

Exploiting symmetry in network analysis

Rubén J. Sánchez García^{1*}

¹Mathematical Sciences, University of Southampton, University Road, Southampton SO17 1BJ, United Kingdom

*R.Sanchez-Garcia@soton.ac.uk

ABSTRACT

Virtually all network analyses involve structural measures between pairs of vertices, or of the vertices themselves, and the large amount of symmetry present in real-world complex networks is inherited by such measures. This has practical consequences which have not yet been explored in full generality, nor systematically exploited by network practitioners. Here we study the effect of network symmetry on arbitrary network measures, and show how this can be exploited in practice in a number of ways, from redundancy compression, to computational reduction. We also uncover the spectral signatures of symmetry for an arbitrary network measure such as the graph Laplacian. Computing network symmetries is very efficient in practice, and we test real-world examples up to several million nodes. Since network models are ubiquitous in the Applied Sciences, and typically contain a large degree of structural redundancy, our results are not only significant, but widely applicable.

Network models of real-world complex systems have been extremely successful at revealing structural and dynamical properties of these systems¹. The success of this approach is due to its simplicity, versatility, and surprising universality, with common properties and principles shared by many disparate systems^{2–4}.

One property of interest is the presence of structural redundancies, which manifest themselves as symmetries in a network model. Symmetries relate to system robustness^{5,6}, as they identify structurally equivalent nodes, and can arise from replicative growth processes such as duplication⁷, evolution from basic principles⁸, or functional optimisation⁹, and can be arbitrarily generated in model graphs¹⁰. It has been shown that real-world networks possess a large number of symmetries^{8,11–14}, and that this has important consequences for network structural¹¹, spectral¹³ and dynamical^{15–19} properties, for instance cluster synchronisation^{14,20–25}.

Crucially, network symmetries are inherited by any measure or metric on the network, that is, any structural measurement between pairs of vertices (such as distances), vertex-valued measurements (such as centrality) or even matrices derived from the network (such as the graph Laplacian). However, the effects of symmetry on arbitrary network measures is not yet fully understood nor exploited in network analysis, even though the network symmetry of the large but sparse graphs typically found in applications can be effectively computed and manipulated.

In this article, we show how a network representation of an arbitrary pairwise measure inherits the same symmetries of the original network, and uncovers the structural and spectral signatures of symmetry on this network representation. Namely, for an arbitrary network measure, we identify sub-graphs where the symmetry is generated (symmetric motifs) and their structure, use the network quotient to quantify the redundancy due to symmetry, develop general compression algorithms that eliminate this redundancy, and study the reduction in computational time obtained by exploiting the presence of symmetries. The eigenvalues and eigenvectors of a

network measure also reflect the presence of symmetry: we show how symmetry explains most of the discrete spectrum of an arbitrary network measure, predict the most significant eigenvalues due to symmetry, and use this to develop a fast symmetry-based eigendecomposition algorithm. We achieve remarkable empirical results in our real-world test networks: compression factors up to 26% of the original size, over 90% of the discrete spectrum explained by symmetry, and full eigendecomposition computations in up to 13% of the original time, demonstrating the practical use of symmetry in network analysis. We also discuss the implications of network symmetry in vertex measures. We illustrate our approach in several network measures, providing novel results of independent interest for the shortest path distance, communicability, the graph Laplacian, closeness centrality and eigenvector centrality. To facilitate dissemination, we provide full implementations of all the algorithms described in this article²⁶. Our results supersede^{11,13} and help to understand other network symmetry results thereafter^{12,27–31}. We focus on structural and spectral properties, and symmetries commonly found in real-world networks: For a more general study of arbitrary symmetry in (networks of) dynamical systems, see^{15–18}. To keep our account as self-contained as possible, we include material well known in the algebraic graph theory literature e.g.^{32–35}, without any originality claim.

Results

Symmetry in complex networks

The notion of network symmetry is captured by the mathematical concept of *graph automorphism*³². This is a permutation of the vertices (nodes) preserving adjacency, and can be expressed in matrix form using the adjacency matrix of the network. If a network (mathematically, a finite simple graph) \mathcal{G} has n vertices, labelled 1 to n , its *adjacency matrix* $A = (a_{ij})$ is an $n \times n$ matrix with (i, j) -entry $a_{ij} = 1$ if there is an edge between nodes i and j , and zero otherwise. A graph automorphism σ is then a permutation, or relabelling, of the

vertices $v \mapsto \sigma(v)$ such that $(\sigma(i), \sigma(j))$ is an edge only if (i, j) is an edge, or, equivalently, $a_{ij} = a_{\sigma(i)\sigma(j)}$ for all i, j . In matrix terms, this can be written as

$$AP = PA, \quad (1)$$

where P is the permutation matrix corresponding to σ , that is, the matrix with (i, j) -entry 1 if $\sigma(i) = j$, and 0 otherwise. The automorphisms of a graph form a mathematical structure called a group, the *automorphism group* of \mathcal{G} . In principle, any (finite) group G is the automorphism group of some graph \mathcal{G} ³², but, in practice, real-world networks exhibit very specific types of symmetries generated at some small subgraphs called *symmetric motifs*¹¹. Namely, we can partition the vertex set into the *asymmetric core* of fixed points V_0 (an automorphism σ moves a vertex $i \in V$ if $\sigma(i) \neq i$, and fixes it otherwise), and the vertex sets M_i of the symmetric motifs,

$$V = V_0 \cup M_1 \cup \dots \cup M_m, \quad (2)$$

as shown in Fig. 1a for a toy example. Equation (2) is called the *geometric decomposition* of the network¹¹.

Real-world networks typically exhibit a core of fixed points (*asymmetric core*), and a large number of relatively small symmetric motifs, where all the network symmetry is generated, and hence the size of the automorphism group is often extremely large, in stark contrast to random graphs, typically asymmetric³⁸. However, each symmetry is the product (composition) of automorphisms permuting a very small number of vertices within a symmetric motif. For example, the toy graph in Fig. 1a has $2^7 \times 3! \times 4! = 18,432$ symmetries (size of the automorphism group) but they generated by (all combinations of) just 10 permutations, each permuting a few vertices within a symmetric motif (one permutation per motif except two for M_4 , M_5 and M_7).

Each symmetric motif can be further subdivided into orbits of structurally indistinguishable nodes (shown by colour in Fig. 1a), which play the same structural role in the network and, therefore, contribute to network redundancy and thus to the robustness of the underlying system. Our notion of structurally indistinguishable nodes (nodes in the same orbit of the automorphism group) extends the notion of structurally equivalent nodes found in the social sciences³⁹, that is, nodes with the same set of neighbours. It is not equivalent: nodes in the same orbit may not have the same neighbours (e.g. M_1 , M_6 or M_7 in Fig. 1a).

Network symmetries of (possibly very large) real-world networks can be effectively computed, stored and manipulated (see Methods). For instance, we computed generators of the automorphism group, and the subsequent geometric decomposition, for real-world networks up to several million nodes and edges in a few seconds (see t_1 and t_2 in Table 1).

Most symmetric motifs in real-world networks (typically over 90%, see the *bsm* column in Table 1) are of a very specific type, called *basic*¹¹: they are made of one or more orbits of the same size, and every permutation of the vertices in each orbit is realisable, that is, can be extended to a network

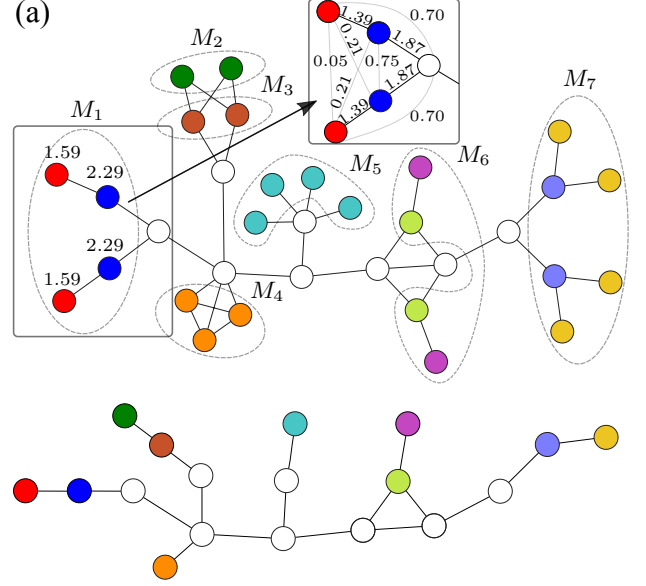


Figure 1. Toy example of a symmetric network. (a)

Decomposition into asymmetric core (white nodes) and 7 symmetric motifs M_1 to M_7 . Each motif can be decomposed into one or more orbits (shown by colour) of structurally indistinguishable nodes. Motifs M_1 to M_6 are *basic*: one or more orbits of the same size, and every permutation of the vertices in an orbit can be extended to a network automorphism. Motif M_7 , a tree, is not basic: in the orbit of yellow vertices, labelled 1 to 4 (top to bottom), the permutation sending 1 to 3 must move 2 to 4 (as 1 and 2 share a common neighbour). A vertex measure (here subgraph centrality³⁶ shown on M_1) is constant on orbits. Inset: A pairwise measure (here communicability³⁷ shown near M_1) inherits all the network symmetries. (b) Quotient network (no loops, edge directions, or weights shown) consisting on one vertex per fixed point (white node) and per orbit (coloured node).

automorphism (see Fig. 1). Basic symmetric motifs (BSMs) have a very constrained structure¹³, which we will generalise to arbitrary network measures and exploit throughout this article. Non-basic symmetric motifs (typically branched trees, as M_7 in Fig. 1) are called *complex*; they are rare and can either be studied on a case-by-case basis, or removed from the symmetry computation altogether (by ignoring the symmetries generated by them).

The definition of network automorphism Eq. (1) carries to an arbitrary $n \times n$ real matrix $A = (a_{ij})$. Any such matrix can be seen as the adjacency matrix of a network with n vertices labelled 1 to n , and an edge (link) from node i to node j with weight a_{ij} if $a_{ij} \neq 0$, and no such edge if $a_{ij} = 0$. This means that an automorphism does not only preserve edges, but also their weights and directions. This may not be a realistic assumption for real-world weighted networks, where the weights often come from observational or experimental data, but it applies to the matrix representing a network structural measure, as we illustrate in Fig. 2 and explain next.

Name	n_g	m_g	gen	t_1	t_2	sm	bsm	mv	\tilde{n}_g	\tilde{m}_g	ext_s	int_f	c_{full}	sp
HumanDisease	1,419	2,738	713	0.00	0.16	272	96.0	71.0	48.3	50.4	83.3	10^{-3}	27.2	11.3
Yeast	1,647	2,736	380	0.00	0.01	149	99.3	33.3	76.3	83.5	98.8	10^{-3}	58.4	44.4
OpenFlights	3,397	19,230	732	0.00	0.11	321	93.5	32.4	77.3	94.4	99.3	10^{-3}	63.5	46.2
USPowerGrid	4,941	6,594	414	0.00	0.09	302	97.4	16.7	90.2	91.3	97.6	10^{-4}	83.9	73.3
HumanPPI	9,270	36,918	972	0.00	0.12	437	100	15.3	89.5	97.0	99.9	10^{-4}	80.1	71.6
Astro-Ph	17,903	196,972	3,232	0.01	0.21	1,682	99.4	27.5	81.9	80.4	95.5	10^{-4}	67.4	54.9
InternetAS	34,761	107,720	15,587	0.03	0.29	3,189	99.9	54.3	55.0	78.2	99.9	10^{-5}	30.3	16.7
WordNet	145,145	656,230	52,152	0.18	0.62	28,456	92.0	60.0	60.1	58.0	89.9	10^{-5}	49.3	21.6
Amazon	334,863	925,872	32,098	0.20	0.39	23,302	99.8	16.8	90.3	89.0	99.0	10^{-6}	81.6	73.6
Actors	374,511	15,014,839	182,803	0.95	1.38	36,703	99.9	58.6	51.2	66.4	90.4	10^{-5}	26.2	13.4
InternetAS-skitter	1,694,616	11,094,209	319,738	1.71	4.17	84,675	99.1	19.7	85.4	92.8	99.9	10^{-6}	73.5	62.3
CaliforniaRoads	1,957,027	2,760,388	36,430	0.47	0.16	35,210	98.8	4.0	97.9	98.4	99.7	10^{-7}	96.3	93.9
LiveJournal	5,189,808	48,687,945	410,575	8.02	3.59	245,211	99.9	12.7	92.1	96.5	99.7	10^{-7}	84.8	78.0

Table 1. Symmetry in some real-world networks. For each test network, we show the number of vertices (n_g), edges (m_g), number of generators (gen) of the automorphism group (sizes, 10^{153} to $10^{197,552}$, not shown), computing times of generators (t_1) and geometric decomposition (t_2), in seconds, number of symmetric motifs (sm) and proportion of basic symmetric motifs (bsm), proportion of vertices moved by an automorphism (mv), proportion of vertices ($\tilde{n}_g = n_g/n_g$) and edges ($\tilde{m}_g = m_g/m_g$) in the quotient, proportion of external edges in the sparse case (ext_s , in percentage), and of internal edges in the full case (int_f , closest power of 10), full compression ratio ($c_{full} = \tilde{n}_g^2$), and spectral computational reduction ($sp = \tilde{n}_g^3$), all for the largest connected component. The proportion of vertices in the *basic* quotient ($\tilde{n}_{g_{basic}}$, not shown) is within 1% of \tilde{n}_g except for HumanDisease ($\tilde{n}_{g_{basic}} = 52.2\%$), OpenFlights (79.7%), USPowerGrid (91.6%) and WordNet (79.2%), and similar results hold for $\tilde{m}_{g_{basic}}$. Datasets available at⁴⁰, except HumanDisease⁴¹, Yeast⁴², and HumanPPI⁴³. Computations on a desktop computer (3.2 GHz Intel Core i5 processor, 16 GB 1.6 GHz DDR3 memory). All networks are symmetric, although the amount of symmetry (as measured by mv or \tilde{n}_g) ranges from several networks with 50% quotient reduction, to CaliforniaRoads with only 4% of vertices participating in any symmetry. However, the effect of compression and computational reduction multiplies as e.g. $c_{full} = \tilde{n}_g^2$ and $sp = \tilde{n}_g^3$, achieving significant results for most of our test networks.

Structural network measures

A (pairwise) structural network measure is a function $F(i, j)$ on pairs of vertices which satisfies

$$F(\sigma(i), \sigma(j)) = F(i, j) \text{ for all } i, j \in V \quad (3)$$

for all automorphisms $\sigma \in \text{Aut}(\mathcal{G})$. Since automorphisms identify structurally indistinguishable vertices (i and $\sigma(i)$) and, similarly, edges ((i, j) and $(\sigma(i), \sigma(j))$), structural network measures are (edge) functions that depend on the network structure alone, and not, for example, on node or edge labels, or other meta-data. Most network measures are structural, including graph metrics (e.g. shortest path), and matrices algebraically derived from the adjacency matrix (e.g. Laplacian matrix). (We identify matrices M with pairwise measures via $F(i, j) = [M]_{ij}$.) In particular, structural measures are independent of the ordering or labelling of the vertices. In contrast, functions depending, explicitly or implicitly, on some vertex ordering or labelling, are not structural, for example the shortest path length through a given node. Our results can be adapted to the presence of node or edge labels, or weights, by restricting to automorphisms preserving the additional structure. For simplicity, here we discuss the unlabelled case only.

