

Non-backtracking random walks and a weighted Ihara's theorem

Mark Kempton*

Abstract

We study the mixing rate of non-backtracking random walks on graphs by looking at non-backtracking walks as walks on the directed edges of a graph. A result known as Ihara's Theorem relates the adjacency matrix of a graph to a matrix related to non-backtracking walks on the directed edges. We prove a weighted version of Ihara's Theorem which relates the transition probability matrix of a non-backtracking walk to the transition matrix for the usual random walk. This allows us to determine the spectrum of the transition probability matrix of a non-backtracking random walk in the case of regular graphs and biregular graphs. As a corollary, we obtain a result of Alon et. al. in [1] that in most cases, a non-backtracking random walk on a regular graph has a faster mixing rate than the usual random walk. In addition, we obtain an analogous result for biregular graphs.

1 Introduction

A random walk on a graph G is a random process on the vertices of G in which, at each step in the walk, we choose uniformly at random among the neighbors of the current vertex. Random walks have been studied extensively, and are used in a variety of algorithms involving graphs. For a comprehensive survey on random walks on graphs, see [13], and for applications of spectral techniques to random walk theory, see [5]. Random walks on graphs have the useful property that given any initial distribution on the vertex set, the random walk converges to a unique stationary distribution as long as the graph is connected and not bipartite. The speed at which this convergence takes place is referred to as the *mixing rate* of the random walk. In a graph where a random walk has a fast mixing rate, vertices can be sampled quickly using this random process, making this a useful tool in theoretical computer science.

A non-backtracking random walk on a graph is a random walk with the added condition that, on a given step, we are not allowed to return to the vertex visited on the previous step. Viewed as a walk on vertices, a non-backtracking random walk loses the property of being a Markov chain, making its analysis somewhat more difficult. However, their study has received increased interest in recent years. Recently, Angel, Friedman, and Hoory [2] studied non-backtracking walks on the universal cover of a graph. Fitzner and Hofstad [7] studied the convergence of non-backtracking random walks on lattices and tori. Krzakala et. al. [11] use a matrix related to non-backtracking walks to study spectral clustering algorithms. Most pertinent to the current paper, Alon, Benjamini, Lubetzky, and Sodin [1] studied the mixing rate of a non-backtracking walk for regular graphs. In particular, they prove that in most cases, a non-backtracking random walk on a regular graph has a faster mixing rate than a random walk allowing backtracking.

In this paper, we study the mixing rate for a non-backtracking random walk, with the goal of removing the condition of regularity needed in the results of Alon et. al. in [1]. We take a different approach than Alon et. al. by looking at the non-backtracking walk as a walk along directed edges of a graph, as is done in [2]. This allows us to turn the non-backtracking random walk into a Markov chain, but on a larger state space, which in turn allows us to determine the stationary distribution to which a non-backtracking walk converges for a general graph, whether or not it is regular. In the case of regular graphs, our approach allows us to compute the spectrum of the transition probability matrix for a non-backtracking random walk, expressed in terms of the eigenvalues of the adjacency matrix. This allows for easy comparison of the mixing rates of a non-backtracking random walk, and an ordinary random walk. As a corollary, this gives us an alternate proof of the result in [1] for regular graphs. Our approach gives more information than the approach in [1],

*Center of Mathematical Sciences and Applications, Harvard University, Cambridge, MA. Email: mkempton@cmsa.fas.harvard.edu

in that the full spectrum of the transition probability matrix is given. In addition, we are able to compute the spectrum of the non-backtracking transition probability matrix for biregular graphs. As a corollary, we generalize the result in [1] for regular graphs to an analogous result for biregular graphs.

A key component in our proof is a weighted version of a result known as Ihara's Theorem, also called the Ihara zeta identity, which relates an operator indexed by the directed edge set of a graph to an operator indexed by the vertex set of the graph. Ihara's Theorem was first considered in the study of number theoretic zeta functions on graphs, and was first proved for regular graphs by Ihara in 1966 (see [9]). Numerous other proofs have been given since, along with generalizations to irregular graphs, by Hashimoto ([8], 1989), Bass ([3], 1992), Stark and Terras ([15], 1996), Kotani and Sunada ([10], 2000), and others. We will give an elementary proof of Ihara's Theorem that, to our knowledge, is original. In addition, we follow ideas similar to those in [10] to obtain a version of Ihara's Theorem with weights that allows us to study the relevant transition probability matrices for random walks.

The remainder of this paper is organized as follows. In section 2, we give the necessary background and preliminary information on random walks, and develop the corresponding theory for non-backtracking walks, including the convergence of a non-backtracking walk to a stationary distribution for a general graph. We accomplish this via walks on the directed edges of a graph. We also investigate bounds obtained from the normalized Laplacian for a directed graph. We also give the relevant background on Ihara's Theorem, and a new elementary proof. In section 3, we prove our weighted version of Ihara's formula. Finally, in section 4, we use this formula to obtain the spectrum of the transition probability matrix for a non-backtracking random walk for regular and biregular graphs. This gives a new proof of the result of Alon et. al. concerning the mixing rate of a non-backtracking random walk on a regular graph, and generalizes this result to the class of biregular graphs.

2 Preliminaries

2.1 Random walks

Throughout this paper, we will let $G = (V, E)$ denote a graph with vertex set V and (undirected) edge set E , and we will let $n = |V|$ and $m = |E|$. A *random walk* on a graph is a sequence (v_0, v_1, \dots, v_k) of vertices $v_i \in V$ where v_i is chosen uniformly at random among the neighbors of v_{i-1} . Random walks on graphs are well-studied, and considerable literature exists about them. See in particular [5] and [13] for good surveys, especially in the use of spectral techniques in studying random walks on graphs.

The *adjacency matrix* A of G is the $n \times n$ matrix with rows and columns indexed by V given by

$$A(u, v) = \begin{cases} 1 & \text{if } u \sim v \\ 0 & \text{otherwise.} \end{cases}$$

It is a well-known fact that the (u, v) entry of A^k is the number of walks of length k starting at vertex u and ending at vertex v . Define D to be the $n \times n$ diagonal matrix with rows and columns indexed by V with $D(v, v) = d_v$, where d_v denotes the degree of vertex v . A random walk on a graph G is a Markov process with transition probability matrix $P = D^{-1}A$, so

$$P(u, v) = \begin{cases} \frac{1}{d_u} & \text{if } u \sim v \\ 0 & \text{otherwise.} \end{cases}$$

Given any starting probability distribution f_0 on the vertex set V , the resulting expected distribution f_k after applying k random walk steps is given by $f_k = f_0 P^k$. Here we are considering f_0 and f_k as row vectors in \mathbb{R}^n .

