522

Markov-Chain based centralities and Space Syntax' Angular Analysis:

an initial overview and application.

DIEGO ALTAFINI, FEDERICO POLONI, BEATRICE MEINI, DARIO BINI & VALERIO CUTINI,

UNIVERSITY OF PISA, ITALY

ABSTRACT

Centrality measures of Integration and Choice have performed a crucial role for Space Syntax in depicting complex relations among form, function, and movement within cities. However, while still relevant, those measures are unable to address certain innate network properties regarding the *relative importance* of certain road elements, essential for urban analyses focused on road-network resilience. The overreliance on Integration and Choice metrics to explain urban phenomena left several configurational patterns derived from connectivity rather unaddressed by Space Syntax and currently constitutes the methodology's main limitation. With those points in consideration, this paper proposes an initial overview regarding the adaptation of *Markov-based* centrality measures to the Space Syntax framework and graph representation. These measures, often computable only in primal graphs, are adapted to the Road-Centre Line representation, in a first approach that aims to further integrate them into the *Angular Analysis'* framework. We use the measures of Normalized PageRank Centrality and Normalized Kemeny-based centrality that estimate the connective *relative importance* of individual road-elements within the system. These measures are based on the notions of *strong-ties* and *weak-ties,* both well-known concepts in social networks; *Weak-ties* are important to establish bridges among interconnected communities of *strong-tied* individuals. In the urban configuration, *weak-ties* give information about crucial bridges among spaces characterized by *strong-ties,* areas that possess a high number of interconnected road elements – common pattern found in urban settlements. Results indicate that adapting Markov-based centralities to Space Syntax is feasible and maintains a configurational and spatial sense, hence it introduces new dimensions to be evaluated in urban-regional analysis.

KEYWORDS

Network Analysis, Markov-Chains, Space Syntax, Kemeny-based Centrality, Pagerank centrality

1 INTRODUCTION

Since their introduction in *Axial Analysis* (Hillier & Hanson, 1984) and their further development throughout the conception of *Angular Analysis* (Turner, 2001; 2005; 2007), the *syntactic* measures of *Integration* and *Choice* constituted a solid base for the *Space Syntax'* understanding of how movement patterns were distributed within urban and regional road-circulation networks.

As stated by Hillier (2012), *Integration* and *Choice* are based on a patchwork of mathematical concepts that were adapted from studies on social networks. *Integration*, represents the patterns of *to-movement*, or the *relative accessibility* derived from proximity among elements and originated from *closeness centrality's* notion of *farness* (Bavelas, 1950). *Choice*, instead, demonstrates the patterns of *through-movement* – or the *preferential routes of movement* – across the urban system, a notion that derives from *betweenness centrality* (Freeman, 1977; 1978). These *syntactic* centrality measures are related to the systems' *Total Depth* (which is a metric of system size), and these differences in depth can be spatialized to give a notion of the hierarchical *importance* of each road element to the system. The logic of *depth* is the basis of *Space Syntax'* configurational analyses, as it serves to estimate the overall differences and tendencies of human movement.

Although *closeness* and *betweenness* centralities are suitable to determine, respectively, the relations of *proximity* and *preferential routes/control* on road-circulation networks, these measures are somewhat limited in their explanation of *positional importance* – or how relevant a particular road-element or connection is to the system integrity. Even though these general centrality measures can be interpreted as hierarchical, and this established order may indicate the road elements' overall *importance* within the network, they do not have an inherent component that is capable to denote if a road-element has *strong* or *weak ties* – connections – towards others or groups of others, neither how important those connections are.

Hence, *syntactic* measures, in general, reveal no information about a road-element's *relative importance* to the network structure. In networks, these logics of *relative importance* can be observed through the interpretation of *Markov-processes* (also known as *Markov-chains),* which are stochastic processes that describe a sequence of events where their probability to happen is dependent only on the knowledge of the system's present state, and not on past events (Markov, 1971). *Markov-processes* can be analysed through linear algebra and used to compute centralities measures that can consider the *relative importance* of road elements, metrics that, currently, are inexistent on *Space Syntax'* methodology.

Considering this gap in *Space Syntax* configurational models. we propose a methodological adaptation of two *Markov-chain*-based centrality measures – the PageRank Centrality and the novel Kemeny-Based Centrality – in an initial approximation to principles of the *Space Syntax'* Angular Segmented Analysis (ASA), rendering these measures suited for Road-centre line representation. Our objective in this first approach is not to make comparisons between *Markov* and *Syntactic* measures, nor to define which is the most efficient, but to construct a methodological bridge between the spatial representations of intrinsically different network metrics. Therefore, we discuss the methods to convert the results from primal to dual graphs, the mathematical definitions of each *Markov-based* measure, as well as spatialize and interpret their outcomes in terms of spatial configuration, to furthermore address the concepts of *strong* and *weak ties* within a territorial context.

2 A REVIEW OF LITERATURE AND CONCEPTS

2.1 Strong Ties and Weak Ties

First addressed in mathematical sociology's social network analysis, the notions about the *strength of ties* were devised to explain the different natures of interpersonal connections among individuals, and their *relative importance* for social groups' network structure. Defined by their *degree* of connection, interpersonal ties establish the manner and the pathways on which information moves across social groups and the main linkages that ensure these relationships. Granovetter, in his essay: *The Strength of Weak Ties* (1973) delved into the characteristics of the micro-level interactions among individuals – what differentiates close friendships from mere acquaintances – to understand how those connections shaped the macro-level relationships and community organization. This exercise had the objective to interpret and discuss how different kinds of ties among individuals impacted how information transmission occurred, and, beyond that, how novel information was shared across the social groups that composed the network.

Granovetter (1973, p.1361) conceptualizes that the *strength* of an interpersonal tie is defined by the amount of time, emotional intensity, and the reciprocal actions that exist between individuals. The presence or absence of such aspects characterizes a relationship as being among "familiar strangers", acquaintances, or friends; in that regard, he recognizes that there are three main varieties of interpersonal ties, that can be divided: *absent, strong* and *weak* (Figure 1):

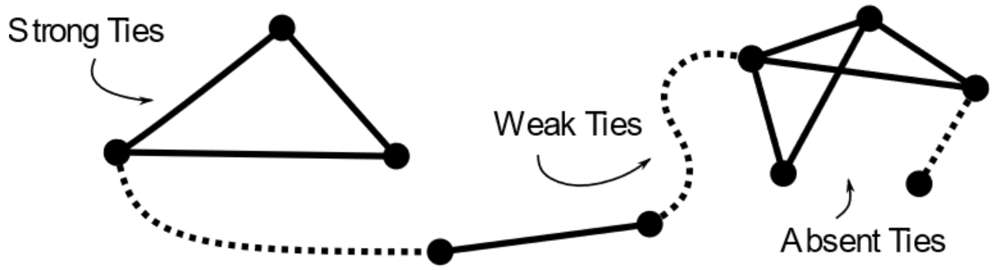


Figure 1: Visualization of the *Absent, Strong* and *Weak ties* dynamics within a social network.

- *Absent ties* are relationships without substantial significance, defined as infrequent interactions among "familiar strangers" that may or may not be reciprocal – as in example: a customer that seldom returns to the same vendor to make an acquisition, or the act of greeting people that live on the same street or apartment block.
- *Strong ties* are characterized, instead, by those relationships that are frequent and reciprocal, between individuals that have a significative degree of mutual knowledge, commitment, and often, emotional attachments with each other – as in example: marriages, sibling relations, or childhood friendships developed among people that attended the same school.
- *Weak ties* can be defined as an intermediate form of interpersonal tie that exists between *absent ties* and *strong ties*, characterized by occasional interactions among individuals that are, at least, acquainted, thus, have a certain degree of mutual knowledge about each other – as in example: friends of friends, classmates, or co-workers.

While concepts of *absent, strong,* and *weak ties* were refined in Granovetter's later works (1973; 1983), his doctoral thesis (1970) provided empirical data that endorsed these notions. Interviews among individuals that changed jobs through personal contact revealed that 16.7% of them had seen their contact person often, while 55.6% had only occasional contacts with that person (1970; 1973). These results unveiled that information about job positions was more often transmitted through acquaintances, instead of through close friends. Granovetter (1970) then concluded that, while *strong ties* might be more prone to disseminate known information inside the established social groups, *weak ties* among individuals that pertain to different circles tend to disseminate novel information across various groups within the social network, as acquaintances serve as means to bridge information beyond established circles. In that aspect, Granovetter (1973) justifies that the *weak ties* have a greater weight in terms of strengthening the underlying social network structure as they, not only give access to broader sources of information but also to more efficient channels to diffuse the information.

