ARTICLE

Check for updates

First-passage times to quantify and compare structural correlations and heterogeneity in complex systems

Aleix Bassolas[1] & Vincenzo Nicosia [1✉]

Virtually all the emergent properties of complex systems are rooted in the non-homogeneous nature of the behaviours of their elements and of the interactions among them. However, heterogeneity and correlations appear simultaneously at multiple relevant scales, making it hard to devise a systematic approach to quantify them. We develop here a scalable and non-parametric framework to characterise the presence of heterogeneity and correlations in a complex system, based on normalised mean first passage times between preassigned classes of nodes. We showcase a variety of concrete applications, including the quantification of polarisation in the UK Brexit referendum and the roll-call votes in the US Congress, the identification of key players in disease spreading, and the comparison of spatial segregation of US cities. These results show that the diffusion structure of a system is indeed a defining aspect of the complexity of its organisation and functioning.

---

[1] School of Mathematical Sciences, Queen Mary University of London, London, UK. ✉email: v.nicosia@qmul.ac.uk

The elements of a variety of complex systems can be naturally associated with one of a small number of classes or categories. Typical examples include the organisation to which an individual belongs[1], the political party of a voter[2], the income level of a household[3] or the functional group of a neuron[4]. Quite often, the co-existence of nodes belonging to different classes and the interactions among those classes play a fundamental role in the functioning of a system. For instance, economic and ethnic segregation in cities is known to be associated with the emergence of social inequalities[5,6]. Similarly, the organisation of neural cells in the brain and the intricate patterns of relations among different functional areas are known to be responsible for the large variety of cognitive tasks that we as humans are able to perform[7,8]. However, obtaining a robust and non-parametric quantification of the heterogeneity of class distributions, especially in systems consisting of a large number of interconnected discrete components, is still an outstanding problem.

The statistics of random walks on interaction graphs have been successfully employed to study the structure of a variety of complex systems[9–12], and have been used to model and characterise transportation[13,14], biological[15,16] and financial systems[17]. And it is now established that the diffusion structure of a graph is intimately connected with many of the meaningful aspects of a complex network[18,19], including the existence of communities[20–22], the distribution of node roles[23], the global navigability of a system[24] and the variability of structural properties in temporal graphs[25]. Quite interesting insights about the relation between the structure and dynamics on a network have come from the analysis of transient and long trajectories of random walks on graphs, including the behaviour of entropy rate[26–28], the statistics of first passage and coverage times[12,29–32] and the systematic study of their fluctuations[33]. However, the potential usefulness of random walks to quantify the heterogeneity of class distributions on networks, either in spatially embedded systems or in high-dimensional networks, has only recently been hinted to[33–35].

We propose here a principled methodology to quantify the presence of correlations and heterogeneity in the distribution of classes or categories in a complex system. The method is based on the statistics of passage times of a uniform random walk on the graph of interconnections among the units of the system. For instance, in the case of a social system we can construct a graph

among individuals based on the observed relations or contacts among them. Similarly, in an urban system we can consider the network of adjacency between census tracts or the connections among census tracts due to human mobility flows. By normalising class mean first passage times (CMFPT) with respect to a null-model where classes are reassigned to nodes uniformly at random, we can effectively quantify and compare the heterogeneity of class distributions in systems of different nature, size and shape.

It is important to stress that, although random walks have been widely used to identify modules and communities of nodes based on how tightly connected they are[20–22,36], here we focus on systems whose nodes have preassigned classes that do not necessarily coincide with their location in the graph. And we study the dual problem of quantifying the heterogeneity and correlations induced by a fixed and exogenous assignment of nodes to classes. We test our framework on a variety of systems with simple geometries and ad-hoc class assignments, and then we use it in three real-world scenarios, namely the quantification of polarisation in the Brexit referendum and in the US Congress since 1926, the role of face-to-face interactions among individuals in the spread of an epidemics and the relation between economic segregation and prevalence of crime in the 53 US cities with more than one million citizens.

Results

Let us consider a graph $G(V, E)$ with $|E| = K$ edges on $|V| = N$ nodes and adjacency matrix $A = \{a_{ij}\}$, and a given colouring function $f : V \to \mathcal{C}$, $\mathcal{C} = \{1, 2, \ldots, C\}$, which associates each node i of G to a discrete label $f(i) = c_i$. Let us also consider a random walk on G, defined by the transition matrix $\Pi = \{\pi_{ij}\}$ where π_{ij} is the probability that a walker at node i jumps from node i to node j in one step (see Fig. 1a). In general, Π could be any row-stochastic transition matrix, but in the following we will consider only uniform random walks.

The method we propose moves from the classical research on mean first passage times (MFPT) between pairs of nodes in a graph[9–12], and focuses instead on the distribution of CMFPT, i.e., the expected number of steps $\tau_{\alpha\beta}$ needed to a random walker to visit for the first time a node of a certain class β when it starts from a node of class α (see Fig. 1a). Consider as an example the two regular square lattices shown in Fig. 1b, c. In both graphs
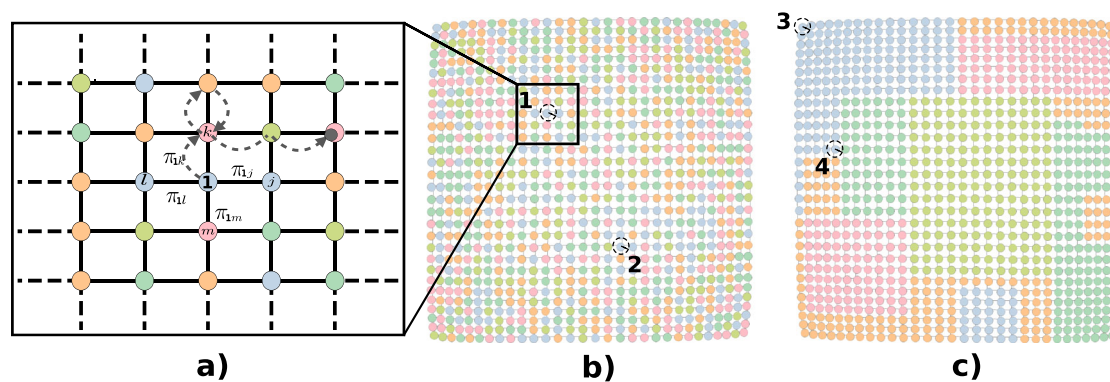

Fig. 1 Characterising colour distributions in regular lattices. The statistics of the amount of time needed to a random walker to hit a node of a given class when starting from another node of the graph encodes useful information about the presence of heterogeneity and correlations in the distributions of node classes. **a** We show a sketch of the trajectory of a random walker over the grid. We indicate by π_{1k}, π_{1l}, π_{1m} and π_{1j} the one-step transition probabilities from node 1 to nodes k, l, m, j, respectively. **b** The nodes of a square lattice are assigned to one of five available classes uniformly at random. As a consequence, the time needed to a random walker to hit an orange node when starting by node 1 is comparable to the time needed when starting from node 2. **c** The assignment of nodes to classes creates homogeneous clusters of nodes of the same colour. As a consequence, the time needed to a random walker to hit an orange node when starting from node 3 is much higher than the time needed when starting from node 4. In all the panels the colour of a node corresponds to the class it is assigned to.

nodes are divided into five classes of equal size. In the graph in Fig. 1b the class to which a node belongs is chosen uniformly at random, while in Fig. 1c, instead, the nodes are associated with a small number of homogeneous clusters. We are interested here in the statistical properties of the symbolic dynamics $W_i = \{c_{i_0}, c_{i_1}, \dots\}$ of node labels or classes visited by the walk W when starting from i. Intuitively, we expect that, for long-enough times, all the trajectories of random walks starting from each of the N nodes of the graph in Fig. 1b will be associated with the same symbolic dynamics, thus becoming indistinguishable. Indeed, that system has neither marked structural heterogeneity, since all the nodes have the same degree (except for the nodes at the border of the grid), nor inhomogeneity in class assignments, since the probability for a node to belong to a certain class does not depend on its position in the graph or on the classes of its neighbours. In particular, a random walker starting at any node belonging to, say, the light-blue class (see nodes 1 and 2 in Fig. 1b) will require on average a small number of steps to hit a node in the orange cluster, since orange nodes can be found in the vicinity of any blue node.

If the association of nodes to classes induces compact clusters, as in the lattice shown in Fig. 1c, then the statistical properties of the symbolic dynamics W_i will heavily depend on the starting node i, despite the fact that almost all the nodes have identical degree. In particular, a random walker starting in the blue cluster at the top-left corner of the graph (node 3) will in general require a very large number of steps before hitting for the first time an orange node. Conversely, a random walker starting at node 4 will hit a node in the orange cluster in a much smaller number of steps, just because one of its immediate neighbours is indeed orange.

Class mean first passage times. A quantity of interest in the study of a symbolic dynamics over graphs is the expected time needed to hit a certain node for the first time, usually known in the literature as the hitting time or MFPT[29]. We denote as $T_{i,\alpha}$ the average MFPT from node i to any node of class α, i.e., the expected number of steps needed to a random walk starting on i to visit for the first time any node j such that $f(j) = \alpha$. Following the formalism to derive the MFPTs between nodes in a graph, we can write a set of self-consistent equations for $T_{i,\alpha}$[18]:

$$T_{i,\alpha} = 1 + \sum_{j=1}^{N} \left(1 - \delta_{f(j),\alpha}\right) \pi_{ij} T_{j,\alpha}. \tag{1}$$

The analytic solution for Eq. (1) depends only on the structure of the graph and on the colouring function f. Let us denote as T_α the column vector of hitting times from nodes of class $\beta \neq \alpha$ to nodes of class α. By convention we set $\{T_\alpha\}_i = 0$ if $f(i) = \alpha$. The self-consistent equation for hitting times can be written as follows:

$$T_\alpha = D_{\bar{\alpha}} + \Pi_{\bar{\alpha}} T_\alpha$$

where $\Pi_{\bar{\alpha}}$ is the transition matrix of the walk where all the rows and columns corresponding to nodes of class α are set to zero. We denote by D_α the indicator vector of nodes belonging to class α, and by $D_{\bar{\alpha}} = \mathbf{1}_N - D_\alpha$ the indicator vector of nodes not belonging to class α, i.e., $\{D_{\bar{\alpha}}\}_i = 1 - \delta_{f(i),\alpha}$. This leads to the solution:

$$T_\alpha = [I - \Pi_{\bar{\alpha}}]^{-1} D_{\bar{\alpha}} \tag{2}$$

The average Class MFPT $\tau_{\alpha\beta}$ from class α to class β is computed as follows:

$$\tau_{\alpha\beta} = \frac{1}{N_\alpha} \sum_{j=1}^{N} D_\alpha^\top T_\beta \tag{3}$$

where N_α is the number of nodes belonging to class α.