Note that, in general, P is not symmetric for an irregular graph, but is similar to the symmetric matrix $D^{-1/2}AD^{-1/2}$. Thus, the eigenvalues of P are real, and if we order them as $\mu_1 \geq \mu_2 \geq \dots \geq \mu_n$, then it is easy to see that $\mu_1 = 1$ with eigenvector $\mathbf{1}$, and $\mu_n \geq -1$. By Perron-Frobenius theory, if the matrix P is irreducible, then we have that $\mu_2 < 1$, and if P is aperiodic, then $\mu_n > -1$. The matrix P being irreducible and aperiodic corresponds to the graph G being connected and non-bipartite.

The *stationary distribution* for a random walk on G is given by

$$\pi(v) = \frac{d_v}{\text{vol}(G)}.$$

The stationary distribution has the important property the $\pi P = \pi$, so that a random walk with initial distribution π will stay at π at each step. An important fact about the stationary distribution is that if G is a connected graph that is not bipartite, then for any initial distribution f_0 on $V(G)$, we have

$$\lim_{t \rightarrow \infty} (f_0 P^t)(v) = \pi(v)$$

for all v (see [13]).

Knowing that a random walk will converge to some stationary distribution, a fundamental question to consider is to determine how quickly the random walk approaches the stationary distribution, or in other words, to determine the *mixing rate*. In order to make this question precise, we need to consider how to measure the distance between two distribution vectors.

Several measures for defining the mixing rate of a random walk have been given (see [5]). Classically, the mixing rate is defined in terms of the pointwise distance (see [13]). That is, the mixing rate is

$$\rho = \limsup_{t \rightarrow \infty} \max_{u,v} |P^t(u,v) - \pi(v)|^{1/t}.$$

Note that a small mixing rate corresponds to fast mixing. Alternatively, the mixing rate can be considered in terms of the standard L_2 (Euclidean) norm, the *relative pointwise distance*, the *total variation distance*, or the χ -*squared distance*. In general, these measures can yield different distances, but spectral bounds on the mixing rate are essentially the same for each. See [5] for a detailed comparison of each. For our purposes, we will primarily be concerned with the χ -squared distance, which will be defined below.

The mixing rate of a random walk is directly related to the eigenvalues of P .

Theorem 1 (Corollary 5.2 of [13]). *Let G be a connected non-bipartite graph with transition probability matrix P , and let the eigenvalues of P be $1 = \mu_1 > \mu_2 \geq \dots \geq \mu_n > -1$. Then the mixing rate is $\max\{\mu_2, |\mu_n|\}$.*

Thus, the smaller the eigenvalues of P , the faster the random walk converges to its stationary distribution.

2.2 Non-backtracking random walks

A *non-backtracking random walk* on G is a sequence (v_0, v_1, \dots, v_k) of vertices $v_i \in V$ where v_{i+1} is chosen randomly among the neighbors of v_i such that $v_{i+1} \neq v_{i-1}$ for $i = 1, \dots, k-1$. In other words, a non-backtracking random walk is a random walk in which a step is not allowed to go back to the immediately previous state. A non-backtracking random walk on a graph is not a Markov chain since, in any given state, we need to remember the previous step in order to take the next step. In order for this to be well-defined, we assume throughout the remainder of the paper that the minimum degree of G is at least 2.

Define $P^{(k)}$ to be the $n \times n$ transition probability matrix for a k -step non-backtracking random walk on the vertices. That is $P^{(k)}(u,v)$ is the probability that a non-backtracking random walk starting at vertex u ends up at vertex v after k steps. Note that $P^{(1)} = P$, where $P = D^{-1}A$ is the transition matrix for an ordinary random walk on G . However, $P^{(k)}$ is not simply P^k since a non-backtracking random walk is not a Markov chain.

This process can be turned into a Markov chain, however, by changing the state space from the vertices of the graph to the directed edges of the graph. That is, replace each edge in E with two directed edges (one in each direction). Then the non-backtracking random walk is a sequence of directed edges (e_1, e_2, \dots, e_k) where if $e_i = (v_j, v_k)$, and $e_{i+1} = (v_r, v_s)$ then $v_k = v_r$ and $v_s \neq v_j$. That is, the non-backtracking condition restricts the walk from moving from an edge to the edge going in the opposite direction. Denote the set of directed edges by \vec{E} . The transition probability matrix for this process we will call \tilde{P} . Observe that

$$\tilde{P}((u,v), (x,y)) = \begin{cases} \frac{1}{d_v-1} & \text{if } v = x \text{ and } y \neq u \\ 0 & \text{otherwise.} \end{cases}$$

Note that \tilde{P} is a $2m \times 2m$ matrix. Note also that \tilde{P}^k is the transition matrix for a walk with k steps on the directed edges.

Lemma 1. *Given any graph G , the matrix \tilde{P} as defined above is doubly stochastic.*

Proof. Observe first that the rows of the matrix \tilde{P} sum to 1, as it is a transition probability matrix. In addition, the columns of \tilde{P} sum to 1. To see this, consider the column indexed by the directed edge (u, v) . The entry of this column corresponding to the row indexed by (x, y) is $\frac{1}{d_y-1}$ if $y = u$ and if $v \neq x$. Since $y = u$ this is equal to $\frac{1}{d_u-1}$. Otherwise, the entry is 0. Thus the column sum is

$$\sum_{\substack{x \sim u \\ x \neq v}} \frac{1}{d_u-1} = \frac{d_u-1}{d_u-1} = 1$$

as claimed. □

Define the distribution $\tilde{\pi} : \vec{E} \rightarrow \mathbb{R}$ by

$$\tilde{\pi} = \frac{\mathbb{1}}{\text{vol}(G)}$$

where $\mathbb{1}$ is the vector of length $2m$ with each entry equal to 1.

Lemma 2. *Let $\tilde{f}_0 : \vec{E} \rightarrow \mathbb{R}$ be any distribution on the directed edges of G . If the matrix \tilde{P} is irreducible and aperiodic, then*

$$\tilde{f}_0 \tilde{P}^k \longrightarrow \tilde{\pi}$$

as $k \rightarrow \infty$.