From a mathematical standpoint, Granovetter's *weak tie hypothesis* (1973) states that if, in a social group, three random individuals are selected (for example: A, B and C), and A has *strong ties* with both B and C, then, in this case, according to probability arguments (Newcomb,1961; Davis,1970), interpersonal ties among individuals B and C are also expected to be present, whether being *strong* or *weak.* If ties between B and C are *absent* in the group, a situation defined by Granovetter (1973) as the *Forbidden Triad,* due to its unlikeness to happen, occurs.
At this state, a bridge, here defined as an arch in the network that provides the sole path between two points can be formed only between the points AB and AC; in these conditions, there are single separated routes on which information can be diffused, as information must always pass-through A to arrive from or to either B or C, which is less efficient in terms of 90ological distance due to the missing connections. Granovetter (1973) then assumes that the information flows tend to be proportional to the number of connections between individuals *(degree)*, and inversely proportional to the connection paths' length. Within this setting, bridges assume a crucial role in the network structure, as their absence creates, on average, longer paths that

information must follow to move across the network – a paragon that can be made also to movement. Hence, removing bridges – thus, eliminating the *weak ties* of a system – tends to damage the network structure in terms of efficiency than the removal of a *strong tied* element. W*eak tied* elements also have greater importance to the network structure from the resilience standpoint, as their removal often leads to the decrease of the system's overall connectivity and path redundancy, which in turn, increases the general network fragility (Figure 2).

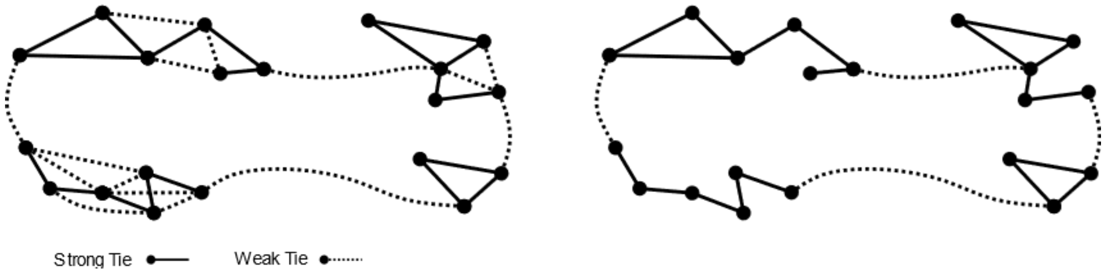


Figure 2: Network organization examples and the impact of *weak ties* removal in the system.

Transposing *The Strength of Ties* (1973) concepts from social networks to other kinds of networks is a dauting task. Non-social networks often do not have direct means to attribute weights between the nodes' linkages that characterize the properties found in the interpersonal relationships, determine their degree of importance, and, in turn, enable to differentiate a *strong* from a *weak tie* in the network. In that regard, correspondences must be found, not on network elements' data, but through the properties in the underlying mathematical structure that defines the network graph.

In recent years, the study of complex networks has become an important topic, stemming from an interdisciplinary field known as *network theory*. Estrada (2012) offers an overview of the recent developments in *network theory*: that range from social networks analysis to protein folding. Estrada (2012) also mentions that *Spectral graph theory* (Biggs, 1974; Cvetković et al. 1980) is a fundamental field of the applied *network theories*, which relies on linear algebra and matrix theory as its main tools for exploring the characteristics and evaluating the importance of nodes or edges – and their underlying properties – in a network graph. *Spectral graph theory* decomposes network graphs into matrix representations, where non-zero values represent connections amongst elements – that tend to be few and far between –, and zero values represent the more prevalent inexistence of connections between elements. For example, in networks with many elements and relatively low connectivity, such as road-circulation networks at urban and regional scales, each road element tends to border only a limited number of other road-elements. Exploring this property of sparsity is crucial to obtaining low computation times, and it is one of the main instruments used in modern numerical linear algebra to allow for more efficient implementations of primitive operations such as solving linear systems associated with networks. fundamental to estimate centrality metrics based on graph properties. In modern network theory, these kinds of centrality metrics have their roots in the analysis of random processes such as the *Markov-chains* and use techniques and concepts from applied probability. Albeit not a novelty, as *markov-based* centrality measures that analyse the properties of *graph's spectra* have been proposed in the literature by several authors, both oriented to general and to road-networks, although most of them are aspatial approaches (Pinski & Narin, 1976; Brin & Page, 1998; Jiang, 2009; Bonacich, 2007; Crisostomi, Kirkland & Shorten, 2011; Estrada & Benzi, 2014; Agryzkov et.al, 2019; Berkhout & Heidergott, 2019).

Adapting these *Markov-based* centrality measures to spatial networks, such as those that represent road-infrastructure elements, may introduce some issues. Software able to estimate *Markov-based* centrality measures operates using primal graph representations, where road elements would be set as nodes. Common datasets that represent urban roads, such as the Road-Centre Line (RCL), are instead constructed using a dual graph network representation, where road-elements assume the position of network edges. This requires extensive database conversions to calculate *Markov-based* metrics and address the *relative importance* of nodes. Due to such issues, the concepts of *strong* and *weak ties,* which can explain aspects about road-elements' *relative importance* in both urban and regional road-circulation networks, remain relatively unexplored. Still, these parameters can give important information on the overall resilience patterns of a road system and can provide insight into the patterns of economic activities' placement, therefore, is useful for risk assessment and decision-making prospects.

2.2 Methods to highlight strong ties and weak ties in road-circulation networks: PageRank Centrality (PRC) and Kemeny-Based-Centrality (KBC)

PageRank (Brin & Page, 1998; Page et al., 1999) was not the first ranking algorithm, being a result of the advancements in *scientometrics* (Pinski & Narin, 1976; Garfield, 1995) and previous experiments in the construction of search engines (Marchiori, 1997; Li, 2002). However, its simplicity and the short computation time-lapses of this algorithm in ranking large groups of elements, together with its validation through the Google Web Search Engine, have consolidated it as an important network analysis method. The PageRank algorithm (Page et al. 1999) computes a centrality score μ_j for each page j in a set of n web pages by considering a certain Markov process that models the behavior of a "random surfer" following links between them. From a mathematical standpoint, consider the matrix P whose element P_{ij} is equal to $1/N_i$ if there is a link from the page i to the page j and 0 otherwise; here N_i is the number of outgoing links from page i. P is called *transition matrix* of the *Markov-process* associated with the movement of the random surfer. Given a parameter α with $0 < \alpha < 1$ (typically $\alpha = 0.85$), the Pagerank vector is the solution of the system of linear equations

$$\mu_j = \frac{1-\alpha}{n} + \sum_{i=1}^{n} \alpha \mu_i P_{ij}\,, \qquad j = 1,2,\dots n. \tag{1}$$

These equations admit two different interpretations (Page et al. 1999):

- μ_j is the amount of time spent on page j by a person who surfs the internet according to the following rule: when they are on a given page, with probability α they follow a random link on the page, and with probability $1-\alpha$ they 'teleport' to another page chosen at random.

- μ_j is an importance score; each page receives a base score (1- α)/n, plus α times an equal share $1/N_i$ of the score μ_i of each page linking to it (see Figure 3 for a visualization of this score propagation).

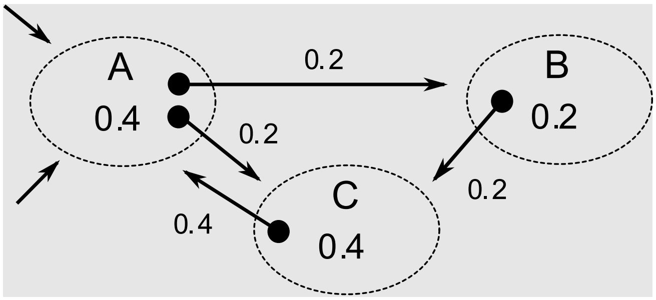


Figure 3: Score propagation in the second interpretation of PageRank: page A "propagates" equal shares of its score to B and C.