We can encode a structural measure F as a network with adjacency matrix $[F(A)]_{ij} = F(i, j)$ (see Fig. 2b and c for two examples), and write (3) in matrix form as

$$F(A)P = PF(A), \quad (4)$$

where P is the permutation matrix corresponding to σ . Comparing this to Eq. (1), we see that the network representation

of F , $F(\mathcal{G})$, with adjacency matrix $F(A)$, inherits all the symmetries of \mathcal{G} . In particular, the network $F(\mathcal{G})$ has the same decomposition into symmetric motifs Eq. (2), and orbits, as \mathcal{G} . The BSMs in $F(\mathcal{G})$ must occur on the same vertices M_i , although they are now all-to-all weighted subgraphs in general (Fig. 2b). Nevertheless, they have a very constrained structure: the intra and inter orbit connectivity depends on two parameters only. Namely, each orbit in a BSM is uniquely determined by $\beta = F(v_i, v_i)$ (the connectivity of a vertex with itself) and $\alpha = F(v_i, v_j)$, $i \neq j$ (the connectivity of a vertex with every other vertex in the orbit), for all v_i, v_j in the orbit. Similarly, the connectivity between two orbits Δ_1 and Δ_2 in the same BSM also depends on two parameters: after a suitable reordering $\Delta_1 = \{v_1, \dots, v_n\}$ and $\Delta_2 = \{w_1, \dots, w_n\}$, we have $\delta = F(v_i, w_i)$ and $\gamma = F(v_i, w_j)$ for all $1 \leq i, j \leq n$. (For a proof, see Theorem 1 in Methods.) This can be observed in Fig. 2c and is represented schematically in Fig. 3a and b. In particular, each BSM takes a very constrained form in the quotient, as shown schematically in Fig. 3c and d.

The results in this article apply to arbitrary structural measures, although the two most common cases in practice are the following. We call F *full* if $F(i, j) \neq 0$ for all $i \neq j \in V$ (e.g. a graph metric), and *sparse* if $F(i, j) = 0$ if $a_{ij} = 0$, for all $i \neq j \in V$ (e.g. the graph Laplacian). The graph representation of $F(\mathcal{G})$ is an all-to-all weighted graph if F is full, and has a sparsity similar to \mathcal{G} if F is sparse (cf. Fig. 2c).

From now on, we will assume that \mathcal{G} is undirected and F is symmetric, $F(i, j) = F(j, i)$, which may not be the case even

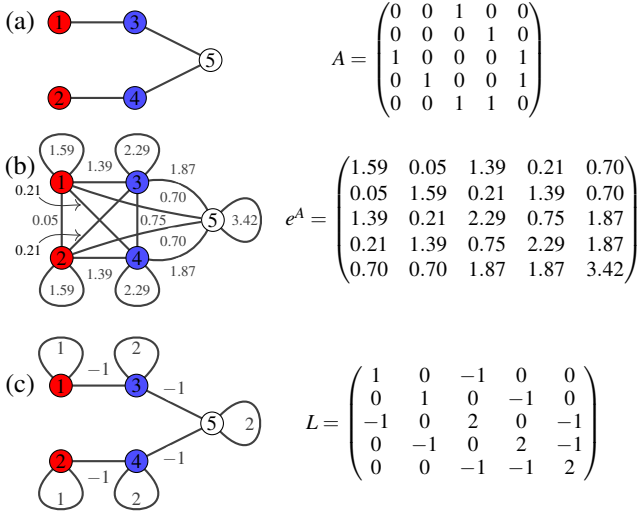


Figure 2. Structural network measures. (a) Neighbourhood of motif M_1 (Fig. 1) and its adjacency matrix. Network and matrix representation of two structural (full, respectively sparse) network measures, (b) the exponential matrix (a measure of vertex communicability³⁷), and (c) the graph Laplacian. In each case, the vertices are numbered 1 to 5 and coloured by orbit, and a non-zero (i, j) -entry in the matrix corresponds to the edge weight between vertices i and j . Note that the network automorphism swapping vertices 1 and 2, and 3 and 4, preserves the network and matrix structure in all three cases. In general, any symmetry of a network is also a symmetry of the network representation of any structural network measure.

if \mathcal{G} is undirected (e.g. the transition probability of a random walker $F(i, j) = \frac{a_{ij}}{\deg(i)}$), and discuss directed networks and asymmetric measures in the Methods section.

Quotient network

The formal procedure to quantify and eliminate structural redundancies in a network is via its *quotient network*. This is the graph with one vertex per orbit or fixed point (see Fig. 1b) and edges representing average connectivity. Formally, if A is the $n \times n$ adjacency matrix of a graph \mathcal{G} , the *quotient network* with respect to a partition of the vertex set $V = V_1 \cup \dots \cup V_m$ is the graph \mathcal{Q} with $m \times m$ adjacency matrix the *quotient matrix* $Q(A) = (b_{kl})$ defined by

$$b_{kl} = \frac{1}{|V_k|} \sum_{i \in V_k} \sum_{j \in V_l} a_{ij}, \quad (5)$$

the average connectivity from a vertex in V_k to all vertices in V_l . There is an explicit matrix equation for the quotient. Consider the $n \times m$ *characteristic matrix* S of the partition, that is, $[S]_{ik} = 1$ if $i \in V_k$, and zero otherwise, and the diagonal matrix $\Lambda = \text{diag}(n_1, \dots, n_m)$, where $n_k = |V_k|$. Then

$$Q(A) = \Lambda^{-1} S^T A S. \quad (6)$$

The quotient network is a directed and weighted network in general. An alternative is to use the *symmetric quotient*,

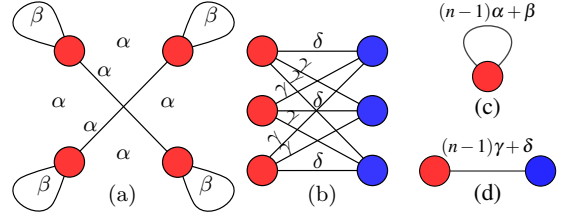


Figure 3. Structure of a basic symmetry motif (BSM) for an arbitrary network measure F . (a) Every orbit in a

BSM is an (α, β) -uniform graph $K_n^{\alpha, \beta}$, the graph with n vertices and adjacency matrix $A = (a_{ij})$ with $a_{ij} = \alpha = F(i, j)$ if $i \neq j$ and $a_{ii} = \beta = F(i, i)$ for some constants α and β . Here we show an example of one orbit with four nodes and edges labelled by their weights. (b) The connectivity between two orbits Δ_1 and Δ_2 in the same BSM (after a suitable relabelling $\Delta_1 = \{v_1, \dots, v_n\}$, $\Delta_2 = \{w_1, \dots, w_n\}$) is given by $\gamma = F(v_i, w_j)$ for $i \neq j$, and $\delta = F(v_i, w_i)$, the (δ, γ) -uniform join of the two orbits. Here we show an example of two orbits (shown by colour) with three vertices each and edges labelled by their weights. (c) In the quotient, the BSM orbit becomes a single vertex with a self-loop weighted by $(n-1)\alpha + \beta$, and (d) the two orbits are joined by an edge weighted by $(n-1)\gamma + \delta$. Here we show the quotients (c) and (d) of the previous BSMs (a) and (b) respectively. Note that, by annotating each orbit in the quotient by n and α (or β), and each intra-motif edge by γ (or δ), we can recover each BSM from such annotated quotient.

with adjacency matrix $Q_{\text{sym}}(A) = \Lambda^{-1/2} S^T A S \Lambda^{-1/2}$, which is weighted but undirected. Note that $Q(A)$ and $Q_{\text{sym}}(A)$ are spectrally equivalent matrices: they have the same eigenvalues, with eigenvectors related by the transformation $\mathbf{v} \mapsto \Lambda^{1/2} \mathbf{w}$.

In the context of symmetries, we will always refer to the quotient with respect to the partition of the vertex set into orbits. This quotient removes all the original symmetries from the network: if $\sigma(v_i) = v_j$, then v_i and v_j are in the same orbit and hence represented by the same vertex in the quotient network, which is then fixed by σ . We can, therefore, infer and quantify properties arising from redundancy alone by comparing a network to its quotient. The quotients of real-world networks are often significantly smaller (in vertex and edge size) than the original networks^{11,12} (see $\tilde{n}_{\mathcal{Q}}$ and $\tilde{m}_{\mathcal{Q}}$ in Table 1), and this reduction quantifies the structural redundancy present in an empirical network. Not every real-world network is equally symmetric, and, in our test networks, we give examples of network quotient reductions ranging from about 50% to just 2%. Computing the network quotient involves multiplication by very sparse matrices (Λ is diagonal and S has one non-zero element per row) and hence is computationally efficient (a few seconds in all our test networks).

Redundancy in network measures

The amount structural redundancy on a network (measured by $\tilde{n}_{\mathcal{Q}} = n_{\mathcal{Q}}/n_{\mathcal{Q}}$) is amplified in the computation of a typical (full) network measure (see Eq. (7) below). It is therefore

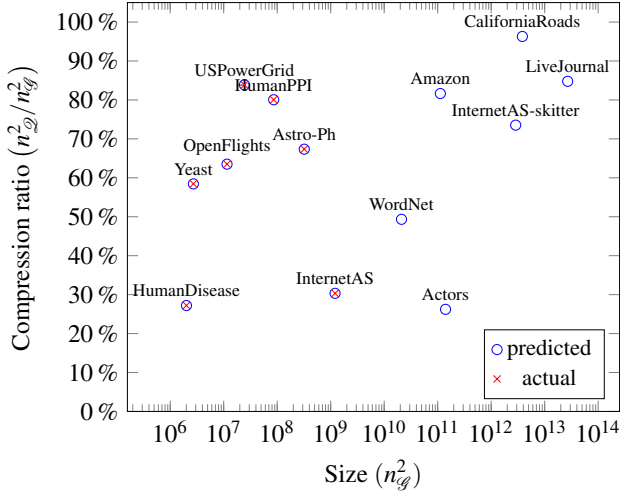


Figure 4. Redundancy in some real-world networks.

Amount of symmetry present on a typical (full) network measure for our test networks (Table 1). The predicted values, $c_{\text{full}} = (n_g/n_g)^2$ coincide (up to 0.01%) with the compression ratio (shown as ‘actual’) obtained by our lossless compression algorithm for the shortest path distance, and communicability (exponential matrix), for our smallest seven networks (memory limit in our computer). After decompression, we recover the original matrix exactly for the shortest path distance, and up to a small numerical error (1.16×10^{-4} mean relative error) for communicability.

natural to ask how to quantify, and eliminate, the symmetry-induced redundancy. If a network has n_g vertices and n_g orbits, there are n_g^2 pairs of vertices but only n_g pairs of orbits, achieving a reduction, or compression ratio, of

$$c_{\text{full}} = \left(\frac{n_g}{n_g} \right)^2 \quad (7)$$

for a full network measure, typically much smaller than the ratio $\tilde{n}_g = n_g/n_g$. On the other hand, for a sparse network measure, we only need to consider edge values, hence the reduction is the ratio between the number of edges in the graph and in its quotient

$$c_{\text{sparse}} = \frac{m_g}{m_g}. \quad (8)$$

For an arbitrary network measure, its compression ratio, which measures the redundancy present (zero values excluded), will range between c_{full} and c_{sparse} . The compression ratios c_{full} and $c_{\text{sparse}} = \tilde{m}_g$ are shown on Table 1 for our test networks. We found a remarkable amount of redundancy (up to 70%) due to symmetry alone (Fig. 4).

Symmetry compression

A natural question, with practical consequences for network analysis, is whether we can easily ‘eliminate’ the symmetry-induced redundancies. This means storing only one value of a network function for each orbit of structurally indistinguishable nodes or edges, all sharing the same such value.

Although this has been explored in particular cases, such as shortest path distances²⁷, here we present a general treatment. A simple method is to use the quotient matrix

$$B = S^T A S, \quad (9)$$

which is easier to store than $\Lambda^{-1} S^T A S$. This matrix achieves a compression ratio between c_{full} and c_{sparse} (by using a sparse representation of B), as explained before. From this matrix, we can recover all but the internal connectivity inside a symmetric motif, which is replaced by the average connectivity. Namely, let us define

$$\bar{a}_{ij} = \frac{1}{n_i} \frac{1}{n_j} b_{kl}, \quad (10)$$

where n_i , respectively n_j , is the size of the orbit containing v_i , respectively v_j (note that these orbit sizes can be obtained as the row sums of the characteristic matrix S). Then one can show (Methods, Theorem 2) that

$$\bar{a}_{ij} = \begin{cases} a_{ij} & \text{if } v_i \text{ and } v_j \text{ are external,} \\ \frac{1}{n_i} \frac{1}{n_j} \sum_{v_k \in \Delta_1} a_{kl} & \text{if } v_i \text{ and } v_j \text{ are internal,} \end{cases} \quad (11)$$

where we call a pair of vertices *external* if they belong to two different symmetric motifs, and *internal* otherwise, and $v_i \in \Delta_1$ and $v_l \in \Delta_l$ are orbits. Hence, if we are not interested in the exact internal connectivity (inside a symmetric motif), or it can be recovered easily by other means (e.g. one motif at a time), we can use this simple method to eliminate all the symmetry-induced redundancies on an arbitrary network measure encoded as a matrix A . We have included simple *average symmetry compression* and *decompression* algorithms (Algs. 1, 2), where A_{avg} is the matrix with entries \bar{a}_{ij} . The original $n_g \times n_g$ matrix A is stored using the $n_g \times n_g$ quotient matrix B plus a very sparse (n non-zero elements) characteristic matrix S .

Input: adjacency matrix A , characteristic matrix S
Output: quotient matrix B
 $B \leftarrow S^T A S$

Algorithm 1: Average symmetry compression.

Input: quotient matrix B , characteristic matrix S
Output: adjacency matrix A_{avg}
 $\Lambda \leftarrow \text{diag}(\text{sum}(S))$
 $R \leftarrow S \Lambda^{-1}$
 $A_{\text{avg}} \leftarrow R B R^T$

Algorithm 2: Average symmetry decompression.

The vast majority of edges in the network representation of a network measure are external (at least 99.999% for a full measure in our test networks, see *int_f* in Table 1), and hence the information loss by using A_{avg} instead of A is minimal. We can nevertheless enforce *lossless compression*, by storing the intra-motif connectivity separately. Indeed, we can exploit

the fact that most symmetric motifs in empirical networks are basic, and hence each orbit, or pair of orbits, is uniquely determined by two parameters (Fig. 3). If we disregard the symmetries generated at non-basic symmetric motifs, the corresponding quotient, called *basic quotient*, written $\mathcal{Q}_{\text{basic}}$, leaves non-basic motifs unchanged and retains most of the symmetry in a typical real-world network. By annotating this quotient, we can recover the original network representation of the network measure exactly. We have implemented lossless compression and decompression algorithms (Methods, Algs. 6 and 7), and evaluated them in our test networks (Fig. 4).

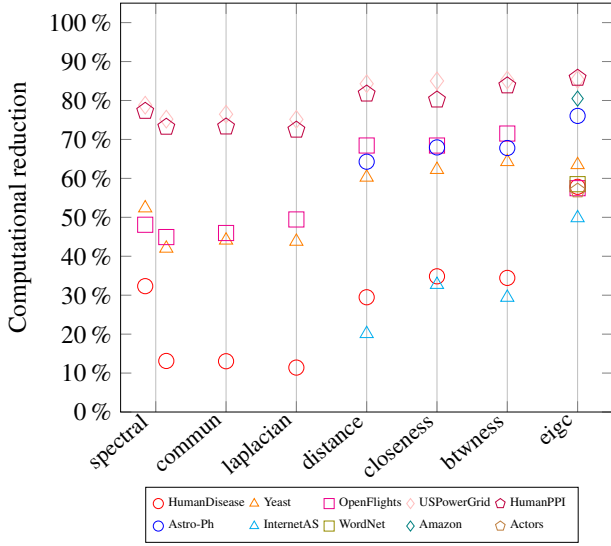


Figure 5. Quotient computational reduction.

