Two Characterizations of Success of Ergodic Markov Chain Families for Combinatorial Optimization

Swagato Sanyal[1] and Somenath Biswas[2]

[1] School of Technology and Computer Science
Tata Institute of Fundamental Research, Mumbai 400 005, India
swagatos@tcs.tifr.res.in

[2] Somenath Biswas
Department of Computer Science and Engineering
Indian Institute of Technology Kanpur, Kanpur 208 016, India
sb@cse.iitk.ac.in

Abstract. Randomized search heuristics are widely used for solving combinatorial optimization problems. Randomized search heuristics can often be modeled, either directly, or after some simple modifications, as families of ergodic Markov chains, where each chain has a goal state that corresponds to the optimum. Such a family is termed successful if every chain, starting from any state, reaches its goal state in a small number of steps. In this paper, we provide two characterizations of success of families of ergodic Markov chains. The first characterization shows that a necessary and sufficient condition for success for the family is that for every chain in the family, for every nonempty subset A of state space of the chain, such that A does not contain the goal state, the ratio $\Phi(A)$ of the ergodic flow out of A to the capacity of A is high. The second characterization shows that a family of ergodic Markov chains will be successful iff the conductance of the family, as well as the stationary probability of the goal state in each chain, are both high.

The present paper generalizes an earlier work of ours which had considered a restricted version, viz., families of time reversible and strongly aperiodic Markov chains.

1 Introduction

Randomized search heuristics are widely used in practice for solving combinatorial optimization problems; example are: the Metropolis algorithm [MRR53], various evolutionary algorithms [AD11], etc. These heuristics often perform very well even for hard problems. For example, it is known that the Metropolis algorithm provably works efficiently for random instances of the graph bisection problem [JS93]. Many of these heuristics can be seen either directly, or after certain modifications, as families of ergodic Markov chains. We will term such a family *successful* for a set of instances L of a combinatorial optimization problem Π if the family results in a polynomial time algorithm that solves Π on L. We provide in this paper two characterizations of such success.

Let $L \subseteq \Sigma^*$ be the set of instances of interest of Π. Associated with every $x \in L$ is a finite set S_x of *candidate solutions*. Usually, the association of S_x is implicit, and $|S_x|$ is exponential in $|x|$. Each candidate solution has a *cost*. The computational problem is, given x as input, to find the minimum cost element of S_x. (W.l.o.g., we assume the optimization problem to be a minimization problem, and that every instance has a unique minimum.) A typical randomized search heuristics for such a problem can be seen as a definition of a family of Markov chains, one chain M_x for each $x \in L$. The chain M_x has as its state space the set S_x of the candidate solutions of x. The state with the minimum cost is termed as the *goal state*.The chain is so defined that its state-to-state transitions can be simulated efficiently. Given x, we run M_x, starting at an arbitrary state, with the aim of reaching the goal state. Clearly, the family of chains is successful if for every $x \in L$, the first passage time to reach the goal state in M_x, starting from any state, is small in an appropriate probabilistic sense.

We consider families of ergodic Markov chains. Although Markov chain families resulting from typical evolutionary algorithms are not egodic, as such chains are not irreducible, (one never moves from a better candidate solutions to worse candidate solutions), one can make such chains irreducible without affecting the first passage time to the goal state by simply adding a transition from every trap state like the goal state to the initial state(s). The aperiodicity condition is usually met, if not, one can make the chain aperiodic by adding a small self loop transition probability, say p, to each state. This will only increase the first passage time to the goal state by a factor of $1/p$. The advantage that one obtains by considering families of ergodic chains is that then the structural notion of conductance is defined, which our characterizations make use of.

In an earlier paper [SRB10], we had proved characterizations similar to the ones presented here for the success of families of *strongly aperiodic, time reversible* Markov chains. The present paper generalizes the results as now we deal with *all* families of ergodic Markov chains. The major motivation for the generalization is that Markov chain families for many randomized search heuristics are not time reversible, though ergodic (or can be made so). A simple example is when we employ random walk on a directed graph when the walk has a stationary distribution. The proof techniques we had used for the restricted families cannot provide the generalization that we prove here. We had used the fact that a family of strongly aperiodic, time reversible Markov chains is rapidly mixing iff a fixed inverse polynomial in instance size lower bounds the conductance of each chain [SJ89]. Although, for the present results too, we use conductance, the result above does not hold for families of unrestricted ergodic Markov chains; one can provide examples of such families with high conductance which are not rapidly mixing. Instead, we use the fact that high conductance does imply rapid mixing for families of strongly aperiodic Markov chains, which are *not necessarily time reversible*. Mihail states this result in [M89]. We provide a proof of the result in the Appendix as we did not find its proof in published literature. We use the result in the following way. To prove that a family X of unrestricted ergodic Markov chains satisfies the first characterization iff it satisfies the second one, we define another family Y, through a simple modification of X. Y is so defined that (a) the first first passage time is one half of that of X, (b) so is its conductance, and (c) it is strongly aperiodic. We then show that the two conditions are equivalent for Y, which in turn will imply equivalence of the two conditions for X.

The notion of conductance is a well-studied structural notion, and there are techniques like canonical paths and resistance which are used to argue about conductance. As our success characterizations are in terms of conductance, we hope that these techniques can be used to argue about the success of the Markov chains approach for combinatorial optimization problems.

Our results also clarify how the success conditions for families of chains to solve combinatorial optimization problems relate to the success conditions for families of ergodic chains used for solving another class of problems, viz., generating almost uniformly at random an element of an implicitly defined finite set associated with the input instance [SJ89]. For the latter, success requires the family of chains to be *rapidly mixing*. Rapid mixing, along with the high stationary probabilities of each goal state, can together be easily seen to constitute a sufficient condition for the success of families for solving optimization problems. Our results show that, for families of strongly aperiodic ergodic chains, this sufficient condition for success is also a *necessary* condition.

2 Preliminaries: notations and basic notions

Let the language $L \subseteq \Sigma^* = \{0,1\}^*$ be a set of instances of interest of a combinatorial optimization problem Π. To use the Markov chains approach to solve Π on L, one defines a family $X = (X^{(s)} | s \in L)$ of Markov chains parameterized on strings in L. The random variable $X_t^{(s)}$ denotes the state in which the chain $X^{(s)}$ is at time t. Let $S^{(s)}$ denote the state space of the chain $X^{(s)}$ which is the same as the set of *candidate solutions* for the instance s in the context of the problem Π; thus we have a state in $X^{(s)}$ for each candidate solution. Each candidate solution,i.e., each state of the chain, has a *cost*. The problem Π requires one to find, given s, the candidate solution of s which has the minimum cost. (W.l.o.g., we assume Π to be a minimization problem and that every instance has a unique minimum.) The state of $X^{(s)}$ which corresponds to this unique minimum cost solution for s is termed as the *goal state* of the chain, and is denoted as $g^{(s)}$. We will use $P^{(s)} = (p_{i,j}^{(s)} : i, j \in S^{(s)})$ to denote the transition probability matrix of the chain $X^{(s)}$. All our chains will be ergodic and therefore each has a unique stationary probability distribution. As we have explained in the Introduction, the ergodicity assumption does not restrict the applicability of our results. We shall denote by $(\pi_i^{(s)} : i \in S^{(s)})$ the unique stationary probability distribution of the chain $X^{(s)}$. We further assume that for all sufficiently large $|s|$ and for some $k > 0$, $\pi_i^{(s)} \geq \frac{1}{2^{|s|^k}}$. In the following, for convenience, when we refer to this condition, we do so by saying that *no stationary probability is pathologically small.* We require this condition later in some of our proofs. The condition is met in natural examples and it is usual to assume it, for example, see [SJ89].

