Dimension-reduction of dynamics on real-world networks with symmetry

Jonathan A. Ward

j.a.ward@leeds.ac.uk

Abstract

We derive explicit formulae to quantify the Markov chain state-space compression, or lumping, that can be achieved in a broad range of dynamical processes on real-world networks, including models of epidemics and voting behaviour, by exploiting redundancies due to symmetries. These formulae are applied in a large-scale study of such symmetry induced lumping in real-world networks, from which we identify specific networks for which lumping enables exact analysis that could not have been done on the full state-space. For most networks, lumping gives a state-space compression ratio of up to 10^7, but the largest compression ratio identified is nearly 10^{12}. Many of the highest compression ratios occur in animal social networks. We also present examples of types of symmetry found in real-world networks that have not been previously reported.

1 Introduction

A wide range of phenomena can be modelled as dynamical processes on networks [72], including epidemic spreading [47, 73], opinion dynamics [29, 88, 92], the diffusion of innovations [8, 58, 59, 103], the evolution of languages [6, 10, 15] and cultural polarisation [4, 14]. Mathematical models of such processes can be formulated as Markov chains [101, 102], where the future evolution is determined by the current state, but to analyse such models it is often necessary to resort to low-dimensional approximations [7, 70, 76]. The standard approach is to make use of mean-field approximations, which at the simplest level ignore network topology completely, i.e. the system is assumed to be well mixed [76]. More complicated approximations that make use of network topology include pair approximations [98], degree-based and heterogeneous mean-field [74], moment closures [47] and approximate master equations [32, 27]. Such approximations are typically based on intuitive probabilistic reasoning rather probabilistic reasoning rather than rigorous mathematics, so it is generally difficult to quantify how well a given approximation will do, given the network or dynamical process [33].

In contrast to approximate mean-field theories, there are exact studies of full Markov chain dynamics on small networks [39, 40, 41, 52, 24] where it is possible to store the full state-space in computer memory. Constructing the full Markov chain gives access to the complete time-evolving probability distribution over state-space, which can be used to derive

detailed statistical information [102], including quantification of stochastic variation (unlike mean-field theories that only provide averages), to conduct sensitivity analysis [54, 35] and to consider more complicated models of specific scenarios [55] that include agent heterogeneity [24, 52]. Furthermore, the full probability distribution can be used in Bayesian methods, e.g. for parameter estimation and model selection.

Exact analysis of larger networks can be achieved by lumping states together to reduce the state-space size [44] using network symmetries [86, 47, 45, 101, 102]. Network symmetries result in redundancies that can be exploited in many applications, for example to reduce network size via quotients [107], or to perform efficient simulations [90]. Symmetries are intimately connected to the spectral peaks of the adjacency and Laplacian matrices [20, 21, 56], and hence they impact on a wide range of network 'structural measures' [84] and dynamical phenomena. In particular, network symmetries facilitate cluster synchronisation [75, 89] and can be used to control group consensus [49]. Symmetry in complex networks has received increasing interest recently, in part because it has been shown that many real-world networks have a significant amount of symmetry [108, 57, 5]. This is unexpected since large networks chosen at random are typically asymmetric [25]. The most common symmetries found in real-world networks are of particular types associated with "basic symmetric motifs" [56], which includes subgraphs made up of leaves, cliques and bicliques [108].

In Sections 2 and 3 we set up the mathematics and derive explicit formulae, (3–6), for the size of the reduced state-space that results from symmetry induced lumping of Markov chain dynamics on real-world networks, a problem that is extremely difficult for networks and graphs in general [43]. Practitioners need only consider Figure 1 and equations (3–6) before moving straight to the summary of the mathematical results in Section 4. In Section 5 we use the formulae to analyse over 1500 real-world networks with one hundred vertices or less, obtained from the website `www.networkrepository.com` [77]. We find that over 80% of the networks analysed have non-trivial symmetry, and the symmetry in 94% of these cases is entirely due to leaves, cliques, bicliques and repeated isomorphic components. We highlight other types of symmetry that are more complex than those previously reported in real-world networks and show that regardless of the computer being used, specific networks can always be identified whose full state-space cannot be stored in memory but whose lumped state-space can. Section 5 includes a detailed discussion of potential applications, limitations and open problems.

2   Network dynamics and symmetry

We focus on dynamical processes on finite networks described by Markov chains in which each vertex can be in one of a finite number of vertex-states and only one vertex can change vertex-state at any instant in time. We refer to such processes as single-vertex transition (SVT) models [101, 102]. An example of an SVT model is the SIR model of epidemics [47, 70], where the network captures social interactions and the vertex-states are susceptible, infected and recovered. Let $\mathcal{V}$ denote the set of vertices and $\mathcal{W}$ the set of vertex-states, then the state-space of an SVT model is the set of all functions from $\mathcal{V}$ to $\mathcal{W}$, denoted $\mathcal{S} = \mathcal{W}^{\mathcal{V}}$. Thus the vertex-

state of vertex $v \in \mathcal{V}$ in state $s \in \mathcal{S}$ is $s(v) \in \mathcal{W}$. The number of states in state-space is M^N, where M is the number of vertex-states and N is the number of vertices. While the number of states increases exponentially with the number of vertices, it is finite so we can enumerate states in state-space, i.e. $\mathcal{S} = \{s_1, s_2, \ldots, s_{M^N}\}$. Let $X(t) = (x_1(t), x_2(t), \ldots, x_{M^N}(t))^{\mathrm{T}}$ be the corresponding time-dependent probability distribution over states in $\mathcal{S}$, then the evolution of $X(t)$ is given by the forward Kolmogorov or master equation [46]

$$\dot{X} = Q^{\mathrm{T}} X,$$

where Q is an $M^N \times M^N$ matrix called the infinitesimal generator whose ijth component describes the transition rate from the state s_i to the state s_j for $i \neq j$, and whose diagonal entries ensure the rows sum to zero (i.e. the magnitude of Q_{ii} is the transition rate out of state s_i). The transition rates in SVT models only depend on the vertex-states of nearest neighbours [101, 102]. In this paper we will focus on the structure of the network and state-space and we will not need the infinitesimal generator directly.

A finite Markov chain is lumpable if there is a partition $\mathcal{L} = \{\mathcal{L}_1, \mathcal{L}_2, \ldots, \mathcal{L}_r\}$ of state-space on which the Markov property is preserved [44] and it has been shown that network symmetries can be used to lump SVT models [86, 101, 102]. Network symmetries, or graph automorphisms, are permutations of the vertices that leave the edge set unchanged. More precisely, a symmetry of a network $\mathcal{G}$ with vertices $\mathcal{V}$ and edges $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$, is a bijection $g : \mathcal{V} \to \mathcal{V}$ such that $(u, v) \in \mathcal{E}$ if and only if $(g(u), g(v)) \in \mathcal{E}$. We use the shorthand $gv = g(v)$. The set of symmetries of a network $\mathcal{G}$ form a permutation group $G = \mathrm{Aut}(\mathcal{G})$ called the automorphism group of $\mathcal{G}$ [34], which we refer to as the symmetry group of the network. It has been shown that the symmetries in typical real-world complex networks can be decomposed into a product of subgroups that act independently of one another [57]. Mathematically this means the symmetry group G of a network can be written as a direct product

$$G \cong H_1 \times H_2 \times \cdots \times H_m. \tag{1}$$

The right-hand-side of (1) is known as the geometric decomposition of G and each H_i as a geometric factor of G. Each geometric factor H_i is associated with a distinct subset of vertices $\mathcal{V}_i$, so that all pairs of subsets $\mathcal{V}_i$ and $\mathcal{V}_j$, $i \neq j$, are disjoint. We refer to the set of vertices $\mathcal{V}_i$ as a geometric component and the induced subgraph on $\mathcal{V}_i$ as a Symmetric Motif (SM). We use $\mathcal{V}_0$ to denote the subset of vertices that are fixed by the symmetry group, i.e. not moved by any network symmetry.

The symmetry group G of a network can also act on states in state-space in such a way that vertex-states are permuted rather than vertices. More precisely if $u \in \mathcal{V}$, $g \in G$ and $s \in \mathcal{S}$ then

$$(gs)(u) = s(g^{-1}u), \tag{2}$$

i.e. the vertex-state of u in gs is the same as the vertex-state of $g^{-1}u$ in s. This action defines an equivalence relation on states that partitions state-space into disjoint sets of states, called orbits [28], and this partition, denoted $\mathcal{S}/G$, is a lumping of $\mathcal{S}$ [101, 45, 102]. In Appendix A we prove that there is a decomposition of state-space that reflects the geometric

decomposition of the symmetry group. This means that we can focus on the states of each geometric component rather than the network as a whole. The states of a geometric component are $\mathcal{S}_i = \mathcal{W}^{\mathcal{V}_i}$ and we refer to each $\mathcal{S}_i$ as a state-space factor. Thus if we can compute the orbit partition of each state-space factor then we can combine these together to determine the orbit partition of the whole network. It follows from the results in Appendix A that if ρ_i is the number of orbits of the ith state-space factor, then the number of orbits of $\mathcal{S}$ is

$$\rho := |\mathcal{S}/G| = \prod_{i=0}^{m} \rho_i. \tag{3}$$

Thus we can compute the size of the lumped state-space, ρ, and the compression ratio, M^N/ρ, from the lumping of each geometric component.

3 Orbit representatives of typical symmetric motifs

The symmetries of typical real-world networks are restricted to only a few different types, examples of which are illustrated in Fig. 1. We refer to the SM illustrated in Figure 1 (a) as SM (a) and similarly for the other SMs in the figure. We refer to SMs (a)–(f) collectively as typical SMs and any other type of SM as atypical. Typical SMs include: (a) leaves, (b) cliques, (c) 'n-stars', (d) mirror symmetries and (e) regular trees. Repeated isomorphic components consisting of a single SM of type (a)–(e), an example of which is illustrated in Fig. 1 (f), are also common in the data that we analyse so we include this as a type of typical SM. For each of these typical SMs, we now describe how to construct a set of orbit representatives, i.e. a sub-set of states such that each is in a distinct state-space orbit, from which we can determine the number of orbits of the corresponding state-space factor, ρ_i, and construct the lumped Markov chain. The construction of orbit representatives can be broken down into three distinct categories of typical SMs, namely Basic Symmetric Motifs (BSMs), height regular trees and isomorphic components of typical SMs. These are addressed in the corresponding sections below. In the following we make repeated use of the number of combinations of n items chosen from k possibilities (an n-combination) with repetition, which is given by $C(n + k - 1, n)$ [2], where $C(n, k)$ is the binomial coefficient n choose k. For context, $C(n + k - 1, n)$ is the number of non-negative integer solutions to the linear Diophantine equation $x_1 + x_2 + \cdots + x_k = n$ [12] and hence it is also the number of states in a stochastic compartmental model with k compartments and population size n.

