

FASTEST MIXING MARKOV CHAIN ON GRAPHS WITH SYMMETRIES*

STEPHEN BOYD[†], PERSI DIACONIS[‡], PABLO PARRILO[§], AND LIN XIAO[¶]

Abstract. We show how to exploit symmetries of a graph to efficiently compute the fastest mixing Markov chain on the graph (i.e., find the transition probabilities on the edges to minimize the second-largest eigenvalue modulus of the transition probability matrix). Exploiting symmetry can lead to significant reduction in both the number of variables and the size of matrices in the corresponding semidefinite program, and thus enable numerical solution of large-scale instances that are otherwise computationally infeasible. We obtain analytic or semianalytic results for particular classes of graphs, such as edge-transitive and distance-transitive graphs. We describe two general approaches for symmetry exploitation, based on orbit theory and block-diagonalization, respectively, and establish a formal connection between them.

Key words. Markov chains, fast mixing, eigenvalue optimization, semidefinite programming, graph automorphism, group representation

AMS subject classifications. 05C25, 20C30, 60J22, 65F15, 90C22, 90C51

DOI. 10.1137/070689413

1. Introduction. In the fastest mixing Markov chain problem [3], we choose the transition probabilities on the edges of a graph to minimize the second-largest eigenvalue modulus of the transition probability matrix. As shown in [3], this can be formulated as a convex optimization problem, in particular a semidefinite program. Thus it can be solved, up to any given precision, in polynomial time by interior-point methods. In this paper, we show how to exploit symmetries of a graph to make the computation much more efficient.

1.1. The fastest mixing Markov chain problem. We consider an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with vertex set $\mathcal{V} = \{1, \dots, n\}$ and edge set \mathcal{E} , and assume that \mathcal{G} is connected. We define a discrete-time Markov chain by associating with each edge $\{i, j\} \in \mathcal{E}$ a transition probability P_{ij} (P_{ii} denotes the holding probability at vertex i). We assume the transition between two vertices connected by an edge is symmetric, i.e., $P_{ij} = P_{ji}$. Thus the transition probability matrix, $P \in \mathbf{R}^{n \times n}$, satisfies

$$P = P^T, \quad P \geq 0, \quad P\mathbf{1} = \mathbf{1},$$

where the superscript T denotes the transpose of a matrix, the inequality $P \geq 0$ means elementwise, and $\mathbf{1}$ denotes the vector of all ones.

Since P is symmetric and stochastic, the uniform distribution $(1/n)\mathbf{1}^T$ is stationary. In addition, the eigenvalues of P are real, and no more than one in modulus. We

*Received by the editors April 24, 2007; accepted for publication (in revised form) March 31, 2009; published electronically June 17, 2009.

<http://www.siam.org/journals/siopt/20-2/68941.html>

[†]Information Systems Laboratory, Department of Electrical Engineering, Stanford University, Stanford, CA 94305 (boyd@stanford.edu).

[‡]Department of Statistics and Department of Mathematics, Stanford University, Stanford, CA 94305.

[§]Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139 (parrilo@mit.edu).

[¶]Microsoft Research, 1 Microsoft Way, Redmond, WA 98052 (lin.xiao@microsoft.com).

list them in nonincreasing order as

$$1 = \lambda_1(P) \geq \lambda_2(P) \geq \dots \geq \lambda_n(P) \geq -1.$$

We denote by $\mu(P)$ the second-largest eigenvalue modulus (SLEM) of P , i.e.,

$$\mu(P) = \max_{i=2,\dots,n} |\lambda_i(P)| = \max \{ \lambda_2(P), -\lambda_n(P) \}.$$

This quantity is widely used to bound the asymptotic convergence rate of the Markov chain to its stationary distribution, in the total variation distance or chi-squared distance (see, e.g., [16, 19]). In general the smaller $\mu(P)$ is, the faster the Markov chain converges. For more background on Markov chains, eigenvalues, and rapid mixing, see, e.g., the text [8].

The *fastest mixing Markov chain* (FMMC) problem [3] is to find the optimal P that minimizes $\mu(P)$. This can be posed as the following optimization problem:

$$(1.1) \quad \begin{aligned} & \text{minimize} && \mu(P) \\ & \text{subject to} && P = P^T, \quad P \geq 0, \quad P\mathbf{1} = \mathbf{1}, \\ & && P_{ij} = 0, \quad i \neq j \text{ and } \{i, j\} \notin \mathcal{E}. \end{aligned}$$

It turns out that this is a convex optimization problem [3]. This can be seen, e.g., by expressing the objective function as $\mu(P) = \|P - (1/n)\mathbf{1}\mathbf{1}^T\|_2$, where $\|\cdot\|_2$ denotes the spectral norm of a matrix. Moreover, it can be transformed into a semidefinite program (SDP):

$$(1.2) \quad \begin{aligned} & \text{minimize} && s \\ & \text{subject to} && -sI \preceq P - (1/n)\mathbf{1}\mathbf{1}^T \preceq sI, \\ & && P = P^T, \quad P \geq 0, \quad P\mathbf{1} = \mathbf{1}, \\ & && P_{ij} = 0, \quad i \neq j \text{ and } \{i, j\} \notin \mathcal{E}. \end{aligned}$$

Here I denotes the identity matrix, and the variables are the matrix P and the scalar s . The symbol \preceq denotes matrix inequality, i.e., $X \preceq Y$ means $Y - X$ is positive semidefinite.

We should note that there are other important measures of rapid mixing, e.g., the log-Sobolev constant [33, 55, 17], and other methods to speed up mixing, e.g., lifting [11]. We focus on the approach of minimizing the SLEM on a fixed graph topology. In addition to its direct connection to Markov chain Monte Carlo simulation, the FMMC problem has found many practical applications in fast load balancing of parallel computing systems (often with highly symmetric configurations, as discussed in this paper) [56, 59], and in average consensus and gossip algorithms in sensor networks [58, 7].

There has been some follow-up work on the FMMC problem. Boyd et al. [6] proved analytically that on an n -path the fastest mixing chain can be obtained by assigning the same transition probability half at the $n - 1$ edges and two loops at the two ends. Roch [51] used standard mixing-time analysis techniques (variational characterizations, conductance, canonical paths) to bound the fastest mixing time. Gade and Overton [23] have considered the fastest mixing problem for a nonreversible Markov chain. Here, the problem is nonconvex and much remains to be done. Finally, closed-form solutions of fastest mixing problems have recently been applied in statistics to give a generalization of the usual spectral analysis of time series for more general discrete data; see [53].

1.2. Exploiting problem structure. When solving the SDP (1.2) by interior-point methods, in each iteration, we need to compute the first and second derivatives of the logarithmic barrier functions (or potential functions) for the matrix inequalities, and assemble and solve a linear system of equations (the Newton system). Let n be the number of vertices and let m be the number of edges in the graph (m is the number of variables in the optimization problem). The Newton system is a set of m linear equations with m unknowns. Without exploiting any structure, the number of flops per iteration in a typical barrier method is on the order $\max\{mn^3, m^2n^2, m^3\}$, where the first two terms come from computing and assembling the Newton system, and the third term amounts to solving it (see, e.g., [2, section 11.8.3]). (Other variants of interior-point methods have similar orders of flop count.)

Exploiting problem structure can significantly improve solution efficiency. As for many other problems defined on a graph, sparsity is the most obvious structure to consider here. In fact, many current SDP solvers already exploit sparsity. However, it is a well-known fact that exploiting sparsity alone in interior-point methods for SDP has limited effectiveness. The sparsity of P , and the sparsity plus rank-one structure of $P - (1/n)\mathbf{1}\mathbf{1}^T$, can be exploited to greatly reduce the complexity of assembling the Newton system, but typically the Newton system itself is dense. The computational cost per iteration is still at the order $O(m^3)$, dominated by solving the dense linear system (see analysis in [58]).

We can also solve the FMMC problem in the form (1.1) by subgradient-type (first-order) methods. The subgradients of $\mu(P)$ can be obtained by computing the SLEM of the matrix P and the associated eigenvectors. This can be done very efficiently by iterative methods, specifically the Lanczos method, for large sparse symmetric matrices (see, e.g., [30, 52]). Compared with interior-point methods, subgradient-type methods can solve much larger problems but only to a moderate accuracy; they also don't have polynomial-time worst-case complexity. In [3], we used a simple subgradient method to solve the FMMC problem on graphs with up to a few hundred thousand edges. More sophisticated first-order methods, for solving large-scale eigenvalue optimization and SDPs, have been developed in, e.g., [10, 35, 38, 45, 46]. A successive partial linear programming method was developed in [47].

In this paper, we focus on the FMMC problem on graphs with large symmetry groups, and we show how to exploit symmetries of the graph to make the computation more efficient. A result by Erdős and Rényi [21] states that with high probability (asymptotically with probability one), the symmetry group of a suitably defined random graph is trivial; i.e., it contains only the identity element. Nevertheless, many of the graphs of theoretical and practical interest, particularly in engineering applications, have very interesting, and sometimes very large, symmetry groups. Symmetry reduction techniques have been explored in several different contexts, e.g., dynamical systems and bifurcation theory [31], polynomial system solving [25, 57], numerical solution of partial differential equations [22], and Lie symmetry analysis in geometric mechanics [40]. In the context of optimization, a class of SDPs with symmetry has been defined in [36], where the authors study the invariance properties of the search directions of primal-dual interior-point methods. In addition, symmetry has been exploited to prune the enumeration tree in branch-and-cut algorithms for integer programming [39] and to reduce matrix size in a spectral radius optimization problem [34].

Closely related to our approach in this paper, the recent work [14] considered general SDPs that are invariant under the action of a permutation group and developed a technique based on matrix $*$ -representation to reduce problem size. This

technique has been applied to simplify computations in SDP relaxations for graph coloring and maximal clique problems [20] and to strengthen SDP bounds for some coding problems [37].

1.3. Contents. The paper is organized as follows. In section 2, we first explain some basic background on graph automorphisms and symmetry groups. We show that the FMMC problem always attains its optimum in the fixed-point subset of the feasible set under the automorphism group. This allows us to consider only a number of distinct transition probabilities that equals the number of orbits of the edges.

In section 3, we give closed-form solutions for the FMMC problem on some special classes of graphs, namely, edge-transitive graphs and distance-transitive graphs. Along the way we also discuss FMMC on graphs formed by taking Cartesian products of simple graphs.

In section 4, we first review the orbit theory for reversible Markov chains developed in [4], which gives sufficient conditions on constructing an orbit chain that contains all distinct eigenvalues of the original chain. This orbit chain is usually no longer symmetric but always reversible. We then solve the fastest reversible Markov chain problem on the orbit graph, from which we immediately obtain an optimal solution to the original FMMC problem.

In section 5, we focus on the approach developed in [26], which block-diagonalizes the linear matrix inequalities in the FMMC problem by constructing a symmetry-adapted basis. The resulting blocks usually have much smaller sizes, and repeated blocks can be discarded in computation. We establish a formal connection between this approach and the orbit theory, and demonstrate their connection on several examples. More examples can be found in [5].

In section 6, we conclude the paper by pointing out some possible future work.

2. Symmetry analysis. In this section we explain the basic concepts that are essential in exploiting graph symmetry, and we derive our result on reducing the number of optimization variables.

2.1. Graph automorphisms and classification. The study of graphs that possess particular kinds of symmetry properties has a long history (see, e.g., [1, 9]). The basic object of study is the *automorphism group* of a graph, and different classes can be defined depending on the specific form in which the group acts on the vertices and edges.

An *automorphism* of a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is a permutation σ of \mathcal{V} such that $\{i, j\} \in \mathcal{E}$ if and only if $\{\sigma(i), \sigma(j)\} \in \mathcal{E}$. The (full) *automorphism group* of the graph, denoted by $\text{Aut}(\mathcal{G})$, is the set of all such permutations, with the group operation being composition. For a vertex $i \in \mathcal{V}$, the set of all images $\sigma(i)$, as σ varies through a subgroup $G \subseteq \text{Aut}(\mathcal{G})$, is called the *orbit* of i under the action of G . Distinct orbits form equivalent classes and they partition the set \mathcal{V} . The action is *transitive* if there is only one single orbit in \mathcal{V} .

A graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is said to be *vertex-transitive* if $\text{Aut}(\mathcal{G})$ acts transitively on \mathcal{V} . The action of a permutation σ on \mathcal{V} induces an action on \mathcal{E} with the rule $\sigma(\{i, j\}) = \{\sigma(i), \sigma(j)\}$. A graph \mathcal{G} is *edge-transitive* if $\text{Aut}(\mathcal{G})$ acts transitively on \mathcal{E} . Graphs can be edge-transitive without being vertex-transitive, and vice versa; simple examples are shown in Figure 2.1.

A graph is called *1-arc-transitive* if, given any four vertices u, v, x, y such that $\{u, v\}, \{x, y\} \in \mathcal{E}$, there exists an automorphism $\sigma \in \text{Aut}(\mathcal{G})$ such that $\sigma(u) = x$ and $\sigma(v) = y$. Notice that, as opposed to edge-transitivity, here the ordering of the

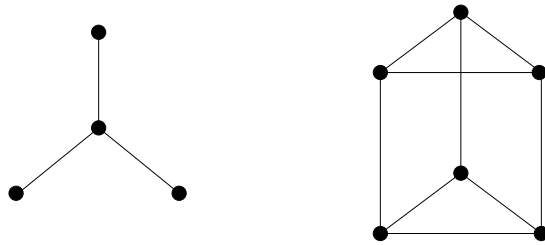


FIG. 2.1. The graph on the left side is edge-transitive but not vertex-transitive. The one on the right side is vertex-transitive but not edge-transitive.

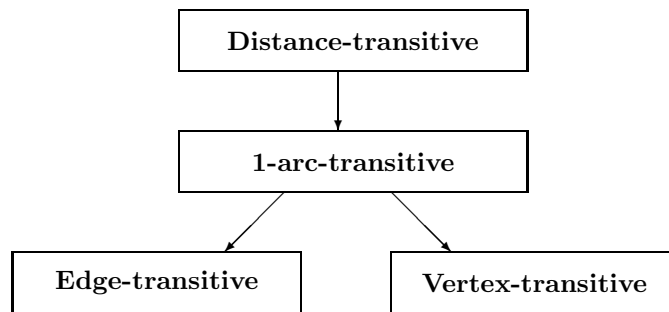


FIG. 2.2. Classes of symmetric graphs and their inclusion relationship.

vertices is important, even for undirected graphs. In fact, a 1-arc-transitive graph must be both vertex-transitive and edge-transitive, and the reverse may not be true. The 1-arc-transitive graphs are called *symmetric graphs* in [1], but the modern use extends this term to all graphs that are simultaneously edge- and vertex-transitive. Finally, let $\delta(u, v)$ denote the distance between two vertices $u, v \in \mathcal{V}$. A graph is called *distance-transitive* if, for any four vertices u, v, x, y with $\delta(u, v) = \delta(x, y)$, there is an automorphism $\sigma \in \text{Aut}(\mathcal{G})$ such that $\sigma(u) = x$ and $\sigma(v) = y$.

The containment relationship among the four classes of graphs described above is illustrated in Figure 2.2. Explicit counterexamples are known for each of the noninclusions. It is generally believed that distance-transitive graphs have been completely classified. This work has been done by classifying the distance-regular graphs. It would take us too far afield to give a complete discussion. See the survey in [18, section 7].

The concept of graph automorphism can be naturally extended to weighted graphs by requiring that the permutation also preserve the weights on edges (see, e.g., [4]). This extension allows us to exploit symmetry in more general reversible Markov chains, where the transition probability matrix is not necessarily symmetric.

2.2. FMMC with symmetry constraints. A permutation $\sigma \in \text{Aut}(\mathcal{G})$ can be represented by a permutation matrix Q , where $Q_{ij} = 1$ if $i = \sigma(j)$ and $Q_{ij} = 0$ otherwise. The permutation σ induces an action on the transition probability matrix by $\sigma(P) = QPQ^T$.

We denote the feasible set of the FMMC problem (1.1) by \mathcal{C} , i.e.,

$$\mathcal{C} = \{P \in \mathbf{R}^{n \times n} \mid P \geq 0, P\mathbf{1} = \mathbf{1}, P = P^T, P_{ij} = 0 \text{ for } \{i, j\} \notin \mathcal{E}\}.$$

This set is invariant under the action of graph automorphism. To see this, let $h = \sigma(i)$

and $k = \sigma(j)$. Then we have

$$(\sigma(P))_{hk} = (QPQ^T)_{hk} = \sum_l (QP)_{hl}Q_{kl} = (QP)_{hj} = \sum_l Q_{hl}P_{lj} = P_{ij}.$$

Since σ is a graph automorphism, we have $\{h, k\} \in \mathcal{E}$ if and only if $\{i, j\} \in \mathcal{E}$, so the sparsity pattern of the probability transition matrix is preserved. It is straightforward to verify that the conditions $P \geq 0$, $P\mathbf{1} = \mathbf{1}$, and $P = P^T$ are also preserved under this action.

Let \mathcal{F} denote the fixed-point subset of \mathcal{C} under the action of $\text{Aut}(\mathcal{G})$; i.e.,

$$(2.1) \quad \mathcal{F} = \{P \in \mathcal{C} \mid \sigma(P) = P, \sigma \in \text{Aut}(\mathcal{G})\}.$$

We have the following theorem

THEOREM 2.1. *The FMMC problem always has an optimal solution in \mathcal{F} .*

Similar results have appeared in, e.g., [14, 26]. Here we include the proof for completeness.

Proof. Let μ^* denote the optimal value of the FMMC problem (1.1), i.e., $\mu^* = \inf\{\mu(P) \mid P \in \mathcal{C}\}$. Since the objective function μ is continuous and the feasible set \mathcal{C} is compact, there is at least one optimal transition matrix P^* such that $\mu(P^*) = \mu^*$. Let \bar{P} denote the average over the orbit of P^* under $\text{Aut}(\mathcal{G})$:

$$\bar{P} = \frac{1}{|\text{Aut}(\mathcal{G})|} \sum_{\sigma \in \text{Aut}(\mathcal{G})} \sigma(P^*).$$

This matrix is feasible because each $\sigma(P^*)$ is feasible and the feasible set is convex. By construction, it is also invariant under the actions of $\text{Aut}(\mathcal{G})$. Moreover, using the convexity of μ , we have $\mu(\bar{P}) \leq \mu(P^*)$. It follows that $\bar{P} \in \mathcal{F}$ and $\mu(\bar{P}) = \mu^*$. \square

As a result of Theorem 2.1, we can replace the constraint $P \in \mathcal{C}$ by $P \in \mathcal{F}$ in the FMMC problem and get the same optimal value. In the fixed-point subset \mathcal{F} , the transition probabilities on the edges within an orbit must be the same. So we have the following corollaries.

COROLLARY 2.2. *The number of distinct edge transition probabilities we need to consider in the FMMC problem is at most equal to the number of orbits of \mathcal{E} under $\text{Aut}(\mathcal{G})$.*

COROLLARY 2.3. *If \mathcal{G} is edge-transitive, then all the edge transition probabilities can be assigned the same value.*

Note that the holding probabilities at the vertices can always be eliminated using $P_{ii} = 1 - \sum_j P_{ij}$ (of course we also need to add the constraint $\sum_j P_{ij} \leq 1$; see section 2.3). So it suffices to consider only the edge transition probabilities.

2.3. Formulation with reduced number of variables. With the results from the previous section, we can give an explicit parametrization of the FMMC problem with a reduced number of variables.

Recall that the *adjacency matrix* of a graph with n vertices is an $n \times n$ matrix A whose entries are given by $A_{ij} = 1$ if $\{i, j\} \in \mathcal{E}$ and $A_{ij} = 0$ otherwise. Let ν_i be the valency (degree) of vertex i . The *Laplacian matrix* of the graph is given by

$$L = \mathbf{Diag}(\nu_1, \nu_2, \dots, \nu_n) - A,$$

where $\mathbf{Diag}(\nu)$ denotes a diagonal matrix with the vector ν on its diagonal. An extensive account of the Laplacian matrix and its use in algebraic graph theory are provided in, e.g., [12, 28, 42].

Suppose that there are N orbits of edges under the action of $\text{Aut}(\mathcal{G})$. For each orbit, we define an orbit graph $\mathcal{G}_k = (\mathcal{V}, \mathcal{E}_k)$, where \mathcal{E}_k is the set of edges in the k th orbit. Note that the orbit graphs are disconnected if the original graph is not edge-transitive. Let L_k be the Laplacian matrix of \mathcal{G}_k . The diagonal entries $(L_k)_{ii}$ equal the valency of node i in \mathcal{G}_k (which is zero if vertex i is disconnected with all other vertices in \mathcal{G}_k).

By Corollary 2.2, we can assign the same transition probability on all the edges in the k th orbit. Denote this transition probability by p_k and let $p = (p_1, \dots, p_N)$. Then the transition probability matrix can be written as

$$(2.2) \quad P(p) = I - \sum_{k=1}^N p_k L_k.$$

This parametrization of the transition probability matrix automatically satisfies the constraints $P = P^T$, $P\mathbf{1} = \mathbf{1}$, and $P_{ij} = 0$ for $\{i, j\} \notin \mathcal{E}$. The entrywise nonnegativity constraint $P \geq 0$ now translates into

$$\begin{aligned} p_k &\geq 0, & k &= 1, \dots, N, \\ \sum_{k=1}^N (L_k)_{ii} p_k &\leq 1, & i &= 1, \dots, n, \end{aligned}$$

where the second set of constraints comes from the nonnegativity of the diagonal entries of P .

It can be verified that the parametrization (2.2), together with the above inequality constraints, is the precise characterization of the fixed-point subset \mathcal{F} . Therefore we can explicitly write the FMMC problem restricted to the fixed-point subset as

$$(2.3) \quad \begin{aligned} &\text{minimize} && \mu \left(I - \sum_{k=1}^N p_k L_k \right) \\ &\text{subject to} && p_k \geq 0, \quad k = 1, \dots, N, \\ &&& \sum_{k=1}^N (L_k)_{ii} p_k \leq 1, \quad i = 1, \dots, n. \end{aligned}$$

3. Some analytic results. For some special classes of graphs, the FMMC problem can be considerably simplified and often solved by exploiting symmetry only. In this section, we give some analytic results for the FMMC problem on edge-transitive graphs, Cartesian product of simple graphs, and distance-transitive graphs. The optimal solution is often expressed in terms of the eigenvalues of the Laplacian matrix of the graph. It is interesting to notice that even for such highly structured classes of graphs, neither the maximum-degree nor the Metropolis–Hastings heuristics discussed in [3] gives the optimal solution. Throughout, we use α^* to denote the optimal transition probability on all the edges, and μ^* to denote the optimal SLEM.

3.1. FMCM on edge-transitive graphs.

THEOREM 3.1. *Suppose the graph \mathcal{G} is edge-transitive, and let α be the transition probability assigned on all the edges. Then the optimal solution of the FMCM problem is*

$$(3.1) \quad \alpha^* = \min \left\{ \frac{1}{\nu_{\max}}, \frac{2}{\lambda_1(L) + \lambda_{n-1}(L)} \right\},$$

$$(3.2) \quad \mu^* = \max \left\{ 1 - \frac{\lambda_{n-1}(L)}{\nu_{\max}}, \frac{\lambda_1(L) - \lambda_{n-1}(L)}{\lambda_1(L) + \lambda_{n-1}(L)} \right\},$$

where $\nu_{\max} = \max_{i \in \mathcal{V}} \nu_i$ is the maximum valency of the vertices in the graph, and L is the Laplacian matrix defined in section 2.3.

