

RESEARCH

Open Access

The time varying network of urban space uses in Milan



Alba Bernini¹, Amadou Lamine Toure^{1,2} and Renato Casagrandi^{1*}

*Correspondence:

renato.casagrandi@polimi.it

¹Dipartimento di Elettronica, Informazione e Bioingegneria (DEIB), Politecnico di Milano, Via Ponzio 34, 20133 Milan, Italy
Full list of author information is available at the end of the article

Abstract

In a metropolis, people movements design intricate patterns that change on very short temporal scales. Population mobility obviously is not random, but driven by the land uses of the city. Such an urban ecosystem can interestingly be explored by integrating the spatial analysis of land uses (through ecological indicators commonly used to characterize natural environments) with the temporal analysis of human mobility (reconstructed from anonymized mobile phone data). Considering the city of Milan (Italy) as a case study, here we aimed to identify the complex relations occurring between the land-use composition of its neighborhoods and the spatio-temporal patterns of occupation made by citizens. We generated two spatially explicit networks, one static and the other temporal, based on the analysis of land uses and mobile phone data, respectively. The comparison between the results of community detection performed on both networks revealed that neighborhoods that are similar in terms of land-use composition are not necessarily characterized by analogous temporal fluctuations of human activities. In particular, the historical concentric urban structure of Milan is still under play. Our big data driven approach to characterize urban diversity provides outcomes that could be important (i) to better understand how and when urban spaces are actually used, and (ii) to allow policy makers improving strategic development plans that account for the needs of metropolis-like permanently changing cities.

Keywords: Community detection, Land use, Mobile phone data, Complex network, Temporal network, Spatial network

Introduction

Understanding land use occupation and dynamics is crucial for urban and territorial planning, since it allows to capture the results of past modifications, monitor ongoing changes and predict future impacts and opportunities. Since the end of the 20th century, land use and land cover maps have been extensively generated at different spatial and temporal scales, especially in the framework of the European CORINE program (Feranec et al. 2016). Land use maps are commonly utilized to inform decision makers on the spatial organization of the territory under their government.

Especially in the urban context, the compresence of different land uses can generate intricate patterns that can interestingly be explored with tools that ecologists typically adopt to analyze biodiversity in a natural environment. Biodiversity can be defined as the heterogeneity of living organisms and processes found in the environment (e.g., DeLong (1996)), and can be evaluated at different scales. According to Whittaker's idea (1972),

α -diversity is a local measure and refers to the richness in species of a site or habitat, β -diversity measures the differentiation between two sites or habitats. The total species diversity in a landscape (i.e., its γ -diversity) is a combination of both α - and β -diversity. These concepts can be transferred to an urban context to measure the "richness" of land uses of the different neighborhoods (corresponding to α -diversity) and to quantify their dissimilarities in terms of land-use mix (corresponding to β -diversity). The presence of not contiguous areas with similar usage and needs could suggest possible planning measures at higher levels, that cross the borders of the single neighborhoods.

Land use classification is commonly defined in terms of physical characteristics, such as reflectivity and texture, that can be well captured from remote sensing. Gong et al. (1990); Fisher (1997); Shaban and Dikshit (2001); Lu and Weng (2006). However, these technologies are unable to identify the actual utilization by people and to discriminate between some land use types (i.e., residential and commercial) (Louail et al. 2014; Pei et al. 2014). Information about human movement and activities can nowadays be derived thanks to data on digital human footprints, i.e. the digital traces that people leave while placing phone calls, interacting on the Internet or on digital devices (Howison et al. 2011). This enormous amount of data can be used to provide new fundamental and quantitative insights on the actual occupation of a certain urban area and, as a consequence, on its possible social function (Gonzalez et al. 2008; Cho et al. 2011; Hawelka et al. 2014; Hoteit et al. 2014; Secchi et al. 2015). The aim of this study is to investigate the complex relations occurring between the land-use composition of the areas inside a city and the spatio-temporal occupation made by citizens. As an important case study, we focused on the city of Milan, which is located in Northern-Italy and, with its 1.40 million inhabitants, is the second-most populous Italian city. Specifically, we aimed at identifying which areas are more similar/dissimilar in terms of either land-use composition or because of the daily trend of human occupation. These outcomes could be important not only to better understand how and when urban spaces are used but also for allowing policy makers to improve strategic development plans accounting for the needs of an "active" city.

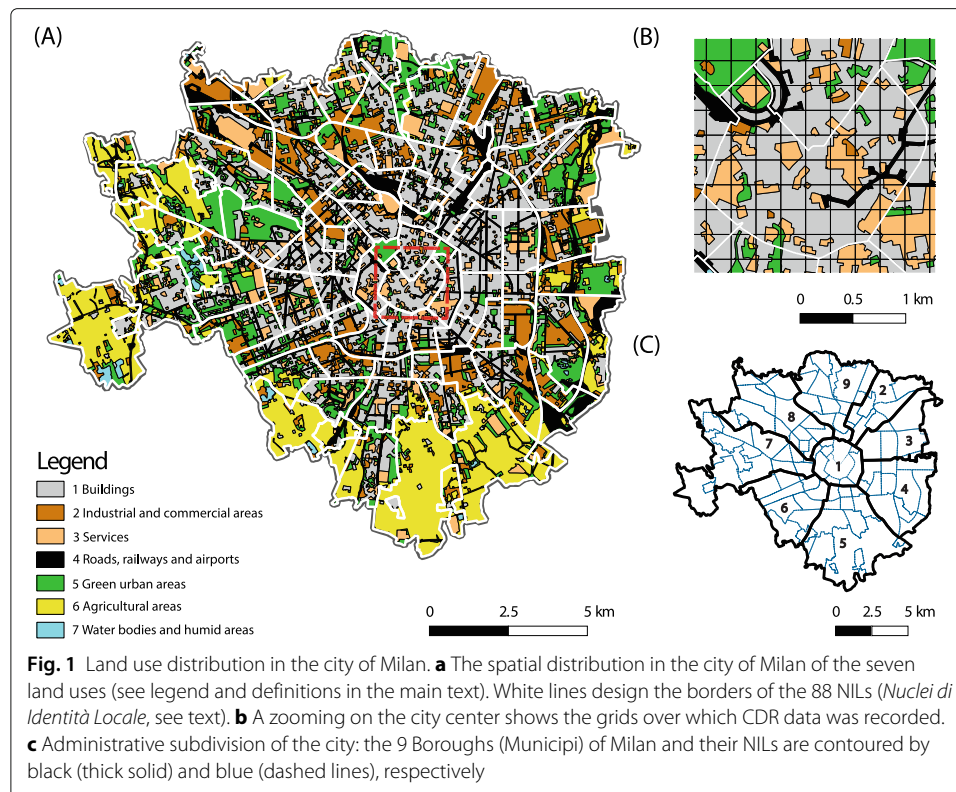
Materials and methods

Data on land use and land cover in the study area of Milan

In this study we focused on the urban area of the city of Milan, the chief seat of the Lombardy Region. The development of Milan occurred through subsequent expansions, which designed concentric circles in the urban structure (Morandi 2007). The recognition of a circular structure of Milan dates back to the XIII century, as documented by the Milanese poet Bonvesin de la Riva (1288) in his book entitled "De magnalibus urbis Mediolani" ("On the greatness of the city of Milan"). At section IV of chapter II, he writes in fact

