



OPEN

Uncovering structural diversity in commuting networks: global and local entropy

Valentina Marin[✉], Carlos Molinero & Elsa Arcaute

In this paper we revisit the concept of mobility entropy. Over time, the structure of spatial interactions among urban centres tends to become more complex and evolves from centralised models to more scattered origin and destination patterns. Entropy measures can be used to explore this complexity, and to quantify the degree of structural diversity of in- and out-flows at different scales and across the system. We use toy models of commuting networks to examine global and local measures, allowing the comparison to occur between different parts of the system. We show that entropy at the link and node level give different insights on the characteristics of the systems, enabling us to identify employment hubs and interdependencies between and within different parts of the system. We compute the measures in the commuting networks of the Northern Powerhouse and Greater South East regions in the UK to examine their relevance when studying real systems of cities. Finally we discuss how these can be used to inform planning and policy decisions oriented towards decentralisation and resilience.

Cities are essentially relational, they are defined by the nature of interactions that holds them together. In the same way, they could also be defined by how they are connected to other cities within a system of intricate relationships. Different types of interactions lead to relationships of dominance, dependency or cooperation between cities, and in doing so, they characterise the functioning and dynamics of the whole system. Systems of cities are interdependent, a significant change in one of its components could impact or disrupt the functioning of other urban entities within the system, or even the structure of the system as a whole.

In this context, systems of cities can be modelled as networks, where different types of links represent different interdependencies, giving rise to different structures^{1–3}. Within a myriad of relationships that link urban systems, the connection between workplaces and home has been a central part of studies for understanding the dynamics occurring within systems of cities. Mobility patterns have been widely studied to examine the structure of mobility and its relation with socio-demographic variables^{4,5}, to define categories of cities according to their commuting structure⁶, to investigate the evolution of mobility patterns over time^{7,8}, to define boundaries of functional areas⁹, and to study the spread of infectious diseases¹⁰, to name a few.

When commuting is seen as a network, cities are represented as nodes, and the flows of commuters constitute the links. The structure and characteristics of such a network can give us insights into the different roles that cities play within the system. In particular, diversity and dispersion of flows across the system can inform about the cohesiveness and balance of the relative importance of urban centres. On the other hand, the over concentration of flows can reveal subordination and high dependence of the system in few specific centres, exhibiting the potential susceptibility of the system as a whole. For example, the extent to which in-commuting is concentrated seems crucial to understand potential labour market centralisation and disparities in the distribution of job opportunities.

Diversity of patterns of labour supply and demand across the territory, is a critical attribute for the resilience of the commuting network. Diverse mobility patterns contribute to the re-organisational capacity of the system¹¹. A diverse system has multiple responses and alternatives of meeting a given need, making space for adaptation and innovation to maintain the functioning of the system across different conditions and change^{12–16}. In this context, the relationship between diversity and complexity has gained great attention, particularly from resilience theory¹⁷, being a central matter in many fields of science. However widely used, and despite the noted relevance of diversity as a crucial characteristic of urban systems^{14,18–20}, it remains a difficult concept to define and measure. The difficulty lies in the many different methodological approaches to measure it across disciplines, encountering great semantic variation^{21,22}.

Centre for Advanced Spatial Analysis, University College London, London W1T 4TJ, UK. ✉email: v.marin@ucl.ac.uk

Entropy is one of the most common ways of quantifying diversity²³. The concept of entropy was first coined in thermodynamics, and then widely used in other fields such as physics, statistics, information theory and ecology. Depending on the research context, entropy is generally addressed as a measure of disorder in a system, or as the level of uncertainty and information²⁴. The latter was introduced by Shannon (1948) in the context of information theory, referring to the amount of information within probability distributions²⁵. Shannon's entropy measures the degree of uncertainty in predicting the types of elements randomly chosen from a sample. It depends on both, the number of types and the relative abundance of them, also known in the field of ecology as richness and evenness respectively. The greater the amount of types (richness) and the more equally abundant they are (evenness), the more difficult it is to predict^{21,26}. In such a way, when applied as a diversity measure, one can say that the more uncertainty, the greater the diversity.

The entropy of network-based systems refers to the heterogeneity in the arrangement of its components. Entropy measures on graphs were first used by Rashevsky (1955) and Mowshowitz (1968) as a measure of relative complexity. First approaches studied the topological information content in unweighted and undirected graphs^{27,28}. A common way to measure entropy in graphs is based on the degree distribution $P(deg)$, measuring the probability of having a node with a certain number of links. It is a local measure with a focus on node's connectivity which ignores to some degree the weight of the different links. Although it is a useful measure to characterise important aspects of the network, on its own it is unable to describe the complexity of the network structure, from both local and global perspectives. More recent studies have deepened the understanding of entropy in weighted and directed graphs by extending information theory concepts to networks^{29,30}. This is crucial for studying diversity of commuting networks, which are described by in-commuting and out-commuting flows.

In commuting networks, entropy is commonly used as a relative measure of the distribution of commuters amongst employment locations. This is achieved by looking at the relative abundance and volume of flows embodied in the weights of the links. Although the study of flows by means of entropy has not been widely adopted, certain studies have validated its use for addressing key urban matters such as the analysis of patterns of spatial dispersion to inform choice models for urban transportation³¹; the use of in- and out-commuting entropy on different cities to explain variations in economic growth³²; the use of entropy of individual users trajectories to study the correspondence of mobility diversity to social behaviour and socio-economic indicators^{33,34}, as a measure of spatial inequality and attractiveness³⁵, or the use of entropy of individual vehicular mobility to characterise spatio-temporal patterns of activities along the day³⁶.

Typically, to the best of our knowledge, measures are carried out at a local level, that is, entropy is calculated for individual trajectories or specific nodes within the network. Cities or administrative areas are the local units of analysis, and there is no wider consideration of the global structure/context to understand their role within the overall regional or national system. On this basis, we think that certain properties of the global network in terms of its structural diversity may have been left unexplored by focusing on the performance of local elements.

In this paper we test different measures of entropy on commuting networks at global and local scales. We aim to explore if the results offered by the different measures are complementary and relevant for the study of the structural diversity of spatial interactions. First we use toy models of networks with different patterns to examine the different measures and compare the outcomes across systems and their constituent parts. We examine the diversity of the global commuting structure by applying a set of measures to both the group of all nodes and the ensemble of links, considering node strength $f(s)$ and link weight distribution $f(w)$ respectively. We look at nodes based on the workforce patterns of in- and out-commuting flows in all urban units as a result of the spatial distribution of labour supply and demand. The measure depicts patterns of centralisation and dispersion in the urban spatial structure. When studying the network from the perspective of its links, we focus on the diversity of distribution of origin-destination pairs, considering intensity and density of flows. Normalisation of the measure is achieved by comparing the entropy of the sampled links with the maximum possible links of a fully connected network. This measure describes the level of dispersion of commuter trips in the territory, outlining potential functional dependencies when many trips take place in few dominant Origin-Destination (OD) pairs. In a following section we address local entropy at nodes individually. We look at the structural diversity of the sub-graph made up of the subset of interactions that a given urban unit establishes directly with its neighbours. Then we compare the results from the general equation of nodal entropy with a normalised measure that considers its maximum potential if connected to all other nodes in the network. Finally, we comment on the outputs obtained by measuring real networks and discuss the relevance of the twofold analysis of the system comparing global and local approaches, guided by the following questions: Is it the same to be a non-diverse unit within a structural dispersed system, than being a non-diverse unit within a structural concentrated one? To what extent the diversity of the individual elements could describe the diversity of the overall system? What can we learn about the system by comparing the outcomes of local and global scales?

Methods

The entropy measures presented in this paper are based on the information-theoretic approach to networks^{29,30}. This approach considers entropy as a measure of uncertainty related to the information content transmitted from sender to receiver. When applied to networks, an analogy is made so effluxes of nodes correspond to the sender and influxes of nodes to the receiver³⁰. Then, the uncertainty of the transmission of a certain flux in the network depends on the probability of its occurrence between sender and receiver. In our case study, commuting networks are constituted by origin and destination nodes representing urban units with in- and out-flows. Measures of uncertainty in Information Theory derive from the Shannon Entropy H formula²⁵, (also known as Shannon's diversity index in ecology) which is defined as:

$$H = - \sum_{v_i} p_i \log p_i \quad (1)$$

where p_i is the probability of occurrence of the i th type within the total sample.