Proof. By definition of an edge-transitive graph, there is a single orbit of edges under the actions of its automorphism group. Therefore we can assign the same transition probability α on all the edges in the graph (Corollary 2.3), and the parametrization (2.2) becomes $P = I - \alpha L$. So we have

$$\lambda_i(P) = 1 - \alpha \lambda_{n+1-i}(L), \quad i = 1, \dots, n,$$

and the SLEM

$$\begin{aligned} \mu(P) &= \max\{\lambda_2(P), -\lambda_n(P)\} \\ &= \max\{1 - \alpha \lambda_{n-1}(L), \alpha \lambda_1(L) - 1\}. \end{aligned}$$

To minimize $\mu(P)$, we let $1 - \alpha \lambda_{n-1}(L) = \alpha \lambda_1(L) - 1$ and get $\alpha = 2/(\lambda_{n-1}(L) + \lambda_{n-1}(L))$. But the nonnegativity constraint $P \geq 0$ requires that the transition probability must also satisfy $0 < \alpha \leq 1/\nu_{\max}$. Combining these two conditions gives the optimal solution (3.1) and (3.2). \square

3.1.1. Example: Cycle graphs. The cycle (or ring) graph C_n is a connected graph with $n \geq 3$ vertices, where each vertex has exactly two neighbors. Its Laplacian matrix is

$$L = \begin{bmatrix} 2 & -1 & 0 & \dots & 0 & -1 \\ -1 & 2 & -1 & \dots & 0 & 0 \\ 0 & -1 & 2 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 2 & -1 \\ -1 & 0 & 0 & \dots & -1 & 2 \end{bmatrix},$$

which has eigenvalues $2 - 2 \cos(2k\pi/n)$, $k = 1, \dots, n$. The two extreme eigenvalues are

$$\lambda_1(L) = 2 - 2 \cos \frac{2\lfloor n/2 \rfloor \pi}{n}, \quad \lambda_{n-1}(L) = 2 - 2 \cos \frac{2\pi}{n},$$

where $\lfloor n/2 \rfloor$ denotes the largest integer that is no larger than $n/2$, which is $n/2$ for n even or $(n - 1)/2$ for n odd. By Theorem 3.1, the optimal solution to the FMCM problem is

$$(3.3) \quad \alpha^* = \frac{1}{2 - \cos \frac{2\pi}{n} - \cos \frac{2\lfloor n/2 \rfloor \pi}{n}},$$

$$(3.4) \quad \mu^* = \frac{\cos \frac{2\pi}{n} - \cos \frac{2\lfloor n/2 \rfloor \pi}{n}}{2 - \cos \frac{2\pi}{n} - \cos \frac{2\lfloor n/2 \rfloor \pi}{n}}.$$

When $n \rightarrow \infty$, the transition probability $\alpha^* \rightarrow 1/2$ and the SLEM $\mu^* \rightarrow 1 - 2\pi^2/n^2$.

3.2. Cartesian product of graphs. Many graphs we consider can be constructed by taking *Cartesian product* of simpler graphs. The Cartesian product of two graphs $\mathcal{G}_1 = (\mathcal{V}_1, \mathcal{E}_1)$ and $\mathcal{G}_2 = (\mathcal{V}_2, \mathcal{E}_2)$ is a graph with vertex set $\mathcal{V}_1 \times \mathcal{V}_2$, where two vertices (u_1, u_2) and (v_1, v_2) are connected by an edge if and only if $u_1 = v_1$ and $\{u_2, v_2\} \in \mathcal{E}_2$, or $u_2 = v_2$ and $\{u_1, v_1\} \in \mathcal{E}_1$. Let $\mathcal{G}_1 \square \mathcal{G}_2$ denote this Cartesian product. Its Laplacian matrix is given by

$$(3.5) \quad L_{\mathcal{G}_1 \square \mathcal{G}_2} = L_{\mathcal{G}_1} \otimes I_{|\mathcal{V}_2|} + I_{|\mathcal{V}_1|} \otimes L_{\mathcal{G}_2},$$

where \otimes denotes the matrix Kronecker product [32]. The eigenvalues of $L_{\mathcal{G}_1 \square \mathcal{G}_2}$ are

$$(3.6) \quad \lambda_i(L_{\mathcal{G}_1}) + \lambda_j(L_{\mathcal{G}_2}), \quad i = 1, \dots, |\mathcal{V}_1|, \quad j = 1, \dots, |\mathcal{V}_2|,$$

where each eigenvalue is obtained as many times as its multiplicity (see, e.g., [43]).

Combining Theorem 3.1 and the above expression for eigenvalues, we can easily obtain solutions to the FMMC problem on graphs formed by taking Cartesian products.

3.2.1. Example: Mesh on a torus. Mesh on a torus is the Cartesian product of two copies of C_n . We write it as $M_n = C_n \square C_n$. By (3.6), its Laplacian matrix has eigenvalues

$$4 - 2 \cos \frac{2i\pi}{n} - 2 \cos \frac{2j\pi}{n}, \quad i, j = 1, \dots, n.$$

By Theorem 3.1, we obtain the optimal transition probability

$$\alpha^* = \frac{1}{3 - 2 \cos \frac{2\lfloor n/2 \rfloor \pi}{n} - \cos \frac{2\pi}{n}}$$

and the smallest SLEM

$$\mu^* = \frac{1 - 2 \cos \frac{2\lfloor n/2 \rfloor \pi}{n} + \cos \frac{2\pi}{n}}{3 - 2 \cos \frac{2\lfloor n/2 \rfloor \pi}{n} - \cos \frac{2\pi}{n}}.$$

When $n \rightarrow \infty$, the transition probability $\alpha^* \rightarrow 1/4$ and the SLEM $\mu^* \rightarrow 1 - \pi^2/n^2$.

3.2.2. Example: Hypercubes. The d -dimensional hypercube, denoted Q_d , has 2^d vertices, each labeled with a binary word with length d . Two vertices are connected by an edge if their words differ in exactly one component. This graph is isomorphic to the Cartesian product of d copies of K_2 , the complete graph with two vertices. The Laplacian of K_2 is

$$L_{K_2} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix},$$

whose two eigenvalues are 0 and 2. The one-dimensional hypercube Q_1 is just K_2 . Higher-dimensional hypercubes are defined recursively:

$$Q_{k+1} = Q_k \square K_2, \quad k = 1, 2, \dots$$

Using (3.5) and (3.6) recursively, the Laplacian of Q_d has eigenvalues $2k$, $k = 0, 1, \dots, d$, each with multiplicity $\binom{d}{k}$. The FMMC is achieved for

$$\alpha^* = \frac{1}{d+1}, \quad \mu^* = \frac{d-1}{d+1}.$$

This solution has also been obtained, for example, in [43].

3.3. FMMC on distance-transitive graphs. Distance-transitive graphs have been studied extensively in the literature (see, e.g., [9]). In particular, they are both edge- and vertex-transitive. In previous examples, the cycles and the hypercubes are actually distance-transitive graphs.

In a distance-transitive graph, all vertices have the same valency, which we denote by ν . The Laplacian matrix is $L = \nu I - A$, with A being the adjacency matrix. Therefore

$$\lambda_i(L) = \nu - \lambda_{n+1-i}(A), \quad i = 1, \dots, n.$$

We can substitute the above equation in (3.1) and (3.2) to obtain the optimal solution in terms of $\lambda_2(A)$ and $\lambda_n(A)$. For distance-transitive graphs, it is more convenient to use the *intersection matrix*, which has all the distinct eigenvalues of the adjacency matrix.

Let d be the *diameter* of the graph. For a nonnegative integer $k \leq d$, choose any two vertices u and v such that their distance satisfies $\delta(u, v) = k$. Let $a_k, b_k,$ and c_k be the number of vertices that are adjacent to u and whose distance from v are $k, k + 1,$ and $k - 1,$ respectively. That is,

$$\begin{aligned} a_k &= |\{w \in \mathcal{V} \mid \delta(u, w) = 1, \delta(w, v) = k\}|, \\ b_k &= |\{w \in \mathcal{V} \mid \delta(u, w) = 1, \delta(w, v) = k + 1\}|, \\ c_k &= |\{w \in \mathcal{V} \mid \delta(u, w) = 1, \delta(w, v) = k - 1\}|. \end{aligned}$$

For distance-transitive graphs, these numbers are independent of the particular pair of vertices u and v chosen. Clearly, we have $a_0 = 0, b_0 = \nu,$ and $c_1 = 1.$ The *intersection matrix* B is the following tridiagonal $(d + 1) \times (d + 1)$ matrix:

$$B = \begin{bmatrix} a_0 & b_0 & & & & \\ c_1 & a_1 & b_1 & & & \\ & c_2 & a_2 & \ddots & & \\ & & \ddots & \ddots & b_{d-1} & \\ & & & c_d & a_d & \end{bmatrix}.$$

Denote the eigenvalues of the intersection matrix, arranged in decreasing order, as $\eta_0, \eta_1, \dots, \eta_d.$ These are precisely the $(d + 1)$ distinct eigenvalues of the adjacency matrix A (see, e.g., [1]). In particular, we have

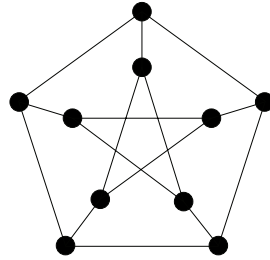
$$\lambda_1(A) = \eta_0 = \nu, \quad \lambda_2(A) = \eta_1, \quad \lambda_n(A) = \eta_d.$$

The following corollary is a direct consequence of Theorem 3.1.

COROLLARY 3.2. *The optimal solution to the FMMC problem on distance-transitive graphs is*

$$(3.7) \quad \alpha^* = \min \left\{ \frac{1}{\nu}, \frac{2}{2\nu - (\eta_1 + \eta_d)} \right\},$$

$$(3.8) \quad \mu^* = \max \left\{ \frac{\eta_1}{\nu}, \frac{\eta_1 - \eta_d}{2\nu - (\eta_1 + \eta_d)} \right\}.$$

FIG. 3.1. *The Petersen graph.*

3.3.1. Example: Petersen graph. The Petersen graph, shown in Figure 3.1, is a well-known distance-transitive graph with 10 vertices and 15 edges. The diameter of the graph is $d = 2$, and the intersection matrix is

$$B = \begin{bmatrix} 0 & 3 & 0 \\ 1 & 0 & 2 \\ 0 & 1 & 2 \end{bmatrix},$$

with eigenvalues $\eta_0 = 3$, $\eta_1 = 1$, and $\eta_2 = -2$. Applying (3.7) and (3.8), we obtain

$$\alpha^* = \frac{2}{7}, \quad \mu^* = \frac{3}{7}.$$

3.3.2. Example: Hamming graphs. The Hamming graphs, denoted $H(d, n)$, have vertices labeled by elements in the Cartesian product $\{1, \dots, n\}^d$, with two vertices being adjacent if they differ in exactly one component. By the definition, it is clear that Hamming graphs are isomorphic to the Cartesian product of d copies of the complete graph K_n . Hamming graphs are distance-transitive, with diameter d and valency $\nu = d(n - 1)$. Their eigenvalues are given by $\eta_k = d(n - 1) - kn$ for $k = 0, \dots, d$. These can be obtained using an equation for eigenvalues of adjacency matrices, similar to (3.6), with the eigenvalues of K_n being $n - 1$ and -1 . Therefore we have

$$\alpha^* = \min \left\{ \frac{1}{d(n-1)}, \frac{2}{n(d+1)} \right\},$$

$$\mu^* = \max \left\{ 1 - \frac{n}{d(n-1)}, \frac{d-1}{d+1} \right\}.$$

We note that hypercubes (see section 3.2.2) are special Hamming graphs with $n = 2$.

