Check for
updates

# Dimension Reduction in the Topology of Multilayer Spatial Networks: The Case of the Interregional Commuting in Greece

**Dimitrios Tsiotas[1]** · **Vassilis Tselios[2]**

## Abstract

The daily commuting out of the city of residence for labor purposes is a complex phenomenon driven by various geographical, transportation, economic, and social forces. Due to its high level of complexity, current modeling approaches are mainly disciplinary and thus lack a multifaceted approach incorporating diverse conceptualization of this phenomenon. To contribute to this demand, this paper develops a methodological framework integrating aspects of commuting of different contexts, functionality, and level of geographical resolution. The proposed framework builds on a complex network and principal component analysis first to develop a multilayer graph of interregional commuting, next to configure variables representing aspects of network topology, and finally to decompose the total model into principal components expressing uncorrelated aspects in the overall conceptualization of interregional commuting. The proposed framework conceptualizes spatial distance as a major force of network topology and develops a quantitative framework evaluating the consistency of network topology across layers of different geographical resolutions and functionality. The overall approach provides insights into the structure and functionality of interregional commuting and promotes the integrated macroeconomic approach in regional research.

**Keywords** Transport geography · Spatial networks · Complex network analysis · Principal component analysis

✉ Dimitrios Tsiotas
tsiotas@aua.gr

Vassilis Tselios
v.tselios@panteion.gr

[1] Department of Regional and Economic Development, Agricultural University of Athens, Nea Poli, 33100 Amfissa, Greece

[2] Department of Economic and Regional Development, Panteion University of Social and Political Sciences, 17671 Athens, Greece

# 1 Introduction

Geographical space, transportation, and regional development lie under a symbiotic relation (Tsiotas and Ducruet 2021): from one side, geographical space imposes friction (cost) to movements between places, affecting the structure and functionality of transportation networks and their relevant ability to serve spatial economic development (Barthelemy 2011; Rodrigue et al. 2013). From another side, regional development contributes to the increase (or change) of demand for geographical space and therefore supports the further evolvement of spatial and transportation networks (Rodrigue et al. 2013). In the light of this interdependence, the spatial property is by default determinative for the components of this symbiotic scheme, affecting (a) the shape, form, and usage (urban, rural, regional, etc.) of geographical space; (b) the topology and functionality of transportation networks (Ducruet and Beauguitte 2014); and (c) the centripetal or centrifugal developmental dynamics of market places (Brakman et al. 2005), in the context of the new economic geography. Spatial constraints are immanent at all levels of network aggregation (neighborhood, local, regional, national, and global), assigning distance costs in the development of connections (Barthelemy 2011; Rodrigue et al. 2013). Moreover, spatiality suggests a major force in the configuration of the regional problem (Capello 2016), which regards the asymmetric development observed either between different geographical places, at a fixed time, or at different times for the same place, or jointly. Although spatial constraints are (either directly or indirectly) evident in all aspects of socioeconomic activity, their modeling and study is a complex task submitted to high level of relativity. Such relativity can be witnessed by (a) the uneven spatial distribution of regional variables and indicators observed at various geographical scales (Garretsen et al 2013); (b) the differences in the topology of transportation networks observed due to node aggregation (Tsiotas and Polyzos 2018) or geographical scale (Tsiotas and Ducruet 2021); (c) the emergence of top-down and bottom-up theories (Crescenzi and Rodriguez-Pose 2011) of regional economics and policy; (d) the need of applying diversified regional policies at different levels of geographical agglomeration (Brakman et al. 2005) and market integration (Ottaviano 2003); and more.

Recently, network science (Barabasi 2013; Brandes et al. 2013), a discipline studying communication systems with the use of network paradigm, contributed to geographical and regional research by describing the structure of spatial networks beyond their geometry (Barthelemy 2011; Ducruet and Beauguitte 2014; Tsiotas and Ducruet et al. 2021), allowing thus incorporating topological variables in the modeling of spatial systems. This modern discipline has been already proven fertile in the modeling of spatial networks (Barthelemy 2011) and has promoted relevant research in transportation (road, rail, maritime, air transport) and other infrastructure networks. In the context of network paradigm, the regional science, economic geography, and relevant disciplines can deeper conceptualize topological aspects of interconnected systems configuring spatial hierarchies (Ducruet et al. 2011; Ducruet and Beauguitte 2014), and have more options available for the multilayer modeling of spatial and regional economic systems.

However, although the coupling of network and regional science is very promising, Marshall et al. (2018) note that network topology is yet a new and undiscovered concept in the context of regional and geographical sciences, and there is still a way to go to integrate topological analysis into current spatial analysis protocols. For instance, Tsiotas and Ducruet (2021) observe that decomposition techniques in network analysis are not yet applied in a comprehensive context to unravel the relationship between space and network topology.

Such reasoning motivates this paper to question whether there are also norms or common features ruling the effect of spatial constraints on the configuration of network topology at different degrees of spatial or administrative aggregation, contrary to the undoubted variability describing spatial inequalities at different levels of geographical scale. This research question is inspired by recent approaches (Tsiotas and Polyzos 2018; Tsiotas and Ducruet 2021) examining (one major finding in economics) the "puzzling effect of distance" on spatial economic interaction using the network paradigm. However, this study goes one step further by examining this research question in the context of land (instead of maritime) transportation. Towards answering to this research question, this paper models the interregional commuting system in Greece (GCN) into a multilayer spatial network (Barthelemy 2011; Kivela et al. 2014; Boccaletti et al. 2014) of national geographical scale. It further studies its multilayer topology both (a) within and (b) between its layers, which are configured at different levels of geographical and administrative resolution (at the national, regional, and capital city level) and different types of structure and functionality (geometric, accessibility, and commuting flow). The study builds on methods of complex network analysis (Barthelemy 2011; Barabasi 2013; Kivela et al. 2014; Boccaletti et al. 2014) and dimension reduction (Norusis 2008; Walpole et al. 2012) to examine the topological properties of GCN across its layers and extract the major topological components of this multilayer model. This is done by applying a principal component analysis (PCA) on a collection of topological variables from all layers. This dimension reduction method is chosen because (Wold et al. 1987; Norusis 2008): (a) it properly serves the research question of detecting common topological features amongst different layers of the GCN; (b) is popular, well-established, and empirically tested; and (c) has been already proven useful in multidisciplinary applications.

The overall approach in this paper conceptualizes the utility of commuting flows and the cost of spatial distance as major determinants of network topology and performs an interlayer analysis to detect the important topological features across the network layers, which represent different aspects of utility and cost of the interregional commuting market. In the context of a spatial-economic interpretation, the GCN is an integrated utility-cost model with multifaceted structural reference and economic functionality. In this light, the construction and study of a multilayer model of interregional commuting can promote scientific research and policy practice. For instance, the modeling approach proposed by this study can provide insights into spatial and transportation planning, and sustainable development, where it is quite assisting for the planners to conduct strategic national plans based on integrated models incorporating utility-cost (Kulmer et al. 2014) and spatial-economic information (Clinch and O'Neil 2009; Vigar 2009; Kulmer et al. 2014). As a market of immanent (and periodically alternating) supply and demand forces, the interregional commuting

network can provide insights into the detection of dipole or polycentric structures in urban and regional systems (Li et al. 2018; Tsiotas et al. 2021). Such structures illustrate either corporate or competitive relations in the geographical space and thus can point out axes and directions toward spatial development (Davoudi 2003), applicable at all geographical scales (Tsiotas et al. 2021). Also, in the model, the availability of layers at different levels of geographical and administrative resolution, can enlighten into a better comprehension of how spatial distance and spatial aggregation (Wegener 2001; Jeon et al. 2012; Tsiotas and Polyzos 2018) affects the structure and functionality of transportation systems. This asset can facilitate more efficient strategic planning based on more accurate and better targeted geographical and administrative levels (Albrechts 2006; Cavaco and Costa 2020). Generally, this paper promotes regional and transportation research by providing an empirical case study and methodological approach examining the interrelation between geographical resolution and spatial hierarchy in transportation networks, which is a major concern in spatial and transportation planning (Rodrigue et al. 2013; Capello 2016). Finally, provided that relevant empirical research mainly builds on comparisons between network layers, this study advances computational network science by configuring a methodological framework for dimension reduction in multilayer networks to detect principal components of network topology and to examine topological consistency across networks.

The remainder of this paper is organized as follows; Sect. 2 provides a literature review. Section 3 presents the methodological and conceptual framework of the study. Section 4 shows the results of the analysis and discusses them within the context of regional and geographical sciences, and finally, in Sect. 4 conclusions are given.

## 2 Literature Review

Commuting is the phenomenon of daily mobility for labor purposes outside the place of residence and has a multidisciplinary framework consisting of geographical, social, economic, technological, and political aspects (Green and Meyer 1997; Evans et al. 2002; Van Ommeren and Rietveld 2005; Kanaroglou et al. 2015). This phenomenon is an act of spatial and socioeconomic interaction between neighbor places and regions and is of great importance for regional and urban research because it involves spatial-socioeconomic structures, at different geographical scales, operating over urban and regional development (McArthur et al. 2011; Drobne et al. 2012; Horak et al. 2014; Rodrigue et al. 2013; Polyzos 2019). Due to its complexity, commuting research has been fruitful and broad, covering many topics, such as transportation-cost assessment (Hamilton 1989; Stutzer and Frey 2008; Van Ommeren and Fosgerau 2009); assessment of travel duration (Hamilton and Roell 1982); psychology of mobility (Koslowsky et al. 1995); traffic and accident analysis (Ozbay et al. 2007); selection of transportation modes and alternative routing (Murphy 2009; Liu and Nie 2011; Bwire and Zengo 2020); productivity (Van Ommeren and Rietveld 2005); land use (Tan et al. 2019; Zhao et al. 2020); and commuting behavior (Ma and Ye 2019); among others (e.g. see Kanaroglou et al. 2015). On the one hand, such approaches promote

polyphony and provide new insights into commuting research. On the other hand, they restrictively conceptualize commuting within the disciplinary framework of certain research fields, such as urban and regional planning, geography, transportation analysis, environmental assessment, and even econophysics (Barthelemy 2011; De Montis et al. 2011; Ren et al. 2014; Marshall et al. 2018). This context of polyphony raises the demand for integration amongst the disciplinary approaches of commuting, especially in the current era of "big data" facilitating the availability, management, and analysis of big datasets. Toward this demand, the common context that can be found in almost every aspect of commuting research is a reference to the spatial impedance of transportation (Rodrigue et al. 2013). In particular, against the demand derived by the attractiveness of places to develop movements of socioeconomic interest, space imposes constraints that are translated to transportation costs (Rodrigue et al. 2013). Transportation costs are immanent in every aspect of commuting research, either directly (Hamilton 1989; Stutzer and Frey 2008; Van Ommeren and Fosgerau 2009), in terms of distance, space–time and trip duration, fuel and ticket cost; or indirectly, affecting the choice of transportation mode (Murphy 2009; Liu and Nie 2011; Bwire and Zengo 2020), the routing, the travel behavior (Koslowsky et al. 1995; Ozbay et al. 2007), the infrastructure quality, and land use (Tan et al. 2019; Zhao et al. 2020).

Network science (Barabasi 2013; Brandes et al. 2013), a modern discipline that emerged from the multidisciplinary study of connectedness, uses the network paradigm to study communication systems (Easley and Kleinberg 2010; Newman 2010; De Montis et al. 2011); and to model network structures to graphs consisting of nodes (interconnected units) and edges (their connections). Within the context of network science, geographical systems of socioeconomic interaction can be modeled to spatial networks, which are graphs embedded in the geographical space (Barthelemy 2011), capable of incorporating topological, geometric, environmental, and socioeconomic information, in a single model. Although is a modern discipline, network science has been proven very effective in the modeling of spatial networks (Barthelemy 2011) and has already provided interesting insights into the structure and functionality of various geographical systems of economic interaction, such as road (Porta et al. 2006a, b; Marshall et al. 2018); rail (Barthelemy 2011); maritime (Ducruet 2013; Tsiotas and Polyzos 2015a); air transport (Cardillo et al. 2013; Tsiotas and Polyzos 2015b); and other infrastructure (Debrie 2010; Barthelemy 2011); networks. Of course, a major part of such effectiveness should be attributed to graph theory (Diestel 2005), which is an established (for almost three centuries) field of discrete mathematics, where quantitative network analysis builds on; and particularly to the past work of Kansky (1963), who succeeded a coupling of graph theoretic and probabilistic approach with transport geography (Rozenblat and Melancon 2013). Another portion of such effectiveness should be attributed to the so-called "gravitation models" (Wilson 1967), which contributed to a discrete conceptualization of geographical systems, expressed as population nodes, facilitating their modeling into graphs (Rozenblat and Melancon 2013).

