Edge Expansion and Spectral Gap of Nonnegative Matrices

Jenish C. Mehta
California Institute of Technology
jenishc@gmail.com

Leonard J. Schulman
California Institute of Technology
schulman@caltech.edu

Abstract

The classic graphical Cheeger inequalities state that if M is an $n \times n$ *symmetric* doubly stochastic matrix, then

$$\frac{1-\lambda_2(M)}{2} \leq \phi(M) \leq \sqrt{2 \cdot (1-\lambda_2(M))}$$

where $\phi(M) = \min_{S \subseteq [n], |S| \leq n/2} \left(\frac{1}{|S|} \sum_{i \in S, j \notin S} M_{i,j} \right)$ is the edge expansion of M, and $\lambda_2(M)$ is the second largest eigenvalue of M. We study the relationship between $\phi(A)$ and the spectral gap $1-\text{Re}\,\lambda_2(A)$ for *any* doubly stochastic matrix A (not necessarily symmetric), where $\lambda_2(A)$ is a nontrivial eigenvalue of A with maximum real part. Fiedler showed that the upper bound on $\phi(A)$ is unaffected, i.e., $\phi(A) \leq \sqrt{2 \cdot (1-\text{Re}\,\lambda_2(A))}$. With regards to the lower bound on $\phi(A)$, there are known constructions with

$$\phi(A) \in \Theta\left(\frac{1-\text{Re}\,\lambda_2(A)}{\log n}\right),$$

indicating that at least a mild dependence on n is necessary to lower bound $\phi(A)$.

In our first result, we provide an *exponentially* better construction of $n \times n$ doubly stochastic matrices A_n, for which

$$\phi(A_n) \leq \frac{1-\text{Re}\,\lambda_2(A_n)}{\sqrt{n}}.$$

In fact, *all* nontrivial eigenvalues of our matrices are 0, even though the matrices are highly *non-expanding*. We further show that this bound is in the correct range (up to the exponent of n), by showing that for any doubly stochastic matrix A,

$$\phi(A) \geq \frac{1-\text{Re}\,\lambda_2(A)}{35 \cdot n}.$$

As a consequence, unlike the symmetric case, there is a (necessary) loss of a factor of n^{α} for $\frac{1}{2} \leq \alpha \leq 1$ in lower bounding ϕ by the spectral gap in the nonsymmetric setting.

Our second result extends these bounds to general matrices R with nonnegative entries, to obtain a two-sided *gapped* refinement of the Perron-Frobenius theorem. Recall from the Perron-Frobenius theorem that for such R, there is a nonnegative eigenvalue r such that all eigenvalues of R lie within the closed disk of radius r about 0. Further, if R is irreducible, which means $\phi(R) > 0$ (for suitably defined ϕ), then r is positive and all other eigenvalues lie within the *open* disk, so (with eigenvalues sorted by real part), $\text{Re}\,\lambda_2(R) < r$. An extension of Fiedler's result provides an upper bound and our result provides the corresponding lower bound on $\phi(R)$ in terms of $r - \text{Re}\,\lambda_2(R)$, obtaining a two-sided quantitative version of the Perron-Frobenius theorem.

1 Introduction

1.1 Motivation and main result

We study the relationship between edge expansion and second eigenvalue of nonnegative matrices. We restrict to doubly stochastic matrices for exposition in the introduction, since the definitions are simpler and it captures most of the key ideas. The extension to general nonnegative matrices is treated in Section 1.2. Let A be an $n\times n$ doubly stochastic matrix, equivalently interpreted as a bi-regular weighted digraph. The *edge expansion* of A, denoted as $\phi(A)$, is defined as

$$\phi(A) = \min_{S\subseteq[n],|S|\leq n/2} \frac{\sum\limits_{i\in S, j\notin S} A_{i,j}}{|S|}.$$

The fact that A is doubly stochastic implies that $\phi(A) = \phi(A^T)$. ϕ is a measure of *how much* the graph would have to be modified for it to lose strong-connectedness; it also lower bounds *how frequently*, in steady state, the associated Markov chain switches between the blocks of any bi-partition; thus, it fundamentally expresses the *extent* to which the graph is connected.

The connection between edge expansion and the second eigenvalue has been of central importance in the case of *symmetric* doubly stochastic matrices M (equivalently, reversible Markov chains with uniform stationary distribution). For such M, let the eigenvalues be $1 = \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq -1$. The Cheeger inequalities give two-sided bounds between the edge expansion $\phi(M)$ and the *spectral gap* $1-\lambda_2(M)$.

Theorem 1. [Dod84, AM85, Alo86] (Cheeger's inequalities for symmetric doubly stochastic matrices) *Let M be a* symmetric *doubly stochastic matrix, then*

$$\frac{1-\lambda_2(M)}{2} \leq \phi(M) \leq \sqrt{2\cdot(1-\lambda_2(M))}.$$

This is closely related to earlier versions for Riemannian manifolds [Che70, Bus82]. Notably, the inequalities in Theorem 1 do not depend on n. Further, they are tight up to constants – the upper bound on ϕ is achieved by the cycle and the lower bound by the hypercube.

The key question we address in this work is whether or to what extent the Cheeger inequalities survive for *nonsymmetric* doubly stochastic matrices (the question was already asked, for e.g., in [MT06]). Let A be a doubly stochastic matrix, not necessarily symmetric. The eigenvalues of A lie in the unit disk around the origin in the complex plane, with an eigenvalue 1 called the *trivial* or *stochastic* eigenvalue corresponding to the eigenvector $\mathbf{1}$ (the all 1's vector), and all other eigenvalues considered *nontrivial*. Let the eigenvalues of A be ordered so that $1 = \lambda_1 \geq \operatorname{Re}\lambda_2 \geq \ldots \geq \operatorname{Re}\lambda_n \geq -1$. The spectral gap will be defined as $1-\operatorname{Re}\lambda_2(A)$. There are three motivations for this definition: first, the continuous-time Markov chain based on A is $\exp(t\cdot(A-I))$, and this spectral gap specifies the largest norm of any of the latter's nontrivial eigenvalues; second, $\operatorname{Re}\lambda_2(A) = 1$ if and only if $\phi = 0$ (i.e., if the matrix is reducible); finally, this gap lower bounds the distance (in the complex plane) between the trivial and any nontrivial eigenvalue.

It was noted by Fiedler [Fie95] that the upper bound on ϕ in Cheeger's inequality (Theorem 1) carries over easily to general doubly stochastic matrices, because for $M = (A + A^T)/2$, $\phi(A) = \phi(M)$ and $\operatorname{Re}\lambda_2(A) \leq \lambda_2(M)$. (In fact, Fiedler made a slightly different conclusion with these observations, but they immediately give the upper bound on ϕ, see Appendix C for an extension and proof).

Lemma 2. (Fiedler [Fie95]) *Let A be* any $n \times n$ *doubly stochastic matrix, then*

$$\phi(A) \le \sqrt{2 \cdot (1 - \operatorname{Re} \lambda_2(A))}.$$

It remained to investigate the other direction, i.e., the lower bound on $\phi(A)$ in terms of the spectral gap $1 - \operatorname{Re} \lambda_2(A)$. Towards this end, we define the function

Definition 3. $\Gamma(n) = \min \left\{ \frac{\phi(A)}{1 - \operatorname{Re} \lambda_2(A)} : A \text{ is an } n \times n \text{ doubly stochastic matrix} \right\}.$

For symmetric doubly stochastic matrices this minimum is no less than $1/2$. However, for doubly stochastic matrices A that are not necessarily symmetric, there are known (but perhaps not widely known) examples demonstrating that it is impossible to have a function Γ entirely independent of n. These examples, discussed in Section 3.1, show that

$$\Gamma(n) \le \frac{1}{\log n}. \tag{1}$$

One reason that ϕ and $1 - \operatorname{Re} \lambda_2$ are important is their connection to mixing time τ – the number of steps after which a random walk starting at any vertex converges to the uniform distribution over the vertices. For the case of symmetric doubly stochastic matrices – or in general reversible Markov chains – it is simple to show that $\tau \in O\left(\frac{\log n}{1-\lambda_2}\right)$, and by Cheeger's inequality (Theorem 1), it further gives $\tau \in O\left(\frac{\log n}{\phi^2}\right)$ where n is the number of vertices. For the case of general doubly stochastic matrices – or in general not-necessarily-reversible Markov chains – it still holds that $\tau \in O\left(\frac{\log n}{\phi^2}\right)$ by a result of Mihail ([Mih89], and see [Fil91]). Depending on the situation, either $\operatorname{Re} \lambda_2$ or ϕ may be easier to estimate. Most often, one is interested in concisely-specified chains on exponential-size sets, i.e., the number of vertices is n but the complexity parameter is $\log n$. In this case, either ϕ or λ_2 (for the reversible case) can be used to estimate τ. However, if the matrix or chain is given explicitly, then one reason the Cheeger inequalites are useful is because it is simpler to estimate τ using λ_2 which is computable in P while computing ϕ is NP-hard [GJS74].

From the point of view of applications to Markov Chain Monte Carlo (MCMC) algorithms, a $\log n$ loss in the relation between ϕ and $1 - \operatorname{Re} \lambda_2$ as implied by (1), is not a large factor. For MCMC algorithms, since "n" is the size of the state space being sampled or counted and the underlying complexity parameter is $\log n$, if it were true that $\Gamma(n) \ge \log^{-c} n$ for some constant c, then the loss in mixing time estimates would be polynomial in the complexity parameter, and thus, the quantity $1 - \operatorname{Re} \lambda_2$ could still be used to obtain a reasonable estimate of the mixing time even in the case of nonsymmetric doubly stochastic matrices.

However, the truth is much different. Our main result is that $\Gamma(n)$ does not scale as $\log^{-c} n$, but is *exponentially* smaller.

Theorem 4. (Bounds on Γ)

$$\frac{1}{35 \cdot n} \le \Gamma(n) \le \frac{1}{\sqrt{n}}.$$

We give an *explicit* construction of doubly stochastic matrices A_n for the upper bound on Γ. This construction of highly nonexpanding doubly stochastic matrices has, in addition, the surprising property that *every* nontrivial eigenvalue is 0. Thus, for non-reversible Markov chains, the connection between ϕ

and $\operatorname{Re}\lambda_2$ breaks down substantially, in that the upper bound on mixing time obtained by lower bounding ϕ by the spectral gap $1-\operatorname{Re}\lambda_2$ can be be exponentially weaker (when the complexity parameter is $\log n$) than the actual mixing time, whereas for reversible chains there is at worst a quadratic loss.

This theorem has a very natural extension to general nonnegative matrices, as we next describe.

1.2 General nonnegative matrices and a two-sided quantitative refinement of the Perron-Frobenius theorem

We extend our results to general nonnegative matrices R. By the Perron-Frobenius theorem (see Theorem 8), since R is nonnegative, it has a nonnegative eigenvalue r (called the PF eigenvalue) that is also largest in magnitude amongst all eigenvalues, and the corresponding left and right eigenvectors u and v have all nonnegative entries. Further, if R is irreducible, i.e., the underlying weighted digraph on edges with positive weight is strongly connected (for every (i,j) there is a k such that $R^k(i,j)>0$), then r is a simple eigenvalue, and u and v have all positive entries. We henceforth assume that nonzero nonnegative R has been scaled (to $\frac{1}{r}R$) so that $r=1$. Thus we can again write the eigenvalues of R as $1=\lambda_1(R)\geq \operatorname{Re}\lambda_2(R)\geq\cdots\geq\operatorname{Re}\lambda_n(R)\geq -1$. Scaling changes the spectrum but not the edge expansion (which we define next), so this canonical scaling is necessary before the two quantities can be compared.

Defining the edge expansion of R is slightly delicate so we explain the reasoning after Definition 11, and present only the final definition (see Definition 11) here. Consider u and v as left and right eigenvectors for eigenvalue 1 of R. If *every* such pair (u,v) for R has some i such that $u(i)=0$ or $v(i)=0$, then define $\phi(R)=0$. Otherwise, let u and v be some positive eigenvectors for eigenvalue 1 of R, normalized so that $\langle u,v\rangle=1$, and define the edge expansion of R as

$$\phi(R)=\min_{S\subseteq[n],\sum_{i\in S}u_i\cdot v_i\leq\frac{1}{2}}\frac{\sum\limits_{i\in S,j\in\overline{S}}R_{i,j}\cdot u_i\cdot v_j}{\sum\limits_{i\in S}u_i\cdot v_i}.$$

Given this definition of $\phi(R)$, we show the following.

Theorem 5. *Let R be a nonnegative matrix with PF eigenvalue 1, and let u and v be any corresponding left and right eigenvectors. Let $\kappa=\min_i u_i\cdot v_i$, and if $\kappa>0$, let u and v be normalized so that $\langle u,v\rangle=1$. Then*

$$\frac{1}{30}\cdot\frac{1-\operatorname{Re}\lambda_2(R)}{n+\ln\left(\frac{1}{\kappa}\right)}\;\leq\;\phi(R)\;\leq\;\sqrt{2\cdot(1-\operatorname{Re}\lambda_2(R))}.$$

The upper bound in Theorem 4 is a straightforward extension of Fiedler's bound (Lemma 2) based on the above mentioned definition of ϕ (Definition 11). Also note that the lower bound in Theorem 4 can be obtained by setting $\kappa=\frac{1}{n}$ in Theorem 5. The upper bound is proven in Appendix C and the lower bound is shown in Section 4.

Since Theorem 5 gives a two-sided relation between the *second* eigenvalue of nonnegative matrices and their edge expansion, it gives a two-sided quantitative refinement of the Perron-Frobenius theorem. Although the Perron-Frobenius theorem implies that the nontrivial eigenvalues of an irreducible nonnegative matrix R with PF eigenvalue 1 have real part strictly less than 1, it does not give any concrete separation. Further, it also does not provide a qualitative (or quantitative) implication in the other direction – whether a nonnegative matrix R with all nontrivial eigenvalues having real part strictly less than 1 *implies* that R is irreducible. Theorem 5 comes to remedy this by giving a lower and upper bound on the spectral gap in terms of ϕ, a quantitative measure of the irreducibility of R. We are not aware of any previous result of this form.

1.3 Mixing time

The third quantity we study is the mixing time of general nonnegative matrices, and we relate it to their singular values, edge expansion, and spectral gap. This helps us obtain new bounds on mixing time, and also obtain elementary proofs for known results. These results are treated in detail in the second part of the paper, in Section 5.

1.4 Perspectives

1.4.1 Matrix perturbations

Let A be a nonnegative matrix with PF eigenvalue 1 and corresponding positive left and right eigenvector w with $\langle w, w\rangle = 1$, and let $\kappa = \min_i w_i^2$. Given such A, it is certainly easier to calculate its spectrum than its edge expansion. However, in other cases, e.g., if the matrix is implicit in a nicely structured randomized algorithm (as in the canonical paths method [SJ89]), the edge expansion may actually be easier to bound. From this point of view, a lower bound on $\phi(A)$ in terms of the spectral gap is an *eigenvalue perturbation bound.* Specifically, one might write a nonnegative matrix A with small edge expansion $\phi(A)$ as a perturbation of another nonnegative matrix A_0, i.e.,

$$A = A_0 + \delta \cdot B$$

where A_0 has disconnected components S, S^c (for S achieving $\phi(A)$), and B is a matrix such that $\|B\|_2 \leq 1$, $Bw = 0$ and $B^T w = 0$. Due to the conditions on A and B, A_0 has PF eigenvalue 1 with left and right eigenvector w, and since $BD_w\mathbf{1} = 0$, writing $\mathbf{1}_{\overline{S}} = \mathbf{1} - \mathbf{1}_S$, we have,

$$\left|\frac{\langle \mathbf{1}_S, D_w B D_w \mathbf{1}_{\overline{S}}\rangle}{\langle \mathbf{1}_S, D_w D_w \mathbf{1}_S\rangle}\right| = \left|\frac{\langle \mathbf{1}_S, D_w B D_w \mathbf{1}_S\rangle}{\langle \mathbf{1}_S, D_w D_w \mathbf{1}_S\rangle}\right| = \left|\frac{\langle D_w\mathbf{1}_S, B D_w \mathbf{1}_S\rangle}{\langle D_w\mathbf{1}_S, D_w \mathbf{1}_S\rangle}\right| \leq \|B\|_2 \leq 1,$$

and it follows then that $\phi(R_0) - \delta \leq \phi(R) \leq \phi(R_0) + \delta$, and since in this case $\phi(R_0) = 0$, so $\phi(A) \leq \delta$. Edge expansion is therefore stable with respect to perturbation by B. What about the spectral gap?

A_0 has (at least) a double eigenvalue at 1, and A retains a simple eigenvalue at 1, so it is natural to try to apply eigenvalue stability results, specifically Bauer-Fike ([BF60], [Bha97] §VIII), to obtain an upper bound on $|1 - \lambda_2(A)|$ (and therefore also on $1 - \operatorname{Re}\lambda_2(A)$). However, Bauer-Fike requires A_0 to be diagonalizable, and the quality of the bound depends upon the condition number of the diagonalizing change of basis[1]. There are extensions of Bauer-Fike which do not require diagonalizability, but deteriorate exponentially in the size of the Jordan block, and the bound still depends on the condition number of the (now Jordan-izing) change of basis ([Saa11] Corollary 3.2). Since there is no a priori (i.e., function of n) bound on these condition numbers, these tools unfortunately do not imply any, even weak, result analogous to Theorem 4.

In summary, the lower bound in Theorem 4 should be viewed as a new eigenvalue perturbation bound:

$$1 - \operatorname{Re}\lambda_2(A) \leq 30 \cdot \delta \cdot \left(n + \ln\left(\frac{1}{\kappa}\right)\right).$$

A novel aspect of this bound, in comparison with the prior literature on eigenvalue perturbations, is that it does not depend on the condition number (or even the diagonalizability) of A or of A_0.

[1] The condition number is $\inf(\|B\| \cdot \|B^{-1}\|)$ over B such that BAB^{-1} is diagonal, with $\|\cdot\|$ being operator norm.

1.4.2 Relating eigenvalues of a nonnegative matrix and its additive symmetrization

Let A be any nonnegative matrix with PF eigenvalue 1, and corresponding positive left and right eigenvector w. Our result helps to give the following bounds between the second eigenvalue of a nonnegative matrix A and that of its *additive symmetrization* $\frac{1}{2}(A+A^T)$.

