Ranking Trees Based on Global Centrality Measures

Amir Barghi* and Daryl DeFord†

Abstract

Trees, or connected graphs with no cycles, are a commonly studied combinatorial family. When many natural metrics on networks are applied to the set of all trees of a fixed order, the extremal values are realized at the star and path graphs. In this paper, we prove inequalities for several global centrality measures, such as global closeness and betweenness centralities, and for graphical Stirling numbers of the first kind that interpolate these extremes. Moreover, we provide two algorithms that allow us to traverse the space of non-isomorphic trees of a fixed order, one towards the star graph of the same order and the other towards the path. Furthermore, we investigate the relationship between these global centrality measures on the one hand and the $(n-2)$nd Stirling numbers of the first kind for small trees on the other hand, demonstrating a strong association between them, in particular with respect to the hierarchical structures obtained from applying our two interpolating algorithms. Based on our observations from these small trees, we prove general bounds that relate the $(n-2)$nd Stirling numbers of the first kind of trees of order n to these global centrality measures. Finally, we provide two related approaches to totally order the set of all non-isomorphic trees of fixed order. We show that the totally ordering obtained from one of these approaches is consistent with the hierarchical structure obtained from our two tree interpolation algorithms in addition to being one of the features to use for predicting the $(n-2)$nd Stirling numbers of the first kind for small trees.

Keywords: degree centrality, closeness centrality, betweenness centrality, Freeman centralization, Stirling numbers of the first kind for graphs

1 Introduction

Let G be a graph of order n. The kth Stirling number of the first kind for G, denoted by $\left[{G \atop k}\right]$, is the number of partitions of G into exactly k disjoint cycles, where a single vertex is a 1-cycle, an edge is a 2-cycle, and cycles of order three or higher have two orientations. Summing over k, we get the graphical factorial of G, denoted by $G!$.

One motivation behind the above definition is a graphical model of seating rearrangements as discussed by Honsberger [11] and Kennedy and Cooper [12] and Otake, Kennedy, and Cooper [15]. In this model, individuals are seated on the vertices of a graph G, initially. Then an individual can stay put at their seat, a pair of individuals whose seats are adjacent can switch seats, or a group of individuals on a cycle in G can move one seat over on that cycle in one direction or the other. Provided that each individual seated on a vertex in G is involved in one and exactly one of the above actions, we call an instance of these collective actions a seating rearrangement. It is not hard not to see that a single seating rearrangement partitions G into k cycles, i.e., a cycle cover, where 1-cycles and 2-cycles are allowed and cycles of order three or higher have two orientations [6].

Another motivation behind the above definition for Stirling numbers of the first kind for graphs is the definition of Stirling numbers of the second kind for graphs [13, 7]: Let G be a graph. The kth Stirling number of the second kind for G, denoted by $\left\{{G \atop k}\right\}$, is the number of ways to partition the vertex set into k independent sets. Summing over k, one obtains the Bell number for graph G, denoted by $B(G)$. Let E_n and K_n be the empty graph and the complete graph of order n, respectively. It is not hard to see that $\left\{{E_n \atop k}\right\} = \left\{{n \atop k}\right\}$ and $B(E_n) = B_n$. As we saw above, a cycle cover of G with k cycles is the Stirling numbers of the first kind equivalent of partitioning G into k independent sets in the case of Stirling numbers of the

*Department of Mathematics and Statistics, Saint Michael's College, Colchester, VT, USA; email: `abarghi@smcvt.edu`
†Department of Mathematics and Statistics, Washington State University, Pullman, WA, USA; email: `daryl.deford@wsu.edu`

second kind for graphs. This provides us with the equivalent identities $\left[{K_n \atop k}\right] = \left[{n \atop k}\right]$ and $K_n! = n!$. For a more detailed discussion, see a paper by Barghi [1].

A natural problem in enumerative graph theory is to take a sequence of related graphs $G_1, G_2, \ldots$ and compute the associated sequence of graphical kth Stirling numbers of the first kind $\left[{G_1 \atop k}\right], \left[{G_2 \atop k}\right], \ldots$. Previous research on graph families has shown that these sequences are often related to common combinatorial sequences, such as the Fibonacci and Pell numbers [1, 5, 6].

An extension of this work on graphical Stirling numbers of the first kind is to consider the set of trees on a fixed order n and then study the scaling behavior as the number of vertices grows. For trees, the graphical Stirling numbers of the first kind simply count the number of matchings with a prescribed number of edges. These measurements give us a partial ordering of trees, whose extreme values are realized by the path and star graphs of order n. In this paper, we prove some initial results describing this partial order and relate the ordering to Freeman centralization measures applied to the same set of trees [9, 10].

Centrality measures have used in the study of social networks to identify vertices that are most important to the structure or dynamics of the network. A classification of such measures is given in a paper by Borgatti [3]. Here, we show that rankings defined by global versions of these metrics reproduces similar behavior to that of graphical Stirling numbers of the first kind, providing an interpolation between the values for the path and the star of a given order.

One motivation behind our work is the difficulty of computing graphical Stirling numbers of the first kind, even for trees. For a fixed n and k, the kth Stirling numbers of the first kind for trees for all trees of order n have a hierarchical structure which some other graph invariants also posses. We are interested in finding ways to approximate graphical Stirling numbers of the first kind using easier-to-compute graph invariants such as global degree centrality, global closeness centrality, and global between centrality—we will define these global centrality measures later in this section. For example, in Figure 1, we see the association between $\left[{T \atop n-2}\right]$ and the mean value of closeness and betweenness global centralities grouped by single values of $\left[{T \atop n-2}\right]$, where T is a tree or order 11 and 12. In these plots, the error bars represent the range of closeness and betweenness global centralities grouped by single values of $\left[{T \atop n-2}\right]$. As we see in Figure 1, we can predict $\left[{T \atop n-2}\right]$ using global betweenness or closeness centrality and statistical methods such as linear regression. We have explored this idea in detail in a recent paper [2].

Our focus on trees in this paper stems from the fact that for graphs in general, and sparse graphs in particular, spanning trees provide a way of finding lower bounds in general, or reasonable approximations in particular, for the Stirling numbers of the first kind for graphs.

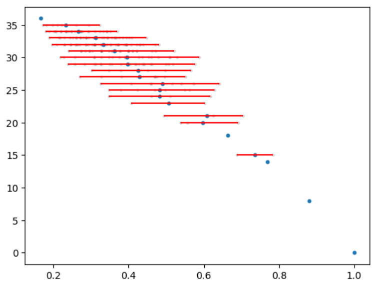
In this paper, we prove monotonicity results with respect to two classes of algorithms for iteratively moving between trees of a given order, interpolating extremes of these measures. This allows us to understand the relationships between the metrics and to compute multiplicative bounds that allow us to estimate the values of the more complex measurements from the easiest to compute metrics. A key tool for our investigation of these functions is analyzing the hierarchical structure of the trees of a fixed order that arises from the interpolation algorithms.

1.1 Mathematical Preliminaries

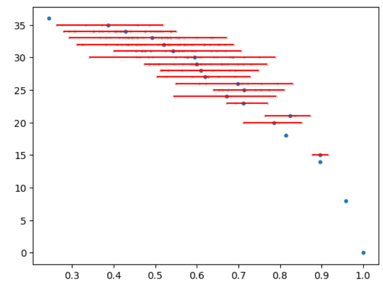
We will denote the empty graph, the path, the cycle, and the star graph of order n by E_n, P_n, C_n, and S_n, respectively. The complement of a graph G, denoted by $\bar{G}$, is the graph such that $V(\bar{G}) = V(G)$ and $uv \in E(\bar{G})$ if and only if $uv \notin E(G)$. The degree of a vertex v in a graph is denoted by $\deg_G(v)$; if it is clear from the context what the underlying graph is, we drop the subscript. We also denote the maximum and minimum degrees in a graph G, by $\Delta(G)$ and $\delta(G)$, respectively. If the underlying graph is clear from context, we simplify these notations to Δ and δ, respectively. The distance between two vertices u, v in a graph G is denoted by $d_G(u, v)$; again, we drop the subscript when it is clear what t he underlying graph is. The eccentricity of a vertex is defined as $\epsilon(v) = \max_{u \in V(G)} d(u, v)$. The diameter and radius of a graph G is the maximum and minimum eccentricity is the graph, respectively. In other words,

$$\operatorname{diam}(G) = \max_{v \in V(G)} \epsilon(v), \qquad \operatorname{radius}(G) = \min_{v \in V(G)} \epsilon(v).$$

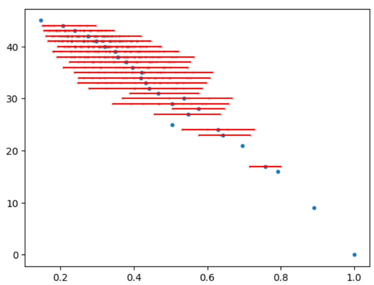
The center $\operatorname{ctr}(G)$ of a graph G is the set of vertices with minimum eccentricity. For a tree T, $\operatorname{ctr}(T)$ consists of either a single vertex or a pair of vertices. We denote the tree obtained from P_{n-1} by connecting a



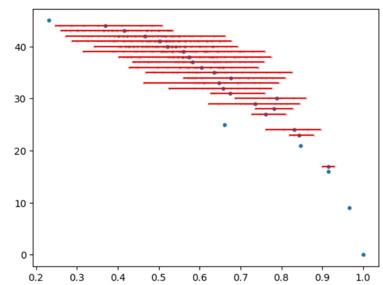
(a) Closeness, $n = 11$



(b) Betweenness, $n = 11$



(c) Closeness, $n = 12$



(d) Betweenness, $n = 12$

Figure 1: The association between $\left[{T \atop n-2}\right]$ (the y-axis) and the mean value of closeness and betweenness global centralities grouped by single values of $\left[{T \atop n-2}\right]$, where T is a tree or order 11 and 12; the error bars represent the range of closeness and betweenness global centralities grouped by single values of $\left[{T \atop n-2}\right]$

new vertex to a vertex in $\mathrm{ctr}(P_{n-1})$ by $\widetilde{P}_n$. Moreover, we denote the tree obtained from S_{n-1} by connecting a new vertex to one of its leaves by $\widetilde{S}_n$.

Regarding the radii and diameters of trees, we have the following result whose proof is straightforward:

Proposition 1. *Let $n \geq 2$. For any tree T of order n,*

$$\mathit{diam}(S_n) \leq \mathit{diam}(T) \leq \mathit{diam}(P_n) \qquad \text{and} \qquad \mathit{radius}(S_n) \leq \mathit{radius}(T) \leq \mathit{radius}(P_n).$$

1.1.1 Global Centrality Measures

Both degree and eccentricity are examples of local measures of centrality, i.e., measures that gauge the importance of a vertex in a graph relative to other vertices in the graph. The use of measure in this context does not refer to a measure as in real analysis but rather a function defined on vertices or graphs [10]. One way to define a global centrality measure is by considering the maximum and minimum values a local centrality measure takes, as is the case with $\Delta(G)$ and $\delta(G)$ on the one hand, and $\mathrm{diam}(G)$ and $\mathrm{radius}(G)$ on the other hand. However, there is an alternative way to define a global centrality measure based on a local centrality measure given by Freeman in 1978 [10]: Let $\mu_G(v)$ be a centrality measure defined for every vertex v in a graph G of order n. Then the global centrality measure associated with μ, denoted by $C_\mu(G)$, is defined as

$$C_\mu(G) = \frac{\sum_{i=1}^{n} (\mu_G(v^*) - \mu_G(v_i))}{H},$$

where $V = \{v_1, \ldots, v_n\}$, v^* is a vertex in G such that $\mu_G(v^*) = \max_{v \in V} \mu_G(v)$, and H is the maximum value the numerator of $C_\mu(G)$ realizes for all graphs of order n for which μ is defined for every vertex. We apply this definition to three local centrality measures to obtain the associated global centrality measure, which are the focus of this paper: degree centrality, closeness centrality, betweenness centrality. According to Freeman [10], the rationale behind this definition is:

> Ideally, all indexes of graph centralization, regardless of the point-base upon which they are built, should have certain features in common: (1) they should measure the degree to which the

centrality of the most central point exceeds the centrality of all other points, and (2) they should each be expressed as a ratio of that excess to its maximum possible value for a graph containing the observed number of points.

The global degree centrality of a graph G is defined as

$$C_{\deg}(G) = \frac{\sum_{i=1}^{n}(\deg(v_\Delta) - \deg(v_i))}{H_{\deg}},$$

where $n = |V(G)|$, v_Δ is a vertex in G such that $\Delta(G) = \deg(v_\Delta)$, and $H_{\deg}$ is the maximum value the numerator of $C_{\deg}(G)$ realizes for all graphs of order n, namely $H_{\deg} = (n-1)(n-2) = n^2 - 3n + 2$ which is realized when the graph is the star graph of order n [10].

The next global centrality measure we define here is closeness centrality. Let G be a connected graph. For a vertex v, closeness is defined as

$$\mathrm{cls}_G(v) = \frac{n-1}{\sum_{u \in V} d(u,v)},$$

where $d(u,v)$ is the distance between u and v. The global closeness centrality is defined as

$$C_{\mathrm{cls}}(G) = \frac{\sum_{i=1}^{n}(\mathrm{cls}_G(v^*) - \mathrm{cls}_G(v_i))}{H_{\mathrm{cls}}},$$

where $n = |V(G)|$, v^* is a vertex in G such that $\mathrm{cls}(v^*) = \max_{v \in V} \mathrm{cls}(v)$, and H_{cls} is the maximum value the numerator of $C_{\mathrm{cls}}(G)$ realizes for all connected graphs of order n. As a result, $C_{\mathrm{cls}}(G)$ takes values between 0 and 1.

