On the relationship between PageRank and automorphisms of a graph

Modjtaba Ghorbani[a,1], Matthias Dehmer[b,c], Abdullah Lotfi[a], Najaf Amraei[a], Abbe Mowshowitz[d], Frank Emmert-Streib[e,f]

[a] Department of Mathematics, Faculty of Science, Shahid Rajaee Teacher Training University, Tehran, 16785-136, I. R. Iran

[b]Swiss Distance University of Applied Sciences, Department of Computer Science, Brig, Switzerland

[c]Department of Biomedical Computer Science and Mechatronics, UMIT, Hall in Tyrol, Austria

[d]Department of Computer Science, The City College of New York (CUNY), New York, NY, USA

[e]Predictive Society and Data Analytics Lab, Faculty of Information Technology and Communication Sciences, Tampere University, Finland

[f]Institute of Biosciences and Medical Technology, Tampere 33520, Finland

---

Abstract

PageRank is an algorithm used in Internet search to score the importance of web pages. The aim of this paper is demonstrate some new results concerning the relationship between the concept of PageRank and automorphisms of a graph. In particular, we show that if vertices u and v are similar in a graph G (i.e., there is an automorphism mapping u to v), then u and v have the same PageRank score. More generally, we prove that if the PageRanks of all vertices in G are distinct, then the automorphism group of G consists of the identity alone. Finally, the PageRank entropy measure of several kinds of real-world networks and all trees of orders 10-13 and 22 is investigated.

Keywords: PageRank, eigenvalue, graph automorphism, web page.

2010 AMS: 05C70, 05C07, 05C35, 92E10.

---

[1]Author to whom correspondence should be addressed(E-mail: mghorbani@sru.ac.ir)

1. Introduction

The eigenvalues and eigenvectors of the adjacency matrix of a graph offer necessary conditions for a graph to possess certain properties. In particular, they have been found useful in studies of graphs associated with web searches. The world wide web can be modeled as a directed graph in a natural way by interpreting web pages as vertices and links between web pages as directed edges in the graph.This model provides a basis for ranking web pages by means of the PageRank (PR) algorithm. The algorithm was developed by Brin and Page in 1998 [2].

The PageRank (PR) algorithm provides a mechanism for scoring the importance of web pages. PR has applications in such diverse fields such as neuroscience [32], bioinformatics [16, 30], sports [3, 25], traffic modeling [5, 29], chemistry [28] and social network analysis [12, 23], as well as others [22, 26]. Also, PR has been used extensively for improving the quality of search engines such as google and so forth, see [5]

The research reported here is especially relevant for chemical database applications. Searching for compounds with special properties can be aided by making use of page rank, and the automorphism group is useful for computing page rank. For other possible applications of the results in this paper (see [4]).

In this paper, we establish connections between the PageRank concept and automorphisms of a graph. The motivation to do so is to get deeper insights into graph-theoretical properties of graphs (here symmetry) in conjunction with PR. First, we define the PageRank (PR) vector and show how it can be computed. In Section 3, we establish new results concerning the concept of PageRank and automorphisms of a graph. In section 4, the PageRank entropy measure is defined In other words, analyzing the reported data shows that the PR-entropy measure is not highly correlated with the size of automorphism group and hence it can be regarded as a new measure to study the algebraic properties of the automorphism group.

Finally, in Section 5, we define the notion of a Co-PageRank graph and offer a conjecture concerning PageRank scores of vertices in non-Co-PageRank graphs. The notation used in this paper mainly follows [24].

2. PageRank Vector

The following discussion makes use of the model of the web as a directed graph. Let n be the number of all web pages, and suppose $P_{n\times n}$ is the Markov transitions matrix associated with the web graph defined as follows:

$$p_{ij} = \begin{cases} \frac{1}{d_i} & \text{if page } i \text{ and page } j \text{ are linked} \\ 0 & \text{otherwise} \end{cases},$$

where d_i is the degree of vertex i. In other words, p_{ij} is the probability of navigating from vertex i to vertex j. For a dangling vertex (one with outdegree 0), a zero row appears in the matrix P which violates the condition of a transition matrix. To overcome this violation and obtain a transition matrix, we define $P + lu^T$ where u is the probability distribution vector, $u = [1/n, 1/n, ..., 1/n]^T$, and l is an n-dimensional vector as follows:

$$l_i = \begin{cases} 1 & \text{if } i \text{ is a dangling node} \\ 0 & \text{otherwise} \end{cases}.$$

A PR vector [24], is an n-dimensional vector π satisfying the following:

$$\begin{cases} \pi^T = \pi^T \widetilde{G} \\ \pi^T \mathbf{j} = 1 \end{cases}, \tag{1}$$

where $\widetilde{G} = \alpha(P + lu^T) + (1-\alpha)\mathbf{j}v^T$, $\mathbf{j} = [1, 1, \ldots, 1]$ and $\alpha \in (0, 1)$ (typically $\alpha = 0.85$). In the present paper, we focus on graphs without dangling vertices. Hence, the vector π can be derived from the following equation:

$$\pi^T = \alpha\pi^T P + (1-\alpha)v^T, \tag{2}$$

or equivalently,

$$(I - \alpha P^T)\pi = (1-\alpha)v, \tag{3}$$

in which $v = [1/n, 1/n, ..., 1/n]^T$.

The PageRank (PR) score of vertex i is the ith entry of the vector π [6]. An example will help to fix ideas.

The Google matrix $\widetilde{G}$ of a directed network is a stochastic square matrix with non-negative matrix elements and the sum of elements in each column being equal to unity. By above notation, the elements of the Google matrix are defined as

$$\widetilde{G}_{ij} = \alpha P_{ij} + (1-\alpha)\frac{1}{n}.$$

Proposition 2.1. *[24] If $\{1, \mu_2, \ldots, \mu_n\}$ are all eigenvalues of transitions matrix P, then $\{1, \alpha\mu_2, \ldots, \alpha\mu_n\}$ are all eigenvalues of $\widetilde{G}$.*

Let G be a graph with adjacency eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_n$. The graph energy of G is defined as

$$\mathcal{E}(G) = \sum_{i=1}^{n} |\lambda_i|,$$

see [18, 19, 20, 21]. Following Gutman definition, if $\{1, \mu_2, \ldots, \mu_n\}$ are all eigenvalues of transitions matrix P, then the transition energy can be defined as

$$\mathcal{E}(P) = \sum_{i=1}^{n} |\mu_i|.$$

Corollary 2.1. *Suppose G is a graph with transitions matrix P. Then*

$$\mathcal{E}(\widetilde{G}) = \alpha(\mathcal{E}(P) - 1) + 1.$$

Proof. By Proposition 2.1, the proof is straightforward.