We use different toy models of commuting networks to compare different measures on both local and global scales applied to the links and nodes of the network. Different forms of normalisation are presented in each case. The following measures consider the commuting flows as directed and weighted graphs (G) represented by a set of n nodes $V(G)$ and m links $E(G)$, each representing a tuple of nodes. Each node attracts in-commuting flows, and releases out-commuting flows in different proportions depending on its role within the system. For every link, a weight w_{ij} is assigned, representing the total flow from origin i to destination j .

Results

Global diversity. Global measures quantifying diversity as a function of the overall structure of the commuting network are computed across the whole graph. To measure entropy globally, we look at how flows are distributed either on nodes or on links, considering every component of the network. Given that all elements are interdependent in the overall structure, any local change in the commuting network will modify the global entropy.

Spatial distribution of labour supply and demand. Labour supply and demand are not evenly distributed in the geographic space, giving rise to complex patterns of spatial interactions which are reflected in the structure of the commuting network.

Urban units have different functional roles within the system. Some cities for example function as employment hubs, attracting large numbers of workers, other cities mostly supply workers to other areas, while some others are able to find a balance between labour supply and demand. Entropy measures enable us to explore whether the flows in a system tend to be concentrated in dominant areas, or evenly dispersed from many origins to many destinations. The former is characterised by a monocentric pattern where the flows come from many origins to very few destinations, and the latter is characterised by a more polycentric pattern, which indicates a greater balance in the importance of urban units.

Identifying monocentricity or polycentricity through commuting patterns can provide an initial insight into diagnosing possible functional dependencies due to disproportionate concentrations in some central cities. These can serve to inform planning to overcome spatial disparities, with interventions related to labour decentralisation and transport infrastructure aiming at encouraging growth of subordinate areas and more balanced and diverse spatial interaction patterns.

Let us start by characterising origins and destinations through the diversity of locations from/to which workers go/arrive to work. This can be captured through the following global entropy measures:

- Global out-flow entropy at node level:

$$H_{GN}^{out} = - \sum_{v_i} \left(\sum_{v_j} p_{ij} \right) \log \left(\sum_{v_j} p_{ij} \right) \quad (2)$$

where $\sum_j p_{ij}$ is the probability of out-flow from node v_i , considering the sum of all flows departing from v_i to every possible node v_j within the total commuting in the system.

- Global in-flow entropy at node level:

$$H_{GN}^{in} = - \sum_{v_j} \left(\sum_{v_i} p_{ij} \right) \log \left(\sum_{v_i} p_{ij} \right) \quad (3)$$

where $\sum_i p_{ij}$ is the probability of in-flow to node v_j , considering the sum of all flows arriving to v_j from every possible node v_i within the total commuting in the system.

Both measures, reveal structural patterns of the network according to commuting origin or destination, providing information about the concentration of flows. These measures examine the distribution of node strength in the network, which accounts for the total in- or out-trips in every node. In the case when one node concentrates most of the flows, the system will exhibit a skewed probability distribution, indicating that if a location for labour supply or demand is taken at random, it will most likely correspond to that node. This reduced uncertainty of knowing where an individual randomly selected might go to work (or come from), is represented through a lower entropy. If on the other hand, there is a similar probability distribution of flows across nodes, such that the system has no node dominating over others, the uncertainty of ascertain the work or home location of an individual will be higher and hence the entropy will also be higher. The latter is maximal when there is equiprobability across space. Figure 1 exemplifies the centralisation of system b with respect to in-commuters, concentrating most of the flows in v_5 , while for the same configuration and different directions, system a does not concentrates job, attaining hence a higher in-flow entropy than b. The inverse occurs for the out-flows.

In general, commuting destinations tend to be more highly concentrated than the commuting origins. This is because employment opportunities tend to cluster in few locations. In those cases the out-entropy will be higher than the in-entropy. However, this is not always the case, and exploring whether the origins or destinations of the commuter flows are more or less diverse, by looking at whether $H_{GN}^{out} > H_{GN}^{in}$, or $H_{GN}^{in} > H_{GN}^{out}$, can give a better understanding of the urban system. Figure 1 presents a strong case of monocentricity with respect to jobs in

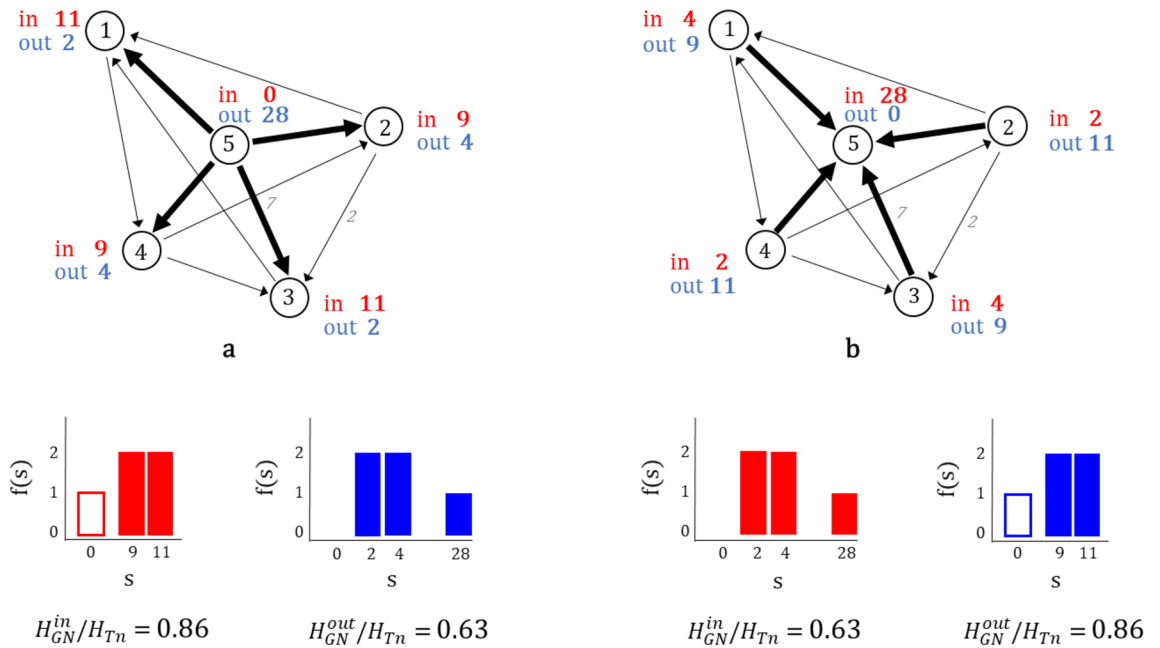


Figure 1. Global nodal entropy. Example of two different commuting networks. The node strength (s) measures the strength of nodes by means of the total weight of their connections, which in this case corresponds to the total amount of in- or out- flow at every node. The distribution $f(s)$ for in-commuting is in red and for out-commuting is in blue. In the example above, when flows are more uniformly distributed in nodes, the H_{GN} is higher, and when there are fewer nodes concentrating the flows, the H_{GN} is lower.

system b, where $H_{GN}^{out} > H_{GN}^{in}$, and the odd case from which a single location provides most workers for several different locations in system a, with $H_{GN}^{in} > H_{GN}^{out}$.

To normalise the results and make them comparable between systems of different sizes (different number of nodes), we look for the total n value for each system, then $H_{Tn} = \log(n)$, where n is the total number of nodes in the system. The normalised entropies can be written as: H_{GN}^{out}/H_{Tn} and H_{GN}^{in}/H_{Tn} .

Commuter trips distribution. In the previous section we characterised the origins and destinations of commuting flows according to their diversity, and considered how such an approach can give insights into the polycentricity of cities. Let us now look at the trips that are being generated, and measure the diversity of the flows along the links of the commuting network.

The concept of diversity is associated in this case with the dispersion of commuter trips in the territory taking into account the distribution of flow intensity $f(w)$ and network density. Therefore, the measure considers the equivalence between the flow capacity of their interactions, as well as the variety of areas that are connected to each other. A system will be more diverse if there are many combinations of origin-destination pairs (higher link density), and if the amount of flows between these pairs is evenly distributed (more uniform link weight distribution).

The framework presented here is relevant to inform infrastructure planning, given that the provision of transport infrastructure is intertwined with the spatial distribution of flows. The more dispersed the pattern of origins and destinations in the territory are, the more challenging is the planning of the physical transport structure that allows these trips to occur more efficiently³¹. In addition, such a framework also allows us to identify functional dependencies between urban units within the system. If the relationships are scattered it means that the operation of the system relies on various labour and economic relationships between its different components. The opposite occurs with the existence of dominant flows where most of the trips occur between few pairs of urban areas, and the overall system is constrained to these specific relationships.