Basic symmetric motifs

BSMs have $k \geq 1$ orbits each with n vertices that are permuted simultaneously by the symmetry group S_n, so that the vertices in pairs of orbits are either connected in a one-to-one or one-to-all-but-one fashion [57, 56]. This includes SMs (a)–(d) in Figure 1 and the set of orbit representatives can be constructed in essentially the same way for each of these. Assign a vertex from each of the k orbits to the set $\mathcal{U}$, then the set of possible states for the vertices in $\mathcal{U}$ is $\mathcal{W}^{\mathcal{U}}$. A set of orbit representatives can be constructed by selecting all
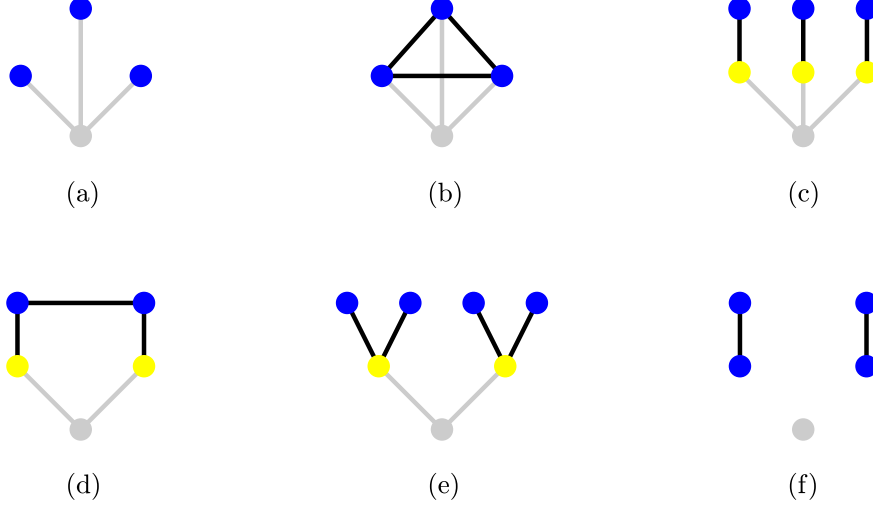
Figure 1: Examples of the main types of symmetry observed in real-world networks, referred to here as typical SMs, indicated by coloured vertices and black edges. Grey vertices are fixed (not permuted by any symmetry) and represent the point of connection (if any) to a larger network. SM vertices are coloured by orbit, i.e. there is a permutation that moves one vertex to any other with the same colour. Grey edges link to the fixed vertices. The number of leaf vertices in SM (a) and the size of the clique in SM (b) are denoted by n; (a) and (b) illustrate the case $n = 3$ and $k = 1$ in (4). SM (c) is an n-star, having k vertices on each of the n paths from the centre of the star (the fixed vertex); (c) illustrates the case $n = 3$ and $k = 2$ in (4). The number of vertices permuted in SM (d) is $2k$; (d) corresponds to the case $n = 2$ and $k = 2$ in (4). The number of branches connected to the fixed node in SM (e) is denoted η_2 and the number of leaves on each branch is denoted η_1 in (5); (e) is the case $\eta_1 = \eta_2 = 2$. The number of isomorphic components in SM (f) is denoted n in (6); (f) illustrates the case $n = 2$.

possible combinations of n states chosen with repetition from the M^k possibilities in $\mathcal{W}^{\mathcal{U}}$. Thus the number of orbits of a BSM is

$$\rho_i = C(n + M^k - 1, n). \tag{4}$$

For clarity, we now relate the general construction of orbit representatives to each of the BSMs in turn. The symmetry groups of SMs (a) and (b) are both the symmetric group S_n acting naturally on n vertices (i.e. all possible permutations of the n vertices), so $k = 1$ and both SMs (a) and (b) correspond to the case $n = 3$. There is a permutation in S_n that maps a state to any other with the same number of vertices in each vertex-state. Thus the set of all possible combinations of n vertex-states chosen with repetition from the M possibilities in $\mathcal{W}$ forms a set of orbit representatives. The symmetry group of SM (c) is the symmetric group S_n having k orbits on vertices, where each orbit has n vertices. SM (c) corresponds to the case $k = 2$ and $n = 3$. The symmetry group of SM (d) is the cyclic group of order two, C_2, which is isomorphic to S_2, hence $n = 2$ and (4) reduces to $M^k(M^k + 1)/2$.

5

Height-regular trees

The symmetry group of SM (e) is $S_{\eta_1} \wr S_{\eta_2}$, where $\wr$ denotes the wreath product [81]. This SM consists of η_2 stars, each having η_1 leaves, and the central vertex of each star is connected to a fixed vertex. We call this a height-regular tree of height two. SM (e) is the case where $\eta_1 = \eta_2 = 2$. The wreath product in the symmetry group $S_{\eta_1} \wr S_{\eta_2}$ captures the fact that the η_1 leaves of each of the stars can be permuted according to the symmetric group S_{η_1}, but also that the η_2 stars can be permuted according to S_{η_2}. Thus $S_{\eta_1} \wr S_{\eta_2}$ consists of $(\eta_1!)^{\eta_2}\eta_2!$ permutations. A set of orbit representatives can be identified via a recursive procedure described in detail in Appendix B. In short, this procedure computes the orbit representatives for one of the stars, and then η_2 of these are chosen with repetition to form the set of orbit representatives for SM (e). Thus the number of orbits of this SM is

$$\rho_i = C(\eta_2 + MC(\eta_1 + M - 1, \eta_1) - 1, \eta_2). \tag{5}$$

Isomorphic components of typical symmetric motifs

Suppose that there are n isomorphic components and that H is the symmetry group of any one of the components in isolation (note that H could be the trivial group), then the symmetry group of the n isomorphic components together is $H \wr S_n$. This captures the fact that it is possible to permute any of the vertices in an isomorphic component according to H independently of the other components, and one can also permute the n components according to S_n. SM (f) is the case $n = 2$ and $H \cong C_2$. Let $\mathcal{R}_{\cong}$ denote a set of orbit representatives for a single component, then a set of orbit representatives can be determined for the collection of isomorphic components by choosing n states from $\mathcal{R}_{\cong}$ with repetition. Thus if $R = |\mathcal{R}_{\cong}|$ then the number of orbits of this SM is

$$\rho_i = C(n + R - 1, n). \tag{6}$$

We consider symmetry due to repeated isomorphic components as typical if each repeated component only has symmetry due to SMs (a)–(e), in which case we can determine $\mathcal{R}_{\cong}$ using the methods described for BSMs and height regular trees.

4 Summary of mathematical results

The mathematical results that we have presented use group theory and combinatorics, however a practitioner need only make use of the formulae (3–6) that we have derived. Figure 2 illustrates how these results may be used in practice. Given a network, the first step is to compute the network symmetries and identify typical SM geometric components (see [57] for computational tools). Once these have been identified, the formulae (3–6) can be applied directly to the corresponding SMs to compute the total number of orbit representatives and the corresponding states. Figure 2 illustrates this for a network with 2 fixed vertices and SMs (a), (b), (c) and (f).
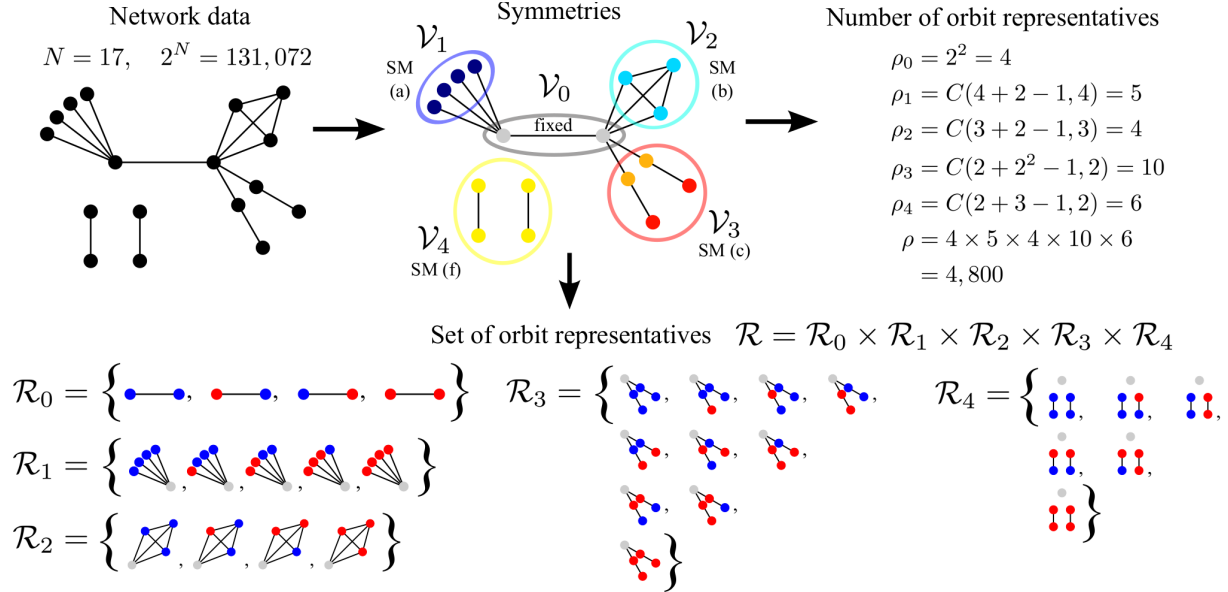
Figure 2: Example application of the mathematical results. The network we consider has $N = 17$ nodes, which with $M = 2$ corresponds to 131,072 states. There are two fixed vertices in $\mathcal{V}_0$ and four SMs: geometric components $\mathcal{V}_1, \mathcal{V}_2, \mathcal{V}_3, \mathcal{V}_4$ correspond to SMs (a), (b), (c) and (f) respectively. Having identified the SM symmetries, we can compute the number of orbit representatives using Eq. (3–6), yielding $\rho = 4,800$. We also illustrate the orbit representatives for each state-space factor, the Cartesian product of which yields the orbit representatives of the whole network.

