Z_n] = \frac{(3m - 4)n^2 + (4m - 3)(3m - 4)n}{(2m - 1)(m - 1)} + 2(2m - 3).$$

As $n \rightarrow \infty$, we have

$$\frac{Z_n}{n^2} \xrightarrow{L_1} \frac{3m - 4}{(2m - 1)(m - 1)}.$$

This convergence takes place in probability as well.

Proof. We start with an almost-sure relation between Z_n and Z_{n-1} , conditional on \mathbb{F}_{n-1} and $\mathbf{1}_{n,j}$; that is,

$$Z_n = Z_{n-1} + (D_{n-1,j} + 1)^2 - D_{n-1,j}^2 + 1 = Z_{n-1} + 2D_{n-1,j} + 2. \quad (4.2)$$

This is identical to the almost-sure relation in Equation (3.5). Taking the expectation with respect to $\mathbf{1}_{n,j}$, we get

$$\begin{aligned} \mathbb{E}[Z_n | \mathbb{F}_{n-1}] &= Z_{n-1} + 2 \sum_{j=1}^m D_{n-1,j} \frac{D_{n-1,j}}{\sum_{j=1}^m D_{n-1,j}} + 2 \\ &= Z_{n-1} + \frac{2}{n + 2m - 3} \sum_{j=1}^m D_{n-1,j}^2 + 2 \\ &= Z_{n-1} + \frac{2}{n + 2m - 3} (Z_{n-1} - (n - 1)) + 2 \\ &= \frac{n + 2m - 1}{n + 2m - 3} Z_{n-1} + \frac{4(m - 1)}{n + 2m - 3}. \end{aligned} \quad (4.3)$$

Taking another expectation with respect to \mathbb{F}_{n-1} , we obtain a recurrence for $\mathbb{E}[Z_n]$:

$$\mathbb{E}[Z_n] = \frac{n+2m-1}{n+2m-3} \mathbb{E}[Z_{n-1}] + \frac{4(m-1)}{n+2m-3}. \quad (4.4)$$

We solve the recurrence with initial condition $\mathbb{E}[Z_0] = Z_0 = 4m - 6$, to get the result stated in the proposition. Both L_1 convergence and in-probability convergence of Z_n/n^2 are obtained immediately. \square

Towards the second moment of Z_n , we need the mean of a new topological index—the total of cubic degrees of all nodes in C_n , denoted by Y_n :

$$Y_n := \sum_{j=1}^m D_{n,j}^3 + n.$$

The mean of Y_n is given in the next lemma.

Lemma 4.2. *For $n \geq 1$, the mean of Y_n is given by*

$$\begin{aligned} \mathbb{E}[Y_n] &= \frac{3(2m-3)n^3 + (27m^2 - 60m + 27)n^2}{(2m-1)m(m-1)} \\ &\quad + \frac{(40m^3 - 111m^2 + 86m - 18)n}{(2m-1)m(m-1)} + 2(4m-7). \end{aligned}$$

Proof. We consider an almost-sure relation between Y_n and Y_{n-1} analogous to Equation (3.6):

$$Y_n = Y_{n-1} + 3D_{n-1,j}^2 + 3D_{n-1,j} + 2. \quad (4.5)$$

Taking the expectation with respect to $\mathbf{1}_{n,j}$, we get

$$\begin{aligned} \mathbb{E}[Y_n | \mathbb{F}_{n-1}] &= Y_{n-1} + 3 \sum_{j=1}^m D_{n-1,j}^2 \frac{D_{n-1,j}}{\sum_{j=1}^m D_{n-1,j}} + 3 \sum_{j=1}^m D_{n-1,j} \frac{D_{n-1,j}}{\sum_{j=1}^m D_{n-1,j}} \\ &\quad + 2 \\ &= Y_{n-1} + \frac{3}{n+2m-3} \sum_{j=1}^m D_{n-1,j}^3 + \frac{3}{n+2m-3} \sum_{j=1}^m D_{n-1,j}^2 + 2 \\ &= Y_{n-1} + \frac{3(Y_{n-1} - (n-1))}{n+2m-3} + \frac{3(Z_{n-1} - (n-1))}{n+2m-3} + 2. \end{aligned} \quad (4.7)$$

The recurrence for $\mathbb{E}[Y_n]$ is obtained by taking another expectation with respect to \mathbb{F}_{n-1} , and by plugging the result of $\mathbb{E}[Z_{n-1}]$. We solve the recurrence for $\mathbb{E}[Y_n]$ with initial condition $\mathbb{E}[Y_0] = Y_0 = 8m - 14$, and obtain the stated result. \square

We are now ready to calculate the second moment of Z_n in the next proposition.

Proposition 4.3. *For $n \geq 1$, the second moment of the Zagreb index of caterpillars is given by*

$$\mathbb{E}[Z_n^2] = \frac{(9m^2 - 3m - 16)n^4 + 2(3m - 2)(12m^2 - 7m - 17)n^3 + O(n^2)}{(2m+1)(2m-1)m(m-1)}.$$

Proof. We begin with squaring the almost-sure relation (between Z_n and Z_{n-1}).

$$\begin{aligned} Z_n^2 &= (Z_{n-1} + 2D_{n-1,j} + 2)^2 \\ &= Z_{n-1}^2 + 4D_{n-1,j}^2 + 4 + 4Z_{n-1}D_{n-1,j} + 4Z_{n-1} + 8D_{n-1,j}. \end{aligned} \quad (4.8)$$

