Physica A 617 (2023) 128672

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Quantification and statistical analysis of topological features of recursive trees

Balázs Király [a,*], István Borsos [a], György Szabó [a,b]

[a] Institute for Technical Physics and Materials Science, Centre for Energy Research, Konkoly-Thege M. út 29-33, H-1525, Budapest, Hungary
[b] Institute of Evolution, Centre for Ecological Research, Konkoly-Thege M. út 29-33, H-1525, Budapest, Hungary

ARTICLE INFO

Article history:
Received 11 January 2023
Received in revised form 8 March 2023
Available online 15 March 2023

Keywords:
Recursive trees
Topological features of graphs
Matrix decomposition
Network analysis
Adjacency matrix
Game theory

ABSTRACT

Some topological features of recursive trees are quantified by exploiting the decomposition of directed graphs into a suitable combination of starlike hierarchical and three-edge cyclic components. This approach requires the adoption of the formalism of weighted directed graphs and allows us to quantify the proportion of hierarchical and hidden cyclic components. Using this concept, we can introduce new local parameters and global measures that quantify certain topological features of recursive trees. The average values of some of these measures over the general set of same-sized recursive trees are also determined.

© 2023 Elsevier B.V. All rights reserved.

1. Introduction

The recent uptick in network analysis is related to the success of combining the approaches of graph theory and statistical physics [1–5]. Within the wide range of graphs and networks, recursive trees characterize river networks, the evolution of biological species via consecutive mutations, and numerous other processes and algorithms [6–9]. Due to their simple structure, many of their important features (e.g., the number of isomorphism classes) have already been clarified in the literature of graph theory. Now we add to earlier analyses by utilizing the concepts of matrix decomposition, which has revealed the existence of four fundamentally different types of pair interactions in (evolutionary) game theory [10–12] and provided new insight into the existence and degeneracy of pure and mixed Nash equilibria by facilitating a systematic analysis of interactions through the linear combinations of elementary interactions [13].

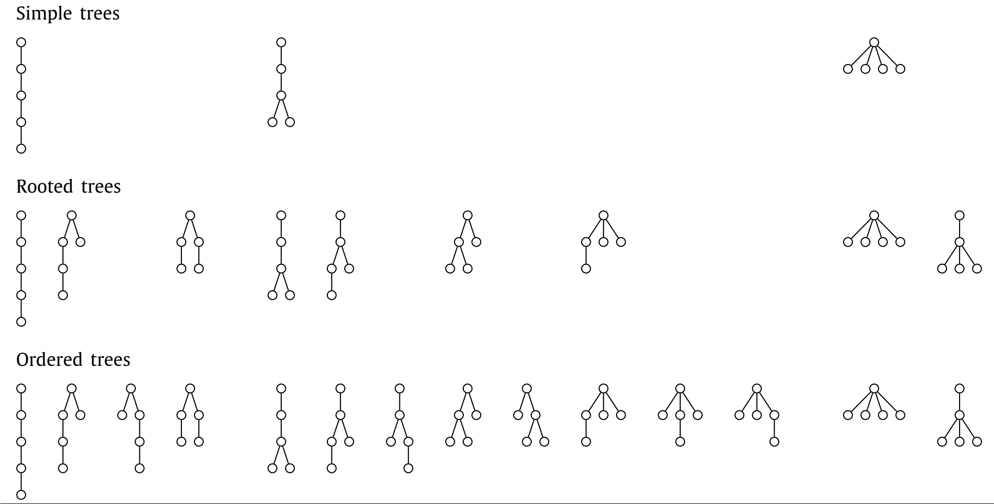
The present investigation is motivated by the simple relationship between directed graphs and games with antisymmetric payoff matrices [14]. The payoff matrices of two-person symmetric zero-sum games [15–18] can be decomposed into two orthogonal sets of elementary interactions that represent hierarchical and cyclic dominance. The adjacency matrices of directed starlike graphs correspond to the basis matrices of the hierarchical component, while a suitable set of three-edge directed graphs that each describe rock–paper–scissors-type cyclic dominance among three strategies span the cyclic dominance component. Previous investigations [12] have already found that quantities measuring the proportion of the hierarchical and cyclic components and the asymmetry of the hierarchical component can be successfully used to characterize graphs describing the outcomes of round robin tournaments.

Now – utilizing the concept of matrix decomposition – we explore the possibilities of quantifying local and global topological features of directed graphs. Our investigations are restricted to graph invariants, that is, quantities providing

* Corresponding author.
E-mail addresses: kiraly.balazs@ek-cer.hu (B. Király), borsos.istvan@ek-cer.hu (I. Borsos), szabo.gyorgy@ek-cer.hu (G. Szabó).

https://doi.org/10.1016/j.physa.2023.128672
0378-4371/© 2023 Elsevier B.V. All rights reserved.

Table 1
Simple, rooted and ordered trees.



the same value for isomorphic graphs, and focus on the properties of their distributions and ability to separate graphs according to isomorphism class. We test the capabilities of this approach on recursive trees, the properties of which simplify some of the necessary numerical calculations.

2. Survey of concepts and previous results

As a large number of different kinds of trees has been proposed and studied in the rich history of trees, we summarize those trees that are related to our investigation.

A simple tree is an acyclic connected undirected graph.

A rooted tree is a simple tree with a node selected as root. The root implies a natural orientation of the edges of the tree when either all edges are directed to the root or away from it depending on the application of the tree. In our case, they are all directed away from the root and thus the root is the only vertex without incoming edges. The vertices without outgoing edges are called leaves. Sometimes the orientation is implicit, not displayed in the drawing. In this case, the root is always the top node and the edges are oriented downwards.

