Calculating the correlation coefficients of graph-theoretical indices

Stephan G. Wagner*

Abstract

Using a generating function approach, the correlation coefficients of four different graph-theoretical indices, namely the number of independent vertex subsets, the number of matchings, the number of subtrees and the Wiener index, are asymptotically determined for random rooted ordered trees.

1 Introduction

In [11], the author investigates correlation measures for graph-theoretical indices which are of interest in theoretical chemistry. In particular, the correlation coefficients for these indices are asymptotically determined. Since the necessary calculations are rather lengthy and tedious, only a few of them are explicitly provided there. This additional note fills the gap by explaining the involved details. See [11] for further references and applications.

The underlying stochastic model is the following: of all rooted ordered trees on n vertices, a tree T_n is selected uniformly at random. The parameter we are interested in is the correlation coefficient of two indices $X_n = X(T_n)$ and $Y_n = Y(T_n)$, defined by

$$r(X_n, Y_n) = \frac{E(X_nY_n) - E(X_n)E(Y_n)}{\sqrt{\operatorname{Var}(X_n)\operatorname{Var}(Y_n)}}. \tag{1}$$

Here, X and Y are two of the following four indices:

(1) The *Merrifield-Simmons-* or σ-index is defined to be the number of independent vertex subsets of a graph, i.e. the number of vertex subsets in which no two vertices are adjacent, including the empty set. Merrifield and Simmons investigated the σ-index in their work [8] and pointed out its correlation to boiling points of molecules.

*Department of Mathematics, Graz University of Technology, Steyrergasse 30, A-8010 Graz, Austria (wagner@finanz.math.tugraz.at).

(2) The *Hosoya-* or Z-index ([5]) is defined as the number of independent edge subsets (also referred to as "matchings"), i.e. the number of edge subsets in which no two edges are adjacent, including the empty set again.

(3) The *number of subtrees* is called ρ-index in [8] and was discussed lately in a paper of Székely and Wang [10].

(4) The *Wiener index* is probably the most popular topological index (s. [1, 2, 12]). It is defined as the sum of all the distances between pairs of vertices, i.e.

$$W(G) = \sum_{v,w \in V(G)} d_G(v, w). \tag{2}$$

It will be shown how one can derive functional equations for generating functions yielding to asymptotic formulas for the quantities of interest by means of the Flajolet-Odlyzko singularity analysis [3]. It turns out that the approach is a slightly different one for the correlation with the Wiener index in view of its different growth structure. Therefore, the correlation coefficients of the σ-, Z- and ρ-index are determined in the following section, whereas the correlation to the Wiener index is investigated in Section 3.

2 σ-, Z- and ρ-index

In this section, we want to determine the asymptotic behavior of the generating function

$$\sum_T X(T)Y(T)z^{|T|},$$

where X, Y stand for σ-, Z- or ρ-index (possibly, X and Y are the same). They count, respectively, the number of independent vertex subsets, edge subsets and subtrees of a tree T. This is done by distinguishing between two cases for each of the indices:

- the root vertex belongs to the independent vertex subset/edge subset/subtree,
- the root does not belong to it.

We denote the corresponding quantities by $\sigma_1(T), \sigma_2(T)$ resp. $Z_1(T), Z_2(T)$ and $\rho_1(T), \rho_2(T)$. If $T_1, \ldots, T_k$ are the branches of the rooted tree T, it is easy to see that

$$\sigma_1(T) = \prod_{i=1}^{k} \sigma_2(T_i),$$

$$\sigma_2(T) = \prod_{i=1}^{k} (\sigma_1(T_i) + \sigma_2(T_i)),$$

$$Z_1(T) = \sum_{j=1}^{k} Z_2(T_j) \prod_{\substack{i=1\\ i\neq j}}^{k} (Z_1(T_i) + Z_2(T_i)),$$

$$Z_2(T) = \prod_{i=1}^{k} (Z_1(T_i) + Z_2(T_i)),$$

$$\rho_1(T) = \prod_{i=1}^{k} (1 + \rho_1(T_i)),$$

$$\rho_2(T) = \sum_{i=1}^{k} (\rho_1(T_i) + \rho_2(T_i)).$$

The corresponding generating functions are called S_1, S_2 resp. Z_1, Z_2 and R_1, R_2. Functional equations for these functions which follow from the recursive relations given above and lead to asymptotic formulas for the average indices have already been presented by Klazar [7] and others. For the sake of completeness, these are given here as well:

$$S_1(z) = \frac{z}{1 - S_2(z)},$$

$$S_2(z) = \frac{z}{1 - S_1(z) - S_2(z)},$$

$$Z_1(z) = \frac{zZ_2(z)}{(1 - Z_1(z) - Z_2(z))^2},$$

$$Z_2(z) = \frac{z}{1 - Z_1(z) - Z_2(z)},$$

$$R_1(z) = \frac{z}{1 - R_1(z) - T(z)},$$

$$R_2(z) = \frac{z}{(1 - T(z))^2}(R_1(z) + R_2(z)).$$

Here, $T(z)$ is the generating function for the number of rooted ordered trees. It is well-known that $T(z)$ satisfies the functional equation

$$T(z) = \frac{z}{1 - T(z)},$$

which leads to an explicit formula for the number of rooted ordered trees:

$$t_n = \frac{1}{n}\binom{2n-2}{n-1} \sim \frac{1}{4\sqrt{\pi}} n^{-3/2} 4^n.$$

As determined by Klazar [7], the functional equations presented above yield the following asymptotic formulas for the expected values of our indices:

$$E(\sigma_n) \sim \sqrt{3}\left(\frac{27}{16}\right)^{n-1} \approx (1.02640) \cdot (1.6875)^n.$$

$$E(Z_n) \sim \sqrt{\frac{65-\sqrt{13}}{78}} \left(\frac{35+13\sqrt{13}}{54}\right)^n \approx (0.88719) \cdot (1.51615)^n.$$

$$E(\rho_n) \sim \frac{16}{3\sqrt{15}} \left(\frac{25}{16}\right)^n \approx (1.37706) \cdot (1.5625)^n.$$

If we combine $\sigma_{1,2}$ and $Z_{1,2}$, we obtain generating functions $\mathrm{SZ}_{11}, \mathrm{SZ}_{12}, \mathrm{SZ}_{21}, \mathrm{SZ}_{22}$ for the correlation between σ- and Z-index:

$$\mathrm{SZ}_{ij}(z) = \sum_T \sigma_i(T) Z_j(T) z^{|T|}.$$

The total is denoted by SZ. In an analogous manner, we define SR_{ij},ZR_{ij},SS_{ij},ZZ_{ij} and RR_{ij}.