Proof. It follows from Lemma 1 that $\tilde{\pi}$ is a stationary distribution for \tilde{P} . This follows because, since the columns of \tilde{P} sum to 1, we have

$$\tilde{\pi} \tilde{P} = \tilde{\pi}.$$

Therefore, if the sequence $\tilde{f}_0 \tilde{P}^k$ converges, it must converge to $\tilde{\pi}$. Now, \tilde{P} being irreducible and aperiodic are precisely the conditions for this to converge. □

Let f be a probability distribution on the vertices of G . Then f can be turned into a distribution \tilde{f} on \vec{E} as follows. Define

$$\tilde{f}((u, v)) = \frac{1}{d_u} f(u).$$

Conversely, given a distribution \tilde{g} on \vec{E} , define a distribution g on the vertices by

$$g(u) = \sum_{(u,v) \in \vec{E}} \tilde{g}(u, v).$$

Thus, given any starting distribution $f_0 : V \rightarrow \mathbb{R}$ on the vertex set of G , we can compute the distribution after k non-backtracking random walk steps $f_k : V \rightarrow \mathbb{R}$ as follows. First compute the distribution \tilde{f}_0 on the directed edges as above, then compute $\tilde{f}_k = \tilde{f}_0 \tilde{P}^k$, then f_k is given by $f_k(u) = \sum_{v \sim u} \tilde{f}_k(u, v)$. The following proposition tells us that this converges to the same stationary distribution as an ordinary random walk on a graph.

Theorem 2. *Given a graph G and a starting distribution $f_0 : V \rightarrow \mathbb{R}$ on the vertices of G , define $f_k = f_0 P^{(k)}$ to be the distribution on the vertices after k non-backtracking random walk steps. Define the distribution $\pi : V \rightarrow \mathbb{R}$ by $\pi(v) = \frac{d_v}{\text{vol}(G)}$ (note that this is the stationary distribution for an ordinary random walk on G). Then if the matrix \tilde{P} is irreducible and aperiodic, then for any starting distribution f_0 on V , we have*

$$f_k \longrightarrow \pi \text{ as } k \rightarrow \infty.$$

Proof. As described above, take the distribution f_0 on vertices to the corresponding distribution \tilde{f}_0 on directed edges. Then define $\tilde{f}_k = \tilde{f}_0 \tilde{P}^k$. Then by Lemma 2, \tilde{f}_k converges to $\tilde{\pi}$. Now $\tilde{\pi} = \frac{\mathbb{1}}{\text{vol}G}$, and observe that

$$\pi(u) = \frac{d_u}{\text{vol}(G)} = \sum_{v \sim u} \frac{1}{\text{vol}(G)} = \sum_{v \sim u} \tilde{\pi}((u, v)).$$

So pulling the distribution $\tilde{\pi}$ on directed edges back to a distribution on the vertices yields π . Thus the result follows. \square

Definition 1. The χ -squared distance for measuring convergence of a random walk is defined by

$$\Delta'(t) = \max_{y \in V(G)} \left(\sum_{x \in V(G)} \frac{(\tilde{P}^t(y, x) - \tilde{\pi}(x))^2}{\tilde{\pi}(x)} \right)^{1/2}.$$

Notice that since $\tilde{\pi} = \mathbb{1}/\text{vol}(G)$,

$$\begin{aligned} \Delta'(t)^2 &= \max_y \frac{1}{2m} \|(\chi_y \tilde{P}^t - \tilde{\pi})\|^2 \\ &= \max_y \frac{1}{2m} \|(\chi_y - \tilde{\pi}) \tilde{P}^t\|^2 \end{aligned}$$

Theorem 3. Let $\mu_1 = 1, \mu_2, \dots, \mu_{2m}$ be the eigenvalues of \tilde{P} . Then the convergence rate for the non-backtracking random walk with respect to the χ -squared distance is bounded above by $\max_{i \neq 1} |\mu_i|$.

Proof. We have

$$\Delta'(t)^2 = \max_y \frac{1}{2m} \|(\chi_y - \tilde{\pi}) \tilde{P}^t\|^2.$$

Observe that $\chi_u - \tilde{\pi}$ is orthogonal $\tilde{\pi}$, which is the eigenvector for μ_1 , so we see that

$$\Delta'(t) \leq \frac{1}{2m} \max_{i \neq 1} |\mu_i|^t.$$

Therefore,

$$\lim_{t \rightarrow \infty} (\Delta'(t))^{1/t} \leq \max_{i \neq 1} |\mu_i|.$$

\square

2.3 Non-backtracking Walks as Walks on a Directed Graph

The transition probability matrix \tilde{P} for the walk on directed edges can be thought of as a transition matrix for a random walk on a directed line graph of the graph G . In this way, theory for random walks on directed graphs can be applied to analyze non-backtracking random walks. Random walks on directed graphs have been studied by Chung in [4] by way of a directed version of the normalized graph Laplacian matrix. In [4], the Laplacian for a directed graph is defined as follows. Let P be the transition probability matrix for a random walk on the directed graph, and let ϕ be its Perron vector, that is, $\phi P = \phi$. Then let Φ be the diagonal matrix with the entries of ϕ along the diagonal. Then the *Laplacian* for the directed graph is defined as

$$\mathcal{L} = I - \frac{\Phi^{1/2} P \Phi^{-1/2} + \Phi^{-1/2} P^* \Phi^{1/2}}{2}.$$

This produces a symmetric matrix that thus has real eigenvalues. Those eigenvalues are then related to the convergence rate of a random walk on the directed graph. In particular, the convergence rate is bounded above by $2\lambda_1^{-1}(-\log \min_x \phi(x))$, where λ_1 is the second smallest eigenvalue of \mathcal{L} (see Theorem 7 of [4]).