The ranking provided by sorting the elements by their PageRank value can be used as a centrality measure to indicate the *relative importance* of an element within the network.
Special provisions are needed to treat pages with $N_i = 0$, but we will not need them in the following since this case does not appear in our model. Other PageRank algorithm versions also include weights, as it is assumed that the propagation should be related to the inlink connectivity between the elements, therefore, being proportional to each link *importance* (Xing, Ghorbani, 2004).

A form of weighted PageRank centrality (PRC) was applied to road-circulation networks – with an axial line representation – by Jiang (2009). His objective was to rank the urban spaces and predict where areas with high localized movement, given the overall system configuration, are located. In a spatial context, his PageRank application considers the importance of a road element for local movement as derived from its connectivity to nearby road elements – and how well-connected those other road elements are. Therefore, the movement distribution probabilities in a system are denoted by the *relative importance* of a road element in relation to the others.
Although Jiang (2009) associated the PRC to the *space syntax'* integration measure (Hillier, 2006), to demonstrate that the *Markov-based* relations have better correspondence with human movement at a local scale, when compared to the *closeness-derived* measure, we argue, instead, that those algorithms are not directly comparable, as they represent fundamentally distinct spatial logics.

Integration aims to describe the *to-movement* potential of a road element in relation to all other elements in the system, thus, how much the overall movement tends to be concentrated in a determined area, given the configuration and mean depth of a road network – it is, in effect, a measure of *relative accessibility*. The PRC, on the other hand, assumes that the general movement probability within a road element is dependent on its connectivity to other well-connected road elements. In that sense, the PRC is not a measure of *nearness*, but instead, it describes the *strength* of a connection among road-elements. In other words, the PRC gives greater importance to the *strong ties* of a network, with its basic assumption that *important* road elements are those well connected to other important ones. In an urban setting, areas with high connectivity – often associated with urban cores – will have a significant presence of *strong-tied* road elements and will be highlighted as central in the PRC; in that manner, results tend to be similar to those from *space syntax'* integration measure (Hillier, 2006).

Further applications were studied by Agryzkov et al., who tested models of PageRank centrality (2012) and *eigenvector centrality* (2019) – both using primal networks – in urban settlements, to address the displacement of key urban activities within the cities' road-infrastructure. While Jiang (2009) and Agryzkov et al (2012; 2019) have attained noteworthy correlations that validate the use of *Markov-based* measures for urban movement analysis, the additional step of addressing the proper value of the road elements using segmented dual networks – instead of only position, through a primal network – is still to be made.

While the PageRank centrality (PRC) highlights the *strong ties* amongst elements within a network, it is unable to address their counterpart, the *weak ties*, as studied by Granovetter in interpersonal networks. *Weak ties* correspond here to network elements that constitute the most important bridges within a system, providing one of very few links between groups of well-connected – *strong tied* – elements. These *weak tied elements* represent critical points of connection within a system; thus, they can be related to the overall resilience of the network: once these elements are removed, the connectivity of the system is reduced, rendering it more fragile; a relevant characteristic to be addressed in urban and regional road-circulation networks movement dynamics.

We then introduce another measure that can highlight the weak ties; it is derived from the so-called *Kemeny constant* of a *Markov-process*. This constant, introduced in (Kemeny, Snell, 1960), gives a measure of how "well-connected" a *Markov-process* is: the higher it is, the more time is needed to "navigate" between its states. To introduce it as a centrality measure, we first need to recall two different concepts from *Markov-chain* theory. The first is the *invariant distribution* of a Markov-chain with transition matrix P, which is a solution of the system of linear equations:

$$\pi_j = \sum_{i=1}^{n} \pi_i P_{ij} \qquad j = 1, \dots, n. \tag{2}$$

While this is a singular system, it can be proved that there is a unique solution such that $\pi_1 + \pi_2 + \cdots + \pi_n = 1$, under mild assumptions that are verified in our setting. Moreover, the second property is the *mean first passage time* m_{ij} which is the average amount of time required to reach state j from state i (with the provision that $m_{ii} = 0$). Based on these properties we can then define the Kemeny constant as:

$$k(P) = \sum_{j} m_{ij}\pi_j, \quad (3)$$

Which represents, the *mean amount of time* needed to reach a state chosen according to the *invariant distribution.* It is a surprising fact explained by Doyle in 1983 that $k(P)$ is independent of the starting vector i, making it a function of the transition matrix of the Markov process P only.

The variation of the Kemeny constant was proposed by Crisostomi, Kirkland & Shorten (2011) as a manner to study the road-circulation network movement dynamics, by identifying the critical nodes within a system. Other applications, in terms of the identification of bridge elements, can be seen in Yilmaz et al. (2020), who proposed the use of the Kemeny constant to trace spreaders amongst individuals during the COVID-19 pandemic. Based on their ideas, we propose a Kemeny-based centrality (KBC), that can be computed without additional information just from a road network and can be applied in territorial analyses. In urban and regional systems, the KBC gives higher importance to the road elements that constitute *weak ties* within the road-circulation network. Therefore, a configurational measure of centrality based on this constant is able to identify crucial road elements that connect groups of *strong ties* – the urban cores – being also related to the systems' structure of *preferential routes*

3 DATASETS AND METHODS

PageRank centralities (PRC) and Kemeny-based centralities (KBC) were computed using MATLAB (Mathworks Inc., 2020). An abbreviation for *Matrix Laboratory*, MATLAB is a proprietary programming language and computing environment tailored to numerical linear algebra and matrix manipulations. This software allowed us to represent the road-structure data as graphs and run algorithms that compute various centrality measures. Several adaptations were needed to import the dataset – the *Grafo Iter.net* (Regione Toscana, 2019), an RCL dual graph of Tuscany's road-circulation network – into the MATLAB environment. Other than a conversion, partial adaptations were performed to the dataset prior configurational analysis, using both DepthMapX 0.8 (2018) and QGIS 3.16 (2020) as auxiliary software. These pre-conversion steps added certain parameters, used to compare the *Markov-based* measures to the *space syntax'* Angular Segmented Analysis (ASA) (Turner, 2001; 2007) and to incorporate the modelled data from MATLAB into the dual graph road-circulation network representation spatialized in QGIS 3.16.

Pre-conversion steps involve the exportation of the RCL graph dataset – a .shp *(shapefile)* set in QGIS 3.16 – into a .dxf (*drawing exchange format*) map; this drawing map is then loaded on DepthMapX 0.8 and converted into a segment map. This step, not only performs the *angular segmentation*, dividing the RCL continuous polylines into individual road elements when there are changes in direction but also attributes an angular coefficient to each road-element relative to the turn's angle. We use the *angular segmentation* to approximate the *Markov-based* measures to the ASA logic, albeit at this first moment we do not weigh the novel centrality measures by the angular coefficient.

This segmented map is then exported from DepthMapX 0.8 (2018) as a .mif (*MapInfo*) file and loaded once again into QGIS 3.16 (2020) where a geometrical tool is used to extract the endpoints of each segmented road element in the network, which converts the dual network into a primal network. The coordinates of these nodes, defined as X1, Y1, X2, and Y2 are then calculated. From the same map, we extracted another set of nodes that comprises the mid-points (centroids) for each road element. We attributed an ID to these nodes and calculate both X and Y coordinates (Figure 4). The ID values are later allowed to join the results into the RCL database, since both PRC and KBC are calculated to this mid-point, based on the primal network. The result of both procedures is shown in Figure 5. Datafiles containing the ID's and the coordinates for both mid-points and endpoints are then exported from QGIS 3.16 as *comma separated values* (.csv) tables. The datasets are then imported on MATLAB, where scripts are run calculate *Markov-based centralities*.