An ordered tree is a rooted tree where the order of the children of its nodes are given for each node.

Albeit the underlying structure of all of the above-mentioned kinds of trees are the same, the trees may differ from each other by the assignment of the root and the ordering of the children. For example, there are three non-isomorphic simple trees of five nodes: the path, the fork, and the star, as in the first row of Table 1. The path can be assigned a root in three different ways. The fork has four possible roots. The star has only two. The selection of the root increases the number of variants a tree has, whereas the symmetries of the underlying tree limit the number of variants that can be obtained this way. We can derive various ordered trees from a rooted tree. Again, the symmetries of the rooted tree limit the number of possible ordered variants. Thus, we can only get four ordered trees from the three rooted variants of the five-node path.

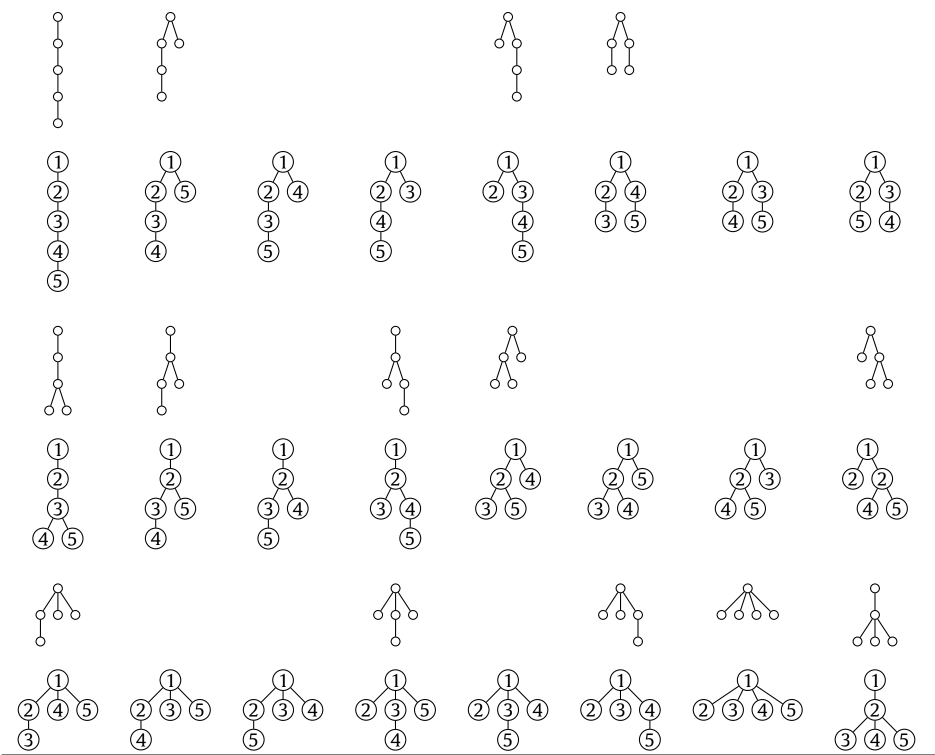
From the fork, we can get 4 rooted trees and from them 8 ordered trees.Table 1 depicts all variants of five-node trees in these three categories.

Random recursive trees are a widely studied subject [19]. Their various average features as random trees, such as depth [20], profile (the distribution of nodes on the levels of the trees) [21], number of leaves [22], isomorphism [23,24], and cycle analysis [25] have been investigated.

Starting from a fixed ordered list of nodes, the first of which is the root, we build a tree by sequentially adding nodes to the tree, connecting it to a previously added node. If the selection is random, then we obtain a random recursive tree. The ordering of the children of a node may or may not be important.

Historically, if the ordering of the children is important, it is represented by drawing the tree in the plane with the root at the top and the other nodes drawn below their ancestors putting the children in the required order, and this is reflected by their traditional naming: plane oriented recursive trees (or PORTs). They may have different embeddings depending on the ordering of the children of the nodes. In non-plane trees, where the ordering is irrelevant, plane embeddings are used only to visualize the trees, but for convenience an increasing order is usually still adopted for the children. Plane embeddings, however, are not directly usable for computer processing, so other representations are needed. Non-plane tree can be represented by an adjacency matrix, the rows of which contain the sets of child nodes for each node. This is not possible for PORTs as the adjacency matrix stores no ordering information. Adjacency lists are capable of containing

Table 2
Ordered and recursive trees.



this additional information without having to refer to plane embeddings. When adding a new node to a recursive tree, it is either appended to or inserted into the list of its would-be siblings. If it is appended (non-plane tree case), the number of new connection slots are increased by one, if it is inserted (PORT case) the number of slots is increased by 2, regardless of whether the parent is a leaf, internal node, or the root. So the total number of non-plane recursive trees is $(n-1)!$, and that of PORTs is $(2n-3)!!$.

The indices of the nodes are frequently considered labels of the nodes. By construction, in a recursive tree the labels of all nodes (except the root) are greater than the labels of any of their ancestors, so going down the tree we find increasing labels. In non-plane trees this is also true for the siblings, they are in increasing order. As recursive trees are labeled ordered trees, more than one recursive tree can be derived from an ordered tree with appropriate labeling, as shown by Table 2.

We see that on all levels of this tree hierarchy the more refined sets of trees are non-uniformly assigned to the trees of the levels above them. If, for example, we uniformly generate random recursive trees and discard the labels, then the ordered trees, rooted trees, or simple trees thus obtained will not be uniformly distributed.

The enumeration or counting of various families of trees, both labeled trees and isomorphic variants, also has a long history starting with Kirchhoff's and Cayley's classical results for labeled undirected trees. Some families of trees have been counted in the field of evolutionary phylogenetics, for details see Refs. [26,27] and the references therein.

