Multiscale dynamical embeddings of complex networks

Michael T. Schaub,[1, 2, a)] Jean-Charles Delvenne,[3, 4] Renaud Lambiotte,[5] and Mauricio Barahona[6, b)]

[1)] *Institute for Data, Systems and Society, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

[2)] *Department of Engineering Science, University of Oxford, Oxford, UK*

[3)] *ICTEAM, Université catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium*

[4)] *CORE, Université catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium*

[5)] *Mathematical Institute, University of Oxford, Oxford, UK*

[6)] *Department of Mathematics, Imperial College London, London SW7 2AZ, UK*

Complex systems and relational data are often abstracted as dynamical processes on networks. To understand, predict and control their behavior a crucial step is to extract reduced descriptions of such networks. Inspired by notions from Control Theory, here we propose a time-dependent dynamical similarity measure between nodes, which quantifies the effect that a node input has on the network over time. This dynamical similarity induces an embedding that can be employed for several analysis tasks. Here we focus on (i) dimensionality reduction, by projecting nodes onto a low dimensional space capturing dynamic similarity at different time scales, and (ii) how to exploit our embeddings to uncover functional modules. We exemplify our ideas through case studies focusing on directed networks without strong connectivity, and signed networks. We further highlight how several ideas from community detection can be generalized in terms of our embedding perspective and linked to ideas from Control Theory.

Complex systems comprising a large number of interacting dynamical elements commonly display a rich repertoire of behaviors across different time and length scales. Viewed as collections of coupled dynamical entities, the dynamical trajectories of such systems reflect how the topology of the underlying graph constrains and moulds the local dynamics. Even for networks without an intrinsically defined dynamics, such as networks derived from relational data, a dynamics is often associated to the network data to serve as a proxy for a process of functional interest, e.g., in the form of a diffusion process. Comprehending how the network connectivity influences a dynamics is thus a task arising across many different scientific domains[1–3].

However, it is often impractical to keep a full description of a dynamics and the network for system analysis. In many cases it may be unclear how such an exhaustive description could be interpreted, or whether such finely detailed data is necessary to understand the phenomena of interest. Accordingly, many studies aim to reduce the complexity of the system by extracting lower dimensional or coarse-grained descriptions, which explain the behavior of interest in a simpler manner with fewer, aggregated variables.

This reductionist paradigm may be illustrated with the process of opinion formation in a social network. In general, there will be many actors in the network, organized in different social circles and influenced by various agents, media, etc. While the full dynamics is highly complex and variable, the globally emerging dynamics may still evolve on an effective subspace of low dimensionality, such that a coarse-grained description at the aggregated level of social circles may be sufficient to describe the process.

A classical source for dimensionality reduction is the presence of symmetries in the system[4,5], or the presence of homogeneously connected blocks of nodes. Yet, *strict, global* symmetries are rare in complex systems describing real applications, and while a statistical approach can be used to interpret the irregularities as random fluctuations from an ideal model, e.g., stochastic blockmodels[6,7], such models often posit strong locality assumptions such as i.i.d. edges. In particular, many global features, such as cyclic structures and higher order dynamical couplings, cannot be captured within such a block structure paradigm[8,9].

Embedding techniques, which define an (often low-dimensional) representation of the network and its nodes in a metric vector-space, have thus gained prominence recently for analyzing network data[10,11], as they allow us to use a plethora of computational techniques that have been developed for analysing data in such domains. Thus far most of these techniques have focussed squarely on representing mainly topological information (structural symmetries or communities), e.g., using a geometry induced by diffusion processes. However, networks often come equipped with a more general dynamics than diffusion, or contain signed and directed edges, for which it is not clear how to define an appropriate diffusion.

Inspired by notions from control theory, here we propose a dynamical embedding of networks that can account for such cases. Our embedding associates to each node the trajectory of its (zero-state) impulse response. As shown in Figure 1 we thereby can construct, for each time t, a representation of the nodes in signal (vector)-space, which provides us with a dynamics-based, geometric representation of the system, and associated sim-

---

[a)] Electronic mail: mschaub@mit.edu

[b)] Electronic mail: m.barahona@imperial.ac.uk

ilarity and distance measures. Nodes that are close in this embedding induce a similar state in the network at a particular time scale t following the application of an impulse.

We can exploit this vector space representation and the associated similarity and dual distance measures for various analysis tasks. While such representations are amenable for general learning tasks, in the following we focus on two examples that highlight particular features of this approach of interest for dynamical network analysis. First, we illustrate how *low-dimensional* embeddings of the system can be constructed — providing a dimensionality reduction of the system in continuous space. Second, we illustrate how these ideas can be exploited to uncover dynamical modules in the system, i.e., groups of nodes that act approximately as a *dynamical unit over a given time scale*, and discuss how these modules can be related to notions from Control Theory — an important topic that has gained prominence recently in network theory. Surprisingly, we find that when viewed under the prism of embeddings, we can find striking similarities between ideas from model order reduction and control theory on the one hand, and notions from network analysis and low-dimensional embeddings on the other hand.

RESULTS

Dynamical embeddings of networks and node distance metrics

To fix ideas, let us envision our system in the form of a set of coupled differential (or difference) equations with n variables. Each variable is identified with a node of a graph, which is defined by the coupling interactions, and we assess whether two nodes play a similar dynamical role. Let us inject an impulse at node i at time $t = 0$ and observe the response of the system $\mathbf{y}_i(t) \in \mathbb{R}^n$. We can then define a mapping $i \mapsto \mathbf{y}_i(t)$, which associates to each node its zero-state impulse response. This mapping embeds the nodes into the space of all signals, and we may use any suitable similarity measure between the signals $\mathbf{y}_i(t)$ and $\mathbf{y}_j(t)$ to define a *node similarity*.

To quantify whether the impact of node i in the network is aligned with the impact of node j at a particular time t, a wide variety of similarity functions between $\mathbf{y}_i(t)$ are possible, including nonlinear kernels[12]. However, we find it convenient to use the standard bilinear inner product:

$$\Psi(t) = \big[\psi_{ij}(t)\big]_{i,j=1,\ldots,n} \qquad (1)$$
$$\text{with} \quad \psi_{ij}(t) = \langle \mathbf{y}_i(t), \mathbf{y}_j(t)\rangle_{\mathcal{W}} = \mathbf{y}_i(t)^\top \mathcal{W}\, \mathbf{y}_j(t).$$

where any positive (semi-)definite weighting matrix $\mathcal{W}$ adapted to the geometry of the state space may be used to our advantage. For instance, we may choose $\mathcal{W}$ to correspond to a degree weighting, or even to project out

certain components of $\mathbf{y}$, e.g., by projecting out the average of $\mathbf{y}$ (thus $\mathcal{W}$ can be used akin to a 'null model' term; see Methods and SI). The above defined similarity using inner products is rich enough to capture the overall behavior of many systems, and moreover allows us to draw parallels to a variety of approaches in the literature. In particular, for diffusion processes we can recover structural concepts such as the modularity matrix and generalisations, which emerge as specific cases (see Methods and SI). Nevertheless, other measures of similarity could be considered such as different correlations, or even information theoretic measures (in particular, if the state vectors can be interpreted in terms of a probability distribution).

Figure 1 illustrates some of the key aspects of this dynamical viewpoint. Notably, a high dynamical similarity does not necessitate direct proximity in the underlying graph, in contrast to most methods used to detect graph communities based on diffusion[13–16] or on the propagation of a perturbation[17,18]. The key difference is that Ψ does not signal the presence of regions in which the flow is trapped; instead, the similarity between two nodes i, j is defined by how aligned the influence of an impulse emanating from nodes i, j is after a time t. This enables us to treat various kinds of connectivities under the same framework, as seen in Figure 1, where cyclic and bipartite structures are identified in a naturally interpretable manner over particular time scales. It is worth emphasizing here once more, however, that we do not seek a reduced description of the network topology alone, but of the conjunction of the network and the dynamics defined on top of the network structure (see also Methods).

Formal definition of dynamical similarity and distance measures

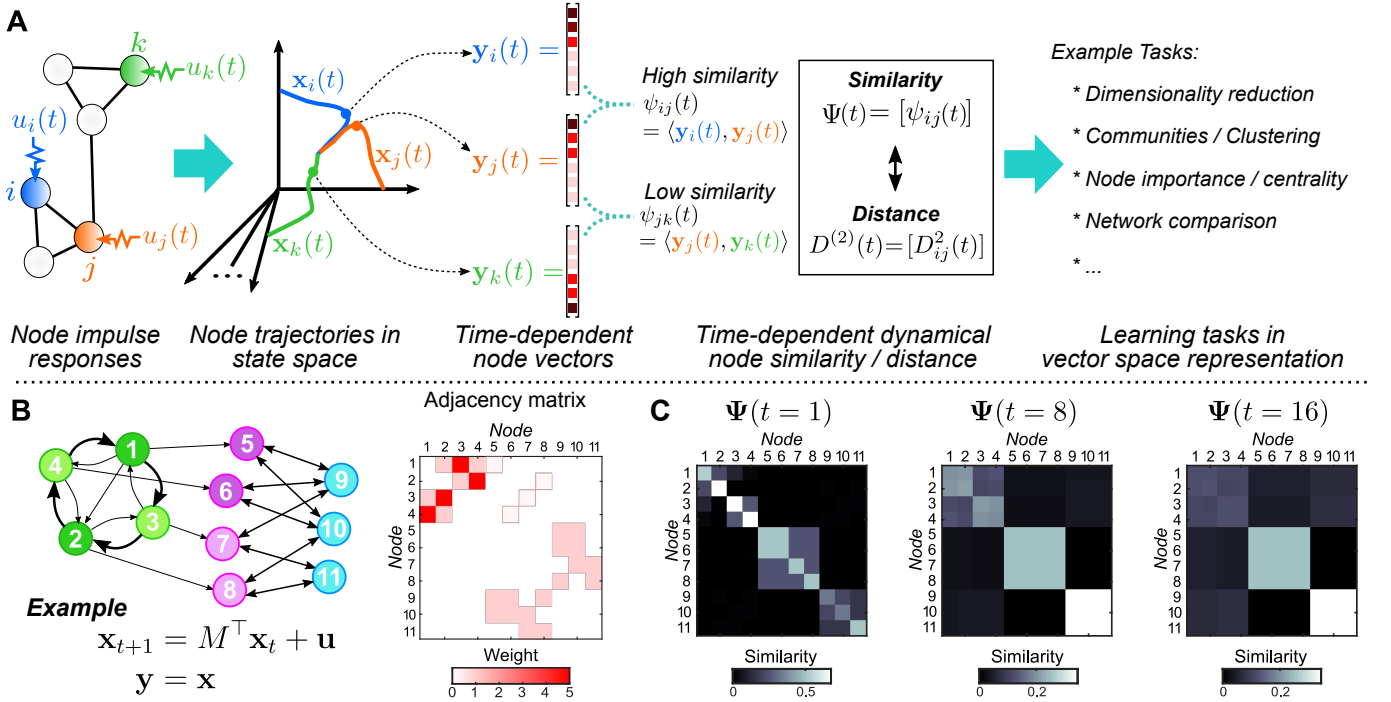More specifically, let us consider the general linear dynamics:

$$\dot{\mathbf{x}} = \mathcal{A}\mathbf{x} + \mathcal{B}\mathbf{u} \quad \mathcal{A} \in \mathbb{R}^{m \times m}, \quad \mathcal{B} \in \mathbb{R}^{m \times p} \qquad (2a)$$
$$\mathbf{y} = \mathcal{C}\mathbf{x} \qquad \mathcal{C} \in \mathbb{R}^{n \times m}, \qquad (2b)$$

where $\mathbf{x} \in \mathbb{R}^m, \mathbf{y} \in \mathbb{R}^n, \mathbf{u} \in \mathbb{R}^p$ are the state, the observed state, and the input vectors, respectively. Discrete-time versions of the same form are also of interest (Fig. 1). Assume a set of inputs given by node impulses at time $t = 0$. Note that a cornerstone of the theory of linear systems is the fact that the impulse response of a system fully characterizes the system (i.e. its transfer function). Indeed, we could obtain the complete frequency response of the system based on its impulse response, which is intrinsic to the system.

We now collect the (zero state) impulse responses $\mathbf{y}_i$ into the matrix $Y(t) = [\mathbf{y}_1, \ldots, \mathbf{y}_n] = \mathcal{C}\exp(\mathcal{A}t)\mathcal{B}$, leading to the inner product matrix:

$$\Psi(t) = Y^\top \mathcal{W} Y = \mathcal{B}^\top \exp\left(\mathcal{A}t\right)^\top \mathcal{C}^\top \mathcal{W}\, \mathcal{C} \exp(\mathcal{A}t)\mathcal{B}. \quad (3)$$

Figure 1. **Constructing dynamical similarity measures. A** Schematic: Impulses are applied as inputs to different nodes of the network. The responses in time are interpreted as node vectors evolving in state space, and can be compared, e.g., via an inner product from which we construct the similarity matrix $\Psi(t)$, or alternatively, its associated distance D^2. Nodes that drive the system similarly (differently) within the projected subspace are assigned a high (low) similarity score. The thus derived vector space representation of the nodes can be used for a number of different learning tasks. **B-C** Illustrative example for the construction of our similarity measures, focusing on the measure Ψ (the construction of D is analogous). **B** Adjacency matrix and visualization of an asymmetric directed network (not strongly connected) which contains a bipartite (disassortative) substructure. We consider a discrete time random walk dynamics on this network $\mathbf{x}_{t+1} = M^\top \mathbf{x}_t$, where $M = K^{-1}A$ is the transition matrix of an unbiased random walker, and K is the diagonal out-degree matrix. **C** We construct $\Psi(t) = M^t[M^t]^\top$ for times $t = \{1, 8, 16\}$. As can been seen, the nodes 5 and 6 in the pink group behave similarly over short time-scales, whereas the other nodes behave more distinctly. Over intermediate time scales, the similarity of the nodes converges into four blocks: the pink group, the cyan group, and two subgroups within the green cycle subgraph (1–2, 3–4). At longer time scales the similarity of the nodes, may in this example be approximated by three dynamical blocks (green, pink, cyan) — hinting at the fact that our embeddings can be employed for the detection of dynamically cohesive modules.