□

Example 2.1. *The following is the adjacency matrix of the graph G_1 in Figure 1.*

$$A = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \end{bmatrix},$$

The transition matrix of this graph is

$$P = \begin{bmatrix} 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 1/3 & 0 & 1/3 & 0 & 1/3 \\ 1/2 & 0 & 0 & 1/2 & 0 \end{bmatrix}.$$

With $\alpha = 0.85$ Eq. 3 gives the following linear system

$$\begin{cases} \pi_1 - \frac{0.85}{3}\pi_4 - \frac{0.85}{2}\pi_5 &= 0.03 \\ \pi_2 - \frac{0.85}{2}\pi_3 &= 0.03 \\ -0.85\pi_2 + \pi_3 - \frac{0.85}{3}\pi_4 &= 0.03 \\ -\frac{0.85}{2}\pi_1 - (\frac{0.85}{2})\pi_3 + \pi_4 - (\frac{0.85}{2})\pi_5 &= 0.03 \\ -(\frac{0.85}{2})\pi_1 - (\frac{0.85}{3})\pi_4 + \pi_5 &= 0.03 \end{cases}. \tag{4}$$

Solving Eq. 4 we obtain the PR vector of G_1:

$$PR = [0.1918, 0.1204, 0.2126, 0.2834, 0.1918].$$

2.1. PageRank score of a vertex

The concept of PageRank score at a vertex is needed to determine the relationship between PageRank and automorphisms of a graph.

Definition 2.1. *Let $A = (a_{ij})$ be an $n \times n$ matrix. Then the* 1*-Norm of matrix A is defined as [27]*

$$||A||_1 = Max_j \sum_{i=1}^{n} |a_{ij}|.$$

Definition 2.2. *The spectral radius $\rho(A)$ of an square matrix A is the largest absolute value of eigenvalues of A, see [27].*

Theorem 2.1. *[1] Let A be an arbitrary square matrix. Then*

$$\rho(A) \leq ||A||_1.$$

Theorem 2.2. *[1] (Geometric series) Let A be an square matrix. If $\rho(A) < 1$, then $(I - A)^{-1}$ exists, and it can be expressed as a convergent series,*

$$(I - A)^{-1} = I + A + A^2 + \cdots + A^m + \cdots = \sum_{k=0}^{\infty} A^k. \tag{5}$$

Lemma 2.1. *Let G be a graph of order n and π be the PR vector of G. The PR of vertex v_i can be determined from the following equation:*

$$\pi_i = \frac{(1-\alpha)}{n} \sum_{k=0}^{\infty} \alpha^k \sum_{t=1}^{n} P_{ti}^k, \tag{6}$$

where $\alpha \in (0, 1)$ and P is the transition matrix.

Proof. Since $\sum_{i=1}^{n} P_{ij}^T = 1$, we see that $||P^T||_1 = 1$. Consequently, we have $||\alpha P^T||_1 = \alpha \cdot ||P^T||_1 = \alpha < 1$. Theorem 2.1 implies that $\rho(\alpha P^T) < 1$, and Theorem 2.2 implies that the inverse matrix $(I - \alpha P^T)^{-1}$ exists and thus

$$(I - \alpha P^T)^{-1} = \sum_{k=0}^{\infty} (\alpha P^T)^k. \tag{7}$$

From Eq. 3 and Eq. 7 we conclude that,

$$\begin{aligned}\pi &= (1-\alpha)(I - \alpha P^T)^{-1} v = (1-\alpha)(\sum_{k=0}^{\infty} (\alpha P^T)^k) v \\ &= (1-\alpha)(I + \alpha P^T + \alpha^2 {P^T}^2 + ...)v.\end{aligned}$$

Since π_i is the ith row of the matrix $(1-\alpha)(I - \alpha P^T)^{-1} v$, it is clear that π_i is the ith row of column matrix $\frac{(1-\alpha)}{n}(I - \alpha P^T)^{-1} e$. This means that

$$\pi_i = \frac{(1-\alpha)}{n} \sum_{t=1}^{n} \left[(I - \alpha P^T)_{it}^{-1} \right] = \frac{(1-\alpha)}{n} \sum_{t=1}^{n} \sum_{k=0}^{\infty} (\alpha P^T)_{it}^k. \tag{8}$$

Hence

$$\pi_i = \frac{(1-\alpha)}{n} \sum_{k=0}^{\infty} \alpha^k \sum_{t=1}^{n} P_{ti}^k. \tag{9}$$

□

According to the definition of matrix P, P_{ti}^{k} is the transition probability from vertex t to vertex i in k steps.

Theorem 2.3. *Let G be a graph and $i, j \in V(G)$. If $\sum\limits_{t=1}^{n} P_{ti}^{k} = \sum\limits_{t=1}^{n} P_{tj}^{k}$, (for all $k \in \mathbb{N}$), then $\pi_i = \pi_j$.*

Proof. Suppose $\sum\limits_{t=1}^{n} P_{ti}^{k} = \sum\limits_{t=1}^{n} P_{tj}^{k}$ and $\alpha \in (0,1)$. Then

$$\sum_{t=1}^{n} \alpha^k P_{ti}^{k} = \sum_{t=1}^{n} \alpha^k P_{tj}^{k} \quad \text{(for all } k \in \mathbb{N}),$$

and consequently

$$\sum_{k=0}^{\infty} \alpha^k \sum_{t=1}^{n} P_{ti}^{k} = \sum_{k=0}^{\infty} \alpha^k \sum_{t=1}^{n} P_{tj}^{k}.$$

Eq. 9 implies that

$$\pi_i = \frac{(1-\alpha)}{n} \sum_{k=0}^{\infty} \alpha^k \sum_{t=1}^{n} P_{ti}^{k} = \frac{(1-\alpha)}{n} \sum_{k=0}^{\infty} \alpha^k \sum_{t=1}^{n} P_{tj}^{k} = \pi_j.$$

□

In light of Theorem 2.3, consider the tree T_1 shown in Figure 2.

Example 2.2. *The sums of the entries in each column of matrices P, P^2, P^3, respectively, of graph T_1, are shown in the end of each column. Consider also the vertices $1, 2$ or $3, 4$ of T_1 and their corresponding columns in matrices P, P^2, P^3 as follows:*

$$A = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}, \quad P = \left[\begin{array}{cccccc} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 1/2 & 0 & 0 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 0 & 1/2 & 0 \\ 0 & 0 & 1/3 & 1/3 & 0 & 1/3 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ \hdashline 1/2 & 1/2 & 4/3 & 4/3 & 2 & 1/3 \end{array}\right],$$

$$P^2 = \left[\begin{array}{cccccc} 1/2 & 0 & 0 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 0 & 1/2 & 0 \\ 0 & 0 & 2/3 & 1/6 & 0 & 1/6 \\ 0 & 0 & 1/6 & 2/3 & 0 & 1/6 \\ 1/6 & 1/6 & 0 & 0 & 2/3 & 0 \\ 0 & 0 & 1/3 & 1/3 & 0 & 1/3 \\ \hdashline 2/3 & 2/3 & 7/6 & 7/6 & 5/3 & 2/3 \end{array}\right],$$

$$P^3 = \left[\begin{array}{cccccc} 0 & 0 & 2/3 & 1/6 & 0 & 1/6 \\ 0 & 0 & 1/6 & 2/3 & 0 & 1/6 \\ 1/3 & 1/12 & 0 & 0 & 7/12 & 0 \\ 1/12 & 1/3 & 0 & 0 & 7/12 & 0 \\ 0 & 0 & 7/18 & 7/18 & 0 & 2/9 \\ 1/6 & 1/6 & 0 & 0 & 5/9 & 0 \\ \hdashline 7/12 & 7/12 & 22/18 & 22/18 & 72/36 & 5/9 \end{array}\right].$$

The sums of columns $1, 2$ *or* $3, 4$ *of* P^k *(for all* k *) are the same, and thus the PR scores of corresponding vertices are the same. This means that*

$$\sum_{t=1}^{6} P_{t1}^k = \sum_{t=1}^{6} P_{t2}^k,$$

and thus $\pi_1 = \pi_2$. *A similar argument shows that* $\pi_3 = \pi_4$. *Hence, the PR vector of this tree is*

$$\pi^T = [0.1090, 0.1090, 0.1975, 0.1975, 0.2821, 0.1049].$$

3. PageRank Vector and Graph Automorphisms

An identity graph or asymmetric graph is a graph whose automorphism group consists of the identity element alone. An example of such a graph is T_2 shown in Figure 4. Note that all entries of the PR vector π of this graph are distinct. The aim of this section is to prove that if the PageRank scores of all vertices are distinct, then the graph must be asymmetric.

