Stratified cohesiveness in complex business networks

Roy Cerqueti[a,b,*], Gian Paolo Clemente[c], Rosanna Grassi[d]

[a] Sapienza University of Rome, Department of Social and Economic Sciences, Piazzale Aldo Moro, 5 - I-00185, Rome, Italy
[b] London South Bank University, School of Business, London SE1 0AA, UK
[c] Catholic University of Milan, Department of Mathematics for Economics, Financial and Actuarial Sciences, Largo Gemelli, 1, 20123 Milan, Italy
[d] University of Milano-Bicocca, Department of Statistics and Quantitative Methods, Via Bicocca degli Arcimboldi, 8, 20126 Milan, Italy

*Corresponding author
Email addresses: roy.cerqueti@uniroma1.it (Roy Cerqueti), gianpaolo.clemente@unicatt.it (Gian Paolo Clemente), rosanna.grassi@unimib.it (Rosanna Grassi)

Stratified cohesiveness in complex business networks

Abstract

In this work, we propose a measure that aims at assessing the position of a node with respect to the interconnected groups of nodes existing in a network. In particular, since the nodes of a network can be placed at different distances from cohesive groups, we extend the standard concept of clustering coefficient and provide the local l-adjacency clustering coefficient of a node i as an opportunely weighted mean of the clustering coefficients of nodes which are at distance l from i. Thus, the standard clustering coefficient is a peculiar local l-adjacency clustering coefficient for $l = 0$. As l varies, the local l-adjacency clustering coefficient is then used to infer insights on the position of each node in the overall structure. Empirical experiments on special business networks are carried out. In particular, the analysis of air traffic networks validate the theoretical proposal and provide supporting arguments on its usefulness.

Keywords: Structural cohesion, Complex business networks, Geodesic distance in networks, Cohesive stratification.

1. Introduction

The analysis of inter-relationships between entities is a strategic aspect to be investigated in business research. In organizations, the individual decisions are affected by the position of the decision maker in structures describing mutual relationships: in the contect of entrepreneurship, this is of particular relevance when evaluating firms' performance. Thus, business literature has largely focused on discovering and analysing clusters of organizations (see e.g. [1] and, more recently, [2]). In general, a managerial architecture is grounded on the relationships among agents, so that network structures are particularly effective in offering a representation of the complexity of the relationship among the involved actors. Firms and managers are usually interconnected through an intricate weave. Hence, discovering how they are cohesive at every level of the structure can give useful insights into entrepreneurial strategic choices (see e.g. [3, 4, 5, 6]). In the area of financial economics, the authors in [7] deal with the clusters of networks of economies by using data on the Gross Domestic Product of a group of countries.

In this perspective, it becomes important to assess at which level, or how deep an individual is involved in cohesive structures. At a local level, the interconnections around a node are generally represented and measured through its clustering coefficient. Such a coefficient can be defined operatively, being computed as the relative number of actual triangles to which the vertex belongs over the hypothetical ones. It has been developed in all the cases of weighted, unweighted, directed and undirected networks. In this respect, we mention the classical contributions of [8, 9, 10, 11] and the non recent but highly informative monograph [12]. Recently, [13] contains a relevant extension of the clustering coefficient proposed by [9]; [14] discusses the clustering coefficient in presence of already established communities for directed networks; [15] presents a concept of clustering coefficient which also includes the presence of missing indirect links in the construction of the triangles.

In socio-economic contexts, the idea of cohesion is strictly related to the so-called "embeddedness", that refers to the fact that economic action and outcomes are affected by both actors' pairwise relations and the structure of the overall network ([16, 17]). This idea focuses on the mutual relationships between the neighbours of the node of interest, that is, in other words, the concept of transitivity. Such a concept is, at the end, what the clustering coefficient mathematically represents. It is worth to underline that the clustering coefficient is a local measure; hence, it does not capture the topology of the whole network. Moreover, the global clustering coefficient of the network – which is simply obtained by taking the average of the local versions over all the nodes of the network – often is not very informative. Indeed, a network can be highly dense at a local level but not at a global one, and this suggests that the average of local clustering could not well represent the global characteristics of the network. This drawback is peculiarly relevant, since it is often crucial to assess the position of a node with respect to cohesive groups in which the node itself is not directly involved (see [18]).

In this paper we propose a measure that aims at assessing the position of a node in a cohesive group, by considering the role and the relevance of the cohesiveness of nodes which are placed at a geodesic distance $l \geq 1$ from the considered one. In details, we define the local l-adjacency clustering coefficient of a node i as an opportunely weighted mean of the clustering coefficients of the nodes at geodesic distance l from i.

The proposed approach fits well with the existing literature. Indeed, in order to overcome the analysis of the structural embeddedness limited to an actor's direct neighborhood, several attempts have been made; in this respect, it is worth mentioning the nested k-connected sets (see e.g. [16, 19]). Under this perspective, our method could be seen as an alternative generalization of the "direct embeddedness", not through nesting but through higher-order clustering coefficients. Precisely, in this work we interpret the l-th stratus associated with a node i as the group of nodes that are at distance l from i. Assessing the stratified cohesiveness structure, or, simply, the stratified structure of a network (see below for a formal definition of this concept), represents a crucial point for understanding the contextualization of the nodes within the overall system. For instance, assume that a node i has a low clustering coefficient but the nodes at a distance $l > 1$ from it (i.e., the nodes that belong to the stratus l of the node i) exhibit a strong cohesiveness. In this case, the node i is not embedded in a powerful cohesive group of nodes but it is surrounded by a high level of mutual interconnections for the nodes quite far from it. This situation can be interpreted under the perspective of shocks propagation. Indeed, if a node located at a geodesic distance from i greater than l receives a shock, such a shock could hardly reach i, since i is surrounded by highly cohesive nodes at a large distance from it and by an empty space – in terms of cohesiveness –

around it. In other words, the highly interconnected nodes at distance l might be viewed as forming a sort of barrier for i, absorbing the external solicitations at larger distances.

It is crucial to point out that we do not aim here at providing a new model describing hierarchical communities in the network. Indeed, our perspective is to provide a concept of stratification which does not nest one stratus in the ones at higher levels, like in the standard dendrogram analysis; rather, we present a framework where all the strati are individual entities, and none of them is included in another one. Thus, strati cannot be intepreted as "hierarchies". For a detailed analysis of this aspect, please refer to dendrogram hierarchical clustering (see e.g. [20, 21, 22])

Furthermore, hierarchical communities represent the core of a wide strand of research focusing on community detection (see e.g. [23, 24, 25, 26, 27]). Undoubtedly, such a research theme is of high relevance either under a methodological point of view as well as for practical applications. However, the present paper is quite far from this type of literature. In fact, communities detection models are basically grounded on suitably defined partition problems over the entire set of the nodes, and communities are represented by the classes of the partition. In so doing, communities detection problems have a global analysis approach over the set of the nodes. Differently, we here explore the cohesiveness property of the individual nodes when considering their geodesic distance from the other nodes of the network; therefore, our approach is here of a purely local type.

With this proposal, we are placed in the strand of literature that aims at catching the effect of interconnection patterns that go beyond the local neighbours (see the interesting point of view in [28] on this). However, we provide a different measure with respect to [29], that assesses how much a node is involved in larger cliques linking the result to community-detection methods (see also [30]). The proposed approach is also different with respect to [31], whose authors provide a clustering coefficient based on the probability, computed under the assumption of the Barabási-Albert model, that there is a specific distance between two neighbours of the node. Since we assess the role of nodes, we also work differently from [32] and [33], where edge centrality scores are proposed providing different weights to local centres and global bridges that connect different regions.

To test the proposed measure we perform an empirical application to the paradigmatic business network related to the U.S. domestic air traffic. In this field, the emergence of competition and the transition in ownership requires the adoption of a different perspective in airport management (see, e.g., [34]). This explains the presence of several studies on airport business. For instance, as stressed in [35], the description of the topological and metric structure of the network is of great importance for understanding the business strategies adopted by different airlines or by different airports, for assessing passengers' mobility in the presence of direct and indirect connections or for investigating the evolution of airports to changes in the passengers' demand and reactions to economical external forces, such as deregulation. Additionally, the so-called "network connectivity", i.e. the extent to which an airport or a network of airports connects users of aviation to the outside world ([36]), is particularly useful to policymakers. Indeed, the authors would like to get insights into how well the national or local airport system responds to the connectivity needs of a country and understand how to enhance air connectivity to support economic growth. As emphasized by the International Transport Forum (see, e.g., [37]), the analysis of network connectivity is relevant for both airports and government in order to make strategic business decisions.

Therefore, we test, with our proposal, how a node is near/far from a strong cohesiveness structure or, in other words, how much an airport is embedded in the airport network. Findings confirm the effectiveness of these measures in seizing the peculiar characteristics of different hubs in the airport network. We observe that larger hubs are not only highly interconnected with the other nodes of the network but also connected to strong cohesive groups. When higher distances are analysed, the effects of indirect connections with remote airports are emphasized. Referring to large and medium hubs, the proposed approach allows to emphasize the strategic role of such nodes in the airport network system. Finally, although the considered empirical network is only weakly asymmetric and it is typically analysed as an undirected network in the literature (see, for instance, [8, 13, 38, 39, 40, 41]), we show that a separate evaluation of directed paths at high levels can be useful to identify specific patterns in terms of in- and out-cohesiveness – i.e. in terms of the property of being part of cohesive groups when considering arcs directions of in- and out-flows.