Definition 1 (strong aperiodicity). *A Markov chain X with state space S and transition probability matrix $P = (p_{i,j} : i, j \in S)$ is said to be strongly aperiodic if for each $i \in S$, $p_{i,i} \geq 1/2$, that is, the self loop probability is at least $1/2$.*

2.1 Notion of success

We assume that the family X is so defined that for each chain in the family, both the simulation of state-to-state transitions as well as finding the cost associated with each state can be done efficiently. Clearly then, the chain X will be able to solve the problem Π on L efficiently, in other words, X will be successful for Π on L, if it is the case that for each chain, the number of steps required to reach its goal state starting at any arbitrary state, is small, in an appropriate probabilistic state. One finds in the literature two notions of this idea of success, which we had formalized in [SRB10] as *S-success* and *W-success* respectively.

Definition 2 (S-success). *The family $X = \{X^{(s)} | s \in L\}$ is defined as S-successful if there exist constants $k, n_0 > 0$ such that $\forall s \in L$ such that $|s| \geq n_0$,*

$$\max_{u \in S^{(s)}} \boldsymbol{E}[\min\{t \geq 0 | X_t^{(s)} = g^{(s)}\} \mid X_0^{(s)} = u] \leq n_i^k$$

To put it in words, X is S-successful if on expectation it takes at most polynomially many steps to reach $g^{(s)}$ from any start state. However, it might not be enough for the number of steps to be small on expectation, it is also important that the behaviour of the chain is close to its expected behaviour sufficiently often. This leads to the notion of W-success. A family X is W-successful if the probability that the chain reaches $g^{(s)}$ in at most polynomially many steps is at least some inverse polynomial of $|s|$, and this holds for every start state. To put it formally,

Definition 3 (W-success). *The family $X = \{X^{(s)} | s \in L\}$ is defined as W-successful if there exist constants $k_1, k_2, n_0 > 0$ such that $\forall s \in L$ such that $|s| \geq n_0$,*

$$\min_{u \in S^{(s)}} \boldsymbol{Pr}[\min\{t \geq 0 | X_t^{(s)} = g^{(s)}\} \leq |s|^{k_1} \mid X_0^{(s)} = u] \geq |s|^{-k_2}$$

It turns out, as we had shown in [SRB10], that these two definitions of success are actually equivalent.

Proposition 1. *The family $X = \{X^{(s)} | s \in L\}$ is S-successful if and only if it is W-successful.*

Therefore, when we need the notion of *success*, we can use either S-success or W-success, whichever in convenient at the point of use.

2.2 Rapid mixing, Conductance

When we use a family of ergodic Markov chains to solve computational problems like combinatorial optimization or almost uniform generation, the efficiency of the approach most often depends on whether or not each chain in the family reaches quickly a distribution close to its stationary one. If that is the case for a family of chains, we say that the family is *rapidly mixing*. The standard definition of rapid mixing is due to Sinclair and Jerrum, as given in [SJ89]. The definition is in terms of *relative pointwise distance*.

Let $X = \{X^{(s)}$ be a family of ergodic Markov chains as defined in the beginning of this Section. The relative pointwise distance (r.p.d.) $\Delta(t)$ of $X^{(s)}$ at the t'th step is defined as

$$\Delta(t) = \max_{i,j \in S^{(s)}} \frac{|(P^{(s)})^t_{i,j} - \pi_j|}{\pi_j}$$

We can see r.p.d. is a measure of how far any state is from its stationary probability at the tth step. Sinclair and Jerrum gave the following definition:

Definition 4 (Sinclair and Jerrum). *$\forall \epsilon > 0$, we define $\tau_\epsilon = \min\{t > 0 \mid \forall t' \geq t, \Delta(t) < \epsilon\}$. A family $X = \{X^{(s)} | s \in L\}$ of Markov chains is said to be rapidly mixing if there exist constants n_0, $k > 0$ such that for all $X^{(s)}$ with $|s| \geq n_0$, $\forall \epsilon > 0$, $\tau_\epsilon \leq (|s| \lg \frac{1}{\epsilon})^k$.*

Thus, according to this definition, a Markov chain family is rapidly mixing if the r.p.d's for the chains fall below ϵ within a number of steps not more than some fixed polynomial in $|s|$ and $\lg \frac{1}{\epsilon}$.

Mihail [M89] provides another definition. As an ergodic Markov chain runs, with each step it comes closer and closer to its stationary distribution. Suppose we define the *error vector* at a step t to be the difference between the distribution at step t and the stationary distribution. Mihail defines a kind of L_2 norm for such error vectors. For a family to be rapidly mixing, for each chain the norm should come close to zero in polynomially many steps. Formally, let X be an ergodic Markov chain with state space S and stationary distribution $\pi = (\pi_i : i \in S)$. Let the distribution after t steps be $d^{(t)} = (d_i^{(t)} : i \in S)$. Then, $e^{(t)} = (d_i^{(t)} - \pi_i^{(t)} : i \in S)$ is the error vector after t steps. Clearly $\sum_{i \in S} e_i^{(t)} = 0$. For any vector $v = (v_i : i \in S)$, Mihail [M89] defines $||v||$ as follows:

$$||v|| = \sum_{i \in S} \frac{v_i^2}{\pi_i}$$

Definition 5 (Mihail). $\forall \epsilon > 0$, *we define* $\tau_\epsilon = \sup_{e^{(0)}} \min\{t > 0 \mid \forall t' \geq t, ||e_s^{(t')}|| < \epsilon\}$*. A family* $X = \{X^{(s)} | s \in L\}$ *of Markov chains is defined to be rapidly mixing if there exist constants* n_0*,* $k > 0$ *such that for all* $X^{(s)}$ *with* $|s| \geq n_0$*,* $\forall \epsilon > 0$*,* $\tau_\epsilon \leq (|s| \lg \frac{1}{\epsilon})^k$*.*

(In the above definition, $||e_s^{(t')}||$ denotes the error vector for the chain $X^{(s)}$ at step t'.

For a result in the next section, we make use of Mihail's definition of rapid mixing. We will also require the Sinclair-Jerrum definition at another place. However, this does not create any inconsistency; we prove that both the definitions above are equivalent.

Lemma 1. *The family of Markov chains* X *as desribed in the beginning of this section is rapidly mixing according to Definitiom 5* if and only if *it is rapid mixing according to Definition 4.*

The proof is provided in the Appendix.

Conductance
Conductance is an important structural property of ergodic Markov chains, as well as of families of Markov chains.

Definition 6 (Sinclair and Jerrum, [SJ89]). *Let* X *be an ergodic Markov chain with state space* S*, transition matrix* $P = (p_{i,j} : i, j \in S)$ *and stationary distribution* $\pi = (\pi_i : i \in S)$*. Then for all non empty proper subsets* A *of* S*, one defines* $Flow(A)$ *as* $\sum_{i \in A, j \in \overline{A}} \pi_i p_{ij}$*,* $Capacity(A)$ *as* $\sum_{i \in A} \pi_i$*, and* $\Phi(A)$ *as* $\frac{Flow(A)}{Capacity(A)}$*. The conductance of the chain is defined as*

$$\Phi(X) = \min_{A \subset S, A \neq \phi, Capacity(A) \leq \frac{1}{2}} \Phi(A)$$

One of the major contributions in the computational aspect of families of Markov chains is the result below:

Theorem 1 (Sinclair-Jerrum Conductance Theorem [SJ89]). *A family* $X = \{X^{(s)} | s \in L\}$ *of time reversible and strongly aperiodic Markov chains, with stationary probability of no state in each chain being pathologically small, is rapidly mixing if and only if there exist constants* n, k *such that for each* s*,* $|s| \geq n$*, the conductance* $\Phi(X^{(s)}$ *of the chain* $X^{(s)}$ *is greater than or equal to* $1/|s|^k$*.*

3 Two characterizations of success

In this section we state and prove the two characterizations for a family of ergodic Markov chains to be successful for a combinatorial optimization problems. Before the main result, we provide what will be needed in the main result proof.

