



Unraveling the hidden organisation of urban systems and their mobility flows

Riccardo Gallotti^{1*} , Giulia Bertagnolli^{1,2} and Manlio De Domenico¹

*Correspondence:

rgallotti@gmail.com

¹CoMuNe Lab, Fondazione Bruno Kessler, Via Sommarive 18, 38123 Povo (TN), Italy

Full list of author information is available at the end of the article

Abstract

Increasing evidence suggests that cities are complex systems, with structural and dynamical features responsible for a broad spectrum of emerging phenomena. Here we use a unique data set of human flows and couple it with information on the underlying street network to study, simultaneously, the structural and functional organisation of 10 world megacities. We quantify the efficiency of flow exchange between areas of a city in terms of integration and segregation using well defined measures. Results reveal unexpected complex patterns that shed new light on urban organisation. Large cities tend to be more segregated and less integrated, while their overall topological organisation resembles that of small world networks. At the same time, the heterogeneity of flows distribution might act as a catalyst for further integrating a city. Our analysis unravels how human behaviour influences, and is influenced by, the urban environment, suggesting quantitative indicators to control integration and segregation of human flows that can be used, among others, for restriction policies to adopt during emergencies and, as an interesting byproduct, allows us to characterise functional (dis)similarities of different metropolitan areas, countries, and cultures.

Keywords: Complex networks; Integration; Segregation; Human mobility; Urban systems

1 Introduction

Cities are complex systems embedded in the physical space which process information, evolve and adapt to their environment [1]. To understand how complex systems – and cities more specifically – operate, it is thus important to quantify how information is processed in terms of integration and segregation. To this aim, on the one hand many relevant network descriptors have been introduced, based either on topological features or on dynamical ones, or both. On the other hand, integration has been reflected either in how information flow is accounted for by more complex topological models where multiple relationships co-exist simultaneously [2–5], namely multilayer systems [6, 7], or in causal effects observed in the time course of systems' units [8–17].

Concerning the topological analysis of classical single-layer networks, to date a clear definition of integrated and segregated information flow is still debated and many proxies

© The Author(s) 2021. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

are used across a broad spectrum of disciplines, ranging from neuroscience to social and urban sciences [18–33], often indicating with the same name very different concepts.

The recent availability of a large amount of human-generated data enables the analysis of urban systems from different perspectives which could not be even considered until a few years ago [34]. Consequently, models and analytical tools inspired by complexity science are proliferating. More and more examples are providing convincing evidences of their fruitful application to real cities [35–40]. Applications range from human mobility [41–44] and traffic congestion [45–49], to energy consumption [50], air quality [51, 52] and climate [53], health and well being [54–57], and the associated topic of accessibility to important facilities like hospitals [58]. Indeed, the city can be seen as a growing complex system [59, 60] whose spatial organisation [61, 62] dynamically experiences a transition from monocentric to polycentric [63, 64].

The relative ease of accessing large and detailed data sources describing at the same time the structure and the function of urban systems, puts them in the position of becoming a paradigmatically example over which we can identify the right methodologies allowing us to understand the behaviour of spatially embedded complex systems. A particularly relevant perspective is offered by activity-aware information [65], such as the one provided by users of Foursquare – a leading location intelligence platform – which allows people to investigate human flows at different scales and thus to reconstruct the functional network of cities with great level of detail [66] and to classify existing activities into a few representative macro-categories (see Methods for details).

In this work, we stratify those human activities to build the functional networks describing the human movements across the urban space of 10 different metropolitan systems spread over three continents. To gain novel insights about the functional organisation of the underlying urban ecosystem, we build a multilayer network [4, 7], where the flows encode how users move between venues of the same macro-category (e.g., from a pub to another one) and between venues of different macro-categories (e.g., from a pub to a cinema). In the following, we will refer to *intra-layer* flow to indicate movements of the first type, and to *inter-layer* flow to indicate movements of the second type.

Our main goal is to better characterise the functional organisation of a city through the lens of network science. To this aim we measure to which extent different areas of the city facilitate human flows – i.e., functional integration – and to which extent there are separate clusters of areas characterised by within-cluster flows larger than between-cluster flows – i.e., functional segregation – (see Methods for details) [67]. By considering those measures simultaneously, it is possible to characterise how well human flows mix through the city according to the existing distribution of venues and the way residents use them. In fact, the dichotomy between integration and segregation – often improperly used as antonyms – is relevant for improving our understanding of the interplay between the urban structure, social relationships and human behaviour.

At the same time, to investigate the coupling between the structure of a city and the dynamics of its inhabitants, we also study the integration and segregation of the structural networks of these cities reconstructed from Open Street Map [68]. See Fig. 1 and Methods for more details on the definition of the structural and functional networks.

2 Results

2.1 Overview of the data sets

The Foursquare data made available for the Future Cities Challenge [69] describe 24 months of check-ins collected between April 2017 and March 2019 (included). The use of these dataset faces multiple limitations, discussed in details in the Methods section.

The 10 world mega-cities included in the challenge are Chicago, Istanbul, Jakarta, London, Los Angeles, Tokyo, Paris, Seoul, Singapore and New York City (represented as example in Fig. 1 right). The extensive characteristics of the datasets are shown in Table 1. The flows between different areas are derived by subsequent anonymised check-ins to the Foursquare's location-based services and coarse grained with a $500\text{ m} \times 500\text{ m}$ granularity

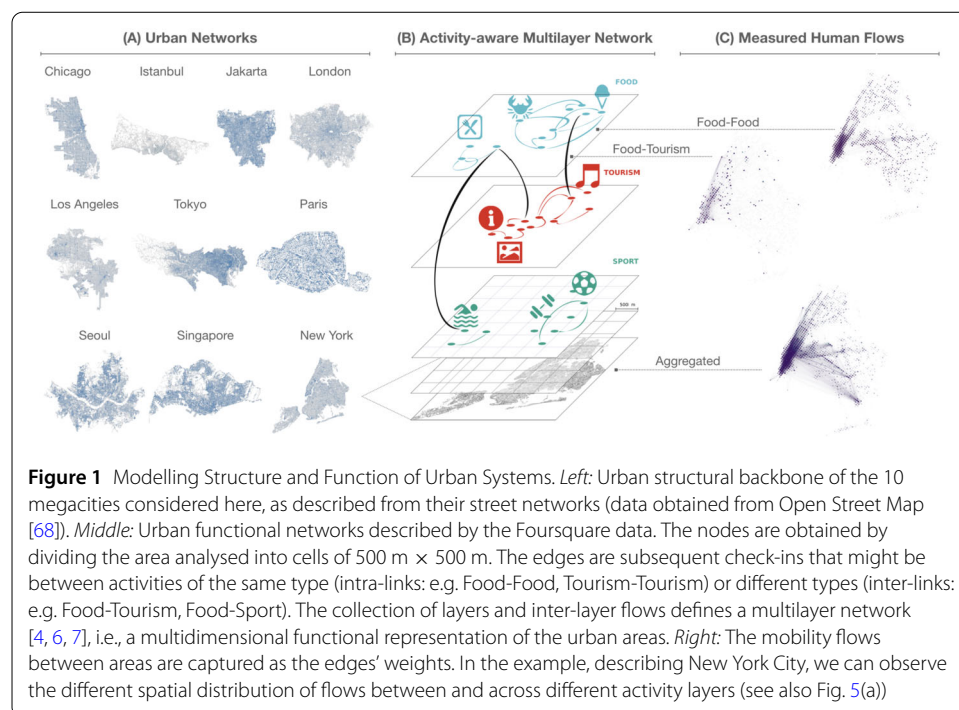


Table 1 Foursquare data set extensive characteristics. The figures here are aggregated for all layers and comprise all 24 months. The linear size L is here estimated as the square root of the total area covered by the data after the aggregation into squares of $500\text{ m} \times 500\text{ m}$. Please note that the value of population for the city of Paris here corresponds to the Grand Paris Metropolitan areas that is the territory roughly covered by the data. Other population correspond to the municipality area (or the national area for the case of Singapore)

City	#Venues	#check-ins	L (km)	Population
Chicago	13,904	10,629,110	18.9	2,705,994
Istanbul	113,752	13,083,383	39.3	15,029,231
Jakarta	21,813	9,281,181	27.6	10,517,000
London	22,689	10,146,880	24.0	18,825,001
Los Angeles	15,868	10,362,146	23.6	3,990,456
New York City	32,971	11,048,584	19.3	8,398,748
Paris	13,588	9,521,723	16.6	7,020,210
Seoul	15,545	9,347,489	18.1	9,806,000
Singapore	23,324	9,691,517	23.5	5,638,700
Tokyo	57,810	11,545,155	30.4	13,515,271

(see Fig. 1 middle, and Methods). In the data provided, check-ins are already aggregated by couple of venues (origin and destination), month.

