A DECOMPOSITION BASED PROOF FOR FAST MIXING OF A MARKOV CHAIN OVER BALANCED REALIZATIONS OF A JOINT DEGREE MATRIX

PÉTER L. ERDŐS†, ISTVÁN MIKLÓS¶∥∗∗, AND ZOLTÁN TOROCZKAI‡§††

Abstract. A joint degree matrix (JDM) specifies the number of connections between nodes of given degrees in a graph, for all degree pairs and uniquely determines the degree sequence of the graph. We consider the space of all balanced realizations of an arbitrary JDM, realizations in which the links between any two degree groups are placed as uniformly as possible. We prove that a swap Markov Chain Monte Carlo (MCMC) algorithm in the space of all balanced realizations of an arbitrary graphical JDM mixes rapidly, i.e., the relaxation time of the chain is bounded from above by a polynomial in the number of nodes \( n \). To prove fast mixing, we first prove a general factorization theorem similar to the Martin-Randall method for disjoint decompositions (partitions). This theorem can be used to bound from below the spectral gap with the help of fast mixing subchains within every partition and a bound on an auxiliary Markov chain between the partitions. Our proof of the general factorization theorem is direct and uses conductance based methods (Cheeger inequality).

Key words. graph sampling, joint degree matrix, rapidly mixing Markov Chains

AMS subject classifications. 05C30, 05C81, 68R10

1. Introduction. The sampling of simple graphs on fixed number of nodes and with given degree sequence is a well studied problem both by the statistics community (binary contingency tables [8], [3]) and the computer science community. Lately the interest in this sampling problem has been widening with applications ranging from social sciences, physics, biology to engineering. The sampling approaches can be classified roughly into two types, one using Markov Chain Monte Carlo (MCMC) algorithms [21] based on simple moves such as edge swaps and the other using direct construction methods [7, 17] and importance sampling [7, 14].

The MCMC approach, while conceptually simple, presents a notoriously difficult question, namely proving (or disproving) the fast mixing nature of a proposed Markov chain. A Markov chain mixes fast if its mixing time (expressed via the inverse of the spectral gap; for a summary of other measures see [12]) has a polynomial upper bound in the size of the problem, usually taken as the number of nodes \( n \). In a seminal paper Kannan, Tetali and Vempala [16] conjectured that the natural MCMC algorithm based on edge swaps mixes rapidly for arbitrary graphical degree sequences and they proved it for regular bipartite graphs. In 2007 Copper, Dyer and Greenhill [5] proved fast mixing for regular degree sequences of undirected graphs and subsequently Greenhill gave a proof for regular directed graphs [13]. Recently, Miklós, Erdős and Soukup [20] proved fast mixing for half-regular bipartite graphs (nodes have to have
the same degree on only one side of the bipartition), using a modified version of Sinclair's multicommodity flow method \cite{22}. In spite of these results on particular degree sequences for simple graphs, the proof for the general case, however, remains elusive. Recall that in a simple graph no multiple edges exist between any pairs of nodes, and there are no self-loops either.

To gain a better understanding of the mixing problem for arbitrary degree sequences of simple graphs we may follow a different approach. In particular, we may consider the sampling problem on a class of graphs obeying stronger constraints than the degree sequence, however, constraints that can be used to implicitly define an arbitrary degree sequence.

One such possibility is the joint degree matrix problem first introduced by Patrinos and Hakimi \cite{1} and subsequently studied by Amanatidis, Green and Mihail \cite{2}, Stanton and Pinar \cite{23} and then by Czabarka et al \cite{6}. A joint degree matrix (JDM) fixes the number of edges between nodes of given degrees, for all degree pairs. This constraint is stronger than just fixing the degree sequence, however, it uniquely defines the degree sequence (see the next section). Thus, we may study the MCMC sampling problem and the associated mixing time question on the space of all the graphical realizations of a given JDM. This means that we study the degree based sampling problem on a subspace of it, restricted by the degree-degree correlations imposed through the JDM. In a different context, Stanton and Pinar \cite{23} propose a restricted swap operation based Markov chain over the space of realizations of a JDM, which was recently shown to be irreducible by Czabarka et.al \cite{6}, a necessary condition of a MCMC sampling algorithm. Note that the JDM problem is interesting in its own right, with applications in social sciences (the degree assortativity problem) and data driven modeling of real-world networks \cite{23}.

However, proving that a suitable MCMC algorithm is fast mixing over the space of all realizations of a JDM is also difficult. Here, instead, we prove fast mixing on a subspace of realizations of an arbitrary JDM, namely the space of balanced realizations. As shown in \cite{6} and \cite{23}, all graphical JDMs admit balanced realizations, in which the partial degrees of nodes from a given degree class towards another degree class are as uniformly distributed as possible. To prove fast mixing, we first introduce a disjoint partition of the space of balanced realizations. The elements in each partition can be expressed as the union of almost regular graphs and almost semi-regular bipartite graphs (see Def 3.3), while the structure of the partitioning can be also described through the collection of half-regular bipartite graphs (see Section 3 for definitions). We then develop a modified version of the factorization theorem (\cite{19}, Thm 3.2) in the disjoint decomposition method first introduced by Martin and Randall \cite{19} to provide a lower bound for the spectral gap that is more readily computable. The factorization theorem of Martin and Randall for disjoint decompositions is based on a result by Caracciolo, Pelissetto and Sokal originally introduced in the framework of simulated tempering \cite{4}. The case of non-disjoint, large-overlap decompositions has been treated earlier by Madras and Randall \cite{18}. The Martin-Randall method for disjoint decompositions introduces a projection Markov chain between the partitions defined with the help of average probabilities. Unfortunately it is difficult to estimate/bound the spectral gap in this chain, and they resort to a lower bound defined via a Metropolis-Hastings chain \cite{18}. Our main Theorem 4.3 also works with a lower bound, however, it is more general and we provide a direct and short proof for it using conductance based arguments.