5 Application to real-world network data

We now apply the methods of computing orbit representatives described in the previous section to real-world network data. In Table 1 in Appendix C we present information about fifteen large real-world networks. The key observation is that the number of vertices fixed by the symmetry group, N_{fixed}, is large in all cases, and the lumped state-space will be larger than $M^{N_{\text{fixed}}}$ when $N_{\text{fixed}} < N$. Thus the size of the lumped state-space of these networks remains far beyond practical computation.

We focus instead on 1524 real-world networks with one hundred vertices or less obtained from the website `networkrepository.com` [77]. We downloaded all networks from the repository having $N \leq 100$ vertices on three occasions during 2018 and 2019, resulting in a total of 1524 networks. For each network, we removed all isolated nodes and self-edges, and made the networks undirected and unweighted. The format of the repository has changed since our first download and naturally new graphs get added over time, thus we have made the processed network data available online [100]. We used the program saucy [42] to compute generators of the symmetry groups of the networks and processed the symmetry groups using the computational algebra program GAP [30]. We found that 1227 of the 1524 networks, more than 80%, have non-trivial symmetry. This supports the conclusions from previous
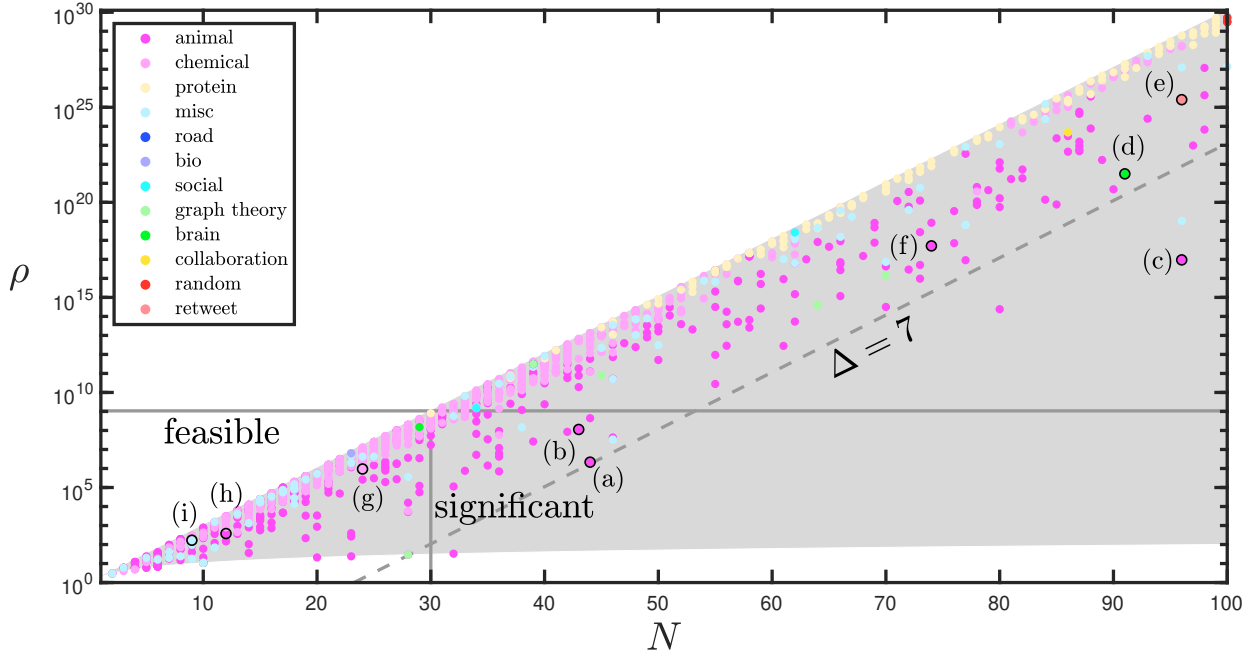
Figure 3: Number of orbit representatives ρ against number of vertices N. Light grey shading indicates the range of possible values; markers indicate individual networks, coloured by type. The vertical axis is logarithmic with base 10. The horizontal dark grey solid line indicates the threshold $\tau = 10^9$ and the vertical dark grey solid line indicates the corresponding number of vertices. The dark grey dashed line corresponds to significance $\Delta = 7$. (Note that the network considered in Figure 2 has $\Delta = 1.44$.) Highlighted markers correspond to networks in Figure 4.

studies [57, 5] that most real-world networks have symmetry.

Our data set contains twelve different types of networks according to the classification on `networkrepository.com`, including 'social', 'brain' and 'protein', but the majority of networks (79%) were either chemical or animal social networks. The animal social networks can also be found on the 'Animal Social Network Repository'[1] [83]. Most of the networks have between 10 and 50 vertices, with an average of 38 vertices. Of the 1227 networks with non-trivial symmetry, 1151 (nearly 94%) have symmetry entirely due to SMs (a)–(f), of which roughly 93% have SMs of types (a) and (b), 25% have type (d) and 27% have type (e); only 4 networks had type (c). The average fraction of vertices moved in networks with nontrivial symmetry is 0.4. More detailed network statistics can be found in Appendix D.

We now turn to the computation of the number of orbit representatives, ρ, for the case where the number of vertex-states is $M = 2$, which corresponds to most SVT models [101, 102] and includes the SIS model of epidemics [47] and the voter model [88] of opinion dynamics. In Figure 3 we plot (N, ρ) for each network with non-trivial symmetry, where the colour indicates the type of network. The ρ-axis is scaled logarithmically and

---

[1]`https://github.com/bansallab/asnr/`

8

the light grey region indicates the area of possible values of (N, ρ), the largest value of ρ being 2^N, corresponding to no symmetry, and the smallest being $N + 1$, corresponding to a complete graph. On a particular computer there will be a limit to the size of the state-space that can be used, which we denote by τ, and we call cases where $\rho \leq \tau$ feasible and we call feasible cases where $2^N > \tau$ significant. The dark grey horizontal line in Figure 3 corresponds to the feasible threshold $\tau = 10^9$, which is indicative of the size of state-space that can be stored in memory on a typical laptop computer at the time of writing[2]. The vertical dark grey line indicates the number of vertices that corresponds to the threshold $\tau = 10^9$, which is roughly $N = 30$. At the feasible threshold level of $\tau = 10^9$, there are a total of 62 networks with significant lumping. To quantify the amount of lumping, we introduce the relative significance $\Delta = N \log_{10}(2) - \log_{10}(\rho)$, which measures the logarithmic reduction in the size of state-space. Thus 10^Δ is the compression ratio. The dark grey dashed line in Figure 3 corresponds to $\Delta = 7$, i.e. the full state-space is 10^7 times larger than the lumped state-space. Under exponentially increasing computer power and memory (e.g. Moore's law), we would expect the threshold τ to increase at a constant rate and so Δ can be thought of as a proxy for how long a particular network would be classed as having significant lumping. Moreover, any network with non-trivial symmetry will at some point in time (possibly in the past) have significant lumping. Crucially, Figure 3 shows that for any value of τ within the limits of our data, we can find specific examples of networks that have significant lumping.

In Figure 4 we plot examples of some of the networks in our data set and these are labelled (a)–(i) in correspondence to those labels in Figure 3. In each network, fixed vertices are coloured light grey and other vertices of the same colour are in the same vertex orbit (i.e. there is a permutation that takes a vertex to any other in the same orbit). The majority of networks with significant lumping for $\tau = 10^9$ are animal social networks [83] and networks (a) and (b) exemplify the sorts of structures that give rise to this significant lumping. Network (a) corresponds to the animal social network 'mammalia-voles-bhp-trapping-63' whose large amount of symmetry comes from repeated isomorphic components, including pairs, triangles and paths of length two. Network (b) is the animal social network 'mammalia-bat-roosting-indiana' whose symmetry is due to cliques. Network (c) corresponds to the animal social network 'mammalia-voles-rob-trapping-51' and has the largest relative significance of $\Delta = 11.93$, i.e. the full state-space of this network is nearly 10^{12} times as large as the lumped state-space. Many of the other examples of networks with high Δ are from animal social networks with symmetry due to isomorphic components, but networks (d) and (e) are two other examples with relatively high Δ. Network (d) is the brain network 'bn-macaque-rhesus_brain_2' [1, 13, 99], which has several instances of S_n symmetry due to bicliques and has $\Delta = 5.90$. Network (e) is the retweet network 'rt-retweet' [78, 79] whose symmetry is due to leaves and has $\Delta = 3.50$. This is perhaps to be expected in retweet networks, which capture the spread of information, and so are likely to be tree-like with few short cycles.

---

[2]If each state in state-space is stored as a long integer (4 bytes), then 4GB of memory are needed to store 10^9 states.

(a) mammalia-voles-bhp-trapping-63
$N = 44$, $\log_{10}(\rho) = 6.3$, $\Delta = 6.9$

(b) mammalia-bat-roosting-indiana
$N = 43$, $\log_{10}(\rho) = 8.1$, $\Delta = 4.9$

(c) mammalia-voles-rob-trapping-51
$N = 96$, $\log_{10}(\rho) = 17.0$, $\Delta = 11.9$

(d) bn-macaque-rhesus_brain_2
$N = 91$, $\log_{10}(\rho) = 21.5$, $\Delta = 5.9$

(e) rt-retweet
$N = 96$, $\log_{10}(\rho) = 25.4$, $\Delta = 3.5$

(f) mammalia-voles-bhp-trapping-47
$N = 74$, $\log_{10}(\rho) = 17.7$, $\Delta = 4.6$

$S_3 \times S_3 \wr C_2$

$C_2 \wr C_2$

(g) ENZYMES_g186
$N = 24$, $\log_{10}(\rho) = 6.0$, $\Delta = 1.3$

(h) reptilia-tortoise-network-pv-2010
$N = 12$, $\log_{10}(\rho) = 2.6$, $\Delta = 1.0$