Associated with this embedding, we define the associated Euclidean distance matrix $D(t)$, whose (squared) entries are:

$$D_{ij}^2(t) = \|\mathcal{W}^{1/2}\left(\mathbf{y}_i(t) - \mathbf{y}_j(t)\right)\|^2 = \psi_{ii} + \psi_{jj} - 2\psi_{ij}. \quad (4)$$

Note that $\Psi(t)$ and $D(t)$ are both time-dependent, allowing us to focus on different time scales of interest. For instance, we can ignore fast paced transients τ and consider only long-time behaviors for $t \geq \tau$, or consider instead quasi-steady behaviors that extend over particular time scales.

Hence, the time-parameter t may be understood is as a sampling of the network dynamics at a particular time-horizon, which can be chosen according to the functional behavior of interest. As a concrete example, consider again a diffusion dynamics in discrete time as shown in Figure 1B-C, where different time-scales provide different meaningful descriptions. Setting $t = 1$ amounts effectively to a structural analysis in which merely the direct coupling is considered; setting $t > 1$ amounts to integrating information over multi-step pathways[8,19]. As features should be persistent over a range of times, we may also integrate over t, thereby eliminating the time-dependence, and recovering another connection to Gramian matrices considered in control (see Methods). The similarity matrices Ψ and the associated distances D can be utilized in different ways for system analysis as we illustrate next.

Dimensionality reduction: low-dimensional node representations derived from dynamical distance D

Which universities are the most prestigious in North America? In a recent study, Clauset et al.[20] provided a data-driven assessment of this question by examining the hiring patterns US-universities for computer science (CS), history and business schools by means of a minimum violation ranking. This ranking aims to order universities such that the fewest number of directed links,
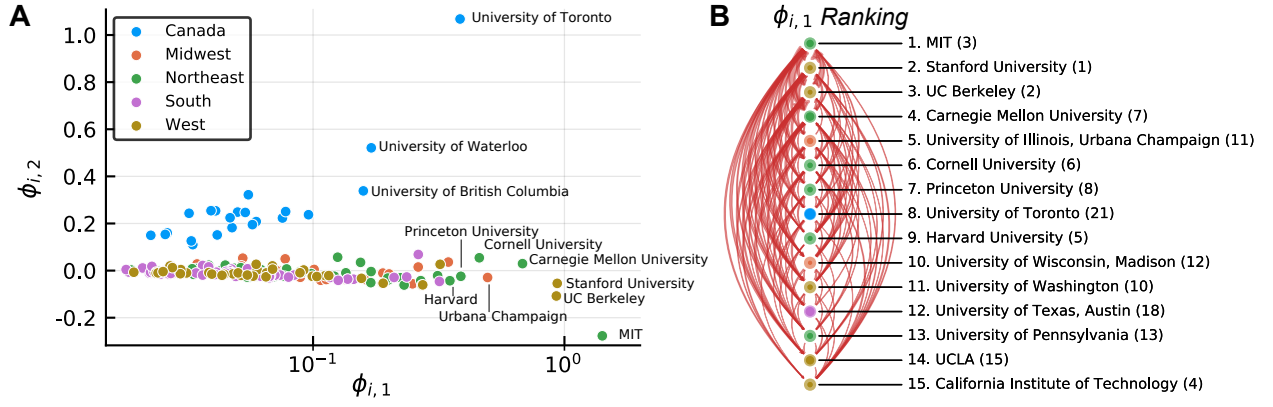
A

Canada
Midwest
Northeast
South
West

University of Toronto

University of Waterloo

University of British Columbia

Princeton University
Cornell University
Carnegie Mellon University
Stanford University
UC Berkeley
Harvard
Urbana Champaign
MIT

B $\phi_{i,1}$ *Ranking*

1. MIT (3)
2. Stanford University (1)
3. UC Berkeley (2)
4. Carnegie Mellon University (7)
5. University of Illinois, Urbana Champaign (11)
6. Cornell University (6)
7. Princeton University (8)
8. University of Toronto (21)
9. Harvard University (5)
10. University of Wisconsin, Madison (12)
11. University of Washington (10)
12. University of Texas, Austin (18)
13. University of Pennsylvania (13)
14. UCLA (15)
15. California Institute of Technology (4)

Figure 2. **Analysing academic influence using low-dimensional embeddings A** We develop a low-dimensional embeddings based on the influence dynamics in the hiring network, as described in the text. The first two dimensions of this embedding are plotted. The first coordinate $\phi_{i,1}$ is strongly correlated with the prestige ranking of Clauset et al.[20], highlighting the influential role played by the universities at the top. Interestingly, the second dimension distinguishes the Canadian universities from the US universities, showing that although these universities are well integrated within the faculty hiring market[20], they play a different role and exert a different type of influence on the system. **B** The ranking obtained when projecting onto the first coordinate only and the associated subgraph of faculty hirings. The numbers in parenthesis correspond to the rankings obtained by Clauset et al[20]. The arrows are proportional to the number of faculty moving between the institutions. Arrows pointing downwards in the ranking are plotted on the left, arrows point upward in terms of the ranking are plotted on the right. The number of hirings from inside the institution itself (self-loops) are indicated by size of the core (darker color) of each node. The area of the core is proportional to the number of self-loops compared to the total out-degree within this subnetwork.

corresponding to faculty hirings, move from lower-ranked to higher-ranked universities. Stated differently, universities with a higher prestige are assumed to act as sources of faculty for lower-ranked universities.

Here we reconsider this question from a more control-theoretic angle, using our above defined distance measure. Specifically, consider the adjacency matrix A of the graph of hiring patterns for CS, where A_{ij} denotes the number of faculty moving from university i to university j. This provides us with a directed, weighted network with 205 nodes corresponding to CS units at the departmental or school level, where we ignore movements of faculty to/from entities outside this set of 205 units. We now posit that a university i exerts an influence on another university j by sending faculty members to it. To normalize for the size of the universities we divide this influence by the in-degree of each university, i.e., an influence of size 1 may be exerted on each university.

This leads us to consider the influence dynamics $\dot{\mathbf{x}} = K_{\text{in}}^{-1} A^T \mathbf{x}$ among the universities, where $K_{\text{in}} = \text{diag}(A^\top \mathbf{1})$. We assume this influence is primarily mediated via direct influences; moreover, the system is unstable and the long-term behavior will be effectively determined by the most unstable mode. We will thus consider shorter time-scales in our analysis here. Note that one could also consider alternative dynamics here, e.g., the relaxation dynamics of the recently proposed 'spring rank' formalism (whose long-term behavior would then correspond to the spring rank)[21].

To avoid having to choose a particular time parameter, we integrate with respect to $t \in [0, 1]$, which ensures

that short lived features arising from the focus of a single time scale are integrated out. We remark that, interestingly, while the underlying network is not strongly connected, there is no need to introduce a teleportation into the dynamics as is commonly the case in diffusion based methods (see SI Appendix E).

To derive a low-dimensional embedding we approximate the resulting (squared) dynamical distance matrix $D^{(2)}_{[0,1]}$ by deriving a low rank approximation of the spectral decomposition of $\Psi_{[0,1]} = V \Lambda V^T$. To this end we define the 'dynamical fingerprints' $\phi_i^{[0,1]}$ as $[\phi_1^{[0,1]}, \ldots, \phi_n^{[0,1]}] = \Lambda^{1/2} V^T =: \Phi_{[0,1]}$. The vectors $\phi_i^{[0,1]}$ define a new coordinate system, whose coordinates are ranked according to their importance to the dynamics. We note that $[D^{(2)}_{[0,1]}]_{ij} = \|\phi_i^{[0,1]} - \phi_j^{[0,1]}\|^2$, and thus our dynamical distance can be approximated by truncating our coordinate system to the first few components of the vectors $\phi_i(t)$, thereby reducing the dimensionality of the system description in terms of D (see Methods).

Figure 2 shows the results of this procedure when applied to the CS dataset of Clauset et al.[20]. Interestingly we find that the first coordinates $\phi_{i,1}^{[0,1]}$ are strongly correlated with the previously obtained ranking[20] (Spearman rank-correlation $\rho \approx 0.91$), i.e., our influence dynamics based dimensionality reduction maintains the essential features of the identified prestige hierarchy. However, in addition our embedding reveals that the Canadian universities play a somewhat different role in the system. Indeed the second coordinate $\phi_{i,2}^{[0,1]}$ is singling out Canadian universities, highlighting that not all features of the

influence dynamics are captured well by a unidimensional ranking (see Figure 2). Note that when symmetrizing the network, the Spearman correlation of the first dimension with the minimum violation ranking of Clauset et al.[20] drops markedly to $\rho \approx 0.8$, emphasizing again that the directionality in this network is an essential feature.

Dynamical blocks: using Ψ to find functional modules in neuronal spiking networks.

Let us now consider the dynamics of a network of leaky-integrate-and-fire (LIF) neurons (see SI Appendix B). Due to their computational simplicity yet complex dynamics, networks of LIF neurons are widely used as scalable prototypes of neural activity. Recently, it has been shown that LIF networks can display "slow switching activity"[22,23], sustained in-group spiking that switches from group to group across the network. Importantly, the cell assemblies of coherently spiking neurons in this context can include both excitatory neurons and inhibitory neurons, and dense clusters of connections are not necessary to give rise to such dynamics (see Figure 3).

As has been shown previously, key insights into the nonlinear LIF dynamics can be obtained from linear rate models of the form[23]:

$$\dot{\mathbf{x}} = (-I + W_N)\mathbf{x} + \mathbf{u}, \qquad (5)$$

where $\mathbf{x}$ describes the n-dimensional firing rate vector relative to baseline; $\mathbf{u}$ is the input; and W_N is the asymmetric synaptic connectivity matrix containing excitatory (positive) and inhibitory (negative) connections between the neurons. The asymmetry of W_N follows from Dale's principle[24], which states that each neuron acts either completely inhibitory or completely excitatory on its efferent neighbours. Clearly, the rate dynamics (5) is of the form (2).

In this example we consider a LIF network, whose coupling matrix W_N is shown by the signed network in Figure 3B. The network structure can be described by a block-partition into 20 blocks: 10 groups of excitatory neurons, and 10 groups of inhibitory neurons, whose ordering is consistent with with the network drawing in Figure 3B. The connectivity patterns between these blocks are homogeneous in terms of the probability of observing a connection and their connection link-strengths. Stated differently, if we were to partition this coupling matrix into homogeneously connected blocks (both in terms of weights and number of connections), we would like to find these 20 blocks.

However, it turns out that this arrangement corresponds to a structure of only 10 planted cell assemblies, each consisting of a mixture of inhibitory and excitatory neurons. Thus, while structurally we may conclude from an inspection of W_N that there should be 20 groups, we know from our design that there only 10 dynamically relevant groupings, which moreover consist of a mixture of the original 20 blocks. As our similarity measure takes into account the dynamical description of the network as well, a natural question is thus whether we can find these modules that are aligned both with the structure and with the dynamics of the network: in other words, whether we can find the 10 cell assemblies.

In order to assess which dynamical role is played by the different neurons, we thus consider our dynamical similarity measure, this time however not with a focus on deriving an embedding, but with an eye towards identifying the planted functional groups in the (nonlinear) dynamics.

While similar results can be obtained by directly analysing Ψ (or D), as outlined in the Methods section, here we take the opportunity to highlight some interesting connections of our dynamical similarity measures to formulations of quality function more commonly employed in network analysis. As outlined above, cell assemblies are defined via a relative firing increase/decrease with respect to the population mean. Instead of using our full Ψ matrix, we may thus construct the following alternative centered similarity matrix by chosing weighting matrix of the form $\mathcal{W} = I - \mathbf{1}\mathbf{1}^\top/n$ (see also Methods):
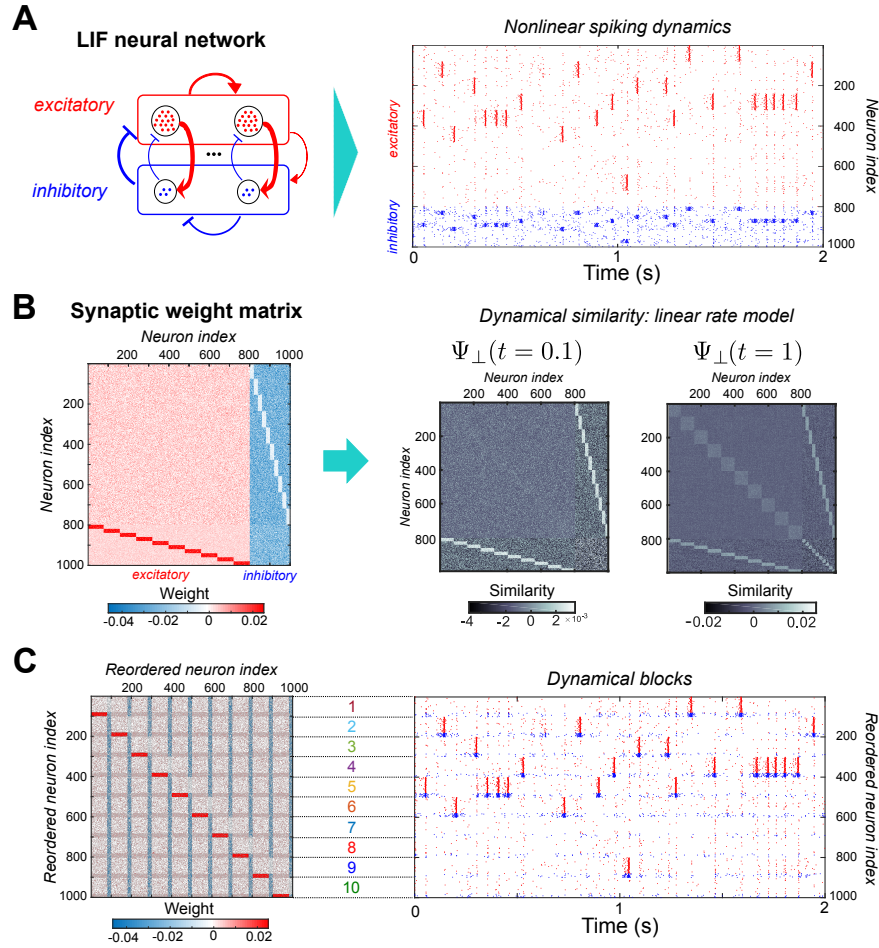
$$\Psi_\perp(t) = \exp(\mathcal{A}t)^\top \left(I - \frac{\mathbf{1}\mathbf{1}^\top}{n} \right) \exp(\mathcal{A}t). \qquad (6)$$

Note that, as we discuss in the supplementary material, this can be interpreted as a choice of a null model, or as introducing a relaxation on the distance matrix different to the low-rank approximation discussed in the previous section.