Lemma 6. *Let A be any nonnegative matrix with PF eigenvalue 1 and corresponding positive left and right eigenvector w, let $\kappa = \min_i w_i^2$ and $M = \frac{1}{2}(A+A^T)$.Then*

$$\frac{1}{1800}\left(\frac{1-\operatorname{Re}\lambda_2(A)}{n+\ln\left(\frac{1}{\kappa}\right)}\right)^2 \leq 1-\lambda_2(M) \leq 1-\operatorname{Re}\lambda_2(A).$$

The bounds immediately follows from Theorem 5 and the fact that ϕ is unchanged by additive symmetrization. We remark that any improved lower bound on $1-\lambda_2(M)$ in terms of $1-\operatorname{Re}\lambda_2(A)$ will help to improve the lower bound in Theorem 5. As a consequence of Lemma 6, any bound based on the second eigenvalue of symmetric nonnegative matrices can be applied, with dimension-dependent loss, to nonnegative matrices that have identical left and right eigenvector for the PF eigenvalue.

An example application of Lemma 6 is the following. For some doubly stochastic A (not necessarily symmetric), consider the continuous time Markov Chain associated with it, $\exp(t\cdot(A-I))$. It is well-known (for instance, [DSC96]) that for any standard basis vector x_i,

$$\left\|\exp(t\cdot(A-I))x_i - \frac{1}{n}\mathbf{1}\right\|_1^2 \leq n\cdot\exp\left(-2\cdot(1-\lambda_2(M))\cdot t\right).$$

Thus, using Lemma 6, we immediately get the bound in terms of the second eigenvalue of A itself (instead of its additive symmetrization),

$$\left\|\exp(t\cdot(A-I))x_i - \frac{1}{n}\mathbf{1}\right\|_1^2 \leq n\cdot\exp\left(-\frac{1}{1800}\left(\frac{1-\operatorname{Re}\lambda_2(A)}{n+\ln\left(\frac{1}{\kappa}\right)}\right)^2\cdot t\right).$$

1.4.3 The role of singular values

Although in the symmetric case singular values are simply the absolute values of the eigenvalues, the two sets can be much less related in the nonsymmetric case. It is not difficult to show the following (see Appendix A).

Lemma 7. *Let A be a nonnegative matrix with PF eigenvalue 1, and let w be the positive vector such that $Aw = w$ and $A^T w = w$. Then*

$$\frac{1-\sigma_2(A)}{2} \leq \phi(A),$$

where $\sigma_2(A)$ is the second largest singular value of A.

Proof. The proof is given in Appendix A. □

Despite appearances, this tells us little about edge expansion. To see this, consider the directed cycle on n vertices, for which every singular value (and in particular σ_2) is 1, so Lemma 7 gives a lower bound of 0 for ϕ, although $\phi = 2/n$. A meaningful lower bound should be 0 if and only if the graph is disconnected, i.e. it should be continuous in ϕ. An even more striking example is that of de Bruijn graphs (described in Section 3.1), for which half the singular values are 1, although $\phi = \Theta(1/\log n)$. Eigenvalues, on the other hand, are more informative, since a nonnegative matrix with PF eigenvalue 1 has a multiple *eigenvalue* at 1 if and only if, as a weighted graph, it is disconnected.

Despite these observations, singular values can be useful to infer information about mixing time, and can be used to recover all known upper bounds on mixing time using ϕ, as discussed in Section 5 and Lemma 24.

Outline: We state the preliminary definitions, theorems, and notations in Section 2. We give the construction for the upper bound on Γ in Theorem 4 in Section 3, and the lower bound on Γ will follow from the general lower bound on ϕ in Theorem 5. We show the upper bound on ϕ in Theorem 5 in Appendix C, and show the lower bound on ϕ in Section 4. We relate mixing time to singular values, edge expansion, and the spectral gap in Section 5 respectively. We defer all proofs to the Appendix.

2 Preliminaries

We consider *doubly stochastic* matrices $A \in \mathbb{R}^{n\times n}$ which are nonnegative matrices with entries in every row and column summing to 1. We also consider general nonnegative matrices R, and say that R is *strongly connected* or *irreducible*, if there is a path from s to t for every pair (s,t) of vertices in the underlying digraph on edges with positive weight, i.e. for every (s,t) there exists $k > 0$ such that $R^k(s,t) > 0$. We say R is *weakly connected*, if there is a pair of vertices (s,t) such that there is a path from s to t but no path from t to s in the underlying digraph (on edges with positive weight). We restate the Perron-Frobenius theorem for convenience.

Theorem 8. (Perron-Frobenius theorem [Per07, Fro12]) *Let $R \in \mathbb{R}^{n\times n}$ be a nonnegative matrix. Then the following hold for R.*

1. *R has some nonnegative eigenvalue r, such that all other eigenvalues have magnitude at most r, and R has nonnegative left and right eigenvectors u and v for r.*
2. *If R has some* positive *left and right eigenvectors u and v for some eigenvalue λ, then $\lambda = r$.*
3. *If R is irreducible, then r is positive and simple (unique), u and v are positive and unique, and all other eigenvalues have real part strictly less than r.*

We denote the all 1's vector by $\mathbf{1}$, and note that for a doubly stochastic matrix A, $\mathbf{1}$ is both the left and right eigenvector with eigenvalue 1. We say that 1 is the *trivial (or stochastic) eigenvalue*, and $\mathbf{1}$ is the *trivial (or stochastic) eigenvector,* of A. All other eigenvalues of A (corresponding to eigenvectors that are not trivial) will be called *nontrivial (or nonstochastic) eigenvalues* of A. Similarly, by Perron-Frobenius (Theorem 8, part 1), a nonnegative matrix R will have a simple nonnegative eigenvalue r such that all eigenvalues have magnitude at most r, and it will be called the *trivial* or PF eigenvalue of R, and all other eigenvalues of R will be called *nontrivial.* The left and right eigenvectors corresponding to r will be called the *trivial or PF left eigenvector* and *trivial or PF right eigenvector*. This leads us to the following definition.

Definition 9. *(Second eigenvalue)* If A is a doubly stochastic matrix, then $\lambda_2(A)$ is the nontrivial eigenvalue of A with the maximum real part, and $\lambda_m(A)$ is the nontrivial eigenvalue that is largest in magnitude. Similarly, if R is any general nonnegative matrix with PF eigenvalue 1, then $\lambda_2(R)$ is the nontrivial eigenvalue with the maximum real part, i.e., $1 = \lambda_1(R) \geq \text{Re}\,\lambda_2(R) \geq \cdots \geq \text{Re}\,\lambda_n(R) \geq -1$, and $\lambda_m(R)$ is the nontrivial eigenvalue that is largest in magnitude.

We will also consider singular values of nonnegative matrices A with identical positive left and right eigenvector w for PF eigenvalue 1, and denote them as $1 = \sigma_1(A) \geq \sigma_2(A) \geq \cdots \geq \sigma_n(A) \geq 0$ (see Lemma 20 for proof of $\sigma_1(A) = 1$). We denote (i,j)'th entry of $M \in \mathbb{C}^{n\times n}$ by $M(i,j)$ or $M_{i,j}$ depending on the importance of indices in context, denote the conjugate-transpose of M as M^* and the transpose of M as M^T. Any $M \in \mathbb{C}^{n\times n}$ has a *Schur decomposition* (see, e.g., [Lax07]) $M = UTU^*$ where T is an upper triangular matrix whose diagonal entries are the eigenvalues of M, and U is a unitary matrix. When we write "vector" we mean by default a column vector. For a vector v, we write $v(i)$ or v_i to denote its i'th entry. For any two vectors $x, y \in \mathbb{C}^n$, we use the standard *inner product* $\langle x, y\rangle = \sum_{i=1}^n x_i^* \cdot y_i$ defining the norm $\|x\|_2 = \sqrt{\langle x, x\rangle}$. We write $u \perp v$ to indicate that $\langle u, v\rangle = 0$. Note that $\langle x, My\rangle = \langle M^*x, y\rangle$. We denote the operator norm of M by $\|M\|_2 = \max_{u:\|u\|_2=1} \|Mu\|_2$, and recall that the operator norm is at most the Frobenius norm, i.e., $\|M\|_2 \leq \sqrt{\sum_{i,j} |M_{i,j}|^2}$. We write D_u for the diagonal matrix whose diagonal contains the vector u. Recall the Courant-Fischer variational characterization of eigenvalues for symmetric real matrices, applied to the second eigenvalue:

$$\max_{u\perp v_1} \frac{\langle u, Mu\rangle}{\langle u, u\rangle} = \lambda_2(M),$$

where v_1 is the eigenvector for the largest eigenvalue of M. We will use the symbol J for the all 1's matrix divided by n, i.e., $J = \frac{1}{n}\mathbf{1}\cdot\mathbf{1}^T$.

We say that any subset $S \subseteq [n]$ is a *cut*, denote its complement by $\overline{S}$, and denote the *characteristic vector of a cut* as $\mathbf{1}_S$, where $\mathbf{1}_S(i) = 1$ if $i \in S$ and 0 otherwise.

Definition 10. *(Edge expansion of doubly stochastic matrices)* For a doubly stochastic matrix A, the *edge expansion of the cut S* is defined as

$$\phi_S(A) := \frac{\langle \mathbf{1}_S, A\mathbf{1}_{\overline{S}}\rangle}{\min\{\langle \mathbf{1}_S, A\mathbf{1}\rangle, \langle \mathbf{1}_{\overline{S}}, A\mathbf{1}\rangle\}}$$

and the *edge expansion of A* is defined as

$$\phi(A) = \min_{S\subseteq[n]} \phi_S(A) = \min_{S\subseteq[n]} \frac{\langle \mathbf{1}_S, A\mathbf{1}_{\overline{S}}\rangle}{\min\{\langle \mathbf{1}_S, A\mathbf{1}\rangle, \langle \mathbf{1}_{\overline{S}}, A\mathbf{1}\rangle\}} = \min_{S,|S|\leq n/2} \frac{\sum_{i\in S, j\in\overline{S}} A_{i,j}}{|S|}.$$

We wish to extend these notions to general nonnegative matrices R. Since eigenvalues and singular values of real matrices remain unchanged whether we consider R or R^T, the same should hold of a meaningful definition of edge expansion. However, note that Definition 10 has this independence only if the matrix is Eulerian, i.e., $R\mathbf{1} = R^T\mathbf{1}$. Thus, to define edge expansion for general matrices, we transform R using its left and right eigenvectors u and v for eigenvalue 1 to obtain D_uRD_v, which is indeed Eulerian, since

$$D_uRD_v\mathbf{1} = D_uRv = D_uv = D_uD_v\mathbf{1} = D_vD_u\mathbf{1} = D_vu = D_vR^Tu = D_vR^TD_u\mathbf{1}.$$

Since D_uRD_v is Eulerian, we can define the edge expansion of R similar to that for doubly stochastic matrices:

Definition 11. *(Edge expansion of nonnegative matrices)* Let $R \in \mathbb{R}^{n\times n}$ be a nonnegative matrix with PF eigenvalue 1. If there are no positive (i.e., *everywhere* positive) left and right eigenvectors u and v for eigenvalue 1, then define the edge expansion $\phi(R) = 0$. Else, let u and v be any (see Lemma 12 for justification) positive left and right eigenvectors for eigenvalue 1, normalized so that $\langle u, v\rangle = 1$. The *edge expansion of the cut S* is defined as

$$\phi_S(R) := \frac{\langle \mathbf{1}_S, D_u R D_v \mathbf{1}_{\overline{S}}\rangle}{\min\{\langle \mathbf{1}_S, D_u R D_v \mathbf{1}\rangle, \langle \mathbf{1}_{\overline{S}}, D_u R D_v \mathbf{1}\rangle\}} \tag{2}$$

and the *edge expansion of R* is defined as

$$\phi(R) = \min_{S\subseteq[n]} \phi_S(R) = \min_{S\subseteq[n], \sum_{i\in S} u_i\cdot v_i \le \frac{1}{2}} \frac{\langle 1_S, D_u R D_v 1_{\overline{S}}\rangle}{\langle 1_S, D_u D_v \mathbf{1}\rangle} = \min_{S\subseteq[n], \sum_{i\in S} u_i\cdot v_i \le \frac{1}{2}} \frac{\sum\limits_{i\in S, j\in \overline{S}} R_{i,j}\cdot u_i \cdot v_j}{\sum\limits_{i\in S} u_i \cdot v_i}.$$

Lemma 12. *Let $R \in \mathbb{R}^{n\times n}$ be a nonnegative matrix with PF eigenvalue 1. Then the value of $\phi(R)$ according to Definition 11 is independent of the choice of the specific (left and right) eigenvectors u and v for eigenvalue 1 of R.*

Proof. Let G be the underlying unweighted directed graph for R, where there is an edge (u, v) in G if and only if $R_{u,v} > 0$. We prove the lemma based on the structure of G. Let G be maximally partitioned into k weakly connected components.

1. If G has some weakly connected component which does *not* have a 1 eigenvalue, then for any pair of left and right eigenvectors u and v for eigenvalue 1, there will be at least one entry i such that $u_i = 0$ or $v_i = 0$. For such matrices R, from Definition 11, $\phi(R) = 0$.

2. If all weakly connected components of G have eigenvalue 1, but there is some weakly connected component S that is not strongly connected, then there is no positive pair of eigenvectors u and v for R, or even for S. Observe that $S = \begin{bmatrix} A & B \\ 0 & C \end{bmatrix}$ with $B \neq 0$, else G is not maximally partitioned. For the sake of contradiction, let x and y be positive left and right eigenvectors for eigenvalue 1 of S. From $Sy = y$ and $S^T x = x$, we get that $Ay_1 + By_2 = y_1$ and $A^T x_1 = x_1$ with x and y partitioned into x_1, x_2 and y_1, y_2 based on the sizes of A, B, C. Thus,

$$\langle x_1, y_1\rangle = \langle x_1, Ay_1 + By_2\rangle = \langle A^T x_1, y_1\rangle + \langle x_1, By_2\rangle = \langle x_1, y_1\rangle + \langle x_1, By_2\rangle$$

   implying $\langle x_1, By_2\rangle = 0$ which is a contradiction since x_1 and y_1 are positive and there is some entry in B which is not 0. Thus, since every pair of eigenvectors x and y for eigenvalue 1 of S has some entry i with $x_i = 0$ or $y_i = 0$, every pair of (left and right) eigenvectors u and v for R (for eigenvalue 1) has some entry i which is 0, and so by Definition 11, $\phi(R) = 0$.

3. If G consists of *one* strongly connected component (i.e., $k = 1$), then by Perron-Frobenius (Theorem 8, part 3), there is a *unique* (up to scaling) pair of positive eigenvectors u and and v for eigenvalue 1.

4. The remaining case is that G has $k \ge 2$ strongly connected components each with eigenvalue 1. By Perron-Frobenius (Theorem 8), there is some pair of positive left and right eigenvectors u and

v for eigenvalue 1 (obtained by concatenating the positive left and right eigenvectors for each component individually). Choose the set S to be one of the components (the one for which the denominator of equation (2) is at most $\frac{1}{2}$), then the numerator of equation (2) will be 0, and thus $\phi(R) = 0$ even in this case, corresponding to the existence of a strict subset of vertices with no expansion.

Thus, Definition 11 for $\phi(R)$ does not depend on the specific choice of u and v. ☐

The Perron-Frobenius theorem (Theorem 8, part 3) can now be restated in terms of $\phi(R)$ and $\text{Re}\,\lambda_2(R)$ as follows.

Lemma 13. *(Perron-Frobenius, part 3 of Theorem 8, restated) Let R be a nonnegative matrix with PF eigenvalue 1. If $\phi(R) > 0$, then $\text{Re}\,\lambda_2(R) < 1$.*

Further, we obtain the following converse of Lemma 13.

Lemma 14. *(Converse of Lemma 13) Let R be a nonnegative matrix with PF eigenvalue 1. If $\text{Re}\,\lambda_2(R) < 1$,* and *there exists a pair of positive left and right eigenvectors u and v for eigenvalue 1 of R, then $\phi(R) > 0$.*

Proof. We show the contrapositive. Let R be as stated and let $\phi(R) = 0$. From Definition 11, if there are no positive eigenvectors u and v for eigenvalue 1, the lemma holds. So assume there are some positive u and v for eigenvalue 1 of R. Since $\phi(R) = 0$, there is some set S for which $\phi_S(R) = 0$, or $\sum_{i\in S, j\in \overline{S}} R_{i,j} \cdot u_i \cdot v_j = 0$. But since $u_i > 0$ and $v_i > 0$ for each i, and since R is nonnegative, it implies that for each $i \in S$, $j \in \overline{S}$, $R_{i,j} = 0$. Further, since $D_u R D_v$ is Eulerian, i.e. $D_u R D_v \mathbf{1} = D_v R^T D_u \mathbf{1}$, it implies that $\sum_{i\in \overline{S}, j\in S} R_{i,j} \cdot u_i \cdot v_j = 0$, further implying that $R_{i,j} = 0$ for each $i \in \overline{S}$, $j \in S$. As a consequence, v can be rewritten as two vectors v_S and $v_{\overline{S}}$, where $v_S(i) = v(i)$ if $i \in S$ and $v_S(i) = 0$ otherwise, and similarly $v_{\overline{S}}$. Similarly, split u into u_S and $u_{\overline{S}}$. Note that v_S and $v_{\overline{S}}$ are linearly independent (in fact, orthogonal), and both are right eigenvectors for eigenvalue 1 (similarly u_S and $u_{\overline{S}}$ as left eigenvectors). Thus, this implies that eigenvalue 1 for R has multiplicity at least 2, and thus $\lambda_2(R) = 1$, as required. ☐

The upper and lower bounds on $\phi(R)$ in Theorem 5 are *quantitative* versions of Lemmas 13 and 14 respectively.

We note that Cheeger's inequalities hold not only for any symmetric doubly stochastic matrix, but also for any nonnegative matrix R which satisfies detailed balance. We say that a nonnegative matrix R with positive left and right eigenvectors u and v for PF eigenvalue 1 satisfies *detailed balance* if $D_u R D_v$ is symmetric, which generalizes the usual definition of detailed balance (or reversibility) for stochastic matrices. We first note that if R satisfies the condition of detailed balance, then R has all real eigenvalues. To see this, let $W = D_u^{\frac{1}{2}} D_v^{-\frac{1}{2}}$ where the inverses are well-defined since we assume u and v are positive (else $\phi = 0$ by definition), and $A = WRW^{-1}$ where A has same eigenvalues as R. For $w = D_u^{\frac{1}{2}} D_v^{\frac{1}{2}} \mathbf{1}$ which is both the left and right eigenvector of A for eigenvalue 1, the detailed balance condition translates to $D_w A D_w = D_w A^T D_w$, which implies $A = A^T$, which further implies that all eigenvalues of A (and R) are real. We can thus state Cheeger inequalities for nonnegative matrices satisfying detailed balance.