Empirical research in spatial networks (Barthelemy 2011; Ducruet 2013; Rodrigue et al. 2013) has shown that (a) spatial constraints are reflected on the shape of the degree distribution, usually configuring a bell-shaped (expressing that the major load of connectivity is undertaken by the majority of nodes) than a power-law pattern

(expressing that the major load of connectivity is undertaken by a few nodes); (b) the transport mode affects network topology (e.g. air transport network are usually scale-free networks, whereas road or railway networks are not); (c) spatial and geographical constraints are more intense in the topology of land transportation networks than of maritime networks, (d) distant communication is mainly undertaken by the network hubs; (e) gateways between network layers (defined by different transport modes) are usually related to hubs within layers; and much more. In the relevant literature, spatial networks of multifaceted structure and functionality are effectively modeled into multilayer network models (Barthelemy 2011; De Domenico et al. 2013; Kivela et al. 2014; Boccaletti et al. 2014), which are collections of various graphs considered as a single model composed by different layers. In multilayer networks, their structure and functionality is applicable toward a double direction; within network layers, where each one is considered as a separate graph of certain topological properties; and between layers, where interlayer links are responsible for the functionality of the whole system are modeled. Indicative works in multilayer networks favoring the study of transportation systems are of De Domenico et al. (2013), which configured a prime framework towards this approach, and of Boccaletti et al. (2014) and Kivela et al. (2014), which are milestones for their formalism and detail. Relevant empirical research on multilayer transportation networks mainly focuses on the multimodal configuration of the network layers (Gallotti and Barthelemy 2014; Alessandretti et al. 2016). For instance, Gallotti and Barthelemy (2014) performed a path analysis on the multilayer, temporal network of the national, multimodal, British public transportation system, and proposed a method of statistical decomposition of trips in urban areas, in terms of riding, waiting, and walking times. The analysis illustrated the dependence of the temporal structure of trips to distance and allowed comparisons between different cities, highlighting the need for better optimization strategies adapted to short, long uni-modal, or multimodal trips. The same authors (Gallotti and Barthelemy 2015), modeled the national UK public transport system as a multilayer network, which consisted of multi-modal layers of an airport, ferry dock, rail, metro, coach, and bus station transportation, and used a coarse-graining procedure to define coupling between different transport modes. Aleta et al. (2016) modeled the multimodal transportation systems of 9 different cities in Europe, ranging from small towns to mega-cities like London and Berlin, to study their interconnected structure for highlighting their vulnerabilities and possible improvements. Their approach allowed creating a simple yet realistic model for urban mobility, able to reproduce real-world facts and to test for network improvements. Alessandretti et al. (2016) developed a multilayer representation of public transportation systems, which was built on different representations, multiple lines, schedule variability, and diversity of transfers. The analysis was implemented to the public transportation systems of several French municipal areas and revealed hidden patterns of privileged connections along with their efficiency compared to commuting flows.

Another strand of studies is interested in examining how space affects network topology across layers in multilayer transportation networks (Ducruet et al. 2011; Tsiotas and Polyzos 2018; Ghavasieh and De Domenico 2020). The work of Ducruet et al. (2011) is indicative, who studied the effect of node aggregation between two different types of transportation networks, air, and sea flow, through three different levels of urban delineation; cities, regional areas, and megalopolises. The study

revealed complementarities between air and sea transport networks in the urban hierarchy, showed that aggregation grasped some important topological and geographical properties of these networks, and found through topological similarity that global players in the transport industry tend to follow the main paths of urban development. Tsiotas and Polyzos (2018) studied the topological consistency of a national (Greek) maritime network due to node grouping between layers of the same socioeconomic system. The analysis highlighted that degree is the most consistent aspect of this multilayer network by preserving both the topological and the socioeconomic information through node aggregation. Ducruet (2017) investigated the degree of overlap among different layers of circulation composing global maritime flows in the period 1977–2008. The analysis confirmed the strong and path-dependent influence of multiplexity on traffic volume, range of interaction, and centrality; and revealed that network grows and concentration around large hubs over time causes a place-dependent traffic distribution, due to the reinforced position of already established nodes. Finally, Ghavasieh and De Domenico (2020), by conceptualizing that transport systems management is related to the structural cost and acceleration of information flows, introduced a framework for functional reducibility of costly information immanent in multilayer transport systems, by coupling layers together concerning dynamics rather than structure. The authors found that the optimal configuration is submitted to maximization of the deviation of the system's entropy from the limit of free and non-interacting layers. Their approach provided a transparent procedure toward reducing diffusion time and optimizing information flow in empirical multilayer systems, without the cost of altering the underlying structure.

The previous short review illustrates that network science contributed to geographical and regional research by describing the structure of spatial networks beyond their geometry, allowing thus incorporating topological variables in the modeling of spatial systems (Barthelemy 2011; Ducruet et al. 2011; Tsiotas and Polyzos 2018). Through the network paradigm, regional science and geographical disciplines can now more in-depth conceptualize the "a-spatial" structural aspects related to systems' topological features; and through the multilayer modeling, can more sufficiently incorporate the diverse geographical, structural, socioeconomic, policy, and other functional approaches of transportation in a single model (Ducruet and Beauguitte 2014). However, Marshall et al. (2018) noted that network topology is yet considered as a new and undiscovered concept in the context of regional and geographical sciences, and there is still a way to go to integrate topological analysis into current spatial analysis protocols. Aiming at serving this demand, this paper models the interregional commuting system in Greece (GCN) to a multilayer spatial network of national geographical scale and studies its multilayer topology both within and between its layers, configured at different levels of geographical and administrative resolution (at the national, regional, and capital city level) and different types of structure and functionality (geometric, accessibility, and commuting flow). In terms of spatial and regional economics (Fujita 2005; Capello 2016), commuting is a spatial-economic phenomenon emerging by differential labor supply and demand derived at geographical places (Capello 2016; Rodrigue et al. 2013; Tsiotas et al. 2021). When a city offers better occupational opportunities than its neighbors, it configures a more attractive labor market, and consequently attracts commuting

flows of greater volume (Rodrigue et al. 2013). This perspective translates an inter-regional network of incoming commuting flows (modeled at the national geographi-cal scale) into an interconnected universal market of labor attractiveness, which rep-resents a closed economic system (economy) configured both at the microeconomic (concerning nodes and their neighborhoods) and macroeconomic (concerning the total system) level. On the other hand, a city of high outgoing commuting flows has the merit to offer better living conditions than its neighbors enjoying its commut-ers (Capello 2016). This perspective translates an interregional network of outgoing commuting flows into an interconnected universal market of housing attractiveness. Further, comparatively to other gravitational (McArthur et al. 2011; Horak et al. 2014) or productivity (Ma and Ye 2019) aspects of interregional systems, commut-ing is equipped by daily periodicity and round-trip directionality (Capello 2016), which describe immanent and not temporary trends of economic interaction, which is translated to round-trip transportation flows. Within this context, the layer of commuting flows in the GCN is a multifaceted model representing a closed eco-nomic system of dual structural configuration (microeconomic/macroeconomic) and dual economic functionality (labor/housing). Moreover, by including layers of distance-weighted graphs in the multilayer model of the GCN, we can incorporate different aspects of the spatial impedance immanent in the commuting market. Pro-vided that the spatial impedance is translated to transportation costs (Rodrigue et al. 2013; Capello 2016), the layers of different geographical resolutions in the GCN allow incorporating diverse aspects of transportation costs applicable in this multi-layer model.

According to the previous literature review, the multilayer GCN has the merit to simultaneously represent a hybrid utility-cost model, to the extent that commuting flows are an aspect of economic utility (Capello 2016), and geographical distance of cost (Rodrigue et al. 2013); in this interregional market. Moreover, this dual utility-cost configuration of CSN goes beyond concordant econometric models (expressed in math form), to the extent that is expressed in graph-theoretic terms by matrices (adjacency, connectivity, and weights) including higher-order tensor information.

## 3 Methodology and Data

Building on the network paradigm, this paper applies techniques of dimension reduction to decompose network topology to a set of principal components. To do so, the study conceives network topology in multivariable terms and not as a single property. In particular, we loosely consider in this paper as identical the concepts of topological space, where a graph is embedded to, and of what in literature is called "network topology" (Newman 2010; Barabasi 2013). To conceive this in a compre-hensive context, we should refer to the concept of mathematical topology, which is specific. According to the Set Theory (Hausdorff 1957), a pair $\langle X, \aleph \rangle$ consisting of a set $X$ and a class $\aleph$, where certain set operations (such as union and intersec-tion with the null-set and other subsets) on $X$ are well-defined, is called topologi-cal space; and, further, the class $\aleph$ is called "a topology on $X$". Within a network context, we should correspond the set $X$ to a graph model $G$, namely $X \equiv G$, which

is also a pair-set $G(V,E)$ of nodes $V$ and edges $E$. Under this correspondence, the term network topology rigorously refers to the class $\aleph_G$, which loosely describes an algebraic ability to apply well-defined set operations on $G$. In the research practice, the set-class $\aleph_G$ defining a network topology is empirically conceived as the relational arrangement that the network elements display in a non-metric layout (Tsiotas 2019), and can be either random-like, or lattice-like, or small-world, or hub-and-spoke, or other distinguishable patterns. To succeed in a better computational approximation of the network topology, many metrics, measures, and indicators have been developed in the literature (Newman 2010; Barthelemy 2011; Barabasi 2013; Boccaletti et al. 2014), capturing computational aspects of the topological properties in a network. Within this context, we define in this paper network topology as the composition of fundamental topological measures and attributes describing the topological features of a network. This approach allows studying network topology in a multivariable context, expressed by the collection of various measurable attributes and not by a single topological characterization. The overall approach goes beyond the typical comprehension of network topology in the literature and provides insights into a multivariable conceptualization of this concept. Within this context, we conceive the network topology of the Greek multilayer commuting network as a set of variables describing various topological properties in each layer and we apply a principal component analysis (PCA) method, which allows reducing the dimension of the available information to the least necessary to describe the multivariable notion of network topology, within a desired level of variability. This approach proposes a novel decomposition method in multilayer network analysis and is expected to reveal either competitive or synergetic roles between different aspects of network structure, geometry, and functionality of commuting.

The analysis is based on a multilevel methodological framework consisting of five discrete steps. The first regards modeling the interregional commuting network in Greece (Greek Commuting Network–GCN) into a multilayer graph model $\mathscr{M}(\mathscr{G})$ (De Domenico et al. 2013; Kivela et al. 2014; Boccaletti et al. 2014). In this multilayer model, the first layer ($G_1$) represents the infrastructure road network at the national scale, the second one ($G_2$) the prefectural grouping of $G_1$, the third ($G_3$) the road accessibility network between Greek prefecture (NUTS III), and the fourth ($G_4$) the directed interregional network of commuting flows. The second step includes the network analysis (Koschutzki et al. 2005; Newman 2010; Aleta et al. 2016), where major measures of network topology and geometry are computed to provide insights into the structure and functionality of this multilayer commuting network. At the third step, a set of network variables extracted from each layer are converted at a standard reference (at the regional scale) and are included in a principal component analysis (PCA), aiming to detect similarities or differences according to this variance optimization method. At the fourth step, the results of PCA are tabulated through a "min–max" criterion, which filters the minimum and maximum cases per principal component and per variable, aiming at distinguishing the most important performances at different geographical and administrative levels of resolution. Finally, at the fifth step, conclusions are formulated within the context of transport geography and regional science.

The construction of the GCN builds on the multilayer network paradigm (De Domenico et al. 2013; Kivela et al. 2014; Barthelemy 2011), which provides a rigorous and sufficient framework for the modeling and topological analysis of multifaceted communication systems, both within and between the network layers. However, in current literature (Barthelemy 2011; Boccaletti et al. 2014; Kivela et al. 2014), the relationship between layers is expressed in terms of interlayer linkages, and therefore the interlayer attributes in such networks are studied through a "within-the-box" rationale, namely by examining interlayer links that are by default construction elements of a multilayer graph. In this paper, the analysis goes beyond the current multilayer network conceptualization by proposing a PCA-driven multivariable framework for the study of interlayer relationships. In particular, this paper constructs a multilayer graph, where layers represent diverse aspects of the same socioeconomic system, the interregional transportation market of road commuting flows in Greece. By defining the available network layers concerning the fixed geographical scale of the whole country (at the national level), the differences between layers detected through comparisons in layer topological features can provide insights into the effect of the geographical and administrative resolution; and network functionality; on the configuration of the socioeconomic system of the Greek commuting. This is a novel approach to the extent it provides a framework for measuring topological differences between layers expressing different aspects of the same socioeconomic system that can be overall seen as the effect of each layer's modeling rule on the reference socioeconomic system. This approach advances current conceptualization on multilayer transportation networks, which currently focuses on the multimodal configuration of the network layers (Gallotti and Barthelemy 2014, 2015; Alessandretti et al. 2016) rather than on differences in geographical resolution and administrative level (Ducruet et al. 2011; Tsiotas and Polyzos 2018; Ghavasieh and De Domenico 2020).

Within this context, the proposed framework can be applicable across jurisdictions due to its methodological configuration, namely to the extent that it suggests a multilayer network modeling at different geographical scales, where all layers refer to a single (transportation or more broadly communication) system, and that it develops a quantitative approach evaluating the consistency of network topology mainly across these different geographical scales. Although the proposed methodological framework can be generalizable to every real-world application, its results are not by default universal to other realities because they are dependent on the framework's empirical specialization, within the context of the certain case study. However, the universal contribution of the proposed methodological framework is that it conceptualizes spatial distance as a major force of network topology and it develops a quantitative framework evaluating the consistency of network topology across different geographical scales and functionality. In the following paragraphs, the steps of the proposed methodological framework are described in more detail.

## 3.1 Graph Modeling

The interregional commuting network in Greece (GCN) is modeled to a multilayer graph $\mathscr{M}(\mathscr{G}, \mathscr{C} = \emptyset)$ consisting of four (4) layers $= \mathscr{G}\{G_p\} = \{V_p, E_p, \; p = 1, \ldots, 4\}$

without interlayer connections $= \mathscr{C}\{E_{ij} \subseteq V_i \times V_j\} = \emptyset$ (De Domenico et al. 2013; Kivela et al. 2014; Aleta et al. 2016), as it is shown in Fig. 1. The first layer $G_1(4,851;5,956)$ represents the infrastructure national road network in Greece and is modeled to a geo-referenced primal graph (Jiang and Claramunt 2004; Porta et al. 2006a; Marshall et al. 2018), where the $n_1 = 4,851$ nodes express route intersections and the $m_1 = 5,956$ links (edges) express road-paths between nodes. The data for the construction of this layer was extracted from the Hellenic Land and Mapping Organization (OKXE 2005) and refer to the primary, secondary, and tertiary levels of the national road network along with the primary and secondary levels of the regional road network in Greece, as they are defined at the Presidential Decree 401/93. Due to the insular Greek morphology, $G_1$ is not a connective graph, but an aggregate network (Tsiotas and Polyzos 2015a, b) including 156 components, the biggest of which is the road network of mainland Greece, whereas the other 155 are the local road networks of the Greek islands that are not connected with the mainland through road transport.

