

RESEARCH

Open Access



Intersection of random spanning trees in complex networks

András London^{1,2*} and András Pluhár¹

*Correspondence:
london@inf.u-szeged.hu

¹ Institute of Informatics,
University of Szeged, Szeged,
Hungary

² Department of Operations
Research and Mathematical
Economics, Poznań University
of Economics and Business,
Poznan, Poland

Abstract

In their previous work, the authors considered the concept of random spanning tree intersection of complex networks (London and Pluhár, in: Cherifi, Mantegna, Rocha, Cherifi, Micciche (eds) *Complex networks and their applications XI*, Springer, Cham, 2023). A simple formula was derived for the size of the minimum expected intersection of two spanning trees chosen uniformly at random. Monte Carlo experiments were run for real networks. In this paper, we provide a broader context and motivations for the concept, discussing its game theoretic origins, examples, its applications to network optimization problems, and its potential use in quantifying the resilience and modular structure of complex networks.

Keywords: Random spanning trees, Small-world networks, Modularity, Network resilience, Spanning tree game

Introduction

Given an undirected connected graph G , a *spanning tree* T of G is a subgraph that is a tree containing all the vertices of G . In the case of weighted graphs, the concept of *minimum spanning tree*, that is, a spanning tree with the smallest possible total edge weight, is of extraordinary importance. Spanning trees play a key role in many applications, such as network design, including computer networks, telecommunication networks, transportation networks, water supply networks, and electrical grids, see for example (Wu and Chao 2004); clustering, see for example the single linkage method and applications in finance (Mantegna 1999; Tola et al. 2008); or image registration and segmentation (Ma et al. 2000; Xu and Uberbacher 1997), just to mention a few, without being exhaustive.

Spanning trees in complex networks have been studied from various perspectives. They have been considered as skeletons of the network (Kim et al. 2004), used for dimension reduction (Mantegna 1999), or used for efficient visualization of evolving networks (Chen and Morris 2003).

In this work, we provide a different perspective of using spanning trees to get a deeper understanding of the structure of small-world networks. We start with a closely related game-theoretic concept.

An edge-tree game

Alon et al. (1995) studied the following two-player zero-sum game on a connected graph G . In a round, *tree player* chooses a spanning tree T of G , while *edge player* chooses an edge e of G . The payoff to the edge player is defined by the function $\text{cost}(T, e)$ as follows. If e is an edge of T , then $\text{cost}(T, e) = 0$, while if e is not in T , then the cost is the length of the unique cycle (also called fundamental cycle) formed by adding e to T . They derived bounds for $\text{cost}(T, e)$ for various classes of graphs and pointed out that the game arises in the context of the k -server problem, an online optimization problem on road networks.

Remarks

Motivated by the work of Alon et al., Bartal (1996) introduced the idea of *hierarchically well-separated tree*, or HST, which proved to be extremely useful in studying randomized algorithms on finite metric spaces. Essentially, a metric space \mathcal{M} can be approximated with polynomially many randomly chosen weighted trees of a special property. (The property is that starting from the root the edge weights decrease exponentially.) Note that later Fakcharoenphol et al. gave the optimal form of this approximation (Fakcharoenphol et al. 2003).

Tree intersection game

Inspired by the ideas above we introduce a similar one replacing the edge player with another tree player. In this scenario, we are given a connected graph G , two players, Big and Small, choose a spanning tree, T_B and T_S of G , respectively, without knowing of each other's choices. The goal of Big is to maximize the number of common edges of T_B and T_S , i.e. the intersection of the two trees, while the goal of Small is the opposite. Finding optimal strategies and the value of this game seems far from obvious for general graphs. Optimal strategies can be found among mixed strategies, meaning that players choose each spanning tree with a probability given by a probability distribution over all spanning trees of the graph.

For some classes of graphs, the optimal strategy for *both* players is to choose a uniform random spanning tree. To compute the payoff (i.e. the intersection of the chosen spanning trees) in this case, we come to the problem of *intersection of random spanning trees*. This is the number of common edges of two spanning trees chosen uniformly at random.

Although this choice is not an optimal strategy (of any player) of the game in general, the parameter it provides (the size of the intersection) captures a lot about the structure of a graph. For real small-world networks, the size of the mean intersection provides powerful insights into their structure. It seems that mainly the modular structure of the network and the number of weak links (i.e. links between communities) Granovetter (1973) determine the value of the parameter, since links between communities tend to be part of the spanning trees of the network. Roughly speaking, one expects large spanning tree intersections for networks with high Newman modularity, since the edges between communities, the weak links, are often overrepresented in the spanning trees. On the other hand, interestingly, our results suggest that

network heterogeneity, i.e. heterogeneous degree distribution, does not itself indicate higher spanning tree overlap as the expected minimum.