3.3.3. Example: Johnson graphs. The Johnson graph $J(n, q)$ (for $1 \leq q \leq n/2$) is defined as follows: the vertices are the q -element subsets of $\{1, \dots, n\}$, with two vertices being connected with an edge if and only if the subsets differ exactly by one element. It is a distance-transitive graph, with $\binom{n}{q}$ vertices and $\frac{1}{2}q(n-q)\binom{n}{q}$ edges. It has valency $\nu = q(n-q)$ and diameter q . The eigenvalues of the intersection matrix can be computed analytically, and they are

$$\eta_k = q(n-q) + k(k-n-1), \quad k = 0, \dots, q.$$

Therefore, by Corollary 3.2, we obtain the optimal transition probability

$$\alpha^* = \min \left\{ \frac{1}{q(n-q)}, \frac{2}{qn + n + q - q^2} \right\}$$

and the smallest SLEM

$$\mu^* = \max \left\{ 1 - \frac{n}{q(n-q)}, 1 - \frac{2n}{qn + n + q - q^2} \right\}.$$

4. FMCMC on orbit graphs. For graphs with large automorphism groups, the eigenvalues of the transition probability matrix often have very high multiplicities. To solve the FMCMC problem, it suffices to work with only the distinct eigenvalues without consideration of their multiplicities. This is exactly what the intersection matrix does for distance-transitive graphs. In this section we develop similar tools for more general graphs, based on the orbit theory developed in [4]. More specifically, we show how to construct an orbit chain which is much smaller in size than the original Markov chain, but contains all its distinct eigenvalues (with much fewer multiplicities). The FMCMC on the original graph can be found by solving a much smaller problem on the orbit chain.

4.1. Orbit theory. Let P be a symmetric Markov chain on the graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, and let H be a group of automorphisms of the graph. Often, it is a subgroup of the full automorphism group $\text{Aut}(\mathcal{G})$. The vertex set \mathcal{V} partitions into orbits $O_v = \{\sigma(v) : \sigma \in H\}$. For notational convenience, in this section we use $P(v, u)$, for $v, u \in \mathcal{V}$, to denote entries of the transition probability matrix. We define the *orbit chain* by specifying the transition probabilities between orbits:

$$(4.1) \quad P_H(O_v, O_u) = P(v, O_u) = \sum_{u' \in O_u} P(v, u').$$

This transition probability is independent of which $v \in O_v$ is chosen, so it is well defined and the lumped orbit chain is indeed Markov (see [4]).

The orbit chain is in general no longer symmetric, but it is always reversible. Let $\pi(i)$, $i \in \mathcal{V}$, be the stationary distribution of the original Markov chain. Then the stationary distribution on the orbit chain is obtained as

$$(4.2) \quad \pi_H(O_v) = \sum_{i \in O_v} \pi(i).$$

It can be verified that

$$(4.3) \quad \pi_H(O_v)P_H(O_v, O_u) = \pi_H(O_u)P_H(O_u, O_v),$$

which is the detailed balance condition to test reversibility.

The following is a summary of the orbit theory developed in [4], which relate the eigenvalues and eigenvectors of the orbit chain P_H to those of the original chain P .

- *Lifting.* If $\bar{\lambda}$ is an eigenvalue of P_H with associated eigenvector \bar{f} , then $\bar{\lambda}$ is an eigenvalue of P with H -invariant eigenfunction $f(v) = \bar{f}(O_v)$. Conversely, every H -invariant eigenfunction appears uniquely from this construction.
- *Projection.* Let λ be an eigenvalue of P with eigenvector f . Define a function on the orbits: $\bar{f}(O_v) = \sum_{\sigma \in H} f(\sigma^{-1}(v))$. Then λ appears as an eigenvalue of P_H , with eigenvector \bar{f} , if either of the following two conditions holds:
 - (a) H has a fixed point v^* and $f(v^*) \neq 0$.
 - (b) f is nonzero at a vertex v^* in an $\text{Aut}(\mathcal{G})$ -orbit which contains a fixed point of H .

Equipped with this orbit theory, we would like to construct one or multiple orbit chains that retain all the distinct eigenvalues of the original Markov chain. The following theorem (Theorem 3.7 in [4]) gives sufficient conditions for this to happen.

THEOREM 4.1. *Suppose that $\mathcal{V} = O_1 \cup \dots \cup O_K$ is a disjoint union of the orbits under $\text{Aut}(\mathcal{G})$. Let H_i be the subgroup of $\text{Aut}(\mathcal{G})$ that has a fixed point in O_i . Then all eigenvalues of P occur among the eigenvalues of $\{P_{H_i}\}_{i=1}^K$. Further, every eigenvector of P occurs by lifting an eigenvector of some P_{H_i} .*

Observe that if $H \subseteq G \subseteq \text{Aut}(\mathcal{G})$, then the eigenvalues of P_H contain all eigenvalues of P_G . This allows disregarding some of the H_i in Theorem 4.1. In particular, it is possible to construct a single orbit chain that contains all distinct eigenvalues of the original chain. Therefore we have the following corollary.

COROLLARY 4.2. *Suppose that $\mathcal{V} = O_1 \cup \dots \cup O_k$ is a disjoint union of the orbits under $\text{Aut}(\mathcal{G})$, and H is a subgroup of $\text{Aut}(\mathcal{G})$. If H has a fixed point in every O_i , then all distinct eigenvalues of P occur among the eigenvalues of P_H .*

Remark. To find H in the above corollary, we can just compute the corresponding stabilizer, i.e., compute the largest subgroup of $\text{Aut}(\mathcal{G})$ that fixes one point in each orbit. Note that the H promised by the corollary may be trivial in some cases; see [4, Remark 3.10].

4.2. Fastest mixing reversible Markov chain on orbit graph. Since in general the orbit chain is no longer symmetric, we cannot directly use the convex optimization formulation (1.1) or (1.2) to minimize $\mu(P_H)$. Fortunately, the detailed balance condition (4.3) leads to a simple transformation that allows us to formulate the problem of finding the fastest reversible Markov chain as a convex program [3].

Suppose the orbit chain P_H contains all distinct eigenvalues of the original chain. Let π_H be the stationary distribution of the orbits, and let $\Pi = \mathbf{Diag}(\pi_H)$. The detailed balance condition (4.3) can be written as $\Pi P_H = P_H^T \Pi$, which implies that the matrix $\Pi^{1/2} P_H \Pi^{-1/2}$ is symmetric (and, of course, has the same eigenvalues as P_H). The eigenvector of $\Pi^{1/2} P_H \Pi^{-1/2}$ associated with the maximum eigenvalue 1 is

$$q = \left(\sqrt{\pi_H(O_1)}, \dots, \sqrt{\pi_H(O_k)} \right).$$

The SLEM $\mu(P_H)$ equals the spectral norm of $\Pi^{1/2} P_H \Pi^{-1/2}$ restricted to the orthogonal complement of the subspace spanned by q . This can be written as

$$\mu(P_H) = \|(I - qq^T)\Pi^{1/2} P_H \Pi^{-1/2} (I - qq^T)\|_2 = \|\Pi^{1/2} P_H \Pi^{-1/2} - qq^T\|_2.$$

Introducing a scalar variable s to bound the above spectral norm, we can formulate the fastest mixing reversible Markov chain problem as an SDP:

$$(4.4) \quad \begin{aligned} & \text{minimize} && s \\ & \text{subject to} && -sI \preceq \Pi^{1/2} P_H \Pi^{-1/2} - qq^T \preceq sI, \\ & && P_H \geq 0, \quad P_H \mathbf{1} = \mathbf{1}, \quad \Pi P_H = P_H^T \Pi, \\ & && P_H(O, O') = 0, \quad (O, O') \notin \mathcal{E}_H. \end{aligned}$$

The optimization variables are the matrix P_H and scalar s , and problem data are given by the orbit graph and the stationary distribution π_H . Note that the reversibility constraint $\Pi P_H = P_H^T \Pi$ can be dropped since it is always satisfied by the construction of the orbit chain; see (4.3). By pre- and postmultiplying the matrix inequality by

$\Pi^{1/2}$, we can then write another equivalent formulation:

$$\begin{aligned}
 & \text{minimize} && s \\
 & \text{subject to} && -s\Pi \preceq \Pi P_H - \pi_H \pi_H^T \preceq s\Pi, \\
 (4.5) & && P_H \geq 0, \quad P_H \mathbf{1} = \mathbf{1}, \\
 & && P_H(O, O') = 0, \quad (O, O') \notin \mathcal{E}_H.
 \end{aligned}$$

To solve the fastest mixing reversible Markov chain problem on the orbit graph, we need the following three steps:

1. Conduct symmetry analysis on the original graph: identify the automorphism group $\text{Aut}(\mathcal{G})$ and determine the number of orbits of edges N . By Corollary 2.2, this is the number of transition probabilities we need to consider.
2. Find a group of automorphisms H that satisfies the conditions in Corollary 4.2. Construct its orbit chain by computing the transition probabilities using (4.1), and compute the stationary distribution using (4.2). Note that the entries of P_H are multiples of the transition probabilities on the original graph.
3. Solve the fastest mixing reversible Markov chain problem (4.4). The optimal SLEM $\mu(P_H^*)$ is also the optimal SLEM for the original chain, and the optimal transition probabilities on the original chain can be obtained by simple scaling of the optimal orbit transition probabilities.

We have assumed a single orbit chain that contains all distinct eigenvalues of the original chain. Sometimes it is more convenient to use multiple orbit chains. Let P_{H_i} , $i = 1, \dots, K$, be the collection of orbit chains in Theorem 4.1. In this case we need to minimize $\max_i \mu(P_{H_i})$. This can be done by simply adding the set of constraints in (4.4) for every matrix P_{H_i} .

Remark. The main challenge of implementing the above procedure is the identification of automorphism groups and construction of the orbit chains. Discussions on efficient algorithms or software that can automate these computational tasks are beyond the scope of this paper. We will give further remarks in our conclusions in section 6.

4.3. Example: K_n - K_n . We demonstrate the above computational procedure on the graph K_n - K_n . This graph consists of two copies of the complete graph K_n joined by a bridge (see Figure 4.1(a)). We follow the three steps described in section 4.2.

First, it is clear by inspection that the full automorphism group of K_n - K_n is $C_2 \times (S_{n-1} \times S_{n-1})$. The actions of $S_{n-1} \times S_{n-1}$ are all possible permutations of the two sets of $n - 1$ vertices, distinct from the two center vertices x and y , among themselves. The group C_2 acts on the graph by switching the two halves. The semidirect product symbol \times means that the actions of $S_{n-1} \times S_{n-1}$ and C_2 do not commute.

By symmetry analysis in section 2, there are three edge orbits under the full automorphism group: the bridging edge between vertices x and y , the edges connecting x and y to all other vertices, and the edges connecting all other vertices. Thus it suffices to consider just three transition probabilities p_0 , p_1 , and p_2 , each labeled in Figure 4.1(a) on one representative of the three edge orbits.

In the second step, we construct the orbit chains. The orbit chain of K_n - K_n under the full automorphism group is depicted in Figure 4.1(b). The orbit O_x includes vertices x and y , and the orbit O_z consists of all other $2(n - 1)$ vertices. The

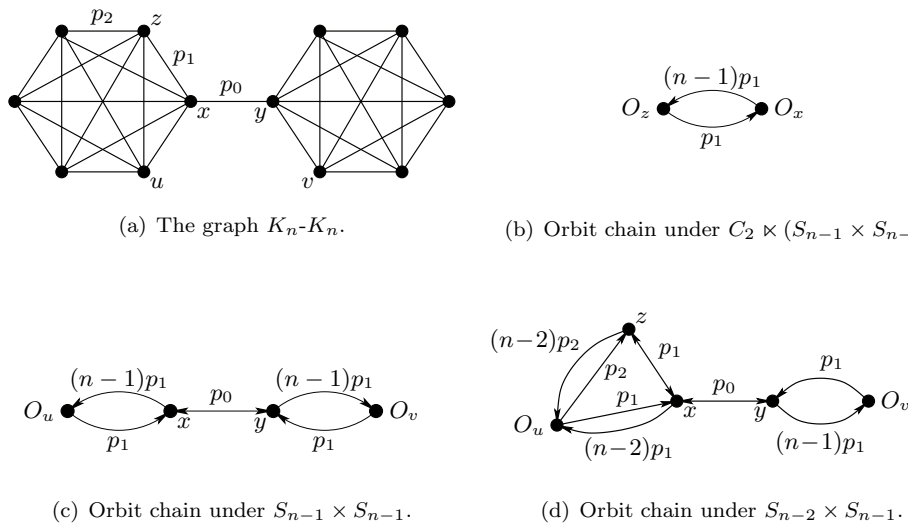


FIG. 4.1. The graph K_n-K_n and its orbit chains under different automorphism groups. Here O_x, O_z, O_u, O_v represent orbits of the vertices x, z, u, v labeled in Figure 4.1(a), respectively, under the corresponding automorphism groups in each subgroup.

transition probabilities of this orbit chain are calculated from (4.1) and are labeled on the directed edges in Figure 4.1(b). Similarly, the orbit chain under the subgroup $S_{n-1} \times S_{n-1}$ is depicted in Figure 4.1(c). While these two orbit chains are the most obvious to construct, none of them contains all eigenvalues of the original chain, nor does their combination. For the one in Figure 4.1(b), the full automorphism group does not have a fixed point in either its orbit O_x or O_z . For the one in Figure 4.1(c), the automorphism group $S_{n-1} \times S_{n-1}$ has a fixed point in O_x (either x or y), but does not have a fixed point in O_z (note here that O_z is the orbit of z under the full automorphism group). To fix the problem, we consider the orbit chain under the group $S_{n-2} \times S_{n-1}$, which leaves the vertices x, y , and z fixed, while permuting the remaining $n - 2$ vertices on the left and the $n - 1$ points on the right, respectively. The corresponding orbit chain is shown in Figure 4.1(d). By Corollary 4.2, all distinct eigenvalues of the original Markov chain on K_n-K_n appear as eigenvalues of this orbit chain. Thus there are at most five distinct eigenvalues in the original chain no matter how large n is.