Based on this transformed similarity measure, we can try to identify dynamical modules by grouping nodes that are similar in a greedy way. A common technique to do so, would be by optimizing a quality function of the form $\text{trace}(H^\top \Psi_\perp H)$, where H denotes a partition indicator matrix (see (17) in Methods). Such an optimization can be done again spectrally or using a Louvain-like optimization (see Methods and SI Appendix C), a method which we will employ in the following. As can be seen in Figure 3, in this case, this procedure can reveal mixed groups of excitatory and inhibitory neurons that exhibit synchronized firing in the fully non-linear LIF network simulations. Again, the detected groups here do not correspond to tightly knit groups in the topology; rather they reflect dynamical similarity.

Note that by following this route we have made an approximation (transformation) of the original Ψ matrix. While here we could our physical intuition about the behavior of cell assemblies to propose an appropriate null model projection, in a general context, the results of such a procedure will of course be dependent on the specific choice made.

The point of our analysis here is however not to suggest a particular null model, or a generic optimization method. Rather, as we discuss in more detail in the supplementary information, too, our work suggests a framework with which one can construct a whole range of dynamical quality functions, each of which inspired by a

Figure 3. **Finding dynamical groups in a neural network description. A** Schematic of the connectivity of the leaky integrate and fire neuronal network, which shows its disassortative feedback structure between inhibitory and excitatory neurons. The exemplar raster plot illustrating its spiking dynamics shows that the system is characterized by slow switching between coherent spiking activity of 10 groups of neurons (each containing both inhibitory and excitatory units). **B** Left: The weighted, signed and directed synaptic connectivity matrix (W_N) of the network does not contain groups of nodes with high internal connection density. However, the analysis of the linear rate model governed by this connectivity matrix (5) using the dynamical similarity (6) in conjunction with a Louvain-type optimization reveals the presence of 10 dynamical modules. Right: Visualization of the centered similarity (6). For visualization purposes only the diagonal of $\Psi_\perp$ has been removed. Note that how after an initial short transient period the block structure into 10 groups becomes apparent, in comparison to the original weight matrix. **C** The blocks revealed from the linear rate model coincide with the dynamically co-activated groups of neurons in the full LIF dynamics, as shown by reordering the neuron indices. On the original weight matrix, they correspond however to a mixture of the blocks inside the weight-matrix W_N.

dynamics tailored to the system of interest. As in many cases we are not looking at a network in isolation, but are interested in some kind of process on it, this can in many instance lead to a more meaningful analysis. In addition the perspective developed here provides us a way in which we can give dynamical interpretations to certain choices of null models from a dynamical perspective (see supplementary information D). Interestingly, for the particular case of diffusion dynamics on undirected graphs, we can precisely recover some well-known methods for community detection, but the formalism developed here easily extends beyond this, opening the door for future research in this direction.

DISCUSSION

Building on ideas from systems and control theory, we have presented a framework that provides dynamical embeddings of complex networks, including signed, weighted and directed networks. These embeddings can be used in a variety of analysis tasks for network data. We have focused here on applications to dimensionality reduction as well as the detection of dynamical modules, as these choices enable us to highlight important features of our embedding framework.

Our dynamical similarity measures $\Psi(t)$ and $D(t)$ may however be used as a starting point for other techniques,

too. For instance, we could consider the (functional) networks induced by our dynamical similarity measures, and employ generative models[6,25,26] for their analysis. One way to approach this would be to define a (negative) Hamiltonian based on our similarity matrix, e.g., in a form similar to (17), and a Boltzmann distribution of the corresponding form. In this view the state-variables of the node (or other labels defined on the nodes) would correspond to latent variables that are coupled via the Hamiltonian.

One may further consider the extension to kernels computed directly from nonlinear dynamical systems, akin to the perturbation modularity recently introduced by Kolchinsky et al.[18], or consider linearisations around a particular state of interest. Alternatively, $\Psi(t)$ could be extended to represent nonlinear systems through an inner product in a higher dimensional space, e.g, by using the 'kernel trick'[12].

For simplicity, we assumed in the examples above the number of state variables equals the number of nodes in the network. Nevertheless, our derivations remain valid when there is more than one state variable per node. For instance, it may be applied to multiplex networks[27] or networks with temporal memory[28,29], that feature expanded state space descriptions and have gained considerable interest recently.

We remark that the measures presented here are different from correlation analysis of time-series data, as considered, e.g., by[30]. Instead of interpreting a correlation matrix as a functional network from n scalar valued time-series and then analysing this correlation matrix, we start with the joint description of a network and a dynamics. While the integrated measures $\Psi_{[0,t]}$ ($D_{[0,t]}^{(2)}$) presented here takes the form of an integrated (possibly transformed) inner-product which is at the core of linear correlation as well, we consider inner-products between the *vector-valued, deterministic* impulse responses of each node over time, rather than a scalar time-series per node. Note further, that while we have not explored this issue here one may vary the similarity measure employed for our analysis.

Conceptually, the similarity measure $\Psi(t)$ has strong theoretical links to model reduction and controllability, which provide meaningful interpretations of dynamic blocks in terms of coarse-grained representations. Classic model reduction[31–33] aims to find reduced models that approximate the input-output behavior of the system; yet the states of the reduced model do not usually have a sparse support in terms of the states of the original system. In contrast, the dynamical blocks found using Ψ are directly associated with particular sets of nodes and can thus be *localized* on the original graph, an important requirement for many applications. Future work will investigate alternative measures to $\Psi(t)$ based on the duality between controllability and observability Gramians from control, as well as measuring the quality of the dynamical blocks in a model reduction sense (see Methods).

Finally, our measure also provides links with other notions of similarity in networks including structural, equivalence, diffusion-based[34,35] and iterative node similarity in networks[36,37]. Such connections highlight the potential applications of our work in machine learning, where a good measure of similarity is central to solving problems such as link prediction[38] and node classification[39].

METHODS

Network embeddings based on dynamics

As discussed in the main text, we define our dynamical similarity measure as:

$$\Psi(t) = \mathcal{B}^{\top} \exp(\mathcal{A}t)^{\top} \mathcal{C}^{\top} \mathcal{W} \mathcal{C} \exp(\mathcal{A}t)\mathcal{B}. \tag{7}$$

Note that this can be written in the form:

$$\Psi(t) = \mathcal{B}^{\top} \Xi(t)\mathcal{B} \tag{8}$$

where $\Xi(t)$ is governed by the following Lyapunov matrix differential equation[40]:

$$\frac{d\,\Xi}{dt} = \mathcal{A}^{\top}\Xi + \Xi\mathcal{A}, \quad \text{with} \quad \Xi(0) = \mathcal{C}^{\top}\mathcal{W}\mathcal{C}. \tag{9}$$

It is well known that the same type of Lyapunov equation also governs the evolution of the covariance matrix of the system (2) driven by white Gaussian noise[40,41], which yields another interpretation of the above similarity measure.

Thus, $\Psi(t)$ is a dynamically evolving positive semidefinite Gram matrix, or a dynamic kernel matrix. This formulation allows us to relate $\Psi(t)$ to other kernel matrices developed for graphs, including the commute-time[39] kernel, or the heat kernel at the core of diffusion maps[42]. Particular relevant cases of Eq. (2) which have been explored in the literature are diffusive dynamics on undirected networks (e.g., where $-\mathcal{A} = L_{\mathrm{rw}}^{\top}$ is the random-walk Laplacian, and $\mathcal{B} = \mathcal{C} = I$, $\|x\|_1 = 1$), and consensus dynamics (where $-\mathcal{A} = L = L^{\top}$ is the combinatorial Laplacian), as will be discussed in the following sections. Instead of fixing a particular time t, we may further integrate over time and thus define the integrated similarity measure

$$\Psi_{[0,t]} := \int_0^t \Psi(t)dt \tag{10}$$

and, analogously to above, the associated elementwise squared distance $D_{[0,t]}^{(2)}$ matrix, which can be written in matrix form as:

$$D_{[0,t]}^{(2)} = \mathbf{1}\mathbf{z}^{\top} + \mathbf{z}\mathbf{1}^{\top} - 2\Psi_{[0,t]}$$

where $\mathbf{z} = \mathrm{diag}(\Psi_{[0,t]})$ is the column vector containing the diagonal entries of $\Psi_{[0,t]}$ (see also (4)). An advantage of this integration procedure is that only longer lived

features will contribute significantly to the this integral, and thus certain short-lived features are integrated out. This is in line with our viewpoint that meaningful dynamical roles should be robust, i.e., the similarities (and associated analysis) should be insensitive to perturbations of the time-parameter. If we are only interested in a particular time-scale t_0, which may well be a short time-scale, we may use this parameter choice and check in addition that the results are robust with respect to small perturbations δ in the interval $(t - \delta, t + \delta)$. We remark further that in the case of a diffusive dynamics $(\mathcal{A} = -L, \mathcal{C} = I - \mathbf{1}\mathbf{1}^\top/n)$, the integral variant (10) is related to the commute time or resistance distance, i.e., for $t \to \infty$, $[D_{[0,\infty]}]_{ij}$ is exactly the commute time distance[39] between nodes i and j.

Control-theoretic interpretations of $\Psi(t)$***.*** The integrated Gramian $\Psi_{[0,t]}$ in (10) may in fact be interpreted in terms of an observability / controllability Gramian matrix considered in Control Theory as we discuss next. Specifically, consider the Gram matrix based on the $\mathcal{L}_2$ inner product:

$$\langle \mathbf{f_i}, \mathbf{f_j} \rangle_{\mathcal{L}_2} = \int_0^t \mathbf{f_i}^\top \mathbf{f_j} \, dt$$

between the vector functions $\mathbf{f_i} : [0,t] \to \mathbb{R}^n$ defined via the mapping $\mathbf{f_i} : \ t \mapsto \mathcal{C} \, e^{\mathcal{A}t} \, \mathbf{e_i}$, where $\mathbf{e_i}$ is the i-th unit vector.

This is precisely the finite-time observability Gramian $G_O(t)$ of the linear system (2), which is defined as:

$$G_O(t) = \int_0^t e^{\mathcal{A}^\top t} \, \mathcal{C}^\top \mathcal{C} \, e^{\mathcal{A}t} \, dt. \tag{11}$$

$[G_O]_{ij}$ quantifies how inferable the initial state at node i is from output j. Hence, a high value of the entry $[G_O]_{ij}$ signifies that node j is highly observable from node i, when $C = I$. More precisely, each entry reflects how the energy of the initial states (localized on the nodes) spread to the outputs[43]. From the perspective developed above we may alternatively say that the classic observability Gramian (11) measures the similarity between two nodes in terms of their dynamical response over the interval $[0,t]$.

Instead, $\Psi(t)$ is the instantaneous Gramian corresponding to a particular t, i.e., $\Psi(t)$ may be understood as computing inner products between *sampled* zero-state impulse response trajectories $t \mapsto \mathcal{C} \, e^{\mathcal{A}t} \mathcal{B} \mathbf{e_i}$ at a particular time t. Indeed, our measure $\Psi(t)$ can be rewritten as

$$\Psi(t) = \mathcal{B}^\top \frac{dG_O(t)}{dt} \mathcal{B}.$$

As G_O has the interpretation of an energy, the entries of $\Psi(t)$ may thus be interpreted as a power transferred between the nodes.

It is well known that there exists a duality between the observability of a system and the controllability of the system governed by transposed matrices. We may

thus also view $\Psi(t)$ as assessing the instantaneous controllability of a dual system to (2) obtained by making the transformation $(\mathcal{A}, \mathcal{B}, \mathcal{C}) \to (\mathcal{A}^\top, \mathcal{C}^\top, \mathcal{B}^\top)$. In a similar vein, we can explore the dual controllability measure in that system. A more detailed investigation of these directions will be the object of future work.

Time-scale separation, low-rank structure and model reduction. Asymptotically, the dynamics of many networked systems converges to a lower dimensional manifold. Think, for instance, of synchronisation processes. In structured networks, however, one typically observes that the state transition matrix, and therefore the similarity matrix $\Psi(t)$, becomes numerically low-rank at much early times. Stated differently, in many structured networks we observe time scale separation linked to low dimensional subspaces of slowly decaying metastable states. Therefore, the system can be effectively described by a small set of slow modes that govern the dynamics over some time scale. A feature specific to networked systems is the fact that these slow modes can be *localized* on the space of nodes. It then follows, that instead of having to account for the whole system, we may just keep track of a few aggregated 'metanodes', whose state is governed by the slow modes, thereby reducing the complexity of the dynamics.

For a Laplacian diffusion dynamics $(\mathcal{A} = -L)$ this idea can be made more precise using so-called externally equitable partitions[44], which explicitly relate our similarity measure to model reduction. Consider an external equitable partition (EEP) characterized by the relation

$$LH_{\mathrm{EE}} = H_{\mathrm{EE}}\widehat{L}, \tag{12}$$

where H_{EE} is an indicator matrix encoding the EEP, and

$$\widehat{L} = (H_{\mathrm{EE}}^\top H_{\mathrm{EE}})^{-1} H_{\mathrm{EE}}^\top L H_{\mathrm{EE}} = H_{\mathrm{EE}}^+ L H_{\mathrm{EE}} \tag{13}$$

is the Laplacian of the quotient graph, the graph in which each group of the partition becomes a 'metanode'.

Now it can be shown that if we observe such a system through its projection onto this external equitable partition (i.e., we set $\mathcal{C} = H_{\mathrm{EE}}^+$), then every node within a group will have exactly the same influence on the observed output trajectories. In other words the similarity matrix $\Psi(t)$ can be written in terms of the quotient graph as:

$$\Psi(t) = \exp(-Lt)^\top (H_{\mathrm{EE}}^+)^\top H_{\mathrm{EE}}^+ \exp(-Lt)$$
$$= \left[\exp(-\widehat{L}t)H_{\mathrm{EE}}^+\right]^\top \exp(-\widehat{L}t)H_{\mathrm{EE}}^+,$$

which shows that Ψ will be *block-structured*. Consequently the dynamics of the full system within the subspace spanned by the partition can be described *exactly* by a reduced model[44,45], which is governed here by $\mathcal{A} = \widetilde{L}, \mathcal{C} = I$ and has only a single input per group, equal to the average input within the original group.