(i) cage4
$N = 9$, $\log_{10}(\rho) = 2.2$, $\Delta = 0.5$
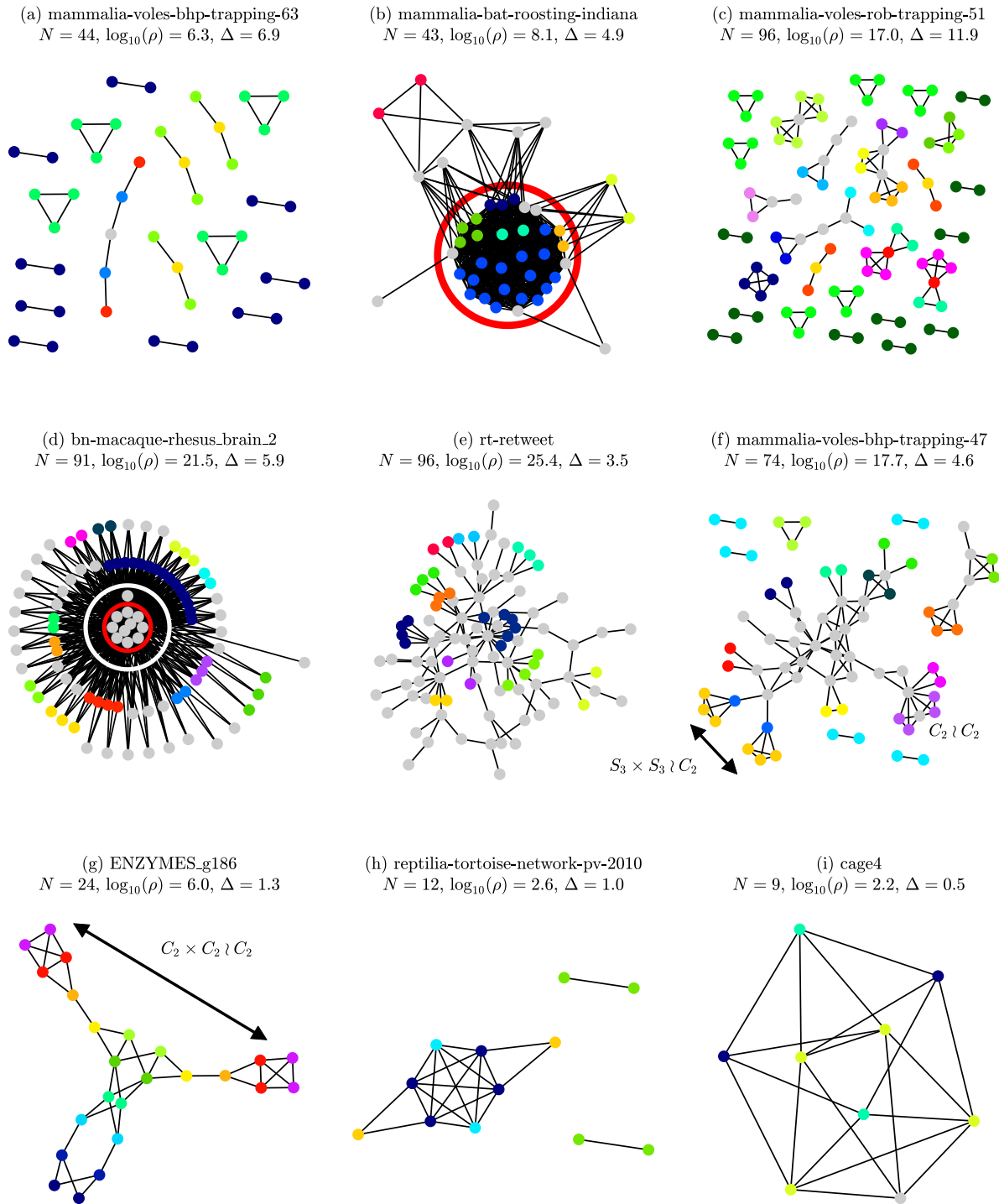
$C_2 \times C_2 \wr C_2$

Figure 4: Examples of networks with the most significant lumping and complex symmetry. Grey vertices are fixed and other vertices are coloured by orbit. The labels (a)–(i) of these networks correspond to those in Figure 3. Networks (a) and (b) are two of the most significant for $\tau = 10^9$. Nodes inside the red circle in (b) are fully connected. Network (c) has the largest relative significance. Networks (d) (brain) and (e) (Twitter) have large relative significance out of those that are not animal social networks. Nodes inside the red circle in (d) are fully connected; there are no connections between nodes outside of the white circle. Networks (f)–(i) have complex symmetries that have not been reported before in real-world networks.

10

We now give some examples of networks with significantly more complex symmetric structures than those previously reported [57]. There are 54 real-world networks with atypical SMs, of which 41 have geometric components with a "mirror" symmetry, i.e. the geometric component can be partitioned into a pair of isomorphic motifs with symmetric connections to one or more fixed vertices, and in some cases with connections between the mirror symmetric motifs. The geometric factor for such components is of the form $H \wr C_2$, where we have found that H is either C_2, S_3 or $C_2 \times C_2$, examples of which can be seen in networks (f) and (g), the animal social network 'mammalia-voles-bhp-trapping-47' and the chemical network 'ENZYMES_g186' respectively. There are also 15 networks with other types of complex symmetry. Network (h) is the animal social network 'reptilia-tortoise-network-pv-2010', in which the geometric factor with blue vertices has D_8 symmetry with three vertex orbits (D_8 symmetry on one vertex orbit also arises via a four-cycle in other networks). Network (i) is the network 'cage4' [97] and is derived from a model of DNA electrophoresis. This network has $C_2 \times C_2$ symmetry, in which the blue vertices can be permuted independently of the green vertices, but the yellow vertices are moved by all permutations.

Discussion

In this paper, we have shown how network symmetries can be used to reduce the size of the state-space of dynamical processes on real-world networks described by Markov chains. This approach makes use of the special structure of the symmetries present in real-world networks, which allowed us to obtain explicit formulae for the size of the reduction, something that is extremely difficult to compute for networks in general [43]. We applied this method to more than 1500 real-world networks that have one-hundred vertices or less, we found 62 networks with significant lumping and illustrated examples of more complex types of real-world network symmetry than previously reported.

We have observed that the most significant lumping arises in animal social networks (although these are also the most prevalent type of network in our data). Dynamical processes on animal social networks have been used to model the spread of diseases [82, 106] and parasites [96] using epidemic models, and the spread of social information via diffusion [3]. There are also network dynamics models of evolutionary dynamics [51, 105, 11], social evolution, co-evolution, population stability, dispersal and invasion [50]. Studies that use observational data typically involve relatively small population sizes and Monte Carlo simulations [106]. Thus such studies could benefit from the lumping techniques discussed in this paper, particularly the use of exact distributions for Bayesian statistics.

We stress that there are limitations to the use of exact lumping, in particular it is not a technique that can be applied directly to typical large networks and when it can be used, our study suggests that it may only allow one to consider an additional 20 nodes at most. However, there is real value in studying small networks exactly, where these gains are significant, particularly for modelling hospital- and healthcare-acquired infections [54, 53], and the effects of peripatetic health-care workers [93]. We also note that often the most significant lumping arises in networks which have multiple isomorphic components. In such cases, using

the lumped state-space to directly compute the evolution of the full probability distribution may not be the most efficient approach, particularly if only basic summary statistics are required, like the mean and variance of the number of infected individuals, since the independence of each component can be exploited. However, the lumping approach described facilitates the computation of more detailed statistics, for example the probability of a given number of infected individuals, and can be used to construct efficient algorithms.

This work suggests a number of unsolved mathematical problems. Firstly, what sort of networks or graphs give rise to highly significant lumping? It is easy to show from Pólya enumeration [81] that a lower bound on the number of state-space orbits for a network with N vertices and symmetry group G is $M^N/|G|$. Thus clearly networks with significant lumping must have very large symmetry groups. We conjecture that the next smallest lumping after the complete graph is the star graph, but how does this sequence continue? Note that the hierarchy of graph symmetry [34], namely vertex-transitive, arc-transitive and distance-transitive, does not necessarily correspond to significant lumping. For example, cycle graphs are vertex-transitive and the symmetry group of an N cycle is the dihedral group of order $2N$. For these graphs it can be shown that the number of state-space orbits is asymptotic to $M^N/2N$ for large N, i.e. the lumped state-space is a similar size to the unlumped state-space. We have found that this is also qualitatively true for other vertex-transitive graphs.

For certain families of graph it is possible to determine formulae for the number of state-space orbits [86, 45]. Moreover, using the results for SM (e), we can construct networks for which the number of state-space orbits can scale like a polynomial in the number of nodes N with any degree m, by having m isomorphic cliques of size N/m. Related asymptotic results might shed light on the structure of the (N, ρ) space, e.g. the sparsity of 'realisable' pairs. In addition to complete graphs, star graphs and multiple isomorphic components, other families of graph that have relatively small numbers of state-space orbits include bipartite graphs, multipartite graphs and lexicographic products of complete graphs [38]. The symmetry groups of these graphs consist of direct and wreath products of symmetric groups — similar to what we have observed in real-world networks. This suggests that the symmetries of real-world networks are exactly the right sorts of symmetry to give rise to significant lumping.

We might also consider whether there are computationally efficient algorithms to compute the number of state-space orbits and a set of orbit representatives for an arbitrary network. We found that a naive application of Pólya enumeration in the computational algebra package GAP [30] was limited by the size of the group, so it was not possible to compute the number of state-space orbits for precisely those networks that have significant lumping. Whilst computing the number of state-space orbits may be a very difficult problem in general, recent advances in algorithms to compute generators for symmetry groups of graphs are encouraging [42]. The development of general algorithms to compute orbit representatives may benefit from consideration of how do to this theoretically, for example by making use of methods from permutation group theory. When analysing a permutation group, one can reduce the scale of the problem through the connections between intransitive, imprimitive and primitive groups, and it may be possible to use this approach to construct a set of orbit representatives.

So far we have also focused on reductions in state-space with no loss of information. However, one can also derive reduced dimensional Markov chains by lumping states together in a non-exact way, but where the transition rates between lumped states is optimal in some sense. Recent attempts to do this have made use of 'local symmetries' [45] and reductions to 'Markov Population Models' [36]. We have also focused on dynamics on networks that are Markovian, but non-Markovian models are important when modelling epidemics on networks, since the distribution of recovery times of real diseases are not necessarily exponential [19]. There is a significant body of work on approximations of non-Markovian epidemic dynamics on networks [48, 91, 80, 26, 85] and there are also non-Markovian models of temporal networks [60] and infectious diseases on temporal networks [19]. We expect that network symmetries will also be relevant to non-Markovian models since network symmetry is a consequence of invariance under relabelling of vertices. Alternative notions of symmetry such as 'stochastic invariance' [31] and isospectral 'latent symmetries' [87] may also be useful when modelling dynamics on networks.

In summary, we have derived explicit formulae for the number of state-space orbits of typical symmetric motifs and used these to study a large number of real-world networks, finding many examples with significant state-space compression. This is a remarkable property of real-world networks and is due to the special types of symmetry present.