The paper is organized as follows: after preliminaries related to JDMs we present
the main theorem (Thm 2.3) for fast mixing and our proof strategy. We then describe in Section 3 the structure of the space of balanced realizations of a JDM and present a disjoint partitioning of this space into subspaces whose elements are formed by almost regular and almost semi-regular subgraphs. In Section 4 using conductance based arguments, we prove our general theorem for fast mixing over state spaces that can be disjointly partitioned as described in Section 3. In Section 5 we briefly recall earlier results on fast mixing over realizations of regular degree sequences and half-regular bipartite degree sequences, then extend their proofs (in the Appendix) to almost regular and almost half-regular cases. We then introduce our Markov chain over the space of balanced realizations, and complete the proof of our main theorem.

2. Preliminaries. A symmetric matrix $J$ with non-negative integer elements is the joint degree matrix (JDM) of an undirected simple graph $G$ iff the element $J_{i,j}$ gives the number of edges between the class $V_i$ of vertices all having degree $i > 0$ and the class $V_j$ of vertices all with degree $j > 0$ in the graph. In this case we also say that $J$ is graphical and that $G$ is a graphical realization of $J$. Note that there can be many different graphical realizations of the same JDM.

Given a JDM, the number of vertices $n_i = |V_i|$ in class $i$ is obtained from:

$$n_i = J_{i,i} + \sum_{j=1}^{k} J_{i,j},$$

where $k$ denotes the number of distinct degree classes. This implies that a JDM also uniquely determines the degree sequence, since we have obtained the number of nodes of given degrees for all possible degrees. Clearly $k \leq \Delta$, where $\Delta$ is the maximum degree. A necessary condition for $J$ to be graphical is that all the $n_i$-s are integers. Let $n$ denote the total number of vertices. Naturally, $n = \sum_i n_i$ and it is uniquely determined via Eq (2.1) for a given graphical JDM. The necessary and sufficient conditions for a given JDM to be graphical are provided, for e.g., in Ref [6].

Let $d_j(v)$ denote the number of edges of a node $v$ incident on class $V_j$. The vector consisting of the $d_j(v)$-s for all $j$ is called the degree spectrum of vertex $v$. A realization of the JDM is balanced iff for every $i$ and all $v \in V_i$ and all $j$, we have

$$\left| d_j(v) - \frac{J_{i,j}}{n_i} \right| < 1.$$

The following theorem is proven in Ref [6] as Corollary 5:

**Theorem 2.1.** Every graphical JDM admits a balanced realization.

A restricted swap operation (RSO) takes two edges $(x,y)$ and $(u,v)$ with $x$ and $u$ from the same vertex class and swaps them with two non-edges $(x,v)$ and $(u,y)$. The RSO preserves the JDM, and in fact forms an irreducible Markov chain over all its realizations [6]. The RSO Markov chain on balanced realizations can thus be defined as follows:

**Definition 2.2.** Let $J$ be a JDM. The state space of the RSO Markov chain on balanced realizations consists of all the balanced realizations of $J$. The transitions of the Markov chain are defined in the following way. With probability $1/2$, the chain remains in the current state (namely, the Markov chain is lazy), and with probability $1/2$, 4 vertices, $v_1, v_2, v_3, v_4$ are chosen from the vertex set of the current realization (the order of the vertices is considered). A restricted swap operation

$$E \setminus \{(v_1, v_3), (v_2, v_4)\} \cup \{(v_1, v_4), (v_2, v_3)\}$$
is performed if such swap exists, and furthermore, if it leads to another balanced JDM realization, otherwise the Markov chain remains in the same state. Note that exactly two different orders of the selected vertices will provide the same swap operation, since the roles of \( v_1 \) and \( v_2 \) are symmetric. Then there is a transition with probability

\[
\frac{1}{n(n-1)(n-2)(n-3)}
\]

between two realizations iff there is a RSO transforming one into the other.

In this paper, we prove that such a Markov chain is rapidly mixing. The convergence of a Markov chain is measured as a function of the input data size. Here we note that the size of the data is the number of vertices (or number of edges, they are polynomially bounded functions of each other) and not the number of digits to describe the JDM. This distinction is important as, for example, one can create a \( 2 \times 2 \) JDM with values \( J_{2,2} = J_{3,3} = 0 \) and \( J_{2,3} = J_{3,2} = 6n \), which has \( \Omega(n) \) number of vertices (or edges) but it needs only \( O(\log(n)) \) number of digits to describe (except in the unary number system). Alternatively, one might consider the input is given in unary.

Formally, we state the rapid mixing property via the following theorem:

**Theorem 2.3.** The RSO Markov chain on balanced JDM realizations is a rapidly mixing Markov chain, namely, for the second largest eigenvalue \( \lambda_2 \) of this chain, it holds that

\[
\frac{1}{1 - \lambda_2} = O(\text{poly}(n))
\]

where \( n \) is the number of vertices in the realizations of the JDM.

The proof is based on the special structure of the state space of the balanced JDM realizations. This special structure allows the following proof strategy: if we can prove that some auxiliary Markov chains are rapidly mixing on some sub-spaces obtained from decomposing the above-mentioned specially structured state space, then the Markov chain on the whole space is also rapidly mixing. We are going to prove the rapid mixing of these auxiliary Markov chains, as well as give the proof of the general theorem, that a Markov chain on this special structure is rapidly mixing, hence proving our main Theorem 2.3.

3. The structure of the space of balanced JDM realizations, and the Markov chain over this space. In order to describe the structure of the space of balanced JDM realizations, we first define the almost semi-regular bipartite and almost regular graphs.