Next, we observe that $1 \leq \sigma(T), Z(T), \rho(T) \leq 2^{|T|}$ for trivial reasons. This helps us to restrict the range of the radius of convergence. In all our cases, it must lie within the interval $\left[\frac{1}{16}, \frac{1}{4}\right]$. Even more can be told about it: the radius of convergence of XY can be at most the minimum of the radii of X and Y (which are $\frac{4}{27}$, $\frac{13\sqrt{13}-35}{72}$ and $\frac{4}{25}$ for σ-, Z- and ρ-index respectively). Furthermore, since $\sigma(T) \geq F_{|T|+2}$, where F_n denotes the n-th Fibonacci number (a result due to Prodinger and Tichy [9]), we even know that, for example, the radius of convergence of SR is $\leq \frac{4}{25} \cdot \frac{\sqrt{5}-1}{2} = \frac{2(\sqrt{5}-1)}{25}$. For the generating functions SS, ZZ and RR, we may apply the Cauchy-Schwarz-inequality to obtain an upper bound for the convergence radius easily: if X has a convergence radius r, then XX has convergence radius $\leq 4r^2$. Summing up, we have the following bounds for the convergence radii:

- SZ: interval $\left[\frac{1}{16}, \frac{(13\sqrt{13}-35)(\sqrt{5}-1)}{144}\right]$,
- SR: interval $\left[\frac{1}{16}, \frac{2(\sqrt{5}-1)}{25}\right]$,
- ZR: interval $\left[\frac{1}{16}, \frac{4}{25}\right]$,
- SS: interval $\left[\frac{1}{16}, \frac{64}{729}\right]$,
- ZZ: interval $\left[\frac{1}{16}, \frac{1711-455\sqrt{13}}{648}\right]$,
- RR: interval $\left[\frac{1}{16}, \frac{64}{625}\right]$.

In all the cases, we will see that these estimates are sufficient to determine the correct dominating singularity.

2.1 σ- and Z-index

The recursive relations for σ- and Z-index lead to the following system of functional equations:

$$\begin{aligned}
\mathrm{SZ}_{11}(z) &= \frac{z\,\mathrm{SZ}_{22}(z)}{(1-\mathrm{SZ}_{21}(z)-\mathrm{SZ}_{22}(z))^2},\\
\mathrm{SZ}_{12}(z) &= \frac{z}{1-\mathrm{SZ}_{21}(z)-\mathrm{SZ}_{22}(z)},\\
\mathrm{SZ}_{21}(z) &= \frac{z(\mathrm{SZ}_{12}(z)+\mathrm{SZ}_{22}(z))}{(1-\mathrm{SZ}_{11}(z)-\mathrm{SZ}_{12}(z)-\mathrm{SZ}_{21}(z)-\mathrm{SZ}_{22}(z))^2},\\
\mathrm{SZ}_{22}(z) &= \frac{z}{1-\mathrm{SZ}_{11}(z)-\mathrm{SZ}_{12}(z)-\mathrm{SZ}_{21}(z)-\mathrm{SZ}_{22}(z)}.
\end{aligned}$$

For instance, the functional equation for SZ_{11} can be deduced as follows:

$$\begin{aligned}
\mathrm{SZ}_{11}(z) &= \sum_T \sigma_1(T) Z_1(T) z^{|T|}\\
&= \sum_{k\geq 0}\sum_{j=1}^{k}\sum_{T_1}\sum_{T_2}\cdots\sum_{T_k}\left(\sigma_2(T_j)Z_2(T_j)\prod_{i\neq j}\sigma_2(T_i)(Z_1(T_i)+Z_2(T_i))\right)\\
&\quad\cdot z^{|T_1|+\ldots+|T_k|+1}\\
&= z\sum_{k\geq 0} k\,\mathrm{SZ}_{22}(z)(\mathrm{SZ}_{21}(z)+\mathrm{SZ}_{22}(z))^{k-1}\\
&= \frac{z\,\mathrm{SZ}_{22}(z)}{(1-SZ_{21}(z)-\mathrm{SZ}_{22}(z))^2}.
\end{aligned}$$

From these, a single equation for SZ_{22} can be worked out by means of Gröbner bases [4], and this can be used to determine the dominating singularity of SZ. All computations are given in the accompanying Mathematica® files, which can be found on `http://finanz.math.tugraz.at/~wagner/Correlation`. From the equation

$$s^{10}+2zs^8-3zs^7+z^2s^6-4z^2s^5+3z^2s^4-z^3s^3+2z^3s^2-z^3s+z^4=0$$

that is satisfied by $s=\mathrm{SZ}_{22}(z)$, we find that

$$\mathrm{SZ}_{22}(z)\sim 0.171502-0.138532\sqrt{1-\frac{z}{z_0}}$$

around the singularity $z_0 = 0.0982673$. From the relation $\mathrm{SZ}(z) = 1-\frac{z}{\mathrm{SZ}_{22}(z)}$ we obtain

$$\mathrm{SZ}(z)\sim 0.427020-0.462827\sqrt{1-\frac{z}{z_0}},$$

which gives us the asymptotic formula

$$E(\sigma_n Z_n)\sim(0.92565)\cdot(2.54408)^n.$$

by a simple application of the Flajolet-Odlyzko singularity analysis [3], as it is also explained in [11].