In Table 3, we list the number of trees in various categories related to our work. Where no formula is given the numbers were obtained by the Pólya–Redfield advanced combinatorial method for counting isomorphism classes [7]. From the table, we see that the number of trees grows steeply as the number of nodes and the degrees of freedom of the various tree concepts increase.

The present analysis is restricted to a type of directed trees which can be used to illustrate the appearance of new species in ecological systems through consecutive mutations [26,27]. In these systems, the species are labeled by i ($i = 1, 2, \dots, n$) and arise from a common ancestor ($i = 1$) via consecutive mutations. Such an evolutionary process is well described by a rooted tree where node i is connected by an incoming edge to node j if the ith species is a mutant (direct successor) of species j ($j < i$). This directed tree graph can be represented by an $(n-1)$-dimensional vector of integers $[a_2, \dots, a_n]$ where a_i is the label of the direct ancestor of species i and $0 < a_i < i$. This construction describes the recursive trees mentioned above, and thus the total number of different scenarios is $(n-1)!$.

Another alternative description of these systems tabulates the connections between the nodes in an antisymmetric adjacency matrix $\mathbf{A}$ where $A_{ij} = -A_{ji} = 1$ if a directed edge runs from node i to node j (if i is the direct ancestor of j in

Table 3
Number of trees, closed formulas are given where known.

Nodes	Trees [7]	Rooted trees [7]	Ordered trees *Catalan numbers* $\frac{1}{n}\binom{2n-2}{n-1}$	Recursive trees $(n-1)!$	PORTs $(2n-3)!!$
1	1	1	1	1	1
2	1	1	1	1	3
3	1	2	2	2	15
4	2	4	5	6	105
5	3	9	14	24	945
6	6	20	42	120	10 395
7	11	48	132	720	135 135
8	23	115	429	5 040	2 027 025
9	47	286	1430	40 320	34 459 425
10	106	719	4862	362 880	654 729 075

the biological terminology) and $A_{ij} = 0$ otherwise. Adopting the concepts of matrix decomposition [11,12], the adjacency matrix **A** can be built up as the sum of a hierarchical (**H**) and a cyclic (**C**) component as

$$\mathbf{A} = \frac{1}{n}\mathbf{H} + \frac{1}{n}\mathbf{C}\,, \tag{1}$$

where the $1/n$ coefficients ensure that the elements of **H** and **C** are integers. The matrices **H** and **C** are orthogonal, that is, their scalar product is zero:

$$\mathbf{H}\cdot\mathbf{C} = \sum_{i,j} H_{ij}C_{ij} = 0\,. \tag{2}$$

In the present approach, the sum of two or more graphs is defined by the sum of their corresponding adjacency matrices. In graph theory, this extension is well defined for weighted directed graphs, and we will consider all graphs as such henceforth. In Ref. [14], the reader can find examples illustrating how a single directed edge or loop can be built up from two sets of elementary basis matrices. These basis matrices can be considered as axes of a coordinate system in the $\binom{n}{2}$-dimensional parameter space of $n \times n$ antisymmetric matrices.

The hierarchical component can be built up as a linear combination of elementary hierarchical basis matrices as

$$\mathbf{H} = \sum_i h_i \mathbf{H}^{(i)}\ \text{, where } h_i = \sum_j A_{ij}\,. \tag{3}$$

It is worth emphasizing here that consequently

$$\sum_i C_{ij} = \sum_j C_{ij} = 0\,. \tag{4}$$

For simple directed graphs, h_i is the difference between the number of the outgoing and incoming edges of node i, and $\mathbf{H}^{(i)}$ is the adjacency matrix of a starlike directed graph having edges that run from node i to all others. This set of matrices has the following two inherent features: The elementary hierarchical components are not orthogonal to each other and $\sum_i \mathbf{H}^{(i)} = 0$. As a result, only $(n-1)$ of the h_i coefficients that define **H** are independent ($\sum_i h_i = 0$). This reflects the equivalence of elementary starlike components within the whole set of directed graphs. The above-mentioned relations give a simple expression for the elements of **H**, namely,

$$H_{ij} = h_i - h_j\,. \tag{5}$$

This feature also guarantees that the values of h_i are determined by the hierarchical component, that is,

$$h_i = \frac{1}{n}\sum_j H_{ij}\,. \tag{6}$$

As a result, the hierarchical component is well characterized by the components of the vector **h** of the expansion coefficients h_i ($\mathbf{h}^T = (h_1, h_2, \ldots, h_n)$).

Straightforward calculations lead to simple relations between some scalar products of the components, such as

$$n\mathbf{H}\cdot\mathbf{A} = \mathbf{H}\cdot\mathbf{H} = \sum_{i,j}(h_i - h_j)^2 = 2n\sum_i h_i^2 = 2n\mathbf{h}\cdot\mathbf{h}\,. \tag{7}$$

For all simple directed graphs, the scalar product $\mathbf{A}\cdot\mathbf{A}$ is determined by the number of the graph's edges. For recursive trees, however, this quantity is directly related to the number of nodes, namely,

$$\mathbf{A}\cdot\mathbf{A} = 2(n-1)\,. \tag{8}$$

The proportion of hierarchical and cyclic components [12] can be quantified by their projections to **A**,

$$\Lambda = \frac{1}{n^2}\frac{\mathbf{H}\cdot\mathbf{H}}{\mathbf{A}\cdot\mathbf{A}} \text{ and } \Theta = \frac{1}{n^2}\frac{\mathbf{C}\cdot\mathbf{C}}{\mathbf{A}\cdot\mathbf{A}} = 1-\Lambda\,. \tag{9}$$