To finish the second step, we calculate the transition probabilities of the orbit chain under $H = S_{n-2} \times S_{n-1}$ using (4.1) and label them in Figure 4.1(d). If we order the vertices of this orbit chain as (x, y, z, O_u, O_v) , then the transition probability matrix is

$$P_H = \begin{bmatrix} 1-p_0-(n-1)p_1 & p_0 & p_1 & (n-2)p_1 & 0 \\ p_0 & 1-p_0-(n-1)p_1 & 0 & 0 & (n-1)p_1 \\ p_1 & 0 & 1-p_1-(n-2)p_2 & (n-2)p_2 & 0 \\ p_1 & 0 & p_2 & 1-p_1-p_2 & 0 \\ 0 & p_1 & 0 & 0 & 1-p_1 \end{bmatrix}.$$

By (4.2), the stationary distribution of the orbit chain is

$$\pi_H = \left(\frac{1}{2n}, \frac{1}{2n}, \frac{1}{2n}, \frac{n-2}{2n}, \frac{n-1}{2n} \right).$$

As the third step, we solve the SDP (4.4) with the above parametrization. Here we only need to solve an SDP with 4 variables (three transition probabilities p_0, p_1, p_2 , and the extra scalar s) and 5×5 matrices no matter how large the graph (n) is.

We will revisit this example in section 5.4.4, where we present an analytic expression for the exact optimal SLEM and corresponding transition probabilities.

5. Symmetry reduction by block-diagonalization. By definition of the fixed-point subset \mathcal{F} in (2.1), any transition probability matrix $P \in \mathcal{F}$ is invariant under the actions of $\text{Aut}(\mathcal{G})$. More specifically, for any permutation matrix Q given by $\sigma \in \text{Aut}(\mathcal{G})$, we have $QPQ^T = P$, equivalently $QP = PQ$. In this section we show that this property allows the construction of a coordinate transformation matrix that can block-diagonalize every $P \in \mathcal{F}$. The resulting blocks usually have much smaller sizes, and repeated blocks can be discarded in computation.

The method we use in this section is based on classical group representation theory (see, e.g., [54]). It was developed for more general SDPs in [26] and has found applications in sum-of-squares decomposition for minimizing polynomial functions [48, 49, 50] and controller design for symmetric dynamical systems [13]. A closely related approach is developed in [14], which is based on a low-order representation of the commutant (collection of invariant matrices) of the matrix algebra generated by the permutation matrices.

5.1. Some group representation theory. Let G be a group. A *representation* ρ of G assigns an invertible matrix $\rho(g)$ to each $g \in G$ in such a way that the matrix assigned to the product of two elements in G is the product of the matrices assigned to each element: $\rho(gh) = \rho(g)\rho(h)$. The matrices we work with are all invertible and are considered over the real or complex numbers. We thus regard ρ as a homomorphism from G to the linear maps on a vector space V . The dimension of ρ is the dimension of V . Two representations are *equivalent* if they are related by a fixed similarity transformation.

If W is a subspace of V invariant under G , then ρ restricted to W gives a *subrepresentation*. Of course the zero subspace and the subspace $W = V$ are trivial subrepresentations. If the representation ρ admits no nontrivial subrepresentation, then ρ is called *irreducible*.

We consider first complex representations, as the theory is considerably simpler in this case. For a finite group G there are only finitely many inequivalent irreducible representations $\vartheta_1, \dots, \vartheta_h$ of dimensions n_1, \dots, n_h , respectively. The degrees n_i divide the group order $|G|$ and satisfy the condition $\sum_{i=1}^h n_i^2 = |G|$. Every linear representation of G has a canonical decomposition as a direct sum of irreducible representations

$$\rho = m_1\vartheta_1 \oplus m_2\vartheta_2 \oplus \dots \oplus m_h\vartheta_h,$$

where m_1, \dots, m_h are the multiplicities. Accordingly, the representation space \mathbf{C}^n has an *isotypic decomposition*

$$(5.1) \quad \mathbf{C}^n = V_1 \oplus \dots \oplus V_h,$$

where each isotypic components consists of m_i invariant subspaces

$$(5.2) \quad V_i = V_i^1 \oplus \dots \oplus V_i^{m_i},$$

each of which has dimension n_i and transforms after the manner of ϑ_i . A basis of this decomposition transforming with respect to the matrices $\vartheta_i(g)$ is called *symmetry-adapted* and can be computed using the algorithm presented in [54, sections 2.6–2.7]

or [22, section 5.2]. This basis defines a change of coordinates by a matrix T collecting the basis as columns. By Schur's lemma (see, e.g., [54]), if a matrix P satisfies

$$(5.3) \quad \rho(g)P = P\rho(g) \quad \forall g \in G,$$

then $T^{-1}PT$ has block-diagonal form with one block P_i for each isotypic component of dimension $m_i n_i$, which further decomposes into n_i equal blocks B_i of dimension m_i . That is,

$$(5.4) \quad T^{-1}PT = \begin{bmatrix} P_1 & & 0 \\ & \ddots & \\ 0 & & P_h \end{bmatrix}, \quad P_i = \begin{bmatrix} B_i & & 0 \\ & \ddots & \\ 0 & & B_i \end{bmatrix}.$$

For our application of semidefinite programs, the problems are presented in terms of real matrices, and therefore we would like to use real coordinate transformations. In fact a generalization of the classical theory to the real case is presented in [54, section 13.2]. If all $\vartheta_i(g)$ are real matrices, the irreducible representation is called *absolutely irreducible*. Otherwise, for each ϑ_i with complex character its complex conjugate will also appear in the canonical decomposition. Since ρ is real, both will have the same multiplicity, and real bases of $V_i + \bar{V}_i$ can be constructed. So two complex conjugate irreducible representations form one real irreducible representation of *complex type*. There is a third case, real irreducible representations of *quaternion type*, rarely seen in practical examples.

In this paper, we assume that the representation ρ is orthogonal, i.e., $\rho(g)^T \rho(g) = \rho(g)\rho(g)^T = I$ for all $g \in G$. As a result, the transformation matrix T can also be chosen to be orthogonal. Thus $T^{-1} = T^T$ (for complex matrices, it is the conjugate transpose). For symmetric matrices the block corresponding to a representation of complex type or quaternion type simplifies to a collection of equal subblocks. For the special case of circulant matrices, complete diagonalization reveals all the eigenvalues [15, p. 50].

5.2. Block-diagonalization of SDP constraint. As in section 2.2, for every $\sigma \in \text{Aut}(\mathcal{G})$ we assign a permutation matrix $Q(\sigma)$ by letting $Q_{ij}(\sigma) = 1$ if $i = \sigma(j)$ and $Q_{ij}(\sigma) = 0$ otherwise. This is an n -dimensional representation of $\text{Aut}(\mathcal{G})$, which is often called the *natural representation*. As mentioned in the beginning of this section, every matrix P in the fixed-point subset \mathcal{F} has the symmetry of $\text{Aut}(\mathcal{G})$; i.e., it satisfies the condition (5.3) with $\rho = Q$. Thus a coordinate transformation matrix T can be constructed such that P can be block-diagonalized into the form (5.4).

Now we consider the FMMC problem (2.3), which can be formulated as the following SDP:

$$(5.5) \quad \begin{aligned} & \text{minimize} && s \\ & \text{subject to} && -sI \preceq I - \sum_{k=1}^N p_k L_k - (1/n)\mathbf{1}\mathbf{1}^T \preceq sI, \\ & && p_k \geq 0, \quad k = 1, \dots, N, \\ & && \sum_{k=1}^N (L_k)_{ii} p_k \leq 1, \quad i = 1, \dots, n. \end{aligned}$$

Here we have expressed the transition probability matrix as $P(p) = I - \sum_{k=1}^N p_k L_k$, where L_k is the Laplacian matrix for the k th orbit graph, and p_k is the common

transition probability assigned on all edges in the k th orbit graph. Since the matrix $P(p)$ has the symmetry of $\text{Aut}(\mathcal{G})$, we can find a coordinate transformation T to block-diagonalize the linear matrix inequalities. Thus we obtain the following equivalent problem:

$$\begin{aligned}
 & \text{minimize} && s \\
 & \text{subject to} && -sI_{m_i} \preceq B_i(p) - J_i \preceq sI_{m_i}, \quad i = 1, \dots, h, \\
 (5.6) \quad & && p_k \geq 0, \quad k = 1, \dots, N, \\
 & && \sum_{k=1}^N (L_k)_{ii} p_k \leq 1, \quad i = 1, \dots, n,
 \end{aligned}$$

where $B_i(p)$ correspond to the small blocks B_i in (5.4) of the transformed matrix $T^T P(p) T$, and J_i are the corresponding diagonal blocks of $T^T (1/n) \mathbf{1} \mathbf{1}^T T$. The number of matrix inequalities h is the number of inequivalent irreducible representations, and the size of each matrix inequality m_i is the multiplicity of the corresponding irreducible representation. Note that we only need one out of n_i copies of each B_i in the decomposition (5.4). Since m_i can be much smaller than n (the number of vertices in the graph), the improvement in computational complexity over the SDP formulation (5.5) can be significant (see the flop counts discussed in section 1.2). This is especially the case when there are high-dimensional irreducible representations (i.e., when n_i is large; see, e.g., K_n - K_n defined in section 4.3).

5.3. Connection between block-diagonalization and orbit theory. With the following theorem, we establish an interesting connection between the block-diagonalization approach and the orbit theory in section 4.

THEOREM 5.1. *Let H be a subgroup of $\text{Aut}(\mathcal{G})$, and let T be the coordinate transformation matrix whose columns are a symmetry-adapted basis for the natural representation of H . Suppose a Markov chain P defined on the graph has the symmetry of H . Then the matrix $T^T (1/n) \mathbf{1} \mathbf{1}^T T$ has the same block-diagonal form as $T^T P T$. Moreover, there is only one nonzero block. Without loss of generality, let this nonzero block be J_1 and the corresponding block of $T^T P T$ be B_1 . These two blocks relate to the orbit chain P_H by*

$$(5.7) \quad B_1 = \Pi^{1/2} P_H \Pi^{-1/2},$$

$$(5.8) \quad J_1 = q q^T,$$

where $\Pi = \mathbf{Diag}(\pi_H)$, $q = \sqrt{\pi_H}$, and π_H is the stationary distribution of P_H .

Proof. First we note that P always has a single eigenvalue 1 with associated eigenvector $\mathbf{1}$. Thus $\mathbf{1}$ spans an invariant subspace of the natural representation, which is obviously irreducible. The corresponding irreducible representation is isomorphic to the trivial representation (which assigns the scalar 1 to every element in the group). Without loss of generality, let V_1 be the isotypic component that contains the vector $\mathbf{1}$. Thus V_1 is a direct product of H -fixed vectors (each corresponds to a copy of the trivial representation), and $\mathbf{1}$ is a linear combination of these vectors.