Lemma 3.1. *Every vertex v_i in a regular graph G of order n has PR score $\pi_i = \frac{1}{n}$.*

Proof. Let G be a regular graph of degree r. Then for every vertex $v_i \in V(G)$, we have

$$\sum_{t=1}^{n} P_{ti} = r.\frac{1}{r} = 1,$$

and

$$\sum_{t=1}^{n} P_{ti}^2 = 1.\frac{1}{r}.r = 1.$$

Hence, for each $k \in \mathbb{N}$,

$$\sum_{t=1}^{n} P_{ti}^k = 1. \tag{10}$$

Using Eq. 10 and Lemma 2.1 implies that

$$\begin{aligned} \pi_i &= \frac{1-\alpha}{n} \sum_{k=0}^{\infty} \alpha^k \sum_{t=1}^{n} P_{ti}^k \\ &= \frac{1-\alpha}{n} \sum_{k=0}^{\infty} \alpha^k = \frac{1-\alpha}{n}(\frac{1}{1-\alpha}) \\ &= \frac{1}{n} \end{aligned}$$

This completes the proof. □

Let T be a tree on n vertices, and denote the degree of a vertex v by d_v. A non-pendant vertex v of T is adjacent to $d_v > 1$ vertices in T.

Theorem 3.1. *Let i, j be two vertices in a graph G. If there exists an automorphism $\psi \in Aut(G)$ such that $\psi(i) = j$, then $\pi_i = \pi_j$.*

Proof. Suppose N_i denotes the set of neighbors of vertex i, namely $N_i = \{t \in V | ti \in E\}$, see Figure 5. For every vertex i_1 in N_i there is a vertex

$j_1 \in N_j$ such that $\psi(i_1) = j_1$. Since i_1 and j_1 are similar, $d_{i_1} = d_{j_1}$, and $P_{i_1 i} = \frac{1}{d_{i_1}} = \frac{1}{d_{j_1}} = P_{j_1 j}$. Hence,

$$\sum_{t=1}^{n} P_{ti} = \sum_{t=1}^{n} P_{tj}.$$

Continuing the method illustrated in Figure 6, for given vertex $i_2 \in N_{i_1}$, there exists a vertex $j_2 \in N_{j_1}$ such that $\psi(i_2) = j_2$, since ψ maps the edge B_i to B_j. This implies that $d_{i_2} = d_{j_2}$ and thus $P_{i_2 i_1} = \frac{1}{d_{i_2}} = \frac{1}{d_{j_2}} = P_{j_2 j_1}$. Therefore, $(P^2)_{i_2 i} = (P^2)_{j_2 j}$ and thus,

$$\sum_{t=1}^{n} P_{ti}^2 = \sum_{t=1}^{n} P_{tj}^2. \tag{11}$$

In general, we have,

$$\sum_{t=1}^{n} P_{ti}^k = \sum_{t=1}^{n} P_{tj}^k \quad (k \in N). \tag{12}$$

From Eq. 12 and Theorem 2.3 it follows that $\pi_i = \pi_j$, and the assertion is proved. □

Theorem 3.1 says that if an automorphism maps a vertex x to vertex y, they must have the same PR score. However, the converse does not hold. A counterexample is the Frucht graph shown in Figure 7. The Frucht graph is regular of degree 3 with 12 vertices and 18 edges and is asymmetric, see [13]. Since it is a regular graph, Lemma 3.1 shows the PR-vector is $[1/12, \ldots, 1/12]$, while the automorphism group of this graph consists of the identity element alone.

In what follows, we prove that a graph whose vertices have distinct PageRank scores is asymmetric. First, consider the following example.

Example 3.1. *The following is the adjacency matrix of the tree T_2 shown in Figure 4:*

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix},$$

In the matrix P associated with A, the sums of $4th$ and $7th$ columns are equal, but in P^2 and P^3 these column sums are not equal.

$$P = \left[\begin{array}{ccccccc} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1/3 & 0 & 1/3 & 0 & 1/3 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ \hdashline 1/3 & 2 & 4/3 & 1/2 & 5/6 & 3/2 & 1/2 \end{array}\right],$$

$$P^2 = \left[\begin{array}{ccccccc} 1/3 & 0 & 1/3 & 0 & 1/3 & 0 & 0 \\ 0 & 2/3 & 0 & 1/6 & 0 & 1/6 & 0 \\ 1/6 & 0 & 2/3 & 0 & 1/6 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 & 0 & 0 \\ 1/6 & 0 & 1/6 & 0 & 5/12 & 0 & 1/4 \\ 0 & 1/4 & 0 & 0 & 0 & 3/4 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 0 & 1/2 \\ \hdashline 2/3 & 7/12 & 7/6 & 4/6 & 17/12 & 11/12 & 3/4 \end{array}\right],$$

$$P^3 = \left[\begin{array}{ccccccc} 0 & 2/3 & 0 & 1/6 & 0 & 1/6 & 0 \\ 2/9 & 0 & 5/6 & 0 & 11/36 & 0 & 1/12 \\ 0 & 7/12 & 0 & 1/3 & 0 & 1/12 & 0 \\ 1/6 & 0 & 2/3 & 0 & 1/6 & 0 & 0 \\ 0 & 11/24 & 0 & 1/12 & 0 & 11/24 & 0 \\ 1/12 & 0 & 1/12 & 0 & 11/24 & 0 & 3/8 \\ 0 & 1/4 & 0 & 0 & 0 & 3/4 & 0 \\ \hdashline 17/36 & 47/24 & 19/12 & 7/12 & 67/72 & 35/24 & 11/24 \end{array}\right].$$

On the other hand, we have,

$$\sum_{t=1}^{7} P_{t4} = \sum_{t=1}^{7} P_{t7},$$

while

$$\sum_{t=1}^{7} P^2_{t4} \neq \sum_{t=1}^{7} P^2_{t7} \quad and \quad \sum_{t=1}^{7} P^3_{t4} \neq \sum_{t=1}^{7} P^3_{t7}.$$

The graph T_2 has no vertices for which corresponding column sums are the same. This means that their PR scores are not equal and the entries of the PR vector are all distinct. Finally, the PR vector of this tree is

$$\pi^T = [0.0878, 0.2343, 0.1660, 0.0920, 0.1592, 0.1680, 0.0928].$$

On the other hand, the automorphism group of T_2 consists of the identity element alone.