Applying this now to non-backtracking random walks, define \tilde{P} as before. Then as seen above, ϕ is the constant vector with $\phi(v) = 1/\text{vol}(G)$ for all v . Then the directed Laplacian for a non-backtracking walk becomes

$$\tilde{\mathcal{L}} = I_{2m} = \frac{\tilde{P} + \tilde{P}^*}{2}.$$

Then Theorem 1 of [4], applied to the matrix $\tilde{\mathcal{L}}$ as defined, gives the Rayleigh quotient for a function $f : \vec{E} \rightarrow \mathbb{C}$ by

$$\tilde{R}(f) = \frac{f^* \tilde{\mathcal{L}} f}{f^* f} = \frac{1}{2} \frac{\sum_{(u,v) \in \vec{E}(G)} \sum_{\substack{(v,w) \\ w \neq u}} (f(u,v) - f(v,w))^2 \tilde{P}((u,v), (v,w))}{\sum_{(u,v) \in \vec{E}(G)} f(u,v)^2}.$$

From this it is clear that $\tilde{\mathcal{L}}$ is positive semidefinite with smallest eigenvalue $\lambda_0 = 0$. If $0 = \lambda_0 \leq \lambda_1 \leq \dots \leq \lambda_{2m-1}$ are the eigenvalues of $\tilde{\mathcal{L}}$, then Theorem 7 from [4] implies that the convergence rate for the corresponding random walk is bounded above by

$$\frac{2 \log \text{vol}(G)}{\lambda_1}.$$

We remark that for an ordinary random walk on an undirected graph G , the convergence rate is also on the order of $1/\lambda_1(\mathcal{L})$, where \mathcal{L} now denotes the normalized Laplacian of the undirected graph G . Note that

$$\lambda_1(\mathcal{L}) = \inf_{\substack{f: V(G) \rightarrow \mathbb{R} \\ f \perp D\mathbb{1}}} R(f)$$

where $R(f) = \frac{\sum_{uv \in E(G)} (f(u) - f(v))^2}{\sum_{v \in V(G)} f(v)^2 d_v}$ denotes the Rayleigh quotient with respect to \mathcal{L} , and

$$\lambda_1(\tilde{\mathcal{L}}) = \inf_{\substack{f: \vec{E}(G) \rightarrow \mathbb{R} \\ f \perp \mathbb{1}}} \tilde{R}(f)$$

with \tilde{R} given above.

The following result shows that the Laplacian bound does not give an improvement for non-backtracking random walks over ordinary random walks.

Proposition 1. *Let G be any graph, and let \mathcal{L} be the normalized graph Laplacian and $\tilde{\mathcal{L}}$ the non-backtracking Laplacian defined above. Then we have*

$$\lambda_1(\tilde{\mathcal{L}}) \leq \lambda_1(\mathcal{L}).$$

Proof. Let $f : V(G) \rightarrow \mathbb{R}$ be the function orthogonal to $D\mathbb{1}$ that achieves the minimum in the Rayleigh quotient for \mathcal{L} . So

$$\sum_{v \in V(G)} f(v) d_v = 0 \text{ and } \lambda_1(\mathcal{L}) = \frac{\sum_{uv \in E(G)} (f(u) - f(v))^2}{\sum_{v \in V(G)} f(v)^2 d_v}.$$

Define $f' : \vec{E} \rightarrow \mathbb{R}$ by $f'(u,v) = f(u)$. Observe that

$$\sum_{(u,v) \in \vec{E}(G)} f'(u,v) = \sum_{(u,v) \in \vec{E}(G)} f(u) = \sum_{u \in V(G)} f(u) d_u = 0.$$

So f' is orthogonal to $\mathbf{1}$. Therefore

$$\begin{aligned}
\lambda_1(\tilde{\mathcal{L}}) \leq \tilde{R}(f') &= \frac{1}{2} \frac{\sum_{(u,v) \in \tilde{E}(G)} \sum_{\substack{(v,w) \\ w \neq u}} (f'(u,v) - f'(v,w))^2 \tilde{P}((u,v), (v,w))}{\sum_{(u,v) \in \tilde{E}(G)} f'(u,v)^2} \\
&= \frac{1}{2} \frac{\sum_{(u,v)} \sum_{\substack{(v,w) \\ w \neq u}} (f(u) - f(v))^2 \frac{1}{d_v - 1}}{\sum_{(u,v)} f(u)^2} \\
&= \frac{1}{2} \frac{\sum_{(u,v)} (f(u) - f(v))^2}{\sum_{u \in V(G)} f(u)^2 d_u} \\
&= \frac{\sum_{\{u,v\} \in E(G)} (f(u) - f(v))^2}{\sum_{u \in V(G)} f(u)^2 d_u} = R(f) = \lambda_1(\mathcal{L}).
\end{aligned}$$

□

2.4 Ihara's Theorem

The transition probability matrix \tilde{P} defined above is a weighted version of an important matrix that comes up in the study of zeta functions on finite graphs. We define B to be the $2m \times 2m$ matrix with rows and columns indexed by the set of directed edges of G as follows.

$$B((u,v), (x,y)) = \begin{cases} 1 & \text{if } v = x \text{ and } y \neq u \\ 0 & \text{otherwise.} \end{cases}$$

The matrix B can be thought of as a non-backtracking edge adjacency matrix, and the entries of B^k describe the number of non-backtracking walks of length k from one directed edge to another, in the same way that the entries of powers of the adjacency matrix, A^k , count the number of walks of length k from one vertex to another. The expression $\det(I - uB)$ is closely related to zeta functions on finite graphs which. A result known as Ihara's Theorem further relates such zeta functions to a determinant expression involving the adjacency matrix. While we will not go into zeta functions on finite graphs in this paper, the following result equivalent to Ihara's theorem will be of interest to us.

Ihara's Theorem. *For a graph G on n vertices and m edges, let B be the matrix defined above, let A denote the adjacency matrix, D the diagonal degree matrix, and I the identity. Then*

$$\det(I - uB) = (1 - u^2)^{m-n} \det(I - uA + u^2(D - I)).$$

We remark that the expression $\det(I - uB)$ is the characteristic polynomial of B evaluated at $1/u$. In this way the complete spectrum of the matrix B is given by the reciprocals of the roots of the polynomial $(1 - u^2)^{m-n} \det(I - uA + u^2(D - I))$. Numerous proofs of this result exist in the literature [9, 8, 3, 15, 10]. For completeness, we will include here an elementary proof that uses only basic linear algebra. To the knowledge of the author, this proof is original. To begin, we will need a lemma giving a well-known property of determinants.