These global measures of topological features are determined by the vector **h** and n for all recursive trees. Any more detailed quantitative analysis of the topological features requires the introduction of further measures that take into account the cyclic component **C**, which can be built up as a linear combination of elementary cyclic components ($\mathbf{C}^{(\mathrm{i,j,k})}$) representing single three-edge directed loops with edges $i \to j \to k \to i$ under conditions $i < j < k$. Unfortunately, the total number of these cyclic components, $\binom{n}{3}$, exceeds the number of independent components, that is, $\binom{n-1}{2}$ [12,14]. However, a convenient complete and independent subset of $\mathbf{C}^{(\mathrm{i,j,k})}$s can be selected by setting $i = 1$, which is a natural choice for rooted trees. Then the cyclic component obeys a simple form:

$$\mathbf{C} = \sum_{\substack{j\\ j>1}}\sum_{\substack{k\\ k>j}} c(1,j,k)\mathbf{C}^{(\mathrm{i,j,k})}\,, \tag{10}$$

where the off-diagonal elements of **C** directly determine the strength of the elementary components, that is, $C_{jk} = c(1,j,k)$. The remaining elements in the first row and column of **C** are not independent from this set as a result of Eq. (4).

Based on the above, one can easily come to the conclusion that $C_{ij} = 0$ for recursive trees if the nodes i and j are both leaves, since then $A_{ij} = 0$ and $H_{ij} = 0$ as $h_i = h_j = -1$, which through Eq. (1) indicates the disappearance of these root–leaf–leaf cyclic components.

Most of the observations in this section concerning the decomposition of adjacency matrices, unless explicitly labeled as being specific to the adjacency matrices of recursive trees, are directly applicable to general square matrices, too, with the caveat that its hierarchical and cyclic components by definition only make up the antisymmetric part of a matrix, so the definitions in Eqs. (1), (3), and (10) have to be adjusted accordingly.

3. Local and global measures

In the approach outlined above, the vector **h** can be interpreted as a local measure quantifying a relevant topological feature of the hierarchical component for each node.

On the analogy of Eq. (6), we can introduce further similar local topological measures that are derived from **H** and **C**. A straightforward extension of Eq. (6) leads to vectors $\mathbf{h}^{(\mathrm{m})}$ with components

$$h_i^{(\mathrm{m})} = \sum_j H_{ij}^m\,, \tag{11}$$

where m is a positive integer. In this notation, $n\mathbf{h} = \mathbf{h}^{(1)}$. We can also derive a similar vector, $\mathbf{c}^{(\mathrm{m})}$, with elements

$$c_i^{(\mathrm{m})} = \sum_j C_{ij}^m\,, \tag{12}$$

which quantify the local topological features related to the cyclic component **C**. Note that $\mathbf{c}^{(1)} = 0$.

The quantities $\mathbf{h}^{(\mathrm{m})}$ and $\mathbf{c}^{(\mathrm{m})}$ do not contain explicit information about the special role of the root. This shortcoming can be eliminated by adopting further local quantities. For recursive trees, the distance (d_i) from the root ($i = 1$) to node i is usually defined as the number of edges connecting these nodes [7]. The collection of these local measures can also be thought of as an n-dimensional vector **d** with elements d_i ($d_1 = 0$). In the following numerical investigations, we have also distinguished a restricted distance vector **f** whose non-zero components are the distances of the leaves, that is, $f_i = d_i$ if $h_i = -1$ and $f_i = 0$ otherwise. On the analogy of Eqs. (11) and (12), we can derive further local vector measures $\mathbf{d}^{(\mathrm{m})}$ and $\mathbf{f}^{(\mathrm{m})}$ with components $d_i^{(\mathrm{m})} = d_i^m$ and $f_i^{(\mathrm{m})} = f_i^m$.

The scalar products of the above-mentioned $\mathbf{h}^{(\mathrm{m})}$, $\mathbf{c}^{(\mathrm{m})}$, $\mathbf{d}^{(\mathrm{m})}$, and $\mathbf{f}^{(\mathrm{m})}$ vectors with each other and the all-one vector **1** can be considered as global measures that quantify relevant topological features involving the cyclic and hierarchical and leaf structures of a graph and the correlations between them. Using scalar products as global measures ensures that their value does not depend on the labeling of the nodes. Consequently, these global measures have the same value for all recursive trees that belong to the same isomorphism class when considered as rooted trees. The converse of this statement, however, is not true, such global measures do not necessarily render unique values to different isomorphism classes. In the rest of this paper, we consider the efficiency of these measures in distinguishing graphs that belong to different isomorphism classes. The numerical investigation is conveniently simplified by the fact that these measures take integer values for recursive trees.

4. Examples

First, we consider some examples to illustrate the applicability and efficiency of the measures introduced above. In the first example, all successors arise directly from the same ancestor ($i = 1$). The corresponding graph is indexed as $[1, 1, \ldots, 1]$ and it belongs to an extreme isomorphism class. It is a starlike directed graph with an adjacency matrix

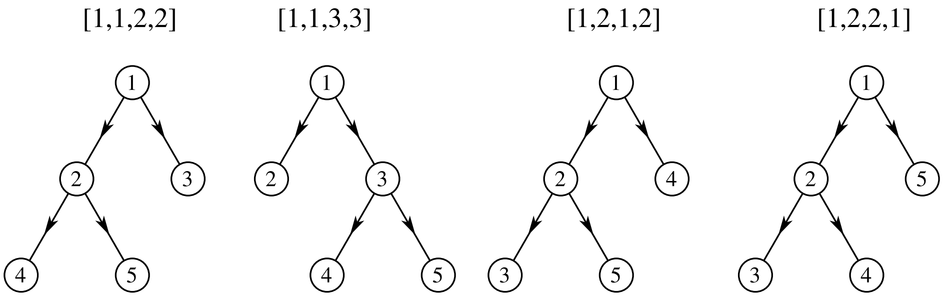


Fig. 1. Four $n = 5$ recursive trees that belong to the same isomorphism class when considered as rooted trees.