**Definition 3.1.** A bipartite graph \( G(U, V, E) \) is almost semi-regular if for any \( u_1, u_2 \in U \) and \( v_1, v_2 \in V \)

\[
|d(u_1) - d(u_2)| \leq 1
\]

and

\[
|d(v_1) - d(v_2)| \leq 1
\]

**Definition 3.2.** A graph \( G(V, E) \) is almost regular, if for any \( v_1, v_2 \in V \)

\[
|d(v_1) - d(v_2)| \leq 1
\]
Note that a balanced JDM realization is the union of almost semi-regular bipartite and almost regular graphs. The restriction of any graphical realization of the JDM to two vertex classes is the union of an almost semi-regular bipartite graph and two almost regular graphs. The collection of these almost semi-regular bipartite and almost regular graphs completely describe the balanced JDM realizations. Formally:

**Definition 3.3 (Labeled union).** Any balanced JDM realization can be represented as a set of almost semi-regular bipartite graphs and almost regular graphs. The realization can be constructed from these factor graphs as their labeled union: the vertices with the same labels are collapsed, and the edge set of the union is the union of the edge sets of the factor graphs.

It is useful to construct the following auxiliary graphs. For each vertex class $V_i$, we create an auxiliary bipartite graph, $G_i(V_i, U; E)$, where $U$ is a set of “super-nodes” representing all vertex classes $V_j$, including $V_i$. There is an edge between $v \in V_i$ and super-node $u_j$ representing vertex class $V_j$ iff

$$d_j(v) = \left\lceil \frac{J_{ij}}{n_i} \right\rceil,$$

i.e., iff node $v$ carries the ceiling of the average degree of its class $i$ towards the other class $j$. We claim that these $k$ auxiliary graphs are half-regular, i.e., each vertex in $V_i$ has the same degree (the degrees in the vertex class $U$ might be arbitrary). Indeed, the vertices in $V_i$ all have the same degree in the JDM realization, therefore, the number of times they have the ceiling of the average degree towards a vertex class is constant in a balanced realization.

Let $Y$ denote the space of all balanced realizations of a JDM and just as before, let $k$ denote the number of distinct vertex classes. We will represent the elements of $Y$ via the $[k(k + 1)/2]$-component vector $y$ whose elements are the $k(k - 1)/2$ almost semi-regular bipartite graphs between the vertex classes and the $k$ almost regular graphs for each vertex class. Given an element $y \in Y$ it has $k$ associated auxiliary graphs, one for every class. We will represent this collection of auxiliary graphs via the $k$-dimensional vector $x$ such that its $i$-th coordinate contains the auxiliary bipartite graph for vertex class $V_i$. For any given $y$ we can determine the corresponding $x$, however, for a given $x$ there can be several $y$-s with the same $x$. We will denote by $Y_x$ the subset of $Y$ containing all the $y$-s with the same $x$ and by $X$ the set of all
possible induced $x$ vectors. Clearly, the $x$ vectors can be used to define a disjoint partition on $Y$: $Y = \bigcup_{x \in X} Y_x$. For notation convenience we will consider the space $Y$ as pairs $(x, y)$. A RSO might fix $x$ in which case it will make a move only within $Y_x$, but if it does not fix $x$ it can change both $x$ and $y$. For any $x$, the RSOs moving only within $Y_x$ form a Markov chain. On the other hand, tracing only the $x$s from the pairs $(x, y)$ is not a Markov chain: the probability that an RSO changes $x$ (and thus also $y$) depends on the current $y$ not only on $x$. However, the following theorem holds:

**Theorem 3.4.** Let $(x_1, y_1)$ be a balanced realization of a JDM in the above mentioned representation. Let $x_2$ be a vector differing only in one coordinate from $x_1$, and furthermore, only in one swap within this coordinate, namely, one swap within one coordinate is sufficient to transform $x_1$ into $x_2$. Then there exists at least one $y_2$ such that $(x_2, y_2)$ is a balanced JDM realization and $(x_1, y_1)$ can be transformed into $(x_2, y_2)$ with a single RSO.

**Proof.** Let $v_1, v_2 \in V_i$ and let the swap transforming $x_1$ into $x_2$ remove edges $(v_1, u_j)$ and $(v_2, u_k)$ (with $j \neq k$) and add edges $(v_1, u_k)$ and $(v_2, u_j)$. Since the edge $(v_1, u_j)$ exists in $x_1$ and $(v_2, u_j)$ does not exist in $x_1$, $d_j(v_1) > d_j(v_2)$. This means that there is at least one vertex $w \in V_j$ such that $w$ is connected to $v_1$ but not to $v_2$. Similarly, there is at least one vertex $r \in V_k$ such that $r$ is connected to $v_2$ but not to $v_1$. Therefore, we have a required RSO on nodes $v_1, v_2, w, r$. 

4. Proving rapid mixing of Markov chains over factorized state spaces.

In this section we will prove a general factorization theorem (Thm 4.3) on fast convergence of a Markov chain. The proof of this theorem will lead to the proof of our main result, Theorem 2.3. Our theorem is similar to a theorem by Martin-Randall and the comparison between the theorems is given at the end of this section. In the following we will need the Cheeger inequality and a slight modification of it.

Let $M$ be a discrete time, discrete space, reversible Markov chain over set $I$ with stationary distribution $\pi$ and transition probabilities $T(b|a)$, where $a, b \in I$. The probability of a subset is denoted by

$$\pi(S) := \sum_{a \in S} \pi(a)$$

The conditional flow out of a subset of the state space $S \subset I$ is defined by

$$\Psi(S) := \sum_{a \in S, b \in \bar{S}} \frac{\pi(a)T(b|a)}{\pi(S)},$$

(4.1)

where $\bar{S}$ denotes the complementary set of $S$ in $I$. The conductance of the state space is defined as

$$\Phi := \min_{S \subset I} \left\{ \Psi(S) \left| 0 < \pi(S) \leq \frac{1}{2} \right. \right\} .$$

The Cheeger inequality quoted in Theorem 4.1 gives lower and upper bounds on the second largest eigenvalue of the Markov chain, see for example Ref [10] for a proof.