The return time to class α, which is the expected number of steps needed to a walker starting on a node of class α to hit a node of class α (including its starting point), can be computed in a similar way. The forward equation for the hitting time to class α from a node of class α reads:

$$R_{i,\alpha} = \sum_{j=1}^{N} \delta_{f(j),\alpha} \pi_{ij} + \sum_{j=1}^{N} (1 - \delta_{f(j),\alpha}) \pi_{ij} (1 + T_{j,\alpha}) \tag{4}$$

where the first contribution accounts for the neighbours of node i that actually belong to class α, while the second contribution corresponds to walks passing through immediate neighbours of i not belonging to class α. The equation can be written in a compact form as follows:

$$R_\alpha = \Pi_{\alpha\alpha} D_\alpha + \Pi_{\alpha\bar{\alpha}} [T_\alpha + D_{\bar{\alpha}}] \tag{5}$$

where R_α is the vector of return times to class α, such that $\{R_\alpha\}_i = 0$ if $f(i) \neq \alpha$, and T_α is the vector of MFPT to class α from nodes that do not belong to class α, as above. Here we denote by $\Pi_{\alpha\alpha}$ the transition matrix of the walk restricted to nodes of class α, i.e., whose generic element π_{ij} is set to 0 if either i or j does not belong to class α. Similarly, $\Pi_{\alpha\bar{\alpha}}$ is the transition matrix restricted to links from nodes of class α to nodes not in class α.

By solving Eqs. (2) and (5) for the grid lattice shown in Fig. 1b we obtain the distribution of CMFPTs provided in Fig. 2a, while for the lattice with clusters in Fig. 1c we obtain the values shown in Fig. 2b. Notice that, as expected, the CMFPTs in the lattice with compact clusters are in general much higher than those observed in the same lattice with uniformly random class assignments. Moreover, both cases we have in general $\tau_{\alpha\beta} \neq \tau_{\beta\alpha}$, since CMFPTs depend primarily on the shape and size of clusters and on the actual fine-grain arrangement of colours.

One could argue that the situation shown in Fig. 1b could be potentially represented by an equivalent mean-field network consisting of a clique of just five nodes, one for each class, and where the probability to jump in one step from a class to any of the another ones is equal for all the classes. However, such a mean-field approximation (see Methods for details) only retains information about the average probability of jumping between two classes in one time-step, which indeed discards all the (possibly relevant) structural information contained in the underlying graph. In particular, in a mean-field model of C nodes in which each class is represented as a supernode the actual distribution of distances among classes is grossly underestimated and flattened, thus resulting in a consequently inaccurate underestimation of CMFPT. For instance, if we use the mean-field approximation for the lattice with random colour assignments shown in Fig. 1b, we get the solutions shown in Fig. 2c, where the values of CMFPT between different classes are substantially smaller than those computed above using Eq. (2). The analysis of suitable mean-field approximations for the computation of CMFPTs is actually quite intricate, and will be thoroughly addressed in a forthcoming paper.

Simple geometries. We compute here the distribution of CMFPT for some simple geometries with simple class assignments. These examples aim at showing that CMFPT depends heavily on class assignment and on the way nodes belonging to the same class are arranged. In all these cases, the symmetric nature of colour assignments will allow us to perform the computations on a minimal weighted graph whose distribution of CMFPT is identical to that of the original system. Notice that those minimal weighted graphs provide exact solutions for the original colour assignment they represent, and should not be confused with mean-field approximations.
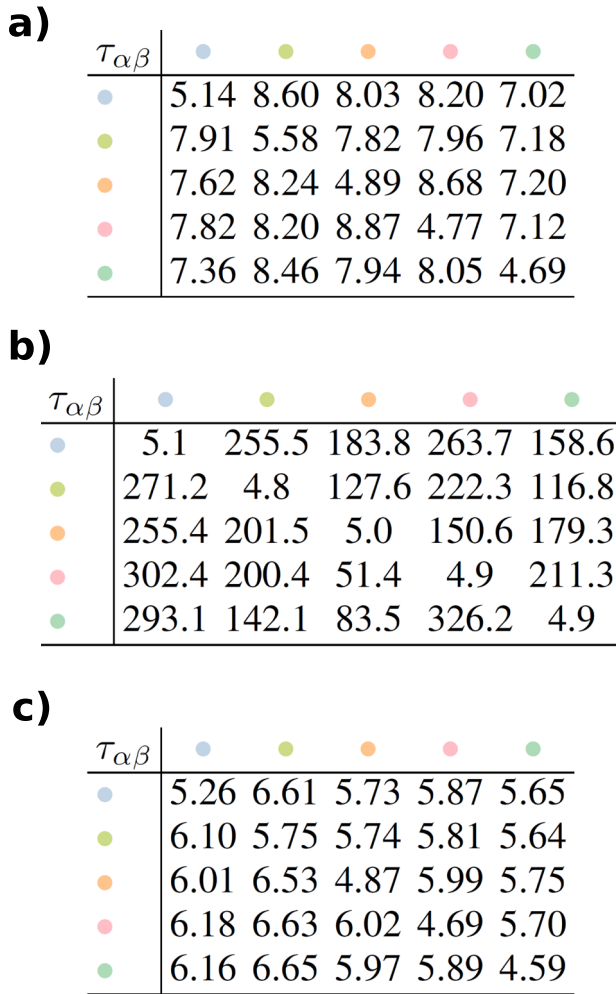
a)

$\tau_{\alpha\beta}$	🔵	🟢	🟠	🔴	🟢
🔵	5.14	8.60	8.03	8.20	7.02
🟢	7.91	5.58	7.82	7.96	7.18
🟠	7.62	8.24	4.89	8.68	7.20
🔴	7.82	8.20	8.87	4.77	7.12
🟢	7.36	8.46	7.94	8.05	4.69

b)

$\tau_{\alpha\beta}$	🔵	🟢	🟠	🔴	🟢
🔵	5.1	255.5	183.8	263.7	158.6
🟢	271.2	4.8	127.6	222.3	116.8
🟠	255.4	201.5	5.0	150.6	179.3
🔴	302.4	200.4	51.4	4.9	211.3
🟢	293.1	142.1	83.5	326.2	4.9

c)

$\tau_{\alpha\beta}$	🔵	🟢	🟠	🔴	🟢
🔵	5.26	6.61	5.73	5.87	5.65
🟢	6.10	5.75	5.74	5.81	5.64
🟠	6.01	6.53	4.87	5.99	5.75
🔴	6.18	6.63	6.02	4.69	5.70
🟢	6.16	6.65	5.97	5.89	4.59

Fig. 2 Class mean first passage times (CMFPT) in synthetic colour distributions. The class mean first passage time (CMFPT) $\tau_{\alpha\beta}$ is the expected number of steps a random walk needs to reach a node of class β when it starts from a node of class α. **a** The CMFPT from each class (left) to each other one (top) in the lattice shown in Fig. 1b, while in **b** we have the CMFPT between classes in the lattice with clusters of Fig. 1c. These values are computed from Eqs. (2) and (5). **c** Mean-field approximation of CMFPT in Fig. 1b based on a clique of five nodes and uniform transition probabilities (see Methods for details).

The first example is that of a two-dimensional square lattice with periodic boundary conditions (a torus), whose nodes are organised in alternate stripes of black and white nodes, as shown in Fig. 3a. Thanks to the symmetric nature of this specific colour assignment, a uniform random walk on that graph is effectively equivalent to a random walk on the weighted minimal two-node graph shown in Fig. 3b, with the transition matrix:

$$\Pi = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

The hitting time to the black node in the minimal equivalent graph when starting from the white node can be written as follows:

$$T_{\circ,\bullet} = 1 + \frac{1}{2}T_{\circ,\bullet}$$

which gives $T_{\circ,\bullet} = 2$ and, by symmetry, also $T_{\bullet,\circ} = 2$.

As a second example we consider a finite chain of white nodes surrounded by black nodes, as shown in Fig. 3c. We are interested here in showing how the length L of a linear cluster of a given colour influences the distribution of CMFPT across the cluster, so we will focus on the CMFPT from nodes of class $\circ$ to nodes of class $\bullet$. In this case the system has a mirror symmetry, which effectively allows us to focus on the $\lceil\frac{L}{2}\rceil$ nodes on either side of the chain. The only caveat is that weighted minimal graphs associated with chains with even length L are slightly different from those associated with chains with odd length, as shown in Fig. 3d. For $L \geq 6$, the closed expression for the distribution of CMFPT from each node in the chain is:

$$T_{k,\bullet} = \left[\frac{4 + A_{M-1}}{4 - B_{M-1}}\right] \prod_{\ell=1}^{k} B_{M-\ell} + \sum_{j=1}^{k-1} A_{M-k+j} \prod_{\ell=0}^{j-1} B_{M-k+\ell} \quad (6)$$

where $M = \lfloor\frac{L-1}{2}\rfloor$, and A_k and B_k are two rational sequences whose form depends on whether the length of the chain is even or odd. More details about the derivation are provided in Supplementary Note 1, where we also compute the distributions for $1 \leq L < 6$. In Fig. 3e we report both the distribution of $T_{k,\bullet}$ across the chain and the average CMFPT to nodes of class $\bullet$. Notice that when $L \gg 1$, $\langle T_\bullet \rangle$ converges to 2, which is the same value obtained in the case of a lattice with rows of alternate colours seen above, as expected.

As a final example we consider the infinite cylinder shown in Fig. 3f, where the nodes in the upper $M + 1$ rows are of class $\circ$ and those in the bottom row are of class $\bullet$. The aim of this example is to show the behaviour of CMFPT as a cluster becomes deeper, i.e., as the nodes in the cluster of a certain class are placed farther away from the frontier with the other cluster. For our purposes, this geometry is effectively equivalent to the linear chain of nodes shown in Fig. 3g, where each row of the original graph is represented by a single node in the minimal graph, with an appropriately weighted self-loop. We can write a set of self-consistent equations for the hitting time to class $\bullet$ for a walk started on each of the rows $k = 0, 1, \ldots, M$:

$$T_{0,\bullet} = 2 + \frac{1}{2}T_{1,\bullet}$$
$$T_{k,\bullet} = 2 + \frac{1}{2}\left[T_{k-1,\bullet} + T_{k+1,\bullet}\right], k = 1, \ldots, M-1 \quad (7)$$
$$T_{M,\bullet} = 3 + T_{M-1,\bullet}$$

whose solution is:

$$T_{k,\bullet} = (k+1)(3 + 4M - 2k), \quad k = 0, 1, \ldots M-1 \quad (8)$$

and

$$T_{M,\bullet} = 3 + M(5 + 2M) \quad (9)$$

The full derivation is reported in Supplementary Note 1.