Data Accessibility

Acknowledgements

A   State-Space Decomposition

The support of an automorphism g is the set of vertices permuted by g, denoted $\text{supp}(g) = \{v \in \mathcal{V} \mid gv \neq v\}$, and similarly the support of the automorphism group of a network is the union of the supports of its automorphisms. We say that two automorphisms g and h are support disjoint if their respective supports are disjoint. Similarly, if H_1 and H_2 are subgroups of G, then we we say that H_1 and H_2 are support disjoint if all pairs of automorphisms $h_1 \in H_1$ and $h_2 \in H_2$ are support disjoint. Thus the geometric decomposition (1) is a direct product of support disjoint subgroups.

We can decompose state-space in a way that reflects the geometric decomposition of the automorphism group, allowing us to focus on the state-space associated with each individual

geometric component. Let $\mathcal{V}_0$ be the set of vertices fixed by G, then $\{\mathcal{V}_0, \mathcal{V}_1, \ldots, \mathcal{V}_m\}$ is a partition of $\mathcal{V}$, i.e. $\mathcal{V}_i \cap \mathcal{V}_j = \emptyset$ for each $i \neq j$. For $0 \leq i \leq m$, let $\mathcal{S}_i = \mathcal{W}^{\mathcal{V}_i}$, and we refer to each $\mathcal{S}_i$ as a state-space factor. Mirroring the geometric decomposition of G into a direct product of geometric factors in (1), we can similarly decompose $\mathcal{S}$ into a Cartesian product of state-space factors,

$$\mathcal{P} = \mathcal{S}_0 \times \mathcal{S}_1 \times \cdots \times \mathcal{S}_m. \tag{7}$$

The automorphism group G then acts on $\mathcal{P}$ in the natural way. Specifically, for every $g \in G$ we can write $g = (h_0, h_1, \cdots, h_m)$, where $h_i \in H_i$ and H_0 is the trivial group consisting of just the identity permutation (since the vertices in $\mathcal{V}_0$ are fixed by every $g \in G$). Then for $g \in G$ and $p = (p_0, p_1, \ldots, p_m) \in \mathcal{P}$, we define the action of G on $\mathcal{P}$ to be

$$gp = (h_0 p_0, h_1 p_1, \ldots, h_m p_m), \tag{8}$$

where the action of h_i on p_i is as in (2).

We now show that there is an equivalence between the action of G on the full-state space $\mathcal{S}$ and on the Cartesian product decomposition $\mathcal{P}$. The notion of equivalence we need comes from permutation group theory [22].

Definition A.1. *Let X and Y be finite sets and let G be a permutation group with an action defined on X and an action defined on Y. We say that the actions of G on X and Y are equivalent if there is a bijection $f : X \to Y$ that satisfies*

$$f(gx) = gf(x)$$

for all $g \in G$ and $x \in X$.

Theorem A.1. *Let G be the automorphism group of a network with geometric decomposition given by (1), state-space $\mathcal{S}$ and Cartesian product of state-space factors $\mathcal{P}$, as defined in (7). Then the actions of G defined on $\mathcal{S}$ by (2) and on $\mathcal{P}$ by (8) are equivalent.*

Proof. Let $f : \mathcal{S} \to \mathcal{P}$ such that for $s \in \mathcal{S}$,

$$f(s) = (s_0, s_1, \ldots, s_m),$$

where for each $0 \leq i \leq m$ and $v \in \mathcal{V}_i$, $s(v) = s_i(v)$. The map f is one-to-one since if $x, y \in \mathcal{S}$ and $f(x) = f(y)$, then for each $0 \leq i \leq m$ and $v \in \mathcal{V}_i$, $x_i(v) = y_i(v)$ and so $x = y$. The map f is onto since $|\mathcal{S}| = |\mathcal{P}|$ and f is one-to-one. For $g \in G$, we can write $g = (h_0, h_1, \ldots, h_m)$ and so

$$\begin{aligned}
f(gx) &= (h_0 x_0, h_1 x_1, \ldots, h_m x_m), \\
&= g(x_0, x_1, \ldots, x_m), \\
&= gf(x).
\end{aligned}$$

□

14

It follows that there is a one-to-one correspondence between the orbits of $\mathcal{S}$ and the orbits of $\mathcal{P}$. Moreover, each orbit in $\mathcal{P}$ corresponds to the Cartesian product of an orbit from each of the state-space factors.

Theorem A.2. *For $s \in \mathcal{S}$ and f defined in the proof of Theorem A.1, let $f(s) = (s_0, s_1, \ldots, s_m)$, $\mathcal{C} = Gs$ be the orbit of s, and $\mathcal{C}_i = H_i s_i$ be the orbit of s_i for each $0 \leq i \leq m$. Then*

$$\{f(x) \mid x \in \mathcal{C}\} = \mathcal{C}_0 \times \mathcal{C}_1 \times \cdots \times \mathcal{C}_m.$$

Thus the orbit of $s \in \mathcal{S}$ corresponds to the Cartesian product of an orbit from each of the state-space factors.

Proof. Let $\mathcal{C}' = \prod_i \mathcal{C}_i$ and
$$F(\mathcal{C}) = \{f(x) \mid x \in \mathcal{C}\}.$$

Suppose $x \in \mathcal{C}$ and $f(x) = (x_0, x_1, \ldots, x_m)$. Thus there is $g \in G$ such that $x = gs$ and consequently $f(x) = gf(s)$. From the geometric decomposition of G, we can write $g = (h_0, h_1, \ldots, h_m)$ and so for each $0 \leq i \leq m$, we have $x_i \in \mathcal{C}_i$ since $x_i = h_i s_i$. Thus $f(x) \in \mathcal{C}'$ and so $F(\mathcal{C}) \subset \mathcal{C}'$. Suppose $(x_0, x_1, \ldots, x_m) \in \mathcal{C}'$ then, from the geometric decomposition of G, for each $0 \leq i \leq m$ there is $h_i \in H_i$ such that $x_i = h_i s_i$ and from the definition of the action of G on $\mathcal{P}$

$$(x_0, x_1, \ldots, x_m) = (h_0 s_0, h_1 s_1, \ldots, h_m s_m),$$
$$= gs.$$

Since $gs \in \mathcal{C}$ we have that $\mathcal{C}' \subset F(\mathcal{C})$ and consequently $F(\mathcal{C}) = \mathcal{C}'$. $\qquad \square$

B Orbit representatives of height-regular trees

We now give a detailed description of how a set of orbit representatives can be constructed for height-regular trees via a recursive procedure that makes use of the results for SM (a). This process is illustrated in Figure 5 for SM (e). Recall that the leaves of a tree have height zero. In step (i) of the recursive procedure, we pick one vertex at height one and consider the state-space of its η_1 children. Since the child vertices are only connected to the parent vertex, we can use the method described for SM (a) to determine the orbit representatives of their state-space, of which there are $C(\eta_1 + M - 1, \eta_1)$. We denote this set of orbit representatives $\mathcal{R}_1$. In step (ii), we include the parent node, resulting in a star graph, and determine the orbit representatives of the corresponding state-space. This is simply each possible vertex-state appended to each of the states in $\mathcal{R}_1$, so there are $MC(\eta_1 + M - 1, \eta_1)$ orbit representatives. We denote this set of orbit representatives by $\mathcal{R}_1^*$. This star can be permuted with any of the $\eta_2 - 1$ other isomorphic stars at the same height. In step (iii) we determine the orbit representatives of the SM by choosing all possible combinations of η_2

Orbit representatives

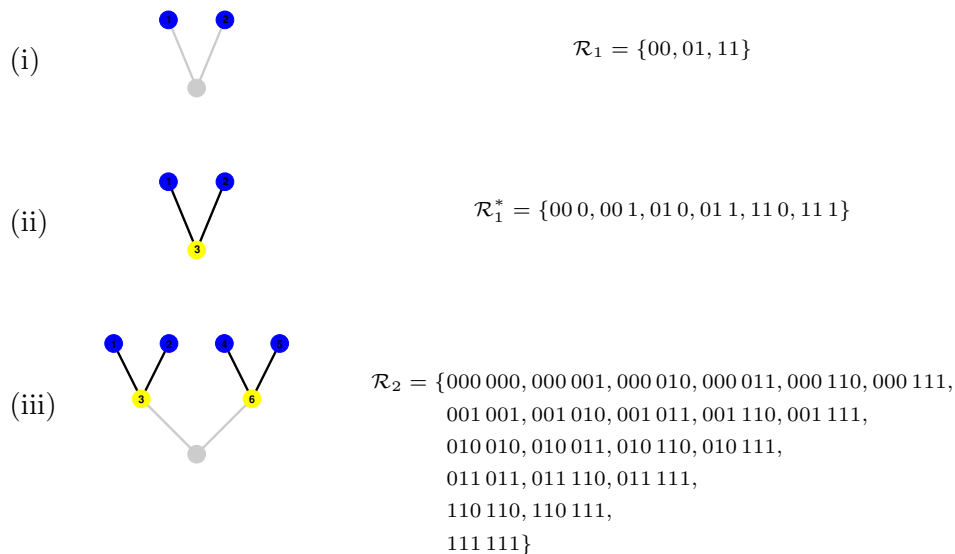(i) $\mathcal{R}_1 = \{00, 01, 11\}$

(ii) $\mathcal{R}_1^* = \{00\,0, 00\,1, 01\,0, 01\,1, 11\,0, 11\,1\}$

(iii) $\mathcal{R}_2 = \{000\,000, 000\,001, 000\,010, 000\,011, 000\,110, 000\,111,$
$001\,001, 001\,010, 001\,011, 001\,110, 001\,111,$
$010\,010, 010\,011, 010\,110, 010\,111,$
$011\,011, 011\,110, 011\,111,$
$110\,110, 110\,111,$
$111\,111\}$

Figure 5: Recursive method of obtaining orbit representatives for SM (e). The orbit representatives listed are for binary vertex-states, i.e. $\mathcal{W} = \{0, 1\}$. States correspond to vertex labels numbered from left to right, so state 01 means vertex 1 has vertex-state 0 and vertex 2 has vertex-state 1.

states chosen from $\mathcal{R}_1^*$ with repetition. Consequently, if the ith geometric factor is $S_{\eta_1} \wr S_{\eta_2}$ then the number of orbits of the corresponding state-space factor is

$$\rho_i = C(\eta_2 + MC(\eta_1 + M - 1, \eta_1) - 1, \eta_2).$$

This procedure can be easily extended to height-regular trees of arbitrary height, although such subgraphs are not common in real-world networks.

