Two-walks degree assortativity in graphs and networks

CrossMark

Alfonso Allen-Perkins [a], Juan Manuel Pastor [a,b], Ernesto Estrada [c,*]

[a] *Complex System Group, Universidad Politécnica de Madrid, 28040 Madrid, Spain*
[b] *E.T.S.I.A.A.B, Universidad Politécnica de Madrid, Avd. Puerta de Hierro 4, 28040 Madrid, Spain*
[c] *Department of Mathematics and Statistics, University of Strathclyde, Glasgow G1 1XQ, United Kingdom*

ARTICLE INFO

ABSTRACT

Degree assortativity is the tendency for nodes of high degree (resp. low degree) in a graph to be connected to high degree nodes (resp. to low degree ones). It is usually quantified by the Pearson correlation coefficient of the degree–degree correlation. Here we extend this concept to account for the effect of second neighbours to a given node in a graph. That is, we consider the two-walks degree of a node as the sum of all the degrees of its adjacent nodes. The two-walks degree assortativity of a graph is then the Pearson correlation coefficient of the two-walks degree–degree correlation. We found here analytical expression for this two-walks degree assortativity index as a function of contributing subgraphs. We then study all the 261,000 connected graphs with 9 nodes and observe the existence of assortative–assortative and disassortative–disassortative graphs according to degree and two-walks degree, respectively. More surprisingly, we observe a class of graphs which are degree disassortative and two-walks degree assortative. We explain the existence of some of these graphs due to the presence of certain topological features, such as a node of low-degree connected to high-degree ones. More importantly, we study a series of 49 real-world networks, where we observe the existence of the disassortative–assortative class in several of them. In particular, all biological networks studied here were in this class. We also conclude that no graphs/networks are possible with assortative–disassortative structure.

1. Introduction

Networks represent the topological skeleton of a wide range of systems in nature and society [1–4]. The characterization of their structure is crucial since it shapes the evolutionary, functional, and dynamical processes that take place in those systems [4–6].

It is well known that links generally do not connect nodes regardless of their characteristics. In social networks, for instance, evidence suggests that individuals prefer to associate with others of similar age, religion, education or occupation as themselves [7]. Assortativity or assortative mixing is a graph metric that refers to the tendency for nodes in networks to be connected to other nodes that are similar (or different) to themselves in some way [8]. Typically, it is determined for the *degree* (i.e. the number of direct neighbours, k) of the nodes in the network [9–12]. The tendency for high-degree nodes to associate preferentially with other high-degree nodes plays a major role in many important processes, such as epidemic

spreading, synchronization or network robustness, among others [9,13–16]. However, assortativity may be applied to any characteristics of a node, including non-topological vertex properties, such as language or race [8]. Most of the research done in this area has been summarized in the review of Noldus et al. [17]. Other extensions to account for interactions beyond the nearest-neighbours have also been proposed in the recent literature [18].

The aim of this work is to define an assortativity index that captures the influence of first and second neighbours of a node. We then express this two-walks assortativity in terms of the subgraphs contributing to it.

The paper is organized as follows. In Section 2, the preliminaries are presented. In Section 3, the concept of two-walks degree assortativity is introduced and analysed. Main result is demonstrated in Section 4. Numerical results are presented in Section 5. Conclusions are summarized in Section 6.

2. Preliminaries

Here we consider simple, undirected graphs $G = (V, E)$, i.e., graphs without multiple edges, self-loops, directions or weights in their edges. The notation used is standard and the reader can check for instance [19]. Let us define some of the measures used in this work in order to make it self-contained. First, we define the degree assortativity index [8]. Mathematically, it is written as:

$$r_k = \frac{\frac{1}{m}\sum_{(i,j)\in E} k_i k_j - \left\{\frac{1}{m}\sum_{(i,j)\in E}\frac{1}{2}\left[k_i + k_j\right]\right\}^2}{\frac{1}{m}\sum_{(i,j)\in E}\frac{1}{2}\left[k_i^2 + k_j^2\right] - \left\{\frac{1}{m}\sum_{(i,j)\in E}\frac{1}{2}\left[k_i + k_j\right]\right\}^2} \tag{2.1}$$

where k_i and k_j are the degrees at both ends of link $(i, j) \in E$ and m is the number of links. A positive assortativity index $r_k > 0$ indicates the tendency of higher degree nodes in the graph to be connected to other higher degree nodes. On the other hand, $r_k < 0$ indicates the tendency of higher degree nodes to be connected to lower degree nodes. It was previously proved the following result [11].

Lemma 1. *Let $G = (V, E)$ be a simple graph and let k_i be the degree of the vertex i. Let $|P_1|$, $|P_2|$ and $|P_3|$ be the number of edges, of paths of length two and of paths of length three, respectively. Finally, let $|C_3|$ be the number of triangles in G. Then, the assortativity coefficient can be written combinatorially as:*

$$r_k = \frac{|P_3| + 3|C_3| - \frac{|P_2|^2}{|P_1|}}{|P_2| + 3|S_{1,3}| - \frac{|P_2|^2}{|P_1|}} \tag{2.2}$$

Let $|P_{r/s}|$ be the ratio $|P_r|/|P_s|$, $|S_{1,3}|$ the number of star graphs of four nodes, and $C = 3|C_3|/|P_2|$. Then G is:

(1) *assortative ($r > 0$): if and only if $|P_{3/2}| + C > |P_{2/1}|$,*
(2) *neutral ($r = 0$): if and only if $|P_{3/2}| + C = |P_{2/1}|$, and $3|S_{1,3}| - |P_2|(|P_{2/1}| - 1) \neq 0$, and*
(3) *disassortative ($r < 0$): if and only if $|P_{3/2}| + C < |P_{2/1}|$*

It is worth mentioning that the denominator of Eq. (2.2) is non-negative. Consequently, the sign of r_k depends only upon the sign of the numerator, which is determined by the following structural factors: the global clustering coefficient (i.e. $C = 3|C_3|/|P_2|$), the intermodular connectivity (i.e. $|P_{3/2}| = |P_3|/|P_2|$) and the branching (i.e. $|P_{2/1}| = |P_2|/|P_1|$) [11].

The number of subgraphs contributing to the degree assortativity can be obtained using the following results [20].