In Fig. 3h we show the scaling of average MFPT to class $\bullet$, defined as follows:

$$\langle T_\bullet \rangle = \frac{1}{M+1}\sum_{k=0}^{M} T_{k,\bullet} = \frac{8M^3 + 33M^2 + 43M + 18}{6(M+1)} \quad (10)$$

$$\simeq \frac{1}{6}\left(8M^2 + 25M\right) \quad (11)$$

where the approximation is accurate for $M \gg 1$. This means that the average MFPT to class $\bullet$ scales as the square of the height of the cylinder. In Fig. 3h we also show the scaling of the second moment of $T_{k,\bullet}$:

$$\langle T_\bullet^2 \rangle = \frac{32M^5 + 355M^4}{15(M+1)} + o(M^2) \simeq \frac{32M^4 + 323M^3}{15} \quad (12)$$

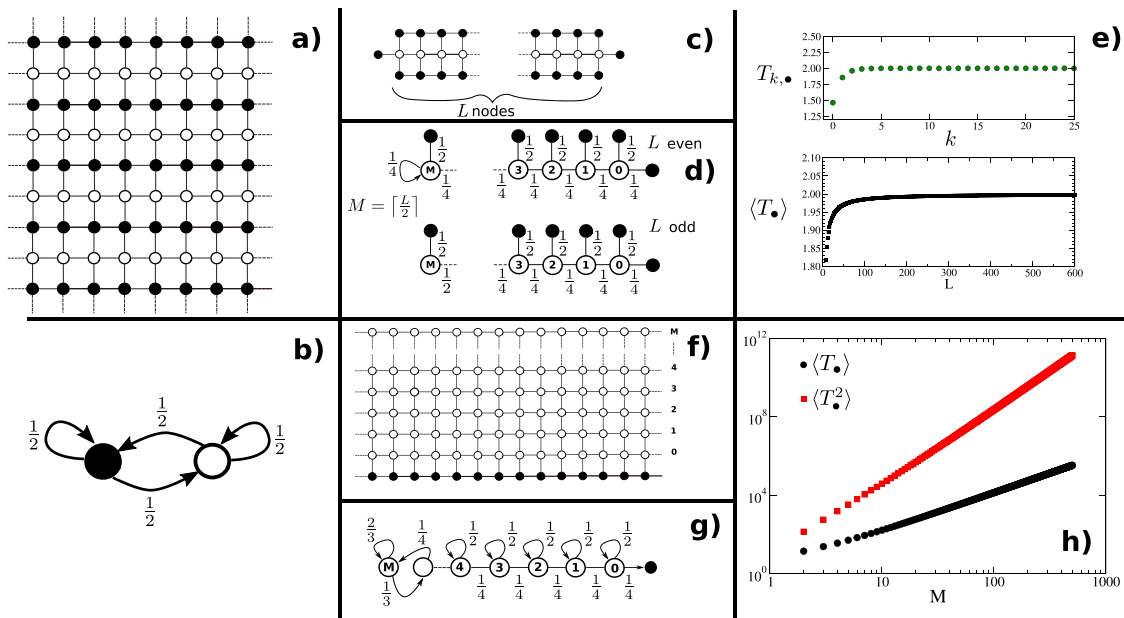which indicates that the standard deviation of $T_{k,\bullet}$ is a function

Fig. 3 Characterisation of colour distributions in simple geometries through class mean first passage times (CMFPT). a Two-colour geometry where the nodes of an infinite lattice are organised in rows of alternate colours. Its equivalent minimal graph (**b**) contains only two nodes with symmetric connections. **c** Two-colour geometry where a chain of L nodes of class ∘ is surrounded by ● nodes. The corresponding minimal associated graph is shown in **d** and is equivalent to a comb graph (the two cases of even and odd values of L are reported for clarity). **e** The CMFPT $T_{k,●}$ of a ∘ node in the comb graph is a function of its distance k from the endpoint, and converges to 2 as k increases . Similarly, the average CMFPT to ● (denoted as $\langle T_● \rangle$) converges to 2 when the length L of the chain increases. **f** Two-colour geometry consisting of an infinite cylinder with $M+1$ rows of ∘ nodes, connected to a row of ● nodes. Its minimal graph (**g**) reduces to a chain with $M+1$ ∘ nodes and one ● node. **h** We show the first two moments of $T_{k,●}$ in an infinite cylinder, as a function of the height of the cylinder.

that grows as $O(M^2)$ as well. In other words, the deeper a cluster the much higher the values of CMFPT to its border, and the much wider the distribution of the CMFPT from any single node of the cluster to the border.

Synthetic colourings in two-dimensional lattices. We study here the distribution of CMFPTs in a two-dimensional square lattice with $N = L \times L$ nodes. Here each node is associated with one of two possible classes, namely ● or ∘, depending on their position in the lattice. Without lack of generality, we set the relative abundance of ● nodes $r = \frac{N_●}{N}$. Then we assign $N_●$ nodes to class ● sampling their coordinates (x, y) from a symmetric two-dimensional Gaussian distribution centred in the middle of the lattice, with standard deviation equal to σ:

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-L/2)^2/2\sigma^2}, \quad (13)$$

By tuning r and σ this model allows for a continuous transition between homogeneous distributions of colours and patterns with strong class segregation. In Fig. 4a–f we show sample sketches of how the spatial distribution of colours looks like as a function of σ and r. For very large values of σ, the picture approaches a homogeneous distribution, regardless of the relative abundance of the two colours. As σ decreases, instead, the nodes of class ● will become more strongly clustered around the centre. The role of the relative abundance of colours is evident from the comparison of Fig. 4b and c, which are two configurations with the same σ respectively for $r = 0.9$ and $r = 0.5$. In particular, we note that the relative abundance of the two colours also has a non-trivial role in determining the degree of mixing between the two classes, since more ∘ nodes can be found inside the ● cluster for $r = 0.5$ than for $r = 0.9$.

A summary of the interplay between these two parameters is reported in Fig. 4g. Since different values of r produce systems with a different level of colour imbalance, we need to properly normalise the values of CMFPT measured in each configuration. We use the normalised CMFPT $\widetilde{\tau}_{●→∘}$ and $\widetilde{\tau}_{●→∘}/\widetilde{\tau}_{∘→●}$, where $\widetilde{\tau}_{\alpha\beta} = \frac{\tau_{\alpha\beta}}{\tau_{\alpha\beta}^{null}}$, and $\tau_{\alpha\beta}^{null}$ is the CMFPT from class α to class β in a null-model where the structure of the graph is preserved but nodes are assigned to classes uniformly at random, preserving the relative abundance of each class (see Methods). The symbols in Fig. 4g correspond to the values of $\widetilde{\tau}_{●→∘}$ and $\widetilde{\tau}_{●→∘}/\widetilde{\tau}_{∘→●}$ obtained for a variety of values of r and σ (the points corresponding to the configurations in Fig. 4a–f are labelled accordingly).

For large values of σ, we expect a more homogeneous distributions of the two colours (see Fig. 4a), and indeed we have $\widetilde{\tau}_{●→∘} \simeq 1$, meaning that the relative distribution of CMFPT from ● nodes is compatible with the one observed in the corresponding null-model. At the same time, the ratio $\widetilde{\tau}_{●→∘}/\widetilde{\tau}_{∘→●}$ is close to 1 as well, meaning that the normalised CMFPTs of the two classes are indistinguishable. As σ decreases, the ● cluster becomes more prominent, but the actual relation between $\widetilde{\tau}_{●→∘}$ and $\widetilde{\tau}_{●→∘}/\widetilde{\tau}_{∘→●}$ will depend on the value of r. In particular, if $r > 0.5$, i.e., nodes of class ● are the majority (see Fig. 4b, d), the ratio $\widetilde{\tau}_{●→∘}/\widetilde{\tau}_{∘→●}$ normally remains larger than 1. This is again expected, since a deeper cluster of ● nodes causes an increase in the CMFPT from ● to ∘ nodes, along the same lines of the increase in CMFPT observed in the simple geometry shown in Fig. 3c. Conversely, if $r < 0.5$ then ● nodes are interspersed within a large cluster of ∘ nodes (see Fig. 4e), making it harder for a walker started at a ∘ node to find a ● node. This results in values of $\widetilde{\tau}_{∘→●}$ larger than 1, i.e., much longer than in the corresponding null-model, and in $\widetilde{\tau}_{●→∘}/\widetilde{\tau}_{∘→●} < 1$. There are some particular situations (see Fig. 4c, f) in which despite the increase in $\widetilde{\tau}_{●→∘}$, the ratio $\widetilde{\tau}_{●→∘}/\widetilde{\tau}_{∘→●}$ remains close to 1. This is due to the fact that
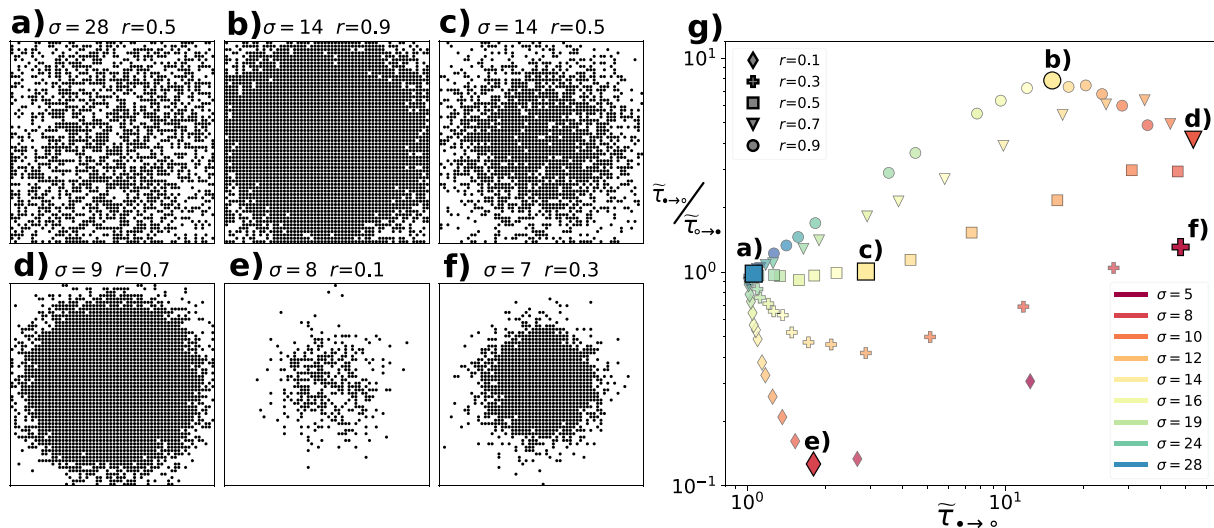
Fig. 4 Characterisation of colour distributions in 2D lattices. a–f Snapshots of the Gaussian cluster model for different values of the parameters σ and r in a regular 2D lattice with 60 × 60 nodes and two colours ● and ○. **a** $\sigma = 28$ and ratio = 0.5; **b** $\sigma = 14$ and ratio = 0.9; **c** $\sigma = 14$ and ratio = 0.5; **d** $\sigma = 9$ and ratio = 0.7; **e** $\sigma = 8$ and ratio = 0.1; **f** $\sigma = 7$ and ratio = 0.3. **g** Transition in the coordinates $\widetilde{\tau}_{\bullet\rightarrow\circ}/\widetilde{\tau}_{\circ\rightarrow\bullet}$ as a function of $\widetilde{\tau}_{\bullet\rightarrow\circ}$ for different values of σ (marker colour) and the ratio between colours (type of marker). Note that σ accounts for the dispersion of ● nodes, with higher values corresponding to a larger dispersion and low clustering.

in these cases the two classes are distributed in a similar fashion, and the relative depths of the two clusters are indeed comparable.

It is worth noting that the $(\widetilde{\tau}_{\bullet\rightarrow\circ}, \widetilde{\tau}_{\bullet\rightarrow\circ}/\widetilde{\tau}_{\circ\rightarrow\bullet})$ phase space provides a very intuitive interpretation and a powerful visualisation of the heterogeneity of distributions of two classes, and it would thus be quite useful to characterise the spatial heterogeneity of generic binary feature distributions and point statistics[37,38].

Polarisation and segregation in voting dynamics. Among the variety of social dynamics that exhibit complex behaviours, voting is possibly one of the most interesting. And not just because anything concerning politics can spur endless and ferocious discussions, but also because voting patterns are the result of the interplay of a variety of factors that are generally difficult to model in an accurate way, including socio-economic and cultural background and spatial and temporal correlations[2,39]. We focus here on two examples where voting dynamics result in the emergence of heterogeneity and correlations, namely the spatial clustering of Leave/Remain voters in the Brexit referendum and the polarisation of opinions of roll-call votes in the US Congress.