C   Symmetries in large real-world networks

Information about fifteen large real-world networks is presented in Table 1. The information for each network includes the number of vertices N, the number of edges $|\mathcal{E}|$, the order (number of permutations) of the symmetry group $|G|$, the number of vertices moved by a permutation in the symmetry group N_{moved}, the number of vertices fixed by all permutations in the symmetry group N_{fixed}, the number of geometric factors m, the number of orbit representatives of the moved vertices ρ_{moved} and the relative significance of the moved vertices $\Delta_{\text{moved}} = N_{\text{moved}} \log_{10}(2) - \log_{10}(\rho_{\text{moved}})$. In all cases the number of fixed vertices means that the lumped state-space is beyond practical computation. While the relative significance is high in several cases (and hence the compression ratio is very high), the number of orbit representatives of the vertices moved by the symmetry group are also very large in most cases.

Name	N	$|\mathcal{E}|$	$|G|$	N_{moved}	N_{fixed}	m	ρ_{moved}	Δ_{moved}
C. Elegans Neural [104, 63]	297	2,148	8.709×10^7	14	283	2	55	2.474
C. Elegans metabolic [23, 62]	453	2,025	1.932×10^{10}	62	391	27	1.898×10^{14}	4.386
Airports500 [18, 61]	500	2,980	8.404×10^{35}	142	358	47	1.190×10^{29}	13.671
Facebook Caltech [94, 95]	762	16,651	4	4	758	2	9	0.250
Interacting proteins [16, 17, 67]	4,713	14,846	2.157×10^{430}	857	3,856	189 (1)	$*8.780 \times 10^{122}$	*135.038
Power grid [69]	4,941	6,594	5.185×10^{152}	823	4,118	302	1.104×10^{184}	63.705
Facebook Georgetown [94, 95]	9,388	425,619	5.898×10^5	31	9,357	14	1.063×10^7	2.305
PGP Network [37, 9, 68]	10,680	24,316	$4.496 \times 10^{1,251}$	4,196	6,484	1,386	1.325×10^{819}	444.000
Facebook Oklahoma [94, 95]	17,420	892,524	1.887×10^7	44	17,376	21	1.860×10^{10}	2.976
Facebook North Carolina [94, 95]	18,158	766,796	9.437×10^6	42	18,116	20	6.199×10^9	2.851
AS Internet [64]	28,311	56,688	$1.038 \times 10^{14,570}$	17,065	11,246	2,703	$1.074 \times 10^{1,872}$	3,265.046
Facebook Florida [94, 95]	35,111	1,465,654	5.242×10^5	38	35,073	19	1.162×10^9	2.374
Co-authorships [71, 66]	39,577	175,693	$5.554 \times 10^{5,271}$	15,746	23,831	5,124 (11)	$*2.455 \times 10^{2,816}$	1,923.628
ITDK Internet [65]	192,244	609,066	$1.600 \times 10^{20,948}$	62,364	129,880	1,9184 (4)	$*1.875 \times 10^{11,693}$	*7,080.162
World-Wide Web	325,729	1,117,563	$1.354 \times 10^{246,428}$	235,052	90,677	19,110 (230)	$*2.329 \times 10^{18,312}$	*52,445.335

Table 1: Large network information: number of vertices N, number of edges $|\mathcal{E}|$, symmetry group order $|G|$, number of vertices moved by the symmetry group N_{moved}, number of vertices fixed by the symmetry group N_{fixed}, the number of geometric factors m, the number of orbit representatives of the moved vertices ρ_{moved} and the relative significance of the moved vertices $\Delta_{\mathrm{moved}} = N_{\mathrm{moved}} \log_{10}(2) - \log_{10}(\rho_{\mathrm{moved}})$. Where present, the numbers in brackets in the m column are the number of geometric factors whose number of orbit representatives could not be computed; all other cases were computed in full. The corresponding ρ_{moved} and Δ_{moved} are labelled with an asterisk to indicate that these are lower and upper bounds respectively.

D   Network Statistics

The following statistics are computed over the networks with non-trivial symmetry. The network repository website includes labels indicating the type of network and in Figure 6 we illustrate how the number of networks breaks down according to these types. Each coloured
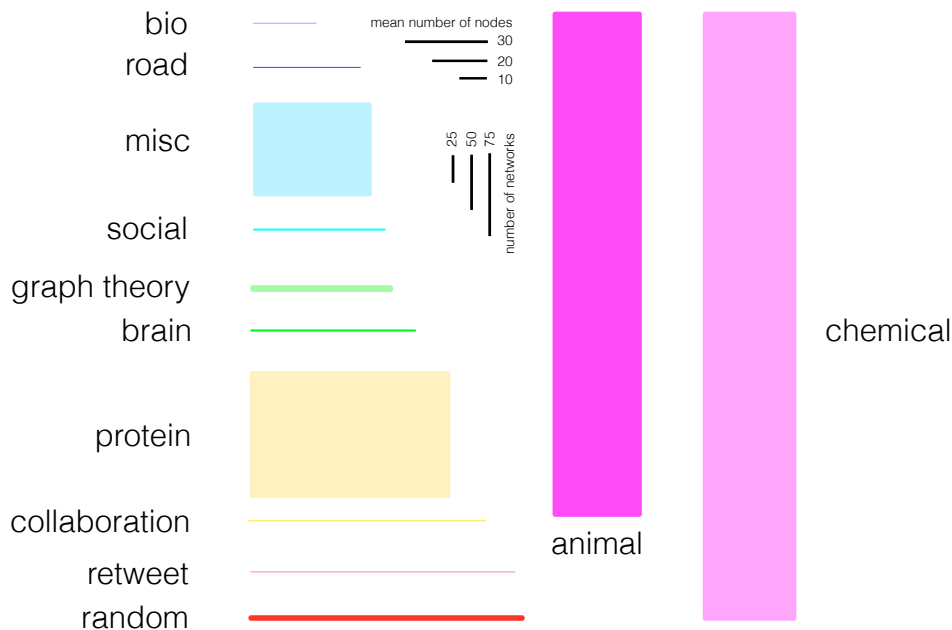

Figure 6: Number of network datasets by type. The vertical extent of each coloured bar indicates the number of networks of the corresponding type and the horizontal extent indicates the mean number of nodes across those networks. Thus the area of each bar corresponds to the total number of nodes of all networks of each type.

rectangle indicates one type of network; the height of the rectangle is scaled to the number of networks of that type and the width is scaled to the mean number of nodes over the networks of that type. The majority of the networks correspond to chemical networks and animal social networks. Whilst most of the networks collected represent real-world networks, our data set does include 'synthetic' networks, which can generally be identified via their type or information provided on the website.

Figure 7(a) is a histogram of the number of nodes in each network with non-trivial symmetry, coloured by type, and shows that the majority of networks have between 10 and 50 nodes. In Figure 7(b) we plot a histogram of mean degree, z, coloured by network type up to $z = 6$ and in Figure 7(c) we plot the same for $z > 6$. The mean degree of the animal social networks is peaked around $z = 2$, and we saw that many of these networks include multiple small components. The mean degree of the chemical networks is peaked at around 4, while the networks from graph theory tend to have relatively large degrees.
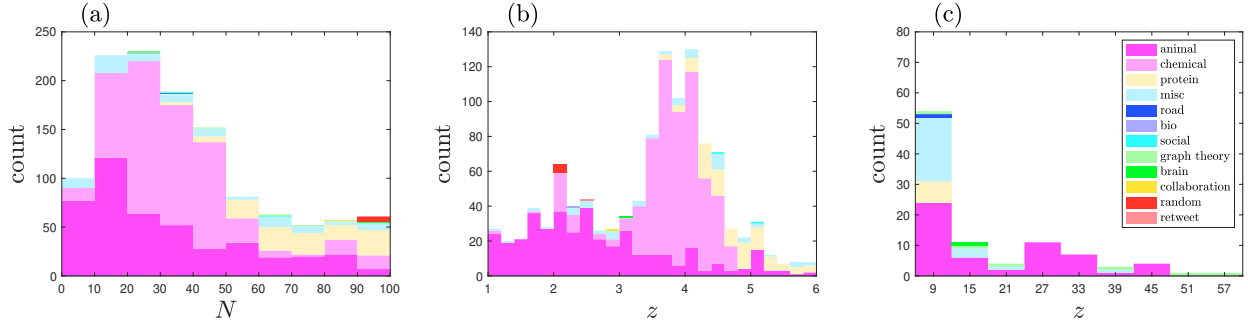
18

Figure 7: Panel (a): histogram of the number of vertices N in each network, coloured by type. Panel (b): histogram of mean degree z for $z < 6$, coloured by type. Panel (c): same as panel (b) but for $z > 6$.

Figure 8 illustrates the fraction of vertices moved by the automorphism group, specifically $\theta_{\text{moved}} = |\text{supp}(G)|/N$, coloured by type. Of the animal social networks, there is a roughly
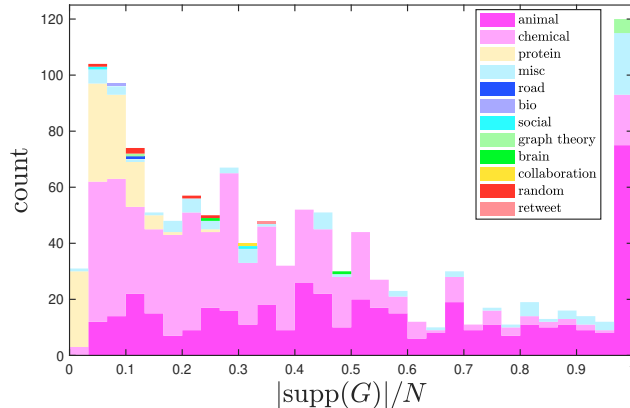

Figure 8: Histogram of the fraction of vertices moved by each graph's automorphism group, $\theta_{\text{moved}} = |\text{supp}(G)|/N$, coloured by type.

constant number of networks for values of θ_{moved} between 0 and 1, but with a larger fraction at $\theta_{\text{moved}} = 1$. In contrast, the number of chemical and protein networks decays with increasing θ_{moved}, more quickly in the latter case. All but one of the graph theoretic networks have $\theta_{\text{moved}} = 1$, i.e. these graphs are vertex-transitive.