Lemma 2. *Let $G = (V, E)$ be a simple graph with n nodes. Let k_i be the degree of the vertex i. Let $|C_3|$ be the number of triangles in G. Then, the number of edges $|P_1|$, path of length two $|P_2|$ and three $|P_3|$ are given, respectively by*

$$|P_1| = \frac{1}{2}\sum_{i=1}^{n} k_i,$$

$$|P_2| = \frac{1}{2}\sum_{i=1}^{n} k_i(k_i - 1),$$

$$|P_3| = \sum_{(i,j)\in E} (k_i - 1)(k_j - 1) - 3|C_3|.$$

Lemma 3. *Let $G = (V, E)$ be a simple graph. Let k_i be the degree of the vertex i in G. Let A be the adjacency matrix of G. Let $|P_1|$ and $|P_2|$ be respectively the number of edges and the number of paths of length two in G. Let $|S_{T1D}|$ be the number of subgraphs S_{T1D} in G (see Table 1). Let $|C_i|$ be the number of cycles of i nodes in G. Then, $|C_3|$, $|C_4|$ and $|C_5|$ are given, respectively by*

$$|C_3| = \frac{1}{6}tr(A^3),$$

$$|C_4| = \frac{1}{8}tr(A^4) - 2|P_1| - 4|P_2|,$$

$$|C_5| = \frac{1}{10}tr(A^5) - 3|C_3| - |S_{T1S}|.$$

Table 1
Collection of subgraphs in Eq. (4.1), excluding the paths P_2, P_3, P_4, P_5, and the cycle C_3.
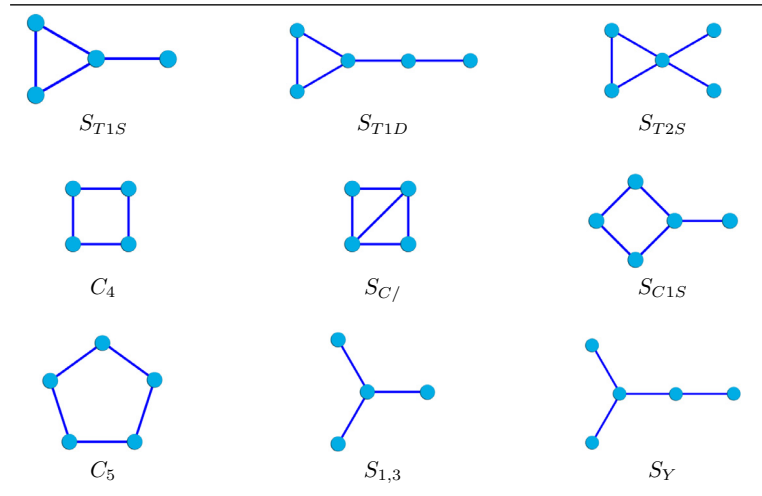
Lemma 4. *Let $G = (V, E)$ be a simple graph. Let k_i be the degree of the vertex i. Let A be the adjacency matrix of G. Let $|S_{1,3}|$, $|S_{T1S}|$, $|S_{T2S}|$, $|S_{T1D}|$, $|S_{C/}|$ and $|S_{C1S}|$ be the number of subgraphs $S_{1,3}$, S_{T1S}, S_{T2S}, S_{T1D}, $S_{C/}$ and S_{C1S}, respectively, in G (see Table 1). Then,*

$$|S_{1,3}| = \frac{1}{6} \sum_{i=1}^{n} k_i (k_i - 1)(k_i - 2)$$

$$|S_{T1S}| = \frac{1}{2} (\vec{k} - \vec{2})^T diag(A^3)$$

$$|S_{T2S}| = \frac{1}{4} ((\vec{k} - \vec{2}) \circ (\vec{k} - \vec{3}))^T diag(A^3)$$

$$|S_{T1D}| = \frac{1}{2} \vec{1}^T (A^2 - A^{2(D)}) diag(A^3) - 6|C_3| - 2|S_{T1S}| - 4|S_{C/}|$$

$$|S_{C/}| = \vec{1}^T (Q \circ (Q - A))(\vec{k} - \vec{2})$$

$$|S_{C1S}| = \vec{1}^T (P - P^{(D)})(\vec{k} - \vec{2}) - 2|S_{C/}|,$$

where $Q = A^2 \circ A$, $P = \frac{1}{2}(A^2 \circ (A^2 - 1))$, $\vec{x}$ is an all-x vector and $\circ$ denotes the Hadamard product.

3. Two-walks degree assortativity in graphs and networks

Let us start by the definition of the degree of a node i, k_i. The intuition behind this index is very simple. Every nearest neighbour of the node i receives an identical weight of 1. Then, we sum the weights of every node adjacent to i to obtain k_i. Mathematically, this corresponds to obtaining the following vector $\vec{k}$ after assigning the unit weights to every node:

$$\vec{k} = A\vec{1}, \tag{3.1}$$

where $\vec{1}$ is an all-ones vector.

It is customary to consider that not all the neighbours of one particular node are equally important. This is the basis for instance of Katz centrality index [21], eigenvector centrality [22], PageRank [23], subgraph centrality [24] and so for. Then, we can consider that every neighbour of the node i is weighted according to its "importance". Of course, the definition of that importance will define the way in which we will proceed. In order to consider the current development as an extension of the concept of node degree we simply weight every node by its own degree. That is, now we consider the vector $\vec{k}$ as the weighting vector for the nodes of the graph. Consequently, an extension of the concept of degree is given by applying a similar procedure as in (3.1) to $\vec{k}$,

$$\vec{k}_2 = A\vec{k}. \tag{3.2}$$

It is straightforward to realize that $\vec{k}_2 = A(A\vec{1}) = A^2\vec{1}$. Then, obviously, the entries of this new vector represent a new kind of centrality of the nodes which counts the number of two-walks starting at the corresponding node. Consequently, we suggest the name of "two-walks" degree for the entries of $\vec{k}_2$. Let us call $\tilde{k}_i$ the ith entry of $\vec{k}_2$ in a graph. Notice that $\tilde{k}_i$
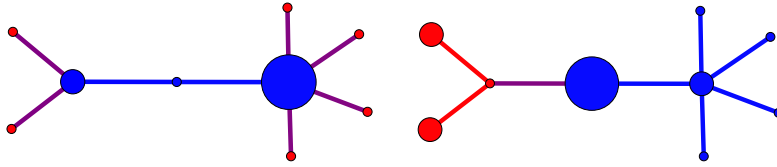
Fig. 1. Example of the structural effect that may produce a change from degree disassortative (left panel) to two-walks degree assortative (right panel) in a simple graph. Here the nodes are drawn in red if their degree (resp. two-walks degree) is smaller than the average degree, or blue otherwise. The size of the node is proportional to the magnitude of this difference. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

accounts for the degree of the node i, i.e., closed walks of length two, as well as for the number of second neighbours of this node. Then,

$$\tilde{k}_i = \sum_{j \in N(i)} k_j, \tag{3.3}$$

where $N(i)$ is the neighbourhood of the node i, i.e., $N(i) = \{j | (j, i) \in E\}$. That is, the two-walks degree $\tilde{k}_i$ represents the number of weighted neighbours that the node i has, where the weight of the nodes is given by its own degree.