2.2 σ- and ρ-index

The recursive relations for σ- and ρ-index lead to the following system of functional equations:

$$\begin{aligned}
\mathrm{SR}_{11}(z) &= \frac{z}{1-\mathrm{SR}_{21}(z)-S_2(z)},\\
\mathrm{SR}_{12}(z) &= \frac{z(\mathrm{SR}_{21}(z)+\mathrm{SR}_{22}(z))}{(1-S_2(z))^2} = \frac{S_1(z)^2(\mathrm{SR}_{21}(z)+\mathrm{SR}_{22}(z))}{z},\\
\mathrm{SR}_{21}(z) &= \frac{z}{1-\mathrm{SR}_{11}(z)-\mathrm{SR}_{21}(z)-S_1(z)-S_2(z)},\\
\mathrm{SR}_{22}(z) &= \frac{z(\mathrm{SR}_{11}(z)+\mathrm{SR}_{12}(z)+\mathrm{SR}_{21}(z)+\mathrm{SR}_{22}(z))}{(1-S_1(z)-S_2(z))^2}\\
&= \frac{S_2(z)^2(\mathrm{SR}_{11}(z)+\mathrm{SR}_{12}(z)+\mathrm{SR}_{21}(z)+\mathrm{SR}_{22}(z))}{z}.
\end{aligned}$$

It would be possible to carry out the same procedure as in the previous case; however, it saves computing time to consider SR_{11} and SR_{21} first. Then, SR_{12} and SR_{22} are given by simple linear equations which result in the formula

$$\mathrm{SR}(z) = \frac{z(S_1(z)^2\,\mathrm{SR}_{21}(z) + z\,\mathrm{SR}_{11}(z) + z\,\mathrm{SR}_{21}(z))}{z^2 - zS_2(z)^2 - S_1(z)^2S_2(z)^2}.$$

We know that S_1 and S_2 are holomorphic in the region of interest, so the dominating singularity of SR is either a singularity of SR_{11} and SR_{21} or a zero of the denominator. However, using the functional equations for S_1 and S_2, we find that the denominator only vanishes at $z = 0$ and at $z = \frac{4}{27}$. Since $\frac{4}{27}$ does not lie in our estimated interval, we have to determine the singularities of SR_{11} and SR_{21}. By the Gröbner basis approach, we find that $s = \mathrm{SR}_{11}(z)$ satisfies the polynomial equation

$$zs^9 - 6z^2s^7 - 4z^3s^6 + 7z^3s^5 - 2z^4s^4 - (z^5+3z^4)s^3 + z^5s^2 + z^5s - z^6 = 0.$$

0 is clearly not a singularity, and the other common zeroes of the polynomial and its derivative satisfy the polynomial equation

$$(27z-4)(8z^2+81)^2(125z^3-412z^2-40936z+3844) = 0.$$

The only value that lies within our bounds and satisfies the equation is $z_0 \approx 0.0938166$. Since $\mathrm{SR}_{21}(z) = 1 - S_2(z) - \frac{z}{\mathrm{SR}_{11}(z)}$ and SR_{11} only vanishes at $z = 0$, the smallest singularity of SR_{21} is the same. Expanding around z_0 gives us

$$\mathrm{SR}(z) \sim 0.560623 - 0.683264\sqrt{1-\frac{z}{z_0}}$$

and thus

$$E(\sigma_n\rho_n) \sim (1.36653)\cdot(2.66477)^n.$$

2.3 Z- and ρ-index

The recursive relations for Z- and ρ-index lead to the following system of functional equations:

$$\begin{aligned}
\mathrm{ZR}_{11}(z) &= \frac{z(\mathrm{ZR}_{21}(z)+Z_2(z))}{(1-\mathrm{ZR}_{11}(z)-\mathrm{ZR}_{21}(z)-Z_1(z)-Z_2(z))^2},\\
\mathrm{ZR}_{12}(z) &= \frac{z(\mathrm{ZR}_{21}(z)+\mathrm{ZR}_{22}(z))}{(1-Z_1(z)-Z_2(z))^2}+\frac{2zZ_2(z)\,\mathrm{ZR}(z)}{(1-Z_1(z)-Z_2(z))^3}\\
&= \frac{Z_2(z)^2(\mathrm{ZR}_{21}(z)+\mathrm{ZR}_{22}(z))}{z}+\frac{2Z_1(z)Z_2(z)\,\mathrm{ZR}(z)}{z},\\
\mathrm{ZR}_{21}(z) &= \frac{z}{1-\mathrm{ZR}_{11}(z)-\mathrm{ZR}_{21}(z)-Z_1(z)-Z_2(z)},\\
\mathrm{ZR}_{22}(z) &= \frac{z\,\mathrm{ZR}(z)}{(1-Z_1(z)-Z_2(z))^2}=\frac{Z_2(z)^2\,\mathrm{ZR}(z)}{z}.
\end{aligned}$$

Using the same approach as in the previous example, we arrive at

$$\mathrm{ZR}(z)=\frac{z(Z_2(z)^2\,\mathrm{ZR}_{21}(z)+z\,\mathrm{ZR}_{11}(z)+z\,\mathrm{ZR}_{21}(z))}{z^2-2zZ_1(z)Z_2(z)-zZ_2(z)^2-Z_2(z)^4}.$$

Again, Z_1 and Z_2 are holomorphic in the region of interest, and the denominator only vanishes at $z=0$ and at $z=\frac{13\sqrt{13}-35}{72}$, which does not lie in our estimated interval, so we have to determine the singularities of ZR_{11} and ZR_{21}. We observe first that $\mathrm{ZR}_{11}(z)=\frac{\mathrm{ZR}_{21}(z)^2(\mathrm{ZR}_{21}(z)+Z_2(z))}{z}$ and $Z_1(z)=\frac{Z_2(z)^3}{z}$. Inserting yields

$$\begin{aligned}
z^2-z\,\mathrm{ZR}_{21}(z)+zZ_2(z)\,\mathrm{ZR}_{21}(z)+Z_2(z)^3\,\mathrm{ZR}_{21}(z)&\\
+z\,\mathrm{ZR}_{21}(z)^2+Z_2(z)\,\mathrm{ZR}_{21}(z)^3+\mathrm{ZR}_{21}(z)^4&=0.
\end{aligned}$$

