Autorrelation and Cross-Relation of Graphs and Networks

Luciano da Fontoura Costa

Autorrelation and *Cross-Relation* of Graphs and Networks

Luciano da Fontoura Costa
luciano@ifsc.usp.br

São Carlos Institute of Physics – DFCM/USP

15th May 2022

Abstract

The concepts of self- and cross-correlation plays a key role in several areas, including signal processing and analysis, pattern recognition, multivariate statistics, as well as physics in general, as these operations underly several real-world structures and dynamics. In the present work, the concept of multiset similarity, more specifically the coincidence similarity index, is used as the basis for defining two operations between a same network, or two distinct networks, which will be respectively called autorrelation and cross-relation. In analogous manner to the self-correlation and cross-correlation counterparts, which are defined in terms of inner products between signals, the two operations suggested here allow the comparison of the similarity of graphs respectively to successive displacements along the neighborhoods of the constituent nodes, which therefore implements a role that is analogue to the lag in the class correlation. In addition to presenting these approaches, this work also illustrates their potential respectively to applications to the characterization of several model-theoretic and real world networks, providing a rich description of the specific properties of each analyzed structure. The possibility of analyzing the obtained individual autorrelation signatures in terms of their respective coincidence networks is also addressed and illustrated.

1 Introduction

Thanks to continuing advances from graph theory (e.g. [1, 2]) and network science (e.g. [3, 4, 5, 6]), graphs and complex networks now underly a wide range of concepts and applications, extending from systems biology to transportation systems. In this work, we will understand complex networks, or *networks* for short, as corresponding to particularly intricate types of *graphs* (e.g. [7, 8]).

One of the most interesting characteristics of graphs and complex networks is their ability to represent virtually every discrete system and phenomenon. Indeed, from the point of view of data structures (e.g. [9, 10, 11]), graphs can be understood as providing the most general possible manner to represent other structures, including lists, trees, lattices, etc.

The intrinsic representational generality of graphs and networks has motivated a directly related interesting issue, namely how to mathematically handle, combine and compare the represented structures. This bears a direct analogy with the manipulation of mathematical structures such as functions, fields, vectors and matrices by using arithmetic and algebraic concepts and methods including addition, product, correlation, etc.

Among the several applications operations between mathematical structures underlain by respective data structures (e.g. vectors and matrices), the operations of self- and cross-correlation, as well as the closely related self- and cross-convolution counterparts (e.g. [12, 13, 14, 15, 16]), play a key role in several important areas, including signal and image processing and analysis, mathematical physics, computer vision, control theory, and electronic engineering, to name but a few possibilities. For instance, the blurring of an image is frequently obtained by convolving the image with a gaussian kernel (e.g. [13, 17]), and the recognition of an instance of a pattern within a signal can be approached in terms of the respective cross-correlation (i.e. template matching, e.g. [13, 18, 19]).

While self- and cross-correlations are well defined respectively to vectors and matrices, their extension to graphs and networks is not so straightforward. Related approaches include the concept of *graph Fourier transform*, which involves the spectral analysis (eigenvalues and eigenvectors) of the Laplacian matrix of a given graph or network (e.g. [20, 21, 22]). The term *Fourier transform* here is used as an analogy, not literally. Another approach related to the relationship between portions of a graph is *correlation clustering*, aimed at finding cuts that optimize the allocation of edges with large weights to result within respective clusters (e.g. [23, 24, 25]). Statistical-based approaches to correlation between graphs have also been suggested (e.g. [26, 27]).

The present work sets out at investigating the possibility to use multiset similarities, in particular the coincidence similarity index [28, 29, 30], as a means for obtaining analogues of the self- and cross-correlation operations respectively to graphs and networks. The main justification for this approach is the fact that the coincidence similarity provides a selective, sensitive, and stable/robust means for comparing any mathematical structures (e.g. [29, 31, 32, 30]).

In particular, the coincidence operation is here employed in direct analogy to the inner product between two signals as the means for gauging the similarity between two graphs or networks. Then, in order to implement the operation analogous to the relative shifting of one signal respectively to the other, we consider the mean coincidence similarity between each node of the networks and the nodes belonging to successively respective neighborhoods, or hierarchies (e.g. [33, 34, 35]), at topological distance δ from the reference node. Thus, in this work the variable d plays an analogous role to the *lag* variable in traditional correlation analysis.

The resulting operations between two graphs are here called *autorrelation* and *corss-relation* so as to indicate their analogy with the classic *autocorrelation* and *cross-correlation*, but at the same time distinguishing the counterparts because they are based on different concepts: while the latter two are based on the *inner product* between functions or vectors, the *autorrelation* and *corss-relation* are founded on the *coincidence similarity* between features representing the topological (or other types) of characteristics of the considered network nodes.

In addition to presenting the thus motivated self- and cross-relation between two graphs or network (a graph with itself in the case of autorrelation), we illustrate the potential of the suggested concepts and methods respectively to the identification of lag-based relationships between model-theoretic and real-world networks. In addition, we also address the possibility to derive coincidence networks from the individual autorrelation signatures, with remarkable results that can reflect in a particularly effective manner the intricacy of the topology around the nodes of the original networks.

The obtained results indicate that the autorrelation and cross-relation between networks reflect several global and local properties of the analyzed networks, therefore constituting an interesting resource for network characterization.

We start this work by presenting the several involved concepts, presenting the suggested autorrelation and cross-relation operations, and illustrating their potential for the characterization of global and local properties of several model-theoretical and real-world networks.

2 Basic Concepts

Given two vectors $\vec{x}$ and $\vec{y}$ with N elements, their *inner product* can be expressed as:

$$\langle \vec{x}, \vec{y} \rangle = \sum_{i=1}^{N} x[i][i] \tag{1}$$

If the magnitudes $|\vec{x}|$ and $|\vec{y}|$ of the two vectors are kept fixed, the respective inner product will provide a measurement of the similarity between these two vectors.

The *cross-correlation* between two vectors (or signals) $\vec{x}$ and $\vec{y}$ with sizes N can be written as:

$$(\vec{x} \otimes \vec{y})[\ell] = \sum_{i=1}^{N-1} x[i]\ y[i-\ell] \tag{2}$$

where vector $\vec{y}$ is suitably padded with zeros and $\ell = 0, 1, \ldots, N-1$, so that the result has size $N-1$.

The *convolution* between two vectors (or signals) $\vec{x}$ and $\vec{y}$ with sizes N can be written as:

$$(\vec{x} * \vec{y})[\ell] = \sum_{i=1}^{N-1} x[i]\ y[\ell-i] \tag{3}$$

where vector $\vec{y}$ is suitably padded with zeros and $\ell = 0, 1, \ldots, N-1$, so that the result has size $N-1$.

Let $g(t)$ be a function, with respective Fourier transform $G(f)$. Its *autocorrelation* (e.g. [12, 13]) can be calculated as:

$$g(t) \otimes g(t) \Longleftrightarrow G(f)^{*} G(f) = |G(f)|^2 \tag{4}$$

where $|G(f)|^2$ is often called the *power spectrum* of $f(t)$.