The rest of the paper is organized as follows. In Section 2, we present the mathematical preliminaries and the basic notation used in the article. In Section 3, the concept of local l-adjacency clustering coefficient is defined for both undirected and directed case. In Section 4 the new conceptualization of clustering coefficients is applied to the U.S. airport network, in order to discuss its stratified structure. Conclusions are in Section 5.

2. Preliminary definitions and notations

We briefly present the mathematical definitions used in the paper. A graph $G = (V, E)$ is identified by a set V of n vertices and a set E of m unordered pairs of vertices (called edges). Vertices i and j are said to be adjacent if $(i, j) \in E$. Graphs considered here are without loops. The degree d_i of the vertex i is the number of its adjacent nodes. A path between two vertices i and j is a sequence of distinct vertices and edges between i and j. In this case, i and j are connected. G is connected if every pair of vertices is connected. The distance $d(i, j)$ is the length of any shortest path (or geodesic) between i and j. If i and j are not connected, then $d(i, j) = \infty$. The diameter $diam(G)$ of a connected graph is the length of any longest geodesic.

Given a connected graph G, we define the set:

$$N_i(l) = \{j \in V | d(i, j) = l\}$$

of nodes $j \in V$ which are at distance l with respect to the node i, where

$l = 1, \dots, diam(G)$. The cardinality $|N_i(l)|$ is denoted as $d_i(l)$. Notice that $N_i(1)$ is the set of nodes adjacent to i, so that $d_i(1) = d_i$.

A graph G is weighted when a positive real number $w_{ij} > 0$ is associated with the edge (i, j). Therefore, $w_{ij} = 0$ if nodes i and j are not adjacent. In particular, when $w_{ij} = 1$ if $(i, j) \in E$, then the graph is unweighted. Thus, the unweighted case can be viewed as a particular weighted one. For this reason, we use in this paper only the general concept of weighted graphs and we denote a weighted graph with its weights simply as weighted network.

The strength of the vertex i is defined as $s_i = \sum_{j=1}^{n} w_{ij}$.

In general, both adjacency relationships between vertices of G and weights on the edges are described by a nonnegative, real n-square matrix $\mathbf{W}$ (the *weighted adjacency matrix*), with entries w_{ij}.

A weight can be also associated with every geodesic of a connected graph G in the following way: let $\mathcal{G}_{ij}$ be the set of the geodesics connecting the vertices i and j; the generic element of $\mathcal{G}_{ij}$ is $g = g_{ij}$.

We observe that more than one geodesic connecting i and j could exist, so that, in general, $|\mathcal{G}_{ij}| \geq 1$. Recalling that all the geodesics in $\mathcal{G}_{ij}$ have the same length, a unique integer $l = l(i, j)$ exists such that the length of all the paths in $\mathcal{G}_{ij}$ is l. The role of l is crucial for the following arguments. Therefore, l will be explicitly added in the notation when needed so that, for instance, $\mathcal{G}_{ij}$ will be $\mathcal{G}_{ij}(l)$, g will become $g(l)$ and so on.

The weight of $g_{ij}(l)$ is the sum of the weights of its edges and we denote it with $w_{ij}(l, g)$. This allows us to define the l-th order strength of the node i in this setting as

$$s_i(l) = \sum_{j \in N_i(l)} w_{ij}(l),$$

where $w_{ij}(l) = \min_{g \in \mathcal{G}_{ij}(l)} \{w_{ij}(l, g)\}$. Notice that when $j \in N_i(1)$, then $\mathcal{G}_{ij}(1) = \{(i, j)\}$; hence $w_{ij}(1) = w_{ij}$ and $s_i(1)$ is the strength s_i of the vertex i.

A directed graph $D = (V, E)$ is obtained from G by adding to its edges a direction, where G is the underlying graph of D. In this case, the links between couples of nodes are called directed edges or arcs. In a weighted directed graph, a weight $w_{ij} > 0$ is associated with the directed edge (i, j) and, in general, the matrix $\mathbf{W}$ is not symmetric. In fact, since bidirectional edges between a pair of nodes can exist, both w_{ij} and w_{ji} can be positive with $w_{ij} \neq w_{ji}$. For the sake of notation, we will denote by $\overrightarrow{w}_{ij}$ and $\overleftarrow{w}_{ij}$ the weight w_{ij} of the arc directed from i to j and the weight w_{ji} of the arc from j to i, respectively.

A directed path from i to j is a sequence of distinct vertices and arcs from i to j such that every arc has the same direction; in this case, we say that j is reachable from i and we call this *out-path* of the node i. The distance $\overrightarrow{d}(i, j)$ from i to j is the length of such shortest out-path (or *out-geodesic*) if any, otherwise $\overrightarrow{d}(i, j) = \infty$.

Since directed paths from j to i can also exist, we define the *in-path* of the node i as the directed path from j to i and we denote with $\overleftarrow{d}(i, j)$ the length of any shortest such in-path (or *in-geodesic*). If i is not reachable from j, then $\overleftarrow{d}(i, j) = \infty$.

If i and j are mutually reachable, both in and out-geodesics of i exist, although the distances $\overleftarrow{d}(i, j)$ and $\overrightarrow{d}(i, j)$ can be different. D is strongly connected if every two vertices are mutually reachable.

D is weakly connected if the underlying graph G is connected. That it means that a geodesic g between i and j exists in the underlying graph G. In this case, distance $d(i, j)$ is finite for all $i, j \in V$.

In addition, we define the following sets:

1. $\overrightarrow{N}_i(l) = \{j \in V | \overrightarrow{d}(i, j) = l\}$, for each $l = 1, \ldots, diam(G)$;
2. $\overleftarrow{N}_i(l) = \{j \in V | \overleftarrow{d}(i, j) = l\}$, for each $l = 1, \ldots, diam(G)$.

Moreover, according to what we did above, we can define $\overrightarrow{\mathcal{G}}_{ij}$ and $\overleftarrow{\mathcal{G}}_{ij}$ the sets of the out-geodesics and in-geodesics connecting the vertices i and j, respectively; the generic elements of $\overrightarrow{\mathcal{G}}_{ij}$ and $\overleftarrow{\mathcal{G}}_{ij}$ are $\overrightarrow{g} = \overrightarrow{g}_{ij}$ and $\overleftarrow{g} = \overleftarrow{g}_{ij}$, respectively.

Hence, the definition of weighted in- and out-geodesics can be easily given by setting:

1. $\overrightarrow{w}_{ij}(l) = \min_{\overrightarrow{g} \in \overrightarrow{\mathcal{G}}_{ij}(l)} \{\overrightarrow{w}_{ij}(l, \overrightarrow{g})\}$;
2. $\overleftarrow{w}_{ij}(l) = \min_{\overleftarrow{g} \in \overleftarrow{\mathcal{G}}_{ij}(l)} \{\overleftarrow{w}_{ij}(l, \overleftarrow{g})\}$,

where $\overleftarrow{w}_{ij}(l, \overleftarrow{g})$ ($\overrightarrow{w}_{ij}(l, \overrightarrow{g})$) is the sum of all the weights of the arcs of the geodesic $\overleftarrow{g}$ ($\overrightarrow{g}$) of length l connecting i and j.

This allows us to define the l-th order in and out-strength of the node i as

$$\overrightarrow{s}_i(l) = \sum_{j \in \overrightarrow{N}_i(l)} \overrightarrow{w}_{ij}(l),$$
$$\overleftarrow{s}_i(l) = \sum_{j \in \overleftarrow{N}_i(l)} \overleftarrow{w}_{ij}(l).$$

3. Stratified cohesiveness

The aim of this section is to define a new indicator of structural cohesiveness around a node i based on the mutual interconnections between nodes at different distances from i. This new indicator uses an extended idea of clustering coefficient, moving along shortest paths. Hence, we are providing a conceptualization of the stratified structure around the nodes of a network.

We introduce the indicator discussing separately the undirected and directed case.

3.1. Local l-adjacency clustering coefficient: undirected case

Let $\mathbf{P}(l) = [p_{ij}(l)]_{i,j \in V}$ for $l = 1, \ldots, diam(G)$ be the matrix such that:

$$p_{ij}(l) = \begin{cases} \frac{w_{ij}(l)}{s_i(l)} & \text{if } j \in N_i(l), \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

For the sake of completeness, the definition of the matrix $\mathbf{P}(l)$ can be extended to the case $l = 0$, by setting $\mathbf{P}(0) = \mathbf{I}$, being $\mathbf{I}$ the identity matrix.

We define the vector of the *local l-adjacency clustering coefficients* of the nodes of the network $\mathbf{c}(l) = [c_i(l)]_{i \in V}$, as:

$$\mathbf{c}(l) = \mathbf{P}(l)\mathbf{c}, \tag{2}$$

where $\mathbf{c} = [c_i]_{i \in V}$ is the vector whose element c_i is the weighted clustering coefficient of node i, defined in Barrat et al. ([8]) as

$$c_i = \frac{1}{s_i(d_i - 1)} \sum_j \sum_{k \neq j} \frac{w_{ij} + w_{ik}}{2} a_{ij} a_{ik} a_{jk}, \tag{3}$$

being a_{ij} equal to 1 if $(i, j) \in E$ and 0 otherwise. Notice that, when $l = 0$, formula (2) gives $\mathbf{c}(0) = \mathbf{c}$, and then we recover the weighted local clustering coefficient defined in [8].