Corollary 3.1. *Let G be a graph. If the PR scores of all the vertices are distinct, then G is asymmetric.*

Proof. For two arbitrary vertices $u, v \in V(G)$, if $\pi_u \neq \pi_v$, then by Theorem 3.1, there is no an automorphism that maps u to v and the assertion follows. □

Corollary 3.2. *Let T be a tree in which no two pendant vertices have the same PR scores. Then the automorphism group of T consists of the identity element alone.*

Proof. For the non-identity automorphism ψ of $Aut(T)$, there are at least two pendant vertices i, j such that $\psi(i) = j$ and thus $\pi_i = \pi_j$. But the pendant vertices have different PR scores from which the result follows. $\square$

Definition 3.1. *Let G be a graph with automorphism group $Aut(G)$, and denote the orbit of a vertex $u \in V(G)$ by $u^{Aut(G)}$ or $[u]$. Note that $u^{Aut(G)}$ is the set $\{\alpha(u) : \alpha \in Aut(G)\}$.*

A graph G is called vertex-transitive, if it has exactly one orbit. In other words, for any two vertices $u, v \in V(G)$, there is an automorphism $\alpha \in Aut(G)$ such that $\alpha(u) = v$.

The PR complexity, $PR_C(G)$, is the number of different values of PR vector.

Theorem 3.2. *Let V_1, V_2, $V_3, \ldots, V_k$ be all the orbits of $Aut(G)$. Then for two vertices $x, y \in V_i (1 \leqslant i \leqslant k)$, $\pi_x = \pi_y$. In particular, if G is vertex-transitive, then $PR_C(G) = 1$.*

Proof. If two vertices are in the same orbit, there is an automorphism mapping one to the other. The assertion follows from Theorem 3.1. $\square$

Corollary 3.3. *Let $\#O$ be the number of distinct orbits of a graph G. Then*

$$PR_C(G) \leq \#O.$$

An illustration of this corollary is given by the tree T_1 shown in Figure 2. This graph has four orbits $\{1, 2\}$, $\{3, 4\}$, $\{5\}$ and $\{6\}$. By Theorem 3.2, $\pi_1 = \pi_2$ and $\pi_3 = \pi_4$. This means that the PR vector π has at most four distinct entries.

Example 3.2. *Suppose t denotes the number of orbits of graph G. It should be noted here that there are graphs with $k < t$. For example consider the graph K in Figure 3. This graphs has three orbits while $k = 2$, the vertices in an orbit are colored by the same colors.*

This example shows that determining graphs with $k = t$ is a hard task. We Solve this problem for graphs with exactly two orbits.

Lemma 3.2. *The connected graph G is regular if and only if $\pi = \lambda\mathbf{j}$, where $\lambda \in \mathbb{R}$.*

Proof. If G is regular, then by Lemma 3.1, $\pi = \frac{1}{n}\mathbf{j}$. Conversely, if $\pi = \lambda\mathbf{j}$ for a scaler $\lambda \in \mathbb{R}$, then all entries of π are the same. Since for two vertices v_i and v_j, we have

$$\pi_i - \pi_j = \alpha(\frac{\pi_j}{d_j} - \frac{\pi_i}{d_i}),$$

necessarily $d_i = d_j$ and thus the graph is regular. □

Theorem 3.3. *Let G be a graph with two distinct orbits. Then either G is a regular graph or $k = 2$.*

Proof. Since G has two orbits, it follows that $k \leq 2$. If $k \neq 2$, then by Lemma 3.2, G is regular. This completes the proof. □

Corollary 3.4. *Let G be an edge-transitive graph. Then either G is a regular graph or $k = 2$.*

Example 3.3. *Consider the complete graph $K_{m,n}(m \neq n)$. It is a well-known fact that $K_{m,n}$ has two orbitse. Since, $m \neq n$, by Theorem 3.3, we obtain $k = 2$. In addition, the matrix P associated to the adjacency matrix of G is*

$$P = \begin{pmatrix} 0_{n\times n} & \frac{1}{m}\mathbf{j}_{n\times m} \\ \frac{1}{n}\mathbf{j}_{m\times n} & 0_{m\times m} \end{pmatrix}.$$

Hence,

$$Spec(P) = \{-1, 0, 0, ..., 0, 1\}$$

and thus for the Google matrix, we have

$$Spec(\widetilde{G}) = \{1, 0, 0, ..., 0, -\alpha\}.$$

Example 3.4. *Let S_n denotes to the star graph with n vertices. The bistar graph $B_{m,n}$ is a graph obtained from union of S_{n+1} and S_{m+1} by joining their central vertices. For the star graph, we obtain*

$$P(S_{n+1}) = \begin{pmatrix} 0_{1\times 1} & \frac{1}{n}\mathbf{j}_{1\times n} \\ \mathbf{j}_{n\times 1} & 0_{n\times n} \end{pmatrix}.$$

This yields that $PR = [\pi_1, \pi_2, \ldots, \pi_2, \pi_2]$, *where* $\pi_1 = (\frac{1-\alpha}{n+1} + \alpha) \times \frac{1}{1+\alpha}$ *and* $\pi_2 = \frac{n+\alpha}{n(n+1)(1+\alpha)}$. *Also, for the bistar graph, it yields*

$$P(B_{m,n}) = \begin{pmatrix} \mathcal{A} & \mathcal{B} \\ \mathcal{B}^T & \mathcal{C} \end{pmatrix},$$

where $\mathcal{C} = \mathbf{0}_{m+n}$,

$$\mathcal{A} = \begin{pmatrix} 0 & \frac{1}{n+1} \\ \frac{1}{m+1} & 0 \end{pmatrix}, \text{and } \mathcal{B} = \begin{pmatrix} \frac{1}{n+1}\mathbf{j}_{1\times n} & \mathbf{0}_{1\times m} \\ \mathbf{0}_{1\times n} & \frac{1}{m+1}\mathbf{j}_{1\times m} \end{pmatrix}.$$

Lemma 3.3. *Let G be a graph and i, j be two distinct vertices having the same neighbors. Then $\pi_i = \pi_j$.*

Proof. Two following cases hold:

a) Suppose vertices i and j are adjacent. According to the definition of PR score, we have,

$$\pi_i = \alpha \sum_{k \in N_i - \{j\}} \frac{\pi_k}{d_k} + \alpha \frac{\pi_j}{d_j} + \frac{1-\alpha}{n},$$

and

$$\pi_j = \alpha \sum_{k \in N_j - \{i\}} \frac{\pi_k}{d_k} + \alpha \frac{\pi_i}{d_i} + \frac{1-\alpha}{n}.$$

Thus

$$\pi_i - \pi_j = \alpha(\frac{\pi_j}{d_j} - \frac{\pi_i}{d_i}),$$

and therefore

$$\pi_i(1 + \frac{\alpha}{d_i}) = \pi_j(1 + \frac{\alpha}{d_j}).$$

Since $|N_i| = |N_j|$, we have $d_i = d_j$ which implies $\pi_i = \pi_j$.