Given a undirected graph g, the *degree* of each of its nodes corresponds to the respective number of links. The degree of node, in general, possibly corresponds to the most important topological information about the local topology around that node. The *clustering coefficient* or *transitivity* of a node corresponds to the ratio between the total number of links existing between the neighbors of that node and the total maximum number of links that would be possible (e.g. [5, 3, 4])

Given a graph (or network) g, and one of its nodes i, the $\mathcal{N}_\delta(i)$ *neighborhood* of i corresponds to the set of all nodes that are at topological distance smaller or equal to δ. These successive neighborhoods can be understood as defining a respective hierarchy respectively to i.

Given two networks g and h with N nodes each, we say they are *aligned* whenever for each node i of g there corresponds an associated node i in h. Observe that the two networks can have distinct interconnections.

The multiset coincidence index has been suggested as a manner to generalize the Jaccard index (e.g. [36, 37, 38, 39, 40]) in order to take into account the interiority

between the two compared mathematical structures [28], being based on multiset theory (e.g. [41, 9, 42, 43, 44, 45]). The real-valued coincidence has been found to provide more strict comparison between two multisets than the Jaccard or overlap indices taken separately, which has paved the way to enhanced selectivity, sensitivity and robustness (e.g. [29, 31, 32]). More recently, the average of the coincidence similarity established between each node of the network and all the others was suggested [46] as an feature that can be used to effectively complement to the characterization of the topology around each node.

It is henceforth assumed that all network measurements are non-negative. Otherwise, the real-valued version of the coincidence index (e.g. [28, 29]) should be considered.

If $\vec{x}$ and $\vec{y}$ are two vectors taking non-negative real values with size N, their multiset Jaccard similarity (e.g [28, 29, 30]) can be defined as:

$$\mathcal{J}(\vec{x},\vec{y}) = \frac{\sum_{i=1}^{N} \min\{x[i], y[i]\}}{\sum_{i=1}^{N} \max\{x[i], y[i]\}} \tag{5}$$

Observe that $0\mathcal{J}(\vec{x},\vec{y}) \leq 1$.

Their respective multiset *interiority* index (also overlap, e.g. [47]) can be expressed as;

$$\mathcal{I}(\vec{x},\vec{y}) = \frac{\sum_{i=1}^{N} \min\{x[i], y[i]\}}{\min\{S_x, S_y\}} \tag{6}$$

with $0 \leq \mathcal{I}(\vec{x},\vec{y}) \leq 1$.

where:

$$S_x = \sum_{i=1}^{N} x[i]; \quad S_y = \sum_{i=1}^{N} y[i] \tag{7}$$

The *multiset coincidence similarity* between the vectors $\vec{x}$ and $\vec{y}$ corresponds to the product of the respective real-valued Jaccard and interiority indices, i.e.: Their respective real-valued *interiority* index (also overlap, e.g. [47]) can be expressed as;

$$\mathcal{C}(\vec{x},\vec{y}) = \mathcal{J}(\vec{x},\vec{y})\ \mathcal{I}(\vec{x},\vec{y}) \tag{8}$$

with $0 \leq \mathcal{C}(\vec{x},\vec{y}) \leq 1$.

The real-valued coincidence similarity between the two vectors $\vec{x}$ and $\vec{y}$ can be understood as being analogous the Pearson correlation coefficient between the two vectors.

3 Autorrelation and Cross-Relation between Two Graphs

Having revised the main basic concepts, we are now in position to mathematically formalize the concepts *Autorrelation* and *Cross-Relation* between two graphs or networks.

Given two aligned networks g and h with N nodes, as well as a set of M topological features x_j, $j = 1, 2, \ldots, M$, we can associate each node $i = 1, 2, \ldots, N$ of those networks to the respective features, therefore obtaining the following *feature vectors*:

$$\vec{X}_{g,i} = [x_{g,1}[i]\ x_{g,2}[i]\ \ldots x_{g,M}[i]\ ]^T \tag{9}$$

$$\vec{X}_{h,i} = [x_{h,1}[i]\ x_{h,2}[i]\ \ldots x_{h,M}[i]\ ]^T \tag{10}$$

where $x_{g,j}[i]$ means the value of the feature x_j of node i in graph g.

The respective *cross-relation* of those two aligned vectors can now be defined as:

$$(g \boxtimes h)[i,\delta] = \frac{1}{|\mathcal{N}_\delta(i)|} \sum_{k \in \mathcal{N}_\delta(i)} \mathcal{C}(\vec{X}_{g,i}, \vec{X}_{h,k}) \tag{11}$$

where $\delta = 1, 2, \ldots, \Delta$ and $|\mathcal{N}_\delta(i)|$ is the number of neighbors of node i at topological distance δ.

In case $g = h$, the above expression becomes the *autorrelation* of a node i of the network g:

$$(g \boxtimes g)[i,\delta] = \frac{1}{|\mathcal{N}_\delta(i)|} \sum_{k \in \mathcal{N}_\delta(i)} \mathcal{C}(\vec{X}_{g,i}, \vec{X}_{g,k}) \tag{12}$$

Observe that $(g \boxtimes h)[i,\delta]$ with $\delta = 1, 2, \ldots, \Delta$ can be understood as a kind of *signature* of each of the network nodes $i = 1, 2, \ldots, N$, corresponding to how much that node is similar in the average, as far as the coincidence index is concerned, to the nodes at each successive neighborhood $\mathcal{N}_\delta(i)$. The basic underlying assumption is that the system of progressive neighborhoods around each node i can establish a coordinate axis associated to that node, and with origin at that node when $\delta = 0$. Therefore, N coordinate axes are established in this manner, allowing one of the networks to be shifted respectively to the other in this highly dimensional space.

As such, the autorrelation of a network g can be understood in analogy to the autocorrelation of a function or signal, while a similar relationship can be established between the cross-relation between two networks g and h and the cross-correlation between two signals or functions.

As defined above the autorrelation and cross-relation between two graphs will involve the coincidence comparison between the features of each node with those of successive neighborhoods, which tend to increase steadily as each node in the neighborhoods radiates more links, and then decrease as a consequence of reaching the border of the network. This implies that the average coincidence is taken respectively to varying numbers of nodes in each successive neighborhood. There are several ways in which the autorrelation and cross-relation between two graphs can be instantiated that can be used to keep this number more commensurate or even constant.

One such manner consists in comparing between each node in one network and a *walk* along the nodes in the other network. These walks can be specified in a vast number of ways, including the more traditional random walks and self-avoiding random walks (e.g. [48, 49]).

The autorrelation (cross-relation) presents potential for characterizing how much one (two) graphs are similar respectively to a sequence of respective *topological lags* δ. In the remainder of this work, we illustrate the application of the autorrelation and cross-relation concepts respectively to several model-theoretical and real-world sections.

An important aspect of the autorrelation and cross-relation between graphs as suggested above concerns the fact that these operations are relative to the set of adopted node features, which can be of topological or complementary type. This flexibility to define specific set of features allow the lag-based comparison of networks to be investigated from the point of view of several choices of measurements, aimed at reflecting network properties of specific interest in each research program. For simplicity's sake, the measurements of node degree and clustering coefficient will be adopted henceforth.

4 Results and Discussion

In this section, the concepts of autorrelation and cross-relation between two graphs are illustrates respectively to model-theoretical and real-world examples. In all cases, networks that happen to be directed are first symmetrized, and only the largest connected component is taken into account. In addition, given that maximum autorrelation is always observed for $\delta = 0$, these values are not shown for simplicity's sake. Observe that this property does not hold for the cross-relation.