**Theorem 4.1 (Cheeger inequality).**

$$1 - 2\Phi \leq \lambda_2 \leq 1 - \frac{\Phi^2}{2} .$$
Here we prove a variant of the left Cheeger inequality (the lower-bound).

**Theorem 4.2.** For any reversible Markov chain, and any subset $S$ of its state space,

$$\frac{1 - \lambda_2}{2} \min\{\pi(S), \pi(\bar{S})\} \leq \sum_{a \in S, b \in \bar{S}} \pi(a)T(b|a).$$  

(4.2)

**Proof.** We distinguish two cases, either $\pi(S) \leq 1/2$ or $\pi(S) > 1/2$. If $\pi(S) \leq 1/2$ then the inequality is read as

$$\frac{1 - \lambda_2}{2} \pi(S) \leq \sum_{a \in S, b \in \bar{S}} \pi(a)T(b|a).$$

After rearranging, we obtain that

$$\frac{1 - \lambda_2}{2} \leq \frac{\sum_{a \in S, b \in \bar{S}} \pi(a)T(b|a)}{\pi(S)}.$$

But this is merely the Cheeger inequality since

$$\frac{1 - \lambda_2}{2} \leq \Phi \leq \frac{\sum_{a \in S, b \in \bar{S}} \pi(a)T(b|a)}{\pi(S)}.$$

Here the second inequality holds due to the definition of $\Phi$.

If $\pi(S) > 1/2$, then the inequality is

$$\frac{1 - \lambda_2}{2} \pi(\bar{S}) \leq \sum_{a \in S, b \in \bar{S}} \pi(a)T(b|a),$$

and then using the reversibility property, we obtain

$$\frac{1 - \lambda_2}{2} \leq \Phi \leq \frac{\sum_{a \in \bar{S}, b \in S} \pi(b)T(a|b)}{\pi(S)} = \Psi(\bar{S}).$$

But if $\pi(S) > 1/2$ then $\pi(\bar{S}) < 1/2$, and the inequality holds just like in the previous case.

Now we are ready to state and prove a general theorem on rapidly mixing Markov chains.

**Theorem 4.3.** Let $\mathcal{M}$ be a class of irreducible, aperiodic, reversible Markov chains whose state space $Y$ can be partitioned into disjoint classes $Y = \cup_{x \in X} Y_x$ by the elements of some set $X$. For notation convenience we also denote the element $y \in Y_x$ via the pair $(x, y)$ to also indicate the partition it belongs to. The problem size of a particular chain is denoted by $n$. Let $T$ be the transition matrix of $M \in \mathcal{M}$, and let $\pi$ denote the globally stable stationary distribution of $M$. Let $\pi_X$ denote the marginal of $\pi$ on the first coordinate, and for any $x$, let $\pi_{Y_x}$ denote the restriction of the distribution for fixed $x$. Assume that the following properties hold:
(i) For all \( x \), the transitions with \( x \) fixed form an aperiodic, irreducible and reversible Markov chain denoted by \( M_x \) with stationary distribution \( \pi_{Y_x} \). Furthermore, this Markov chain is rapidly mixing, i.e., for its second largest eigenvalue \( \lambda_{M_x, 2} \) it holds that
\[
\frac{1}{1 - \lambda_{M_x, 2}} \leq \text{poly}_1(n).
\]

(ii) There exists a Markov chain \( M' \) with transition matrix \( T' \) which is aperiodic, irreducible and reversible w.r.t. \( \pi_X \), and for all \( x_1, y_1, x_2 \) it holds that
\[
\sum_{y_2 \in Y_{x_2}} T((x_2, y_2)|(x_1, y_1)) \geq T'(x_2|x_1).
\]
Furthermore, this Markov chain is rapidly mixing, namely, for its second largest eigenvalue \( \lambda_{M', 2} \) it holds that
\[
\frac{1}{1 - \lambda_{M', 2}} \leq \text{poly}_2(n).
\]

Then \( M \) is also rapidly mixing as its second largest eigenvalue obeys:
\[
\frac{1}{1 - \lambda_{M, 2}} = O(\text{poly}(n)).
\]

Proof. For any subset \( S \) of the state space \( Y = \bigcup_x Y_x \) of \( M \) we define
\[
X(S) := \{ x \in X \mid \exists y, (x, y) \in S \}
\]
and for any given \( x \in X \) we have
\[
Y_x(S) := \{ (x, y) \in Y \mid (x, y) \in S \}.
\]

We are going to prove that the conditional flow \( \Psi(S) \) (see equation (4.1)) from any \( S \subset Y \) with \( \pi(S) \leq 1/2 \) cannot be too small and therefore, neither the conductance of the Markov chain. We cut the state space into two parts \( Y = Y^l \cup Y^u \), namely the lower and upper parts using the following definitions (see also Fig. 4.1): the partition \( X = L \cup U \) is defined as
\[
L := \left\{ x \in X \left| \frac{\pi(Y_x(S))}{\pi(Y_x)} \leq \frac{1}{\sqrt{2}} \right. \right\},
\]
\[
U := \left\{ x \in X \left| \frac{\pi(Y_x(S))}{\pi(Y_x)} > \frac{1}{\sqrt{2}} \right. \right\}.
\]

Furthermore, we introduce:
\[
Y^l := \bigcup_{x \in L} Y_x \quad \text{and} \quad Y^u := \bigcup_{x \in U} Y_x,
\]
and finally let
\[
S_l := S \cap Y^l \quad \text{and} \quad S_u := S \cap Y^u.
\]
Since $M'$ is rapidly mixing we can write (based on Theorem 4.1):

$$1 - 2\Phi_{M'} \leq \lambda_{M',2} \leq 1 - \frac{1}{\text{poly}_2(n)},$$

or

$$\Phi_{M'} \geq \frac{1}{2\text{poly}_2(n)}.$$

We use this lower bound of conductance to define two cases regarding the lower and upper part of $S$.