Let us introduce global entropy at link level, as a measure of flow diversity, considering every OD pair in the system. We normalise the measure with respect to its maximum, so that comparisons with other systems can be made. In this case, we need a *joint entropy* encompassing the uncertainty associated with both origin and destination, through the link probability. Such a measure can be interpreted as an average diversity of the system as a whole²⁹. The entropy of trips can be defined as:

- Global entropy at link level:

$$H_{GL} = - \sum_{\forall i} \sum_{\forall j} p_{ij} \log p_{ij} \tag{4}$$

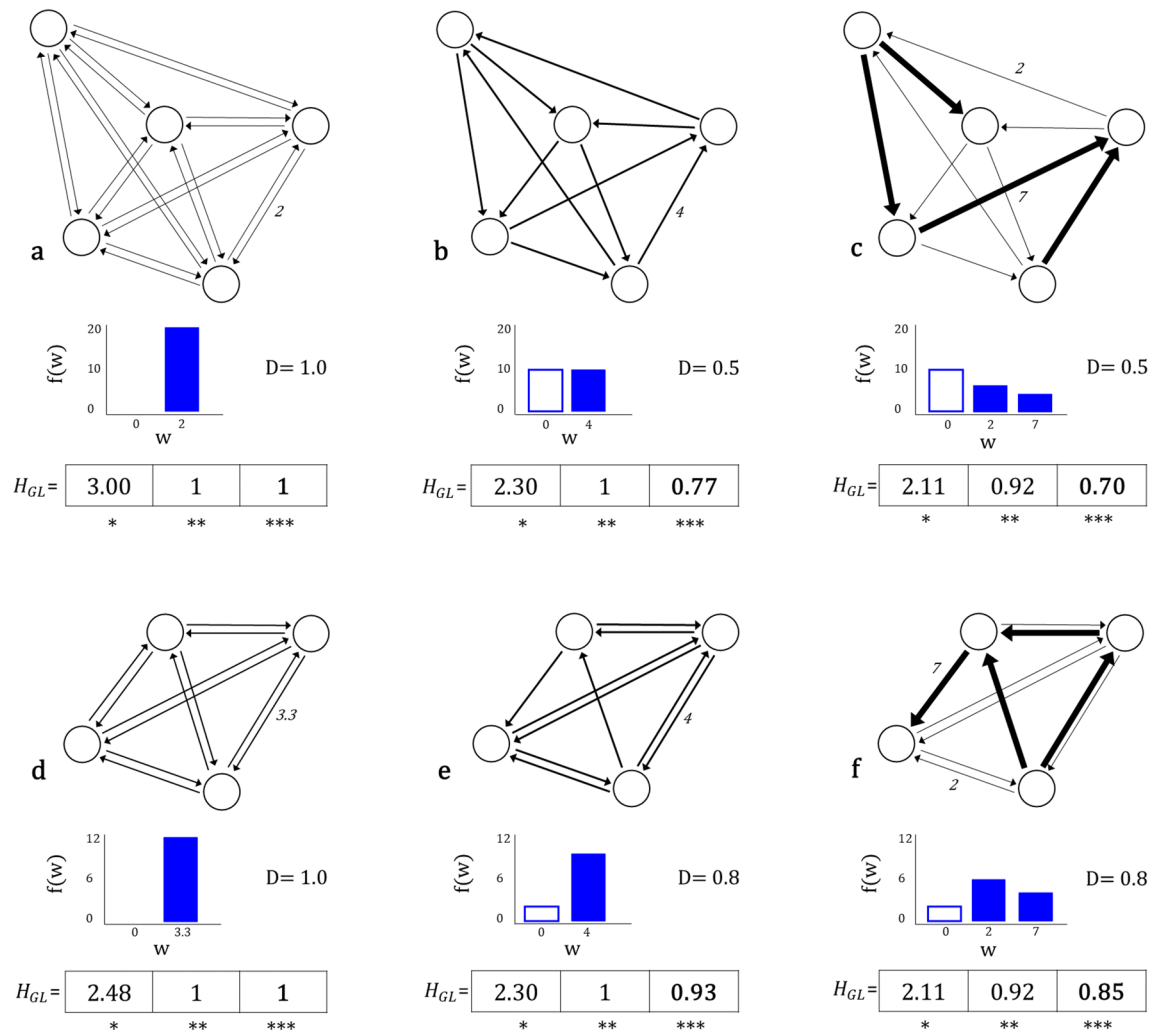


Figure 2. Global links entropy. Example of commuting networks with the distribution of link weights $f(w)$, the link density $D = m/n(n - 1)$ and the corresponding general entropy with normalised values: (*) = H_{GL} , (**) = H_{GL}/H_{Tm} , and (***) = H_{GL}/H_{Mpm} . A uniform distribution of weights and high density results in a high H_{GL} diversity (systems a and d), while when few links concentrate many flows, and the density is low, the system will exhibit a low H_{GL} diversity (system c). Only the normalisation (***) is able to differentiate between systems b) and e), and c) and f), where higher link density results in higher entropy values.

where p_{ij} is the probability that a commuting flow from v_i to v_j occurs in the system, hence $p_{ij} = \frac{w_{ij}}{\sum_i \sum_j w_{ij}}$, where w_{ij} is the number of trips from v_i to v_j . H_{GL} takes higher values when flow weights are evenly distributed, so every commuting flow is equally relevant in the commuting network. Conversely, if only few OD links contain the large majority of commuting flows, the diversity of the system is low. Then, the dominance of some flows in the network reduces the entropy H_{GL} . This is clear when looking at networks b and c in Fig. 2. With the same total flow count and the same amount of links, but with a different distribution of flows among them, the global entropy at link level is higher in b than in c. In b flows are evenly distributed, while in c certain links have a much higher density than in the rest of the system. In general, entropy values tend to be higher when the number of elements in the system increases, so the more the links m or nodes n in the network, the higher the entropy values. In Fig. 2, we can confirm this by looking at networks a and d. In both systems flows are evenly distributed across links, however entropy (*) is higher in a than d because $m_a > m_d$. This means that the comparison of different systems is not a straightforward task. To address this issue we need to normalise our measures of entropy. A common way of addressing normalisation when studying entropy is by looking at its maximum value, which occurs when all elements are equally abundant. Then, normalisation is done by dividing the entropy value by the entropy of the total number of elements present in the system. Accordingly, diversity of link weights in a network with a total of m links would be normalised by $H_{Tm} = \log(m)$, leading to $H^{**} = H_{GL}/H_{Tm}$ ³⁷. A more suitable normalisation when comparing diversity of commuting flows between system of cities should consider a notion of density. Network density in this context, is understood as the ratio of the total number of links in the network to the number of links in its theoretical fully connected network³⁸. The maximum possible number of links (Mpm) in a graph is given by $n(n - 1)$, n being the number of nodes in the graph. The proposed

normalisation in this work is then by $H_{Mpm} = \log(n(n-1))$, leading to $H^{***} = H_{GL}/H_{Mpm}$. This process takes into account the diversity of flows in an existing level of interaction in a system, in comparison to its own maximum potential of connectivity. The latter, allows a meaningful comparison of flow diversity between systems of different sizes (different number of commuting OD pairs or different number of cities).

A comparison between both forms of normalisation could be easily done by looking at examples b and e in Fig. 2. Taking the common form of normalisation: $H^{**}(b) = H^{**}(e) = 1$, giving the maximum value to both systems which have the same amount of existing links ($m = 10$) and even distribution of weights. By this, we could conclude that both systems are equally diverse. However, when taking the proposed normalisation (***) the results for each systems are $1 > H^{***}(e) > H^{***}(b)$. As this measure considers density, in this case none of the systems meet the maximum entropy value of 1, because they do not present the maximum possible number of OD pairs according to their own potential, as in the cases of a and d. We observe that the density of links in e is closer to its maximum potential, in comparison to b.

Local diversity. In addition to identifying general characteristics of the network, we are also interested in understanding the role of individual locations. Local measures serve this purpose, and local diversity can be thought of as a sub-graph entropy of the node in question, where every in or out link directly connected to it is taken into account. This measure considers the intensity and density of the trips that are released (outflow) or attracted (inflow) by each unit (node). The distribution of flow intensity informs whether the relationships are organised in a scattered or polarised manner. Density, on the other hand, looks at the variety of urban areas with which the unit in question interacts.