*Civitas ista ipsa orbicularis est ad circulli modum, cuius mirabilis rotonditas perfectio-
nis eius est signum*

(which reads as "This same city has an orbital shape, in the shape of a circle: this admirable roundness is the sign of its perfection"). From an administrative perspective, the municipality of Milan is nowadays subdivided into nine Boroughs, the so-called *Municipi* (see Fig. 1C), each one having its own Council and President. The nine Borough Councils are coordinated at the city level by the City Council, which can decide over the general rules



for the use of goods and services. On the other hand, Borough Councils have independent administrative power and responsibility on some local but important matters, such as schools, social services, waste collection, roads, parks, libraries and local commerce. It is not guaranteed that the nine boroughs, however, cannot capture the richness of the city. Indeed, each Borough includes districts with different social and cultural aspects. Accounting for these characteristics, the city was further subdivided into 88 districts with unique social and cultural identity, the so called *Nuclei di Identità Locale* (NILs), i.e. Local Identity Neighborhoods, which are bounded by white lines in Fig. 1. Since their distinctive social identity is often reflected in their urban characteristics, they are often well recognizable by both inhabitants and tourists. Although they do not have an administrative function, NILs are adopted as reference system in the *Piano di Governo del Territorio* (PGT), i.e. the Territorial Administration Plan, to analyze and manage the local demand of services (www.pgt.comune.milano.it). The social relevance of NILs have been once again confirmed since one of the five objectives of 2019 PGT is to enhance them through the regeneration of their public spaces, as a first step towards a new polycentric city model. Given their importance in Milan urban planning, we adopted NILs as units in the analysis we performed, as also done in other studies on this city (Arnaboldi et al. 2017; Mariotti et al. 2017, 2018).

Land use maps for the city of Milan were obtained from the DUSAF geographical database, which is managed by E.R.S.A.F. (Regional Agency for Services to Agriculture and Forestry) and available and freely accessible at the GEOPortale (www.cartografia.regione.lombardia.it). We used the latest version of the database (DUSAF 5.0), which dates back to 2015.

In the DUSAF database, the territory of the whole Lombardy Region (thus also of Milan) is partitioned into geographical polygons, each representing an uniform area in terms of land use, e.g. a group of contiguous buildings, a road, a park. Polygons are classified according to a legend with 5 hierarchical levels detailing their nature. From coarsest to finest scales, the first three levels refer to the guidelines of CORINE Land Cover, while the last two levels were defined by E.R.S.A.F. using auxiliary databases to better categorize some land uses.

Each polygon in the DUSAF is labeled with a string made of a minimum 1 characters to a maximum of 5 characters. According to its position within the string, each character aims at qualifying the land use of the polygon at a specific level of detail: from left to right, the first character refers to level 1, the second to level 2 and so on. For example, hospitals are labeled as 12121, and each character represents (left to right, again):

- Level 1 1 Anthropic areas
- Level 2 12 Productive sites, big plants and communication networks
- Level 3 121 Industrial sites and public and private services
- Level 4 1212 Public and private services (hospitals, schools, ...)
- Level 5 12121 Hospitals

As a first step of the analysis we needed to select the categories that could be interesting for our study, reconciling the level of detail of the DUSAF description to the relevance of particular categories of land use. More specifically, we filtered only those categories that could be related to human occupation patterns during the day, such as industries, services and parks. We then grouped the rarest (e.g. water bodies and humid areas) or less relevant ones (e.g. agricultural areas) under new reference names. This preliminary analysis led to the selection of the following 7 categories:

1. *Buildings*, including all the subcategories in Urbanized areas (11);
2. *Industrial and commercial areas*, where we grouped Industrial and commercial sites (1211) and Construction sites, landfill, abandoned areas (13);
3. *Services*, including Public and private services (1212) such as hospitals, schools, courts and cemeteries;
4. *Roads, railways and airports*, bringing together Roads and railways (122) and Airports (124);
5. *Green urban areas*, including Forested and semi-natural areas (3), often located in public parks or gardens, and Green urban areas (14);
6. *Agricultural areas*, where we grouped all the subcategories belonging to the macro-category Agricultural areas (2);
7. *Water bodies and humid areas*, including the last two macro-categories Humid areas (4) and Water bodies (5).

Figure 1 shows the distribution of the land uses according to the grouping performed above in the study area of Milan. The city center is mainly occupied by Buildings and Services (see the grey and pink polygons in Panel (B)). Moving from the center to the periphery, Industrial and commercial areas (brown) and Green urban areas (green) are more abundant. The agricultural fields (yellow) are part of the rural area of the Parco Agricolo Sud Milano and are located in the external south and west part of the municipal area. Roads and railways (black) start revealing the urban concentric structure of the city.

Indeed, nowadays the historic city limits can be recognized since main roads traveled by cars and public transports follow them: the transportation network design a series of ring roads, crossed by radial roads that connect the center and the peripheries.

Biodiversity indicators based on land uses

The quantification of diversity is quite an explored field in Ecology. Since the studies of pioneers (Whittaker 1972), biodiversity in a given ecosystem has been measured by accounting for both its richness in terms of species and heterogeneity in the individuals representing them. Biodiversity within an ecosystem is commonly named α -diversity and it is high (*ceteris paribus*) when the number of present species is high and/or when the distribution of individuals among the species is evenly distributed (Simpson (1949); Shannon (1948), see for example). While α -diversity is a local measure of differentiation, β -diversity is used to quantify the differences between ecosystems or habitats (Whittaker 1972) and it has been mainly measured via dissimilarity indices, such as Jaccard and Sørensen indices, which account for the number of species they share or not (Jaccard 1901; Sørensen 1948): *ceteris paribus*, the larger is the fraction of shared species (possibly weighted by the number of individuals), the lower is β -diversity. In our study, we proposed to elaborate these concepts developed for natural environments to an urban context. Considering as “species” each of the seven land uses defined in the previous paragraph, first we measured the richness within NILs, so as to obtain α -diversity, and then we computed the pairwise dissimilarity between them, corresponding to their β -diversity.