The second ($G_2$) layer is configured by the grouping of $G_1$ into 51 sub-graphs $G2 = \{G_{2p}\} = \{V_{2p}, E_{2p}, p = 1, …, 39\}$ expressing the same in number prefectures of Greece, as they were defined by the act of Law 2539/97, which is commonly known as the "*Kapodestrian*" administrative division of Greece. In practice, layer $G_2$ is the division of $G_1$ layer into 51 sub-graphs, where each is the part of $G_1$ included within the area of a certain Greek prefecture (NUTS III level). In the third layer $G_3(39;71)$, the $n_3 = 39$ nodes express the non-insular capital cities of the Greek prefectures, and the $m_3 = 71$ edges express direct road connectivity between prefectures. In particular, two nodes (prefecture capital cities) in this layer are connected if no administrative part of any other prefecture intermediates (or includes part of) their road-path. In this representation, nodes are geo-referenced at the coordinates of the city centers (WGS Web Mercator)
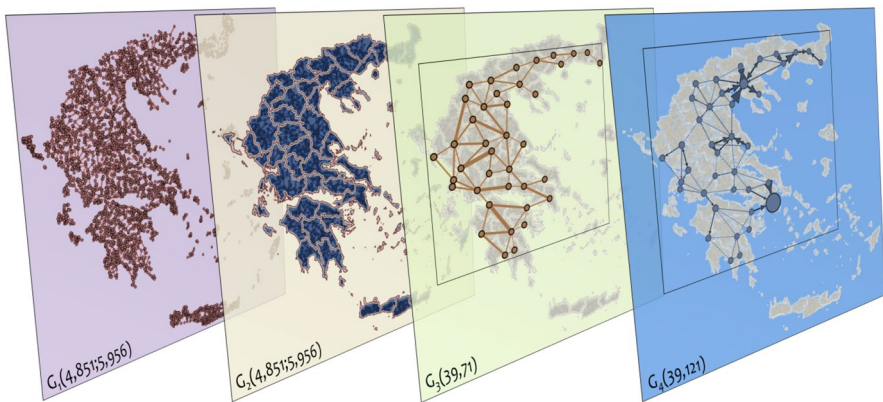


**Fig. 1** The interregional commuting network in Greece (GCN), modeled to a multilayer graph $\mathscr{M}(\mathscr{G})$. The first ($G_1$) layer is the infrastructure road network at the national scale, the second ($G_2$) is the prefectural grouping of $G_1$, the third ($G_3$) expresses road accessibility between Greek regions (NUTS III), and the fourth ($G_4$) is the directed interregional network of commuting flows. Data are extracted from OKXE (2005) and Google Maps (2019) and network layouts are produced by using the open-source software of Bastian et al. (2009)

and this layer expresses the direct road accessibility between prefectures in Greece. Data of this layer were extracted from the Google Digital Mapping Services (Google Maps 2019). Finally, the fourth layer $G_4$(39;121) builds on the $G_3$ layer and models the directed flows of interregional commuting between the non-insular Greek capital cities, where edge-weights $w_{ij}$ express the number of commuters moving daily from city $i = 1,2,\ldots,39$ to city $j = 1,2,\ldots,39$ ($i \neq j$) to work. The available commuting data concern employed persons with residence in the area by place of work, and were extracted from the 2011 national census of Greece (Hellenic Statistical Service – ELSTAT 2011).

## 3.2 Network Analysis

After the multilayer graph modeling, the topology of the GCN is studied by using a set of network measures shown in Table 1. In general, network measures capture a certain aspect of network topology (e.g. connectivity, intermediacy, path length, clustering, centrality, etc., see Koschutzki et al. 2005 and Newman 2010) and their total consideration can be seen as an approximation of the overall network topology related to pattern structures in complex networks, such as the random, lattice, small-world, and scale-free topologies (Tsiotas 2019). Some measures refer to the total network $G$ and are considered as global (or aggregate), whereas others refer to network components (nodes, edges) and are considered as local (Koschutzki et al. 2005).

Direct comparisons of network measures between layers are possible only for cases capturing aspects of network topology and thus for those that are measured by unit-free or dimensionless numbers. For instance, measures of connectivity, neighborhood, shortest-paths, and centralities, such as node degree, clustering coefficient, (binary) average path length and network diameter, graph density, and modularity are directly comparable between layers (Koschutzki et al. 2005; Newman 2010; Boccaletti et al. 2014; Kivela et al. 2014). On the other hand, measures that are defined within a weighted (edge-dependent) context, such as average path length, edge length, weighted network diameter, etc. (Barthelemy 2011; Newman 2010), are not directly comparable and should be de-escalated by normalization or rescaling (Walpole et al. 2012) techniques to be comparable (they are indirectly comparable). In general, most of the network measures are well-defined within connected components and thus calculations can be made without restrictions. However, in networks with more than one component, network measures are not well-defined, and consequent computational and interpretation issues emerge (Koschutzki et al. 2005). For instance, it is impossible to compute the distance between two disconnected nodes and it is not easy to evaluate the importance of two nodes with the same degree that is included in components of different sizes. This problem is generally known as "*insufficient connectivity*" in complex networks and several repairing methods were proposed to overcome its restrictions each being appropriate within a certain context depending on the modeling and the purpose of the research. For this study, we use the simplest *local restriction method* – LRM (see Koschutzki et al. 2005; Tsiotas and Polyzos 2015a, b), where measures computed within connected components (local) are converted to global (i.e. referring to the aggregate network) without

**Table 1** Measures of network topology used for the analysis of the GCN

| Measure | Description | Math Formula | Reference |
|---|---|---|---|
| Graph | A pair set consisting of a node-set V and an edge-set E. In graph G(V,E), n expresses the number of nodes and m the number of links | $G(V,E)$ | Newman (2010), Barthelemy (2011) |
| Graph density (ρ) | The fraction of the existing connections (m) to the number of the possible connections. It expresses the probability to meet a link between two randomly chosen nodes in the network | $\rho = m / \binom{n}{2} = \dfrac{2m}{n\cdot(n-1)}$ | Newman (2010) |
| Network diameter (dG) | The longest path p(i,j) in the network | $d(G) = \max\{p(i,j)\|i,j \in V\}$ | Koschutzki et al. (2005) |
| Eccentricity (e(u)) | The longest path p(u,j) in the network from a given node u ∈ V | $e(u) = \max\{p(u,j)\|j \in V\}$ | Koschutzki et al. (2005) |
| Node Degree (k) | The number of graph edges being adjacent to a given node i. It expresses the communication potential of a node | $k_i = m(i) = m_i = \sum_{j\in V(G)} \delta_{ij} = \sum_{j\in V(G)} \delta_{ij},$ where $\delta_{ij} = \begin{cases} 1, \text{if } e_{ij} \in E(G) \\ 0, \text{otherwise} \end{cases}$ | Newman (2010) |
| In-degree (k−) | The number of incoming edges being adjacent to a given node i (this measure is applicable in directed networks) | $k_i- = m_i- = \sum_{j\in V(G)} \delta_{ij}-,$ where $\delta_{ij} = \begin{cases} 1, \text{if } e_{ij} \in E(G) \\ 0, \text{otherwise} \end{cases}$ | Newman (2010), Barthelemy (2011) |
| Out-degree (k+) | The number of outgoing edges being adjacent to a given node i (this measure is applicable in directed networks) | $k_i- = m_i- = \sum_{j\in V(G)} \delta_{ji}-,$ where $\delta_{ij} = \begin{cases} 1, \text{if } e_{ji} \in E(G) \\ 0, \text{otherwise} \end{cases}$ | Newman (2010) |
| Node strength (s) | The sum of weights (w_ij) of the links (e_ij) being adjacent to a given node i. The δ_ij operator is the Kronecker delta function yielding a true output for links belonging to graph G | $s_i = s(i) = \sum_{j\in V(G)} \delta_{ij} \cdot d_{ij},$ where $d_{ij} = w(e_{ij})$ in km | Newman (2010), Barthelemy (2011) |
| Average degree (⟨k⟩) | The mean value of node degrees $k_i$, where i represents a network node | $\langle k \rangle = \dfrac{1}{n} \cdot \sum_{i=1}^{n} k_i$ | Albert and Barabási (2002) |

**Table 1** (continued)

| Measure | Description | Math Formula | Reference |
|---|---|---|---|
| *Closeness Centrality (CC)* | Is computed on the average path-lengths $d(i,j)$ originating from a given node $i \in V$ to all other nodes $jV$ in the network. It is a measure of accessibility | $CC(i) = \frac{1}{n-1} \cdot \sum\limits_{j=1, i \neq j}^{n} d_{ij} = \bar{d}_i$ | Koschutzki et al. (2005) |
| *Betweenness Centrality (CB)* | The proportion that is defined by $\sigma(i)$ shortest-paths that pass through a given node $i$ to the total shortest-paths $\sigma$ in the network. It expresses intermediacy | $CB(i) = \sigma(i)/\sigma$ | Koschutzki et al. (2005) |
| *Local Clustering Coefficient (C(i))* | The probability a node $i$ to have $E(i)$ neighbors connected. It is computed on the number of triangles configured by node $i$ to the number of the total triplets $k_i(k_i-1)$ shaped by this node | $C(i) = \frac{E(i)}{k_i \cdot (k_i - 1)}$ | Newman (2010) |
| *Modularity (Q)* | Objective function expressing the potential of a network to be subdivided into communities. In the mathematical formula, $g_i$ is the community of node $i \in V(G)$, $[A_{ij} - P_{ij}]$ is the difference of the actual minus the expected number of edges falling between a particular pair of vertices $i,j \in V(G)$, and $\delta(g_i, g_j)$ is an indicator function returning 1 when $g_i = g_j$ | $Q = \dfrac{\sum\limits_{ij} [A_{ij} - P_{ij}] \cdot \delta(g_i, g_j)}{2m}$ | Blondel et al. (2008), Fortunato (2010) |
| *Average path length ⟨l⟩* | The average of the path length $d[p(i,j)]$ computed for all accessible pairs $(i,j)$ of network nodes | $\langle l \rangle = \dfrac{\sum\limits_{i \in V} d(p(i,j))}{n \cdot (n-1)}$ | Barthelemy (2011) |

modifications (as they are). This approach is considered satisfactory for the case of GCN because, first, the network is already complex due to its multilayer configuration, which implies that insufficient connectivity should be managed most simply, and, second, because restriction due to model selection of each layer is already applicable implying that seeking for a high level of resolution in the analysis is aimless.

### 3.3 Principal Component Analysis

The study aims to extract the major topological components of the multilayer GCN. To do so, we apply a principal component analysis (PCA) (Wold et al. 1987; Norusis 2008) between various topological variables of different GCN layers. Generally, the PCA is used to reduce the dimension of a set of possibly correlated (original or source) variables, by converting them into a set of linearly uncorrelated variables, called principal components. The process applies an orthogonal transformation to the variables, which can be perceived as fitting a *p*-dimensional ellipsoid to the data, where each axis of the ellipsoid is a principal component. When some axes are relatively small (in terms of explained variance) they can be removed from the dataset and thus the overall dimension of the system is reduced. The resulting principal components configure an uncorrelated orthogonal basis and are arranged into an ascending order, where the first principal component has the largest possible variance and so on. The choice of the number of principal components can be illustrated to a scree-plot displaying a distinct break between the steep slope of the large factors and the gradual trailing of the rest (the scree), according to their eigenvalue size. The point at which a component adds relatively little variance to the total variance defines the proper number of principal components. The specifications of the algorithm used to perform the analysis are defined as follows: (a) computations are applied to the correlation (instead of the covariance) matrix, which is preferable for variables measured on different scales (Norusis 2008); (b) 25 maximum iterations is set for convergence; (c) extraction is applied when eigenvalues are greater than one ($>1$); and (d) rotation is applied according to the "Varimax" method, which minimizes the number of variables that have high loadings on each factor, simplifying though the interpretation of the factors.

In transportation research, PCA enjoys a variety of applications, such as the study of structural efficiency of airline networks (Adler and Golany 2001), measurement of transportation impacts of urban sprawl (Ewing et al. 2003), analysis of gas-emission and meteorological impacts of transportation (Shiva Nagendra and Khare 2003), evaluation of public-transport customers' satisfaction (Sezhian et al. 2011), origin–destination assessment (Djukic et al. 2012), ecological assessment of transportation manufacturing (Park et al. 2015), and many others. For instance, Chen et al. (2011) studied by using PCA the road-safety of three roads in Beijing, China, and they further evaluated the effectiveness of safety programs of traffic flow functionality. The analysis was applied to traffic data extracted from these roads and provided insights about the security level of each road, highlighting aspects of best and worst performance. The study concluded that safety programs cannot affect the normal operation of traffic flow but they

are applicable on a cost–benefit basis. Chung and Song (2018) studied, by using PCA, the determinative factors of motorcycle crash-severity in Korea, on data extracted from the year 2009. The analysis distinguished the most critical factors of motorcycle crash-severity and provided a framework for formulating remedial policy measures to decrease this phenomenon on roadways. Tang et al. (2018) studied, by using PCA, some pivotal influencing factors of fatal traffic accidents recorded in the United States, for the period 2010–2016. The analysis examined the traffic conditions and several pivotal influencing factors of fatal traffic accidents, concluding that tire wear, rim damage, exhaust system failure, and coupling failure are the most important factors. Bham et al. (2019) developed a composite rank-measure overcoming limitations of existing measures for network screening of hotspots or high-crash locations on highways. The PCA contributed to the assignment of non-arbitrary weights to different individual measures that are incorporated in a composite measure maximizing the variance without building on assumptions of the crash-data statistical distribution. As it can be observed, all such applications are described by the multidimensionality of available data, requiring data-management through dimension reduction, where PCA is an established method. Within this framework, the PCA is applied to the set of network variables of the GCN to restructure its topology into a set of uncorrelated topological (principal) components, where items within each component are correlated in terms of variability. The overall approach proposes a novel method for linking network variables across layers, which are defined at different geographical scales and functionality but refer to the same socioeconomic system.

