Chaos, Solitons and Fractals 107 (2018) 26–38

Contents lists available at ScienceDirect

Chaos, Solitons and Fractals

Nonlinear Science, and Nonequilibrium and Complex Phenomena

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

Directed clustering in weighted networks: A new perspective

G.P. Clemente [a], R. Grassi [b,*]

[a] Department of Mathematics, Finance and Econometrics, Catholic University of Milan, Italy
[b] Department of Statistics and Quantitative Methods, University of Milano–Bicocca, Italy

ARTICLE INFO

Article history:
Received 28 September 2017
Revised 4 December 2017
Accepted 11 December 2017

Keywords:
Complex networks
Clustering coefficient
Weighted networks
Directed graphs

ABSTRACT

Several definitions of clustering coefficient for weighted networks have been proposed in literature, but less attention has been paid to both weighted and directed networks. We provide a new local clustering coefficient for this kind of networks, starting from those already existing in the literature for the weighted and undirected case. Furthermore, we extract from our coefficient four specific components, in order to separately consider different link patterns of triangles. Empirical applications on several real networks from different frameworks and with different order are provided. The performance of our coefficient is also compared with that of existing coefficients.

1. Introduction

Literature in network theory mainly focused on unweighted undirected networks and several topological properties of networks have been identified through useful indicators, which enhance the efficiency of a network in carrying out its essential functionality ([1–3]). Among these is the case of clustering coefficient that measures the tendency to which nodes in a graph tend to cluster together. Indeed, in most real networks empirical evidence shows that nodes tend to form tightly-knit groups characterized by a relatively high density of ties. In other words, the clustering coefficient is a measure of cohesion and it was developed with the aim to quantify the level to which a network manifests this property.

Different definitions of clustering coefficient have been proposed for binary undirected networks (BUN). A global coefficient, often referred as transitivity, gives an overall indication of the clustering in the network being measured as the fraction of triplets (i.e. three nodes with at least two ties among them) that are closed (i.e. they form a triangle) (see [4,5]). A local version has been also introduced in [3] in order to quantify how close the node's neighbours are from being a clique. Although it suffers from a number of limitations[1](see [6,7]), the local coefficient is capable to capture the degree of social embeddedness of single nodes and for instance it is used by several mainstream indicators to assess small-world property of a network (see [3,8,9]). Unlike the local clustering coefficient, the transitivity does not suffer from the same type of limitations because it is not an average of individual fractions calculated for each node. However, in many context this additional information is needed for each node. Indeed networks could be highly clustered at local level, despite showing a transitivity coefficient significantly low (see [10] page 83). Hence a more node-oriented analysis is often required to better investigate the network cliquishness.

It is to be mentioned that other measures of neighbours' interconnectedness have also been provided in the literature, but focusing on different topological aspects. For instance, the overlapping coefficient (see [7,11–14]) considers the number of triangles to which an edge belong. The aim is to catch the relative topological overlap of the neighborhood of two users, representing the proportion of their common friends. A node overlapping index has also been obtained as the ratio of the sum of these overlapping indices to the number of neighbours (see [11]). This measure can be interpreted as an other form of clustering attachment measure (see [15]). Overlapping concept has been also applied in community detection context. In this framework, the intuition behind overlapping clustering is based on the fact that real complex networks usually are not divided into sharp sub-networks, but typically nodes may naturally belong to more than one communities. Thus, being able to identify the overlapping communities of directed networks, could give fruitful insights about network structure (see [16–18]).

* Corresponding author.
E-mail addresses: gianpaolo.clemente@unicatt.it (G.P. Clemente), rosanna.grassi@unimib.it (R. Grassi).

[1] For instance, the local clustering coefficient can be biased by correlation with node's degree.

https://doi.org/10.1016/j.chaos.2017.12.007

Furthermore, while binary networks allowed to properly model many real-world phenomena, further complexity is often needed to adequately catch heterogeneous strengths and asymmetric connections between pairs of nodes. In these contexts weighted and directed networks are fruitful tools. Furthermore, it is well known that many real-world complex systems involve non-mutual relationships, which imply non-symmetric adjacency or weighted matrices. As regard to this issue, the transitivity coefficient has been extended to both binary and weighted directed networks in [19]. The proposed generalization retains the information encoded in the weights of ties. At the same time, local clustering coefficient have been also generalized to weighted undirected networks (WUN) by considering different ways to weight the neighbours of a node (see [20,21], etc.). See [22,23] for a review of such definitions in the literature.

In this context Fagiolo ([24]) attempts to bridge different approaches (proposed in [21,25]) in order to present a unifying framework for computing local clustering for weighted directed networks (WDN). In addition to the measures already discussed in [21,25], the coefficient proposed in [24] allows to explicitly account for directed and weighted links and to define a specific clustering coefficient for any type of triangle pattern. However, as partially[2] noticed also in [26], this coefficient does not properly account for the strength of a node, resulting in a clustering coefficient too affected by weights.

To overcome this issue, we propose a new local clustering coefficient for weighted and directed networks based on a generalization of the clustering coefficient developed in [20]. On one hand, our proposal takes into account the triangles that the neighbours of a node i form, completely preserving the initial idea of the clustering coefficient. On the other hand, the weights of these triangles also affect the coefficient. In our proposed clustering coefficient we do not consider the weight of the closing link of a triangle (i.e. the link between adjacent neighbours of i). This is because the aim of the clustering coefficient is to assess the likelihood of the occurrence of this link that closes the triangle, and not its weight. A proper normalization of the local coefficient is assured by considering the strength of the node. Hence, both the number of triangles and their weights are captured by our coefficient, that in this way well replicate, for weighted and directed networks, the idea of nodes to be “clustered together”.