Spatial heterogeneity of Brexit vote. Here we show how inter-class MFPTs can be used to quantify the spatial segregation of voting patterns. We consider the results of the so-called Brexit referendum, held in the United Kingdom in 2016 to decide whether to leave the European Union. The referendum had a turnout of 72.2%, and 51.9% of the voters expressed the preference to leave the EU. The results of the referendum have been analysed in several works[40] that have outlined interesting correlations of voting preference with a variety of socio-economic indicators, including age, income level, unemployment and level of education[41]. One of the most intriguing aspects of the results was that the vote was highly segregated. Indeed, highly urbanised areas, as well as the majority of constituencies in Scotland, voted preferentially for Remain, while the rest of the country expressed a preference to Leave.

We constructed the planar graph of constituencies in Great Britain (i.e., all the mainland constituencies in England, Wales and Scotland, leaving out the few constituencies in Northern

Table 1 Normalised class mean first massage times (CMFPT) between constituencies that voted for Leave (L) or Remain (R) in the Brexit referendum if the class assignment is done according to the option with the majority of the votes.

$\widetilde{\tau}_{\alpha\beta}$	L	R
L	0.976	2.857
R	12.304	1.008

Note that $\widetilde{\tau}_{LR}$ is almost five times smaller than $\widetilde{\tau}_{RL}$.

Ireland), where each node is associated with a constituency and a link between two nodes exists if the corresponding constituencies border each other. We assigned each node to either Leave or Remain according to which party won the majority of votes in the corresponding constituency. We then computed the normalised average CMFPT pattern for Remain (R) and Leave (L), we obtained the values shown in Table 1. It is worth noting that while the normalised return time to each class is close to 1 in both cases, i.e., it is consistent with the corresponding null-model where classes are reassigned to nodes uniformly at random, the proper inter-class MFPTs exhibit a pronounced disparity between the two classes. In particular, the normalised CMFPT from Leave to Remain $\widetilde{\tau}_{LR}$ is much smaller (2.827) than its counterpart $\widetilde{\tau}_{RL}$ (12.304). This means that, on average, in this graph is much easier for a random walker starting at a node whose citizens expressed a majority of votes for Leave to arrive at a node where people preferentially voted for Remain, than the other way around. Or, putting it in another way, it was much easier for a Leave supporter wandering through the graph to meet a Remain supporter than the other way around.

A possible interpretation of this result is the presence of a structural reinforcement of segregation. Indeed, if we assume that voters could influence each other by discussing the matter of the referendum with other voters holding opposite opinions, then a person determined to vote for Remain would have had a much harder time finding a Leave supporter to convince. On the

Table 2 Normalised class mean first massage times (CMFPT) between constituencies that voted for Leave (L) or Remain (R) in the Brexit referendum if we consider the ensemble of colour assignments where a node is assigned to Leave (Remain) with a probability equal to the fraction of voters supporting Leave (Remain) in that constituency.

$\widetilde{\tau}_{\alpha\beta}$	L	R
L	1.018	1.057
R	1.090	0.981

The resulting normalised CMFPT is not distinguishable from the null-model expectation.

contrary, Leave supporters would have been able to find Remain supporters much more easily, as Leave constituencies are on average closer (in terms of MFPT) to other Remain constituencies than the other way around.

However, this picture is completely reversed if we take into account the actual percentage of votes for Leave and Remain in each constituency, instead of noting only which party got a majority. We considered an ensemble of colour assignments obtained by assigning colour L to node i with probability $p_L(i)$ equal to the percentage of Leave voters in constituency i. For instance, if in a given constituency i we had 45% of votes for Leave, then node i will be assigned to class L with probability 0.45, and to class R with probability 0.55. We computed the inter-class MFPTs in this ensemble of colourings, and normalised them by the corresponding values in a null-model where we reassigned vote proportions among nodes uniformly at random. The results are reported in Table 2.

It is evident that, by taking into account the actual distribution of voters in each constituency, the pattern of inter-class MFPT becomes practically indistinguishable from the one we would observe in the corresponding null-model. This means that there was indeed no significative spatial segregation effect in the Brexit vote, and indicates that indeed the reasons in support for Leave or Remain were most probably linked to socio-economic characteristics, rather than to geographical ones.q

Polarisation of roll-call votes. In this section we show how $\widetilde{\tau}_{\alpha\beta}$ can be used to quantify the level of polarisation in roll-call votes in the US congress[42–44], and to keep track of its evolution over time. We considered the full data set of affiliation and single roll-call votes of the members of the US Congress, and we built a weighted graph among members of each chamber in each term between 1929 and 2016, where the weight of the edge connecting two nodes is equal to the number of times the votes of the corresponding members in a roll call coincided. We assigned each node to either Republicans, Democrats or Others, according to the party to which they belonged, but we focused exclusively on the CMFPT between Republicans and Democrats, since members of Other parties are normally a rather small minority, if present at all.

It is worth noting that, at difference with the synthetic networks we have studied so far, these networks do not admit a natural embedding in a metric space. Intuitively, we expect that members of both the Senate and the House of Representatives would, in general, be more likely to vote as other members of their party, giving rise to somehow definite clusters. However, the situation is not always that clear. In Fig. 5a–c we show the networks of Senate members observed in three different terms from the last 30 years. Indeed, it is evident that stronger connections between members of the same party appear as time passes. Moreover, when comparing the 112th and 115th congresses, we can see that the party with majority tends to be

more heavily connected than the other one. This is also reflected in the relative size of nodes, which is proportional to the corresponding CMFPT to the other party. Just by looking at these three graphs we would be inclined to think that the level of polarisation in the US Congress seems to increase over time.

For each graph, we also show the distribution of normalised MFPTs $\widetilde{T}_{i,\alpha}$ from each node i to each of the parties, respectively, for Democrats and Republicans. By looking at these distributions on the same scale, it is evident that the polarisation, intended as the relative distance between nodes of different parties as measured by inter-class MFPT, has increased substantially in the last 30 years. In particular, the separation between the intra- and inter-class MFPT distributions has increased dramatically, to the point that in the 115th Congress the distributions of intra-class and inter-class MFPT are clearly separated. When comparing the 112th and 115th legislatures we also observe that the changes in the shape and position of the MFPT distributions seem to depend on which party holds the majority of the seats. In particular, when Democrats have the majority, the CMFPT to Republicans across nodes is higher than from Republicans to Democrats. Conversely, when Republicans are ruling the situation is inverted, pointing out that the party that holds the majority seems to be the main driver of polarisation.

In Fig. 5d, e, we show the evolution of the average $\widetilde{\tau}_{\alpha\beta}$ between the two main US parties in both the Senate and the House of Representatives. The dashed lines indicate the intra-class MFPT, and indeed support the intuition that polarisation has increased dramatically in recent years. The plots of return times $\widetilde{\tau}_{\alpha\alpha}$ instead (solid lines) are far more stable, and the small oscillations we observe depend only on which is the ruling party, since that one will normally yield lower values of $\widetilde{\tau}_{\alpha\alpha}$.

We quantify the overall polarisation in the Senate and the US House of Representatives by computing the average between the inter-class passage times $(\widetilde{\tau}_{D\to R} + \widetilde{\tau}_{R\to D})/2$, which provides an estimate of how close are the voting behaviours of the two main parties in each Congress. The temporal evolution of $(\widetilde{\tau}_{D\to R} + \widetilde{\tau}_{R\to D})/2$ is shown in Fig. 5f, where each point is coloured according to the party holding the majority of seats in that term. Again, we observe a clear increase of polarisation after the 1970s in both branches of the Congress, which is in very good agreement with prior works[42–44].

We inspect next if, as hypothesised, the increasing polarisation observed in Fig. 5f is due to the party holding the majority of seats in term. We computed the ratio of inter-class passage times $\widetilde{\tau}_{Maj\to Min}/\widetilde{\tau}_{Min\to Maj}$, where $\widetilde{\tau}_{Maj\to Min}$ is the normalised inter-class MFPT from the party with the majority to the party with the minority and $\widetilde{\tau}_{Min\to Maj}$ is the normalised inter-class MFPT from the minority to the majority. The values of $\widetilde{\tau}_{Maj\to Min}/\widetilde{\tau}_{Min\to Maj}$ (Fig. 5g) are indeed relatively stable over time, with the vast majority of points lying along or above the solid grid line corresponding to absence of polarisation. This indicates that in most of the terms over the last 80 years the party holding the majority has been the main driver of polarisation. These results demonstrate that $\widetilde{\tau}_{\alpha\beta}$ is indeed a robust measure of polarisation. We argue that the same framework could be easily used in other contexts, including the polarisation of discussions in (online) social networks[45,46] or the flip of candidates between different political parties[47]. In particular, it could be also possible to identify those agents or individuals who contribute more to polarisation by looking at the ranking of nodes by their values of CMFPT to other classes.

Contact assortativity and relation with epidemics spreading. Understanding the mixing between groups of individuals in a

Fig. 5 Roll-call polarisation in the Senate and the US House of Representatives. a–c Three networks among the members of the Senate and the corresponding distribution of normalised class mean first passage times (CMFPT) from the each node (member) i to each of the parties α. **a** 101st Congress (1989–1990), **b** 112th Congress (2011–2012) and **c** 115th Congress (2017–2018). Nodes correspond to republican (red) and democrat (blue) senators, while the width of each edge is proportional to the number of calls in which the corresponding two members voted the same. **d–e** Temporal evolution of the normalised CMFPT $\widetilde{\tau}_{\alpha\beta}$ from class α to class β (in **d** the Senate and in **e** the US House of Representatives, respectively). **f** Temporal evolution of polarisation in the Senate and the US House of Representatives, as measured by $(\widetilde{\tau}_{D\rightarrow R} + \widetilde{\tau}_{R\rightarrow D})/2$ where D stands for Democrats and R for Republicans. Triangles correspond to the House of Representatives and circles to the Senate and their colour is set to red when Republicans hold the majority and blue when Democrats hold it. **g** Temporal evolution of $\widetilde{\tau}_{Maj\rightarrow Min}/\widetilde{\tau}_{Min\rightarrow Maj}$ where $\widetilde{\tau}_{Maj\rightarrow Min}$ is the normalised CMFPT from the party holding the majority to the minority one, while $\widetilde{\tau}_{Min\rightarrow Maj}$ is the normalised CMFPT from the party in minority to the one holding the majority. The fact that this quantity is larger than one in most of the terms indicates that the party with a majority is the main driver of polarisation.

network can provide a lot of information about the properties of social dynamics, including the role played by different individuals in the transmission of diseases[1,48–50]. As a second case study, we use CMFPTs to identify how different groups of people interact in three face-to-face contact networks, namely the contacts in a hospital, in a school and in an enterprise, obtained from the SocioPatterns project data set[1,51–53].

The definition of groups or classes is specific of each system, e.g., the role of a person in the case of hospitals, the class in the case of schools and the department in which a person works for the network of contacts in an enterprise environment. The weight of the undirected edge connecting two nodes in each graph is equal to the number of contacts between the corresponding individuals. By looking at the dynamics of two simple epidemic models, namely a susceptible–infected–susceptible (SIS) and a susceptible–infected–recovered (SIR), we show here that the distribution of CMFPT in each graph provides relevant information about the dynamics of disease spreading in the system. For both epidemic models, if an individual i is infected it selects one of its neighbours j with probability $w_{ij}/\sum_j w_{ij}$ and infects it with a probability β. Afterwards, each of the infected individuals will either recover with probability μ in the case of the SIR model, or become susceptible again in the case of the SIS model.