3.1 Preamble to main result

Lemma 2. *Let X be an ergodic Markov chain with state space S, transition probability matrix $(P_{i,j} : i,j \in S$ and stationary distribution $\pi = (\pi_i : i \in S)$. Then, for all non-empty proper subsets A of S, $\sum_{i \in A, j \in \overline{A}} \pi_i P_{i,j} = \sum_{i \in A, j \in \overline{A}} \pi_j P_{j,i}$.*

The Lemma states the intuitively obvious fact that when the chain reaches its stationary distribution, the ergodic flow out of any subset of the state space equals the ergodic flow into that subset. The proof of the Lemma is given in the Appendix.

At a step in the proof of our characterizations, it would have been convenient to conclude for a family of ergodic chains its rapid mixing from its high conductance. However, unlike the case of time reversible, strongly aperiodic families of chains, this need not hold for general families, we show this through an example given in the Appendix. Such examples show that we will need some condition on the extent of aperiodicity. However, even if we assume strong aperiodicity, we cannot invoke the Sinclair-Jerrum Conductance Theorem stated in the last section, because its proof uses time reversibility in an essential manner, whereas we are concerned with general ergodic families, not necessarily time reversible. We use instead

Theorem 2. *Let X be a strongly aperiodic ergodic Markov chain with state space S, transition probability matrix $P = (P_{i,j} : i,j \in S)$, and the stationary distribution $\pi = (\pi_i : i \in S)$. Let the distribution after t steps be $d^{(t)} = (d_i^{(t)} : i \in S)$, and therefore, let $e^{(t)} = (d_i^{(t)} - \pi_i^{(t)} : i \in S)$ denote the error vector after t steps. Let Φ be the conductance of X. Then, $||e^{(t)}|| \leq (1 - \frac{\Phi^2}{2})^t \, ||e^{(0)}||$*

This was stated by Mihail in [M89] without proof. As we did not find its proof in published literature, we provide a proof in the Appendix. The proof we provide follows steps similar to the proof Mihail gives for the time reversible case, but we need to use Lemma 2. One possible point of interest in the proof is how the extent of aperiodicity is made use of. Actually, it can be seen from the proof that one could use a condition weaker than strong aperiodicity, viz., in each chain $X^{(s)}$ in the family, the self loop probability of $1/p(|s|)$ for some fixed polynomial $p(\cdot)$.

The Corollary below follows easily from Theorem 2:

Corollary 1. *If no stationary probability of any chain of X in the statement of the Theorem above is pathologically small then X is rapidly mixing.*

In the proof of the above it will be natural to use Definition 5 of rapid mixing. However, this is not a limitation as we have proved in Lemma 1 the equivalence of this definition with the more standard definition.

In the proof of our main result, to argue about chains which are not necessarily strongly aperiodic by employing Corollary 1, we shall invoke the following Lemma.

Lemma 3. *Let $Y^{(s)}$ be the chain with state space $S^{(s)}$ and transition probability matrix $Q^{(s)} = \frac{1}{2}(P^{(s)} + I) = (q_{i,j}^{(s)} : i, j \in S^{(s)})$, where I is the identity matrix and $P^{(s)}$ is the transition probability matrix of X. The family $Y = (Y^{(s)} : s \in L)$ is successful if and only if the family X is successful.*

Intuitively, it is clear that Y is like X except it moves at half the speed. The proof is provided in the Appendix.

3.2 Main result

We now state and prove a theorem which is our main result. The theorem proves the equivalence of three statements about families of ergodic Markov chains. The first statement asserts that the family is successful, the second and third staments are conditions on the family. Thus, *each of these two conditions is a characterization of success for families of ergodic Markov chains.*

Theorem 3. *Let $X = \{X^{(s)} : s \in L\}$ be a family of ergodic Markov chains, where for each chain $X^{(s)}$, let $S^{(s)}$ denote its state space, $g^{(s)}$ its unique goal state, $P^{(s)} = (p_{i,j}^{(s)} : i, j \in S^{(s)})$ its transition probability matrix, $(\pi_i^{(s)} : i \in S^{(s)})$ its stationary probability distribution, and $\Phi^{(s)}$ its conductance. We assume that none of the π_i's is pathologically small. The following three statements about the family X are equivalent:*
(a) The family X of Markov chains is successful.
(b) There exist constants $k, n_0 > 0$, such that for all s in L with $|s| \geq n_0$, for all non-empty subsets A of $S^{(s)} - \{g^{(s)}\}$, the condition

$$\Phi(A) \geq |s|^{-k}$$

is satisfied.
(c) There exist constants $n_0, c_1, c_2 > 0$ such that for all s in L with $|s| \geq n_0$ we have $\pi_{g^{(s)}}^{(s)} \geq |s|^{-c_1}$ and $\Phi^{(s)} \geq |s|^{-c_2}$.

Proof. ((a) *implies* (b))
Let X be successful and hence W-successful. Therefore, there exist constants $k_1, k_2, n_0 > 0$ such that $\forall s \in L$ with $|s| \geq n_0$,

$$\min_{u \in S^{(s)}} \mathbf{Pr}[\min\{t \geq 0 | X_t^{(i)} = g^{(s)}\} \leq |s|^{k_1} \mid X_0^{(i)} = u] \geq |s|^{-k_2}$$

For each s such that $|s| \geq n_0$, suppose that the start state is chosen from $S^{(i)}$ according to some probability distribution $(f_i^{(s)} : i \in S^{(s)})$ on states. Then
$\mathbf{Pr}[\min\{t \geq 0 | X_t^{(i)} = g^{(s)}\} \leq |s|^{k_1}]$

$$= \sum_{u \in S^{(i)}} \mathbf{Pr}[\min\{t \geq 0 | X_t^{(i)} = s_{opt}^{(i)}\} \leq |s|^{k_1} \mid X_0^{(i)} = u] \mathbf{Pr}[X_0^{(i)} = u]$$

$$\geq |s|^{-k_2} \sum_{u \in S^{(i)}} f_u^{(s)}$$

$$= |s|^{-k_2} \tag{1}$$

Our proof is obtained by showing that we reach a contradiction if we assume that the successful family X does not satisfy the condition of (b).

Let us assume that for X it holds that for all constants $k, m > 0, \exists s \in L$ with $|s| > m$, and an $A \subseteq S^{(s)} - \{g^{(s)}\}$ such that

$$\Phi(A) < |s|^{-k} \tag{2}$$

We fix constants $k', m' > 0$. Then $\exists s' \in L$ with $|s'| > m'$,and an $A \subseteq S^{(s')} - \{g^{(s')}\}$, such that

$$\Phi(A) < |s'|^{-k'} \tag{3}$$

For the chain $X^{(s')}$, using the A satisfying 3, we define the following initial distribution $f_s^{(i')} : s \in S^{(i')}$:

$$f_i^{(s')} = \begin{cases} \frac{\pi_i^{(s')}}{cap(A)} & \text{if } i \in A; \\ 0 & \text{otherwise;} \end{cases}$$

With this as the initial distribution, we derive a contradiction to (a) above for the chain $X^{(s')}$. First, we prove by induction on t that the probability that $X^{(s')}$ makes a transition from some state in A to some state in $\overline{A}$ for the first time in the t-th step, with initial distribution as $f_i^{(s')}$, is less than $|s'|^{-k'}$. Let $P^{(s')} = (p_{i,j}^{(s')} : i, j \in S^{(s')})$ be the transition matrix of $X^{(s')}$.