Identifying important actors in the distribution of flows is important to be able to construct decentralised solutions. These are favoured to increase the resilience of the network. Decentralisation can be achieved by diversifying the dependence between nodes. Looking at the specific case of commuting networks, the distribution of inflows is determined by the areas of provision of labour for the internal employment market. On the other hand, the distribution of workforce outflows accounts for dependencies between residents in a certain area, and the provision of jobs in other locations. Within this proposed framework, areas of similar interaction and dependency patterns can be identified, from which a categorisation of cities can be constructed to inform planning decisions. For in- and out-commuting scenarios, local entropy is defined as:

- Local in-flow entropy:

$$H_L^{in} = - \sum_{v_i} p_{(i|j)} \log p_{i|j} = - \sum_{v_i} \frac{p_{ij}}{p_j} \log \frac{p_{ij}}{p_j} \quad (5)$$

- Local out-flow entropy:

$$H_L^{out} = - \sum_{v_j} p_{j|i} \log p_{j|i} = - \sum_{v_j} \frac{p_{ij}}{p_i} \log \frac{p_{ij}}{p_i} \quad (6)$$

where $p_j = \frac{\sum_i w_{ij}}{\sum_i \sum_j w_{ij}}$ represents the sum of every flow arriving at v_j divided by the total flow of the system,

and $p_i = \frac{\sum_j w_{ij}}{\sum_i \sum_j w_{ij}}$ refers to the sum of every flow leaving v_i divided by the total flow of the system. $p_{(i|j)}$ and $p_{(j|i)}$ represent the probability that a flow within the system is received or sent by a specific node respectively.

These measures give information about the node diversity in terms of the flows that are sent or received by its direct neighbours (one-hop neighbours of the target node). In this case, entropy functions are applied to the distribution of flow weights to or from a given node. Thus, the dominance of an origin-destination pair at a given node reduces the entropy, while an equal distribution of flows results in higher values of entropy. Figure 3 shows the example of a node v_1 whose in-links have equal weights, while the out-links are dominated by flows commuting to v_5 , leading to $H_L^{in} > H_L^{out}$. An entropy equal to zero occurs when there is no link arriving or departing from a node (e.g. out-commuting from v_5 in Figure 3). But, this will also be the case when there is only one link connected to the node, since there will be no uncertainty (e.g. in-commuting from v_2 or v_4 in Fig. 3).

Normalisation can be achieved by dividing by the maximum value of entropy given the in or out degree of the node: $H_{deg}^{in} = \log(\deg_{in}(v_j))$ and $H_{deg}^{out} = \log(\deg_{out}(v_i))$. Another approach for normalising local nodal entropy could be done by looking at its maximum possible degree value, which in this case would be given by $H_{Mpd} = \log(n-1)$. As explained previously for Eq. (4), the proposed framework takes into consideration the network density and the maximum potential of connectivity of nodes in a network. The normalised diversity measures are: H_L^{in}/H_{Mpd} and H_L^{out}/H_{Mpd} . The relevance of the normalisation is illustrated in Fig. 3. With the first normalisation (**), looking at the in-flows, both nodes v_5 and v_1 present the highest value, since flows are equally distributed among the existing links. In the second normalisation (***), node v_5 has a higher value since it receives flows from every possible node in the network, while v_1 only receives flows from half of the potential origins.

If we want to describe the system based on its local relationships, we can compute the average among every local entropy. The following equations measure the weighted mean where every local measure in the system is considered based on its different probability of occurrence. In Information Theory this measure is known as *conditional entropy*, and it quantifies the uncertainty about a variable when another variable is known³⁰. The

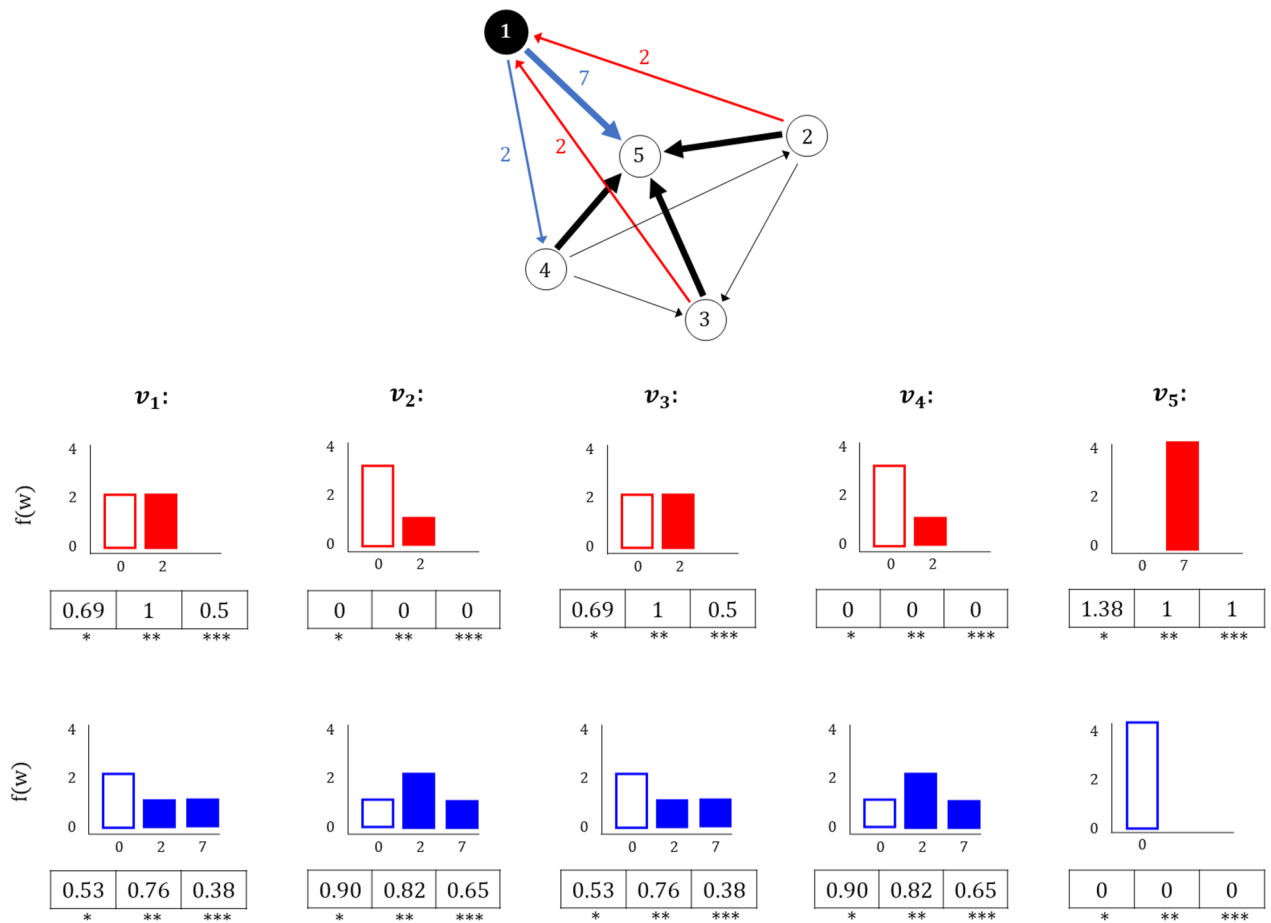


Figure 3. Local node entropy. A commuting network where the sub-graph at node v_1 is highlighted with in-flows in red, and out-flows in blue. The network consists of nodes with different degree and link weight distributions to explore the different results and normalisation approaches. Diagrams for in-flows frequency distribution $f(w)$ at every node are in red, and out-commuting $f(w)$ are in blue. Entropy and normalised values corresponds to: (*) = H_L^{in} , (**) = H_L^{in}/H_{deg}^{in} or H_L^{out}/H_{deg}^{out} , and (***) = H_L^{in}/H_{Mpd} or H_L^{out}/H_{Mpd} . As an example, a uniform distribution of in-flows and a $deg_{in} = 2$ on v_1 results in a maximum 1 for (**) and 0.5 for (***), whereas a uniform weight distribution and a maximum possible $deg_{in} = 4$ of node v_5 results in a maximum 1 for (**) and (***).

first measure corresponds to the uncertainty of in-commuting when destination is known, the second one corresponds to the uncertainty of out-commuting when origin is known:

- Average local in-flow entropy:

$$H_{L\mu}^{in} = \sum_{\forall j} p_j (H_L^{in})_j = - \sum_{\forall i} \sum_{\forall j} p_{ij} \log p_{(ij)} = - \sum_{\forall i} \sum_{\forall j} p_{ij} \log \frac{p_{ij}}{p_j} \tag{7}$$

- Average local out-flow entropy:

$$H_{L\mu}^{out} = \sum_{\forall i} p_i (H_L^{out})_i = - \sum_{\forall i} \sum_{\forall j} p_{ij} \log p_{(ji)} = - \sum_{\forall i} \sum_{\forall j} p_{ij} \log \frac{p_{ij}}{p_i} \tag{8}$$

Note that the average is obtained by considering all possible values of i or j given by each probability of occurrence p_i or p_j . Normalisation of these measures could be done by dividing the results by $\log(n - 1)$, in the same way as for Eqs. (5) and (6), allowing us to compare diversity in different systems at local scales in terms of their own maximum potential of connectivity.