$\mathbf{A} = \mathbf{H}^{(1)}$. This graph has no cyclic components ($\mathbf{C} = 0$, and consequently $\mathbf{c}^{(m)} = 0\ \forall m$) and its hierarchical elementary components are given by the vector $\mathbf{h}$ ($\mathbf{h}^T = (n-1, -1, \dots, -1)$). This graph maximizes several global measures for arbitrary n, including the proportion of the hierarchical component ($\Lambda = 1$), the number of leaves ($n-1$), and $\mathbf{h} \cdot \mathbf{h}$ ($n(n-1)$).

All non-starlike recursive tree graphs have a non-zero cyclic component. The proportion of the hierarchical component is minimal for line graphs indexed by $[1, 2, \dots, n-1]$. Simple calculations give $\mathbf{h}^T = (1, 0, \dots, 0, -1)$, $\mathbf{h} \cdot \mathbf{h} = 2$, and the proportion of the hierarchical component ($\Lambda = \mathbf{H} \cdot \mathbf{H}/(n^2 \mathbf{A} \cdot \mathbf{A}) = 1/n$) goes to zero as $n \to \infty$. Line graphs minimize the number of leaves and the value of $\mathbf{h} \cdot \mathbf{h}$, too.

After the two extreme recursive trees above, consider now a five-node ($n = 5$) graph indexed as $[1, 1, 2, 2]$. The 24 $n = 5$ recursive trees can be sorted into 9 isomorphism classes when considered as rooted trees [7]. Fig. 1 shows three other recursive trees isomorphic to $[1, 1, 2, 2]$: $[1, 1, 3, 3]$, $[1, 2, 1, 2]$, and $[1, 2, 2, 1]$. Here we have to emphasize that the correspondence between the indices and the portrayed two-dimensional representations of the graphs follows simple rules. Namely, the arrows point downward, the labels increase from left to right, and nodes that share their ancestor are placed equidistantly along the same horizontal line.

The adjacency matrix of the graph $[1, 1, 2, 2]$ is

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 \\ -1 & 0 & 0 & 1 & 1 \\ -1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \end{pmatrix}. \tag{13}$$

The non-zero elements of $\mathbf{A}$ are directly defined by the index, that is, $A_{ij} = -A_{ji} = 1$ if $i = a_j$ for all $j > 1$. This relationship can be utilized in systematic numerical investigations. According to Eq. (6), $\mathbf{h}^T = (2, 1, -1, -1, -1)$ for the graph $[1, 1, 2, 2]$, and the corresponding hierarchical and cyclic components are

$$\mathbf{H} = \begin{pmatrix} 0 & 1 & 3 & 3 & 3 \\ -1 & 0 & 2 & 2 & 2 \\ -3 & -2 & 0 & 0 & 0 \\ -3 & -2 & 0 & 0 & 0 \\ -3 & -2 & 0 & 0 & 0 \end{pmatrix} \text{ and } \mathbf{C} = \begin{pmatrix} 0 & 4 & 2 & -3 & -3 \\ -4 & 0 & -2 & 3 & 3 \\ -2 & 2 & 0 & 0 & 0 \\ 3 & -3 & 0 & 0 & 0 \\ 3 & -3 & 0 & 0 & 0 \end{pmatrix}. \tag{14}$$

The other three isomorphic recursive trees have different $\mathbf{h}$ and $\mathbf{c}$ vectors and $\mathbf{H}$ and $\mathbf{C}$ matrices. Nonetheless, their global measures are identical. For example, $\mathbf{A} \cdot \mathbf{A} = 8$ and $\mathbf{h} \cdot \mathbf{h} = 8$ for all of them.

Visually, these parameters translate to the following: The $[1, 1, 2, 2]$ recursive tree can be built up from all five starlike hierarchical and just three of the six independent three-edge cyclic components, with weight factors proportional to h_i and C_{2j} for $j > 2$. The $[1, 1, 2, 2]$ graph has three leaves ($i = 3$, 4 and 5), and the vanishing elements of its $\mathbf{C}$ matrix in Eq. (14) make explicit the absence of elementary cyclic components of three-edge loops involving the root and any two leaves.

The set of $n = 5$-node recursive trees has 9 isomorphism classes. These obviously cannot be identified by those single global measures that take fewer distinct values. For example: The number of leaves can only be 1, 2, 3, or 4; $\mathbf{h} \cdot \mathbf{h}$ exhibits just 6 different values; there are only 8 possible outcomes when measuring $\mathbf{d} \cdot \mathbf{d}$ or $\mathbf{h}^{(1)} \cdot \mathbf{c}^{(2)}$. At the same time, $\mathbf{f} \cdot \mathbf{f}$ and $\mathbf{h}^{(1)} \cdot \mathbf{c}^{(3)}$ both take 9 different values and are able to correctly identify the isomorphism classes. In the next section, we consider these capabilities for larger n.