Base step(t=1): The probability that the 1st step of $X^{(s')}$ is from some state in A to some state in $\overline{A}$ is clearly

$$\sum_{i \in A, j \in \overline{A}} f_i^{(s')} p_{i,j}^{(s')} = \Phi(A) < |s'|^{-k'}$$

by invoking 3.

Induction step: We assume the induction hypothesis to be true for all $t \leq t'$. The probability that $X^{(s')}$ makes a move from some state in A to some state in $\overline{A}$ for the first time in $(t'+1)$-th step is $\mathcal{T}_{t'+1}/Capacity(A)$, where

$$\mathcal{T}_{t'+1} = \sum_{s_1, \ldots, s_{t'+1} \in A, s_{t'+2} \in \overline{A}} \pi_{s_1}^{(s')} p_{s_1,s_2}^{(s')} \ldots p_{s_{t'+1},s_{t'+2}}^{(s')}$$

$$= \sum_{s_2, \ldots, s_{t'+1} \in A, s_{t'+2} \in \overline{A}} \left(\left(\sum_{s_1 \in A} \pi_{s_1}^{(s')} p_{s_1,s_2}^{(s')} \right) p_{s_2,s_3}^{(s')} \ldots p_{s_{t'+1},s_{t'+2}}^{(s')} \right)$$

$$\leq \sum_{s_2, \ldots, s_{t'+1} \in A, s_{t'+2} \in \overline{A}} \pi_{s_2}^{(s')} p_{s_2,s_3}^{(s')} \ldots p_{s_{t'+1},s_{t'+2}}^{(s')}$$

$$= \mathcal{T}_{t'}$$

Thus, the probability that $X^{(s')}$ makes a move from some state in A to some state in $\overline{A}$ for the first time in the $(t'+1)$-th step

$$= \mathcal{T}_{t'+1}/Capacity(A)$$

$$\leq \mathcal{T}_{t'}/Capacity(A)$$

This is the probability that $X^{(s')}$ makes a move from some state in A to some state in $\overline{A}$ for the first time in the t'-th step and this probability is less than $|s'|^{-k'}$ (from the induction hypothesis). For the same initial distribution, $\forall k'' > 0$,

$$\mathbf{Pr}[\min\{t \geq 0 | X_t^{(s')} = g^{(s')}\} \leq |s'|^{k''}]$$

$$\leq \mathbf{P}[\min\{t \geq 0 | X_t^{(s')} \in \overline{A}\} \leq |s'|^{k''}]$$

$$< |s'|^{k''}.|s'|^{-k'} \quad \text{(using union bound over } t \leq |s'|^{k''}\text{)}$$

$$= |s'|^{-(k'-k'')}$$

Since we are free to choose k', k'' and m', we have, therefore, proved that $\forall c, c', m'' > 0, \exists s \in L$ with $|s| > m''$ and an initial distribution, such that
$\mathbf{Pr}[\min\{t \geq 0 | X_t^{(s)} = g^{(s)}\} \leq |s|^c] < |s|^{-c'}$
under that initial distribution. This contradicts 1, hence the family X could not have been successful and we have a contradiction.

((b) *implies* (c))
Let there exist constants $k, n_0 > 0$, such that $\forall s$ such that $|s| \geq n_0$,

$$\min_{A \subseteq S^{(s)} - \{g^{(s)}\}, A \neq \phi} \Phi(A) \geq |s|^{-k}$$

We first show that this implies at least inverse polynomial conductance.
Let m be a constant such that $\forall x \geq m$, $x^k + 1 \leq x^{k+1}$. Let $s \in L$ be an instance where $|s| \geq \max\{n_0, m\}$. From the definition of conductance and Lemma 2, we have $\Phi(X^{(s)}) = \min_{A \subseteq S^{(s)}, A \neq \phi} \max\{\Phi(A), \Phi(\overline{A})\}$. As one of A and $\overline{A}$ does not contain $g^{(s)}$, $\Phi(X^{(s)})$ is at least $|s|^{-k}$.
Next, we prove that the goal state has a high stationary distribution probability. Let $A' = S^{(s)} - \{g^{(s)}\}$. Thus,

$$\Phi(A') = \frac{\sum_{i \in S^{(s)} - \{g^{(s)}\}} \pi_i^{(s)} p_{i,g^{(s)}}^{(s)}}{\sum_{i \in S^{(s)} - \{g^{(s)}\}} \pi_i^{(s)}}$$

$$= \frac{\sum_{i \in S^{(s)} - \{g^{(s)}\}} \pi_{g^{(s)}}^{(s)} p_{g^{(s)},i}^{(s)}}{1 - \pi_{g^{(s)}}^{(s)}} \quad \text{(From Lemma 2)}$$

$$= \frac{\pi^{(s)}_{g^{(s)}} \sum_{i \in S^{(s)} - \{g^{(s)}\}} p^{(s)}_{g^{(s)},i}}{1 - \pi^{(s)}_{g^{(s)}}}$$

$$\leq \frac{\pi^{(s)}_{g^{(s)}}}{1 - \pi^{(s)}_{g^{(s)}}}$$

Also $\Phi(A') \geq n_i^{-k}$. Thus, we have

$$n_i^{-k} \leq \frac{\pi^{(s)}_{g^{(s)}}}{1 - \pi^{(s)}_{g^{(s)}}}$$

$$\Rightarrow 1 - \pi^{(s)}_{g^{(s)}} \leq |s|^k \pi^{(s)}_{g^{(s)}}$$

$$\Rightarrow \pi^{(s)}_{g^{(s)}} (1 + |s|^k) \geq 1$$

$$\Rightarrow \pi^{(s)}_{g^{(s)}} \geq \frac{1}{1 + |s|^k} \geq \frac{1}{|s|^{k+1}} \text{ (as } |s| \geq m)$$

((c) *implies* (a))
Let $Y^{(s)}$ be the chain with state space $S^{(s)}$ and transition probability matrix $Q^{(s)} = \frac{1}{2}(P^{(s)} + I) = (q^{(s)}_{i,j} : i, j \in S^{(s)})$, where I is the identity matrix, (as before) $P^{(s)}$ is the transition probability matrix of $X^{(s)}$. We observe that $\forall j \in S^{(s)}, q^{(s)}_{j,j} \geq \frac{1}{2}$. The stationary probabilioty distribution of $X^{(s)}$ and $Y^{(s)}$ are the same. Also, the conductance of $Y^{(s)}$ is half of that of $X^{(s)}$ and hence is at least $\frac{1}{2}|s|^{-k}$. Since each stationary probability is at least some inverse exponential in $|s|$, it follows from Corollary 1 that the family Y is rapidly mixing. Let $\forall s \in L$, such that $|s|$ is large enough, $g^{(s)} \geq |s|^{-k_1}$ for some constant $k_1 > 0$. Let $k_2 > 0$ be a constant such that if $(f^{(s)}_i : i \in S^{(s)})$ be the distribution of $Y^{(s)}$ after n^{k_2} steps, and, $\forall i \in S^{(s)}, f^{(s)}_i \geq \frac{\pi^{(s)}_j}{2}$, for any start state (since Y is a rapidly mixing family of Markov chains, there exists such a constant). Then, the expected number of steps till the chain hits the goal state for the first time is at most $2|s|^{k_1+k_2}$ (in other words, on expectation, the chain requires at most $2|s|^{k_1}$ blocks, each of $|s|^{k_2}$ steps, to reach $g^{(s)}$ for the first time). This shows that Y is S-successful and hence successful. From Lemma 3, we conclude that X is also successful. ■

We observe that the assumption of no stationary probability being pathologically small is used only in the '(c) implies (a)' part of the proof. Therefore, the statement (b) is a necessary condition for success even without this assumption. The assumption actually allows us to talk of rapid mixing according to either of the two definitions that we have provided. In the '(c) implies (a)' part, we note that high conductance and high stationary probability of $g^{(s)}$ imply low first passage time of $g^{(s)}$, but high conductance does not necessarily imply rapid mixing for chains which do not have sufficiently large self loop probability at each step, as shown by the example in the Appendix. We sidestepped this problem by defining family Y, used its rapid mixing to conclude it reaches goal states soon enough, and then we argued about success of X by invoking Lemma 3.