Theorem 15. (Cheeger's inequalities for nonnegative matrices satisfying detailed balance) *Let R be a nonnegative matrix with PF eigenvalue 1 and positive left and right eigenvectors u and v (else $\phi(R) = 0$ by definition), and let R satisfy the condition of detailed balance, i.e. $D_u R D_v = D_v R^T D_u$. Then*

$$\frac{1 - \lambda_2(R)}{2} \leq \phi(R) \leq \sqrt{2 \cdot (1 - \lambda_2(R))}.$$

3 Construction of doubly stochastic matrices with small edge expansion and large spectral gap

As discussed, it might have been tempting to think that $\Gamma(n)$ (see Definition 3) should be independent of the matrix size n, since this holds for the symmetric case, and also for the upper bound on ϕ for the general nonsymmetric doubly stochastic case (Lemma 2). However, two known examples showed that a mild dependence on n cannot be avoided.

3.1 Known Constructions

Klawe-Vazirani construction: The first is a construction of Klawe [Kla84] – these families of d-regular undirected graphs have edge expansion $(\log n)^{-\gamma}$ for various $0 < \gamma < 1$. However, there is a natural way in which to direct the edges of these graphs and obtain a $(d/2)$-in, $(d/2)$-out -regular graph, and it was noted by U. Vazirani [Vaz17] in the 1980s that for this digraph A, which shares the same edge expansion as Klawe's, all eigenvalues (except the stochastic eigenvalue) have norm $\leq 1/2$. Specifically, one construction is as follows: let n be an odd prime, and create the graph on n vertices with in-degree and out-degree 2 by connecting every vertex $v \in \mathbb{Z}/n$ to two vertices, $1+v$ and $2v$. Dividing the adjacency matrix of the graph by 2 gives a doubly stochastic matrix A_{KV}. It is simple to see (by transforming to the Fourier basis over $\mathbb{Z}/n$) that the characteristic polynomial of A_{KV} is $x(x-1)((2x)^{n-1}-1)/(2x-1)$, so apart from the trivial eigenvalue 1, A_{KV} has $n-2$ nontrivial eigenvalues λ such that $|\lambda| = \frac{1}{2}$ and one eigenvalue 0, and thus, $\text{Re}\,\lambda_2(A_{KV}) \leq \frac{1}{2}$. Further, upper bounding the edge expansion ϕ by the vertex expansion bound (Theorem 2.1 in [Kla84]), it follows that for some constant c,

$$\Gamma(n) \leq \frac{\phi(A_{KV})}{1-\text{Re}\,\lambda_2(A_{KV})} \leq c \cdot \left(\frac{\log\log n}{\log n}\right)^{1/5}.$$

de Bruijn construction: A second example is the de Bruijn digraph [Bru46]. This is if anything even more striking: the doubly stochastic matrix (again representing random walk along an Eulerian digraph) has edge expansion $\Theta(1/\log n)$ [DT98], yet *all* the nontrivial eigenvalues are 0. More specifically, define a special case of de Bruijn [Bru46] graphs as follows: Let $n = 2^k$ for some integer k, and create the graph of degree 2 on n vertices by directing edges from each vertex $v = (v_1, v_2, \ldots, v_k) \in \{0,1\}^k$ to two vertices, $(v_2, v_3, \ldots, v_k, 0)$ and $(v_2, v_3, \ldots, v_k, 1)$. Dividing the adjacency matrix of the graph by 2 gives a doubly stochastic matrix A_{dB}. Since this random walk completely forgets its starting point after k steps, every nontrivial eigenvalue of A_{dB} is 0 (and each of its Jordan blocks is of dimension at most k). Further, it was shown in [DT98] that $\phi(A_{dB}) \in \Theta(1/k)$, and thus,

$$\Gamma(n) \leq \frac{\phi(A_{dB})}{1-\text{Re}\,\lambda_2(A_{dB})} \leq \frac{1}{k} = \frac{1}{\log n}.$$

Other literature – Feng-Li construction: We round out this discussion by recalling that Alon and Boppana [Alo86, Nil91] showed that for any infinite family of d-regular undirected graphs, the adjacency matrices, normalized to be doubly stochastic and with eigenvalues $1 = \lambda_1 \geq \lambda_2 \geq \ldots \geq -1$, have $\lambda_2 \geq \frac{2\sqrt{d-1}}{d} - o(1)$. Feng and Li [Li92] showed that undirectedness is essential to this bound: they provide a construction of cyclically-directed r-partite ($r \geq 2$) d-regular digraphs (with $n = kr$ vertices for $k > d$, $\gcd(k,d) = 1$), whose normalized adjacency matrices have (apart from r "trivial" eigenvalues), only eigenvalues of norm $\leq 1/d$. The construction is of an affine-linear nature quite similar to the preceding two, and to our knowledge does not give an upper bound on Γ any stronger than those.

3.2 A new construction

To achieve the upper bound in Theorem 4, we need a construction that is *exponentially* better than known examples. We give an explicit construction of $n \times n$ doubly stochastic matrices A_n (for every n) that are highly *nonexpanding*, since they contain sets with edge expansion less than $1/\sqrt{n}$, even though *every* nontrivial eigenvalue is 0. The construction might seem nonintuitive, but in Appendix F we give some explanation of how to arrive at it.

Theorem 16. *Let $m = \sqrt{n}$,*

$$a_n = \frac{m^2+m-1}{m\cdot(m+2)}, \qquad b_n = \frac{m+1}{m\cdot(m+2)}, \qquad c_n = \frac{1}{m\cdot(m+1)},$$

$$d_n = \frac{m^3+2m^2+m+1}{m\cdot(m+1)\cdot(m+2)}, \qquad e_n = \frac{1}{m\cdot(m+1)\cdot(m+2)}, \qquad f_n = \frac{2m+3}{m\cdot(m+1)\cdot(m+2)},$$

and define the $n \times n$ matrix

$$A_n = \begin{bmatrix} a_n & b_n & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & c_n & d_n & e_n & e_n & e_n & e_n & \cdots & e_n \\ 0 & c_n & e_n & d_n & e_n & e_n & e_n & \cdots & e_n \\ 0 & c_n & e_n & e_n & d_n & e_n & e_n & \cdots & e_n \\ 0 & c_n & e_n & e_n & e_n & d_n & e_n & \cdots & e_n \\ 0 & c_n & e_n & e_n & e_n & e_n & d_n & \cdots & e_n \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & c_n & e_n & e_n & e_n & e_n & e_n & \dots & d_n \\ b_n & f_n & c_n & c_n & c_n & c_n & c_n & \cdots & c_n \end{bmatrix}.$$

Then the following hold for A_n:

1. *A_n is doubly stochastic.*
2. *Every nontrivial eigenvalue of A_n is 0.*
3. *The edge expansion is bounded as*
$$\frac{1}{6\sqrt{n}} \leq \phi(A_n) \leq \frac{1}{\sqrt{n}}.$$
4. *As a consequence of 1,2,3,*
$$\phi(A_n) \leq \frac{1-\mathrm{Re}\,\lambda_2(A_n)}{\sqrt{n}}$$
and thus
$$\Gamma(n) \leq \frac{1}{\sqrt{n}}.$$

Proof. The proof is given in Appendix B. □

This completes the proof of the upper bound in Theorem 4. We remark that for the matrices A_n constructed in Theorem 16, it holds that

$$\phi(A_n) \leq \frac{1-|\lambda_i(A)|}{\sqrt{n}}$$

for any $i \neq 1$, giving a stronger guarantee than that required for Theorem 4.

We also remark that it would be unlikely to arrive at such a construction by algorithmic simulation, since the eigenvalues of the matrices A_n are extremely sensitive. Although $\lambda_2(A_n) = 0$, if we shift only $O(1/\sqrt{n})$ of the mass in the matrix A_n to create a matrix A'_n, by replacing a_n with $a'_n = a_n + b_n$, b_n with $b'_n = 0$, f_n with $f'_n = f_n + b_n$ and keeping c_n, d_n, e_n the same, then $\lambda_2(A'_n) = 1$. Thus, since perturbations of $O(1/\sqrt{n})$ (which is tiny for large n) cause the second eigenvalue to jump from 0 to 1 (and the spectral gap from 1 to 0), it would not be possible to make tiny changes to random matrices to arrive at a construction satisfying the required properties in Theorem 16.

4 Lower bound on the edge expansion ϕ in terms of the spectral gap

In this section, we prove the lower bound on ϕ in Theorem 5, and the lower bound on ϕ in Theorem 4 will follow as a special case. The proof is a result of a sequence of lemmas that we state next. The first lemma states that ϕ is sub-multiplicative in the following sense.

Lemma 17. *Let $R \in \mathbb{R}^{n\times n}$ be a nonnegative matrix with left and right eigenvectors u and v for the PF eigenvalue 1. Then*

$$\phi(R^k) \leq k \cdot \phi(R).$$

Proof. The proof is given in Appendix D.1. □

For the case of symmetric doubly stochastic matrices R, Lemma 17 follows from a theorem of Blakley and Roy [BR65]. (It does not fall into the framework of an extension of that result to the nonsymmetric case [Pat12]). Lemma 17 helps to lower bound $\phi(R)$ by taking powers of R, which is useful since we can take sufficient powers in order to make the matrix simple enough that its edge expansion is easily calculated. The next two lemmas follow by technical calculations.

Lemma 18. *Let $T \in \mathbb{C}^{n\times n}$ be an upper triangular matrix with $\|T\|_2 = \sigma$ and for every i, $|T_{i,i}| \leq \beta$. Then*

$$\|T^k\|_2 \leq n \cdot \sigma^n \cdot \binom{k+n}{n} \cdot \beta^{k-n}.$$

Proof. The proof is given in Appendix D.2. □

Using Lemma 18, we can show the following lemma for the special case of upper triangular matrices with operator norm at most 1.

Lemma 19. *Let $T \in \mathbb{C}^{n\times n}$ be an upper triangular matrix with $\|T\|_2 \leq 1$ and $|T_{i,i}| \leq \alpha < 1$ for every i. Then $\|T^k\| \leq \epsilon$ for*

$$k \geq \frac{4n + 2\ln(\frac{n}{\epsilon})}{1-\alpha}.$$

Proof. The proof is given in Appendix D.3. □

Given lemmas 17 and 19, we can lower bound $\phi(R)$ in terms of $1-|\lambda_m(R)|$ (where λ_m is the nontrivial eigenvalue that is maximum in magnitude). Our aim is to lower bound $\phi(R)$ by $\phi(R^k)$, but since the norm of R^k increases by powering, we cannot use the lemmas directly, since we do not want a dependence on $\sigma(R)$ in the final bound. To handle this, we transform R to A, such that $\phi(R)=\phi(A)$, the eigenvalues of R and A are the same, but $\sigma(A)=\|A\|_2=1$ irrespective of the norm of R.

Lemma 20. *Let R be a nonnegative matrix with positive (left and right) eigenvectors u and v for the PF eigenvalue 1, normalized so that $\langle u,v\rangle=1$. Define $A=D_u^{\frac{1}{2}}D_v^{-\frac{1}{2}}RD_u^{-\frac{1}{2}}D_v^{\frac{1}{2}}$. Then the following hold for A:*

1. *$\phi(A)=\phi(R)$.*
2. *For every i, $\lambda_i(A)=\lambda_i(R)$.*
3. *$\|A\|_2=1$.*

Proof. The proof is given in Appendix D.4. □

Given Lemma 20, we lower bound $\phi(A)$ using $\phi(A^k)$ in terms of $1-|\lambda_m(A)|$, to obtain the corresponding bounds for R.

Lemma 21. *Let R be a nonnegative matrix with positive (left and right) eigenvectors u and v for the PF eigenvalue 1, normalized so that $\langle u,v\rangle=1$. Let λ_m be the nontrivial eigenvalue of R that is maximum in magnitude and let $\kappa=\min_i u_i\cdot v_i$. Then*

$$\frac{1}{20}\cdot\frac{1-|\lambda_m|}{n+\ln\left(\frac{1}{\kappa}\right)}\leq\phi(R).$$

Proof. The proof is given in Appendix D.5. □

Given Lemma 21, we use the trick of lazy random walks to get a bound on $1-\operatorname{Re}\lambda_2(R)$ from a bound on $1-|\lambda_m(R)|$.

Lemma 22. *Let R be a nonnegative matrix with positive (left and right) eigenvectors u and v for the PF eigenvalue 1, normalized so that $\langle u,v\rangle=1$. Let $\kappa=\min_i u_i\cdot v_i$. Then*

$$\frac{1}{30}\cdot\frac{1-\operatorname{Re}\lambda_2(R)}{n+\ln\left(\frac{1}{\kappa}\right)}\leq\phi(R).$$

For any doubly stochastic matrix A,

$$\frac{1-\operatorname{Re}\lambda_2(A)}{35\cdot n}\leq\phi(A),$$

and thus

$$\frac{1}{35\cdot n}\leq\Gamma(n).$$

Proof. The proof is given in Appendix D.6. □

This completes the proof of the lower bound on ϕ in Theorem 5, and the upper bound on ϕ in Theorem 5 is shown in Appendix C. Combined with Theorem 16, this also completes the proof of Theorem 4.

5 Mixing time

We now study the mixing time of nonnegative matrices, and relate it to all the quantities we have studied so far. To motivate the definition of mixing time for general nonnegative matrices, we first consider the mixing time of doubly stochastic matrices. The mixing time of a doubly stochastic matrix A (i.e., of the underlying Markov chain) is the worst-case number of steps required for a random walk starting at any vertex to reach a distribution approximately uniform over the vertices. To avoid complications of periodic chains, we assume that A is $\frac{1}{2}$-lazy, meaning that for every i, $A_{i,i} \geq \frac{1}{2}$. Given any doubly stochastic matrix A, it can be easily converted to the lazy random walk $\frac{1}{2}I + \frac{1}{2}A$. This is still doubly stochastic and in the conversion both $\phi(A)$ and the spectral gap are halved. The mixing time will be finite provided only that the chain is connected. Consider the indicator vector $\mathbf{1}_{\{i\}}$ for any vertex i. We want to find the smallest τ such that $A^\tau \mathbf{1}_{\{i\}} \approx \frac{1}{n}\mathbf{1}$ or $A^\tau \mathbf{1}_{\{i\}} - \frac{1}{n}\mathbf{1} \approx 0$, which can further be written as $\left(A^\tau - \frac{1}{n}\mathbf{1}\cdot\mathbf{1}^T\right)\mathbf{1}_{\{i\}} \approx 0$. Concretely, for any ϵ, we want to find $\tau = \tau_\epsilon(A)$ such that for any i,

$$\left\|\left(A^\tau - \frac{1}{n}\mathbf{1}\cdot\mathbf{1}^T\right)\mathbf{1}_{\{i\}}\right\|_1 \leq \epsilon.$$

Given such a value of τ, for any vector x such that $\|x\|_1 = 1$, we get

$$\left\|\left(A^\tau - \frac{1}{n}\mathbf{1}\cdot\mathbf{1}^T\right)x\right\|_1 = \left\|\sum_i \left(A^\tau - \frac{1}{n}\mathbf{1}\cdot\mathbf{1}^T\right)x_i\mathbf{1}_{\{i\}}\right\|_1 \leq \sum_i |x_i| \left\|\left(A^\tau - \frac{1}{n}\mathbf{1}\cdot\mathbf{1}^T\right)\mathbf{1}_{\{i\}}\right\|_1 \leq \sum_i |x_i|\cdot\epsilon = \epsilon.$$

Thus, the mixing time $\tau_\epsilon(A)$ is the number τ for which $\|(A^\tau - J)\cdot x\|_1 \leq \epsilon$ for any x such that $\|x\|_1 = 1$.

We want to extend this definition to any nonnegative matrix R with PF eigenvalue 1 and corresponding positive left and right eigenvectors u and v. Note that if R is reducible (i.e., $\phi(R) = 0$), then the mixing time is infinite. Further, if R is periodic, then mixing time is again ill-defined. Thus, we again assume that R is irreducible and $\frac{1}{2}$-lazy, i.e. $R_{i,i} \geq \frac{1}{2}$ for every i. Let x be any nonnegative vector for the sake of exposition, although our final definition will not require nonnegativity and will hold for any x. We want to find τ such that $R^\tau x$ about the same as the component of x along the direction of v. Further, since we are right-multiplying and want convergence to the right eigenvector v, we will define the ℓ_1-norm using the left eigenvector u. Thus, for the starting vector x, instead of requiring $\|x\|_1 = 1$ as in the doubly stochastic case, we will require $\|D_u x\|_1 = 1$. Since x is nonnegative, $\|D_u x\|_1 = \langle u, x\rangle = 1$. Thus, we want to find τ such that $R^\tau x \approx v$, or $\left(R^\tau - v\cdot u^T\right)x \approx 0$. Since we measured the norm of the starting vector x with respect to u, we will also measure the norm of the final vector $\left(R^\tau - v\cdot u^T\right)x$ with respect to u. Thus we arrive at the following definition.

Definition 23. (Mixing time of general nonnegative matrices R) Let R be a $\frac{1}{2}$-lazy, irreducible nonnegative matrix with PF eigenvalue 1 with u and v as the corresponding positive left and right eigenvectors, where u and v are normalized so that $\langle u, v\rangle = \|D_u v\|_1 = 1$. Then the mixing time $\tau_\epsilon(R)$ is the smallest number τ such that $\left\|D_u\left(R^\tau - v\cdot u^T\right)x\right\|_1 \leq \epsilon$ for every vector x with $\|D_u x\|_1 = 1$.

We remark that similar to the doubly stochastic case, using the triangle inequality, it is sufficient to find mixing time of standard basis vectors $\mathbf{1}_{\{i\}}$. Let $y_i = \frac{\mathbf{1}_{\{i\}}}{\|D_u\mathbf{1}_{\{i\}}\|_1}$, then y_i is nonnegative, $\|D_u y_i\|_1 = \langle u, y_i\rangle = 1$, then for any x, such that $\|D_u x\|_1 = 1$, we can write

$$x = \sum_i c_i \mathbf{1}_{\{i\}} = \sum_i c_i \|D_u \mathbf{1}_{\{i\}}\|_1 y_i$$

with

$$\|D_u x\|_1 = \left\| D_u \sum_i c_i \mathbf{1}_{\{i\}} \right\|_1 = \sum_i |c_i| \|D_u \mathbf{1}_{\{i\}}\|_1 = 1.$$

Thus, if for every i, $\left\|D_u\left(R^\tau - v\cdot u^T\right)y_i\right\|_1 \leq \epsilon$, then

$$\left\|D_u\left(R^\tau - v\cdot u^T\right)x\right\|_1 = \left\|D_u\left(R^\tau - v\cdot u^T\right)\sum_i c_i \|D_u\mathbf{1}_{\{i\}}\|_1 y_i\right\|_1 \leq \sum_i c_i \|D_u\mathbf{1}_{\{i\}}\|_1 \left\|D_u\left(R^\tau - v\cdot u^T\right)y_i\right\|_1 \leq \epsilon.$$

Thus, it is sufficient to find mixing time for every nonnegative x with $\|D_u x\|_1 = \langle u, x\rangle = 1$, and it will hold for all x.