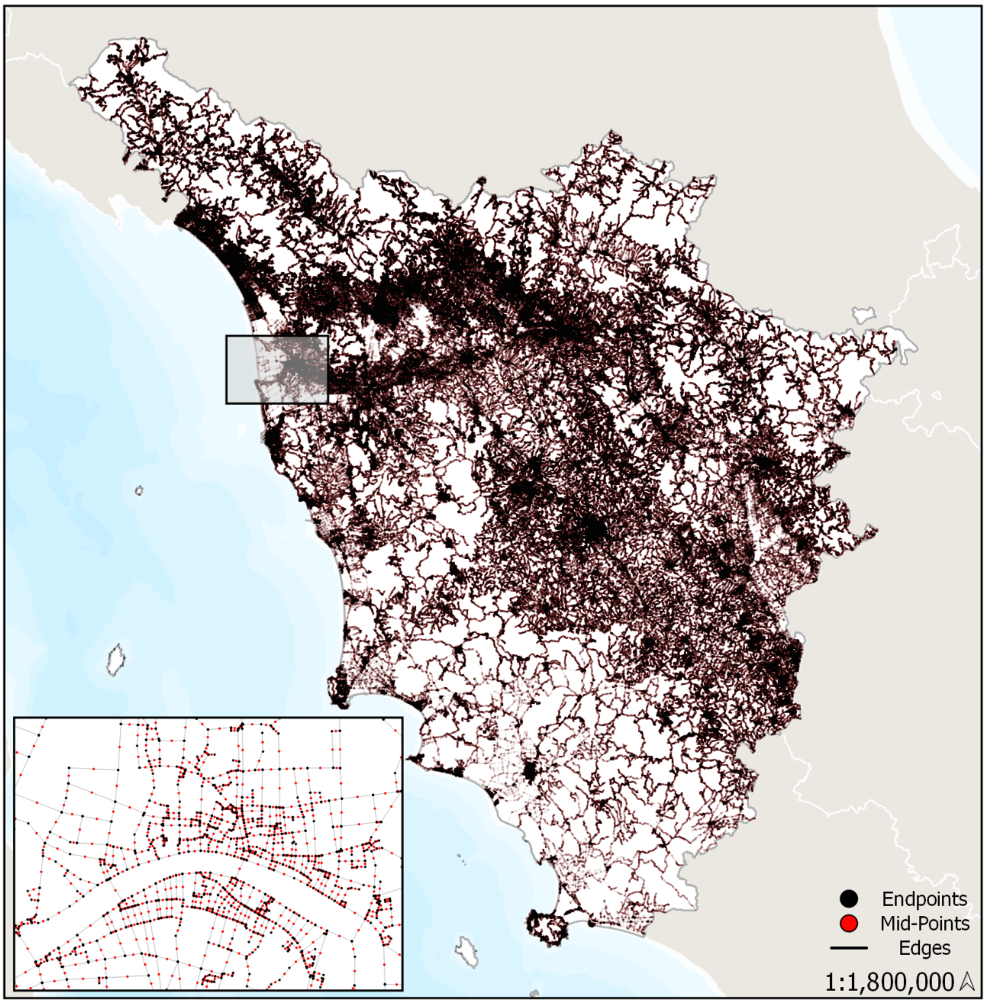


Figure 4: Tuscany road-circulation network as nodes and edges.

The PRC and KBC models are estimated directly for Tuscany mainland and at the regional scale as, due to being based on invariant network properties, they tend to present a *homothetic behaviour* across scales (Altafini, Cutini, 2020a).

The mathematical definitions used to compute the PageRank centrality (PRC) and Kemeny-based centrality (KBC) can be summarized as follows. We assumed an undirected network, where the edges represent the road elements and the nodes (states) are their endpoints (primal graph). We define that when there is an intersection between the elements, the probability of each of the elements to be followed (traversed) is proportional to (Equation 8):

$$w_\ell = e^{\ell/-\ell_{max}} \tag{4}$$

Where ℓ is the length of the road element, and ℓ_{max} is the maximum length of a road element in the network. Therefore, w_ℓ is a decreasing function of ℓ, which has values between 0 and 1. The role of this weight function, which is applied in both measures, is to discourage a random individual from choosing the longer segments, to compensate for the fact that they take more time to traverse it. With this in consideration we can define the transition matrix P of the *Markov process* attributed to the Tuscany road-circulation network as (Equation 5):

$$P_{ij} = \frac{W_{ij}}{\sum_{.k} W_{ijk}} \tag{5}$$

If there is a road element of (Euclidean) length ℓ, which joins the intersection points in i and j, then $W_{ij} = w_\ell$, and otherwise $W_{ij} = 0$. Note that $\sum_{j=1}^{n} P_{ij} = 1$ for each i, as the total probability of traversing a road element departing from it must be 1. We then compute the measure according to the probability of finding a random surfer in each intersection point (node). Based on these, we can then compute the corresponding probability that the random surfer traverses a road-element *(i,j)* as

$$PRC_{ij} = \mu_i P_{ij} + \mu_j P_{ji.} \tag{6}$$

Indeed, since our model is based on an undirected graph, each road element can be traversed in two directions, starting from i or j. Computed for each road element, – the PageRank centrality (PRC) – is then a *relative importance* measure that is independent of the network depth.

As the PRC probabilities are divided across each road element, and their sum is equal to 1, PRC overall values tend to assume quite small numbers for very large networks. To address this issue, a normalization is made by multiplying the PRC results by the node count (the total number of road elements in the network), as presented in the formula (Equation 12); the Normalized PageRank centrality (NPRC) also brings the PRC to values between 0 and 1:

$$NPRC = \frac{(PRC * NC) - \wedge(PRC * NC)}{\vee(PRC * NC) - \wedge(PRC * NC)} \tag{7}$$

Using MATLAB, the computational time required to estimate the PageRank centralities (PRC) for the Tuscany Road-Graph (composed of 1.2M road elements) is of a few seconds only on a modern computer. Indeed, the algorithm has a cost that scales as O(n^2) for a typical road network. Still, is considerably faster when compared to *Space Syntax'* angular integration for the same network, (~2.5 months) (Altafini & Cutini, 2020b) even though DepthMapX 0.8 (2018) also uses an O(n^2) algorithm to calculate mean-depth and integration (Turner, 2004), which may be an indicator that the sparsity principles are not considered in DepthMapX.

We use a *Markov-process* with the same transition matrix P to compute a centrality score based on the Kemeny constant $k(P)$. In detail, the Kemeny-based centrality (KBC) score of a road element *(i,j)* can then be defined as (Equation 8):

$$KBC_{ij} = k(\widehat{P}) - k(P) \tag{8}$$

Where $\hat{P}$ is the transition matrix of the *Markov chain* obtained after removing the road element *(i,j)* from the network. The matrix $\hat{P}$ must be re-built without the removed road element according to the formulas above and, in particular, it has $\sum \hat{P}_{ij} = 1$. Therefore, for each road segment (i,j) that has its KBC evaluated, we need to compute the Kemeny constant of another network $\hat{P}$, with one fewer road-element. Due to this iterative process the computation costs, in terms of time, may become prohibitive for large networks. Other formulas for the Kemeny constant can be found in Wang et al. (2017), using them, one can reduce the cost of computing the Kemeny-based centrality for all road segments in a network, by considering the fact that $\hat{P}$ is a small-rank update of the initial transition matrix P. The various possible optimizations, together with further modifications which are currently under analysis, will be described in a more technical upcoming paper, in which we detail the algorithms that can be used to speed up this computation. In the end, the estimation time of the KBC model for the Tuscany Road Graph resulted was ~3.5 days on a computing server with 12 computing cores. Nevertheless, the time-lapses are still minimal when compared to the *space syntax'* angular integration measure modelled in DepthMapX 0.8 (2018).

The KBC highlights the location of the *weak ties* within the network. Due to the nature of the measure, values have a high amplitude between its maximum – positive – and minimum – negative – scores. To partially address this issue, we calculate KBC', defined by the sum of each KBC value to the modulus of the KBC's minimum value. As a result, the maximum KBC' (∧KBC') becomes equal to 0 (Equation 9).

$$KBC' = KBC + |\wedge KBC| \tag{9}$$

As the amplitude between 0 and the maximum positive value is still high, we normalize KBC' to values between 0 and 1 (Equation 10):

$$NKBC = \frac{KBC' - \wedge KBC'}{\vee KBC' - \wedge KBC'} \tag{10}$$

Even if the Normalized Kemeny-Based centrality (NKBC) reduces the numerical difference between $\vee KBC'$ and $\wedge KBC'$, the difference between zero and the first non-zero value is still quite high, due to the measures' intrinsic properties.

Once the measures are modelled, we export the results in a .csv *(comma separated values)* file, which has both NPRC and NKBC attributed for each road element, according to the previously established ID for the mid-point. This table is then imported into QGIS 3.16 and joined with the graph dataset shapefile, from where the endpoints and midpoints were extracted. This completes the process of estimating and integrating the *Markov-based* centrality measures to the *Angular Analysis* framework.