Elimination of Z_2 via the relation $z^2-zZ_2(z)+zZ_2(z)^2+Z_2(z)^4=0$ gives a polynomial equation of degree 16. For computational purposes, however, it is much faster to find a common zero of the equation above together with its derivative and the condition for Z_2. Then we find that a singularity of ZR_{21} (and thus also of ZR_{11}) and ZR must satisfy the polynomial equation

$$(4096z^2-448z+1)(2560000z^2+2894400z+531441)=0.$$

The only value that lies within our bounds and satisfies the equation is $z_0\approx 0.107095$. Expanding around $z_0=\frac{7+3\sqrt{5}}{128}$ gives us

$$\mathrm{ZR}(z)\sim\frac{1}{928}(211+93\sqrt{5})-\frac{1}{232}\sqrt{\frac{5(128985+57683\sqrt{5})}{58}}\cdot\sqrt{1-\frac{z}{z_0}}$$

and thus

$$E(Z_n\rho_n)\sim\frac{1}{116}\sqrt{\frac{5(128985+57683\sqrt{5})}{58}}\cdot\Big(8(7-3\sqrt{5})\Big)^n.$$

2.4 Variance of the σ-index

For the variances, the calculations are even a little simpler, since we have one variable less to cope with. In this case, we obtain the functional equations

$$\begin{aligned}
\mathrm{SS}_{11}(z) &= \frac{z}{1-\mathrm{SS}_{22}(z)},\\
\mathrm{SS}_{12}(z) &= \frac{z}{1-\mathrm{SS}_{12}(z)-\mathrm{SS}_{22}(z)},\\
\mathrm{SS}_{22}(z) &= \frac{z}{1-\mathrm{SS}_{11}(z)-2\,\mathrm{SS}_{12}(z)-\mathrm{SS}_{22}(z)},
\end{aligned}$$

which result in a single equation for $s = \mathrm{SS}(z) = \mathrm{SS}_{11}(z) + 2\,\mathrm{SS}_{12}(z) + \mathrm{SS}_{22}(z)$:

$$\begin{aligned}
s^6 - 6s^5 + (4z+14)s^4 + (8z^2-20z-16)s^3 + (4z^3-30z^2+36z+9)s^2&\\
-(12z^3-36z^2+28z+2)s - (z^4-8z^3+14z^2-8z) &= 0.
\end{aligned}$$

We use Gröbner bases once again and find the only possible singularity which lies within our bounds: its value is $z_0 \approx 0.0873832$. We calculate the expansion of $\mathrm{SS}(z)$ around z_0:

$$\mathrm{SS}(z) \sim 0.614803 - 0.519010\sqrt{1-\frac{z}{z_0}},$$

and finally obtain the asymptotics of the variance:

$$\mathrm{Var}(\sigma_n) \sim (1.03802)\cdot(2.86096)^n.$$

2.5 Variance of the Z-index

We proceed in the same way in the case of the Z-index:

$$\begin{aligned}
\mathrm{ZZ}_{11}(z) &= \frac{z\,\mathrm{ZZ}_{22}(z)}{(1-\mathrm{ZZ}(z))^2} + \frac{2z(\mathrm{ZZ}_{12}(z)+\mathrm{ZZ}_{22}(z))^2}{(1-\mathrm{ZZ}(z))^3},\\
\mathrm{ZZ}_{12}(z) &= \frac{z(\mathrm{ZZ}_{12}(z)+\mathrm{ZZ}_{22}(z))}{(1-\mathrm{ZZ}(z))^2},\\
\mathrm{ZZ}_{22}(z) &= \frac{z}{1-\mathrm{ZZ}(z)},
\end{aligned}$$

which results in a single equation for $s = \mathrm{ZZ}(z) = \mathrm{ZZ}_{11}(z) + 2\,\mathrm{ZZ}_{12}(z) + \mathrm{ZZ}_{22}(z)$:

$$\begin{aligned}
s^8 - 7s^7 - (z-21)s^6 + (4z-35)s^5 + (2z^2-5z+35)s^4 - (7z^2+21)s^3&\\
-(z^3-9z^2-5z-7)s^2 + (2z^3-5z^2-4z-1)s + (z^4-z^3+z^2+z) &= 0.
\end{aligned}$$

Here, the value of the singularity is $z_0 \approx 0.107969$. We calculate the expansion of $\mathrm{ZZ}(z)$ around z_0:

$$\mathrm{ZZ}(z) \sim 0.296221 - 0.386136\sqrt{1-\frac{z}{z_0}},$$

and finally obtain the asymptotics of the variance:

$$\mathrm{Var}(Z_n) \sim (0.77227)\cdot(2.31549)^n.$$

2.6 Variance of the ρ-index

In this case, we obtain the following system of equations:

$$\begin{aligned}
\mathrm{RR}_{11}(z) &= \frac{z}{(1-\mathrm{RR}_{11}(z)-2R_1(z)-T(z)}, \\
\mathrm{RR}_{12}(z) &= \frac{z(\mathrm{RR}_{11}(z)+\mathrm{RR}_{12}(z)+R_1(z)+R_2(z))}{(1-R_1(z)-T(z))^2} = \\
&= \frac{R_1(z)^2(\mathrm{RR}_{11}(z)+\mathrm{RR}_{12}(z)+R_1(z)+R_2(z))}{z}, \\
\mathrm{RR}_{22}(z) &= \frac{z\,\mathrm{RR}(z)}{(1-T(z))^2}+\frac{2z(R_1(z)+R_2(z))^2}{(1-T(z))^3} \\
&= \frac{T(z)^2\,\mathrm{RR}(z)}{z}+\frac{2R_2(z)^2(1-T(z))}{z}.
\end{aligned}$$