4 Concluding remarks

It will be of interest to see if we can make use of the success characterizations provided here to argue about the performance of randomized search heuristics which can be modeled as families of non-time reversible ergodic Markov chains. As our characterizations are in terms of conductance, and as there are techniques like canonical paths and resistance to obtain bounds for conductance, we do hope such uses will be possible. Another issue of interest is if we can provide success characterizations based on the notion of *merging conductance* instead of the conductance. We recall Mihaili [M89] had introduced the notion of merging conductance to argue about rapid mixing of general families of ergodic Markov chains, the same families that we have dealt with here.

References

[AD11] ANNE AUGER AND BENJAMIN DOERR,*Theory of Randomized Search Heuristics* World Scientific, 2011.

[JS93] M. JERRUM AND G.B. SORKIN, Simulated annealing for graph bisection *Proc.of 34th IEEE FOCS Conf.*, 94–103, 1993.

[MRR53] N. METROPOLIS, A. ROSENBLUTH, M. ROSENBLUTH, A. TELLER, AND E. TELLER, Equations of state calculations by fast computing machines *Jl. of Chemical Physics*, 21:1087–1092, 1953.

[M89] MILENA MIHAIL, Conductance and Convergence of Markov Chains– A Combinatorial Treatment of Expanders– *Proc.of 30th IEEE FOCS Conf.*, 526–531, 1989.

[SJ89] ALISTAIR SINCLAIR AND MARK JERRUM, Approximate Counting, Uniform Generation and Rapidly Mixing Markov Chains *Information and Computation*, 82:93–133, 1989.

[SRB10] SWAGATO SANYAL, RAJA S. AND SOMENATH BISWAS, Necessary and Sufficient Conditions for Success of the Metropolis Algorithm for Optimization *Proc.of 12th ACM GECCO Conf.*, 1417–1424, 2010.

Appendix

Proof of Lemma 7

Proof. Let $d^{(t)} = (d_i^{(t)} : i \in S^{(s)})$ be the distribution of $X^{(s)}$ after t steps.
(*if*)Let $\varDelta(t) < \sqrt{\epsilon}$. Now,

$$\begin{aligned}
\frac{|d_i^{(t)} - \pi_i|}{\pi_i} &= \frac{|\sum_{j \in S^{(s)}} d_j^{(0)} (P^{(s)})^t_{j,i} - \pi_i|}{\pi_i} \\
&= \frac{|\sum_{j \in S^{(s)}} d_j^{(0)} ((P^{(s)})^t_{j,i} - \pi_i)|}{\pi_i} \\
&\leq \frac{\sum_{j \in S^{(s)}} d_j^{(0)} |(P^{(s)})^t_{j,i} - \pi_i|}{\pi_i} \\
&\leq \sum_{j \in S^{(s)}} d_j^{(0)} \Delta(t) \\
&= \Delta(t)
\end{aligned}$$

Thus we have,

$$\begin{aligned}
||e^{(t)}|| &= \sum_{i \in S^{(s)}} \frac{|d_i^{(t)} - \pi_i|^2}{\pi_i} \\
&= \sum_{i \in S^{(s)}} \pi_i \frac{|d_i^{(t)} - \pi_i|^2}{\pi_i^2} \\
&\leq \sum_{i \in S^{(s)}} \pi_i \Delta(t)^2 \\
&< \epsilon
\end{aligned}$$

Which implies that if a family of Markov chains is rapid mixing according to Definition 4, it is also rapid mixing according Definition 5.
*(only if)*Here we make use of the assumption that for all sufficiently large $|s|$, for all $i \in S^{(s)}$, $\pi_i^{(s)}$ is at least $2^{-|s|^k}$. Let $||e^{(t)}|| \leq 2^{-|s|^k}\epsilon^2$ for all $||e^{(0)}||$. Then, each $\frac{|d_i^{(t)} - \pi_i|^2}{\pi_i}$ is at most $2^{-|s|^k}\epsilon^2$. Since each π_i is at least $2^{-|s|^k}$ for some $k > 0$, we have that each $\frac{|d_i^{(t)} - \pi_i|}{\pi_i}$ is at most ϵ. Since this holds for any $||e^{(0)}||$, that essentially means $\Delta(t) \leq \epsilon$. If X is rapidly mixing according to our definition, this is achieved in number of steps at most some fixed polynomial in $|s|$ and $\lg(2^{|s|^k} \times \epsilon^{-2})$ which is also polynomial in $|s|$ and $\lg \epsilon^{-1}$. Hence the implication follows. ■

Proof of Lemma 10
Proof. $\sum_{i \in S} \pi_i P_{i,j} = \sum_{i \in S} \pi_j P_{j,i}$ (since both sides are equal to π_j)
$\Rightarrow \sum_{j \in \overline{A}} \sum_{i \in S} \pi P_{i,j} = \sum_{j \in \overline{A}} \sum_{i \in S} \pi_j P_{j,i}$
Subtracting $\sum_{i \in \overline{A}, j \in \overline{A}} \pi_i P_{i,j}$ from both sides, and observing that $\sum_{i \in \overline{A}, j \in \overline{A}} \pi_i P_{i,j} = \sum_{i \in \overline{A}, j \in \overline{A}} \pi_j P_{j,i}$, the Lemma follows. ■

Proof of Theorem 11
Proof. Our proof proceeds in ithe same lines with the proof of Theorem 2.1 in [M89].

stage 1: PROBABILITY CHARGES
We observe that the error vector $e^{(t)}$ evolves over time exactly as any distribution vector.

$$e^{(t+1)} = d^{(t+1)} - \pi$$
$$= d^{(t)}P - \pi P$$
$$= e^{(t)}P$$

stage 2: AVERAGING ALONG EDGES
We focus on the decrease $||e^{(t)}|| - ||e^{(t+1)}||$. As it turns out, each edge serves to distribute the charges lying at its end vertices so as to decrease the error.

$$||e^{(t+1)}|| = \sum_{i \in S} \frac{(\sum_{j \in S} e_j^{(t)} P_{j,i})^2}{\pi_i}$$
$$= \sum_{i \in S} \frac{(e_i^{(t)} P_{i,i} + \sum_{j \in S, j \neq i} e_j^{(t)} P_{j,i})^2}{\pi_i}$$
$$= \sum_{i \in S} \frac{[e_i^{(t)}(1 - \sum_{j \in S, j \neq i} P_{i,j}) + \sum_{j \in S, j \neq i} e_j^{(t)} P_{j,i}]^2}{\pi_i}$$
$$= \sum_{i \in S} \frac{[e_i^{(t)} - \sum_{j \in S, j \neq i}(e_i^{(t)} P_{i,j} - e_j^{(t)} P_{j,i})]^2}{\pi_i}$$
$$= \sum_{i \in S} \frac{(e_i^{(t)})^2}{\pi_i} - 2 \sum_{i \in S} e_i^{(t)} \frac{\sum_{j \in S, j \neq i}(e_i^{(t)} P_{i,j} - e_j^{(t)} P_{j,i})}{\pi_i} + \sum_{i \in S} \frac{[\sum_{j \in S, j \neq i}(e_i^{(t)} P_{i,j} - e_j^{(t)} P_{j,i})]^2}{\pi_i}$$