Lemma 3. *Let M be a $k \times l$ matrix, N a $l \times k$ matrix, and A an invertible $k \times k$ matrix. Then*

$$\det(A + MN) = \det(A) \det(I + NA^{-1}M).$$

Proof. Note that

$$\begin{bmatrix} I & N \\ 0 & I \end{bmatrix} \begin{bmatrix} I & 0 \\ A^{-1}M & A^{-1}(A + MN) \end{bmatrix} \begin{bmatrix} I & -N \\ 0 & I \end{bmatrix} = \begin{bmatrix} I + NA^{-1}M & 0 \\ A^{-1}M & I \end{bmatrix}.$$

Taking determinants of both sides gives the result. \square

Proof of Ihara's Theorem. Define S to be the $2m \times n$ matrix

$$S((u, v), x) = \begin{cases} 1 & \text{if } v = x \\ 0 & \text{otherwise} \end{cases}$$

so S is the endpoint incidence operator. Define T to be the $n \times 2m$ matrix given by

$$T(x, (u, v)) = \begin{cases} 1 & \text{if } u = x \\ 0 & \text{otherwise} \end{cases}$$

so T is the starting point incidence operator. We will also define τ to be the $2m \times 2m$ matrix giving the reversal operator that switches a directed edge with its opposite. That is,

$$\tau((a, b), (c, d)) = \begin{cases} 1 & \text{if } b = c, a = d \\ 0 & \text{otherwise} \end{cases}$$

Now, a straightforward computation verifies that

$$B = ST - \tau, \tag{1}$$

$$A = TS, \tag{2}$$

and

$$D = T\tau S. \tag{3}$$

Then from Lemma 3 and (1) we obtain

$$\begin{aligned} \det(I - uB) &= \det(I - u(ST - \tau)) \\ &= \det(I + u\tau - uST) \\ &= \det(I + u\tau) \det(I - uT(I + u\tau)^{-1}S) \end{aligned}$$

where u is chosen so that the matrix $I + u\tau$ is invertible.

Observe that $\tau^2 = I$, so that $(I - u\tau)(I + u\tau) = (1 - u^2)I$, so $(I + u\tau)^{-1} = \frac{1}{1 - u^2}(I - u\tau)$. Thus, applying (2) and (3), the above becomes

$$\begin{aligned} \det(I - uB) &= \det(I + u\tau) \det\left(I - \frac{u}{1 - u^2}T(I - u\tau)S\right) \\ &= \det(I + u\tau) \det\left(I - \frac{u}{1 - u^2}(TS - uT\tau S)\right) \\ &= \det(I + u\tau) \frac{1}{(1 - u^2)^n} \det((1 - u^2)I - uA + u^2D) \\ &= (1 - u^2)^{m-n} \det(I - uA + u^2(D - I)) \end{aligned}$$

where the last step is obtained by observing that $\det(I + u\tau) = (1 - u^2)^m$. This is the desired equality for our choice of u . This is a polynomial of finite degree in u , and there are infinitely many u that make $I + u\tau$ invertible, so the equality holds for all u . \square

3 A weighted Ihara's theorem

In this section, we will give a weighted version of Ihara's Theorem. The proof presented in the previous section does not lend itself well to generalization to the weighted setting, so we will not follow that strategy. Rather, we will follow the main ideas of the proof of Ihara's theorem found in [10] to obtain our weighted version of this result.

To each vertex $x \in V(G)$ we assign a weight $w(x) \neq 0$, and let W be the $n \times n$ diagonal matrix given by $W(x, x) = w(x)$. Define S and T to be the matrices from the proof of Ihara's Theorem in the previous section, and define $\tilde{S} = SW$ and $\tilde{T} = WT$. So \tilde{S} is the weighted version of the endpoint vertex-edge incidence operator, and \tilde{T} is the weighted version of the starting point vertex-edge incidence operator. Define τ from the proof of Ihara's Theorem, and define $\tilde{\tau}$ to be the weighted version of τ , that is

$$\tilde{\tau}((a, b), (c, d)) = \begin{cases} w(b)^2 & \text{if } b = c, a = d \\ 0 & \text{otherwise} \end{cases}$$

Finally, define the $2m \times 2m$ matrix \tilde{P} by

$$\tilde{P}((a, b), (c, d)) = \begin{cases} w(b)^2 & \text{if } b = c, a \neq d \\ 0 & \text{otherwise.} \end{cases} \quad (4)$$

Then \tilde{P} is the weighted version of the non-backtracking edge adjacency matrix B seen above in Ihara's theorem, with $w(b)^2$ the weight on edge (a, b) . We remark that if we take $w(x) = 1/\sqrt{d_x - 1}$ for each $x \in V(G)$, then \tilde{P} is exactly the transition probability matrix for a non-backtracking random walk on the directed edges of G defined in Section 2.2. This case is our primary focus, but we note that our computations apply for any arbitrary positive weights assigned to the vertices.

Now, a straightforward computation verifies that

$$\tilde{P} = \tilde{S}\tilde{T} - \tilde{\tau} \quad (5)$$

and

$$\tilde{T}\tilde{S} = WAW. \quad (6)$$

We will define $\tilde{A} = WAW$. Note that $\tilde{A}(u, v) = w(u)w(v)$, so this is the adjacency matrix for the weighted graph with edge weights $w(u)w(v)$. The matrix \tilde{A} is similar to W^2A , so when $w(x) = 1/\sqrt{d_x - 1}$, this is the matrix whose entries are the transition probabilities for a single step of a non-backtracking random walk G .