When $l = 1$, the local l-adjacency clustering coefficient $\mathbf{c}(1) = [c_i(1)]_{i \in V}$ is the vector of elements:

$$c_i(1) = \frac{1}{s_i} \sum_{j \in V} w_{ij} c_j, \tag{4}$$

where each element represents the weighted average of the clustering coefficients c_j of the nodes j which are adjacent to i. This is also true for $l > 1$, as in general, formula (2) states that $c_i(l)$ is the weighted average, with weights $w_{ij}(l)$, of clustering coefficients c_j of nodes j which are at distance l with respect to the node i. Furthermore, it is noteworthy that, in case of an unweighted graph, the coefficient $c_i(l)$ reduces to a classic arithmetic mean.

The elements of the vector defined in (2) give insights about the network structure at a specific distance from nodes of the graph. Indeed, the element l associated with the proposed definition of the clustering coefficient explains how the nodes at distance l from i are interconnected in the graph.

Large values of such clustering coefficients of i at high levels l suggest that i is connected with well-established highly cohesive nodes which are far from the node itself. In other words, the analysis of $\mathbf{c}(l)$ with $l = 0, 1, \ldots, diam(G)$ leads to a complete view of the graphs in terms of cohesiveness, and this might give insights on how shocks propagate. The quantity $c_i(l)$ describes the *cohesiveness at stratus l around the node i* and the set of vectors $\mathbf{c}(l)$ defines the *stratified cohesiveness structure – or, as already said before, the stratified structure – of the network* as the value of l varies.

In order to measure the overall cohesiveness structure around a node i, we then introduce the vector $\mathbf{h} = [h_i]_{i=0,1,\ldots,n}$ such that

$$h_i = \sum_{l=0}^{diam(G)} x_l c_i(l), \tag{5}$$

where $x_l \geq 0$, for each l, and $\sum_{l=0}^{diam(G)} x_l = 1$.

The vector $\mathbf{h} = [h_i]_{i=0,1,\ldots,n}$ allows tracking the whole network structure around a single node i, and takes into account all the strati. Notice that the selection of a peculiar distribution of the weights $(x_0, x_1, \ldots, x_{diam(G)})$ provides the meaning of the concept of stratified cohesiveness for all the nodes of the graph. In particular, high polarization of such a distribution at low (high) level l leads to core-based (periphery-based) identification of the cohesiveness structure. The special case $x_l = 1$ focuses attention only to interconnected nodes at stratus l.

3.2. Local l-adjacency clustering coefficients: directed case

We consider a directed, weighted and weakly connected graph D.
As already pointed out in Section 2, in addition to weighted paths, also weighted in- and out-paths can exist and we can focus only on a specific pattern (out-path or in-path for the node i), or we can consider all edges' directions. Each choice is reasonable and depends on the kind of problem we deal with.

For $l = 1, \ldots, diam(G)$, we define the matrix $\bar{\mathbf{P}}(l)$ with the following entries:

$$\bar{p}_{ij}(l) = \begin{cases} \frac{\bar{w}_{ij}(l)}{\bar{s}_i(l)} & \text{if } j \in \bar{N}_i(l) \text{ and } \bar{N}_i(l) \neq \emptyset, \\ 0 & \text{otherwise,} \end{cases} \tag{6}$$

where:

(*a*) $\bar{N}_i(l) = \overrightarrow{N}_i(l)$, $\bar{w}_{i,j}(l) = \overrightarrow{w}_{ij}(l)$ and $\bar{s}_i(l) = \overrightarrow{s}_i(l)$ if only out-paths of node i are considered. In this case, $\bar{\mathbf{P}}(l)$ will be denoted by $\overrightarrow{\mathbf{P}}(l)$;

(*b*) $\bar{N}_i(l) = \overleftarrow{N}_i(l)$, $\bar{w}_{i,j}(l) = \overleftarrow{w}_{ij}(l)$ and $\bar{s}_i(l) = \overleftarrow{s}_i(l)$ if only in-paths of node i are considered. In this case, $\bar{\mathbf{P}}(l)$ will be denoted by $\overleftarrow{\mathbf{P}}(l)$;

(c) $\bar{N}_i(l) = N_i(l)$, $\bar{w}_{i,j}(l) = w_{ij}(l)$ and $\bar{s}_i(l) = s_i(l)$ if all the directions of the edges are taken into account. In this case, $\bar{\mathbf{P}}(l)$ will be denoted by $\mathbf{P}(l)$.

Also in this case, we set $\bar{\mathbf{P}}(0) = \mathbf{I}$.

The definition of local l-adjacency clustering coefficients, introduced in formula (2) has to be extended to the three cases (a), (b) and (c).

Indeed, in the specific context of directed graphs, edges pointing in different directions have a completely different interpretation in terms of the resulting flow pattern. To this aim, alternative in-type or out-type local l-adjacency clustering coefficients can be also obtained as:

$$\mathbf{c}^{in}(l) = \overleftarrow{\mathbf{P}}(l)\mathbf{c}^{in}. \tag{7}$$

$$\mathbf{c}^{out}(l) = \overrightarrow{\mathbf{P}}(l)\mathbf{c}^{out}. \tag{8}$$

where $\overleftarrow{\mathbf{P}}(l)$ and $\overrightarrow{\mathbf{P}}(l)$ are matrices defined in formula (6) whose elements are computed considering only in-paths (case (b)) or out-paths (case (a)), respectively. $\mathbf{c}^{in} = [c_i^{in}]_{i\in V}$ and $\mathbf{c}^{out} = [c_i^{out}]_{i\in V}$ are the vectors whose elements c_i^{in} and c_i^{out} are the in and out local clustering coefficients defined in [13] as:

$$c_i^{in} = \frac{\frac{1}{2}[\mathbf{W}^T(\mathbf{A}+\mathbf{A}^T)\mathbf{A}]_{ii}}{\overleftarrow{s}_i(1)\left(\overleftarrow{d}_i(1)-1\right)}$$

and

$$c_i^{out} = \frac{\frac{1}{2}[\mathbf{W}(\mathbf{A}+\mathbf{A}^T)\mathbf{A}^T]_{ii}}{\overrightarrow{s}_i(1)\left(\overrightarrow{d}_i(1)-1\right)}.$$

These two coefficients convey information about clustering of two different patterns (in or out) within tightly connected directed neighbourhoods.

According to the case (c), we define the local l-adjacency clustering coefficients as:

$$\mathbf{c}^{all}(l) = \mathbf{P}(l)\mathbf{c}^{all}, \tag{9}$$

where $\mathbf{c}^{all} = [c_i^{all}]_{i\in V}$ is the vector whose elements c_i^{all} are the weighted and directed clustering coefficients provided in [13].

The difference between the vectors of local l-adjacency clustering coefficients in (2) and (9) lies in the considered definition of triangles. The same triple of nodes might be associated with one triangle (no directions of the arcs to be taken into account) in the former case and two of them (two possible directions for the arcs) in the latter one. In the special case of absence of bilateral arcs, the following result holds true:

Proposition 1. *Let D be a directed graph. If the graph D has not bilateral arcs, then $\mathbf{c}^{all}(l) = \frac{\mathbf{c}(l)}{2}$, where $\mathbf{c}(l)$ is the vector of the local l-adjacency clustering coefficients of the undirected underlying graph G.*

Proof. We define $\mathbf{W}(D)$ and $\mathbf{W}(G)$ the weighted adjacency matrices of the graphs D and G. The local clustering coefficient for weighted and directed network is defined as (see [13]):

$$c_i^{all} = \frac{\frac{1}{2}\left[\left(\mathbf{W}(D) + \mathbf{W}^T(D)\right)\left(\mathbf{A}(D) + \mathbf{A}^T(D)\right)^2\right]_{ii}}{s_i(1,D)\left(d_i(1,D) - 1\right) - 2s_i^{\leftrightarrow}(1,D)} \tag{10}$$

where $s_i(1,D)$ and $d_i(1,D)$ are respectively the total strength and the degree of the node i, whereas $s_i^{\leftrightarrow}(1,D)$ is the strength related to bilateral arcs between i and its adjacent nodes.

Since the graph D has not bilateral arcs, $s_i^{\leftrightarrow}(1,D) = 0$. Additionally, if nodes i and j are adjacent, a (unique) weighted arc between i and j exists so that if $\overrightarrow{w}_{ij}(1,D) > 0$ then $\overleftarrow{w}_{ij}(1,D) = 0$ or vice versa. As a consequence, $\mathbf{W}(D) + \mathbf{W}^T(D) = \mathbf{W}(G)$ and $\mathbf{A}(D) + \mathbf{A}^T(D) = \mathbf{A}(G)$.

Furthermore,

$$s_i(1,D) = \overrightarrow{s}_i(1,D) + \overleftarrow{s}_i(1,D) = \sum_{j \in \overrightarrow{N}_i(1)} \overrightarrow{w}_{ij}(1,D) + \sum_{j \in \overleftarrow{N}_i(1)} \overleftarrow{w}_{ji}(1,D) = \sum_{j \in N_i(1)} w_{i,j}(1,G) = s_i(1,G).$$

A similar chain of equalities entails that $d_i(1,D) = d_i(1,G)$.