We organized the data about land uses in a table **X** of 88 rows (the NILs) and 7 columns (the land uses). The element x_{ij} of the matrix was equal to the fraction of the area of NIL i occupied by the land use j . The row X_i of the table represented then the land-use mix of a NIL i (see some examples in Table 1). We measured the α -diversity within each NIL using the Simpson index (Simpson 1949), computed as:

$$H(X_i) = 1 - \sum_{j=1}^7 (x_{ij})^2 \quad (1)$$

The minimum value of the Simpson index is 0 and is obtained when only one land use is present in the NIL. On the contrary, its maximum possible value occurs when all land uses are equally distributed within the NIL i , i.e. $x_{ij} = 1/7 \forall j$, so that $H_i^{\max} = 6/7$.

To evaluate β -diversity between two natural ecosystems, typically adopted measures quantify their differences in terms of number of shared/unshared species. Examples of

Table 1 Land-use mixes of NILs 1, 18, 35 and 87, selected here as representative of Communities 1, 2, 3 and 4 respectively shown in Fig. 3, which have been identified by community detection performed on land uses

NIL	Buildings	Industrial and commercial areas	Services	Roads, railways and airports	Green urban areas	Agricultural areas	Water bodies and humid areas
1. DUOMO	0.646	0.015	0.246	0.052	0.041	0.000	0.000
18. PARCO LAMBRO - CIMIANO	0.194	0.073	0.169	0.125	0.281	0.138	0.020
35. LODI - CORVETTO	0.276	0.218	0.052	0.173	0.130	0.144	0.007
87. PARCO AGRICOLO SUD	0.029	0.066	0.000	0.034	0.084	0.723	0.064

Each row of the table, say X_i , represents the land-use mix of NIL i and cell x_{ij} contains the relative abundance of land use j in NIL i

β -diversity indices are the Jaccard (1901) and Sørensen (1948) indices, which have their minimum value when the two ecosystems do not share any species and their maximum when all species are shared. In our case study, the fact that all the seven categories of land use (i.e., our “urban species”) were present in almost all NILs (i.e., our “urban ecosystems”) posed some limitations in applying traditional β -diversity indices. Indeed, they would have returned an unrealistically uniform picture of the city, with all NILs being similar. To overcome such limitations, we decided to focus on the differences between distributions of land uses, rather than on their turnover along the territory. Thus, instead of using traditional β -diversity indices, we compared each couple of NILs, say i and j by computing the (Euclidean) distance between their land-use mixes, represented by two vectors X_i and X_j in a seven-dimensional parameter space where each axis is a land use type. We then normalized the Euclidean distance by dividing it by the maximum potential value it can reach, amounting to $\sqrt{2}$ (a value that is reached when the two NILs are both occupied by a unique, yet different land use). Our newly defined metric for computing the dissimilarity of land uses thus reads as

$$d_{LU}(X_i, X_j) = \sqrt{\frac{\sum_{k=1}^7 (x_{ik} - x_{jk})^2}{2}} \quad (2)$$

We note that, although Eq. 2 was here defined in the context of urban ecology, it might in principle represent an interesting metrics also to quantitatively compare at a glance natural simple ecosystems.

Community detection based on land uses

To detect the presence of groups of NILs with similar land-use mixes, whose number was not *a priori* known, we referred to network theory (Newman 2010) and, specifically, to community detection (Fortunato 2010). We built a complete, undirected yet weighted network with 88 nodes representing the NILs. Each link between two NILs, say i and j , was assigned a weight equal to the similarity between them. According to the metric defined by Eq. 2, we computed similarity as:

$$s_{LU}(X_i, X_j) = 1 - d_{LU}(X_i, X_j) \quad (3)$$

Over this network, we performed community detection based on modularity optimization (Newman 2006), which does not require to fix the optimal number of clusters in advance. In order to assess the robustness of the results, we repeated community detection also using the Louvain method (Blondel et al. 2008). In addition to applying network-based approaches, we also tried to use hierarchical clustering algorithms based on Ward’s method (Ward 1963) to detect groups of NILs with similar land-use mixes. To that end, we utilized as distance metric the measure $d_{LU}(X_i, X_j)$, defined in Eq. 2.

Data on CDR: Milan Telecommunications

To account for the effective use of the urban territory by people, we used an anonymized dataset on phone calls and internet connections. This dataset was part of the “2014 Telecom Italia Big Data Challenge”, which provides various geo-referenced information, released to the research community under the Open Database License (OdbL). The whole dataset (available at www.dataverse.harvard.edu) contains information about the use over time of the mobile-phone network in the urban area of Milan over a period of two months,

from November 1, 2013 to December 31, 2013 and it was obtained from Call Detail Records (CDRs).

Technically, CDRs log the details of calls made over a phone service and are registered by operators for billing purposes and network management. Every time a user engages a telecommunication interaction (either voice, text or other ICT connections), a Radio Base Station (RBS) is assigned by the operator and delivers the communication through the network. The time and duration of the interaction, together with its nature and the RBS, which handled it are recorded in a new CDR. From the information on the RBS it is possible to obtain the approximate user's geographical location at the time of the activity which must be within the area of influence of that RBS. The Telecom Italia Big Data Challenge dataset that we used for our analysis was the result of a computation over the CDRs, which were aggregated in a spatial grid with pixels 235m x 235m and in time slots of 10 min, as detailed in Barlacchi et al. (2015). The aggregation process generated a dataset, where each record describes the network activity occurred in a square of the spatial grid during a time slot in terms of incoming/outgoing SMSs and calls and Internet traffic activity.

Since we aimed at analyzing human presence in the different NILs during time, we considered the total activity as the sum of all types of activities (incoming/outgoing SMS and calls together with Internet connections) and we aggregated it by NIL and by hour of the day. Moreover, we arbitrarily decided to only consider in the analysis the four weeks from November 4, 2013 to December 1, 2013 in order not to include the religious holidays – such as for the patron Saint of Milan, Sant'Ambrogio (December 7, 2013), the Immaculate Conception (December 8, 2013) and Christmas. In these periods, indeed, people movement habits could have been considerably different from the rest of the year.

We therefore organized the data in a series of 28 matrices $\mathbf{Y}(t)$, with $t = 1, 2, \dots, 28$ (one for each day in the study period), each one made by 88 rows (the NILs) and 24 columns (the hours of the day). Each element $y_{ij}(t)$ represented the total number of mobile phone/network activities within NIL i occurring at the daily hour j of the day t divided by the surface area of NIL i . Each row $Y_i(t)$ of the table represented the network activity time series within NIL i in the day t .