Remarks

A somewhat similar game was introduced and analyzed by Gueye et al. (2010), often referred to as the *secure broadcast game*, see e.g. Kottegoda (2020). In this game, a *broadcaster* B located in a network node wants to broadcast a message to all other nodes. This is done by choosing a spanning tree. The other player, called *eavesdropper* E, can observe the transmission along a single link. B wins the game if the spanning tree avoids E's edge, while E wins if the tree includes it. The game has been used for an interpretation of the *p-modulus theory* of graphs, see (Kottegoda 2020; Albin et al. 2021).

The rest of the paper is organized as follows. At first, we briefly introduce spanning tree games on graphs and show simple examples. Then we discuss the main concepts and definitions we are dealing with. We derive a lower bound on the expected intersection of two random spanning trees. This is done by considering a more general problem using hypergraphs. Finally, we present experimental results on both synthetic network models and real-world networks, as well as some possible directions for future work.

Throughout this paper, $G = (V, E)$ will be a finite, connected, undirected and unweighted graph with $|V| = n$ and $|E| = m$.

Spanning tree games on graphs

The games mentioned in the Introduction, (Alon et al. 1995; Kottegoda 2020), and the tree overlap game are so-called *matrix games*, which are in a sense perfectly described. The available (pure) strategies of Big can be listed as R_1, \dots, R_k , while those of Small as O_1, \dots, O_ℓ . Then matrix A such that a_{ij} , the element in the i th row and j th column is the payoff of Big playing strategy R_i , assuming Small plays the strategy O_j .

The Minimax theorem, see e.g. at Chvátal's book (Chvatal 1983), states that there are probability distributions \mathbf{x} and \mathbf{y} on the rows and columns, which are the optimal strategies for the players. These strategies can be computed efficiently by linear programming. However, in the case of combinatorial games the number of strategies is usually too large to consider this approach.

Case of G is a union of two disjoint spanning trees

This construction plays an important role in the solution of the Shannon switching game by Lehman (1964). For a given simple graph G two players, Maker and Breaker, alternately claim the edges of G by taking turns. Maker wins by claiming a connected spanning subgraph, otherwise Breaker wins.

Theorem 1 (Lehman 1964) *For a graph G , Maker as a second player wins the Shannon switching game on G if and only if G contains two edge-disjoint spanning trees.*

Observation 1 *If a graph G on n vertices is the union of two disjoint spanning trees, then the number of different trees in G is at least 2^{n-1} .*

Proof For a simple argument we recall the celebrated Erdős-Selfridge theorem (Erdős and Selfridge 1973). A hypergraph $\mathcal{F} = (V, \mathcal{H})$ is a ground set V (nodes) and $\mathcal{H} \subset 2^V$ (hyperedges). A Maker-Breaker game on \mathcal{F} is as before, the players take the elements of V and Maker wins by taking all elements of an $A \in \mathcal{H}$.

Theorem 2 (Erdős and Selfridge 1973) *Breaker can win a Maker-Breaker game $\mathcal{F} = (V, \mathcal{H})$, as a first player, if $\Phi(\mathcal{F}) := \sum_{A \in \mathcal{H}} 2^{-|A|} < 1$.*¹

Now let $V = E(G)$, the edge set of G and \mathcal{H} be the set of spanning trees. Every spanning tree has $n - 1$ edges, that is $\Phi(\mathcal{F}) = 2^{n-1}|\mathcal{H}|$. Since Maker wins by Theorem 1, by Theorem 2 we have $\Phi(\mathcal{F}) \geq 1$, which gives $2^{n-1} \leq |\mathcal{H}|$. □

Observation 2 *If a graph G on n vertices is the union of two disjoint spanning trees, then the value of the tree intersection game is $(n - 1)/2$.*

Proof Let G be the union of T_1 and T_2 . Big takes both T_1 and T_2 with probability $1/2 - 1/2$. If for a spanning tree T , $|T \cap T_1| = i$, then $|T \cap T_2| = n - i - 1$, so the expected payoff of Big is $(n - 1)/2$. On the other hand, Small can follow the same strategy with expected payoff $(n - 1)/2$. □

Some other cases

Observation 3 *The value of the tree intersection game is $2(n - 1)/n$, if $G = K_n$.*