In Figure 9 we plot a Venn diagram of the number of networks that have at least one geometric factor corresponding to S_n acting naturally [SMs (a) and (b)], C_2 [SM (d)] or multiple isomorphic components [SM (f)], and no other types of SMs. Figure 9 shows that the majority of symmetry is S_n, i.e. SMs (a) and (b), and there is roughly an equal number of networks that have geometric factors corresponding to C_2 or repeated isomorphic components.
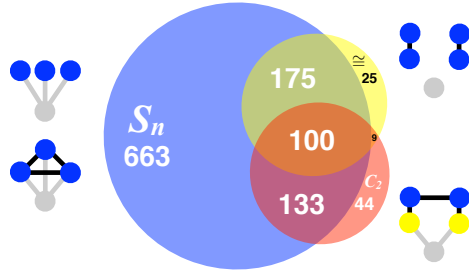
Figure 9: Venn diagram illustrating the main types of symmetry in networks whose symmetry is entirely due to SMs (a), (b), (d) and (f). The blue circle represents symmetric group S_n where symmetry is due to leaves and cliques, the red circle represents the cyclic group of order two, C_2, and the yellow circle represents symmetry due to multiple isomorphic components, indicated by the symbol $\cong$. Example SMs are illustrated next to the corresponding circles.

References

[1] K. Amunts, C. Lepage, L. Borgeat, H. Mohlberg, T. Dickscheid, M.-É. Rousseau, S. Bludau, P.-L. Bazin, L. B. Lewis, A.-M. Oros-Peusquens, N. J. Shah, T. Lippert, K. Zilles, and A. C. Evans. Bigbrain: An ultrahigh-resolution 3d human brain model. *Science*, 340(6139):1472–1475, 2013.

[2] I. Anderson. *A first course in combinatorial mathematics*. Clarendon Press, second edition, 1989.

[3] L. M. Aplin, D. R. Farine, J. Morand-Ferron, and B. C. Sheldon. Social networks predict patch discovery in a wild population of songbirds. *Proceedings of the Royal Society B: Biological Sciences*, 279(1745):4199–4205, 2012.

[4] R. Axelrod. The dissemination of culture: A model with local convergence and global polarization. *Journal of Conflict Resolution*, 41(2):203–226, 1997.

[5] F. Ball and A. Geyer-Schulz. How symmetric are real-world graphs? a large-scale study. *Symmetry*, 10(1):29, 2018.

[6] A. Baronchelli, M. Felici, V. Loreto, E. Caglioti, and L. Steels. Sharp transition towards shared vocabularies in multi-agent systems. *Journal of Statistical Mechanics*, 2006(06):P06014, 2006.

[7] A. Barrat, M. Barthelemy, and A. Vespignani. *Dynamical processes on complex networks*. Cambridge University Press, Cambridge, 2008.

[8] F. M. Bass. A new product growth for model consumer durables. *Management Science*, 15(5):215–227, 1969.

[9] M. Boguná, R. Pastor-Satorras, A. Díaz-Guilera, and A. Arenas. Models of social networks based on social distance attachment. *Physical Review E*, 70(5):056122, 2004.

[10] E. Bonabeau, G. Theraulaz, and J.-L. Deneubourg. Phase diagram of a model of self-organizing hierarchies. *Physica A*, 217(3-4):373–392, 1995.

[11] M. Broom, J. Rychtář, and B. T. Stadler. Evolutionary dynamics on graphs-the effect of graph structure and initial placement on mutant spread. *Journal of Statistical Theory and Practice*, 5(3):369–381, 2011.

[12] R. A. Brualdi. *Introductory combinatorics*. Pearson Education, Inc, 1977.

[13] R. Burns, K. Lillaney, D. R. Berger, L. Grosenick, K. Deisseroth, R. C. Reid, W. G. Roncal, P. Manavalan, D. D. Bock, N. Kasthuri, et al. The open connectome project data cluster: scalable analysis and vision for high-throughput neuroscience. In *Proceedings of the 25th International Conference on Scientific and Statistical Database Management*, pages 1–11, 2013.

[14] C. Castellano, M. Marsili, and A. Vespignani. Nonequilibrium phase transition in a model for social influence. *Physical Review Letters*, 85(16):3536, 2000.

[15] X. Castelló, V. M. Eguíluz, and M. San Miguel. Ordering dynamics with two non-excluding options: bilingualism in language competition. *New Journal of Physics*, 8(12):308, 2006.

[16] V. Colizza, A. Flammini, A. Maritan, and A. Vespignani. Characterization and modeling of protein–protein interaction networks. *Physica A: Statistical Mechanics and its Applications*, 352(1):1–27, 2005.

[17] V. Colizza, A. Flammini, M. A. Serrano, and A. Vespignani. Detecting rich-club ordering in complex networks. *Nature Physics*, 2(2):110–115, 2006.

[18] V. Colizza, R. Pastor-Satorras, and A. Vespignani. Reaction–diffusion processes and metapopulation models in heterogeneous networks. *Nature Physics*, 3(4):276–282, 2007.

[19] A. Darbon, D. Colombi, E. Valdano, L. Savini, A. Giovannini, and V. Colizza. Disease persistence on temporal contact networks accounting for heterogeneous infectious periods. *Royal Society open science*, 6(1):181404, 2019.

[20] S. C. de Lange, M. P. van den Heuvel, and M. A. de Reus. The role of symmetry in neural networks and their laplacian spectra. *NeuroImage*, 141:357–365, 2016.

[21] C. P. Dettmann and G. Knight. Symmetric motifs in random geometric graphs. *Journal of Complex Networks*, 6(1):95–105, 2018.

[22] J. D. Dixon and B. Mortimer. *Permutation groups*, volume 163. Springer Science & Business Media, 1996.

[23] J. Duch and A. Arenas. Scaling of fluctuations in traffic on complex networks. *Physical Review Letters*, 96(21):218702, 2006.

[24] A. Economou, A. Gómez-Corral, and M. López-García. A stochastic sis epidemic model with heterogeneous contacts. *Physica A: Statistical Mechanics and its Applications*, 421:78–97, 2015.

[25] P. Erdős and A. Rényi. Asymmetric graphs. *Acta Mathematica Hungarica*, 14(3-4):295–315, 1963.

[26] M. Feng, S.-M. Cai, M. Tang, and Y.-C. Lai. Equivalence and its invalidation between non-markovian and markovian spreading dynamics on complex networks. *Nature communications*, 10(1):1–10, 2019.

[27] P. G. Fennell and J. P. Gleeson. Multistate dynamical processes on networks: analysis through degree-based approximation frameworks. *SIAM Review*, 61(1):92–118, 2019.

[28] J. B. Fraleigh. *A first course in abstract algebra*. Pearson Education India, 2003.

[29] S. Galam. Minority opinion spreading in random geometry. *The European Physical Journal B*, 25(4):403–406, 2002.

[30] The GAP Group. *GAP – Groups, Algorithms, and Programming, Version 4.8.8*, 2017.

[31] D. Garlaschelli, F. Ruzzenenti, and R. Basosi. Complex networks and symmetry i: A review. *Symmetry*, 2(3):1683–1709, 2010.

[32] J. P. Gleeson. High-accuracy approximation of binary-state dynamics on networks. *Physical Review Letters*, 107(6):68701, 2011.

[33] J. P. Gleeson, S. Melnik, J. A. Ward, M. A. Porter, and P. J. Mucha. Accuracy of mean-field theory for dynamics on real-world networks. *Physical Review E*, 85(2):026106, 2012.

[34] C. Godsil and G. F. Royle. *Algebraic graph theory*, volume 207. Springer Science & Business Media, 2013.

[35] A. Gómez-Corral and M. López-García. Perturbation analysis in finite ld-qbd processes and applications to epidemic models. *Numerical Linear Algebra with Applications*, 25(5):e2160, 2018.

[36] G. Großmann and L. Bortolussi. Reducing spreading processes on networks to markov population models. In *International Conference on Quantitative Evaluation of Systems*, pages 292–309. Springer, 2019.

[37] X. Guardiola, R. Guimera, A. Arenas, A. Diaz-Guilera, D. Streib, and L. Amaral. Macro-and micro-structure of trust networks. *arXiv preprint cond-mat/0206240*, 2002.

[38] R. Hammack, W. Imrich, and S. Klavžar. *Handbook of product graphs.* CRC press, 2011.

[39] P. Holme. Shadows of the susceptible-infectious-susceptible immortality transition in small networks. *Physical Review E*, 92(1):012804, 2015.

[40] P. Holme. Three faces of node importance in network epidemiology: Exact results for small graphs. *Physical Review E*, 96(6):062305, 2017.

[41] P. Holme and L. Tupikina. Epidemic extinction in networks: insights from the 12,110 smallest graphs. *New Journal of Physics*, 20(11):113042, 2018.

[42] H. Katebi, K. Sakallah, and I. Markov. Symmetry and satisfiability: An update. *Theory and Applications of Satisfiability Testing–SAT 2010*, pages 113–127, 2010.

[43] T. M. Keller. Orbits in finite group actions. *Groups St. Andrews 2001 in Oxford*, 2:306–331, 2003.

[44] J. G. Kemeny, J. L. Snell, et al. *Finite markov chains*, volume 356. van Nostrand Princeton, NJ, 1960.

[45] W. R. KhudaBukhsh, A. Auddy, Y. Disser, and H. Koeppl. Approximate lumpability for markovian agent-based models using local symmetries. *Journal of Applied Probability*, 56(3):647–671, 2019.

[46] M. Kijima. *Markov processes for stochastic modeling*, volume 6. CRC Press, 1997.

[47] I. Z. Kiss, J. C. Miller, and P. L. Simon. *Mathematics of Epidemics on Networks.* Springer, 2017.

[48] I. Z. Kiss, G. Röst, and Z. Vizi. Generalization of pairwise models to non-Markovian epidemics on networks. *Physical Review Letters*, 115(7):078701, 2015.

[49] I. Klickstein, L. Pecora, and F. Sorrentino. Symmetry induced group consensus. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 29(7):073101, 2019.

[50] R. H. Kurvers, J. Krause, D. P. Croft, A. D. Wilson, and M. Wolf. The evolutionary and ecological consequences of animal social networks: emerging issues. *Trends in Ecology & Evolution*, 29(6):326–335, 2014.

[51] E. Lieberman, C. Hauert, and M. A. Nowak. Evolutionary dynamics on graphs. *Nature*, 433(7023):312–316, 2005.

[52] M. López-García. Stochastic descriptors in an sir epidemic model for heterogeneous individuals in small networks. *Mathematical Biosciences*, 271:42 – 61, 2016.