In Fig. 6a–c, we report the matrices of $\widetilde{\tau}_{\alpha\beta}$ respectively for the school, the enterprise and the hospital. In the school network, we observe a consistent pattern of lower values of normalised intra-CMFPT $\widetilde{\tau}_{\alpha\alpha}$, which is most probably due to the much higher number of contacts among individuals in the same classroom compared to individuals in other classrooms. Notwithstanding this general pattern, some of the classes are more tightly

connected than other close-by classes, as in the case of ce1b and ce2b. Also in the case of the enterprise network we distinguish a clear pattern where for each class α the value of $\widetilde{\tau}_{\alpha\alpha}$ is normally much smaller than $\widetilde{\tau}_{\alpha\beta}$ for $\alpha \neq \beta$. Yet, there are some interesting deviations, as in the case of SFLE. This department plays a similar role to that of teachers in the case of the school network (i.e., contain people who tend to interact with more than one class), and all the other departments exhibit similar values of inter-CMFPT to that class. Finally, in the hospital, contact network patients seem to be the most isolated, while the paramedical staff and administratives display lower CMFPT to all the other classes.

As we show in the following, we found a quite interesting relation between the pattern of CMFPT in each network and the spreading dynamics on the same graph. We ran a large number of simulations of the SIR and SIS models, seeding the disease in each of the nodes of a network, and calculating the number of time steps needed to the spread to reach the peak in each of the classes, as a function of the group to which the seed node belongs. In Fig. 6d–f we report, as a function of $\widetilde{\tau}_{\alpha\beta}$, the average number of steps until the peak of the epidemic t_{peak} in class β is reached in a SIR model where the seed is a node in class α. Interestingly, we found that the time to the peak in class β is an increasing function of $\widetilde{\tau}_{\alpha\beta}$, and the rank correlation between the two variables is always pretty large and significant in the three systems.

The results on the SIS model somehow complement the picture observed in the case of SIR. In Fig. 6g–i we show that the class mean return times $\widetilde{\tau}_{\alpha\alpha}$ are strongly correlated with the fraction ρ_α of infected individuals of that class in the stationary state. In
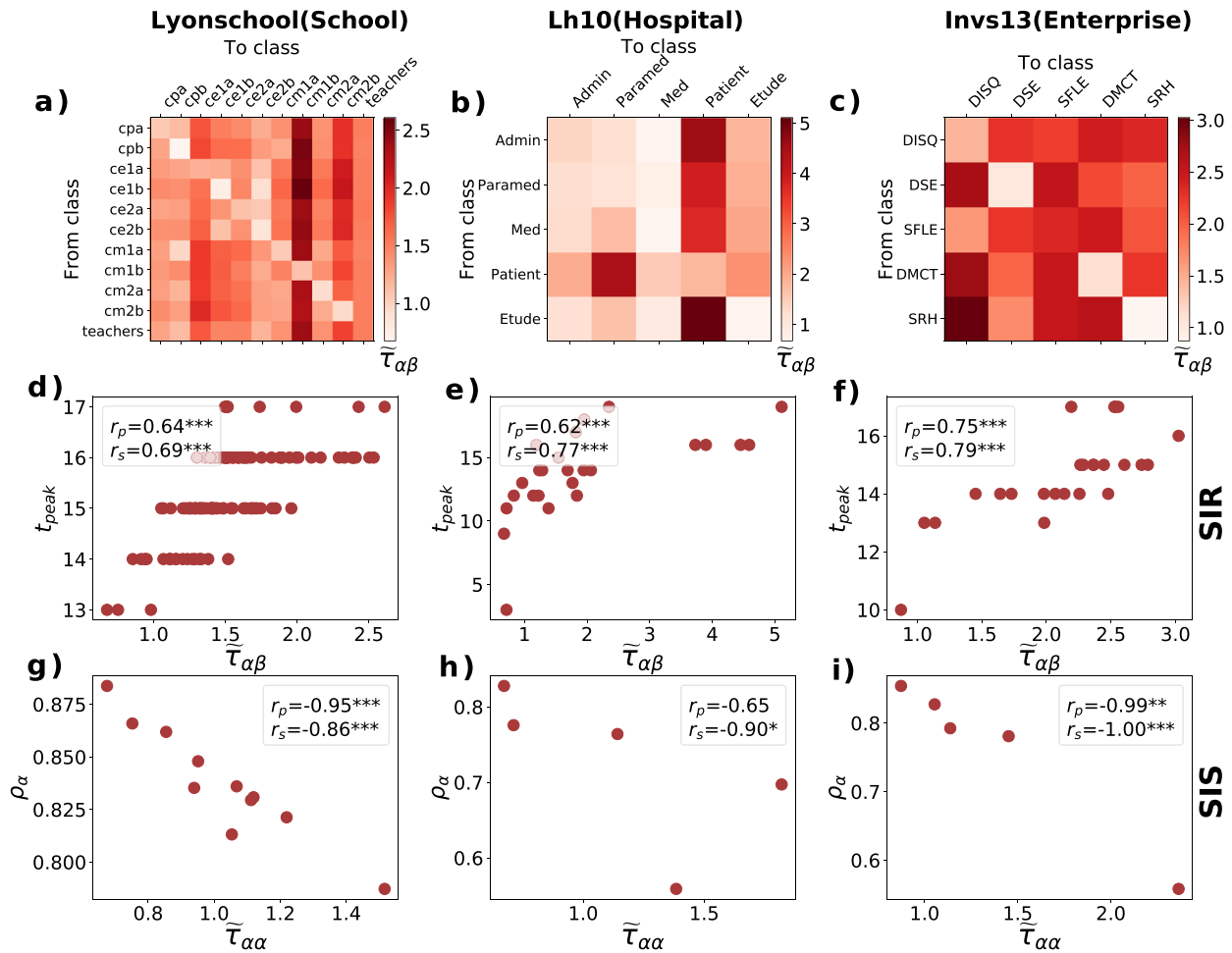
Fig. 6 Class mean first passage times and the spread of epidemics in contact networks. a–c Class mean first passage times in contact networks corresponding to a school, a hospital and an enterprise. Each acronym corresponds to a class in the case of the school, a role in the case of the hospital and a department in the case of the enterprise. **d–f** Relation between the normalised class mean first passage times from class α to class β $\widetilde{\tau}_{\alpha\beta}$ and the time steps until the peak of the epidemic t_{peak} in a susceptible–infected–recovered (SIR) model. **g–i** Relation between the return times $\widetilde{\tau}_{\alpha\alpha}$ and the fraction of infected individuals in each class in the stationary state ρ_α in a susceptible–infected–susceptible (SIS) model. Simulations were performed using the parameters $\beta = 0.7$ and $\mu = 0.1$. r_p stands for the Pearson correlation coefficient while r_s for the Spearman correlation coefficient. The significance is determined by the p value (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$).

particular, the lower the value of $\widetilde{\tau}_{\alpha\alpha}$, the larger the fraction of infected individuals in class α in the endemic state, suggesting that the steady-state dynamics is indeed predominantly driven by interactions among individuals in the same class. In Supplementary Table 1 we report these correlations for a wider range of β and μ. As shown in detail in Supplementary Figs. 1 and 2, the observed correlations between ρ_α and $\widetilde{\tau}_{\alpha\beta}$ are consistently higher than the correlation with either the total number of edges between two classes or the total fraction of edges from class α to class β. Those results as well as the correlations with a wider range of parameters and contact networks can be found in Supplementary Note 2.

Residential and dynamical urban economic segregation. Socio-economic segregation has an enormous impact on city livability, and many different measures to quantify it have been proposed in the past years. However, most of those measures, with a few notable exceptions[54], focus on first-neighbour information, and disregard the role of the mobility of citizens[55–58]. We test here the potential of CMFPTs to quantify urban economic segregation in large metropolitan areas, taking into account the daily mobility patterns of individuals.

We considered the 53 US cities with more than one million inhabitants, and census information about the number of households in each of the 16 income categories defined by the US Census Borough (see Supplementary Table 2 and the 2017 American Community survey[59]), where Class 1 is lowest income and Class 16 is the highest one. We constructed two different graphs among census tracts, namely the graph of tract adjacency and the graph of daily workplace commuting[60]. The former graph is undirected and unweighted, and is better suited to measure the so-called residential segregation, i.e., the extent to which people with similar levels of income tend to live in close-by areas. The commuting graph, instead, is directed and weighted so that the weight of a link going from i to j is given by the sum of the residents in i working in j and the residents of j working in i in order to mimic the daily mobility of citizens.

The fact that households of more than one category are present in each block group prevents us from assigning a single class to each unit, so we computed the distribution of CMFPT by averaging over a large number of realisations of class assignments. In each realisation, the class of each node in the graph is sampled from the distribution of household in the corresponding census tract, so that the probability that node i is assigned to class α in a
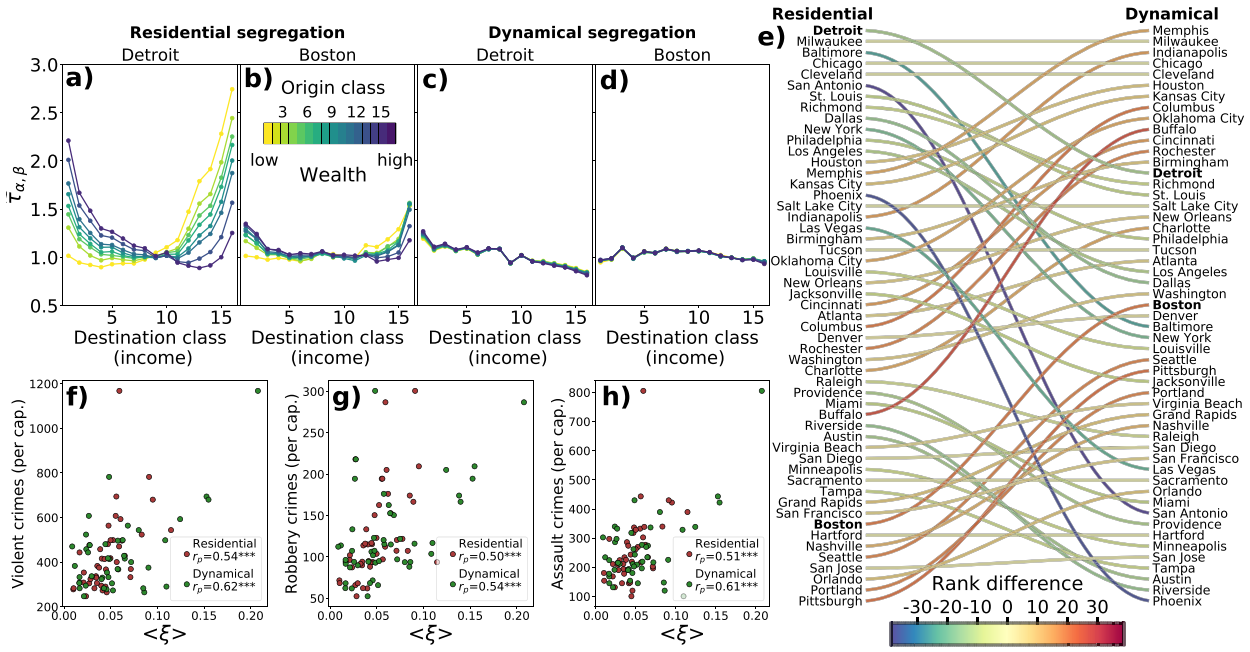
Fig. 7 Class mean first passage times and urban income inequality. a–d Normalised class mean first passage times from class α to class β $\widetilde{\tau}_{\alpha\beta}$ between each of the income categories in Detroit and Boston when computed over the adjacency graphs (residential segregation, **a** and **b**) and the commuting graphs, respectively (dynamical segregation, **c** and **d**). Each class corresponds to an income bracket provided by the US census: Class 1 is the poorest, while Class 16 is the wealthiest. See Supplementary Table 2 for details on the average income of each class. **e** Changes in the ranking of the residential and dynamical $\langle \xi \rangle$ starting with higher values on top and lower values at the bottom. **f–h** Correlation between the index of segregation $\langle \xi \rangle$ in US cities and the levels of **f**) violent crimes, **g**) robbery crimes and **h**) assault crimes per capita. r_p stands for the Pearson correlation coefficient and the significance is determined by the p value (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$).