The third global centrality measure we consider in this paper is betweenness centrality. Let G be a connected graph. For a vertex v in a graph G, the betweenness centrality is defined as

$$\mathrm{btw}(v) = \frac{1}{\binom{n-1}{2}} \sum_{u,w \in V-v} \frac{P(u,w;v)}{P(u,w)},$$

where $P(u,w)$ is the number of shortest paths from u to w and $P(u,w;v)$ are such paths that go through v. If necessary, we will use $P_G(u,w)$ and $P_G(u,w;v)$ to highlight the underlying graph G. We divide by $\binom{n-1}{2}$ to normalize this centrality measure. The global betweenness centrality is defined as

$$C_{\mathrm{btw}}(G) = \frac{\sum_{i=1}^{n}(\mathrm{btw}_G(v^*) - \mathrm{btw}_G(v_i))}{H_{\mathrm{btw}}},$$

where $n = |V(G)|$, v^* is a vertex in G such that $\mathrm{btw}_G(v^*) = \max_{v \in V} \mathrm{btw}_G(v)$, and H_{btw} is the maximum value the numerator of $C_{\mathrm{btw}}(G)$ realizes for all connected graphs of order n. For example, $H_{\mathrm{btw}} C_{\mathrm{btw}}(S_n) = n-1$. Freeman shows that this maximum is realized for S_n [9]. In the case of $C_{\mathrm{btw}}(G)$ it does not matter whether we use a normalized or non-normalized definition for $C_{\mathrm{btw}}(v)$. It should be noted that $C_{\mathrm{btw}}(G)$ also takes values between 0 and 1.

1.2 Outline

In Section 2, we present two tree interpolation algorithms (the star-to-path and the path-to-star algorithm). We use these two algorithms to explore the space of all non-isomorphic trees of a fixed order. In Section 3, we prove results regarding Stirling numbers of the first kind for trees and forests. In particular, we show that for any tree T of order n,

$$\begin{bmatrix} S_n \\ k \end{bmatrix} \leq \begin{bmatrix} T \\ k \end{bmatrix} \leq \begin{bmatrix} P_n \\ k \end{bmatrix}.$$

As we saw in Proposition 1, both the radius and diameter of a tree T of a fixed order satisfy similar inequalities as does the kth Stirling number of the first kind for T. These simple observations suggest a

(a) Path-to-star

(b) Star-to-path

Figure 2: The hierarchical structure for trees of order 11 based on the path-to-star and star-to-path algorithms; the colors in each subfigure represent the distance from S_{11} and P_{11}, respectively

natural ordering on trees based on different structural measures, something we explore in Sections 4, 5, and 6. In particular, in these sections, we show that for any tree of order $n \geq 2$,

$$C_{\text{deg}}(P_n) \leq C_{\text{deg}}(T) \leq C_{\text{deg}}(S_n),$$

$$C_{\text{cls}}(P_n) \leq C_{\text{cls}}(T) \leq C_{\text{cls}}(S_n),$$

and

$$C_{\text{btw}}(P_n) \leq C_{\text{btw}}(T) \leq C_{\text{btw}}(S_n),$$

respectively. This shows that a number of global graph statistics obtain their extrema at P_n and S_n for trees of fixed order. In Section 7, we provide two related methods of defining a total ordering on the space of all non-isomorphic trees of fixed order, based on distinguishing polynomials introduced by Liu [14]. Finally, in Section 9, we discuss future directions for this paper.

2 Tree Interpolation Algorithms

Throughout our analysis, we will use two main algorithms for traversing the space of non-isomorphic trees of a fixed order, providing us with a hierarchical structure similar to the ones we see in Figure 2. These algorithms are intended to interpolate between the path and the star graph, while monotonically increasing or decreasing the relevant centrality measure. Each of these algorithms proceeds by changing the attachment of a leaf at every step of the process, terminating when the desired path or star graph is reached.

The methods we introduce are not the only possible algorithms for interpolating these sets of trees but they are sufficient for proving our bounds on the global centrality measures below. Additionally, these methods introduce a stochastic component for simplicity but deterministic versions could be implemented by ordering the vertices in the tree and always selecting the vertex of lowest order among all possible candidates. This can arise when there are ties for the various measurements, such as centralities or selecting a longest path. As we will see in our proofs below, this randomness does not impact the results and since the number of leaves decreases (increases) monotonically the algorithms are guaranteed to terminate.

Our first algorithm transforms a given tree T to be more star-like at every step by identifying a most central vertex v and then selecting a leaf u at maximum distance from v. We then u from T and reattach it as a leaf at v. Repeating this process eventually transforms T into the star graph. This algorithm is represented in pseudo-code in Algorithm 1. To make trees more path-like we take the opposite approach. For a given tree T we begin by identifying a subpath Q of T realizing the diameter and a distinguished leaf w on Q. At each step, we remove a leaf u closest to the central vertex that does not lie on Q and reattach it to w, making u the target for the next step. This is described in pseudo-code in Algorithm 2. We use the idea behind Algorithm 2 in the proofs of Theorem 7 and 9.

Algorithm 1: Path-to-Star

Input: A tree T
Output: `TreeList`, d

Initialize: `TreeList` $= \emptyset$, d=0
while $T \ncong S_n$ **do**
  Select: A most central vertex v
  Select: A leaf u at maximum distance from v
  Remove: u from T
  Attach: u to v as a leaf in T
  Replace: d with $d + 1$
  Add: T to `TreeList`
end

Return: `TreeList`, d

Algorithm 2: Star-to-Path

Input: A tree T
Output: `TreeList`, d

Initialize: `TreeList` $= \emptyset$, d=0
  Select: A subpath Q of T of maximum length
  Select: A leaf w of Q
while $T \ncong P_n$ **do**
  Select: A most central vertex v
  Select: A leaf u at minimum distance from v that does not lie on Q
  Remove: u from T
  Attach: u to w in T
  Replace: w with u
  Replace: d with $d + 1$
  Add: T to `TreeList`
end

Return: `TreeList`, d

3 Stirling Numbers of the First Kind for Trees and Forests

In order to compute the complete list of Stirling numbers of the first kind for trees and forests, we apply a modified version of the standard deletion-contraction algorithm, which we call the deletion-inclusion algorithm described in Algorithms 3 and 4 of Barghi and DeFord [2]. For an implementation of this algorithm, see the function `deletion_inclusion_alg_trees(G)` in the Python script `tree_functions_1.py` on our corresponding GitHub page. The modification requires fewer operations than the original because we can remove all edges incident to the end points of the contracted edges, but this speedup is not enough to overcome the growth rate of graphical Stirling numbers of the first kind over all k, as n increases for trees of order n. For large k this could be combined with a pruning algorithm to further increase the speedup but we do not consider this question further in this paper.

Regarding forests, we have the following partial results that connect graphical Stirling numbers of the first kind with graphical Stirling numbers of the second kind for graphs, and consequently, the graphical factorials with the graphical Bell numbers.

Theorem 1. *If F is a forest, then $\left[{F \atop k}\right] = \left\{{\overline{F} \atop k}\right\}$ for all k.*

Proof. Suppose F is a tree. Since forests are acyclic in the conventional definition of a cycle, i.e., cycles of length three or higher, the only cyclic partitions of F are those that partition F into 1-cycles and 2-cycles, i.e., vertices and edges respectively. Let σ be a cyclic partition of F into k cycles. For every edge $uv \in F$, let $I_{uv}(\sigma) \in \{0,1\}$ be the variable indicating whether uv is an edge in σ. On the other hand, in $\overline{F}$, the only independent sets are single vertices and pairs of none adjacent vertices (which are edges in F). So, $1 - I_{uv}(\sigma)$ would be the variable indicating whether the pair of vertices u and v form a single independent set or two independent sets in $\bar{F}$. Clearly, this mapping is reversible. Therefore, $\left[{F \atop k}\right] = \left\{{\overline{F} \atop k}\right\}$ for all k. □

Summing over k, we have:

Corollary 1. *If F is a forest, then $F! = B(\bar{F})$.* □

Theorem 2. *If G is a simple graph, then $\left[{G \atop k}\right] \geq \left\{{\overline{G} \atop k}\right\}$ for all k. Moreover, if the equality holds for all k, then G is a forest.*

Proof. Let G be a graph. Let A be a partition of $\overline{G}$ into k independent sets. If a part in A consists of a single vertex or a pair of vertices, then the contribution this part makes to the associated cyclic partitions of G (into k parts) would be a 1-cycle or a 2-cycle, respectively. On the other hand, if a part of A contains three or more vertices, say l, then the contribution this part makes to the associated cyclic partitions of G (into k parts) would be $(l-1)!$ many l-cycles. It is not hard to see that if A and B are two distinct partitions of $\overline{G}$ into k independent sets, their associated cyclic partitions of G into k parts are disjoint. Therefore, $\left[{G \atop k}\right] \geq \left\{{\overline{G} \atop k}\right\}$ for all k.

Now suppose G is not a forest. This means that G contains at least one cycle of length three or higher. Suppose C is a (conventional) cycle of shortest length in G. This means that C contains no chords. Let us denote the length of this cycle by l. Now let $k = 1 + (n - l)$. If $\{v_1, \ldots, v_{k-1}\}$ (possibly empty) is the set of vertices not in C, then consider the following partition of G into k cycles: $\sigma = \{C, \{v_1\}, \ldots, \{v_{k-1}\}\}$. On the one hand, σ contributes twice to $\left[{G \atop k}\right]$. On the other hand, σ's contribution to $\left\{{\overline{G} \atop k}\right\}$ is at most 1, when C is a cycle of length 3. This proves that $\left[{G \atop k}\right] > \left\{{\overline{G} \atop k}\right\}$ for $k = 1 + (n - l)$, where l is the length of the shortest cycle in G. □

Corollary 2. *If G is a simple graph, then $G! \geq B(\bar{G})$.* □

For F with c connected components, we have

$$\begin{bmatrix} F \\ k \end{bmatrix} = \sum_{k_1 + \ldots + k_c = k} \begin{bmatrix} F_1 \\ k_1 \end{bmatrix} \cdots \begin{bmatrix} F_c \\ k_c \end{bmatrix},$$

where $F_1, \ldots, F_c$ are the connected components of F. Because of this fact, we can recover information for the Stirling numbers of the first kind for forests by focusing their connected components which are trees. That said, the following theorem is easier to prove for forests than for trees, but we will use a specification of it in Corollary 3 below.

Theorem 3. *For any forest F of order n and any integer $n \geq k \geq \lceil \frac{n}{2} \rceil$*

$$\begin{bmatrix} F \\ k \end{bmatrix} \leq \begin{bmatrix} P_n \\ k \end{bmatrix}.$$

Proof. We proceed by strong induction. For the base case, there are two forests of order 2, namely $K_2 = P_2$ and E_2. In this case, $\left[{P_2 \atop 1}\right] = 1 \geq 0 = \left[{E_2 \atop 1}\right]$ and $\left[{P_2 \atop 2}\right] = 1 = \left[{E_2 \atop 2}\right]$.

More generally, for $n \geq 2$, $\left[{E_n \atop n}\right] = 1$ and $\left[{E_n \atop k}\right] = 0$ for all $n > k \geq \lceil \frac{n}{2} \rceil$. Since $\left[{P_n \atop n}\right] = 1$, we have the inequalities for when $F = E_n$.

For the inductive step, let F be any forests of order n that is not E_n. Let v be an arbitrary leaf of F

and uv be the only edge incident with v. We have

$$\begin{aligned}\begin{bmatrix} F \\ k \end{bmatrix} &= \begin{bmatrix} F - \{v\} \\ k-1 \end{bmatrix} + \begin{bmatrix} F - \{u,v\} \\ k-1 \end{bmatrix} \\ &\leq \begin{bmatrix} P_{n-1} \\ k-1 \end{bmatrix} + \begin{bmatrix} P_{n-2} \\ k-1 \end{bmatrix} \\ &= \begin{bmatrix} P_n \\ k \end{bmatrix},\end{aligned}$$

where the final equality comes from considering the behavior of an endpoint of the path. $\square$

Now we will find a lower bound for $\left[\begin{smallmatrix} T \\ k \end{smallmatrix}\right]$, where T is a tree of order n:

Theorem 4. *For any tree T of order n and any integer $n \geq k \geq \lceil \frac{n}{2} \rceil$,*

$$\begin{bmatrix} S_n \\ k \end{bmatrix} \leq \begin{bmatrix} T \\ k \end{bmatrix}.$$

Proof. We know that for any graph G of order n, $\left[\begin{smallmatrix} G \\ n \end{smallmatrix}\right]$ is 1 and $\left[\begin{smallmatrix} G \\ n-1 \end{smallmatrix}\right]$ is equal to the size of the graph. Therefore, $\left[\begin{smallmatrix} S_n \\ n \end{smallmatrix}\right] = 1$ and $\left[\begin{smallmatrix} S_n \\ n-1 \end{smallmatrix}\right] = n-1$, and $\left[\begin{smallmatrix} T \\ n \end{smallmatrix}\right] = n$ and $\left[\begin{smallmatrix} T \\ n-1 \end{smallmatrix}\right] = n-1$. Moreover, $\left[\begin{smallmatrix} S_n \\ k \end{smallmatrix}\right] = 0$ for $n-1 > k \geq \lceil \frac{n}{2} \rceil$ since every 2-cycle in S_n involves the center vertex which prevents S_n from having more than one 2-cycle. $\square$

Corollary 3. *For any tree T of order n and any integer $n \geq k \geq \lceil \frac{n}{2} \rceil$*

$$\begin{bmatrix} S_n \\ k \end{bmatrix} \leq \begin{bmatrix} T \\ k \end{bmatrix} \leq \begin{bmatrix} P_n \\ k \end{bmatrix}.$$

Stirling numbers of the first kind for a tree are parameterized by the number of edges in a vertex-disjoint matching in the tree. Therefore, $k = n-2$ is the first non-trivial and interesting case to consider. For the graphs on the two ends of tree spectrum, we have the following results which are obtained using simple combinatorial arguments involving the inclusion/exclusion principle:

Proposition 2. *For $n \geq 3$, we have*

- $\begin{bmatrix} S_n \\ n-2 \end{bmatrix} = 0;$
- $\begin{bmatrix} \widetilde{S}_n \\ n-2 \end{bmatrix} = n-3;$
- $\begin{bmatrix} P_n \\ n-2 \end{bmatrix} = \binom{n-1}{2} - (n-2);$
- $\begin{bmatrix} \widetilde{P}_n \\ n-2 \end{bmatrix} = \binom{n-1}{2} - (n-1).$

4 Degree Centrality

We will show that for trees, $C_{\deg}(G)$ has a closed form in the following theorem. This result was shown in Lemma 3.2, using a different proof, in a paper by Di Cerbo and Taylor [4] in the context of applying degree centrality in studying equity measures in finance.