Hence, (10) yields, $\forall i = 1, ..., n$:

$$c_i^{all} = \frac{\frac{1}{2}\left[\mathbf{W}(G)\mathbf{A}^2(G)\right]_{ii}}{s_i(1,G)\left(d_i(1,G) - 1\right)} = \frac{c_i}{2} \tag{11}$$

and, by formula (9):

$$\mathbf{c}^{all}(l) = \mathbf{P}(l)\mathbf{c}^{all} = \frac{1}{2}\mathbf{P}(l)\mathbf{c} = \frac{1}{2}\mathbf{c}(l). \tag{12}$$

□

Analogously to formula (5), we can define $\mathbf{h}^{in} = [h_i^{in}]_{i\in V}$, $\mathbf{h}^{out} = [h_i^{out}]_{i\in V}$, and $\mathbf{h}^{all} = [h_i^{all}]_{i\in V}$, where:

$$h_i^{in} = \sum_{l=0}^{diam(G)} x_l c_i^{in}(l), \qquad h_i^{out} = \sum_{l=0}^{diam(G)} x_l c_i^{out}(l), \qquad h_i^{all} = \sum_{l=0}^{diam(G)} x_l c_i^{all}(l). \tag{13}$$

3.3. How the local l-adjacency clustering coefficients work: an illustrative example

The classical clustering coefficient does not give insights on both the topological structure and the stratified structure of the whole network, being a measure of the local cohesiveness structure concentrated around the nodes of the network. For the same reason also the global clustering coefficient of the network, which is given by the average of the local version around the nodes, is often not very informative. Furthermore, the nodes of a network can be highly cohesive at a local level but not on a global level, so that the average of the local clustering could not well represent the global characteristics of the network (see [18]).

To illustrate how the local l-adjacency clustering coefficients effectively map the cohesiveness structure in the network, we show how they work by comparing two small graphs. For the sake of simplicity, we limit our investigation to undirected and unweighted graphs, as interesting remarks can be done also in this very simplified, but meaningful, case.

Let us consider the two graphs G and G', sharing the same number of nodes (namely, 9) and same diameter (namely, 4), but different topology, since graph G' have more arcs than G (hence, showing a stronger cohesiveness structure, see Figure 1). In both graphs, node 1 shares the same neighbours and, being part of the same triangles, it has the same clustering coefficient $c_1(0)$. This coefficient is then not effective in capturing the actual position of the node with respect to the cohesiveness in the network. Indeed, the classical clustering coefficient gives insights about how the node 1 is embedded in a cohesive group only respect to its neighbours.

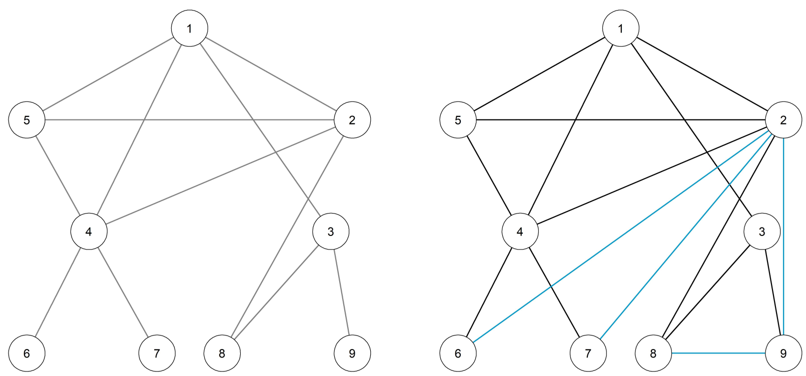

Figure 1: Graphs G (left side) and G' (right side).

To have a more reliable information about how the node is located with respect to the whole structure, in particular to cohesiveness properties of existing nodes, we need to analyse its position looking deeper than its neighbours. Hence, the l-adjacency clustering coefficients (with $l = 1, .., 4$) in this sense are meaningful.

Node	$c_i(0)$	$c_i(1)$	$c_i(2)$	$c_i(3)$	$c_i(4)$	h_i Decreasing Weights	h_i Uniform Weights	h_i Increasing Weights
1	0.50	0.45	0	0	0	0.32	0.19	0.09
2	0.50	0.45	0	0	0	0.32	0.19	0.09
3	0	0.17	0.60	0	0	0.12	0.15	0.11
4	0.30	0.40	0	0	0	0.22	0.14	0.07
5	1	0.43	0	0	0	0.53	0.29	0.14
6	0	0.30	0.50	0	0	0.14	0.16	0.11
7	0	0.30	0.50	0	0	0.14	0.16	0.11
8	0	0.25	0.45	0	0	0.12	0.14	0.09
9	0	0	0.25	0.60	0	0.10	0.17	0.17

Table 1: Graph G: l-adjacency clustering coefficients and elements of vector **h** computed for different choices of weights.

Comparing the values of the clustering coefficients in Tables 1 and 2, the coefficient of node 1 decreases with respect to l in the case of G (from 0.5 to 0)

Node	$c_i(0)$	$c_i(1)$	$c_i(2)$	$c_i(3)$	h_i Decreasing Weights	h_i Uniform Weights	h_i Increasing Weights
1	0.50	0.53	0.83	0	0.50	0.47	0.34
2	0.29	0.76	0.33	0	0.37	0.35	0.24
3	0.33	0.61	0.60	1	0.52	0.63	0.76
4	0.50	0.76	0.56	0	0.51	0.45	0.31
5	1	0.43	0.73	0	0.70	0.54	0.36
6	1	0.39	0.77	0.33	0.74	0.62	0.53
7	1	0.39	0.77	0.33	0.74	0.62	0.53
8	0.67	0.43	0.80	0	0.55	0.47	0.34
9	0.67	0.43	0.80	0	0.55	0.47	0.34

Table 2: Graph G': l-adjacency clustering coefficients and elements of vector $\mathbf{h}$ computed for different choices of weights.

moving through paths of length greater than 1, whereas in case of graph G' it increases till $l = 2$ (from 0.5 to 0.83) and vanishes for $l = 3$.

Therefore, the different strati of the cohesiveness structure around the node 1 well reflect the position of such a node with respect to the way the other nodes are connected in the structure.

Through elements h_i we are able to simultaneously consider all the cohesiveness properties of the nodes at the different strati. We control the impact of the coefficients through their weights x_l. We consider here three possible scenarios for the weights x's:

- Decreasing weights: $x_l = \frac{(l+1)^{-1}}{\sum_{h=0}^{diam(G)}(h+1)^{-1}} = \frac{(l+1)^{-1}}{H_G}$ where H_G is the harmonic number of order $diam(G) + 1$, for each l;
- Uniform weights: $x_l = \frac{1}{diam(G)+1}$, for each l;
- Increasing weights: $x_l = \frac{(l+1)}{\sum_{h=0}^{diam(G)}(l+1)} = \frac{2(l+1)}{(diam(G)+1)(diam(G)+2)}$, for each l.

Natural interpretations of the weights arise. Decreasing weights, for instance, reduce the impact on the node of high distances when assessing the cohesiveness. Notice that, the elements in $\mathbf{h}$ do not provide similar information of the classical average clustering coefficient. Instead, we are measuring the position of the node inside the network looking at each stratus. These indicators then provide an overall look and, at the same time, they track the node distances from the more cohesive groups of nodes in the network.

4. Empirical experiments

In order to see how the proposed indicator is effective in assessing how much a node is connected to a cohesive group, we test it on the peculiar business network of the U.S. airports, where the nodes are the airports and the arcs are weighted on the basis of the number of passengers among them in a given year. The analysis is of particular interest in this specific field since the hub-and-spoke operations have changed the competition between airlines and airports in a structural way (see, e.g., [36, 42]). Through their networks, airlines compete both directly and indirectly. On the one hand, the relevance of an airport in the network depends on direct routes. On the other hand, they compete indirectly with a transfer at a hub. Hence, to have a global view of the strategic role of the airports we measure how an airport is located with respect to groups of interconnected airports.

The considered reference year in the proposed experiments is 2017. The network is constructed by using the Air Carrier Statistics database (available on the U.S. Department of Transportation[1]), also known as the T-100 data bank, that contains domestic and international airline market and segment data. Both certificated U.S. air carriers and foreign carriers (having at least one point of service in the United States or one of its territories) report monthly traffic information. The weight of an arc corresponds to the number of emplaned passengers[2]. It considers revenue emplaned passengers within the U.S., and passengers emplaned outside U.S. but deplaned within the U.S. as well.

In the reference year 2017, the airport network has 1701 nodes and 27005

[1] Data are collected by the Office of Airline Information, Bureau of Transportation Statistics, Research and Innovative Technology Administration.

[2] The term "emplaned passengers", widely used in the aviation industry, refers to passengers boarding a plane at a particular airport. Since the majority of the revenue of an airport are generated, directly or indirectly, by emplaned passengers, this number is the most important air traffic metric. Data consider the total number of revenue passengers boarding an aircraft (including originating, stopover, and transfer passengers) in both scheduled and non-scheduled services.

arcs, considering both domestic and international flights. Density is around 0.001, showing a very sparse network. Moreover, significant differences are observed between big and small airports. To give a preliminary idea of the network, Figure 2 depicts the U.S. domestic airport network. In order to preserve the clarity of the figure, we reported only arcs with weights greater than 95th percentile (equal to 198,540) of weights' distribution. In other words, for the sake of simplicity we are displaying only routes with more than about 200,000 enplaned passengers.