Let m_1 be the dimension of V_1 , which is the number of H -fixed vectors. We can calculate m_1 by Frobenius reciprocity, or ‘‘Burnside’s lemma’’; see, e.g., [54]. To do so, we note that the character χ of the natural representation $Q(g)$, $g \in H$, is the number of fixed points of g , i.e.,

$$\chi(g) = \mathbf{Tr} Q(g) = \text{FP}(g) = \#\{v \in \mathcal{V} : g(v) = v\}.$$

Burnside’s lemma says that

$$\frac{1}{|H|} \sum_{g \in H} \text{FP}(g) = \#\text{orbits}.$$

The left-hand side is the inner product of χ with the trivial representation. It thus counts the number of H -fixed vectors in V . So m_1 equals the number of orbits under H .

Suppose that $\mathcal{V} = O_1 \cup \dots \cup O_{m_1}$ as a disjoint union of H -orbits. Let $b_i(v) = 1/\sqrt{|O_i|}$ if $v \in O_i$ and zero otherwise. Then b_1, \dots, b_{m_1} are H -fixed vectors, and they form an orthonormal symmetry-adapted basis for V_1 (these are not unique). Let $T_1 = [b_1 \dots b_{m_1}]$ be the first m_1 columns of T . They are orthogonal to all other columns of T . Since $\mathbf{1}$ is a linear combination of b_1, \dots, b_{m_1} , it is also orthogonal to other columns of T . Therefore the matrix $T^T(1/n)\mathbf{1}\mathbf{1}^T T$ has all its elements zero except for the first $m_1 \times m_1$ diagonal block, which we denote as J_1 . More specifically, $J_1 = qq^T$, where

$$\begin{aligned} q &= \frac{1}{\sqrt{n}} T_1^T \mathbf{1} = \frac{1}{\sqrt{n}} [b_1^T \mathbf{1} \dots b_{m_1}^T \mathbf{1}]^T \\ &= \frac{1}{\sqrt{n}} \left[\frac{|O_1|}{\sqrt{|O_1|}} \dots \frac{|O_{m_1}|}{\sqrt{|O_{m_1}|}} \right]^T = \left[\sqrt{\frac{|O_1|}{n}} \dots \sqrt{\frac{|O_{m_1}|}{n}} \right]^T. \end{aligned}$$

Note that by (4.2) the stationary distribution of the orbit chain P_H is

$$\pi_H = \left[\frac{|O_1|}{n} \dots \frac{|O_{m_1}|}{n} \right]^T.$$

Thus we have $q = \sqrt{\pi_H}$. This proves (5.8).

Finally we consider the relationship between the two matrices $B_1 = T_1^T P T_1$ and P_H . We prove (5.7) by showing

$$\Pi^{-1/2} B_1 \Pi^{1/2} = \Pi^{-1/2} T_1^T P T_1 \Pi^{1/2} = P_H.$$

It is straightforward to verify that

$$\begin{aligned} \Pi^{-1/2} T_1^T &= \sqrt{n} \begin{bmatrix} b_1^T \\ \vdots \\ b_{m_1}^T \end{bmatrix}, & b_i'(v) &= \begin{cases} \frac{1}{|O_i|} & \text{if } v \in O_i, \\ 0 & \text{if } v \notin O_i, \end{cases} \\ T_1 \Pi^{1/2} &= \frac{1}{\sqrt{n}} [b_1'' \dots b_{m_1}''], & b_i''(v) &= \begin{cases} 1 & \text{if } v \in O_i, \\ 0 & \text{if } v \notin O_i. \end{cases} \end{aligned}$$

The entry at the i th row and j th column of the matrix $\Pi^{-1/2} T_1^T P T_1 \Pi^{1/2}$ is given by

$$b_i^T P b_j'' = \frac{1}{|O_i|} \sum_{v \in O_i} \sum_{u \in O_j} P(v, u) = \frac{1}{|O_i|} \sum_{v \in O_i} P_H(O_i, O_j) = P_H(O_i, O_j).$$

In the last equation, we have used the fact that $P_H(O_i, O_j)$ is independent of which $v \in O_i$ is chosen. This completes the proof. \square

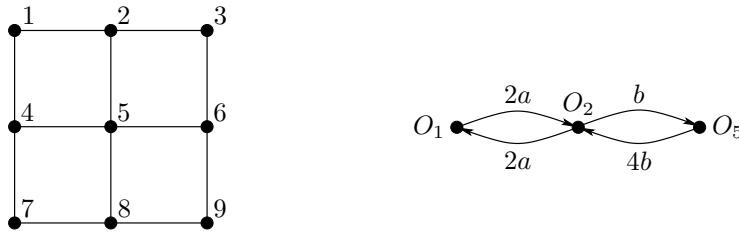


FIG. 5.1. Left: the 3×3 grid. Right: its orbit chain under D_4 .

From Theorem 5.1, we know that B_1 contains the eigenvalues of the orbit chain under H . Other blocks B_i contain additional eigenvalues (not including those of P_H) of the orbit chains under various subgroups of H . (Note that the eigenvalues of the orbit chain under H are always contained in the orbit chain under its subgroups.) With this observation, it is possible to identify the multiplicities of eigenvalues in orbit chains under various subgroups of $\text{Aut}(\mathcal{G})$ by relating to the decompositions (5.1), (5.2), and (5.4) (some preliminary results are discussed in [4]).

5.4. Examples. We present several examples that use the block-diagonalization method and draw connections to the method based on orbit theory in section 4. Some of the examples may be difficult if one uses the orbit theory alone, but are nicely handled by block-diagonalization.

5.4.1. The 3×3 grid. Consider the symmetric Markov chain on a 3×3 grid \mathcal{G} ; see Figure 5.1(left). The automorphism group $\text{Aut}(\mathcal{G})$ is isomorphic to the 8-element dihedral group D_4 , and corresponds to flips and 90-degree rotations of the graph. The orbits of $\text{Aut}(\mathcal{G})$ acting on the vertices are

$$\{1, 3, 7, 9\}, \quad \{5\}, \quad \{2, 4, 6, 8\},$$

and there are two orbits of edges

$$\{\{1, 2\}, \{1, 4\}, \{2, 3\}, \{3, 6\}, \{4, 7\}, \{7, 8\}, \{6, 9\}, \{8, 9\}\},$$

$$\{\{2, 5\}, \{4, 5\}, \{5, 6\}, \{5, 8\}\}.$$

So \mathcal{G} is neither vertex- nor edge-transitive.

By Corollary 2.2, we associate transition probabilities a and b to the two edge orbits, respectively. The transition probability matrix has the form

$$P = \begin{bmatrix} 1-2a & a & 0 & a & 0 & 0 & 0 & 0 & 0 \\ a & 1-2a-b & a & 0 & b & 0 & 0 & 0 & 0 \\ 0 & a & 1-2a & 0 & 0 & a & 0 & 0 & 0 \\ a & 0 & 0 & 1-2a-b & b & 0 & a & 0 & 0 \\ 0 & b & 0 & b & 1-4b & b & 0 & b & 0 \\ 0 & 0 & a & 0 & b & 1-2a-b & 0 & 0 & a \\ 0 & 0 & 0 & a & 0 & 0 & 1-2a & a & 0 \\ 0 & 0 & 0 & 0 & b & 0 & a & 1-2a-b & a \\ 0 & 0 & 0 & 0 & 0 & a & 0 & a & 1-2a \end{bmatrix}.$$

The matrix P satisfies $Q(\sigma)P = PQ(\sigma)$ for every $\sigma \in \text{Aut}(\mathcal{G})$. Using the algorithm in [22, section 5.2], we found a symmetry-adapted basis for the representation Q , which

we take as columns to form

$$T = \frac{1}{2} \begin{bmatrix} 0 & 1 & 0 & 1 & 0 & \sqrt{2} & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & -1 & 0 & 0 & 0 & \sqrt{2} & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & -1 \\ 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & -1 & 0 & 1 \\ 0 & 1 & 0 & -1 & 0 & 0 & 0 & -\sqrt{2} & 0 \\ 0 & 0 & 1 & 0 & -1 & 0 & -1 & 0 & -1 \\ 0 & 1 & 0 & 1 & 0 & -\sqrt{2} & 0 & 0 & 0 \end{bmatrix}.$$

With this coordinate transformation matrix, we obtain

$$T^T P T = \begin{bmatrix} 1-4b & 0 & 2b & & & & & & \\ 0 & 1-2a & 2a & & & & & & \\ 2b & 2a & 1-2a-b & & & & & & \\ & & & 1-2a & & & & & \\ & & & & 1-2a-b & & & & \\ & & & & & 1-2a & \sqrt{2}a & & \\ & & & & & \sqrt{2}a & 1-2a-b & & \\ & & & & & & & 1-2a & \sqrt{2}a \\ & & & & & & & \sqrt{2}a & 1-2a-b \end{bmatrix}.$$

The three-dimensional block B_1 contains the eigenvalue 1, and it is related to the orbit chain in Figure 5.1 (right) by (5.7). The corresponding nonzero block of $T^T(1/n)\mathbf{1}\mathbf{1}^T T$ is

$$J_1 = \frac{1}{9} \begin{bmatrix} 1 & 2 & 2 \\ 2 & 4 & 4 \\ 2 & 4 & 4 \end{bmatrix}.$$

Next, we substitute the above expressions into the SDP (5.6) and solve it numerically. Since there are repeated 2×2 blocks, the original 9×9 matrix is replaced by four smaller blocks of dimension 3, 1, 1, 2. The optimal solutions are

$$a^* \approx 0.363, \quad b^* \approx 0.2111, \quad \mu^* \approx 0.6926.$$

Interestingly, it can be shown that these optimal values are not rational, but instead algebraic numbers with defining minimal polynomials:

$$\begin{aligned} 18157 a^5 - 17020 a^4 + 6060 a^3 - 1200 a^2 + 180 a - 16 &= 0, \\ 1252833 b^5 - 1625651 b^4 + 791936 b^3 - 173536 b^2 + 15360 b - 256 &= 0, \\ 54471 \mu^5 - 121430 \mu^4 + 88474 \mu^3 - 18216 \mu^2 - 2393 \mu + 262 &= 0. \end{aligned}$$

5.4.2. Complete k -partite graphs. The complete k -partite graph, denoted K_{n_1, \dots, n_k} , has k subsets of vertices with cardinalities n_1, \dots, n_k , respectively. Each vertex is connected to all the vertices in a different subset, and is not connected to any of the vertices in the same subset. In this case, the transition probability matrix has dimensions $\sum_i n_i$ and the structure

$$P(p) = \begin{bmatrix} (1 - \sum_{j \neq 1} n_j p_{1j}) I_{n_1} & p_{12} \mathbf{1}_{n_1 \times n_2} & \cdots & p_{1k} \mathbf{1}_{n_1 \times n_k} \\ p_{21} \mathbf{1}_{n_2 \times n_1} & (1 - \sum_{j \neq 2} n_j p_{2j}) I_{n_2} & \cdots & p_{2k} \mathbf{1}_{n_2 \times n_k} \\ \vdots & \vdots & \ddots & \vdots \\ p_{k1} \mathbf{1}_{n_k \times n_1} & p_{k2} \mathbf{1}_{n_k \times n_2} & \cdots & (1 - \sum_{j \neq k} n_j p_{kj}) I_{n_k} \end{bmatrix},$$

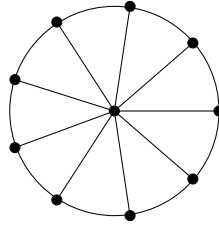


FIG. 5.2. The wheel graph with $n = 9$ (total of 10 nodes).

where the probabilities satisfy $p_{ij} = p_{ji}$. There are a total of $\binom{k}{2}$ independent variables.