Computational time reduction of several structural measures in some of our test networks (Table 1) obtained by performing the calculation in the quotient network versus the original network. The computations are: spectral decomposition of the adjacency matrix A (spectral), exponential matrix $\exp(A)$ (commun), pseudoinverse of the Laplacian matrix (laplacian), shortest path distance (distance), closeness centrality (closeness), betweenness centrality (btwness) and eigenvector centrality (eigc), using MATLAB R2018a built-in functions. For spectral, we also show (left column) the reduction including the (sequential) symmetric motif calculation. In each case, median computational reduction over at least 10 iterations shown.

Computational reduction

Network symmetries can also reduce the computational time of evaluating an arbitrary network measure F . By Eq. (3), we only need to evaluate F on orbits, resulting in a computational reduction ratio of between $\tilde{m}_{\mathcal{Q}}$ and $\tilde{n}_{\mathcal{Q}}^2$ (Table 1) for sparse, respectively full, network measures. Of course, this assumes that the computation on each pair of vertices $F(i, j)$ is independent of one another, which is often not the case. Moreover, the calculation of $F(i, j)$ is still performed on the whole network \mathcal{G} .

A more substantial computational reduction can be obtained by evaluating F on the (often much smaller) quotient network

instead. We call F *quotient recoverable* if it can be applied to the quotient network \mathcal{Q} , and $F(\mathcal{G})$ can be recovered from $F(\mathcal{Q})$, for all networks \mathcal{G} . Note that this may involve, beyond evaluating $F(\mathcal{Q})$, an independent (hence parallelizable) computation on each symmetric motif (typically a very small graph). By evaluating F in the quotient network, we can obtain very substantial computational time savings, depending on the amount of symmetry present and the computational complexity of F . Depending on the network measure, it may not be possible to recover $F(\mathcal{G})$ exactly from $F(\mathcal{Q})$, but only partially. We call a network measure F *partially quotient recoverable* if it can be applied to a quotient network \mathcal{Q} of a network \mathcal{G} , and all the external edges of $F(\mathcal{G})$ can be recovered from $F(\mathcal{Q})$, for all networks \mathcal{G} . Since the quotient averages the network connectivity, we can often recover the average values of F within symmetric motifs. We call F *average quotient recoverable* if, in addition to external edges, the average intra-motif edges can be recovered from $F(\mathcal{Q})$. A typical situation is when $F(\mathcal{Q})$ equals the quotient of F , that is, in symbols,

$$F(\mathcal{Q}) = Q(F(\mathcal{G})). \quad (12)$$

In the Applications section, we will show that communicability is average quotient recoverable, and shortest path distance is partially, but not average, quotient recoverable. Not every measure can be (partially) recovered from the quotient, for example the number of distinct paths between two vertices, as the internal connectivity within each symmetric motif is lost, and replaced by its average connectivity, in the quotient. Note that the word ‘partially’ can be misleading: typically almost all edges are external (see ext_s and int_f in Table 1). The resulting computational time reduction obtained by evaluating F in the quotient can be very substantial, as illustrated by several popular network measures in our test networks (Fig. 5).

Spectral signatures of symmetry

The spectrum of the network’s adjacency matrix relates to a multitude of structural and dynamical properties¹. The presence of symmetries is reflected in the spectrum of the network¹³, and indeed in the spectrum of any network measure. Symmetries give rise to high-multiplicity eigenvalues (shown as ‘peaks’ in the spectral density) and, in fact, we can explain and predict most of the discrete part of the spectrum of an arbitrary network measure on a typical real-world network.

Let A be the $n \times n$ adjacency matrix of a (possibly weighted) network (such as the network representation of a network measure). First, note that symmetry naturally produces high-multiplicity eigenvalues, since

$$AP\mathbf{v} = PA\mathbf{v} = \lambda P\mathbf{v}. \quad (13)$$

where (λ, \mathbf{v}) is an eigenpair of A and P the permutation matrix of a network automorphism (Eq. (1)). This gives another eigenpair (λ, \mathbf{v}) whenever \mathbf{v} and $P\mathbf{v}$ are linearly independent (obviously not always the case).

Let $B = Q(A)$ be the $m \times m$ quotient of A (Eq. (6)) with respect to the partition of the vertex set into orbits. This partition satisfies a regularity condition called *equitability*³⁵, which can be written in matrix form as $AS = SB$, where S is the characteristic matrix of the partition. In particular, if (λ, \mathbf{v}) is a quotient eigenpair, then $(\lambda, S\mathbf{v})$ is a parent eigenpair,

$$A(S\mathbf{v}) = SB\mathbf{v} = \lambda(S\mathbf{v}). \quad (14)$$

In fact, one can show (Methods, Theorem 3) that A has an eigenbasis of the form

$$\{S\mathbf{v}_1, \dots, S\mathbf{v}_m, \mathbf{w}_1, \dots, \mathbf{w}_{n-m}\}, \quad (15)$$

where $\{\mathbf{v}_1, \dots, \mathbf{v}_m\}$ is any eigenbasis of B , and $S^T \mathbf{w}_j = 0$ for all j . We can think of a vector $\mathbf{v} \in \mathbb{R}^m$, respectively $\mathbf{w} \in \mathbb{R}^n$, as a vector on (the vertices of) the quotient, respectively the parent, network. Then, each vector $S\mathbf{v}_i$ equals the vector \mathbf{v}_i *lifted* to the parent network by repeating the value on each orbit. Similarly, $S^T \mathbf{w}_j = 0$ means that the sum of the entries of \mathbf{w}_j on each orbit is 0. All in all, we can always find an eigenbasis of A consisting of *non-redundant* eigenvectors $\{S\mathbf{v}_1, \dots, S\mathbf{v}_m\}$ arising from a quotient eigenbasis by repeating values on each orbit, and *redundant* eigenvectors $\{\mathbf{w}_1, \dots, \mathbf{w}_{n-m}\}$ arising from the network symmetries, which add up to zero on each orbit (hence ‘disappearing’ in the quotient). Similarly, we call their respective eigenvalues *redundant* and *non-redundant*.

Analogous to the way that symmetry is generated at symmetric motifs, the redundant eigenvectors and eigenvalues arise directly from certain eigenvectors and eigenvalues of the symmetric motifs, considered as networks on their own (Fig. 6). In fact, each symmetric motif \mathcal{M} contributes the same (called *redundant*) eigenpairs to *any* network containing \mathcal{M} as a symmetric motif: One can show (Methods, Theorem 4) that if \mathcal{M} is a symmetric motif of a network \mathcal{G} and (λ, \mathbf{w}) is a redundant eigenpair of \mathcal{M} (that is, the values of \mathbf{w} add up to zero on each orbit of \mathcal{M}), then $(\lambda, \tilde{\mathbf{w}})$ is an eigenpair of \mathcal{G} , where $\tilde{\mathbf{w}}$ is equal to \mathbf{w} on (the vertices of) \mathcal{M} , and zero elsewhere. We call such a vector $\tilde{\mathbf{w}}$ *localised* on the motif \mathcal{M} ¹³, as it is zero outside the motif. Moreover, if \mathcal{M} has n vertices and k orbits, then it has an eigenbasis consisting of $n - k$ redundant eigenpairs, which are inherited by any network containing \mathcal{M} as a symmetric motif (Fig. 6, Theorem 4 in Methods).

Furthermore, since most symmetric motifs in real-world networks are basic, thus have a very constrained structure (Fig. 3), we can in fact determine the redundant spectrum of BSMs with up to a few orbits, that is, we can predict where the most significant ‘peaks’ in the spectral density of an arbitrary network function will occur. The formulae for the redundant spectra for BSMs of one or two orbits (which covers most BSMs, up to 99% of them in our test networks) is given on Table 2. For example, for the graph Laplacian, symmetry explains between 89% and 97% of the discrete spectrum in our test networks, with most of the high-multiplicity eigenvalues (‘peaks’ in the eigenvalue histograms) occurring at the predicted values (resulting from our analysis of the spectrum

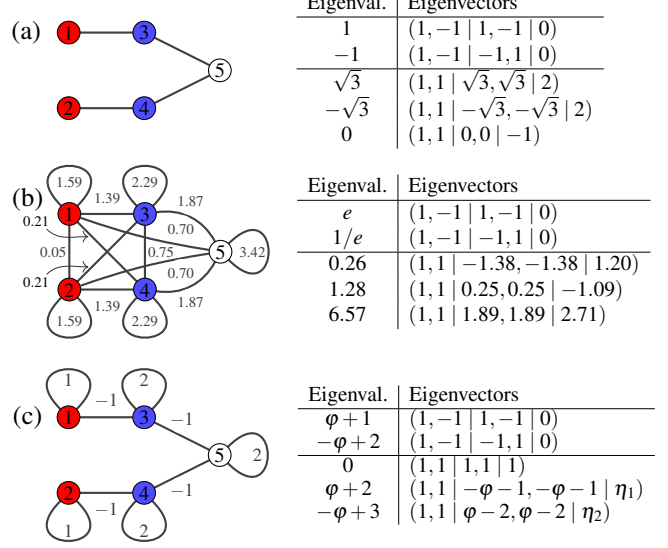


Figure 6. Redundant spectrum. Eigenvalues and eigenvectors near the symmetric motif M_1 (Fig. 1 (inset)) for three network measures: (a) adjacency A , (b) communicability e^A , and (c) Laplacian L , represented as weighted networks with weights shown as edge labels (as in Fig. 2). Eigenvector coordinates are separated by orbit, for convenience. In all three cases, there are two redundant, and three non-redundant, eigenpairs (separated here by a horizontal line), whose eigenvectors add up to zero, respectively are constant, on each orbit. Redundant eigenvectors are localised (zero outside the motif) and ‘survive’ in any network where this graph is a symmetric motif. The redundant eigenvalues and eigenvectors agree with those predicted by the formulae in Table 2. Here, $\varphi = \frac{1+\sqrt{5}}{2}$, the golden ratio, $\eta_1 = 2 + \frac{2}{\varphi}$ and $\eta_2 = 2 + \frac{2}{1-\varphi}$.

of the most common symmetric motifs, namely BSMs up to 2 orbits, see below and Table 2), namely positive integers (Fig. 7).

We now give more details of the computation of the redundant spectrum of BSMs up to two orbits (Table 2), with full details in the Methods section. A BSM with one orbit is an (α, β) -uniform graph $K_n^{\alpha, \beta}$ with adjacency matrix $A_n^{\alpha, \beta} = (a_{ij})$ given by $a_{ij} = \alpha$ and $a_{ii} = \beta$ for all $i \neq j$. Then $K_n^{\alpha, \beta}$ has eigenvalues $(n-1)\alpha + \beta$ (non-redundant), with multiplicity 1, and $-\alpha + \beta$ (redundant), with multiplicity $n-1$. The corresponding eigenvectors are $\mathbf{1}$, the constant vector 1 (non-redundant), and \mathbf{e}_i , the vectors with non-zero entries 1 at position 1, and -1 at position i , $2 \leq i \leq n$ (redundant). For unweighted graphs without loops ($\beta = 0$, $\alpha \in \{0, 1\}$), we recover the redundant eigenvalues 0 and -1 predicted in¹³.

A BSM with two orbits must be a uniform join of the form $K_n^{\alpha_1, \beta_1} \xrightarrow{\gamma, \delta} K_n^{\alpha_2, \beta_2}$ (Fig. 3). Let κ_1 and κ_2 be the two solutions of the quadratic equation $c\kappa^2 + (b-a)\kappa - c = 0$, where $a = \alpha_1 - \beta_1$, $b = \alpha_2 - \beta_2$ and $c = \gamma - \delta$. Then, the

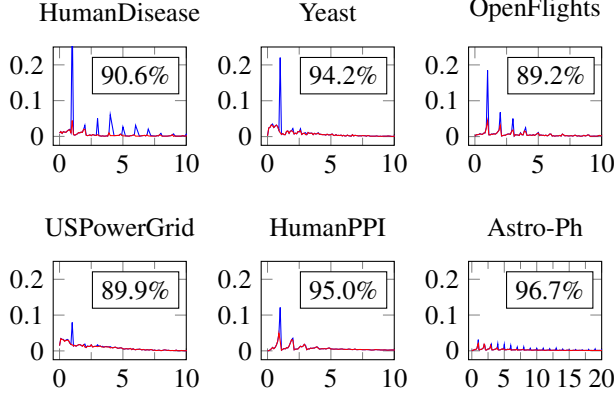


Figure 7. Spectral signatures of network symmetry.

Laplacian spectrum of six test networks (blue) and of their quotient (red), given as relative probability of eigenvalue count, with multiplicity, in bins of size 0.1. Only the most significant part of the spectrum is shown. Most of the ‘peaks’ observed in the spectral density occur at positive integers, as predicted. (Insets) Percentage of the high-multiplicity spectrum explained by the symmetry, as the ratio of $\sum_{m_\lambda > 1} m_\lambda$ for the quotient eigenvalues, and for the Laplacian eigenvalues, where m_λ is the multiplicity of an eigenvalue λ rounded to 8 decimal places.

redundant eigenvalues of this BSM are (Methods, Theorem 5)

$$\lambda_1 = -b - c\kappa_1 = \frac{-(a+b) + \sqrt{(a-b)^2 + 4c^2}}{2}, \text{ and } (16)$$

$$\lambda_2 = -b - c\kappa_2 = \frac{-(a+b) - \sqrt{(a-b)^2 + 4c^2}}{2}, \quad (17)$$

each with multiplicity $n-1$, with eigenvectors $(\kappa_1 \mathbf{e}_i | \mathbf{e}_i)$ and $(\kappa_2 \mathbf{e}_i | \mathbf{e}_i)$ respectively, $2 \leq i \leq n$. For unweighted graphs without loops, we recover the redundant eigenvalues predicted in¹³, that is,

$$-2, -\varphi, -1, 0, \varphi - 1 \text{ and } 1, \quad (18)$$

where $\varphi = \frac{1+\sqrt{5}}{2}$, the golden ratio.

Eigendecomposition algorithm

Decoupling the contribution to the network spectrum from the symmetric motifs and from the quotient network, as explained above, naturally leads to an eigendecomposition algorithm that exploits the presence of symmetries: The spectrum and eigenbasis of an undirected network (equivalently, a diagonalisation of its adjacency matrix $A = UDU^T$) can be obtained from those of the quotient, and of the symmetric motifs, reducing the computational time (cubic on the number of vertices) to up to a third in our test networks (Fig. 5, left column of the spectral case), in line with our predictions ($sp = n_{\mathcal{Q}}^3$ in Table 1). The algorithm is shown and explained below. A MATLAB implementation is available at a public repository²⁶.

Our eigendecomposition algorithm (Alg. 3) applies to any undirected matrix with symmetries (identifying a matrix with the network it represents). It first computes the

BSM	eigenvalues	mult	eigenvectors
$K_n^{\alpha, \beta}$	$-\alpha + \beta$	$n-1$	\mathbf{e}_i
$K_n^{\alpha_1, \beta_1} \xleftrightarrow{\gamma, \delta} K_n^{\alpha_2, \beta_2}$	$-b - \kappa_1 c$	$n-1$	$(\kappa_1 \mathbf{e}_i \mathbf{e}_i)$
	$-b - \kappa_2 c$	$n-1$	$(\kappa_2 \mathbf{e}_i \mathbf{e}_i)$

Table 2. Redundant spectra of basic symmetric motifs (BSMs) with one or two orbits.