Multiple measures analysis. In this paper we have explored some entropy functions at the global and local level for directed networks, aiming at capturing different relationships between system components. When looking at Fig. 4, where all revised measures are computed across 10 toy models with different flow distributions

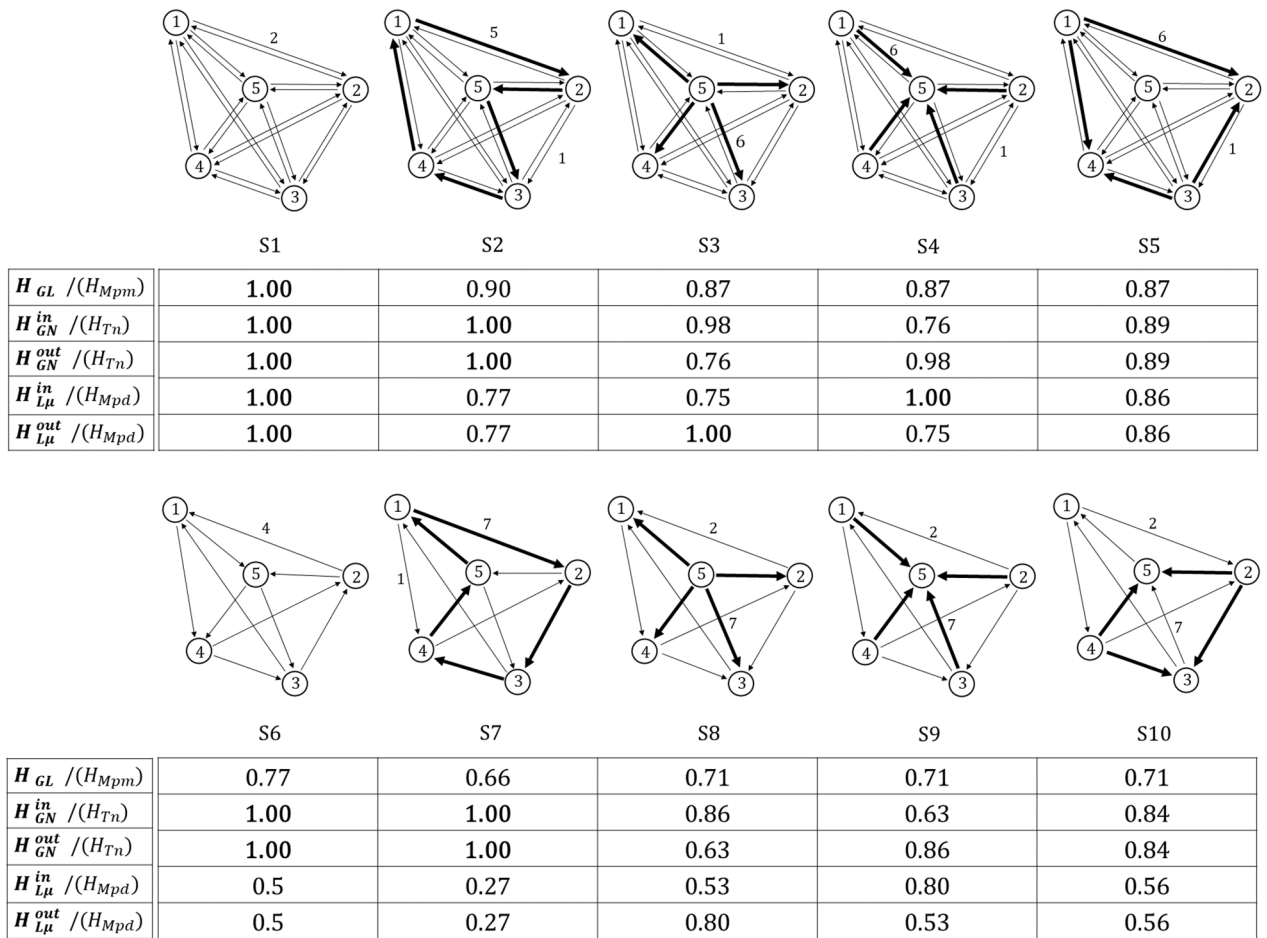


Figure 4. Global and local measures across different commuting networks. Values correspond to normalised entropy. The total volume of trips is 40 flows in all systems. The link density D for S1 to S5 is 1.0 and for S6 to S10 is 0.5.

but same topology, we can observe the high variability of outputs. This tells us how important it is to choose the proper measure to describe the pattern of interest. A single measure will not be able to capture the complex structure of flow diversity in the system. This is clearly shown in system S7 in Fig. 4, where values of entropy vary between 0.27, the minimum value in the whole table, and the maximum 1.

H_{GN}^{out} and H_{GN}^{in} are able to capture concentration of flows which could potentially inform about the presence of predominant centres. If we look at the first row of networks in Fig. 4, H_{GN}^{in} results can be sorted as $S1 = S2 > S3 > S5 > S4$. In the first two systems, flows are equally distributed among nodes, having the maximum diversity. The network S5 presents polycentricity, where flows are mainly clustered in two nodes. While network S4 presents monocentricity, with a concentration of flows at one destination. This system has a lower H_{GN}^{in} value. On the other hand, in order to differentiate patterns between S1 and S2 we need another measure. By looking at the H_{GL} values, we can observe that $S1 > S2$. This means that while in both systems nodes are equally relevant, the relationships between nodes are different. In S2 the measure is able to capture that some of the origin-destination pairs dominate over others.

Local measures for commuting networks are more commonly used to better understand the dynamics occurring at place level. In a regional system for example, local diversity allows us to capture how heterogeneous is the interconnection of a city with other cities in the system. In this context, the average of all local measures among cities, is expected to reflect the dynamics occurring among all interconnections between the whole system components. However, we believe that this is not a straightforward assumption, and it is necessary to question to what extent local measures could capture the global diversity of mobility, and how both local and global relates to each other. In Fig. 5, we explore those relationships between local and global measures of diversity across toy models presented in Fig. 4. Figure 5b shows the relationship between global H_{GN}^{in} and local $H_{L\mu}^{in}$ in-commuting measures. In general we can see that both reflect different aspects of diversity in the system, and they do not present any obvious correlation. For example, system S4 has maximum local diversity but a relative low global diversity.

A city is not fully characterised by the relationships established with its immediate neighbours, its role within the wider context and dynamics of the region, country or trade network it belongs to, play an important role in its characterisation. For example, a city with a high local mobility diversity within a diverse region, will not have the same role as a city with the same local diversity but within a non-diverse regional system. In Fig. 5d,