Proof If n is even, we can divide the edge set of K_n to $n/2$ disjoint trees $T_1, \dots, T_{n/2}$. Taking those by probability $2/n$ each, is optimal strategy for both players, with value $2(n - 1)/n$. If n is odd, we divide the edge set into $(n - 1)/2$ Hamiltonian circuits $C_1, \dots, C_{(n-1)/2}$, and within each C_i we consider the Hamiltonian paths P_1^i, \dots, P_n^i , where $P_j^i \subset C_i$ missing the j th edge of C_i . Taking the appropriate uniform distribution on $\{P_j^i\}_{i,j}$ one observes that this is optimal for both players and gives expected intersection $2(n - 1)/n$. □

Observation 4 *If graph G has an edge-transitive automorphism group,² then the value of the tree overlap game is $(n - 1)^2/m$, where $m = |E(G)|$.*

Proof Take the uniform distribution of a arbitrary fixed tree T over the automorphism group. We claim that this is optimal for both players. Every edge $e \in E(G)$ is covered by probability $(n - 1)/m$, which gives the expected intersection $(n - 1)^2/m$ of T with any fixed tree T' . □

¹ In the original form Maker is the first player, and the $\Phi(\mathcal{F}) < 1/2$ condition is needed.

² Automorphism of a graph is mapping onto itself that preserves the edge-vertex connectivity. The set of automorphisms of a given graph, under the composition operation, forms a group, the automorphism group of the graph.

Remarks

One can argue that Observation 3 is a simple consequence of Observation 4. It is still worth separating them, since the proof of Observation 3 uses only linear number of trees, while the proof of the other may need super exponential number of trees. It is not clear if this number can be reduced in general.

Random spanning trees

Let \mathcal{T}_G be the set of all spanning trees of a graph G . The cardinality of \mathcal{T}_G , i.e. the number of spanning trees of G , is explicitly known by Kirchhoff's matrix tree theorem and can be calculated as the product of the positive eigenvalues of the graph Laplacian matrix L , divided by n . Moreover, it can be calculated in polynomial ($\sim n^3$) time, by removing the i th row and i th column (for any i) of L and calculate the determinant of the remaining $(n - 1) \times (n - 1)$ matrix. This allows us to precisely define the *uniform random spanning trees*, i.e. the uniform probability distribution over all spanning trees. Note also that for an edge $e \in E(G)$ we can compute the probability $p(e)$ that the edge e is in a uniform random spanning tree of G in polynomial time. This will be important in the use of Observation 5.

Another results of Kirchhoff states that the probability of $e \in T$ for any edge e and a random tree T equals to the effective resistance of that edge, when considering the graph as an electrical network with unit edge conductance. For more details and computation methods see (Ellens et al. 2011).

Although we are able to determine what the probability of picking a spanning tree uniformly at random is, it is not straightforward how to actually generate one. Several algorithms have been provided to generate random spanning trees of an undirected graph, see, for instance Broder's (1989), Wilson's (1996) algorithms. In our experiments we utilized Wilson's algorithm available in Python package `DPPY` (Gautier et al. 2019).

We should note here that more general random spanning trees can be considered given any probability mass function over \mathcal{T}_G . Related problems and results are discussed e.g. in Albin et al. (2021). In this paper, we consider only the uniform distribution case.

Minimum expected intersection of random spanning trees

In order to provide a lower bound on the expected value of the number of common edges of two random spanning trees of G we consider a more general problem on hypergraphs. We point out that the calculations can be easily done directly for graphs and its random spanning trees.

Now, let $\mathcal{F} = (V, \mathcal{H})$ be a *uniform* hypergraph, i.e. all hyperedge $A \in \mathcal{H}$ contains the same number of nodes. Let $|V| = \mu$, $|\mathcal{H}| = \kappa$ and $|A| = \nu$ for each hyperedge $A \in \mathcal{H}$.

Let A and B be two hyperedges of \mathcal{F} chosen independently with probability $1/\kappa$. The *expected hyperedge intersection* of hypergraph \mathcal{F} is $\mathbb{E}(|A \cap B|)$.

The key theorem that works for any hypergraph \mathcal{H} is the following.