Theorem 5. *Let T be a tree. If $m = n + 2 - \Delta(T)$, then*

$$C_{\deg}(T) = \frac{n^2 - m \cdot n + 2}{n^2 - 3n + 2}.$$

Proof. We will prove this result by inducing on m. When T is the star of order n, which is the only tree of order n with maximum degree of $n-1$, we have $m = 3$ and

$$C_{\deg}(S_n) = \frac{n^2 - 3n + 2}{n^2 - 3n + 2} = 1,$$

demonstrating the base case.

Now assume that the statement of this theorem is true for all tree T such that $m = k$. Let v_Δ be a vertex in T such that $\deg(v_\Delta) = \Delta(T)$. Note that v_Δ may not be unique but we can choose any vertex that satisfies this condition. Now let us assume that there is a leaf u in T such that u is not adjacent to v_Δ. If all the leaves in T are adjacent to v_Δ, then T is the star graph of order n, which is the base case. So there is a leaf u_1 in T that is adjacent to $u_2 \neq v_\Delta$. We will remove the u_1u_2 and add the edge u_2v_Δ, and call this new tree T'. In T', relative to T, the degree of v_Δ increases by one, the degree of u_2 decreases by one, and the degrees of the rest of the vertices remains the same. And $\Delta(T') = \deg_{T'}(v_\Delta) = \deg_T(v_\Delta) + 1 = \Delta(T) + 1$, and we have

$$C_{\deg}(T) = \frac{\sum_{i=1}^{n}(\deg_T(v_\Delta) - \deg_T(v_i))}{n^2 - 3n + 2} = \frac{\left(\sum_{v \neq u_2, v \neq v_\Delta} \deg_T(v_\Delta) - \deg_T(v)\right) + (\deg_T(v_\Delta) - \deg_T(u_2))}{n^2 - 3n + 2} =$$

$$\frac{\left(\sum_{v \neq u_2, v \neq v_\Delta} \deg_{T'}(v_\Delta) - 1 - \deg_{T'}(v)\right) + ((\deg_{T'}(v_\Delta) - 1) - (\deg_{T'}(u_2) + 1)}{n^2 - 3n + 2} =$$

$$\frac{\left(\sum_{v \neq u_2, v \neq v_\Delta} \deg_{T'}(v_\Delta) - \deg_{T'}(v)\right) - (n-2) + (\deg_{T'}(v_\Delta) - \deg_{T'}(u_2)) - 2}{n^2 - 3n + 2} =$$

$$\frac{(\sum_{i=1}^{n} \deg_{T'}(v_\Delta) - \deg_{T'}(v_i)) - n}{n^2 - 3n + 2} = \frac{n^2 - kn + 2 - n}{n^2 - 3n + 2} = \frac{n^2 - (k+1)n + 2}{n^2 - 3n + 2}.$$

Since $k = n + 2 - \Delta(T')$, we have $k + 1 = n + 2 - (\Delta(T') - 1) = n + 2 - \Delta(T)$, which finishes the proof. □

Corollary 4. *For any tree T of order n, we have*

$$C_{\deg}(P_n) \leq C_{\deg}(T) \leq C_{\deg}(S_n).$$

Moreover, since P_n and S_n are the only trees with maximum degree of 2 and $n-1$, respectively, the inequalities are sharp.

5 Closeness Centrality

Lemma 1. *Let G be a connected graph and let T be a spanning tree of G. Then $cls_G(v) \geq cls_T(v)$ for all v in $V = V(G) = V(T)$.*

Proof. For all $u, v \in V$, $d_G(u,v) \leq d_T(u,v)$. It follows that

$$\text{cls}_G(v) = \frac{n-1}{\sum_{u \in V} d_G(u,v)} \geq \frac{n-1}{\sum_{u \in V} d_T(u,v)} = \text{cls}_T(v)$$

and

$$\text{cls}_G(v^*) = \max_{v \in V} \text{cls}_G(v) \geq \max_{v \in V} \text{cls}_T(v) = \text{cls}_T(v^{**}).$$ □

Lemma 2. *Let G be a connected graph. There exists a spanning tree T of G such that*

$$\max_{v \in V} cls_G(v) = \max_{v \in V} cls_T(v).$$

Proof. Let v^* be the vertex in G such that $\mathrm{cls}_G(v^*) = \max_{v\in V} \mathrm{cls}_G(v)$. In other words, v^* is the vertex for which $\sum_{u\in V(G)} d_G(u,v)$ minimizes. We will construct a spanning tree as follows: First, we will include all the vertices that are in $N(v^*)$. These are vertices whose distance is one from v^*. Then, we include vertices whose distance from v^* is two, three, etc. without creating any cycles by connecting them to existing paths to v^*. The end result is a spanning tree of G.

Since $d_G(u,v^*) = d_T(u,v^*)$ for all $u \in V$,

$$\mathrm{cls}_G(v^*) = \sum_{u\in V(G)} d_G(u,v^*) = \sum_{u\in V(G)} d_T(u,v^*) = \mathrm{cls}_T(v^*)$$

and

$$\max_{v\in V} \mathrm{cls}_G(v) = C_G(v^*) = \mathrm{cls}_T(v^*) \leq \max_{v\in V} \mathrm{cls}_T(v).$$

Along with the inequality in Lemma 1, we have equality. $\square$

Lemma 3. *Let G be a connected graph. There exists a spanning tree T of G such that $C_{cls}(G) \leq C_{cls}(T)$.*

Proof. Let T be the spanning tree of G such that

$$\max_{v\in V} \mathrm{cls}_G(v) = \max_{v\in V} \mathrm{cls}_T(v).$$

Since $\mathrm{cls}_G(v) \geq \mathrm{cls}_T(v)$ for all v and $\mathrm{cls}_G(v^*) = \mathrm{cls}_T(v^*)$, we have

$$C_{\mathrm{cls}}(G) = \frac{\sum_{i=1}^{n}(\mathrm{cls}_G(v^*) - \mathrm{cls}_G(v_i))}{H_{\mathrm{cls}}} \leq \frac{\sum_{i=1}^{n}(\mathrm{cls}_T(v^*) - \mathrm{cls}_T(v_i))}{H_{\mathrm{cls}}} = \mathrm{cls}_T(G).$$ $\square$

This means that in order to find the graph that realizes H_{cls}, we can restrict our search to trees. It is easy to show that $H_{\mathrm{cls}}C_{\mathrm{cls}}(S_n) = (n-1)(n-2)/(2n-3)$. We will show that H_{cls} is realized by S_n and as a result $C_{\mathrm{cls}}(S_n) = 1$. For an intuitive justification of the upper bound in Theorem 6, see a paper by Freeman [10].

Theorem 6. *Let T be a tree of order n, where $n \geq 2$. Then*

$$C_{cls}(T) \leq C_{cls}(S_n).$$

Proof. The cases where $n = 2$ and $n = 3$ are trivial since $S_2 = P_2$ and $S_3 = P_3$ are the only trees of those orders. So, we assume that $n \geq 4$.

We will use induction on the number of non-leaf vertices, denoted of $N(T)$. When $N(T) = 1$, then $T = S_n$ and we are done. Let us assume that $N(T) > 1$. Let v^* be the vertex in T such that $\mathrm{cls}_T(v^*) = \max_{v\in V(T)} \mathrm{cls}_T(v)$. Let v' be a leaf that is not adjacent to v^* (if such a leaf does not exist, then $T = S_n$.) and $d_T(v^*,v') \leq d_T(v^*,u)$ for all $u \in V(T)$ (select the leaf with the maximum distance from v^*). Let w be v''s neighbor and L the set of leaves that are adjacent to w. It is not hard to see that all of w neighbors, except one which we denote by w', are leaves since $d_T(v^*,v') \leq d_T(v^*,u)$ for all $u \in V(T)$. It follows that $|V(L)| = \deg_T(w) - 1$. Clearly, the path from v' to v^* is $v', w, w', \ldots, v^*$, with the possibility of $w' = v^*$.

We will remove the edges connecting vertices in L with w and connect them with w'. We will denote the resulting tree by T'. By doing this process, w becomes a leaf in T'; hence, $N(T') = N(T) - 1$.

Here are how the distances between vertices are affected by changes from T to T':

- For $u, v \in V - L$, $d_{T'}(u,v) = d_T(u,v)$.
- For $u \in V - L - \{w, w'\}$ and $v \in L$, $d_{T'}(u,v) = d_T(u,v) - 1$.
- For $v \in L$, $d_{T'}(w,v) = d_T(w,v) + 1 = 2$ and $d_{T'}(w',v) = d_T(w',v) - 1 = 1$.
- For $u, v \in L$, $d_{T'}(u,v) = d_T(u,v) = 2$.

Let $r = \frac{\deg_T(w)-1}{n-1}$. Clearly, $0 < r < 1$. It is easy to check that for $v \neq w$,

$$\frac{1}{\mathrm{cls}_{T'}(v)} = \frac{\sum_{u\in V} d_{T'}(u,v)}{n-1} = \frac{\sum_{u\in V} d_T(u,v) - \deg_T(w) + 1}{n-1}.$$

It follows that

$$\frac{1}{\mathrm{cls}_{T'}(v)} = \frac{1}{\mathrm{cls}_T(v)} - \frac{\deg_T(v)+1}{n-1} = \frac{1}{\mathrm{cls}_T(v)} - r$$

As a result,

$$\mathrm{cls}_{T'}(v) = \frac{\mathrm{cls}_T(v)}{1 - r\mathrm{cls}_T(v)}.$$

Moreover,

$$\frac{1}{\mathrm{cls}_{T'}(w)} = \frac{\sum_{u\in V} d_{T'}(u,w)}{n-1} = \frac{\sum_{u\in V} d_T(u,w) + \deg_T(w) - 1}{n-1}.$$

Hence,

$$\frac{1}{\mathrm{cls}_{T'}(w)} = \frac{1}{\mathrm{cls}_T(w)} + \frac{\deg_T(w)-1}{n-1} = \frac{1}{\mathrm{cls}_T(w')} + r.$$

Therefore,

$$\mathrm{cls}_{T'}(w') = \frac{\mathrm{cls}_T(w')}{1 + r\mathrm{cls}_T(w')}.$$

Since $\mathrm{cls}_T(v^*) \geq \mathrm{cls}_T(v)$ for all $v \in V$, we have

$$0 < 1 - r\mathrm{cls}_T(v^*) \leq 1 - r\mathrm{cls}_T(v),$$

and for $v \neq w$,

$$\mathrm{cls}_{T'}(v^*) = \frac{\mathrm{cls}_T(v^*)}{1 - r\mathrm{cls}_T(v^*)} \geq \frac{\mathrm{cls}_{T'}(v)}{1 - r\mathrm{cls}_{T'}(v)} = \mathrm{cls}_{T'}(v)$$

On the other hand,

$$\mathrm{cls}_{T'}(w) = \frac{n-1}{\sum_{u\in V} d_{T'}(u,w)} = \frac{n-1}{\sum_{u\in V} d_T(u,w) + \deg_T(w) - 1} \leq \mathrm{cls}_T(w) \leq \mathrm{cls}_T(v^*) \leq \mathrm{cls}_{T'}(v^*)$$

This proves that $\mathrm{cls}_{T'}(v^*) = \max_{v\in V} \mathrm{cls}_{T'}(v)$. It follows that

$$\sum_{v\in V} \mathrm{cls}_{T'}(v^*) - \mathrm{cls}_{T'}(v) = \sum_{v\neq w} \frac{\mathrm{cls}_T(v^*)}{1 - r\mathrm{cls}_T(v^*)} - \frac{\mathrm{cls}_T(v)}{1 - r\mathrm{cls}_T(v)} + \left[\frac{\mathrm{cls}_T(v^*)}{1 - r\mathrm{cls}_T(v^*)} - \frac{\mathrm{cls}_T(w)}{1 + r\mathrm{cls}_T(w)}\right] \geq$$

$$\sum_{v\neq w} \frac{\mathrm{cls}_T(v^*)}{1 - r\mathrm{cls}_T(v^*)} - \frac{\mathrm{cls}_T(v)}{1 - r\mathrm{cls}_T(v^*)} + \left[\frac{\mathrm{cls}_T(v^*)}{1 - r\mathrm{cls}_T(v^*)} - \frac{\mathrm{cls}_T(w)}{1 + r\mathrm{cls}_T(w)}\right] \geq$$

$$\sum_{v\neq w} \frac{\mathrm{cls}_T(v^*) - \mathrm{cls}_T(v)}{1 - r\mathrm{cls}_T(v^*)} + [\mathrm{cls}_T(v^*) - \mathrm{cls}_T(w)] \geq$$

$$\sum_{v\neq w} \mathrm{cls}_T(v^*) - \mathrm{cls}_T(v) + [\mathrm{cls}_T(v^*) - \mathrm{cls}_T(w)] = \sum_{v\in V} \mathrm{cls}_T(v^*) - \mathrm{cls}_T(v).$$

This finishes the proof by induction. □

Next, we will look at the universal lower bound for trees of order n, which is realized by the path P_n.

Theorem 7. *Let T be a tree of order n, where $n \geq 2$. Then*

$$C_{cls}(P_n) \leq C_{cls}(T).$$

Proof. The proof is algorithmic: Remove the closest leaf to the central vertex and reattach it to the tree at one of the endpoints of a path whose length is the diameter of the original tree. This takes a leaf with minimal distance to the central vertex and increases the distance to the central vertex and, as a result, decreases the overall closeness measure. Note that this terminates when the algorithm reaches P_n. □

For an implementation of the algorithm in the proof of Theorem 7, see the script `tree_functions_1.py` and the Jupyter notebook `Ranking_Trees--Centrality_Bounds_1.0.ipynb` on our corresponding GitHub page.