We can easily find a decomposition of the associated matrix algebra. Using the orthogonal coordinate transformation matrix

$$T = \begin{bmatrix} (1/\sqrt{n_1})\mathbf{1}_{n_1 \times 1} & \cdots & \mathbf{0} & F_{n_1} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & (1/\sqrt{n_k})\mathbf{1}_{n_k \times 1} & \mathbf{0} & \cdots & F_{n_k} \end{bmatrix},$$

the matrix $T^T P(p) T$ decomposes into $k + 1$ blocks: one of dimension k , with the remaining k blocks each having dimension $n_i - 1$. The decomposition is

$$\begin{bmatrix} (1 - \sum_{j \neq 1} n_j p_{1j}) & p_{12} \sqrt{n_1 n_2} & \cdots & p_{1k} \sqrt{n_1 n_k} \\ p_{21} \sqrt{n_2 n_1} & (1 - \sum_{j \neq 2} n_j p_{2j}) & \cdots & p_{2k} \sqrt{n_2 n_k} \\ \vdots & \vdots & \ddots & \vdots \\ p_{k1} \sqrt{n_k n_1} & p_{k2} \sqrt{n_k n_2} & \cdots & (1 - \sum_{j \neq k} n_j p_{kj}) \end{bmatrix},$$

$$I_{n_i-1} \otimes \left(1 - \sum_{j \neq i} n_j p_{ij} \right), \quad i = 1, \dots, k.$$

These blocks can be substituted into the SDP (5.6) to solve the FMMC problem.

For the complete bipartite graph $K_{m,n}$, there is only one transition probability p , and the matrix $T^T P(p) T$ has the following diagonal blocks:

$$\begin{bmatrix} 1 - mp & p\sqrt{nm} \\ p\sqrt{nm} & 1 - np \end{bmatrix}, \quad I_{n-1} \otimes (1 - mp), \quad I_{m-1} \otimes (1 - np).$$

The 2×2 block has eigenvalues 1 and $1 - (m + n)p$. The other diagonals reveal the eigenvalue $1 - mp$ and $1 - np$, with multiplicities $n - 1$ and $m - 1$, respectively. By Theorem 3.1, we have (assuming $m \leq n$)

$$p^* = \min \left\{ \frac{1}{n}, \frac{2}{n + 2m} \right\}, \quad \mu^* = \max \left\{ \frac{n - m}{n}, \frac{n}{n + 2m} \right\}.$$

5.4.3. Wheel graph. The wheel graph consists of a center vertex (the *hub*) and a ring of n peripheral vertices, each connected to the hub; see Figure 5.2. It has a total of $n + 1$ nodes. Its automorphism group is isomorphic to the dihedral group D_n

with order $2n$. The transition probability matrix has the structure

$$(5.9) \quad P = \begin{bmatrix} 1 - np & p & p & \dots & p & p \\ p & 1 - p - 2q & q & \dots & 0 & q \\ p & q & 1 - p - 2q & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ p & 0 & 0 & \dots & 1 - p - 2q & q \\ p & q & 0 & \dots & q & 1 - p - 2q \end{bmatrix},$$

where p and q are the transition probabilities between the hub and each peripheral vertex, and between adjacent peripheral vertices, respectively.

For this structure, the block-diagonalizing transformation is given by

$$T = \mathbf{Diag}(1, \mathcal{F}_n), \quad [\mathcal{F}_n]_{jk} = \frac{1}{\sqrt{n}} e^{\frac{2\pi i(j-1)(k-1)}{n}},$$

where \mathcal{F}_n is the unitary Fourier matrix of size $n \times n$. As a consequence, the matrix $T^{-1}PT$ is block-diagonal with a 2×2 matrix and $n - 1$ scalars on its diagonal, given by

$$\begin{bmatrix} 1 - np & \sqrt{np} \\ \sqrt{np} & 1 - p \end{bmatrix}$$

and

$$1 - p + (\omega_n^k + \omega_n^{-k} - 2) \cdot q, \quad k = 1, \dots, n - 1,$$

where $\omega_n = e^{\frac{2\pi i}{n}}$ is an elementary n th root of unity. The 2×2 block is B_1 , which contains eigenvalues of the orbit chain under D_n (it has only two orbits).

With the above decomposition, we obtain the optimal solution to the FMMC problem in closed form:

$$p^* = \frac{1}{n}, \quad q_* = \frac{1 - \frac{1}{n}}{2 - \cos \frac{2\pi}{n} - \cos \frac{2|n/2|\pi}{n}}.$$

The optimal value of the SLEM is

$$\mu^* = \left(1 - \frac{1}{n}\right) \frac{\cos \frac{2\pi}{n} - \cos \frac{2|n/2|\pi}{n}}{2 - \cos \frac{2\pi}{n} - \cos \frac{2|n/2|\pi}{n}}.$$

Compared with the optimal solution for the cycle graph in (3.3) and (3.4), we see an extra factor of $1 - 1/n$ in both the SLEM and the transition probability between peripheral vertices. This is exactly the factor improved by adding the central hub over the pure n -cycle case.

The wheel graph is an example for which the block-diagonalization technique works out nicely, while the orbit theory leads to much less reduction. Although there are only two orbits under the full automorphism group, any orbit graph that has a fixed peripheral vertex will have at least $(n + 1)/2$ orbits (corresponding symmetry is reflection through that vertex).

5.4.4. K_n - K_n . We did careful symmetry analysis for the graph K_n - K_n in section 4.3; see Figure 4.1. The transition probability matrix on this graph has the structure

$$P = \begin{bmatrix} C & p_1 \mathbf{1} & 0 & 0 \\ p_1 \mathbf{1}^T & 1 - p_0 - (n-1)p_1 & p_0 & 0 \\ 0 & p_0 & 1 - p_0 - (n-1)p_1 & p_1 \mathbf{1}^T \\ 0 & 0 & p_1 \mathbf{1} & C \end{bmatrix},$$

where C is a circulant matrix

$$C = (1 - p_1 - (n-3)p_2)I_{n-1} + p_2 \mathbf{1}_{(n-1) \times (n-1)}.$$

Since circulant matrices are diagonalized by Fourier matrices, we first use the transformation matrix

$$T_1 = \begin{bmatrix} \mathcal{F}_{n-1} & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & \mathcal{F}_{n-1} \end{bmatrix},$$

where \mathcal{F}_{n-1} is the unitary Fourier matrix of dimension $n-1$. This corresponds to block diagonalization using the symmetry group $S_{n-1} \times S_{n-1}$, which is a subgroup of $\text{Aut}(K_n$ - $K_n)$. The matrix $T_1^{-1}PT_1$ has diagonal blocks

$$B'_1 = \begin{bmatrix} 1 - p_1 & \sqrt{n-1}p_1 & 0 & 0 \\ \sqrt{n-1}p_1 & 1 - p_0 - (n-1)p_1 & p_0 & 0 \\ 0 & p_0 & 1 - p_0 - (n-1)p_1 & \sqrt{n-1}p_1 \\ 0 & 0 & \sqrt{n-1}p_1 & 1 - p_1 \end{bmatrix}$$

and

$$(5.10) \quad I_{2n-4} \otimes (1 - p_1 - (n-1)p_2).$$

From this we know that P has an eigenvalue $1 - p_1 - (n-1)p_2$ with multiplicity $2n-4$, and the remaining four eigenvalues are the eigenvalues of the above 4×4 block B'_1 . The block B'_1 corresponds to the orbit chain under the symmetry group $H = S_{n-1} \times S_{n-1}$. More precisely, $B'_1 = \Pi^{1/2}P_H\Pi^{-1/2}$, where $\Pi = \text{Diag}(\pi_H)$, and P_H and π_H are the transition probability matrix and stationary distribution of the orbit chain shown in Figure 4.1(c), respectively.

Exploring the full automorphism group of K_n - K_n , we can further block-diagonalize B'_1 . Let

$$T = T_1 \begin{bmatrix} I_{n-2} & & \\ & T_2 & \\ & & I_{n-2} \end{bmatrix}, \quad T_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & -1 & 0 \\ 1 & 0 & 0 & -1 \end{bmatrix}.$$

The 4×4 block B'_1 is decomposed into

$$\begin{bmatrix} 1 - p_1 & \sqrt{n-1}p_1 \\ \sqrt{n-1}p_1 & 1 - (n-1)p_1 \end{bmatrix}, \quad \begin{bmatrix} 1 - 2p_0 - (n-1)p_1 & \sqrt{n-1}p_1 \\ \sqrt{n-1}p_1 & 1 - p_1 \end{bmatrix}.$$

The first block is B_1 , which has eigenvalues 1 and $1 - np_1$. By Theorem 5.1, B_1 is related to the orbit chain under $\text{Aut}(K_n - K_n)$ (see Figure 4.1(b)) by (5.7). The second 2×2 block has eigenvalues

$$1 - p_0 - (1/2)np_1 \pm \sqrt{(p_0 + (1/2)np_1)^2 - 2p_0p_1}.$$

These are the eigenvalues contained in the orbit chain of Figure 4.1(c) but not in Figure 4.1(b).

In summary, the distinct eigenvalues of the Markov chain on $K_n - K_n$ are

$$1, \quad 1 - np_1, \quad 1 - p_0 - (1/2)np_1 \pm \sqrt{(p_0 + (1/2)np_1)^2 - 2p_0p_1}, \quad 1 - p_1 - (n-1)p_2,$$

where the last one has multiplicity $2n-4$. As we mentioned before, the huge reduction for $K_n - K_n$ is due to the fact that it has an irreducible representation with high dimension $2n-4$ and multiplicity 1 (see [4, Proposition 2.4]). In the decomposition (5.4), this means a block of size 1 repeated $2n-4$ times; see (5.10).

Since now the problem has been reduced to something much more tractable, we can even obtain an analytic expression for the optimal transition probabilities. The optimal solution for the $K_n - K_n$ graph (for $n \geq 2$) is given by

$$p_0^* = (\sqrt{2} - 1) \frac{n + \sqrt{2} - 2}{n + 2 - 2\sqrt{2}}, \quad p_1^* = \frac{2 - \sqrt{2}}{n + 2 - 2\sqrt{2}}, \quad p_2^* = \frac{n - \sqrt{2}}{(n-1)(n + 2 - 2\sqrt{2})}.$$

The corresponding optimal convergence rate is

$$\mu^* = \frac{n - 4 + 2\sqrt{2}}{n + 2 - 2\sqrt{2}}.$$

For large n , we have $\mu^* = 1 - \frac{6-4\sqrt{2}}{n} + O(\frac{1}{n^2})$. The limiting value of the optimal transition probability between the two clusters is $\sqrt{2} - 1 \approx 0.4142$. The optimal μ^* is quite close to the SLEM of a suboptimal construction with transition probabilities

$$(5.11) \quad p_0 = \frac{1}{2}, \quad p_1 = p_2 = \frac{1}{2(n-1)}.$$

As shown in [4], the corresponding SLEM of (5.11) is of the order $\mu = 1 - \frac{1}{3n} + O(\frac{1}{n^2})$. Here we compare $\frac{1}{3}$ with $6 - 4\sqrt{2} \approx 0.3431$.

6. Conclusions. We have shown that exploiting graph symmetry can lead to significant reduction in both the number of variables and the size of matrices, in solving the FMMC problem. For special classes of graphs such as edge-transitive and distance-transitive graphs, symmetry reduction leads to closed-form solutions in terms of the eigenvalues of the Laplacian matrix or the intersection matrix. For more general graphs, we gave two symmetry reduction methods, based on orbit theory and block-diagonalization, respectively.

The method based on orbit theory is very intuitive, but the construction of “good” orbit chains can be of more art than technique. The method of block-diagonalization can be mostly automated once the irreducible representations of the automorphism groups are generated (for small graphs, they can be generated using software for computational discrete algebra such as GAP [24]). These two approaches have an interesting connection: orbit theory gives nice interpretation of the diagonal blocks, while

the block-diagonalization approach offers theoretical insights about the construction of the orbit chains.

The symmetry reduction method developed in this paper can be very useful in many combinatorial optimization problems where the graph has rich symmetry properties, in particular, problems that can be formulated as or approximated by SDP or eigenvalue optimization problems involving weighted Laplacian matrices (see, e.g., [29, 44]). In addition to the reduction of problem size, other advantages of symmetry exploitation includes degeneracy removal, better conditioning, and reliability [26].