We start the study of the application of the autorrelation and cross-relation concepts respective to three model-theoretical complex networks (e.g. [5, 3, 4]), namely: Erdős-Rényi – ER, Barabási-Albert – BA, and Watts-Strogatz – WS in its ring configuration. Consequently, we have the possibility to consider uniformly random interconnectivity (ER), scale free degree distribution (BA), as well as the model WS which varies, in terms of a respective rewiring probability, from a perfect lattice to an ER configuration. The ER, BA and WS with high reconnection probability are all characterized by having relatively small average shortest path distances.

Figure 1(a) presents the autorrelation obtained for an ER network with $N = 300$ nodes and average degree $\langle k \rangle = 4$, considering topological lags $\delta = 1, 2, \ldots, 10$. Shown are the several autorrelation signatures obtained for each of the 300 nodes (several random colors), as well as the respective mean signature (shown in black). Interestingly, the autorrelation function of this network can be found to decrease monotonically with two negative slopes: one smaller taking place from δ equal to 1 to 4, and another steeper slope thereon. The autorrelation function for this network can then be subdivided into two portions, the latter starting when the neighborhoods reach the border of the network. Indeed, the lower curves in this figures correspond to some of the border nodes, implying individual autorrelation that is initially small (the border elements tend not to be similar to the inner nodes) and then increases as the neighborhoods reach the core of the network.

The autorrelation function of a BA network also with $N = 300$ nodes and $\langle k \rangle = 4$ is presented in Figure 1(b). Compared to the ER case, a completely different result has been obtained. Now, the autorrelation increases progressively, and then undergoes a steep decrease. Similarity to the ER, though, the latter regime again coincides with the neighborhoods hitting the network border. The autorrelation increase, however, is a phenomenon specific to this type of network. It follows from the fact that, in a BA network, although only a few nodes are hubs, they are soon reached by the neighborhood progressions starting from the other nodes. Because the hubs are not similar to the other nodes, while the other nodes are mutually similar, the autorrelation tends to increase emphafter hitting the hubs, until the border is reached. The lower individual signatures are associated to the hubs of the network, which have markedly distinct topological features respectively to the other nodes.

Figure 1(c) depicts the autorrelation signatures obtained for a WS network (ring) with similar parameters as before, obtained by using a rewiring probability of $p_w = 0.05$. A third type of autorrelation function can be observed for this type of network, characterized by an initial plateau, followed by the a steep decrease which again is a consequence of the finite size of the network. The obtained plateau derives from the intrinsic mutual similarity between most nodes of this network. Interestingly, the nodes corresponding to the shortcuts implemented by the rewirings can be readily observed as corresponding to the lower signatures in this figure.

In order to explore in more detail the autorrelation structure of the WS networks, Figure 2 presents the results respectively obtained for WS networks with $N = 300$ nodes and rewiring probabilities $p_w = 0.01, 0.02$, and 0.03.

It can be readily observed that the increase of the rewiring probability not only reduces the range of lags characterized by non-zero autorrelations, but also influences the sharpness of the transition from core to border,

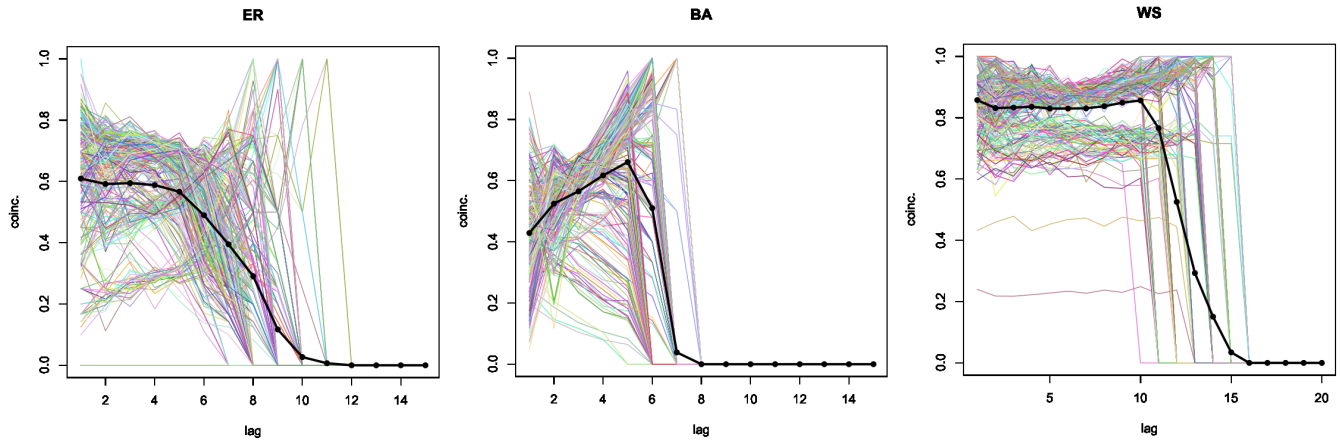


Figure 1: The autorrelation functions in terms of $\delta = 1, 2, \ldots, 30$ for ER, BA and WS networks with $N = 300$ nodes and $\langle k \rangle = 4$. The WS network was obtained by using rewiring probability $p_w = 0.05$.

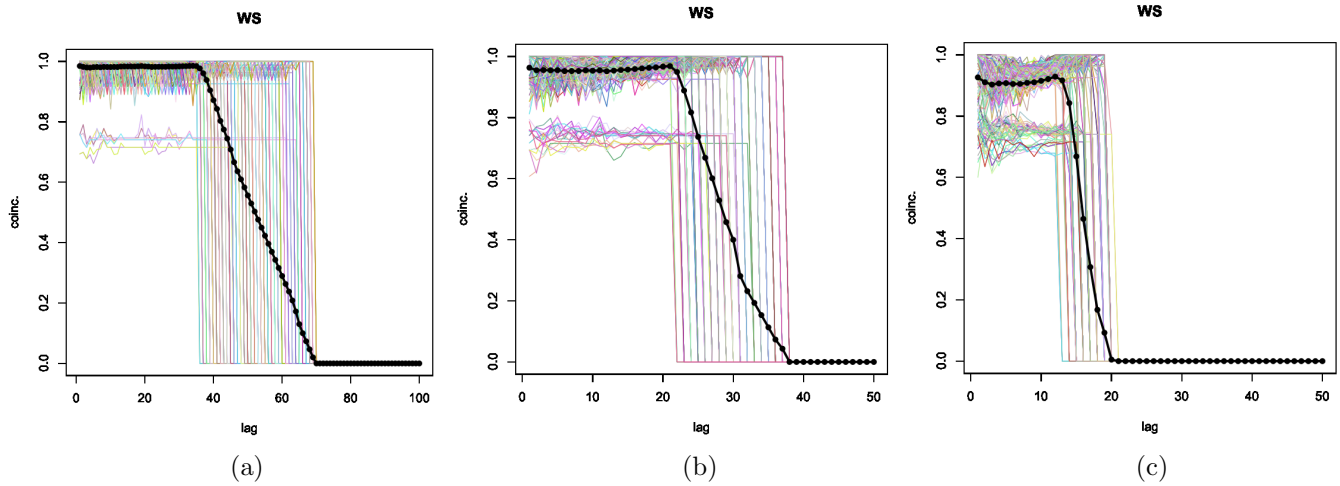


Figure 2: The autorrelation functions of an WS networks with $N = 300$ nodes, $p = 4$ and $p_w = 0.01(a), 0.02(b)$, and 0.03 (c), respectively to topological lags $\delta = 1, 2, \ldots, 30$.

as well as the height and extent of the plateaux. Also of interest are the lower signatures, which correspond to the nodes corresponding to the network rewirings.