Corollary 5. *Let T be a tree of order n, where $n \geq 2$. Then*

$$C_{cls}(P_n) \leq C_{cls}(T) \leq C_{cls}(S_n).$$

The closeness centrality of important graphs are as follows:

Proposition 3. *For $n \geq 3$, $H_{cls} = (n-1)(n-2)/(2n-3)$ and we have*

- $C_{cls}(S_n) = 1$;
- $C_{cls}(\widetilde{S}_n) = \dfrac{1}{H_{cls}} \dfrac{3n^3 - 14n^2 + 17n - 12}{6n(n-2)}$;
- *when n is even,*

$$C_{cls}(P_n) = \frac{1}{H_{cls}} \left(\frac{4(n-1)}{n} - \sum_{k=1}^{n} \frac{2(n-1)}{(n-k)^2 + (k-1)^2 + n - 1} \right); \tag{1}$$

- *when n is odd,*

$$C_{cls}(P_n) = \frac{1}{H_{cls}} \left(\frac{4n}{n+1} - \sum_{k=1}^{n} \frac{2(n-1)}{(n-k)^2 + (k-1)^2 + n - 1} \right); \tag{2}$$

- *when n is even,*

$$C_{cls}(\widetilde{P}_n) = \frac{1}{H_{cls}} \left(\frac{4(n-1)^2}{n^2 - 2n + 4} - 4 \sum_{k=1}^{n/2-1} \frac{n-1}{(n-k)^2 + (k-1)^2 + 1} - \frac{4(n-1)}{n^2 + 2n - 4} \right); \tag{3}$$

- *when n is odd,*

$$\begin{aligned} C_{cls}(\widetilde{P}_n) = \frac{1}{H_{cls}} \Bigg(& \frac{4n(n-1)}{n^2 - 2n + 5} - \sum_{k=1}^{(n-1)/2} \frac{2(n-1)}{(n-k)^2 + (k-1)^2} \\ & - \sum_{k=(n+1)/2}^{n-1} \frac{2(n-1)}{(n-k)^2 + (k+1)^2 - 2n + 2} - \frac{4(n-1)}{n^2 + 2n - 3} \Bigg). \end{aligned} \tag{4}$$

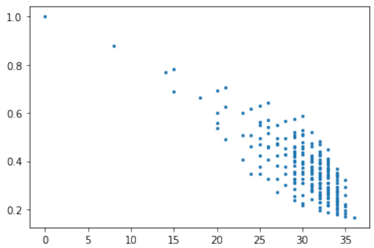
For proofs of (1)–(4), see Appendix B. It is shown in B that, for large enough n,

$$C_{\text{cls}}(P_n) \approx \frac{8 - 2\pi}{n} \quad \text{and} \quad C_{\text{cls}}(\widetilde{P}_n) \approx \frac{8 - 2\pi}{n}.$$

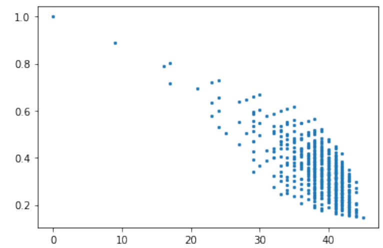
We also know from Proposition 2 that

$$\begin{bmatrix} P_n \\ n-2 \end{bmatrix} = \binom{n-1}{2} - (n-2) \quad \text{and} \quad \begin{bmatrix} \widetilde{P}_n \\ n-2 \end{bmatrix} = \binom{n-1}{2} - (n-1).$$

The line that goes through $(\left[\begin{smallmatrix} P_n \\ n-2 \end{smallmatrix}\right], C_{\text{cls}}(P_n))$ and $(\left[\begin{smallmatrix} \widetilde{P}_n \\ n-2 \end{smallmatrix}\right], C_{\text{cls}}(\widetilde{P}_n))$ is the line that goes through the lower-right-most two points in Figure 3a and in Figure 3b. So the lower bound for the points in each figure is

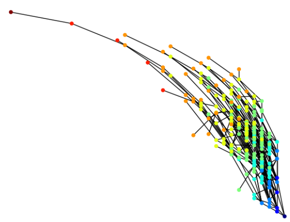


(a) $n = 11$

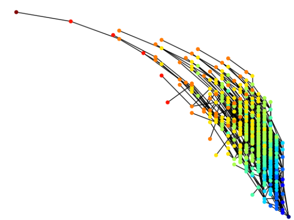


(b) $n = 12$

Figure 3: The scatterplot for the association between global closeness centrality and $\left[{T \atop n-2}\right]$, where T is a tree of order $n \in \{11, 12\}$



(a) $n = 11$



(b) $n = 12$

Figure 4: The scatter plot for the association between global betweenness centrality and $\left[{T \atop n-2}\right]$, where T is a tree of order $n \in \{11, 12\}$ with the embedded hierarchical structure based on the path-to-star algorithm; the colors in each subfigure represent the distance from S_{11} and S_{12}, respectively

bounded below by a line that asymptotically approaches the horizontal line $y = \frac{8-2\pi}{n}$. Moreover, the line that goes through the points $(\left[{S_n \atop n-2}\right], C_{\text{cls}}(S_n))$ and $(\left[{\widetilde{S}_n \atop n-2}\right], C_{\text{cls}}(\widetilde{S}_n))$ has a slope of

$$\frac{-7n^3 + 28n^2 - 51n + 36}{6n(n-1)(n-2)^2(n-3)}$$

and intercept of 1. This is the line that goes through the upper-left-most two points in Figure 3a and in Figure 3b. This line provides an upper bound for the points in each figure and we expect this to be true as n increases. These two observations provide us with a band for where $(\left[{T \atop n-2}\right], C_{\text{cls}}(T))$ is located for all non-isomorphic trees T of order n. The code in `Ranking_Trees--Checking_Bounds_1.0.ipynb` validates these results computationally for when n is between 7 and 12. Figures 4 and 5 contain the same scatter plots as do Figures 3a and 3b but with the embedded hierarchical structure based on the star-to-path and path-to-star algorithms. The colors in each figure represent the distance from S_n and P_n, respectively, for $n \in \{11, 12\}$. As these figures suggest, there are strong associations between the hierarchical structures interpolated from the path-to-star and star-to-path algorithms on the one hand and $\left[{T \atop n-2}\right]$ and $C_{\text{cls}}(T))$ on the other hand.

6 Betweenness Centrality

Theorem 8. *Let T be a tree of order n, where $n \geq 2$. Then*

$$C_{btw}(T) \leq C_{btw}(S_n).$$

Proof. The cases where $n = 2$ and $n = 3$ are trivial since $S_2 = P_2$ and $S_3 = P_3$ are the only trees of those orders. So, we assume that $n \geq 4$.

(a) $n = 11$ (b) $n = 12$

Figure 5: The scatter plot for the association between global betweenness centrality and $\left[{T \atop n-2}\right]$, where T is a tree of order $n \in \{11, 12\}$ with the embedded hierarchical structure based on the star-to-path algorithm; the colors in each subfigure represent the distance from P_{11} and P_{12}, respectively

We will use induction on the number of non-leaf vertices, denoted of $N(T)$. When $N(T) = 1$, then $T = S_n$ and we are done. Let us assume that $N(T) > 1$. Let v^* be the vertex in T such that $\mathrm{btw}_T(v^*) = \max_{v \in V(T)} \mathrm{btw}_T(v)$. Let v' be a leaf that is not adjacent to v^* (if such a leaf does not exist, then $T = S_n$) and $d_T(v^*, v') \le d_T(v^*, u)$ for all $u \in V(T)$ (select the leaf with the maximum distance from v^*). Let w be v''s neighbor and L the set of leaves that are adjacent to w. It is not hard to see that all of w neighbors, except one which we denote by w', are leaves since $d_T(v^*, v') \le d_T(v^*, u)$ for all $u \in V(T)$. It follows that $|V(L)| = \deg_T(w) - 1$. Clearly, the path from v' to v^* is $v', w, w', \ldots, v^*$, with the possibility of $w' = v^*$.

We will remove the edges connecting vertices in L with w and connect them with w'. We will denote the resulting tree by T'. By doing this process, w becomes a leaf in T'; hence, $N(T') = N(T) - 1$.

Here are how the betweenness centrality between vertices are affected by changes from T to T':

- We have $\mathrm{btw}_T(w) > 0$ while $\mathrm{btw}_{T'}(w) = 0$.
- For $v \neq w$, $\mathrm{btw}_T(v) = \mathrm{btw}_{T'}(v)$.

It follows that

$$C_{\mathrm{btw}}(T) = \frac{\sum_{v \in V}(\mathrm{btw}_T(v^*) - \mathrm{btw}_T(v))}{H_{\mathrm{btw}}} = \frac{\sum_{v \neq w}(\mathrm{btw}_T(v^*) - \mathrm{btw}_T(v)) + (\mathrm{btw}_T(v^*) - \mathrm{btw}_T(w))}{H_{\mathrm{btw}}} <$$

$$\frac{\sum_{v \neq w}(\mathrm{btw}_{T'}(v^*) - \mathrm{btw}_{T'}(v)) + (\mathrm{btw}_{T'}(v^*) - \mathrm{btw}_{T'}(w))}{H_{\mathrm{btw}}} = \frac{\sum_{v \in V}(\mathrm{btw}_{T'}(v^*) - \mathrm{btw}_{T'}(v))}{H_{\mathrm{btw}}} = C_{\mathrm{btw}}(T').$$

This finishes the proof by induction. □

Next, we will look at the universal lower bound for trees of order n, which is realized by the path P_n.

Theorem 9. *Let T be a tree of order n, where $n \ge 2$. Then*

$$C_{btw}(P_n) \le C_{btw}(T).$$

Proof. The proof is algorithmic: Remove the closest leaf to the central vertex and reattach it to the tree at one of the endpoints of a path whose length is the diameter of the original tree. This takes a leaf with minimal distance to the central vertex and decreases the number of shortest paths that goes through the central vertex, and as a result, decreases the overall betweenness measure. Note that this terminates when it reaches P_n. □

For an implementation of the algorithm in the proof of Theorem 9, see the script `tree_functions_1.py` and the Jupyter notebook `Ranking_Trees--Centrality_Bounds_1.0.ipynb` on our corresponding GitHub page.

Corollary 6. *Let T be a tree of order n, where $n \ge 2$. Then*

$$C_{btw}(P_n) \le C_{btw}(T) \le C_{btw}(S_n).$$

The betweenness centrality of important graphs are as follows:

Proposition 4. *For $n \geq 3$, $H_{btw} = n - 1$ and we have*

- $C_{btw}(S_n) = 1$;
- $C_{btw}(\widetilde{S}_n) = \dfrac{1}{H_{btw}} \left(\dfrac{n^3 - 4n^2 + n + 4}{(n-1)(n-2)} \right)$;
- *when n is even,*
$$C_{btw}(P_n) = \frac{1}{H_{btw}} \left(\frac{n(n+2)}{6(n-1)} \right); \tag{5}$$
- *when n is odd,*
$$C_{btw}(P_n) = \frac{1}{H_{btw}} \left(\frac{n(n+1)}{6(n-2)} \right); \tag{6}$$
- *when n is even,*
$$C_{btw}(\widetilde{P}_n) = \frac{1}{H_{btw}} \left(\frac{n^3 + 9n^2 - 28n + 12}{6(n-1)(n-2)} \right); \tag{7}$$
- *when n is odd,*
$$C_{btw}(\widetilde{P}_n) = \frac{1}{H_{btw}} \left(\frac{n^3 + 9n^2 - 31n + 9}{6(n-1)(n-2)} \right). \tag{8}$$

For proofs of (5)–(8), see Appendix C. It is easy to see, using (5)–(8), that

$$C_{\text{btw}}(P_n) \approx \frac{1}{6} \quad \text{and} \quad C_{\text{btw}}(\widetilde{P}_n) \approx \frac{1}{6}.$$

We also know from Proposition 2 that

$$\begin{bmatrix} P_n \\ n-2 \end{bmatrix} = \binom{n-1}{2} - (n-2) \quad \text{and} \quad \begin{bmatrix} \widetilde{P}_n \\ n-2 \end{bmatrix} = \binom{n-1}{2} - (n-1).$$

The line that goes through $(\left[\begin{smallmatrix} P_n \\ n-2 \end{smallmatrix}\right], C_{\text{btw}}(P_n))$ and $(\left[\begin{smallmatrix} \widetilde{P}_n \\ n-2 \end{smallmatrix}\right], C_{\text{btw}}(\widetilde{P}_n))$ is the line that goes through the lower-right-most two points in Figure 3a and in Figure 3b. So the lower bound for the points in each figure is bounded below by a line that asymptotically approaches the horizontal line $y = 1/6$. The line that goes through the points $(\left[\begin{smallmatrix} S_n \\ n-2 \end{smallmatrix}\right], C_{\text{btw}}(S_n))$ and $(\left[\begin{smallmatrix} \widetilde{S}_n \\ n-2 \end{smallmatrix}\right], C_{\text{btw}}(\widetilde{S}_n))$ has a slope of

$$\frac{-4n + 6}{(n-1)^2(n-2)(n-3)}$$

and intercept of 1. This is the line that goes through the upper-left-most line segment in Figure 6a and Figure 6b. This line provides an upper bound for the points in each figure and we expect this to be true as n increases. These two observations provide us with a band for where $(\left[\begin{smallmatrix} T \\ n-2 \end{smallmatrix}\right], C_{\text{btw}}(T))$ is located for all non-isomorphic trees T of order n. The code in `Ranking_Trees--Checking_Bounds_1.0.ipynb` validates these results computationally for when n is between 7 and 12. Figures 7 and 8 contain the same scatter plots as do Figures 6a and 6b but with the embedded hierarchical structure based on the star-to-path and path-to-star algorithms. The colors in each figure represent the distance from S_n and P_n, respectively, for $n \in \{11, 12\}$. As these figures suggest, there are strong associations between the hierarchical structures interpolated from the path-to-star and star-to-path algorithms on the one hand and $\left[\begin{smallmatrix} T \\ n-2 \end{smallmatrix}\right]$ and $C_{\text{btw}}(T))$ on the other hand.