Theorem 3 *For any hypergraph \mathcal{F} the expected hyperedge intersection is at least ν^2/μ .*

Proof The degree $d_{\mathcal{F}}(x)$ of a node $x \in V$ is the number of hyperedges containing x . Then $\sum_{x \in V} d_{\mathcal{F}}(x) = \kappa v$, while the average degree is $\bar{d}_{\mathcal{F}} = \kappa v / \mu$. If we pick two hyperedges A and B randomly, then the probability that they both contain a node x with degree $d_{\mathcal{F}}(x)$ is

$$\Pr(x \in A \cap B) = \frac{\binom{d_{\mathcal{F}}(x)}{2}}{\kappa^2}$$

and hence the expected number of edges in the intersection is

$$\mathbb{E}(|A \cap B|) = \sum_{x \in V} \frac{\binom{d_{\mathcal{F}}(x)}{2}}{\kappa^2} \geq \mu \frac{\left(\frac{\kappa v}{\mu}\right)^2}{\kappa^2} = \frac{v^2}{\mu},$$

by noting that the expectation is minimized when all nodes have the same degree. \square

Observation 5 Let $\mathcal{F} = (V, \mathcal{H})$ and $p(x) = \Pr(x \in A)$, where $A \in \mathcal{H}$ chosen uniformly at random. The exact hyperedge intersection value of \mathcal{F} is $\sum_{x \in V} p^2(x)$.³

Now let us define $\mathcal{F} = (V, \mathcal{H})$ as follows. Each node $x \in V$ of \mathcal{F} corresponds to an edge $e \in E$ of a graph G and a hyperedge $A \in \mathcal{H}$ corresponds to a spanning tree T of G . That is, $V(\mathcal{F}) = E(G)$ and $\mathcal{H} = \mathcal{T}_G$. Hence, \mathcal{F} is an $(n - 1)$ -uniform hypergraph with $|V| = m$, and using Theorem 3 we get the following lower bound on the expected spanning tree intersection for any graph.

Corollary Given a connected graph G of n nodes and m edges and $T_1, T_2 \in \mathcal{T}_G$ two random spanning trees of G . The minimum expected intersection of T_1 and T_2 , i.e. the minimum expected spanning tree intersection of G , is $(n - 1)^2 / m$.

Remarks

In the special case where G is the union of two spanning trees the minimum expected intersection $(n - 1)/2$ is easily obtained using Theorem 3, by noting that in this case $v = n - 1$ and $\mu = 2(n - 1)$. Observe that it is equal to the value of the game in the 2-player game on this graph introduced above.

For $G = K_n$, we have $v = n - 1$ and $\mu = \binom{n}{2}$, according to Theorem 3 the minimum expected intersection is $2(n - 1)/n$, which is the value of the game in the 2-player game on this graph introduced above.

Experiments

In this section, through numerical experiments, we investigate how the experimentally calculated intersection values (using Monte-Carlo simulations) differ from the previously derived minimum expected intersection. To provide a simple metric that shows how likely a network’s random spanning trees are to intersect, compared to the minimum expected intersection, we use the following normalized score:

³ As we mentioned, the value $p(x)$ is polynomially computable if \mathcal{F} is the set of spanning trees of a graph G .

$$RTI = \frac{\text{observed mean} - \text{min. expected}}{\text{maximum} - \text{min. expected}} = \frac{\text{observed mean} - \frac{(n-1)^2}{m}}{n - \frac{(n-1)^2}{m}}$$

RTI takes values between 0 and 1. $RTI = 0$ means that the empirical mean intersection is equal to the minimum expected intersection, while $RTI = 1$ if and only if the network is a tree. Note that the maximum possible intersection of two spanning trees is $n - 1$ (when they are identical), but we used n in the formula to avoid division by zero.

We conducted our experiments using the following setup. For each network we randomly sampled 100 pairs of spanning trees with bootstrap sampling (i.e. a sample may be used multiple times) and computed the empirical mean intersection and the corresponding standard deviation value. This is called the *observed mean intersection*. The *minimum expected intersection*, according to the results of the previous section was calculated as $(n - 1)^2/m$.