A BSM with one orbit is a uniform graph $K_n^{\alpha, \beta}$ with n vertices and adjacency matrix $A_n^{\alpha, \beta} = (a_{ij})$ where $a_{ij} = \alpha$ if $i \neq j$ and $a_{ii} = \beta$, for all i, j and some constants α and β . A BSM with two orbits consists of the (γ, δ) -uniform join of two uniform graphs $K_n^{\alpha_1, \beta_1}$ and $K_n^{\alpha_2, \beta_2}$, that is, the graph with $2n$ vertices and block adjacency matrix (after a suitable labelling of the vertices) of the form $\begin{pmatrix} A & C \\ C^T & B \end{pmatrix}$ where $A = A_n^{\alpha_1, \beta_1}$, $B = A_n^{\alpha_2, \beta_2}$ and $C = A_n^{\gamma, \delta}$, each defined as above. We write \mathbf{e}_i for the vector with non-zero entries 1 at position 1, and -1 at position i ($2 \leq i \leq n$), κ_1 and κ_2 for the two solutions of the quadratic equation $c\kappa^2 + (-a+b)\kappa - c = 0$ where $a = \alpha_1 - \beta_1$, $b = \alpha_2 - \beta_2$ and $c = \gamma - \delta$, and use $(v|w)$ to represent the concatenation of two vectors.

eigendecomposition of the quotient matrix, then, for each motif, the redundant eigenpairs. Namely, it first computes the spectral decomposition eig of the symmetric quotient $B_{\text{sym}} = \Lambda^{-1/2} S^T A S \Lambda^{-1/2}$ where Λ is the diagonal matrix of the orbit sizes (which can be obtained as the column sums of S). This matrix is symmetric and has the same eigenvalues as the left quotient. Moreover, if $B_{\text{sym}} = U_q D_q U_q^{-1}$ then the left quotient eigenvectors are the columns of ΛU_q . These become, in turn, eigenvectors of A by repeating their values on each orbit, and can be obtained mathematically by left multiplying by the characteristic matrix S . Then, for each motif, we compute the redundant eigenpairs using a null space matrix (explained below), storing eigenvalues and localised (zero outside the motif) eigenvectors.

Only redundant eigenvectors of a symmetric motif (that is, those which add up to zero on each orbit) become eigenvectors of A by extending them as zero outside the symmetric motif. Therefore, we need to construct redundant eigenvectors from the output of eig on each motif (the spectral decomposition of the corresponding submatrix). If $U_\lambda = (\mathbf{v}_1 | \dots | \mathbf{v}_k)$ are λ -eigenvectors of a symmetric motif with characteristic matrix of the orbit partition S_{sm} , we need to find linear combinations such that

$$S_{\text{sm}}^T (\alpha_1 \mathbf{v}_1 + \dots + \alpha_k \mathbf{v}_k) = \mathbf{0} \iff S_{\text{sm}}^T U_\lambda \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_k \end{pmatrix} = \mathbf{0}. \quad (19)$$

Therefore, if the matrix $Z \neq 0$ represents the null space of $S_{\text{sm}}^T U_\lambda$, that is, $S_{\text{sm}}^T U_\lambda Z = 0$ and $Z^T Z = 0$, then the columns of $U_\lambda Z$ are precisely the redundant eigenvectors. This is implemented in Alg. 3 within the innermost **for** loop.

Vertex measures

We have so far considered network measures of the form $F(i, j)$, where i and j are vertices. However, many important

Input: adjacency matrix A , characteristic matrix S , list of motifs
Output: spectral decomposition $A = UDU^T$

```

initialise  $U, D$  to zero matrices
 $\Lambda \leftarrow \text{diag}(\text{sum}(S))$ 
 $B_{\text{sym}} \leftarrow \Lambda^{-1/2} S^T A S \Lambda^{-1/2}$ 
 $[U_q, D_q] \leftarrow \text{eig}(B_{\text{sym}})$  so that  $B_{\text{sym}} = U_q D_q U_q^{-1}$ 
 $U_q \leftarrow \Lambda U_q$ 
 $U \leftarrow (SU_q \mid 0)$ 
 $D \leftarrow \begin{pmatrix} D_q & 0 \\ 0 & 0 \end{pmatrix}$ 
foreach motif do
   $A_{\text{sm}} \leftarrow A(\text{motif}, \text{motif})$ 
  compute orbits from motif and  $S$ 
   $S_{\text{sm}} \leftarrow S(\text{motif}, \text{orbits})$ 
   $[U_{\text{sm}}, D_{\text{sm}}] \leftarrow \text{eig}(A_{\text{sm}})$ 
  for  $\lambda \in \text{unique}(\text{diag}(D_{\text{sm}}))$  do
     $U_\lambda \leftarrow \lambda$ -eigenvectors from  $U_{\text{sm}}$ 
     $Z \leftarrow \text{null}(S_{\text{sm}}^T U_\lambda)$ 
     $d \leftarrow \text{ncol}(Z)$ 
    if  $d > 0$  then
      store  $U_\lambda Z$  in  $U$ 
      store  $\lambda$  in  $D$  with multiplicity  $d$ 
    end
  end
end

```

Algorithm 3: Eigendecomposition algorithm.

network measurements are vertex based, that is, of the form $G(i)$ for each vertex i . We say that a vertex measure G is *structural* if it only depends on the network structure and, therefore, satisfies

$$G(i) = G(\sigma(i)) \quad (20)$$

for each automorphism $\sigma \in \text{Aut}(\mathcal{G})$, that is, it is constant on orbits (Fig. 1).

Although for vertex measures we do not have a network representation, we can still exploit the network symmetries. First, G needs only to be computed/stored once per orbit, resulting on a reduction/compression ratio of $\tilde{n}_{\mathcal{Q}} = n_{\mathcal{Q}}/n_{\mathcal{G}}$ (Table 1). Secondly, when quotient recovery holds (that is, we can recover G from its values on the quotient and symmetry information alone), it amounts to a further computational reduction (Fig. 5), depending on the computational complexity of G . Finally, many vertex measures arise nevertheless from a pairwise function, such as $G(i) = F(i, i)$ (subgraph centrality from communicability), or $G(i) = \frac{1}{n} \sum_j F(i, j)$ (closeness centrality from shortest path distance), allowing the symmetry-induced results on F to carry over to G .

Applications

We illustrate our methods on several popular pairwise and vertex-based network measures. Although novel and of independent interest, these are example applications: Our methods are general and the reader should be able to adapt our results to the network measure of their interest.

Adjacency matrix. The methods in this paper can be applied to the network itself, that is, to its adjacency matrix. We

recover the structural and spectral results in^{11,13}, and the quotient compression ratio reported in¹², here $c_{\text{sparse}} = \tilde{m}_{\mathcal{Q}}$ in Table 1. The network (adjacency) eigendecomposition can be significantly sped up by exploiting symmetries (Fig. 5).

Communicability. Communicability is a very general choice of structural measure, consisting on any analytical function $f(x) = \sum a_n x^n$ applied to the adjacency matrix, $f(A) = \sum_{n=0}^{\infty} a_n A^n$, and it is a natural measure of network connectivity, since the matrix power A^k counts walks of length k ³⁷. The most common choice of coefficients is $a_n = \frac{1}{n!}$, which gives the exponential matrix $e^A = \sum_{n=0}^{\infty} \frac{A^n}{n!}$. Communicability is a structural network measure and its network representation, the graph $f(\mathcal{G})$ with adjacency matrix $f(A)$, inherits all the symmetries of \mathcal{G} and thus it has the same symmetric motifs and orbits. The BSMs are uniform joins of orbits, and each orbit is a uniform graph (Figs. 3 and 2b) characterised by the communicability of a vertex to itself (a natural measure of centrality³⁶), and the communicability between distinct vertices. As a full network measure, the compression ratio c_{full} applies (Table 1), indicating the fraction of storage needed by using the quotient to eliminate redundancies (Fig. 4). Moreover, average quotient recovery holds for communicability since $f(Q(A)) = Q(f(A))$ (Methods, Theorem 6). Alternatively, we can use the spectral decomposition algorithm on the adjacency matrix ($A = UDU^T$ implies $f(A) = Uf(D)U^T$) reducing the computation, typically cubic on the number of vertices, by $sp = \tilde{n}_{\mathcal{Q}}^3$ (Table 1, Fig. 5). For the spectral results, note that $f(A) = Uf(D)U^T$ has eigenvalues $f(\lambda)$, and same eigenvectors, as A . Thus

$$f(-2), f(-\varphi), f(-1), f(0), f(\varphi - 1), \text{ and } f(1) \quad (21)$$

account for most of the discrete part of the spectrum $f(A)$, for the adjacency matrix A of a typical (undirected, unweighted) real-world network (Eq. (18)).

Shortest path distance. This is the simplest metric on a (connected) network, namely the length of a shortest path between vertices. A *path* of length n is a sequence $(v_1, v_2, \dots, v_{n+1})$ of distinct vertices, except possibly $v_1 = v_{n+1}$, such that v_i is connected to v_{i+1} for all $1 \leq i \leq n-1$. The *shortest path distance* $d^{\mathcal{G}}(u, v)$ is the length of the shortest (minimal length) path from u to v . If $p = (v_1, v_2, \dots, v_n)$ is a path and $\sigma \in \text{Aut}(\mathcal{G})$, we define $\sigma(p) = (\sigma(v_1), \sigma(v_2), \dots, \sigma(v_n))$, also a path since σ is a bijection. A path p is a *shortest path* if it is of minimal length between its endpoints. One can show that (i) automorphisms preserve shortest paths and their lengths; (ii) shortest paths between vertices in different symmetric motifs do not contain intra-orbit edges; and (iii) shortest path distance is a partially quotient recoverable structural measure (Methods, Theorem 7). In particular, automorphisms σ preserve the shortest path metric, $d(i, j) = d(\sigma(i), \sigma(j))$, and we can compute shortest distances from the quotient,

$$d^{\mathcal{Q}}(\alpha, \beta) = d^{\mathcal{Q}}(i, j), \quad \alpha \in V_i, \beta \in V_j, \quad (22)$$

whenever V_i and V_j are orbits in different symmetric motifs.

This accounts for all but the (small) intra-motif distances and reduces the computation as shown in Fig. 5.

Distances between points within the same motif cannot in general be directly recovered from the quotient, not even for BSMs. (Consider for instance the double star, motif M_1 , in Fig. 1: The distance from the top red to the bottom blue vertex is three, while in the quotient is one.) In general, therefore, the shortest path distance is partially, but not average, quotient recoverable. Intra-motif distances, if needed, could still be recovered one motif at a time.

Note that these results can be exploited for other graph-theoretic notions defined in terms of distance, for example eccentricity (and thus radius or diameter), which only depends on maximal distances and thus it can be computed directly in the quotient.

In terms of symmetry compression, the compression ratio c_{full} applies, accounting for the amount of structural redundancy due solely to symmetries. The spectral results, although perhaps less relevant, still apply for $d(\mathcal{G})$, the graph encoding pairwise shortest path distances. The adjacency matrix $d(A) = (d^{\mathcal{G}}(i, j))$ is nonzero outside the diagonal, hence $d(\mathcal{G})$ is a all-to-all weighted network without self-loops and integer weights, and so is each symmetric motif. Using the formula in Table 2, we can easily compute values of the most significant part of the discrete spectrum (redundant eigenvalues) of $d(A)$, namely $-3, -2, -1, 0, -2 \pm \sqrt{2}, -3 \pm \sqrt{2}, \frac{-3 \pm \sqrt{5}}{2}, \frac{-5 \pm \sqrt{5}}{2}$ and $\frac{-5 \pm \sqrt{13}}{2}$.

Laplacian matrix. The Laplacian matrix of a network $L = D - A$, where D is the diagonal matrix of vertex degrees, is a (sparse) network measure and therefore inherits all the symmetries of the network. The matrix L can be seen as the adjacency matrix of a network \mathcal{L} with identical symmetric motifs, except that all edges are weighted by -1 and all vertices have self-loops weighted by their degrees in \mathcal{G} (Fig. 2c). In particular, the motif structure (namely, the self-loop weights) depends on the how the motif is embedded in the network \mathcal{G} .

Quotient compression and computational reduction are less useful in this case, however the spectral results are more interesting. The spectral decomposition applies, and we can compute redundant Laplacian eigenvalues directly from Table 2, for instance positive integers for BSMs with one orbit (Methods, Corollary 2). This explains and predicts most of the ‘peaks’ (high multiplicity eigenvalues) in the Laplacian spectral density, confirmed on our test networks (Fig. 7). Using the formula in Table 2, one can similarly compute the redundant spectrum for 2-orbit BSMs, and for other versions of the Laplacian (e.g. normalised, vertex weighted). Finally, observe that the spectral decomposition applies, thus Algorithm 3 provides an efficient way of computing the Laplacian eigendecomposition with an expected $sp = \tilde{n}_{\mathcal{Q}}^3$ (see Table 1) computational time reduction.

Commute distance and matrix inversion. The commute distance is the expected time for a random walker to travel

between two vertices and back⁴⁴. In contrast to the shortest path distance, it is a global metric which takes into account all possible paths between two vertices. The commute distance is equal up to a constant (the volume of the network) to the resistance metric r ⁴⁵, which can be expressed in terms of $L^{\dagger} = (l_{ij}^{\dagger})$, the pseudoinverse (or Moore-Penrose inverse) of the Laplacian, as $r(i, j) = l_{ii}^{\dagger} + l_{jj}^{\dagger} - 2l_{ij}^{\dagger}$. The commute (or resistance) distance is a (full) structural measure, and all our structural and spectral results apply. Crucially, we can use eigendecomposition algorithm to obtain $L = UDU^T$ (and hence $L^{\dagger} = UD^{\dagger}U^T$, and r) from the quotient and symmetric motifs, resulting in significant computational gains (Fig. 5). More generally, if M_F is the matrix representation of a network measure, its pseudoinverse M_F^{\dagger} is also a network measure, and the comments above apply. Note that M_F^{\dagger} is generally a full measure even if M_F is sparse (the inverse of a sparse matrix is not generally sparse).

Vertex symmetry compression. As a vertex measure G is constant on orbits, we only need to store one value per orbit. Let S be the characteristic matrix of the partition of the vertex set into orbits, and Λ the diagonal matrix of orbit sizes (column sums of S). If G is represented by a vector $v = (G(i))$ of length $n_{\mathcal{G}}$, we can store one value per orbit by taking $\mathbf{w} = \Lambda^{-1}S^T \mathbf{v}$, a vector of length $n_{\mathcal{Q}}$, and recover $\mathbf{v} = S^T \mathbf{w}$ (Methods, Theorem 9).

Degree Centrality. The degree of a node (in- or out-degree if the network is directed) is a natural measure of vertex centrality. As expected, the degree is preserved by any automorphism σ , which can also be checked directly,

$$d_i = \sum_{j \in V} a_{ij} = \sum_{j \in V} a_{\sigma(i)\sigma(j)} = \sum_{j \in V} a_{\sigma(i)j} = d_{\sigma(i)}, \quad (23)$$

as automorphisms permute orbits (so $j \in V$ and $\sigma(j) \in V$ are the same elements but in a different order). In particular, the degree is constants on orbits. We recover the degree centrality from the quotient as the out-degree (Methods, Proposition 2).

Closeness centrality. The closeness centrality of a node i in a graph \mathcal{G} , $cc^{\mathcal{G}}(i)$, is the average shortest path length to every node in the graph. As symmetries preserve distance, they also preserve closeness centrality, explicitly,

$$\begin{aligned} cc(i) &= \frac{1}{n_{\mathcal{G}}} \sum_{j \in V} d(i, j) = \frac{1}{n_{\mathcal{G}}} \sum_{j \in V} d(\sigma(i), \sigma(j)) \\ &= \frac{1}{n_{\mathcal{G}}} \sum_{j \in V} d(\sigma(i), j) = cc(\sigma(i)), \end{aligned} \quad (24)$$

and centrality is constant on each orbit, as expected. Moreover, closeness centrality can be recovered from the quotient (shortest paths does not contain intra-orbit edges, except between vertices in the same symmetric motif, see above), as

$$cc^{\mathcal{G}}(i) = \sum_{l \neq k} \frac{n_l}{n_{\mathcal{G}}} d^{\mathcal{Q}}(k, l) + \frac{n_i}{n_{\mathcal{G}}} d_k \quad (25)$$

if i belongs to the orbit V_k and d_k is the average intra-motif distance, that is, the average distances of a vertex in V_k to any vertex in \mathcal{M} , the motif containing V_k . By annotating each orbit by d_k , we can recover betweenness centrality exactly. Alternatively, as $d_k \ll n$ (note that $d_k \leq m$ if \mathcal{M} has m orbits), we can approximate $cc^{\mathcal{G}}(i)$ by the first summand, or simply by the quotient centrality $cc^{\mathcal{Q}}(\alpha)$, in most practical situations.