For the variables of different layers to be compatible with the PCA, they should have the same length (number of cases) (Norusis 2008). To comply with this criterion, all GCN variables were converted under a list-wise exclusion criterion, to have a length of 39 cases, equal to the number of the non-insular Greek prefectures (defined by $G_3$ and $G_4$ layers). This restriction drove the exclusion of layer $G_1$ from the analysis because it coincides with the $G_2$ layer. However, in the other parts of the analysis, the available variables were managed by default pair-wisely. Table 2 shows the 30 available topological variables of length 39 participating in the PCA of GCN.

On the other hand, no further data management, such as rescaling, number density computation, or normalization (Norusis 2008; Walpole et al. 2012) was applicable to improve consistency between variables. Such approaches were not considered to significantly contribute to the analysis, because, although there is high-detailed information available on network structure, the commuting flow data is only available at low-resolution level (i.e. at the $G_4$ layer that is configured at the interregional scale), a fact that inevitably also restricts the resolution of the computations. Being aware of this by default low-resolution context of modeling conditions on network flows, the analysis builds on a "minimax" consideration, which takes into account only the minimum and maximum values per case, and not on accurate enumerations, a fact that sets any attempt for improving the computational resolution as redundant. However, these restrictions do not deteriorate the added value of the proposed methodological framework to conceptualize a multilayer network modeling on layers configured at different geographical scales, which can be scalable to other

**Table 2** Network variables participating in the PCA of GCN

| Layer | Variable Code | Variable Symbol | Description |
|---|---|---|---|
| $G_2$ | $G_{201}$ | RDD | Region's road density |
| | $G_{202}$ | RDL | Region's road length |
| | $G_{203}$ | NODES | Number of nodes |
| | $G_{204}$ | EDGES | Number of edges |
| | $G_{205}$ | MAX.DEG | Maximum degree |
| | $G_{206}$ | MIN.DEG | Minimum degree |
| | $G_{207}$ | AV.DEG | Node average degree |
| | $G_{208}$ | STR | Strength |
| | $G_{209}$ | COMP | Number of connected components |
| | $G_{210}$ | AV.E.LNG | Average edge length |
| | $G_{211}$ | DIAM | Network diameter |
| | $G_{212}$ | AV.C | Average clustering coefficient |
| | $G_{213}$ | AV.PL | Average path length |
| | $G_{214}$ | AV.CC | Average closeness centrality |
| | $G_{215}$ | AVG.CB | Average betweenness centrality |
| | $G_{216}$ | AV.ECC | Average eccentricity |
| $G_3$ | $G_{301}$ | DEG | Node degree |
| | $G_{302}$ | STR | Node strength |
| | $G_{303}$ | C | Node clustering coefficient |
| | $G_{304}$ | ECC | Node eccentricity |
| | $G_{305}$ | CC | Node closeness centrality |
| $G_4$ | $G_{306}$ | CB | Node betweenness centrality |
| | $G_{401}$ | COMMUTERS | Number of commuters in each node |
| | $G_{402}$ | IN.DEG | In-degree |
| | $G_{403}$ | OUT.DEG | Out-degree |
| | $G_{404}$ | DEG | Node degree |
| | $G_{405}$ | ECC | Node eccentricity |
| | $G_{406}$ | CC | Node closeness centrality |
| | $G_{407}$ | CB | Node betweenness centrality |
| | $G_{408}$ | C | Node clustering coefficient |
| | $G_{409}$ | STR | Node strength |

realities and therefore the restrictions applied to this study do not seem to suggest a considerable concern. The overall approach proposes a PCA-based methodological framework for the in common modeling of various topological aspects in complex networks defined at different geographical scales. In general, the PCA-driven analysis is applied to GCN to restructure its topology into a set of uncorrelated topological (principal) components, where items within each component are relevant in terms of their variability and of their contribution to the semiology of the principal component.

## 4 Results and Discussion

### 4.1 Network Analysis

#### 4.1.1 Network Measures

At the first step of the analysis, network measures are computed for each layer and their results are shown in Table 3. Layer $G_1$ appears almost a hundred times ($100\times$) greater in size ($n$, $m$), and $G_2$ is almost two times ($2\times$) greater in size than the other two ($G_3$, $G_4$). Differences in the number of components seem to comply with the graph-model configuration, where layers $G_1$ and $G_2$ include the total (mainland and insular) national geographical space, while the other two ($G_3$, $G_4$) just the mainland. In terms of average degree $\langle k \rangle$, layers $G_3$ and $G_4$ have the greatest values, implying that these networks enjoy higher average connectivity than the others. A similar picture is shaped for the edge-distance measures ($\langle s \rangle, \langle d(e_{ij}) \rangle, \sum_{ij} d(e_{ij})$) where the $G_3$ scores are greater than these of $G_1$ and $G_2$ (weighted scores of layer $G_4$ are not comparable and thus not included). This is a reasonable finding since $G_3$ is constructed at the interregional scale, whereas the other two layers also include intraregional connections.

For the binary path-based measures ($\langle l \rangle_{\text{bin}}$, $d_{\text{bin}}(G)$), we can observe that they follow a ranking $x(G_1) > x(G_2) > x(G_3) \approx x(G_4)$ similar to the node ordering $n(G_1) = 4{,}851 > n(G_2) = 94.59$ (a rational number due to average computations) $> n(G_3) = n(G_4) = 39$. This observation describes a ranking relationship between network size (nodes) and these path-based measures. To the extent that the (binary) average path length and network diameter express aspects of accessibility in networks, this observation illustrates a gravitational configuration of these measures due to their ranking relevance to network size. In terms of interpretation, this observation implies that accessibility becomes more perplexed as network size increases. However, the weighted expressions of $\langle l \rangle_{\text{wei}}$, $d_{\text{wei}}(G)$ shape a different picture, where $G_3$ has greater scores than $G_1$ and $G_2$, which seems to be an effect of geographical resolution (since $G_3$ is constructed at the interregional level, whereas $G_1$ and $G_2$ include intraregional connections). In terms of graph density, layers $G_3$ and $G_4$ have the highest scores, which seem also to be an effect of network size. This is also the case for clustering coefficient (scores of $G_3$ and $G_4$ are greater than the other layers), implying that models at the interregional scale enjoy higher clustering (circulation of information/ traffic) than the models at the intraregional scale. To the extent that clustering in transportation networks is translated to traffic decentralization (through the development of proportionally more triangular connections), this observation points out the advantage of the interregional administrative setting for the development of peripheral connectivity, which is a major goal towards regional development and regional inequalities convergence. Finally, modularity scores seem to follow a (network) size distribution pattern, according to which greater in size networks ($G_1 > G_2 > G_4$, $G_3$) tend to divide into communities more easily. This observation may also point out an advantage of the interregional administrative setting toward improving structural consistency in transport management configurations.

**Table 3** Network measures of the multilayer GCN

| Measure/ Metric | Symbol | Unit | $\mathcal{G}$ | | | |
|---|---|---|---|---|---|---|
| | | | $G_1$ | $G_2{}^{\text{a}}$ | $G_3$ | $G_4$ |
| Nodes | $n$ | #[b] | 4,851 | 94.59 | 39 | 39 |
| Edges/links | $m$ | # | 5,956 | 116.98 | 71 | 121 |
| Components | $p$ | # | 164 | 2.82 | 1 | 1 |
| Maximum degree | $k_{\text{max}}$ | # | 8 | 8 | 7 | 13 |
| Minimum degree | $k_{\text{min}}$ | # | 1 | 0 | 1 | 1 |
| Average degree | $\langle k \rangle$ | # | 2.456 | 1.22 | 3.641 | 3.103 |
| Average strength[c] | $\langle s \rangle$ | wei | 13.979 (km) | 6.171 (km) | 322.26 (km) | 1,506.4 (commuters) |
| Average edge length[c] | $\langle d(e_{ij}) \rangle$ | wei | 5.757 (km) | 5.226 (km) | 85.497 (km) | 506.74 (commuters) |
| Total edge length[c] (km) | $\sum_{ij} d(e_{ij})$ | wei | 23,377 (km) 3334.4 (km) | 588 58,754 (commuters) | | |
| Average path length (binary) | $\langle l \rangle_{\text{bin}}$ | # | 47.29 | 6.87 | 4.58 | 4.196 (5.213[d]) |
| Average path length[c] | $d(\langle l \rangle)$ | wei | 272.25 (km) | 35.9 (km) | 389.045 (km) | 2641.6 (commuters) |
| Network diameter (binary) | $d_{\text{bin}}(G)$ | # | 146 | 18.27 | 14 | 14 |
| Network diameter[c] | $d_w(G)$ | wei | 840.52 (km) | 95.47 (km) | 1,124.4 (km) | 7,094.4 (commuters) |
| Graph density (non-planar) | $\rho$ | net | 0.001 | 0.018[c] | 0.097 | 0.097 (0.082 [d]) |
| Clustering coefficient | $C$ | net | 0.046 | n/a[e] | 0.47 | 0.474 (0.073[d]) |
| Average clustering coefficient | $\langle C \rangle$ | net | 0.07 | 0.04 | 0.422 | 0.514 (0.405[d]) |
| Modularity | $Q$ | net | 0.964 | 0.76 | 0.566 | 0.621 (0.719 [c]) |

[a]Average values computed on the regional sub-networks

[b]Dimensionless number (unit-free measure)

[c]Computed on weighted network

[d]Computed on directed network

[e]Not available

### 4.1.2 Degree Distributions and Strengths

The degree distributions of the GCN layers are shown in Fig. 2. For the $G_2$ layer, box-plots (Walpole et al. 2012) are also used to display the distribution variables $p(k=1)$, $p(k=2)$, $p(k=3)$, $p(k=4)$, and $p(k=5)$, configured within each available case of degree, namely from $k=1$, …, to $k=5$ (maximum degree). Substantially, instead of one degree distribution, Fig. 2b illustrates 51-degree distributions (shown

in dots) configured by the concordant sub-networks included in this layer. Therefore, the available box plots provide a collective picture of the 51 degree distributions of the sub-networks in layer $G_2$, and include information about the range, interquartile range, average, and outliers of the available distribution variables. For layer $G_4$, the in-degree (id) (Fig. 2e) and out-degree (od) (Fig. 2f) distributions are also shown. As it can be observed, all distributions follow a bell-shaped pattern implying that connectivity is undertaken by the majority of nodes in their corresponding networks. This implies that all layers of GCN are structured under the effect
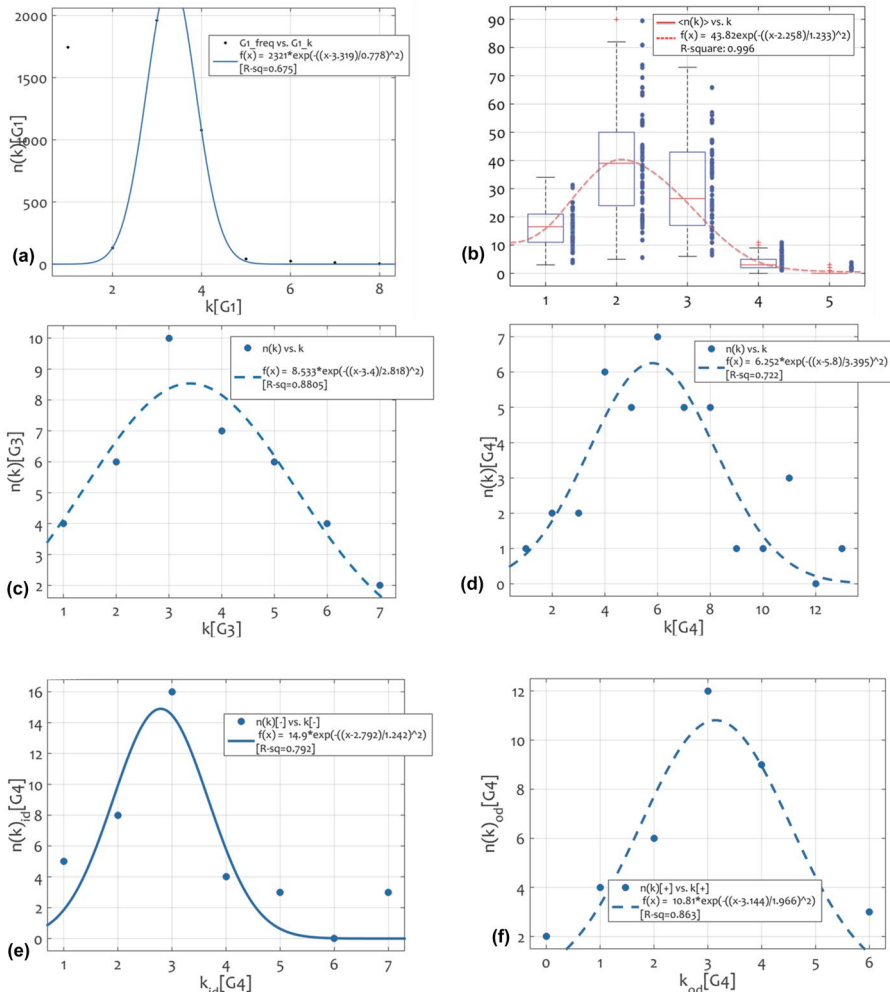


**Fig. 2** Scatter plots $(k, n(k))$ of the degree distributions for: **a** $G_1$ layer; **b** $G_2$ layer, including box-plots also illustrating the distributions within each degree $k=1,\dots,5$, for the available 51 sub-networks included in this layer; **c** $G_3$ layer; **d** $G_4$ layer (undirected); **e** $G_4$ layer, in-degree $(k-)$; and **f** $G_4$ layer, out-degree $(k+)$. A normal fitting curve is fitted, where possible

of spatial constraints (Barthelemy 2011), where hubs frequently emerge and do not undertake the major load of connectivity. In terms of kurtosis, the degree distribution of $G_1$ is the most peaked (Fig. 2a), whereas $G_3$ is the smoothest one (Fig. 2c), indicating that $G_1$ is leptokurtic and $G_3$ platykurtic. This observation implies that $G_3$ may show more distinguishable structures of hierarchy since more highly (hubs) and less connected (spokes) nodes exist in this layer proportionally to the others. Also, degree distributions of layers $G_3$ (Fig. 2c) and $G_4$ (Fig. 2d) appear more symmetric than the others, implying a better symmetry in the lattice-like configuration of their networks. Finally, the out-degree distribution (Fig. 2f) of $G_4$ appears more similar to the undirected case (Fig. 2d), implying that the outgoing configuration of layer $G_4$ is possibly more affected by network connectivity.