From (5) and (6) we obtain the following equations.

$$(I - u\tilde{P})(I - u\tilde{\tau}) = I - u\tilde{S}\tilde{T} + u^2\tilde{S}\tilde{T}\tilde{\tau} - u^2\tilde{\tau}^2 \quad (7)$$

$$(I - u\tilde{\tau})(I - u\tilde{P}) = I - u\tilde{S}\tilde{T} + u^2\tilde{\tau}\tilde{S}\tilde{T} - u^2\tilde{\tau}^2 \quad (8)$$

We define \tilde{D} to be the diagonal $n \times n$ matrix $\tilde{D}(x, x) = \sum_{v \sim x} w(x)^2 w(v)^2$ and observe that $\tilde{T}\tilde{\tau}\tilde{S} = \tilde{D}$. It then follows that

$$\left((I - u\tilde{P})(I - u\tilde{\tau}) + u^2\tilde{\tau}^2 \right) \tilde{S} = \tilde{S} \left(I - u\tilde{A} + u^2\tilde{D} \right) \quad (9)$$

$$\tilde{T} \left((I - u\tilde{\tau})(I - u\tilde{P}) + u^2\tilde{\tau}^2 \right) = \left(I - u\tilde{A} + u^2\tilde{D} \right) \tilde{T} \quad (10)$$

We remark that in the proof in [10], they use the unweighted versions of each of these matrices, so τ rather than $\tilde{\tau}$ yields $\tau^2 = I$. Hence S and T will factor through τ^2 , so that the $u^2\tau^2$ term stays on the right hand side of the above equations. Here we have $\tilde{\tau}^2$ is a $2m \times 2m$ diagonal matrix with $\tilde{\tau}^2((u, v), (u, v)) = w(u)^2 w(v)^2$. Depending on the $w(u)$'s this matrix might not behave nicely with respect to the action of S and T , hence the extra terms that need to stay on the left-hand side above. This difference from [10] is one of the primary difficulties in generalizing this result.

We will now perform a change of basis to see how the operator $(I - u\tilde{P})(I - u\tilde{\tau}) + u^2\tilde{\tau}^2$ behaves with respect to the decomposition of the space of functions $f : \vec{E} \rightarrow \mathbb{C}$ as the direct sum of Image \tilde{S} and Ker \tilde{S}^T .

To this end, fix any basis of the subspace $\text{Ker } \tilde{S}^T$, and let R be the $2m \times (2m - n)$ matrix whose columns are the vectors of that basis (note that \tilde{S} has rank n). Define $M = \begin{bmatrix} \tilde{S} & R \end{bmatrix}$. This will be our change of basis matrix. To obtain the inverse of M , form the matrix $\begin{bmatrix} (\tilde{S}^T \tilde{S})^{-1} \tilde{S}^T \\ (R^T R)^{-1} R^T \end{bmatrix}$ and observe that

$$\begin{bmatrix} (\tilde{S}^T \tilde{S})^{-1} \tilde{S}^T \\ (R^T R)^{-1} R^T \end{bmatrix} \begin{bmatrix} \tilde{S} & R \end{bmatrix} = \begin{bmatrix} (\tilde{S}^T \tilde{S})^{-1} \tilde{S}^T \tilde{S} & (\tilde{S}^T \tilde{S})^{-1} \tilde{S}^T R \\ (R^T R)^{-1} R^T \tilde{S} & (R^T R)^{-1} R^T R \end{bmatrix} = \begin{bmatrix} I_n & 0 \\ 0 & I_{2m-n} \end{bmatrix}.$$

Therefore we have that $M^{-1} = \begin{bmatrix} (\tilde{S}^T \tilde{S})^{-1} \tilde{S}^T \\ (R^T R)^{-1} R^T \end{bmatrix}$.

Applying this change of basis, direct computation, applying (7) and (9), yields

$$\begin{bmatrix} (\tilde{S}^T \tilde{S})^{-1} \tilde{S}^T \\ (R^T R)^{-1} R^T \end{bmatrix} \left((I - u\tilde{P})(I - u\tilde{\tau}) + u^2\tilde{\tau}^2 \right) \begin{bmatrix} \tilde{S} & R \end{bmatrix} = \begin{bmatrix} I - u\tilde{A} + u^2\tilde{D} & -u\tilde{T}R + u^2\tilde{T}\tilde{\tau}R \\ 0 & I \end{bmatrix}. \quad (11)$$

Therefore, the matrix $(I - u\tilde{P})(I - u\tilde{\tau}) + u^2\tilde{\tau}^2$ is similar to the matrix $\begin{bmatrix} I - u\tilde{A} + u^2\tilde{D} & -u\tilde{T}R + u^2\tilde{T}\tilde{\tau}R \\ 0 & I \end{bmatrix}$, so they have the same determinant. Thus, we have proven a weighted version of Ihara's Theorem, which we state as the following.

Theorem 4. *Let G be a graph on n vertices and m edges, and assign an arbitrary positive weight $w(x) > 0$ assigned to each vertex x . Let \tilde{P} be the $2m \times 2m$ weighted non-backtracking edge adjacency matrix with edge weight $w(v)^2$ assigned to edge (u, v) as defined in (4). Let \tilde{A} be the weighted $n \times n$ adjacency matrix with edge weight $w(u)w(v)$ assigned to each edge. Let $\tilde{\tau}$ be the weighted reversal operator defined above, and \tilde{D} the $n \times n$ diagonal matrix with $\tilde{D}(x, x) = \sum_{v \sim x} w(x)^2 w(v)^2$ as defined above. Then we have*

$$\det \left((I - u\tilde{P})(I - u\tilde{\tau}) + u^2\tilde{\tau}^2 \right) = \det(I - u\tilde{A} + u^2\tilde{D}).$$

As a corollary to the decomposition in equation (11), if we take $w(x) = 1$ for all x , then $\tilde{\tau}^2 = I$, and the usual unweighted Ihara's Theorem falls out immediately.

If we take $w(x) = \frac{1}{\sqrt{d_x - 1}}$, then \tilde{P} becomes the transition probability matrix for the non-backtracking walk on directed edges, and $\tilde{A}(u, v) = \frac{1}{\sqrt{(d_u - 1)(d_v - 1)}}$. This is clearly similar to the matrix $(D - I)^{-1}A$. So in this case \tilde{A} is similar to the matrix whose entries are the transition probabilities for a single step in a non-backtracking random walk. (Note, however, that $(D - I)^{-1}A$ is not the transition probability matrix for a non-backtracking random walk.)

4 The mixing rate of non-backtracking random walks

4.1 An alternate proof for regular graphs

Applying the results of the previous section to regular graphs yields a different proof of the results from [1] on the mixing rate of non-backtracking random walks on regular graphs.