b) Now suppose i and j are not adjacent. Then $\pi_i = \alpha \sum_{k \in N_i} \frac{\pi_k}{d_k} + \frac{(1-\alpha)}{n}$ and $\pi_j = \alpha \sum_{k \in N_j} \frac{\pi_k}{d_k} + \frac{(1-\alpha)}{n}$. Since $N_i = N_j$, we conclude $\pi_i - \pi_j = 0$ and thus $\pi_i = \pi_j$. □

Lemma 3.4. *Let i, j be two adjacent vertices of a graph G. If $\pi_i < \pi_j$, then $N_j \nsubseteq N_i$*

Proof. Suppose to the contrary that $N_j \subseteq N_i$. Hence, we obtain

$$\pi_j = \alpha \sum_{k \in N_j} \frac{\pi_k}{d_k} + \frac{1-\alpha}{n} \leq \alpha \sum_{k \in N_i} \frac{\pi_k}{d_k} + \frac{1-\alpha}{n} = \pi_i,$$

a contradiction. □

Lemma 3.5. *Let G be a graph. If i is a pendant vertex adjacent to vertex j, then $\pi_i < \pi_j$.*

Proof. Clearly $d_j \geq 2$ and thus $-\frac{1}{d_j} \geq -\frac{1}{2}$. This implies

$$(\pi_i - \frac{1}{d_j}\pi_j) \geq (\pi_i - \frac{1}{2}\pi_j). \tag{13}$$

From the definition of PR and Eq. 13, we have

$$\pi_j = \alpha(\frac{\pi_i}{1}) + \alpha\Big(\sum_{\substack{k \in N_j \\ k \neq i}} \frac{\pi_k}{d_k}\Big) + \frac{1-\alpha}{n}, \quad \pi_i = \alpha(\frac{\pi_j}{d_j}) + \frac{1-\alpha}{n}.$$

Hence

$$
\begin{aligned}
\pi_j - \pi_i &= \alpha(\pi_i - \frac{1}{d_j}\pi_j) + \alpha \sum_{\substack{k \in N_j \\ k \neq i}} \frac{\pi_k}{d_k} + (\frac{1-\alpha}{n} - \frac{1-\alpha}{n}) \\
&\geq \alpha(\pi_i - \frac{1}{2}\pi_j) + \alpha \sum_{\substack{k \in N_j \\ k \neq i}} \frac{\pi_k}{d_k} \\
&= \alpha(\pi_i - \pi_j) + \frac{1}{2}\alpha\pi_j + \alpha \sum_{\substack{k \in N_j \\ k \neq i}} \frac{\pi_k}{d_k} \\
&> \alpha(\pi_i - \pi_j) + \frac{1}{2}\alpha\pi_j,
\end{aligned}
$$

and thus

$$(\pi_j - \pi_i) > \frac{\frac{1}{2}\alpha\pi_j}{1+\alpha} > 0. \tag{14}$$

□

4. Graph Entropy Measure

The general Shannon entropy [5] is defined by $I(p) = -\sum_{i=1}^{n} p_i \log(p_i)$ for finite probability vector p and the symbol log is the logarithm on the basis 2. Let $\Lambda = \sum_{j=1}^{n} \Lambda_j$ and $p_i = \Lambda_i/\Lambda, (i = 1, 2, \ldots, n)$. Generally, the entropy of an n-tuple $(\Lambda_1, \Lambda_2, \ldots, \Lambda_n)$ of real numbers is given by

$$I(\Lambda_1, \Lambda_2, \ldots, \Lambda_n) = \log\left(\sum_{i=1}^{n} \Lambda_i\right) - \sum_{i=1}^{n} \frac{\Lambda_i}{\sum_{j=1}^{n} \Lambda_j} \log \Lambda_i. \tag{15}$$

There are many different ways to associate an n-tuple $(\Lambda_1, \Lambda_2, \ldots, \Lambda_n)$ to a graph G (see [1, 6, 7, 8, 9, 10, 11, 14, 22, 24, 32]). A graph entropy measure due to PageRank vector [15] is defined as

$$I_\pi(G) = \log\left(\sum_{i=1}^{n} \pi_i\right) - \sum_{i=1}^{n} \frac{\pi_i}{\sum_{j=1}^{n} \pi_j} \log \pi_i. \tag{16}$$

This phrase reduces the complexity of the graph G into a single quantity: $I_\pi(G)$ bits of information. This means that the PR-entropy I_π, forms a simple and graceful discriminant statistic for determining the topology of a graph. This metric is the subject of the present section. The entropy function maximizes the freedom in choosing the p_{ij}'s. The theory tell us that the entropy function gives the best unbiased probability assignment to the variables given the restriction.

Example 4.1. *Consider the Karate graph $\mathcal{K}$ [31] as depicted in Figure 10. It has 34 vertices and 78 edges and the PageRank vector is as follows:*

$$\begin{aligned}\pi = \quad [\quad & 0.097, 0.053, 0.057, 0.036, 0.022, 0.029, 0.029, 0.024, 0.029, 0.014, 0.022,\\ & 0.009, 0.015, 0.029, 0.014, 0.014, 0.017, 0.014, 0.014, 0.019, 0.014, 0.015,\\ & 0.014, 0.031, 0.021, 0.021, 0.015, 0.026, 0.019, 0.026, 0.025, 0.037, 0.072,\\ & 0.101].\end{aligned}$$

The interpretation of $\pi_1 = 0.097$ is that 9.7 percent of the time the random surfer visits page 1. Therefore, the pages in this tiny web can be ranked by their importance. Hence, page 34 is the most important page and page 12 by $\pi_{12} = 0.009$ is the least important page, according to the PageRank definition of importance. Also its PR-entropy is $I_\pi(\mathcal{K}) = 4.78$.

Example 4.2. *Consider the graph $\mathcal{G}$ as depicted in Figure 11. It presents a typical arrangement of symmetric subgraphs found in many real world networks. It has 33 vertices and 37 edges. The PageRank vector is as follows:*

$$\begin{aligned}\pi = \quad [\quad & 0.04, 0.031, 0.018, 0.031, 0.018, 0.064, 0.031, 0.031, 0.031, 0.04, 0.031,\\ & 0.016, 0.018, 0.035, 0.027, 0.075, 0.017, 0.017, 0.017, 0.017, 0.045, 0.046,\\ & 0.037, 0.015, 0.037, 0.015, 0.04, 0.046, 0.046, 0.017, 0.017, 0.017, 0.017].\end{aligned}$$

The PR-entropy for graph $\mathcal{G}$ is $I_\pi(\mathcal{G}) = 4.89$.

In continuing, five classes of trees of orders 10-13, and 22, were choosen and the results indicated a weak correlation between $|Aut(G)|$ and $I_\pi(G)$. These values are given in Figures 12,13, Figure 14, 15, and Figure 16. In other words, *analyzing the reported data shows that the PR-entropy measure is not highly correlated with the size of automorphism group and hence it can be regarded as a new measure to study the algebraic properties of the automorphism group.*

It is clear that if in the Shannon entropy definition, all p_i's are equal, then I_π achieves the maximum value which is $log(n)$. By Lemma 3.2, if G is regular, then $I_\pi = log(n)$. Graphs with minimum value of PR-entropy are more difficult to characterize. We conjecture that for a given number n, the star graph S_n has the minimum PR-entropy. To do this, three classes of graphs, namely all graphs of orders 5-6 and all trees of order 12 were choosen and the results confirm our following conjecture.