The Open Street Map data has been obtained using the OSMNX python library [68] (see Fig. 1 left). The urban area selected has been set to match the cells covered by the Foursquare venues. The structural network has been reduced to a lattice-like form of the same granularity as the urban flow, so that all nodes in the structural network find their correspondence in the functional network. Differently from the functional one, the structural network is purely topological, as an undirected link between two cells exists if at least one street connects the two areas.

2.2 Quantifying integration and segregation

As previously mentioned, we characterise the organisation of the city through measures of integration and segregation. To avoid confusion in the reader, it is worth remarking that our measures of integration and segregation are those established in the field of network neuroscience [28], rather than being associated to the traditional social concepts, and are thus not related to population or cultural mixing [70], but only to how cities are lived by their users. Integration quantifies, in terms of information exchange efficiency, the ability of a city to favour the flow of people across its areas, and is measured by means of the global communication efficiency GCE, specifically normalised to correctly compare the efficiencies of weighted and un-weighted networks [71]. Segregation, on the other hand, evaluates the strength of segregated communities, areas of the city with strong flows inside the area and weak inter-areas flows and is estimated as the maximal modularity Q^* [72] of the network (see Methods for further details).

2.3 Structural vs functional networks

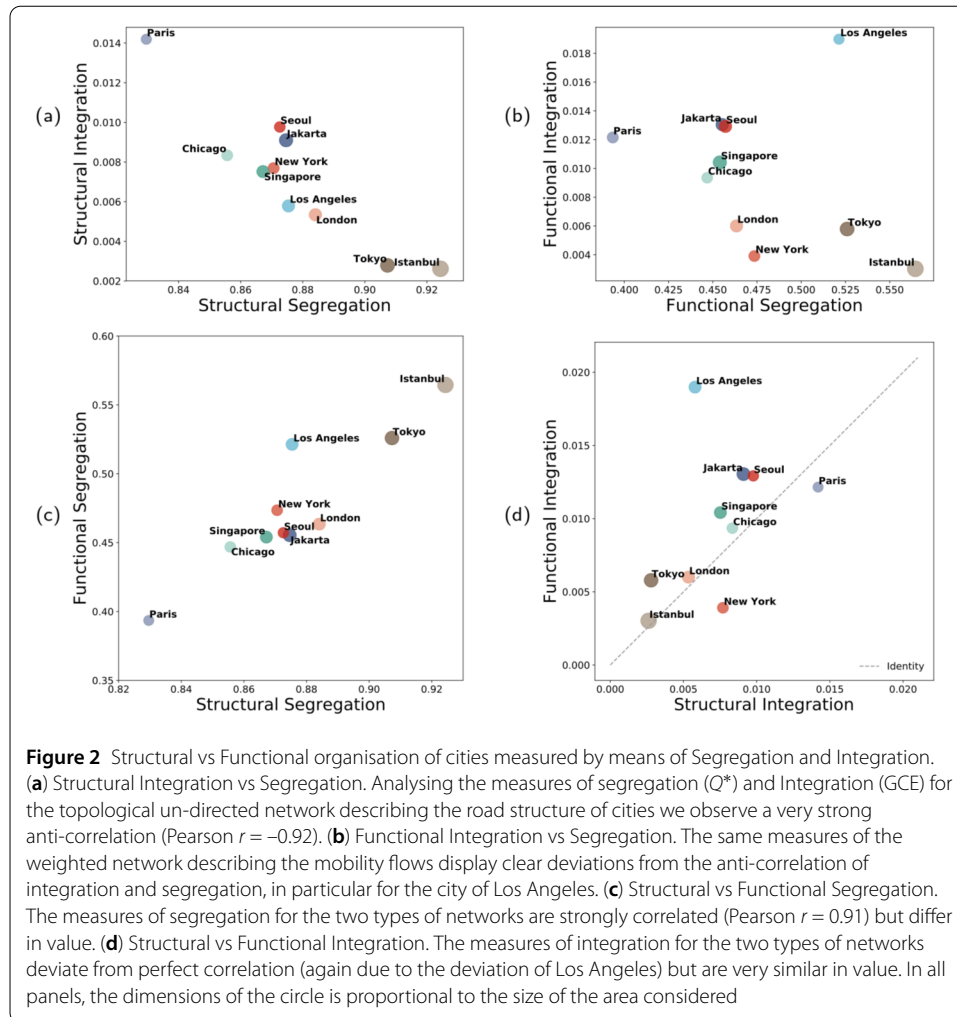
Having identified two measures suitable for comparing different cities and types of networks, we begin our analysis by mapping the link between integration and segregation in both the structural road networks and the single layer flow networks, obtained aggregating for each city inter-layer and intra-layer flows over the whole temporal extension of the dataset, which describe the functional use of the city by individuals.

The results, displayed in panels (a) and (b) of Fig. 2, suggest that, in general, higher values of segregation are associated to lower values of integration, as common sense would suggest. However, we also observe clear deviations from this trend, the major one being the functional network for the city of Los Angeles appearing to be much more integrated than what would be expected by its relatively high level of segregation.

Of particular interest is the comparison of structural and functional properties of the same systems (panels (c) and (d) of Fig. 2). The segregation, estimated through the lens of modularity, seems to systematically deviate, with the functional flow network being less segregated than the structural network even if the values for the different cities are highly correlated. The integration instead, studied with an indicator specifically developed for allowing this type of comparisons [71] corresponds also numerically for the very different structural and functional network, and this perfect correspondence reveals a divergence between structural and functional properties of the city of Los Angeles.