given realisation is equal to $m_{i,\alpha}/\sum_{\forall\beta}m_{i,\beta}$, where $m_{i,\alpha}$ is the number of people of category α living in node i. We took a slightly different approach for the commuting graph as described in[34], where the population attributed to node i is a combination of the resident population at i and the number of commuters working at i:

$$\widetilde{m}_{i,\alpha} = m_{i,\alpha} + \sum_j \omega_{ji} \frac{m_{j,\alpha}}{\sum_{\forall\beta}m_{j,\beta}}, \qquad (14)$$

where ω_{ji} is the weight from node j to node i in the commuting graph, which is equal to the total number of people who live in j and commute to their workplace in i. By doing so we take into account the fact that commuters effectively contribute to the diversity of an area, as they spend a considerable amount of time in there and actually interact with other commuters coming from different areas as well. For each realisation of a class assignment, we run 10^4–10^6 random walkers from each of the nodes. Then, we obtain the CMFPT $\tau_{\alpha\beta}$ for each ordered pairs of classes (α, β) by averaging over all nodes and realisations of class assignments, and we analyse the normalised CMFPT $\widetilde{\tau}_{\alpha\beta} = \tau_{\alpha\beta}/\tau_{\alpha\beta}^{null}$.

In Fig. 7a, b we show the profiles of $\widetilde{\tau}_{\alpha\beta}$ from each of the 16 classes computed over the adjacency graph, respectively, for Detroit and Boston. It is worth noting that the categories at the two extremes (i.e., the poorest and the wealthiest ones) exhibit a quite similar pattern in the two cities. They are both characterised by larger values of CMFPT from any of the other classes, meaning that those two classes are in general more isolated from the rest of the population, with most of the high-income classes appearing slightly more isolated in Detroit than in Boston. Moreover, in both cities all classes have a virtually identical value of CMFPT to Class 9, and very similar values to Classes 8 and 10, which indicates that these middle-income classes play a pivotal role in the spatial distribution of income. However, despite the fact that the qualitative behaviour is similar in the two cities, there are

some noticeable quantitative differences. First of all, the values of normalised CMFPT are significantly larger in Detroit than in Boston. Second, in Detroit we observe a strong dependence of CMFPT on the class of the starting node, while in Boston the average number of steps required to reach a class below 12 is almost constant regardless of the category of the origin node. These important quantitative differences would suggest that the spatial distribution of income in Detroit is more heterogeneous than in Boston, a conclusion which is in line with the classical literature about economic segregation in the US[3,61–63].

However, in large metropolitan areas most of the daily activities of individuals happen far away from their home, due to urbanisation pressure and to decentralisation of productive sectors, so that residential segregation can hardly tell the whole story. Indeed, in the last years there has been an increasing interest in the quantification of social and economic segregation by taking into account mobility patterns[64–66]. This is easily doable within our framework by letting the walkers move on the mobility graph instead of the adjacency network between census tracts. We have computed $\widetilde{\tau}_{\alpha\beta}$ upon the mobility graph of each US city in the data set, as obtained from workplace commuting information. The results are shown in Fig. 7c, d, and provide an interesting picture of the differences between residential and mobility-focused segregation. First of all, in both cities $\widetilde{\tau}_{\alpha\beta}$ does not depend too much on the origin class α but instead on the destination class β. This is most likely due to the fact that people of different backgrounds commute to similar areas, i.e., the city centre and industrial sites. Still, both cities display a distinct organisation of CMFPTs. In Detroit low-income classes are much more isolated (higher $\widetilde{\tau}_{\alpha\beta}$) compared to high-income classes, which exhibit systematically lower values of $\widetilde{\tau}_{\alpha\beta}$. In the case of Boston, instead, $\widetilde{\tau}_{\alpha\beta}$ is almost flat with no important dependence on either the source or the destination class.

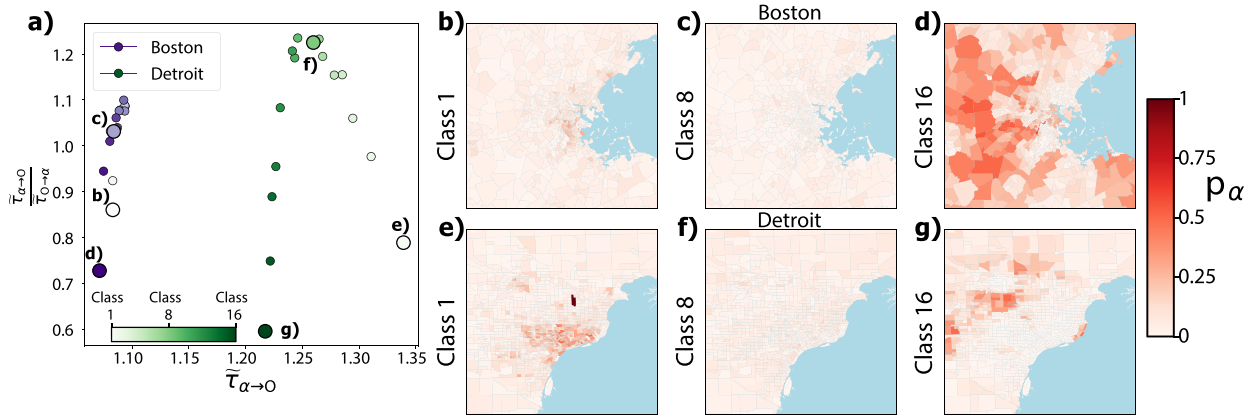**Fig. 8 Characterisation of residential segregation in Boston and Detroit through class mean first passage times (CMFPT). a** Position in the $(\widetilde{\tau}_{\alpha\to O}, \widetilde{\tau}_{\alpha\to O}/\widetilde{\tau}_{O\to\alpha})$ phase space of each of the income classes in Boston (shades of blue) and Detroit (shades of green). Here $\widetilde{\tau}_{\alpha\to O}$ stands for the average normalised CMFPT from a class α to the rest of classes and $\widetilde{\tau}_{O\to\alpha}$ from the rest of classes to a class α. **b–d** Distribution of the probability of finding a class α in each census tract of Boston, respectively for Class 1 (**b**), Class 8 (**c**) and class 16 (**d**). **e–g** Distribution of the probability of finding a given class α in each census tract of Detroit, again for Class 1 (**e**), Class 8 (**f**) and Class 16 (**g**). Class numbering goes from the most deprived (Class 1) to the wealthiest one (Class 16).

The profiles of $\widetilde{\tau}_{\alpha\beta}$ that we show in Fig. 7a–d provide an overall clear picture of the distribution of CMFPT in a metropolitan area, but do not allow us to easily compare two cities in a systematic manner. Hence, we devised two synthetic indices $\overline{\xi^{out}}$ and $\overline{\xi^{in}}$ that summarise the information on spatial income heterogeneity in a single number. The idea behind these quantities is that an income class α is more heterogeneously distributed if there is a large difference between the CMFPT from α to the income classes immediately adjacent to α and the median CMFPT from α to any other class. We consider the discrepancy between the local and global median of CMFPT from class α:

$$\xi^{out}_\alpha = |\overline{\widetilde{\tau}}^{out}_{\alpha_{nn}} - \overline{\overline{\widetilde{\tau}}}^{out}_\alpha|$$

where $\overline{\overline{\widetilde{\tau}}}^{out}_\alpha$ is the median of $\widetilde{\tau}_{\alpha\beta}$ when $\alpha \neq \beta$ and $\overline{\widetilde{\tau}}^{out}_{\alpha_{nn}}$ is the median of $\widetilde{\tau}_{\alpha\beta}$ to its nearest neighbours $\beta \in \{\alpha - 1, \alpha, \alpha + 1\}$, if $\alpha = 1$ or $\alpha = 16$ we only consider the median between, respectively, $\{\alpha, \alpha + 1\}$ or $\{\alpha - 1, \alpha\}$. Similarly for the discrepancy between local and global median of CMFPT to class α:

$$\xi^{in}_\alpha = |\overline{\widetilde{\tau}}^{in}_{\alpha_{nn}} - \overline{\overline{\widetilde{\tau}}}^{in}_\alpha|$$

where $\overline{\overline{\widetilde{\tau}}}^{in}_\alpha$ and $\overline{\widetilde{\tau}}^{in}_{\alpha_{nn}}$ are now the median of $\widetilde{\tau}_{\beta\alpha}$ when $\alpha \neq \beta$ and the median of $\widetilde{\tau}_{\beta\alpha}$ from its nearest neighbours $\beta \in \{\alpha - 1, \alpha, \alpha + 1\}$, respectively.

In Fig. 7e we show the ranking of US cities induced by $\langle\xi\rangle = (\overline{\xi^{in}} + \overline{\xi^{out}})/2$, which is the average discrepancy between local and global CMFPT from/to each income class. In general, larger values of $\langle\xi\rangle$ indicate more pronounced levels of segregation. On the left-hand side of Fig. 7e the cities are ranked according to $\langle\xi\rangle$ in the adjacency graph of census tracts, while on the right-hand side the ranking is based on $\langle\xi\rangle$ in the commuting graph. Interestingly, Detroit is the first US city by residential segregation, with other cities traditionally known for their high levels of segregation like Milwaukee and Cleveland following closely. Conversely, Boston is at the bottom of the ranking. However, the ranking changes substantially if we consider instead the mobility graph, and what we call dynamic segregation, as reported in the right-hand side of Fig. 7e. For instance, Baltimore (which is ranked pretty high for residential segregation) gets relegated to a mid-rank position, while cities like Buffalo or

Indianapolis, where residential segregation is not that high, get to the top of the ranking of dynamic segregation.

The most interesting aspect of $\langle\xi\rangle$ is that it captures some of the most undesirable consequences of income segregation, namely the incidence of different types of crimes obtained from[67]. In Fig. 7f–h we report the correlation between incidence per capita of assaults, violent crimes and robberies with the levels of residential segregation measured on the adjacency and on the commuting graph of the cities in our data set. Interestingly, both indices display a significant correlation with all three types of crime. Moreover, the indices computed over the commuting network display a stronger correlation in all three cases, reinforcing the idea that quantifying segregation by disregarding mobility can indeed lead to distorted conclusions. For instance, the city that appears on the top of the dynamical segregation ranking (Memphis) is the one displaying the highest incidence of violent crimes per capita among the largest US cities, although it is placed in the second quartile of the ranking by residential segregation. As a comparison, we report in Supplementary Note 3 and Supplementary Figs. 3–5 the correlations of traditional metrics of segregation (i.e., the Spatial Gini coefficient and the Moran's I index) with the same crime indicators, showing that they are not able to attain values as high as those obtained with $\langle\xi\rangle$. Besides the lower correlations, it is also important to note that both the Spatial Gini Coefficient and the Moran's I index are symmetric quantities, and as such they fail to capture the intrinsic asymmetry between income classes revealed by Fig. 7a–d.