Conjecture 4.1. *Among all connected graphs on n vertices, the star graph S_n has the minimum value of PR-entropy.*

In [14], it is proved that if T is a tree with two orbits and $n \geq 3$ vertices, then T is isomorphic with either the star graph S_n or bistar graph $B_{m,m}$. By Example 3.4, we conclude the following result.

Theorem 4.1. *Let T be a tree with two orbits and $n \geq 3$ vertices. Then one of the following cases hold:*

i) $T \cong S_n$ *and* $I_\pi(T) \approx 0.55 \log n + 0.91$.

ii) $T \cong B_{m,m}$ *and* $I_\pi(T) \approx 0.6 \log n + 0.93$.

Many networks can be modeled as a star graph. For example, an inwardly directed star graph may be used to represent retweet activity on Twitter and an outwardly directed star graph can be used to represent a hub authority. One may see that the star graph is a special case of $G + \{u\}$ in which G is a vertex-transitive graph. Here, we explain how one can the PR-vector of $G + \{u\}$ by having the PR-vector of G.

Lemma 4.1. *Let G be an r-regular graph on n vertices. Then the PageRank vector of graph $G + \{u\}$ is $\pi = [\pi_1, \ldots, \pi_n, \pi_{n+1}]$, where $\pi_{n+1} = (\frac{1-\alpha}{n+1} + \frac{\alpha}{r+1}) \times (\frac{r+1}{\alpha+r+1})$ and $\pi_1 = \ldots = \pi_n = \frac{1-\pi_{n+1}}{n}$.*

Proof. Suppose G is a regular graph with $P(G)$ associated to its adjacency matrix. For an arbitrary vertex u, the matrix $\widetilde{P} = P(G + \{u\})$ can be regarded as follows:

$$\widetilde{P} = \begin{pmatrix} \frac{1}{r+1}A & \frac{1}{r+1}\mathbf{j}_{n\times 1} \\ \frac{1}{n}\mathbf{j}_{1\times n} & 0_{1\times 1} \end{pmatrix},$$

where A is the adjacency matrix of G. By replacing $\widetilde{P}$ with P in Eq. 3 the result follows. □

5. Co-PageRank Graphs

There exist non-isomorphic graphs with the same PR vectors; these graphs are said to be Co-PageRank (or Co-PR). For example, the two graphs G and H shown in Figure 8 have the same PR-vector, namely,

$$[0.185065, 0.185065, 0.129870, 0.185065, 0.185065, 0.129870].$$

but they are not isomorphic. In general, suppose $\alpha = \alpha_1, \cdots, \alpha_n$ and $\beta = \beta_1, \cdots, \beta_n$ the PR vectors of two graphs G and H, respectively, where $\alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_n$ and $\beta_1 \leq \beta_2 \leq \cdots \leq \beta_n$. If $\alpha = \beta$, then G and H are Co-PR; if , on the other hand, α and β differ in at least one entry, then G and H are non-Co-PR. Two graphs G and H are completely non-Co-PR if for each i $(1 \leq i \leq n)$ $\alpha_i \neq \beta_i$. For example, the two graphs L and K shown in Figure 9, are non-Co-PR, with

$$PR(L) = [0.143736, 0.209536, 0.143736, 0.209536, 0.146727, 0.146727],$$

$$PR(K) = [0.161121, 0.237500, 0.177757, 0.100546, 0.161121, 0.161954].$$

We end this paper with the following conjecture.

Conjecture 5.1. *Suppose G and H are two non-Co-PR graphs. Then for each vertex $u \in V(G)$ and each vertex $v \in V(H)$, $\pi_u \neq \pi_v$. More generally G and H are completely non-Co-PR.*

Conclusion

In this paper, we have investigated the relationship between the concept of PageRank and automorpisms of a graph. In particular, we proved that if the pendant vertices of a tree T have distinct PRs, then T is asymmetric. Results regarding symmetry relations for trees as well as graphs can be useful to design new graph measures. Moreover, we established conditions for which two distinct vertices of a graph have the same PageRank. The main result in this paper is that two vertices in the same orbit have the same PR score. As future work, we hope to determine the structure of automorphism groups of well-known graphs in terms of PR vectors.

Acknowledgements

The authors would like to thank Prof. Paolo Boldi for his valuable comments and suggestions to improve the quality of the paper.

References

[1] K. E. Atkinson, An introduction to numerical analysis, John Wiley & Sons, 2008.

[2] S. Brin, L. Page, Reprint of: The anatomy of a large-scale hypertextual web search engine, Comput. Netw. 56 (2012) 3825–3833.

[3] S. Brown, A pagerank model for player performance assessment in basketball, soccer and hockey, arXiv preprint arXiv:1704.00583.

[4] K. J. Burch, Chemical applications of graph theory, in Mathematical Physics in Theoretical Chemistry, S. M. Blinder and J. E. House, eds., Elsevier, 2018. https://doi.org/10.1016/B978-0-12-813651-5.00008-5.

[5] E. Crisostomi, S. Kirkland, R. Shorten, A Google-like model of road network dynamics and its application to regulation and control, Int. J. Control. 84 (2011) 633–651.

[6] M. Dehmer, F. Emmert-Streib, Y. Shi, Graph distance measures based on topological indices revisited, Appl. Math. Comput. 266 (2015) 623–633.

[7] M. Dehmer, F. Emmert-Streib, Y. Shi, Interrelations of graph distance measures based on topological indices, PLoS ONE. 9 (2014) e94985.

[8] M. Dehmer, A. Mehler, A new method of measuring similarity for a special class of directed graphs, Tatra Mt. Math. Publ. 36 (2007) 39–59.

[9] M. Dehmer, A. Mowshowitz, A history of graph entropy measures, Inf. Sci. Int. J. 181 (2011) 57–78.

[10] M. Dehmer, A. Mowshowitz, F. Emmert-Streib, Connections between classical and parametric network entropies, PLoS ONE, 6 (2011) e15733.

[11] M. Dehmer,Y. Shi, F. Emmert-Streib, Structural differentiation of graphs using Hosoya-based indices, PLoS ONE. 9 (2014) e102459.

[12] K. M. Frahm, D. L. Shepelyansky, Ising-PageRank model of opinion formation on social networks, Physica A. 526 (2019) 121069.

[13] R. A. Frucht, One-Regular Graph of Degree Three, Canad. J. Math. 4 (1952) 240-247.

[14] M. Ghorbani, M. Dehmer, M. Mowshowitz, J. Tao, F. Emmert-Streib, The Hosoya Entropy of Graphs Revisited, Symmetry . 9 (2019) 11, 1013; doi:10.3390/sym11081013.

[15] Giller, Graham L., The entropy of PageRank vectors and the construction of a discriminatory statistic for the identification of network topologies, Available at SSRN: https://ssrn.com/abstract=2239394.

[16] I. GJbor, V. Grolmusz, When the Web meets the cell: using personalized PageRank for analyzing protein interaction networks, Bioinformatics. 27 (2010) 405-407.

[17] V. Grolmusz, A note on the pagerank of undirected graphs, Inf. Process. Lett. 115 (2012) 633-634.

[18] I. Gutman, The energy of a graph, Ber. Math. Statist. Sekt. Forschungszentram Graz. 103 (1978) 1–22.