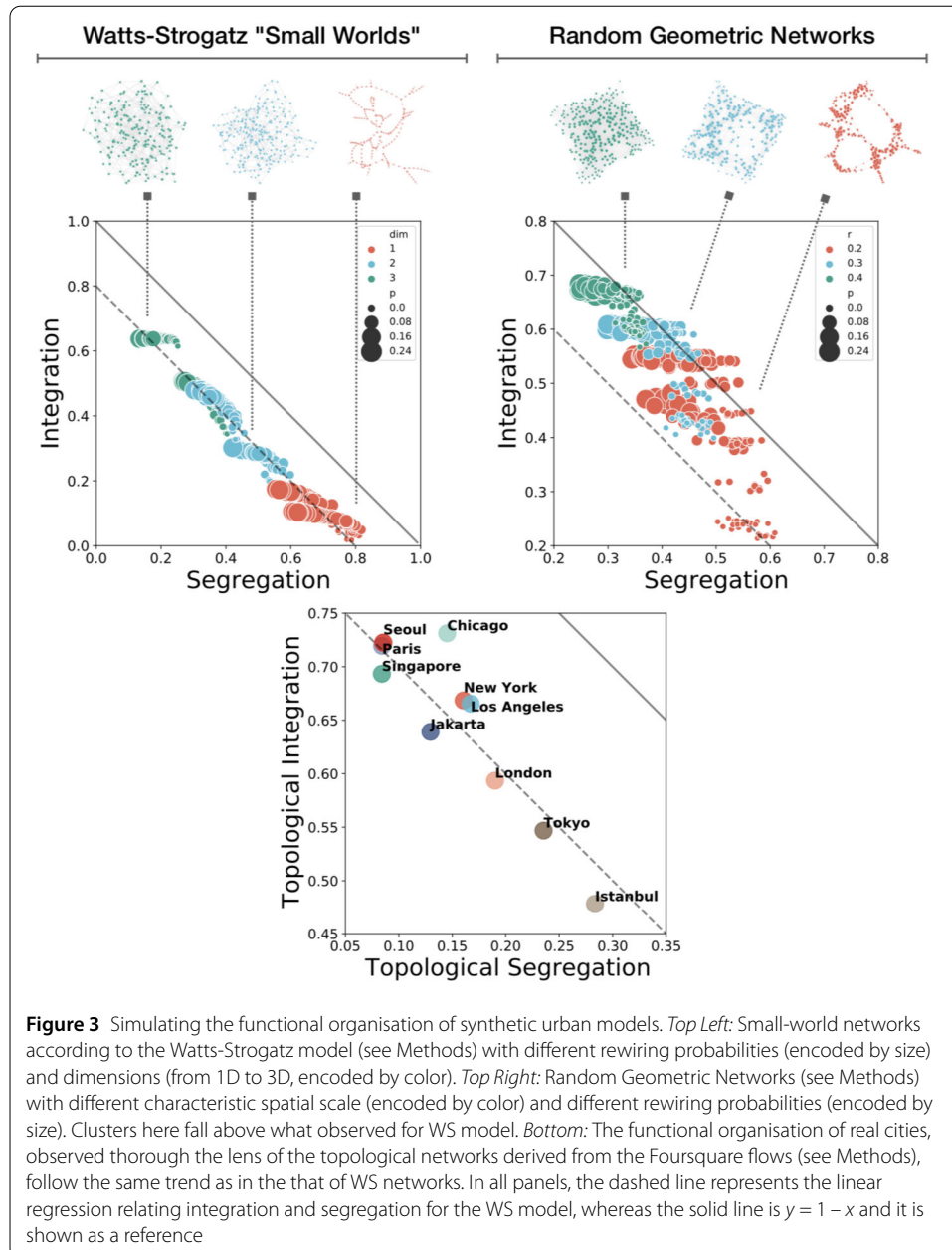
2.4 What determines integration and segregation

In order to understand what lies behind the pattern of anti-correlation between integration and segregation observed in Fig. 2, we generate spatially embedded networks that



attempt at reproducing the key feature of the urban functional networks using two widely used null models: (i) the Watts-Strograts (WS) small world networks obtained through rewiring of a regular lattice; (ii) the Random Geometric Networks (RGN) obtained by linking two randomly placed points if their distance falls below a fixed threshold r (see Methods). Also for the RGNs we proceeded with random rewiring and, in both cases, the probability of rewiring is indicated by p .

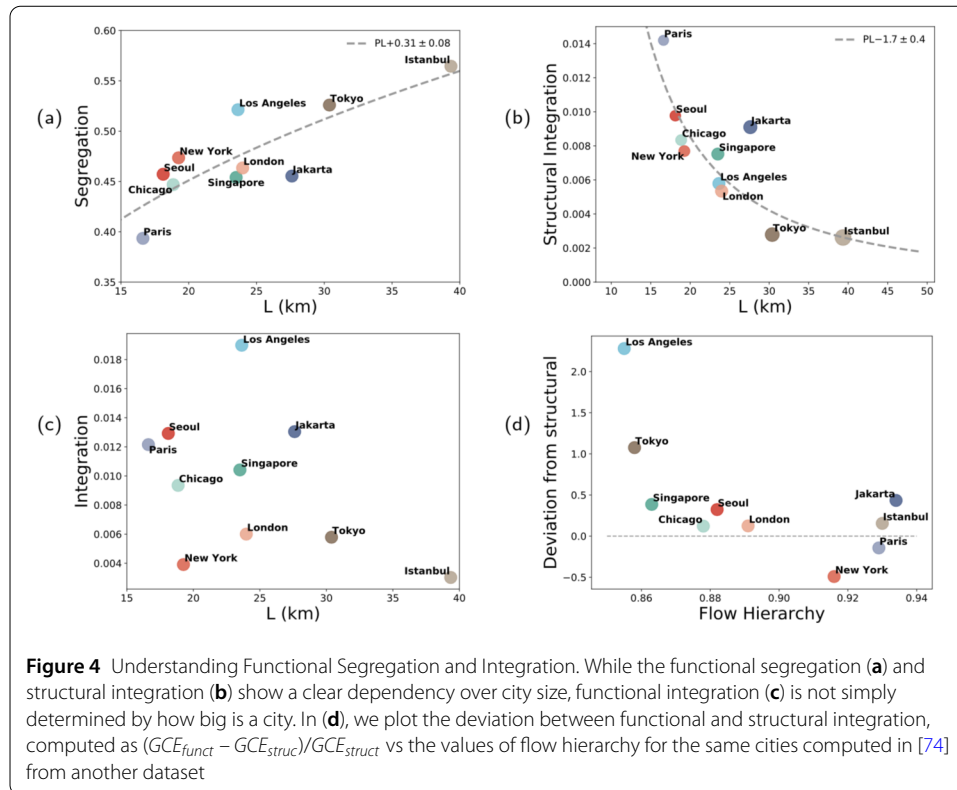
In Fig. 3 we observe that for both null models we reproduce the same anti-correlation pattern observed for real networks, but also see that rewiring is strongly reducing segregation and increasing integration in a way that breaks the linear relationship between the two quantities. Moreover, since by generating them we can control all features of the WS and RGN networks considered, we are able to isolate the leading factors behind this pattern. For WS, integration grows and segregation drops as the network dimensionality grows. The same happens for RGN as the radius r grows. Indeed, both increased dimensionality and r leads to generating networks with a higher edge density, allowing us to isolate the important role played by edge density in dictating the state of integration and segregation of spatial networks. For topological (i.e. not weighted) networks the Global Communication Efficiency, used to estimate integration, grows as the edge density grows.



This is indeed what we observe in Additional file 1, Fig. 1 while a less tight correlation can be observed for segregation in Additional file 1, Fig. 2.

However, the values observed in Fig. 2(b) deviate sensibly by those describing the networks we generated in Fig. 3. This because the urban functional networks are defined as weighted networks, while our null models do not describe weights. Indeed, if we reduce the urban functional networks to a purely topological undirected network, we see in Fig. 3 (right) that the numerical values of topological urban functional networks correspond to those described by WS model (dashed line).