At next, correlations between incoming ($s_-$) and outgoing ($s_+$) strengths are examined according to the scatterplots shown in Fig. 3. Similar to the in- and out-degree defined in Table 1, incoming strength is the sum of incoming commuting flows and outgoing strength is the sum of the outgoing flows, for a given node. Reference lines are set to the median values and divide the plane into four quadrants. Cases above the line are considered as high (H) and below the line as low (L). This division produces four distinct commuting profiles corresponding to labor markets that can be described as *isolated* (LL), *exporting* (LH), *importing* (HL), and *interactive* (HH). In particular, the first (LL) profile includes cities with the low incoming and outgoing commuting flow and it can be considered as "isolated" to the extent that it describes cases with small interaction (both incoming and outgoing) with their neighborhood regions and thus those referring to more self-sufficient economies. The second profile (LH) includes cases with low incoming and high outgoing commuting flows and illustrates "exporting" economies leaking more labor force than this they receive. The third (HL) profile includes cases with high incoming and low outgoing commuting flows and illustrates "importing" economies to the extent that these cases are receiving more labor force than they leak, describing thus markets of great occupation opportunities. Finally, the fourth (HH) profile includes cases with high incoming and outgoing commuting flows and illustrates "interactive" economies to the extent that its cases have high interaction (both incoming and outgoing) with their neighborhood regions. According to this classification, the HL quadrant describes optimum cases benefited high incoming flows with small loss of labor force.

Further, different classifications in Fig. 3 between the unadjusted (s) and per-capita ($s^*$) cases, which are computed by dividing the commuting flows by the total city population, illustrate the effect of the population in the commuting flows of the GCN. As it can be observed, the cities of *Amfissa*, *Athens*, *Florina*, *Grevena*, *Igoumenitsa*, *Karpenission*, *Kastoria*, *Komotini*, *Lamia*, *Lefkada*, *Levadia*, *Patra*, *Tripolis*, and *Preveza* belong to different classes between the unadjusted and per-capita cases. Among these cities, *Athens* and *Patra* (which are amongst the most populated, with almost 4 million and over 300 k population, respectively) move from HL(s) to LL($s^*$), implying that they are considered as stable labor-markets when the control of the population is removed. On the other hand, *Amfissa*, *Karpenission*, *Kastoria*, *Lefkada*, and *Tripolis* move from LL(s) to HL($s^*$) when the control of the population is removed from the commuting strengths, implying that they are relatively attractive labor markets. Finally, Fig. 3c shows that group LL has a (statistically) smaller
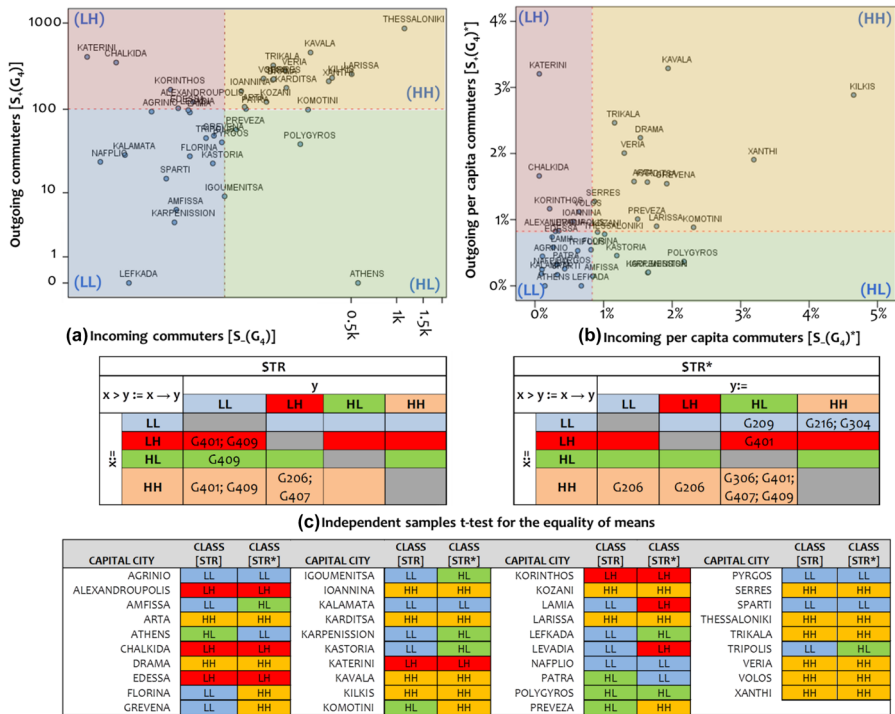
**STR**

| x > y := x → y | y | | | |
|---|---|---|---|---|
| | **LL** | **LH** | **HL** | **HH** |
| **LL** | | | | |
| **LH** | G401; G409 | | | |
| **HL** | G409 | | | |
| **HH** | G401; G409 | G206; G407 | | |

(x:: column on left side)

**STR\***

| x > y := x → y | y:= | | | |
|---|---|---|---|---|
| | **LL** | **LH** | **HL** | **HH** |
| **LL** | | | G209 | G216; G304 |
| **LH** | | | G401 | |
| **HL** | | | | |
| **HH** | G206 | G206 | G306; G401; G407; G409 | |

**(c)** Independent samples t-test for the equality of means

| CAPITAL CITY | CLASS [STR] | CLASS [STR*] | CAPITAL CITY | CLASS [STR] | CLASS [STR*] | CAPITAL CITY | CLASS [STR] | CLASS [STR*] | CAPITAL CITY | CLASS [STR] | CLASS [STR*] |
|---|---|---|---|---|---|---|---|---|---|---|---|
| AGRINIO | LL | LL | IGOUMENITSA | LL | HL | KORINTHOS | LH | LH | PYRGOS | LL | LL |
| ALEXANDROUPOLIS | LH | LH | IOANNINA | HH | HH | KOZANI | HH | HH | SERRES | HH | HH |
| AMFISSA | LL | HL | KALAMATA | LL | LL | LAMIA | LL | LH | SPARTI | LL | LL |
| ARTA | HH | HH | KARDITSA | HH | HH | LARISSA | HH | HH | THESSALONIKI | HH | HH |
| ATHENS | HL | LL | KARPENISSION | LL | HL | LEFKADA | LL | HL | TRIKALA | HH | HH |
| CHALKIDA | LH | LH | KASTORIA | LL | HL | LEVADIA | LL | LH | TRIPOLIS | LL | HL |
| DRAMA | HH | HH | KATERINI | LH | LH | NAFPLIO | LL | LL | VERIA | HH | HH |
| EDESSA | LH | LH | KAVALA | HH | HH | PATRA | HL | LL | VOLOS | HH | HH |
| FLORINA | LL | HH | KILKIS | HH | HH | POLYGYROS | HL | HL | XANTHI | HH | HH |
| GREVENA | LL | HH | KOMOTINI | HL | HH | PREVEZA | HL | HH | | | |

**Fig. 3** Scatter-plots of correlations between the incoming ($S_{(-)}$) and outgoing ($S_{(+)}$) strengths of the Greek commuting network (GCN), in **a** numeric (log–log scale), and **b** per capita (divided by the total city population) expression (metric scale). Reference lines express the median values and divide the plane into four quadrants (where H=high and L=low). Optimum cases appear in the HL quadrant, where incoming flows are high and outgoing flows are low for a given node. **c** Results of independent samples t-tests for the equality of means between the LL, LH, HL, and HH groups for the topological variables are shown in Table 2

average than the other three groups (LH, HL, and HH) in variable $G_{409}$ (strength). It also has a smaller average than the groups LH, and HH in $G_{401}$ (commuters). Group HH has a greater average than group LH in $G_{206}$ (min degree) and $G_{407}$ (betweenness centrality). Accordingly, for the per capita ($s^*$) case, we can observe that group HH has a greater average than LL and LH in variable $G_{206}$ (min degree); group HH has a greater average than HL in $G_{306}$ (betweenness centrality), $G_{401}$ (commuters), $G_{407}$ (betweenness centrality), $G_{409}$ (strength). Group LL has a greater average than HL in $G_{209}$ (no. of connected components); and a greater average than HH in $G_{216}$ (average eccentricity) and $G_{304}$ (eccentricity). Group LH has a greater average than HH in $G_{401}$ (commuters). This incoming and outgoing strength correlation analysis generates groups (LL, LH, HL, and HH) with pair-wisely distinguishable topological attributes regarding commuting flows ($G_{401}$, $G_{409}$), intermediacy ($G_{306}$, $G_{407}$), eccentricity ($G_{216}$, $G_{304}$), and connectivity ($G_{206}$, $G_{209}$), addressing avenues for further research towards the topological configuration of the relationship between strength directionality ($s_+$, $s_-$).

## 4.2 Principal Component Analysis

The results of the PCA applied to the network variables of Table 2 are shown in Fig. 4. According to the scree-plot (Fig. 4a), 8 components are significant to the analysis and can be considered as principal, explaining an amount of 85% of the total variance. This result translates 74% of dimension reduction (8 principal components out of 31 available items) with only 15% loss of information (total variance loss). Provided that principal components are uncorrelated variables configuring an orthogonal basis for the total system's variance (Norusis 2008), the resulting principal components have similarly separate semiology. That is, to the extent that each principal component includes a unique combination of loadings (Fig. 4b) and thus of information linked to the 31 available items, each combination is also uncorrelated with the others, in terms of its semiology composed by the physical meanings of the items. Within this context, items with the highest loadings within each component can be considered as relevant in terms of their great importance in the configuration of the unique semiology of a principal component. To this end, Fig. 4c and d display respectively the maximum and minimum values of the PCA loadings (Fig. 4b), where cases included within a component are considered as relevant in terms of semiology.
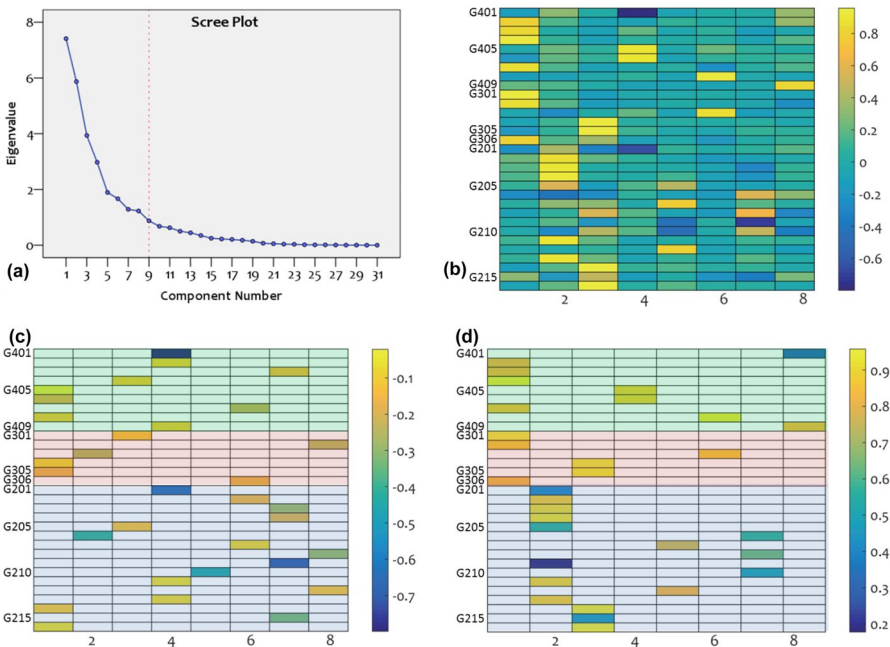


**Fig. 4** Results of the principal component analysis (PCA) applied to the network variables (shown in Table 2) of GCN, where **a** shows the scree plot with a reference line to the 8 principal components, **b** shows the rotated coefficients matrix, **c** shows the minimum, and **d** shows the maximum values of the rotated coefficients matrix. Rows 1:9 correspond to $G_4$ layer (variables $G_{401}:G_{409}$), 10:15 to $G_3$ layer (variables $G_{301}:G_{306}$), and 16:31 to $G_2$ layer (variables $G_{201}:G_{216}$)

To facilitate the interpretation of the principal components, the first two layers ($G_1$ and $G_2$) are considered as aspects of the GCN's geometry (commuting cost/intra-regional level); the third layer ($G_3$) as an aspect of the GCN's accessibility (commuting cost/interregional level); and the fourth ($G_4$) as an aspect of the GCN's functionality (commuting utility/interregional level). A summary of the principal components resulted from the PCA and of their semiology is shown in Table 4. In particular, the first principal component (PC#1) is positively related (Fig. 4b, c) to the node eccentricity ($G_{216}, G_{304}, G_{405}$) and closeness centrality ($G_{214}, G_{305}, G_{406}$) of all three ($G_2, G_3$, and $G_4$) layers. On the other hand, it is negatively related (Fig. 4b, d) to the node degree ($G_{301}$), strength ($G_{302}$), and betweenness ($G_{306}$) of the $G_3$ layer, to the in- ($G_{402}$), out- ($G_{403}$), and total ($G_{404}$) degree of the $G_4$ layer, and the betweenness centrality ($G_{407}$) of $G_4$ layer. In terms of interpretation, PC#1 is positively related to network-accessibility (as expressed by the measures of eccentricity and closeness centrality) across all layers and negatively related to connectivity (degree) and betweenness of the GCN's accessibility and functionality.