Numerical results point out that the proposed coefficient proves effective in capturing both the number of closed triangles and the presence of strong neighbours (i.e. with higher weights), that classical indicators fail to correctly detect. The coefficient treats all possible directed triangles as they were the same, as if directions of edges were irrelevant. Furthermore, as in [24], we are able to provide alternative coefficients that only consider particular types of directed triangles. In other words, the proposed measure is capable to distinguish different patterns of directed triangles from a node perspective. In this way, we allow for different interpretation in terms of the resulting patterns.

The paper is organized as follows. Section 2 introduces some basic definitions and notations used in the paper; Section 2.2 briefly reviews some local clustering coefficients provided in the literature for weighted undirected networks. Also the clustering coefficient, given in [24] for weighted directed networks, is reported. Section 3 describes a new clustering coefficient for weighted directed networks proposed in order to overcome some pitfalls of the existing one. Additionally, we look at directed networks at different “observation scales” in order to separately catch different patterns. Four clustering coefficients are proposed, whose weighted average coincides with the overall coefficient. A toy example compares our proposal with the existing coefficient given in [24]. Section 4 provides numerical results, to compare our procedure to classical coefficients on the empirical networks considered. Section 5 concludes.

[2] [26] compares alternative clustering coefficients for complete weighted graphs.

2. Preliminaries

2.1. Basic notations

We assume that the reader is familiar with standard graph theory definitions. We only remind here the notations we use in the rest of the text. Formally, a directed graph (or digraph) $D=(V,A)$ is a pair of sets V and A, where V is the set of n vertices (or nodes) and A is the ordered set of m pairs (arcs) of vertices of V; if (i, j) or $(j, i) \in A$, then vertices i and j are adjacent.

A weight $w_{ij}>0$ can be associated with each link (i, j) so that a weighted directed graph is obtained; we assume that, if omitted, the weight w_{ij} of an arc (i, j) is equal to 1 (binary case). In general, both adjacency relationships between vertices of D and weights on the arcs are described by a nonnegative, real n-square matrix $\mathbf{W}$ (the weighted adjacency matrix). In the unweighted case, matrix $\mathbf{W}$ is simply the classical binary matrix $\mathbf{A}$ (the adjacency matrix). In the next, we will consider the case of either unweighted or weighted graphs with no loops (i.e. $a_{ii}=0,\ w_{ii}=0\ \forall i$).

We call $G=(V,E)$ the graph in which every edge corresponds to an arc (i, j) or (j, i) in $D=(V,A)$. Observe that G is a weighted graph and to every arc (i, j) with weight $w_{ij}>0$ corresponds an edge (i, j) with weights $w_{ij}=w_{ji}$. G represents the “symmetric case” of D. The matrix $\mathbf{W}$ (or $\mathbf{A}$, for the unweighted case) associated to G is, of course, a symmetric matrix. The (i, j) element of the $k-$power of the $\mathbf{A}$ is the number of walks of length k from i to j.

Using the same notation as in Fagiolo ([24]), we define the in-degree (respectively out-degree) of a node i as the number of arcs pointing towards (respectively starting from) i:

$$d_i^{in}=\sum_{j\neq i}a_{ji}=\mathbf{A}_i^T\mathbf{1} \tag{1}$$

$$d_i^{out}=\sum_{j\neq i}a_{ij}=\mathbf{A}_i\mathbf{1}. \tag{2}$$

where $\mathbf{A}_i$ and $\mathbf{A}_i^T$ are respectively the i-th row of $\mathbf{A}$ and of its transpose, $\mathbf{1}$ is the unit column vector of n elements. The degree d_i^{tot} of a vertex is then:

$$d_i^{tot}=d_i^{in}+d_i^{out}=(\mathbf{A}^T+\mathbf{A})_i\mathbf{1}. \tag{3}$$

Bilateral arcs between the node i and its adjacent nodes, if any, are represented as:

$$d_i^{\leftrightarrow}=\sum_{j\neq i}a_{ij}a_{ji}=\mathbf{A}_{ii}^2. \tag{4}$$

Moving to the weighted case, the previous definitions can be replaced by the strength of a node i:

$$s_i^{in}=\sum_{j\neq i}a_{ji}w_{ji}=(\mathbf{A}^T\mathbf{W})_{ii}=\mathbf{W}_i^T\mathbf{1} \tag{5}$$

$$s_i^{out}=\sum_{j\neq i}a_{ij}w_{ij}=(\mathbf{A}\mathbf{W}^T)_{ii}=\mathbf{W}_i\mathbf{1}. \tag{6}$$

The total strength of i is then:

$$\begin{aligned} s_i^{tot}&=s_i^{in}+s_i^{out}=\sum_{j\neq i}\left(a_{ji}w_{ji}+a_{ij}w_{ij}\right)\\ &=\left(\mathbf{A}^T\mathbf{W}+\mathbf{A}\mathbf{W}^T\right)_{ii}=\left(\mathbf{W}^T+\mathbf{W}\right)_i\mathbf{1}. \end{aligned} \tag{7}$$