4 RESULTS OF THE PAGERANK CETRALITY (PRC) AND KEMENY-BASED CENTRALITY (KBC) MODELS

Results for the Normalized PageRank centrality model (NPRC) at the regional scale are spatialized in QGIS 3.16 (Figure 5). As a first visual impression, it can be observed that the NPRC highest values (red and orange ranges) tend to be concentred around the areas that constitute urban cores throughout the region. While there are similarities with the *space syntax*' integration, these are restricted to this visual aspect, and even then, only to a certain degree and scale.

In quantitative terms, since the PRC scores are probabilities to travel through each road element, and their sum is equal to 1, their values tend to be quite small, which hinders comparisons among the ranges (Table 1). After normalization (NPRC – Table 1), however, it is revealed that the higher ranges – the road elements in the top 20% of the distribution (Figure 6) – are set in the interval between 0.39 and 1, whereas when the top 10% of the distribution is considered, the interval set on 0.44 to 1.

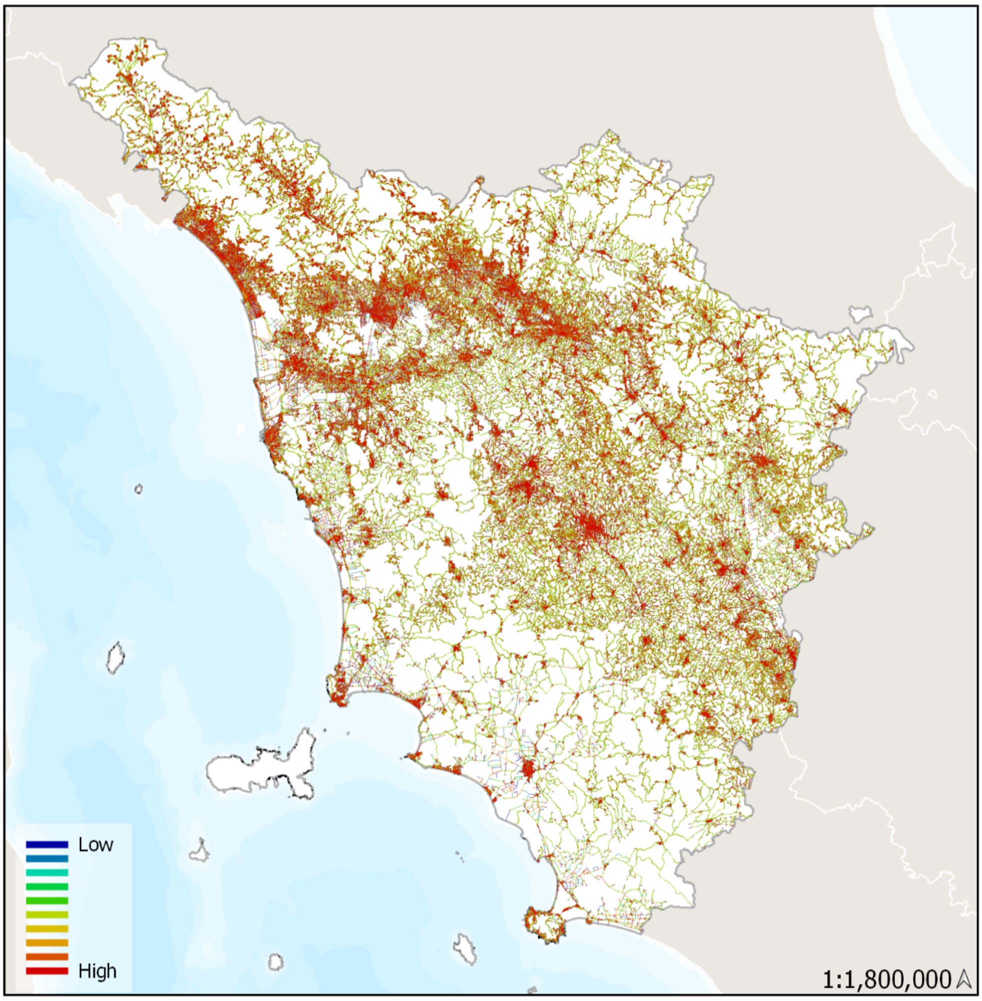
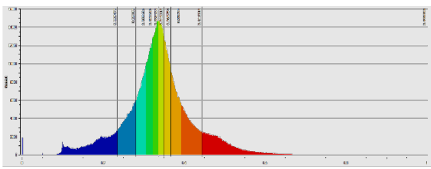


Figure 5: Normalized PageRank centrality (NPRC) for Tuscany regional road-circulation network.

This indicates a large amplitude among the values for the road-elements in this range, as well as that there are many road elements with very high values. Still, most road-elements tend to be set around values in the interval of 0.31 and 0.37, whereas lower values range from 0 to 0.31, which indicates that the distribution is not tail-heavy as in other centrality measures – in effect, the PRC distribution is bell-shaped (Table 1).

Table 1: PageRank centrality (PRC) and Normalized PageRank Centrality (NPRC) value ranges, configurational statistics, and value distribution for Tuscany regional road-circulation network.

PRC - Value Ranges	NPRC – Value Ranges	Configurational Statistics	
0.00000000 – 0.00000057	0.00 – 0.24 (Low)	Connectivity Max.	Node Count
0.00000057 – 0.00000068	0.24 – 0.28	12	1,225,140
0.00000068 – 0.00000074	0.28 – 0.31	Connectivity Avg.	
0.00000074 – 0.00000078	0.31 – 0.32	2.56	
0.00000078 – 0.00000081	0.32 – 0.34	Distribution (PRC):	
0.00000081 – 0.00000085	0.34 – 0.35		
0.00000085 – 0.00000089	0.35 – 0.37		
0.00000089 – 0.00000095	0.37 – 0.39		
0.00000095 – 0.00000107	0.39 – 0.44		
0.00000107 – 0.00000241	0.44 – 1.00 (High)		

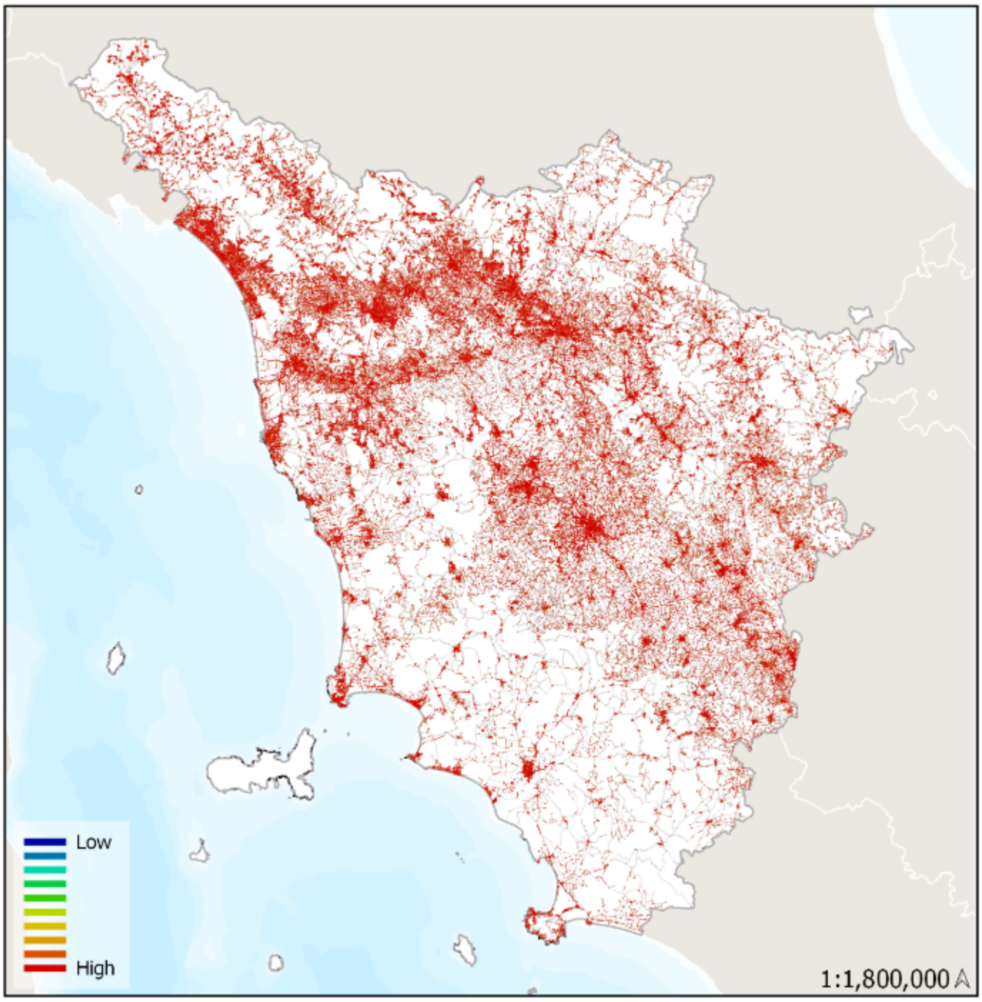


Figure 6: Normalized PageRank centrality (NPRC) restricted to the 20% of the road elements within the higher ranges for the Tuscany regional road-circulation network.