To isolate the driving factors determining a city integration and segregation we have to expand from the ideal world of synthetic models and find instead guidance from the methods commonly adopted to investigate the physics of cities. Many properties of cities are



known to be power law functions of population size [59]. Here, we are not in the position of deriving with precision the population in the area defined by the Foursquare data, and we use instead as measure of the city size the square root of the area covered ($L = \sqrt{A}$) which is also a proxy for the average length of a trip in a city [63]. We therefore plot in Fig. 4(a), (b), (c) the values of Functional Segregation and Structural and Functional integration against L (see Additional file 1, Fig. 3 to see how other network indicators scale). In our case, the sizes of the cities considered are not diverse enough for initiating a meaningful discussion based on the value of the exponents observed (that are reported in panels (a) and (b) only to support future studies on the matter). We focus indeed on the fact that a power law scaling is able to explain most of the variance observed for Functional Segregation ($R^2 = 0.67$) and Structural Integration ($R^2 = 0.71$) but totally fails at predicting the values of Functional Integration ($R^2 = 0.05$). In other words, size matters. In particular it matters for functional segregation, also linked to the total flow circulating over the network (Additional file 1, Fig. 2(c)): in fact, as observed in [73], it can be expected to grow proportionally with population. However, there is something more that is strongly influencing functional integration and makes it deviate from the structural integration (as seen in Fig. 2(d)). This extra factor is determined by how flows are distributed in the network. To show this, in Fig. 4(d) we compute how much the weighted functional networks deviate from the values estimated from the structural network as $(GCE_{funct} - GCE_{struc})/GCE_{struc}$, and plot it against the flow hierarchy estimated for the same city from another dataset (numerical values computed and obtained from [74]). A low flow hierarchy indicates that larger fraction of movements are expected to be between strong mobility hubs and less active areas. This means that, in general, excess of integration is expected when marginal areas are more strongly connected. This appears similar to what observed in hierarchi-

cal modular brain networks, which are locally segregated, but global neuronal operation integrate segregated functions [75].

Lastly, using the RGN model we also measured the importance of the spatial extension of the network. Fixing the radius below which nodes are connected, we find (see Additional file 1, Fig. 4) that the largest the area ($A = L^2$) covered by a square RGN the more the network is segregated and the less it is, at the same time, integrated. Indeed, here again integration and segregation seem to be very strongly correlated and increasing the radius have a similar effect as reducing the spatial extension.

2.5 Cities within a city

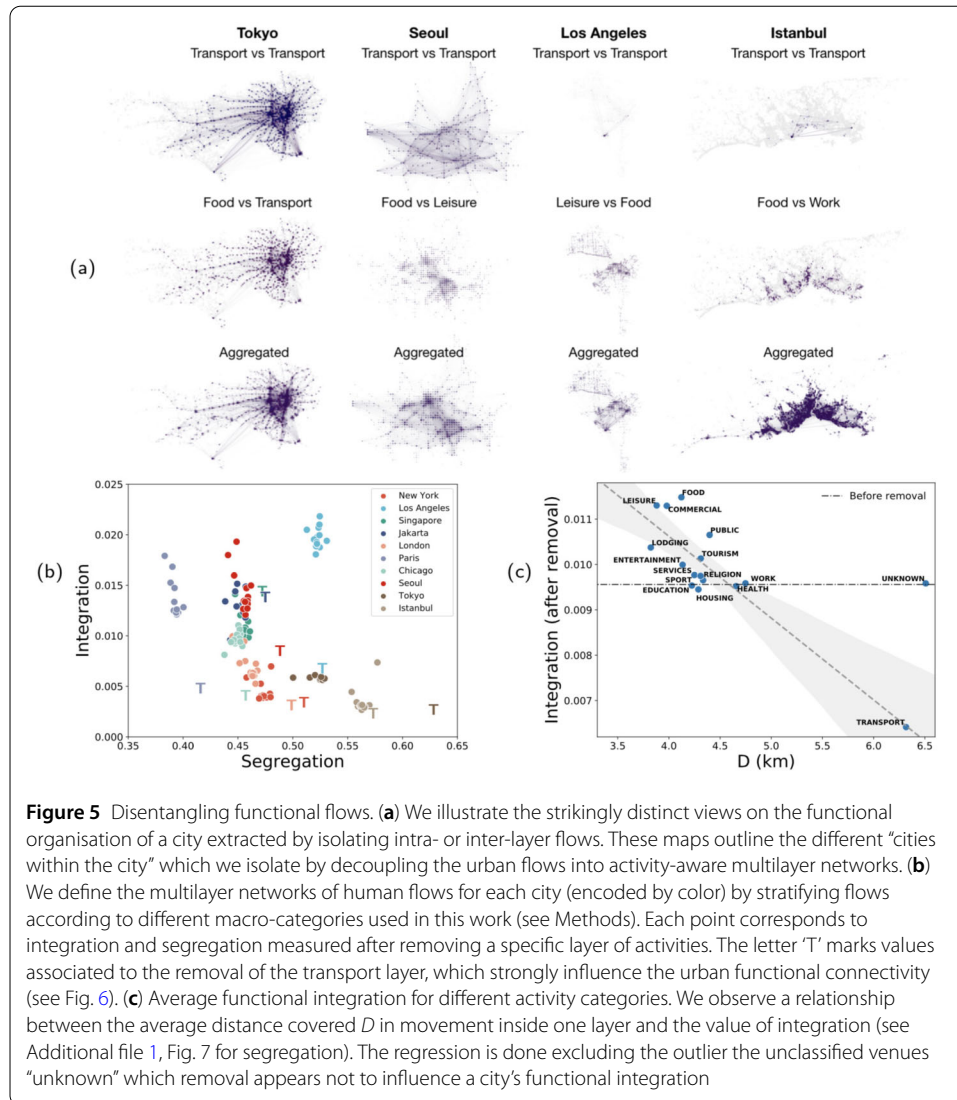
Having understood the behaviours of integration and segregation of cities at an aggregated level, is worth checking if this pattern is an intrinsic feature of urban systems or if it is proper of some specific activity layers. Indeed, the metadata of the venues include a *category* field which describes the type of venue in great detail (e.g.: Knitting Stores, Mini Golf Courses, Rock Clubs, ...). We defined a set of macro-categories we used to aggregate categories in limited number of layers (see Methods and Fig. 1 middle). Statistical information about the number of nodes and links in the different layers are provided in Additional file 1, Table I.

In Fig. 5(a) we can visually inspect some examples of activity-aware layers. Remarkably, for all the cities considered in this study, the intra-layer connectivity characterizing the transport layer provides a natural link between our functional analysis and the underlying structure of the city. In the data, however, it can be clearly seen in cities where public transport is well developed and largely used, such as Tokyo or Seoul, way more than cities where private transportation is dominant, such as Los Angeles and Istanbul.

By disentangling the mobility flows into a multilayer network structure (see Methods and Fig. 1 right), we are able to quantify the differences in the functional organisation of human flows between different types of activities or different month (see Additional file 1, Fig. 5) enabling the identification of different “cities within the city” which indeed shows clear dissimilarities in terms of both functional integration and segregation.