(a) $n = 11$

(b) $n = 12$

Figure 6: The scatter plot between betweenness centrality and the $(n-2)$nd Stirling number of the first kind, i.e., $\left[{T \atop n-2}\right]$, when T is a tree of order $n \in \{11, 12\}$

(a) $n = 11$

(b) $n = 12$

Figure 7: The scatter plot for the association between global closeness centrality and $\left[{T \atop n-2}\right]$, where T is a tree of order $n \in \{11, 12\}$ with the embedded hierarchical structure based on the path-to-star algorithm; the colors in each subfigure represent the distance from S_{11} and S_{12}, respectively

(a) $n = 11$

(b) $n = 12$

Figure 8: The scatter plot for the association between global closeness centrality and $\left[{T \atop n-2}\right]$, where T is a tree of order $n \in \{11, 12\}$ with the embedded hierarchical structure based on the star-to-path algorithm; the colors in each subfigure represent the distance from P_{11} and P_{12}, respectively

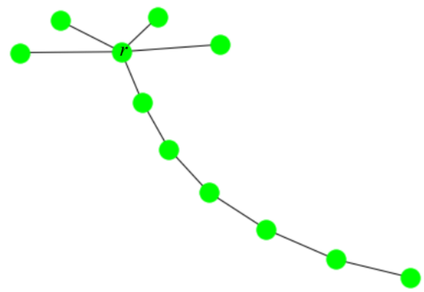


Figure 9: A rooted and unlabelled tree of order 11, where the root is indicated by r

7 Ordering Unrooted and Unlabelled Trees and Forests

In this section we will use the polynomials introduced in a paper by Liu [14] to define a total order structure on the set of unrooted and unlabelled trees.

For a rooted tree T_r, where r is the root, a primary subtree is a subtree S of T_r such that S has the same root as T_r and every leaf of T_r is either a leaf of S or is a descendant of a leaf of S. For a primary subtree S of T_r, let α_S be the number of leaves of S that are leaves in T_r as well and let β_S be the number of leaves of S that are internal vertices in T_r. By convention, the root r is considered an internal vertex even though it might be a leaf. The polynomial $P(T_r; x, y) = \sum_S x^{\alpha_S} y^{\beta_S}$, where the sum is over all primary subtrees of T_r, is the distinguishing polynomial for T_r. Liu shows that $P(.; x, y)$ is a complete isomorphism invariant for rooted unlabelled trees [14]. As pointed out by Liu [14], an alternative way of calculating $P(.; x, y)$ is to find the Dyck word (for information on Dyck words, see pages 22-25 of the monograph by Ghys [8]) associated with the rooted tree and placing an 'x' for every leave (a consecutive paired '(' and ')') and a '$+y$' for every interval vertex (a closed pair of parentheses). The Python script `Ranking_Trees--tree_functions_1.py` on our corresponding GitHub page contains a function `find_poly(T, r)` that uses this approach to compute $P(.; x, y)$ for rooted unlabelled trees.

For example, for the rooted and unlabelled tree in T_r Figure 9, where the root is indicated by the vertex r, the associated Dyck word is (((((((())))))()()()()), which gives us

$$((((((((x) + y) + y) + y) + y) + y)(x)(x)(x)(x) + y)$$

and

$$P(T_r; x, y) = x^4(x + 5y) + y = x^5 + 5x^4y + y.$$

Liu shows that for a rooted tree T_r, $P(T_r; x, y)$ is irreducible in $\mathbb{Z}[x, y]$ [14]. Now, suppose $\mathcal{T}_{r,n}$ be the set of rooted trees of order n. One can also see that for $T_r, T_r' \in \mathcal{T}_{r,n}$, $P(T_r; x, 1) - P(T_r'; x, 1)$ has a finite number of roots in $\mathbb{N}$. Let $R(T_r, T_r')$ be the set of such roots. The set

$$R(\mathcal{T}_{r,n}) = \bigcup_{T_r, T_r' \in \mathcal{T}_{r,n}} R(T_r, T_r')$$

is also finite and $\mathbb{N} - R(\mathcal{T}_{r,n})$ has a least element, say $\mu_{r,n}$. It follows that $P(T_r; \mu_{r,n}, 1) \neq P(T_r'; \mu_{r,n}, 1)$ for any $T_r, T_r' \in \mathcal{T}_{r,n}$ such that $T_r \not\simeq T_r'$.

For an unrooted and unlabelled tree T, $P(T; x, y)$ is the product of $P(T_1; x, y), \ldots, P(T_l; x, y)$, where l is the number of leaves in T and T_i is a rooted tree obtained from T by contracting an edge incident with a leave and declaring the resulting vertex the root of T_i—note that the order of each T_i is $n - 1$. Liu also shows that $P(.; x, y)$ is an isomorphism invariant polynomial for unrooted and unlabeled trees [14]. The Python script `Ranking_Trees--tree_functions_1.py` on our corresponding GitHub page contains a function `find_poly(T)` on our corresponding GitHub page uses this approach for unrooted and unlabelled trees.

For example, if we remove the root in the tree T_r in Figure 9, the distinguishing polynomial of the resulting unrooted and unlabelled tree T is

$$P(T; x, y) = (x^4 + 6y)(x^4 + 5x^3y + y)^4,$$

which can be rewritten as

$$\begin{aligned}P(T;x,y) =& x^{20}+20x^{19}y+150x^{18}y^2+500x^{17}y^3+625x^{16}y^4+10x^{16}y+180x^{15}y^2\\&+1200x^{14}y^3+3500x^{13}y^4+3750x^{12}y^5+30x^{12}y^2+420x^{11}y^3+1950x^{10}y^4\\&+3000x^9y^5+40x^8y^3+380x^7y^4+900x^6y^5+25x^4y^4+120x^3y^5+6y^5.\end{aligned} \tag{9}$$

We define the degree of a monomial $m(x,y) = x^iy^j$ as $i+j$ while we say the degree of $m(x,y)$ in x and y are i and j, respectively. The degree of a polynomial $p(x,y)$ is defined as the highest degree of the monomial terms and the degree of $p(x,y)$ in x and y are the highest degrees of the monomial terms in x and y, respectively. When we expand $P(T;x,y)$, we order the terms based on their degrees in a descending order, with the degrees in x descending and the degrees in y ascending when the degrees are equal among terms, as we see in (9). Note that the degree of $P(T;x,y)$ is $l(l-1)$, where l is the number of leaves, and this degree coincides with its degree in x. On the other hand, the degree of $P(T;x,y)$ in y is l since each $P(T_i;x,y)$ has degree 1 in y based on how distinguishing polynomials are constructed for rooted trees and that $P(T;x,y) = \prod_{i=1}^{l} P(T_i;x,y)$. Moreover, all the terms in $P(T;x,y)$ have a degree of at most l in y. Since there are at least two leaves in a tree and the only tree with two leaves is the path, the degree of $P(T;x,y)$ is $l(l-1)$.

One way to define a total ordering on trees of order n using $P(.;x,y)$ is based on the degrees in an ascending order and breaking the ties by comparing the degrees in x in an ascending order first and then the degrees in y in a descending order. For an implementation of an algorithm that creates this total ordering, see the function `get_total_list_degree_based(Tree_List)` in the Python script `tree_functions_1.py` and the Jupyter notebook `Ranking_Trees--Distinguishing_Polynomials_1.0.ipynb` on our corresponding GitHub page. We will call this the degree-based total ordering. It is not hard to show that the first two elements in the degree-based total ordering are

$$P(P_n;x,y) = (x+(n-2)y)^2 \text{ and } P(P_{n-1}+e;x,y) = (x^2+(n-2)y)(x^2+(n-3)xy+y)^2,$$

where $P_{n-1}+e$ is the tree of order n obtained from P_{n-1} by connecting a new leaf to one of its degree two vertices that already has a leaf as a neighbor, and the last two are

$$P(S_n;x,y) = (x^{n-2}+y)^{n-1} \text{ and } P(\widetilde{S}_n;x,y) = (x^{n-3}+2y)(x^{n-4}(x+y)+y)^{n-3}.$$

We can also define a total ordering by evaluating $\log_{10}(P(T;x,y))$ at appropriate values of x and y. Our approach is similar to the one we used for rooted and unlabelled trees: find $x=\mu_n$ for which $P(T;\mu_n,1) \neq P(T';\mu_n,1)$ for any unrooted and unlabeled trees T, T' of order n such that $T \not\simeq T'$. For example, we show computationally that for any $7 \le n \le 11$, $\mu_n = 2$ and $\mu_{12} = 3$ using the code in the Python script `tree_functions_1.py` (see the function `get_total_list_evaluation_based(tree_list, a = 2, b = 1)`) and the Jupyter notebook `Ranking_Trees--Distinguishing_Polynomials_1.0.ipynb` on our corresponding GitHub page. We call this method of ordering trees of fixed order, the evaluation-based total ordering.

8 Discussion

One can check that the total orderings obtained from the degree-based and valuation-based total orderings in Section 7 are not equivalent. This can be checked by running the code in the Jupyter notebook `Ranking_Trees--Distinguishing_Polynomials_1.0.ipynb` on our corresponding GitHub page. And even though the degree-based approach seems to be more natural as it considers all the information encoded in the distinguishing polynomial, the evaluation-based approach has the advantage that it has a single numeric output (rather than a polynomial) for each tree which make studying its relation with other graph statistics easier. For example, in Figure 10 and 11, we have visualized the association between $\log_{10}(P(T;x_0,1))$, radius(T), diam(T), $C_{\text{deg}}(T)$, $C_{\text{cls}}(T)$, $C_{\text{btw}}(T)$, and $\left[{T \atop n-2}\right]$ for trees T of order 11 when $x_0 = 2$ and of order 12 when $x_0 = 3$, respectively, in two pair-plots. As we see in the last rows of Figures 10 and 11, in addition to $\log_{10}(P(T;\mu_n,1))$ for $n \in \{11,12\}$ which gives us total orderings for trees of order n, we see strong associations between radius(T), diam(T), $C_{\text{deg}}(T)$, $C_{\text{cls}}(T)$, $C_{\text{btw}}(T)$, and leaf number as single predictors on the one hand and $\left[{T \atop n-2}\right]$ as the response variable on the other hand. Figures 12 and 13 in Appendix A show

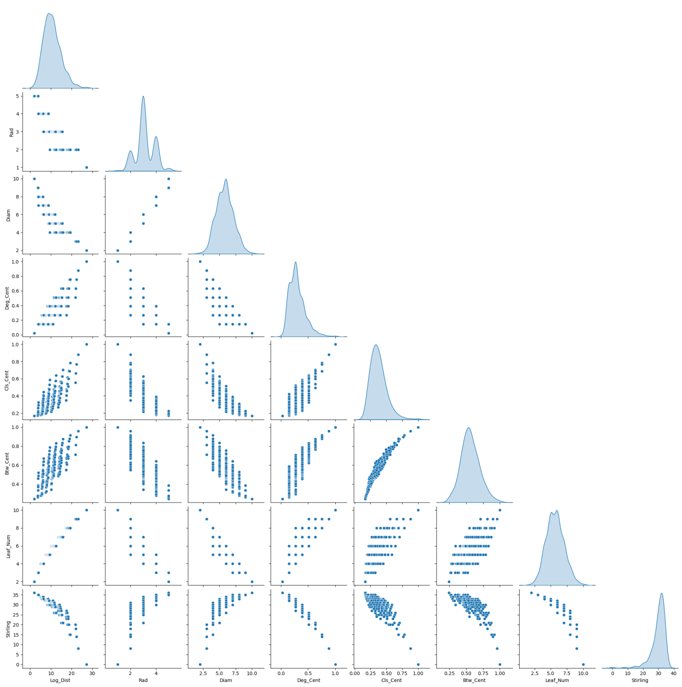


Figure 10: Pairplot for the association between $\log_{10}(P(T;2,1))$, radius(T), diam(T), $C_{\text{deg}}(T)$, $C_{\text{cls}}(T)$, $C_{\text{btw}}(T)$, leaf number, and $\left[{T \atop n-2}\right]$ for trees T of order $n = 11$

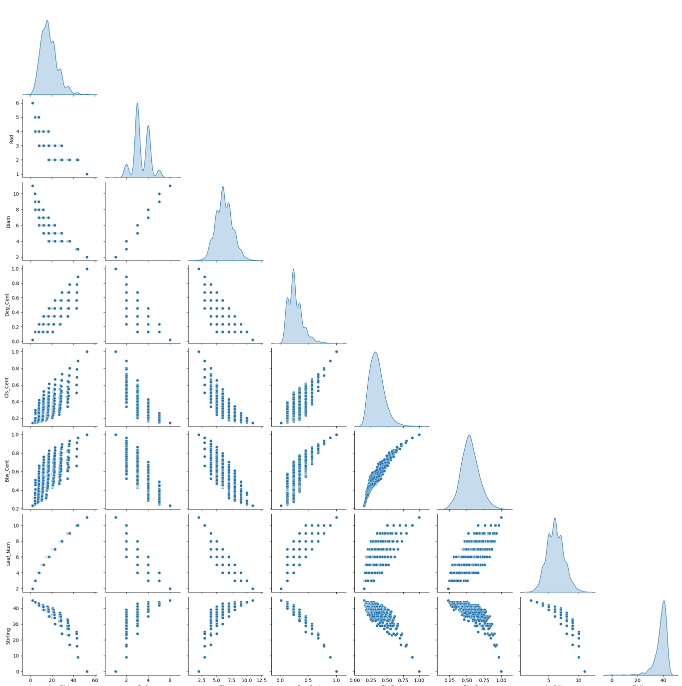


Figure 11: Pairplot for the association between $\log_{10}(P(T;3,1))$, radius(T), diam(T), $C_{\text{deg}}(T)$, $C_{\text{cls}}(T)$, $C_{\text{btw}}(T)$, leaf number, and $\left[{T \atop n-2}\right]$ for trees T of order $n = 12$

the same information but based on the number of leaves in each tree, which showcases the role of this graph statistics.

Since the total ordering that we obtain from these two methods are not equivalent, one natural question is how different the the total orderings are. On the one hand, $P(T;x,y)$ are ordered in an ascending order based on their degree that is $l(l-1)$ where l is the number of leaves in T. Since the number of leaves determines the degree of $P(T;x,y)$ and its degree in x and y, it plays an essential role in the degree-based total ordering. This role become less influential in the evaluation-based total ordering because letting $x=2$ (or $x=3$ when $n=12$) and $y=1$ changes the influence and contribution of each term in $P(T;x,y)$. Because of this observation, we will check if some structural similarities are preserved between the two approaches. In Figures 14, 15, and 16 in Appendix A, we have the scatter plots of measures of non-isomorphic trees of order n in the two total orderings when $n \in \{8,9,\ldots,12\}$ and $P(T;x,1)$ evaluated at $x \in \{2,3,4,5\}$ with colors based on the number of leaves in each tree. As indicated on each scatter plot, the correlation between the measures in these two total orderings does not decrease as we increase x when n is fixed. Moreover, when x is fixed and we increase n, the correlation between the measures in these total orderings decreases. Given the strong correlation that we see between the degree-based total ordering and the evaluation-based total ordering, we are confident that using the degree-based total ordering would have yielded similar pairplots as in Figure 10 and Figure 11 if we used $\log_{10}(P(T;\mu_n,1))$ as a quantifying statistic (and not as a ranking statistic that is used in the evaluation-based method).