Note that the system can even be solved explicitly by successive solution of quadratic equations. We apply the method that was also used for the covariance of σ- and ρ- resp. Z- and ρ-index: RR_{12} and RR_{22} are expressed in terms of the other functions. We obtain

$$\mathrm{RR}(z) = \mathrm{RR}_{11}(z) + 2\,\mathrm{RR}_{12}(z) + \mathrm{RR}_{22}(z) = \frac{N}{(T(z)^2-z)(R_1(z)^2-z)},$$

where the numerator N is given by

$$\begin{aligned}
N = z\,\mathrm{RR}_{11}(z)(z+R_1(z)^2) &- 2R_1(z)^2R_2(z)^2(1-T(z)) \\
&+ 2zR_1(z)^2(R_1(z)+R_2(z)) + 2zR_2(z)^2(1-T(z)).
\end{aligned}$$

Next, we prove that the denominator doesn't vanish within the bounds of interest: an easy calculation shows that it can only vanish at $z=0$, $z=\frac{1}{4}$ or $z=\frac{4}{25}$.

So it suffices to determine the singularities of $\mathrm{RR}_{11}(z)$. The application of Gröbner bases shows that the smallest singularity of $\mathrm{RR}_{11}(z)$ is the surprisingly nice value $z_0=\frac{8}{81}$. Expansion of RR_{11} around z_0 yields

$$\mathrm{RR}_{11}(z) \sim \frac{2\sqrt{2}}{9} - \frac{4}{3\sqrt{7}}\sqrt{1-\frac{z}{z_0}},$$

so after calculating the values of R_1, R_2 and T at z_0, we see that

$$\mathrm{RR}(z) \sim \frac{2432\sqrt{2}-1632}{3087} - \frac{32\sqrt{14}}{147}\sqrt{1-\frac{z}{z_0}},$$

giving us the asymptotics

$$\mathrm{Var}(\rho_n) \sim \frac{64\sqrt{14}}{147}\cdot\left(\frac{81}{32}\right)^n.$$

3 The Wiener index

Now, we want to determine the asymptotic behavior of the generating function

$$\sum_T X(T)W(T)z^{|T|},$$

where X stands for σ-, Z- or ρ-index,$W(T)$ is the Wiener index and T runs over all rooted ordered trees. Again, we have to distinguish sets containing and not containing the root. For the Wiener index, on the other hand, we have the following recursive relations:

$$D(T) = \sum_{i=1}^{k} D(T_i) + |T| - 1$$

and

$$W(T) = D(T) + \sum_{i=1}^{k} W(T_i) + \sum_{i \neq j} \Big(D(T_i) + |T_i| \Big) |T_j|.$$

Here, D is the total height or internal path length, the sum of the distances to the root. From these relations, one finds the following functional equations for the respective generating functions:

$$D(z) = \frac{zD(z)}{(1-T(z))^2} + zT'(z) - T(z),$$
$$W(z) = D(z) + \frac{zW(z)}{(1-T(z))^2} + \frac{2z^2T'(z)(D(z)+zT'(z))}{(1-T(z))^3},$$

from which $D(z)$ and $W(z)$ can be determined without difficulty. The asymptotic behavior of the average Wiener index follows at once:

$$E(W_n) \sim \frac{\sqrt{\pi}}{4} n^{5/2}.$$

The variance of the Wiener index is given in a paper of Janson [6]:

$$\mathrm{Var}(W_n) \sim \frac{16-5\pi}{80} n^5.$$

Now, we define generating functions $\mathrm{DS}_1, \mathrm{DS}_2, \mathrm{WS}_1, \mathrm{WS}_2$ for the product of $D(T)$ resp. $W(T)$ with the number of independent vertex subsets containing resp. not containing the root, e.g.

$$\mathrm{DS}_1(z) = \sum_T D(T)\sigma_1(T)z^{|T|}.$$

In an analogous manner, we define the functions $\mathrm{DZ}_i, \mathrm{WZ}_i, \mathrm{DR}_i, \mathrm{WR}_i$. Here, we obtain linear functional equations for the generating functions, which can be solved explicitly.

3.1 σ- and Wiener index

The recursive relations give us the following system of functional equations, which can be simplified by means of the functional equations for S_1 and S_2, especially the facts that $\frac{z}{1-S_2(z)} = S_1(z)$, $\frac{z}{1-S_1-S_2(z)} = S_2(z)$ and $\frac{z}{(1-S_2(z))^2} = S_2(z)$ (the latter follows after some simple algebraic manipulations):

$$\begin{aligned}
\mathrm{DS}_1(z) &= \frac{z\,\mathrm{DS}_2(z)}{(1-S_2(z))^2} + zS_1'(z) - S_1(z) \\
&= \mathrm{DS}_2(z)S_2(z) + zS_1'(z) - S_1(z), \\
\mathrm{DS}_2(z) &= \frac{z(\mathrm{DS}_1(z)+\mathrm{DS}_2(z))}{(1-S_1(z)-S_2(z))^2} + zS_2'(z) - S_2(z) \\
&= \frac{S_2(z)^2(\mathrm{DS}_1(z)+\mathrm{DS}_2(z))}{z} + zS_2'(z) - S_2(z), \\
\mathrm{WS}_1(z) &= \mathrm{DS}_1(z) + \frac{z\,\mathrm{WS}_2(z)}{(1-S_2(z))^2} + \frac{2z^2S_2'(z)(\mathrm{DS}_2(z)+zS_2'(z))}{(1-S_2(z))^3} \\
&= \mathrm{DS}_1(z) + S_2(z)\,\mathrm{WS}_2(z) + 2S_1(z)S_2(z)S_2'(z)(\mathrm{DS}_2(z)+zS_2'(z)), \\
\mathrm{WS}_2(z) &= \mathrm{DS}_2(z) + \frac{z(\mathrm{WS}_1(z)+\mathrm{WS}_2(z))}{(1-S_1(z)-S_2(z))^2} \\
&\quad + \frac{2z(zS_1'(z)+zS_2'(z))(\mathrm{DS}_1(z)+\mathrm{DS}_2(z)+zS_1'(z)+zS_2'(z))}{(1-S_1(z)-S_2(z))^3} \\
&= \mathrm{DS}_2(z) + \frac{S_2(z)^2(\mathrm{WS}_1(z)+\mathrm{WS}_2(z))}{z} \\
&\quad + \frac{2S_2(z)^3(S_1'(z)+S_2'(z))(\mathrm{DS}_1(z)+\mathrm{DS}_2(z)+zS_1'(z)+zS_2'(z))}{z}.
\end{aligned}$$