For the case of *reversible* nonnegative matrices M with PF eigenvalue 1, the mixing time is well-understood, and it is easily shown that

$$\tau_\epsilon(M) \leq \frac{\ln\left(\frac{n}{\kappa\cdot\epsilon}\right)}{1-\lambda_2(M)} \overset{\text{(Theorem 1)}}{\leq} \frac{2\cdot\ln\left(\frac{n}{\kappa\cdot\epsilon}\right)}{\phi^2(M)}. \tag{3}$$

We will give corresponding bounds for the mixing time of general nonnegative matrices.

5.1 Mixing time and singular values

We first show a simple lemma relating the mixing time of nonnegative matrices to the second singular value. This lemma is powerful enough to recover the bounds obtained by Fill [Fil91] and Mihail [Mih89] in an elementary way. Since the largest singular value of any general nonnegative matrix R with PF eigenvalue 1 could be much larger than 1, the relation between mixing time and second singular value makes sense only for nonnegative matrices with the same left and right eigenvector for eigenvalue 1, which have largest singular value 1 by Lemma 20.

Lemma 24. *(Mixing time and second singular value) Let A be a nonnegative matrix (not necessarily lazy) with PF eigenvalue 1, such that $Aw = w$ and $A^T w = w$ for some w with $\langle w, w\rangle = 1$, and let $\kappa = \min_i w_i^2$. Then for every $c > 0$,*

$$\tau_\epsilon(A) \leq \frac{c\cdot\ln\left(\frac{\sqrt{n}}{\sqrt{\kappa\cdot\epsilon}}\right)}{1-\sigma_2^c(A)} \leq \frac{c\cdot\ln\left(\frac{n}{\kappa\cdot\epsilon}\right)}{1-\sigma_2^c(A)}.$$

Proof. The proof is given in Appendix E.1. ☐

For the case of $c = 2$, Lemma 24 was obtained by Fill [Fil91], but we find our proof simpler.

5.2 Mixing time and edge expansion

We now relate the mixing time of general nonnegative matrices R to its edge expansion $\phi(R)$. The upper bound for row stochastic matrices R in terms of $\phi(R)$ were obtained by Mihail [Mih89] and simplified by Fill [Fil91] using Lemma 24 for $c = 2$. Thus, the following lemma is not new, but we prove it in Appendix E.2 for completeness, since our proof is simpler and holds for any nonnegative matrix R.

Lemma 25. *(Mixing time and edge expansion) Let $\tau_\epsilon(R)$ be the mixing time of a $\frac{1}{2}$-lazy nonnegative matrix R with PF eigenvalue 1 and corresponding positive left and right eigenvectors u and v, and let $\kappa = \min_i u_i \cdot v_i$. Then*

$$\frac{\frac{1}{2}-\epsilon}{\phi(R)} \;\leq\; \tau_\epsilon(R) \;\leq\; \frac{4\cdot\ln\left(\frac{n}{\kappa\cdot\epsilon}\right)}{\phi^2(R)}.$$

Proof. The proof is given in Appendix E.2. □

5.3 Mixing time and spectral gap

We obtain bounds for the mixing time of nonnegative matrices in terms of the spectral gap, using methods similar to the ones used to obtain the upper bound on ϕ in Theorem 5.

Lemma 26. *(Mixing time and spectral gap) Let $\tau_\epsilon(R)$ be the mixing time of a $\frac{1}{2}$-lazy nonnegative matrix R with PF eigenvalue 1 and corresponding positive left and right eigenvectors u and v, and let $\kappa = \min_i u_i \cdot v_i$. Then*

$$\frac{\frac{1}{2}-\epsilon}{\sqrt{2\cdot(1-\operatorname{Re}\lambda_2(R))}} \;\leq\; \tau_\epsilon(R) \;\leq\; 20\cdot\frac{n+\ln\left(\frac{1}{\kappa\cdot\epsilon}\right)}{1-\operatorname{Re}\lambda_2(R)}.$$

Proof. The proof is given in Appendix E.3 □

We remark that there is only *additive* and not multiplicative dependence on $\ln\left(\frac{n}{\kappa\cdot\epsilon}\right)$. Further, our construction for the upper bound in Theorem 4 also shows that the upper bound on τ using $\operatorname{Re}\lambda_2$ in Lemma 26 is also (almost) tight. For the construction of A_n in Theorem 4, letting the columns of U_n be $u_1,\ldots,u_n$, for $x=u_2$, $(A_n^k - J)u_2 = (1-(2+\sqrt{n})^{-1})^k u_3$, and so for $k=O(\sqrt{n})$, the triangular block of $A^{O(\sqrt{n})}$ has norm about $1/e$, which further becomes less than ϵ after about $\ln\left(\frac{n}{\epsilon}\right)$ powers. Thus for the matrices A_n, $\tau_\epsilon(A_n)\in O\left(\sqrt{n}\cdot\ln\left(\frac{n}{\epsilon}\right)\right)$. This shows Lemma 26 is also (almost) tight since $\lambda_2(A_n)=0$.

5.4 Mixing time of a nonnegative matrix and its additive symmetrization

We can also bound the mixing time of a nonnegative matrix A with the same left and right eigenvector w for PF eigenvalue 1, with the mixing time of its *additive symmetrization* $M=\frac{1}{2}(A+A^T)$. Note that we obtained a similar bound on the spectral gaps of A and M in Lemma 6. Since $\phi(A)=\phi(M)$, we can bound $\tau_\epsilon(A)$ and $\tau_\epsilon(M)$ using the two sided bounds between edge expansion and mixing time in Lemma 25. For the lower bound, we get $\gamma_1\cdot\sqrt{\tau_\epsilon(M)}\leq\tau_\epsilon(A)$, and for the upper bound, we get

$$\tau_\epsilon(A)\leq\gamma_2\cdot\tau_\epsilon^2(M),$$

where γ_1 and γ_2 are some functions polylogarithmic in $n,\kappa,\frac{1}{\epsilon}$. However, by bounding the appropriate operator, we can show a tighter upper bound on $\tau_\epsilon(A)$, with only a *linear* instead of quadratic dependence on $\tau_\epsilon(M)$.

Lemma 27. *Let A be a $\frac{1}{2}$-lazy nonnegative matrix with positive left and right eigenvector w for PF eigenvalue 1, let $M=\frac{1}{2}(A+A^T)$, and $\kappa=\min_i w_i^2$. Then*

$$\frac{1-2\epsilon}{4\cdot\ln^{\frac{1}{2}}\left(\frac{n}{\kappa\cdot\epsilon}\right)}\cdot\tau_\epsilon^{\frac{1}{2}}(M) \;\leq\; \tau_\epsilon(A) \;\leq\; \frac{2\cdot\ln\left(\frac{n}{\kappa\cdot\epsilon}\right)}{\ln\left(\frac{1}{\epsilon}\right)}\cdot\tau_\epsilon(M).$$

Proof. The proof is given in Appendix E.4. □

One example application of Lemma 27 is the following: given any undirected graph G such that each vertex has degree d, *any* manner of orienting the edges of G to obtain a graph in which every vertex has in-degree and out-degree $d/2$ cannot increase the mixing time of a random walk (up to a factor of $\ln\left(\frac{n}{\kappa\cdot\epsilon}\right)$).

5.5 Mixing time of the continuous operator

Let R be a nonnegative matrix with PF eigenvalue 1 and associated positive left and right eigenvectors u and v. The continuous time operator associated with R is defined as $\exp(t\cdot(R-I))$, where for any matrix M, we formally define $\exp(M)=\sum_{i=0}\frac{1}{i!}M^i$. The reason this operator is considered continuous, is that starting with any vector x_0, the vector x_t at time $t\in\mathbb{R}_{\geq 0}$ is defined as $x_t=\exp(t\cdot(R-I))\,x_0$. Since

$$\exp(t\cdot(R-I))=\exp(t\cdot R)\cdot\exp(-t\cdot I)=e^{-t}\sum_{i=0}^{\infty}\frac{1}{i!}t^iR^i$$

where we split the operator into two terms since R and I commute, it follows that $\exp(t\cdot(R-I))$ is non-negative, and if λ is any eigenvalue of R for eigenvector y, then $e^{t(\lambda-1)}$ is an eigenvalue of $\exp(t\cdot(R-I))$ for the same eigenvector y. Thus, it further follows that u and v are the left and right eigenvectors for $\exp(t\cdot(R-I))$ with PF eigenvalue 1. The mixing time of $\exp(t\cdot(R-I))$, is the value of t for which

$$\left\|D_u\left(\exp(t\cdot(R-I))-v\cdot u^T\right)v_0\right\|_1\leq\epsilon$$

for every v_0 such that $\|D_uv_0\|_1=1$, and thus, it is exactly same as considering the mixing time of $\exp(R-I)$ in the sense of Definition 23.

Lemma 28. *Let R be a nonnegative matrix (not necessarily lazy) with positive left and right eigenvectors u and v for PF eigenvalue 1, normalized so that $\langle u,v\rangle=1$ and let $\kappa=\min_i u_i\cdot v_i$. Then the mixing time of* $\exp(t\cdot(R-I))$*, or* $\tau_\epsilon(\exp(R-I))$ *is bounded as*

$$\frac{\frac{1}{2}-\epsilon}{\phi(R)}\;\leq\;\tau_\epsilon(\exp(R-I))\;\leq\;\frac{100\cdot\ln\left(\frac{n}{\kappa\cdot\epsilon}\right)}{\phi^2(R)}.$$

Proof. The proof is given in Appendix E.5. □

5.6 Bounds using the canonical paths method

For the case of symmetric nonnegative matrices M with PF eigenvalue 1, as stated earlier in this section in equation 3, since τ varies inversely with $1-\lambda_2$ (up to a loss of a factor of $\ln(\frac{n}{\kappa\cdot\epsilon})$), it follows that any lower bound on the spectral gap can be used to upper bound $\tau_\epsilon(M)$. Further, since $1-\lambda_2$ can be written as a minimization problem for symmetric matrices (see Section 2), any relaxation of the optimization problem can be used to obtain a lower bound on $1-\lambda_2$, and inequalities obtained thus are referred to as *Poincare inequalities*. One such method is to use *canonical paths* [Sin92] in the underlying weighted graph, which helps to bound mixing time in certain cases in which computing λ_2 or ϕ is infeasible. However, since it is possible to define canonical paths in many different ways, it leads to multiple relaxations to bound $1-\lambda_2$, each useful in a different context. We remark one particular definition and lemma here, since it is relevant to our construction in Theorem 16, after suitably modifying it for the doubly stochastic case.

Lemma 29. *[Sin92] Let M represent a symmetric doubly stochastic matrix. Let W be a set of paths in M, one between every pair of vertices. For any path $\gamma_{u,v} \in S$ between vertices (u,v) where $\gamma_{u,v}$ is simply a set of edges between u and v, let the number of edges or the (unweighted) length of the path be $|\gamma_{u,v}|$. Let*

$$\rho_W(M) = \max_{e=(x,y)} \frac{\sum\limits_{(u,v):e\in\gamma_{u,v}} |\gamma_{u,v}|}{n \cdot M_{x,y}}.$$

Then for any W,

$$1 - \lambda_2(M) \geq \frac{1}{\rho_W(M)}$$

and thus,

$$\tau_\epsilon(M) \leq \rho_W(M) \cdot \ln\left(\frac{n}{\epsilon}\right).$$

Corollary 30. *Combining Lemma 7 and Lemma 29, it follows that for any doubly stochastic matrix A, and any set W of paths in the underlying graph of AA^T,*

$$\tau_\epsilon(A) \leq \frac{2 \cdot \ln\left(\frac{n}{\epsilon}\right)}{1 - \sigma_2^2(A)} = \frac{2 \cdot \ln\left(\frac{n}{\epsilon}\right)}{1 - \lambda_2(AA^T)} \leq 2 \cdot \rho_W(AA^T) \cdot \ln\left(\frac{n}{\epsilon}\right).$$

Consider the example A_n in Theorem 16. It is not difficult to see that

$$\tau_\epsilon(A_n) \in O\left(\sqrt{n} \cdot \ln\left(\frac{n}{\epsilon}\right)\right). \tag{4}$$

This follows since the factor of $\sqrt{n}$ ensures that the only non zero entries in the triangular matrix T_n (see Appendix B) in the Schur form of $A^{\lceil\sqrt{n}\rceil}$ are about e^{-1}, and the factor of $\ln\left(\frac{n}{\epsilon}\right)$ further converts these entries to have magnitude at most $\frac{\epsilon}{n}$ in A^τ. Thus, the operator norm becomes about $\frac{\epsilon}{n}$, and the ℓ_1 norm gets upper bounded by ϵ. However, from Theorem 16, since $\phi(A_n) \geq \frac{1}{6\sqrt{n}}$, it follows from Lemma 25 that $\tau_\epsilon(A_n) \in O\left(n \cdot \ln\left(\frac{n}{\epsilon}\right)\right)$, about a quadratic factor off from the actual upper bound in equation 4. Further, from Theorem 16, the second eigenvalue of A_n is 0, and even employing Lemma 26 leads to a quadratic factor loss from the actual bound. However, Lemma 24 and Corollary 30 do give correct bounds. Since $\sigma_2(A_n) = 1 - \frac{1}{\sqrt{n}+2}$ from Theorem 16 (see the proof in Appendix B), it follows from Lemma 24 for $c = 1$ that $\tau_\epsilon(A_n) \in O\left(\sqrt{n} \cdot \ln\left(\frac{n}{\epsilon}\right)\right)$, matching the bound in equation 4. Now to see the bound given by canonical paths and corollary 30, consider the matrix $M = A_n A_n^T$. Every entry of M turns out to be positive, and the set W is thus chosen so that the path between any pair of vertices is simply the edge between the vertices. Further for r_n, α_n, β_n defined in the proof (Appendix B) of Theorem 16, $M = J + r_n^2 B$, where

$$B_{1,1} = \frac{n-2}{n}, \quad B_{n,n} = (n-2) \cdot \beta_n^2, \quad B_{i,i} = \alpha_n^2 + (n-3) \cdot \beta_n^2, \quad B_{1,n} = B_{n,1} = \frac{n-2}{\sqrt{n}} \cdot \beta_n,$$

$$B_{n,j} = B_{j,n} = \alpha_n \cdot \beta_n + (n-3) \cdot \beta_n^2, \quad B_{1,j} = B_{j,1} = \frac{1}{\sqrt{n}} \cdot (\alpha_n + (n-3) \cdot \beta_n), \quad B_{i,j} = 2 \cdot \alpha_n \cdot \beta_n + (n-4) \cdot \beta_n^2,$$

and $2 \leq i, j \leq n-1$. It follows that any entry of the matrix M is at least $c \cdot n^{-\frac{3}{2}}$ (for some constant c), and from Corollary 30, we get that $\tau_\epsilon(A_n) \in O\left(\sqrt{n} \cdot \ln\left(\frac{n}{\epsilon}\right)\right)$, matching the bound in equation 4.

Acknowledgements

We thank Shanghua Teng, Ori Parzanchevski and Umesh Vazirani for illuminating conversations. JCM would like to thank Luca Trevisan for his course "Graph Partitioning and Expanders" on VentureLabs, which got the author interested in Spectral Graph Theory. Research was supported by NSF grants 1319745, 1553477, 1618795, and 1909972; BSF grant 2012333; and, during a residency of LJS at the Israel Institute for Advanced Studies, by a EURIAS Senior Fellowship co-funded by the Marie Skłodowska-Curie Actions under the 7th Framework Programme.

References

[Alo86] N. Alon. Eigenvalues and expanders. *Combinatorica*, 6(2):83–96, 1986. 1, 3.1

[AM85] N. Alon and V. Milman. λ_1, isoperimetric inequalities for graphs, and superconcentrators. *Journal of Combinatorial Theory, Series B*, 38:73–88, 1985. 1

[BF60] F. L. Bauer and C. T. Fike. Norms and exclusion theorems. *Numerische Mathematik*, 2(1):137–141, 1960. 1.4.1

[Bha97] R. Bhatia. *Matrix Analysis*. Springer, 1997. 1.4.1

[BR65] G. R. Blakley and P. Roy. A Hölder type inequality for symmetric matrices with nonnegative entries. *Proc. American Mathematical Society*, 6(16):1244–1245, 1965. 4

[Bru46] N. G. de Bruijn. A combinatorial problem. *Proceedings of the Section of Sciences of the Koninklijke Nederlandse Akademie van Wetenschappen te Amsterdam*, 49(7):758–764, 1946. 3.1

[Bus82] P. Buser. A note on the isoperimetric constant. *Ann. Sci. École Norm. Sup.*, 15(2):213–230, 1982. 1.1

[Che70] J. Cheeger. A lower bound for the smallest eigenvalue of the Laplacian. In R. C. Gunning, editor, *Problems in analysis, a symposium in honor of S. Bochner*, pages 195–199. Princeton Univ. Press, 1970. 1.1

[Dod84] J. Dodziuk. Difference equations, isoperimetric inequality and transience of certain random walks. *Trans. Amer. Math. Soc.*, 284(2):787–794, 1984. 1

[DSC96] P. Diaconis and L. Saloff-Coste. Logarithmic Sobolev inequalities for finite Markov chains. *Annals of Applied Probability*, 6(3):695–750, 1996. 1.4.2

[DT98] C. Delorme and J.-P. Tillich. The spectrum of de Bruijn and Kautz graphs. *European Journal of Combinatorics*, 19(3):307–319, 1998. 3.1

[Fie95] M. Fiedler. An estimate for the nonstochastic eigenvalues of doubly stochastic matrices. *Linear algebra and its applications*, 214:133–143, 1995. 1.1, 2

[Fil91] J. A. Fill. Eigenvalue bounds on convergence to stationarity for nonreversible Markov chains, with an application to the exclusion process. *Annals of Applied Probability*, 1(1):62–87, 1991. 1.1, 5.1, 5.1, 5.2, E.2

[Fro12] G. Frobenius. Über matrizen aus nicht negativen elementen. 1912. 8

[GJS74] M. R. Garey, D. S. Johnson, and L. Stockmeyer. Some simplified NP-complete problems. In *Proceedings of the sixth annual ACM symposium on Theory of computing*, pages 47–63. ACM, 1974. 1.1

[Kla84] M. Klawe. Limitations on explicit constructions of expanding graphs. *SIAM Journal on Computing*, 13(1):156–166, 1984. 3.1