A noticeably interesting possibility is to illustrate the autorrelations between the nodes in a given network in terms of a respective coincidence network [30]. Examples of this type of coincidence network – obtained for the above ER, BA and WS networks, is shown in Figure 3 (a), (b), and (c), respectively. The size of the nodes reflect the degree of the nodes in the original network.

As could be expected, a relatively uniform autorrelation network has been obtained for the ER network, other than by statistical fluctuations. In addition, most of the nodes are shown in cyan, reflecting the small transitivity of the original respective reference nodes.

A more interesting structure can be observed in the case of the BA structure, with the hubs either resulting in one of the extremities of the larger connected component, which are followed by nodes of signatures obtained for medium and small degree original nodes. This type of structure reflects the fact that, in a BA network, the hubs are different from the other nodes regarding their topological properties along the successive neighborhoods, but are mutually similar. As could be expected, most original nodes have very small transitivity (most nodes are in cyan).

A strongly uniform autorrelation network has been obtained for the WS case, reflecting the predominant regularity of the original network. Interestingly, some nodes in cyan can be seen around the uniform mass of node, most of which with significant transitivity values (in magenta). The former nodes correspond to the nodes the

nodes establishing shortcuts in the original network.

All in all, the autorrelation signatures obtained for the three model-theoretical networks confirm the ability of this function to reflect several of the critical properties of the respective networks, as well as their parameter configurations. The individual signatures also provide means for identifying special types of nodes.

We now proceed to verifying the types of autorrelation functions as obtained from some real-world networks.

Figure 4(a) presents the autorrelation signature obtained for the Zachary karate club network [50, 51]. Despite the relatively small size of this network, the respective autorrelation signatures suggest that it has hubs, possibly possessing a scale-free node distribution. The obtained autorrelation signatures indicate similarity with the ER model, being characterized by monotonically decreasing autorrelation values.

The autorrelation signatures obtained for the *US airports* network [51] is presented in Figure 4(d). The average signature, which involves small values, indicates a relatively more complex autorrelation structure for this network, involving the three following regimes: (a) an initial plateaux that indicates mutual similarity between each node and those at the nearest respective neighborhoods; (b) a following peak of autorrelation values, possibly related to presence of hubs in this network; and (c) a moderate decrease implied by the neighborhood progressions hitting the border of this network. An impressive variety of individual autorrelation signatures can be identified, some characterized by particularly small autorrelation values, indicating the presence of several interesting outlier nodes.

Figure 5 depicts the coincidence network obtained by considering the individual autorrelation signatures obtained for the *US airports* network as input to the coincidence methodology for transforming datasets into respective networks [30]. The resulting coincidence network has been thresholded at $T = 0.87$ for the sake of improved visualization.

Several interesting aspects can be observed from the obtained coincidence network. First, we have that the signatures have been organized along a sequence of interconnections related to the degree of the original nodes. The hubs, being most different from the other nodes, but tending to be similar each other, resulted at one of the extremities of the main obtained structure. As the degree of the nodes becomes smaller, the observed structure initially undergoes a kind of mostly sequential unfolding, up to a point where several branching and more intricate structures starts to emerge.

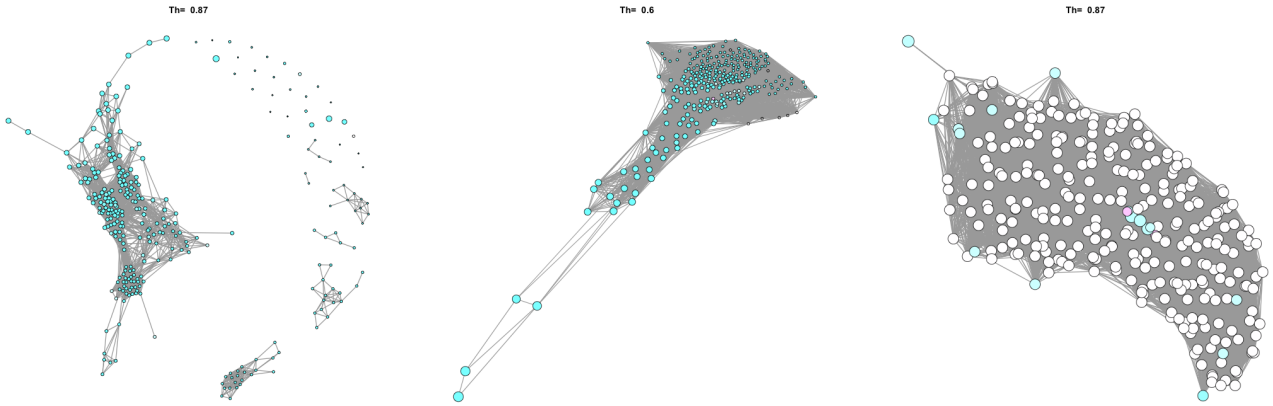
Of particular interest is the separation of this initially string-like structure into two major branches, each of which being presenting complex interconnections, including ring-like substructures. At the same time, several smaller disconnected components involving strongly interconnected nodes with smaller degree degree can be also observed. Also of particular interest are the small string component observed at the top left-hand side of the visualization in Figure 5, as well as the nodes that remained disconnected. These nodes, the vast majority of which being characterized by low transitivity, can therefore be understood as *outliers* in this network as far as their hierarchical similarity structure is concerned respectively to those of the other network nodes.

It is also interesting to compare the coincidence network obtained for the *US airports* individual autorrelation signatures with that of a BA network, as illustrated in Figure 3(b). Though these two networks share some important properties, such as the sequential connection in terms of the degree of the original nodes, they also present important differences, including the markedly more intricate structure of the *US airports* network for smaller values of original node degree, possibly as a consequence of the more local and geographical nature of the interconnections between smaller airports, thus suggesting that this network involves a mixture of structural organization principles.

In order to verify the effect of the extension of the considered range of lags taken into account while building the coincidence network from the individual autorrelation signatures, we shown in Figure 6 this type of network derived from the *US airports* individual signatures considering successive lags. The effect of taking into account longer lag extensions contributes to obtaining progressively more detailed coincidence networks. Interestingly, the complex topological structure of this network could not be discerned from the coincidence networks considering narrow lag extensions. At the same time, the effect of incorporating additional lags tends to be larger for smaller lag extensions, with progressively more similar coincidence networks being obtained as the maximum lag approaches the borders of the network. Interestingly, the enhanced selectivity and sensitivity of the coincidence similarity index provides allows the intricate level of details of the autorrelation networks to be effectively revealed.