Experiments on random model networks

At first we performed experiments on some synthetic networks including Erdős-Rényi random graphs $G(n, p)$ Erdős and Rényi (1960) (i.e. a graph of n nodes where the probability of drawing an edge between any pair of nodes is p) with $n = 100, 200, 500, 1000$ and $p = 0.1, 0.2, \dots, 1$; random regular graphs $RR(n, d)$ (i.e. a random graph of n nodes where each node has degree d) with $n = 100, 200, 500, 1000$ and $d = 3, 4, \dots, 9$; Preferential attachment networks $PA(n, k)$ Barabási and Albert (1999) (i.e. a network of n nodes created by a process where at each step a new node is added to the network and connected to k already existing nodes) with $n = 100, 200, 500, 1000$ and $k = 1, 2, \dots, 10$; Watts-Strogatz networks $WS(n, k, p)$ Watts and Strogatz (1998) (i.e. starting from a 4-regular network of n nodes each edge is rewired with probability p) with $n = 100, 200, 500, 1000, k = 4$ and $p = 0.1, 0.2, \dots, 1$. For each parameter we generated 10 graphs (e.g. 10 graphs for $G(1000, 0.1)$ or $BA(1000, 2)$, etc.). For each graph we sampled 100 pairs of spanning trees. Then the sample means and standard deviations were calculated.

The results for $n = 1000$ are shown in Fig. 1 and suggest that in case of random network models the minimum expected intersection value is equal to the observed (empirical) intersection value obtained by the simulations. Similar conclusions can be drawn for $n = 100, 200, 500$. For almost all cases the standard deviation was close to zero confirming the robustness of the simulations. An interesting observation is that heterogeneous degree distribution (e.g. the power-law), by itself does not indicate a higher spanning tree intersection value as minimally expected. A deeper investigation of the intersection value as the function of the model graph’s parameter (such as the parameter p for $WS(n, k, p)$, or power-law exponent of PA graphs) could be the subject of another study. For example, since $m \propto \binom{n}{2} p$ for $G(n, p)$ we expect an intersection size $2/p$, which is confirmed by our simulations (Fig. 1 top right).

In this work we are more interested in that how the minimum expected value differs from the real one in the case of real complex networks. In the next section we present our experiments considering networks with various global structural characteristics.

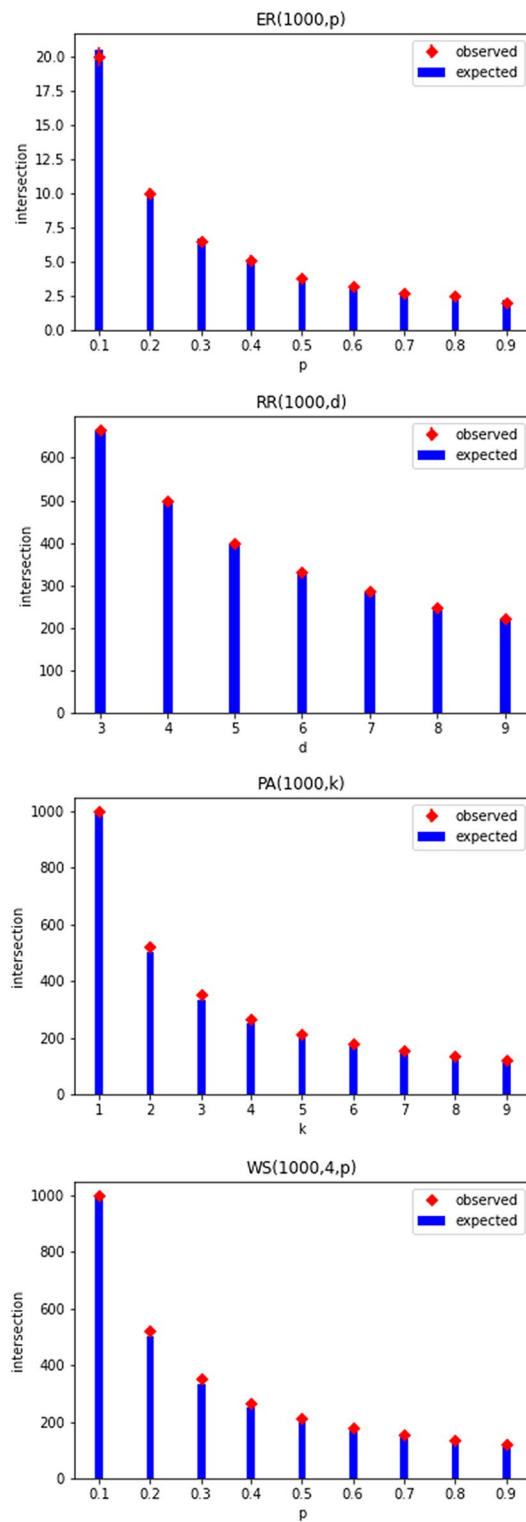


Fig. 1 Experimental results on synthetic networks: Erdős–Rényi, random regular, Barabási–Albert and Watts–Strogatz (top-down) with 1000 nodes each. Red diamonds (with red lines indicating standard deviation) show the observed results based on bootstrap Monte-Carlo experiments, while blue bars show the minimum expected intersection values