**Case 1** We say that the lower part $S_l$ is not a negligible part of $S$ when

$$\frac{\pi(S_l)}{\pi(S_u)} \geq \frac{1}{4\sqrt{2}\text{poly}_2(n)} \left(1 - \frac{1}{\sqrt{2}}\right).$$

(4.4)

**Case 2** We say that the lower part $S_l$ is a negligible part of $S$ when

$$\frac{\pi(S_l)}{\pi(S_u)} < \frac{1}{4\sqrt{2}\text{poly}_2(n)} \left(1 - \frac{1}{\sqrt{2}}\right).$$

(4.5)

Our plan is the following: the conditional flow $\Psi(S)$ is positive on any subset and it obeys: $\Psi(S) = \Psi(S_l)/\pi(S_l) + \Psi(S_u)/\pi(S_u)$. It cannot be too small, if at least one of $\Psi(S_l)$ or $\Psi(S_u)$ is big enough (and the associate fraction $\pi(S_l)/\pi(S)$). In Case 1 we will show that $\Psi(S_l)$ itself is big enough. To that end it will be sufficient to consider the part which leaves $S_l$ but not $Y^l$ (see also Fig. 4.2). For Case 2 we will consider $\Psi(S_u)$, particularly that part of it which goes from $S_u$ to $Y^l \setminus S_l$ (see also Fig. 4.3).

In Case 1, the flow going out from $S_l$ within $Y^l$ is sufficient to prove that the conditional flow going out from $S$ is not negligible. We know that for any particular $x$, we have a rapidly mixing Markov chain $M_x$ over the second coordinate $y$. Let their smallest conductance be denoted by $\Phi_x$. Since all these Markov chains are rapidly mixing, we have that

$$\min_x \lambda_{M_x,2} \leq 1 - \frac{1}{\text{poly}_4(n)}$$

**Fig. 4.1. The structure of $Y = Y^l \cup Y^u$.** A non-filled ellipse (with a simple line boundary) represents the space $Y_x$ for a given $x$. The solid black ellipses represent the set $S$ with some of them (the $S_l$) belonging to the lower part $Y^l$, and the rest (the $S_u$) belonging to the upper part ($Y^u$).
Fig. 4.2. When $S_l$ is not a negligible part of $S$, there is a considerable flow going out from $S_l$ to within $Y^l$, implying that the conditional flow going out from $S$ cannot be small. See text for details and rigorous calculations.

or, equivalently:

$$\frac{1}{2\text{poly}_1(n)} \leq \Phi_x.$$  

However, in the lower part, for any particular $x$ one has:

$$\pi_{Y_x}(Y_x(S)) = \frac{\pi(Y_x(S))}{\pi(Y_x)} \leq \frac{1}{\sqrt{2}}$$

so for any fixed $x$, belonging to $S_l$ it holds that

$$\frac{1}{2\text{poly}_1(n)} \min \left\{ \pi_{Y_x}(Y_x(S)), \left(1 - \frac{1}{\sqrt{2}}\right) \right\} \leq \sum_{(x,y) \in S, (x,y') \in \bar{S}} \pi_{Y_x}((x,y)) T((x,y')|(x,y))$$

using the modified Cheeger inequality (Theorem 4.2). Observing that

$$\pi_{Y_x}((x,y)) = \frac{\pi((x,y))}{\pi(Y_x)},$$

we obtain:

$$\frac{1}{2\text{poly}_1(n)} \min \left\{ \pi(Y_x(S)), \pi(Y_x) \left(1 - \frac{1}{\sqrt{2}}\right) \right\} \leq \frac{1}{2\text{poly}_1(n)} \pi(Y_x(S)) \left(1 - \frac{1}{\sqrt{2}}\right) \leq \sum_{(x,y) \in S, (x,y') \in \bar{S}} \pi((x,y)) T((x,y')|(x,y)).$$

Summing this for all the $x$s belonging to $S_l$, we deduce that

$$\pi(S_l) \frac{1}{2\text{poly}_1(n)} \left(1 - \frac{1}{\sqrt{2}}\right) \leq \sum_{x | Y_x(S) \subseteq S_l} \sum_{(x,y) \in S, (x,y') \in \bar{S}} \pi((x,y)) T((x,y')|(x,y)).$$
Note that this flow is not only going out from $S_l$ but also from the entire $S$. Therefore, we have that

$$\Psi(S) \geq \frac{\pi(S)}{\pi(S)} \times \frac{1}{2\text{poly}_1(n)} \left(1 - \frac{1}{\sqrt{2}}\right).$$

Either $\pi(S_l) \leq \pi(S_u)$, which then yields

$$\frac{\pi(S_l)}{\pi(S)} = \frac{\pi(S_l)}{\pi(S_l) + \pi(S_u)} \geq \frac{\pi(S_l)}{2\pi(S_u)} \geq \frac{1}{8\sqrt{2}\text{poly}_2(n)} \left(1 - \frac{1}{\sqrt{2}}\right)$$

after using Equation 4.4, or $\pi(S_l) > \pi(S_u)$, in which case we have

$$\frac{\pi(S_l)}{\pi(S)} > \frac{1}{2} \geq \frac{1}{8\sqrt{2}\text{poly}_2(n)} \left(1 - \frac{1}{\sqrt{2}}\right).$$

Thus in both cases the following inequality holds:

$$\Psi(S) \geq \frac{1}{8\sqrt{2}\text{poly}_2(n)} \left(1 - \frac{1}{\sqrt{2}}\right) \times \frac{1}{2\text{poly}_1(n)} \left(1 - \frac{1}{\sqrt{2}}\right).$$

In Case 2, the lower part of $S$ is a negligible part of $S$. In this case, we have that

$$\pi_X(X(S_u)) \leq \frac{1}{\sqrt{2}}$$

since $\pi(S)$ should be smaller than $1/2$.