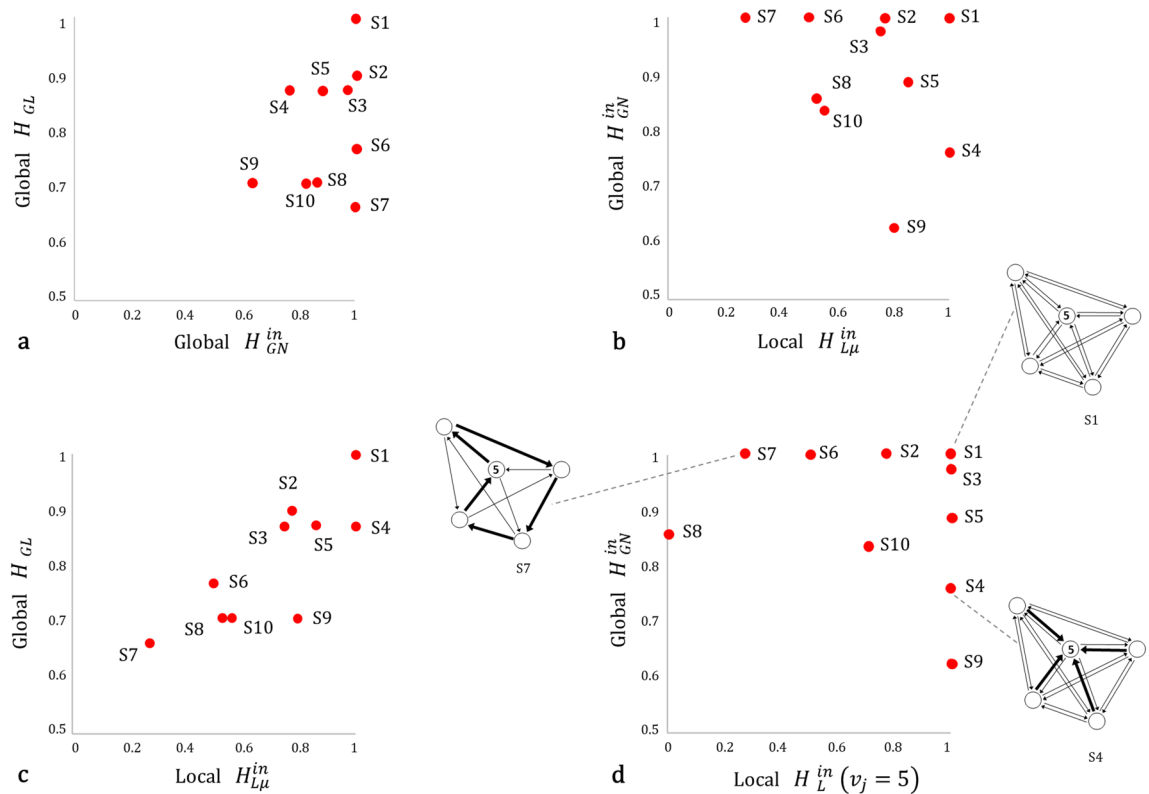


Figure 5. Comparison between local and global proposed normalised measures of networks in Fig. 4. (a) compares both global measures, (b) and (c) compare global with average local entropy, and finally (d) compares global with local entropy at node v_5 . System S1 is a fully connected network with evenly distributed flows with the maximum entropy in every measure, while other systems present varying patterns across measures depending on their structural organisation. As an example we can see that in (d) node v_5 in S1 has high entropy within a globally diverse system, and in contrast node v_5 in S4 has high local entropy within a system which is centrally organised. Although individual node interactions are equally distributed among neighbours in S4, node v_5 has a preponderant role globally. In S7 node v_5 has low local entropy but is part of a fully diverse system.

we compute local in-commuting node entropy H_L^{in} for node v_5 in each system in Fig. 4, and plot it against the global entropy H_{GN}^{in} of its correspondent network. This shows that the node v_5 in networks S1, S3, S4, S5 and S9 is fully diverse locally, with the same entropy value, nevertheless, it belongs to systems that behave completely differently globally. Taking S1 and S4, as opposite examples, we can clearly see that in network S4, the node v_5 has a dominant role, functioning as a centre of destinations. Conversely in S1, node v_5 has the same role as every other component within the overall network.

Northern Powerhouse and Greater South East structural diversity

Building on the discussion in previous sections, we took the Greater South East (GSE) and the Northern Powerhouse (NP) regions in the UK and compared the structure of their commuting patterns by computing the different entropy measures (Fig. 6). The dataset corresponds to the aggregated commuter flows at the Local Authority District (LAD) level from the 2011 Census for England and Wales. Origin-Destination flows with weights $w_{ij} > 10$ within the regions of the North East, North West and Yorkshire and Humber comprise the NP commuter network with $n = 72$ nodes, while the ones within the East, the South East and the London regions conform the GSE network with $n = 135$ nodes. The total flow within the GSE and the NP is 2,449,781 and 1,163,861 commuters respectively.

Different aspects of the functioning of each super-region as a whole can be described by the global entropies (Fig. 6). Global in-flow entropy at node level H_{GN}^{in} , depicted a greater balance in the importance of employment hubs in the NP with an output of 0.91 versus a 0.79 in the GSE. The diversity of the spatial distribution of labour supply in various urban units in the north, supports the interpretation of the Northern Powerhouse as a polycentric region in the global scale, contrasting with a more monocentric structure of the Greater South East where Inner London attracts a large part of the workforce trips (Fig. 6c). The global entropy at links H_{GL} captured a lower diversity of commuter flows along links in both networks. The 50% of journeys to work are distributed within the 5% of the OD links in the NP; and within the 7% of the OD links in the GSE. The dominance of few links with high volume (Fig. 6d) and the relatively low link density D in both networks explain the similar results. The average of the local measures $H_{L\mu}^{in}$ and $H_{L\mu}^{out}$ show different characteristics of the systems. Although the GSE has a higher $H_{L\mu}^{in}$ value due to the significant weight of London, for both networks the local average entropies are low. The case of the NP is interesting when comparing a more polycentric structure at the global scale and a low

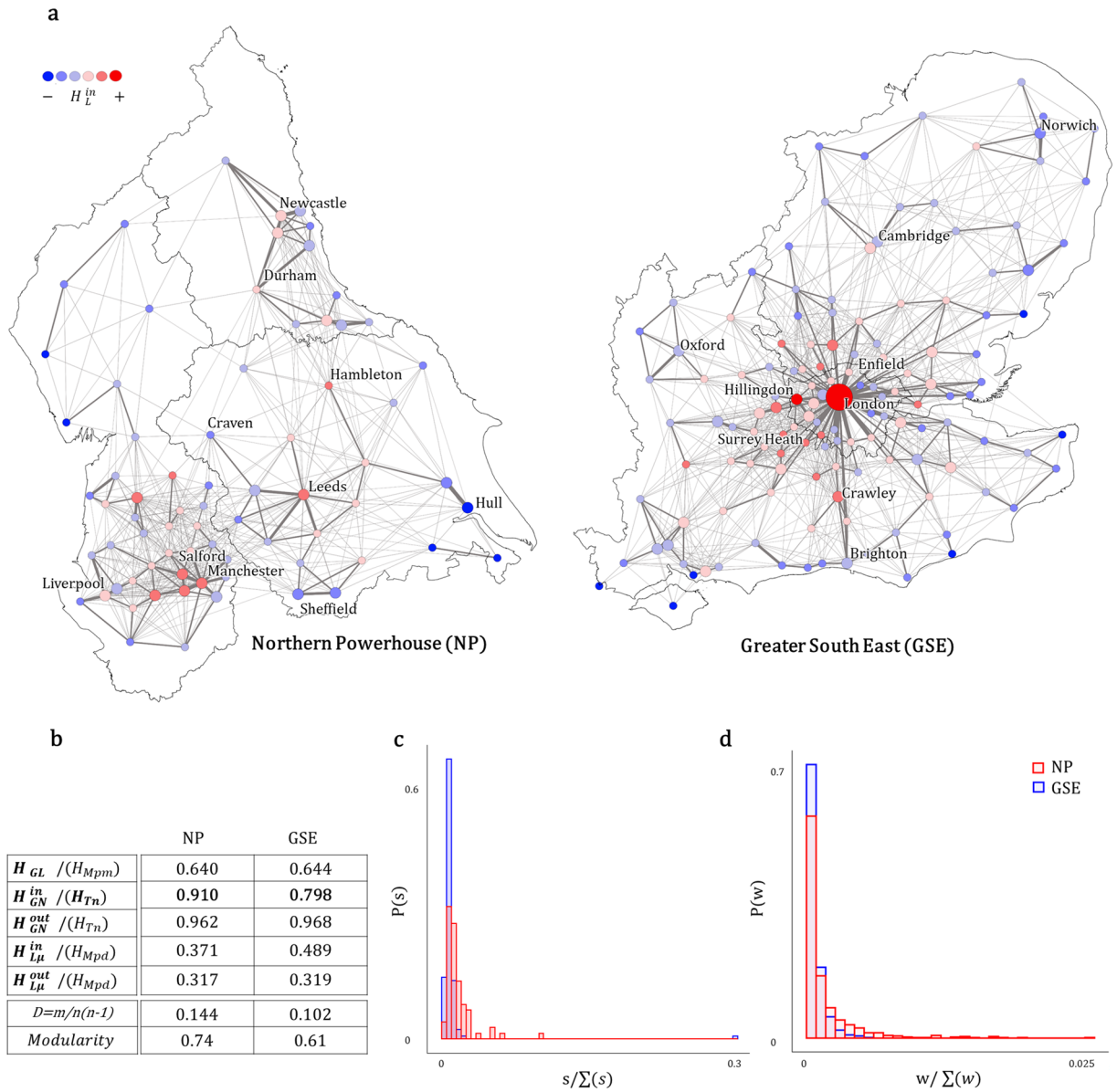


Figure 6. Northern Powerhouse (NP) and Greater South East (GSE) regions commuting patterns. **(a)** GSE and NP commuting networks with nodes corresponding to Local Authority District (LAD) levels. Values of H_L^{in} / H_{Mpd} are indicated in the colour scale of nodes, going from the lowest values in blue to the highest in red. The size of nodes reflects the total count of in-commuting flows (in-strength s). The width of links denotes the flow weight. **(b)** Outputs for normalised entropy measures for both systems, including link density (D) and modularity. **(c)** Normalised node in-strength distribution $P(s)$ and **(d)** normalised link weight distribution $P(w)$. (Map created by the authors with QGIS 2.18.15 software. URL www.qgis.org).

diverse commuting exchange across local entities. We observe a relation of this with the modularity of the NP network, which shows denser connections between the LADs within regions and sparse connection with LADs within other regions or “modules”. These results, however, reflect the structure of the system in 2011, before the Northern Powerhouse initiative in 2015, whose agenda promoted strategies to improve transport connectivity and more diffuse movement patterns to encourage cross-regional exchanges. Future analyses of the 2021 census data could reflect the level of impact of these strategies, and may eventually show greater diversity at the local scale consistent with the polycentric potential of the system at the global level.