Betweenness centrality. This is the sum of proportions of shortest paths between pairs of vertices containing a given vertex. It can be computed from shortest path distances and number of shortest paths⁴⁶, both pairwise structural measures, reducing the computation of a naive $O(n^3)$ time, $O(n^2)$ space implementation by $\tilde{n}_{\mathcal{Q}}^3$ and $\tilde{n}_{\mathcal{Q}}^2$. It would be interesting to adapt a faster algorithm e.g.⁴⁶ to exploit symmetries, but this is beyond our scope.

Eigenvector centrality. Eigenvector centrality is obtained from a Perron-Frobenius eigenvector (i.e. of the largest eigenvalue) of the adjacency matrix of a connected graph¹. Since this eigenvalue must be simple, it cannot be a redundant eigenvalue. Hence it is a quotient eigenvalue, and, as those are a subset of the parent eigenvalues, it must still be the largest (hence the Perron-Frobenius) eigenvalue of the quotient. Its eigenvector can then be lifted to the parent network, by repeating entries on orbits. That is, if (λ, v) is the Perron-Frobenius eigenpair of the quotient, then (λ, Sv) is the Perron-Frobenius eigenpair of the parent network. In practice, we use the symmetric quotient $B_{\text{sym}} = \Lambda^{-1/2} S^T A S \Lambda^{-1/2}$ for numerical reasons (Algorithm 4). Hence the computation (quadratic time by power iteration) can be reduced by $\tilde{n}_{\mathcal{Q}}^2$ (Fig. 5).

Input: adjacency matrix A , characteristic matrix S

Output: (right) Perron-Frobenius eigenpair (λ, v) of A

$\Lambda \leftarrow \text{diag}(\text{sum}(S))$

$R \leftarrow S \Lambda^{-1/2}$

$B_{\text{sym}} \leftarrow R^T A R$

$(\lambda, w) \leftarrow \text{eig}(B_{\text{sym}}, 1)$ eigenpair of the largest eigenvalue

$v \leftarrow R w$

Algorithm 4: Eigenvector centrality from the quotient network.

Discussion

We have presented a general theory to describe and quantify the effects of network symmetry on arbitrary network measures, and explained how this can be exploited in practice in a number of ways.

Network symmetry of the large but sparse graphs typically found in applications can be effectively computed and manipulated, making it an inexpensive pre-processing step. We showed that the amount of symmetry is amplified in a pairwise network measure but can be easily discounted using the quotient network. We can for instance eliminate the symmetry-induced redundancies, or use them to simplify the calculation by avoiding unnecessary computations. Symmetry has also a profound effect on the spectrum, explaining the character-

istic ‘peaks’ observed in the spectral densities of empirical networks, and occurring at values we are able to predict.

Our framework is very general and apply to any pairwise or vertex-based network measure beyond the ones we discuss as examples. We emphasised practical and algorithmic aspects throughout, and provide pseudocode and full implementations²⁶. Since real-world network models and data are very common, and typically contain a large degree of structural redundancy, our results should be relevant to any network practitioner.

Methods

Geometric decomposition and symmetric motifs

We write $\text{Aut}(\mathcal{G})$ for the automorphism group of an (unweighted, undirected, possibly very large) network $\mathcal{G} = (V, E)$ (see below for a discussion of directed and weighted networks). Each automorphism (symmetry) $\sigma \in \text{Aut}(\mathcal{G})$ is a permutation of the vertices and its *support* is the set of vertices moved by σ ,

$$\text{supp}(\sigma) = \{i \in V \text{ such that } \sigma(i) \neq i\}. \quad (26)$$

Two automorphisms σ and τ are *support-disjoint* if the intersection of their supports is empty, $\text{supp}(\sigma) \cap \text{supp}(\tau) = \emptyset$. The *orbit* of a vertex i is the set of vertices to which i can be moved to by an automorphism, that is,

$$\{\sigma(i) \text{ such that } \sigma \in \text{Aut}(\mathcal{G})\}. \quad (27)$$

One can show¹¹ that there is a partition a set X of generators of $\text{Aut}(\mathcal{G})$ into its finest support-disjoint classes $X = X_1 \cup \dots \cup X_m$ which is unique up to permutation of the sets X_i . The vertex sets $M_i = \bigcup_{\sigma \in X_i} \text{supp}(\sigma)$ give the geometric decomposition Eq. (2), and the subgraphs induced by them are, by definition, the *symmetric motifs* of \mathcal{G} . (The next section explains how to compute the geometric decomposition in practice.) Since support-disjoint automorphisms must commute (the order in which they are composed is irrelevant), the subgroups of $\text{Aut}(\mathcal{G})$ generated by X_1 to X_m , call them H_1 to H_m , give a direct product decomposition $\text{Aut}(\mathcal{G}) = H_1 \times \dots \times H_m$. The geometric decomposition is defined from the finest support-disjoint partition of a special set of generators (called *essential*), as explained in¹¹. However, the results in this article are valid for any support-disjoint decomposition of any set of generators (essential or not) of $\text{Aut}(\mathcal{G})$.

If all the orbits of a symmetric motif have the same size k and every permutation of the vertices in each orbit can be extended to a network automorphism supported on the motif, we call the symmetric motif *basic* (or BSM) of *type* k . (In particular, the corresponding subgroup H_i must be $\text{Sym}(k)$, the symmetric group of all permutations of k elements.) If a symmetric motif is not basic, we call it *complex* or of *type* 0.

Network symmetry computation

First, we compute a list of generators of the automorphism group from an edge list (we use `saucy3`⁴⁷, which is extremely fast for the large but sparse networks typically found in applications). Then, we partition the set of generators X into support-disjoint classes $X = X_1 \cup \dots \cup X_m$, that is, σ and τ are support-disjoint whenever $\sigma \in X_i$, $\tau \in X_j$ and $i \neq j$. To find the finest such partition, we use a bipartite graph representation of vertices V and generators X . Namely, let \mathcal{B} be the graph with vertex set $V \cup X$ and edges between i and σ whenever $i \in \text{supp}(\sigma)$. Then X_1, \dots, X_m are the connected components of \mathcal{B} (as vertex sets intersected with X). Each X_i corresponds to the vertex set M_i of a symmetric motif \mathcal{M}_i , as $M_i = \bigcup_{\sigma \in X_i} \text{supp}(\sigma)$. Finally, we use GAP⁴⁸ to compute the orbits and type of each symmetric motif (Alg. 5). Full implementations of all the procedures outlined above are available at a public repository²⁶.

Structural network measures

We prove below the structural result for BSMs for arbitrary graphs and network measures. The proof is a generalisation of the argument on [49, p.48] to weighted directed graphs with symmetries.

Theorem 1. *Let M be the vertex set of a BSM of a network \mathcal{G} , and F a structural network measure. Then the graph induced by M in $F(\mathcal{G})$ is a BSM of $F(\mathcal{G})$, and satisfies:*

```

Input:  $X$  a set of permutations of a symmetric motif
Output:  $O_1, \dots, O_k$  orbits, and type  $m$ , of the symmetric motif
 $H \leftarrow \text{Group}(X)$ 
 $\{O_1, \dots, O_k\} \leftarrow \text{Orbits}(H)$ 
 $m \leftarrow \min(\text{size}(O_1), \dots, \text{size}(O_k))$ 
if  $m = \max(\text{size}(O_1), \dots, \text{size}(O_k))$  then
  for  $i \leftarrow 1$  to  $k$  do
    if not
       $\text{IsNaturalSymmetricGroup}(\text{Action}(H, O_i))$ 
    then
       $m \leftarrow 0$ 
      break
    end
  end
else
   $m \leftarrow 0$ 
end

```

Algorithm 5: Orbits and type of a symmetric motif.

- (i) for each orbit $\Delta = \{v_1, \dots, v_n\}$, there are constants α and β such that the orbit internal connectivity is given by $\alpha = F(v_i, v_j)$ for all $i \neq j$ and $\beta = F(v_i, v_i)$ for all i ;
- (ii) for every pair of orbits Δ_1 and Δ_2 , there is a labelling $\Delta_1 = \{v_1, \dots, v_n\}$, $\Delta_2 = \{w_1, \dots, w_n\}$ and constants $\gamma_1, \gamma_2, \delta_1, \delta_2$ such that $\gamma_1 = F(v_i, w_j)$, $\gamma_2 = F(w_j, v_i)$, $\delta_1 = F(v_i, w_i)$, and $\delta_2 = F(w_i, v_i)$, for all $i \neq j$;
- (iii) every vertex v not in the BSM is joined uniformly to all the vertices in each orbit $\{v_1, \dots, v_n\}$ in the BSM, that is, $F(v, v_i) = F(v, v_j)$ and $F(v_i, v) = F(v_j, v)$ for all i, j .

Moreover, property (iii) holds in general for any symmetric motif.

If \mathcal{G} is undirected and F is symmetric, $\gamma_1 = \gamma_2$ and $\delta_1 = \delta_2$ and each orbit is a (α, β) -uniform graph $K_n^{\alpha, \beta}$ and each pair of orbits form a (γ, δ) -uniform join, explaining Figure 3(a, b).

Proof of Theorem. As $F(\mathcal{G})$ inherits all the symmetries of \mathcal{G} , M has the same orbit decomposition and the symmetric group S_n acts in the same way, hence M induces a BSM in $F(\mathcal{G})$ too. For the internal connectivity, note that every permutation of the vertices v_i is realisable. Thus, given arbitrary $1 \leq i, j, k, l \leq n$, we can find $\sigma \in \text{Aut}(\mathcal{G})$ such that $\sigma(v_k) = v_i$ and, if $j \neq i$ and $l \neq k$, additionally satisfies $\sigma(v_l) = v_j$. This gives

$$F(v_i, v_j) = F(\sigma(v_k), \sigma(v_l)) = F(v_k, v_l),$$

as F is a structural network measure. The other case, $i = j$ and $k = l$, gives

$$F(v_i, v_i) = F(\sigma(v_k), \sigma(v_k)) = F(v_k, v_k).$$

For the orbit connectivity result (ii), we generalise the argument in [49, p.48] to weighted directed graphs with symmetries, particularly $F(\mathcal{G})$. We assume some basic knowledge and terminology about group actions⁵⁰ and symmetric groups S_n . Given two orbits $\Delta_1 = \{v_1, \dots, v_n\}$ and $\Delta_2 = \{w_1, \dots, w_n\}$ and $1 \leq i \leq n$, define

$$\Gamma_i = \{w_j \in \Delta_2 \mid F(v_i, w_j) \neq 0\},$$

the vertices in Δ_2 joined to v_i in $F(\mathcal{G})$. If a finite group G acts on a set X , the stabiliser of a point $G_x = \{g \in G \mid gx = x\}$ is a subgroup of G of index $[G : H] = \frac{|G|}{|H|}$ equals to the size of the orbit of x . Hence the stabilisers G_{v_i} or G_{w_j} are subgroups of S_n of index n , for all i, j . The group S_n has a unique, up to conjugation, subgroup of index n if $n \neq 6$. In this case, G_{v_i} is conjugate to G_{w_1} so $G_{v_i} = \sigma G_{w_1} \sigma^{-1} = G_{\sigma w_1}$ for some $\sigma \in S_n$. Relabelling σw_1 as w_1 we have $G_{v_i} = G_{w_1}$. Similarly, we can relabel the remaining vertices in Δ_2 so that $G_{v_i} = G_{w_i}$ for all i : write $v_2 = \sigma_2 v_1, v_3 = \sigma_3 v_1, \dots$ and relabel $w_2 = \sigma_2 w_1, w_3 = \sigma_3 w_1, \dots$, noticing there cannot be repetitions as $\sigma_k w_1 = \sigma_l w_1$ for $k \neq l$ implies $\sigma_k \sigma_l^{-1} \in G_{w_1} = G_{v_1}$ and thus $v_k = \sigma_k v_1 = \sigma_l v_1 = v_l$, a contradiction. Fix $1 \leq i \leq n$. The stabiliser G_{v_i} fixes v_i but it may permute vertices in Δ_2 . In fact, the set Γ_i above must be a union of orbits of G_{v_i} on Δ_2 : if $w \in \Gamma_i$ and $\sigma \in G_{v_i}$ then

$$0 \neq F(v_i, w) = F(\sigma v_i, \sigma w) = F(v_i, \sigma w)$$

so σw also belongs to Γ_i . The orbits of $G_{v_i} = G_{w_i}$ in Δ_2 are $\{w_i\}$ and $\Delta_2 \setminus \{w_i\}$, as G_{w_i} fixes w_i and freely permutes all other vertices in Δ_2 . The case $n = 6$ is similar, except that S_6 has two conjugacy classes of subgroups of index 6, one as above, and the other a subgroup acting transitively on the 6 vertices, which gives a unique orbit Δ_2 . In all cases, the set $\Delta_2 \setminus \{w_i\}$ is part of an G_{v_i} -orbit, which gives the connectivity result, as follows. Fix $1 \leq i \leq n$. For $1 \leq j, k \leq n$ different from i , the vertices w_j and w_k are in the same G_{v_i} -orbit so there is $\sigma \in G_{v_i}$ with $\sigma w_j = w_k$ and, therefore,

$$F(v_i, w_j) = F(\sigma v_i, \sigma w_j) = F(v_i, w_k).$$

The argument is general, so we have shown $a_i = F(v_i, w_j)$ is constant for all $j \neq i$. It is enough to show $a_i = a_1$ for all i . Choose $j \neq i$, then

$$a_i = F(v_i, w_j) = F(\sigma_i v_1, \sigma_j w_1) = F(v_1, \sigma_i^{-1} \sigma_j w_1) = a_1$$

as long as $\sigma_i^{-1} \sigma_j w_1 \neq w_1$, which cannot happen as otherwise $\sigma_i^{-1} \sigma_j \in G_{w_1} = G_{v_1}$ implies $\sigma_i^{-1} \sigma_j v_1 = v_1$ or $v_j = \sigma_j v_1 = \sigma_i v_1 = v_i$, a contradiction. Hence we have shown $F(v_i, w_j)$ is a constant, call it γ_i , for all $i \neq j$. In addition,

$$F(v_i, w_i) = F(\sigma_i v_1, \sigma_i w_1) = F(v_1, w_1)$$

is also a constant, call it δ_i , for all i . The cases $\gamma_2 = F(w_j, v_i)$ and $\delta_2 = F(w_i, v_i)$ are identical, reversing the roles of Δ_1 and Δ_2 .