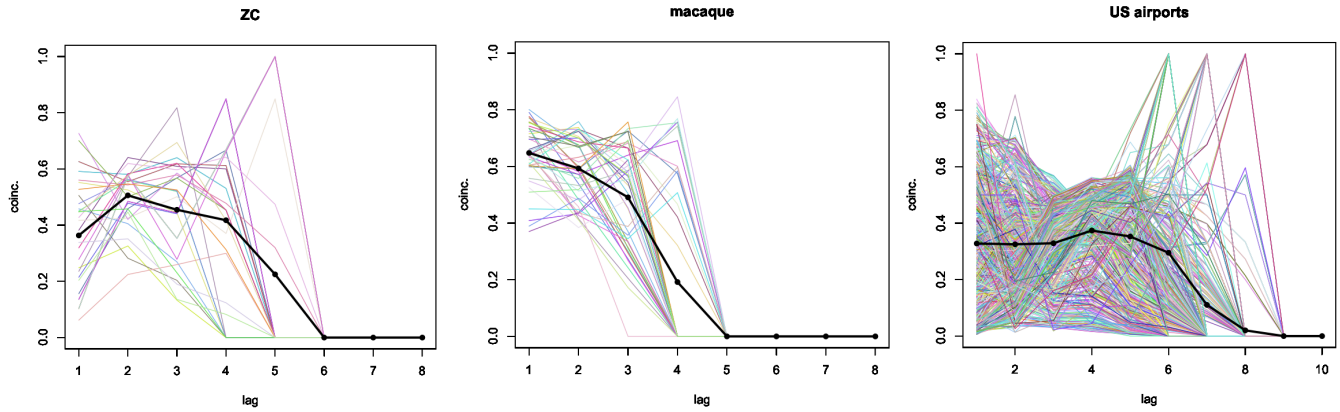
The above example illustrates the potential of the autorrelation approach for revealing impressively detailed information about the network being analyzed from the perspective of its hierarchical similarity structure.

Examples of *cross-relations* between two graphs are de-



(a) (b) (c)

Figure 3: The coincidence networks obtained from the individual autorrelation signatures of the *ER* (a), *BA* (b), and *WS* (c) networks, shown after thresholding by $T = 0.87$ ($T = 0.6$ in the BA case) and $\alpha = 0.5$. The size of the nodes corresponds to the degree of the original network nodes providing the reference for the autorrelation analysis. The width of the edges is proportional to the respective coincidence value between the respective individual autorrelation signatures. The colors (from cyan to magenta) indicate the transitivity of the original nodes.



(a) (b) (c)

Figure 4: The average (in black) and individual (random colors) autorrelation signatures obtained for the Zachary karate club network [50, 51](a); *macaque* network [52, 51](b); and *US airports* network [51] (c).

picted in Figure 7, referring to the cross-relation between original and rewired versions of a same network. First, we have the ER network in Figure 1(a) before and after 500 rewirings. Interestingly, a noticeable effect on the average signature can be observed, which is a consequence of random fluctuations given the relatively small size of this network. The individual and average autorrelation signatures obtained for the BA network are shown in Figure 7(b). Interestingly, the obtained average cross-relation signature is characterized by a less steep initial increase, which is a direct consequence of the partial loss of the scale free property, as a consequence of the hubs being rewired into more uniform configurations, which is corroborated by the associated reduction of the number of lower individual signatures. The rewiring effects on the *US airports* network, shown in Figure 7(c), included making the intermediate peak less salient, reducing the extent of the signatures at the same time as they became less dispersed, as well as making the final transition sharper.

The cross-relation between networks and their rewired versions is strongly affected by the necessarily implied reduction of the average shortest path length, which substantially narrows the extent of the signatures along the lag δ. Another interesting study of cross-relations between two graphs consists of considering the original graph and a respective version with a given percentage of hubs removed from the network. Figure 8 illustrates this type of analysis respectively to the removal of 5% of the nodes with the highest degrees respectively to the ER,

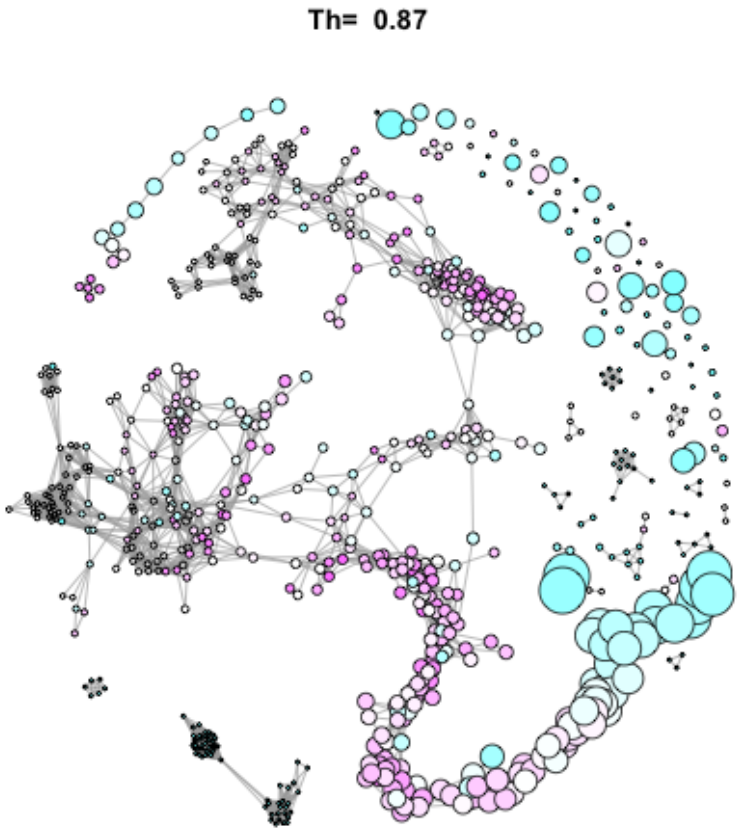


Figure 5: The coincidence network obtained from the individual autorrelation signatures of the *US airports* network, shown after thresholding by $T = 0.87$ and $\alpha = 0.5$. The size of the nodes corresponds to the degree of the original network nodes providing the reference for the autorrelation analysis. The width of the edges is proportional to the respective coincidence value between the respective individual autorrelation signatures. The colors (from cyan to magenta) indicate the transitivity of the original nodes.

BA and *US airports* networks.

As could be expected, this modification had almost no effect in the case of the ER network, which does not tend to have nodes with substantially higher degrees (hubs) except for statistical fluctuations. However, in the case of the BA network, several alterations can be verified as a consequence of the removal of only 5 hubs. We can observe a reduction of the extent of the signatures, as well as a substantial shape of the average signature, which nows corresponds mostly to a plateau instead of a peak. The effects on the *US airports* network are also strong, implying longer signatures as well as a reduction of the peak and change of the shape of most signatures. This result indicates that the hubs in this network have a strong role in defining its respective topological properties.

Figure 9 presents the cross-relation analysis of networks under another type of modification, now involving the complete permutation of the labels of all nodes respectively to the original network. The effect of this alteration was similar in the case of the ER, BA, and *US airports* networks, consisting in making the signatures more similar, therefore revealing the fact that the respective networks were perceived as being substantially more regular when compared under nodes permutation.

The above cross-relation examples above illustrate the potential of this approach for comparing between two networks while considering respective displacements along the distance lags. The obtained results indicate that the cross-relation average and individual signatures can provide interesting comprehensive information about the topological properties of the analyzed structures, especially regarding the presence of hubs and the uniformity of the topology around the nodes at successive topological scales.