Let us now define a quantity analogous to the degree assortativity index based on the two-walks degrees instead of on the node degrees.

Definition 5. Let G be a connected simple graph with adjacency matrix A and let $\tilde{k}_i$ be the two-walks degree of the vertex i. The two-walks degree assortativity index of a graph is defined as

$$r_{\tilde{k}} = \frac{\frac{1}{m}\sum_{(i,j)\in E}\tilde{k}_i\tilde{k}_j - \left\{\frac{1}{m}\sum_{(i,j)\in E}\frac{1}{2}\left[\tilde{k}_i + \tilde{k}_j\right]\right\}^2}{\frac{1}{m}\sum_{(i,j)\in E}\frac{1}{2}\left[\tilde{k}_i^2 + \tilde{k}_j^2\right] - \left\{\frac{1}{m}\sum_{(i,j)\in E}\frac{1}{2}\left[\tilde{k}_i + \tilde{k}_j\right]\right\}^2}. \tag{3.4}$$

Obviously, this quantity tell us how well connected the most important nodes in a graph are. That is, if $r_{\tilde{k}} > 0$, the graph is two-walks degree assortative, which means that the most weighted nodes in terms of the degree of their neighbours tend to be connected to each other. On the other hand, if $r_{\tilde{k}} < 0$, the graph is two-walks degree disassortative, which means that the most weighted nodes in terms of the degree of their neighbours tend to be connected to those with least weight. If $r_{\tilde{k}} = 0$, neither of these two tendencies is observed and we shall call such graphs neutral.

In Fig. 1 we represent a graph which is strongly disassortative for the degree ($r_k \approx -0.822$) but it is assortative for the two-walks degree index ($r_{\tilde{k}} \approx 0.212$). We plot the graph with the nodes weighted by the difference between the degree (resp. two-walks degree) minus the average degree (resp. average two-walks degree). The negative values are coloured in red and the positive contributions in blue. The size of the nodes is proportional to the absolute value of this difference. As can be seen in this picture the degree–degree interaction between the nodes (left panel) is dominated by red–blue interactions, which indicates a large number of interactions between high degree nodes (blue ones) with low degree ones (red nodes). This of course results in a negative degree assortativity coefficient. On the other hand, for the two-walks degree plot the graph is dominated by blue–blue and red–red interations (right panel). That is, nodes of high two-walks degree interact with each other, and low two-walks degree nodes also interact preferentially among them. This effects result in a two-walks degree assortativity coefficient.

With the new correlation coefficient introduced here we assess the tendency of neighbourhoods with many interactions to be connected to other "high-connected" neighbourhoods. However, in order for a graph to display a transition from degree diassortative to two-walks degree assortative it is necessary that there are *separator nodes* between the high-degree nodes. The graph in Fig. 1 has a separator, which is the node of degree 2 connecting both nodes of degree 3 and 5. A separator must be a low-degree node which connects two or more high-degree ones. Notice that if the number of high-degree nodes connected to the separator is too high, it will produce an increase in its own degree, which decreases its chances of being a proper separator. This characteristic – a separator connected to two high-degree nodes – introduces degree disassortativity to the graph. However, in terms of the second-order correlation a separator allow the two-steps interactions between hubs, which results in two-walks degree assortativity. Mathematically, it is not difficult to see that the two-walks degree is related to walks of length two between nodes.

It is easy to realize that the two-walks degree assortativity can be written in matrix-vector form in the following way:

$$r_{\tilde{k}} = \frac{\vec{1}^T A^5 \vec{1} - \left(\vec{1}^T A \vec{1}\right)^{-1}\left(\vec{1}^T A^3 \vec{1}\right)^2}{\left(\vec{k}_2 \circ \vec{k}_2\right)^T A \vec{1} - \left(\vec{1}^T A \vec{1}\right)^{-1}\left(\vec{1}^T A^3 \vec{1}\right)^2} \tag{3.5}$$

4. Main result

Our main result here consists on the determination of the two-walks degree assortativity of a graph in terms of contributing subgraphs of the graph. This allows us to understand this quantity in structural terms for the analysis of real world systems in further sections of this work.

Theorem 6. *Let G be a simple graph. Then, G is, in terms of two-walks degree,*

(i) *assortative if $R > 2|P_2||P_{2/1}|\left(|P_{3/2}| + C + 2\right)^2$,*

(ii) *neutral if $R = 2|P_2||P_{2/1}|\left(|P_{3/2}| + C + 2\right)^2$,*

(iii) *disassortative if $R < 2|P_2||P_{2/1}|\left(|P_{3/2}| + C + 2\right)^2$,*

where

$$
\begin{aligned}
R = {}& 4|P_2| + 8|P_3| + 4|P_4| + 2|P_5| + 42|C_3| + 24|C_4| + 10|C_5| \\
& + 12|S_{1,3}| + 4|S_Y| + 22|S_{T1S}| + 4|S_{T2S}| + 4|S_{T1D}| + 12|S_{C/}| + 4|S_{C1S}|,
\end{aligned}
\tag{4.1}
$$

$|P_{r/s}| = |P_r|/|P_s|$ *and* $C = 3|C_3|/|P_2|$.

First, we prove that the denominator of expression (3.5) is always non-negative.