We can extend the framework developed in Fig. 4 to the case of multiple classes or, in the present case, income categories. To do so we average the profiles $\widetilde{\tau}_{\alpha\beta}$ to obtain the following quantities:

$$\widetilde{\tau}_{\alpha\to O} = \frac{\sum_{\beta\neq\alpha}\widetilde{\tau}_{\alpha\beta}}{N_i - 1},$$
$$\widetilde{\tau}_{O\to\alpha} = \frac{\sum_{\beta\neq\alpha}\widetilde{\tau}_{\beta\alpha}}{N_i - 1},$$

that are, respectively, the average CMFPT from class α to all the other classes and from all the other classes to class α. Here N_i corresponds to the total number of income branches in a city. For a given city, we can put each class α in the $(\widetilde{\tau}_{\alpha\to O}, \widetilde{\tau}_{\alpha\to O}/\widetilde{\tau}_{O\to\alpha})$ phase space, as shown in Fig. 8a for each of the income classes in Detroit and Boston (here $\widetilde{\tau}_{\alpha\to\beta}$ is calculated on the adjacency graphs). As noted previously, Boston exhibits lower levels of segregation than Detroit, and indeed most of the classes in Boston

(in shades of blue) are clustered around the point $(1, 1)$, and are associated with lower values of $\widetilde{\tau}_{\alpha \to O}$ that in Detroit (shades of green). The classes in Detroit, instead, display larger values of $\widetilde{\tau}_{\alpha \to O}$ and a larger variability along the y-axis. Still, both cities show a similar qualitative behaviour: $\widetilde{\tau}_{\alpha \to O}/\widetilde{\tau}_{O \to \alpha}$ is smaller than 1 for low-income classes, increasing above one for intermediate income classes, and decreases again for higher income classes. Such a transition could be explained by the fact that middle-income classes are in general more homogeneously distributed across a city, while low-income and high-income ones tend to form isolated clusters. If middle-income classes are fairly distributed across the city, they are easy to reach from the rest of classes but for them reaching low- and high-income classes is much harder, which leads to $\widetilde{\tau}_{\alpha \to O} > \widetilde{\tau}_{O \to \alpha}$. To validate this hypothesis, we show in Fig. 8b–g the probability of finding a given class α in a spatial unit p_α, respectively for Class 1 (b–e), 8 (c–f) and 16 (d–g) in Boston (b–d) and Detroit (e–g). Indeed, Class 1 (extremely low income) seems to be clustered around the centre of both cities, and especially so in Detroit, while Class 16 is mostly concentrated in the peripheries. Notice that Class 8 (median income levels) does not display any clearly visible isolation. We provide the location of the cities studied in the $(\widetilde{\tau}_{\alpha \to O}, \widetilde{\tau}_{\alpha \to O}/\widetilde{\tau}_{O \to \alpha})$ for several income classes in Supplementary Fig. 6.

Overall, our approach to spatial segregation based on the diffusion of random walks is not only a natural extension of the latter multi-scalar approaches introduced to characterise residential segregation[68], but it also allows us to define a dynamical segregation that includes mobility into the analysis as it has been recently discussed for instance in refs. [64,66].

Conclusions

We have shown here that the information captured by the distribution of inter-class MFPTs can be used not just as a way to detect the presence of anisotropy and correlations in the properties of nodes, but also as a reliable proxy for the dynamics and emergent behaviours of a complex system. One of the most interesting aspects of the measures of heterogeneity, polarisation and segregation that we have introduced in this work is that they take into account microscopic, meso-scopic and global relations among classes, due to the fact that in principle random walks integrate information about paths of all possible lengths. Another relevant property of the measures of segregation based on CMFPT is that they are non-parametric and correctly normalised with respect to a meaningful null-model, hence allowing us to compare on equal grounds the heterogeneity of class distributions in systems of different sizes, which is where most of the classical indices of segregation fail[56,68]. Even more importantly, the profiles of inter-class MFPTs are not symmetric with respect to classes, and provide fine-grained information about which classes are most responsible for the emergence of polarisation and heterogeneity. In this respect, it would be worthy exploring how the simple measure of polarisation that we proposed can be extended to the case of more than two classes.

The fact that measures of class heterogeneity based on random walk statistics correlate quite well with some of the intrinsic dynamics happening in social networks (i.e., the spread of an epidemic) and with some other exogenous processes mediated by the underlying graph (i.e., the incidence of crime in a city) confirms that the profiles of CMFPT are a useful toolbox for targeted mitigation of the undesired effects of these dynamics. For instance, the groups of a social network that are more central according to $\widetilde{\tau}_{\alpha \beta}$ might be the best candidates for early vaccinations aiming at slowing-down an epidemic. At the same time, the definition of dynamical segregation based on walks on the

mobility graphs, and the fact that it correlates quite substantially with crimes, potentially paves the way for a re-definition of the traditional role attributed to residential segregation, in favour of a more balanced view that takes into account the activity patterns of citizens together with the spatial distribution of their dwellings.

The generality of the methodology proposed in this paper and its applicability to different classical problems in complexity science establish a concrete link between classical statistical physics and modern complexity science, and have the potential to provide new interesting insights about the relation between structure and dynamics of complex systems.

Methods

Mean-field approximation. The expressions for CMFPT and class return time provided in Eqs. (2) and (5) are exact, but they have the drawback of being computationally intensive for graphs with a large number of nodes. It is possible to construct C-class mean-field approximations of these expressions, by representing the behaviour of all the nodes of a class with a single node, and looking at the graph of node classes. If we denote by $\pi_{\alpha\beta}$ the probability for a walker to jump in one step from any node of class α to any node of class β, the general mean-field equation for CMFPT reads:

$$T_{\beta\alpha}^{\mathrm{MF}} = D_{\overline{\alpha}} + \sum_{\gamma \neq \alpha} \pi_{\beta\gamma} T_{\gamma\alpha}$$

which can be written in compact form as follows:

$$T_\alpha^{\mathrm{MF}} = D_{\overline{\alpha}} + \Pi_{\overline{\alpha}} T_\alpha^{\mathrm{MF}} \tag{15}$$

where by definition $\{T_\alpha^{\mathrm{MF}}\}_\alpha = 0$ and $\Pi_{\overline{\alpha}}$ is the transition matrix where the row and column corresponding to class α are set equal to zero as above. Notice that $\pi_{\alpha\beta}$ is the total fraction of edges from nodes of class α to nodes of class β, $\pi_{\alpha\alpha}$, defined as follows:

$$\pi_{\alpha\beta} = \begin{cases} \frac{e_{\alpha\beta}}{\sum_\beta \pi_{\alpha\beta}} & \alpha \neq \beta \\ \frac{2e_{\alpha\beta}}{\sum_\beta \pi_{\alpha\beta}} & \alpha = \beta \end{cases} \tag{16}$$

where $e_{\alpha\beta}$ is the total number of edges from nodes of class α to nodes of class β. Solving Eq. (15) for T_α^{MF} we obtain:

$$T_\alpha^{\mathrm{MF}} = \left[I - \Pi_{\overline{\alpha}}\right]^{-1} D_{\overline{\alpha}}$$

Notice that this equation is formally identical to Eq. (2), with the only difference that it deals with MFPTs from all other *classes* to class α, while Eq. (2) provides the MFPTs from all the nodes in other classes to class α.

Similarly, the return times in the C-class mean-field approximation can be computed staring from Eq. (4) for each of the classes, obtaining:

$$R_\alpha = 1 + \sum_{\beta \neq \alpha} \pi_{\alpha\beta} T_{\beta\alpha} \tag{17}$$

Normalisation of CMFPT distributions. As we will see in the following sections, the statistics of CMFPT depend substantially on the shape, size, and organisation of class assignments. To allow a fair comparison of CMFPT in systems with different sizes and shapes, we compute the normalised inter-class MFPT $\widetilde{\tau}_{\alpha\beta} = \tau_{\alpha\beta}/\tau_{\alpha\beta}^{\mathrm{null}}$. This quantity is the ratio between the expected number of time steps needed to a walk that starts on a node of class α to reach a node of class β for the first time, and the corresponding value in a null-model where classes are assigned to nodes uniformly at random, by preserving their relative abundance. Notice that if classes are distributed uniformly across the system, then $\widetilde{\tau}_{\alpha\beta} \approx 1, \forall \alpha, \beta$, while in the presence of spatial correlations $\widetilde{\tau}_{\alpha\beta}$ will deviate from 1. By using $\widetilde{\tau}_{\alpha\beta}$ we properly take into account several common confounding factors, including an uneven abundance of colours, differences in size and shape and the effect of borders, thus making it possible to compare different systems on common grounds. Depending on the size of the system, the computation of $\tau_{\alpha\beta}^{\mathrm{null}}$ is based on the average over 10^3–10^5 independent assignments of classes to the nodes of the original graph.

Computation of CMFPT distributions in real-world networks. In the following we will compute and use the distribution of CMFPT in a variety of synthetic and real-world networks. If a system is naturally represented by an unweighted network, the elements of the transition matrix of the random walker will be set to $\pi_{ij} = \frac{a_{ij}}{k_i}$, where $a_{ij} = 1$ is an edge from node i to node j, and $k_i = \sum_j a_{ij}$ is the (out) degree of node i. If the network is weighted, instead, the elements of the transition matrix will be $\pi_{ij} = \frac{w_{ij}}{s_i}$, where w_{ij} is the weight of the edge connecting node i to node j, and $s_i = \sum_j w_{ij}$ is the (out)strength of node i. For networks with up to $N \sim 10^4$ nodes, we solved Eqs. (2) and (5) exactly, using standard linear algebra packages. For larger graphs we reverted instead to Monte-Carlo simulations, where

we estimated the value of $T_{i,\alpha}$ for each node i of the graph as the average over 10^4–10^6 random walks originating at that node, and then obtained $\tau_{\alpha\beta}$ using Eq. (3).