[19] I. Gutman, B. Furtula, Survey of graph energies, Math. Interdisc. Res. 2 (2017) 85–129.

[20] I. Gutman, Y. Hou, H. B. Walikar, H. S. Ramane and P. R. Hampiholi, No Hückel graph is hyperenergetic, J. Serb. Chem. Soc. 65 (2000) 799–801.

[21] I. Gutman, H. Ramane, Research on graph energies in 2019, MATCH Commun. Math. Comput. Chem. 84 (2020) 277-292.

[22] B. Jiang, K. kloster, D. F. Gleich, M. Gribskov, Aptrank: an adaptive pagerank model for protein function prediction on bi-relational graph, Bioinformatics. 33 (2017) 1829-1836.

[23] V. Kandiah, D. L. Shepelyansky, Pagerank model of opinion formation on social networks, Physica A. 391 (2012) 5779-5793.

[24] A. N. Langville, C. D. Meyer, Google's pagerank and beyond, Princeton University Press, 2005.

[25] V. Lazova, L. Basnarkov, Pagerank approach to ranking national football teams, arXiv preprint arXiv:1503.01331.

[26] Ch. Liao, K. Lu, M. Baym, R. Singh, B. Berger, Isorank: spectral methods for global alignment of multiple protein networks, Bioinformatics. 25 (2009) 253-258.

[27] C. D. Meyer, Matrix analysis and applied linear algebra, SIAM. Philadelphia, 2000.

[28] B. L. Mooney, L. R. Corrales, A. E. Clark, Molecularnetworks: An integrated graph theoretic and data mining tool to explore solvent organization in molecular simulation, J. Comput. Chem. 33 (2012) 853-860.

[29] N. Mukai, Pagerank-based traffic simulation using taxi probe data, Procedia Comput Sci. 22 (2013) 1156–1163.

[30] C. Winter, G. Kristiansen, S. Kersting, J. Roy, D. Aust, T. Knsel, P. Rummele, B. Jahnke, V. Hentrich, F. Ruckert, M. Niedergethmann, Google goes cancer: improving outcome prediction for cancer patients by network-based ranking of marker genes, Plos Comput. Biol. 2 (2012) e1002511.

[31] W. Zachary, An information flow model for conflict and fission in small groups, J. Anthropol. Res. 33 (1977) 452–473 .

[32] X. N. Zuo, R. Ehmke, M. Mennes, D. Imperati, F.X. Castellanos, O, Sporns, M. P. Milham, Network centrality in the human functional connectome, Cerebral cortex. 22 (2011) 1862-1875.

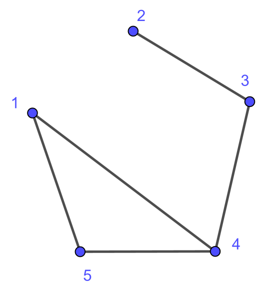

Figure 1: Graph G_1 in Example 2.1.

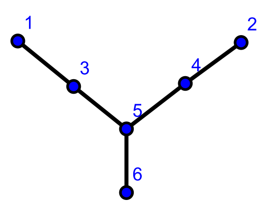

Figure 2: The tree T_1 in Example 2.2.

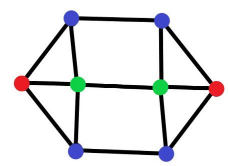

Figure 3: The graph K with three orbits and $k = 2$.

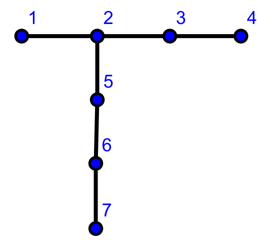

Figure 4: The tree T_2 in Example 3.1.

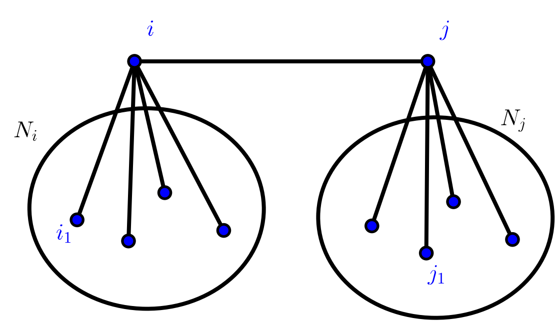

Figure 5: The neighbors of two adjacent vertices i, j.

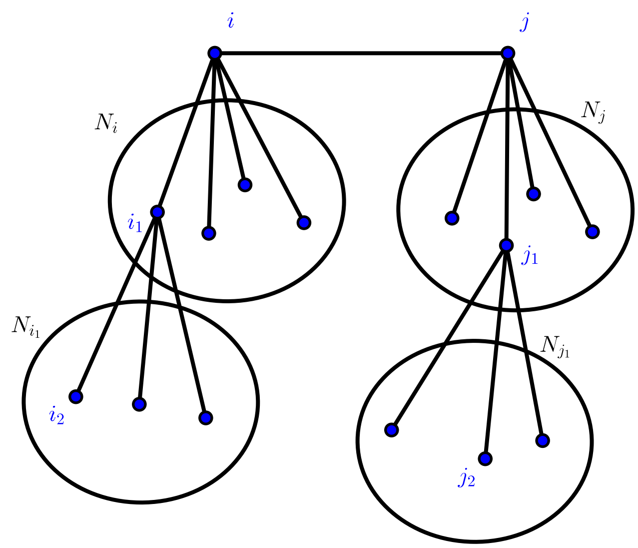

Figure 6: The neighbors of neighbors of vertices i, j.

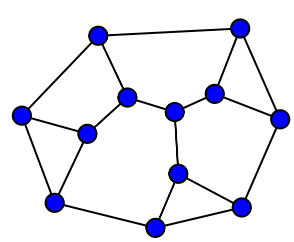

Figure 7: The Frucht graph.

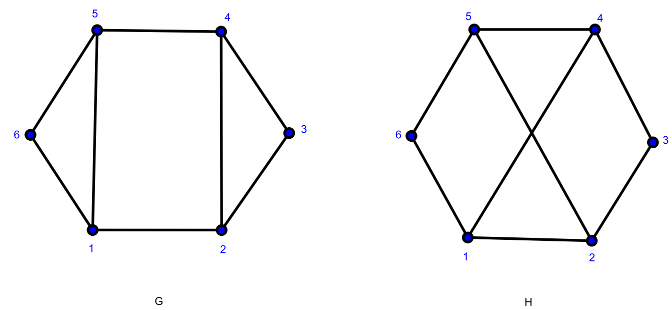

Figure 8: Two Co-PR graphs.

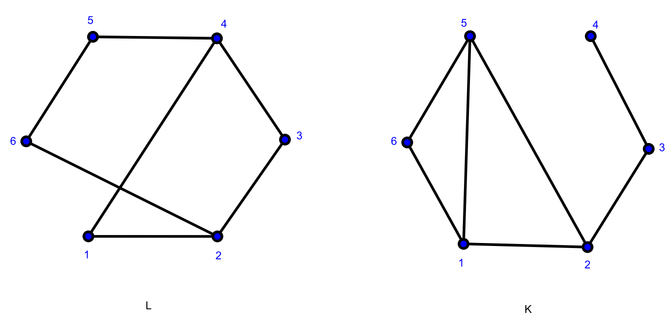

Figure 9: Two non-Co-PR graphs.