5. Distributions and averages

We have performed numerical investigations to determine the different values a wide range of global measures mentioned above can take. Fig. 2 shows the number of $n = 7$-node recursive trees characterized by the same values

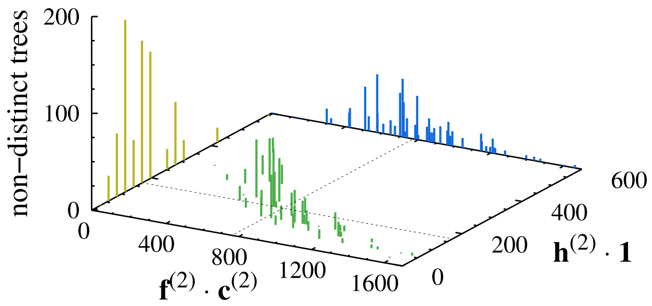


Fig. 2. Distribution of the number of recursive trees characterized by two measures, $\mathbf{h}^{(2)} \cdot \mathbf{1}$ and $\mathbf{f}^{(2)} \cdot \mathbf{c}^{(2)}$, for $n = 7$. The marginal distributions are shown by vertical columns along the axes and the dashed lines indicate their average values (quantitatively 168 and 770) over the whole set.

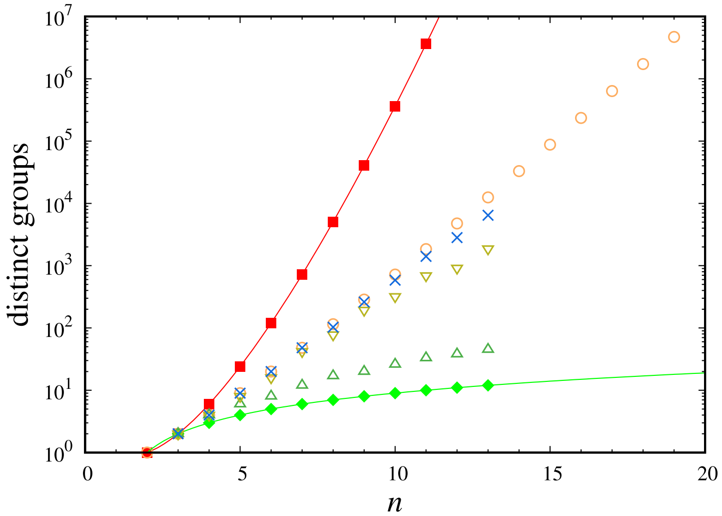


Fig. 3. Lin-log plot of the number of recursive trees (■), their isomorphism classes (◯) and the different values for the number of leaves (♦), $\mathbf{h}^{(2)} \cdot \mathbf{1}$ (△), $\mathbf{f}^{(2)} \cdot \mathbf{h}^{(2)}$ (▽), and $\mathbf{f}^{(2)} \cdot \mathbf{c}^{(2)}$ (×) can take for recursive trees, as a function of the number of nodes n. The continuous lines denote analytical results.

of $\mathbf{h}^{(2)} \cdot \mathbf{1} = \mathbf{H} \cdot \mathbf{H}$ and $\mathbf{f}^{(2)} \cdot \mathbf{c}^{(2)}$. There are 720 such graphs in total, and they can be sorted into 48 isomorphism classes. Six of the isomorphism classes contain only a single graph. All these graphs have an obvious symmetry reflected by their indices: [1, 1, 1, 1, 1, 1], [1, 2, 2, 2, 2, 2], [1, 2, 3, 3, 3, 3], [1, 2, 3, 4, 4, 4], [1, 2, 3, 4, 5, 5], and [1, 2, 3, 4, 5, 6]. Most of the global measures are not able to identify the isomorphism classes on their own. For example, the quantity $\mathbf{h} \cdot \mathbf{h}$ and global measures strongly related to it (e.g., $\mathbf{H} \cdot \mathbf{H}$ and $\mathbf{C} \cdot \mathbf{C}$) take only 12 different values as shown in Fig. 2. On the other hand, the maximum number of different values (48) is replicated by the quantification of $\mathbf{f}^{(2)} \cdot \mathbf{c}^{(2)}$. For other global measures, the number of different values varied between a minimum of 6 and the maximum. For example, $\mathbf{h}^{(1)} \cdot \mathbf{c}^{(3)}$ is able to distinguish 38 classes. Evidently, using combinations of two or three global measures enhances their capability of distinguishing isomorphism classes.

Fig. 3 shows the number of different possible values of some global measures we studied, as a function of n ($n < 14$). The total number of recursive trees (given in the book by Harary [7]) is also plotted for comparison. These numerical analyses are limited by the technical difficulties related to the exponentially increasing numbers of both the recursive trees themselves and the different values the global measures can take. Disregarding weak irregularities, most of these quantities increase exponentially with n. Notable exceptions are the number of different possible values of leaves and $\mathbf{h} \cdot \mathbf{h}$, which both increase linearly with n.

It is noteworthy that the quantity $\mathbf{f}^{(2)} \cdot \mathbf{c}^{(2)}$ is able to distinguish the largest number of isomorphism classes. Fig. 3, however, shows clearly that even this quantity is not capable of identifying all isomorphism classes on its own when $n > 7$. Combinations of two or more global measures can be used to refine the quantification of topological features, as illustrated by Fig. 2, but these methods still suffer from the same technical difficulties imposed by exponential increases.

The iteratively generated nature of trees allows us to calculate the averages of certain topological quantities over the general set of same-sized recursive trees without detailed knowledge of their distribution. Each $(n+1)$-node recursive tree can be uniquely obtained by attaching a new leaf to an n-node recursive tree. If we can track how this process changes the quantity in question, we can derive the n-dependence of its average by induction. For example, the number of leaves remains the same when the new leaf is attached to a (former) leaf, and increases by 1 otherwise. This means that the increases contribute as many additional leaves to the $(n+1)$-node set as there are non-leaf nodes in the n-node set on top of the 'original' leaves that are carried over to all n successors of each tree for a total of

$$l(n+1) = nl(n) + [n(n-1)! - l(n)] = (n-1)l(n) + n!\,. \tag{15}$$

Dividing by $n!$, we get for the averages:

$$\bar{l}(n+1) = \frac{n-1}{n}\bar{l}(n) + 1\,. \tag{16}$$

It is easy to check that this result is consistent with $\bar{l}(n) = n/2$, which obviously holds for $n = 2$.