Hence in the Markov chain $M'$, based on the modified Cheeger inequality, we obtain for $X(S_u)$ that

$$\frac{1}{2\text{poly}_2(n)} \min \left\{ \pi_X(X(S_u)), \left(1 - \frac{1}{\sqrt{2}}\right) \right\} \leq \sum_{x' \in X(S_u), x \in X(\delta_u)} \pi_X(x)T'(x'|x).$$

Fig. 4.3. When $S_l$ is a negligible part of $S$, there is a considerable flow going out from $S_u$ into $Y^l \setminus S_l$. See text for details and rigorous calculations.
For all \( y \) for which \((x, y) \in S_u\) we can write:

\[
T'(x'|x) \leq \sum_{y'} T((x', y')|(x, y)).
\]

Multiplying this with \( \pi((x, y)) \) then summing for all suitable \( y \):

\[
\pi(Y_x(S)) T'(x'|x) \leq \sum_{y'|(x, y) \in S_u} \sum_{y'} \pi(x, y) T((x', y')|(x, y))
\]

and thus

\[
T'(x'|x) \leq \frac{\sum_{y|(x, y) \in S_u} \sum_{y'} \pi(x, y) T((x', y')|(x, y))}{\pi(Y_x(S))}.
\]

Inserting this into Equation 4.6, we find that

\[
\frac{1}{2\text{poly}_2(n)} \min \left\{ \frac{\pi_X(X(S_u))}{\pi(Y_x(S))}, \left(1 - \frac{1}{\sqrt{2}}\right) \right\} \leq \sum_{x \in X(S_u), x' \in X(S_u)} \frac{\pi_X(x)}{\pi(Y_x(S))} \sum_{y|(x, y) \in S_u} \sum_{y'} \pi(x, y) T((x', y')|(x, y)).
\]

We have \( \pi_X(x) := \pi(Y_x) \), and thus \( \frac{\pi_X(x)}{\pi(Y_x(S))} \leq \sqrt{2} \) for all \( x \in X(S_u) \). Therefore we can write that

\[
\frac{1}{2\text{poly}_2(n)} \min \left\{ \frac{\pi_X(X(S_u))}{\pi(Y_x(S))}, \left(1 - \frac{1}{\sqrt{2}}\right) \right\} \leq \sqrt{2} \sum_{(x, y) \in S_u} \left( \sum_{(x', y')|x' \in X(S_u)} \pi(x, y) T((x', y')|(x, y)) \right).
\]

Note that \( \pi(S_u) \leq \pi_X(X(S_u)) < 1 \), and thus we have

\[
\frac{1}{2\sqrt{2}\text{poly}_2(n)} \pi(S_u) \left(1 - \frac{1}{\sqrt{2}}\right) \leq \sqrt{2} \sum_{(x, y) \in S_u} \left( \sum_{(x', y')|x' \in X(S_u)} \pi(x, y) T((x', y')|(x, y)) \right).
\]

This flow is going out from \( S_u \), and it has such a magnitude that at most half of it can go to the lower part of \( S \) (due to reversibility and due to Equation 4.5), so at least the half will go out of \( S \). Therefore:

\[
\Phi(S) \geq \frac{\pi(S_u)}{\pi(S)} \times \frac{1}{4\sqrt{2}\text{poly}_2(n)} \left(1 - \frac{1}{\sqrt{2}}\right).
\]

However, since \( S_u \) dominates \( S \), namely, \( \pi(S_u) > \frac{\pi(S)}{2} \), we have that

\[
\Phi(S) \geq \frac{1}{8\sqrt{2}\text{poly}_2(n)} \left(1 - \frac{1}{\sqrt{2}}\right).
\]
Namely, for all $S$ satisfying $0 < \pi(S) \leq \frac{1}{2}$, we can write:

$$\Psi(S) \geq \frac{1}{16\sqrt{2}\text{poly}_2(n)\text{poly}_1(n)} \left(1 - \frac{1}{\sqrt{2}}\right)^2.$$ 

And thus, for the conductance of the Markov chain $M$

$$\Phi_M \geq \frac{1}{16\sqrt{2}\text{poly}_2(n)\text{poly}_1(n)} \left(1 - \frac{1}{\sqrt{2}}\right)^2.$$ 

Applying this for the Cheeger inequality, one obtains

$$\lambda_{M,2} \leq 1 - \left(\frac{1}{16\sqrt{2}\text{poly}_2(n)\text{poly}_1(n)} \left(1 - \frac{1}{\sqrt{2}}\right)^2\right)^2$$

and thus

$$\frac{1}{1 - \lambda_{M,2}} \leq \frac{256\text{poly}_1^2(n)\text{poly}_2^1(n)}{\left(1 - \frac{1}{\sqrt{2}}\right)^4},$$

which is what we wanted to prove. 

Martin and Randall [19] have developed a similar theorem. They assume a disjoint decomposition of the state space $\Omega$ of an irreducible and reversible Markov chain defined via the transition probabilities $P(y|x)$. They require that the Markov chain be rapidly mixing when restricted onto each partition $\Omega_i$ ($\Omega = \bigcup \Omega_i$) and furthermore, another Markov chain, the so-called projection Markov chain $\tilde{P}(i|j)$ defined over the indices of the partitions be also rapidly mixing. If all these hold, then the original Markov chain is also rapidly mixing. For the projection Markov chain they use the normalized conditional flow

$$\tilde{P}(i|j) = \frac{1}{\pi(\Omega_i)} \sum_{x \in \Omega_i, y \in \Omega_j} \pi(x) P(y|x)$$