Lemma 7. *Let G be a connected simple graph with adjacency matrix A. Let $\vec{k}$ and $\vec{k}_2$ be vectors of the nodes degrees and a vector of nodes two-walks degrees, respectively. Then,*

$$
\left(\vec{k}_2 \circ \vec{k}_2\right)^T A \vec{1} - \left(\vec{1}^T A \vec{1}\right)^{-1} \left(\vec{1}^T A^3 \vec{1}\right)^2 \geq 0
\tag{4.2}
$$

where m is the network's number of edges, $\vec{1}$ is an all-ones vector and $\circ$ denotes the Hadamard product.

Proof. By the Cauchy–Bunyakovsky–Schwarz inequality:

$$
\begin{aligned}
\left(\vec{1}^T A \vec{1}\right)^{-1}\left(\vec{1}^T A^3 \vec{1}\right)^2 &= \left(\vec{1}^T A \vec{1}\right)^{-1}\left(\sum_{i=1}^n k_i \tilde{k}_i\right)^2 \leq \left(\vec{1}^T A \vec{1}\right)^{-1}\left(\sum_{i=1}^n k_i^2\right)\left(\sum_{i=1}^n \tilde{k}_i^2\right) \\
&= \frac{2|P_1| + 2|P_2|}{2|P_1|}\sum_{i=1}^n \tilde{k}_i^2 = \sum_{i=1}^n \tilde{k}_i^2\left(\frac{1}{n}\frac{\sum_{i=1}^n k_i^2}{\sum_{i=1}^n k_i}\right).
\end{aligned}
\tag{4.3}
$$

Then, we have

$$
\begin{aligned}
\left(\vec{k}_2 \circ \vec{k}_2\right)^T A \vec{1} - \left(\vec{1}^T A \vec{1}\right)^{-1}\left(\vec{1}^T A^3 \vec{1}\right)^2 &\geq \sum_{i=1}^n \tilde{k}_i^2\left(k_i - \frac{1}{n}\frac{\sum_{j=1}^n k_j^2}{\sum_{j=1}^n k_j}\right) \\
&= \sum_{i=1}^n \tilde{k}_i^2 \frac{n k_i \sum_{j=1}^n k_j \left(n k_i - k_j\right)}{n \sum_{j=1}^n k_j}.
\end{aligned}
\tag{4.4}
$$

As G is a connected simple graph, $k_i \geq 1$ and the maximum degree in the graph is $n - 1$, then, $n k_i - k_j \geq 0$, and hence the last term is always greater than or equal to zero, which proves the result. □

What remains now for the proof of the main result is to express the numerator $N(r)$ of the Pearson coefficient of the two-walks degree – two-walks degree correlation in terms of subgraphs of the graph (reminding that when the denominator is equal to zero, the Pearson Correlation coefficient is not defined). We can write $N(r)$ as follows:

$$
N(r) = \frac{1}{m}\sum_{(i,j)\in E}\tilde{k}_i\tilde{k}_j - \left\{\frac{1}{m}\sum_{(i,j)\in E}\frac{1}{2}\left[\tilde{k}_i + \tilde{k}_j\right]\right\}^2
\tag{4.5}
$$

where $\tilde{k}_i$ and $\tilde{k}_j$ are the two-walks degrees of nodes i and j, respectively, located at both ends of link $(i, j) \in E$. We can now rewrite the sums in Eq. (4.5) as:

$$
\sum_{(i,j)\in E}\tilde{k}_i\tilde{k}_j = \frac{1}{2}\vec{k}_2^T A \vec{k}_2,
\tag{4.6}
$$

$$
\sum_{(i,j)\in E}\left(\tilde{k}_i + \tilde{k}_j\right) = \vec{k}_2^T \vec{k}.
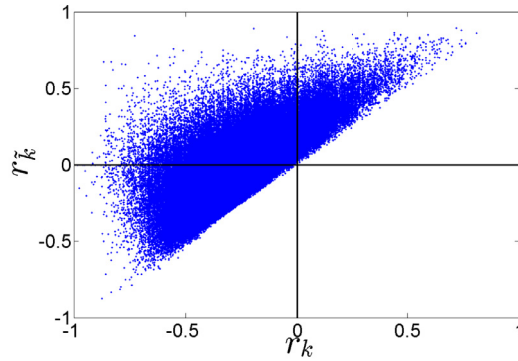\tag{4.7}
$$

Fig. 2. Degree and two-walks degree assortativities for all the connected graphs with 9 unlabelled nodes.

Let us now find the expressions for the two terms contributing to $N(r)$. The first is given by

$$\frac{1}{m}\sum_{(i,j)\in E}\tilde{k}_i\tilde{k}_j = \frac{1}{2m}(2|P_1| + 12|P_2| + 12|P_3| + 4|P_4| + 2|P_5| + 54|C_3| + 24|C_4| + 10|C_5|$$
$$+ 12|S_{1,3}| + 22|S_{T1S}| + 4|S_Y| + 4|S_{T2S}| + 4|S_{T1D}| + 12|S_{C/}| + 4|S_{C1S}|),\tag{4.8}$$

where $|P_4|$ and $|P_5|$ are the number of paths of order 4 and 5, respectively, and $|S_Y|$ is the number of fragments S_Y which are illustrated in Table 1. We will give formulas for calculating these fragments for the sake of completeness of the paper.

For the second term contributing to $N(r)$ we have

$$\left\{\frac{1}{m}\sum_{(i,j)\in E}\frac{1}{2}[\tilde{k}_i + \tilde{k}_j]\right\}^2 = \frac{1}{(2m)^2}\left(2|P_1| + 4|P_2| + 2|P_3| + 6|C_3|^2\right).\tag{4.9}$$

Thus, we can rewrite $N(r)$ as:

$$N(r) = \frac{1}{2m}\left(R - 2|P_2||P_{2/1}|\left(|P_{3/2}| + C + 2\right)^2\right),\tag{4.10}$$

which proves the main result.

Let us now give the formulas for calculating the subgraphs remaining in the expression of the two-walks degree assortativity which have not been previously defined. The proofs of these results are based on the strategy developed and explained in [25] and are not given here as they are lengthy and technical.