Community detection based on CDR data

We used the CDRs to evaluate the spatio-temporal fluctuations of human activities within the city and to reveal the presence of human generated connections between NILs that are active at the same time. In particular, for each day t , we performed a pairwise time series comparison of CDR activities between the 88 NILs using an adaptive dissimilarity index covering both proximity on values and on behavior, as defined by Chouakria et al. (2007). Following their methodology, we computed the dissimilarity index as the product of a function μ of the temporal correlation ρ between two time series, say $Y_i(t)$ and $Y_j(t)$, and a measure of their (Euclidean) distance.

$$\rho(Y_i(t), Y_j(t)) = \frac{\sum_{k=1}^{23} (y_{i(k+1)}(t) - y_{ik}(t))(y_{j(k+1)}(t) - y_{jk}(t))}{\sqrt{\sum_{k=1}^{23} (y_{i(k+1)}(t) - y_{ik}(t))^2} \sqrt{\sum_{k=1}^{23} (y_{j(k+1)}(t) - y_{jk}(t))^2}} \quad (4)$$

$$\mu(\rho(Y_i(t), Y_j(t))) = \frac{2}{1 + \exp(2\rho(Y_i(t), Y_j(t)))} \quad (5)$$

$$d_{CDR}(Y_i(t), Y_j(t)) = \mu(\rho(Y_i(t), Y_j(t))) \times \sqrt{\sum_{k=1}^{24} (y_{ik}(t) - y_{jk}(t))^2} \quad (6)$$

where the subscript $k = 1, 2, \dots, 24$ indicates the hour within the day t . We then obtained 28 distance matrices, one for each day.

To evaluate the emergence of groups of NILs characterized by similar patterns of daily human occupation, we also performed network community detection on each distance matrix but using an agglomerative hierarchical clustering algorithm, based on Ward's method (Ward 1963). The motivation for using a clustering algorithm in this case, rather than a network-based algorithm as before, was to a priori fix the number of clusters to be identified within the city, a number that we wanted to keep equal to the one obtained from the analysis on the land-use mixes for comparison.

To measure the robustness through time of the identified communities, we built a network with 88 nodes representing the NILs and we traced links between nodes that had been grouped at least once in the same community within the time horizon under study (28 days). Then, we provided a weight w_{ij} to each link of the network connecting every node/NIL i with node/NIL $j \neq i$ equal to the fraction of days the two nodes were grouped in the same community. We finally defined as measure of distance between NILs i and j the difference $1 - w_{ij}$. Using this measure of distance, we performed thus community detection using hierarchical clustering with the Ward's agglomerative method on the weighted network.

As a last step, we measured to what extent the partition in communities generated by the analysis of the CDR dataset was coherent with the one generated with the analysis of land uses. Following the methodology presented by Fortunato et al. (2016), we computed the similarity of the two partitions through the Jaccard index, i.e. the ratio between the number of node pairs classified in the same cluster in both partitions and the number of node pairs classified in the same cluster in at least one partition. To assess whether the obtained similarity score was significantly different from random, we compared it with the values obtained comparing random partitions, following the methodology of Hubert and Arabie (1985). We generated 1000 randomizations of the land-use partition, grouping NILs into clusters with the same size and in the same number as in the land-use partition based on community detection. Analogously, we produced 1000 randomizations of the CDR partition. We then pairwise compared the randomizations through the Jaccard index and we assessed the differences between the 1000 values obtained in this way and the similarity score based on the land-use and the CDR partitions.

To assess the robustness of our outcomes, we repeated the analysis on CDRs on both the daily networks and the temporally-aggregated network by using community detection via modularity maximization (Newman 2006), a method that does not require to fix the number of desired communities in advance. For each day, we built a similarity matrix $S_{CDR}(t)$, whose elements were computed as:

$$s_{CDR}(Y_i(t), Y_j(t)) = 1 - \frac{d_{CDR}(Y_i(t), Y_j(t))}{\max[d_{CDR}(Y_i(t), Y_j(t))]} \quad (7)$$

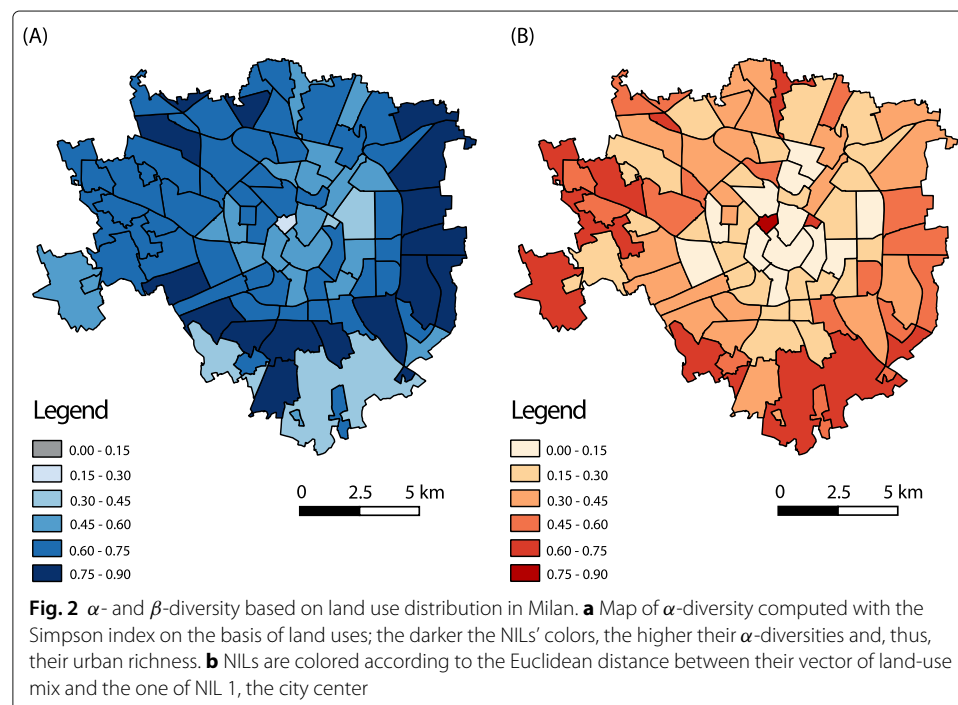
where $\max[d_{CDR}(Y_i(t), Y_j(t))]$ is the day-dependent maximum possible value of $d_{CDR}(Y_i(t), Y_j(t))$. Considering the matrices $S_{CDR}(t)$ as adjacency matrices, we generated 28 weighted, undirected networks over which we performed community detection. Finally, we aggregated the results so obtained in a network of 88 nodes (one per NIL) with

an undirected, yet weighted link between each pair of nodes that had been grouped in the same community at least once. Each weight was set equal to the fraction of days the two linked nodes were grouped in the same community. We last investigated the emergence of communities in such weighted network using again the modularity optimization algorithm.