To complement this section, we estimate an analogue of the "power spectrum" (see Equation 4) of the mean au-

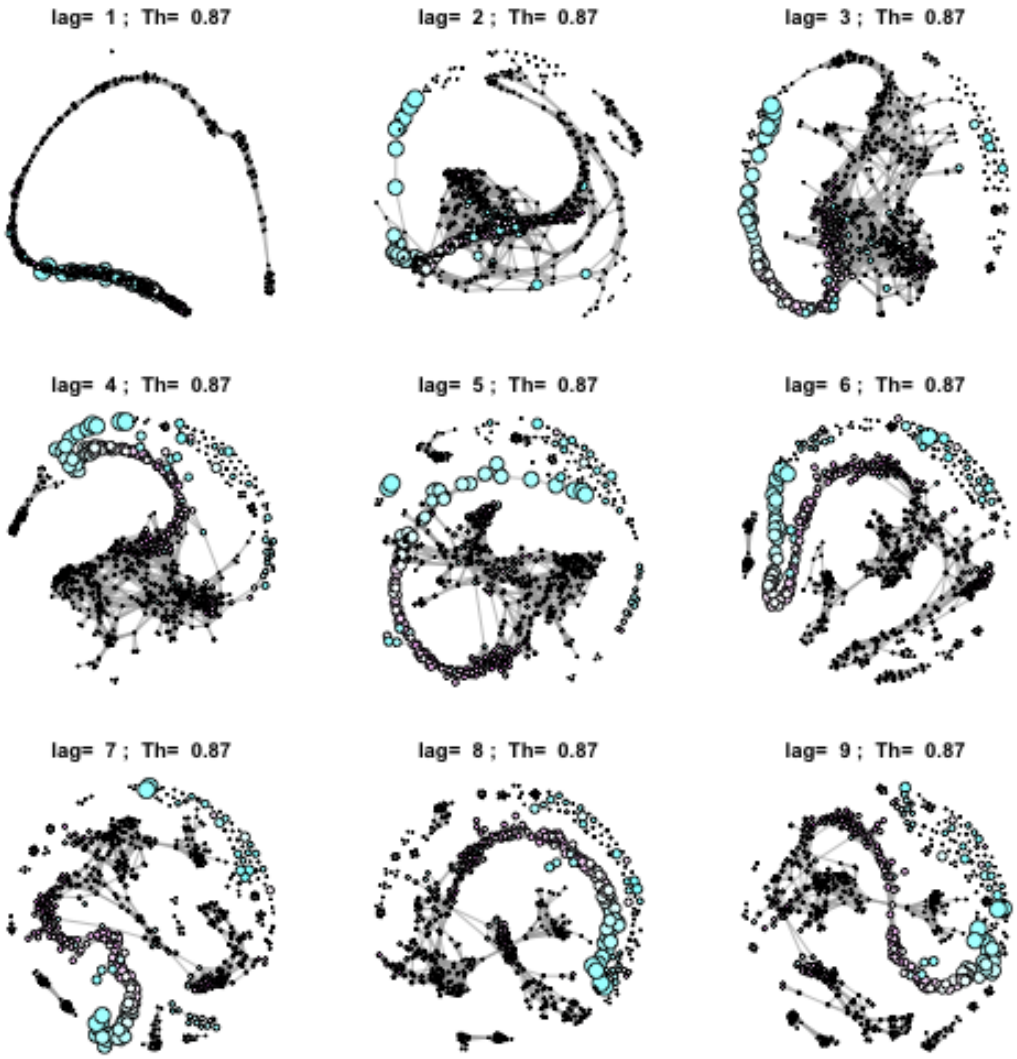


Figure 6: The coincidence networks obtained from the individual autorrelation signatures of the *US airports* network considering successive lags varying fro 1 to 9. By supplying additional information about the topology of the original network, networks derived from individual signatures with larger lag values yield more detailed coincidence networks.

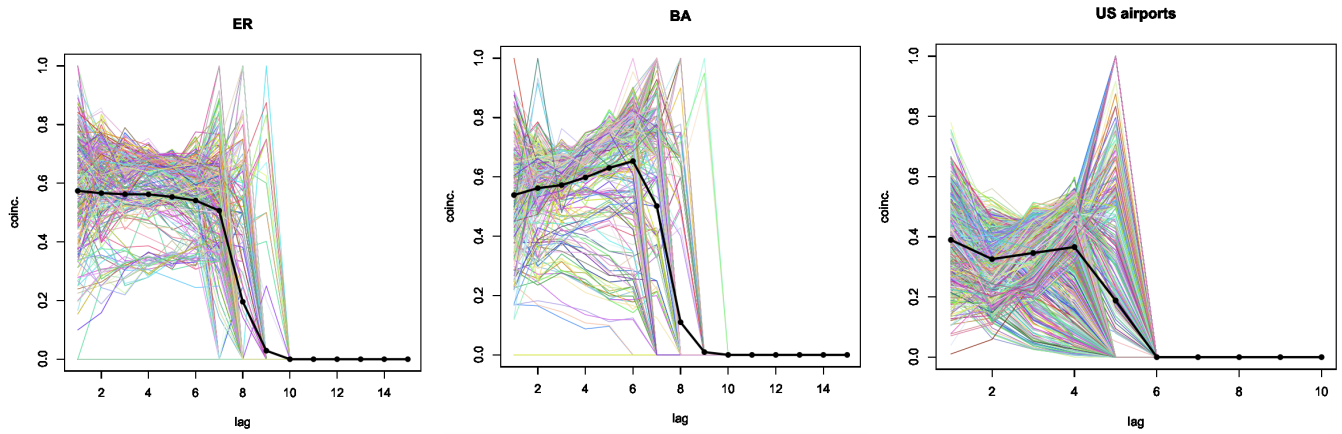
torrelation signature of the *US airport* network by using . The magnitude of the Fourier transform of the average autorrelation function obtained for the *US airports* network is shown in Figure 10.

5 Concluding Remarks

An important class of relationship between a pair of mathematical structures regards how much they are interrelated in terms of a relative displacement, controlled by respective lags. The important of this type of relationship is reflect in the key roles these operations play in areas including signal and image processing and analysis, control systems, computer vision, neuronal networks and AI at large.

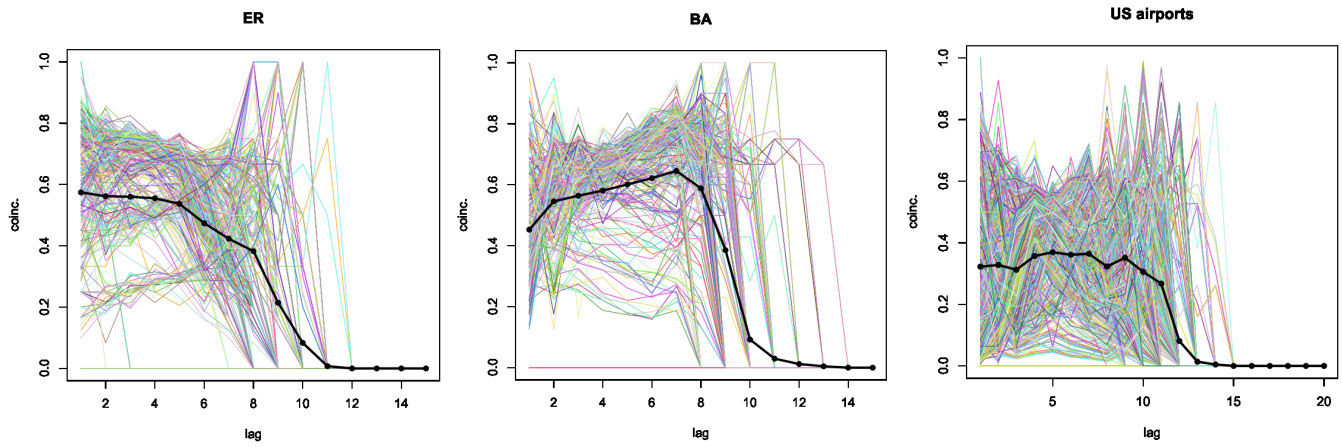
In the present work, we aimed at developing operations that are analogue to the autocorrelation and cross-correlation between functions, fields, vectors and matrices, but adapted to graphs and networks. In order to do so, we resourced to the multiset coincidence similarity, which implements a strict and stable quantification of the similarity between any two mathematical structures.