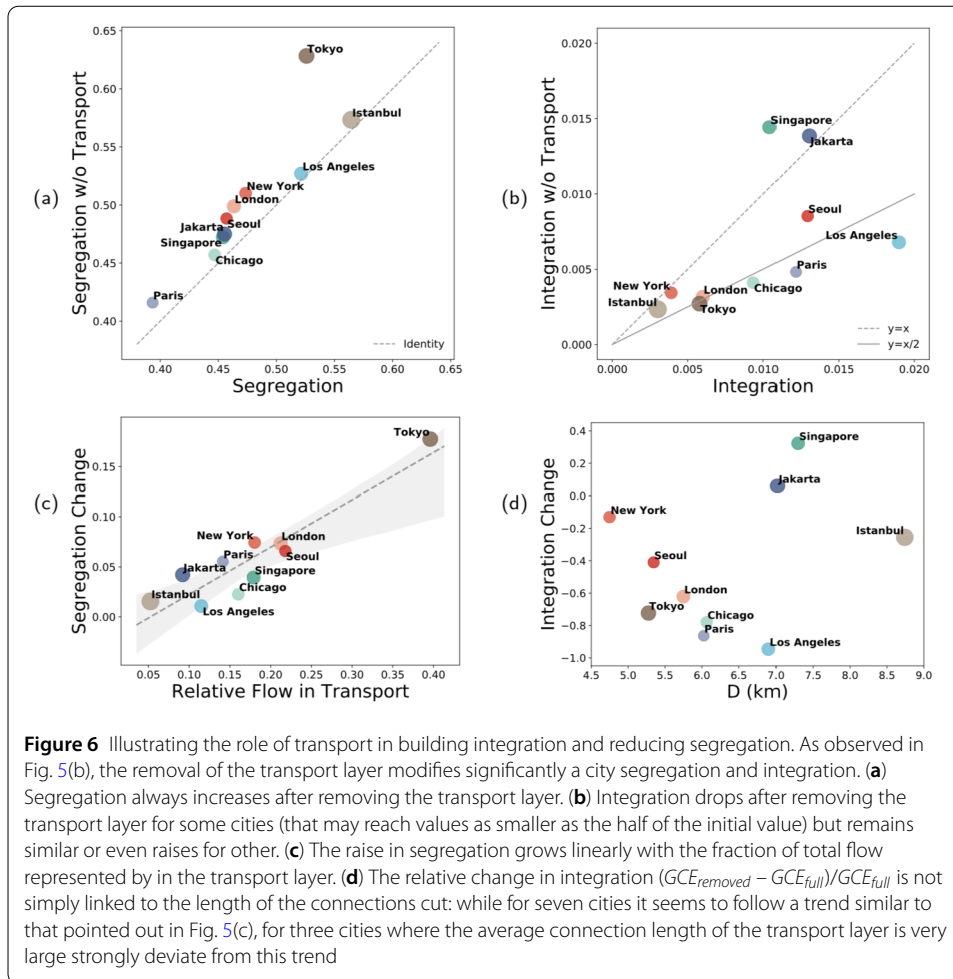
To this aim, we perform targeted attacks on each layer of the corresponding multilayer network and measure the response of the systems in terms of changes in segregation and integration. In Fig. 5(b) we observe how removing those flows coming from a specific activity type significantly changes urban functional segregation and integration. This is especially true if the activity is Transport, whose removal yields the rightmost outliers in the figure. An even stronger variation is observed in the integration and segregation restricted to movements between similar layers (see Additional file 1, Fig. 6).

To better understand these differences, in Fig. 5(c) we link the average values of integration measured for flows between the same categories across all cities with the corresponding weighted average of geographical distances between nodes. We observe a bulk of correlated points and two outliers: one the natural long-range linking layer of transportation, the other the locations not associated to a macro category and left as “unknown” (see Methods). Excluding “unknown” that does not seem to influence integration at all, we observe a clear effect: removing the transport layer strongly disrupts integration, while removing short range layers actually improves it. In Additional file 1, Fig. 7 we could conversely see how, again with the notable exception of the removal of the Transport layer, the segregation of cities remains relatively unchanged after single layer removal. The results of



this analysis points out that is possible to close restaurants, leisure and commercial activities while keeping a city functional and, possibly, even more integrated. This perspective provides new insight on the effects of restriction policies adopted during emergencies by quantifying a hidden, systemic, social costs and benefits associated to the closure of different kind of activities in time of a pandemic emergency.

It is natural observing how the transport layer represents the backbone of a city organisation, but for some cities this effect is stronger than in others. To understand these differences, in Fig. 6 we explore with more depth the difference in segregation and integration consequent to the removal of the transport layer. The effect is clear for the change in segregation (panels (a) and (c)): the increase in segregation. consequent to layer removal is proportional to how much flow pass through that layer. Things are, again, more complicated when we observe integration: for some cities, the integration drops of $\approx 50\%$ without the transport layer, while for others (notably Singapore, Jakarta and Istanbul) integration is unchanged, or even slightly increased, by the layer removal (panel (b)). These three cities have also the transport layer characterised by the longest average link distance



(panel (d)), and while for the other seven cities one might have dared to see a trend, similar to that of Fig. 5(c), linking higher drop in integration to longer connections, the presence of these three outliers suggests, another time, that microscopic details in the distribution of flows of a functional network can play a major role in determining its robustness and more general its organisation.

3 Discussion

Understanding how cities process information, here encoded by human flows, is of paramount importance for designing more efficient and smart urban systems and communities. By characterising at the same the structural and the functional organisation of 10 large urban systems in terms of well defined and normalised measures of network integration and segregation, we have shown how network-based analysis can support, and further expand, ongoing discussions about and novel understanding provided by the ICT-data driven quantitative urbanism [38].

From a modelling perspective, going beyond the antonymic dichotomy between integration and segregation by studying the Segregation/Integration diagrams allowed us to expand our understanding of the interplay between the urban structure, social relationships and human behaviour. This can be exemplified by three clear results. First, the identification of the dominant factor dominating this negative correlation (the edge density,

which is in turn a function of a city size) and forcing the deviations from it (the hierarchical structure of flows). Second, the correspondence of the empirical results with those of Small World networks shows that for modelling urban system one has necessarily to go beyond “first neighbour” transmission as long range interactions are extremely relevant to reproduce the many salient features measured from empirical data. Third, we were able to rightfully isolate, using this approach, the essential role played by the transportation layer that is pivotal for both integration (thanks to its long distance connectivity) and segregation (thanks to its large flows).

Under this lens, many features of complex megacities can be therefore understood from simple mechanisms related to geometric constraints and city’s characteristic size, with larger cities tending to be more segregated and less integrated. More in details, for growing cities, it is expected a transition from a monocentric to a polycentric organisation, characterised by a sub-linear growth of the number of hotspots with population [63]. Similarly, for both urban structural and functional networks, we provide evidence that large polycentric cities, which are characterised by a larger number of hotspots (although being the growth sub-linear they have a smaller fraction of hotspots as shown in Additional file 1, Fig. 3(d)), appear to be more segregated and less integrated than smaller, and monocentric, cities. We have highlighted, however, that a city can be much more integrated than what expected by its size if it display a low flow-hierarchy [74] and thus has more direct connections between central and marginal areas. However, the interplay between heterogeneities in the distribution of flows, spatial constraints, and the layered structure of flows, might be responsible for the emergence of peculiar integrated/segregated structures that might be reflected in the functional organisation of the city. Future research in this direction, including a wider spectrum of urban and non urban systems, is required to gain more insights on this matter.

Finally, from a more methodological point of view, our analysis highlights the importance of data sources for the analysis of the interplay between the city and its main users, i.e., the citizens. Thanks to the unique dataset of anonymised movements provided by Foursquare and the easy access to street data [68], we have been able to gain novel insights on urban and human behaviour in terms of interaction between structure and functional organisation of the system. The availability of activity-aware information, in particular, allowed the analysis of attacks targeted towards specific types of activities which unraveled the fundamental importance of transport as integrator an urban system. This result is specially relevant for policy and decision-making in time of crisis, provide new quantitative tools that allow one to identify a limited set of activities (commercial, restaurants, leisure) which can be prioritised or temporary limited to achieve a desired amount of human flows integrated across the city.

4 Methods

4.1 Limitations of this study

Our study is based on a large collection of user-generated access data to public venues. As all sources of automatically collected social data, it is affected by a series of biases that might influence our observations [76].

- *Representativeness.* The Foursquare user-base does not cover, naturally, the totality of a city population. Some public figures are available online [77], from which we can both get indirect estimates that about 13% [78] of adult social media users in the USA

used Foursquare in 2018. Since the United States about 79% of adults used social media in 2019 [79], that would make our samples for Chicago, Los Angeles and New York City covering $\approx 10\%$ of the total adult population. Naturally not all users use it regularly (see Inhomogeneity of users' behaviour), and also the representativeness will surely vary from country to country. To estimate how representativeness may translate to other cities, we can use as a proxy the check-ins per capita in the cities (see Table 1), which is more or less homogeneous, ranging between 0.8–0.9 for Asian cities to the higher values of American cities (2.5 in Los Angeles and 3.7 in Chicago). Using these proportions we can estimate that the total user base can be of the order of 2% in Asian cities.