All these equations are obtained as in the following example:

$$\begin{aligned}
\mathrm{DS}_1(z) &= \sum_T D(T)\sigma_1(T)z^{|T|} \\
&= \sum_{k\ge 0}\sum_{T_1}\cdots\sum_{T_k}\left(\sum_{i=1}^k D(T_i)\prod_{j=1}^k \sigma_2(T_j)\right) z^{|T_1|+\ldots+|T_k|+1} \\
&\quad + \sum_T (|T|-1)\sigma_1(T)z^{|T|} \\
&= \sum_{k\ge 0}\sum_{T_1}\cdots\sum_{T_k}\left(\sum_{i=1}^k D(T_i)\sigma_2(T_i)\prod_{j\ne i}\sigma_2(T_j)\right) z^{|T_1|+\ldots+|T_k|+1} \\
&\quad + zS_1'(z) - S_1(z) \\
&= z\sum_{k\ge 0} k\,\mathrm{DS}_2(z)S_2(z)^{k-1} + zS_1'(z) - S_1(z) \\
&= \frac{z\,\mathrm{DS}_2(z)}{(1-S_2(z))^2} + zS_1'(z) - S_1(z).
\end{aligned}$$

We solve the system of linear equations for WS_1 and WS_2 and obtain an expression for $\mathrm{WS}(z) = \mathrm{WS}_1(z) + \mathrm{WS}_2(z)$ in terms of S_1 and S_2. Then, we replace $S_1(z)$ by $\frac{z}{1-S_2(z)}$ and $S_1'(z)$ by

$$S_1'(z) = \frac{d}{dz}\frac{z}{1-S_2(z)} = \frac{1}{1-S_2(z)} + \frac{zS_2'(z)}{(1-S_2(z))^2} = \frac{1}{1-S_2(z)} + S_2(z)S_2'(z).$$

Finally, implicit differentiation of the equation $S_2(z)^3 - 2S_2(z)^2 + S_2(z) - z = 0$ yields

$$S_2'(z) = \frac{1}{3S_2(z)^2 - 4S_2(z) + 1},$$

so WS can be written in terms of S_2 and z only. We obtain an expression of the form

$$\mathrm{WS}(z) = \frac{N}{(1-3S_2(z))^2(1-S_2(z))^3(S_2(z)^2 + S_2(z)^3 - z)^2},$$

where N is a polynomial in S_2 and z. The denominator only vanishes at $z = 0$ (which is clearly no singularity) and at $z = \frac{4}{27}$, which is the dominating singularity of S_2. Therefore, we only have to consider the expansion of WS around $\frac{4}{27}$, which is given by

$$\mathrm{WS}(z) \sim \frac{5}{81\left(1 - \frac{27z}{4}\right)^2}.$$

This yields

$$E(W_n\sigma_n) \sim \frac{20\sqrt{\pi}}{81} n^{5/2} \left(\frac{27}{16}\right)^n.$$

3.2 Z- and Wiener index

All steps are analogous to the previous section. We start from the following functional equations, which are simplified by the relations $Z_1(z) = \frac{z^2}{(1-Z_1(z)-Z_2(z))^3}$ and $Z_2(z) = \frac{z}{1-Z_1(z)-Z_2(z)}$:

$$\begin{aligned}
\mathrm{DZ}_1(z) &= \frac{2zZ_2(z)(\mathrm{DZ}_1(z) + \mathrm{DZ}_2(z))}{(1-Z_1(z)-Z_2(z))^3} + \frac{z\,\mathrm{DZ}_2(z)}{(1-Z_1(z)-Z_2(z))^2} + zZ_1'(z) - Z_1(z) \\
&= \frac{2Z_1(z)Z_2(z)(\mathrm{DZ}_1(z) + \mathrm{DZ}_2(z))}{z} + \frac{Z_2(z)^2\,\mathrm{DZ}_2(z)}{z} + zZ_1'(z) - Z_1(z), \\
\mathrm{DZ}_2(z) &= \frac{z(\mathrm{DZ}_1(z) + \mathrm{DZ}_2(z))}{(1-Z_1(z)-Z_2(z))^2} + zZ_2'(z) - Z_2(z) \\
&= \frac{Z_2(z)^2(\mathrm{DZ}_1(z) + \mathrm{DZ}_2(z))}{z} + zZ_2'(z) - Z_2(z),
\end{aligned}$$