Next, the second principal component (PC#2) is positively related (Fig. 4b, c) to the betweenness centrality ($G_{206}$) of the $G_2$ layer and clustering coefficient ($G_{303}$) of the $G_3$ layer. On the other hand, it is negatively related (Fig. 4b, d) to the road density and length ($G_{201}, G_{202}$), to network size ($G_{203}, G_{204}$), maximum degree ($G_{205}$), number of connected components ($G_{209}$), network diameter ($G_{211}$), and average path length ($G_{211}$) of layer $G_2$. In terms of interpretation, PC#2 seems to be positively related to the clustering of GCN's accessibility and the intermediacy (betweenness) of GCN's geometry and negatively related to measures describing network size (nodes, edges, connected components, diameter, and average path length) of the geometry of GCN. The third principal component (PC#3) is positively related (Fig. 4b, c) to the maximum node degree ($G_{205}$) of $G_2$ layer, to node degree ($G_{301}, G_{404}$) of $G_3$ and $G_4$ layer, and negatively related (Fig. 4b, d) to closeness centrality and eccentricity ($G_{214}, G_{216}, G_{304}, G_{305}$) of $G_2$ and $G_3$ layers and betweenness centrality ($G_{215}$) of $G_2$ layer. In terms of interpretation, PC#3 appears positively related to connectivity (maximum degree, degree) across all layers and negatively related to network accessibility and the betweenness of the GCN's geometry.

The fourth principal component (PC#4) is positively related (Fig. 4b, c) to road density ($G_{201}$), network diameter ($G_{211}$), and average path length ($G_{213}$) of $G_2$ layer and commuters ($G_{401}$), in degree ($G_{402}$), and node strength ($G_{409}$) of $G_4$ layer and negatively related (Fig. 4b, d) to eccentricity ($G_{405}$) and closeness centrality ($G_{406}$) of $G_4$ layer. In terms of interpretation, PC#4 includes positive information about road density and network length of the GCN's geometry, as well as connectivity and flow information of GCN's functionality. On the other hand, it is negatively related to network accessibility (closeness and eccentricity) of the GCN's functionality. The fifth principal component (PC#5) is positively related (Fig. 4b, c) to edge length ($G_{210}$) and negatively related (Fig. 4b, d) to node degree ($G_{207}$) and clustering coefficient ($G_{212}$) of the $G_2$ layer. In terms of interpretation, PC#5 includes positive information about edge length and negative information about connectivity and clustering of the GCN's geometry. Next, the sixth principal component (PC#6) is positively related (Fig. 4b, c) to road length ($G_{202}$) and average degree ($G_{207}$) of $G_2$ layer, to betweenness centrality ($G_{306}$) of $G_3$ and $G_4$ layer, and it is negatively related

**Table 4** Summary and semiology of the principal components resulted from the PCA

| Principal Component | Correlated source variables [Semiology] | |
|---|---|---|
| | POSITIVELY (+) | NEGATIVELY (−) |
| PC#1 | G214,G216,G304,G305,G405,G406,G408 [Network-accessibility (eccentricity and closeness) across all aspects of GCN] | G301,G302,G306,G402,G403,G404,G407 [Connectivity and betweenness of the GCN's accessibility[a] and functionality[b]] |
| PC#2 | G206,G303 [Clustering of the GCN's accessibility[a] and betweenness of the GCN's geometry[c]] | G201,G202,G203,G204,G205,G209,G211,G213 [Network size (nodes, edges, connected components, diameter, and average path length) of GCN's geometry[c]] |
| PC#3 | G205,G301,G404 [Connectivity across all layers] | G214,G215,G216,G304,G305 [Network-accessibility (closeness and eccentricity) and betweenness of the GCN's geometry[c]] |
| PC#4 | G201,G211,G213,G401,G402,G409 [Road density and network length (diameter and average path length) of GCN's geometry[c] and connectivity and flow information of GCN's functionality[b]] | G405,G406 [Network accessibility (closeness and eccentricity) of GCN's functionality[b]] |
| PC#5 | G210 [Length of GCN's geometry[c]] | G207,G212 [Connectivity and clustering of the GCN's geometry[c]] |
| PC#6 | G202,G207,G306,G407 [Road length and average degree of the GCN's geometry[c] and betweenness centrality of the GCN's accessibility[a] and functionality[b]] | G303,G408 [Clustering of the GCN's accessibility[a] and functionality[b]] |
| PC#7 | G203,G204,G209,G215,G403 [Network size (nodes, edges, number of connected components) and betweenness of the GCN's geometry[c]] | G206,G208,G210 [Connectivity and length of the GCN's geometry[c]] |
| PC#8 | G208,G212,G302 [Network distance of the GCN's geometry[(c)] and accessibility[(a)] and clustering of the GCN's accessibility[a]] | G401,G409 [Commuting flows[b]] |

[a] The third layer ($G_3$) is considered as an aspect of the GCN's accessibility

[b] The fourth layer ($G_4$) is considered as an aspect of the GCN's functionality

[c] The first two layers ($G_1$ and $G_2$) are considered as aspects of the GCN's geometry

(Fig. 4b, d) to clustering coefficient ($G_{303}$) of $G_3$ and $G_4$ layer. In terms of interpretation, PC#6 appears positively related to road length and average degree of GCN's geometry, to betweenness centrality of the GCN's accessibility and functionality, and is negatively related to clustering of the GCN's accessibility and functionality.

The seventh principal component (PC#7) is positively related (Fig. 4b, c) to the number of nodes and edges ($G_{203}, G_{204}$), the number of connected components ($G_{209}$), and betweenness centrality ($G_{215}$) of the $G_2$ layer, to out-degree ($G_{403}$) of the $G_4$ layer, and is negatively related (Fig. 4b, d) to the minimum degree ($G_{206}$), node strength ($G_{208}$), and edge length ($G_{210}$) of the $G_2$ layer. In terms of interpretation, PC#7 includes information about network size (nodes, edges, number of connected components) and betweenness of GCN's geometry, and negative information about connectivity and length of the GCN's geometry. Finally, the eighth principal component (PC#8) is positively related (Fig. 4b, c) to node strength ($G_{208}, G_{302}$) of the $G_2$ and $G_3$ layers, to clustering ($G_{212}$) of the $G_2$ layer, and is negatively related (Fig. 4b, d) to the number of commuters and node strength ($G_{401}, G_{409}$) of $G_4$ layer. In terms of interpretation, PC#8 includes information about network distance of GCN's geometry and accessibility, about the clustering of the GCN's accessibility, and negative information about the functionality of GCN.

Overall, the PCA allowed decomposing the complexity of the multilayer GCN's topological information into principal components of distinct semiology defined by interlayer topological measures. However, based on the empirical nature of the analysis, these results cannot be considered as time-invariant and therefore they are not by default stable overtime for this network. Although the interpretation of these principal components is insightful within the spatiotemporal context of the available data, the high sunk-cost of road-transportation networks can provide a significantly higher endurance to the semiology of these principal components in comparison with other transport applications that are not infrastructure-dependent and thus are more flexible to rerouting or rescheduling (e.g. air-transport network constructed on flight timetables, trade courier networks, etc.). Within this context, the results of the analysis can shape a satisfactory picture of the medium-term or even long-term structure of the Greek commuting market, depending on the data restrictions and resolution. Besides, the proposed methodological framework is indifferent to the case study because it conceptualizes network topology as a composition of layers defined at different geographical resolutions and functionality of the same socioeconomic system, and thus within a context-free of spatiotemporal restrictions. This approach is novel to be used as a decomposition method in the analysis of multilayer networks, where, in the case of the multilayer GCN, showed distinct and sometimes competitive roles between network connectivity and accessibility, network length and connectivity, network length and clustering, and network distance and commuting flows.

## 5 Further Analysis

On further analysis, we examine the association amongst the principal components' grouping, the layer configuration of the GCN, and the grouping produced by an alternative established classification method. To do so, we further classify the available

GCN variables of Table 2 according to the hierarchical clustering method (Revelle 1979; Murtagh and Contreras 2012), and afterward, we apply a Pearson's chi-square test (McHugh 2013; Sharpe 2015) to detect any relevance amongst the diverse available groupings. At first, hierarchical clustering or cluster analysis (Revelle 1979; Murtagh and Contreras 2012) is an iterative method for grouping either cases or variables at different levels of hierarchy, resulting in a hierarchy of clusters. In this method, clustering is achieved under either an agglomerative or divisive algorithm, with a clustering criterion based on pairwise dissimilarity measures (chosen amongst a collection of available metrics), which quantify the linkage between observations (Norusis 2008). To perform this analysis, we apply to non-standardized data (to comply with a relevant specification of the PCA algorithm) an agglomeration algorithm capturing between-groups linkage through a Pearson's correlation dissimilarity measure. For the sake of comparisons, 8 clusters are extracted from this process to comply with the PCA outcome. Therefore, the results of the hierarchical clustering are shown in Fig. 5, where the configuration of the 8 clusters is shown in dashed-line colored windows. As it can be observed, there are 3 main clusters (HC#1, HC#2, and HC#7) including respectively 8, 9, and 6 items; there are 3 other clusters (HC#3, HC#5, and HC#6) including two items each; and other 2 (HC#4 and HC#8) including one item each. This composition configures an ascending rank-size ordering {9, 8, 6, 2, 2, 2, 1, 1}. Computing the correlation coefficients (Norusis 2008) with the concordant rank-size orderings, {7, 6, 5, 4, 3, 3, 2, 1}, and {8, 7, 5, 3, 2, 2, 2, 2}, of the groups of positive and negative PCA loadings (Table 4) yields respectively $r_{(HC,PCA+)} = 0.944$, and $r_{(HC,PCA-)} = 0.986$, which are all significant coefficients at the 0.01 level. These results imply that both (positive and negative) groupings generated by the PCA have a good structure of hierarchy in terms of their rank-size configuration. Amongst the PCA groupings, this of negative loadings seems to have a better structure of hierarchy, to the extent that a hierarchical structure is expressed by the hierarchical clustering result.

To better detect the hierarchical association between the available groupings, we further apply a Pearson's chi-square test for independence (McHugh 2013; Sharpe 2015). This method is a statistical hypothesis-testing used to examine whether there is a relationship and specifically whether there is a statistically significant association in the frequencies between the groups of two categorical variables. In particular, the chi-square test measures the discrepancy of the class frequencies between the examined variables comparatively to what would be expected in case they were unrelated. The null hypothesis ($H_o$) states that the class frequencies between the examined variables are not associated, whereas the alternative ($H_o$) states that they are. In terms of interpretation, a valid null hypothesis indicates that frequencies in one variable can be considered as irrelevantly distributed across the classes of the other variable, whereas a valid alternative hypothesis is that there is an association between the distribution of frequencies between these variables (Norusis 2008). To visualize the nature of the association, the class proportions are displayed with bar charts. To perform the Pearson's chi-square test (McHugh 2013; Sharpe 2015), we configure four categorical variables, (a) the layer grouping (LAYER), indicating the layer where each variable of Table 2 belongs to; (b) the hierarchical clustering grouping (CLUSTERS HIER.), indicating the hierarchical cluster (1–8) where each
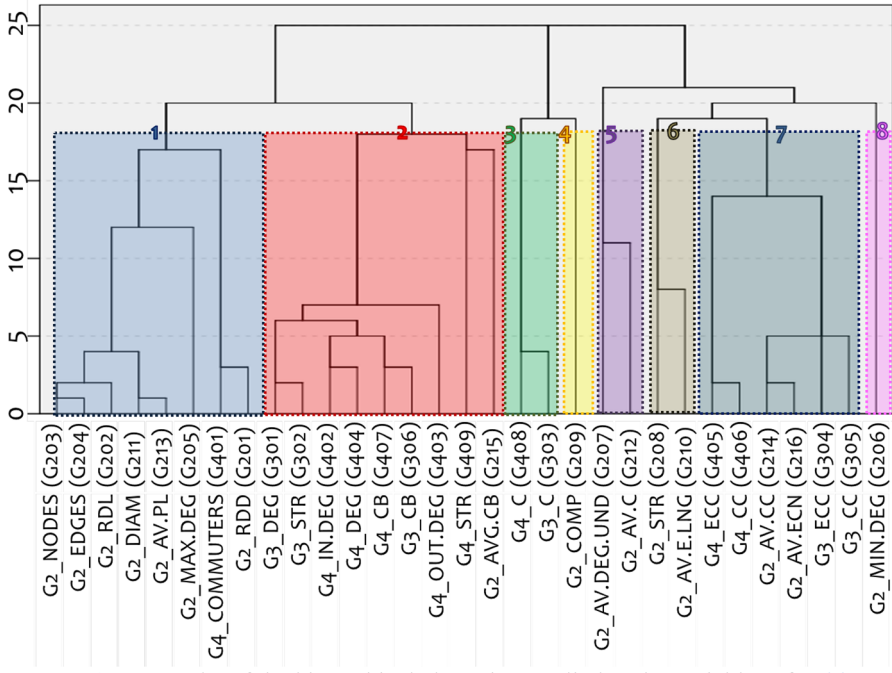
**Fig. 5** Results of the hierarchical clustering applied to the variables of Table 2

variable belongs to (Fig. 5); (c) the positive PCA grouping (PC POSITIVE), indicating the principal component (1–8) where each variable has the highest positive loading (Fig. 4b, c); and (d) the negative PCA grouping (PC NEGATIVE), indicating the principal component (1–8) where each variable has the most negative loading (Fig. 4b, d).