- *Demographic bias.* The Foursquare user-base is mostly cantered around the age 18–34 and the male population is almost the double of females. The foursquare penetration is also greater penetration among users with higher income [77].
- *Inhomogeneity of users' behaviour.* Of course, not all users are active daily on Foursquare. An empirical analysis [80] describing a dataset of Foursquare check-ins collected in 2010 over 4 months via Twitter, with no spatial boundaries set, provides hints for a dishomogeneous, but somehow limited, number of checkins per users.
- *Subsampling and missing stops.* As shown again in [80], the distribution of inter-time between check-ins is long tailed. This can strongly bias the observed displacements [81]. Flows in this analysis will often not correspond to real movement but they have to be taken for what they are: subsequent checkins. For this reason, we opted to avoid focusing on the temporal disaggregation of flows that Foursquare provided on base of the hour-of-the-day and month of the arrival check-in. We decouple the functional use of a city in different months of the year of the network only to test what happen by sub-sampling the flow network.
- *Inhomogeneity of venues.* Venues are not homogeneously distributed across the city, with a larger densities in the city centres. Moreover, venues display a great inhomogeneity in the number of check-ins they capture (see Additional file 1, Fig. 8).
- *Definition of city* It is known that many urban measures may strongly depend on how the city itself is defined [82]. In the dataset provided, cities administrative areas were already selected (with the exception of Paris where it has been selected the “Grand Paris” area). In Additional file 1, Fig. 9, we test robustness of our metrics to the boundary definition by radially reducing the city area.

4.2 Geographic coarse-graining

We reconstruct the flows network by aggregating data over areal units of $500\text{ m} \times 500\text{ m}$, in all 10 cities considered. Flows are reconstructed from subsequent anonymised check-ins into Foursquare venues, ignoring the order (undirected network). Flows inside the same area have been integrated into a self-loop link only if the check-ins were between two different locations. Subsequent check-ins in the same location have been excluded from the analysis. We reconstruct the structural networks using OSMnx [68], a python library which provides a network object where nodes are the street intersection and links are defined as the stretch of road between two subsequent intersections. We coarse grained these street network to match the granularity imposed to the flow network. The short-range nature of the street network provided by OSMnx makes that these coarse grained structural maps are mostly lattice-like.

4.3 Activity stratification

We use Foursquare's rich system of categories and manually associate them to a reduced number of macro-categories (food, lodging, tourism, work, religion, services, education, health, sport, transport, entertainment, leisure, public, housing and commercial). We do not use Foursquare Venue Category Hierarchy [83], except for venue icons in Fig. 1. The few categories that did not fit any macro-category have been labelled as 'unknown'. These categories allow us to build "activity-aware multilayer networks", where activities of different types are associated to different layers of our model. Flows between activities of the same macro-category are encoded by intra-layer links, while flows between different categories are encoded by inter-layer links.

4.4 Measuring functional integration

We measure to which extent a network is integrated in terms of communication, i.e., how efficient nodes are, on average, in exchanging information, using an indicator based on the concept of shortest path. Given two areal units i and j we can reasonably assume that the efficiency ϵ_{ij} in their communication is inversely proportional to their distance d_{ij} . If d_{ij} is a topological distance, counting the number of links in a shortest-path from i to j , our assumption means that the longer the path a piece of information has to travel, the more inefficient will be the communication, since the probability that the message is corrupted along the way increases. A global descriptor of the topological communication efficiency [18] of a city is then the average pairwise efficiency of its nodes is the average shortest path length in the network

$$E = \frac{1}{N(N-1)} \sum_{i \neq j} \frac{1}{d_{ij}}. \quad (1)$$

4.5 Normalising functional integration of flow networks

For flow networks, like those analysed in this paper, given the additional information on the strength of connections distances can be very different. If the flow between two nodes is large, their distance should be, intuitively, small. For this reason, the distance averaged has to be that of weighted shortest-paths, minimising the sum of *costs* along all paths between pairs of nodes. In a flow network with edge weights representing the intensity of the connections, the costs of edges are the inverse of weights.

Unfortunately, (1) cannot be effortlessly generalised to weighted networks, since it depends on the scale of weights. Latora and Marchiori proposed a weighted efficiency descriptor in [84], rescaling the value of efficiency in $[0, 1]$ considering an idealised proxy considering an idealised proxy of G , G_{ideal} , having maximum efficiency. However, that finding the ideal proxy G_{ideal} of a network G for the normalisation of the weighted $E(G)$ is often ambiguous.

An universally valid solution for the normalisation of the global efficiency, capturing at the same time information of link existence and link weights has been proposed in [71], enabling the comparison of communication efficiency of disparate systems. The idea is that each (weighted) shortest-path in the network has a length, which is the sum of links costs along the path, and a total flow, which is the sum of the links weights. These path flows ϕ_{ij} are strictly positive for each pair of nodes (i, j) in a connected network and can be added to the original network as an artificial direct flow between i and j . In other words,

to the network G are added artificial links representing all missing shortcuts between pair of nodes, which allow to deliver the total flux through a shortest-path from origin to destination in one topological-step.

A correct normalisation of E is then possible using this network G_{ideal} resulting from a physically-grounded enrichment procedure independent from the scale of flows and from any metadata or the lack thereof. The normalised Global Communication Efficiency can be then computed as:

$$GCE = E(G)/E(G_{\text{ideal}}). \quad (2)$$

4.6 Measuring functional segregation

A usual measure of network segregation, quantifying how strongly the units are organised in into M non-overlapping blocks, is the modularity [72]

$$Q = \sum_{u \in M} \left[e_{uu} - \left(\sum_{v \in M} e_{uv} \right)^2 \right], \quad (3)$$

where e_{uu} is the proportion of links inside module u , while e_{uv} accounts for the connectivity between two distinct modules u and v . More specifically, our measure of segregation is the maximum value Q^* of the modularity that we find using the Louvain algorithm [85]. We also verify that the observed modularity is significant, by comparison with the values of Q^* computed over an ensemble of configuration models obtained reshuffling the network (see Additional file 1, Tables II and III). Finally, note that here, instead, we used the weights defined by flows. Values of Q^* for weighted and unweighted networks are indeed comparable, as opposite to what discussed above for E , and using weights here allowed us to better discern the characteristics of different layers.

4.7 Synthetic network models

We use two standard spatial network models for our analysis.

We first consider a class of networks characterised by small average geodesic distance: the Watts-Strogatz (WS) model. Starting from a regular graph, e.g., a two-dimensional lattice, each link has a probability p of being *rewired*, that is removed and re-placed randomly in the network. If p is large the resulting WS network will look more like an ER random graph than the original lattice. WS networks are also highly clustered, where nodes tend to form closed triangles. WS model are usually referred to as *small-world* networks.

Alternatively to WS, we study also the simplest network model actively involving the spatial dimension model is the random geometric network (RGN), where nodes randomly distributed in space are connected if they are closer than a fixed threshold distance. The RGNs share many important properties with regular lattices, in particular they are not “small world”. For this reason, similarly to the WS case, here also for the RGN we perform a rewiring with probability α .

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1140/epjds/s13688-020-00258-3>.

[Additional file 1](#). Supplementary information (PDF 677 kB)

Acknowledgements

The authors thank Foursquare for granting access to the data set used in this study and acknowledge Matthew Kamen, Renaud Lambiotte, Jesse Lane, Anastasios Noulas, Cecilia Mascolo, Vsevolod Salnikov, Sarah Spagnolo and Adam Walksman for organising the Future Cities Challenge. The authors acknowledge Giuseppe Lupo and Valeria d'Andrea for fruitful discussions.

Funding

Not applicable.

Abbreviations

OSMnx, Open Street Map NetworkX; GCE, Global Communication Efficiency; WS, Watts–Strogats; RGN, Random Geometric Networks; ICT, Information and Communication Technologies.

Availability of data and materials

The aggregated flow networks obtained are available from the authors upon request.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

RG and MDD designed research. RG, GB and MDD performed the research and wrote the paper. All authors read and approved the final manuscript.