Examining the lower scale (Figure 7), we can observe that the PRC (NPRC) does not have a well-defined distribution in terms of internal hierarchies, despite being a bell-shaped distribution, above all, amongst values that are set in the average-to-lower ranges.

While these values tend to be predominant around the boundaries between the urban cores and the sprawled areas across the hinterlands – where the *relative importance* of the road-elements tends to drop, due to a lower connectivity – road-elements within lower PageRank scores can also be found inside the urban cores.

Figure 7: Normalized PageRank centrality (NPRC) distribution for different urban cores in Tuscany

In that aspect, the attained results corroborate with the conclusions made by Jiang (2009) and Agryzkov et al (2012), which state that PageRank scores tend to function as localized movement indicators, although this also shows that this *Markov-based* measure is essentially different from the *space syntax'* integration (*closeness* centrality), since even if modelled at regional scale, it is a local indicator. It is, then, sensible to assume that the use of the NPRC in an *angular segmented* RCL graph could better highlight the localized movement probabilities than when used in an *Axial representation*, as the differences among local movement are then informed at the road-element level often without visible continuities in value decrements. With this in consideration, it is arguable that the PageRank scores – and the NPRC as a whole – would function better as an indicator at regional scales, if the RCL graphs are not submitted to *angular segmentation*. In that sense, the PRC measure would be able to inform the *relative importance* or the localized movement probabilities for the continuous road-elements, being better distributed across the road-circulation network. However, since the objective posed for the measure is to address and locate where the groups of *strong tied* elements are, the *angular segmented* version of the PRC attains the expected results, as the *relative importance* of each road-element is informed. Based on the results the NPRC, as it is, can be useful both in regional and in urban settings, being adapted for several structural evaluations related to localized movement dynamics as, likewise tested by Jiang (2009) and Agryzkov et al (2012), it provides a good correspondence with both human movements and with the placement of urban equipment – economic activities such as retail and services, and with the public institutions.

Results for the Normalized Kemeny-based centrality model (NKBC) at regional scale are spatialized in QGIS 3.16 (Figure 8). Unlike the NPRC, the NKBC demonstrates visually distinctive spatial patterns at this scale, where a hierarchical distribution amongst higher, average, and lower values is evidenced. Table 2 exhibits the results for both KBC and NKBC.

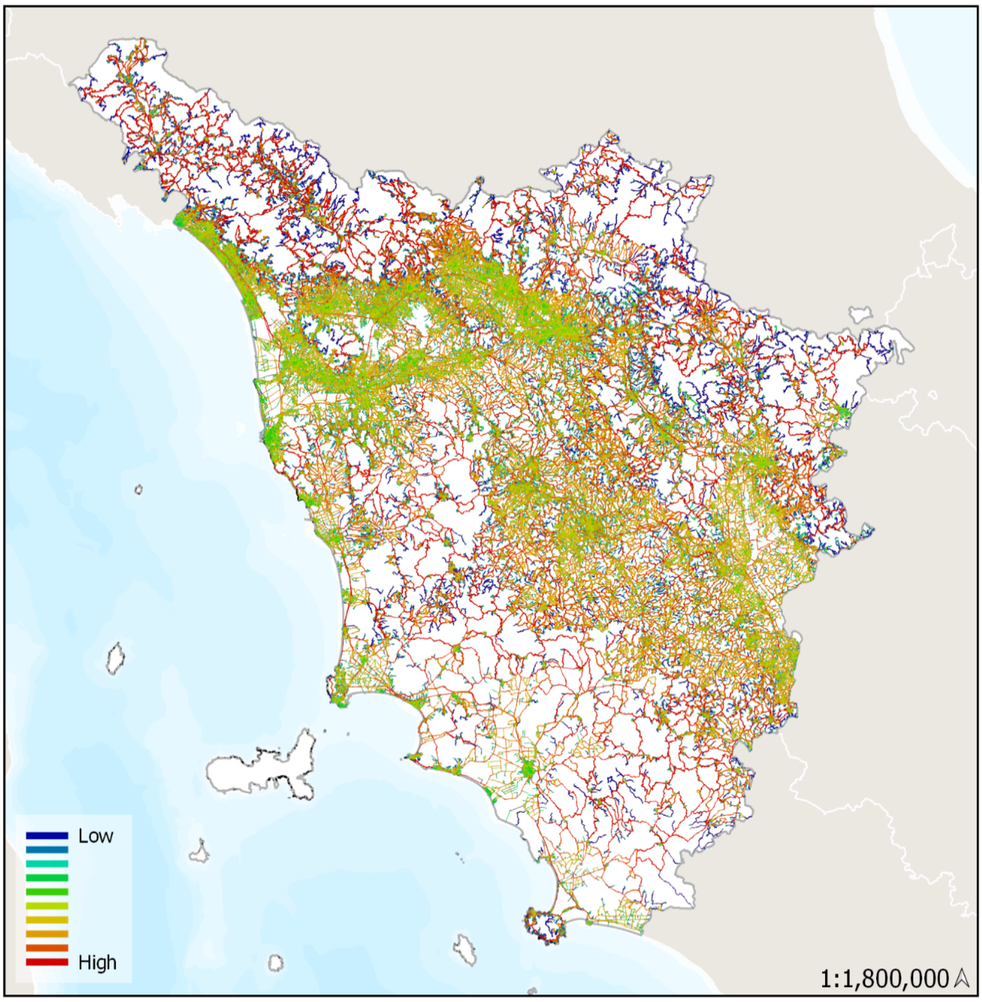
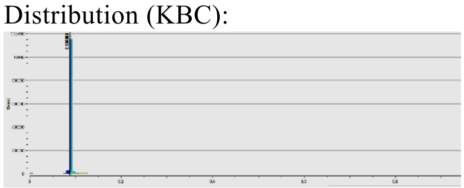


Figure 8: Normalized Kemeny-based centrality (NKBC) for Tuscany regional road-circulation network.

Table 2: Kemeny-based centrality (KBC) and Normalized Kemeny-based centrality (NKBC) value ranges, configurational statistics, and value distribution for the Tuscany regional road-circulation network.

KBC - Value Ranges	NKBC - Value Ranges	Configurational Statistics	
[-1464601] – [-1516]	0.000000 – 0.087664 (Low)	Connectivity Max.	Node Count
[-1516] – [-307]	0.087664 – 0.087739	12	1,225,140
[-307] – [-84]	0.087739 – 0.087753	Connectivity Avg.	
[-84] – [-20]	0.087753 – 0.087756	2.56	
[-20] - 92	0.087756 – 0.087763	Distribution (KBC):	
92 – 447	0.087763 – 0.087784		
447 – 1441	0.087784 – 0.087844		
1441 – 4541	0.087844 – 0.088030		
4541 – 18229	0.088030 – 0.088850		
18229 – 15224546	0.088850 – 1.000000 (High)		

The higher values, set in red and orange ranges, correspond to the road elements that constitute the most important bridges within the system – or the *weak ties* – that are, in turn, set in-between the areas that locate the road elements that correspond to *strong ties* – medium ranged values in green and yellow; the system endpoints, that have no relevance as bridge elements, are instead set in blue. In quantitative terms it can be observed that the NKBC possesses a very heavy-tailed value distribution, as there are many road elements that constitute endpoints in a system (Table

2). In that particular aspect, the NKBC assumes distributional properties that are likewise those found in *betweenness centralities*, even though these measures are fundamentally distinct. The road elements that constitute the system's *weak ties* are especially present in the mountain passes and throughout the hinterland areas, where they constitute the sole connection among small groups of *strong tied* elements. The large clusters of well-connected road elements have, instead, fewer *weak ties* that establish their connection (Figure 8). These distributional patterns, which are arguably the remarkable result attained through the measure, given the practical challenges in individuating those road-elements in the road-circulation network, can be better visualized when we restrict the road elements to the top 20% of the distribution (9° and 10° decile) (Figure 9).