If the leaf number is a loose indicator for whether a tree is more "star-like" or "path-like", we see from Figures 10 and 11 that $\log_{10}(P(T;\mu_n,1))$ provides an alternative but more granular method for classifying a tree as "star-like" or "path-like". We have explored this idea further in another paper [2].

9 Future Work

We will finish this paper with the following questions that we would like to address in the future:

- This paper raises natural questions relating the structural measures we compute to sampling versions or probabilistic bounds. One way to formulate this investigation would be to prove a theorem of the following form:

  Conjecture 1. *Let G be a graph of order n and T a uniformly random spanning tree of G. Select $m \leq \frac{n}{2}$ and $\ell \leq n$ and let $C : V(G) \to \mathbb{R}$ be a centrality measure. Then the probability that a uniformly selected m-matching of T contains the ℓ highest centrality values of $\{C(V(G))\}$ is*

  – *given by a function $f(G,C,m,\ell)$,*
  – *bounded by a function $g(m,\ell)$,*
  – *or approaches 1 as n goes to ∞.*

- An alternative approach to compare total orderings on trees obtained from degree-based and evaluation based methods is to apply the topologically motivated method of edge-smoothing before comparing the two total orderings. This operation appears to closely connect with the critical structural features of a tree of a fixed order to the centrality measure discussed in this paper, especially in the light of the critical role leaf numbers play in these two total ordering methods discussed in 7.

- The algorithms in Section 2 and their variants define a hierarchical structure on trees of a fixed order by defining $T_1 < T_2$ if T_2 is reachable by applying a sequence of path-to-star or star-to-path steps. This relationship is more nuanced and refined than the local version needed for the proofs in this paper with respect to relationships with other metrics. Exploring these connections further will lead to interesting additional research.

10 Code

Replication codes for the material discussed in this paper can be found on the corresponding GitHub page https://github.com/drdeford/Ranking_Trees

11 Acknowledgements

Amir Barghi wishes to thank the Office of the Dean of the Faculty at Saint Michael's College for the Merit-based Course Reduction Award in Fall 2020 that allowed him to continue his collaboration on this paper.

References

[1] Amir Barghi. Stirling numbers of the first kind for graphs. *Australasian Journal of Combinatorics*, 70:253–268, 01 2018.

[2] Amir Barghi and Daryl DeFord. Stirling numbers of uniform trees and related computational experiments. *Algorithms*, 16(5), 2023.

[3] Stephen P. Borgatti. Centrality and network flow. *Social Networks*, 27(1):55–71, 2005.

[4] Luca F. Di Cerbo and Stephen Taylor. Graph theoretical representations of equity indices and their centrality measures. *Quantitative Finance*, 21(4):523–537, 2021.

[5] Daryl DeFord. Counting Rearrangements on Generalized Wheel Graphs. *Fibonacci Quarterly*, 51(3):259–273, 2013.

[6] Daryl DeFord. Seating rearrangements on arbitrary graphs. *Involve: A Journal of Mathematics*, 7(6):787 – 805, 2014.

[7] Bryce Duncan and Rhodes Peele. Bell and stirling numbers for graphs. *Journal of Integer Sequences*, 7:2–3, 01 2009.

[8] Étienne Ghys. *A Singular Mathematical Promenade*. ENS Éditions, Lyon, France, 2017.

[9] Linton C. Freeman. A set of measures of centrality based on betweenness. *Sociometry*, 40(1):35–41, 1977.

[10] Linton C. Freeman. Centrality in social networks conceptual clarification. *Social Networks*, 1(3):215–239, 1978.

[11] Ross Honsberger. *In Pólya's footsteps*, volume 19 of *The Dolciani Mathematical Expositions*. Mathematical Association of America, Washington, DC, 1997. Miscellaneous problems and essays.

[12] R. Kennedy and C. Cooper. Variations on a 5×5 seating rearrangement problem. *Mathematics in College*, pages 59–67, 1993.

[13] Zsófia Kereskényi-Balogh and Gábor Nyul. Stirling numbers of the second kind and bell numbers for graphs. *The Australasian Journal of Combinatorics [electronic only]*, 58, 01 2014.

[14] Pengyu Liu. A tree distinguishing polynomial. *Discrete Applied Mathematics*, 288:1–8, 2021.

[15] T. Otake, R. Kennedy, and C. Cooper. On a seating rearrangement problem. *Mathematics and Informatics Quarterly*, pages 63–71, 1996.

A Figures

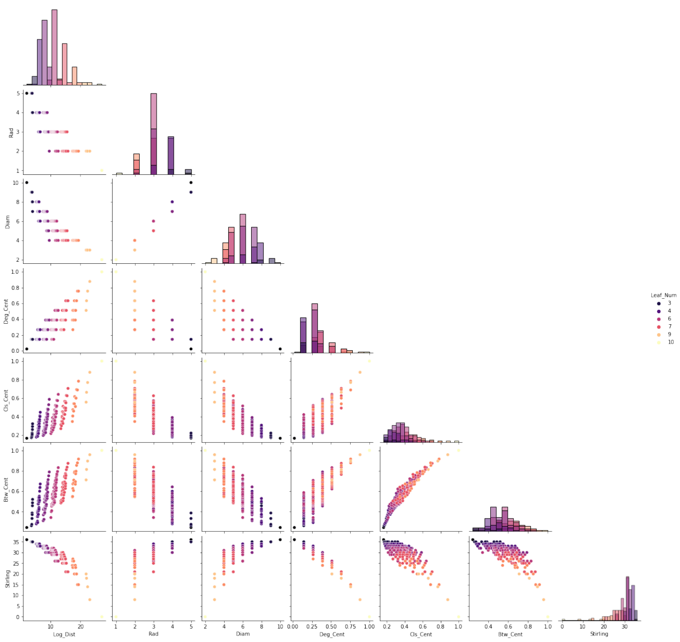


Figure 12: Pairplot for the association between $\log_{10}(P(T;2,1))$, $\text{radius}(T)$, $\text{diam}(T)$, $C_{\text{deg}}(T)$, $C_{\text{cls}}(T)$, $C_{\text{btw}}(T)$, and $\left[{T \atop n-2}\right]$ for trees T of order $n = 11$ with leaf number represented by color

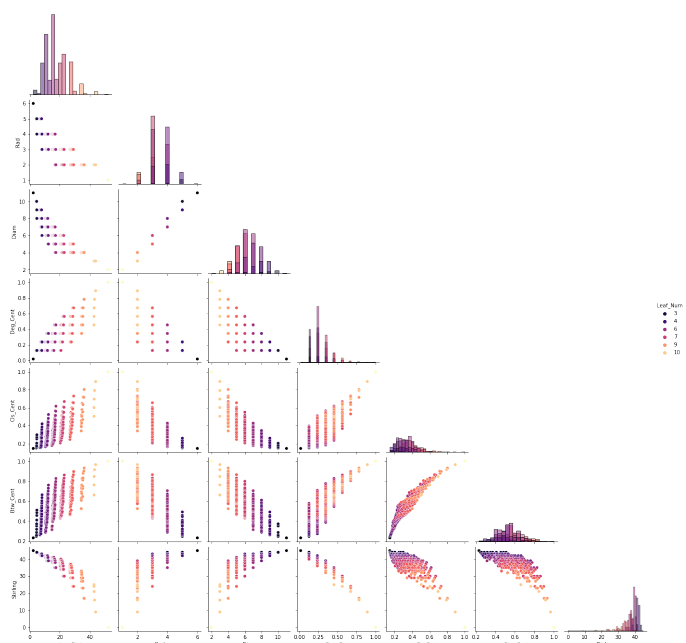
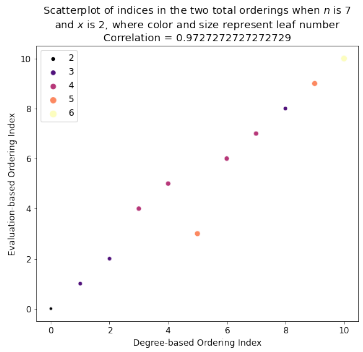
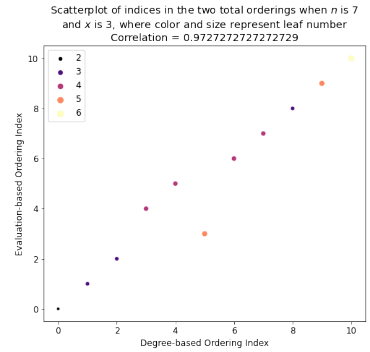


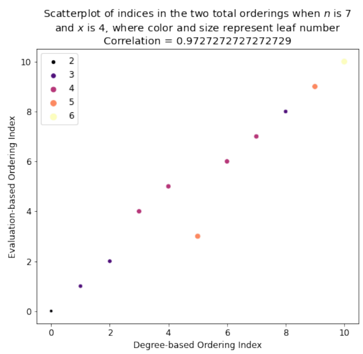
Figure 13: Pairplot for the association between $\log_{10}(P(T;3,1))$, $\text{radius}(T)$, $\text{diam}(T)$, $C_{\text{deg}}(T)$, $C_{\text{cls}}(T)$, $C_{\text{btw}}(T)$, and $\left[{T \atop n-2}\right]$ for trees T of order $n = 12$ with leaf number represented by color



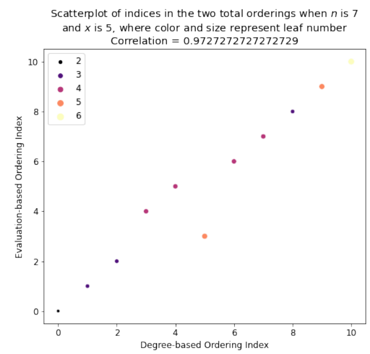
(a)



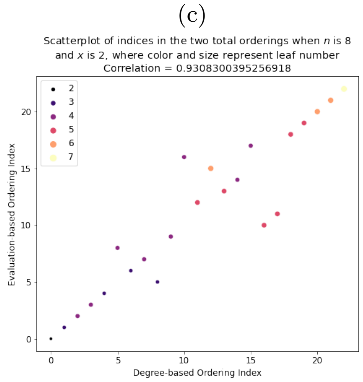
(b)



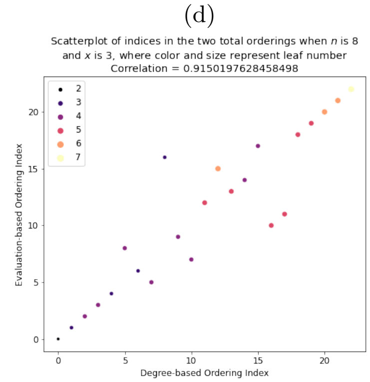
(c)



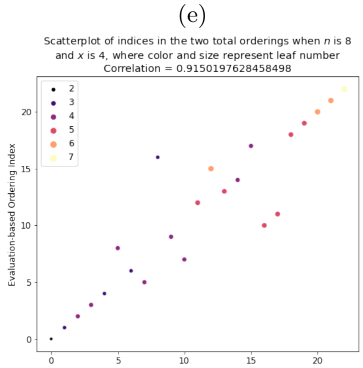
(d)



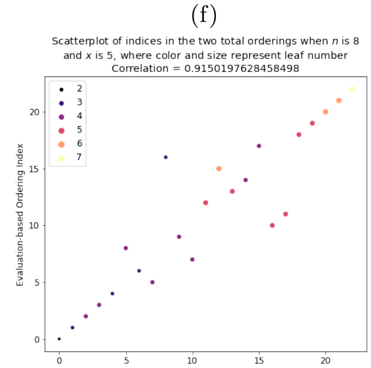
(e)



(f)

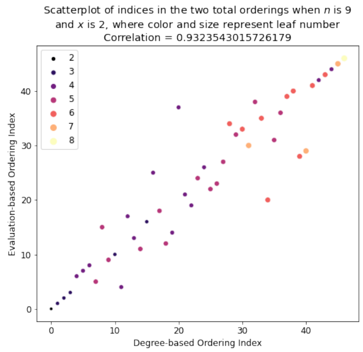


(g)

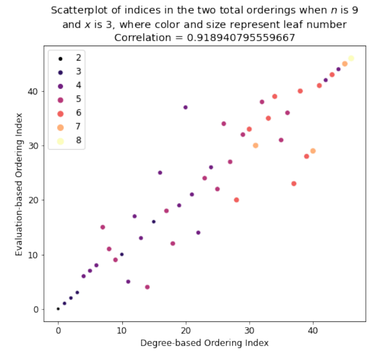


(h)

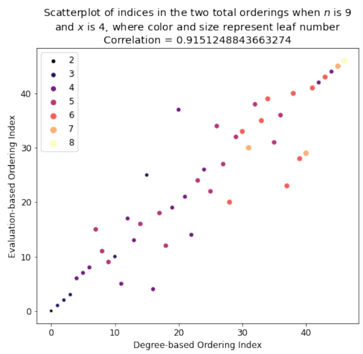
Figure 14: Scatter plots of measures of non-isomorphic trees of order n in the two total orderings when $n \in \{7, 8\}$ and $P(T; x, 1)$ evaluated at $x \in \{2, 3, 4, 5\}$ with color and size represent leaf number



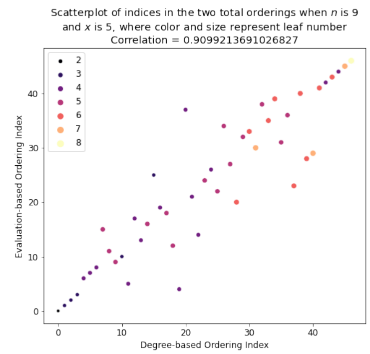
(a)