Lemma 8. *Let $G = (V, E)$ be a simple graph. Let k_i be the degree of the vertex i. Let $|S_Y|$ be the number of subgraphs S_Y (see Table 1). Then,*

$$|S_Y| = \sum_{(i,j)\in E}\left\{\binom{k_i - 1}{2}(k_j - 1) + \binom{k_j - 1}{2}(k_i - 1)\right\} - 2|S_{T1S}|.\tag{4.11}$$

Lemma 9. *Let $G = (V, E)$ be a simple graph. Then, the number of subgraphs $|P_4|$ and $|P_5|$ in G are given by, respectively,*

$$|P_4| = \frac{1}{2}\vec{1}^T A^4 \vec{1} - |P_1| - 4|P_2| - 2|P_3| - 9|C_3| - 4|C_4| - 6|S_{1,3}| - 4|S_{T1S}|,\tag{4.12}$$

$$|P_5| = \frac{1}{2}\vec{1}^T A^5 \vec{1} - |P_1| - 6|P_2| - 6|P_3| - 2|P_4| - 27|C_3|$$
$$- 12|C_4| - 5|C_5| - 6|S_{1,3}| - 2|S_Y| - 11|S_{T1S}| - 2|S_{T2S}|$$
$$- 2|S_{T1D}| - 6|S_{C/}| - 2|S_{C1S}|.\tag{4.13}$$

5. Computational results

5.1. Small graphs

In this section we describe the results obtained for all the 261,000 connected graphs with 9 unlabelled nodes. We have calculated the degree and two-walks degree assortativities for these graphs (see Fig. 2). As we can see there is no trivial correlation between the two indices, which indicates that the new index does not duplicate the structural information contained in the degree assortativity and consequently gives some new structural insights about graphs. This conclusion is also easily obtained by considering the subgraph contributions to both measures.
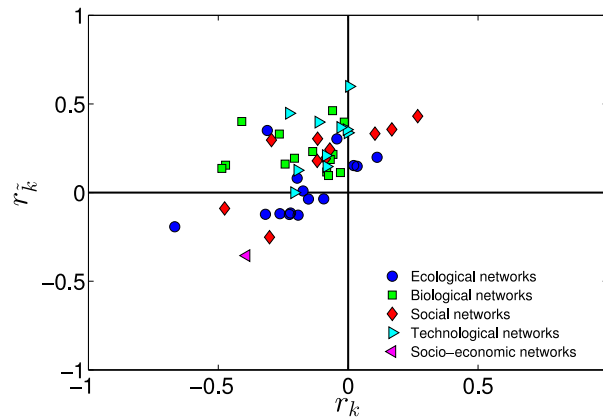
Fig. 3. Degree and two-walks degree assortativities for all the real-world networks studied in this work.

According to computer calculations 7% of the networks are assortative-assortative by both measures (AA), 60% are disassorartive–disassortative (DD) and 33% are disassortative by degree and assortative by two-walks degree (DA). The main observation is that there are no graphs which are degree assortative and two-walks degree disassortative (AD). We conjecture that these graphs cannot exist. Computer calculations show that $R \gtrsim 2|P_2|(|P_{3,2}| + C)(|P_{3,2}| + C + 2)^2$. Therefore, we can express the numerator of the neighbourhood assortativity (Eq. (4.10)) as follows:

$$N(r) \gtrsim \frac{1}{m}|P_2|(|P_{3,2}| + C - |P_{2/1}|)(|P_{3,2}| + C + 2)^2 \qquad (5.1)$$

Using the results from [11], if $r_k \geq 0$, then $N(r) \geq 0$. The intuition behind this result is very simple. Nodes that belong to a degree assortatitivive network tend to be linked to other nodes with similar degree. Therefore, their two-walks degrees tend to be similar too.

Generally, the two-walks degree assortativity depends on the balance between four structural factors: the weighted sum of subgraphs given by $R/|P_2|$, transitivity (C), intermodular connectivity ($|P_{3,2}|$) and relative branching ($|P_{2,1}|$). The first three produce a positive contribution to the two-walks degree assortativity of a network, while branching is more likely associated with disassortative networks.

5.2. Real-world networks

In this subsection we study of group of 49 real-world networks representing systems in ecological (E), biological (B), social (S), technological (T) and socio-economic (SE) environments. The networks are described in the Appendix of this paper. We have calculated the degree and two-walks degree assortativities for these networks (see Fig. 3). According to these results 14% of the networks are assortative-assortative (AA) according to both measures, 24% are disassorartive–disassortative (DD) and the majority of networks analysed (61%) are diassortative–assortative (DA). This confirms our previous observation that there are no graphs/networks which are assortative–disassorartive (AD). The analysis of the networks according to the functions shows the following trends: 53% of the ecological networks analysed are DD, 27% are DA and 20% are AA; 50% of the social networks analysed are DA, 30% are AA and 20% are DD; 80% of technological networks are DA, 10% are AA and 10% are DD. Finally, 100% of biological networks considered are DA. They included 8 protein-protein interaction networks (PINs), 3 transcription networks and 2 brain networks. This is a remarkable observation because it is the only single functional class of networks which is formed by one structural class, i.e., DA.

An important characteristic of our current approach is that we can understand the structural causes for the different kinds of assortativity in networks using the interpretation of these quantities in terms of subgraphs of the graph. As we have seen before an important structural feature of graphs allowing the transition from degree disassortative to two-walks degree assortative is the presence of separators. It has to be stressed that this is not a unique structural feature of this kind of networks and more studies are needed to completely understand the structural characterization of this kind of networks. However, it is easy to visualize the connectors in the small PIN of the bacterium *B. subtilis* (see left panel in Fig. 4). In Fig. 4 we also illustrate the degree and two-walks degree of the nodes in the food web of ScotchBroom and in the transcription network of *E. coli*. All of them displaying DA characteristics.