Let G be a regular graph where each vertex has degree d . Then choosing $w(x) = 1/\sqrt{d - 1}$ for all x yields gives us that \tilde{P} is the transition probability matrix for the non-backtracking random walk on G . We remark that, from the previous section, we have $\tilde{\tau} = \frac{1}{d-1}\tau$, $\tilde{\tau}^2 = \frac{1}{(d-1)^2}I$, $\tilde{A} = \frac{1}{d-1}A$, and $\tilde{D} = \frac{d}{(d-1)^2}I$. Therefore, the decomposition in (11) becomes

$$(I - u\tilde{P})(I - u\tilde{\tau}) \sim \begin{bmatrix} I - \frac{u}{d-1}A + \frac{u^2}{d-1}I & * \\ 0 & \left(1 - \frac{u^2}{(d-1)^2}\right)I \end{bmatrix}.$$

Noting that $\tilde{\tau}$ can be thought of as block diagonal with m blocks of the form $\begin{bmatrix} 0 & 1/(d-1) \\ 1/(d-1) & 0 \end{bmatrix}$, then taking determinants, we find that

$$\det(I - u\tilde{P}) \left(1 - \frac{u^2}{(d-1)^2}\right)^m = \left(1 - \frac{u^2}{(d-1)^2}\right)^{2m-n} \det \left(I - \frac{u}{d-1}A + \frac{u^2}{d-1}I \right)$$

and hence

$$\det(I - u\tilde{P}) = \left(1 - \left(\frac{u}{d-1}\right)^2\right)^{m-n} \prod_{i=1}^n \left(1 - \frac{\lambda_i}{d-1}u + \frac{1}{d-1}u^2\right)$$

where the product ranges over all the eigenvalues λ_i of the adjacency matrix A for $i = 1, \dots, n$. As remarked previously, the left hand side $\det(I - u\tilde{P})$ is the characteristic polynomial of \tilde{P} evaluated at $1/u$, so from this we obtain the spectrum of \tilde{P} .

Theorem 5. *Let G be a d -regular graph with m edges and n vertices, and let \tilde{P} be the $2m \times 2m$ transition probability matrix for a non-backtracking random walk as defined above. Then the eigenvalues of \tilde{P} are*

$$\pm \frac{1}{d-1}, \frac{\lambda_i \pm \sqrt{\lambda_i^2 - 4(d-1)}}{2(d-1)}, (i = 1, \dots, n)$$

where λ_i ranges over the eigenvalues of the adjacency matrix A , and $\pm 1/(d-1)$ each have multiplicity $m-n$.

From this we obtain the result from [1].

Corollary 1. *Let G be a non-bipartite, connected d -regular graph on n vertices for $d \geq 3$, and let ρ and $\tilde{\rho}$ denote the mixing rates of simple and non-backtracking random walk on G , respectively. Let λ be the second largest eigenvalue of the adjacency matrix of G in absolute value.*

If $\lambda \geq 2\sqrt{d-1}$, then

$$\frac{d}{2(d-1)} \leq \frac{\tilde{\rho}}{\rho} \leq 1.$$

If $\lambda < 2\sqrt{d-1}$ and $d = n^{o(1)}$, then

$$\frac{\tilde{\rho}}{\rho} = \frac{d}{2(d-1)} + o(1).$$

Proof. We remark that the expression $\frac{\lambda + \sqrt{\lambda^2 - 4(d-1)}}{2(d-1)}$ is precisely the expression derived by Alon et al. in [1] for the mixing rate of a non-backtracking random walk on a regular graph, and we may proceed with the analysis of the convergence rate in the same way they do. The convergence rate is given by the second largest eigenvalue of \tilde{P} , which will be obtained setting λ to be the second largest eigenvalue of A . Let μ be this eigenvalue.

For $2\sqrt{d-1} \leq \lambda \leq d$ we have

$$\frac{\lambda}{2(d-1)} < \frac{\lambda + \sqrt{\lambda^2 - 4(d-1)}}{2(d-1)} \leq \frac{\lambda}{d}.$$

So $\frac{\lambda}{2(d-1)} \leq \mu \leq \frac{\lambda}{d}$. Since $\frac{\lambda}{d}$ is the second largest eigenvalue of the transition probability matrix P for the usual walk, the first case follows.

For $\lambda < 2\sqrt{d-1}$, μ is complex, and we obtain

$$|\mu|^2 = \left| \frac{\lambda + \sqrt{\lambda^2 - 4(d-1)}}{2(d-1)} \right|^2 = \left(\frac{\lambda}{2(d-1)} \right)^2 + \left(\frac{\sqrt{4(d-1) - \lambda^2}}{2(d-1)} \right)^2 = \frac{1}{d-1}$$

so $|\mu| = \frac{1}{\sqrt{d-1}}$.

We remark that in this case that $\lambda < 2\sqrt{d-1}$, a classic result of Nilli ([14]) related to the Alon-Boppana Theorem implies that we are never too far below this bound. Indeed, the result states that if G is d -regular with diameter at least $2(k+1)$, then $\lambda \geq 2\sqrt{d-1} - \frac{2\sqrt{d-1}-1}{k+1}$. With the restriction that $d = n^{o(1)}$, then the diameter is at least $(1 - o(1)) \log_{d-1} n$, and so $\lambda > (1 - o(1))2\sqrt{d-1}$, and the second case follows. \square

4.2 Biregular graphs

A graph G is called (c, d) -biregular if it is bipartite and each vertex in one part of the bipartition has degree c , and each vertex of the other part has degree d . In the weighted Ihara's Theorem, we have $\tilde{\tau}^2((u, v), (u, v)) = \frac{1}{(d_u-1)(d_v-1)}$, so in the case where G is (c, d) -biregular, then we have $\tilde{\tau}^2 = \frac{1}{(c-1)(d-1)}I$. So since $\tilde{\tau}^2$ is a multiple of the identity, as with regular graphs, in the decomposition (11), the $u^2\tilde{\tau}^2$ term can be taken to the other side of the equation. Note that \tilde{D} is diagonal with $\tilde{D}(u, u) = \sum_{v \sim u} \frac{1}{(d_u-1)(d_v-1)} = \frac{c}{(c-1)(d-1)}$ if u has degree c , or $\frac{d}{(c-1)(d-1)}$ if u has degree d . Then $\tilde{D} - \tilde{\tau}^2$ is diagonal with entry $\frac{c}{(c-1)(d-1)} - \frac{1}{(c-1)(d-1)} = \frac{1}{d-1}$ or $\frac{d}{(c-1)(d-1)} - \frac{1}{(c-1)(d-1)} = \frac{1}{c-1}$. Hence the decomposition (11) becomes

$$(I - u\tilde{P})(I - u\tilde{\tau}) \sim \begin{bmatrix} I - u \begin{bmatrix} 0 & \frac{1}{c-1}M \\ \frac{1}{d-1}M^T & 0 \end{bmatrix} + u^2 \begin{bmatrix} \frac{1}{d-1}I & 0 \\ 0 & \frac{1}{c-1}I \end{bmatrix} & * \\ & \left(1 - \frac{u^2}{(c-1)(d-1)}\right) I \end{bmatrix}$$

where $A = \begin{bmatrix} 0 & M \\ M^T & 0 \end{bmatrix}$ is the adjacency matrix of G .