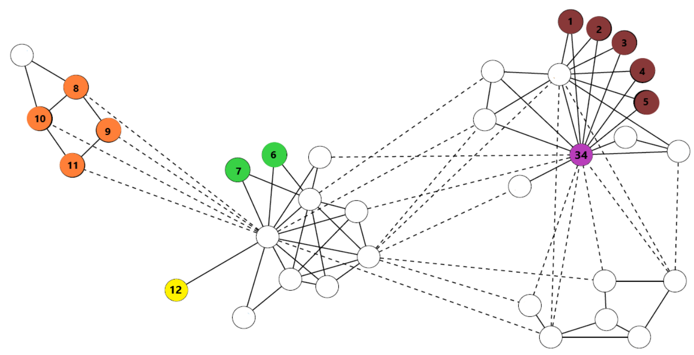

Figure 10: Zachary's Karate graph $\mathcal{K}$.

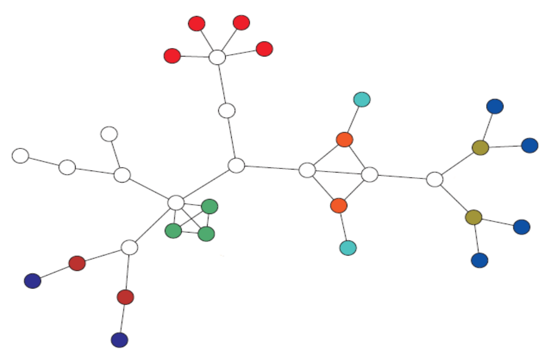

Figure 11: The graph $\mathcal{G}$.

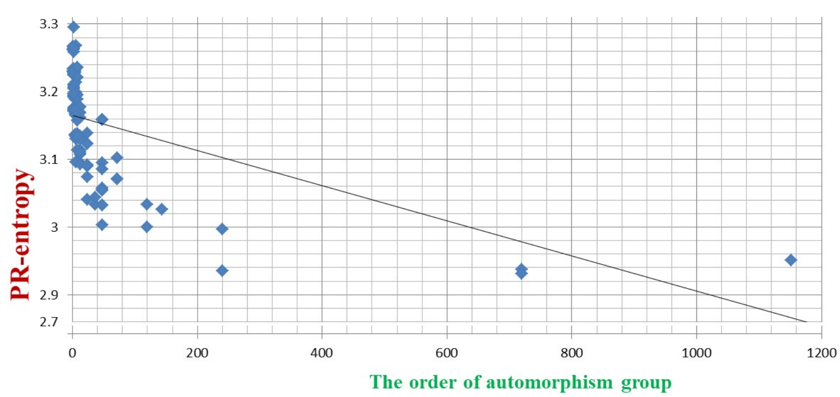

Figure 12: All trees of order 10. The correlation between $|Aut(T)|$ and $I_\pi(T)$ is -0.60.

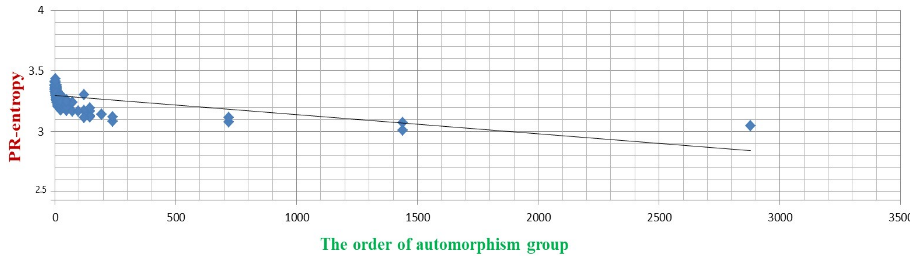

Figure 13: All trees of order 11. The correlation between $|Aut(T)|$ and $I_\pi(T)$ is -0.50.

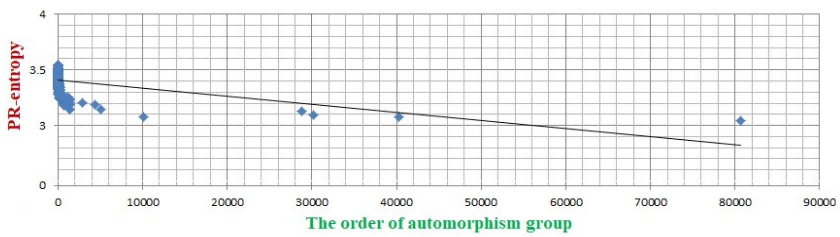

Figure 14: All trees of order 12. The correlation between $|Aut(T)|$ and $I_\pi(T)$ is -0.34.

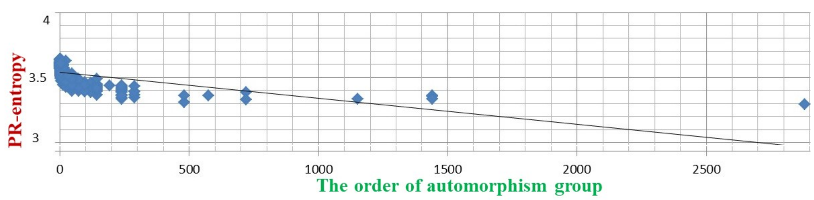

Figure 15: All trees of order 13. The correlation between $|Aut(T)|$ and $I_\pi(T)$ is -0.46.

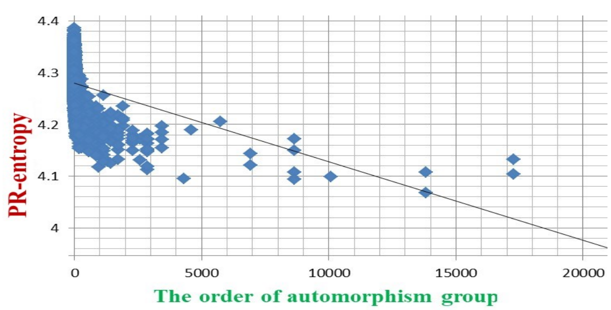

Figure 16: All trees of order 22. The correlation between $|Aut(T)|$ and $I_\pi(T)$ is -0.29.

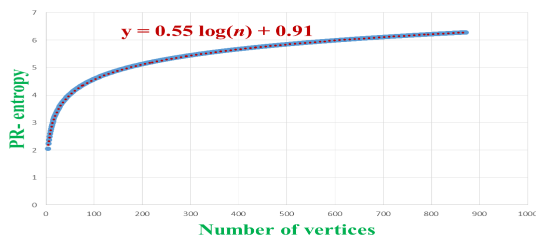

Figure 17: The value of $I_\pi(T)$ for a star graph with at most 872 vertices.

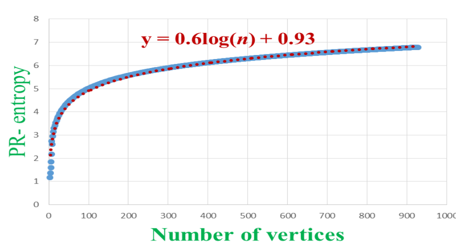

Figure 18: The value of $I_\pi(T)$ for a bistar graph with at most 467 vertices.