6. Conclusions

Here we have proposed an extension of the concept of degree assortativity to one that account for the correlation between the degrees of the nodes and their nearest neighbours in graphs and networks. This measure, here named the
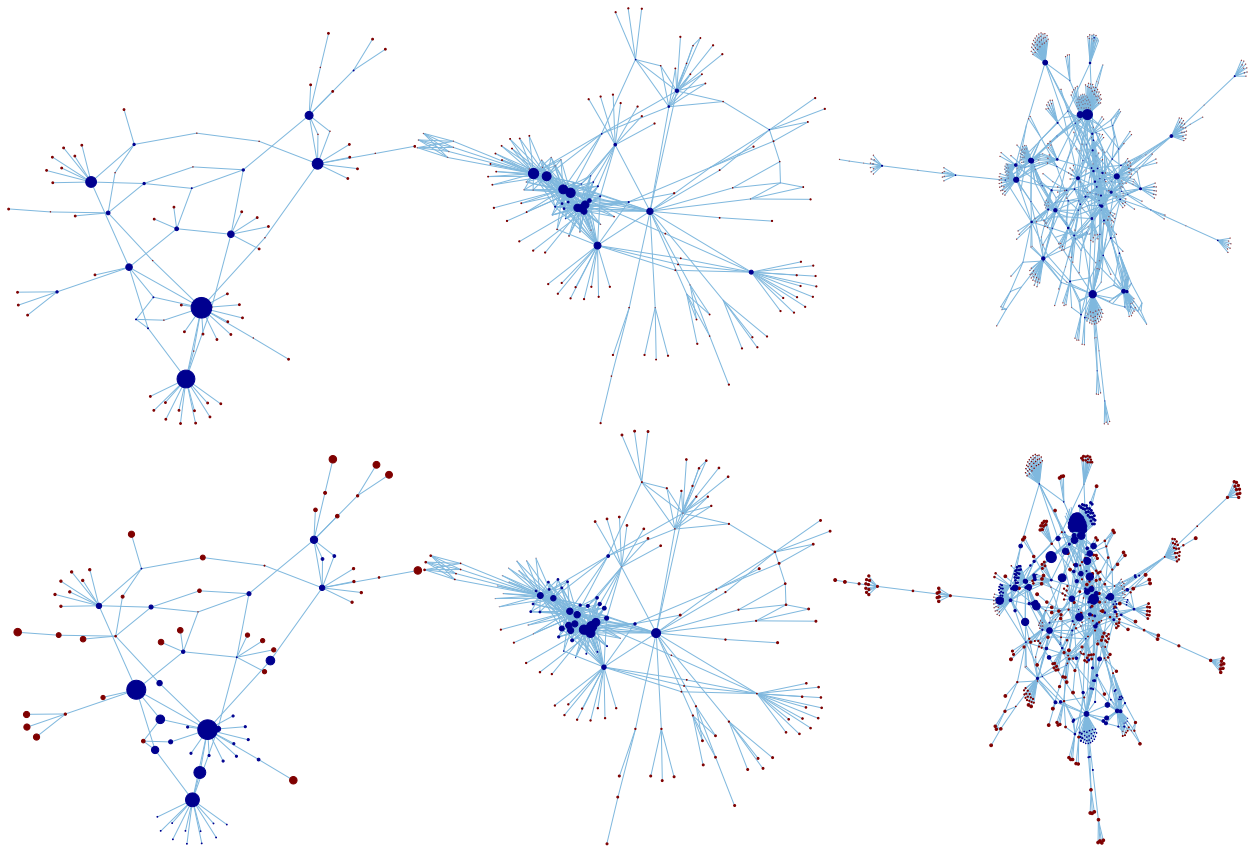
Fig. 4. Illustration of the differences between the degrees and mean degree of every node (top panels) and the same for the two-walks degrees (bottom panels) in the protein interaction network *B. subtilis*, food web of Scotch Broom, and transcription network of yeast from left to right.

two-walks degree assortativity, is expressed in terms of subgraphs of the graph. As we have proved here there are a few more fragments contributing to the two-walks degree assortativity than to the degree assortativity. This clearly indicates that the new quantity accounts for more structural information than the previous one. We have seen that both measures are not linearly correlated neither for all the connected graphs with 9 nodes nor for real-world networks. Further studies are needed to understand the role of this quantity in the study of real-world problems, as we have seen here, there are some apparently universal features of some classes of networks in relation to this quantity. For instance, all real-world biological networks studied here are degree disassortative but two-walks assortative. The implications of this observation for the study of the biological processes taking place on these networks is far beyond the scope of this work.

Acknowledgements

Appendix. Real-world network dataset

The real-world networks used in this paper belong to different domains: ecological (includes food webs and ecosystems), social (networks of friendships, communication networks, corporate relationships), technological (internet, transport, software development networks), informational (vocabulary networks, citations) and biological (protein–protein interaction networks, transcriptional regulation networks). The dataset comprises networks of different sizes, ranging from $n = 29$ to $n = 4941$ nodes. The networks are listed in Table 2.

Table 2

Dataset of real-world networks: network name, domain, n number of nodes, m number of links, reference, degree and two-walks degree assortative coefficients.

No.	Dataset	Domain	n	m	Ref.	r_k	$r_{\tilde{k}}$
1	Drosophila PIN	Biological	3039	3715	[26]	−0.060	0.462
2	Hpyroli	Biological	710	1396	[27]	−0.243	0.161
3	KSHV	Biological	50	122	[28]	−0.058	0.215
4	MacaqueVisualCortex	Biological	30	190	[29]	−0.030	0.113
5	Malaria PIN	Biological	229	604	[30]	−0.083	0.116
6	Neurons	Biological	280	1973	[31]	−0.069	0.187
7	PIN-Afulgidus	Biological	32	38	[32]	−0.472	0.154
8	Pin-Bsubtilis	Biological	84	98	[33]	−0.486	0.136
9	PIN-Ecoli	Biological	230	695	[34]	−0.015	0.397
10	PIN-Human	Biological	2783	6438	[35]	−0.137	0.231
11	Trans-Ecoli	Biological	328	456	[36]	−0.265	0.330
12	Transc-yeast	Biological	662	1062	[36]	−0.410	0.401
13	Trans-urchin	Biological	45	80	[36]	−0.207	0.194
14	Benguela	Ecological	29	191	[37]	0.0211	0.153
15	BridgeBrook	Ecological	75	547	[38]	−0.668	−0.193
16	Canton	Ecological	108	708	[39]	−0.226	−0.123
17	Chesapeake	Ecological	33	72	[40]	−0.196	0.081
18	Coachella	Ecological	30	261	[41]	0.0347	0.148
19	ElVerde	Ecological	156	1441	[42]	−0.174	0.009
20	ReefSmall	Ecological	50	524	[43]	−0.193	−0.127
21	ScotchBroom	Ecological	154	370	[44]	−0.311	0.350
22	Shelf	Ecological	81	1476	[45]	−0.094	−0.035
23	Skipwith	Ecological	35	364	[46]	−0.319	−0.122
24	StMarks	Ecological	48	221	[47]	0.111	0.199
25	StMartin	Ecological	44	218	[48]	−0.153	−0.0365
26	Stony	Ecological	112	832	[49]	−0.222	−0.115
27	Ythan1	Ecological	134	597	[50]	−0.263	−0.119
28	World Trade	Economic	80	875	[51]	−0.392	−0.355
29	SmallW	Informational	233	994	[52]	−0.303	−0.251
30	ColoSPG	Social	324	347	[53]	−0.295	0.296
31	CorporatePeople	Social	1586	13126	[54]	0.268	0.431
32	Dolphins	Social	62	159	[55]	−0.044	0.303
33	Drugs	Social	616	2012	[51]	−0.117	0.304
34	Hi-tech	Social	33	91	[56]	−0.087	0.191
35	Geom	Social	3621	9461	[51]	0.168	0.356
36	PRISON-Sym	Social	67	142	[57]	0.103	0.332
37	Sawmill	Social	36	62	[58]	−0.071	0.243
38	social3	Social	32	80	[59]	−0.119	0.179
39	Zackar	Social	34	78	[60]	−0.476	−0.089
40	electronic1	Technological	122	189	[61]	−0.002	0.337
41	electronic2	Technological	252	399	[61]	−0.006	0.355
42	electronic3	Technological	512	819	[61]	−0.030	0.367
43	Power grid	Technological	4941	6594	[62]	0.003	0.599
44	Software Abi	Technological	1035	1736	[63]	−0.086	0.208
45	Software Digital	Technological	150	198	[63]	−0.228	0.447
46	Software Mysql	Technological	1480	4221	[63]	−0.083	0.147
47	Software-XMMS	Technological	971	1809	[63]	−0.114	0.397
48	Software-VTK	Technological	771	1369	[63]	−0.195	0.126
49	USA Air 97	Technological	332	2126	[52]	−0.208	−0.000