Author details

¹CoMuNe Lab, Fondazione Bruno Kessler, Via Sommarive 18, 38123 Povo (TN), Italy. ²Department of Mathematics, University of Trento, Via Sommarive 14, 38123 Povo (TN), Italy.

Publisher's Note

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

Received: 17 June 2020 Accepted: 21 December 2020 Published online: 15 January 2021

References

1. Barthelemy M (2019) The statistical physics of cities. *Nat Rev Phys* 1:406–415
2. Mucha PJ, Richardson T, Macon K, Porter MA, Onnela J-P (2010) Community structure in time-dependent, multiscale, and multiplex networks. *Science* 328(5980):876–878
3. Szell M, Lambiotte R, Thurner S (2010) Multirelational organization of large-scale social networks in an online world. *Proc Natl Acad Sci* 107(31):13636–13641
4. De Domenico M, Solé-Ribalta A, Cozzo E, Kivela M, Moreno Y, Porter MA, Gómez S, Arenas A (2013) Mathematical formulation of multilayer networks. *Phys Rev X* 3(4):041022
5. De Domenico M (2018) Multilayer network modeling of integrated biological systems: comment on “Network science of biological systems at different scales: a review” by Gosak et al. *Phys Life Rev* 24:149–152
6. Kivela M, Arenas A, Barthelemy M, Gleeson JP, Moreno Y, Porter MA (2014) Multilayer networks. *J Complex Netw* 2(3):203–271
7. De Domenico M, Granell C, Porter MA, Arenas A (2016) The physics of spreading processes in multilayer networks. *Nat Phys* 12(10):901–906
8. Schreiber T (2000) Measuring information transfer. *Phys Rev Lett* 85(2):461
9. Barnett L, Barrett AB, Seth AK (2009) Granger causality and transfer entropy are equivalent for Gaussian variables. *Phys Rev Lett* 103(23):238701
10. Runge J, Heitzig J, Petoukhov V, Kurths J (2012) Escaping the curse of dimensionality in estimating multivariate transfer entropy. *Phys Rev Lett* 108(25):258701
11. Sugihara G, May R, Ye H, Hsieh C-h, Deyle E, Fogarty M, Munch S (2012) Detecting causality in complex ecosystems. *Science* 338(6106):496–500
12. Stramaglia S, Cortes JM, Marinazzo D (2014) Synergy and redundancy in the granger causal analysis of dynamical networks. *New J Phys* 16(10):105003
13. Van Nes EH, Scheffer M, Brovkin V, Lenton TM, Ye H, Deyle E, Sugihara G (2015) Causal feedbacks in climate change. *Nat Clim Change* 5(5):445–448
14. Diez I, Erramuzpe A, Escudero I, Mateos B, Cabrera A, Marinazzo D, Sanz-Arigita EJ, Stramaglia S, Cortes Diaz JM, Initiative ADN (2015) Information flow between resting-state networks. *Brain Connect* 5(9):554–564
15. Tononi G, Boly M, Massimini M, Koch C (2016) Integrated information theory: from consciousness to its physical substrate. *Nat Rev Neurosci* 17(7):450–461
16. James RG, Barnett N, Crutchfield JP (2016) Information flows? A critique of transfer entropies. *Phys Rev Lett* 116(23):238701
17. Ye H, Sugihara G (2016) Information leverage in interconnected ecosystems: overcoming the curse of dimensionality. *Science* 353(6302):922–925
18. Latora V, Marchiori M (2001) Efficient behavior of small-world networks. *Phys Rev Lett* 87(19):198701
19. Newman ME (2004) Analysis of weighted networks. *Phys Rev E* 70(5):056131
20. Guimera R, Amaral LAN (2005) Functional cartography of complex metabolic networks. *Nature* 433(7028):895–900
21. Colizza V, Flammini A, Serrano MA, Vespignani A (2006) Detecting rich-club ordering in complex networks. *Nat Phys* 2(2):110–115
22. Bassett DS, Bullmore ET (2009) Human brain networks in health and disease. *Curr Opin Neurol* 22(4):340–347

23. Rubinov M, Sporns O (2010) Complex network measures of brain connectivity: uses and interpretations. *NeuroImage* 52(3):1059–1069
24. Van Den Heuvel MP, Sporns O (2011) Rich-club organization of the human connectome. *J Neurosci* 31(44):15775–15786
25. Sporns O (2013) Network attributes for segregation and integration in the human brain. *Curr Opin Neurobiol* 23(2):162–171
26. Centola D (2015) The social origins of networks and diffusion. *Am J Sociol* 120(5):1295–1338
27. Deco G, Tononi G, Boly M, Kringelbach ML (2015) Rethinking segregation and integration: contributions of whole-brain modelling. *Nat Rev Neurosci* 16(7):430–439
28. Cohen JR, D'Esposito M (2016) The segregation and integration of distinct brain networks and their relationship to cognition. *J Neurosci* 36(48):12083–12094
29. Aerts H, Fias W, Caeyenberghs K, Marinazzo D (2016) Brain networks under attack: robustness properties and the impact of lesions. *Brain* 139(12):3063–3083
30. Bertolero M, Yeo B, D'Esposito M (2017) The diverse club. *Nat Commun* 8(1):1277
31. Bertolero MA, Yeo BT, Bassett DS, D'Esposito M (2018) A mechanistic model of connector hubs, modularity and cognition. *Nat Hum Behav* 2(10):765–777
32. Yamamoto H, Moriya S, Ide K, Hayakawa T, Akima H, Sato S, Kubota S, Tanii T, Niwano M, Teller S et al (2018) Impact of modular organization on dynamical richness in cortical networks. *Sci Adv* 4(11):4914
33. Stella M, Cristoforetti M, De Domenico M (2019) Influence of augmented humans in online interactions during voting events. *PLoS ONE* 14(5):0214210
34. Batty M (2013) Big data, smart cities and city planning. *Dialogues Hum Geogr* 3(3):274–279
35. Tsai Y-H (2005) Quantifying urban form: compactness versus 'sprawl'. *Urban Stud* 42(1):141–161
36. Guerois M, Pumain D (2008) Built-up encroachment and the urban field: a comparison of forty European cities. *Environ Plann A Econ Space* 40(9):2186–2203. <https://doi.org/10.1068/a39382>
37. Schwarz N (2010) Urban form revisited? selecting indicators for characterising European cities. *Landsc Urban Plan* 96(1):29–47
38. 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
39. Gately CK, Hutyrá LR, Wing IS (2015) Cities, traffic, and CO₂: a multidecadal assessment of trends, drivers, and scaling relationships. *Proc Natl Acad Sci* 112(16):4999–5004. <https://doi.org/10.1073/pnas.1421723112>
40. Ewing R, Hamidi S (2015) Compactness versus sprawl: a review of recent evidence from the United States. *J Plan Lit* 30(4):413–432
41. Song C, Koren T, Wang P, Barabási A-L (2010) Modelling the scaling properties of human mobility. *Nat Phys* 6(10):818–823. <https://doi.org/10.1038/nphys1760>
42. Louail T, Lenormand M, Picornell M, Cantú OG, Herranz R, Frias-Martinez E, Ramasco JJ, Barthelemy M (2015) Uncovering the spatial structure of mobility networks. *Nat Commun* 6:6007
43. Gallotti R, Bazzani A, Rambaldi S, Barthelemy M (2016) A stochastic model of randomly accelerated walkers for human mobility. *Nat Commun* 7(1):12600. <https://doi.org/10.1038/ncomms12600>
44. Barbosa H, Barthelemy M, Ghoshal G, James CR, Lenormand M, Louail T, Menezes R, Ramasco JJ, Simini F, Tomasini M (2018) Human mobility: models and applications. *Phys Rep* 734:1–74
45. Helbing D (2001) Traffic and related self-driven many-particle systems. *Rev Mod Phys* 73(4):1067
46. Li D, Fu B, Wang Y, Lu G, Berezin Y, Stanley HE, Havlin S (2015) Percolation transition in dynamical traffic network with evolving critical bottlenecks. *Proc Natl Acad Sci* 112(3):669–672
47. Çolak S, Lima A, González MC (2016) Understanding congested travel in urban areas. *Nat Commun* 7(1):10793. <https://doi.org/10.1038/ncomms10793>
48. Solé-Ribalta A, Gómez S, Arenas A (2018) Decongestion of urban areas with hotspot pricing. *Netw Spat Econ* 18(1):33–50
49. Depersin J, Barthelemy M (2018) From global scaling to the dynamics of individual cities. *Proc Natl Acad Sci* 115(10):2317–2322
50. Le Néchet F (2012) Urban spatial structure, daily mobility and energy consumption: a study of 34 European cities. *Cybergeo: Eur J Geogr*
51. Stone B (2008) Urban sprawl and air quality in large US cities. *Environ Eng Manag J* 86(4):688–698. <https://doi.org/10.1016/j.jenvman.2006.12.034>
52. Uherek E, Halenka T, Borken-Kleefeld J, Balkanski Y, Berntsen T, Borrego C, Gauss M, Hoor P, Juda-Rezler K, Lelieveld J (2010) Transport impacts on atmosphere and climate: land transport. *Atmos Environ* 44(37):4772–4816. <https://doi.org/10.1016/j.atmosenv.2010.01.002>
53. Martilli A (2014) An idealized study of city structure, urban climate, energy consumption, and air quality. *Urban Clim* 10:430–446. <https://doi.org/10.1016/j.uclim.2014.03.003>
54. Ewing R, Meakins G, Hamidi S, Nelson AC (2014) Relationship between urban sprawl and physical activity, obesity, and morbidity – update and refinement. *Health Place* 26:118–126. <https://doi.org/10.1016/j.healthplace.2013.12.008>
55. Newby DE, Mannucci PM, Tell GS, Baccarelli AA, Brook RD, Donaldson K, Forastiere F, Franchini M, Franco OH, Graham I, Hoek G, Hoffmann B, Hoylaerts MF, Kunzli N, Mills N, Pekkanen J, Peters A, Piepoli MF, Rajagopalan S, Storey RF (2014) Expert position paper on air pollution and cardiovascular disease. *Eur Heart J* 36(2):83–93. <https://doi.org/10.1093/eurheartj/ehu458>
56. Rice MB, Ljungman PL, Wilker EH, Dorans KS, Gold DR, Schwartz J, Koutrakis P, Washko GR, O'Connor GT, Mittleman MA (2015) Long-term exposure to traffic emissions and fine particulate matter and lung function decline in the framingham heart study. *Am J Respir Crit Care Med* 191(6):656–664. <https://doi.org/10.1164/rccm.201410-1875oc>
57. Li W, Dorans KS, Wilker EH, Rice MB, Long MT, Schwartz J, Coull BA, Koutrakis P, Gold DR, Fox CS, Mittleman MA (2017) Residential proximity to major roadways, fine particulate matter, and hepatic steatosis. *Am J Epidemiol* 186(7):857–865. <https://doi.org/10.1093/aje/kwx127>
58. Nicholl J, West J, Goodacre S, Turner J (2007) The relationship between distance to hospital and patient mortality in emergencies: an observational study. *J Emerg Med* 24(9):665–668. <https://doi.org/10.1136/emj.2007.047654>