[Lax07] P. D. Lax. *Linear Algebra and Its Applications*. Wiley-Interscience, Hoboken, NJ, second edition, 2007. 2

[Li92] W.-C. W. Li. Character sums and Abelian Ramanujan graphs. *J. Number Theory*, 41:199–217, 1992. Appendix by K. Feng and W.-C. W. Li. 3.1

[Mih89] M. Mihail. Conductance and convergence of Markov chains - a combinatorial treatment of expanders. In *30th Annual symposium on Foundations of computer science*, pages 526–531. IEEE, 1989. 1.1, 5.1, 5.2

[MT06] R. Montenegro and P. Tetali. Mathematical aspects of mixing times in Markov chains. *Foundations and Trends in Theoretical Computer Science*, 1(3):237–354, 2006. 1.1

[Nil91] A. Nilli. On the second eigenvalue of a graph. *Discrete Math.*, 91:207–210, 1991. 3.1

[Pat12] T. H. Pate. Extending the Hölder type inequality of Blakley and Roy to non-symmetric non-square matrices. *Trans. American Mathematical Society*, 364(8):4267–4281, 2012. 4

[Per07] O. Perron. Zur theorie der matrices. *Mathematische Annalen*, 64(2):248–263, 1907. 8

[Saa11] Y. Saad. *Numerical methods for large eigenvalue problems*. SIAM, second edition, 2011. 1.4.1

[Sin92] A. Sinclair. Improved bounds for mixing rates of Markov chains and multicommodity flow. *Combinatorics, probability and Computing*, 1(4):351–370, 1992. 5.6, 29

[SJ89] A. Sinclair and M. Jerrum. Approximate counting, uniform generation and rapidly mixing Markov chains. *Information and Computation*, 82(1):93 – 133, 1989. 1.4.1

[Vaz17] U. Vazirani. Personal communication, 2017. 3.1

A Proof of Lemma 7

Proof. Let w be the left and right eigenvector of A for eigenvalue 1. Then note from part (3) of Lemma 20, we have that $\|A\|_2 = 1$. We first note that since the left and right eigenvectors for eigenvalue 1 are the same,

$$\max_{v \perp w} \frac{\|Av\|_2}{\|v\|_2} \leq \sigma_2(A). \tag{5}$$

To see this, let W be a unitary matrix with first column w, then since $Aw = w$ and $A^*w = A^T w = w$, W^*AW has two blocks, a 1×1 block containing the entry 1, and an $n-1 \times n-1$ block, and let the second block's singular value decomposition be PDQ with $n-1 \times n-1$ unitaries P and Q and diagonal D. Then

$$A = W \begin{bmatrix} 1 & 0 \\ 0 & P \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & D \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & Q \end{bmatrix} W^*,$$

giving a singular value decomposition for A. Thus, for any $v \perp w$ with $\|v\|_2 = 1$, W^*v has 0 as the first entry, and thus

$$\|Av\|_2 \leq \|W\|_2 \|P\|_2 \|D\|_2 \|Q\|_2 \|W^*\|_2 \|v\|_2 = \|D\|_2 = \sigma_2(A).$$

Thus we have,

$$\phi(A) = \min_{S: \sum_{i \in S} w_i^2 \leq \frac{1}{2}} \frac{\langle \mathbf{1}_S, D_w A D_w (\mathbf{1} - \mathbf{1}_S) \rangle}{\langle \mathbf{1}_s, D_w^2 \mathbf{1} \rangle} = 1 - \max_{S: \sum_{i \in S} w_i^2 \leq \frac{1}{2}} \frac{\langle \mathbf{1}_S, D_w A D_w \mathbf{1}_S \rangle}{\langle \mathbf{1}_s, D_w^2 \mathbf{1} \rangle} \tag{6}$$

where the second equality used the fact that $AD_w\mathbf{1} = Aw = w = D_w\mathbf{1}$. Let $D_w\mathbf{1}_S = c \cdot w + v$, where $\langle w, v \rangle = 0$. Then

$$c = \langle w, D_w \mathbf{1}_S \rangle = \langle \mathbf{1}_s, D_w^2 \mathbf{1} \rangle = \sum_{i \in S} w_i^2$$

and

$$\begin{aligned} \|v\|_2^2 &= \langle v, D_w \mathbf{1}_S - c \cdot w \rangle \\ &= \langle v, D_w \mathbf{1}_S \rangle \\ &= \langle D_w \mathbf{1}_S - c \cdot w, D_w \mathbf{1}_S \rangle \\ &= \langle D_w \mathbf{1}_S, D_w \mathbf{1}_S \rangle - c \cdot \langle D_w \mathbf{1}_S, w \rangle \\ &= c \cdot (1 - c). \end{aligned} \tag{7}$$

Thus,

$$
\begin{aligned}
\max_{S:\sum_{i\in S} w_i^2 \le \frac{1}{2}} \frac{\langle \mathbf{1}_S, D_w A D_w \mathbf{1}_S\rangle}{\langle \mathbf{1}_s, D_w^2 \mathbf{1}\rangle} &= \frac{\langle c\cdot w + v, A(c\cdot w+v)\rangle}{c} \\
&\quad [\text{since } D_w \text{ is diagonal and so } D_w^T = D_w] \\
&= \frac{c^2 + \langle v, Av\rangle}{c} \\
&\quad [\text{since } Aw = w \text{ and } A^T w = w] \\
&\le \frac{c^2 + \|v\|_2 \|Av\|_2}{c} \\
&\le \frac{c^2 + \|v\|_2 \sigma_2(A) \|v\|_2}{c} \\
&\quad [\text{from equation 5 since } v \perp w] \\
&= c + \sigma_2(A)\cdot(1-c) \\
&\quad [\text{using equation 7}] \\
&= \sigma_2(A) + (1-\sigma_2(A))\cdot c \\
&\le \frac{1+\sigma_2(A)}{2} \\
&\quad [\text{since } c = \langle \mathbf{1}_s, D_w^2 \mathbf{1}\rangle \le \frac{1}{2}]
\end{aligned}
$$

which completes the proof after replacing the above upper bound in equation 6.

A simple extension (and alternate proof) is the following. Let $H = AA^T$, then H is a nonnegative matrix that has w as the left and right eigenvector for eigenvalue 1. For any integer k, we know that H^k is nonnegative, and since $\lambda_2(H) = \sigma_2^2(A)$, we get

$$\frac{1-\sigma_2^{2\cdot k}(A)}{2} = \frac{1-\lambda_2^k(H)}{2} \le \phi(H^k) \le k\cdot\phi(H) \le k\cdot(2\cdot\phi(A))$$

where the first inequality used Cheeger's inequality (Theorem 15) and the second and third inequalities were obtained by 17 and its proof in Appendix D.1. Thus, we get that for any integer $c \ge 2$,

$$\frac{1-\sigma_2^c(A)}{2\cdot c} \le \phi(A).$$

□

B Proof of the main construction in Theorem 16

Proof. The following calculations are easy to check, to see that A_n is a doubly stochastic matrix:

1. $a_n \ge 0$, $b_n \ge 0$, $c_n \ge 0$, $d_n \ge 0$, $e_n \ge 0$, $f_n \ge 0$.

2. $a_n + b_n = 1$.

3. $c_n + d_n + (n-3)e_n = 1$.

4. $b_n + f_n + (n-2)c_n = 1.$

This completes the proof of (1).

A_n is triangularized as T_n by the unitary U_n, i.e.

$$A_n = U_n T_n U_n^*,$$

with T_n and U_n defined as follows. Recall that $m = \sqrt{n}$. Let

$$r_n = 1 - \frac{1}{m+2},$$

$$\alpha_n = \frac{-n^2 + 2n - \sqrt{n}}{n \cdot (n-1)} = -1 + \frac{1}{m \cdot (m+1)},$$

$$\beta_n = \frac{n - \sqrt{n}}{n \cdot (n-1)} = \frac{1}{m \cdot (m+1)},$$

$$T_n = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & r_n & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & r_n & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & r_n & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & 0 & r_n & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & r_n & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & r_n \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

and

$$U_n = \begin{bmatrix} \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} \\ \frac{1}{\sqrt{n}} & \alpha_n & \beta_n & \beta_n & \beta_n & \beta_n & \cdots & \beta_n & \beta_n \\ \frac{1}{\sqrt{n}} & \beta_n & \alpha_n & \beta_n & \beta_n & \beta_n & \cdots & \beta_n & \beta_n \\ \frac{1}{\sqrt{n}} & \beta_n & \beta_n & \alpha_n & \beta_n & \beta_n & \cdots & \beta_n & \beta_n \\ \frac{1}{\sqrt{n}} & \beta_n & \beta_n & \beta_n & \alpha_n & \beta_n & \cdots & \beta_n & \beta_n \\ \frac{1}{\sqrt{n}} & \beta_n & \beta_n & \beta_n & \beta_n & \alpha_n & \cdots & \beta_n & \beta_n \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{1}{\sqrt{n}} & \beta_n & \beta_n & \beta_n & \beta_n & \beta_n & \cdots & \alpha_n & \beta_n \\ \frac{1}{\sqrt{n}} & \beta_n & \beta_n & \beta_n & \beta_n & \beta_n & \cdots & \beta_n & \alpha_n \end{bmatrix}.$$

To show that U_n is a unitary, the following calculations can be easily checked:

1. $\frac{1}{n} + \alpha_n^2 + (n-2) \cdot \beta_n^2 = 1.$

2. $\frac{1}{\sqrt{n}} + \alpha_n + (n-2) \cdot \beta_n = 0.$

3. $\frac{1}{n} + 2 \cdot \alpha_n \cdot \beta_n + (n-3) \cdot \beta_n^2 = 0.$

Also, to see that $A_n = U_n T_n U_n^*$, the following calculations are again easy to check:

1.
$$A_n(1,1) = a_n = \langle u_1, Tu_1 \rangle = \frac{1}{n} + \frac{1}{n} \cdot (n-2) \cdot r_n.$$

2.
$$A_n(1,2) = b_n = \langle u_1, Tu_2 \rangle = \frac{1}{n} + \frac{1}{\sqrt{n}} \cdot (n-2) \cdot r_n \cdot \beta_n.$$

3.
$$A_n(n,1) = b_n = \langle u_n, Tu_1 \rangle = \frac{1}{n} + \frac{1}{\sqrt{n}} \cdot (n-2) \cdot r_n \cdot \beta_n.$$

4. For $3 \leq j \leq n$,
$$A_n(1,j) = 0 = \langle u_1, Tu_j \rangle = \frac{1}{n} + \frac{1}{\sqrt{n}} \cdot \alpha_n \cdot r_n + (n-3) \cdot \frac{1}{\sqrt{n}} \cdot \beta_n \cdot r_n.$$

5. For $2 \leq i \leq n-1$,
$$A_n(i,1) = 0 = \langle u_i, Tu_1 \rangle = \frac{1}{n} + \frac{1}{\sqrt{n}} \cdot \alpha_n \cdot r_n + \frac{1}{\sqrt{n}} \cdot (n-3) \cdot \beta_n \cdot r_n.$$

6. For $2 \leq i \leq n-1$,
$$A_n(i,2) = c_n = \langle u_i, Tu_2 \rangle = \frac{1}{n} + \alpha_n \cdot \beta_n \cdot r_n + (n-3) \cdot \beta_n^2 \cdot r_n.$$

7. For $3 \leq j \leq n$,
$$A_n(n,j) = c_n = \langle u_n, Tu_j \rangle = \frac{1}{n} + \alpha_n \cdot \beta_n \cdot r_n + (n-3) \cdot \beta_n^2 \cdot r_n.$$

8. For $2 \leq i \leq n-1$,
$$A_n(i,i+1) = d_n = \langle u_i, Tu_{i+1} \rangle = \frac{1}{n} + \alpha_n^2 \cdot r_n + (n-3) \cdot \beta_n^2 \cdot r_n.$$

9. For $2 \leq i \leq n-2$, $3 \leq j \leq n$, $i+1 \neq j$,
$$A_n(i,j) = e_n = \langle u_i, Tu_j \rangle = \frac{1}{n} + 2 \cdot \alpha_n \cdot \beta_n \cdot r_n + (n-4) \cdot \beta_n^2 \cdot r_n.$$

10.
$$A_n(n,2) = f_n = \langle u_n, Tu_2 \rangle = \frac{1}{n} + (n-2) \cdot r_n \cdot \beta_n^2.$$

We thus get a Schur decomposition for A_n, and since the diagonal of T_n contains only zeros except the trivial eigenvalue 1, we get that all nontrivial eigenvalues of A_n are zero. This completes the proof of (2).

If we let the set $S = \{1\}$, then we get that

$$\phi(A_n) \leq \phi_S(A_n) = b_n < \frac{1}{\sqrt{n}}.$$

Further, since T_n can be written as $\Pi_n D_n$, where $D_n(1,1) = 1$, $D_n(i,i) = r_n$ for $i = 2$ to $n-1$, and $D_n(n,n) = 0$ for some permutation Π_n, we get that $A_n = (U_n \Pi_n) D_n U_n^*$ which gives a singular value decomposition for A_n since $U_n \Pi_n$ and U_n^* are unitaries. Thus, A_n has exactly one singular value that is 1, $n-2$ singular values that are r_n, and one singular value that is 0. Thus, from Lemma 7, we get that

$$\phi(A) \geq \frac{1-r_n}{2} = \frac{1}{2 \cdot \left(\sqrt{n}+2\right)} \geq \frac{1}{6\sqrt{n}}$$

and this completes the proof of (3). □

C Proof of the upper bound on ϕ in Theorem 5

Proof. Let R be the nonnegative matrix as described, and let u and v be the eigenvectors corresponding to the eigenvalue $\lambda_1 = 1$. Let $\kappa = \min_i u_i \cdot v_i$. If $\kappa = 0$, $\phi(R) = 0$ by Definition 11 and the upper bound on $\phi(R)$ in the lemma trivially holds. If $\kappa > 0$, then both u and v are positive, and we normalize them so that $\langle u, v \rangle = 1$, and define

$$A = D_u^{\frac{1}{2}} D_v^{-\frac{1}{2}} R D_u^{-\frac{1}{2}} D_v^{\frac{1}{2}}$$

and

$$w = D_u^{\frac{1}{2}} D_v^{\frac{1}{2}} \mathbf{1},$$

where we use positive square roots for the entries in the diagonal matrices. Further, from Lemma 20, the edge expansion and eigenvalues of R and A are the same, and $\|A\|_2 = 1$. We will show the bound for A and the bound for R will follow. Since A has w as both the left and right eigenvector for eigenvalue 1, so does $M = \frac{A+A^T}{2}$.

As explained before Definition 11 for edge expansion of general nonnegative matrices, $D_w A D_w$ is Eulerian, since $D_w A D_w \mathbf{1} = D_w A w = D_w w = D_w^2 \mathbf{1} = D_w w = D_w A^T w = D_w A^T D_w \mathbf{1}$. Thus, for any S,

$$\langle \mathbf{1}_S, D_w A D_w \mathbf{1}_{\overline{S}} \rangle = \langle \mathbf{1}_{\overline{S}}, D_w A D_w \mathbf{1}_S \rangle = \langle \mathbf{1}_S, D_w A^T D_w \mathbf{1}_{\overline{S}} \rangle,$$

and thus for any set S for which $\sum_{i \in S} w_i^2 \leq \frac{1}{2}$,

$$\phi_S(A) = \frac{\langle \mathbf{1}_S, D_w A D_w \mathbf{1}_{\overline{S}} \rangle}{\langle \mathbf{1}_S, D_w A D_w \mathbf{1} \rangle} = \frac{1}{2} \cdot \frac{\langle \mathbf{1}_S, D_w A D_w \mathbf{1}_{\overline{S}} \rangle + \langle \mathbf{1}_S, D_w A^T D_w \mathbf{1}_{\overline{S}} \rangle}{\langle \mathbf{1}_S, D_w A D_w \mathbf{1} \rangle} = \phi_S(M)$$

and thus

$$\phi(A) = \phi(M). \tag{8}$$

For any matrix H, let

$$R_H(x) = \frac{\langle x, Hx \rangle}{\langle x, x \rangle}.$$

For every $x \in \mathbb{C}^n$,

$$\mathrm{Re} R_A(x) = R_M(x), \tag{9}$$

since A and $\langle x, x \rangle$ are nonnegative and we can write

$$R_M(x) = \frac{1}{2} \frac{\langle x, Ax \rangle}{\langle x, x \rangle} + \frac{1}{2} \frac{\langle x, A^* x \rangle}{\langle x, x \rangle} = \frac{1}{2} \frac{\langle x, Ax \rangle}{\langle x, x \rangle} + \frac{1}{2} \frac{\langle Ax, x \rangle}{\langle x, x \rangle} = \frac{1}{2} \frac{\langle x, Ax \rangle}{\langle x, x \rangle} + \frac{1}{2} \frac{\langle x, Ax \rangle^*}{\langle x, x \rangle} = \mathrm{Re} R_A(x).$$