Results

The results obtained computing the within-NIL α -diversity based on the land uses are mapped in Fig. 2A. The darker is the color of a NIL, the higher is the value of the indicator that represents, we remind to the reader, its richness in terms of land-use mix. A visual comparison between maps in Figs. 1A and 2A, which represents the actual distribution of land uses in the city, is sufficient to reveal that the low values of α -diversity in the NILs located in the city center and in the southern part of the city were due to different urban environments. While the dominance of the land use “Buildings” characterized the city center, the large prevalence of “Agricultural areas” qualified the southern part of the city (Parco Agricolo Sud Milano). Conversely, the “transition” NILs between the city center and the agricultural areas were characterized by more diversified land uses mixes, resulting in high levels of α -diversity.

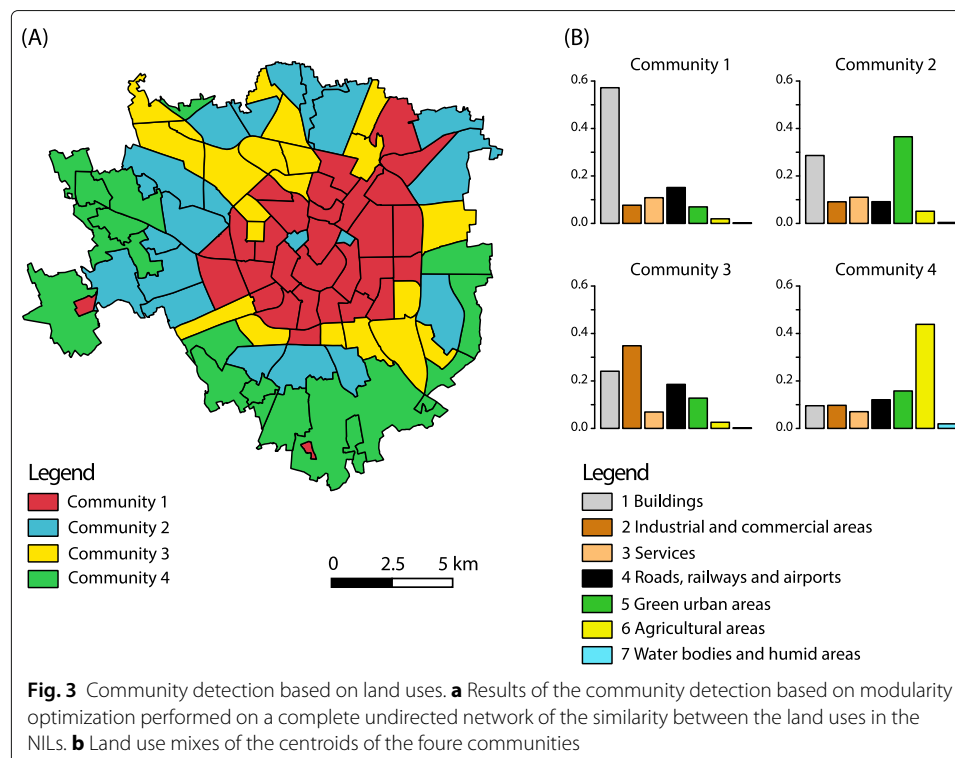
To quantify β -diversity within the city we elaborated the differences between the composition of the NILs in terms of land uses, by calculating the distances between the vectors representing their land-use mixes. Of course, a general assessment of the whole β -diversity of land uses in Milan would have required to compute the pairwise differences between each couple of NILs, so with the evaluation of $(88 \times 87)/2$ indicators. An exercise whose outcomes would not be easily synthesizable. We therefore selected the map here only the comparison between NIL 1, namely *Duomo*, shown in Fig. 1b, and each other



NIL. In Fig. 2b, NILs are colored according to their “land-use distance” from the city center (NIL 1): the darker is the color of NIL i , the higher is the distance between the vectors describing land uses of NIL i and NIL 1, thus, the “urban difference”. It emerged that NILs resulting similar to NIL 1 in terms of α -diversity were, instead, very different in terms of land-use mix. Indeed, the maximum values of distances were obtained comparing the city center with the more peripheral NILs.

Apart from the punctual differences that can be identified through fine scale analyses of distances between vectors of land-use mix, it is worthwhile to investigate whether there is any geographical consistence (e.g. adjacency or proximity) in the patterns of urban compositions of Milan’s NILs. To group the NILs that were more similar between them relative to other NILs, we used techniques from the complex network literature.

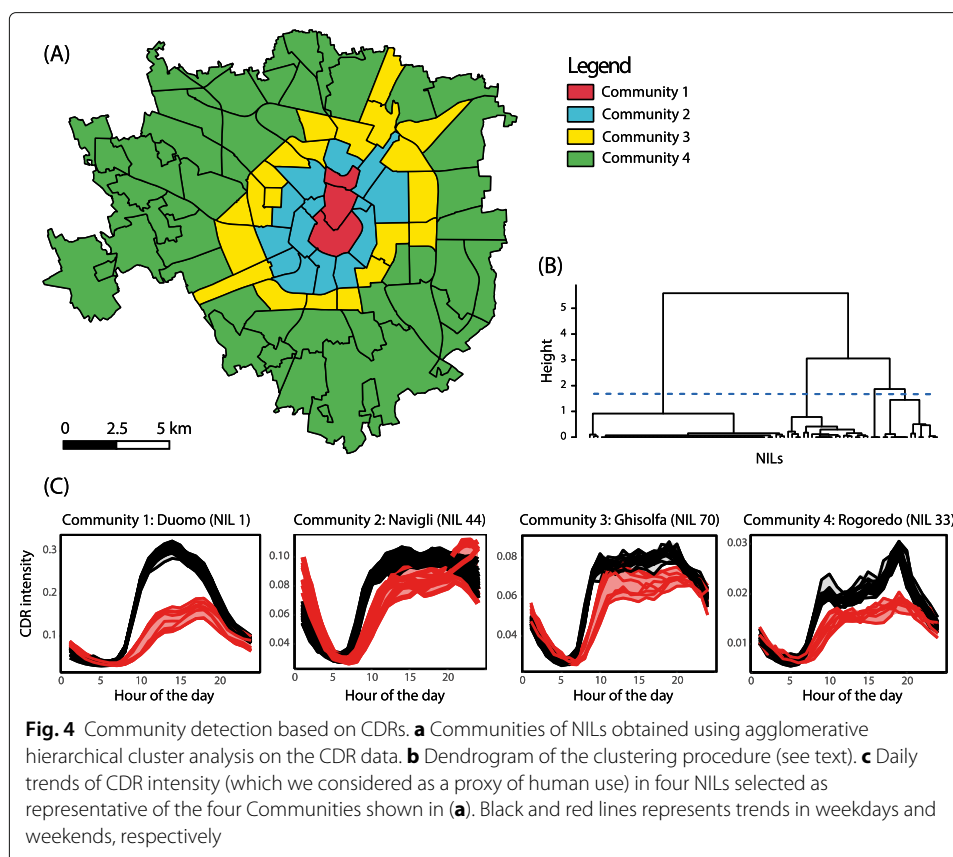
In fact, we created a complete undirected network with 88 nodes, representing each NIL connected by links that were weighted according to the similarity of their land-use mixes. On this *land-use network*, we performed community detection based on modularity optimization. Using the algorithm, we obtained that the modularity is maximized when the land-use network is split into 4 communities, which are shown in Fig. 3. Although the value of modularity obtained was not very high in absolute terms (modularity = 0.06), the a posteriori analysis of the land-use mix of the centroids of such communities and their geographical position in the city led us to consider them reliable. Interestingly, also using another algorithm for community detection, namely the Louvain method, we obtained the same subdivision in communities, which once again strengthened our results. The apparent urban pattern that emerged while looking at the Milan mapping of community detection on the land-use network is that, apart from a few explainable exceptions, the NILs grouped together are contiguous (see Fig. 3A). Second, and analogously important,