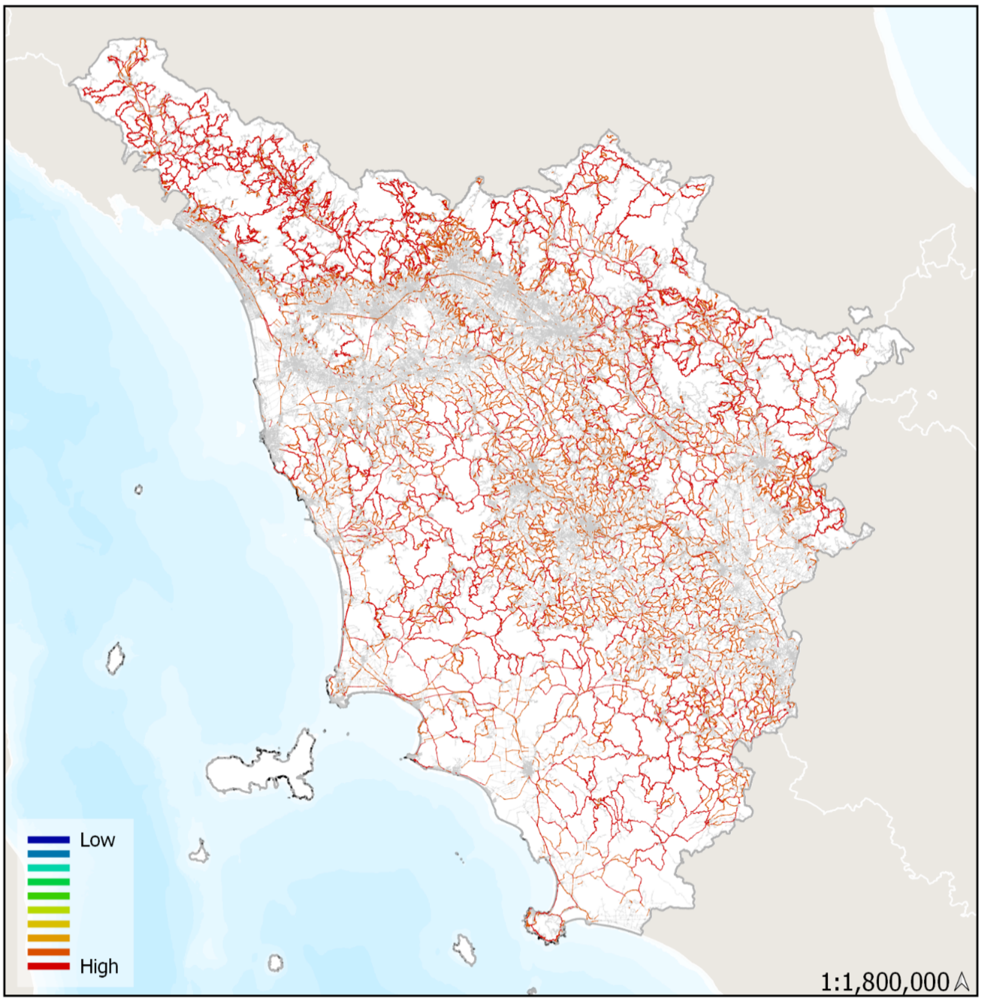


Figure 9: Normalized Kemeny-based centrality (NKBC) restricted to the 20% of the road elements within the higher ranges for the Tuscany regional road-circulation network.

It is important to address that the KBC normalization is only numeric, imperative as the high amplitudes between the minimum and maximum values, associated with the presence of negative values, hindered results' interpretation. Hence, normalization does not consider configurational properties as, in example, the system's Total Depth, in this sense, being distinct from the processes devised by Hillier, et al. (2012) for the *space syntax'* Normalized Angular Integration (NAIN) and Normalized Angular Choice (NACH) measures. While this is not an issue as, in terms of results, the NKBC identifies effectively the *weak ties* within a road-circulation network, we recognize that a further depth-based normalization would be important since the NKBC, as it is, does not present distinguishable differences in *relative importance* between *weak-ties* that connect different hierarchies of *strong-tied* road-elements. In that sense, the bridges that connect larger clusters of *strong tied* elements – the cities and conurbated areas – tend to be similar – or even lesser – in value to those that connect smaller groups that represent the hinterland and mountain towns. This characteristic is, once again, better observed when the 20% of the road elements with the highest values are highlighted (Figure 9).

We assumed that these differences in *weak ties'* hierarchies could be solved through similar methods of normalization as those used for *space syntax'* NACH, as the *betweenness-based* measure likewise does not consider depth (Hillier, et al, 2012). Nevertheless, the attained results demonstrated that the KBC measure seems "unaffected" by the inclusion of Total Depth, as it returns the patterns of distribution – both visual and in value – from the very measure that is used to normalize it. In effect, the same happened when other variables were tested, which demonstrates the apparent KBC invariance. In that aspect, further considerations and tests are in course to address these questions. Even if not yet an ideal measure from the mathematical standpoint, the NKBC has an interesting potential to describe configurational patterns (Figure 9).

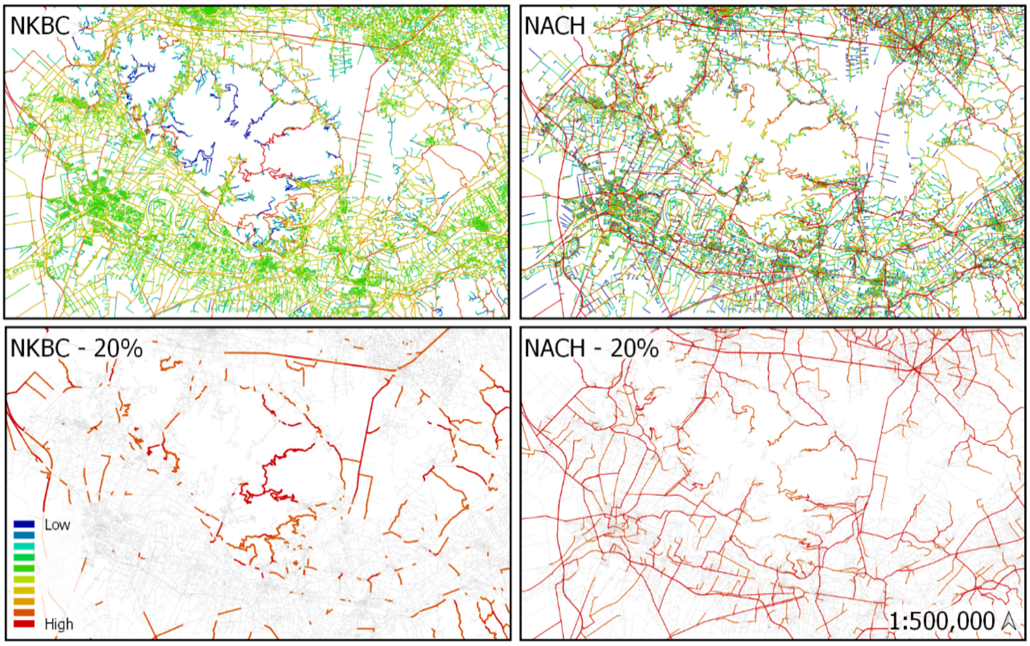


Figure 9: Comparison of the road-element value distribution amongst Normalized Kemeny-based centrality (NKBC) and Normalized Angular Choice (NACH) measures and their restrictions to the top 20% of road-elements.

Analogously to the results attained by Crisostomi, Kirkland & Shorten (2011), the NKBC can be used to individuate where the *fragility points* are located within the network, as the *weak tied* road-elements represent crucial bridges between the different groups of elements – yet, this time, at a road-element scale. If removed or interrupted, these road-elements tend to cause disruptions in the system's *preferential routes*, as they constitute the initial points of the road-elements highlighted in NACH measures (Figure 9, p.18).

Hence, the NKBC can provide an indicator of the overall resilience of the system, as well as be useful as a measure of the *relative importance* of the connections, given their nearness to urban areas and, at the same time to the linkages between urban areas, which is an important aspect to be considered for the location some economic activities focused on distribution, such as logistics or industry, and urban equipment that are dependent on the connection.