On the other hand, local entropy computes diversity of relationships to or from each local authority. In Fig. 6a we can observe values of H_L^{in} in both regions. In the GSE, the LADs with higher diversity are those in the central area, with values decreasing towards the boundaries of the region. In the NP, the highest values are geographically more scattered. Inner London has the highest values for both, H_L^{in} and H_L^{out} , showing significant diversity of flows in both directions (Fig. 7a). In important urban centres such as Manchester, Leeds and Salford, the diversity of workers in-flows is greater than that of their out-flows. Workers coming from other local areas are attracted to these employment hubs with more even volumes, than the ones of their residents commuting to

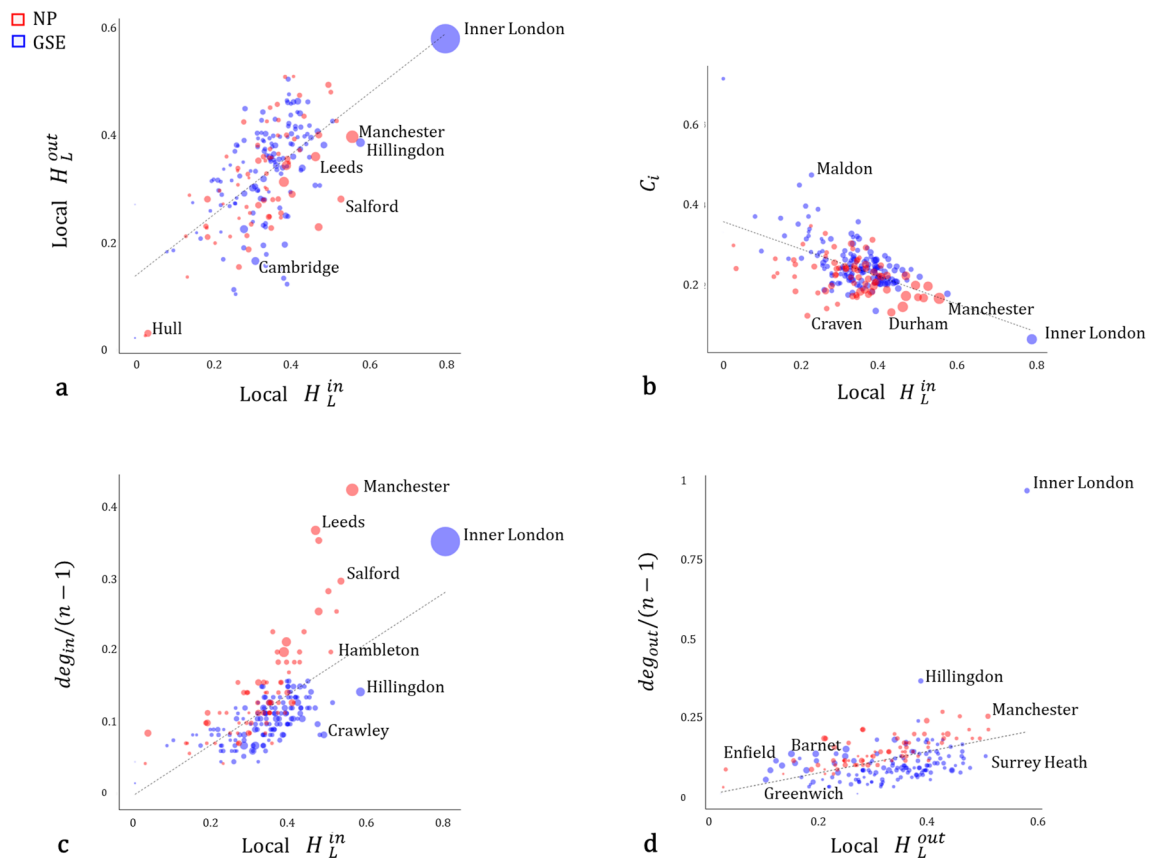


Figure 7. Local measures. Points represent Local Authority Districts (LAD) in the NP and the GSE regions. (a) Normalised local in and out entropy. (b) Clustering coefficient C_i as a function of H_L^{in} . (c) and (d) show in and out-values for degree and local entropy. Degree values are normalised by the total maximum possible degree in each system ($n - 1$). The size of points in (a) and (c) corresponds to the total number of in-commuting (in-strength s), in (d) it corresponds to the total number of out-commuting (out-strength s), and in (b) to the in-degree of each node.

other locations, which tend to be more polarised to fewer centres. In contrast, smaller local authorities tend to have higher out-flow than in-flow diversity, possibly associated with a more prominent role of workers providers. Results in Fig. 7b:d allow us to examine the relationship between values of local entropy, node strength and topological properties such as node degree and clustering coefficient. Figure 7b shows a negative correlation between both measures, and a very similar clustering coefficient for Craven, Durham and London, while our proposed measure of entropy H_L^{in} is able to differentiate between them. In Fig. 7c,d for both cases we observe that the local authorities in the NP have a positive residual in the relationship between degrees and local entropy. For in-commuters (Fig. 7c) this increases notoriously in the most important employment centres like Manchester or Leeds. In terms of the out-commuters (Fig. 7d), Inner London is a particular case. Many workers prefer to live in the central area and commute larger distances, so Inner London is connected to almost every other LAD in the region being its out-degree value notoriously higher than the expected by its entropy. A different case can be found in LADs like Enfield, Greenwich and Barnet, with a high volume of out-flows s and low values of both, H_L^{out} and deg_{out} . Most flows head to inner London than to the rest of the system.

Discussion

Through this work we have shown that the application of entropy theory in the analysis of commuting networks provides relevant information on the distribution of flows in the territory. We explored measures of entropy on global and local scales, as well as on the different constituent elements of the network of flows, links and nodes. Each measure proved relevant in capturing distinct aspects of the spatial interaction patterns.

Link entropy focuses on the interactions between pairs of urban areas, based on the distribution of origin-destination trips. Nodal entropy on the other hand, gives us information on the concentration or dispersion of flows among urban centres. The local analysis examines the relationships between labour supply and demand that a specific area establishes with its most direct context. When extending the analysis to the larger scale, each of the interactions occurring in the system are considered. All the constituent elements, whether they are or not directly or intensely connected to each other, influence the whole-system's entropy.

The latter is particularly useful if for example, we want to analyse the resilience of commuting networks based on the diversity of the structure given by connectivity. Systems can face direct or indirect changes that occur at

different levels. Local entropy will change when endogenous changes in the local labour market alter the structure of relationships of an urban area. On the other hand, changes in global entropy can account for exogenous changes that occur in other local systems. These changes can end up affecting the global structure to a greater or lesser extent, and therefore indirectly modify the structure of relationships between all constituent parts.

It is worth mentioning that with this analysis we are not studying the optimum degree of diversity in the system, nor are we arguing that the maximum possible entropy should be pursued. The functioning of urban systems must be flexible enough to adapt to changes and at the same time efficient enough to optimise resources. As Cabral et al. (2013) argue, if the system falls short of a minimum entropy, the system will be very centralised and therefore vulnerable to changes, while if it exceeds a certain degree of entropy, the system will not be dealing efficiently with resources³⁹. The distribution of workplaces and housing requires a certain degree of concentration to benefit from specialisation and proximity, but at the same time a degree of diversity and dispersion would increase the capacity for resilience and adaptability³². Consequently, the interpretation of the results of the different entropy measures presented in this paper must be made based on the specific criteria of the system under study.