It is instructive to compare the above dynamical block-structure to the notions like stochastic block-models[6,7],

in which each node in a group has statistically the same (static) connection profile. Here we are interested in nodes that have *dynamically* the same effect, and define nodes accordingly. Note however, that the connections formed by each node do not have to be the same, but simply lead to similar dynamical effects: our measures assess the node similarity with respect to some observable $\mathbf{y}$ and not with respect to the connections formed. In other words our objective is to obtain a joint low-dimensional description of the dynamics of the system and localized features of the network structure. Thus for the same network we may have in principle different types of dynamical modules, depending on the dynamics acting on top of it.

Dimensionality reduction

Consider the spectral decomposition of $\Psi(t)$ into its eigenvectors $\mathbf{v}_1(t), \mathbf{v}_2(t), \ldots, \mathbf{v}_n(t)$ with associated eigenvalues $\mu_1(t) \geq \mu_2(t) \geq \ldots \geq \mu_n(t)$. We define the mapping $i \mapsto \phi_i(t)$:

$$\phi_i(t) = [\sqrt{\mu_1}\, v_{1,i}, \sqrt{\mu_2}\, v_{2,i}, \ldots, \sqrt{\mu_n}\, v_{n,i}]^\top. \qquad (14)$$

Using simple algebraic manipulations, it can now be shown that our dynamical distance measure (4) can be written as:

$$D_{ij}^2(t) = \|\phi_i(t) - \phi_j(t)\|^2. \qquad (15)$$

Hence, the vectors ϕ_i map the data into a Euclidean space, in which the (Euclidean) distance is aligned with the dynamical impacts of the nodes at time t. The (entry-wise squared) distance matrix $D^{(2)}(t) = [D_{ij}^2]$ can thus be approximated by keeping only the first c coordinates in each mapping $\phi_i(t)$, thereby producing a low dimensional embedding of the original system. Other distances induced by Ψ, or different dimensionality reduction techniques, e.g., multi-dimensional scaling[46], may also be used.

Relations to diffusion distance, and diffusion maps. In several applications, continuous datasets are represented via undirected similarity graphs, usually constructed as proximity networks based on nearest neighbours or other geometric constructions of an assortative nature (nearby nodes are similar, and thus more likely to be connected).

Consider a diffusive dynamics on such graphs with either $-\mathcal{A} = L$ (the combinatorial Laplacian) or $-\mathcal{A}^\top = L_{\mathrm{rw}}$ (the random walk Laplacian matrix). In this setting, it can be shown that $D(t)$ corresponds to a continuous-time diffusion distance if the weighting matrix $\mathcal{W}$ is chosen appropriately[42,47]. To see this, not that from the orthogonal spectral decomposition $L = \sum_i \lambda_i \mathbf{v}_i \mathbf{v}_i^\top$, it follows that the dynamic fingerprints

$$\phi_i(t) = [e^{-\lambda_1 t} v_{1i}, \ldots, e^{-\lambda_n t} v_{ni}]^\top$$

constitute a time-dependent diffusion map embedding[42,47].

Finding functional modules via dynamical similarity measures

Instead of finding a continuous embedding for our system, in a number of applications we are interested in grouping nodes according to a discrete set of categories, i.e., assigning them into clusters or modules. While the above described dimensionality reduction techniques aim at embedding the original system into a low dimensional continuous space by using information drawn from the dynamical distance $D(t)$, as we will see in the following grouping nodes into clusters can be naturally understood from the perspective of the similarity matrix Ψ.

Spectral partitioning of the dynamic similarity Ψ and dynamical blocks. Perhaps the most straightforward way to split the nodes of the network into blocks with similar dynamical behavior is to use $\Psi(t)$ in approach akin to spectral clustering[48–50].

To split a system into k groups, we may either use the coordinates defined by ϕ_i (which include contributions from the eigenvalues), or alternatively assemble the matrix V_c containing the top c dominant eigenvectors of the similarity matrix Ψ: $V = [\mathbf{v}_1, \ldots, \mathbf{v}_c]$. The rows of V_c can then taken as new c-dimensional coordinate vectors for each node on which a k-means clustering is run to obtain the k modules. Though, in general, the dimension of the coordinate space c and the number of modules k need not be the same, one typically[50] chooses $c = k$ or $c = k - 1$.

A Louvain-like combinatorial optimization to detect relevant dynamical scales. Alternatively, we can use our similarity measures to derive a quality function in form of a trace maximization problem, as often employed in network analysis. This objective can then be optimzed, e.g., by means of a Louvain-like combinatorial optimization.

To this end, we use an orthogonal projection $\mathcal{W}_\perp = I - \boldsymbol{\nu}\boldsymbol{\nu}^\top$ as weighting matrix, and we denote the ensuing dynamic similarity as:

$$\Psi_\perp(t) = \mathcal{B}^\top \exp(\mathcal{A}t)^\top \mathcal{C}^\top \mathcal{W}_\perp \mathcal{C} \exp(\mathcal{A}t)\mathcal{B}. \qquad (16)$$

As discussed in the SI, this may be seen as performing a *different* type of approximation of the similarity / distance matrix. Rather than keeping the dominant modes, here the weighted inner product $\langle \mathbf{y}_i(t), \mathbf{y}_j(t) \rangle_\mathcal{W}$ projects out particular properties associated with $\boldsymbol{\nu}$.

From a network science perspective, the choice of $\mathcal{W}_\perp$ can thus be interpreted as selecting a type of 'null model' for the nodes or, alternatively, as subtracting uninformative components of the data; for instance, choosing $\boldsymbol{\nu} = \mathbf{1}/\sqrt{n}$ is equivalent to subtracting the mean of each of the vectors (i.e., centering the data). Our results thus also provide us with a fresh perspective on the problem of null-model selection from a geometric perspective.

As further discussed in the SI, note that the time-parameter acts nonlinearly on each entry Ψ_{ij}, and should thus not be seen simply as a null model resolution parameter. Indeed as it is already apparent from the above for-

mulation the time-parameter is independent of the projection term; e.g., an additional resolution parameter α could be incorporated by choosing $\mathcal{W}_\perp = I - \alpha \boldsymbol{\nu}\boldsymbol{\nu}^\top$ (see SI for additional discussion).

Having defined a similarity matrix $\Psi_\perp(t)$ as above, we can obtain dynamical blocks with respect to the null model $\boldsymbol{\nu}$ as follows. Let us define the quality function

$$r_{\boldsymbol{\nu}}(t, H) = \text{trace } H^\top \Psi_\perp(t)\, H, \qquad (17)$$

where H is a partition indicator matrix with entries $H_{ij} = 1$ if state i is in group j and $H_{ij} = 0$ otherwise. The combinatorial optimization of $r_{\boldsymbol{\nu}}(t, H)$ over the space of partitions can be performed efficiently for different values of the time t through an augmented version of the Louvain heuristic (see SI Appendix C). For the special case of diffusive Laplacian dynamics on undirected (but not on directed) networks, it can be shown that $r_{\boldsymbol{\nu}}(t, H)$ becomes equivalent to the Markov Stability cost function[13], and thereby encompasses Modularity[51] and various Potts model heuristics (see SI Appendix D).

ACKNOWLEDGMENTS

REFERENCES

[1] M. E. J. Newman, *Networks: An Introduction* (Oxford University Press, USA, 2010).

[2] A. Arenas, A. Díaz-Guilera, J. Kurths, Y. Moreno, and C. Zhou, "Synchronization in complex networks," Physics Reports **469**, 93–153 (2008).

[3] E. Bullmore and O. Sporns, "Complex brain networks: graph theoretical analysis of structural and functional systems," Nature Reviews Neuroscience **10**, 186–198 (2009).

[4] L. M. Pecora, F. Sorrentino, A. M. Hagerstrom, T. E. Murphy, and R. Roy, "Cluster synchronization and isolated desynchronization in complex networks with symmetries." Nature communications **5**, 4079 (2014).

[5] F. Sorrentino, L. M. Pecora, A. M. Hagerstrom, T. E. Murphy, and R. Roy, "Complete characterization of the stability of cluster synchronization in complex dynamical networks," Science advances **2**, e1501737 (2016).

[6] P. W. Holland, K. B. Laskey, and S. Leinhardt, "Stochastic blockmodels: First steps," Social networks **5**, 109–137 (1983).

[7] T. A. Snijders and K. Nowicki, "Estimation and prediction for stochastic blockmodels for graphs with latent block structure," Journal of classification **14**, 75–100 (1997).

[8] M. T. Schaub, J.-C. Delvenne, S. N. Yaliraki, and M. Barahona, "Markov dynamics as a zooming lens for multiscale community detection: non clique-like communities and the field-of-view limit," PloS one **7**, e32210 (2012).

[9] R. Banisch and N. D. Conrad, "Cycle-flow–based module detection in directed recurrence networks," EPL (Europhysics Letters) **108**, 68008 (2015).

[10] B. Perozzi, R. Al-Rfou, and S. Skiena, "Deepwalk: Online learning of social representations," in *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining* (ACM, 2014) pp. 701–710.

[11] A. Grover and J. Leskovec, "node2vec: Scalable feature learning for networks," in *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining* (ACM, 2016) pp. 855–864.

[12] B. Schölkopf and A. J. Smola, *Learning with kernels: support vector machines, regularization, optimization, and beyond* (MIT press, 2002).

[13] J.-C. Delvenne, S. N. Yaliraki, and M. Barahona, "Stability of graph communities across time scales," Proceedings of the National Academy of Sciences **107**, 12755–12760 (2010).

[14] M. Rosvall and C. T. Bergstrom, "Maps of random walks on complex networks reveal community structure," Proceedings of the National Academy of Sciences **105**, 1118–1123 (2008).

[15] P. Pons and M. Latapy, "Computing communities in large networks using random walks," in *Computer and Information Sciences-ISCIS 2005* (Springer, 2005) pp. 284–293.

[16] M. De Domenico, "Diffusion geometry unravels the emergence of functional clusters in collective phenomena," Physical Review Letters **118**, 168301 (2017).

[17] A. Arenas, A. Fernández, and S. Gómez, "Analysis of the structure of complex networks at different resolution levels," New Journal of Physics **10**, 053039 (2008).

[18] A. Kolchinsky, A. J. Gates, and L. M. Rocha, "Modularity and the spread of perturbations in complex dynamical systems," Physical Review E **92**, 060801 (2015).

[19] J.-C. Delvenne, M. T. Schaub, S. N. Yaliraki, and M. Barahona, "The Stability of a Graph Partition: A Dynamics-Based Framework for Community Detection," in *Dynamics On and Of Complex Networks, Volume 2*, Modeling and Simulation in Science, Engineering and Technology, edited by A. Mukherjee, M. Choudhury, F. Peruani, N. Ganguly, and B. Mitra (Springer New York, 2013) pp. 221–242.

[20] A. Clauset, S. Arbesman, and D. B. Larremore, "Systematic inequality and hierarchy in faculty hiring networks," Science Advances **1** (2015), 10.1126/sciadv.1400005.

[21] C. De Bacco, D. B. Larremore, and C. Moore, "A physical model for efficient ranking in networks," arXiv preprint arXiv:1709.09002 (2017).

[22] A. Litwin-Kumar and B. Doiron, "Slow dynamics and high variability in balanced cortical networks with clustered connections," Nature Neuroscince **15**, 1498–1505 (2012).

[23] M. T. Schaub, Y. Billeh, C. A. Anastassiou, C. Koch, and M. Barahona, "Emergence of slow-switching assemblies in structured neuronal networks," PLoS Computational Biology **11**, e1004196 (2015).

[24] P. Strata and R. Harvey, "Dale's principle," Brain Research Bulletin **50**, 349 – 350 (1999).

[25] M. E. Newman and A. Clauset, "Structure and inference in annotated networks," Nature communications **7** (2016).

[26] T. P. Peixoto, "Parsimonious module inference in large networks," Physical review letters **110**, 148701 (2013).

[27] P. J. Mucha, T. Richardson, K. Macon, M. A. Porter, and J.-P. Onnela, "Community structure in time-dependent, multiscale, and multiplex networks," science **328**, 876–878 (2010).

[28] M. Rosvall, A. V. Esquivel, A. Lancichinetti, J. D. West, and R. Lambiotte, "Memory in network flows and its effects on spreading dynamics and community detection," Nature communications **5** (2014).

[29] J.-C. Delvenne, R. Lambiotte, and L. E. Rocha, "Diffusion on

networked systems is a question of time or structure," Nature communications **6** (2015).

[30] M. MacMahon and D. Garlaschelli, "Community detection for correlation matrices," Phys. Rev. X **5**, 021006 (2015).

[31] G. E. Dullerud and F. Paganini, *A course in robust control theory*, Vol. 6 (Springer New York, 2000).

[32] U. Baur, P. Benner, and L. Feng, "Model order reduction for linear and nonlinear systems: a system-theoretic perspective," Archives of Computational Methods in Engineering **21**, 331–358 (2014).

[33] W. H. Schilders, H. A. Van der Vorst, and J. Rommes, *Model order reduction: theory, research aspects and applications*, Vol. 13 (Springer, 2008).

[34] R. I. Kondor and J. D. Lafferty, "Diffusion Kernels on Graphs and Other Discrete Input Spaces," in *Proceedings of the Nineteenth International Conference on Machine Learning*, ICML '02 (Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2002) pp. 315–322.

[35] A. J. Smola and R. Kondor, "Kernels and regularization on graphs," in *Learning theory and kernel machines* (Springer, 2003) pp. 144–158.

[36] V. Blondel, A. Gajardo, M. Heymans, P. Senellart, and P. Van Dooren, "A measure of similarity between graph vertices: Applications to synonym extraction and web searching," SIAM Review **46**, 647–666 (2004).

[37] E. Leicht, P. Holme, and M. Newman, "Vertex similarity in networks," Phys. Rev. E **73**, 026120 (2006).

[38] L. Lü and T. Zhou, "Link prediction in complex networks: A survey," Physica A: Statistical Mechanics and its Applications **390**, 1150–1170 (2011).

[39] F. Fouss, A. Pirotte, J. Renders, and M. Saerens, "Random-walk computation of similarities between nodes of a graph with application to collaborative recommendation," IEEE Transactions on knowledge and data engineering (2007).

[40] H. Abou-Kandil, G. Freiling, V. Ionescu, and G. Jank, *Matrix Riccati equations in control and systems theory* (Birkhäuser, 2012).

[41] R. E. Skelton, T. Iwasaki, and D. E. Grigoriadis, *A unified algebraic approach to control design* (CRC Press, 1997).