Data availability

Code availability

References

1. Vanhems, P. et al. Estimating potential infection transmission routes in hospital wards using wearable proximity sensors. *PloS One* **8**, e73970 (2013).
2. Fernández-Gracia, J., Suchecki, K., Ramasco, J. J., San Miguel, M. & Eguíluz, V. M. Is the voter model a model for voters? *Phys. Rev. Lett.* **112**, 158701 (2014).
3. Jargowsky, P. A. Take the money and run: economic segregation in US metropolitan areas. *Am. Sociol. Rev.* **61**, 984–998 (1996).
4. Yan, G. et al. Network control principles predict neuron function in the *Caenorhabditis elegans* connectome. *Nature* **550**, 519–523 (2017).
5. Barthelemy, M. *The Structure and Dynamics of Cities* (Cambridge University Press, Cambridge, UK, 2016).
6. Batty, M. *The New Science of Cities* (MIT Press, Cambridge, MA, 2017).
7. Bullmore, E. & Sporns, O. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat Rev Neurosci* **10**, 186–198 (2009).
8. Bassett, D. S. & Sporns, O. Network neuroscience. *Nat. Neurosci.* **20**, 353–364 (2017).
9. Noh, J. D. & Rieger, H. Random walks on complex networks. *Phys. Rev. Lett.* **92**, 118701 (2004).
10. Zhang, Z., Julaiti, A., Hou, B., Zhang, H. & Chen, G. Mean first-passage time for random walks on undirected networks. *Eur. Phys. J. B* **84**, 691–697 (2011).
11. Hwang, S., Lee, D.-S. & Kahng, B. First passage time for random walks in heterogeneous networks. *Phys. Rev. Lett.* **109**, 088701 (2012).
12. Bonaventura, M., Nicosia, V. & Latora, V. Characteristic times of biased random walks on complex networks. *Phys. Rev. E* **89**, 012803 (2014).
13. De Domenico, M., Solé-Ribalta, A., Gómez, S. & Arenas, A. Navigability of interconnected networks under random failures. *Proc. Natl Acad. Sci. USA* **111**, 8351–8356 (2014).
14. Bassolas, A., Gallotti, R., Lamanna, F., Lenormand, M. & Ramasco, J. J. Scaling in the recovery of urban transportation systems from massive events. *Sci. Rep.* **10**, 1–13 (2020).
15. Codling, E. A., Plank, M. J. & Benhamou, S. Random walk models in biology. *J. R. Soc. Interface* **5**, 813–834 (2008).
16. Nicosia, V., Skardal, P. S., Arenas, A. & Latora, V. Collective phenomena emerging from the interactions between dynamical processes in multiplex networks. *Phys. Rev. Lett.* **118**, 138302 (2017).
17. Bacry, E., Delour, J. & Muzy, J.-F. Modelling financial time series using multifractal random walks. *Physica A Stat. Mech. Appl.* **299**, 84–92 (2001).
18. Masuda, N., Porter, M. A. & Lambiotte, R. Random walks and diffusion on networks. *Phys. Rep.* **716-717**, 1 – 58 (2017).
19. Zhang, Z., Shan, T. & Chen, G. Random walks on weighted networks. *Phys. Rev. E* **87**, 012112 (2013).
20. Pons, P. & Latapy, M. Computing communities in large networks using random walks. *Int. Symposium Computer Inf. Sci.*, 284–293 (2005).
21. Rosvall, M. & Bergstrom, C. T. Maps of random walks on complex networks reveal community structure. *Proc. Natl Acad. Sci. USA* **105**, 1118–1123 (2008).
22. Lambiotte, R., Delvenne, J.-C. & Barahona, M. Random walks, Markov processes and the multiscale modular organization of complex networks. *IEEE Trans. Netw. Sci. Eng.* **1**, 76–90 (2014).
23. Newman, M. E. A measure of betweenness centrality based on random walks. *Soc. Networks* **27**, 39–54 (2005).
24. De Domenico, M., Lancichinetti, A., Arenas, A. & Rosvall, M. Identifying modular flows on multilayer networks reveals highly overlapping organization in interconnected systems. *Phys. Rev. X* **5**, 011027 (2015).
25. Hoffmann, T., Porter, M. A. & Lambiotte, R. Random walks on stochastic temporal networks. In *Understanding Complex Systems*, pp. 295–313 (Springer, Berlin/Heidelberg, Germany, 2013).
26. Gómez-Gardeñes, J. & Latora, V. Entropy rate of diffusion processes on complex networks. *Phys. Rev. E* **78**, 065102 (2008).
27. Burda, Z., Duda, J., Luck, J. M. & Waclaw, B. Localization of the maximal entropy random walk. *Phys. Rev. Lett.* **102**, 160602 (2009).
28. Sinatra, R., Gómez-Gardeñes, J., Lambiotte, R., Nicosia, V. & Latora, V. Maximal-entropy random walks in complex networks with limited information. *Phys. Rev. E* **83**, 030103 (2011).
29. Redner, S. *A Guide to First-passage Processes* (Cambridge University Press, 2001).
30. Condamin, S., Bénichou, O. & Moreau, M. First-passage times for random walks in bounded domains. *Phys. Revi. Lett.* **95**, 260601 (2005).
31. Condamin, S., Tejedor, V., Voituriez, R., Bénichou, O. & Klafter, J. Probing microscopic origins of confined subdiffusion by first-passage observables. *Proc. Natl Acad. Sci. USA* **105**, 5675–5680 (2008).
32. Fronczak, A. & Fronczak, P. Biased random walks in complex networks: the role of local navigation rules. *Phys. Rev. E* **80**, 016107 (2009).
33. Nicosia, V., Domenico, M. D. & Latora, V. Characteristic exponents of complex networks. *Europhys. Lett.* **106**, 58005 (2014).
34. Bassolas, A., Sousa, S. & Nicosia, V. Diffusion segregation and the disproportionate incidence of covid-19 in African American communities. *J. R Soc. Interface* **18**, 20200961 (2021).
35. Sousa, S. & Nicosia, V. Quantifying ethnic segregation in cities through random walks. *arXiv*. Preprint at http://arxiv.org/abs/2010.10462 (2020).
36. Kuncheva, Z. & Montana, G. Community detection in multiplex networks using locally adaptive random walks. In *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*, 1308–1315 (2015).
37. Marcon, E. & Puech, F. Evaluating the geographic concentration of industries using distance-based methods. *J. Econ. Geogr.* **3**, 409–428 (2003).
38. Marcon, E. & Puech, F. Measures of the geographic concentration of industries: improving distance-based methods. *J. Econ. Geogr.* **10**, 745–762 (2010).
39. Braha, D. & De Aguiar, M. A. Voting contagion: modeling and analysis of a century of us presidential elections. *PloS One* **12**, e0177970 (2017).
40. Stolz, B., Harrington, H. & Porter, M. A. The topological 'shape' of Brexit. Available at *SSRN 2843662* (2016).
41. Dlotko, P., Rudkin, S. & Qiu, W. An economic topology of the Brexit vote. *arXiv*. Preprint at http://arxiv.org/abs/1909.03490 (2019).
42. Waugh, A. S., Pei, L., Fowler, J. H., Mucha, P. J. & Porter, M. A. Party polarization in congress: a network science approach. *arXiv*. Preprint at http://arxiv.org/abs/0907.3509v3 (2011).
43. Hirano, S., Snyder, J. M. Jr, Ansolabehere, S. D. & Hansen, J. M. Primary elections and partisan polarization in the US Congress. *Q. J. Polit. Sci.* **5**, 169–91 (2010).
44. Neal, Z. P. A sign of the times? Weak and strong polarization in the US Congress, 1973–2016. *Soc. Networks* **60**, 103–112 (2020).
45. Guerra, P. H. C., Meira, W. Jr, Cardie, C. & Kleinberg, R. A measure of polarization on social media networks based on community boundaries. In *ICWSM* (2013).
46. Matakos, A., Terzi, E. & Tsaparas, P. Measuring and moderating opinion polarization in social networks. *Data Min. Knowl. Discov.* **31**, 1480–1505 (2017).
47. Faustino, J., Barbosa, H., Ribeiro, E. & Menezes, R. A data-driven network approach for characterization of political parties' ideology dynamics. *Appl. Netw. Sci.* **4**, 48 (2019).
48. Starnini, M., Machens, A., Cattuto, C., Barrat, A. & Pastor-Satorras, R. Immunization strategies for epidemic processes in time-varying contact networks. *J. Theor. Biol.* **337**, 89–100 (2013).
49. Barrat, A., Cattuto, C., Tozzi, A. E., Vanhems, P. & Voirin, N. Measuring contact patterns with wearable sensors: methods, data characteristics and applications to data-driven simulations of infectious diseases. *Clin. Microbiol. Infect.* **20**, 10–16 (2014).
50. Kiti, M. C. et al. Quantifying social contacts in a household setting of rural kenya using wearable proximity sensors. *EPJ Data Sci.* **5**, 1–21 (2016).
51. DATASETS SocioPatterns.org. http://www.sociopatterns.org/datasets/ (2016). Accessed 30 September 2020.
52. Génois, M. et al. Data on face-to-face contacts in an office building suggest a low-cost vaccination strategy based on community linkers. *Netw. Sci.* **3**, 326–347 (2015).
53. G'enois, M. & Barrat, A. Can co-location be used as a proxy for face-to-face contacts? *EPJ Data Sci.* **7**, 11 (2018).
54. Barter, E. & Gross, T. Manifold cities: social variables of urban areas in the UK. *Proc. R. Soc. A* **475**, 20180615 (2019).
55. Winship, C. A revaluation of indexes of residential segregation. *Soc. Forces* **55**, 1058–1066 (1977).

56. Reardon, S. F. & O'Sullivan, D. Measures of spatial segregation. *Sociol. Methodol.* **34**, 121–162 (2004).
57. Ballester, C. & Vorsatz, M. Random walk-based segregation measures. *Rev. Econ. Stat.* **96**, 383–401 (2014).
58. Louf, R. & Barthelemy, M. Patterns of residential segregation. *PloS One* **11**, e0157476 (2016).
59. Manson, S., Schroeder, J., Riper, D. V. & Ruggles, S. IPUMS National Historical Geographic Information System: Version 14.0 [Database]. Minneapolis, MN: IPUMS. 2019. https://doi.org/10.18128/D050.V14.0
60. Longitudinal Employer-Household Dynamics. https://lehd.ces.census.gov/ (2016). Accessed 30 May 2020.
61. Logan, J. R. The persistence of segregation in the 21st century metropolis. *City Community* **12**, 160–168 (2013).
62. The Rise of Residential Segregation by Income|Pew Research Center. https://www.pewsocialtrends.org/2012/08/01/the-rise-of-residential-segregation-by-income/ (2016). Accessed 30 May 2020.
63. Waitzman, N. J. & Smith, K. R. Separate but lethal: the effects of economic segregation on mortality in metropolitan America. *Milbank Q.* **76**, 341–373 (1998).
64. Le Roux, G., Vallée, J. & Commenges, H. Social segregation around the clock in the Paris region (France). *J. Trans. Geogr.* **59**, 134–145 (2017).
65. Petrović, A., van Ham, M. & Manley, D. Multiscale measures of population: within-and between-city variation in exposure to the sociospatial context. *Ann. Am. Assoc. Geogr.* **108**, 1057–1074 (2018).
66. Randon-Furling, J., Olteanu, M. & Lucquiaud, A. From urban segregation to spatial structure detection. *Environ. Plan. B Urban Anal. City Sci.* **47**, 645–661 (2020).
67. Uniform Crime Reporting (UCR) Program – FBI. https://ucr.fbi.gov/crime-in-the-u.s/ (2016). Accessed 30 September 2020.
68. Olteanu, M., Randon-Furling, J. & Clark, W. A. Segregation through the multiscalar lens. *Proc. Natl Acad. Sci. USA* **116**, 12250–12254 (2019).
69. Data—Voteview. https://voteview.com/data (2016). Accessed 30 September 2020.

Acknowledgements

Author contributions

A.B. and V.N. devised the study. A.B. performed the simulations and computations. A.B. and V.N. provided methods, analysed the results, prepared the figures and the visual material, wrote the paper and approved the final submitted version.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42005-021-00580-w.

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

Reprints and permission information is available at http://www.nature.com/reprints

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