References

[1] R. Albert, A.-L. Barabási, Statistical mechanics of complex networks, Rev. Mod. Phys 74 (2002) 47.
[2] M.E.J. Newman, The structure and function of complex networks, SIAM Rev. 45 (2003) 167.
[3] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D.-U. Hwang, Complex networks: structure and dynamics, Phys. Rep. 424 (2006) 175.
[4] S.H. Strogatz, Exploring complex networks, Nature 410 (2001) 268.
[5] A. Barrat, M. Barthélemy, A. Vespignani, Dynamical Processes on Complex Networks, Cambridge University Press, UK, 2008.
[6] L.D.F. Costa, F.A. Rodrigues, G. Travieso, P.R.V. Boas, Characterization of complex networks: a survey of measurements, Adv. Phys. 56 (2007) 167.
[7] M. McPherson, L. Smith-Lovin, J.M. Cook, Birds of a feather: homophily in social networks, Annu. Rev. Sociol. 27 (2001) 415.
[8] M.E.J. Newman, Mixing patterns in networks, Phys. Rev. E 67 (2003) 026126.
[9] M.E.J. Newman, Assortative mixing in networks, Phys. Rev. Lett. 89 (2002) 208701.
[10] R. Pastor-Satorras, A. Vázquez, A. Vespignani, Dynamical and correlation properties of the internet, Phys. Rev. Lett. 87 (2001) 258701.
[11] E. Estrada, Combinatorial study of degree assortativity in networks, Phys. Rev. E 84 (2011) 047101.
[12] M. Piraveenan, M. Prokopenko, A. Zomaya, Local assortativeness in scale-free networks, Europhys. Lett. 84 (2008) 28002.
[13] M.E.J. Newman, J. Park, Why social networks are different from other types of networks, Phys. Rev. E 68 (2003) 036122.
[14] V.M. Eguíluz, K. Klemm, Epidemic threshold in structured scale-free networks, Phys. Rev. Lett. 89 (2002) 108701.