Property (iii) holds for any symmetric motif, not necessarily basic, as follows. By the definition of orbit, for each pair i, j we can find an automorphism σ in the geometric factor such that $\sigma(v_j) = v_i$. Since v is not in the support of that geometric factor, it is fixed by σ , that is, $\sigma(v) = v$. Therefore

$$F(v, v_i) = F(\sigma(v), \sigma(v_j)) = F(v, v_j),$$

and similarly $F(v_i, v) = F(v_j, v)$. \square

Average compression

Theorem 2. Let $A = (a_{ij})$ be the $n \times n$ adjacency matrix of a network with vertex set V . Let S be the $n \times m$ characteristic matrix of the partition of V into orbits of the automorphism group, and Λ the diagonal matrix of column sums of S . Define $B = S^T A S$ and $A_{\text{avg}} = R B R^T = (\bar{a}_{ij})$ where $R = S \Lambda^{-1}$. Then,

- (i) if $i, j \in V$ belong to different symmetric motifs, $\bar{a}_{ij} = a_{ij}$;
- (ii) if $i, j \in V$ belong to orbits $i \in \Delta_1$ and $j \in \Delta_2$ in the same symmetric motif,

$$\bar{a}_{ij} = \frac{1}{|\Delta_1|} \frac{1}{|\Delta_2|} \sum_{\substack{u \in \Delta_1 \\ v \in \Delta_2}} a_{uv}. \quad (28)$$

Before proving this statement, we make a few observations. The column sums of S equal the sizes of the vertex partition sets, hence Λ is the same as in the definition of quotient matrix (6), and can be obtained easily from S . The matrix S is very sparse (each row has a unique non-zero entry) and can be stored very efficiently. Case (i) covers the vast majority of vertex pairs (external edges) for a network measure (see *ext_s* and *int_f* in Table 1). In (ii), the case $\Delta_1 = \Delta_2$ is allowed. The matrix $B = S^T A S$ is symmetric with integer entries if A is too, hence generally easier to store than $Q(A) = \Lambda^{-1} S^T A S$.

Proof of Theorem. Let $V = \Delta_1 \cup \dots \cup \Delta_m$ be the partition into orbits, and write $n_k = |\Delta_k|$. Clearly, the row sums of S equals n_1, \dots, n_m . Writing $[M]_{ij}$ for the (i, j) -entry of a matrix M , matrix multiplication gives

$$[R]_{ik} = \sum_l [S]_{il} [\Lambda^{-1}]_{lk} \stackrel{l=k}{=} [S]_{ik} \frac{1}{n_k} = \begin{cases} \frac{1}{n_k} & \text{if } i \in \Delta_k, \\ 0 & \text{otherwise.} \end{cases}$$

Similarly, assuming $i \in \Delta_k$ and $j \in \Delta_l$, we have

$$\bar{a}_{ij} = [R B R^T]_{ij} = \sum_{\alpha, \beta} [R]_{i\alpha} [B]_{\alpha\beta} [R]_{j\beta} = \frac{1}{n_k} \frac{1}{n_l} [B]_{kl} = \frac{1}{n_k} \frac{1}{n_l} \sum_{\substack{u \in \Delta_k \\ v \in \Delta_l}} a_{uv}.$$

This expression reduces to a_{ij} if the orbits belong to different symmetric motifs, since in this case all the summands in $\sum_{u \in \Delta_k, v \in \Delta_l} a_{uv}$ are equal to one another. Indeed, given $i_1, i_2 \in \Delta_k$ and $j_1, j_2 \in \Delta_l$, we can find, by the definition of orbit and symmetric motif, automorphisms σ and τ such that $\sigma(i_1) = i_2$ while fixing j_1 , and $\tau(j_1) = j_2$ while fixing i_1 . This gives

$$a_{i_1 j_1} = a_{\tau \sigma(i_1) \tau \sigma(j_1)} = a_{\tau(i_2) \tau(j_1)} = a_{i_2 j_2}. \quad \square$$

A similar proof shows that we can recover exact inter-motif connectivity (external edges), and average intra-motif connectivity (average internal edges) from the quotient network, as follows.

Corollary 1. Let $A = (a_{ij})$ be the $n \times n$ adjacency matrix of a network and $Q(A) = (b_{kl})$ its quotient with respect to the partition into orbits of the automorphism group $V = \Delta_1 \cup \dots \cup \Delta_m$. Suppose that $i \in \Delta_k, j \in \Delta_l$. Then,

- (i) if the orbits Δ_k and Δ_l belong to different symmetric motifs, $b_{kl} = a_{ij}$;
- (ii) if the orbits Δ_k and Δ_l belong to the same symmetric motif,

$$b_{kl} = \frac{1}{|\Delta_k|} \sum_{u \in \Delta_k, v \in \Delta_l} a_{uv}. \quad (29)$$

Lossless compression

We can achieve lossless compression by exploiting the structure of BSMs, which account for most of the symmetry in real-world networks. If the motif is basic, we can preserve the exact parent network connectivity in an annotated quotient, as follows. Each orbit in a BSM is a uniform graph $K_n^{\alpha, \beta}$ which appears in the quotient as a single vertex with a self-loop weighted by $(n-1)\alpha + \beta$ (Fig. 3 (c)). Hence if we annotate this vertex in the quotient by not only n but also α , or β , we can recover the internal connectivity. Similarly, the connectivity between two orbits in the same symmetric motif is given by two parameters γ, δ and appears in the quotient as an edge weighted $(n-1)\gamma + \delta$ (Fig. 3 (d)) and thus can also be recovered from a quotient with edges annotated by γ , or δ .

Since there is no general formula for an arbitrary non-basic symmetric motif, we can work with the *basic quotient* $\mathcal{Q}_{\text{basic}}$ instead, that is, the quotient with respect to the partition of the vertex set into orbits in BSMs only (vertices in non-basic symmetric motifs become fixed points hence part of the asymmetric core). The annotated (as above) basic quotient achieves most of the symmetry reduction in a typical empirical network ($\bar{n}_{\mathcal{Q}}^{\text{basic}} \approx \bar{n}_{\mathcal{Q}}, \bar{m}_{\mathcal{Q}}^{\text{basic}} \approx \bar{m}_{\mathcal{Q}}$, (see Table 1, caption). To maintain the same vertex labelling as in the parent network, we record, for each pairs of orbits in the same symmetric motif, the corresponding permutation of the second orbit (else we recover the adjacency matrix only up to permutations of the orbits).

Algorithms for lossless compression and recovery based on the basic quotient are shown below (Algorithms 6 and 7), and MATLAB implementations for BSMs up to two orbits are available at a public repository²⁶. The results reported in Fig. 4 are with respect to these implementations, and the actual compression ratios reported include the size of the annotation data for lossless compression with vertex identity (a very small fraction of the size of the quotient in practice, adding at most 0.02% to the basic full compression ratio in all our test cases).

Spectral signatures of symmetry

The partition into orbits satisfy the following regularity condition^{34,35}. A partition of the vertex set $V = V_1 \cup \dots \cup V_m$ is *equitable* if

$$\sum_{j \in V_l} a_{ij} = \sum_{j \in V_l} a_{i_2j} \quad \text{for all } i_1, i_2 \in V_k, \text{ for all } 1 \leq k, l \leq m, \quad (30)$$

that is, if the connectivity from a node in V_i to all nodes in V_j is independent of the chosen node in V_i . For completeness, we show the following.

Proposition 1. Let $V = V_1 \cup \dots \cup V_m$ be a partition of the vertex set of a graph with adjacency matrix $A = (a_{ij})$, and let S be the characteristic matrix of the partition. Write $Q(A)$ for the quotient with respect to the partition.

- (i) The partition is equitable if and only if $AS = SQ(A)$;
- (ii) The partition into orbits of the automorphism group is equitable.

Proof. (i) Fix $1 \leq i \leq n$ and $1 \leq k \leq m$, and suppose $i \in V_l$. Then

$$[AS]_{ik} = \sum_{j \in V_k} a_{ij},$$

and, using the equitable condition,

$$[SQ(A)]_{ik} = [Q(A)]_{ik} = \frac{1}{|V_l|} \sum_{i_1 \in V_l, i_2 \in V_k} a_{i_1j} = \frac{1}{|V_l|} |V_l| \sum_{j \in V_k} a_{ij} = \sum_{j \in V_k} a_{ij}.$$

Input: adjacency matrix A , characteristic matrix for the basic quotient S , list of BSMs motifs

Output: quotient matrix B , annotation structure a

$B \leftarrow S^T A S$

extract orbits from S

foreach orb in orbits **do**

$\text{rep} \leftarrow \min(\text{orb})$

$\beta \leftarrow A(\text{rep}, \text{rep})$

 store β in annotation structure a

end

$k_{\max} \leftarrow \max(\text{size}(\text{motifs}))$ maximal number of orbits in a motif

for $k \leftarrow 2$ **to** k_{\max} **do**

 extract k -BSM (list of BSMs with k orbits) from motifs

foreach bsm in k -BSM **do**

foreach pairs of distinct orbits V_1, V_2 in bsm **do**

 compute δ and permutation of V_2 perm such that

$A(k, \text{perm}(k)) = \delta$ for all $k \in V_1$

 store orbit numbers (with respect to S), δ and perm in annotation structure a

end

end

end

Algorithm 6: Lossless symmetry compression.

For the converse, note that $[AS]_{il}$ does not depend on i but on the orbit of i . Namely, given $i_1, i_2 \in V_k$,

$$\sum_{j \in V_l} a_{i_1j} = [AS]_{i_1l} = [Q(A)]_{kl} = [AS]_{i_2l} = \sum_{j \in V_l} a_{i_2j}.$$

(ii) Given i_1 and i_2 in the same orbit Δ_k , choose an automorphism σ such that $\sigma(i_1) = i_2$. Then, since automorphisms respect the adjacency matrix, $a_{ij} = a_{\sigma(i)\sigma(j)}$ for all i, j , we have

$$\sum_{j \in \Delta_l} a_{i_1j} = \sum_{j \in \Delta_l} a_{\sigma(i_1)\sigma(j)} = \sum_{j \in \Delta_l} a_{i_2\sigma(j)} = \sum_{j \in \Delta_l} a_{i_2j},$$

where the last equality follows from the fact that an element in a group permutes orbits, in this case, $\{j : j \in \Delta_l\} = \{\sigma(j) : j \in \Delta_l\}$. Hence the partition into orbits is equitable. \square

It follows immediately that the quotient eigenvalues are a subset of the eigenvalues of the parent network,

$$Q(A)v = \lambda v \implies A(Sv) = SQ(A)v = \lambda Sv. \quad (31)$$

(Note that $Sv \neq 0$ if $v \neq 0$.) That is, the spectrum of the quotient is a subset of the spectrum of the graph, with eigenvectors lifted from the quotient by repeating entries on orbits. Moreover, we can complete an eigenbasis with eigenvectors orthogonal to the partition (adding up to zero on each orbit).

Theorem 3. Suppose that A is an $n \times n$ real symmetric matrix and B the $m \times m$ quotient matrix with respect to an equitable partition $V_1 \cup \dots \cup V_m$ of the set $\{1, 2, \dots, n\}$. Let S be the characteristic matrix of the partition. Then A has an eigenbasis of the form

$$\{Sv_1, \dots, Sv_m, w_1, \dots, w_{n-m}\},$$

where $\{v_1, \dots, v_m\}$ is any eigenbasis of B , and $S^T w_i = 0$ for all i .

Proof. Recall that $Sv \neq 0$ if $v \neq 0$ (S lifts the vector v from the quotient by repeating entries on each orbit) so the linear map

$$\mathbb{R}^m \rightarrow \mathbb{R}^n, v \mapsto Sv$$

has trivial kernel and hence it is an isomorphism onto its image. In particular, $\mathcal{B} = \{Sv_1, \dots, Sv_m\}$ is also a linearly independent set, and they are all eigenvectors of A , since $AS = SB$ as the partition is equitable. To finish the proof we need to complete \mathcal{B} to a basis $\{Sv_1, \dots, Sv_m, w_1, \dots, w_{n-m}\}$ such that each w_j is an A -eigenvector orthogonal to all Sv_i . As \mathcal{B} is a basis of $\text{Im}(S)$, this would imply $w_i \in \text{Im}(S)^\perp = \text{Ker}(S^T)$, giving $S^T w_i = 0$ for all i ,

Input: quotient matrix B , characteristic matrix S , annotation structure a

Output: adjacency matrix A

$\Lambda \leftarrow \text{diag}(\text{sum}(S))$

$R \leftarrow S\Lambda^{-1}$

$A \leftarrow RBR^T$

extract orbits from S

foreach orb in orbits **do**

$n \leftarrow \text{size}(\text{orb})$

 extract β from a

 compute α from B , β and n (using

$[B]_{\text{orb}, \text{orb}} = n((n-1)\alpha + \beta)$)

 construct adjacency matrix of the orbit $A_n^{\alpha, \beta}$

$A(\text{orb}, \text{orb}) \leftarrow A_n^{\alpha, \beta}$

end

extract pairs of orbits in the same BSM from a

foreach (V_1, V_2) in pairs **do**

$n \leftarrow \text{size}(V_1)$

 extract δ , perm from a

 compute γ from B , δ and n (using $[B]_{V_1, V_2} = n((n-1)\gamma + \delta)$)

 construct matrix $A_n^{\gamma, \delta}$

$A(V_1, \text{perm}) \leftarrow A_n^{\gamma, \delta}$

$A(\text{perm}, V_1) \leftarrow A_n^{\gamma, \delta}$

end

Algorithm 7: Lossless symmetry decomposition.

as desired. Since A is diagonalisable, \mathbb{R}^n decomposes as an orthogonal direct sum of eigenspaces, $\mathbb{R}^n = \bigoplus_\lambda E_\lambda$. In each E_λ , we can find vectors w_j such that they complete $V_\lambda = \{Sv_i \in \mathcal{B} \mid v_i \text{ } \lambda\text{-eigenvector}\}$ to a basis of E_λ and that are orthogonal to all vectors in V_λ (consider the orthogonal complement of the subspace generated by V_λ in E_λ). Repeating this procedure on each E_λ , we find vectors $\{w_1, \dots, w_{n-m}\}$ as needed. \square

The statement and proof above holds for arbitrary matrices A by replacing ‘eigenbasis’ by ‘maximal linearly independent set’ and removing the condition $S^T w_i = 0$. It would be interesting to know whether the condition $S^T w_i = 0$ holds for motif eigenvectors in the directed case as well (the proof above is no longer valid).

Further to the spectral decomposition theorem above, we can give an even more precise description of the redundant spectrum: it is made of the contributions from the spectrum of each individual symmetric motif.

Theorem 4. Let \mathcal{M} be a symmetric motif of a (possibly weighted) undirected graph \mathcal{G} . If (λ, w) is a redundant eigenpair of \mathcal{M} then (λ, \tilde{w}) is an eigenpair of \mathcal{G} , where \tilde{w} is equal to w on (the vertices of) \mathcal{M} , and zero elsewhere.