Also,

$$\mathrm{Re}\lambda_2(A) \le \lambda_2(M). \tag{10}$$

To see this, first note that if $\lambda_2(A) = 1$, since there exists some positive u and v for PF eigenvalue 1, then from Lemma 12, the underlying graph of A has multiply strongly connected components each with eigenvalue 1, and so does M, and thus $\lambda_2(M) = 1$. If $\lambda_2(A) \neq 1$, then let v be the eigenvector corresponding to $\lambda_2(A)$. Then since $Aw = w$, we have that

$$Av = \lambda_2(A)v \Rightarrow \langle w, Av\rangle = \langle w, \lambda_2(A)v\rangle \Leftrightarrow \langle A^T w, v\rangle = \lambda_2(A)\langle w, v\rangle \Leftrightarrow (1-\lambda_2(A))\langle w, v\rangle = 0$$

which implies that $v \perp w$. Thus, we have that

$$\mathrm{Re}\lambda_2(A) = \mathrm{Re}\frac{\langle v, Av\rangle}{\langle v, v\rangle} = \frac{\langle v, Mv\rangle}{\langle v, v\rangle} \le \max_{u\perp w} \frac{\langle u, Mu\rangle}{\langle u, u\rangle} = \lambda_2(M)$$

where the second equality uses equation 9, and the last equality follows from the variational characterization of eigenvalues stated in the Preliminaries (Appendix 2). Thus, using equation 10, equation 8 and Cheeger's inequality for M (Theorem 15), we get

$$\phi(A) = \phi(M) \le \sqrt{2\cdot(1-\lambda_2(M))} \le \sqrt{2\cdot(1-\mathrm{Re}\lambda_2(A))}$$

as required. □

D Proofs for the lower bound on ϕ in Theorem 5

D.1 Proof of Lemma 17

Proof. For any cut S and non negative R as defined, let $G_R = D_u R D_v$, and note that G_R is Eulerian, since $G_R \mathbf{1} = G_R^T \mathbf{1}$. Let

$$\begin{aligned}\gamma_S(R) &= \langle \mathbf{1}_{\overline{S}}, G_R \mathbf{1}_S\rangle + \langle \mathbf{1}_S, G_R \mathbf{1}_{\overline{S}}\rangle \\ &= \langle \mathbf{1}_{\overline{S}}, D_u R D_v \mathbf{1}_S\rangle + \langle \mathbf{1}_S, D_u R D_v \mathbf{1}_{\overline{S}}\rangle\end{aligned} \tag{11}$$

for any matrix R with left and right eigenvectors u and v. We will show that if R and B are non negative matrices that have the same left and right eigenvectors u and v for eigenvalue 1, then for every cut S,

$$\gamma_S(RB) \le \gamma_S(R+B) = \gamma_S(R) + \gamma_S(B).$$

Fix any cut S. Assume $R = \begin{bmatrix} P & Q \\ H & V \end{bmatrix}$ and $B = \begin{bmatrix} X & Y \\ Z & W \end{bmatrix}$ naturally divided based on cut S. For any vector u, let D_u be the diagonal matrix with u on the diagonal. Since $Rv = v, R^T u = u, Bv = v, B^T u = u$, we have

$$P^T D_{u_S}\mathbf{1} + H^T D_{u_{\overline{S}}}\mathbf{1} = D_{u_S}\mathbf{1}, \tag{12}$$

$$Z D_{v_S}\mathbf{1} + W D_{v_{\overline{S}}}\mathbf{1} = D_{v_{\overline{S}}}\mathbf{1}, \tag{13}$$

$$X D_{v_S}\mathbf{1} + Y D_{v_{\overline{S}}}\mathbf{1} = D_{v_S}\mathbf{1}, \tag{14}$$

$$Q^T D_{u_S}\mathbf{1} + V^T D_{u_{\overline{S}}}\mathbf{1} = D_{u_{\overline{S}}}\mathbf{1}, \tag{15}$$

where u is divided into u_S and $u_{\overline{S}}$ and v into v_S and $v_{\overline{S}}$ naturally based on the cut S. Further, in the equations above and in what follows, the vector $\mathbf{1}$ is the all 1's vector with dimension either $|S|$ or $|\overline{S}|$ which should be clear from the context of the equations, and we avoid using different vectors to keep the notation simpler. Then we have from equation 11,

$$\begin{aligned}\gamma_S(R) &= \langle \mathbf{1}_{\overline{S}}, D_u R D_v \mathbf{1}_S \rangle + \langle \mathbf{1}_S, D_u R D_v \mathbf{1}_{\overline{S}} \rangle \\ &= \langle \mathbf{1}, D_{u_S} Q D_{v_{\overline{S}}} \mathbf{1} \rangle + \langle \mathbf{1}, D_{u_{\overline{S}}} H D_{v_S} \mathbf{1} \rangle\end{aligned}$$

and similarly

$$\gamma_S(B) = \langle \mathbf{1}, D_{u_S} Y D_{v_{\overline{S}}} \mathbf{1} \rangle + \langle \mathbf{1}, D_{u_{\overline{S}}} Z D_{v_S} \mathbf{1} \rangle.$$

The matrix RB also has u and v as the left and right eigenvectors for eigenvalue 1 respectively, and thus,

$$\begin{aligned}\gamma_S(RB) &= \langle \mathbf{1}, D_{u_S} PY D_{v_{\overline{S}}} \mathbf{1} \rangle + \langle \mathbf{1}, D_{u_S} QW D_{v_{\overline{S}}} \mathbf{1} \rangle + \langle \mathbf{1}, D_{u_{\overline{S}}} HX D_{v_S} \mathbf{1} \rangle + \langle \mathbf{1}, D_{u_{\overline{S}}} VZ D_{v_S} \mathbf{1} \rangle \\ &= \langle P^T D_{u_S} \mathbf{1}, Y D_{v_{\overline{S}}} \mathbf{1} \rangle + \langle \mathbf{1}, D_{u_S} QW D_{v_{\overline{S}}} \mathbf{1} \rangle + \langle \mathbf{1}, D_{u_{\overline{S}}} HX D_{v_S} \mathbf{1} \rangle + \langle V^T D_{u_{\overline{S}}} \mathbf{1}, Z D_{v_S} \mathbf{1} \rangle \\ &= \langle D_{u_S} \mathbf{1} - H^T D_{u_{\overline{S}}} \mathbf{1}, Y D_{v_{\overline{S}}} \mathbf{1} \rangle + \langle \mathbf{1}, D_{u_S} Q (D_{v_{\overline{S}}} \mathbf{1} - Z D_{v_S} \mathbf{1}) \rangle \\ &\quad + \langle \mathbf{1}, D_{u_{\overline{S}}} H (D_{v_S} \mathbf{1} - Y D_{v_{\overline{S}}} \mathbf{1} \rangle + \langle D_{u_{\overline{S}}} \mathbf{1} - Q^T D_{u_S} \mathbf{1}, Z D_{v_S} \mathbf{1} \rangle \\ &\qquad \text{[from equations 12, 13, 14, 15 above]} \\ &= \langle D_{u_S} \mathbf{1}, Y D_{v_{\overline{S}}} \mathbf{1} \rangle + \langle \mathbf{1}, D_{u_S} Q D_{v_{\overline{S}}} \mathbf{1} \rangle - \langle H^T D_{u_{\overline{S}}} \mathbf{1}, Y D_{v_{\overline{S}}} \mathbf{1} \rangle - \langle \mathbf{1}, D_{u_S} QZ D_{v_S} \mathbf{1} \rangle \\ &\quad + \langle \mathbf{1}, D_{u_{\overline{S}}} H D_{v_S} \mathbf{1} \rangle + \langle D_{u_{\overline{S}}} \mathbf{1}, Z D_{v_S} \mathbf{1} \rangle - \langle \mathbf{1}, D_{u_{\overline{S}}} HY D_{v_{\overline{S}}} \mathbf{1} \rangle - \langle Q^T D_{u_S} \mathbf{1}, Z D_{v_S} \mathbf{1} \rangle \\ &\le \langle D_{u_S} \mathbf{1}, Y D_{v_{\overline{S}}} \mathbf{1} \rangle + \langle \mathbf{1}, D_{u_S} Q D_{v_{\overline{S}}} \mathbf{1} \rangle + \langle \mathbf{1}, D_{u_{\overline{S}}} H D_{v_S} \mathbf{1} \rangle + \langle D_{u_{\overline{S}}} \mathbf{1}, Z D_{v_S} \mathbf{1} \rangle \\ &\qquad \text{[since every entry of the matrices is nonnegative, and thus each of the terms above]} \\ &= \gamma_S(R) + \gamma_S(B) \qquad (16)\end{aligned}$$

Thus, noting that R^k has u and v as left and right eigenvectors for any k, we inductively get using inequality 16 that

$$\gamma_S(R^k) \le \gamma_S(R) + \gamma_S(R^{k-1}) \le \gamma_S(R) + (k-1) \cdot \gamma_S(R) = k \cdot \gamma_S(R).$$

Note that for any fixed set S, the denominator in the definition of $\phi_S(R)$ is independent of the matrix R and depends only on the set S, and since the numerator is exactly $\gamma_S(R)/2$, i.e.

$$\phi_S(R) = \frac{\gamma_S(R)}{2 \cdot \sum_{i \in S} u_i \cdot v_i}$$

we get by letting S be the set that minimizes $\phi(R)$, that

$$\phi(R^k) \le k \cdot \phi(R).$$

□

D.2 Proof of Lemma 18

Proof. Let $g_r(k)$ denote the maximum of the absolute value of entries at distance r from the diagonal in T^k, where the diagonal is at distance 0 from the diagonal, the off-diagonal is at distance 1 from the diagonal and so on. More formally,

$$g_r(k) = \max_i |T^k(i, i+r)|.$$

We will inductively show that for $\alpha \leq 1$, and $r \geq 1$,

$$g_r(k) \leq \binom{k+r}{r} \cdot \alpha^{k-r} \cdot \sigma^r, \tag{17}$$

where $\sigma = \|T\|_2$. First note that for $r = 0$, since T is upper triangular, the diagonal of T^k is α^k, and thus the hypothesis holds for $r = 0$ and all $k \geq 1$. Further, for $k = 1$, if $r = 0$, then $g_0(1) \leq \alpha$ and if $r \geq 1$, then $g_r(1) \leq \|T\|_2 \leq \sigma$ and the inductive hypothesis holds also in this case, since $r \geq k$ and $\alpha^{k-r} \geq 1$. For the inductive step, assume that for all $r \geq 1$ and all $j \leq k-1$, $g_r(j) \leq \binom{j+r}{r} \cdot \alpha^{j-r} \cdot \sigma^r$. We will show the calculation for $g_r(k)$.

Since $|a+b| \leq |a| + |b|$,

$$\begin{aligned} g_r(k) &\leq \sum_{i=0}^{r} g_{r-i}(1) \cdot g_i(k-1) \\ &= g_0(1) \cdot g_r(k-1) + \sum_{i=0}^{r-1} g_{r-i}(1) \cdot g_i(k-1). \end{aligned}$$

The first term can be written as,

$$\begin{aligned} g_0(1) \cdot g_r(k-1) &= \alpha \cdot \binom{k-1+r}{r} \cdot \alpha^{k-1-r} \cdot \sigma^r \\ &\qquad [\text{using that } g_0(1) \leq \alpha \text{ and the inductive hypothesis for the second term}] \\ &\leq \alpha^{k-r} \cdot \sigma^r \cdot \binom{k+r}{r} \cdot \frac{\binom{k-1+r}{r}}{\binom{k+r}{r}} \\ &\leq \frac{k}{k+r} \cdot \binom{k+r}{r} \cdot \alpha^{k-r} \cdot \sigma^r \end{aligned} \tag{18}$$

and the second term as

$$
\begin{aligned}
g_r(k) &\le \sum_{i=0}^{r-1} g_{r-i}(1)\cdot g_i(k-1) \\
&\le \sigma \cdot \sum_{i=0}^{r-1} \binom{k-1+i}{i}\cdot \alpha^{k-1-i}\cdot \sigma^i \\
&\quad [\text{using } g_{r-i}(1)\le\sigma \text{ and the inductive hypothesis for the second term}] \\
&\le \sigma^r\cdot\alpha^{k-r}\cdot\binom{k+r}{r}\cdot\sum_{i=0}^{r-1}\cdot\frac{\binom{k-1+i}{i}}{\binom{k+r}{r}}\cdot\alpha^{r-1-i} \\
&\quad [\text{using } \sigma^i\le\sigma^{r-1} \text{ since } \sigma\ge 1 \text{ and } i\le r-1] \\
&\le \sigma^r\cdot\alpha^{k-r}\cdot\binom{k+r}{r}\cdot\sum_{i=0}^{r-1}\cdot\frac{\binom{k-1+i}{i}}{\binom{k+r}{r}} \\
&\quad [\text{using } \alpha\le 1 \text{ and } i\le r-1]
\end{aligned}
$$

We will now show that the quantity inside the summation is at most $\frac{r}{k+r}$. Inductively, for $r = 1$, the statement is true, and assume that for any other r,

$$
\sum_{i=0}^{r-1}\frac{\binom{k-1+i}{i}}{\binom{k+r}{r}} \le \frac{r}{k+r}.
$$

Then we have

$$
\begin{aligned}
\sum_{i=0}^{r}\frac{\binom{k-1+i}{i}}{\binom{k+r+1}{r+1}} &= \frac{\binom{k+r}{r}}{\binom{k+r+1}{r+1}}\cdot\sum_{i=0}^{r-1}\frac{\binom{k-1+i}{i}}{\binom{k+r}{r}} + \frac{\binom{k-1+r}{k-1}}{\binom{k+r+1}{k}} \\
&\le \frac{r+1}{k+r+1}\cdot\frac{r}{k+r} + \frac{(r+1)\cdot k}{(k+r+1)\cdot(k+r)} \\
&= \frac{r+1}{k+r+1}
\end{aligned}
$$

Thus we get that the second term is at most $\sigma^r\cdot\alpha^{k-r}\cdot\binom{k+r}{r}\cdot\frac{r}{k+r}$, and combining it with the first term (equation 18), it completes the inductive hypothesis. Noting that the operator norm is at most the Frobenius norm, and since $g_r(k)$ is increasing in r and the maximum value of r is n, we get using equation 17,

$$
\begin{aligned}
\|T^k\|_2 &\le \sqrt{\sum_{i,j}|T^k(i,j)|^2} \\
&\le n\cdot\sigma^n\cdot\alpha^{k-n}\cdot\binom{k+n}{n}
\end{aligned}
$$

as required. □

D.3 Proof of Lemma 19

Proof. Let $X = T^{k_1}$, where $k_1 = \frac{c_1}{1-\alpha}$ for $\alpha < 1$. Then $\|X\|_2 \le \|T\|_2^{k_1} \le 1$, and for every i,

$$|X_{i,i}| \le |T_{i,i}|^{k_1} \le |\lambda_m|^{c_1/(1-\alpha)} \le e^{-c_1}.$$

Using Lemma 18 for X with $\sigma = 1$ and $\beta = e^{-c_1}$, we get that for $k_2 = c_2 \cdot n$,

$$\begin{aligned}
\|X^{k_2}\| &\le n \cdot \binom{k_2+n}{n} \cdot e^{-c_1(k_2-n)} \\
&\le n \cdot e^n \cdot (c_2+1)^n \cdot e^{-c_1(c_2-1)n} \\
&\qquad [\text{using } \binom{a}{b} \le \left(\frac{ea}{b}\right)^b] \\
&= \exp\left(n \cdot \left(\frac{\ln n}{n} + \ln(c_2+1) + 1 + c_1 - c_1 c_2\right)\right)
\end{aligned}$$

and to have this quantity less than ϵ, we require

$$\begin{aligned}
&\exp\left(n \cdot \left(\frac{\ln n}{n} + \ln(c_2+1) + 1 + c_1 - c_1 c_2\right)\right) \le \epsilon \\
\Leftrightarrow &\left(\frac{1}{\epsilon}\right)^{\frac{1}{n}} \le \exp\left(-1 \cdot \left(\frac{\ln n}{n} + \ln(c_2+1) + 1 + c_1 - c_1 c_2\right)\right) \\
\Leftrightarrow &\frac{1}{n} \ln \frac{n}{\epsilon} + 1 + c_1 + \ln(c_2+1) \le c_1 c_2 \qquad (19)
\end{aligned}$$

and we set

$$c_1 = 1 + \frac{1}{2.51} \cdot \frac{1}{n} \ln\left(\frac{n}{\epsilon}\right)$$

and $c_2 = 3.51$ which always satisfies inequality 19. As a consequence, for

$$\begin{aligned}
k &= k_1 \cdot k_2 \\
&= \frac{c_1 \cdot c_2 \cdot n}{1-\alpha} \\
&= \frac{3.51 \cdot n + 1.385 \cdot \ln\left(\frac{n}{\epsilon}\right)}{1-\alpha}
\end{aligned}$$

we get,

$$\|T^k\|_2 \le \|X^{k_2}\| \le \epsilon$$

as required. ☐

D.4 Proof of Lemma 20

Proof. Let the matrix A be as defined, and let $w = D_u^{\frac{1}{2}} D_v^{\frac{1}{2}} \mathbf{1}$. Then it is easily checked that $Aw = w$ and $A^T w = w$. Further,

$$\langle w, w\rangle = \langle D_u^{\frac{1}{2}} D_v^{\frac{1}{2}} \mathbf{1}, D_u^{\frac{1}{2}} D_v^{\frac{1}{2}} \mathbf{1}\rangle = \langle D_u \mathbf{1}, D_v \mathbf{1}\rangle = \langle u, v\rangle = 1$$

where we used the fact that the matrices $D_u^{\frac{1}{2}}$ and $D_v^{\frac{1}{2}}$ are diagonal, and so they commute, and are unchanged by taking transposes. Let S be any set. The condition $\sum_{i\in S} u_i \cdot v_i \le \frac{1}{2}$ translates to $\sum_{i\in S} w_i^2 \le \frac{1}{2}$ since $u_i \cdot v_i = w_i^2$. Thus, for any set S for which $\sum_{i\in S} u_i \cdot v_i = \sum_{i\in S} w_i^2 \le \frac{1}{2}$,

$$\begin{aligned}
\phi_S(R) &= \frac{\langle \mathbf{1}_S, D_u R D_v \mathbf{1}_{\overline{S}}\rangle}{\langle \mathbf{1}_S, D_u D_v \mathbf{1}_S\rangle} \\
&= \frac{\langle \mathbf{1}_S, D_u D_u^{-\frac{1}{2}} D_v^{\frac{1}{2}} A D_u^{\frac{1}{2}} D_v^{-\frac{1}{2}} D_v \mathbf{1}_{\overline{S}}\rangle}{\langle \mathbf{1}_S, D_u^{\frac{1}{2}} D_v^{\frac{1}{2}} D_u^{\frac{1}{2}} D_v^{\frac{1}{2}} \mathbf{1}_S\rangle} \\
&= \frac{\langle \mathbf{1}_S, D_u^{\frac{1}{2}} D_v^{\frac{1}{2}} A D_u^{\frac{1}{2}} D_v^{\frac{1}{2}} \mathbf{1}_{\overline{S}}\rangle}{\langle \mathbf{1}_S, D_u^{\frac{1}{2}} D_v^{\frac{1}{2}} D_u^{\frac{1}{2}} D_v^{\frac{1}{2}} \mathbf{1}_S\rangle} \\
&= \frac{\langle \mathbf{1}_S, D_w A D_w \mathbf{1}_{\overline{S}}\rangle}{\langle \mathbf{1}_S, D_w^2 \mathbf{1}_S\rangle} \\
&= \phi_S(A)
\end{aligned}$$

and (1) holds. Further, since A is a similarity transform of R, all eigenvalues are preserved and (2) holds. For (3), consider the matrix $H = A^T A$. Since w is the positive left and right eigenvector for A, i.e. $Aw = w$ and $A^T w = w$, we have $Hw = w$. But since A was nonnegative, so is H, and since it has a positive eigenvector w for eigenvalue 1, by Perron-Frobenius (Theorem 8, part 2), H has PF eigenvalue 1. But $\lambda_i(H) = \sigma_i^2(A)$, where $\sigma_i(A)$ is the i'th largest singular value of A. Thus, we get $\sigma_1^2(A) = \lambda_1(H) = 1$, and thus $\|A\|_2 = 1$. □

D.5 Proof of Lemma 21

Proof. From the given R with positive left and right eigenvectors u and v for eigenvalue 1 as stated, let $A = D_u^{\frac{1}{2}} D_v^{-\frac{1}{2}} R D_u^{-\frac{1}{2}} D_v^{\frac{1}{2}}$ and $w = D_u^{\frac{1}{2}} D_v^{\frac{1}{2}} \mathbf{1}$ as in Lemma 20. Note that w is positive, and

$$\langle w, w\rangle = \langle D_u^{\frac{1}{2}} D_v^{\frac{1}{2}} \mathbf{1}, D_u^{\frac{1}{2}} D_v^{\frac{1}{2}} \mathbf{1}\rangle = \langle u, v\rangle = 1.$$

Further, $Aw = w$ and $A^T w = w$.