[15] M. Boguñá, R. Pastor-Satorras, A. Vespignani, Absence of epidemic threshold in scale-free networks with degree correlations, Phys. Rev. Lett. 90 (2003) 028701.
[16] M.D. Bernardo, F. Garofalo, F. Sorrentino, Effects of degree correlation on the synchronization of networks of oscillators, Int. J. Bifurc. Chaos 17 (2006) 3499.
[17] R. Noldus, P.V. Mieghem, Assortativity in complex networks, J. Complex Netw. 3 (2015) 507.
[18] A. Allen-Perkins, J. Galeano, J.M. Pastor, Inducing self-organized criticality in a network toy model by neighborhood assortativity, Phys. Rev. E 94 (2016) 052304.
[19] E. Estrada, The Structure of Complex Networks. Theory and Applications, Oxford University Press, UK, 2012.
[20] N. Alon, R. Yuster, U. Zwick, Finding and counting given length cycles, Algorithmica 17 (1997) 3.
[21] L. Katz, A new status index derived from sociometric analysis, Psychometrica 18 (1953) 1.
[22] P. Bonacich, Factoring and weighting approaches to status scores and clique identification, J. Math. Sociol. 2 (1972) 1.
[23] S. Brin, L. Page, The anatomy of a large-scale hypertextual web search engine, Comput. Net. ISDN Syst. 33 (1998) 1–7.
[24] E. Estrada, J.A. Rodríguez-Velázquez, Subgraph centrality in complex networks, Phys. Rev. E 71 (2005) 056103.
[25] E. Estrada, P. Knight, A First Course on Network Theory, Oxford University Press, UK, 2012.
[26] L. Giot, J.S. Bader, C. Brouwer, A. Chaudhuri, B. Kuang, Y. Li, Y. Hao, C. Ooi, B. Godwin, E. Vitols, et al., A protein interaction map of drosophila melanogaster, Science 302 (2003) 5651.
[27] C.-Y. Lin, C.-L. Chen, C.-S. Cho, L.-M. Wang, C.-M. Chang, P.-Y. Chen, C.-Z. Lo, C.A. Hsiung, hp-DPI: Helicobacter pylori database of protein interactome-sembracing experimental and inferred interactions, Bioinformatics 21 (2005) 7.
[28] P. Uetz, Y.-A. Dong, C. Zeretzke, C. Atzler, A. Baiker, B. Berger, S.V. Rajagopala, M. Roupelieva, D. Rose, E. Fossum, et al., Herpesviral protein networks and their interaction with the human proteome, Science 311 (2006) 5758.
[29] O. Sporns, R. Kötter, Motifs in brain networks, PLoS Biol. 2 (2004) 11.
[30] D.J. LaCount, M. Vignali, R. Chettier, A. Phansalkar, R. Bell, J.R. Hesselberth, L.W. Schoenfeld, I. Ota, S. Sahasrabudhe, C. Kurschner, et al., A protein interaction network of the malaria parasite plasmodium falciparum, Nature 438 (2005) 7064.
[31] J.G. White, E. Southgate, J.N. Thomson, S. Brenner, The structure of the nervous system of the nematode caenorhabditis elegans, Philos. Trans. R. Soc. Lond. B: Biol. Sci. 314 (1986) 1165.
[32] M. Motz, I. Kober, C. Girardot, E. Loeser, U. Bauer, M. Albers, G. Moeckel, E. Minch, H. Voss, C. Kilger, et al., Elucidation of an archaeal replication protein network to generate enhanced PCR enzymes, J. Biol. Chem. 277 (2002) 18.
[33] P. Noirot, M.F. Noirot-Gros, Protein interaction networks in bacteria, Curr. Opin. Microbiol. 7 (2004) 5.
[34] G. Butland, J.M. Peregrín-Alvarez, J. Li, W. Yang, X. Yang, V. Canadien, A. Starostine, D. Richards, B. Beattie, N. Krogan, et al., Interaction network containing conserved and essential protein complexes in escherichia coli, Nature 433 (2005) 7025.
[35] J.-F. Rual, K. Venkatesan, T. Hao, T. Hirozane-Kishikawa, A. Dricot, N. Li, G.F. Berriz, F.D. Gibbons, M. Dreze, N. Ayivi-Guedehoussou, et al., Towards a proteome-scale map of the human protein–protein interaction network, Nature 437 (2005) 7062.
[36] R. Milo, S. Itzkovitz, N. Kashtan, R. Levitt, S. Shen-Orr, I. Ayzenshtat, M. Sheffer, U. Alon, Superfamilies of evolved and designed networks, Science 303 (2004) 5663.
[37] P. Yodzis, Diffuse effects in food webs, Ecology 81 (2000) 1.
[38] G.A. Polis, Complex trophic interactions in deserts: an empirical critique of food-web theory, Am. Nat. 138 (1991) 123–155.
[39] C.R. Townsend, Disturbance, resource supply, and food-web architecture in streams, Ecol. Lett. 1 (1998) 200–2009.
[40] R.R. Christian, J.J. Luczkovich, Organizing and understanding a winter's seagrass foodweb network through effective trophic levels, Ecol. Model. 117 (1999) 1.
[41] P.H. Warren, Spatial and temporal variation in the structure of a freshwater food web, Oikos 55 (1989) 299–311.
[42] D.P. Reagan, R.B. Waide, The Food Web of a Tropical Rain Forest, University of Chicago Press, 1996.
[43] S. Opitz, Trophic Interactions in Caribbean Coral Reefs, vol. 1085, WorldFish, 1996.
[44] J. Memmott, N.D. Martinez, J.E. Cohen, Predators, parasitoids and pathogens: species richness, trophic generality and body sizes in a natural food web, J. Anim. Ecol. 69 (2000) 1.
[45] J. Link, Does food web theory work for marine ecosystems? Mar. Ecol. Prog. Ser. 230 (2002) 1–9.
[46] P. Yodzis, Diffuse effects in food webs, Ecology 81 (2000) 1.
[47] L. Goldwasser, J. Roughgarden, Construction and analysis of a large caribbean food web, Ecology 74 (1993) 4.
[48] N.D. Martinez, Artifacts or attributes? Effects of resolution on the little rock lake food web, Ecol. Monogr. 61 (1991) 4.
[49] D. Baird, R.E. Ulanowicz, The seasonal dynamics of the Chesapeake Bay ecosystem, Ecol. Monogr. 59 (1989) 4.
[50] M. Huxham, S. Beaney, D. Raffaelli, Do parasites reduce the chances of triangulation in a real food web? Oikos 76 (1996) 284–300.
[51] V. Batagelj, A. Mrvar, Analysis of large networks, in: Proceedings of Pajek Workshop at XXVI Sunbelt Conference, 2006.
[52] V. Batagelj, A. Mrvar, Pajek datasets, http://vlado.fmf.uni-lj.si/pub/networks/data/, 2001.
[53] J.J. Potterat, L. Phillips-Plummer, S.Q. Muth, R.B. Rothenberg, D.E. Woodhouse, T.S. Maldonado-Long, H.P. Zimmerman, J.B. Muth, Risk network structure in the early epidemic phase of HIV transmission in Colorado Springs, Sex. Transm. Infect. 78 (suppl 1) (2002) i159–i163.
[54] G.F. Davis, M. Yoo, W.E. Baker, The small world of the american corporate elite 1982–2001, Strateg. Organ. 1 (2003) 3.
[55] D. Lusseau, K. Schneider, O. Boisseau, P. Haase, E. Slooten, S.M. Dawson, The Bottlenose Dolphin community of doubtful sound features a large proportion of long-lasting associations, Behav. Ecol. Sociobiol. 54 (2003) 4.
[56] D. Krackhardt, The ties that torture: Simmelian Tie analysis in organizations, Res. Sociol. Organ. 16 (1999) 1.
[57] D. MacRae, Direct factor analysis of sociometric data, Sociometry 23 (1960) 4.
[58] J.H. Michael, J.G. Massey, Modeling the communication network in a sawmill, Forest Prod. J. 47 (1997) 9.
[59] L.D. Zeleny, Adaptation of research findings in social leadership to college classroom procedures, Sociometry 13 (1950) 4.
[60] W.W. Zachary, An information flow model for conflict and fission in small groups, J. Anthropol. Res. 33 (1977) 4.
[61] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, U. Alon, Network motifs: simple building blocks of complex networks, Science 298 (2002) 5594.
[62] D.J. Watts, S.H. Strogatz, Collective dynamics of small-world networks, Nature 393 (1998) 6684.
[63] C.R. Myers, Software systems as complex networks: Structure, function, and evolvability of software collaboration graphs, Phys. Rev. E 68 (2003) 046116.