Table 1 Network statistics and random spanning tree intersection values of some real-world network. Data sources: Mark Newman's network data, Network Repository, Kunegis (2013), Stanford Large Network Dataset Collection

network	n	m	ρ	cc	$\bar{\ell}$	Min. exp.	Obs. mean	RTI
Florentine	15	20	0.190	0.191	2.486	9.8	10.5 (0.888)	0.166
Zachary	34	78	0.140	0.256	2.408	13.961	15.00 (2.291)	0.052
Dolphins	62	159	0.084	0.309	3.357	23.40	28.34 (2.19)	0.131
Iceland	75	114	0.041	0.157	3.200	48.035	57.21 (2.08)	0.353
Adjnoun	112	425	0.068	0.157	2.535	29.00	38.8 (4.3)	0.120
Jazz	198	2742	0.141	0.520	2.235	14.15	23.51 (3.66)	0.051
C-elegans	297	2148	0.049	0.180	2.455	40.79	57.8 (5.5)	0.067
NetSci	379	914	0.013	0.431	6.042	156.328	182.12 (7.98)	0.116
Wiki-Vote	889	2914	0.007	0.127	4.096	270.61	430.47 (9.66)	0.259
EuroRoad	1039	1305	0.002	0.035	18.395	825.63	853.5 (7.07)	0.131
Polblogs	1222	16717	0.022	0.226	2.737	89.18	300 (10.55)	0.186
Arxiv GR-QC	4158	13428	0.002	0.629	6.053	1287	2102 (21.65)	0.284
Erdős	4491	7428	0.0006	0.042	5.51	3352	4160 (17.45)	0.493
Power grid	4941	6594	0.0005	0.103	18.99	3700.87	3921 (16.71)	0.178
Email-Enron	33696	180811	0.0003	0.085	4.025	6279.225	16602 (64.73)	0.376

Experiments on real-networks

To perform experiments on real-world networks we considered the largest connected component in the case of unconnected networks and we did not consider the direction and/or weight of the edges in the case of directed and/or weighted networks, respectively.

We compared the RTI score with some metrics that are commonly used as indicators of the small-world structure of a network. These are the network density $\rho = m/\binom{n}{2}$, the clustering coefficient $cc = 3 \times \text{number of triangles}/\binom{n}{3}$ and the average shortest path length $\bar{\ell} = 1/\binom{n}{2} \cdot \sum_{i \neq j} \ell_{ij}$, where ℓ_{ij} is the length of the shortest path between nodes i and j .

Table 1 shows the results on 15 real-world complex networks including social, biological and technological networks as well (the network data are available on the websites Mark Newman's network data, Network Repository, Kunegis 2013, and Stanford Large Network Dataset Collection). The network size varies from $n = 15$ (marriage ties between Florentine families) to $n = 33,696$ (Enron email communication network). The average shortest path length varies between $\bar{\ell} = 2.235$ (collaboration network of jazz musicians) and $\bar{\ell} = 6.053$ (collaboration network of the fields general relativity and quantum computing based on Arxiv articles), except for the US power grid and EuroRoad technological networks with $\bar{\ell} = 19$ and $\bar{\ell} = 18.4$, respectively. Almost all networks have a relatively high clustering coefficient providing a strong small-world property of the studied networks.

The highest RTI scores were obtained for the Erdős co-authorship network (RTI=0.493) and the Enron email network (RTI=0.376), for which the difference between the empirically observed mean and the minimum expected intersection value is more than 10,000, i.e. more than 5% of the total number of edges. Figure 2 shows the RTI values in terms of the value (observed mean - min. expected)/ m , i.e. the difference between the real tree intersection value and the expected lower bound as a percentage of the total number of edges in

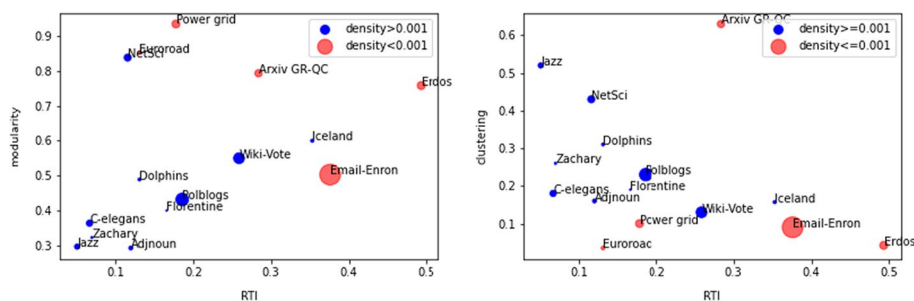


Fig. 2 Relationship between RTI and modularity (left) and RTI and clustering (right) on the investigated real-world dataset. Dot size is adjusted according to the number of nodes

the network. The Iceland sexual contact network, the Arxiv collaboration network and the Wiki-Vote network also provided high RTI (equals 0.353, 0.284 and 0.259, resp.).