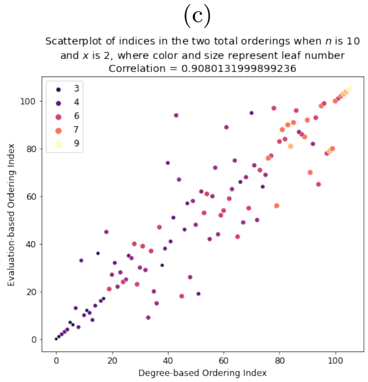
(b)



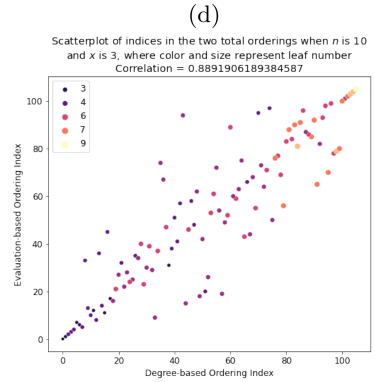
(c)



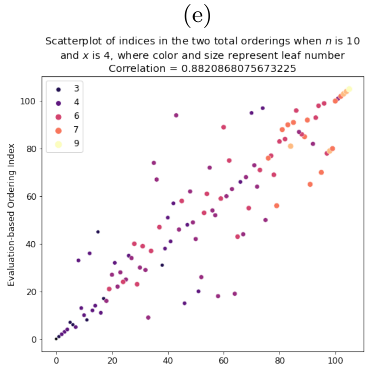
(d)



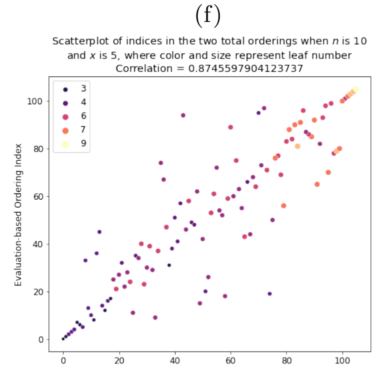
(e)



(f)

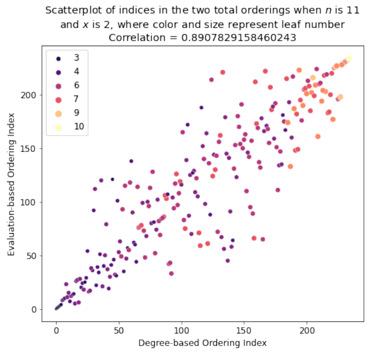


(g)

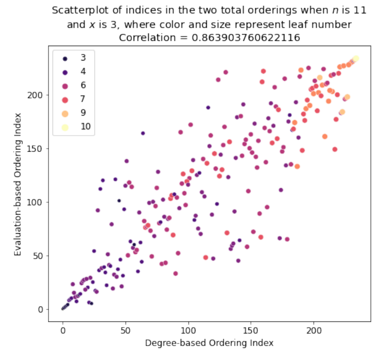


(h)

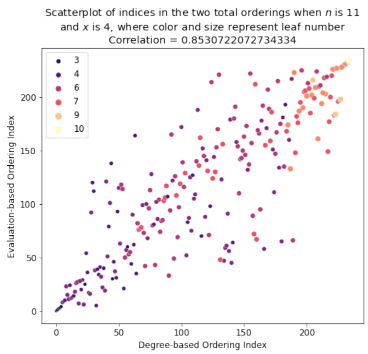
Figure 15: Scatter plots of measures of non-isomorphic trees of order n in the two total orderings when $n \in \{9, 10\}$ and $P(T; x, 1)$ evaluated at $x \in \{2, 3, 4, 5\}$ with color and size represent leaf number



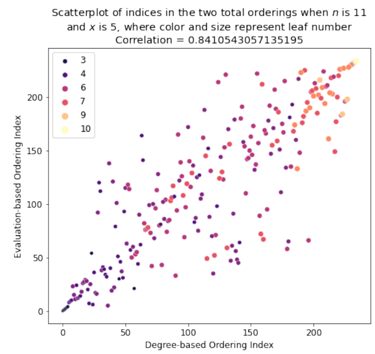
(a)



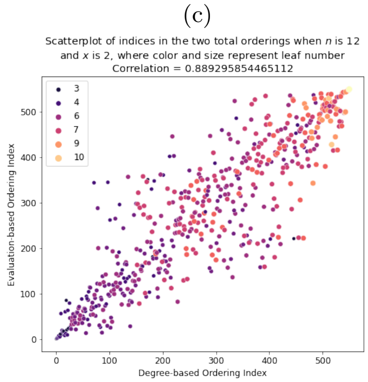
(b)



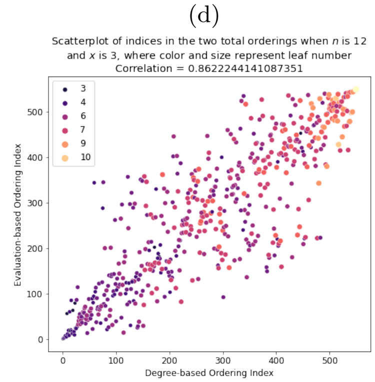
(c)



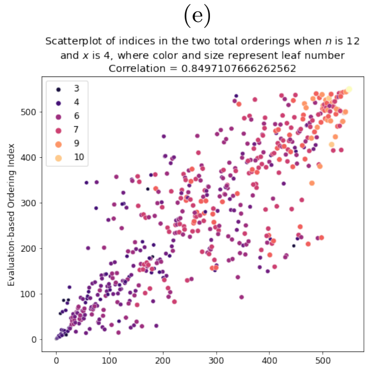
(d)



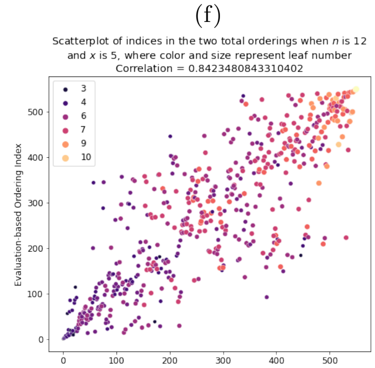
(e)



(f)



(g)



(h)

Figure 16: Scatter plots of measures of non-isomorphic trees of order n in the two total orderings when $n \in \{11, 12\}$ and $P(T; x, 1)$ evaluated at $x \in \{2, 3, 4, 5\}$ with color and size represent leaf number

B Closeness

In P_n, we label the vertices consequently from v_1 to v_n. For v_k and $u \neq v_k$, the distances $d(v_k, u)$ increase from 1 to $k-1$ in one direction and from 1 to $n-k$ in the other direction from v_k, and we have

$$\mathrm{cls}_{P_n}(v_k) = \frac{n-1}{\sum_{i=1}^{k-1} i + \sum_{i=1}^{n-k} i} = \frac{n-1}{\dfrac{k(k-1)}{2} + \dfrac{(n-k+1)(n-k)}{2}} = \frac{2(n-1)}{n(n+1) - 2k(n+1-k)}.$$

It is not hard to check that $\mathrm{cls}_{P_n}(v_k)$ maximizes when k is about $n/2$.

When n is odd, the center of the graph consists of a single vertex v_k, where $k = \frac{n+1}{2}$. This is also the vertex for which $C_{P_n}(v_k)$ is maximized and we will denote it by v^*. It follows that

$$\mathrm{cls}_{P_n}(v^*) = \frac{n-1}{2\sum_{i=1}^{m} i},$$

where m is $\frac{n-1}{2}$. Therefore,

$$\mathrm{cls}_{P_n}(v^*) = \frac{n-1}{\left(\dfrac{n-1}{2}\right)\left(\dfrac{n+1}{2}\right)} = \frac{4}{n+1}.$$

As a result,

$$H_{\mathrm{cls}} C_{\mathrm{cls}}(P_n) = \sum_{k=1}^{n} (\mathrm{cls}_{P_n}(v^*) - \mathrm{cls}_{P_n}(v_k)) = \sum_{k=1}^{n} \frac{4}{n+1} - \frac{2(n-1)}{n(n+1) - 2k(n+1-k)},$$

where $H_{\mathrm{cls}} = (n-1)(n-2)/(2n-3)$. Consequently,

$$H_{\mathrm{cls}} C_{\mathrm{cls}}(P_n) = \frac{4n}{n+1} - \frac{2(n-1)}{n+1} \sum_{k=1}^{n} \frac{\dfrac{1}{n}}{1 - 2\dfrac{k}{n}\left(1 - \dfrac{k}{n+1}\right)}$$

When n is large enough

$$\sum_{k=1}^{n} \frac{\dfrac{1}{n}}{1 - 2\dfrac{k}{n}\left(1 - \dfrac{k}{n+1}\right)} \approx \int_0^1 \frac{1}{1 - 2x + 2x^2}\, dx = \frac{\pi}{2}$$

and

$$C_{\mathrm{cls}}(P_n) \approx \frac{8 - 2\pi}{n}.$$

When n is even, P_n has two vertices in its center, namely vertices indexed by $n/2$ and $n/2+1$. If we denote either of them by v^*, we have

$$\mathrm{cls}_{P_n}(v^*) = \frac{n-1}{\sum_{i=1}^{n/2-1} i + \sum_{i=1}^{n/2} i} = \frac{4(n-1)}{n^2}.$$

Similar to what we did above for when n is odd, we can show that

$$C_{\mathrm{cls}}(P_n) \approx \frac{8 - 2\pi}{n}$$

when n is even and large enough.

Let n be even. In $\widetilde{P}_n$, we label the vertices consequently from v_1 to v_{n-1}, with v_n being the leave in the middle of the tree adjacent to $v_{n/2}$. For $1 \leq k \leq n-1$ and $u \neq v_k$ and $u \neq v_n$, the distances $d(v_k, u)$ increase from 1 to $k-1$ in one direction and from 1 to $n-1-k$ in the other direction from v_k. Moreover,

$d(v_k, v_n) = 1 + d(v_k, v_{n/2})$. So for $1 \le k \le n/2$, $d(v_k, v_n) = n/2 - k + 1$ and for $n/2 < k \le n-1$, $d(v_k, d_n) = 1 + k - n/2$. As a result, for $1 \le k \le n/2$, we have

$$\mathrm{cls}_{\widetilde{P}_n}(v_k) = \frac{n-1}{1 + n/2 - k + \sum_{i=1}^{k-1} i + \sum_{i=1}^{n-1-k} i} =$$

$$\frac{n-1}{1 + n/2 - k + \dfrac{k(k-1)}{2} + \dfrac{(n-k-1)(n-k)}{2}} =$$

$$\frac{2(n-1)}{(n-k)^2 + (k-1)^2 + 1}.$$

On the other hand, for $n/2 < k \le n-1$, we have

$$\mathrm{cls}_{\widetilde{P}_n}(v_k) = \frac{n-1}{1 + k - n/2 + \sum_{i=1}^{k-1} i + \sum_{i=1}^{n-1-k} i} =$$

$$\frac{n-1}{1 + k - n/2 + \dfrac{k(k-1)}{2} + \dfrac{(n-k-1)(n-k)}{2}} =$$

$$\frac{2(n-1)}{(n-k)^2 + (k+1)^2 - 2n + 1}.$$

Finally,

$$\mathrm{cls}_{\widetilde{P}_n}(v_n) = \frac{n-1}{1 + 2\sum_{i=2}^{n/2} i} = \frac{4(n-1)}{n^2 + 2n - 4}$$

It is not hard to check that $\mathrm{cls}_{\widetilde{P}_n}(v_k)$ maximizes when k is $n/2$ and we have

$$\mathrm{cls}_{\widetilde{P}_n}(v^*) = \mathrm{cls}_{\widetilde{P}_n}(v_{n/2}) = \frac{4(n-1)}{n^2 - 2n + 4}.$$

It follows that

$$H_{\mathrm{cls}} C_{\mathrm{cls}}(\widetilde{P}_n) = \sum_{k=1}^{n} (\mathrm{cls}_{\widetilde{P}_n}(v^*) - \mathrm{cls}_{\widetilde{P}_n}(v_k)) =$$

$$\frac{4(n-1)^2}{n^2 - 2n + 4} - \sum_{k=1}^{n/2-1} \frac{2(n-1)}{(n-k)^2 + (k-1)^2 + 1} - \sum_{k=n/2+1}^{n-1} \frac{2(n-1)}{(n-k)^2 + (k+1)^2 - 2n + 1} - \frac{4(n-1)}{n^2 + 2n - 4},$$

where $H = (n-1)(n-2)/(2n-3)$. By symmetry, we have

$$H_{\mathrm{cls}} C_{\mathrm{cls}}(\widetilde{P}_n) = \frac{4(n-1)^2}{n^2 - 2n + 4} - 4\sum_{k=1}^{n/2-1} \frac{n-1}{(n-k)^2 + (k-1)^2 + 1} - \frac{4(n-1)}{n^2 + 2n - 4},$$

and

$$H_{\mathrm{cls}} C_{\mathrm{cls}}(\widetilde{P}_n) = \frac{4(n-1)^2}{n^2 - 2n + 4} - 4\sum_{k=1}^{n/2-1} \frac{n-1}{n^2 + 2 - 2k(n+1-k)} - \frac{4(n-1)}{n^2 + 2n - 4},$$

Consequently,

$$H_{\mathrm{cls}} C_{\mathrm{cls}}(\widetilde{P}_n) = \frac{4(n-1)^2}{n^2 - 2n + 4} - \frac{4(n-1)}{n} \sum_{k=1}^{n/2-1} \frac{\dfrac{1}{n}}{1 + \dfrac{2}{n^2} - 2\dfrac{k}{n}\left(1 - \dfrac{k-1}{n}\right)} - \frac{4(n-1)}{n^2 + 2n - 4}$$