is that land-use composition within each community had a specific urban identity that fingerprinted a major land use of that part of the city (Fig. 3B). Community 1, in fact, mainly included NILs located in the very center of Milan and was mainly characterized by the prevalence of the land use type “Buildings”. In Communities 2 and 3, the predominant categories were “Green urban areas” and “Industrial and commercial sites”, respectively, although the relative abundance of “Buildings” was not negligible either. Notably, the two NILs belonging to Community 2 and surrounded by NILs belonging to Community 1 include indeed two large city parks (Giardini di Porta Venezia and Parco Sempione, respectively).

The Communities 2 and 3 designed a sort of circle around the center and divided it from the last community, labeled as 4, which grouped the NILs in the agricultural area. When we repeated the analysis via hierarchical clustering, the obtained partition of the city in four clusters was consistent to the one presented in Fig. 3, but the obtained clusters had very different size. In particular, almost all NILs in the city periphery (52 NILs in total) were grouped together within a single cluster, comprehensive of Communities 2 and 3 of Fig. 3. To the eyes of people living in Milan, the refinement offered by the network-based approach for city partitioning is relevant and close to citizens’ experience. We therefore focused on the outcomes generated with the modularity maximization algorithm.

As explained in the “Materials and methods” section, in order to analyze the actual presence of citizens within each of the NILs throughout the day, we relied on anonymized phone data. More precisely, we used the CDR dataset to evaluate with a method also mutated by complex networks theory the possible emergence of patterns of occupation. We built a daily temporal network of similarity between the time series of CDR intensities in the NILs. More precisely, we performed, on each daily network, community detection with an agglomerative hierarchical clustering algorithm by fixing a priori the number of clusters equal to 4, so as to be coherent with the results of the analysis just performed on land uses. To measure the robustness in time of the communities, we built a network of 88 nodes representing the NILs and we traced links only between the nodes that had been grouped at least once in the same community. On each link we posed a weight equal to the fraction of times within the period the nodes connected had been in the same community. We then performed community detection with an agglomerative hierarchical clustering algorithm on the *mobile-phone network* built in this way. Consistently with the city partitioning determined above for the land uses, we kept the number of clusters equal to 4 and obtained the communities shown in Fig. 4A. The concentric urban structure of Milan emerges now clearly through four almost circular communities of NILs. The dendrogram in Fig. 4B helps to read behind the performed clustering, since the horizontal line shows where we cut the tree to obtain the 4 communities. On the left part of the dendrogram, Fig. 4B reveals the presence of a conspicuous group of nodes that, according to the used metrics, are very similar between them and were grouped into the same community. The rest of nodes are grouped into smaller (and perhaps less sharply defined) clusters. Figure 4C shows the hourly trend of CDR intensity during workdays (black lines) and weekends (red lines) of four representative NILs, one for each community shown in Fig. 4A. Interestingly, we found that NILs belonging to Community 1 (the most central part of Milan) are utilized by citizens especially during the central hours of workdays, with a unimodal, bell-shaped temporal pattern. This pattern can be explained considering the many offices and shops located in this area of the city as well