Spanning tree intersection and modularity

Comparing RTI and Newman modularity scores provides another interesting aspect of random tree intersection. Modularity is calculated as

$$Q = \frac{1}{2m} \sum_{i < j} \left[a_{ij} - \frac{d_i d_j}{2m} \right] \delta(C_i, C_j),$$

given a partition \mathcal{C} of the nodes, such that $C_i \in \mathcal{C}$ is the class to which node i belongs, a_{ij} is the element of the network’s adjacency matrix indicating whether nodes i and j are connected or not, d_i is the degree (number of connections) of node i , while $\delta(C_i, C_j) = 1$ if i and j belong to the same class and 0 otherwise. Q is computed by considering the community structure (partition) obtained by the “Leuven” fast greedy method (Blondel et al. 2008).

Among the investigated networks the Word adjacencies (Adjnoun), the Jazz musician, the Zacharay and the C -elegans networks show the lowest RTI score (in the range 0.051–0.12) and the lowest modularity score (in the range 0.293–0.364) as well. On the other hand, Erdös, Email-Enron, Iceland and Arxiv GR-QC (have RTI scores that are among the highest) have high modularity scores between 0.504 and 0.793, see Fig. 2. However, Powergrid, Euroroad and NetSci networks have the highest modularities (0.838–0.934), but we need to note here that, according to Good et al. (2010), increasing n (or increasing number of clusters) generally tend to increase $\max Q$. These suggest that not only modularity, but probably also the density and the clustering coefficient correlates with RTI, the former positively (0.34, on the investigated dataset), while the latter two negatively (−0.5 and −0.33, resp.). A Deeper analysis on correlations and relationship of RTI to other metrics remains the topic of a future study.

Conclusions

In this work we investigated that how randomly chosen spanning trees are likely to intersect in order to gain some new insights on the macroscopic structure of complex networks. First, we derived a formula for the lower bound of the expected value of the intersection (i.e. the number of common edges of two randomly chosen spanning trees)

as a function of the number of nodes and edges of the network. We compared this value with the empirical (real) intersection value (obtained by simulation experiments) and found that in case of random model networks, such as the preferential attachment or the Watts-Strogatz small world network model, there is no significant difference between the two values.

On the other hand, and more interestingly, experiments show that for some real networks, the empirical mean intersection is very different from the expected minimum, suggesting the existence of special links that are likely to appear in most spanning trees of the network. Comparing the networks' Newman modularity with the introduced random tree intersection (RTI) score we observed a positive correlation on the investigated network dataset and we hypothesise that network size, density and modularity are the main drivers of the RTI value. This is consistent with the existence of "weak links" (Granovetter 1973), connecting separate clusters (or modules) of the network, as they are often over-represented in the spanning trees of the network. Looking at RTI in this context, it could be seen as a measure of network resilience, as the edges that appear more often in the spanning trees are more crucial for the connectivity of the network. This suggest at least two direction of future work. Firstly, the relationship between the probability of an edge being in a randomly chosen spanning tree and the edge between value seems interesting. Secondly, network connectivity and resilience are worth examining from this perspective. Other possibilities include deriving the exact bounds of spanning tree intersection for random model networks, and examining the relationship between the spanning tree intersection value and widely used global structural metrics of complex networks in more detail.

Acknowledgements

The authors would like to thank the anonymous referee for carefully reading the manuscript.

Author contributions

All authors contributed to the conception and design of the study.

Funding

AL is partially supported by the National Research, Development and Innovation Office - NKFIH, Fund No. SNN-13564.

Availability of data and materials

The data and all implementations in Python language used in the current study are available from the corresponding author on reasonable request.

Declarations

Competing interests

The authors declare that they have no competing interests.