More specifically, we defined the order of neighborhoods around each network node as being analogous to the lag parameter in the classic correlation. Therefore, it became possible to compare the topological features each reference node with those of nodes in successive respective neighorhoods. This operation yields a function of the distance from the successive neighborhoods, which acts as the lag parameter. The possibility to chose among sev-



(a) (b) (c)

Figure 7: The average (in black) and individual (random colors) cross-relation signatures obtained between the ER network in Fig. 1(b) and the respectively network obtained by 500 edge rewirings (a), the original *BA* network and respective version after 4000 rewirings (b), as well as the *US airports* network and respective network obtained for 3000 rewirings (c).



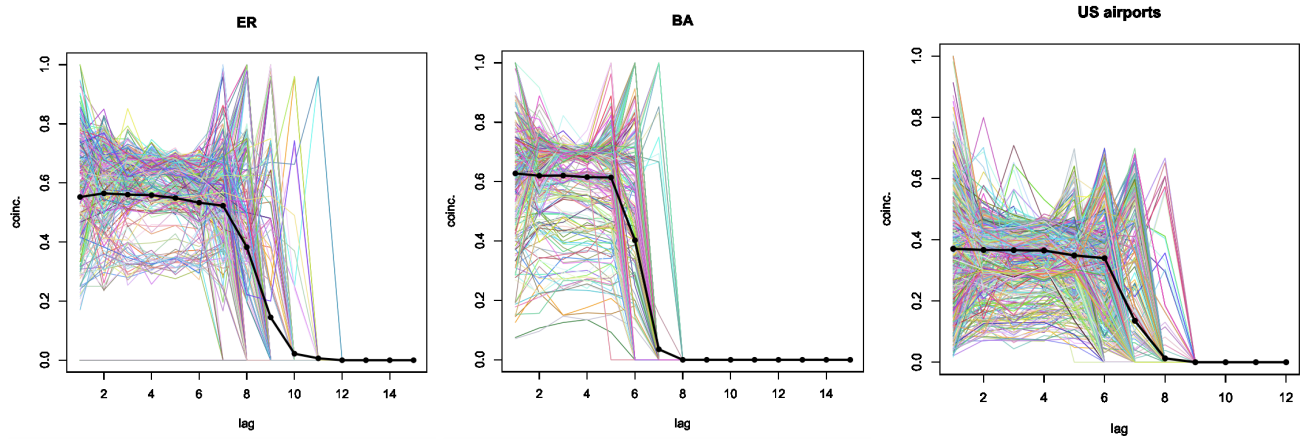
(a) (b) (c)

Figure 8: The cross-relation functions between the original ER, BA, and *US airports* networks and their respective versions after removal of the 5% of nodes with the highest degree (hubs).

eral possible topological, or other types of features accounts for additional flexibility in the implemented analysis, which can be customized to the aspects of specific interest in each research approach. Once these measurements have been obtained, the estimation of the autorrelation and cross-relation involves relatively little computational cost. In the case of the present work, which adopted the node degree and transitivity as features, the overall cost resulted particularly low given that these measurements are local and easy to be calculated.

As with the traditional autocorrelation and cross-correlation between two functions, the autorrelation and cross-relation have their maximum value observed for lag zero. As it has been indicated by several examples involving several model-theoretical and real-world networks, the average and individual autorrelation and cross-relation signatures reflect several aspects of the analyzed networks, including the presence of hubs, communities, the overall regularity, scale-free behavior, presence of outliers, as well as the extent of topological scales along which the network is more or less strongly mutually similar. Of particular interest is the intricate coincidence similarity obtained from the individual autorrelation signatures which, in the case of the *US airports* network revealed highly intricate interrelationships between involved signatures, confirming the particularly complex organization of this network.

The reported concepts and methods paves the way to several interesting further developments, including the consideration of other measurements, other model-theoretical networks (especially heterogeneous modular networks), the investigation of other parameter configurations (e.g. average degree and number of nodes). Other



(a) (b) (c)

Figure 9: The cross-relation functions between the original ER, BA, and *US airports* networks and their respective versions after permutation of nodes.

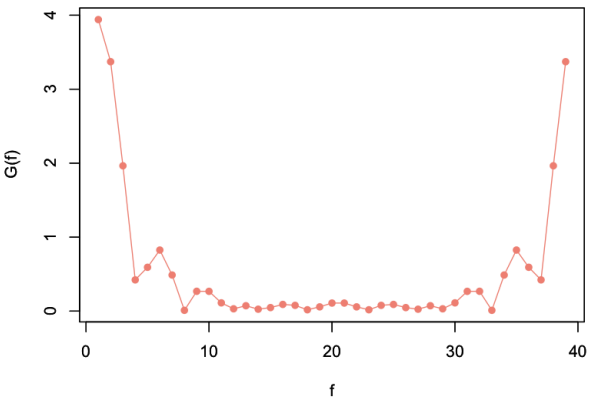


Figure 10: The Fourier transform of the average autorrelation function obtained for the *US airports* network. Though this should not be taken for the respective power spectrum, the obtained Fourier transform has special interest as it highlights eventual oscillations present in the autorrelation function, which would be related to periodical properties eventually found in the original network topology.

promising possibilities include the application to other real-world networks, as well as further relating the autorrelation and cross-relation signatures with other interesting properties of networks such as their spectra and Fourier transforms.

Acknowledgments.

Luciano da F. Costa thanks CNPq (grant no. 307085/2018-0) and FAPESP (grant 15/22308-2).

Note:

As all other preprints by the author, this work is possibly being considered by a scientific journal. Respective modification, commercial use, or distribution of any of its parts are not possible. Many of the preprints by the author are also available in HAL and arXiv. This work can also be cited by using the DOI number or article identification link. Thanks for reading.

References

[1] D. B. et al. West. *Introduction to graph theory*, volume 2. Prentice hall Upper Saddle River, 2001.

[2] B. Bollobás. *Modern graph theory*, volume 184. Springer Science & Business Media, 1998.

[3] A.L. Barabási and Pósfai M. *Network Sience*. Cambridge University Press, 2016.

[4] M. Newman. *Networks: An introduction*. Oxford University Press, 2010.

[5] L. da F. Costa, F. A. Rodrigues, G. Travieso, and P. R. Villas Boas. Characterization of complex networks: A survey of measurements. *Advances in Physics*, 56(1):167–242, 2007.

[6] L. da F. Costa, O.N. Oliveira Jr., G. Travieso, F.A. Rodrigues, P.R. Villas Boas, L. Antiqueira, M.P. Viana, and L.E.C. Rocha. Analyzing and modeling real-world phenomena with complex networks: a survey of applications. *Advances in Physics*, 60(3):329–412, 2011.