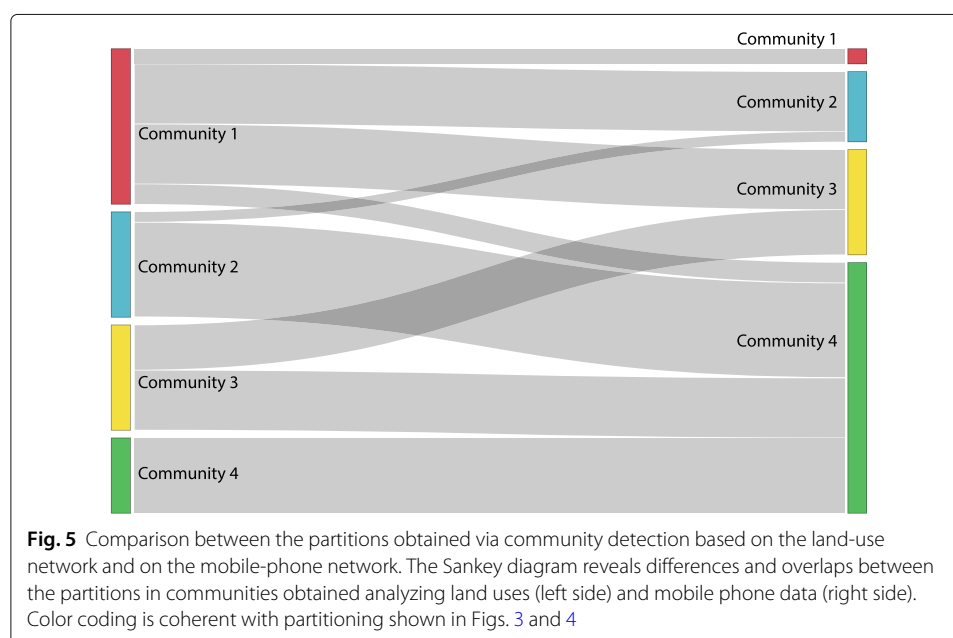


as human mobility that crosses the city center. It is noticeable that the CRD intensity during weekends is considerably lower compared to workdays and that it peaks in the late afternoon, rather than at lunch time. Community 2 circularly surrounds Community 1 and contains the internal ring road of Milan (called “Circonvallazione interna”, at the border of Municipio 1). It includes NILs whose peculiarity is that of being used by people in the second part of the day (during evening and night), especially during weekends, more than other NILs are. Indeed, in NILs like *Navigli* there are many bars and restaurants and most people hang out there at night time. Patterns of occupation in NILs belonging to Community 3 do not display much difference between workdays and weekends and are relatively constant throughout the day. This is probably due to the traffic over the external ring road of Milan (“Circonvallazione esterna”), which is almost always busy, and that surrounds the core of the city. Finally, Community 4 includes NILs that host ways of entrance to/exit from the city for citizens living in the larger metropolitan area (“la Grande Milano”). The trends in CDR activities are characterized there by two peaks in workdays in correspondence to the hours when people commute towards either their workplace (morning peak) or their home (evening peak).

We also tried to obtain a partition of the city with respect to human activities registered via CDRs by using the network-based approach of modularity optimization. Such method (whose results are not shown here) divided the city into three (instead of four) groups that were quite unbalanced in terms of numbers of NILs included: the city central part (34 NILs), the periphery (52 NILs) and a third cluster consisting of only 2 NILs (both

located at the border between the other two larger clusters). The agglomerative hierarchical clustering method used in Fig. 4 thus returned a richer description of the city than the one that could have been obtained via the network-based algorithm, because it was able to detect the presence of concentric clusters whose structure is coherent with the urban texture that Milan has since centuries (as we have seen in the Introduction).

Given the two different partitions of the city, based on community detection in the land-use network (see Fig. 3) and in the mobile-phone network (see Fig. 4), there was a need to reconcile them or at least to understand and to discuss their level of overlapping. A simple, very synthetic index is the Jaccard similarity index described in the “Materials and methods” section, that amounted to 0.26 in the current case. To assess its significance, we compared the measured score with the values generated over 1000 different randomizations of the partitions and we found out that the value we obtained for our case study was systematically (i.e. 100% of times) greater than the random values (mean and standard deviation equal to 0.20 and 0.008, respectively). This result showed the presence of a correlation between the two land-use and the CDR partitions. The Sankey diagram in Fig. 5 graphically illustrates how nodes/NILs were identified to be part of different communities in the two partitions generated by analyzing the (weighted) networks of land uses and mobile phone data (CDRs). Rectangular colored boxes on the left and right extremes of the grey bands indicate, respectively, nodes/NILs belonging to the communities obtained from analysis of the land-use network (shown in Fig. 3) and on the mobile-phone network (Fig. 4). The height of each rectangle is proportional to the number of nodes in the corresponding community. The width of each grey band is instead proportional to the number of NILs shared by the land-use community on the left and the mobile-phone community on the right. As it can be seen, the overlap between the two partitions is only partial. Interestingly, NILs in Community 1 of the land-use partition (on the left in Fig. 5), which are located in the very center of Milan (see the red NILs in Fig. 3), could belong to each of the communities detected in the CDR partitioning (on the right).



This implies that, although pretty similar in terms of land uses, the different NILs in the center of Milan are quite "patchy" and not very homogeneous in terms of temporal patterns of human occupancy through the hours of the day. This situation was analogous to the one emerging for Community 4 in the CDR partition, whose nodes were present in all the communities in the land-use partition. Such NILs (green in Fig. 4) resulted to be characterized by similar patterns in terms of mobile phone data, but different in terms of land uses. Figure 5 also helps to clarify that the NILs were more evenly distributed in the communities obtained from the analysis on land uses than from the CDR partitioning.

Conclusions

The ever-changing urban environment of a metropolis is shaped by phenomena occurring over very different spatial and temporal scales. Rapid, but long-term land-use changes are in fact coupled with the highly intricate patterns of human mobility occurring even on very short (i.e. hourly) temporal scales. Data on land uses are often available with no restrictions nowadays, since many EU countries have developed land use and land cover maps, especially in the framework of the CORINE program. Conversely, the availability of communications and social media data is typically restricted to a few research teams that sign Non-Disclosure Agreements (NDAs) and research contracts with telecommunication and other private companies. The lack of open datasets limits the number of potential studies, although the almost universal adoption of mobile phones, accompanied by the fast advancing ICT technology could provide great opportunities for urban planning and management. Indeed, these data could enable a better understanding of the patterns and mechanisms of human mobility, activities, and their relationship with the urban environment. Interesting and innovative frameworks are under development, like for example the OPAL project (www.opalproject.org).

In this study, we could benefit of the Telecom Italia Big Data Challenge dataset, a multi-source dataset of Call Detail Records about the city of Milan for a period of 4 weeks. We coupled this rich dataset with the land use information and we compared the districts of the city, i.e. the NILs, by integrating land uses and temporal patterns of their human occupation. The goal of our study was to investigate the emergence of similar/dissimilar districts according to either land use or human presence habits (or both) and to reveal the different social functions of areas characterized by similar land-use compositions. Not secondary, our framework fruitfully reveals how indicators that are frequently used in ecological studies on natural ecosystems can be extended and applied to urban settings.

Community detection on similarity networks based on land-use composition of NIL and on temporal trajectories of human occupation revealed that, as obvious, the two aspects generate partitions of the city into communities that are not completely overlapped. Interestingly, our results also showed that NILs that appeared to be very similar in terms of land uses were grouped in different communities when mobile phone data were taken into account. Information on mobile phone data could thus be used to inform land use maps and reveal the actual use of the space. Both partitions in communities showed to be spatially related to the historical urban concentric structure of the city of Milan, although this aspect was more evident in the analysis of CDRs than in the analysis of land uses.

From an administrative point of view, Milan is subdivided into 9 Boroughs (Municipi) with local administrative power and responsibility on specific public services such as

schools, parks and roads. The eight Municipi (numbered as 2, 3, ..., 9 in Fig. 1c) group NILs that are characterized by spatial proximity (adjacent to each other) and are departing radially from the center of Milan (Municipio 1) toward the city periphery and border. However, as our community/cluster detection has evidenced, neighborhoods belonging to the same Municipio can be characterized by quite different land-use mixes and/or human uses. Our results thus indicate that alternative, politically relevant aggregations of neighborhoods might be possible and they might perhaps facilitate governance because of similarity. Indeed, we showed that areas with similar characteristics in terms of either land uses or human occupation (areas that we called “urban communities”) are sometimes geographically distant one from the other. Nevertheless, people living such urban communities probably experience similar problems with respect to city uses and they probably share similar needs. City maps based on land-use mixes and on human use of the urban texture, as those shown in our Figs. 3 and 4, can thus support decision makers to envision and elaborate integrated responses to citizens, possibly providing solutions that go beyond a compartmentalized planning and management (as somehow suggested by the interesting connections among NILs provided by Fig. 5). Indeed, the outcomes of this kind of analysis could reveal areas that should be administratively connected, suggesting possible strategic development measures for the city of Milan.