When n is large enough,

$$\sum_{k=1}^{n/2-1} \frac{\frac{1}{n}}{1+\frac{2}{n^2}-2\frac{k}{n}\left(1-\frac{k-1}{n}\right)} \approx \int_0^{1/2} \frac{1}{1-2x+2x^2}\,dx = \frac{\pi}{4}$$

and

$$C_{\mathrm{cls}}(\widetilde{P}_n) \approx \frac{8-2\pi}{n}.$$

Now let n be odd. In $\widetilde{P}_n$, we label the vertices consequently from v_1 to v_{n-1}, with v_n being the leave in the middle of the tree adjacent to $v_{(n-1)/2}$. For $1 \le k \le n-1$ and $u \ne v_k$ and $u \ne v_n$, the distances $d(v_k, u)$ increase from 1 to $k-1$ in one direction and from 1 to $n-1-k$ in the other direction from v_k. Moreover, $d(v_k, v_n) = 1 + d(v_k, v_{(n-1)/2})$. So for $1 \le k \le (n-1)/2$, $d(v_k, v_n) = (n-1)/2 - k + 1$ and for $(n-1)/2 < k \le n-1$, $d(v_k, d_n) = 1 + k - (n-1)/2$. As a result, for $1 \le k \le (n-1)/2$, we have

$$\mathrm{cls}_{\widetilde{P}_n}(v_k) = \frac{n-1}{1+(n-1)/2-k+\sum_{i=1}^{k-1} i + \sum_{i=1}^{n-1-k} i} =$$

$$\frac{n-1}{1+(n-1)/2-k+\dfrac{k(k-1)}{2}+\dfrac{(n-k-1)(n-k)}{2}} =$$

$$\frac{2(n-1)}{(n-k)^2+(k-1)^2}.$$

On the other hand, for $(n-1)/2 < k \le n-1$, we have

$$\mathrm{cls}_{\widetilde{P}_n}(v_k) = \frac{n-1}{1+k-(n-1)/2+\sum_{i=1}^{k-1} i + \sum_{i=1}^{n-1-k} i} =$$

$$\frac{n-1}{1+k-(n-1)/2+\dfrac{k(k-1)}{2}+\dfrac{(n-k-1)(n-k)}{2}} =$$

$$\frac{2(n-1)}{(n-k)^2+(k+1)^2-2n+2}.$$

Finally,

$$\mathrm{cls}_{\widetilde{P}_n}(v_n) = \frac{n-1}{1+\sum_{i=2}^{(n-1)/2} i + \sum_{i=2}^{(n+1)/2} i} = \frac{4(n-1)}{n^2+2n-3}$$

It is not hard to check that $C_{P_{n-1}+e}(v_k)$ maximizes when k is $(n-1)/2$ and we have

$$\mathrm{cls}_{\widetilde{P}_n}(v^*) = \mathrm{cls}_{\widetilde{P}_n}(v_{(n-1)/2}) = \frac{4(n-1)}{n^2-2n+5}.$$

It follows that

$$H_{\mathrm{cls}} C_{\mathrm{cls}}(\widetilde{P}_n) = \sum_{k=1}^{n} (\mathrm{cls}_{\widetilde{P}_n}(v^*) - \mathrm{cls}_{\widetilde{P}_n}(v_k)) =$$

$$\frac{4n(n-1)}{n^2-2n+5} - \sum_{k=1}^{(n-1)/2} \frac{2(n-1)}{(n-k)^2+(k-1)^2} - \sum_{k=(n+1)/2}^{n-1} \frac{2(n-1)}{(n-k)^2+(k+1)^2-2n+2} - \frac{4(n-1)}{n^2+2n-3},$$

where $H_{\mathrm{cls}} = (n-1)(n-2)/(2n-3)$. On one hand, we have

$$\sum_{k=1}^{(n-1)/2} \frac{2(n-1)}{(n-k)^2+(k-1)^2} = \sum_{k=1}^{(n-1)/2} \frac{2(n-1)}{n^2+1-2k(n+1-k)}$$

$$= \frac{2(n-1)}{n} \sum_{k=1}^{(n-1)/2} \frac{\frac{1}{n}}{1 + \frac{1}{n^2} - 2\frac{k}{n}\left(1 - \frac{k-1}{n}\right)}$$

When n is large enough,

$$\sum_{k=1}^{(n-1)/2} \frac{\frac{1}{n}}{1 + \frac{1}{n^2} - 2\frac{k}{n}\left(1 - \frac{k-1}{n}\right)} \approx \int_0^{1/2} \frac{1}{1 - 2x + 2x^2}\, dx = \frac{\pi}{4}.$$

On the other hand, with a change of variable, we have

$$\sum_{k=(n+1)/2}^{n-1} \frac{2(n-1)}{(n-k)^2 + (k+1)^2 - 2n + 2} = \sum_{k=1}^{(n-1)/2} \frac{2(n-1)}{(n-k)^2 + (k-1)^2 + 2},$$

which is equal to

$$\frac{2(n-1)}{n} \sum_{k=1}^{(n-1)/2} \frac{\frac{1}{n}}{1 + \frac{3}{n^2} - 2\frac{k}{n}\left(1 - \frac{k+1}{n}\right)}.$$

When n is large enough,

$$\sum_{k=1}^{(n-1)/2} \frac{\frac{1}{n}}{1 + \frac{3}{n^2} - 2\frac{k}{n}\left(1 - \frac{k+1}{n}\right)} \approx \int_0^{1/2} \frac{1}{1 - 2x + 2x^2}\, dx = \frac{\pi}{4}.$$

and

$$C_{\text{cls}}(\widetilde{P}_n) \approx \frac{8 - 2\pi}{n}.$$

C Betweenness

In P_n, we label the vertices consequently from v_1 to v_n. For v_k, only v_i and v_j, where $i > k$ and $j > k$ contribute to a nonzero term to $C_{P_n}(v_k)$. It follows that

$$\text{btw}_{P_n}(v_k) = \frac{(k-1)(n-k)}{\binom{n-1}{2}},$$

and as a result, we have

$$\sum_{i=1}^{n} \text{btw}_{P_n}(v_i) = \sum_{i=1}^{n} \frac{(i-1)(n-i)}{\binom{n-1}{2}} = \frac{1}{\binom{n-1}{2}} \sum_{i=1}^{n} \left(n(i-1) + i - i^2\right)$$

$$= \frac{1}{\binom{n-1}{2}} \left(\frac{n^2(n-1)}{2} + \frac{n(n+1)}{2} - \frac{n(n+1)(2n+1)}{6} \right) = \frac{n}{3}.$$

When n is odd, $\text{btw}_{P_n}(v_k) = \frac{(k-1)(n-k)}{\binom{n-1}{2}}$ is maximized when $k = \frac{n+1}{2}$. We denote this vertex by v^* which is also the only vertex in the center of P_n. Since $\text{btw}_{P_n}(v^*) = \frac{n-1}{2(n-2)}$, we have

$$H_{\text{btw}} C_{\text{btw}}(P_n) = \sum_{i=1}^{n} (\text{btw}_{P_n}(v^*) - \text{btw}_{P_n}(v_i)) = \frac{n(n-1)}{2(n-2)} - \frac{n}{3} = \frac{n(n+1)}{6(n-2)}.$$

When n is even, $\mathrm{btw}_{P_n}(v_k) = \frac{(k-1)(n-k)}{\binom{n-1}{2}}$ is maximized when $k = \frac{n}{2}$ and $k = \frac{n}{2}+1$. We can denote either of these vertices, which constitute the center of P_n, by v^*. Since $\mathrm{btw}_{P_n}(v^*) = \frac{n}{2(n-1)}$, we have

$$H_{\mathrm{btw}}C_{\mathrm{btw}}(P_n) = \sum_{i=1}^{n}(\mathrm{btw}_{P_n}(v^*) - \mathrm{btw}_{P_n}(v_i)) = \frac{n^2}{2(n-1)} - \frac{n}{3} = \frac{n(n+2)}{6(n-1)}.$$

Now, let us consider $\widetilde{P}_n$. Let n be even. In $\widetilde{P}_n$, we label the vertices consequently from v_1 to v_{n-1}, with v_n being the leave in the middle of the tree adjacent to $v_{n/2}$. For $1 \le k < n/2$,

$$\mathrm{btw}_{\widetilde{P}_n}(v_k) = \frac{(k-1)(n-1-k)+k-1}{\binom{n-1}{2}} = \frac{(k-1)(n-k)}{\binom{n-1}{2}},$$

and for $n/2 < k \le n-1$,

$$\mathrm{btw}_{\widetilde{P}_n}(v_k) = \frac{(k-1)(n-1-k)+n-1-k}{\binom{n-1}{2}} = \frac{k(n-1-k)}{\binom{n-1}{2}}.$$

Moreover, $\mathrm{btw}_{\widetilde{P}_n}(v_n) = 0$ and

$$\mathrm{btw}_{\widetilde{P}_n}(v_{n/2}) = \frac{(n/2-1)^2+2(n/2-1)}{\binom{n-1}{2}} = \frac{n^2-4}{4\binom{n-1}{2}}.$$

It is not hard to see that $\mathrm{btw}_{\widetilde{P}_n}(v_k)$ maximizes when $k = n/2$. It follows that

$$\begin{aligned}
\sum_{i=1}^{n}\mathrm{btw}_{\widetilde{P}_n}(v_i) &= \sum_{i=1}^{n-1}\frac{(i-1)(n-1-i)}{\binom{n-1}{2}} + \sum_{i=1}^{n/2}\frac{i-1}{\binom{n-1}{2}} + \sum_{i=n/2}^{n-1}\frac{n-1-i}{\binom{n-1}{2}} \\
&= \sum_{i=1}^{n-1}\frac{(i-1)(n-1-i)}{\binom{n-1}{2}} + 2\sum_{i=1}^{n/2}\frac{i-1}{\binom{n-1}{2}} \\
&= \frac{1}{\binom{n-1}{2}}\sum_{i=1}^{n-1}\left(ni-(n-1)-i^2\right) + 2\sum_{i=1}^{n/2}\frac{i-1}{\binom{n-1}{2}} \\
&= \frac{1}{\binom{n-1}{2}}\left(\frac{n^2(n-1)}{2} - (n-1)^2 - \frac{(n-1)n(2n-1)}{6}\right) + \frac{(n/2-1)n/2}{\binom{n-1}{2}} \\
&= \frac{2n^3-9n^2+16n-12}{12\binom{n-1}{2}}.
\end{aligned}$$

Since $\mathrm{btw}_{\widetilde{P}_n}(v_k)$ is maximized when $k = \frac{n}{2}$, we denote the vertex $v_{n/2}$ by v^*, and we have

$$H_{\mathrm{btw}}C_{\mathrm{btw}}(\widetilde{P}_n) = \sum_{i=1}^{n}(\mathrm{btw}_{\widetilde{P}_n}(v^*) - \mathrm{btw}_{\widetilde{P}_n}(v_i)) = \frac{n(n^2-4)}{4\binom{n-1}{2}} - \frac{2n^3-9n^2+16n-12}{12\binom{n-1}{2}} = \frac{n^3+9n^2-28n+12}{6(n-1)(n-2)}.$$

Now let n be odd. In $\widetilde{P}_n$, we label the vertices consequently from v_1 to v_{n-1}, with v_n being the leave in the middle of the tree adjacent to $v_{(n-1)/2}$. For $1 \le k < (n-1)/2$,

$$\mathrm{btw}_{\widetilde{P}_n}(v_k) = \frac{(k-1)(n-1-k)+k-1}{\binom{n-1}{2}} = \frac{(k-1)(n-k)}{\binom{n-1}{2}},$$

and for $(n-1)/2 < k \le n-1$,

$$\mathrm{btw}_{\widetilde{P}_n}(v_k) = \frac{(k-1)(n-1-k)+n-1-k}{\binom{n-1}{2}} = \frac{k(n-1-k)}{\binom{n-1}{2}}.$$

Moreover, $\mathrm{btw}_{\widetilde{P}_n}(v_n) = 0$ and

$$\mathrm{btw}_{\widetilde{P}_n}(v_{(n-1)/2}) = \frac{((n-1)/2-1)((n-1)/2)+(n-1)/2-1+(n-1)/2}{\binom{n-1}{2}} = \frac{n^2-5}{4\binom{n-1}{2}}.$$

It is not hard to see that $\mathrm{btw}_{\widetilde{P}_n}(v_k)$ maximizes when $k=(n-1)/2$. It follows that

$$\begin{aligned}
\sum_{i=1}^{n}\mathrm{btw}_{\widetilde{P}_n}(v_i) &= \sum_{i=1}^{n-1}\frac{(i-1)(n-1-i)}{\binom{n-1}{2}} + \sum_{i=1}^{(n-1)/2}\frac{i-1}{\binom{n-1}{2}} + \sum_{i=(n-1)/2}^{n-1}\frac{n-1-i}{\binom{n-1}{2}} \\
&= \sum_{i=1}^{n-1}\frac{(i-1)(n-1-i)}{\binom{n-1}{2}} + 2\sum_{i=1}^{(n-1)/2}\frac{i-1}{\binom{n-1}{2}} + \frac{(n-1)/2}{\binom{n-1}{2}} \\
&= \frac{1}{\binom{n-1}{2}}\sum_{i=1}^{n-1}\left(ni-(n-1)-i^2\right) + 2\sum_{i=1}^{(n-1)/2}\frac{i-1}{\binom{n-1}{2}} + \frac{(n-1)/2}{\binom{n-1}{2}} \\
&= \frac{1}{\binom{n-1}{2}}\left(\frac{n^2(n-1)}{2}-(n-1)^2-\frac{(n-1)n(2n-1)}{6}\right) + \frac{((n-1)/2-1)(n-1)/2}{\binom{n-1}{2}} + \frac{(n-1)/2}{\binom{n-1}{2}} \\
&= \frac{2n^3-9n^2+16n-9}{12\binom{n-1}{2}}.
\end{aligned}$$

Since $\mathrm{btw}_{\widetilde{P}_n}(v_k)$ is maximized when $k=\frac{n-1}{2}$, we denote the vertex $v_{(n-1)/2}$ by v^*, and we have

$$H_{\mathrm{btw}}C_{\mathrm{btw}}(\widetilde{P}_n) = \sum_{i=1}^{n}(\mathrm{btw}_{\widetilde{P}_n}(v^*)-\mathrm{btw}_{\widetilde{P}_n}(v_i)) = \frac{n(n^2-5)}{4\binom{n-1}{2}} - \frac{2n^3-9n^2+16n-9}{12\binom{n-1}{2}} = \frac{n^3+9n^2-31n+9}{6(n-1)(n-2)}.$$