Within this context, Table 5 shows the results of the Pearson's chi-square test for independence, which are further illustrated in the bar charts of Fig. 6. First, as it can be observed, only variable PC NEGATIVE is significantly associated with LAYER (Table 5). In particular (Fig. 6a), components PC#2(–), PC#5(–), and PC#7(–) exclusively include variables from layer $G_2$; components PC#4(–), PC#8(–) exclusively include variables from layer $G_4$; while the other components include variables from pair layers (but not from all of them). For the pairs LAYER-HIER.CLUSTER and LAYER-PC POSITIVE, the group configuration (Fig. 6b, c) is not as clear as is for PC NEGATIVE. These results indicate that the negative loadings' grouping is relevant to the layer configuration, whereas the other two groupings are not. Next, variables PC POSITIVE and PC NEGATIVE are significantly associated with HIER.CLUSTERS (Table 5). In particular (Fig. 6d), component PC#4(+) is mainly included in cluster HC#1, PC#1(+) is mainly included in cluster HC#2, PC#5(+) is exclusively included in cluster HC#7, cluster HC#5 exclusively includes variables from component PC#2(+), and HC#8 exclusively includes variables from PC#7(+); whereas other hierarchical clusters are configured by 2 or maximum 3 PC(+)

**Table 5** Chi-square tests between the categorical variables of the layer, principal component, and hierarchical analysis grouping

| VARIABLES | Statistics | VARIABLES | | | |
|---|---|---|---|---|---|
| | | LAYER | CLUSTERS (HIER.) | PC POSITIVE | PC NEGATIVE |
| **LAYER** | Pearson Chi-Square[a] | | 19.211 | 9.852 | 34.408 |
| | Degrees of freedom | | 14 | 14 | 14 |
| | Asymp. Sig. (2-sided) | | 0.157 | 0.773 | **0.002**[c] |
| | Phi[b] | | 0.787 | 0.564 | 1.054 |
| | Approx. Sig | | 0.157 | 0.773 | **0.002** |
| | Cramer's V[b] | | 0.557 | 0.399 | 0.745 |
| | Approx. Sig | | 0.157 | 0.773 | **0.002** |
| | N of Valid Cases | | 31 | 31 | 31 |
| **CLUSTERS (HIER.)** | Pearson Chi-Square[a] | 19.211 | | 86.875 | 144.936 |
| | Degrees of freedom | 14 | | 49 | 49 |
| | Asymp. Sig. (2-sided) | 0.157 | | **0.001** | **0** |
| | Phi[b] | 0.787 | | 1.674 | 2.162 |
| | Approx. Sig | 0.157 | | **0.001** | **0** |
| | Cramer's V[b] | 0.557 | | 0.633 | 0.817 |
| | Approx. Sig | 0.157 | | **0.001** | **0** |
| | N of Valid Cases | 31 | | 31 | 31 |
| **PC POSITIVE** | Pearson Chi-Square[a] | 9.852 | 86.875 | | 70.034 |
| | Degrees of freedom | 14 | 49 | | 49 |
| | Asymp. Sig. (2-sided) | 0.773 | **0.001** | | **0.026** |
| | Phi[b] | 0.564 | 1.674 | | 1.503 |
| | Approx. Sig | 0.773 | **0.001** | | **0.026** |
| | Cramer's V[b] | 0.399 | 0.633 | | 0.568 |
| | Approx. Sig | 0.773 | **0.001** | | **0.026** |
| | N of Valid Cases | 31 | 31 | | 31 |
| **PC NEGATIVE** | Pearson Chi-Square[a] | 34.408 | 144.936 | 70.034 | |
| | Degrees of freedom | 14 | 49 | 49 | |
| | Asymp. Sig. (2-sided) | **0.002** | **0** | **0.026** | |
| | Phi[b] | 1.054 | 2.162 | 1.503 | |
| | Approx. Sig | **0.002** | **0** | **0.026** | |
| | Cramer's V[b] | 0.745 | 0.817 | 0.568 | |
| | Approx. Sig | **0.002** | **0** | **0.026** | |
| | N of Valid Cases | 31 | 31 | 31 | |

[a]Chi-Square Tests

[b]Symmetric Measures, Nominal by Nominal

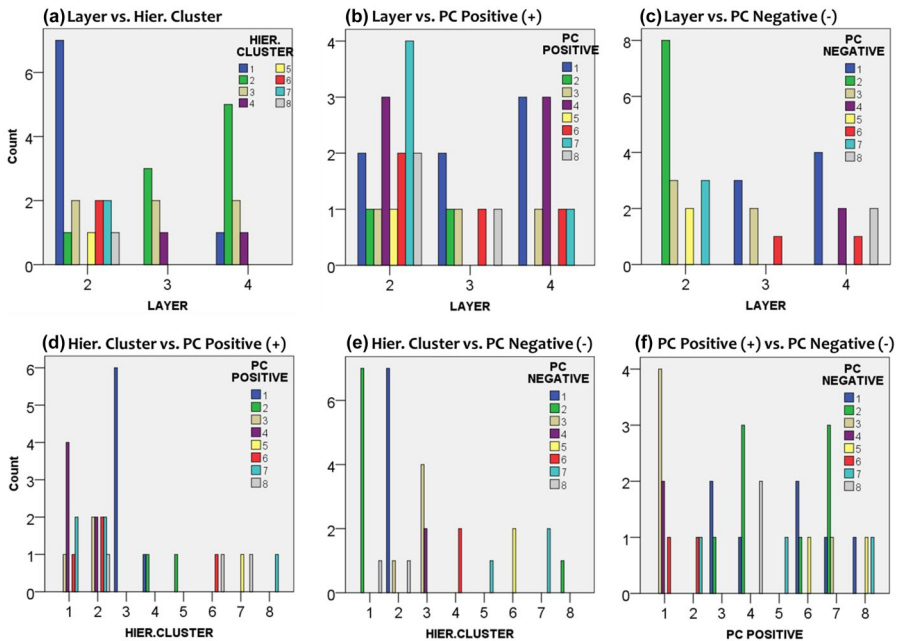Statistically significant results are shown in bold font

**Fig. 6** Bar charts illustrating class frequencies between the available categorical variables referring to the layer (G), principal component (PC), and hierarchical clustering (HC) grouping

components. Also (Fig. 6e), component PC#1(–) is exclusively included in cluster HC#2, PC#6(–) is identical to HC#4, and PC#5(–) is identical to HC#6, whereas other hierarchical clusters are configured by 2 or maximum 3 PC(–) components. Finally, according to a similar reading, we can observe that the configurations of the PC POSITIVE and PC NEGATIVE grouping are associated. These results highlight that the principal components groupings have a good structure of hierarchy.

Overall, this further analysis provided insights into the association amongst the principal components' grouping, the layer configuration, and the hierarchical structure of GCN, as it is captured by hierarchical clustering on the set of the available topological variables. The Pearson's chi-square test for independence showed that the PCA grouping, except for dimension reduction, may provide an effective clustering method that is first consistent with the topological hierarchy (both PCA groupings are significantly associated with hierarchical clustering) and secondly with the layer configuration (the PCA grouping of negative loadings is significantly associated with layer configuration) of the GCN.

# 6 Conclusions

Towards detecting norms or common features ruling the effect of spatial constraints on the configuration of network topology at different degrees of spatial or administrative aggregation, this paper modeled the interregional commuting in Greece into

a multilayer graph (GCN). The study built on methods of complex network analysis and dimension reduction to examine its topological properties across its layers and extract principal topological components of this multilayer model, by applying a Principal Component Analysis (PCA). The GCN layers were defined at different levels of geographical and administrative resolution, and different types of structure and functionality. The study conceptualized the available layers in terms of the utility of commuting flows and cost of spatial distance and performed an interlayer analysis to detect the important topological features across these different aspects of the interregional commuting market. The methodological framework first developed a multilayer graph of interregional commuting, next configured variables representing aspects of network topology, applied network analysis, and finally decomposed the total model into principal components expressing uncorrelated aspects of interregional commuting.

First, network analysis revealed that *the distance-defined topological features* (and particular distance strength, average, and total edge length, and distance-weighted average path length and network diameter) *and the network clustering of GCN increase either with the increase of the administrative level or the decrease of geographical resolution.* To the extent that distance is related to transportation cost, this finding underlined the importance of the administrative level choice in the strategic cost of spatial planning and transportation management. On the other hand, it pointed out the advantage of the interregional administrative setting for the development of peripheral connectivity, which is a major goal towards regional development and regional inequalities convergence. Next, the analysis revealed that *the path-based topological measures* (average path length and binary network diameter) *positively depend on network size.* To the extent that these "a-spatial" topological measures express aspects of network accessibility, this finding underlined the importance of network size in the establishment of accessibility, highlighting thereby that effective spatial planning and transportation management should build on simplicity. Finally, network analysis revealed that *network size is competitive to structural consistency and as it grows up it increases the network's tendency to divide into communities.* This finding also pointed out the importance of structural simplicity in spatial planning and transportation management, along with the advantage of the interregional administrative setting toward improving structural consistency in transport management configurations.

In terms of statistical mechanics, the *degree distributions* analysis revealed that all degree distributions across the GCN layers follow a bell-shaped pattern, implying that *connectivity in this network is submitted to spatial constraints and is mainly undertaken by the majority of nodes rather than by hubs.* A more distinguishable hierarchical structure detected in the layer related to the GCN interregional accessibility, revealed the advantage of this administrative setting towards a more effective planning building on centrality and structures of hierarchy. The correlation analysis applied to incoming and outgoing node strengths developed profiles of commuting performance (configuring the *isolated*, *exporting*, *importing*, and *interactive* labor markets) based on the strength directionality of GCN. This approach revealed that some profiles share common topological attributes across all layers, addressing avenues for further research towards the topological configuration of the strength

directionality of commuting, which expresses either labor or housing attractiveness in the interregional commuting market.

Finally, the PCA resulted in dimension reduction of the GCN topology, from 31 available topological items to 8 principal components. Each component enjoyed a distinct and singular semiology defined by the physical meanings of the included items, *providing thereby interesting insights into the relationship between network connectivity and accessibility, network length and connectivity, network length and clustering, and network distance and network functionality (commuting flows) of the GCN*. On further analysis, the association was examined amongst the grouping produced by the PCA, the layer configuration of GCN, and a hierarchical clustering applied to the set of the available topological variables. The results showed that the PCA grouping was first consistent with the topological hierarchy and secondly with the layer configuration of the GCN, highlighting that the proposed method can also be seen as an effective clustering technique, conservative to the norms describing the total system.

Provided that it was applied to the case of interregional commuting in Greece, the results of the analysis are not universal and thus they do not by default describe other realities. For instance, any analogies related to comparisons of topological features between layers, fitting specializations of the degree distributions, the commuting profiles based on strength directionality correlations, and semiology of the principal components resulted from the analysis, are all related to the case study and the modeling specifications and may vary when computed in other applications, although common patterns may appear due to the common setting provided by the spatial constraints to network connectivity. However, in methodological terms, the conceptual and methodological framework of this study enjoys the universality of its component methods (multilayer modeling, network analysis, correlation, and principal component analysis) and thus the effectiveness and consistency of the overall analysis are indifferent to the case of GCN. One suggestion that the analysis offers to spatial planning and transportation management is *incorporating topological aspects in the description and modeling of transportation systems*, which is an ongoing process not yet fully integrated into spatial planning and transportation research. This is because network topology conceptualizes systems of spatial interaction beyond their geometry, and thus can provide quantitative measurements of relational linkages and features that can facilitate a deeper systems' knowledge and equip to more effective planning. Particularly, being compatible with the multilayer setting of spatial planning and transportation management, the multilayer graph modeling can facilitate quantitative decomposition and interlayer analysis providing profound knowledge of transportation systems and systems of spatial interaction. Within this context, *another suggestion* of this study *toward ensuring simplicity in spatial and transportation planning is the reduction of (horizontal) complexity within layers and the increase of interlayer (vertical) configuration in the modeling of transportation systems*. Also, to the extent that larger geographical scales are related to larger distances (expressing cost), the finding that peripheral connectivity in GCN is better developed at larger geographical scales tells us that *the development of peripheral connectivity is a relatively more costly goal of transportation planning*. Taking into account that the interregional administrative setting was found favorable to the

development of peripheral connectivity, *this study also suggests giving more effort to spatial planning and transportation management at the interregional administrative level*. This approach can facilitate processes towards regional development and regional inequalities convergence.

Overall, the analysis provided insights into the spatial, topological, and functional configuration of interregional commuting in Greece, at different levels of geographical resolution, administrative scale, and functionality; conceptualized network topology in a novel multivariable framework; highlighted the importance of spatial constraints in this multilayer setting of Greek commuting; underlined the importance of geographical scale and resolution in transportation modeling; detected the effectiveness of the interregional administrative scale on the topology of transportation networks; and proposed a methodological framework for dimension reduction in the topology of multilayer spatial networks into principal components. This paper promotes multidisciplinary conceptualization and suggests incorporating topological aspects in the description and modeling of transportation systems in spatial planning, transportation, and regional research.

## Declarations

**Conflict of Interest** The authors have no competing interests to declare that are relevant to the content of this article.