We take the average over j to get

$$\begin{aligned} \mathbb{E}[Z_n^2 | \mathbb{F}_{n-1}] &= Z_{n-1}^2 + 4 \sum_{j=1}^m D_{n-1,j}^2 \frac{D_{n-1,j}}{\sum_{j=1}^m D_{n-1,j}} + 4 \\ &\quad + 4Z_{n-1} \sum_{j=1}^m D_{n-1,j} \frac{D_{n-1,j}}{\sum_{j=1}^m D_{n-1,j}} + 4Z_{n-1} \\ &\quad + 8 \sum_{j=1}^m D_{n-1,j} \frac{D_{n-1,j}}{\sum_{j=1}^m D_{n-1,j}} \\ &= Z_{n-1}^2 + \frac{4(Y_{n-1} - (n-1))}{n+2m-3} + 4 + \frac{4Z_{n-1}(Z_{n-1} - (n-1))}{n+2m-3} \\ &\quad + 4Z_{n-1} + \frac{8(Z_{n-1} - (n-1))}{n+2m-3}. \end{aligned} \quad (4.9)$$

We take another expectation with respect to \mathbb{F}_{n-1} to get the recurrence for $\mathbb{E}[Z_n^2]$. Solving the recurrence with initial condition $\mathbb{E}[Z_0^2] = Z_0^2 = (4m-6)^2$, we obtain the result stated in the proposition. \square

The variance of Z_n is then obtained immediately by taking the difference between $\mathbb{E}[Z_n^2]$ and the square of $\mathbb{E}[Z_n]$.

Corollary 4.4. *For $n \geq 1$, the variance of the Zagreb index of caterpillars is*

$$\text{Var}[Z_n] = \frac{(6m^3 - 22m^2 + 29m - 16)n^4 + O(n^3)}{(2m-1)^2(m-1)^2(2m+1)m}.$$

5. Conclusion

In this article, we investigate the Zagreb index of three classes of random networks, a class of random graphs extended from RRTs, PORTs, and preferential attachment caterpillars. For the first class, we show that the limit distribution of the Zagreb index scaled by \sqrt{n} is normal. For the second class, we prove that the asymptotic distribution of the Zagreb index is not normal by showing that it is always skewed to the right. For the third class, we find that the second moment and the variance of the Zagreb index are both of order n^4 . We thus conjecture that its asymptotic distribution is not normal as well.

One of the possible future work is to study the Zagreb index of more general preferential attachment networks [1], a class of networks extended from PORTs. The recurrence method seems not amenable owing to the non-uniformity of the sampling distribution. One alternative approach is to exploit the degree profile of preferential attachment networks developed in [10]. We will conduct the investigation in this direction and report the results elsewhere.

References

1. Barabási, A. and Albert, R.: Emergence of scaling in random networks, *Science* **286** (1999), no. 5439, 509–512.
2. Devroye, L. and Lu, J.: The strong convergence of maximal degrees in uniform random recursive trees and dags, *Random Structures Algorithms* **7** (1995), no. 1, 1–14.
3. El-Basil, S.: Applications of caterpillar trees in chemistry and physics. *J. Math. Chem.* **1** (1987), no. 2, 153–174.
4. El-Basil, S.: Caterpillar (Gutman) trees in chemical graph theory, in: *Advances in the Theory of Benzenoid Hydrocarbons* (1990) 273–289, Springer, Berlin, Heidelberg.
5. Feng, Q. and Hu, Z.: On the Zagreb index of random recursive trees, *J. Appl. Probab.* **48** (2011), no. 4, 1189–1196.
6. Feng, Q. and Hu, Z.: Asymptotic normality of the Zagreb index of random b -ary recursive trees, *Dal'nevost. Mat. Zh.* **15** (2015), no. 1, 91–101.
7. Gutman, I. and Trinajstić, N.: Graph theory and molecular orbitals. Total ψ -electron energy of alternant hydrocarbons, *Chem. Phys. Lett.* **17** (1972), no. 4, 535–538.
8. Hall, P. and Heyde, C.: *Martingale Limit Theory and Its Application*, Academic Press, New York, NY, 1980.
9. Nikolić, S., Kovačević, G., Miličević, A., and Trinajstić, N: The Zagreb indices 30 years after, *Croat. Chem. Acta* **76** (2003), no. 2, 113–124.
10. Peköz, E., Röllin, A., and Ross, N: Joint degree distributions of preferential attachment random graphs, *Adv. in Appl. Probab.* **49** (2017), no. 2, 368–387.
11. Ren, Y., Zhang, P., and Dey, D.K.: Investigating several fundamental properties of random lobster trees and random spider trees, *Methodol. Comput. Appl. Probab.* (2021), <https://doi.org/10.1007/s11009-021-09863-9>
12. Todeschini, R. and Consonni, V.: *Molecular Descriptors for Chemoinformatics*, Wiley, Hoboken, NJ, 2009.
13. Zhang, P.: On several properties of a class of preferential attachment trees—plane-oriented recursive trees, *Probab. Eng. Inf. Sci.* **35** (2021), no. 4, 839–857.
14. Zhang, P. and Wang, X.: Several topological indices of random caterpillars, *Methodol. Comput. Appl. Probab.* (2021), <https://doi.org/10.1007/s11009-021-09895-1>

PANPAN ZHANG: DEPARTMENT OF BIostatistics, EPIDEMIOLOGY AND INFORMATICS, UNIVERSITY OF PENNSYLVANIA, PHILADELPHIA, PA 19104, USA

Email address: `panpan.zhang@penncmedicine.upenn.edu`