as transition probabilities. This can be interpreted as a weighted average transition probability between two partitions, while in our case, Equation (4.3) requires only that the transition probability of the lower bounding Markov chain is not more than the minimum of the sum of the transition probabilities going out from one member of the partition (subset $Y_{x_1}$) to the other other member of the partition (subset $Y_{x_2}$) with the minimum taken over all the elements of $Y_{x_1}$. Obviously, it is a stronger condition that our Markov chain must be rapidly mixing, since a Markov chain is mixing slower when each transition probability between any two states is smaller. (The latter statement is based on a comparison theorem by Diaconis and Saloff-Coste [9].) Therefore, from that point of view, our theorem is weaker. On the other hand, the average transition probability (4) is usually hard to calculate, and in this sense our theorem is more applicable. Note that Martin and Randall have also resorted in the end to using chain comparison techniques (Sections 2.2 and 3 in their paper) employing a Metropolis-Hastings chain as a lower bounding chain instead of the projection chain above. Our theorem, however, provides a direct proof to a similar statement.
5. The RSO Markov chain on balanced realizations. In this section, via a series of theorems we prove that for the RSO Markov chain on the balanced realizations and for an appropriate auxiliary $M'$ Markov chain, all the conditions in Theorem 4.3 are fulfilled and thus the RSO Markov chain is also rapidly mixing on the balanced JDM realizations.

**Theorem 5.1.** Let $\mathcal{M}$ be a class of Markov chains whose state space is a $k$ dimensional direct product of spaces, and the problem size of a particular chain is denoted by $n$. (We can assume that $k < n$.) Any transition of $M \in \mathcal{M}$ changes only one coordinate, each coordinate with equal probability, and the transition probabilities do not depend on other coordinates. The transitions on each coordinate form an irreducible, aperiodic Markov chain, reversible w.r.t. a distribution $\pi_i$. Furthermore, all of these Markov chains restricted onto a single coordinate are assumed to be rapidly mixing, i.e., with the relaxation time $\frac{1}{1-\lambda_{2,i}}$ being bounded by a $\text{poly}(n)$ for all $i$. Then the Markov chain converges rapidly to the direct product of the $\pi_i$ distributions, and its second largest eigenvalue is

$$\frac{k - 1 + \max_i \{\lambda_{2,i}\}}{k}$$

and thus it is also bounded by a $\text{poly}(n)$.

**Proof.** The transition matrix of $M$ can be described as

$$\sum_{i=1}^k \left[ \bigotimes_{j=i}^{i-1} I_j \right] \otimes M_i \otimes \bigotimes_{j=i+1}^k I_j$$

where $\otimes$ denotes tensor product, $M_i$ denotes the the transition matrix of the Markov chain on the $i$th coordinate, $I_j$ denotes the identical matrix with the same size as $M_j$. Therefore, the eigenvalues of $M$ are

$$\frac{\sum_{i=1}^k \lambda_{j_i,i}}{k}$$

where $j_i$ is an arbitrary index in the $i$th Markov chain. The second largest eigenvalue of $M$ is indeed obtained when the maximal second largest eigenvalue is taken together with the other largest eigenvalues, i.e. 1s, and therefore it is

$$\frac{k - 1 + \max_i \{\lambda_{2,i}\}}{k}$$

If $g$ denotes the smallest spectral gap, i.e. $1 - \max_i \{\lambda_{2,i}\}$, then the second largest eigenvalue of $M$ is

$$\frac{k - g}{k} = 1 - \frac{g}{k}$$

namely, the second largest eigenvalue of $M$ is only $k$ times closer to 1 than the maximal second largest eigenvalue of the individual Markov chains.

Next, we announce two theorems that are direct extensions of statements for fast mixing by swap Markov chains for regular degree sequences [5] and for half-regular bipartite degree sequences [11]. The proofs are shown in the Appendix.

**Theorem 5.2.** The swap Markov chain on the realizations of almost regular degree sequences is rapidly mixing.
Theorem 5.3. The swap Markov chain on the realizations of almost half-regular bipartite degree sequences is rapidly mixing.

We are now ready to prove the main theorem.

Proof. (Theorem 2.3) We show that the RSO Markov chain on balanced realizations fulfills the conditions in Theorem 4.3. First we show that condition (i) of Theorem 4.3 holds. When restricted to the partition $Y_x$ (that is with $x$ fixed), the RSO Markov chain over the balanced realizations walks on the union of almost semi-regular and almost regular graphs. Since an RSO changes only one coordinate at a time, independently of other coordinates, all the conditions in Theorem 5.1 are fulfilled. Thus the relaxation time of the RSO Markov chain restricted onto $Y_x$ is bounded from above by the relaxation time of the chain restricted onto that coordinate (either an almost semi-regular bipartite or an almost regular graph) on which this restricted chain is the slowest (the smallest gap). However, based on Theorems 5.2 and 5.3, all these restrictions are fast mixing, and thus by Theorem 5.1 the polynomial bound in (i) holds. (Note that an almost semi-regular bipartite graph is also an almost half-regular bipartite graph.)

Next we show that condition (ii) of Theorem 4.3 also holds. The first coordinate is the union of auxiliary bipartite graphs, all of which are half-regular. The $M'$ Markov chain corresponding to Theorem 4.3 is the swap Markov chain on these auxiliary graphs. Since, again all conditions of Theorem 5.1 are fulfilled (mixing is fast within any coordinate due to Theorem 5.3), the $M'$ Markov chain is also fast mixing. Condition in Equation (4.3) holds due to Theorem 3.4. Since all conditions in Theorem 4.3 hold, the RSO swap Markov chain on balanced realizations is also rapidly mixing.