## References

Adler N, Golany B (2001) Evaluation of deregulated airline networks using data envelopment analysis combined with principal component analysis with an application to Western Europe. Eur J Oper Res 132(2):260–273

Albert R, Barabási AL (2002) Statistical mechanics of complex networks. Rev Mod Phys 74(1):47

Albrechts L (2006) Bridge the gap: From spatial planning to strategic projects. Eur Plan Stud 14(10):1487–1500

Alessandretti L, Karsail M, Gauvin L (2016) User-based representation of time-resolved multimodal public transportation networks. R Soc Open Sci 3:160156

Aleta A, Meloni S, Moreno Y (2016) A Multilayer perspective for the analysis of urban transportation systems. Sci Rep 7:44359

Barabasi A-L (2013) Network science. Philos Trans Royal Soc Lond A Math Phys Eng Sci 371(198):20120375

Barthelemy M (2011) Spatial networks. Phys Rep 499(1–3):1–101

Bastian M, Heymann S, Jacomy M (2009) Gephi: an open source software for exploring and manipulating networks. In: Proceedings of the international AAAI conference on web and social media, vol 3, no 1. pp 361–362

Bham GH, Manepalli UR, Samaranayke VA (2019) A composite rank measure based on principal component analysis for hotspot identification on highways. J Transp Saf Secur 11(3):225–242

Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. J Stat Mech: Theory Exp 2008(10):P10008

Boccaletti S, Bianconi G, Criado R, del Genio CI, Gomez-Gardenes J, Romance M, Sendina-Nadal I, Wang Z, Zanin M (2014) The structure and dynamics of multilayer networks. Phys Rep 544:1–122

Brakman S, Garretsen H, Gorter J, van der Horst A, Schramm M (2005) New economic geography, empirics, and regional policy, vol 56. Hague, CPB Netherlands Bureau for Economic Policy Analysis

Brandes U, Robins G, McCranie A, Wasserman S (2013) What is Network Science? Network Science 1:1–15

Bwire H, Zengo E (2020) Comparison of efficiency between public and private transport modes using excess commuting: An experience in Dar es Salaam. J Transp Geogr 82:102616

Capello R (2016) Regional Economics, 2nd edn. Routledge, London

Cardillo A, Zanin M, Gomez-Gardenes J, Romance M, Garcia del Amo AJ, Boccaletti S (2013) Modeling the multi-layer nature of the European air transport network: resilience and passengers rescheduling under random failures. Eur Phys J Spec Top 215:23–33

Cavaco C, Costa JP (2020) Administrative organisation and spatial planning in portugal: a push towards soft planning spaces in Europe?. In: Lingua V, Balz V (eds) Shaping regional futures. Springer, Cham. https://doi.org/10.1007/978-3-030-23573-4_7

Chen X, Wei Z, Gao L (2011) Evaluation of road safety based on principal component analysis. Proceedings of the 30th Chinese Control Conference, Yantai, China, p. 5601–5605

Chung Y, Song TJ (2018) Safety analysis of motorcycle crashes in Seoul metropolitan area, South Korea: an application of nonlinear optimal scaling methods. Int J Environ Res Public Health 15(12):2702

Clinch JP, O'Neill E (2009) Applying spatial economics to national spatial planning. Reg Stud 43(2):157–178

Crescenzi R, Rodriguez-Pose A (2011) Reconciling top-down and bottom-up development policies. Environ Plan A 43(4):773–780

Davoudi S (2003) European briefing: polycentricity in European spatial planning: from an analytical tool to a normative agenda. Eur Plan Stud 11(8):979–999

De Domenico M, Sole-Ribalta A, Cozzo E, Kivela M, Moreno Y, Porter MA, Gomez S, Arenas A (2013) Mathematical Formulation of Multilayer Networks. Phys Rev X 3:041022

De Montis A, Caschili S, Chessa A (2011) Spatial complex network analysis and accessibility indicators: The case of Municipal Commuting in Sardinia, Italy. Eur J Transp Infrastruct Res 11(4):405–419

Debrie J (2010) From colonization to national territories in continental West Africa: the historical geography of a transport infrastructure network. J Transp Geogr 18(2):292–300

Diestel R (2005) Graph theory, 3rd edn. Springer, Heidelberg

Djukic T, Flotterod G, Van Lint H, Hoogendoorn S (2012) Efficient real time OD matrix estimation based on Principal Component Analysis. In 2012 15th Int IEEE Conf Intel Transp Syst IEEE, p 115–121

Drobne S, Bogataj M, Lisec A (2012) Dynamics and local policy in labour commuting. Bus Syst Res 3(2):14–26

Ducruet C (2013) Network diversity and maritime flows. J Transp Geogr 30:77–88

Ducruet C (2017) Multilayer dynamics of complex spatial flows: The case of global maritime flows (1977–2008). J Transp Geogr 60:47–58

Ducruet C, Beauguitte L (2014) Spatial science and network science: review and outcomes of a complex relationship. Netw Spat Econ 14:297–316

Ducruet C, Ietri D, Rozenblat C (2011) Cities in worldwide air and sea flows: A multiple networks analysis. Eur J Geogr 528:23603

Easley D, Kleinberg J (2010) Networks, crowds, and markets: Reasoning about a highly connected world. Cambridge University Press, Cambridge, UK

Evans G, Wener R, Phillips D (2002) The Morning Rush Hour: Predictability and Commuter Stress. Environ Behav 34(4):521–530

Ewing R, Pendall R, Chen D (2003) Measuring sprawl and its transportation impacts. Transp Res Rec 1831(1):175–183

Fortunato S (2010) Community detection in graphs. Phys Rep 486(3–5):75–174

Fujita M (2005) Spatial Economics. Edward Elgar Publishing, UK

Gallotti R, Barthelemy M (2014) Anatomy and efficiency of urban multimodal mobility. Sci Rep 4(1):1–9

Gallotti R, Barthelemy M (2015) The multilayer temporal network of public transport in Great Britain. Scientific Data 2:140056

Garretsen H, McCann P, Martin R, Tyler P (2013) The future of regional policy. Camb J Reg Econ Soc 6(2):179–186

Ghavasieh A, De Domenico M (2020) Enhancing transport properties in interconnected systems without altering their structure. Phys Rev Res 2:013155

Google Maps (2019) Google Mapping Services, available at the www.google.gr/maps?hl=el. Accessed 29 Aug 2019

Green M, Meyer S (1997) An overview of commuting in Canada: With special emphasis on rural commuting and employment. J Rural Stud 13(2):163–175

Hamilton B (1989) Wasteful commuting again. J Polit Econ 97(6):1497–1504

Hamilton B, Roell A (1982) Wasteful commuting. J Polit Econ 90(5):1035–1053

Hausdorff F (1957) Set theory. Chelsea Publications Company, New York

Hellenic Land and Mapping Organization - OKXE (2005) Road Network DRPS (Department of Road Project Studies), available at the http://www1.okxe.gr/geonetwork/srv/en/google.kml?uuid=19226 ad1-1297-4041-8cea-3aa45a2b5f4e&layers=rdndmeo. Accessed 07 Mar 2015

Hellenic Statistical Service – ELSTAT (2011) Results of the Census of Population-Habitat 2011 referring to the permanent population of Greece, Newspaper of Government (ΦΕΚ), Second Issue (T-B), Number 3465, 28 December 2012

Horak J, Ivan I, Tesla J (2014) Improved gravity models of commuting conditions: a Czech case study, Proceedings of the Second International Conference on Traffic and Transport Engineering, Belgrade, Serbia, November 27 – 28, 2014, pp. 171 – 178

Jeon JH, Kho SY, Park JJ, Kim DK (2012) Effects of spatial aggregation level on an urban transportation planning model. KSCE J Civ Eng 16(5):835–844

Jiang B, Claramunt C (2004) Topological analysis of urban street-networks. Envioron Plann B 31:151–162

Kanaroglou PS, Higgins CD, Chowdhury TA (2015) Excess commuting: a critical review and comparative analysis of concepts, indices, and policy implications. J Transp Geogr 44:13–23

Kansky KJ (1963) Structure of transportation networks: Relationship between network geometry and regional characteristics. University of Chicago Press, Chicago

Kivela M, Arenas A, Barthelemy M, Gleeson J, Moreno Y, Porter MA (2014) Multilayer networks. J Complex Netw 2:203–271

Koschutzki D, Lehmann K, Peeters L, Richter S (2005) Centrality Indices. In: Brandes U, Erlebach T (eds) Network Analysis. Springer-Verlag Publications, Berlin, Germany, pp 16–61

Koslowsky M, Kluger AN, Reich M (1995) Behavioral, emotional, and attitudinal effects of commuting. In: Commuting stress. The Plenum series on stress and coping. Springer, Boston, MA, pp 79–106

Kulmer V, Koland O, Steininger KW, Fürst B, Käfer A (2014) The interaction of spatial planning and transport policy: A regional perspective on sprawl. J Transp Land Use 7(1):57–77

Li X, Mou Y, Wang H, Yin C, He Q (2018) How does polycentric urban form affect urban commuting? Quantitative measurement using geographical big data of 100 cities in China. Sustainability 10(12):4566

Liu Y, Nie YM (2011) Morning commute problem considering route choice, user heterogeneity and alternative system optima. Trans Res B: Methodol 45(4):619–642

Ma L, Ye R (2019) Does daily commuting behavior matter to employee productivity? J Transp Geogr 76:130–141

Marshall S, Gil J, Kropf K, Tomko M, Figueiredo L (2018) Street-network studies: from networks to models and their representations. Netw Spat Econ 18(3):735–749

McArthur DP, Kleppe G, Thorsen I, Uboe J (2011) The spatial transferability of parameters in a gravity model of commuting flows. J Transp Geogr 19:596–605

McHugh ML (2013) The chi-square test of independence. Biochemia Medica 23(2):143–149

Murphy E (2009) Excess commuting and modal choice. Trans Res A: Policy Pract 43(8):735–743

Murtagh F, Contreras P (2012) Algorithms for hierarchical clustering: an overview. Wiley Interdiscip Rev Data Min Knowl Discov 2(1):86–97

Nagendra SS, Khare M (2003) Principal component analysis of urban traffic characteristics and meteorological data. Transp Res D Transp Environ 8(4):285–297

Newman M (2010) Networks: An introduction, 1st edn. Oxford, Oxford Academic

Norusis MJ (2008) SPSS statistics 17.0: Statistical procedures companion. New Jersey, Prentice Hall/Pearson Education

Ottaviano G (2003) Regional policy in the global economy: Insights from New Economic Geography. Reg Stud 37(6–7):665–673

Ozbay K, Bartin B, Yanmaz-Tuzel O, Berechman J (2007) Alternative methods for estimating full marginal costs of highway transportation. Trans Res Part A: Policy Pract 41(8):768–786

Park YS, Egilmez G, Kucukvar M (2015) A novel life cycle-based principal component analysis framework for eco-efficiency analysis: case of the United States manufacturing and transportation nexus. J Clean Prod 92:327–342

Porta S, Crucitti P, Latora V (2006a) The network analysis of urban streets: a primal approach. Environ Plann B Plann Des 33(5):705–725

Polyzos S (2019) Regional development, 2nd edn. Kritiki Publications, Athens [in Greek]

Porta S, Crucitti P, Latora V (2006b) The network analysis of urban streets: A dual approach. Physica A 369(2):853–866

Ren Y, Ercsey-Ravasz M, Wang P, González MC, Toroczkai Z (2014) Predicting commuter flows in spatial networks using a radiation model based on temporal ranges. Nat Commun 5(1):1–9

Revelle W (1979) Hierarchical cluster analysis and the internal structure of tests. Multivar Behav Res 14(1):57–74

Rodrigue JP, Comtois C, Slack B (2013) The Geography of Transport Systems. Routledge Publications, New York

Rozenblat C, Melancon G (eds) (2013) Methods for Multilevel Analysis and Visualisation of Geographical Networks, vol. 11. Dordrecht, Springer Science+Business Media

Sezhian MV, Muralidharan C, Nambirajan T, Deshmukh SG (2011) Ranking of a public sector passenger bus transport company using principal component analysis: a case study. Manag Res Pract 3(1):62–71

Sharpe D (2015) Chi-square test is statistically significant: Now what? Pract Assess Res Eval 20(1):8

Stutzer A, Frey B (2008) Stress that doesn't pay: The Commuting Paradox. Scand J Econ 110(2):339–366

Tan R, He Q, Zhou K, Xie P (2019) The effect of new metro stations on local land use and housing prices: The case of Wuhan, China. J Transp Geogr 79:102488

Tang Y, Zhong D, Zha X, Na L (2018) Principal Component Analysis of Fatal Traffic Accidents Based on Vehicle Condition Factors. 11th International Conference on Intelligent Computation Technology and Automation (ICICTA), Changsha, China, p 315–317. https://doi.org/10.1109/ICICTA.2018.00078

Tsiotas D (2019) Detecting different topologies immanent in scale-free networks with the same degree distribution. Proc Natl Acad Sci 116(14):6701–6706

Tsiotas D, Axelis N, Polyzos S (2021) A methodological framework for defining city dipoles in urban systems based on a functional attribute. Cities 119:103387

Tsiotas D, Ducruet C (2021) Measuring the effect of distance on the network topology of the Global Container Shipping Network. Sci Rep 11:21250

Tsiotas D, Polyzos S (2015a) Analyzing the Maritime Transportation System in Greece: a Complex Network approach. Netw Spat Econ 15(4):981–1010

Tsiotas D, Polyzos S (2015b) Decomposing multilayer transportation networks using complex network analysis: A case study for the Greek aviation network. J Complex Netw 3(4):642–670. https://doi.org/10.1093/comnet/cnv003

Tsiotas D, Polyzos S (2018) Effects in the network's topology due to node aggregation: empirical evidence from the maritime transportation network in Greece. Physica A 491C:71–88

Van Ommeren J, Fosgerau M (2009) Workers' marginal costs of commuting. J Urban Econ 65(1):38–47

Van Ommeren J, Rietveld P (2005) The commuting time paradox. J Urban Econ 58(3):437–454

Vigar G (2009) Towards an integrated spatial planning? Eur Plan Stud 17(11):1571–1590

Walpole RE, Myers RH, Myers SL, Ye K (2012) Probability & Statistics for Engineers & Scientists, 9th edn. Prentice Hall Publications, New York

Wegener M (2001) New spatial planning models. Int J Appl Earth Obs Geoinf 3(3):224–237

Wilson A (1967) A statistical theory of spatial distribution models. Transp Res 1:253–269

Wold S, Esbensen K, Geladi P (1987) Principal component analysis. Chemom Intell Lab Syst 2(1–3), 37–52

Zhao J, Fan W, Zhai X (2020) Identification of land-use characteristics using bicycle sharing data: A deep learning approach. J Transp Geogr 82:102562

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.