Since $||e^{(t)}|| = \sum_{i \in S} \frac{(e_i^{(t)})^2}{\pi_i}$, we have

$$||e^{(t)}|| - ||e^{(t+1)}|| = 2 \sum_{i,j \in S, i \neq j} \frac{e_i^{(t)}(e_i^{(t)} P_{i,j} - e_j^{(t)} P_{j,i})}{\pi_i} - \sum_{i \in S} \frac{[\sum_{j \in S, j \neq i}(e_i^{(t)} P_{i,j} - e_j^{(t)} P_{j,i})]^2}{\pi_i} \tag{4}$$

Now,

$$\left[\sum_{j \in S, j \neq i}(e_i^{(t)} P_{i,j} - e_j^{(t)} P_{j,i})\right]^2$$
$$= \left[\sum_{j \in S, j \neq i}(\tfrac{e_i^{(t)}}{\pi_i} \pi_i P_{i,j} - \tfrac{e_j^{(t)}}{\pi_j} \pi_j P_{j,i})\right]^2$$
$$= \left[\tfrac{e_i^{(t)}}{\pi_i} \sum_{j \in S, j \neq i} \pi_i P_{i,j} - \sum_{j \in S, j \neq i} \tfrac{e_j^{(t)}}{\pi_j} \pi_j P_{j,i}\right]^2$$
$$= \left[\tfrac{e_i^{(t)}}{\pi_i} \sum_{j \in S, j \neq i} \pi_j P_{j,i} - \sum_{j \in S, j \neq i} \tfrac{e_j^{(t)}}{\pi_j} \pi_j P_{j,i}\right]^2$$

(since $\sum_{j \in S, j \neq i} \pi_i P_{i,j} = \sum_{j \in S, j \neq i} \pi_j P_{j,i} = \pi_i - \pi_i P_{i,i}$)

$$= \left[\sum_{j \in S, j \neq i} \left(\tfrac{e_i^{(t)}}{\pi_i} - \tfrac{e_j^{(t)}}{\pi_j} \right) \pi_j P_{j,i} \right]^2$$

$$= (\pi_i - \pi_i P_{i,i})^2 \left[\sum_{j \in S, j \neq i} \left(\tfrac{e_i^{(t)}}{\pi_i} - \tfrac{e_j^{(t)}}{\pi_j} \right) \tfrac{\pi_j P_{j,i}}{\pi_i - \pi_i P_{i,i}} \right]^2$$

$$\leq (\pi_i - \pi_i P_{i,i})^2 \sum_{j \in S, j \neq i} \left[\tfrac{e_i^{(t)}}{\pi_i} - \tfrac{e_j^{(t)}}{\pi_j} \right]^2 \frac{\pi_j P_{j,i}}{\pi_i - \pi_i P_{i,i}}$$

(since square of mean is always less than or equal to mean of squares)

$$= (\pi_i - \pi_i P_{i,i}) \sum_{j \in S, j \neq i} \left[\tfrac{e_i^{(t)}}{\pi_i} - \tfrac{e_j^{(t)}}{\pi_j} \right]^2 \pi_j P_{j,i}$$

$$\leq \frac{\pi_i}{2} \left[\tfrac{e_i^{(t)}}{\pi_i} - \tfrac{e_j^{(t)}}{\pi_j} \right]^2 \pi_j P_{j,i} \tag{5}$$

(as $P_{i,i} \geq \frac{1}{2}$) From equation 5, we have,

$$\frac{1}{\pi_i} \left[\sum_{j \in S, j \neq i} (e_i^{(t)} P_{i,j} - e_j^{(t)} P_{j,i}) \right]^2 \leq \frac{1}{2} \sum_{j \in S, j \neq i} \left[\tfrac{e_i^{(t)}}{\pi_i} - \tfrac{e_j^{(t)}}{\pi_j} \right]^2 \pi_j P_{j,i}$$

$$\Rightarrow \sum_{i \in S} \frac{\left[\sum_{j \in S, j \neq i} (e_i^{(t)} P_{i,j} - e_j^{(t)} P_{j,i}) \right]^2}{\pi_i} \leq \frac{1}{2} \sum_{i,j \in S, i \neq j} \left[\tfrac{e_i^{(t)}}{\pi_i} - \tfrac{e_j^{(t)}}{\pi_j} \right]^2 \pi_j P_{j,i} \tag{6}$$

We observe that this is the only place where we use the strong aperiodicity condition. Now,

$$2 \sum_{i,j \in S, i \neq j} \frac{e_i^{(t)} (e_i^{(t)} P_{i,j} - e_j^{(t)} P_{j,i})}{\pi_i}$$

$$= \sum_{i,j \in S, i \neq j} \frac{(e_i^{(t)})^2 P_{i,j}}{\pi_i} + \sum_{i,j \in S, i \neq j} \frac{(e_j^{(t)})^2 P_{j,i}}{\pi_j} - \sum_{i,j \in S, i \neq j} \frac{2 e_i^{(t)} e_j^{(t)} P_{j,i}}{\pi_i}$$

$$= \sum_{i \in S} \left[\tfrac{(e_i^{(t)})^2}{\pi_i^2} \sum_{j \in S, j \neq i} \pi_i P_{i,j} \right] + \sum_{i,j \in S, i \neq j} \frac{(e_j^{(t)})^2}{\pi_j^2} \pi_j P_{j,i} - \sum_{i,j \in S, i \neq j} 2 \frac{e_i^{(t)}}{\pi_i} \frac{e_j^{(t)}}{\pi_j} \pi_j P_{j,i}$$

$$= \sum_{i \in S} \left[\tfrac{(e_i^{(t)})^2}{\pi_i^2} \sum_{j \in S, j \neq i} \pi_j P_{j,i} \right] + \sum_{i,j \in S, i \neq j} \frac{(e_j^{(t)})^2}{\pi_j^2} \pi_j P_{j,i} - \sum_{i,j \in S, i \neq j} 2 \frac{e_i^{(t)}}{\pi_i} \frac{e_j^{(t)}}{\pi_j} \pi_j P_{j,i}$$

(since $\sum_{j \in S, j \neq i} \pi_i P_{i,j} = \sum_{j \in S, j \neq i} \pi_j P_{j,i} = \pi_i - \pi_i P_{i,i}$)

$$= \sum_{i,j \in S, i \neq j} \frac{(e_i^{(t)})^2}{\pi_i^2} \pi_j P_{j,i} + \sum_{i,j \in S, i \neq j} \frac{(e_j^{(t)})^2}{\pi_j^2} \pi_j P_{j,i} - \sum_{i,j \in S, i \neq j} 2 \frac{e_i^{(t)}}{\pi_i} \frac{e_j^{(t)}}{\pi_j} \pi_j P_{j,i}$$

$$= \sum_{i,j \in S, i \neq j} \left[\frac{e_i^{(t)}}{\pi_i} - \frac{e_j^{(t)}}{\pi_j} \right]^2 \pi_j P_{j,i} \tag{7}$$

From equations 6 and 7 we have

$$||e_i^{(t)}|| - ||e_i^{(t+1)}|| \geq \frac{1}{2} \sum_{i,j \in S, i \neq j} \left[\frac{e_i^{(t)}}{\pi_i} - \frac{e_j^{(t)}}{\pi_j} \right]^2 \pi_j P_{j,i} \tag{8}$$