[7] L. da F. Costa. What is a complex network? https://www.researchgate.net/publication/324312765_What_is_a_Complex_Network_CDT-2, 2018. [Online; accessed 05-May-2019].

[8] L. da F. Costa and G. Domingues. Cost-based approach to complexity: A common denominator? *Revista Brasileira de Ensino de Física*, 44, 2022.

[9] D. E. Knuth. *The Art of Computing*. Addison Wesley, 1998.

[10] N. Wirth. *Algorithms & data structures*. Prentice-Hall, Inc., 1985.

[11] T. H. Cormen. *Introduction to Algorithms*. MIT Press, 2009.

[12] E. O. Brigham. *Fast Fourier Transform and its Applications*. Pearson, 1988.

[13] L. da F. Costa. *Shape Classification and Analysis: Theory and Practice*. CRC Press, Boca Raton, 2nd edition, 2009.

[14] M. Vetterli, J. Kovačević, and V. K. Goyal. *Foundations of signal processing*. Cambridge University Press, 2014.

[15] S. Haykin. *Neural Networks And Learning Machines*. McGraw-Hill Education, 9th edition, 2013.

[16] R.G. Gallager. *Stochastic Processes: Theory for Applications*. Cambridge University Press, 2014.

[17] R. C. Gonzalez and R. E. Woods and. *Digital Image Processing*. Pearson, New York, 2018.

[18] B. K. P. Horn. *Robot Vision*. McGraw Hill, Cambridge, 1986.

[19] E. R. Davies. *Machine Vision*. Morgan Kaufmann, Amsterdam, 2005.

[20] D. I. Shuman, B. Ricaud, and P. Vandergheynst. A windowed graph fourier transform. In *2012 IEEE Statistical Signal Processing Workshop (SSP)*, pages 133–136. Ieee, 2012.

[21] S. Sardellitti, S. Barbarossa, and P. Di Lorenzo. On the graph fourier transform for directed graphs. *IEEE Journal of Selected Topics in Signal Processing*, 11(6):796–811, 2017.

[22] B. Ricaud, P. Borgnat, N. Tremblay, P. Gonçalves, and P. Vandergheynst. Fourier could be a data scientist: From graph fourier transform to signal processing on graphs. *Comptes Rendus Physique*, 20(5):474–488, 2019.

[23] E. D. Demaine, D. Emanuel, A. Fiat, and N. Immorlica. Correlation clustering in general weighted graphs. *Theoretical Computer Science*, 361(2-3):172–187, 2006.

[24] A. Hero and B. Rajaratnam. Hub discovery in partial correlation graphs. *IEEE Transactions on Information Theory*, 58(9):6064–6078, 2012.

[25] N. Bansal, A. Blum, and S. Chawla. Correlation clustering. *Machine learning*, 56(1):89–113, 2004.

[26] J. Berg and M. Lässig. Correlated random networks. *Physical review letters*, 89(22):228701, 2002.

[27] A. Fujita, D. Y. Takahashi, J. B. Balardin, M. C. Vidal, and J. R. Sato. Correlation between graphs with an application to brain network analysis. *Computational Statistics and Data Analysis*, 109:76–92, 2017.

[28] L. da F. Costa. Further generalizations of the Jaccard index. https://www.researchgate.net/publication/355381945_Further_Generalizations_of_the_Jaccard_Index, 2021. [Online; accessed 21-Aug-2021].

[29] L. da F. Costa. On similarity. https://www.sciencedirect.com/science/article/pii/S037843712200334X, 2022. Physica A: Statistical Mechanics and its Applications, 127456.

[30] L. da F. Costa. Coincidence complex networks. https://iopscience.iop.org/article/10.1088/2632-072X/ac54c3, 2022. J. Phys.: Complexity, (3): 015012.

[31] L. da F. Costa. Comparing cross correlation-based similarities. https://www.researchgate.net/publication/355546016_Comparing_Cross_Correlation-Based_Similarities, 2021. [Online; accessed 21-Oct-2021].

[32] L. da F. Costa. Multiset neurons. https://www.researchgate.net/publication/356042155_Multiset_Neurons, 2021.

[33] B. A. N. Travençolo and L. da F. Costa. Hierarchical spatial organization of geographical networks. *Journal of Physics A: Mathematical and Theoretical*, 41(22):224004, 2008.

[34] L. da F. Costa and F. N. Silva. Hierarchical characterization of complex networks. *Journal of Statistical Physics*, 125(4):841–872, 2006.

[35] L. da F. Costa and R. F. S. Andrade. What are the best concentric descriptors for complex networks? *New Journal of Physics*, 9(9):311, 2007.

[36] P. Jaccard. Distribution de la flore alpine dans le bassin des dranses et dans quelques régions voisines. *Bulletin de la Société vaudoise des sciences naturelles*, 37:241–272, 1901.

[37] P. Jaccard. Étude comparative de la distribution florale dans une portion des alpes et des jura. *Bulletin de la Société vaudoise des sciences naturelles*, 37:547–549, 1901.

[38] B. K. Samanthula and W. Jiang. Secure multiset intersection cardinality and its application to jaccard coefficient. *IEEE Transactions on Dependable and Secure Computing*, 13(5):591–604, 1989.

[39] Wikipedia. Jaccard index. https://en.wikipedia.org/wiki/Jaccard_index. [Online; accessed 10-Oct-2021].

[40] A. Schubert and A. Telcs. A note on the Jaccardized Czekanowski similarity index. *Scientometrics*, 98:1397–1399, 2014.

[41] J. Hein. *Discrete Mathematics*. Jones & Bartlett Pub., 2003.

[42] W. D. Blizard. Multiset theory. *Notre Dame Journal of Formal Logic*, 30:36—66, 1989.

[43] W. D. Blizard. The development of multiset theory. *Modern Logic*, 4:319–352, 1991.

[44] P. M. Mahalakshmi and P. Thangavelu. Properties of multisets. *International Journal of Innovative Technology and Exploring Engineering*, 8:1–4, 2019.

[45] D. Singh, M. Ibrahim, T. Yohana, and J. N. Singh. Complementation in multiset theory. *International Mathematical Forum*, 38:1877–1884, 2011.

[46] L. da F. Costa. On the self-coincidence structure of networks. https://www.researchgate.net/publication/361182049_On_the_Self-Coincidence_Structure_of_Networks, 2022.

[47] M. K. Vijaymeena and K. Kavitha. A survey on similarity measures in text mining. *Machine Learning and Applications*, 3(1):19–28, 2016.

[48] L. Lovász. Random walks on graphs. *Combinatorics, Paul erdos is eighty*, 2(1-46):4, 1993.

[49] J. D. Noh and H. Rieger. Random walks on complex networks. *Physical review letters*, 92(11):118701, 2004.

[50] W. W. Zachary. An information flow model for conflict and fission in small groups. *Journal of anthropological research*, 33(4):452–473, 1977.

[51] G. Csardi and T. Nepusz. The igraph software package for complex network research. *InterJournal*, Complex Systems:1695, 2006.

[52] L. Négyessy, T. Nepusz, L. Kocsis, and F. Bazsó. Prediction of the main cortical areas and connections involved in the tactile function of the visual cortex by network analysis. *European Journal of Neuroscience*, 23(7):1919–1930, 2006.