The different measures of entropy presented here contribute to advancing our understanding of the complexity of spatial flows, to inform policy development and take strategic planning actions. By analysing the entropy relative to a maximum possible number of interactions of the system, instead of the given or existing ones, it is possible to compare the system with itself in terms of its maximum capacity. We believe that this form of normalisation presented in this paper facilitates the study of systems based on their own potentialities, offering a different perspective for planning.

This introductory work contributes to the understanding of real commuting networks across many different scales of organisation, in addition to providing a framework to better understand the interplay between transport infrastructure and the layout of economic opportunities in cities.

Received: 1 July 2021; Accepted: 7 December 2021

Published online: 31 January 2022

References

- Pumain, D. Alternative explanations of hierarchical differentiation in urban systems. In Pumain, D. (ed) *Hierarchy in Natural and Social Sciences* 169–222 (Springer, 2006).
- Bretagnolle, A., Pumain, D. & Vacchiani-Marcuzzo, C. The organization of urban systems. In Lane, D. et al. (eds) *Complexity Perspectives in Innovation and Social Change* 197–220 (Springer, 2009).
- Batty, M. *Inventing Future Cities* (MIT Press, 2018).
- De Montis, A., Barthélemy, M., Chessa, A. & Vespignani, A. The structure of interurban traffic: A weighted network analysis. *Environ. Plan. B Plan. Des.* **34**, 905–924. <https://doi.org/10.1068/b32128> (2007).
- Lenormand, M. et al. Influence of sociodemographic characteristics on human mobility. *Sci. Rep.* **5**, 1–15. <https://doi.org/10.1038/srep10075> (2015).
- Louail, T. et al. Uncovering the spatial structure of mobility networks. *Nat. Commun.* **6**, 1–8. <https://doi.org/10.1038/ncomms7007> (2015).
- Patuelli, R., Reggiani, A., Gorman, S. P., Nijkamp, P. & Bade, F.-J. Network analysis of commuting flows: A comparative static approach to German data. *Netw. Spat. Econ.* **7**, 315–331. <https://doi.org/10.1007/s11067-007-9027-6> (2007).
- Louail, T. et al. From mobile phone data to the spatial structure of cities. *Sci. Rep.* **4**, 1–12. <https://doi.org/10.1038/srep05276> (2014).
- Kropp, P. & Schwengler, B. Three-step method for delineating functional labour market regions. *Reg. Stud.* **50**, 429–445. <https://doi.org/10.1080/00343404.2014.923093> (2016).
- Balcan, D. et al. Multiscale mobility networks and the spatial spreading of infectious diseases. *Proc. Natl. Acad. Sci.* **106**, 21484–21489. <https://doi.org/10.1073/pnas.0906910106> (2009).
- Reggiani, A. & Rietveld, P. Networks, commuting and spatial structures: An introduction guest editorial. *J. Transp. Land Use* **2**, 1–4 (2010).
- Levin, S. A. et al. Resilience in natural and socioeconomic systems. *Environ. Dev. Econ.* **3**, 221–262 (1998).
- Berkes, F., Colding, J. & Folke, C. *Navigating Social-ecological Systems: Building Resilience for Complexity and Change* (Cambridge University Press, 2008).
- Ahern, J. From fail-safe to safe-to-fail: Sustainability and resilience in the new urban world. *Landsch. Urban Plann.* **100**, 341–343. <https://doi.org/10.1016/j.landurbplan.2011.02.021> (2011).
- Cumming, G. S., Olsson, P., Chapin, F. & Holling, C. Resilience, experimentation, and scale mismatches in social-ecological landscapes. *Landsch. Ecol.* **28**, 1139–1150. <https://doi.org/10.1007/s10980-012-9725-4> (2013).
- Marcus, L. & Colding, J. Toward an integrated theory of spatial morphology and resilient urban systems. *Ecol. Soc.* **19**(4), 55 (2014).
- Holling, C. S. Understanding the complexity of economic, ecological, and social systems. *Ecosystems* **4**, 390–405. <https://doi.org/10.1007/s10021-001-0101-5> (2001).
- Jacobs, J. *The Death and Life of Great American Cities* (Penguin Harmondsworth, 1961).
- Batty, M., Besussi, E., Maat, K. & Harts, J. J. Representing multifunctional cities: density and diversity in space and time. *Built Environ.* **30**, 324–337. <https://doi.org/10.2148/benv.30.4.324.57156> (2004).
- Bettencourt, L. M., Samaniego, H. & Youn, H. Professional diversity and the productivity of cities. *Sci. Rep.* **4**, 1–6. <https://doi.org/10.1038/srep05393> (2014).
- Hamilton, A. J. Species diversity or biodiversity?. *J. Environ. Manag.* **75**, 89–92. <https://doi.org/10.1016/j.jenvman.2004.11.012> (2005).
- Jost, L. Entropy and diversity. *Oikos* **113**, 363–375. <https://doi.org/10.1111/j.2006.0030-1299.14714.x> (2006).
- Page, S. E. *Diversity and Complexity* Vol. 2 (Princeton University Press, 2010).
- Ben-Naim, A. *A Farewell to Entropy: Statistical Thermodynamics Based on Information* (World Scientific, 2008).
- Shannon, C. E. A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x> (1948).
- Heip, C., Herman, P. & Soetaert, K. Indices of diversity and evenness. *Oceanis* **24**, 61–88 (1998).
- Rashevsky, N. Life, information theory, and topology. *Bull. Math. Biophys.* **17**, 229–235. <https://doi.org/10.1007/BF02477860> (1955).
- Mowshowitz, A. Entropy and the complexity of graphs: I. An index of the relative complexity of a graph. *Bull. Math. Biophys.* **30**, 175–204 (1968).

29. Solé, R. V. & Valverde, S. Information theory of complex networks: On evolution and architectural constraints. In Ben-Naim, E. *et al.* (eds) *Complex Networks* 189–207 (Springer, 2004).
30. Wilhelm, T. & Hollunder, J. Information theoretic description of networks. *Physica A Stat. Mech. Appl.* **385**, 385–396. <https://doi.org/10.1016/j.physa.2007.06.029> (2007).
31. Lowe, J. C. Patterns of spatial dispersion in metropolitan commuting. *Urban Geogr.* **19**, 232–253. <https://doi.org/10.2747/0272-3638.19.3.232> (1998).
32. Goetz, S. J., Han, Y., Findeis, J. L. & Brasier, K. J. Us commuting networks and economic growth: Measurement and implications for spatial policy. *Growth Change* **41**, 276–302. <https://doi.org/10.1111/j.1468-2257.2010.00527.x> (2010).
33. Pappalardo, L. *et al.* An analytical framework to nowcast well-being using mobile phone data. *Int. J. Data Sci. Anal.* **2**, 75–92. <https://doi.org/10.1007/s41060-016-0013-2> (2016).
34. Cottineau, C. & Vanhoof, M. Mobile phone indicators and their relation to the socioeconomic organisation of cities. *ISPRS Int. J. Geo-Inf.* **8**, 19. <https://doi.org/10.3390/ijgi8010019> (2019).
35. Lenormand, M. *et al.* Entropy as a measure of attractiveness and socioeconomic complexity in Rio de Janeiro metropolitan area. *Entropy* **22**, 368. <https://doi.org/10.3390/e22030368> (2020).
36. Gallotti, R., Bazzani, A., Degli Esposti, M. & Rambaldi, S. Entropic measures of individual mobility patterns. *J. Stat. Mech. Theory Exp.* **2013**, P10022. <https://doi.org/10.1088/1742-5468/2013/10/p10022> (2013).
37. Pielou, E. C. The measurement of diversity in different types of biological collections. *J. Theor. Biol.* **13**, 131–144. [https://doi.org/10.1016/0022-5193\(66\)90013-0](https://doi.org/10.1016/0022-5193(66)90013-0) (1966).
38. Green, N. Functional polycentricity: A formal definition in terms of social network analysis. *Urban Stud.* **44**, 2077–2103. <https://doi.org/10.1080/00420980701518941> (2007).
39. Cabral, P., Augusto, G., Tewolde, M. & Araya, Y. Entropy in urban systems. *Entropy* <https://doi.org/10.3390/e15125223> (2013).

Acknowledgements

VM thanks The National Research and Development Agency (ANID) for the financial support provided as a PhD scholarship “Becas Chile”.

Author contributions

V.M. developed the theoretical framework, designed the experiments and conducted the analysis. C.M. and E.A. contributed to the theoretical framework. V.M. wrote the paper, and all authors contributed to the structure and fine tuning of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to V.M.

Reprints and permissions information is available at www.nature.com/reprints.

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



Open Access 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/>.

© The Author(s) 2022