[42] R. R. Coifman, S. Lafon, A. B. Lee, M. Maggioni, B. Nadler, F. Warner, and S. W. Zucker, "Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps," Proceedings of the National Academy of Sciences of the United States of America **102**, 7426–7431 (2005).

[43] E. I. Verriest, "Time Variant Balancing and Nonlinear Balanced Realizations," in *Model Order Reduction: Theory, Research Aspects and Applications*, edited by W. H. A. Schilders, H. A. van der Vorst, and J. Rommes (Springer Berlin Heidelberg, Berlin, Heidelberg, 2008) pp. 213–250.

[44] N. O'Clery, Y. Yuan, G.-B. Stan, and M. Barahona, "Observability and coarse graining of consensus dynamics through the external equitable partition," Physical Review E **88**, 042805 (2013).

[45] N. Monshizadeh, H. L. Trentelman, and M. K. Camlibel, "Projection-based model reduction of multi-agent systems using graph partitions," Control of Network Systems, IEEE Transactions on **1**, 145–154 (2014).

[46] I. Borg and P. J. Groenen, *Modern multidimensional scaling: Theory and applications* (Springer Science & Business Media, 2005).

[47] S. Lafon and A. Lee, "Diffusion maps and coarse-graining: a unified framework for dimensionality reduction, graph partitioning, and data set parameterization," Pattern Analysis and Machine Intelligence, IEEE Transactions on **28**, 1393 –1403 (2006).

[48] J. Shi and J. Malik, "Normalized cuts and image segmentation," Pattern Analysis and Machine Intelligence, IEEE Transactions on **22**, 888–905 (2000).

[49] A. Y. Ng, M. I. Jordan, Y. Weiss, *et al.*, "On spectral clustering: Analysis and an algorithm," Advances in neural information processing systems **2**, 849–856 (2002).

[50] U. Von Luxburg, "A tutorial on spectral clustering," Statistics and computing **17**, 395–416 (2007).

[51] M. E. Newman and M. Girvan, "Finding and evaluating community structure in networks," Physical review E **69**, 026113 (2004).

[52] C. Altafini, "Consensus Problems on Networks With Antagonistic Interactions," Automatic Control, IEEE Transactions on **58**, 935–946 (2013).

[53] C. Altafini and G. Lini, "Predictable Dynamics of Opinion Forming for Networks With Antagonistic Interactions," Automatic Control, IEEE Transactions on **60**, 342–357 (2015).

[54] E. W. D. Luca, S. Albayrak, J. Kunegis, A. Lommatzsch, S. Schmidt, and J. Lerner, "Spectral Analysis of Signed Graphs for Clustering, Prediction and Visualization," in *Proceedings of the 2010 SIAM International Conference on Data Mining* (2010) Chap. 48, pp. 559–570.

[55] K. E. Read, "Cultures of the Central Highlands, New Guinea," Southwestern Journal of Anthropology **10**, pp. 1–43 (1954).

[56] P. Hage and F. Harary, *Structural Models in Anthropology* (Cambridge University Press, 1983).

[57] J. Kunegis, S. Schmidt, A. Lommatzsch, J. Lerner, E. W. D. Luca, and S. Albayrak, "Spectral Analysis of Signed Graphs for Clustering, Prediction and Visualization," in *Proceedings of the 2010 SIAM International Conference on Data Mining* (2010) pp. 559–570.

[58] D. Cartwright and F. Harary, "Structural balance: a generalization of Heider's theory." Psychological review **63**, 277 (1956).

[59] V. A. Traag and J. Bruggeman, "Community detection in networks with positive and negative links," Phys. Rev. E **80**, 036115 (2009).

[60] S. Fortunato, "Community detection in graphs," Physics Reports **486**, 75 – 174 (2010).

[61] J. Reichardt and S. Bornholdt, "Statistical mechanics of community detection," Phys. Rev. E **74**, 016110 (2006).

[62] R. Campigotto, P. C. Céspedes, and J.-L. Guillaume, "A generalized and adaptive method for community detection," arXiv:1406.2518 (2014).

[63] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," Journal of Statistical Mechanics: Theory and Experiment **2008**, P10008 (2008).

[64] M. T. Schaub, R. Lambiotte, and M. Barahona, "Encoding dynamics for multiscale community detection: Markov time sweeping for the map equation," Phys. Rev. E **86**, 026112 (2012).

[65] S. Fortunato and M. Barthélemy, "Resolution limit in community detection," Proceedings of the National Academy of Sciences **104**, 36–41 (2007).

[66] A. Lancichinetti and S. Fortunato, "Limits of modularity maximization in community detection," Phys. Rev. E **84**, 066122 (2011).

[67] R. Lambiotte, "Multi-scale modularity in complex networks," in *Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt), 2010 Proceedings of the 8th International Symposium on* (IEEE, 2010) pp. 546–553.

[68] A. Delmotte, E. W. Tate, S. N. Yaliraki, and M. Barahona, "Protein multi-scale organization through graph partitioning and robustness analysis: application to the myosin? myosin light chain interaction," Physical biology **8**, 055010 (2011).

[69] M. Meila, "Comparing clusterings–an information based distance," Journal of Multivariate Analysis **98**, 873 – 895 (2007).

[70] K. Cooper and M. Barahona, "Role-similarity based comparison of directed networks," ArXiv , arXiv:1103.5582 (2011).

[71] K. Cooper and M. Barahona, "Role-based similarity in directed networks," (2010).

[72] M. Faccin, M. T. Schaub, and J.-C. Delvenne, "Entrograms and coarse graining of dynamics on complex networks," Journal of Complex Networks , cnx055 (2017).

[73] R. J. Sanchez-Garcia, "Exploiting symmetry in network analysis," arXiv preprint arXiv:1803.06915 (2018).

[74] M. Schaub, N. O'Clery, Y. N. Billeh, J.-C. Delvenne, R. Lambiotte, and M. Barahona, "Graph partitions and cluster synchronization in networks of oscillators," Chaos **26**, 094821 (2016).

[75] R. Lambiotte, J. Delvenne, and M. Barahona, "Random Walks, Markov Processes and the Multiscale Modular Organization of Complex Networks," Network Science and Engineering, IEEE Transactions on **1**, 76–90 (2014).

[76] P. Bremaud, *Markov Chains: Gibbs fields, Monte Carlo simulation, and queues*, corrected edition ed. (Springer, 1999).

[77] R. Gallager, *Stochastic Processes: Theory for Applications*, Stochastic Processes: Theory for Applications (Cambridge University Press, 2013).

[78] J. Reichardt and S. Bornholdt, "Detecting Fuzzy Community Structures in Complex Networks with a Potts Model," Phys. Rev. Lett. **93**, 218701 (2004).

[79] V. A. Traag, P. Van Dooren, and Y. Nesterov, "Narrow scope for resolution-limit-free community detection," Phys. Rev. E **84**, 016114 (2011).

[80] L. Page, S. Brin, R. Motwani, and T. Winograd, "The PageRank citation ranking: bringing order to the Web." (1999).

[81] V. Satuluri and S. Parthasarathy, "Symmetrizations for clustering directed graphs," in *Proceedings of the 14th International Conference on Extending Database Technology* (ACM, 2011) pp. 343–354.

[82] J. M. Kleinberg, "Authoritative sources in a hyperlinked environment," Journal of the ACM (JACM) **46**, 604–632 (1999).

[83] R. Lambiotte and M. Rosvall, "Ranking and clustering of nodes in networks with smart teleportation," Phys. Rev. E **85**, 056107 (2012).

[84] M. T. Schaub, *Unraveling complex networks under the prism of dynamical processes: relations between structure and dynamics*, Ph.D. thesis, Imperial College London (2014).

SUPPLEMENTARY INFORMATION The supplementary information is structured as follows. We first provide additional details on the two main examples discussed in the text in Section A, and Section B. We further describe a third illustrative example in which both the embedding and the detection of modules is exemplified in Section C.

We then expand on various relationships between the here developed dynamical similarity and distance measures and notions from network theory. First, in Section D, we highlight relations and differences between the embedding and module detection tasks. Second, in Section E, we comment on how specific measures can be recovered and extended within the here presented framework, when focussing on diffusion processes.

Appendix A: Analysing academic hiring networks via low-dimensional influence based embeddings

Here we provide additional information on the data and analysis performed in the main text.

The dataset we considered has been originally collected and released by Clauset et al.[20] and has been made available at http://tuvalu.santafe.edu/~aaronc/facultyhiring/. It consists of the placement of nearly 19,000 tenure track or tenured faculty among 461 North American departmental or school level academic units. The hiring data was collected for the disciplines business (112 institutions), history (144 institutions), and computer science (205 institutions). In the main text we analyzed the computer science data, as it contains the largest number of institutions. In contrast to the history and business data, the computer science hiring data moreover included 23 Canadian institutions. These institutions, though well integrated into the North American hiring market[20], appear to play a different role in the hiring market, however (see main text). The dataset also includes the ranking obtained by Clauset et al, as well as the US News ranking and the ranking from the National Research Council (NRC).

Below we show the corresponding analyses for the disciplines history and business. With Canadian institutions absent from this data, the second dimension of the embedding appears to be not correlate clearly with a geographical feature. However, for the business data there appears to be a slight separation along the second dimension between the more coastal regions (West, Northeast), which have a higher $\phi_{i,2}$ projection and the South and Midwest institutions, which tend to have a lower coordinate $\phi_{i,2}$. For the History dataset, the Southern Baptist Theological Seminary appears to be singled out in our second projection coordinate. It turns out that one of the main differences of this unit is its relatively large number of self-loops in the hiring data (11 hirings come from the same institution).

From the first dimension of the embedding we can again derive an influence ranking that is strongly correlated to the results obtained by Clauset et al. (in both cases the Spearman rank correlation is > 0.9). We show the rankings thus obtained in Figure 4. A more detailed investigation of these datasets, potentially linking it to the relaxation dynamics of the recently proposed SpringRank method[21] would be an interesting subject of future investigations.

Appendix B: Leaky-integrate-and-fire neural networks and functional modules

Due to their computational simplicity yet complex dynamics, networks of LIF neurons are widely used as scalable prototypes of neural activity. The non-linear dynamics of LIF models reproduce Poisson-like neuronal firing with refractory periods, among other features. Here, we employ that LIF networks display structured behavior[22,23], in which sustained in-group spiking switches from group to group across the network. Importantly, these cell assemblies of coherently spiking neurons include both excitatory neurons (which exhibit a positive influence on their neighbours) and inhibitory neurons (whose influence is negative), which have different connection profiles. Moreover, we remark that these groups are not densely connected clusters which are only weakly connected to other clusters, but the behavior emerges from the connections between the various groups. Stated differently, the observed grouping is dynamical (functional) rather than structural.

We simulated leaky-integrate-and-fire (LIF) networks with $n = 1000$ neurons (800 excitatory, 200 inhibitory). Using a time step of 0.1ms, we numerically integrated the non-dimensionalized membrane potential of each neuron,

$$\frac{dV_i(t)}{dt} = \frac{1}{\tau_m^{E/I}}(u_i - V_i(t)) + \sum_j [W_N]_{ij}\, g_j^{E/I}(t), \quad \text{(B1)}$$

with a firing threshold of 1 and a reset potential of 0. The input terms u_i were chosen uniformly at random in the interval $[1.1, 1.2]$ for excitatory neurons, and in the interval $[1, 1.05]$ for inhibitory neurons. The membrane time constants for excitatory and inhibitory neurons were set to $\tau_m^E = 15$ ms and $\tau_m^I = 10$ ms, respectively, and the refractory period was fixed at 5 ms for both excitatory and inhibitory neurons. Note that although the constant input term is supra-threshold, balanced inputs guarantee an average sub-threshold membrane potential[22]. The network dynamics is captured by the sum in (B1), which describes the input to neuron i from all other neurons in the network and $[W_N]_{ij}$ denotes the weight of the connection from neuron j to neuron i. Synaptic inputs are modelled by $g_j^{E/I}(t)$, which is increased step-wise instantaneously after a presynaptic spike of neuron j ($g_j^{E/I} \to g_j^{E/I} + 1$) and then decays exponentially according to:

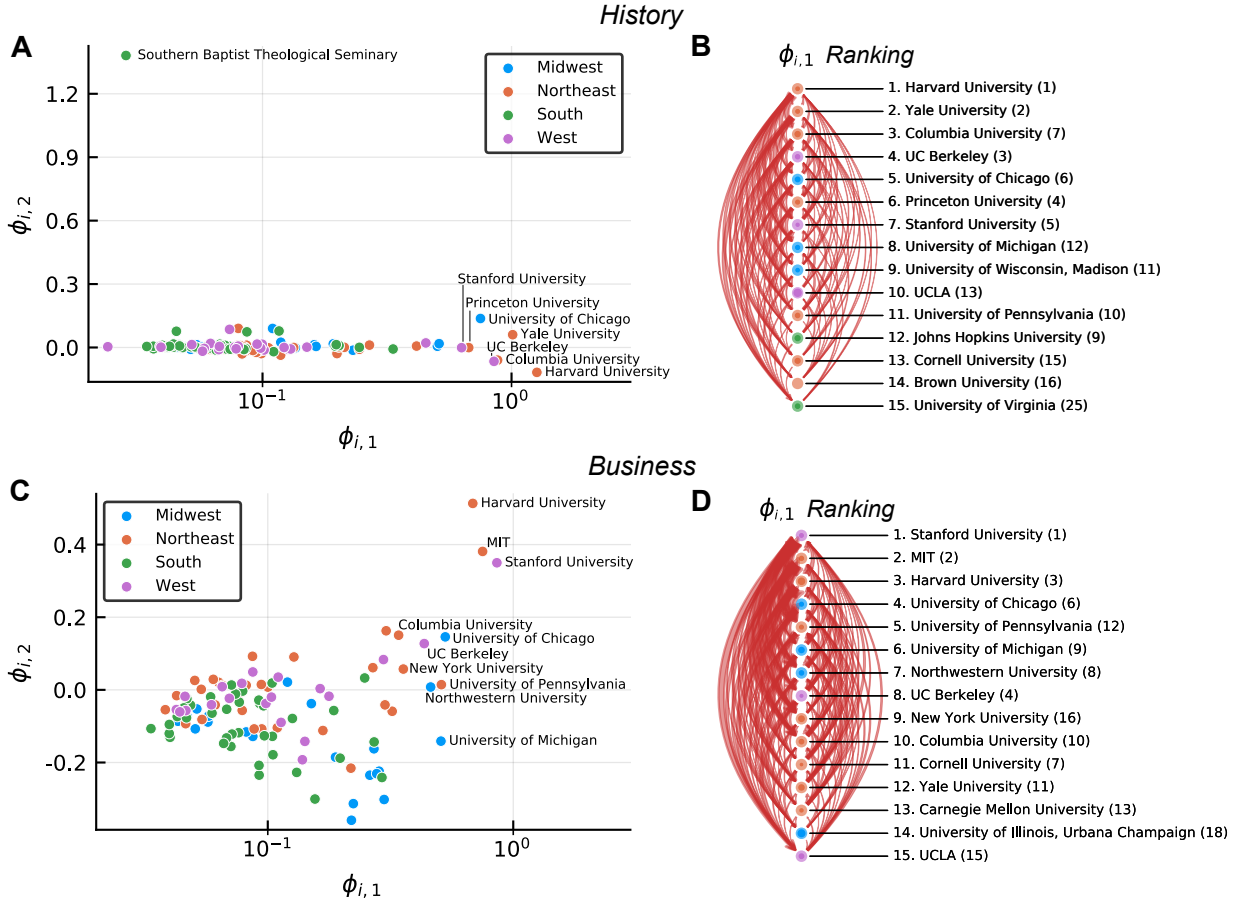$$\tau_s^{E/I}\frac{dg_j^{E/I}}{dt} = -g_j^{E/I}(t), \quad \text{(B2)}$$