6. Conclusions. We have introduced a swap Markov chain over the space of balanced realizations of an arbitrary JDM, and therefore arbitrary degree sequences, and proved that it is fast mixing. Our proof is based on the following observations and intermediate results. Any balanced realization can be represented as the labeled union of almost regular and almost semi-regular bipartite graphs. Every balanced realization induces a collection of auxiliary bipartite graphs that are all half-regular and which can be naturally used to generate a disjoint partition of the state space of all balanced realizations. Using conductance methods we then directly proved a general theorem for fast mixing of Markov chains over such structured state spaces, which is similar to an earlier result by Martin and Randall [19]. We have also provided extensions to the existing proofs for MCMC fast mixing in the spaces of almost regular graphs based on Ref [5], and almost half-regular bipartite graphs based on Refs [11] and [20].

The obvious open question is the existence of a fast mixing Markov chain for sampling from the full space of simple graphs realizing a given JDM. Since a given JDM also uniquely determines the degree sequence, this could provide an important insight towards proving fast mixing for the degree based MCMC problem, which currently is still open.

Acknowledgements. Bélint Tóth is thanked for his comments related to Theorem 5.1. The authors express their thanks to K.E. Bassler and the Max-Planck-Institut für Physik Komplexer Systeme for its kind hospitality within its “Adaptive Networks” advanced study group program, where the writing of this paper has been completed.
Appendix. Proofs for Theorems 5.2 and 5.3. The proofs below are direct continuations of the proofs in Refs [5] and [11]. They can be followed within the language and the context of those two papers, which, however, we do not reproduce here, for reasons of brevity.

Proof of Theorem 5.2:

The proof is based on [5]. In that paper, the authors prove the rapid mixing of the swap Markov chain in case of regular graphs. The only lemma where they use regularity is Lemma 3, which claims the following: consider a graph in which at most 4 edges are “bad”, meaning that they have an assigned value $-1$ or $2$, and they form a subgraph of one of the following 5 configurations shown on Figure A.1. All other edges get a value $1$, and for each vertex, the sum of the assigned values of its edges is a constant $d \leq n/2$, where $n$ is the number of vertices of the graph. Then at most 3 switches is sufficient to transform this graph into a graph that does not contain any bad edges. A switch operates on 4 vertices $v_1, v_2, v_3, v_4$, and increases by 1 the assigned value of edges $(v_1, v_2)$ and $(v_3, v_4)$ (if the edge is not present, then an edge is added with value 1, if the assigned value was $-1$, the edge is deleted) and decreases by 1 the assigned values of edges $(v_2, v_3)$ and $(v_4, v_1)$ (if the modified value is 0, then the edge is deleted).

We prove that a similar lemma holds in the situation when the degrees are both $d \geq 0$ and $d + 1 \leq n$, in which case at most 4 switches are necessary. The 4 switches instead of 3 causes a higher order, but still polynomial upper bound on the relaxation time.

The first observation is that the graphs with bad edges are obtained by a matrix $G + G' - Z$, where all $G$, $G'$ and $Z$ are adjacency matrices of graphs with the same degree sequence. What follows is that whenever a vertex has a bad valued edge, the degree of the edge can be neither 0 nor $n$.

If there is vertex $v_1$ which has both a 2 and a $-1$ edge, then let denote $v_2$ and $v_3$ the corresponding neighbor vertices, respectively. There is a vertex $v_4$ such that $v_2$ is not connected to $v_4$, and $v_3$ might or might not be connected to $v_4$. We apply a switch on $v_1, v_2, v_3$ and $v_4$, which removes two bad edges and creates at most 1 new bad edges, thus decreasing the number of bad edges by at least 1.

If there is no vertex with different types of bad edges, then consider any bad edge $(v_1, v_2)$. If the assigned value is $-1$, then we have to find a $v_3$ which is connected to
v_2. If there is a −1 valued edge (v_3, v_4), then we apply a switch on v_1, v_2, v_3 and v_4, which removes two bad edges and creates at most 1 new bad edge, thus decreasing the number of bad edges by at least 1. Otherwise, there must be a vertex v_4 which is not connected to v_3 but connected to v_1, since d(v_1) ≥ d(v_3) − 1, and the difference on the sum of bad values for v_1 and v_3 is at least 2. We apply a switch on v_1, v_2, v_3 and v_4, which removes the 1 bad edge, (v_1, v_2).

Finally, if the assigned bad value to edge (v_1, v_2) is 2, then there must be a vertex v_3 such that v_2 is not connected to v_3. If there is a 2 valued edge (v_3, v_4), then we apply a switch on v_1, v_2, v_3 and v_4, which removes two bad edges and creates at most 1 new bad edges, thus decreasing the number of bad edges by at least 1. Otherwise, there must be a vertex v_4 which is connected to v_3 but not connected to v_1, since d(v_1) − 1 ≤ d(v_3), and the difference on the sum of bad values for v_1 and v_3 is at least 2. We apply a switch on v_1, v_2, v_3 and v_4, which removes the 1 bad edge (v_1, v_2).

Hence, while there are bad value edges, we can apply a switch that decreases the number of bad edges at least by 1. Since there are at most 4 bad edges, the number of necessary switches is at most 4.

Proof of Theorem 5.3:

The proof is based on [11]. In that paper, the authors prove the rapid mixing of a Markov chain on half-regular degree sequence realizations with a forbidden (possibly empty) star and (also possibly empty) one factor. The only place where they use half-regularity is their Lemma 4.6. In that lemma, the authors prove that a certain 0-1 matrix with at most 3 possible “bad” values, at most two −1 values and at most one 2 value, in the same column can be transformed into a 0−1 matrix using at most 3 switches. Here we prove if there is no forbidden sub-graph, and the degree sequence is almost half-regular (instead of half-regular), then the same lemma holds.

Indeed, in that case, the difference in the row sums can be also at most 1. Consider the row containing a bad value 2. There must be a row containing 0 in the same column where the bad value 2 is. The difference between 2 and 0 is 2, while the difference between the raw sums can be at most 1, therefore, we have to find another column, where the corresponding values are 1 and 0, and a switch on these 4 values eliminates the bad value 2 without creating a new bad value. Similar reasoning holds for the bad value −1.

REFERENCES