Thus, error reduces more if there are more edges $\{i, j\}$ for which the absolute value of $\frac{e_i^{(t)}}{\pi_i} - \frac{e_j^{(t)}}{\pi_j}$ is large. We observe that 8 holds for any vector $e^{(t)}$ whcih evolves over time like a distribution/error vector, not necesarily an error vector.
stage 3: EFFECTIVE AVERAGING
In this stage we prove the Theorem only for vectors $e^{(t)}$ (not necessarily error vectors) where $\sum_{i \in \{s \in S | e_s^{(t)} > 0\}} \pi_i \leq \frac{1}{2}$ and $\sum_{i \in \{s \in S | e_s^{(t)} < 0\}} \pi_i \leq \frac{1}{2}$. Let $u^{(t)}$ and $v^{(t)}$ be two vectors. $\forall i \in S$, $u_i^{(t)}$ and $v_i^{(t)}$ are defined as follows:

$$u_i^{(t)} = \max(e_i^{(t)}, 0)$$

$$v_i^{(t)} = \min(e_i^{(t)}, 0)$$

Clearly,

$$\sum_{i,j \in S, i \neq j} \left[\frac{e_i^{(t)}}{\pi_i} - \frac{e_j^{(t)}}{\pi_j} \right]^2 \pi_j P_{j,i} \geq \sum_{i,j \in S, i \neq j} \left[\frac{u_i^{(t)}}{\pi_i} - \frac{u_j^{(t)}}{\pi_j} \right]^2 \pi_j P_{j,i} + \sum_{i,j \in S, i \neq j} \left[\frac{v_i^{(t)}}{\pi_i} - \frac{v_j^{(t)}}{\pi_j} \right]^2 \pi_j P_{j,i} \tag{9}$$

Without loss of generality we assume that the states are numbered in such a way that $\frac{e_i^{(t)}}{\pi_i} \geq \frac{e_j^{(t)}}{\pi_j}$ if $i < j$. Let there be n states. Now,

$$\left[\frac{u_i^{(t)}}{\pi_i} + \frac{u_j^{(t)}}{\pi_j} \right]^2 \leq 2 \left[\frac{(u_i^{(t)})^2}{\pi_i^2} + \frac{(u_j^{(t)})^2}{\pi_j^2} \right]$$

Hence,

$$\sum_{i,j \in S} \left[\frac{u_i^{(t)}}{\pi_i} + \frac{u_j^{(t)}}{\pi_j} \right]^2 \pi_j P_{j,i} \leq 2 \sum_{i,j \in S} \frac{(u_i^{(t)})^2}{\pi_i^2} \pi_j P_{j,i} + \sum_{i,j \in S} \frac{(u_j^{(t)})^2}{\pi_j^2} \pi_j P_{j,i}$$

$$= 2 \sum_{i \in S} \frac{(u_i^{(t)})^2}{\pi_i} + 2 \sum_{j \in S} \frac{(u_j^{(t)})^2}{\pi_j}$$

(since $\sum_{j \in S} \pi_j P_{j,i} = \pi_i$ and $\sum_{i \in S} \pi_j P_{j,i} = \pi_j$)

$$= 4||u^{(t)}|| \tag{10}$$

Thus,

$$\sum_{i,j\in S, i\neq j} \left[\frac{u_i^{(t)}}{\pi_i} - \frac{u_j^{(t)}}{\pi_j} \right]^2 \pi_j P_{j,i}$$

$$\sum_{i,j\in S} \left[\frac{u_i^{(t)}}{\pi_i} - \frac{u_j^{(t)}}{\pi_j} \right]^2 \pi_j P_{j,i}$$

$$\geq \frac{\left[\sum_{i,j\in S} \left[\frac{u_i^{(t)}}{\pi_i} - \frac{u_j^{(t)}}{\pi_j} \right]^2 \pi_j P_{j,i} \right] \left[\sum_{i,j\in S} \left[\frac{u_i^{(t)}}{\pi_i} + \frac{u_j^{(t)}}{\pi_j} \right]^2 \pi_j P_{j,i} \right]}{4||u^{(t)}||}$$

(From 10)

$$\geq \frac{\left[\sum_{i,j\in S} \left[\left[\frac{u_i^{(t)}}{\pi_i} - \frac{u_j^{(t)}}{\pi_j} \right] \sqrt{\pi_j P_{j,i}} \left[\frac{u_i^{(t)}}{\pi_i} + \frac{u_j^{(t)}}{\pi_j} \right] \sqrt{\pi_j P_{j,i}} \right] \right]^2}{4||u^{(t)}||}$$

(By Cauchy-Schwarz inequality)

$$= \frac{\left[\sum_{i,j\in S} \left| \frac{(u_i^{(t)})^2}{\pi_i^2} - \frac{(u_j^{(t)})^2}{\pi_j^2} \right| \pi_j P_{j,i} \right]^2}{4||u^{(t)}||} \tag{11}$$

Now,

$$\sum_{i,j\in S} \left| \frac{(u_i^{(t)})^2}{\pi_i^2} - \frac{(u_j^{(t)})^2}{\pi_j^2} \right| \pi_j P_{j,i}$$

$$= \sum_{i,j\in S, i\leq j} \left[\frac{(u_i^{(t)})^2}{\pi_i^2} - \frac{(u_j^{(t)})^2}{\pi_j^2} \right] \pi_j P_{j,i} + \sum_{i,j\in S, i>j} \left[\frac{(u_j^{(t)})^2}{\pi_j^2} - \frac{(u_i^{(t)})^2}{\pi_i^2} \right] \pi_j P_{j,i}$$

$$= \sum_{i,j\in S, i\leq j} \left[\frac{(u_i^{(t)})^2}{\pi_i^2} - \frac{(u_j^{(t)})^2}{\pi_j^2} \right] \pi_j P_{j,i} + \sum_{i,j\in S, i\leq j} \left[\frac{(u_i^{(t)})^2}{\pi_i^2} - \frac{(u_j^{(t)})^2}{\pi_j^2} \right] \pi_i P_{i,j}$$

$$= \sum_{i,j\in S, i\leq j} \left[\left[\frac{(u_i^{(t)})^2}{\pi_i^2} - \frac{(u_{i+1}^{(t)})^2}{\pi_{i+1}^2} \right] + \left[\frac{(u_{i+1}^{(t)})^2}{\pi_{i+1}^2} - \frac{(u_{i+2}^{(t)})^2}{\pi_{i+2}^2} \right] + \cdots + \left[\frac{(u_{j-1}^{(t)})^2}{\pi_{j-1}^2} - \frac{(u_j^{(t)})^2}{\pi_j^2} \right] \right] (\pi_i P_{i,j} + \pi_j P_{j,i})$$

$$= \sum_{i,j\in S, i\leq j} \sum_{k=i}^{j-1} \left[\frac{(u_k^{(t)})^2}{\pi_k^2} - \frac{(u_{k+1}^{(t)})^2}{\pi_{k+1}^2} \right] (\pi_i P_{i,j} + \pi_j P_{j,i})$$

$$= \sum_{k=i}^{n-1} \left[\left[\frac{(u_k^{(t)})^2}{\pi_k^2} - \frac{(u_{k+1}^{(t)})^2}{\pi_{k+1}^2} \right] \textstyle\sum_{i=1}^{k} \sum_{j=k+1}^{n} (\pi_i P_{i,j} + \pi_j P_{j,i}) \right]$$

$$= \sum_{k=i}^{n-1} \left[\left[\frac{(u_k^{(t)})^2}{\pi_k^2} - \frac{(u_{k+1}^{(t)})^2}{\pi_{k+1}^2} \right] \textstyle\sum_{i=1}^{k} \sum_{j=k+1}^{n} (2\pi_i P_{i,j}) \right]$$