Figure 4. **Analysing academic influence using low-dimensional embeddings A** Low-dimensional embeddings based on the influence dynamics in the hiring network, as described in the text (see Figure 2) for the discipline History. **B** The ranking obtained when projecting onto the first coordinate only and the associated subgraph of faculty hirings (see Figure 2). The Spearman rank correlation to the results obtained by Clauset et al is $\rho \approx 0.92$. **C** Low-dimensional embeddings based on the influence dynamics in the hiring network, as described in the text (see Figure 2) for the discipline Business. **D** The ranking obtained when projecting onto the first coordinate only and the associated subgraph of faculty hirings (see Figure 2). The Spearman rank correlation to the results obtained by Clauset et al is $\rho \approx 0.96$.

with time constants $\tau_s^E = 3$ ms for an excitatory interaction, and $\tau_s^I = 2$ ms if the presynaptic neuron is inhibitory. Excitatory and inhibitory neurons were connected uniformly with probabilities $p_{EE} = 0.2$, $p_{II} = 0.5$, and weight parameters $W_{EE} = 0.022$ and $W_{II} = 0.042$, respectively.

The network comprised 10 functional groups of neurons, each of which consists of 80 excitatory and 20 inhibitory neurons connected as follows. The excitatory neurons are statistically biased to target the inhibitory neurons in their own assembly with probability $p_{IE}^{in} = 0.90$ and weight $W_{IE}^{in} = 0.0263$, compared to $p_{IE} = 0.4545$ and $W_{IE} = 0.0087$ otherwise. Inhibitory neurons connect to all excitatory neurons with probability $p_{EI} = 0.5263$ and weight $W_{EI} = 0.045$, apart from the excitatory neurons in their own assembly which are connected with probability $p_{EI}^{in} = 0.2632$ and $W_{EI}^{in} = 0.015$. Note that, while from a purely structural point of view we may split this network into 20 groups (10 groups of excitatory neurons, 10 groups of inhibitory neurons; see also

Figure 3), it can be shown that this configuration gives rise to 10 functional groups of neurons firing in synchrony with respect to the rest of the network[23].

Appendix C: The enemy of my enemy is my friend – dynamical embeddings in a signed social interaction network of highland tribes of New Guinea

Many networked systems contain both attractive and repulsive interactions. Examples include social systems, in which people may be friends or foes, or genetic networks, in which inhibitory and excitatory interactions are commonplace. Such systems can be represented as *signed graphs*, with positive and negative edge weights. A simple model for opinion formation on signed networks is given by[52,53]:

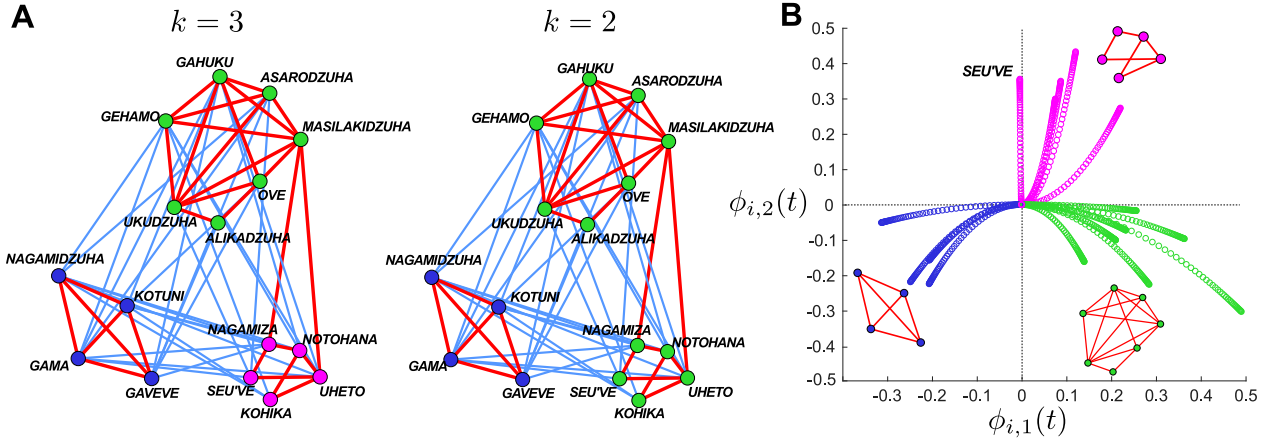$$\dot{\mathbf{x}} = -L_s \mathbf{x} + \mathbf{u}, \qquad (C1)$$

Figure 5. **Analysis of a signed social network: the highland tribes in New Guinea. A** The network of 16 tribes with positive interactions ('hina') in red and negative interactions ('rova') in blue. A spectral clustering of the similarity matrix Ψ using the top $c = 2$ eigenvectors reveals partitions into $k = 3$ and $k = 2$ groups with positive interactions mostly concentrated within groups, and antagonistic interactions across groups. **B** The time evolution of the tribes in state space ('opinion space' of the tribes) under the consensus dynamics (C1) is represented through the dynamical embeddings $\phi_i(t)$. Here we plot only the first two dominant coordinates. As time grows, the Seu've tribe switches from a marginal allegiance to the pink/green groupings to be grouped with the blue block. This is the result of an 'enemy of my enemy is my friend' effect.

where the *signed Laplacian* matrix is defined as $L_s = D_s - A_s$. Here A_s is the adjacency matrix of the network, with positive and negative edge weights, and D_s is the matrix containing the weighted absolute strengths of the nodes on the diagonal, $[D_s]_{ii} = \sum_k |A_{ik}|$ and $[D_s]_{ij} = 0$ for $i \neq j$. The signed Laplacian is positive semidefinite[52,54] and reduces to the standard combinatorial Laplacian if A contains only positive weights. Clearly, this dynamics is of the form (2) discussed in the main text, with $\mathcal{A} = -L_s$ and $\mathcal{B} = \mathcal{C} = I$. In this case, the dynamic similarity

$$\Psi(t) = \exp(-L_s t)^\top \exp(-L_s t), \qquad \text{(C2)}$$

has time-independent eigenvectors $\mathbf{v}_i$ and associated eigenvalues $\mu_i(t) = e^{-\lambda_i t}$, where the $\mathbf{v}_i$ and λ_i are eigenvectors and eigenvalues of L_s.

As an illustrative example, consider the network of relationships between 16 tribal groups in New Guinea chartered by Read[55] and first examined in the social network literature by Hage and Harari[56]. The relationships between the different tribes are either sympathetic ('hina'; red edges in Fig. 5) or antagonistic ('rova'; blue edges in Fig. 5). A 'hina' edge signifies political alignment and limited feuds; 'rova' edges denote relationships in which warfare is commonplace.

We applied spectral clustering to $\Psi(t)$ for large t based on the top 2 eigenvectors ($\mathbf{v}_1$ and $\mathbf{v}_2$) to split the network into $k = 2, 3$ groups (Fig. 5A). The blocks obtained are characterized by high internal density of positive links with negative links placed across groups. These groupings reflect the relative strength of positive links between the groups conataining the Ove and Seu've tribes. It can be easily shown that our procedure recovers a spectral heuristic that optimizes the signed ratio cut[57].

Further information from $\Psi(t)$ can be revealed by the 'dynamical fingerprints' $\phi_i(t)$ (defined in Eq. (14) in Methods), which can be seen as feature vectors that combine the information of the eigenvectors and eigenvalues. The time evolution of these fingerprints provides a dynamical embedding of the signed opinion network, which reflects the relative position of the nodes (tribes) in the state-space of the system, and a more nuanced appreciation of how closely aligned individual tribes are to each other over time (Figure 5B).

Specifically, the role of the Seu've tribe is noteworthy. The Seu've subtribe has direct positive relationships to the Nagamiza and Uheto subtribes, and negative relationships with Asarodzuha and Ukudzuha, and may thus be seen as 'neutral' with respect to the green group in Figure 5. The only direct link of Seu've with the blue group is however a negative interaction with the Gama tribe. From a structural perspective, the Seu've clan is thus favorably placed within the green group, notwithstanding its controversies with Asarodzuha and Ukudzuha.

Our dynamical embedding shows that this placement of the Seu've tribe within the green group is weak, however, and at long times the Seu've tribe becomes more similar to the nodes in the blue module including the Gama tribe, despite their antagonistic relationship. The reason for this switching is the behavior described as "the enemy of my enemy is my friend", which is inherent to signed interaction dynamics. This effect plays a more important role for larger time scales. In terms of direct interactions, the Seu've tribe is more similar to the green group; yet when the indirect dynamical relationships propagated through the signed Laplacian dynamics are taken into account, the Seu've tribe becomes more similar to the blue group. In particular, the mutual an-

tipathy of all three Seu've, Gama and Nagadmidzuha clans against the Asarodzuha tribe implies a net positive interaction between Gama and Nagadmidzuha with Seu've in the long run. Following structural balance theory[58], one may conjecture that the Gama-Seu've relationship could cease to be of 'rova' type in a future observation of the network. In his socio-ethnographic characterization of this tribal system, Read indeed remarked that the system was "relative and dynamic"[55].

While other methods have been proposed to cluster signed networks[27,54,59] that can be used to find the groupings in the here considered setting, all of them follow a combinatorial approach and aim to find dense groupings in the network containing a maximum number of positive links within the groups and most negative links across groups. Our analysis highlights that there is additional information to be gained when adopting a dynamical point of view, as shown by potential of the Seu've tribe to be 'turned around'.

Appendix D: Relations between dimensionality reduction and module detection

In this section we elaborate on the relationship between dimensionality reduction and the detection of dynamical modules as discussed in the main text.

Let us initially consider the problem from the point of view of the squared distance matrix $D^{(2)} := [D_{ij}^2]$, where here we omit writing the time-dependence to emphasize that the derivations below apply to both the integrated $D_{[0,t]}^{(2)}$ as well as the instantaneous distance matrix $D^{(2)}(t)$.

A naive idea to derive a clustering measure would be to simply try and place all nodes into the same group such that the sum of the distances in each group is minimized, which would lead to the following optimization procedure:

$$\min_H \text{trace } H^\top D^{(2)} H,$$

where $H \in \{0,1\}^{n \times k}$ is a partition indicator matrix with $H_{ij} = 1$ if node i is in group j and $H_{ij} = 0$ otherwise. We can rewrite the above using the definition of $D^{(2)}$ as

$$\min_H \text{trace } H^\top \left[\mathbf{1}\mathbf{z}^\top + \mathbf{z}\mathbf{1}^\top - 2\Psi \right] H,$$

where $\mathbf{z} = \text{diag}(\Psi)$ is the vector containing the diagonal entries of Ψ. It is easy to see that if k is not constrained in the above optimization problem, then the best choice will be to trivially put each node in its own group ($k = n$). Stated differently, if we are free to choose any number of groups k, then we can make the distance within each group zero, thus minimizing the above objective.

One potential remedy to fix the above shortcoming would be to fix the number of groups, a priori, and then perform some kind of selection procedure afterwards to pick the number of groups. Another option is to introduce some 'slack' in the distance measurements, thus permitting nodes whose distance is comparably small to contribute negative to the cost function (which is here to be minimized). As we will show in the following this naturally leads to a problem formulation akin to many network partitioning procedures which have been proposed in the literature.

Let us consider the spectral expansion $\Psi = \sum_{i=1}^n \lambda_i \mathbf{v}_i \mathbf{v}_i^\top$, where we assume the eigenvalues to be ordered, such that $\lambda_1 > \ldots > \lambda_n \geq 0$. Then we can rewrite D_{ij}^2 as:

$$D_{ij}^2 = \sum_{k=1}^n \lambda_k (\mathbf{v}_k^\top \mathbf{e_i})^2 + \lambda_k (\mathbf{v}_k^\top \mathbf{e_i})^2 - 2\lambda_k (\mathbf{v}_k^\top \mathbf{e_j})(\mathbf{v}_k^\top \mathbf{e_i}),$$

where $\mathbf{e_i}$ is the i-th unit vector. Let us now introduce some slack variables (multipliers) γ_k for all the modes in the first two terms and rewrite the above expression in terms of the dynamical coordinates $\boldsymbol{\phi}_k$:

$$D_{ij}^2 = \sum_{k=1}^n \gamma_k \lambda_k (\mathbf{v}_k^\top \mathbf{e_i})^2 + \gamma_k \lambda_k (\mathbf{v}_k^\top \mathbf{e_i})^2 - 2\lambda_k (\mathbf{v}_k^\top \mathbf{e_j})(\mathbf{v}_k^\top \mathbf{e_i})$$

$$= \sum_{k=1}^n \gamma_k [(\phi_{i,k})^2 + (\phi_{j,k})^2] - 2\boldsymbol{\phi}_i^\top \boldsymbol{\phi}_j,$$

which shows that γ_k, may be seen as weighting functions for the first k coordinates in the $\boldsymbol{\phi}$ coordinate space for the first 2 (norm) terms in the distance.