There is still much to do in understanding how to exploit symmetry in semidefinite programming. The techniques presented in this paper (and in [14]) require a good understanding of orbit theory, group representation theory, and interior-point methods for SDP. It is of practical importance to develop general purpose methods that can automatically detect symmetries (e.g., the code `nauty` [41] for graph automorphisms), and then exploit them in computations. A good model here is general purpose (but heuristic) methods for exploiting sparsity in numerical linear algebra, where symbolic operations on graphs (e.g., minimum degree permutation) reduce fill-ins in numerical factorization (see, e.g., [27]). As a result of this work, even very large sparse optimization problems are now routinely solved by users who are not experts in sparse matrix methods. For exploiting symmetry in SDPs, the challenges include the development of fast methods to detect large symmetry groups (for computational purposes, it often suffices to recognize parts of the symmetries) and the integration of algebraic methods (e.g., orbit theory and group representations) and numerical algorithms (e.g., interior-point methods).

REFERENCES

- [1] N. BIGGS, *Algebraic Graph Theory*, Cambridge University Press, Cambridge, UK, 1974.
- [2] S. BOYD AND L. VANDENBERGHE, *Convex Optimization*, Cambridge University Press, Cambridge, UK, 2004.
- [3] S. BOYD, P. DIACONIS, AND L. XIAO, *Fastest mixing Markov chain on a graph*, *SIAM Rev.*, 46 (2004), pp. 667–689.
- [4] S. BOYD, P. DIACONIS, P. A. PARRILO, AND L. XIAO, *Symmetry analysis of reversible Markov chains*, *Internet Math.*, 2 (2005), pp. 31–71.
- [5] S. BOYD, P. DIACONIS, P. A. PARRILO, AND L. XIAO, *Fastest Mixing Markov Chain on Graphs with Symmetries*, Tech. Report MSR-TR-2007-52, Microsoft Research, Redmond, VA, 2007.
- [6] S. BOYD, P. DIACONIS, J. SUN, AND L. XIAO, *Fastest mixing Markov chain on a path*, *Amer. Math. Monthly*, 113 (2006), pp. 70–74.
- [7] S. BOYD, A. GHOSH, B. PRABHAKAR, AND D. SHAH, *Randomized gossip algorithms*, *IEEE Trans. Inform. Theory*, 52 (2006), pp. 2508–2530.
- [8] P. BRÉMAUD, *Markov Chains, Gibbs Fields, Monte Carlo Simulation and Queues*, *Texts Appl. Math.* 31, Springer-Verlag, Berlin, Heidelberg, 1999.
- [9] A. E. BROUWER, A. M. COHEN, AND A. NEUMAIER, *Distance-Regular Graphs*, Springer-Verlag, Berlin, 1989.
- [10] S. BURER AND R. D. C. MONTEIRO, *A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization*, *Math. Program.*, 95 (2003), pp. 329–357.
- [11] F. CHEN, L. LOVÁSZ, AND I. PAK, *Lifting Markov chains to speed up mixing*, in *Proceedings of the 31st Annual ACM Symposium on Theory of Computing*, 1999, pp. 275–281.
- [12] F. R. K. CHUNG, *Spectral Graph Theory*, *CBMS Reg. Conf. Ser. Math.* 92, AMS, Providence, RI, 1997.
- [13] R. COGILL, S. LALL, AND P. A. PARRILO, *Structured semidefinite programs for the control of symmetric systems*, *Automatica*, 44 (2008), pp. 1411–1417.
- [14] E. DE KLERK, D. V. PASECHNIK, AND A. SCHRIJVER, *Reduction of symmetric semidefinite programs using the regular $*$ -representation*, *Math. Program.*, 109 (2007), pp. 613–624.
- [15] P. DIACONIS, *Group Representations in Probability and Statistics*, IMS, Hayward, CA, 1988.

- [16] P. DIACONIS AND L. SALOFF-COSTE, *Comparison theorems for reversible Markov chains*, Ann. Appl. Probab., 3 (1993), pp. 696–730.
- [17] P. DIACONIS AND L. SALOFF-COSTE, *Logarithmic Sobolev inequalities for finite Markov chains*, Ann. Appl. Probab., 6 (1996), pp. 695–750.
- [18] P. DIACONIS AND L. SALOFF-COSTE, *Separation cut-offs for birth and death chains*, Ann. Appl. Probab., 16 (2006), pp. 2098–2122.
- [19] P. DIACONIS AND D. STROOCK, *Geometric bounds for eigenvalues of Markov chains*, Ann. Appl. Probab., 1 (1991), pp. 36–61.
- [20] I. DUKANOVIC AND F. RENDL, *Semidefinite programming relaxations for graph coloring and maximal clique problems*, Math. Program., 109 (2007), pp. 345–365.
- [21] P. ERDŐS AND A. RÉNYI, *Asymmetric graphs*, Acta Math. Acad. Sci. Hungar., 14 (1963), pp. 295–315.
- [22] A. FÄSSLER AND E. STIEFEL, *Group Theoretical Methods and Their Applications*, Birkhäuser, Boston, 1992.
- [23] K. K. GADE AND M. L. OVERTON, *Optimizing the asymptotic convergence rate of the Diaconis-Holmes-Neal sampler*, Adv. in Appl. Math., 38 (2007), pp. 382–403.
- [24] THE GAP GROUP, *GAP: Groups, Algorithms, Programming. A System for Computational Discrete Algebra*, version 4.4.6, 2005. <http://www.gap-system.org>.
- [25] K. GATERMANN, *Computer Algebra Methods for Equivariant Dynamical Systems*, Lecture Notes in Math. 1728, Springer-Verlag, Berlin, 2000.
- [26] K. GATERMANN AND P. A. PARRILO, *Symmetry groups, semidefinite programs, and sums of squares*, J. Pure Appl. Algebra, 192 (2004), pp. 95–128.
- [27] A. GEORGE AND J. LIU, *Computer Solution of Large Sparse Positive Definite Systems*, Prentice-Hall, Englewood Cliffs, NJ, 1981.
- [28] C. GODSIL AND G. ROYLE, *Algebraic Graph Theory*, Grad. Texts in Math. 207, Springer-Verlag, New York, 2001.
- [29] M. X. GOEMANS, *Semidefinite programming in combinatorial optimization*, Math. Programming, 79 (1997), pp. 143–161.
- [30] G. H. GOLUB AND C. F. V. LOAN, *Matrix Computations*, 3rd ed., The Johns Hopkins University Press, Baltimore, MD, 1996.
- [31] M. GOLUBITSKY, I. STEWART, AND D. G. SCHAEFFER, *Singularities and Groups in Bifurcation Theory II*, Appl. Math. Sci. 69, Springer-Verlag, New York, 1988.
- [32] A. GRAHAM, *Kronecker Products and Matrix Calculus with Applications*, Ellis Horwoods Ltd., Chichester, UK, 1981.
- [33] L. GROSS, *Logarithmic Sobolev inequalities and contractivity properties of semigroups*, in Dirichlet Forms, Lecture Notes in Math. 1563, Springer-Verlag, Berlin, 1993, pp. 54–88.
- [34] B. HAN, M. L. OVERTON, AND T. P.-Y. YU, *Design of Hermite subdivision schemes aided by spectral radius optimization*, SIAM J. Sci. Comput., 25 (2003), pp. 643–656.
- [35] C. HELMBERG AND F. RENDL, *A spectral bundle method for semidefinite programming*, SIAM J. Optim., 10 (2000), pp. 673–696.
- [36] Y. KANNO, M. OHSAKI, K. MUROTA, AND N. KATOH, *Group symmetry in interior-point methods for semidefinite programming*, Optim. Eng., 2 (2001), pp. 293–320.
- [37] M. LAURENT, *Strengthened semidefinite programming bounds for codes*, Math. Program., 109 (2007), pp. 239–261.
- [38] Z. LU, A. NEMIROVSKI, AND R. D. C. MONTEIRO, *Large-scale semidefinite programming via saddle point mirror-prox algorithm*, Math. Program., 109 (2007), pp. 211–237.
- [39] F. MARGOT, *Exploiting orbits in symmetric ILP*, Math. Program., 98 (2003), pp. 3–21.
- [40] J. E. MARSDEN AND T. RATIU, *Introduction to Mechanics and Symmetry*, 2nd ed., Texts Appl. Math. 17, Springer-Verlag, New York, 1999.
- [41] B. MCKAY, *nauty User's guide (Version 2.2)*, Australian National University, 2003. Available online from <http://cs.anu.edu.au/~bdkm/nauty/>.
- [42] R. MERRIS, *Laplacian matrices of graphs: A survey*, Linear Algebra Appl., 197 (1994), pp. 143–176.
- [43] B. MOHAR, *Some applications of Laplace eigenvalues of graphs*, in Graph Symmetry: Algebraic Methods and Applications, G. Hahn and G. Sabidussi, eds., NATO Sci. Ser. C Math. Phys. Sci. 497, Kluwer Academic, Dordrecht, The Netherlands, 1997, pp. 225–275.
- [44] B. MOHAR AND S. POLJAK, *Eigenvalues in combinatorial optimization*, in Combinatorial and Graph-Theoretical Problems in Linear Algebra, R. A. Brualdi, S. Friedland, and V. Klee, eds., IMA Vol. Math. Appl. 50, Springer-Verlag, New York, 1993, pp. 107–151.
- [45] A. NEMIROVSKI, *Prox-method with rate of convergence $O(1/t)$ for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems*, SIAM J. Optim., 15 (2004), pp. 229–251.

- [46] Y. NESTEROV, *Smooth minimization of non-smooth functions*, Math. Program., 103 (2005), pp. 127–152.
- [47] M. L. OVERTON, *Large-scale optimization of eigenvalues*, SIAM J. Optim., 2 (1992), pp. 88–120.
- [48] P. A. PARRILO, *Structured Semidefinite Programs and Semialgebraic Geometry Methods in Robustness and Optimization*, Ph.D. thesis, California Institute of Technology, Pasadena, CA, 2000.
- [49] P. A. PARRILO, *Semidefinite programming relaxations for semialgebraic problems*, Math. Program., 96 (2003), pp. 293–320.
- [50] P. A. PARRILO AND B. STURMFELS, *Minimizing polynomial functions*, in Algorithmic and Quantitative Real Algebraic Geometry, S. Basu and L. Gonzalez-Vega, eds., DIMACS Ser. Discrete Math. Theoret. Comput. Sci. 60, AMS, Providence, RI, 2003, pp. 83–99.
- [51] S. ROCH, *Bounding fastest mixing*, Electron. Comm. Probab., 10 (2005), pp. 282–296.
- [52] Y. SAAD, *Numerical Methods for Large Eigenvalue Problems*, Manchester University Press, Manchester, UK, 1992.
- [53] J. SALTZMAN, *A Generalization of Spectral Analysis for Discrete Data Using Markov Chains*, Ph.D. thesis, Department of Statistics, Stanford University, Stanford, CA, 2006.
- [54] J.-P. SERRE, *Linear Representations of Finite Groups*, Springer-Verlag, New York, 1977.
- [55] D. STROOCK, *Logarithmic Sobolev inequalities for Gibbs states*, in Dirichlet Forms, Lecture Notes in Math. 1563, Springer-Verlag, Berlin, 1993, pp. 194–228.
- [56] R. SUBRAMANIAN AND I. D. SCHERSON, *An analysis of diffusive load-balancing*, in Proceedings of the 6th Annual ACM Symposium on Parallel Algorithms and Architectures, 1994, pp. 220–225.
- [57] P. WORFOLK, *Zeros of equivariant vector fields: Algorithms for an invariant approach*, J. Symbolic Comput., 17 (1994), pp. 487–511.
- [58] L. XIAO AND S. BOYD, *Fast linear iterations for distributed averaging*, Systems Control Lett., 53 (2004), pp. 65–78.
- [59] L. XIAO, S. BOYD, AND S.-J. KIM, *Distributed average consensus with least-mean-square deviation*, J. Parallel Distrib. Comput., 67 (2007), pp. 33–46.