[53] M. López-García, M. Aruru, and S. Pyne. Health analytics and disease modeling for better understanding of healthcare-associated infections. *BLDE University Journal of Health Sciences*, 3(2):69–74, 2018.

[54] M. López-García, M.-F. King, and C. J. Noakes. A multicompartment SIS stochastic model with zonal ventilation for the spread of nosocomial infections: Detection, outbreak management, and infection control. *Risk Analysis*, 2019.

[55] M. López-García and T. Kypraios. A unified stochastic modelling framework for the spread of nosocomial infections. *Journal of the Royal Society Interface*, 15(143):20180060, 2018.

[56] B. D. MacArthur and R. J. Sánchez-García. Spectral characteristics of network redundancy. *Physical Review E*, 80(2):026117, 2009.

[57] B. D. MacArthur, R. J. Sánchez-García, and J. W. Anderson. Symmetry in complex networks. *Discrete Applied Mathematics*, 156(18):3525–3531, 2008.

[58] A. Mellor, M. Mobilia, S. Redner, A. M. Rucklidge, and J. A. Ward. Influence of Luddism on innovation diffusion. *Physical Review E*, 92(1):012806, 2015.

[59] S. Melnik, J. A. Ward, J. P. Gleeson, and M. A. Porter. Multi-stage complex contagions. *Chaos*, 23(1):013124, 2013.

[60] A. Moinet, M. Starnini, and R. Pastor-Satorras. Burstiness and aging in social temporal networks. *Physical review letters*, 114(10):108701, 2015.

[61] [Network Data] Network of the 500 US airports with the largest amount of traffic from publicly available data. `https://sites.google.com/site/cxnets/usairtransportationnetwork` (accessed Nov 2020).

[62] [Network Data] Metabolic network of C. Elegans. `https://deim.urv.cat/~alexandre.arenas/data/xarxes/celegans_metabolic.zip` (accessed Nov 2020).

[63] [Network Data] Neural network of C. Elegans. `http://www-personal.umich.edu/~mejn/netdata/celegansneural.zip` (accessed Nov 2020).

[64] [Network Data] The CAIDA Autonomous System Relationships Dataset. `http://www.caida.org/data/active/as-relationships`; `https://www.caida.org/data/request_user_info_forms/as_relationships.xml` (accessed Jun 2008).

[65] [Network Data] The CAIDA Macroscopic Internet Topology Data Kit. `https://www.caida.org/data/internet-topology-data-kit/` (accessed Jun 2008).

[66] [Network Data] Network of coauthorships between scientists posting preprints on the Condensed Matter E-Print Archive, includes all preprints posted between 1-Jan-1995 and 31-Mar-2005. `http://www-personal.umich.edu/~mejn/netdata/cond-mat-2005.zip` (accessed Nov 2020).

[67] [Network Data] Protein interaction network of the yeast Saccharomyces cerevisae extracted with different experimental techniques and collected at the Database of Interacting Proteins (http://dip.doe-mbi.ucla.edu/). `https://sites.google.com/site/cxnets/research22` (accessed Nov 2020).

[68] [Network Data] Largest connected component of the network of users of the Pretty-Good-Privacy algorithm for secure information interchange. `https://deim.urv.cat/~alexandre.arenas/data/xarxes/PGP.zip` (accessed Nov 2020).

[69] [Network Data] An undirected, unweighted network representing the topology of the Western States Power Grid of the United States. `http://www-personal.umich.edu/~mejn/netdata/power.zip` (accessed Nov 2020).

[70] M. Newman. *Networks*. Oxford University Press, Oxford, 2010.

[71] M. E. Newman. The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*, 98(2):404–409, 2001.

[72] M. E. Newman. The structure and function of complex networks. *SIAM Review*, 45(2):167–256, 2003.

[73] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani. Epidemic processes in complex networks. *Reviews of Modern Physics*, 87(3):925, 2015.

[74] R. Pastor-Satorras and A. Vespignani. Epidemic spreading in scale-free networks. *Physical Review Letters*, 86(14):3200, 2001.

[75] L. M. Pecora, F. Sorrentino, A. M. Hagerstrom, T. E. Murphy, and R. Roy. Cluster synchronization and isolated desynchronization in complex networks with symmetries. *Nature Communications*, 5(1):1–8, 2014.

[76] M. A. Porter and J. P. Gleeson. Dynamical systems on networks: a tutorial. *Frontiers in Applied Dynamical Systems: Reviews and Tutorials*, 4, 2016.

[77] R. A. Rossi and N. K. Ahmed. The network data repository with interactive graph analytics and visualization. In *AAAI*, 2015.

[78] R. A. Rossi, D. F. Gleich, A. H. Gebremedhin, and M. A. Patwary. What if clique were fast? maximum cliques in information networks and strong components in temporal networks. *arXiv preprint arXiv:1210.5802*, pages 1–11, 2012.

[79] R. A. Rossi, D. F. Gleich, A. H. Gebremedhin, and M. A. Patwary. Fast maximum clique algorithms for large graphs. In *Proceedings of the 23rd International Conference on World Wide Web (WWW)*, 2014.

[80] G. Röst, Z. Vizi, and I. Z. Kiss. Pairwise approximation for sir-type network epidemics with non-markovian recovery. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 474(2210):20170695, 2018.

[81] J. J. Rotman. *An introduction to the theory of groups*, volume 148. Springer Science & Business Media, 2012.

[82] P. Sah, J. Mann, and S. Bansal. Disease implications of animal social network structure: a synthesis across social systems. *Journal of Animal Ecology*, 87(3):546–558, 2018.

[83] P. Sah, J. D. Méndez, and S. Bansal. A multi-species repository of social networks. *Scientific data*, 6(1):1–6, 2019.

[84] R. J. Sánchez-García. Exploiting symmetry in complex network analysis. *arXiv preprint arXiv:1803.06915*, 2018.

[85] N. Sherborne, J. C. Miller, K. B. Blyuss, and I. Z. Kiss. Mean-field models for non-markovian epidemics on networks. *Journal of mathematical biology*, 76(3):755–778, 2018.

[86] P. L. Simon, M. Taylor, and I. Z. Kiss. Exact epidemic models on graphs using graph-automorphism driven lumping. *Journal of mathematical biology*, 62(4):479–508, 2011.

[87] D. Smith and B. Webb. Hidden symmetries in real and theoretical networks. *Physica A: Statistical Mechanics and its Applications*, 514:855–867, 2019.

[88] V. Sood and S. Redner. Voter model on heterogeneous graphs. *Physical Review Letters*, 94(17):178701, 2005.

[89] F. Sorrentino, L. M. Pecora, A. M. Hagerstrom, T. E. Murphy, and R. Roy. Complete characterization of the stability of cluster synchronization in complex dynamical networks. *Science Advances*, 2(4):e1501737, 2016.

[90] F. Sorrentino, A. B. Siddique, and L. M. Pecora. Symmetries in the time-averaged dynamics of networks: Reducing unnecessary complexity through minimal network models. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 29(1):011101, 2019.

[91] M. Starnini, J. P. Gleeson, and M. Boguñá. Equivalence between non-markovian and markovian dynamics in epidemic spreading processes. *Physical review letters*, 118(12):128301, 2017.

[92] K. Sznajd-Weron and J. Sznajd. Opinion evolution in closed community. *International Journal of Modern Physics C*, 11(06):1157–1165, 2000.

[93] L. Temime, L. Opatowski, Y. Pannet, C. Brun-Buisson, P. Y. Boëlle, and D. Guillemot. Peripatetic health-care workers as potential superspreaders. *Proceedings of the National Academy of Sciences*, 106(43):18420–18425, 2009.

[94] A. L. Traud, E. D. Kelsic, P. J. Mucha, and M. A. Porter. Comparing community structure to characteristics in online collegiate social networks. *SIAM Review*, 53(3):526–543, 2011.

[95] A. L. Traud, P. J. Mucha, and M. A. Porter. Social structure of facebook networks. *Physica A: Statistical Mechanics and its Applications*, 391(16):4165–4180, 2012.

[96] M. Van Baalen. Contact networks and the evolution of virulence. *Adaptive dynamics of infectious diseases: in pursuit of virulence management. Cambridge Univ. Press, Cambridge, UK*, pages 85–103, 2002.

[97] A. van Heukelum, G. Barkema, and R. H. Bisseling. DNA electrophoresis studied with the cage model. *Journal of Computational Physics*, 180(1):313–326, 2002.

[98] F. Vazquez and V. M. Eguíluz. Analytical solution of the voter model on uncorrelated networks. *New Journal of Physics*, 10(6):063011, 2008.

[99] J. T. Vogelstein, E. Perlman, B. Falk, A. Baden, W. G. Roncal, V. Chandrashekhar, F. Collman, S. Seshamani, J. L. Patsolic, K. Lillaney, et al. A community-developed open-source computational ecosystem for big neuro data. *Nature Methods*, 15(11):846–847, 2018.

[100] J. A. Ward. Dimension-reduction of dynamics on real-world networks with symmetry - network data. Research Data Leeds Repository, DOI: 10.5518/911, 2020.

[101] J. A. Ward and J. Evans. A general model of dynamics on networks with graph automorphism lumping. In *International Conference on Complex Networks and their Applications*, pages 445–456. Springer, 2018.

[102] J. A. Ward and M. López-García. Exact analysis of summary statistics for continuous-time discrete-state markov processes on networks using graph-automorphism lumping. *Applied Network Science*, 4(1):108, 2019.

[103] D. J. Watts. A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences*, 99(9):5766–5771, 2002.

[104] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world'networks. *Nature*, 393(6684):440, 1998.

[105] P. A. Whigham, G. C. Dick, and H. G. Spencer. Genetic drift on networks: Ploidy and the time to fixation. *Theoretical Population Biology*, 74(4):283–290, 2008.

[106] J. K. Wilson-Aggarwal, L. Ozella, M. Tizzoni, C. Cattuto, G. J. Swan, T. Moundai, M. J. Silk, J. A. Zingeser, and R. A. McDonald. High-resolution contact networks of free-ranging domestic dogs canis familiaris and implications for transmission of infection. *PLoS Neglected Tropical Diseases*, 13(7):e0007565, 2019.

[107] Y. Xiao, B. D. MacArthur, H. Wang, M. Xiong, and W. Wang. Network quotients: Structural skeletons of complex systems. *Physical Review E*, 78(4):046102, 2008.

[108] Y. Xiao, M. Xiong, W. Wang, and H. Wang. Emergence of symmetry in complex networks. *Physical Review E*, 77(6):066108, 2008.