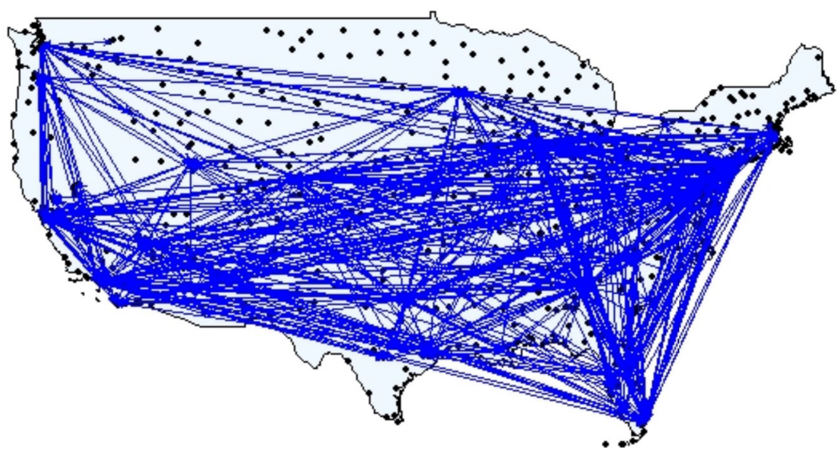

Figure 2: Domestic U.S. airport network built on the basis of 2017 data. Only arcs with weights greater than 95th percentile of weights' distribution have been reported.

Figure 3 reports the distributions of total strength for U.S. airports, capturing the total passenger traffic during 2017. Airports have been split according to Federal Aviation Administration (FAA) categories. According to FAA, a large hub is an airport which accounts for at least 1% of total U.S. passenger enplanements. A medium hub is defined as an airport accounting for a percentage of the total passenger enplanements ranging between 0.25% and 1% (see Tables A.3 and A.4 in the Appendix for the list of large and medium U.S. hubs). A small hub is associated with a percentage ranging between 0.05% and 0.25%. Last categories concern smaller airports, that are divided between non-hub and non-primary if they have respectively more or less than 10,000 annual passengers.

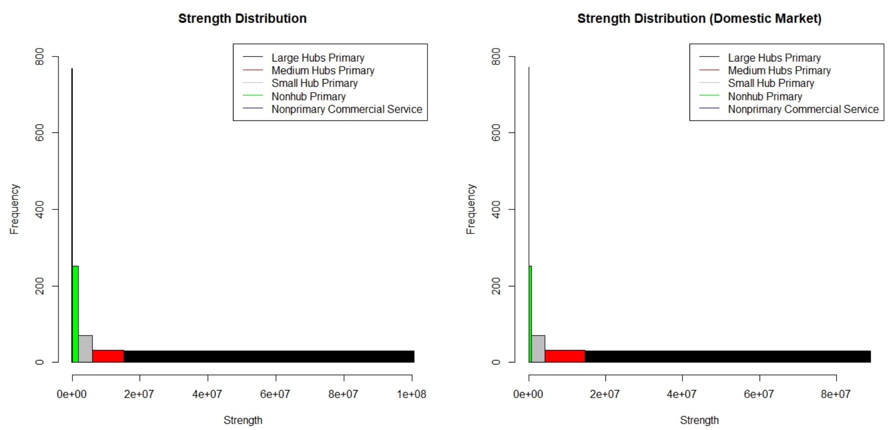

Figure 3: Total strength distribution of U.S. airports for domestic and international markets and only domestic market respectively. Airports are classified according to FAA categories.

System-wide passenger enplanements is not far from 900 millions of passengers. The 30 large hubs move 70% of the passengers, and this ratio becomes higher than 85% if also medium hubs are included. These data are in line with the ones published by [43].

Furthermore, the degrees of the nodes of the network highlight that each U.S. airport is connected on average to 23 airports, unless large hubs are connected to more than 200 airports.

If we focus only on domestic market, we have roughly 740 millions of passengers. In this case, the network is characterized by 1149 U.S. airports and 20445 connections between them. As shown by the strength distribution (Figure 3, right side), a significant proportion of traffic (around 85%) is concentrated around the top 61 airports, considering both large and medium hubs.

For the sake of brevity, we do not report a graphical representation of the strength distributions for in and out-flows. However, it is worth mentioning that, for both indicators, in and out results are strongly correlated with the total strength distribution. In other words, except for some specific airports, we observe similar patterns between the number of passengers departing from and arriving at each airport.

In order to compute the local l-adjacency clustering coefficients, we consider only the U.S. domestic market network, preventing possible distorted effects due to international flights. Indeed, data regarding connections between airports located outside of U.S. territory are not included in the dataset, and triangles could be impossible to be found. Notice that the restriction to the domestic flights does not lead to a noticeable bias of the analysis of the U.S. airport network as a whole, since domestic market covers roughly 80% of total passengers that arrive or depart from the U.S. airports in 2017.

Figure 4 displays the distributions of the components of the local l-adjacency clustering coefficients vector $\mathbf{c}^{all}(l)$ computed at different levels l and considering as nodes either all the airports (on the left side) or only large and medium hubs (on the right side). We also report synthetic measures in $\mathbf{h}(l)$ for alternative choices of weights x's.

As a premise, the classical global clustering coefficient, obtained as mean of the coefficients $c_i^{all}(0)$ over the nodes $i = 1, ..., n$, is equal to 0.56. When referring to the local l-adjacency clustering coefficients, we notice that the distribution shows high volatility, enhancing relevant differences between airports, and negative skewness, showing a median equal to 0.67, significantly greater than the mean.

Focusing on large and medium airports (Figure 4, right side), the average clustering increases, as the mean is equal to 0.69 and the median is equal to 0.71. Except for the Ted Stevens Anchorage International[3] – a medium hub located in Alaska – all relevant airports in terms of passengers traffic have a clustering

[3] $c_i^{all}(0)$ is equal to 0.21 for this airport.

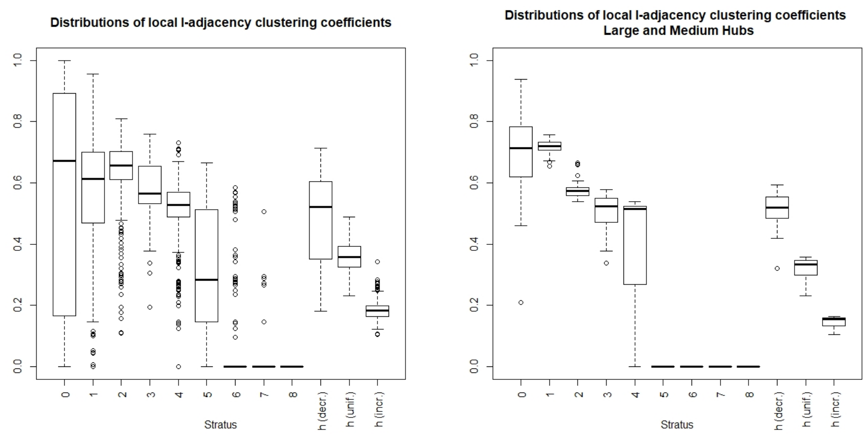

Figure 4: Distributions of the elements of the vector of the l-adjacency clustering coefficients $\mathbf{c}^{all}(l)$ computed at different levels l. Furthermore, the distributions of the elements of $\mathbf{h}^{all}$ are reported for three different choices of weights (decreasing, uniform and increasing respectively). On the right side, the same distributions are computed considering only large and medium hubs.

coefficient not lower than 0.5. The different behavior of the Anchorage airport can be easily justified by the specific characteristics of this hub. The airport is indeed connected to strategic hubs and to some other remote airports in U.S. as well. Among larger hubs, the highest rankings are instead observed for Ontario International (CA) and Southwest Florida International (FL). Both these airports are characterized by a low proportion of direct connections, but they are on average connected to airports that are connected to each other.

Different patterns between larger and smaller hubs – when the classical clustering coefficient is considered – can be partially explained also by the number of geodesic paths moving from the nodes and with a given length. To this aim, we refer to Figure 5, which reports the percentage of the geodesic paths of a fixed length (vertical axis) versus the total strength (horizontal axis) for all the nodes of the network.

On upper left-side, Figure 5 depicts the proportion of geodesics of length 1 for each airport. Large and medium hubs are on average directly connected to 15% and 11% of total airports, while smaller hubs are directly connected to 1%

of the total nodes.