Proof. Since (λ, v) is an \mathcal{M} -eigenpair,

$$\sum_{j \in V(\mathcal{M})} [A_{\mathcal{M}}]_{ij} w_j = \lambda w_i \quad \forall i \in V(\mathcal{M}),$$

where $A_{\mathcal{M}}$ is the adjacency matrix of \mathcal{M} . We can decompose \mathcal{M} into orbits,

$$V(\mathcal{M}) = V_1 \cup \dots \cup V_m,$$

and, by the spectral decomposition theorem above applied to \mathcal{M} , w is orthogonal to each orbit, that is,

$$\sum_{j \in V_i} w_j = 0 \quad \forall 1 \leq i \leq m.$$

We need to show that (λ, \tilde{w}) is a \mathcal{G} -eigenpair. Let us write A for the adjacency matrix of \mathcal{G} (recall \mathcal{M} is a subgraph so A restricts to $A_{\mathcal{M}}$ on \mathcal{M}). We need to show $A\tilde{w} = \lambda\tilde{w}$. Given $i \in V(\mathcal{G})$, we have two cases. First, if $i \in V(\mathcal{M})$,

$$\begin{aligned} \sum_{j \in V(\mathcal{G})} [A]_{ij} \tilde{w}_j &= \sum_{j \in V(\mathcal{M})} [A]_{ij} \tilde{w}_j + \sum_{j \in V(\mathcal{G}) \setminus V(\mathcal{M})} [A]_{ij} \tilde{w}_j \\ &= \sum_{j \in V(\mathcal{M})} [A]_{ij} w_j = \lambda w_i = \lambda \tilde{w}_i, \end{aligned}$$

since \tilde{w} equals w on \mathcal{M} , and is zero outside \mathcal{M} . The second case, when $i \in V(\mathcal{G}) \setminus V(\mathcal{M})$, gives

$$\sum_{j \in V(\mathcal{G})} [A]_{ij} \tilde{w}_j = \sum_{j \in V(\mathcal{M})} [A]_{ij} w_j,$$

as before, and then we use the decomposition of \mathcal{M} into orbits,

$$\sum_{j \in V(\mathcal{M})} [A]_{ij} w_j = \sum_{k=1}^m \sum_{j \in V_k} [A]_{ij} w_j = \sum_{k=1}^m \alpha_k \sum_{j \in V_k} w_j.$$

Here we have used that the vertex i , outside the motif, connects uniformly to each orbit, that is, $A_{ij_1} = A_{ij_2}$ for all $j_1, j_2 \in V_k$, and we call this quantity α_k . Finally, recall that w is orthogonal to each orbit, to conclude

$$\sum_{j \in V(\mathcal{M})} [A]_{ij} w_j = \sum_{k=1}^m \alpha_k \sum_{j \in V_k} w_j = 0 = \lambda \tilde{w}_i. \quad \square$$

Therefore, the redundant spectrum of \mathcal{G} is the union of the redundant eigenvalues of the symmetric motifs, together with their redundant eigenvectors localised on them. Since most symmetric motifs in real-world networks are basic, most symmetric motifs in the network representation of a network measure will be basic too. Given their constrained structure, one can in fact determine the redundant spectrum of BSMs with up to few orbits, for arbitrary undirected networks with symmetry.

Redundant spectrum of a 1-orbit BSM. A BSM with one orbit is an (α, β) -uniform graph $K_n^{\alpha, \beta}$ with adjacency matrix $A_n^{\alpha, \beta} = (a_{ij})$ given by $a_{ij} = \alpha$ and $a_{ii} = \beta$ for all $i \neq j$. Then $K_n^{\alpha, \beta}$ has eigenvalues $(n-1)\alpha + \beta$ (non-redundant), with multiplicity 1, and $-\alpha + \beta$ (redundant), with multiplicity $n-1$. The corresponding eigenvectors are $\mathbf{1}$, the constant vector 1 (non-redundant), and \mathbf{e}_i , the vectors with non-zero entries 1 at position i , and -1 at position i , $2 \leq i \leq n$ (redundant). This can be shown directly by computing $A_n^{\alpha, \beta} \mathbf{1}$ and $A_n^{\alpha, \beta} \mathbf{e}_i$, and noting that $\mathbf{1}, \mathbf{e}_2, \dots, \mathbf{e}_n$ are linearly independent (although not orthogonal) and thus form an eigenbasis. Indeed, $A_n^{\alpha, \beta} \mathbf{1}$ is the vector of column sums of the matrix $A_n^{\alpha, \beta}$, which are constant $(n-1)\alpha + \beta$, and $A_n^{\alpha, \beta} \mathbf{e}_i$ is the constant 0 vector, except possibly at positions 1, which equals $\beta - \alpha$, and i , which equals $\alpha - \beta$.

Redundant spectrum of a 2-orbit BSM. A BSM with two orbits is a uniform join of the form

$$K_n^{\alpha_1, \beta_1} \xleftrightarrow{\gamma, \delta} K_n^{\alpha_2, \beta_2}. \quad (32)$$

Define $a = \alpha_1 - \beta_1$, $b = \alpha_2 - \beta_2$, $c = \gamma - \delta$, and note that $c \neq 0$: otherwise $\gamma = \delta$ and we can freely permute one orbit while fixing the other, that is, this would not be a BSM with two orbits but rather two BSMs with one orbit each. As above, let \mathbf{e}_i be the vector with non-zero entries 1 at position 1, and -1 at position i , for any $2 \leq i \leq n$.

Lemma 1. The following set of vectors is linearly independent

$$\{(\kappa_1 \mathbf{e}_i | \mathbf{e}_i), (\kappa_2 \mathbf{e}_i | \mathbf{e}_i) \mid 2 \leq i \leq n\}$$

for all $\kappa_1 \neq \kappa_2 \in \mathbb{R}$.

Proof. Define the $(n-1) \times n$ matrix

$$B_n = (\mathbf{1} \mid -\text{Id}_{n-1})$$

where $\mathbf{1}$ is a constant 1 column vector, and Id_{n-1} the identity matrix of size $n-1$. The set of vectors in the statement can be arranged in block matrix form as

$$\left(\begin{array}{c|c} \kappa_1 B_n & B_n \\ \kappa_2 B_n & B_n \end{array} \right).$$

This matrix has a minor of order $2(n-1)$,

$$\det \left(\begin{array}{c|c} -\kappa_1 \text{Id}_{n-1} & -\text{Id}_{n-1} \\ -\kappa_2 \text{Id}_{n-1} & -\text{Id}_{n-1} \end{array} \right).$$

Using that $\det \begin{pmatrix} A & B \\ C & D \end{pmatrix} = AD - BC$ whenever A, B, C, D are square blocks of the same size and C commutes with D ⁵¹, this minor equals

$$\det(-\kappa_1 \text{Id}_{n-1} - \kappa_2 \text{Id}_{n-1}) = (-1)^{n-1} (\kappa_1 + \kappa_2)^{n-1} \neq 0 \iff \kappa_1 \neq \kappa_2. \quad \square$$

Next we derive conditions for a vector $v_i = (\kappa e_i | e_i)$ to be an eigenvector of the uniform join (32), that is, $Av_i = \lambda v_i$, for some $\lambda \in \mathbb{R}$, where A is the (symmetric) adjacency matrix of the uniform join,

$$A = \begin{pmatrix} A_n^{\alpha_1, \beta_1} & A_n^{\gamma, \delta} \\ A_n^{\gamma, \delta} & A_n^{\alpha_2, \beta_2} \end{pmatrix}. \quad (33)$$

The j th entry of the vector Av is

$$\begin{aligned} \kappa \beta_1 - \kappa \alpha_1 + \delta - \gamma &= -(\kappa a + c) & j = 1 \\ \kappa \alpha_1 - \kappa \beta_1 + \gamma - \delta &= \kappa a + c & j = i \\ \kappa \delta - \kappa \gamma + \beta_2 - \alpha_2 &= -(\kappa c + b) & j = n+1 \\ \kappa \gamma - \kappa \delta + \alpha_2 - \beta_2 &= \kappa c + b & j = n+i \\ 0 & & \text{otherwise.} \end{aligned}$$

Comparing these with the entries of the vector λv_i , we obtain

$$Av_i = \lambda v_i \iff (\kappa a + c = -\lambda \kappa \text{ and } \kappa c + b = -\lambda), \quad (34)$$

The two equations on the right-hand side are satisfied if and only if $\lambda = -\kappa c - b$ and κ is a solution of the quadratic equation

$$c\kappa^2 + (b-a)\kappa - c = 0, \quad (35)$$

which has two distinct real solutions

$$\frac{(a-b) \pm \sqrt{(a-b)^2 + 4c^2}}{2c}, \quad (36)$$

since $c \neq 0$, as explained above. Together with the lemma, we have shown the following.

Theorem 5. *The redundant spectrum of a symmetric motif with two orbits $K_n^{\alpha_1, \beta_1} \xrightarrow{\gamma, \delta} K_n^{\alpha_2, \beta_2}$ is given by the eigenvalues*

$$\begin{aligned} \lambda_1 &= -b - c\kappa_1 = \frac{-(a+b) + \sqrt{(a-b)^2 + 4c^2}}{2}, \text{ and,} \\ \lambda_2 &= -b - c\kappa_2 = \frac{-(a+b) - \sqrt{(a-b)^2 + 4c^2}}{2}, \end{aligned}$$

each with multiplicity $n-1$, and eigenvectors $(\kappa_1 e_i | e_i)$ and $(\kappa_2 e_i | e_i)$ respectively, where κ_1 and κ_2 are the two solutions of the quadratic equation $c\kappa^2 + (b-a)\kappa - c = 0$, $a = \alpha_1 - \beta_1$, $b = \alpha_2 - \beta_2$ and $c = \gamma - \delta \neq 0$.

For unweighted graphs without loops, we recover the redundant eigenvalues for BSMs with two orbits predicted in¹³, as follows. We have $\beta_1 = \beta_2 = 0$, $\alpha_1, \alpha_2, \gamma, \delta \in \{0, 1\}$ and thus $a, b \in \{0, 1\}$ and $c \in \{-1, 1\}$. If $a = b$, the quadratic equation becomes $\kappa^2 - 1 = 0$ with solutions $\kappa = \pm 1$ and thus $\lambda = -b - c\kappa \in \{-2, -1, 0, 1\}$. If $a \neq b$ we can assume $a = 1$, $b = 0$ and the quadratic $c\kappa^2 - \kappa - c = 0$ has solutions φ and $1 - \varphi$ if $c = 1$, $-\varphi$ and $\varphi - 1$ if $c = -1$, where $\varphi = \frac{1+\sqrt{5}}{2}$ is the golden ratio. In either case, the redundant eigenvalues $\lambda = -b - c\kappa = -c\kappa$ are $-\varphi$ and $\varphi - 1$. Altogether, the redundant eigenvalues for 2-orbit BSMs are $\{-2, -\varphi, -1, 0, \varphi - 1, 1\}$, which equals the redundant eigenvalues RSpec_2 in the notation of¹³.

We omit the calculation of the redundant spectrum of BSMs with three (or more) orbits, as it becomes much more elaborate, and its relevance in real-world networks is less justified (for example, less than 1% of BSMs in each of our test networks, Table 1, has 3 or more orbits).

Applications

Theorem 6 (Communicability). *Let $Q(A)$ be the quotient of the adjacency matrix A of a network with respect to the partition into orbits of the automorphism group. Let $f(x) = \sum a_n x^n$ be an analytic function. Then $f(Q(A)) = Q(f(A))$.*

Proof. Call $B = Q(A)$ and recall that $AS = SB$ by Proposition 1(i), where S is the characteristic matrix of the partition. Therefore, $A^n S = S B^n$ for all $n \geq 0$ and

$$Q(f(A)) = \Lambda^{-1} S^T \left(\sum_{n=0}^{\infty} a_n A^n \right) S = \sum_{n=0}^{\infty} a_n (\Lambda^{-1} S^T A^n S) \quad (37)$$

$$= \sum_{n=0}^{\infty} a_n (\Lambda^{-1} S^T S B^n) = \sum_{n=0}^{\infty} a_n B^n = f(B), \quad (38)$$

since $\Lambda^{-1} S^T S$ is the identity matrix. \square

Theorem 7 (Shortest path distance). *Let $A = (a_{ij})$ be as above. Then*

- (i) *if (v_1, v_2, \dots, v_n) is a shortest path from v_1 to v_n and $\sigma \in \text{Aut}(\mathcal{G})$, then $(\sigma(v_1), \sigma(v_2), \dots, \sigma(v_n))$ is a shortest path from $\sigma(v_1)$ to $\sigma(v_n)$;*
- (ii) *if (v_1, v_2, \dots, v_n) is a shortest path from v_1 to v_n , and v_1 and v_n belong to different symmetric motifs, then v_i and v_{i+1} belong to different orbits, for all $1 \leq i \leq n-1$;*
- (iii) *if u and v belong to orbits U , respectively V , in different symmetric motifs, then the distance from u to v in \mathcal{G} equals the distance from U to V in the unweighted (or skeleton) quotient \mathcal{Q} .*

Proof. (i) Since automorphisms are bijections and preserve adjacency, $(\sigma(v_1), \sigma(v_2), \dots, \sigma(v_n))$ is a path from $\sigma(u)$ to $\sigma(v)$ of the same length. If there were a shorter path $(\sigma(u) = w_1, w_2, \dots, \sigma(v) = w_m)$, $m < n$, the same argument applied to σ^{-1} gives a shorter path $(u = \sigma^{-1}(w_1), \sigma^{-1}(w_2), \dots, v = \sigma^{-1}(w_m))$ from u to v , a contradiction.

(ii) Any subpath of a minimal length path is also of minimal length between its endpoints. Arguing by contradiction, there exists a subpath $p = (w_1, w_2, \dots, w_n)$ (or $p = (w_n, w_{n-1}, \dots, w_1)$), such that w_1 and w_2 belong to the same orbit, and w_n belongs to a different symmetric motif. Hence, we can find $\sigma \in \text{Aut}(\mathcal{G})$ with $\sigma(w_2) = w_1$ and fixing w_n . This implies $\sigma(p) = (\sigma(w_1), \sigma(w_2) = w_1, \sigma(w_3), \dots, \sigma(w_n) = w_n)$, a shortest path by (i), of length $n-1$. The subpath $(w_1, \sigma(w_3), \dots, w_n)$ has length $n-2$, contradicting p being a minimal length path from w_1 to w_n . (The case $p = (w_n, w_{n-1}, \dots, w_1)$ is analogous.)

(iii) Let $p = (u = v_1, v_2, \dots, v_{n+1} = v)$ be a shortest path from u to v , so that $d^{\mathcal{G}}(u, v) = n$. Let V_k be the orbit containing v_k , for all k . By (ii), $V_k \neq V_{k+1}$ for all $1 \leq k \leq n$ thus $q = (U = V_1, V_2, \dots, V_{n+1} = V)$ is a path in \mathcal{Q} and $d^{\mathcal{Q}}(U, V) \leq n$. By contradiction, assume there is a shorter path in \mathcal{Q} from U to V , that is, $(U = W_1, W_2, \dots, W_{m+1} = V)$ with $m < n$. We can construct a path in \mathcal{G} from u to v of length m (a contradiction), as follows. For each $1 \leq i \leq m$, W_i is connected to W_{i+1} in \mathcal{Q} , hence there is a vertex in W_i connected to at least one vertex in W_{i+1} . Since vertices in an orbit are structurally indistinguishable, any vertex in W_i is then connected to at least one vertex in W_{i+1} (formally, if $w \in W_i$ is connected to $w' \in W_{i+1}$ then $\sigma(w) \in W_i$ is connected to $\sigma(w') \in W_{i+1}$). This allows us to construct a path in \mathcal{G} from u to v of length $m < n$, a contradiction. \square

Let us call the *external degree* of a vertex as the number of adjacent vertices outside the motif it belongs to. The proof of the following is straightforward from the definitions.

Theorem 8 (Symmetric motif Laplacian). *A symmetric motif \mathcal{M} in \mathcal{G} induces a symmetric motif in \mathcal{L} with adjacency matrix*

$$L_{\mathcal{M}} + (d_1 I_{m_1} \oplus \dots \oplus d_k I_{m_k}), \quad (39)$$

where $L_{\mathcal{M}}$ is the ordinary Laplacian matrix of \mathcal{M} considered as a graph on its own, and d_1, \dots, d_k are the external degrees of the k orbits of \mathcal{M} of sizes m_1, \dots, m_k . (Here I_n is the identity matrix of size n and we use \oplus to construct a block diagonal matrix.)

Recall that each orbit in a BSM (in an undirected, unweighted graph) is either a complete or an empty graph.