$$
\begin{aligned}
\mathrm{WZ}_1(z) &= \mathrm{DZ}_1(z) + \frac{2zZ_2(z)(\mathrm{WZ}_1(z)+\mathrm{WZ}_2(z))}{(1-Z_1(z)-Z_2(z))^3} + \frac{z\,\mathrm{WZ}_2(z)}{(1-Z_1(z)-Z_2(z))^2} \\
&\quad + \frac{2z}{(1-Z_1(z)-Z_2(z))^3}\Big((\mathrm{DZ}_2(z)+zZ_2'(z))(zZ_1'(z)+zZ_2'(z)) \\
&\qquad + zZ_2'(z)(\mathrm{DZ}_1(z)+\mathrm{DZ}_2(z)+zZ_1'(z)+zZ_2'(z))\Big) \\
&\quad + \frac{6zZ_2(z)(zZ_1'(z)+zZ_2'(z))(\mathrm{DZ}_1(z)+\mathrm{DZ}_2(z)+zZ_1'(z)+zZ_2'(z))}{(1-Z_1(z)-Z_2(z))^4} \\
&= \mathrm{DZ}_1(z) + \frac{2Z_1(z)Z_2(z)(\mathrm{WZ}_1(z)+\mathrm{WZ}_2(z))}{z} + \frac{Z_2(z)^2\,\mathrm{WZ}_2(z)}{z} \\
&\quad + 2Z_1(z)\Big((\mathrm{DZ}_2(z)+zZ_2'(z))(Z_1'(z)+Z_2'(z)) \\
&\qquad + Z_2'(z)(\mathrm{DZ}_1(z)+\mathrm{DZ}_2(z)+zZ_1'(z)+zZ_2'(z))\Big) \\
&\quad + \frac{6Z_1(z)Z_2(z)^2}{z}(Z_1'(z)+Z_2'(z))(\mathrm{DZ}_1(z)+\mathrm{DZ}_2(z)+zZ_1'(z)+zZ_2'(z)), \\
\mathrm{WZ}_2(z) &= \mathrm{DZ}_2(z) + \frac{z(\mathrm{WZ}_1(z)+\mathrm{WZ}_2(z))}{(1-Z_1(z)-Z_2(z))^2} \\
&\quad + \frac{2z^2(Z_1'(z)+Z_2'(z))(\mathrm{DZ}_1(z)+\mathrm{DZ}_2(z)+zZ_1'(z)+zZ_2'(z))}{(1-Z_1(z)-Z_2(z))^3} \\
&= \mathrm{DZ}_2(z) + \frac{Z_2(z)^2(\mathrm{WZ}_1(z)+\mathrm{WZ}_2(z))}{z} \\
&\quad + \frac{2Z_2(z)^3(Z_1'(z)+Z_2'(z))(\mathrm{DZ}_1(z)+\mathrm{DZ}_2(z)+zZ_1'(z)+zZ_2'(z))}{z}.
\end{aligned}
$$

We solve this linear equation for WZ_1 and WZ_2 and obtain an expression for $\mathrm{WZ}(z) = \mathrm{WZ}_1(z) + \mathrm{WZ}_2(z)$ in terms of Z_1 and Z_2. Then, we replace $Z_1(z)$ by $\frac{Z_2(z)^3}{z}$ and $Z_1'(z)$ by $\frac{3Z_2(z)^2Z_2'(z)}{z} - \frac{Z_2(z)^3}{z^2}$. Finally, implicit differentiation of the functional equation $z^2 - zZ_2(z) + zZ_2(z)^2 + Z_2(z)^4 = 0$ yields

$$
Z_2'(z) = \frac{2z - Z_2(z) + Z_2(z)^2}{z - 2zZ_2(z) - 4Z_2(z)^3},
$$

so WZ can be written in terms of Z_2 and z only. We obtain an expression of the form

$$
\mathrm{WZ}(z) = \frac{N}{z^3(z^2 - zZ_2(z)^2 - 3Z_2(z)^4)^2(z - 2zZ_2(z) - 4Z_2(z)^3)^2},
$$

where N is a polynomial in Z_2 and z. The denominator only vanishes at $z = 0$ (which is clearly no singularity) and at $z = \frac{\pm 13\sqrt{13}-35}{72}$, which are singularities of Z_2. Therefore, we only have to consider the expansion of WZ around the dominating singularity $z_0 = \frac{13\sqrt{13}-35}{72}$, which is given by

$$
\mathrm{WZ}(z) \sim \frac{91 - 5\sqrt{13}}{1248\left(1 - \frac{z}{z_0}\right)^2}.
$$

This yields

$$E(W_nZ_n) \sim \frac{(91-5\sqrt{13})\sqrt{\pi}}{312} n^{5/2} \left(\frac{35+13\sqrt{13}}{54}\right)^n .$$

3.3 ρ- and Wiener index

Again, all steps are almost analogous to the previous section. We start from the following functional equations, which are simplified by the relations $T(z) = \frac{z}{1-T(z)}$, $R_1(z) = \frac{z}{1-R_1(z)-T(z)}$ and $R_2(z) = \frac{z(R_1(z)+R_2(z))}{(1-T(z))^2}$:

$$\begin{aligned}
\mathrm{DR}_1(z) &= \frac{z(\mathrm{DR}_1(z)+D(z))}{(1-R_1(z)-T(z))^2} + zR_1'(z) - R_1(z) \\
&= \frac{R_1(z)^2(\mathrm{DR}_1(z)+D(z))}{z} + zR_1'(z) - R_1(z), \\
\mathrm{DR}_2(z) &= \frac{2zD(z)(R_1(z)+R_2(z))}{(1-T(z))^3} + \frac{z(\mathrm{DR}_1(z)+\mathrm{DR}_2(z))}{(1-T)^2} + zR_2'(z) - R_2(z) \\
&= \frac{2D(z)T(z)R_2(z)}{z} + \frac{T(z)^2(\mathrm{DR}_1(z)+\mathrm{DR}_2(z))}{z} + zR_2'(z) - R_2(z), \\
\mathrm{WR}_1(z) &= \mathrm{DR}_1(z) + \frac{z(\mathrm{WR}_1(z)+W(z))}{(1-R_1(z)+T(z))^2} \\
&\quad + \frac{2z(zR_1'(z)+zT'(z))(\mathrm{DR}_1(z)+D(z)+zR_1'(z)+zT'(z))}{(1-R_1(z)-T(z))^3} \\
&= \mathrm{DR}_1(z) + \frac{R_1(z)^2(\mathrm{WR}_1(z)+W(z))}{z} \\
&\quad + \frac{2R_1(z)^3(R_1'(z)+T'(z))(\mathrm{DR}_1(z)+D(z)+zR_1'(z)+zT'(z))}{z}, \\
\mathrm{WR}_2(z) &= \mathrm{DR}_2(z) + \frac{2zW(z)(R_1(z)+R_2(z))}{(1-T(z))^3} + \frac{z(\mathrm{WR}_1(z)+\mathrm{WR}_2(z))}{(1-T(z))^2} \\
&\quad + \frac{2z(\mathrm{DR}_1(z)+\mathrm{DR}_2(z)+zR_1'(z)+zR_2'(z))zT'(z)}{(1-T(z))^3} \\
&\quad + \frac{2z(zR_1'(z)+zR_2'(z))(D(z)+zT'(z))}{(1-T(z))^3} \\
&\quad + \frac{6z^2T'(z)(D(z)+zT'(z))(R_1(z)+R_2(z))}{(1-T(z))^4} \\
&= \mathrm{DR}_2(z) + \frac{2T(z)^3W(z)(R_1(z)+R_2(z))}{z^2} + \frac{T(z)^2(\mathrm{WR}_1(z)+\mathrm{WR}_2(z))}{z} \\
&\quad + \frac{2T(z)^3(\mathrm{DR}_1(z)+\mathrm{DR}_2(z)+zR_1'(z)+zR_2'(z))T'(z)}{z} \\
&\quad + \frac{2T(z)^3(R_1'(z)+R_2'(z))(D(z)+zT'(z))}{z} \\
&\quad + \frac{6T(z)^4T'(z)(D(z)+zT'(z))(R_1(z)+R_2(z))}{z^2}.
\end{aligned}$$