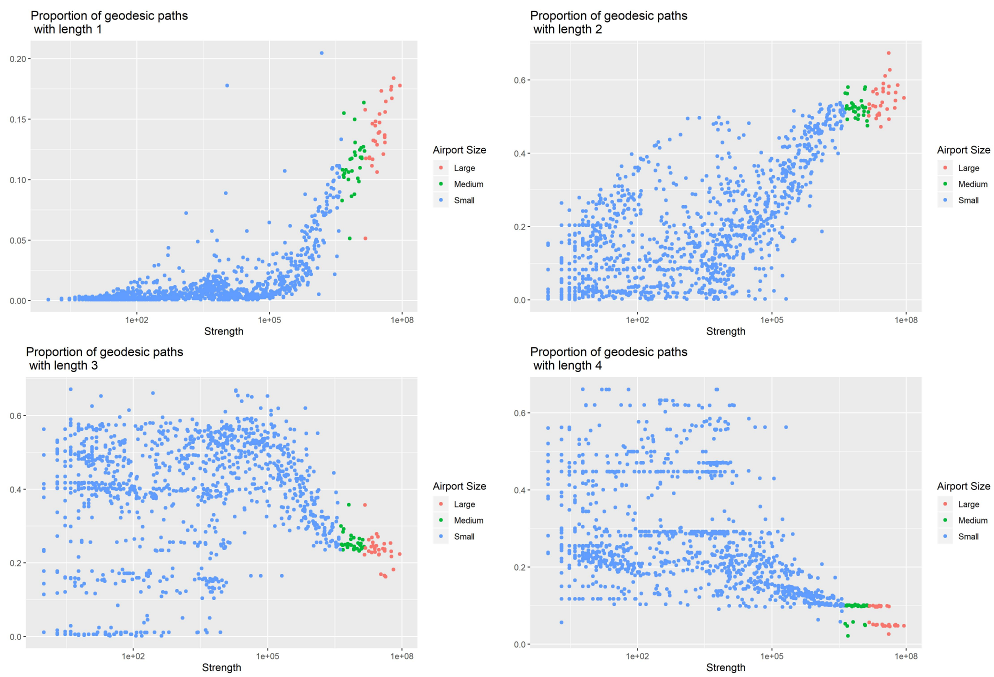

Figure 5: Proportion of geodesic paths with different lengths $(1, 2, 3, 4$, respectively) vs total strength of each airport. Large and medium hubs are reported in red and green respectively. Each plot is a linear-log plot, where x-axis is on a logarithmic scale and y-axis is on a linear scale.

Moving to the analysis of the local l-adjacency clustering coefficient for node i when $l = 1$ of the type $c_i^{all}(1)$, such a coefficient seizes possible connections of the $i - th$ airport with highly cohesive areas, reachable from it with one stopover. In this case, we observe a slight reduction of average clustering and a general decrease of the variability between different airports. Large and medium hubs have instead a different behavior, showing an average increase of the 1-adjacency clustering coefficients (mean and median move respectively from 0.69 and 0.71 to 0.72 and 0.73 respectively). It is worth mentioning the considerably low volatility of the distribution of the components of $\mathbf{c}^{all}(1)$ for these hubs. The elements of the $\mathbf{c}^{all}(1)$ range indeed in the interval $(0.65 - 0.76)$.

Results show that larger hubs are highly cohesive but – according to the local

l-adjacency clustering coefficient with $l = 1$ – they are also directly connected to strongly cohesive nodes; this confirms their strategic role in the airport system.

Focusing on higher levels in terms of distances (see Figure 5, upper right side), we detect a significant proportion of geodesics of length 2 in the network. On average, large and medium hubs are respectively connected to 55% and 52% of total airports via geodesic paths of length 2. Through these paths they reach strongly cohesive nodes as well as non-primary hubs characterized by a low clustering coefficient. Hence, $c_i^{all}(2)$ is lower than $c_i^{all}(1)$ for all airports of label i belonging to this group, with reductions that vary between 2% and 38%. Smaller hubs reach instead 20% of the nodes in two steps. Typically, they are connected to highly cohesive areas showing a $c_i^{all}(2)$ higher than $c_i^{all}(1)$ and $c_i^{all}(0)$.

As regard to higher strati, we observe in Figure 5 a proportion of geodesics of lengths 3 and 4, equal to 40% and 27% for small hubs, respectively. Large and medium hubs reach instead 25% and 8% of airports in 3 and 4 steps, respectively. As a consequence, the local l-adjacency clustering coefficient is slowly decreasing with respect to l for small hubs, while an higher reduction is observed for larger hubs. It is worth noting the high volatility of the components of $\mathbf{c}^{all}(4)$ for the latter category. In particular, roughly an half of relevant hubs has a value of such clustering coefficient higher than 0.5. Seattle-Tacoma International (WA), Ted Stevens Anchorage International (AK), Daniel K. Inouye International (HI) and Kahului (HI) airports show instead very low local 4-adjacency clustering coefficients (≤ 0.2), mainly justified by the fact that geodesics of length 4 usually connect remote airports with weak cohesive structure at stratus 0.

In line with the case $l = 4$, one can notice that small airports are connected on average to the 11% of total airports by geodesics of length 5. However, a very high volatility is observed in this class of airports. Some specific non-primary airports are able to reach more than a half of the airports through geodesics of length 5. These patterns justify the significant volatility and a not negligible average of the elements of $\mathbf{c}^{all}(5)$. Larger hubs have instead very few connections at this stratus ($< 1\%$) leading to a clustering coefficient close to zero. This argument is confirmed and furtherly stressed for strati greater than 5. Only few nonprimary hubs i are connected to some other nodes through geodesic paths with length larger than 5, hence showing values of $c_i^{all}(l)$ greater than zero for $l > 5$. Indeed, typically, these connections regard relations between very remote and without strongly cohesive airports. For instance, Blakely Island (WA) and Tatitlek (AK) airports are connected by a geodesic path of length 8[4].

The values of the elements of $\mathbf{h}^{all}$ in Figure 4 synthesize the overall cohesiveness structure of each node, thus providing a measure of the relevance of the nodes in the network. The choice of weights x_l can modulate the intensity of the elements of $\mathbf{c}(l)$ in contributing to the overall stratified structure, giving to this indicator a high degree of flexibility.

Here, we consider the three possible scenarios already used in Section 3.3 (i.e. decreasing, uniform and increasing weights). For instance, assuming that weights x's are decreasing, we are reducing the impact of the elements of local l-adjacency clustering coefficients $\mathbf{c}(l)$ with respect to the whole system when the distance l increases. In particular, by concentrating the mass of weights over the small values of l, we take into major consideration the cohesiveness structures close to the nodes of the graph. In this case, the average of the components of $\mathbf{h}^{all}$ is equal to 0.48, and it is higher for large and medium hubs (equal to 0.54). Therefore, large and medium airports are confirmed to be strategic hubs in the network. Indeed, on one hand, these airports are involved in highly cohesive nodes at low strati; on the other hand, they are directed connected to highly cohesive areas.

Differently, the cases of either uniform weights or concentration of the x's over large values of l emphasize the relevant role of peripheral nodes of airports which are strongly cohesive. Since these scenarios are more sensitive to cohesive-

[4]The geodesic path is given by the following sequence of edges: Blakely Island – Friday – Kenmore Air Arbor – Roche Harbor Country – Seattle Tacoma International airport – Ted Stevens Anchorage International Airport Country – Beluga airport – Merrill Field Anchorage Airport – Tatitlek airport

ness far from the nodes, we observe a reduction of the average of the components of $\mathbf{h}^{all}$, equal to 0.36 (uniform weights) and 0.18 (increasing weights), and higher values for smaller hubs.

We focus now on computing the in and out local l-adjacency clustering coefficients by means of a separate evaluation of in- and out-paths. As stressed before, the airport network is highly symmetric so that it is usually analysed as an undirected one (see [8])). In our case, we observe a strong positive correlation (close to 1) between in- and out-degree (and between in and out-strength). To assess the symmetry of the network, Fagiolo ([44]) proposes a specific measure S. If the value of S is close to zero, then an empirically-observed network is sufficiently symmetric to justify an undirected network analysis. In our case, we obtain 0.02. This index becomes 0.19 when weights are removed. Hence, network is weakly asymmetric with a more pronounced behaviour when weights are not considered. However, although direct connections are highly correlated, some differences could be observed when we focus on long geodesics.

As a consequence, at stratus $l = 0$, in- and out- local l-adjacency clustering coefficients are very similar (see Figure 6) and lower than c_0^{all}. With the directed (in and out) local l-adjacency clustering coefficients, we are focusing on specific patterns. Indeed we neglect some types of triangles (like cycles and middle-man triangles, according to the classification provided in [9]) and consider only directed paths.

On average we observe slightly higher out-clustering coefficients for larger hubs and lower ones for smaller airports. However, there is not an univocal pattern among the airports, although in many cases differences are negligible. In the class of medium hubs, an interesting case is the node i associated with the Luiz Munoz Marin International Airport in San Juan (Puerto Rico), characterized by a $c_i^{out}(0)$ equal to 0.81 against a $c_i^{in}(0)$ equal to 0.77. In other words, this airport is more involved in weighted triangles of out-type than in-triangles. This evidence is partially justified by a number of passengers departure higher than arrivals, probably motivated by higher movements towards U.S. than vice versa.