Corollary 2 (Redundant Laplacian eigenvalues). *Let \mathcal{G} be an undirected, unweighted network. If \mathcal{M} is a 1-orbit BSM with m vertices of external degree d , then the redundant Laplacian eigenvalue induced by \mathcal{M} is d if \mathcal{M} is an empty graph, and $d+m$ if \mathcal{M} is a complete graph, in both cases with multiplicity $m-1$.*

Proof. By Theorem 8, the Laplacian of the motif in \mathcal{L} is $L_{\mathcal{M}} + dI_m$. The redundant eigenvalues of this matrix are the redundant eigenvalues of $L_{\mathcal{M}}$ (0 if \mathcal{M} is empty and m if \mathcal{M} is a complete graph, in both cases with multiplicity $m-1$) plus d . All in all, the redundant eigenvalues for 1-orbit BSMs occur at the positive integers \mathbb{Z}^+ . \square

Theorem 9 (Vertex compression). *If \mathbf{v} is a vector of length $n_{\mathcal{G}}$ that is constant on orbits, then $S\Lambda^{-1}S^T \mathbf{v} = \mathbf{v}$.*

Proof. First, note that $S^T S = \Lambda$ (this holds for any partition of the vertex set).

$$[S^T S]_{\alpha\beta} = \sum_i [S^T]_{\alpha i} [S]_{i\beta} = \sum_i [S]_{i\alpha} [S]_{i\beta} = \begin{cases} 0 & \text{if } \alpha \neq \beta, \\ |V_\alpha| & \text{if } \alpha = \beta. \end{cases}$$

As v is constant on orbits, it is already of the form $v = Sw$ for some w . Therefore $SA^{-1}S^T v = SA^{-1}S^T Sw = Sw = v$. \square

Proposition 2 (Degree centrality). *Let $B = (b_{\alpha\beta})$ be the adjacency matrix of the quotient, and $V = V_1 \cup \dots \cup V_m$ the partition of the vertex set into orbits. If $i \in V_\alpha$, then $d_i^\mathcal{G} = d_\alpha^{\mathcal{Q},out}$.*

Proof.

$$\begin{aligned} d_i^\mathcal{G} &= \sum_{j \in V} a_{ij} = \sum_{j \in V_1} a_{ij} + \dots + \sum_{j \in V_m} a_{ij} = \frac{1}{n_1} \sum_{\substack{j \in V_1 \\ i \in V_1}} a_{ij} + \dots + \frac{1}{n_m} \sum_{\substack{j \in V_m \\ i \in V_i}} a_{ij} \\ &= b_{i1} + \dots + b_{im} = d_\alpha^{\mathcal{Q},out}. \end{aligned} \quad \square$$

Weighted and directed networks

The adjacency matrix of a network can encode arbitrary weights and directions, making a general $n \times n$ real matrix A the adjacency matrix of some (weighted, directed) network. The definition of automorphism group, geometric decomposition, symmetric motif, and orbit, and their properties, as they are defined only in terms of A , carry verbatim to arbitrarily weighted and directed networks. In this setting, a symmetry (automorphism), respects not only adjacency, but weights and directions. In particular, the automorphism group is smaller than (a subgroup of) the automorphism group of the underlying undirected, unweighted network. By introducing edge weights or directions, some symmetries will disappear, removing (and occasionally subdividing) symmetric motifs and orbits, as the next results shows.

Theorem 10. *Let $A_w = (w_{ij})$ be the adjacency matrix of an arbitrarily weighted and directed network \mathcal{G}_w , and $A = (a_{ij})$ the adjacency matrix of the underlying undirected and unweighted network \mathcal{G} , that is, $a_{ij} = \text{sgn}(|w_{ij}| + |w_{ji}|)$. Consider the symmetric motifs of \mathcal{G} , respectively \mathcal{G}_w , with vertex sets M_1, \dots, M_m , respectively M'_1, \dots, M'_m . Then for every $1 \leq i \leq m'$ there is a unique $1 \leq j \leq m$ such that $M'_i \subseteq M_j$. Similarly, each vertex orbit in \mathcal{G}_w is a subset of a vertex orbit in \mathcal{G} .*

Proof. First we show that the automorphism group of \mathcal{G}_w is a subgroup of the automorphism group of \mathcal{G} . If $\sigma: V \rightarrow V$ is a permutation of the vertices, then

$$w_{\sigma(i)\sigma(j)} = w_{ij} \implies a_{\sigma(i)\sigma(j)} = a_{ij}$$

by considering two cases: $w_{ij} \neq 0$ implies $w_{\sigma(i)\sigma(j)} \neq 0$ which gives $a_{ij} = a_{\sigma(i)\sigma(j)} = 1$; $w_{ij} = 0$ implies $w_{\sigma(i)\sigma(j)} = 0$ which gives $a_{ij} = a_{\sigma(i)\sigma(j)} = 0$ (note $w_{ij} \neq 0 \iff a_{ij} = 1$). Hence $\text{Aut}(\mathcal{G}_w) \subset \text{Aut}(\mathcal{G})$, which immediately gives the result on orbits.

Let us choose essential¹¹ sets of generators S , respectively S' , of $\text{Aut}(\mathcal{G})$, respectively $\text{Aut}(\mathcal{G}_w)$, with support-disjoint partitions

$$X = X_1 \cup \dots \cup X_m, \text{ respectively } X' = X'_1 \cup \dots \cup X'_{m'}.$$

It is enough to prove the statement for these sets: given i , there is unique j such that $X'_i \subseteq X_j$. Let $x' \in X'_i \subseteq \text{Aut}(\mathcal{G}_w) \subseteq \text{Aut}(\mathcal{G})$ thus we can write $x' = h_1 \dots h_m$ with $h_k \in H_k = \langle X_k \rangle$. Since X' is an essential set of generators, there is an index j such that $h_k = 1$ (the identity, or trivial permutation) for all $k \neq j$, so that $x' = h_j$. Given any other $y' \in X'_i$, the same argument gives $y' = h_l$ for some $1 \leq l \leq m$. We claim $j = l$, as follows. The partition of X , respectively X' , above are the equivalence classes of the equivalence relation generated by $\sigma \sim \tau$ if σ and τ are not support-disjoint permutations. Since x', y' are in the same equivalence class, so are h_j and h_l and thus $j = l$. \square

The same result applies to networks with other additional structure, not necessarily expressed in terms of the adjacency matrix, such as arbitrary vertex or edge labels, by restricting to automorphisms preserving the additional structure. We obtain fewer symmetries, and a refinement of the geometric decomposition, symmetric motifs, and orbits as above. The results in this paper, although applicable in theory, become less useful in practice as further restrictions are imposed, reducing the number of available network symmetries.

Asymmetric measures

In the case of an asymmetric network measure ($F(i, j) \neq F(j, i)$), its network representation $F(\mathcal{G})$ is directed even if \mathcal{G} is not. However, $F(\mathcal{G})$ still inherits all the symmetries of \mathcal{G} , that is, every automorphism of \mathcal{G} respects weights and edge directions in $F(\mathcal{G})$. Therefore, $F(\mathcal{G})$ has the same symmetric motifs (as vertex sets) and orbits as \mathcal{G} , and the structural results in this paper apply verbatim.

Data and code availability

The datasets analysed during the current study are available at the locations stated in the caption to Table 1. The datasets generated during the current study can be found at <https://doi.org/10.6084/m9.figshare.11619792> and the code used to process the datasets at <https://bitbucket.org/rubenjsanchezgarcia/networksymmetry/>.

References

- Newman, M. *Networks: An Introduction* (Oxford University Press, Oxford, 2010).
- Watts, D. & Strogatz, S. Collective dynamics of small-world networks. *Nature* **393**, 440–442 (1998).
- Barabási, A.-L. & Albert, R. Emergence of scaling in random networks. *Science* **286**, 509–512 (1999).
- Milo, R. *et al.* Network motifs: simple building blocks of complex networks. *Science* **298**, 824–827 (2002).
- Tononi, G., Sporns, O. & Edelman, G. M. Measures of degeneracy and redundancy in biological networks. *Proc Natl Acad Sci USA* **96**, 3257–3262 (1999).
- Albert, R. & Barabási, A.-L. Statistical mechanics of complex networks. *Rev Mod Phys* **74**, 47 (2002).
- Chung, F., Lu, L., Dewey, T. G. & Galas, D. J. Duplication models for biological networks. *J Comput Biol* **10**, 677–687 (2003).
- Xiao, Y., Xiong, M., Wang, W. & Wang, H. Emergence of symmetry in complex networks. *Phys Rev E* **77**, 066108 (2008).
- Nishikawa, T. & Motter, A. E. Network-complement transitions, symmetries, and cluster synchronization. *Chaos* **26**, 094818 (2016).
- Klickstein, I. & Sorrentino, F. Generating symmetric graphs. *Chaos: An Interdiscip. J. Nonlinear Sci.* **28**, 121102 (2018). DOI 10.1063/1.5064375.
- MacArthur, B. D., Sánchez-García, R. J. & Anderson, J. W. Symmetry in complex networks. *Discrete Appl Math* **156**, 3525–3531 (2008).
- Xiao, Y., MacArthur, B. D., Wang, H., Xiong, M. & Wang, W. Network quotients: Structural skeletons of complex systems. *Phys Rev E* **78**, 046102 (2008).
- MacArthur, B. D. & Sánchez-García, R. J. Spectral characteristics of network redundancy. *Phys Rev E* **80**, 026117 (2009).
- Pecora, L. M., Sorrentino, F., Hagerstrom, A. M., Murphy, T. E. & Roy, R. Cluster synchronization and isolated desynchronization in complex networks with symmetries. *Nat Commun* **5**, 4079 (2014).
- Golubitsky, M. & Stewart, I. *The symmetry perspective: from equilibrium to chaos in phase space and physical space*, vol. 200 (Springer Science & Business Media, Basel, 2003).
- Stewart, I., Golubitsky, M. & Pivato, M. Symmetry groupoids and patterns of synchrony in coupled cell networks. *SIAM J. on Appl. Dyn. Syst.* **2**, 609–646 (2003).
- Golubitsky, M., Stewart, I. & Török, A. Patterns of synchrony in coupled cell networks with multiple arrows. *SIAM J. on Appl. Dyn. Syst.* **4**, 78–100 (2005).
- Aguiar, M. A. & Dias, A. P. S. Synchronization and equitable partitions in weighted networks. *Chaos: An Interdiscip. J. Nonlinear Sci.* **28**, 073105 (2018).
- Sorrentino, F., Siddique, A. B. & Pecora, L. M. Symmetries in the time-averaged dynamics of networks: Reducing unnecessary complexity through minimal network models. *Chaos: An Interdiscip. J. Nonlinear Sci.* **29**, 011101 (2019).
- Nicosia, V., Valencia, M., Chavez, M., Díaz-Guilera, A. & Latora, V. Remote synchronization reveals network symmetries and functional modules. *Phys Rev Lett* **110**, 174102 (2013).
- Sorrentino, F., Pecora, L. M., Hagerstrom, A. M., Murphy, T. E. & Roy, R. Complete characterization of the stability of cluster synchronization in complex dynamical networks. *Sci Adv* **2**, e1501737 (2016).
- Sorrentino, F. & Pecora, L. Approximate cluster synchronization in networks with symmetries and parameter mismatches. *Chaos: An Interdiscip. J. Nonlinear Sci.* **26**, 094823 (2016).
- Schaub, M. T. *et al.* Graph partitions and cluster synchronization in networks of oscillators. *Chaos: An Interdiscip. J. Nonlinear Sci.* **26**, 094821 (2016).
- Pecora, L. M., Sorrentino, F., Hagerstrom, A. M., Murphy, T. E. & Roy, R. Discovering, constructing, and analyzing synchronous clusters of oscillators in a complex network using symmetries. In *Advances in Dynamics, Patterns, Cognition*, 145–160 (Springer, 2017).

25. Siddique, A. B., Pecora, L., Hart, J. D. & Sorrentino, F. Symmetry- and input-cluster synchronization in networks. *Phys. Rev. E* **97**, 042217 (2018).
26. Bitbucket repository. <https://bitbucket.org/rubenjsanchezgarcia/networksymmetry/> (2019).
27. Xiao, Y., Wu, W., Pei, J., Wang, W. & He, Z. Efficiently indexing shortest paths by exploiting symmetry in graphs. In *Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology*, 493–504 (ACM, 2009).
28. Wang, J., Huang, Y., Wu, F.-X. & Pan, Y. Symmetry compression method for discovering network motifs. *IEEE/ACM Trans Comput Biol Bioinf* **9**, 1776–1789 (2012).
29. Karalus, S. & Krug, J. Symmetry-based coarse-graining of evolved dynamical networks. *Europhys Lett* **111**, 38003 (2015).
30. Nyberg, A., Gross, T. & Bassler, K. E. Mesoscopic structures and the laplacian spectra of random geometric graphs. *J Complex Networks* **3**, 543–551 (2015).
31. Do, A.-L., Höfener, J. & Gross, T. Engineering mesoscale structures with distinct dynamical implications. *New J Phys* **14**, 115022 (2012).
32. Biggs, N. *Algebraic graph theory* (Cambridge University Press, Cambridge, 1993).
33. Godsil, C. & Royle, G. F. *Algebraic graph theory* (Springer, New York, 2013).
34. Cvetkovic, D. M., Rowlinson, P. & Simic, S. *An introduction to the theory of graph spectra* (Cambridge University Press, Cambridge, 2010).
35. Brouwer, A. E. & Haemers, W. H. *Spectra of graphs* (Springer, New York, 2011).
36. Estrada, E. & Rodriguez-Velazquez, J. A. Subgraph centrality in complex networks. *Phys Rev E* **71**, 056103 (2005).
37. Estrada, E. & Hatano, N. Communicability in complex networks. *Phys Rev E* **77**, 036111 (2008).
38. Erdős, P. & Rényi, A. Asymmetric graphs. *Acta Math. Hungarica* **14**, 295–315 (1963).
39. Borgatti, S. & Grosser, T. Structural Equivalence: Meaning and Measures. *Int. Encycl. Soc. & Behav. Sci.* (2015).
40. Internet topology network dataset – KONECT. [http://konect.uni-koblenz.de/networks/\[openflights|opsahl|powergrid|ca-AstroPh|topology|wordnet-words|com-amazon|actor-collaboration|as-skitter,|roadNet-CA|livejournal-links\]](http://konect.uni-koblenz.de/networks/[openflights|opsahl|powergrid|ca-AstroPh|topology|wordnet-words|com-amazon|actor-collaboration|as-skitter,|roadNet-CA|livejournal-links]) (2016).
41. A Network Graph for Human Diseases. <http://exploring-data.com/info/human-disease-network/> (2018).
42. Yeast Interactome Project. http://interactome.dfci.harvard.edu/S_cerevisiae/ (2018).
43. Human Protein Reference Database, Release9_062910. <http://www.hprd.org/> (2018).
44. Lovász, L. Random Walks on Graphs: A Survey (1993).
45. Klein, D. J. & Randić, M. Resistance distance. *J Math Chem* **12**, 81–95 (1993).
46. Brandes, U. A faster algorithm for betweenness centrality. *J Math Sociol* **25**, 163–177 (2001).
47. saucy 3.0. <http://vlsicad.eecs.umich.edu/BK/SAUCY/> (2012).
48. GAP – Groups, Algorithms, and Programming, Version 4.8.7. <http://www.gap-system.org> (2017).
49. Liebeck, M. W. Graphs whose full automorphism group is a symmetric group. *J Aust Math Soc* **44**, 46–63 (1988).
50. Rotman, J. J. *An introduction to the theory of groups* (Springer, New York, 2012).
51. Horn, R. A. & Johnson, C. R. *Matrix analysis* (Cambridge University Press, Cambridge, 1990).

Acknowledgements

Special thanks to Ben MacArthur for support and advice during the writing of this article. The Newton Institute in Cambridge supported the author during the programme ‘Theoretical foundations for statistical network analysis’ (EP-SRC grant EP/K032208/1). Conor Wild’s undergraduate project ‘Topics in Network Symmetries’ contains (unpublished) results about the graph Laplacian in the context of symmetries generalised here. Thanks to Yamir Moreno and Emanuele Cozzo, whose ‘simple’ question inadvertently prompted this lengthy answer.

Author contributions

The author conceived the project, carried out the mathematical and numerical analysis, and wrote the article.

Competing interests

The author declares no competing interests.