Received: 21 February 2023 Accepted: 28 September 2023

Published online: 13 October 2023

References

- Albin N, Clemens J, Hoare D, Poggi-Corradini P, Sit B, Tymochko S (2021) Fairest edge usage and minimum expected overlap for random spanning trees. *Discrete Math* 344(5):112282
- Alon N, Karp RM, Peleg D, West D (1995) A graph-theoretic game and its application to the k-server problem. *SIAM J Comput* 24(1):78–100
- Barabási A-L, Albert R (1999) Emergence of scaling in random networks. *Science* 286(5439):509–512
- Bartal Y (1996) Probabilistic approximation of metric spaces and its algorithmic applications. In: *Proceedings of 37th conference on foundations of computer science*. IEEE, pp 184–193
- Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. *J Stat Mech Theory Exp* 2008(10):10008
- Broder AZ (1989) Generating random spanning trees. *FOCS* 89:442–447

- Chen C, Morris S (2003) Visualizing evolving networks: minimum spanning trees versus pathfinder networks. In: IEEE symposium on information visualization 2003 (IEEE Cat. No. 03TH8714). IEEE, pp 67–74
- Chvatal V (1983) Linear programming. Macmillan, New York
- Ellens W, Spieksma FM, Van Mieghem P, Jamakovic A, Kooij RE (2011) Effective graph resistance. *Linear Algebra Appl* 435(10):2491–2506
- Erdős P, Rényi A et al (1960) On the evolution of random graphs. *Publ Math Inst Hung Acad Sci* 5(1):17–60
- Erdős P, Selfridge JL (1973) On a combinatorial game. *J Comb Theory Ser A* 14(3):298–301
- Fakcharoenphol J, Rao S, Talwar K (2003) A tight bound on approximating arbitrary metrics by tree metrics. In: Proceedings of the thirty-fifth annual ACM symposium on theory of computing, pp 448–455
- Gautier G, Polito G, Bardenet R, Valko M (2019) DPPy: DPP sampling with python. *J Mach Learn Res* 20:180–1
- Good BH, De Montjoye Y-A, Clauset A (2010) Performance of modularity maximization in practical contexts. *Phys Rev E* 81(4):046106
- Granovetter MS (1973) The strength of weak ties. *Am J Sociol* 78(6):1360–1380
- Gueye A, Walrand JC, Anantharam V (2010) Design of network topology in an adversarial environment. In: International conference on decision and game theory for security. Springer, pp 1–20
- Kim D-H, Noh JD, Jeong H (2004) Scale-free trees: the skeletons of complex networks. *Phys Rev E* 70(4):046126
- Kottegoda RN (2020) Spanning tree modulus and secure broadcast games. Kansas State University, Manhattan, Kansas
- Kunegis J (2013) KONECT—the Koblenz network collection. In: Proceedings of the international conference on world wide web companion, pp 1343–1350. <http://dl.acm.org/citation.cfm?id=2488173>
- Lehman A (1964) A solution of the Shannon switching game. *J Soc Ind Appl Math* 12(4):687–725
- London A, Pluhár A (2023) Intersection of random spanning trees in small-world networks. In: Cherifi H, Mantegna RN, Rocha LM, Cherifi C, Micciche S (eds) *Complex networks and their applications XI*. Springer, Cham, pp 337–345
- Ma B, Hero A, Gorman J, Michel O (2000) Image registration with minimum spanning tree algorithm. In: Proceedings 2000 international conference on image processing (Cat. No. 00CH37101), vol 1. IEEE, pp 481–484
- Mantegna RN (1999) Hierarchical structure in financial markets. *Eur Phys J B Condens Matter Complex Syst* 11(1):193–197
- Mark Newman's network data. <http://www-personal.umich.edu/~mejn/netdata/>. Accessed 02 Feb 2023
- Network Repository. <https://networkrepository.com/>. Accessed 02 Feb 2023
- Stanford Large Network Dataset Collection. <https://snap.stanford.edu/data/>. Accessed 02 Feb 2023
- Tola V, Lillo F, Gallegati M, Mantegna RN (2008) Cluster analysis for portfolio optimization. *J Econ Dyn Control* 32(1):235–258
- Watts DJ, Strogatz SH (1998) Collective dynamics of 'small-world' networks. *Nature* 393(6684):440–442
- Wilson DB (1996) Generating random spanning trees more quickly than the cover time. In: Proceedings of the twenty-eighth annual ACM symposium on theory of computing, pp 296–303
- Wu BY, Chao K-M (2004) *Spanning trees and optimization problems*. Chapman and Hall, New York
- Xu Y, Uberbacher EC (1997) 2d image segmentation using minimum spanning trees. *Image Vis Comput* 15(1):47–57

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)