By the same token, we can calculate the set average of quantities of the form $h^{(m)} = \sum_i h_i^m$ by observing how adding an additional node to a tree changes its h_i. Since h_i is the net number of incoming and outgoing edges at node i, all attaching a new leaf to node j does is it increases h_j by 1 and introduces $h_{n+1} = -1$. As a result, the total increase in $h^{(m)}$ for a single tree is $\Delta h^{(m)} = (h_j+1)^m + (-1)^m - h_j^m$. This means that each tree in the n-node set contributes its original $h^{(m)}$ to all of its n successors in the $(n+1)$-node set as a baseline, which is modified by the appropriate $\Delta h^{(m)}$ for each node's corresponding successor. Dividing the sum by $n!$, we again get a recursive expression for the averages:

$$\overline{h^{(m)}}(n+1) = \overline{h^{(m)}}(n) + \frac{1}{n}\overline{\Delta h^{(m)}}(n)\,, \tag{17}$$

where $\overline{h^{(m)}}(n)$ is to be calculated as if the changes applied to all nodes of all n-node trees at the same time, that is, as the sum $\sum_i \Delta h_i^{(m)} = \sum_i \left[(h_i+1)^m + (-1)^m - h_i^m\right]$ averaged over all $(n-1)!$ n-node recursive trees. In particular, for $m = 1$, 2, and 3 this means:

$$\overline{h^{(1)}}(n+1) = \overline{h^{(1)}}(n) + 0\,, \tag{18}$$

$$\overline{h^{(2)}}(n+1) = \overline{h^{(2)}}(n) + \frac{2}{n}\overline{h^{(1)}}(n) + 2\,, \tag{19}$$

$$\overline{h^{(3)}}(n+1) = \overline{h^{(3)}}(n) + \frac{3}{n}\overline{h^{(2)}}(n) + \frac{3}{n}\overline{h^{(1)}}(n)\,. \tag{20}$$

Starting from the base case of $n = 2$, when $\mathbf{h}^T = (1, -1)$ and $\overline{h^{(2k+1)}}(2) = 0$ for odd and $\overline{h^{(2k)}}(2) = 2$ for even exponents, it is easy to show by induction that

$$\overline{h^{(1)}}(n) = 0\,, \tag{21}$$

$$\overline{h^{(2)}}(n) = 2(n-1)\,, \tag{22}$$

$$\overline{h^{(3)}}(n) = 6\left(n - 1 - \sum_{k=1}^{n-1}\frac{1}{k}\right)\,. \tag{23}$$

These calculations predict that the average value of Λ, the proportion of hierarchical components, over the set of recursive trees is $2/n$. Similar results are found via numerical simulations when the average value of Λ is determined over the larger set of simple (random) directed graphs [12]. This general result coincides with the feature that a single directed edge can be built up from two starlike hierarchical and $(n-2)$ directed three-edge cyclic components [14].

6. Summary

In this study, the concept of the decomposition of antisymmetric matrices into cyclic and hierarchical components is adopted to extend the quantitative analysis of topological features in recursive trees. This framework is based on the notion of addition (linear combination) of directed graphs. This approach can be considered as the introduction of basis directed graphs from which all directed weighted graphs can be built up. The most striking message of this approach is that it reveals cyclic components that are hidden when directed trees are observed visually. The mathematical relationship between the cyclic and starlike hierarchical components has allowed us to introduce measures that give more detailed information about the topological features of these directed graphs.

We have chosen recursive trees to demonstrate the applicability and capabilities of these measures, because the presence of a unique root naturally designates a convenient independent set of three-edge directed loops whose linear combinations span the cyclic component for all weighted directed graphs. Other advantageous features of this set of directed graphs includes their recursive structure (facilitating analytical calculations) and easy indexability (simplifying the numerical investigations).

We have found that the hierarchical component is well defined by a vector $\mathbf{h}$ that counts the difference between the outgoing and incoming edges for each node. For recursive trees, this vector also determines the scalar products of the adjacency matrices of the cyclic and hierarchical components as well as their proportions defined as projections. Due to the close relationship between these matrix scalar products the number of different values they can take are identical for these quantities.

On the analogy of the **h** vector, we could introduce additional vectors to describe further local topological features. The scalar products of these vectors can also be used as global measures of topological features, as they assign the same integers to recursive trees that belong to the same isomorphism class when considered as rooted trees. These scalar products quantify the correlations between different local measures derived from the leaf structure and the adjacency matrices of the hierarchical and cyclic components. Our efforts are focused on counting the number of different values a limited set of these global measures can take. The closer the number of different values a given measure can take is to the number of isomorphism classes, the more capable it is of providing a detailed classification of the graphs based on their topological features. We found that global measure that combine fundamentally different topological features have the highest resolving power. Although the number of different values one such promising measure involving the cyclic and leaf structures of the graphs can take increases exponentially with n, it still becomes unable to distinguish all isomorphism classes when $n > 7$. Evidently, the resolving power of this approach is increased by considering two or more suitable global measures simultaneously.

In light of our observations, the most promising direct application of these measures may be in the quantitative analysis of evolutionary processes through some of their topological features, as the average and expectation values and (simplified) distributions of global measures over sets of directed graphs could serve as a base of comparison and thus help reveal hidden relationships. One such field of research, where our approach could possibly provide new insights, is the study of phylogenetic trees, whose actual topology and scaling behavior has already been – basically in the spirit outlined above – compared to the extreme cases of balanced and maximally unbalanced binary trees by previous research [28,29].