Using the above derivation, let us rewrite the previously considered minimization as an equivalent maximization problem.

$$\max_H 2 \text{ trace } H^\top [\Psi - \frac{1}{2}(\tilde{\mathbf{z}}\mathbf{1}^\top - \mathbf{1}\tilde{\mathbf{z}}^\top)]H$$
$$\text{with } \tilde{\mathbf{z}} = \text{diag } \Phi^\top \Gamma \Phi \in \mathbb{R}^n,$$
$$\Gamma = \text{diag}(\gamma_1, \ldots, \gamma_n) \in \mathbb{R}^{n \times n}$$

While different weighting schemes $\{\gamma_k\}$ are of potential interest here, let us now consider the specific choice $\gamma_1 = 1$, $\gamma_k = 0$, $(k > 1)$, which corresponds to making a simple low rank-correction of Ψ. This specific scheme is akin to choosing a type of null model in our optimization scheme as we will illustrate next.

For concreteness, let us consider the familiar case of Laplacian dynamics for a symmetric graph, i.e., $\Psi = \exp(-Lt)^\top \exp(-Lt)$. In this case the eigenvectors of Ψ are constant over time, and the first eigenvalue of Ψ is one with an associated constant eigenvector and thus $\phi_{i,1} = 1/\sqrt{n}, \forall i$. This leads to an optimization of the form:

$$\max_H 2 \text{ trace } H^\top [\exp(-2Lt) - \frac{1}{2n}(\mathbf{1}\mathbf{1}^\top - \mathbf{1}\mathbf{1}^\top)]H,$$

which can be simplified to the equivalent optimization:

$$\max_H \text{ trace } H^\top [\exp(-2Lt) - \frac{1}{n}\mathbf{1}\mathbf{1}^\top]H$$

Note that this is just the (rescaled) Markov stability at time $2t$ (see also Section E). Linearising the above expression thus leads to recovering a Potts-model like community detection scheme[19], where the last term can be identified with an Erdős-Rényi null model. Following exactly the same procedure, similar expressions may also be derived for various other null models, such as the configuration model.

Further, as discussed in the next section, the above expression can be rewritten in the form $r_{\boldsymbol{\nu}}(t, H)$ (see Equation (17)), and can thus be interpreted as a quality function that we can optimize using the Louvain optimization scheme. This emphasizes how the Louvain optimization scheme can be interpreted as operating with an approximation of the distance matrix / similarity matrix Ψ. This result may be used in several ways to derive some more general null models by choosing an appropriate weighting scheme.

1. Null models, projections and time-parameter choices

As observed above, the Louvain algorithm might be seen as solving a closely related ('dual') problem to the distance minimization. However, instead of trying to find groups with minimal distance according to some criterion, the optimization operates in terms of the associated similarity measure (inner product). An interesting question that we will not pursue in the following would thus be to investigate equivalent distance based optimization problems. Instead, in the following we will discuss how the here derived formulation ties in with many quality function commonly considered in networks analysis.

Within network science many quality functions for community detection can effectively be written in the form[19,60–62]:

$$r(H) = \text{trace } H^\top \left[G - \alpha N\right] H, \tag{D1}$$

where G is a term customarily related to the network structure, and N is a 'null-model' term, which customarily includes a scalar multiplier α as a free resolution parameter. It is insightful to rewrite the above as:

$$r(H) = \langle H, GH \rangle - \alpha \langle H, NH \rangle. \tag{D2}$$

In particular, if the null model term is a positive semi-definite matrix, we can further simplify this to:

$$r(H) = \langle H, GH \rangle - \alpha \|N^{1/2}H\|_F^2, \tag{D3}$$

where $\| \cdot \|_F$ is the Frobenius norm. This highlights how the null model term acts effectively as a regularization term (similar to regression type-problems) with a weighting factor (Lagrange multiplier) given by the resolution parameter α.

Let us now consider how the above formulations apply to the similarity measures derived above. To link with our above discussion let us concentrate on the case where we (partially) project out a rank-1 term via $\mathcal{W} = I - \alpha \boldsymbol{\nu}\boldsymbol{\nu}^\top$. Note that the same effect can also be achieved by choosing $\mathcal{C}$ appropriately, providing additional interpretations which we will not explore in the following. This would lead to a similarity matrix $\Psi_\perp$ of the form:

$$\Psi_\perp = Y^\top Y - \alpha Y^\top \boldsymbol{\nu}\boldsymbol{\nu}^\top Y, \tag{D4}$$

where $Y = \mathcal{C}\exp(\mathcal{A}t)\mathcal{B}$. This kind of similarity lead to a quality function of the form:

$$r_\nu(t, \alpha, H) = \|Y(t)H\|^2 - \alpha\|\boldsymbol{\nu}^\top Y(t)H\|^2 \tag{D5}$$

$$= \|H\|_{\Psi(t)}^2 - \alpha\|\boldsymbol{\nu}^\top Y(t)H\|^2 \tag{D6}$$

where we explicitly have written out the dependency on t and α and have defined the semi-norms $\|X\|_Y :=$ trace $X^\top Y X$, which is possible in our formulation as all the relevant matrices are positive semi-definite.

From the above we can make the following observations. First, as already alluded to before, the influence of the resolution parameter α is akin to a Lagrange multiplier, which *linearly* scales the regularization term. In contrast the influence of the time-parameter is more subtle as it changes the eigenvalues / eigenvectors of Ψ and thus acts in a nonlinear fashion.

Second, by choosing a particular $\boldsymbol{\nu}$ in the projection term such that a time-independent component is picked out from Y, we can recover classical some classical null model terms. As an example we can again consider the symmetric Laplacian dynamics $Y = \exp(-Lt)$, for which the centering operation $\mathcal{W}_\perp = I - \mathbf{1}\mathbf{1}^\top/n$ corresponds to projecting out the stationary eigenvector (see also Section E, where the case $Y = \exp(-D^{-1}Lt)$ and the configuration null model is discussed as well). Note, however, that in general the second term remains time-dependent and the null model may vary with time, too. The choice of a projection may thus be guided either by simple considerations on what aspect of the dynamics we are interested in (e.g., the relative difference of the influence which would lead to a type of centering operation), or by suppressing some type of mode which we know might be irrelevant for our considerations (e.g., the stationary distribution in a diffusion process[8,19]).

2. Optimizing the quality measure $r_\nu(t, H)$ using an adapted Louvain algorithm

The optimization of the quality measure $r_\nu(t, H)$ given in (17) can be achieved by various means, e.g., via MCMC or spectral techniques. Here we propose to use an augmented version of the Louvain algorithm[63], which was initially proposed as an efficient algorithm to optimize the Newman-Girvan modularity[51]. The algorithm operates as follows:

1. Loop over all nodes in a random order, and assign each node greedily to the community for which the

increase in quality is maximal until no further move is possible.

2. Build a coarse-grained network, in which each node represents a community in the previous network.

3. Repeat steps 1-2, until no further improvement is possible.

As outlined in Ref.[62], this generic procedure can be used to optimize any quality function of the form:
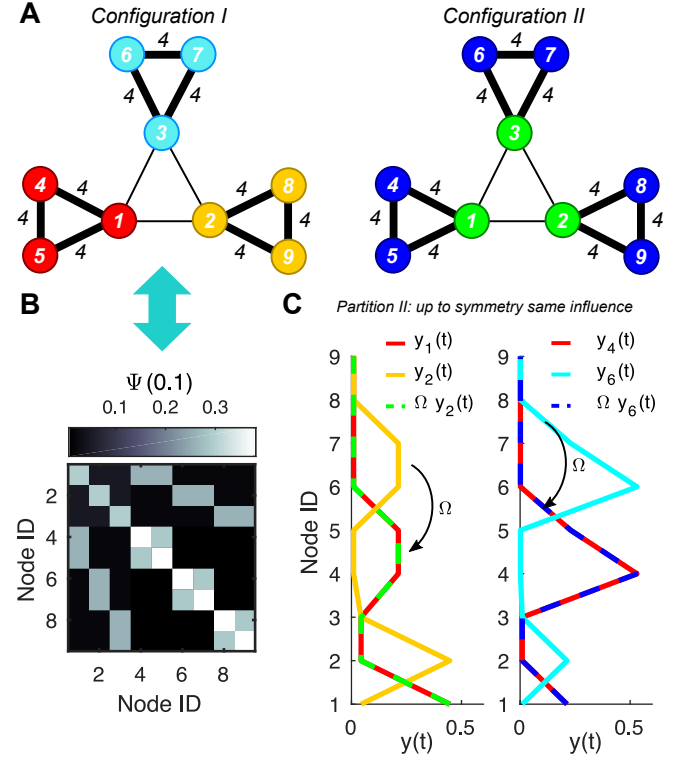
$$\text{trace } H^\top [F - \mathbf{a}\mathbf{b}^\top]H, \qquad (D7)$$

where H is the partition indicator matrix, F is a general matrix derived from the network, and $\mathbf{a}, \mathbf{b}$ are two n dimensional vectors. The quality function (17) is clearly of this form. An inherent problem of many community detection measures is the choice of the relevant resolution, or scale, of the partitioning. In many methods this choice has to be made explicitly a priori, by declaring how many groups are to be found by the method. If this is not the case, then there there is either a free ('resolution') parameter or a regularization scheme, with which the size of the groups found can be controlled explicitly or implicitly, or there is an implicit scale associated with the method, which will determine an upper and lower limit of size the communities to be found[8,64–66].

Instead of choosing and fixing a particular scale, we here identify significant partitions according to the criteria outlined in Refs.[8,67,68]. We advocate to look at the trajectories for all times t and let thereby the dynamical process reveal the important scales of the problem. These scales should be associated with robust partitions over time and relative to the optimization. Thus we are interested in identifying robust partitions as indicated by: (i) a persistence to (small) time-variations, which translates into long plateaux in the number of communities plotted against time; (ii) consistency of partitions obtained from the generalized Louvain algorithm over random initialisation conditions. as measured by the mean distance between partitions using the normalized Variation of Information (VI) metric[69]. A VI of zero results when all iterations of the Louvain algorithm return exactly the same clustering. By computing the matrix $VI(t, t')$, containing the mean variation of information between any two sets of partitions at different times, we can easily identify time-epochs over which we always obtain very similar, robust partitions.

3. Dynamical roles, modules and symmetries

As many notions of dynamical or functional role have been presented in the literature so far, we provide here a short conceptual clarification on what we would consider a 'dynamical' module in this work, in the sense that our similarity measures would assign a high similarity score between each node in a module. To this end consider the example network depicted in Figure 6. For simplicity of

Figure 6. **Dynamical similarities, modules and roles**. **A** A small network with 2 possible partitions with nodes may be described as similar. **B** If we were to group the nodes according to the dynamical similarity measure as defined in the text we would pick out configuration A, as the nodes in each group have a similar impulse response after time t. **C** However, if we consider the impulse responses up to the action of a permutation Ω (isometric mapping), we could also infer the role' partition of configuration B.

our exposition we will consider here a simple consensus dynamics of the form $\dot{\mathbf{x}} = -L x$, where L is the standard graph Laplacian. In this case it is essentially the structure of the network that dictates how our similarity measures evolves through the spectral properties of the Laplacian.

We consider two possible partitions in Figure 6A. Both partition may be seen to correspond to a type of 'role' of the nodes in the network. In this case, as can been seen in Figure 6B, our notion of similarity is commensurate with partition I, as the impulse responses of the nodes in the colored group influence the same parts of the network in essentially the same way. Stated differently, if we denote by $\mathbf{y}_i(t)$ the impulse response of node i after time t, then after a brief transient period an impulse given to any node in the same group results in approximately the same state vector of the network (e.g. $\mathbf{y}_1(t) \approx \mathbf{y}_4(t) \approx \mathbf{y}_5(t)$, in case of the red group).

However, the nodes in partition II are indeed similar in the following sense. If we consider the vectors $\mathbf{y}_i(t)$ up to the action of a symmetry group (permutation), then we can see that indeed partition II groups nodes together

which are similar in this sense. More precisely, call Ω a permutation matrix corresponding to the orbit partition II indicated in Figure 6A. Then for any two nodes i, j in the same group there exist a permutation matrix Ω such that $\mathbf{y}_i = \Omega \mathbf{y}_j$ (see Figure 6C) for an illustration.

One could thus try to search for these kinds of partitions as well from the perspective of our dynamical framework. We postpone a exploration of these tasks for future work. However, see for instance Refs[70–74] for related discussions.

Appendix E: The (centered) dynamic similarity $\Psi(t)$ for diffusive processes

If one considers the particular case of diffusive dynamics, our dynamic similarity $\Psi(t)$ recovers and extends other methods proposed in the literature. We consider below first the undirected case as well as the directed case. However, as discussed in the main text, the dynamical similarity Ψ is applicable to general linear dynamics (including signed networks).

1. The undirected case

To put our approach in the context of diffusion processes, let us first consider an undirected dynamics of the form

$$\dot{\mathbf{p}} = -\mathbf{p}\, L, \tag{E1}$$

where $\mathbf{p}$ is the $1 \times n$ row vector describing the probability of a particle to be present at any node. Note that this diffusion is the dual of the consensus process:

$$\dot{\mathbf{x}} = -L\mathbf{x}, \tag{E2}$$

and indeed, in this case these two dynamics are in fact identical, as transposing Equation (E1) corresponds to a dynamics of the form (2) with $\mathcal{B} = \mathcal{C} = I$ and $\mathcal{A} = -L = -L^\top$, the combinatorial graph Laplacian.

As it is customary in the context of diffusion processes to deal with row vectors, and accordingly many results in the literature are presented in this form we will adopt this convention throughout this section. All these results can be readily transformed into a column vector setup (or in the directed case, may also be interpreted in the light of the dual consensus process).

Consider the random walk associated with the dynamics (E1) described by the row indicator vector $\mathbf{N}(t) \in \{0,1\}^n$, where $N_i(t) = 1$ if the walker is present at node i at time t and zero otherwise, and the $1 \times n$-dimensional vector $\mathbf{p}(t)$ describes the probability of the walker to be at each node at time t. It is well known that if the process takes places on an undirected (i.e., $L = L^\top$) connected graph, it is guaranteed to be wide sense stationary (in fact ergodic) and $\mathbf{p}(t)$ converges to the unique stationary

distribution

$$\boldsymbol{\pi} = \frac{\mathbf{1}^\top}{n},$$

irrespective of the initial condition.