5 CONCLUSIONS

Closeness and *betweenness centralities* respectively determine the *syntactic* measures of *Integration* and *Choic*e, which are useful to describe certain configurational patterns related to movement. Nevertheless, those measures are limited in explaining properties related to positional *relative importance* within the road-circulation networks. In that regard, an overreliance on *Integration* and *Choice* consists of a limitation faced by the current *Space Syntax* methodology, above all when these measures are proposed to address relations between movement and network resilience or fragility. These aspects are important in urban analysis, but unfortunately, tend to lie beyond the explanatory capabilities of the *syntactic* measures. In this initial contribution, we summarized certain concepts of *strong-ties* and *weak-ties* that can be adapted from social network analysis – as once done with *Integration* and *Choice* measures – which can be useful in such assessments. We then devised two measures founded on *Markov-processes* that can estimate those properties of *strong ties* and *weak ties* in road-circulation networks, respectively the Normalized PageRank Centrality (NPRC) and the Normalized Kemeny-based Centrality (NKBC). These measures, originally conceived to estimate these relations in primal graphs, were adapted to a dual graph Road-Centre line-based representation, akin to the one used in *Space Syntax Angular Analysis.* This initial approach aims to merge those *Markov-based* centrality measures with *Space Syntax* to create appropriate instruments to highlight the *relative importance* of road elements. Our results demonstrate that adapting Markov-based centralities to *Space Syntax* is feasible and maintains a configurational and spatial sense, hence it introduces new dimensions to be evaluated in urban-regional analysis. Further research is being conducted to incorporate other parameters of *Angular Analysis* to these measures, as well as to refine them in mathematical terms. Beyond that, more tests are being done to compare and confront the *Markov-based* measures with established centrality measures to address their overall efficiency. This, however, lies outside the scope of this first approach, as it merits a dedicated analysis, given the methodological differences between *Markov* and *Syntactic* measures. Other possibilities of development lie in the *Markov-based* centralities application to empirical cases, to test how well they perform and emphasize the configurational properties in terms of practical usage.

REFERENCES

Agryzkov, T., Tortosa, L., Vicent, J. F., & Wilson, R. (2019). A centrality measure for urban networks based on the eigenvector centrality concept. *Environment and Planning B: Urban Analytics and City Science*, *46*(4), 668–689. https://doi.org/10.1177/2399808317724444

Agryzkov, T; Oliver, J; Tortosa, L; Vicent, J. (2012). An algorithm for ranking the nodes of an urban network based on the concept of PageRank vector, Applied Mathematics and Computation 219(4), –. doi:10.1016/j.amc.2012.08.064

Altafini, D., and Cutini, V. (2020a). "Homothetic Behavior of Betweenness Centralities: A Multiscale Alternative Approach to Relate Cities and Large Regional Structures" *Sustainability* 12, no. 19: 7925. https://doi.org/10.3390/su12197925

Altafini, D; Cutini, V. (2020b) Tuscany Configurational Atlas: a GIS-based multiscale assessment of road-circulation networks centralities hierarchies. Computational Science and Its Applications – ICCSA 2020, 20th International Conference, Cagliari, Italy, July 1–4, 2020, Proceedings, Part III

Berkhout, J., Heidergott, B. F. (2019). Analysis of Markov influence graphs. *Operations Research*, *67*(3), 892–904. https://doi.org/10.1287/opre.2018.1813

Biggs, N. 1974. *Algebraic Graph Theory* (2nd ed., Cambridge Mathematical Library). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511608704

Bonacich, P. (2007). Some unique properties of eigenvector centrality. *Social Networks*, *29*(4), 555–564. https://doi.org/10.1016/j.socnet.2007.04.002

Brin, S.; Page, L. 1998. The anatomy of a large-scale hypertextual Web search engine. Computer Networks and ISDN Systems. 30 (1–7): 107–117.

Crisostomi, E., Kirkland, S., & Shorten, R. (2011). A Google-like model of road network dynamics and its application to regulation and control. *International Journal of Control*, *84*(3), 633–651. https://doi.org/10.1080/00207179.2011.568005

Cvetković, D. M., Doob, M., & Sachs, H. (1980). *Spectra of graphs: theory and application.* New York, Academic Press.

Davis, J., 1970. Clustering and Hierarchy in Interpersonal Relations. *American Sociological Review* 35 (October): 843-52.

depthmapX development team. (2018). depthmapX (Version 0.8.0) [Computer software]. Retrieved from https://github.com/SpaceGroupUCL/depthmapX/

Estrada, E. (2012). *The structure of complex networks: theory and applications*. Oxford University Press.

Estrada, E., & Benzi, M. (2014). Walk-based measure of balance in signed networks: Detecting lack of balance in social networks. *Physical Review E*, *90*(4), 042802.

Garfeld, E. (1995) New international professional society signals the maturing of scientometrics and informetrics. The Scientist, 9(16).

Hillier, B. (2012) The genetic code for cities: Is it simpler than we think? In: J. Portugali, H. Meyer, E. Stolk and E. Tan (eds.) Complexity Theories of Cities Have Come of Age: An Overview with Implications to Urban Planning and Design. London: Springer, pp. 67–89.

Hillier, B., Yang, T. and Turner, A. (2012) Normalising Least Angle Choice in Depthmap and How It Opens up New Perspectives on the Global and Local Analysis of City Space. Journal of Space Syntax, 3, 155-193.

Hillier, B. 2007 Space is the machine: a configurational theory of architecture, Space Syntax, London pp. 368

Jiang, B. (2009). "Ranking spaces for predicting human movement in an urban environment". International Journal of Geographical Information Science. 23 (7): 823–837. arXiv:physics/0612011. Bibcode:2006physics..12011J. doi:10.1080/13658810802022822. S2CID 26880621.

Li, Y. (2002). "Toward a qualitative search engine". *IEEE Internet Computing*. **2** (4): 24–29. doi:10.1109/4236.707687.

Marchiori, M, 1997 "The Quest for Correct Information on the Web: Hyper Search Engines", Proceedings of the Sixth International World Wide Web Conference (WWW6).

Markov A. 1971. "Extension of the limit theorems of probability theory to a sum of variables connected in a chain". reprinted in Appendix B of: R. Howard. *Dynamic Probabilistic Systems,* *volume 1: Markov Chains. John Wiley and Sons.* Markov A. 1906 "Rasprostranenie zakona bol'shih chisel na velichiny, zavisyaschie drug ot druga". *Izvestiya Fiziko-matematicheskogo obschestva pri Kazanskom universitete, 2-ya seriya, tom 15, pp. 135–156.*

Mathworks Inc. (2020). MATLAB *version 7.10.0 (R2020a)*. [Computer software] Natick, Massachusetts: The MathWorks Inc.

Newcomb, T. M. 1961. The Acquaintance Process. New York: Holt, Rinehart & Winston. Pinski, G; Narin, F. (1976). "Citation influence for journal aggregates of scientific publications: Theory, with application to the literature of physics". *Information Processing & Management*. **12** (5): 297–312. doi:10.1016/0306-4573(76)90048-0

Page, L; Brin, S; Motwani, R; Winograd, T. (1999). The PageRank citation ranking: Bringing order to the Web - Technical report on January 29, 1998, Unversity of Standford.

QGIS (2020). QGIS 3.16 [Computer software] Available online: https://www.qgis.org/en/site/forusers/download.html

Regione Toscana, (2019) Direzione Urbanistica e Politiche Abitative - Sistema Informativo Territoriale e Ambientale – *SITA.: Grafo stradario e ferroviario della Regione Toscana* http://www502.regione.toscana.it/geoscopio/cartoteca.html

Turner, A., 2007. From axial to road-centre lines: A new representation for space syntax and a new model of route choice for transport network analysis. Environ. Plan. B Plan. Des. 34:539–555.

Turner, A., 2004, “Depthmap 4 | A Researcher's Handbook", Bartlett School of Graduate Studies, UCL, London.

Turner, A. 2001. Angular Analysis. In: Proceedings of the 3rd International Symposium on Space Syntax, pages 7-11. Georgia Institute of Technology, Atlanta, Georgia.

Wang, X; Dubbeldam, J.; Van Mieghem, P. Kemeny's constant and the effective graph resistance. Linear Algebra Appl. 535 (2017), 231–244.

Xing, W; Ghorbani, A. (2004) "Weighted PageRank algorithm," Proceedings. Second Annual Conference on Communication Networks and Services Research, pp. 305-314, doi: 10.1109/DNSR.2004.1344743.