Let $A = w \cdot w^T + B$. Since $(w \cdot w^T)^2 = w \cdot w^T$ and $Bw \cdot w^T = w \cdot w^T B = 0$, we get

$$A^k = w \cdot w^T + B^k.$$

Let $B = UTU^*$ be the Schur decomposition of B, where the diagonal of T contains all but the stochastic eigenvalue of A, which is replaced by 0, since w is both the left and right eigenvector for eigenvalue 1 of A, and that space is removed in $w \cdot w^T$. Further, the maximum diagonal entry of T is at most $|\lambda_m|$ where λ_m is the nontrivial eigenvalue of A (or R) that is maximum in magnitude. Note that if $|\lambda_m| = 1$, then the lemma is trivially true since $\phi(R) \ge 0$, and thus assume that $|\lambda_m| < 1$. Since $w \cdot w^T B = Bw \cdot w^T = 0$ and $\|A\|_2 \le 1$ from Lemma 20, we have that $\|B\|_2 \le 1$.

Thus, using Lemma 19 (in fact, the last lines in the proof of Lemma 19 in Appendix D.3 above), for

$$k \ge \frac{3.51 \cdot n + 1.385 \cdot \ln\left(\frac{n}{\epsilon}\right)}{1 - |\lambda_m|},$$

we get that

$$\|B^k\|_2 = \|T^k\|_2 \leq \epsilon,$$

and for e_i being the vector with 1 at position i and zeros elsewhere, we get using Cauchy-Schwarz

$$|B^k(i,j)| = \left|\langle e_i, B^k e_j\rangle\right| \leq \|e_i\|_2 \|B\|_2 \|e_j\|_2 \leq \epsilon.$$

Thus, for any set S for which $\sum_{i\in S} w_i^2 \leq \frac{1}{2}$, let

$$\epsilon = c \cdot \kappa = c \cdot \min_i u_i \cdot v_i = c \cdot \min_i w_i^2 \leq c \cdot w_i \cdot w_j$$

for any i, j, where c is some constant to be set later. Then

$$\begin{aligned}
\phi_S(A^k) &= \frac{\langle \mathbf{1}_S, D_w A^k D_w \mathbf{1}_{\overline{S}}\rangle}{\langle \mathbf{1}_S, D_w D_w \mathbf{1}\rangle} \\
&= \frac{\sum_{i\in S, j\in\overline{S}} A^k(i,j) \cdot w_i \cdot w_j}{\sum_{i\in S} w_i^2} \\
&= \frac{\sum_{i\in S, j\in\overline{S}} \left(w \cdot w^T(i,j) + B^k(i,j)\right) \cdot w_i \cdot w_j}{\sum_{i\in S} w_i^2} \\
&\geq \frac{\sum_{i\in S, j\in\overline{S}} \left(w_i \cdot w_j - \epsilon\right) \cdot w_i \cdot w_j}{\sum_{i\in S} w_i^2} \\
&\geq (1-c) \frac{\sum_{i\in S} w_i^2 \sum_{j\in\overline{S}} w_j^2}{\sum_{i\in S} w_i^2} \\
&\geq \frac{1}{2}(1-c)
\end{aligned}$$

since $\sum_{i\in\overline{S}} w_i^2 \geq \frac{1}{2}$. Thus, we get that for

$$k \geq \frac{3.51n + 1.385 \cdot \ln\left(\frac{c\cdot n}{\kappa}\right)}{1-|\lambda_m|}$$

the edge expansion

$$\phi(A^k) \geq \frac{1}{2}(1-c),$$

and thus using Lemma 17, we get that

$$\phi(A) \geq \frac{1}{k} \cdot \phi(A^k) \geq \frac{1-c}{2} \cdot \frac{1-|\lambda_m|}{3.51 \cdot n + 1.385 \cdot \ln(c \cdot n) + 1.385 \cdot \ln\left(\frac{1}{\kappa}\right)}$$

and setting $c = \frac{1}{1.4\cdot e}$, and using $\ln(e \cdot x) \leq x$,we get that

$$\begin{aligned}
\phi(A) &\geq \frac{1-|\lambda_m|}{15 \cdot n + 2 \cdot \ln\left(\frac{1}{\kappa}\right)} \\
&\geq \frac{1}{15} \cdot \frac{1-|\lambda_m|}{n + \ln\left(\frac{1}{\kappa}\right)}
\end{aligned}$$

□

D.6 Proof of Lemma 22

Proof. Let R be a nonnegative matrix with PF eigenvalue 1, and let some eigenvalue $\lambda_r(R) = \alpha + i\beta$, and letting the gap $g = 1-\alpha$, we can rewrite

$$\lambda_r = (1-g) + i\beta$$

with $(1-g)^2 + \beta^2 \leq 1$, or

$$g^2 + \beta^2 \leq 2g. \tag{20}$$

Consider the lazy random walk $\tilde{R} = pI + (1-p)R$ for some $0 \leq p \leq 1$, the eigenvalue modifies as

$$\begin{aligned}\lambda_r(\tilde{R}) &= p + (1-p)(1-g) + i(1-p)\beta \\ &= 1 - g(1-p) + i\beta(1-p)\end{aligned}$$

and for any set S, since $\phi_S(I) = 0$, we have

$$\phi_S(\tilde{R}) = (1-p)\phi_S(R)$$

or

$$\phi(\tilde{R}) = (1-p)\phi(R). \tag{21}$$

Further,

$$\begin{aligned}1 - |\lambda_r(\tilde{R})| &= 1 - \sqrt{(1-g(1-p))^2 + (\beta(1-p))^2} \\ &= 1 - \sqrt{1 + (g^2+\beta^2)(1-p)^2 - 2g(1-p)} \\ &\geq 1 - \sqrt{1 + 2g(1-p)^2 - 2g(1-p)} \\ &\quad [\text{using inequality } 20] \\ &= 1 - \sqrt{1 - 2gp(1-p)} \\ &\geq 1 - e^{-gp(1-p)} \\ &\quad [\text{using } 1-x \leq e^{-x}] \\ &\geq 1 - (1 - \frac{1}{2}gp(1-p)) \\ &\quad [\text{using } e^{-x} \leq 1 - \frac{1}{2}x \text{ for } 0 \leq x \leq 1] \\ &= \frac{1}{2}gp(1-p). \end{aligned} \tag{22}$$

Thus, we have that,

$$\begin{aligned}(1-\text{Re}\lambda_2(R))\cdot p(1-p) &\leq (1-\text{Re}\lambda_m(R))\cdot p(1-p)\\ &\quad[\text{since both } \lambda_2 \text{ and } \lambda_m \text{ are nontrivial eigenvalues}]\\ &\leq 2\cdot(1-|\lambda_m(\tilde{R})|)\\ &\quad[\text{using inequality 22}]\\ &\leq 2\cdot 15\cdot\left(n+\ln\left(\frac{1}{\kappa}\right)\right)\cdot\phi(\tilde{R})\\ &\quad[\text{using Lemma 21 for } \tilde{R}]\\ &\leq 2\cdot 15\cdot\left(n+\ln\left(\frac{1}{\kappa}\right)\right)\cdot(1-p)\phi(R)\\ &\quad[\text{from equation 21}]\end{aligned}$$

and taking the limit as $p\to 1$, we get that

$$\frac{1}{30}\cdot\frac{1-\text{Re}\lambda_2(R)}{n+\ln\left(\frac{1}{\kappa}\right)}\leq\phi(A).$$

Further, if A is doubly stochastic, then $\kappa=\frac{1}{n}$, and using $\ln(en)\leq n$ and the last equations at the end of the proof of Lemma 19 in Appendix D.3 above, we get using the same calculations that

$$\frac{1}{35}\cdot\frac{1-\text{Re}\lambda_2(R)}{n}\leq\phi(A)$$

as required. □

E Proofs of bounds on Mixing Time

E.1 Proof of Lemma 24

Proof. Writing τ as shorthand for $\tau_\epsilon(A)$ and since $A=w\cdot w^T+B$ with $Bw=0$ and $B^Tw=0$, we have that $A^\tau=w\cdot w^T+B^\tau$. Let x be a nonnegative vector such that $\langle w,x\rangle=\|D_wx\|_1=1$. As discussed after Definition 23, this is sufficient for bounding mixing time for all x. Then we have

$$\|D_w(A^\tau-w\cdot w^T)x\|_1=\|D_wB^\tau x\|_1=\|D_wB^\tau D_w^{-1}y\|_1\leq\|D_w\|_1\|B^\tau\|_1\|D_w^{-1}\|_1\|y\|_1$$

where $y=D_wx$ and $\|y\|_1=1$. Further, since $\|w\|_2=1$ and $\kappa=\min_i w_i^2$, we have $\|D_w\|_1\leq 1$ and $\|D_w^{-1}\|_1\leq\frac{1}{\sqrt{\kappa}}$, and using these bounds to continue the inequalities above, we get

$$\|D_w(A^\tau-w\cdot w^T)x\|_1\leq\frac{1}{\sqrt{\kappa}}\|B^\tau\|_1\leq\frac{\sqrt{n}}{\sqrt{\kappa}}\|B^\tau\|_2\leq\frac{\sqrt{n}}{\sqrt{\kappa}}\|B\|_2^\tau\leq\frac{\sqrt{n}}{\sqrt{\kappa}}(\sigma_2(A))^\tau=\frac{\sqrt{n}}{\sqrt{\kappa}}\left(\sigma_2^c(A)\right)^{\frac{\tau}{c}}\leq\epsilon$$

where the second inequality is Cauchy-Schwarz, the fourth inequality used $\|B\|_2\leq\sigma_2(A)$ as shown in the proof (in Appendix A) of Lemma 7, and the last inequality was obtained by setting $\tau=\tau_\epsilon(A)=\frac{c\cdot\ln\left(\frac{\sqrt{n}}{\sqrt{\kappa}\cdot\epsilon}\right)}{1-\sigma_2^c(A)}$. □

E.2 Proof of Lemma 25

Proof. We first show that the mixing time of R and $A = D_u^{\frac{1}{2}} D_v^{-\frac{1}{2}} R D_u^{-\frac{1}{2}} D_v^{\frac{1}{2}}$ are the same. Let $w = D_u^{\frac{1}{2}} D_v^{\frac{1}{2}} \mathbf{1}$, the left and right eigenvector of A for eigenvalue 1. We will show that for every x for which $\|D_u x\|_1 = 1$, there exists some y with $\|D_w y\|_1 = 1$ such that

$$\|D_u(R^\tau - v \cdot u^T)x\|_1 = \|D_w(A^\tau - w \cdot w^T)y\|_1$$

which would imply that $\tau_\epsilon(R) = \tau_\epsilon(A)$ by definition.

Let x be a vector with $\|D_u x\|_1 = 1$ and let $y = D_u^{\frac{1}{2}} D_v^{-\frac{1}{2}} x$. Then since $D_w = D_u^{\frac{1}{2}} D_v^{\frac{1}{2}}$, we get $\|D_w y\|_1 = \|D_u x\|_1 = 1$. Let $R = v \cdot u^T + B_R$ and $A = w \cdot w^T + B_A$ where $B_A = D_u^{\frac{1}{2}} D_v^{-\frac{1}{2}} B_R D_u^{-\frac{1}{2}} D_v^{\frac{1}{2}}$ and $B_A^\tau = D_u^{\frac{1}{2}} D_v^{-\frac{1}{2}} B_R^\tau D_u^{-\frac{1}{2}} D_v^{\frac{1}{2}}$. Then

$$\begin{aligned}
D_u(R^\tau - v \cdot u^T)x &= D_u B_R^\tau x \\
&= D_u(D_u^{-\frac{1}{2}} D_v^{\frac{1}{2}} B_A^\tau D_u^{\frac{1}{2}} D_v^{-\frac{1}{2}})x \\
&= D_w B_A^\tau y \\
&= D_w(A^\tau - w \cdot w^T)y
\end{aligned}$$

as required. Further, from Lemma 20, we have that $\phi(R) = \phi(A)$. Thus, we show the bound for $\tau_\epsilon(A)$ and $\phi(A)$, and the bound for R will follow. Note that if R is $\frac{1}{2}$-lazy, then A is also $\frac{1}{2}$-lazy, since if $R = \frac{1}{2}I + \frac{1}{2}C$ where C is nonnegative, then

$$A = \frac{1}{2} D_u^{\frac{1}{2}} D_v^{-\frac{1}{2}} I D_u^{-\frac{1}{2}} D_v^{\frac{1}{2}} + D_u^{\frac{1}{2}} D_v^{-\frac{1}{2}} C D_u^{-\frac{1}{2}} D_v^{\frac{1}{2}} = \frac{1}{2} I + D_u^{\frac{1}{2}} D_v^{-\frac{1}{2}} C D_u^{-\frac{1}{2}} D_v^{\frac{1}{2}}.$$

We first lower bound $\tau_\epsilon(A)$ using ϕ, showing the lower bound on mixing time in terms of the edge expansion. We will show the bound for nonnegative vectors x, and by Definition 23 and the discussion after, it will hold for all x. By definition of mixing time, we have that for any nonnegative x such that $\|D_w x\|_1 = \langle w, x\rangle = 1$, since $A^\tau = w \cdot w^T + B^\tau$,

$$\|D_w(A^\tau - w \cdot w^T)x\|_1 = \|D_w B^\tau x\|_1 \le \epsilon$$

and letting $y = D_w x$, we get that for any nonnegative y with $\|y\|_1 = 1$, we have

$$\|D_w B^\tau D_w^{-1} y\|_1 \le \epsilon.$$

Plugging the standard basis vectors for i, we get that for every i,

$$\sum_j \left| \frac{1}{w(i)} \cdot B^\tau(j,i) \cdot w(j) \right| = \frac{1}{w(i)} \cdot \sum_j |B^\tau(j,i)| \cdot w(j) \le \epsilon.$$

Thus, for any set S,

$$\sum_{i\in S} w(i)^2 \cdot \frac{1}{w(i)} \cdot \sum_j |B^\tau(j,i)| \cdot w(j) = \sum_{i\in S}\sum_j w(i) \cdot |B^\tau(j,i)| \cdot w(j) \le \sum_{i\in S} w(i)^2 \cdot \epsilon. \tag{23}$$

Thus, for any set S for which $\sum_{i\in S} w_i^2 \leq \frac{1}{2}$,

$$\begin{aligned}
\phi_S(A^\tau) &= \frac{\langle \mathbf{1}_S, D_w A^\tau D_w \mathbf{1}_{\overline{S}}\rangle}{\langle \mathbf{1}_S, D_w^2 \mathbf{1}\rangle} = \frac{\langle \mathbf{1}_{\overline{S}}, D_w A^\tau D_w \mathbf{1}_S\rangle}{\langle \mathbf{1}_S, D_w^2 \mathbf{1}\rangle} \\
&\quad [\text{since } D_w A^\tau D_w \text{ is Eulerian, i.e. } D_w A^\tau D_w \mathbf{1} = D_w (A^\tau)^T D_w \mathbf{1}] \\
&= \frac{\sum_{i\in S}\sum_{j\in\overline{S}} A^\tau(j,i)\cdot w(i)\cdot w(j)}{\sum_{i\in S} w_i^2} \\
&= \frac{\sum_{i\in S}\sum_{j\in\overline{S}} w(j)\cdot w(i)\cdot w(i)\cdot w(j) + \sum_{i\in S}\sum_{j\in\overline{S}} B^\tau(j,i)\cdot w(i)\cdot w(j)}{\sum_{i\in S} w_i^2} \\
&\geq \sum_{j\in\overline{S}} w_j^2 - \frac{\sum_{i\in S}\sum_{j\in\overline{S}} |B^\tau(j,i)|\cdot w(i)\cdot w(j)}{\sum_{i\in S} w_i^2} \\
&\geq \sum_{j\in\overline{S}} w_j^2 - \frac{\sum_{i\in S}\sum_{j} |B^\tau(j,i)|\cdot w(i)\cdot w(j)}{\sum_{i\in S} w_i^2} \\
&\geq \frac{1}{2} - \epsilon \\
&\quad \left[\text{since } \sum_{i\in S} w_i^2 \leq \frac{1}{2} \text{ and } \sum_i w_i^2 = 1, \text{ and the second term follows from equation 23}\right]
\end{aligned}$$

and thus

$$\phi(A^\tau) \geq \frac{1}{2} - \epsilon,$$

and using Lemma 17, we obtain

$$\phi(A^\tau) \leq \tau \cdot \phi(A),$$

or

$$\frac{\frac{1}{2}-\epsilon}{\phi(A)} \leq \tau_\epsilon(A).$$

We now upper bound $\tau_\epsilon(A)$ in terms of $\phi(A)$. From Lemma 24 for $c = 2$, we have that

$$\tau_\epsilon(A) \leq \frac{2\cdot \ln\left(\frac{n}{\kappa\cdot\epsilon}\right)}{1-\sigma_2^2(A)} = \frac{2\cdot \ln\left(\frac{n}{\kappa\cdot\epsilon}\right)}{1-\lambda_2(AA^T)}. \tag{24}$$

The continuing argument is straightforward, and was shown in Fill [Fil91] (albeit for row-stochastic matrices), and we reproduce it here for completeness. Since A is $\frac{1}{2}$-lazy, we have that $2A-I$ is nonnegative, has PF eigenvalue 1, and has the same left and right eigenvector w for eigenvalue 1 implying that the PF eigenvalue is 1 by Perron-Frobenius (Theorem 8, part 2), and also that its largest singular value is 1 from Lemma 20. Further, $\frac{1}{2}(A+A^T)$ also has the same properties. Thus, for any x,

$$\begin{aligned}
AA^T &= \frac{A+A^T}{2} + \frac{(2A-I)(2A^T-I)}{4} - \frac{I}{4} \\
\Rightarrow \langle x, AA^T x\rangle &\leq \langle x, \frac{A+A^T}{2}x\rangle + \|x\|_2 \frac{\|(2A-I)\|_2\|(2A^T-I)\|_2}{4}\|x\|_2 - \frac{\|x\|_2^2}{4} \leq \langle x, \frac{A+A^T}{2}x\rangle \\
\Rightarrow \max_{x\perp w}\langle x, AA^T x\rangle &\leq \max_{x\perp w}\langle x, \frac{A+A^T}{2}x\rangle.
\end{aligned}$$

where the last implication followed by the variational characterization of eigenvalues since AA^T and $\frac{1}{2}(A+A^T)$ are symmetric, and thus

$$\lambda_2(AA^T) \leq \lambda_2\left(\frac{A+A^T}{2}\right)$$

which from equation 24, gives

$$\tau_\epsilon(A) \leq \frac{2 \cdot \ln\left(\frac{n}{\kappa \cdot \epsilon}\right)}{1-\sigma_2^2(A)} = \frac{2 \cdot \ln\left(\frac{n}{\kappa \cdot \epsilon}\right)}{1-\lambda_2(AA^T)} \leq \frac{2 \cdot \ln\left(\frac{n}{\kappa \cdot \epsilon}\right)}{1-\lambda_2\left(\frac{A+A^T}{2}\right)} \leq \frac{4 \cdot \ln\left(\frac{n}{\kappa \cdot \epsilon}\right)}{\phi^2\left(\frac{A+A^T}{2}\right)} = \frac{4 \cdot \ln\left(\frac{n}{\kappa \cdot \epsilon}\right)}{\phi^2(A)}$$

where the second last inequality follows from Cheeger's inequality 15 for the symmetric matrix $(A+A^T)/2$. □

E.3 Proof of Lemma 26

Proof. Since the mixing time (as shown in the proof of Lemma 25 in Appendix E.2), eigenvalues, edge expansion, and value of κ for R and $A = D_u^{\frac{1}{2}} D_v^{-\frac{1}{2}} R D_u^{-\frac{1}{2}} D_v^{\frac{1}{2}}$ are the same, we provide the bounds for A and the bounds for R follow. We also restrict to nonnegative vectors, the bound for general vectors follows by the triangle inequality as discussed after Definition 23.