59. Bettencourt LM, Lobo J, Helbing D, Kühnert C, West GB (2007) Growth, innovation, scaling, and the pace of life in cities. *Proc Natl Acad Sci* 104(17):7301–7306
60. Bettencourt LM (2013) The origins of scaling in cities. *Science* 340(6139):1438–1441
61. Bertaud A (2004) The spatial organization of cities: deliberate outcome or unforeseen consequence? Working Paper Series, UC Berkeley IURD
62. Volpati V, Barthelemy M (2018) The spatial organization of the population density in cities. [arXiv:1804.00855](https://arxiv.org/abs/1804.00855)
63. Louf R, Barthelemy M (2013) Modeling the polycentric transition of cities. *Phys Rev Lett* 111(19):198702
64. Louf R, Barthelemy M (2014) How congestion shapes cities: from mobility patterns to scaling. *Sci Rep* 4(1):5561. <https://doi.org/10.1038/srep05561>
65. Phithakkitnukoon S, Horanont T, Di Lorenzo G, Shibasaki R, Ratti C (2010) Activity-aware map: identifying human daily activity pattern using mobile phone data. In: International workshop on human behavior understanding. Springer, Berlin, pp 14–25
66. Noulas A, Mascolo C, Frias-Martinez E (2013) Exploiting Foursquare and cellular data to infer user activity in urban environments. In: 2013 IEEE 14th international conference on mobile data management. IEEE Comput. Soc., Los Alamitos. <https://doi.org/10.1109/mdm.2013.27>
67. Bullmore E, Sporns O (2012) The economy of brain network organization. *Nat Rev Neurosci* 13(5):336–349
68. Boeing G (2017) Osmnx: new methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Comput Environ Urban Syst* 65:126–139
69. Future Cities Challenge. <https://www.futurecitieschallenge.com>. Accessed 05 Aug 2019
70. Louf R, Barthelemy M (2016) Patterns of residential segregation. *PLoS ONE* 11(6):0157476
71. Bertagnolli G, Gallotti R, De Domenico M (2020) Quantifying efficient information exchange in real network flows. [arXiv:2003.11374](https://arxiv.org/abs/2003.11374)
72. Newman ME (2004) Fast algorithm for detecting community structure in networks. *Phys Rev E* 69(6):066133
73. Gallotti R, Barthelemy M (2014) Anatomy and efficiency of urban multimodal mobility. *Sci Rep* 4(1):1–9
74. Bassolas A, Barbosa-Filho H, Dickinson B, Dotiwalla X, Eastham P, Gallotti R, Ghoshal G, Gipson B, Hazarie SA, Kautz H et al (2019) Hierarchical organization of urban mobility and its connection with city livability. *Nat Commun* 10(1):1–10
75. Park H-J, Friston K (2013) Structural and functional brain networks: from connections to cognition. *Science* 342(6158):1238411
76. Olteanu A, Castillo C, Diaz F, Kiciman E (2019) Social data: biases, methodological pitfalls, and ethical boundaries. *Front Big Data* 2:13
77. Foursquare Statistics. <https://99firms.com/blog/foursquare-statistics>. Accessed 25 Nov 2020
78. We Are Flint. <https://castfromclay.co.uk/models-research/main-findings-social-media-demographics-uk-usa-2018/>. Accessed 25 Nov 2020
79. Our World in Data. <https://ourworldindata.org/rise-of-social-media>. Accessed 25 Nov 2020
80. Noulas A, Scellato S, Mascolo C, Pontil M (2011) An empirical study of geographic user activity patterns in Foursquare. *ICWSM* 11(70–573):2
81. Gallotti R, Louf R, Luck J-M, Barthelemy M (2018) Tracking random walks. *J R Soc Interface* 15(139):20170776
82. Cottineau C, Hatna E, Arcaute E, Batty M (2017) Diverse cities or the systematic paradox of urban scaling laws. *Comput Environ Urban Syst* 63:80–94
83. Foursquare Developers Venue Categories. <https://developer.foursquare.com/docs/api/venues/categories>. Accessed 02 Aug 2019
84. Latora V, Marchiori M (2003) Economic small-world behavior in weighted networks. *Eur Phys J B, Condens Matter Complex Syst* 32(2):249–263
85. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. *J Stat Mech Theory Exp* 2008(10):10008

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)