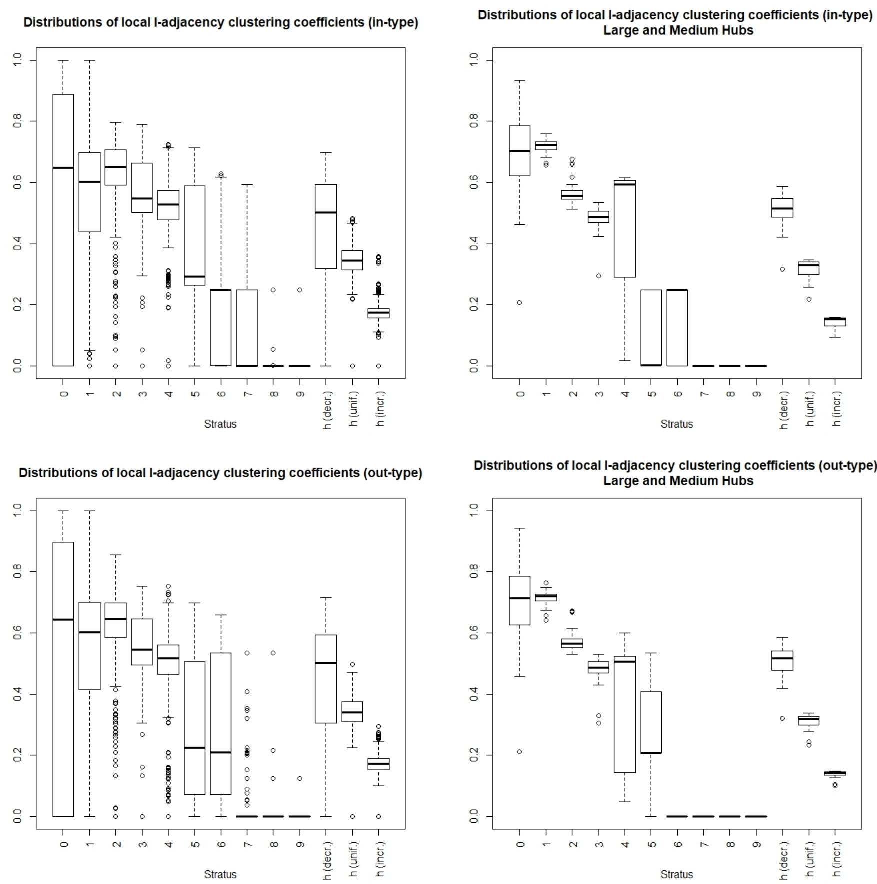

Figure 6: Upper left figure displays the distributions of local l-adjacency clustering coefficients of in-type in $\mathbf{c}^{in}(l)$ and distributions of the elements of $\mathbf{h}^{in}$ for the considered different choices of weights (decreasing, uniform and increasing respectively). On the right side, the same distributions are computed by considering only large and medium hubs. In the bottom part, we report the distributions of the components of $\mathbf{c}^{out}(l)$ and $\mathbf{h}^{out}$ for all the airports (left side) and only for large and medium hubs (right side).

The directed l-adjacency clustering coefficients display similar distributions and, on average, lower values than the elements of $\mathbf{c}^{all}(l)$ until l assumes values equal to 4. Remarkable differences are instead observed for higher strati. In particular, stronger cohesiveness of in-type are observed in peripheral nodes. Since stratus 5, we observe indeed higher values of the components of $\mathbf{c}^{in}(l)$

than of the ones of $\mathbf{c}^{out}(l)$.

The analysis of the synthetic indicators given by the components of $\mathbf{h}^{in}$ and $\mathbf{h}^{out}$ – and when decreasing weights are considered – confirms the relevance of large and medium hubs in terms of cohesiveness structure (both in and out)[5]. Instead, when we base our analysis on increasing weights, as expected the role of peripheral nodes is emphasized. In this case, a low correlation (see Figure 7) is observed depending on the specific behavior of each airport. Furthermore, different patterns of long directed paths are also caught by the synthetic measure although the network is only weakly asymmetric in terms of adjacency matrix. We have indeed a slight prevalence of cohesiveness structures of in-type.

In this application, the local l-adjacency clustering coefficients allow to highlight the role of an airport in the network taking into account direct interconnections as well as indirect ties. This topic assumes relevance since the position of an airport with respect to existing interconnected groups of airports is an important element for evaluating the competition between them. In particular, airports in the same catchment area could be connected directly or indirectly to cohesive structures and favour in a different way flows of passengers or commodities. This issue can also be relevant in other contexts. For instance, in disease propagation analysis, an airport characterized by a high value of l-adjacency clustering coefficient could be exposed to the propagation of a desease even if it is not direcly connected to the region in which the disease under consideration appeared for the first time.

5. Conclusions

Interconnection plays a fundamental role in the business research context. As well-known, in network theory, the level of interconnectivity in the neighbourhood of a node is typically assessed by means of the clustering coefficient

[5]This evidence is a consequence of the high correlation between in and out l-adjacency correlation coefficients computed for low values of l (at this regard, see Figure 7)

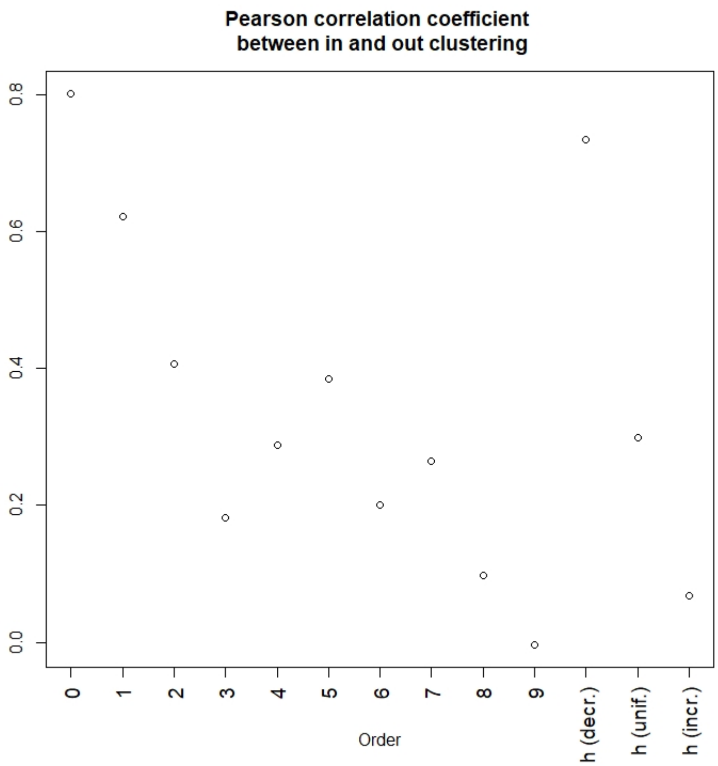

Figure 7: Pearson correlation coefficient between in and out l-adjacency correlation coefficients at different strati l.

and captures the cohesiveness structure associated with the considered node. Moving from this fact, we exploit the concept of cohesiveness structure to understand the contextualization of the nodes within the overall system. In particular, we provide a generalization of the concept of clustering coefficient in order to catch both the presence of cohesive areas around a node and/or high levels of mutual interconnections at different distances from the node itself. With respect to the classical clustering coefficient, we are able to capture in a better way the topological structure of the whole network and to map the presence of stratified structures in the network at different levels. Furthermore, we also define a synthetic indicator for each node in order to simultaneously consider all the coefficients. Being this indicator dependent on a set of weights of the strati, we allow for a degree of flexibility in order to modulate the effects of both adjacent nodes and peripheral nodes.

An empirical application to U.S. domestic air traffic network is developed where the interconnectedness and the effect on connectivity of an airport may play a relevant role in strategic business decisions. Results show the effectiveness of these measures in catching the peculiar characteristics of different nodes in the airport network. In particular, focusing on large and medium hubs, we are able to emphasize their strategic role in the airport system. We observe, indeed, that larger hubs are not only highly cohesive but also at the centre of strongly cohesive nodes. When different cohesive strati are analysed, the effects of indirect connections with remote airports are emphasized. Finally, although the network is only weakly asymmetric and it is typically analysed as an undirected network in the literature, we show that a separate evaluation of directed paths at high levels can be useful to identify specific patterns in terms of in and out-cohesive groups of nodes.

Acknowledgements We would like to thank the Editor, the Guest Editors of the Special Issue "Complexity: challenges and opportunities" and the anonymous Reviewers for their careful reviews on an earlier version of this paper.

References

[1] T. Ritter, H. G. Gemünden, Interorganizational relationships and networks: An overview, Journal of Business Research 56 (2003) 691 – 697. URL: http://www.sciencedirect.com/science/article/pii/S0148296301002545. doi:https://doi.org/10.1016/S0148-2963(01)00254-5.

[2] M. Cinelli, G. Ferraro, A. Iovanella, Evaluating relevance and redundancy to quantify how binary node metadata interplay with the network structure, Scientific Reports 9 (2019) 11404.

[3] H. Håkansson, D. Ford, How should companies interact in business networks?, Journal of Business Research 55 (2002) 133–139.

[4] I. Wilkinson, L. Young, On cooperating: firms, relations and networks, Journal of Business Research 55 (2002) 123–132.

[5] M. E. Zaglia, Brand communities embedded in social networks, Journal of Business Research 66 (2013) 216–223.

[6] M. Cinelli, G. Ferraro, A. Iovanella, Rich-club ordering and the dyadic effect: Two interrelated phenomena, Physica A: Statistical Mechanics and its Applications 490 (2018) 808–818.

[7] M. Ausloos, R. Lambiotte, Clusters or networks of economies? A macroeconomy study through gross domestic product, Physica A: Statistical Mechanics and its applications 382 (2007) 16–21.

[8] A. Barrat, M. Barthélemy, R. Pastor-Satorras, A. Vespignani, The architecture of complex weighted networks, Proceedings of the National Academy of Sciences 101 (2004) 3747–3752.

[9] G. Fagiolo, Clustering in complex directed networks, Physical Review E 76 (2007). doi:10.1103/physreve.76.026107.