Acknowledgements

A.L.T. acknowledges funding from the project MASTR-SLS (PoliSocial Award 2016, Politecnico di Milano), while A.B. acknowledges funding from the project HABITAT@SCUOLA (Fondazione Cariplo). R.C. was the PI of both projects and is very grateful to the funding agencies.

Authors' contributions

RC conceived the study and designed the research performed together with AB, who coded the analysis of the land-use dataset, and with ALT, who coded the analysis of the CDRs. All authors contributed to analyze the results. AB produced the figures and wrote the article with RC. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Dipartimento di Elettronica, Informazione e Bioingegneria (DEIB), Politecnico di Milano, Via Ponzio 34, 20133 Milan, Italy.

²African Institute for Mathematical Sciences, AIMS-Senegal Km2 route de Joal (Centre IRD), 1418 M'bour-Thies, Senegal.

Received: 11 June 2019 Accepted: 10 December 2019

Published online: 30 December 2019

References

- Arnaboldi M, Brambilla M, Cassottana B, Ciuccarelli P, Vantini S (2017) Urbanscope: A lens to observe language mix in cities. *Am Behav Sci* 61(7):774–793
- Barlacchi G, De Nadai M, Larcher R, Casella A, Chitic C, Torrisi G, Antonelli F, Vespignani A, Pentland A, Lepri B (2015) A multi-source dataset of urban life in the city of milan and the province of trentino. *Sci Data* 2:150:055
- Blondel V, Guillaume JL, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. *J Stat Mech Theory Exp* 2008(10). <https://doi.org/10.1088/1742-5468/2008/10/P10008>
- Bonvesin de la Riva (1228) De magnalibus Mediolani. translated by Chiesa P. (2009) as Le meraviglie di Milano. Mondadori
- Cho E, Myers SA, Leskovec J (2011) Friendship and mobility: user movement in location-based social networks. In: Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM. pp 1082–1090. <https://doi.org/10.1145/2020408.2020579>
- Chouakria AD, Nagabhushan PN (2007) Adaptive dissimilarity index for measuring time series proximity. *Adv Data Anal Classif* 1(1):5–21
- DeLong D (1996) Defining biodiversity. *Wildl Soc Bull* 24(4):738–749
- Feranec J, Soukup T, Hazeu G, Jaffrain G (2016) European landscape dynamics: CORINE land cover data. Taylor & Francis. <https://doi.org/10.1201/9781315372860>
- Fisher P (1997) The pixel: a snare and a delusion. *Int J Remote Sens* 18(3):679–685
- Fortunato S (2010) Community detection in graphs. *Phys Rep* 486(3-5):75–174
- Fortunato S, Hric D (2016) Community detection in networks: A user guide. *Phys Rep* 659:1–44
- Gong P, Howart PJ, et al. (1990) The use of structural information for improving land-cover classification accuracies at the rural-urban fringe. *Photogramm Eng Remote Sens* 56(1):67–73
- Gonzalez MC, Hidalgo CA, Barabasi AL (2008) Understanding individual human mobility patterns. *Nature* 453(7196):779

- Hawelka B, Sitko I, Beinat E, Sobolevsky S, Kazakopoulos P, Ratti C (2014) Geo-located twitter as proxy for global mobility patterns. *Cartogr Geogr Inf Sci* 41(3):260–271
- Hoteit S, Secci S, Sobolevsky S, Ratti C, Pujolle G (2014) Estimating human trajectories and hotspots through mobile phone data. *Comput Netw* 64:296–307
- Howison J, Wiggins A, Crowston K (2011) Validity issues in the use of social network analysis with digital trace data. *J Assoc Inf Syst* 12(12):2
- Hubert L, Arabie P (1985) Comparing partitions. *J Classif* 2(1):193–218
- Jaccard P (1901) Distribution de la flore alpine dans le bassin des dranses et dans quelques régions voisines. *Bull Soc Vaudoise Sci Nat* 37:241–272
- Louail T, Lenormand M, Ros OGC, Picornell M, Herranz R, Frias-Martinez E, Ramasco JJ, Barthelemy M (2014) From mobile phone data to the spatial structure of cities. *Sci Rep* 4:5276
- Lu D, Weng Q (2006) Use of impervious surface in urban land-use classification. *Remote Sens Environ* 102(1–2):146–160
- Mariotti I, Pacchi C, Di Vita S (2017) Co-working spaces in milan: Location patterns and urban effects. *J Urban Technol* 24(3):47–66
- Mariotti I, Brouwer AE, Gelormini M (2018) Is milan a city for elderly? mobility for aging in place. *Tema J Land Use Mobil Environ* 2:95–104
- Morandi C (2007) Milan. The great urban transformation. Marsilio
- Newman M (2010) *Networks: An Introduction*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199206650.001.0001>
- Newman ME (2006) Modularity and community structure in networks. *Proc Natl Acad Sci* 103(23):8577–8582
- Pei T, Sobolevsky S, Ratti C, Shaw SL, Li T, Zhou C (2014) A new insight into land use classification based on aggregated mobile phone data. *Int J Geogr Inf Sci* 28(9):1988–2007
- Secchi P, Vantini S, Vitelli V (2015) Analysis of spatio-temporal mobile phone data: a case study in the metropolitan area of milan. *Stat Methods Appl* 24(2):279–300
- Shaban M, Dikshit O (2001) Improvement of classification in urban areas by the use of textural features: the case study of lucknow city, uttar pradesh. *Int J Remote Sens* 22(4):565–593
- Shannon CE (1948) A mathematical theory of communication. *Bell Syst Tech J* 27(3):379–423
- Simpson EH (1949) Measurement of diversity. *Nature* 163(4148):688
- Sørensen T (1948) A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on danish commons. *Biol Skr* 5:1–34
- Ward JH (1963) Hierarchical grouping to optimize an objective function. *J Am Stat Assoc* 58(301):236–244
- Whittaker RH (1972) Evolution and measurement of species diversity. *Taxon*:213–251. <https://doi.org/10.2307/1218190>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)