A similar approach could also prove to be useful in the analysis of payoff matrices, where its core idea of matrix decomposition originated. Typically, especially when the players are assumed to be rational, the order in which its pure strategies are labeled is immaterial to the description of a game. However, choosing a different labeling will obviously lead to a formally different payoff matrix with permutated entries, obscuring the connection. Conversely, numerous seemingly different matrices in seemingly different models (or even fields of research) may actually define the same game theoretic situation. This equivalence of payoff matrices is analogous to the isomorphism of weighted looped graphs, a general version of the simple directed graph isomorphism problem we considered in the special case of recursive trees in this paper. In some situations, certain payoff components or payoff scaling can also be irrelevant to game theoretic outcomes, which could further complicate identifying equivalent descriptions [30].

CRediT authorship contribution statement

Balázs Király: Conceptualization, Methodology, Investigation, Formal analysis, Writing – original draft, Writing – review & editing, Visualization. **István Borsos:** Conceptualization, Methodology, Investigation, Writing – original draft, Writing – review & editing, Visualization. **György Szabó:** Conceptualization, Methodology, Investigation, Formal analysis, Writing – original draft, Writing – review & editing, Visualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work was supported by the National Research, Development and Innovation Office – NKFIH, Hungary under grant number OTKA PD-138571.

References

[1] B. Bollobás, Random Graphs, Academic Press, New York, 1985.
[2] B. Bollobás, Modern Graph Theory, Springer, New York, 1998.
[3] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D. Hwang, Complex networks: Structure and dynamics, Phys. Rep. 424 (2006) 175–308.
[4] A.-L. Barabási, Network Science, Cambridge University Press, Cambridge, UK, 2016.
[5] S. Johnson, Digraphs are different: Why directionality matters in complex systems, J. Phys. Complex. 1 (2020) 015003.
[6] F. Harary, R.Z. Norman, D. Cartwright, Structural Models: An Introduction to the Theory of Directed Graphs, Wiley, New York, 1966.
[7] F. Harary, Graph Theory, Addison-Wesley Publishing Company, Reading, 1969.
[8] V. Modrak, Z. Soltysova, Exploration of the optimal modularity in assembly line design, Sci. Rep. 12 (2022) 20414.
[9] C. Goldschmidt, J.B. Martin, Random recursive trees and the Bolthausen-Sznitman coalescent, Electron. J. Probab. 10 (2005) 718–745.
[10] G. Szabó, K.S. Bodó, B. Allen, M.A. Nowak, Four classes of interactions for evolutionary games, Phys. Rev. E 92 (2015) 022820.
[11] G. Szabó, I. Borsos, Evolutionary potential games on lattices, Phys. Rep. 624 (2016) 1–60.
[12] G. Szabó, I. Borsos, B. Leitner, Statistical analyses of cyclic and hierarchical dominances in directed graphs, Phys. Rev. E 100 (2019) 032301.

[13] G. Szabó, B. Király, General features of Nash equilibria in combinations of elementary interactions in symmetric two-person games, Eur. Phys. J. B 94 (2021) 102.
[14] G. Szabó, K.S. Bodó, K.A. Samani, Separation of cyclic and starlike hierarchical dominance in evolutionary matrix games, Phys. Rev. E 95 (2017) 012320.
[15] J. von Neumann, O. Morgenstern, Theory of Games and Economic Behaviour, Princeton University Press, Princeton, NJ, 1944.
[16] M.J. Osborne, A. Rubinstein, A Course in Game Theory, MIT Press, Cambridge, MA, 1994.
[17] H. Gintis, Game Theory Evolving, Princeton University Press, Princeton, 2000.
[18] K. Sigmund, The Calculus of Selfishness, Princeton University Press, Princeton, NJ, 2010.
[19] M. Drmota, RandOm Trees: An Interplay Between Combinatorics and Probability, Springer, Wien, 2009.
[20] P.P. Pach, G. Pluhár, A. Pongrácz, C. Szabó, The number of rooted trees of given depth, Electon. J. Combin. 20 (2013) 38.
[21] H.-K. Huang, Profiles of random trees: Plane oriented recursive trees, Random Struct. Algorithms 30 (2007) 380–413.
[22] Y. Zhang, On the number of leaves in random recursive trees, Braz. J. Probab. Stat. 29 (2015) 897–908.
[23] M. Bóna, P. Flajolet, Isomorphism and symmetries in random phylogenetic trees, J. Appl. Probab. 46 (2009) 1005–1019.
[24] C. Olsson, The probability of random trees being isomorphic, 2022, URL https://arxiv.org/abs/2205.06625.
[25] V. Vasiliauskaite, T.S. Evans, P. Expert, Cycle analysis of directed acyclic graphs, Physica A 596 (2022) 127097.
[26] J. Felsenstein, The number of evolutionary trees, Syst. Zool. 27 (1978) 27–33.
[27] J. Felsenstein, Counting phylogenetic invariants in some simple cases, J. Theoret. Biol. 152 (1991) 357–376.
[28] E.A. Herrada, C.T. Tessone, K. Klemm, V.M. Eguíluz, E. Hernández-Garcia, C.M. Duarte, Universal scaling in the branching of the tree of life, PLoS ONE 3 (2008) e2757.
[29] C. Xue, Z. Liu, N. Goldenfeld, Scale-invariant topology and bursty branching of evolutionary trees emerge from niche construction, Proc. Natl. Acad. Sci. USA 117 (2020) 7879–7887.
[30] J. Gabarró, A. García, M. Serna, The complexity of game isomorphism, Theoret. Comput. Sci. 412 (2011) 6675–6695.