[10] J. Onnela, J. Saramäki, J. Kertész, K. Kaski, Intensity and coherence of motifs in weighted complex networks, Physical Review E 71 (2005).

[11] D. J. Watts, S. H. Strogatz, Collective dynamics of small-world networks, Nature 393 (1998) 440–442. doi:10.1038/30918.

[12] S. Wasserman, K. Faust, Social Network Analysis: Methods and Applications., Cambridge University Press, New York, NY., 1994. arXiv:1607.00509.

[13] G. P. Clemente, R. Grassi, Directed clustering in weighted networks: A new perspective, Chaos, Solitons & Fractals 107 (2018) 26–38.

[14] G. Rotundo, M. Ausloos, Organization of networks with tagged nodes and biased links: a priori distinct communities, Physica A: Statistical Me-

chanics and its Applications 389 (2010) 5479–5494. doi:10.1016/j.physa.2010.07.029.

[15] R. Cerqueti, G. Ferraro, A. Iovanella, A new measure for community structure through indirect social connections, Expert Systems with Applications 114 (2018) 196–209.

[16] J. Moody, D. R. White, Structural cohesion and embeddedness: A hierarchical concept of social groups, American sociological review (2003) 103–127.

[17] M. Granovetter, Economic action and social structure: The problem of embeddedness, American journal of sociology 91 (1985) 481–510.

[18] E. Estrada, The structure of complex networks: theory and applications, Oxford University Press, 2011.

[19] J. Moody, J. Coleman, Clustering and cohesion in networks: Concepts and measures, in: International Encyclopedia of the Social & Behavioral Sciences, Second ed., Elsevier, 2015, pp. 906–912.

[20] A. Fernandez, S. Gomez, Solving non-uniqueness in agglomerative hierarchical clustering using multidendrograms, Journal of Classification 25 (2008) 43–65.

[21] P. Langfelder, B. Zhang, S. Horvath, Defining clusters from a hierarchical cluster tree: the dynamic tree cut package for r, Bioinformatics 24 (2007) 719–720.

[22] N. Yuruk, M. Mete, X. Xu, T. Schweiger, Ahscan: Agglomerative hierarchical structural clustering algorithm for networks, in: Social Network Analysis and Mining, 2009, pp. 72–77.

[23] A.-L. Barabási, E. Ravasz, ,hierarchical organization in complexnetworks, physical review e67(2003), no. 2, 026112, Physical Review E 67 (2003).

[24] L. Chunlin, B. Jingpan, W. Zhao, X. Yang, Community detection using hierarchical clustering based on edge-weighted similarity in cloud environment, Information Processing & Management 56 (2019) 91–109.

[25] D. Edler, L. Bohlin, M. Rosvall, Mapping higher-order network flows in memory and multilayer networks with Infomap, Algorithms (2017).

[26] H. Lu, M. Halappanavar, A. Kalyanaraman, Parallel heuristics for scalable community detection, Parallel Computing 47 (2015) 19–37.

[27] M. Rosvall, J. Delvenne, M. Schaub, R. Lambiotte, Advances in Network Clustering and Blockmodeling, 2019.

[28] G. Ferraro, A. Iovanella, Organizing collaboration in inter-organizational innovation networks, from orchestration to choreography, International Journal of Engineering Business Management 7 (2015) 7–24.

[29] H. Yin, A. R. Benson, J. Leskovec, Higher-order clustering in networks, Physical Review E 97 (2018).

[30] H. Yin, A. R. Benson, J. Leskovec, D. F. Gleich, Local higher-order graph clustering, in: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '17, ACM, New York, NY, USA, 2017, pp. 555–564. doi:10.1145/3097983.3098069.

[31] A. Fronczak, J. A. Hołyst, M. Jedynak, J. Sienkiewicz, Higher order clustering coefficients in barabási–albert networks, Physica A: Statistical Mechanics and its Applications 316 (2002) 688–694. URL: http://dx.doi.org/10.1016/S0378-4371(02)01336-5. doi:10.1016/s0378-4371(02)01336-5.

[32] P. Jensen, M. Morini, T. Venturini, M. Jacomy, J.-P. Cointet, P. Mercklé, M. Karsai, E. Fleury, Bridgeness: a novel centrality measure to detect global bridges, in: ECCS 2014–European Conference on Complex Systems, 2014.

[33] A. Wu, L. Tian, Y. Liu, Bridges in complex networks, Physical Review E 97 (2018).

[34] E. Jimenez, J. Claro, J. P. de Sousa, The airport business in a competitive environment, Procedia-Social and Behavioral Sciences 111 (2014) 947–954.

[35] M. Zanin, F. Lillo, Modelling the air transport with complex networks: A short review, The European Physical Journal Special Topics 215 (2013) 5–21.

[36] G. Burghouwt, R. Redondi, Connectivity in air transport networks: an assessment of models and applications, Journal of Transport Economics and Policy (JTEP) 47 (2013) 35–53.

[37] OECD, Defining, Measuring and Improving Air Connectivity, Technical Report, OECD Report, International Transport Forum, 2018.

[38] V. Colizza, R. Pastor-Satorras, A. Vespignani, Reaction-diffusion processes and metapopulation models in heterogeneous networks., Nature Physics 3 (2007) 276–282. doi:10.1038/nphys560.

[39] T. Jia, B. Jiang, Building and analyzing the US airport network based on en-route location information, Physica A: Statistical Mechanics and its Applications 391 (2012) 4031–4042. doi:10.1016/j.physa.2012.03.006.

[40] T. Jia, K. Qin, J. Shan, An exploratory analysis on the evolution of the US airport network, Physica A: Statistical Mechanics and its Applications 413 (2014) 266–279.

[41] T. Opsahl, P. Panzarasa, Clustering in weighted networks, Social Network 231 (2009) 155–163. doi:10.1016/j.physa.2016.05.063.

[42] D. Tłoczyński, Connectivity in air transport networks: an assessment of the polish aviation market, Transport Economics and Logistics 70 (2017) 53–62.

[43] Federal Aviation Administration, Passenger Boarding (Enplanement) and All-Cargo Data for U.S. Airports, Technical Report, Federal Aviation Administration, 2017.

[44] G. Fagiolo, Directed or Undirected? A new index to check for directionality of relations in socio-economic networks, Economics Bulletin 3 (2006) 1–12. arXiv:physics/0612017.

Appendix A. List of Airports

Rank	City	Airport Name
1	Atlanta	Hartsfield - Jackson Atlanta International
2	Los Angeles	Los Angeles International
3	Chicago	Chicago O'Hare International
4	Fort Worth	Dallas-Fort Worth International
5	Denver	Denver International
6	New York	John F Kennedy International
7	San Francisco	San Francisco International
8	Las Vegas	McCarran International
9	Seattle	Seattle-Tacoma International
10	Charlotte	Charlotte/Douglas International
11	Newark	Newark Liberty International
12	Orlando	Orlando International
13	Phoenix	Phoenix Sky Harbor International
14	Miami	Miami International
15	Houston	George Bush Intercontinental/Houston
16	Boston	General Edward Lawrence Logan International
17	Minneapolis	Minneapolis-St. Paul International/Wold-Chamberlain
18	Detroit	Detroit Metropolitan Wayne County
19	Fort Lauderdale	Fort Lauderdale/Hollywood International
20	New York	Laguardia
21	Philadelphia	Philadelphia International
22	Glen Burnie	Baltimore/Washington International Thurgood Marshall
23	Salt Lake City	Salt Lake City International
24	Arlington	Ronald Reagan Washington National
25	San Diego	San Diego International
26	Dulles	Washington Dulles International
27	Chicago	Chicago Midway International
28	Honolulu	Daniel K. Inouye International
29	Tampa	Tampa International
30	Portland	Portland International

Table A.3: Large Primary Hubs according to FAA classification (based on enplanements in 2017)

Rank	City	Airport Name
31	Dallas	Dallas Love Field
32	St. Louis	St Louis Lambert International
33	Nashville	Nashville International
34	Austin	Austin-Bergstrom International
35	Houston	William P. Hobby
36	Oakland	Metropolitan Oakland International
37	San Jose	Norman Y. Mineta San Jose International
38	Metairie	Louis Armstrong New Orleans International
39	Raleigh	Raleigh-Durham International
40	Kansas City	Kansas City International
41	Sacramento	Sacramento International
42	Santa Ana	John Wayne Airport-Orange County
43	Cleveland	Cleveland-Hopkins International
44	San Antonio	San Antonio International
45	Fort Myers	Southwest Florida International
46	Indianapolis	Indianapolis International
47	Pittsburgh	Pittsburgh International
48	San Juan	Luis Munoz Marin International
49	Greater Cincinnati	Cincinnati/Northern Kentucky International
50	Columbus	John Glenn Columbus International
51	Kahului	Kahului
52	Milwaukee	General Mitchell International
53	Windsor Locks	Bradley International
54	West Palm Beach	Palm Beach International
55	Jacksonville	Jacksonville International
56	Anchorage	Ted Stevens Anchorage International
57	Albuquerque	Albuquerque International Sunport
58	Burbank	Bob Hope
59	Buffalo	Buffalo Niagara International
60	Ontario	Ontario International
61	Omaha	Eppley Airfield

Table A.4: Medium Primary Hubs according to FAA classification (based on enplanements in 2017)