The lower bound on $\tau_\epsilon(R)$ follows by combining the lower bound on $\tau_\epsilon(R)$ in terms of $\phi(R)$ in Lemma 25, and the upper bound on $\phi(R)$ in terms of $1-\mathrm{Re}\lambda_2(R)$ in Theorem 5. For the upper bound on $\tau_\epsilon(R)$, similar to the proof of Lemma 24, we have for any nonnegative vector x with $\langle w, x\rangle = 1$, for $A = w \cdot w^T + B$,

$$\|D_w(A^\tau - w \cdot w^T)x\|_1 \leq \sqrt{\frac{n}{\kappa}} \cdot \|B^\tau\|_2$$

and having

$$\|B^\tau\|_2 \leq \frac{\epsilon\sqrt{\kappa}}{\sqrt{n}}$$

is sufficient. Let T be the triangular matrix in the Schur form of B, from the proof of Lemma 19 in Appendix D.3, we have that for

$$k \geq \frac{3.51n + 1.385\ln\left(\frac{n}{\delta}\right)}{1-|\lambda_m(A)|},$$

the norm

$$\|B^k\|_2 \leq \delta,$$

and thus setting $\delta = \frac{\epsilon\sqrt{\kappa}}{\sqrt{n}}$, we get that

$$\tau_\epsilon(A) \leq \frac{3.51n + 1.385\ln\left(n \cdot \frac{\sqrt{n}}{\sqrt{\kappa}\cdot\epsilon}\right)}{1-|\lambda_m(A)|}.$$

Further, since A is $\frac{1}{2}$-lazy, $2A-I$ is also nonnegative, with the same positive left and right eigenvector w for PF eigenvalue 1, thus having largest singular value 1, and thus every eigenvalue of $2A-I$ has magnitude at most 1. Similar to the proof of Lemma 22 in Appendix D.6, if $\lambda_r = a + i \cdot b$ is any eigenvalue of A, the corresponding eigenvalue in $2A-I$ is $2a-1+i\cdot 2b$, whose magnitude is at most 1, giving $(2a-1)^2+4b^2 \leq 1$ or $a^2+b^2 \leq a$. It further gives that

$$1-|\lambda_r| = 1-\sqrt{a^2+b^2} \geq 1-\sqrt{1-(1-a)} \geq 1-e^{\frac{1}{2}(1-a)} \geq \frac{1}{4}(1-a)$$

or

$$1-|\lambda_m| \geq \frac{1}{4}(1-\mathrm{Re}\lambda_m) \geq \frac{1}{4}(1-\mathrm{Re}\lambda_2),$$

which gives

$$\begin{aligned}\tau_\epsilon(A) &\leq 4\cdot\frac{3.51n+1.385\ln\left(n\cdot\frac{\sqrt{n}}{\sqrt{\kappa\cdot\epsilon}}\right)}{1-\mathrm{Re}\lambda_2(A)} \\ &\leq 20\cdot\frac{n+\ln\left(\frac{1}{\kappa\cdot\epsilon}\right)}{1-\mathrm{Re}\lambda_2(A)}\end{aligned}$$

completing the proof. □

E.4 Proof of Lemma 27

Proof. Since $\phi(A)=\phi(M)$ and each have the same left and right eigenvector w for PF eigenvalue 1, the lower bound on $\tau_\epsilon(A)$ in terms of $\tau_\epsilon(M)$ follows immediately from Lemma 25, since

$$\sqrt{\tau_\epsilon(M)} \leq \ln^{\frac{1}{2}}\left(\frac{n}{\kappa\cdot\epsilon}\right)\cdot\frac{1}{\phi(M)} = \ln^{\frac{1}{2}}\left(\frac{n}{\kappa\cdot\epsilon}\right)\cdot\frac{1}{\phi(A)} \leq \ln^{\frac{1}{2}}\left(\frac{n}{\kappa\cdot\epsilon}\right)\cdot\frac{2\tau_\epsilon(A)}{1-2\epsilon}.$$

For the upper bound, we first define a new quantity for *positive semidefinite* nonnegative matrices M with PF eigenvalue 1 and w as the corresponding left and right eigenvector. Let $T_\epsilon(M)$ be defined as the smallest number k for which

$$\|M^k - w\cdot w^T\|_2 = \epsilon.$$

Since M is symmetric, we can write $M = w\cdot w^T + UDU^*$ where the first column of the unitary U is w, and the diagonal matrix D contains all eigenvalues of M except the eigenvalue 1 which is replaced by 0. Further, since M is positive semidefinite, $\lambda_2(M)$ is the second largest eigenvalue of M, then for every $i>2$, $0\leq\lambda_i(M)\leq\lambda_2(M)$. Thus we have

$$\|M^k - w\cdot w^T\|_2 = \|UD^kU^*\|_2 = \|D^k\|_2 = \lambda_2^k(M)$$

and

$$T_\epsilon(M) = \frac{\ln\left(\frac{1}{\epsilon}\right)}{\ln\left(\frac{1}{\lambda_2(M)}\right)}.$$

Further, for the eigenvector y of M corresponding to λ_2, we have for $k=T_\epsilon(M)$,

$$(M^k - w\cdot w^T)y = \lambda_2^k\cdot y = \epsilon\cdot y.$$

Setting $x=\frac{y}{\|D_w y\|_1}$, we have $\|D_w x\|_1 = 1$, and we get

$$\|D_w(M^k - w\cdot w^T)x\|_1 = \left\|D_w\frac{\epsilon\cdot y}{\|D_w y\|_1}\right\|_1 = \epsilon,$$

which implies that

$$\tau_\epsilon(M) \geq k = T_\epsilon(M), \tag{25}$$

since there is some vector x with $\|D_w x\| = 1$ that has $\|D_w(M^k - w \cdot w^T)x\|_1 = \epsilon$ and for every $t < k$, it is also the case that $\|D_w(M^t - w \cdot w^T)x\|_1 > \epsilon$.

Now we observe that since A is $\frac{1}{2}$-lazy with PF eigenvalue 1, and the same left and right eigenvector w for eigenvalue 1, we have that $M = \frac{1}{2}(A + A^T)$ is positive semidefinite. We continue to bound $\tau_\epsilon(A)$ similar to the proof method (in Appendix E.1) used for Lemma 24. For any t and x with $\|D_w x\|_1 = 1$, we have that

$$\|D_w(A^t - w \cdot w^T)x\|_1 \leq \sqrt{\frac{n}{\kappa}}\sigma_2(A)^t = \sqrt{\frac{n}{\kappa}}\sigma_2^2(A)^{\frac{t}{2}}$$

and thus,

$$\tau_\epsilon(A) \leq \frac{2 \cdot \ln\left(\frac{\sqrt{n}}{\sqrt{\kappa}\cdot\epsilon}\right)}{\ln\left(\frac{1}{\sigma_2^2(A)}\right)}$$

and since A is $\frac{1}{2}$-lazy, as shown in the proof of Lemma 25 in Appendix E.2,

$$\sigma_2^2(A) = \lambda_2(AA^T) \leq \lambda_2(M),$$

giving

$$\tau_\epsilon(A) \leq \frac{2 \cdot \ln\left(\frac{\sqrt{n}}{\sqrt{\kappa}\cdot\epsilon}\right)}{\ln\left(\frac{1}{\lambda_2(M)}\right)} = \frac{2 \cdot \ln\left(\frac{\sqrt{n}}{\sqrt{\kappa}\cdot\epsilon}\right)}{\ln\left(\frac{1}{\epsilon}\right)} \cdot \frac{\ln\left(\frac{1}{\epsilon}\right)}{\ln\left(\frac{1}{\lambda_2(M)}\right)} = \frac{2 \cdot \ln\left(\frac{\sqrt{n}}{\sqrt{\kappa}\cdot\epsilon}\right)}{\ln\left(\frac{1}{\epsilon}\right)} \cdot T_\epsilon(M) \leq \frac{2 \cdot \ln\left(\frac{\sqrt{n}}{\sqrt{\kappa}\cdot\epsilon}\right)}{\ln\left(\frac{1}{\epsilon}\right)} \cdot \tau_\epsilon(M)$$

where the last inequality followed from equation 25. □

E.5 Proof of Lemma 28

Proof. We will first find the mixing time of the operator $\exp\left(\frac{R-I}{2}\right)$, and the mixing time of the operator $\exp(R-I)$ will simply be twice this number. By expansion of the exp function, it follows that $\exp\left(\frac{R-I}{2}\right)$ has PF eigenvalue 1, and u and v as the corresponding left and right eigenvectors. Further,

$$\exp\left(\frac{R-I}{2}\right) = e^{-\frac{1}{2}}\left(I + \sum_{i\geq 1}\frac{1}{i!}\frac{R^i}{2^i}\right)$$

which is $\frac{1}{2}$-lazy due to the first term since $e^{-\frac{1}{2}} \geq \frac{1}{2}$ and all the other terms are nonnegative. Further, for any set S for which $\sum_{i \in S} u_i v_i \leq \frac{1}{2}$, let $\delta = \frac{1}{2}$, then

$$\begin{aligned}
\phi_S\left(\exp\left(\frac{R-I}{2}\right)\right) &= e^{-\delta} \cdot \frac{\langle \mathbf{1}_S, D_u \exp\left(\frac{R}{2}\right) D_v \mathbf{1}_{\overline{S}} \rangle}{\langle \mathbf{1}_S, D_u D_v \mathbf{1} \rangle} \\
&= e^{-\delta} \cdot \sum_{i \geq 1} \frac{\delta^i}{i!} \cdot \frac{\langle \mathbf{1}_S, D_u R^i D_v \mathbf{1}_{\overline{S}} \rangle}{\langle \mathbf{1}_S, D_u D_v \mathbf{1} \rangle} \\
&\quad [\text{since } \langle \mathbf{1}_S, D_u I D_v \mathbf{1}_{\overline{S}} \rangle = 0] \\
&= e^{-\delta} \cdot \sum_{i \geq 1} \frac{\delta^i}{i!} \cdot \phi_S(R^i) \\
&\quad [\text{since } R^i \text{ also has } u \text{ and } v \text{ as the left and right eigenvectors for eigenvalue } 1] \\
&\leq e^{-\delta} \cdot \sum_{i \geq 1} \frac{\delta^i}{i!} \cdot i \cdot \phi_S(R) \\
&\quad [\text{using Lemma } 17] \\
&= e^{-\delta} \cdot \delta \cdot \phi_S(R) \sum_{i \geq 1} \frac{\delta^{i-1}}{(i-1)!} \\
&= e^{-\delta} \cdot \delta \cdot \phi_S(R) \cdot e^{\delta} \\
&= \delta \cdot \phi_S(R)
\end{aligned}$$

and thus,

$$\phi\left(\exp\left(\frac{R-I}{2}\right)\right) \leq \frac{1}{2}\phi(R). \tag{26}$$

Moreover, considering the first term in the expansion, we get

$$\phi_S\left(\exp\left(\frac{R-I}{2}\right)\right) = e^{-\delta} \sum_{i \geq 1} \frac{\delta^i}{i!} \cdot \phi_S(R^i) \geq e^{-\delta} \cdot \delta \cdot \phi_S(R)$$

or

$$\phi\left(\exp\left(\frac{R-I}{2}\right)\right) \geq \frac{3}{10} \cdot \phi(R). \tag{27}$$

Since $\exp\left(\frac{R-I}{2}\right)$ has left and right eigenvectors u and v, and is $\frac{1}{2}$-lazy, we get from Lemma 25 and equations 26 and 27 that

$$\frac{1}{2} \cdot \frac{\frac{1}{2} - \epsilon}{\frac{1}{2}\phi(R)} \leq \tau_\epsilon(\exp(R-I)) \leq 2 \cdot \frac{4 \cdot \ln\left(\frac{n}{\kappa \cdot \epsilon}\right)}{\left(\frac{3}{10}\right)^2 \phi^2(R)} \leq \frac{100 \cdot \ln\left(\frac{n}{\kappa \cdot \epsilon}\right)}{\phi^2(R)}$$

giving the result. □

F Intuition behind the construction

The main idea behind the construction in Section 3 is to systematically reduce the space of doubly stochastic matrices under consideration, while ensuring that we still have matrices that we care about in our space. We explain some of the key ideas that led to the construction.

All eigenvalues 0. Consider de Bruijn graphs. As stated in Section 3.1, all their nontrivial eigenvalues are 0, yet they have small expansion. This suggests the question whether de Bruijn matrices have the *minimum* expansion amongst *all* doubly stochastic matrices that have eigenvalue 0 (which turned out to be far from the case). As we see in the first steps in the proof of Lemma 19 in Appendix D.3, if for a doubly stochastic matrix A, every nontrivial eigenvalue has magnitude at most $1-c$ for some constant c, then powering just $O(\log n)$ times will make the diagonal entries inverse polynomially small in magnitude, and thus it would seem that the matrix should have behavior similar to matrices with all eigenvalues 0. Thus, the starting point of the construction is to consider only doubly stochastic matrices that have all eigenvalues 0. Note that this also helps us to restrict to real (i.e., orthogonal) diagonalizing unitaries U, and real triangular matrices T in the Schur form.

Schur block that looks like a Jordan block. The next observation again comes from the proof of Lemma 19 (in Appendix D.3). Note that after writing $A = J + B$, the main idea was to observe that if $B^m \approx 0$, then $\phi(A^m) \approx \phi(J)$. Since $B^m = UT^mU^*$ where T is upper triangular, this is equivalent to requiring that $T^m \approx 0$. As we have assumed that all nontrivial eigenvalues of T are 0, it means the entire diagonal of T is 0. How large does m have to be to ensure that $T^m \approx 0$? It is shown in Lemma 19 that for $m \approx O(n)$, $T^m \approx 0$. Observe that for Lemma 19, $\Omega(n)$ is indeed necessary, considering the case of nilpotent T, with all 0 entries, and 1 on every entry above the diagonal. But can a doubly stochastic matrix have such a Schur form? Such matrices would at least be sufficient on the outset, to ensure that Lemma 19 is tight, hopefully giving us the kind of bound we need. Thus, our second restriction is the following: we assume that every entry of T is 0, and every off-diagonal entry (entries $T(i, i+1)$ for $2 \le i \le n-1$) is r. We want to *maximize* r and still have some unitary that transforms T to a doubly stochastic matrix. Note that T will have only 0 in the entries of the first row and column since we have diagonalized-out the projection onto the stochastic eigenvector (i.e., J). In any upper triangular matrix T in which the diagonal has zeros, the off-diagonal entries *affect* the entries in powers of T the most, since entries far from the diagonal will become ineffective after a few powers of T. Thus, choosing T with the non-zeros pattern of a Jordan block will not be far from optimal.

Restricting to hermitian unitaries with a uniform row vector. Next we make several observations about the unitary U, where $A = J + UTU^*$. Note that the first column of U is the vector $\frac{1}{\sqrt{n}}\mathbf{1}$, since it is an eigenvector of A. The choice of U constrains the value of r that we can use in T since it must lead to A becoming a doubly stochastic matrix. From considering the optimal 3×3 and 4×4 cases, which although, clearly, do not give any information about relation between ϕ and $\text{Re}\lambda_2$ or Γ (since it could be off by large constant factors), we note that they do give information about the *unitary* U, since it is a *rotation* and is not a numeric value like ϕ or $\text{Re}\lambda_2$. From these cases, we make the crucial (albeit empirical) observation: the *direction* in which r can be maximized (given our constraints) is such that the first *row* of U is also $\frac{1}{\sqrt{n}}\mathbf{1}$, and this is the second constraint we impose on U. Given that the unitaries have the *same* vector in the first row and column, our next observation (again from considering certain specific cases) is that since the unitary preserves the all 1's eigenvector, and since T has *exactly the same value* above the diagonal by construction and all other entries are 0, U is performing only 2 types of *actions* on T. This type of feature is provided by unitaries that are hermitian, and this is the next assumption we make – we restrict to unitaries that are hermitian.

Optimizing for ϕ and r. The restrictions adopted so far on the unitary U and the Schur matrix T imply a sequence of inequalities to ensure that $A = J + UTU^*$ is doubly stochastic. Subject to the constraints provided by these inequalities, we aim to *minimize* ϕ and *maximize* r. Due to our restrictions on U, the cut $S = \{1\}$ in the resulting matrix A is special, and we aim to *minimize* the edge expansion $1 - A_{1,1}$ of this cut. With the set of possible values that r can take, we note that a set of extreme points of the resulting optimization problem of minimizing $1 - A_{1,1}$ or maximizing $A_{1,1}$ are obtained if we *force* the values of all the entries $A_{1,i}$ for $3 \leq i \leq n$ to 0. We then maximize r for the resulting matrix (indeed, there are exactly two possible doubly stochastic matrices at this point), and the result is the construction given in Section 3.