In this case, we solve the linear equation for WR_1 and WR_2 and insert the exact expressions for T, D, W, R_1 and R_2, which can be determined by simple quadratic equations: we have

$$R_1(z) = \frac{1}{4}\left(1 + \sqrt{1-4z} - \sqrt{2(1-10z+\sqrt{1-4z})}\right)$$

and

$$R_2(z) = \frac{1-\sqrt{1-4z}}{2\sqrt{1-4z}} \cdot R_1(z).$$

Note that we always have to take the branch whose value is 0 at $z = 0$. We will use q_1 as an abbreviation for $\sqrt{1-4z}$ and q_2 as an abbreviation for $\sqrt{2(1-10z+\sqrt{1-4z})}$. Then we also have

$$T(z) = \frac{1-q_1}{2},\ D(z) = \frac{2z^2}{q_1^2(1+q_1)},\ W(z) = \frac{z^2}{q_1^4}.$$

We obtain an exact expression for $\mathrm{WR}(z) = \mathrm{WR}_1(z) + \mathrm{WR}_2(z)$ in terms of a rational function in q_1 and q_2. The denominator of this expression is given by

$$q_1^6(5q_1-3)(3-5q_1+q_2)^2,$$

which only vanishes at $z = 0$, $z = \frac{1}{4}$ and $z = \frac{4}{25}$. Furthermore, q_1 has its only singularity at $\frac{1}{4}$, and q_2 has singularities exactly at $\frac{1}{4}$ and $\frac{4}{25}$. Therefore, the singularity of WR we have to investigate is $\frac{4}{25}$. We obtain the expansion

$$\mathrm{WR}(z) \sim \frac{1}{15\left(1-\frac{25z}{4}\right)^2},$$

which yields

$$E(W_n\sigma_n) \sim \frac{4\sqrt{\pi}}{15} n^{5/2}\left(\frac{25}{16}\right)^n.$$

$r(\sigma_n, Z_n) \sim (-1.01706) \cdot (0.99405)^n$
$r(\sigma_n, \rho_n) \sim (1.05088) \cdot (0.99023)^n$
$r(Z_n, \rho_n) \sim (-1.08924) \cdot (0.97853)^n$
$r(W_n, \sigma_n) \sim (-0.27891) \cdot (0.99767)^n$
$r(W_n, Z_n) \sim (0.40351) \cdot (0.99637)^n$
$r(W_n, \rho_n) \sim (-1.78357) \cdot (0.98209)^n$

Table 1: Asymptotic formulas for the correlation coefficients.

4 Conclusion

From the expected values and variances which were calculated in the preceding sections, it is possible now to deduce the asymptotic correlation coefficients for the investigated indices (Table 1). For their interpretations and further discussion see [11].

References

[1] A. A. Dobrynin, R. Entringer, and I. Gutman. Wiener index of trees: theory and applications. *Acta Appl. Math.*, 66(3):211–249, 2001.

[2] R. C. Entringer, A. Meir, J. W. Moon, and L. A. Székely. The Wiener index of trees from certain families. *Australas. J. Combin.*, 10:211–224, 1994.

[3] P. Flajolet and A. Odlyzko. Singularity analysis of generating functions. *SIAM J. Discrete Math.*, 3(2):216–240, 1990.

[4] R. Fröberg. *An introduction to Gröbner bases.* Pure and Applied Mathematics (New York). John Wiley & Sons Ltd., Chichester, 1997.

[5] H. Hosoya. Topological index as a common tool for quantum chemistry, statistical mechanics, and graph theory. In *Mathematical and computational concepts in chemistry (Dubrovnik, 1985)*, Ellis Horwood Ser. Math. Appl., pages 110–123. Horwood, Chichester, 1986.

[6] S. Janson. The Wiener index of simply generated random trees. *Random Structures Algorithms*, 22(4):337–358, 2003.

[7] M. Klazar. Twelve countings with rooted plane trees. *European J. Combin.*, 18(2):195–210, 1997.

[8] R. E. Merrifield and H. E. Simmons. *Topological Methods in Chemistry.* Wiley, New York, 1989.

[9] H. Prodinger and R. F. Tichy. Fibonacci numbers of graphs. *Fibonacci Quart.*, 20(1):16–21, 1982.

[10] L. A. Székely and H. Wang. On subtrees of trees. *Adv. in Appl. Math.*, 34(1):138–155, 2005.

[11] S. Wagner. Correlation of graph-theoretical indices. *SIAM Journal on Discrete Mathematics*, to appear. Preprint available at `http://finanz.math.tugraz.at/~wagner/`.

[12] H. Wiener. Structural determination of paraffin boiling points. *J. Amer. Chem. Soc.*, 69:17–20, 1947.