(Putting $A = \{1 \cdots n\}$ in Lemma 2 we have $\sum_{i=1}^{k}\sum_{j=k+1}^{n}\pi_i P_{i,j} = \sum_{i=1}^{k}\sum_{j=k+1}^{n}\pi_j P_{j,i}$)

$$\geq 2\sum_{k=i}^{n-1}\left[\left[\frac{(u_k^{(t)})^2}{\pi_k^2} - \frac{(u_{k+1}^{(t)})^2}{\pi_{k+1}^2}\right]\Phi \sum_{i=1}^{k}\pi_i)\right]$$

$$= 2\Phi\sum_{i=1}^{n}\frac{(u_i^{(t)})^2}{\pi_i}]$$

$$= 2\Phi||u^{(t)}||$$

From 11 we have

$$\sum_{i,j\in S, i\neq j}\left[\frac{u_i^{(t)}}{\pi_i} - \frac{u_j^{(t)}}{\pi_j}\right]^2 \pi_j P_{j,i} \geq \Phi^2||u^{(t)}||$$

Similary one can show that

$$\sum_{i,j\in S, i\neq j}\left[\frac{v_i^{(t)}}{\pi_i} - \frac{v_j^{(t)}}{\pi_j}\right]^2 \pi_j P_{j,i} \geq \Phi^2||v^{(t)}||$$

Thus from 9,

$$\sum_{i,j\in S, i\neq j}\left[\frac{e_i^{(t)}}{\pi_i} - \frac{e_j^{(t)}}{\pi_j}\right]^2 \pi_j P_{j,i} \geq \Phi^2(||u^{(t)}|| + ||v^{(t)}||) = \Phi^2||f^{(t)}|| \tag{12}$$

Hence, from 8,

$$||e^{(t)}|| - ||e^{(t+1)}|| \geq \frac{\Phi^2}{2}||e^{(t)}||$$

$$\Rightarrow ||e^{(t+1)}|| \leq ||e^{(t)}||\left(1 - \frac{\Phi^2}{2}\right) \leq ||e^{(0)}||\left(1 - \frac{\Phi^2}{2}\right)^{t+1} \tag{13}$$

stage 4: NORMALIZATION
Here we prove 13 for any arbitrary error vector $e^{(t)}$. Let $u^{(t)}$ and $v^{(t)}$ be as defined in stage 3. We also assume the states to be numberd as described in stage 3. Let m be the smallest integer such that $\sum_{i=1}^{m}\pi_i > \frac{1}{2}$. We define the vector $f^{(t)} = (f_i^{(t)} : 1 \leq i \leq n)$ as follows:

$$f_i^{(t)} = e_i^{(t)} - \frac{e_m^{(t)}\pi_i}{\pi_m}$$

so that

$$\frac{f_i^{(t)}}{\pi_i} = \frac{e_i^{(t)}}{\pi_i} - \frac{e_m^{(t)}}{\pi_m}$$

Clearly, for the vector $f^{(t)}$, the assumptions of stage 3 hold true. From 8 and 12, we have

$$||e^{(t)}||-||e^{t+1)}|| \geq \frac{1}{2}\sum_{i,j\in S,i\neq j}\left[\frac{e_i^{(t)}}{\pi_i}-\frac{e_j^{(t)}}{\pi_j}\right]^2\pi_j P_{j,i} = \frac{1}{2}\sum_{i,j\in S,i\neq j}\left[\frac{f_i^{(t)}}{\pi_i}-\frac{f_j^{(t)}}{\pi_j}\right]^2\pi_j P_{j,i} \geq \Phi^2||f^{(t)}|| \tag{14}$$

Now,

$$||f^{(t)}|| = \sum_{i=1}^{n}\frac{(f_i^{(t)})^2}{\pi_i}$$

$$= ||e^{(t)}|| + \frac{(e_m^{(t)})^2}{\pi_m^2} \quad \text{(as } \textstyle\sum_{i=1}^n \pi_i = 1 \text{ and } \sum_{i=1}^n e_i^{(t)} = 0)$$

$$\geq ||e_i^{(t)}||$$

Thus from 14 we have

$$||e^{(t)}|| - ||e^{t+1)}|| \geq \Phi^2||e^{(t)}||$$

$$\Rightarrow ||e^{(t+1)}|| \leq ||e^{(t)}||\left(1-\tfrac{\Phi^2}{2}\right) \leq ||e^{(0)}||\left(1-\tfrac{\Phi^2}{2}\right)^{t+1} \tag{15}$$

Replacing $t+1$ with t, the Theorem follows. ■

Proof of Lemma 13

Proof. The Markov chain $Y^{(s)}$ is equivalent to a random process $\mathcal{R}$ which,

1. with $\frac{1}{2}$ probability simulates $X^{(s)}$.
2. with remaining $\frac{1}{2}$ probability stays at its current state. For a fixed start state $u \in S^{(s)}$.

Let
$\tau_Y = \min\{t \geq 0 | Y_t^{(s)} = g^{(s)}\}$,
$\tau_X = \min\{t \geq 0 | X_t^{(s)} = g^{(s)}\}$, and
τ_{idle}=Number of times $\mathcal{R}$ chooses to take step 2 before $X^{(s)}$ hits $g^{(s)}$ for the first time.
Clearly,

$$\tau_Y = \tau_{idle} + \tau_X \tag{16}$$

$$\Rightarrow \mathbf{E}[\tau_Y] = \mathbf{E}[\tau_{idle}] + \mathbf{E}[\tau_X]$$

Now, clearly $\mathbf{E}[\tau_{idle} \mid \tau_Y = a] = \frac{a}{2}$. Thus we have

$$\mathbf{E}[\tau_{idle}] = \sum_{a=1}^{\infty}\mathbf{E}[\tau_{idle} \mid \tau_Y = a].P[\tau_Y = a]$$

$$= \sum_{a=1}^{\infty} \frac{a}{2} P[\tau_Y = a]$$
$$= \frac{1}{2}\mathbf{E}[\tau_Y]$$

Substituting in 16 we have $\mathbf{E}[\tau_Y] = 2\mathbf{E}[\tau_X]$, we have the result. ■

Example of a family of ergodic Markov chains which has high conductance but does not mix rapidly

For each $s \in \Sigma^*$, let $X^{(s)}$ be the following chain: $S^{(s)} = \{1, 2, \ldots, 2n\}$ where $|s| = n$. For each $i, j \in S^{(s)}$, the transition probabilities $P_{i,j}^{(s)}$ are as follows:

$$P_{i,j}^{(s)} = \begin{cases} \frac{1}{2^n} \quad (\text{if } i = j) \\ \frac{1}{2} - \frac{1}{2^{n+1}} \quad (\text{if } j = (i+1) \bmod n \text{ or } j = (i-1) \bmod n) \\ 0 \quad (\text{otherwise}) \end{cases}$$

This chain is easily seen to be ergodic and time-reversible with the stationary probability of each state being $\frac{1}{2n}$. Now, consider any non-empty proper subset A of $S^{(s)}$. There has to be at least one cross edge from A to $\overline{A}$, let it be (u, v). Thus $\Phi(A) = \frac{\sum_{i \in A, j \in \overline{A}} \pi_i^{(s)} P_{i,j}^{(s)}}{\sum_{i \in A} \pi_i} \geq \frac{\frac{1}{2n}(\frac{1}{2} - \frac{1}{2^{n+1}})}{1}$. Therefore, the conductance is at least some inverse polynomial in n. If the family is rapidly mixing, there should exist a constant $k > 0$ such that if $X^{(0)}$ is run for at least n^k steps from any starting state, the stationary probability of each state becomes at least $\frac{1}{4n}$ for sufficiently large n. Now we note that for any starting state i, the chain can be in i after an odd number of steps only if it takes any of its self loops at least once. So given it starts at i, probability that it will be in i after $2n^k + 1$ steps is at most $(n^k + 1)\frac{1}{2^n}$ which is much less than $\frac{1}{4n}$ for sufficiently large n. Therefore, the family is not rapidly mixing despite having high conductance.