To derive further results, it is insightful to compute the auto-covariance matrix of this process. Let us assume that we prepare the system at stationarity, $\mathbf{p}(0) \sim \boldsymbol{\pi}$ (i.e., the walker is equally likely to start at any node at time $t = 0$). The expectation of $\mathbf{N}(t)$ remains constant over time and we have

$$\mathbb{E}[\mathbf{N}(t)] = \mathbb{E}[\mathbf{N_0}]\, P(t) = \boldsymbol{\pi} P(t) = \boldsymbol{\pi},$$

where the transition matrix for this process is given by:

$$P(t) = \exp(-Lt).$$

The auto-covariance matrix of the process is then:

$$\Sigma(t) = \mathrm{cov}\left[\mathbf{N}(0)^\top, \mathbf{N}(t)\right] \tag{E3}$$

$$= \mathbb{E}[\mathbf{N_0}^\top \mathbf{N}(t)] - \mathbb{E}[\mathbf{N_0}^\top]\, \mathbb{E}[\mathbf{N_0}] \tag{E4}$$

$$= \Pi P(t) - \boldsymbol{\pi}^\top \boldsymbol{\pi}, \tag{E5}$$

where $\Pi = \mathrm{diag}(\boldsymbol{\pi})$. Defining $\Sigma_0 = \Pi - \boldsymbol{\pi}^\top \boldsymbol{\pi}$, we get that

$$\Sigma(t) = \Sigma_0\, P(t) = \Sigma_0 \exp(-Lt), \tag{E6}$$

and it becomes apparent that $\Sigma(t)$ is governed by the matrix differential equation:

$$\frac{d\Sigma}{dt} = -\Sigma L \quad \text{with} \quad \Sigma(0) = \Sigma_0. \tag{E7}$$

This autocovariance matrix $\Sigma(t)$ has been used as a dynamic similarity matrix in the Markov Stability framework for community detection[8,13,19].

Now, let us compare the autocovariance $\Sigma(t)$ with the dynamic similarity $\Psi(t)$. As shown in (9), when $\mathcal{B} = I$ the dynamic similarity $\Psi(t)$ obeys a Lyapunov matrix differential equation, which in this diffusive case is:

$$\frac{d\Psi}{dt} = -L^\top \Psi - \Psi L \quad \text{with} \quad \Psi(0) = \Psi_0. \tag{E8}$$

Without loss of generality we may pick $\Psi_0 = \Sigma_0$ as initial condition, so that the solution is given by

$$\Psi(t) = P(t)^\top \Sigma_0 P(t) = \exp(-L^\top t)\Sigma_0 \exp(-Lt), \tag{E9}$$

which is to be compared to (E6)

For the case of undirected graphs, we have $L = L^\top$ and $L\mathbf{1} = 0$, hence

$$\Sigma_0 = \Pi - \boldsymbol{\pi}^\top \boldsymbol{\pi} = \frac{1}{n}\left(I - \frac{\mathbf{1}\mathbf{1}^\top}{n}\right).$$

Therefore, we can interpret $\mathcal{W}_\perp = n\Sigma_0$ as a (scaled) projection matrix in Eq. (E9), and (E9) as a centered dynamic similarity:

$$\frac{1}{n}\Psi_\perp(t) = \exp(-Lt)\frac{1}{n}\left(I - \frac{\mathbf{1}\mathbf{1}^\top}{n}\right)\exp(-Lt). \tag{E10}$$

We can then rewrite $\Psi_\perp(t)$ to show the equivalence with $\Sigma(t)$ up to a simple rescaling:

$$\frac{1}{n}\Psi_\perp(t) = \frac{1}{n}\exp(-Lt)\left(I - \frac{\mathbf{11}^\top}{n}\right)\exp(-Lt)$$

$$= \frac{1}{n}\left(I - \frac{\mathbf{11}^\top}{n}\right)\exp(-Lt)\exp(-Lt)$$

$$= \frac{1}{n}\left(I - \frac{\mathbf{11}^\top}{n}\right)\exp(-L(2t)) = \Sigma_0 P(2t) = \Sigma(2t),$$

Hence for the case of diffusion on undirected graphs, the centered dynamic similarity $\Psi_\perp(t)$ is proportional to the autocovariance of the diffusion on a rescaled time.

Although we have exemplified this connection with a particular example, this result applies to any *time-reversible dynamics*. This includes all customary defined diffusion dynamics on undirected graphs like the continuous-time unbiased random walk, the combinatorial Laplacian random walk, or the maximum entropy random walk[75].

This result follows from the reversibility condition for a Markov process[76,77], also known as detailed balance:

$$\pi^{(i)}p_{i\to j} = \pi^{(j)}p_{j\to i} \quad \forall\, i,j, \tag{E11}$$

i.e., at stationarity, the probability to transition from state i to state j is the same as the probability to transitions from j to i (for any i,j). In matrix terms, the detailed balance condition is:

$$\Pi P(t) = P(t)^\top \Pi, \quad \forall t. \tag{E12}$$

Let us consider a centered dynamical similarity measure, in which we choose the weighting matrix

$$\mathcal{W}_\Pi = \Pi - \boldsymbol{\pi}^\top\boldsymbol{\pi},$$

which can be throught of as a generalization of the standard projection matrix $\mathcal{W}_\perp$. It is now easy to see that the analogous relationship between $\Psi_\Pi(t)$ and the autocovariance $\Sigma(t)$ holds also under detailed balance:

$$\Psi_\Pi(t) = P(t)^\top\left(\Pi - \boldsymbol{\pi}^\top\boldsymbol{\pi}\right) P(t)$$

$$= \left(\Pi P(t) - \boldsymbol{\pi}^\top\boldsymbol{\pi}P(t)\right) P(t)$$

$$= \left(\Pi - \boldsymbol{\pi}^\top\boldsymbol{\pi}\right) P(2t) = \Sigma(2t).$$

In the case of diffusive dynamics on undirected networks with detailed balance, we have shown that the dynamical similarity $\Psi_\Pi(t)$ is equivalent to the autocovariance $\Sigma(t)$ up to a rescaling. Therefore the Louvain-like analysis of $\Psi_\Pi(t)$ on the quality function $r_\Pi(t,H) = $ trace $H^\top\Psi_\Pi(t)H$ can be seen as a proper generalization of the Markov Stability framework, which optimizes the quality function $s(t,H) = $ trace $H^\top\Sigma(t)H$, and thus encompasses a wide array of notions of community detection including the classical Newman-Girvan Modularity[51], the self-loop adjusted modularity version of Arenas et al.[17], the Potts model heuristics of Reichardt

and Bornholdt[78], as well as Traag et al.[79], and classical spectral clustering[48]. For details, we refer the reader to the derivations given in Refs.[19,75] in terms of the Markov Stability measure.

Note, however, that Markov Stability only deals with diffusion dynamics, whereas both the centered dynamical similarity Ψ_Π (related to the quality function $r_\Pi(t,H)$), and the kernel Ψ can be applied to general linear models (including signed networks), as discussed in the main text. An interesting case occurs when considering diffusive processes on directed graphs, as discussed in the next section.

2. The directed case

For undirected diffusions, the autocovariance can be used as a dynamic similarity between nodes. However, there are important differences for directed (asymmetric) diffusive processes, as such processes are not guaranteed to be ergodic.

Let us consider a diffusion on a directed graph:

$$\dot{\mathbf{p}} = -\mathbf{p}L$$

so that $\mathcal{B} = \mathcal{C} = I$ and $-\mathcal{A} = L \neq L^\top$, an asymmetric Laplacian. This process is only ergodic if the graph is strongly connected. In many scenarios this is not the case, however, and the process asymptotically concentrates the probability on sink nodes with no outgoing links. Hence node similarities cannot be based on autocovariances at stationarity.

In order to study node similarities based on the diffusive dynamics, the original process is usually modified by adding a small 'teleportation' term (e.g., allowing for the process to diffuse to any node on the graph with a small probability)[13,14]. This approach is also known as the 'Google trick,' as it was popularized through its use in the original computation of Pagerank[80]. In its original form, the introduction of teleportation creates a related strongly connected graph by combining the original graph (with Laplacian L) together with the complete graph. This creates a related (yet different) ergodic process on this surrogate graph which can then be analyzed[75] via dynamic similarities based on autocovariances, as discussed in Section E 1. Specifically, the surrogate, ergodic system is defined by an adjusted Laplacian operator

$$\widetilde{L} = L + L_{\text{teleport}}.$$

Following an analogous calculation as above, the autocovariance $\Sigma(t)$ of this process:

$$\widetilde{\Sigma}(t) = \widetilde{\Pi}\exp(-\widetilde{L}t) - \widetilde{\boldsymbol{\pi}}^\top\widetilde{\boldsymbol{\pi}} = \left(\widetilde{\Pi} - \widetilde{\boldsymbol{\pi}}^\top\widetilde{\boldsymbol{\pi}}\right)\exp(-\widetilde{L}t),$$

where $\widetilde{\boldsymbol{\pi}}$ is the stationary distribution of the surrogate ergodic process (e.g., Page Rank). However, this autocovariance is now asymmetric, in general, and its interpretation as a similarity matrix is problematic.
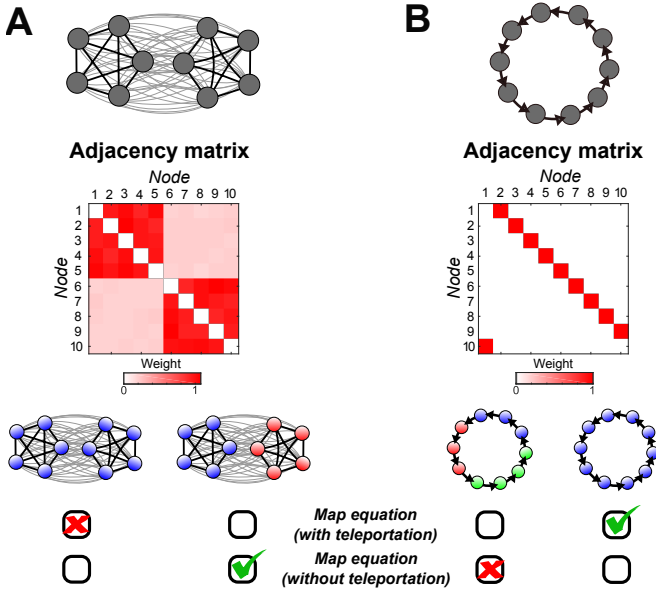
Figure 7. **Unpredictable influence of teleportation on clustering of directed graphs ($A \neq A^\top$).** When studying a directed diffusive dynamics (on a directed graph), a teleportation component is usually added to make the process ergodic. The addition of teleportation can lead to unexpected effects when clustering the network, since teleportation also influences the cut (i.e., the probability flow across group boundaries). For concreteness, we illustrate this effect here the map-equation[14] although this effect is general. **A** A directed network of two cliques. The weights between nodes inside each group are drawn from a uniform distribution with mean 1 ± 0.1; the weights across different groups are drawn from a uniform distribution with mean 0.2 ± 0.02. The network is directional but the asymmetry of A is so weak as to appear visually virtually undirected. The introduction of teleportation leads to a resolution limit effect in which the groups cannot be resolved. **B** A directed cycle network with equal weights. Without teleportation, a split of the ring into multiple groups is found, whereas the introduction of teleportation improves the result in this case so that the whole cycle is detected. Using an analysis based on Ψ without teleportation (E14) finds the dynamical blocks directly in both of these examples.

In contrast, we can use the Lyapunov equation (E8) to define the dynamic similarity (E9) of the ergodic system

$$\widetilde{\Psi}_{\widetilde{\Pi}}(t) = \exp(-\widetilde{L}^\top t)\left(\widetilde{\Pi} - \widetilde{\boldsymbol{\pi}}^\top \widetilde{\boldsymbol{\pi}}\right)\exp(-\widetilde{L}t), \qquad \text{(E13)}$$

where we have chosen the initial condition $\Psi_0 = \widetilde{\Pi} - \widetilde{\boldsymbol{\pi}}^\top \widetilde{\boldsymbol{\pi}}$. Note that Eq. (E13) may alternatively be constructed

from the dual process $\dot{\mathbf{x}} = \mathcal{A}\mathbf{x}$, with the operator $\mathcal{A} = -\widetilde{L}$. The similarity $\widetilde{\Psi}_{\widetilde{\Pi}}(t)$ may now be exploited directly to carry out embeddings, spectral clusterings, or Louvain-like block detection, as described in the Methods section. Note that the analysis based on the dynamic similarity $\widetilde{\Psi}_{\widetilde{\Pi}}(t)$ remains distinct to the symmetrized autocovariance $(\widetilde{\Sigma} + \widetilde{\Sigma}^\top)/2$ which is optimized when using a Louvain-like algorithm[75]. Other symmetrizations have been introduced in the context of transition matrices in directed graphs[81] generalizing Kleinberg's HITS scores[82].

The definition of the dynamic similarity of the associated ergodic process (E13) renders it consistent with our generic framework. However, the introduction of teleportation to create the surrogate process has conceptual disadvantages.

First, teleportation perturbs the dynamics in a non-local manner and induces uncontrolled effects when finding node similarities based on dynamics. In particular, it can reduce overclustering (a positive effect), but can also lead to (unwanted) resolution limits when finding dynamic blocks (Fig. 7). These issues are only at best mitigated by recent teleportation schemes[83].

Second, teleportation creates an ergodic, stationary dynamics when key features of the original system might be fundamentally linked to non-stationary data and non-ergodic processes. This can have a bearing on the conclusions drawn from the surrogate ergodic process with added teleportation.

We illustrate some of these problems in Figure 7 for the particular example of the map-equation. However, these issues are generic and effect other diffusion based clustering measures that are based on a notion of persistence of the flow within a region over time, and require a type of ergodicity assumption. For instance, similar effects will affect the Markov stability measure[13,84].

Importantly, our dynamic similarity $\Psi(t)$ can be directly applied to non-ergodic, directed graphs without the need to add teleportation (i.e., without creating the associated, but distinct stationary process). In this case, the dynamic similarity is

$$\Psi(t) = \exp(-L^\top t)\Sigma_0 \exp(-Lt), \qquad \text{(E14)}$$

which fulfils the Lyapunov equation (E8). The initial condition Σ_0 can be chosen to be any (covariance) matrix which serves as the null model for the process. The analysis of this dynamic similarity can reveal dynamic blocks based on directed flows from the *original* diffusive process, as shown in (Fig. 7). This directed case is another instance where the notion of 'dynamic block' generalizes the idea of modules, originally conceived from a structural perspective.