Note that $\tilde{\tau}$ is similar to a block diagonal matrix with blocks of the form $\begin{bmatrix} 0 & 1/(c-1) \\ 1/(d-1) & 0 \end{bmatrix}$, so taking the determinant above we obtain

$$\det(I - u\tilde{P}) \left(1 - \frac{u^2}{(c-1)(d-1)}\right)^m = \left(1 - \frac{u^2}{(c-1)(d-1)}\right)^{2m-n} \det \left(I - u \begin{bmatrix} 0 & \frac{1}{c-1}M \\ \frac{1}{d-1}M^T & 0 \end{bmatrix} + u^2 \begin{bmatrix} \frac{1}{d-1}I & 0 \\ 0 & \frac{1}{c-1}I \end{bmatrix} \right)$$

so

$$\det(I - u\tilde{P}) = \left(1 + \frac{u^2}{(c-1)(d-1)}\right)^{m-n} \det \begin{bmatrix} \left(1 + \frac{u^2}{d-1}\right)I & \frac{u}{c-1}M \\ \frac{u}{d-1}M^T & \left(1 - \frac{u^2}{c-1}\right)I \end{bmatrix}$$

We will look at the matrix $\begin{bmatrix} \left(1 + \frac{u^2}{d-1}\right)I & \frac{u}{c-1}M \\ \frac{u}{d-1}M^T & \left(1 + \frac{u^2}{c-1}\right)I \end{bmatrix}$. Suppose the first part in the bipartition of G has size r , and the second part has size s , where without loss of generality, $r > s$. By row reduction, this has the same determinant as the matrix

$$\begin{bmatrix} \left(1 + \frac{u^2}{d-1}\right)I & \frac{u}{c-1}M \\ 0 & \left(1 - \frac{u^2}{c-1}\right)I - \frac{1}{1 + \frac{u^2}{d-1}} \frac{u^2}{(c-1)(d-1)} M^T M \end{bmatrix}$$

which is

$$\begin{aligned} & \left(1 + \frac{u^2}{d-1}\right)^r \det \left(\left(1 + \frac{u^2}{c-1}\right)I - \frac{1}{1 + \frac{u^2}{d-1}} \frac{1}{(c-1)(d-1)} M^T M \right) \\ & = \left(1 + \frac{u^2}{d-1}\right)^{r-s} \det \left(\left(1 + \frac{u^2}{c-1}\right) \left(1 + \frac{u^2}{d-1}\right)I - \frac{u^2}{(c-1)(d-1)} M^T M \right). \end{aligned}$$

Now, the above determinant is given by the product of the eigenvalues of the matrix. Observe that if λ is an eigenvalue of the adjacency matrix A , then λ^2 is an eigenvalue of $M^T M$. Therefore, in all we have

$$\begin{aligned} \det(I - u\tilde{P}) & = \left(1 - \frac{u^2}{(c-1)(d-1)}\right)^{m-n} \left(1 + \frac{u^2}{d-1}\right)^{r-s} \\ & \quad \times \prod_{i=1}^s \left(\left(1 + \frac{u^2}{c-1}\right) \left(1 + \frac{u^2}{d-1}\right) - \frac{\lambda_i^2 u^2}{(c-1)(d-1)} \right) \end{aligned}$$

where the product ranges over the s largest eigenvalues of A (or in other words, λ_i^2 ranges of the s eigenvalues of $M^T M$). Therefore the characteristic polynomial is given by

$$\det(uI - \tilde{P}) = \left(u^2 - \frac{1}{(c-1)(d-1)}\right)^{m-n} \left(u^2 + \frac{1}{d-1}\right)^{r-s} \\ \times \prod_{i=1}^s \left(\left(u^2 + \frac{1}{c-1}\right) \left(u^2 + \frac{1}{d-1}\right) - \frac{\lambda_i^2 u^2}{(c-1)(d-1)} \right).$$

Thus we can explicitly obtain the eigenvalues of \tilde{P} .

Theorem 6. *Let G be a (c, d) -biregular graph, let the part with degree c have size r , and the part with degree d have size s , and assume without loss of generality that $r \geq s$. Suppose G has n vertices and m edges. Then the eigenvalues of the non-backtracking transition probability matrix \tilde{P} defined above are*

$$\pm \frac{1}{\sqrt{(c-1)(d-1)}} \text{ with multiplicity } m-n \text{ each, } \pm \frac{1}{\sqrt{d-1}} i \text{ with multiplicity } r-s \text{ each}$$

as well as the 4 roots of the polynomial

$$u^4 + \left(\frac{1}{(c-1)} + \frac{1}{(d-1)} - \frac{\lambda_i^2}{(c-1)(d-1)} \right) u^2 + \frac{1}{(c-1)(d-1)}$$

for each value of λ_i ranging over the s positive eigenvalues of the adjacency matrix A . These roots are

$$\pm \sqrt{\frac{\lambda_i^2 - (c-1) - (d-1) \pm \sqrt{(\lambda_i^2 - (c-1) - (d-1))^2 - 4(c-1)(d-1)}}{2(c-1)(d-1)}}, \quad (i = 1, \dots, s). \quad (12)$$

We can now give a version of Corollary 1 for (c, d) -biregular graphs.

Corollary 2. *Let G be a (c, d) -biregular graph with $c, d \geq 2$. Let $\rho = \lambda^2/cd$ be the square of the second largest eigenvalue of the transition probability matrix P for a random walk on G , and let $\tilde{\rho} = |\mu|^2$ be the square of the second largest modulus of an eigenvalue of \tilde{P} . Let λ be the second largest eigenvalue of the adjacency matrix of G . Then we have the following cases.*

If $\lambda > \sqrt{c-1} + \sqrt{d-1}$, then

$$\frac{cd}{2(c-1)(d-1)} \left(1 - \frac{c-1+d-1}{c-1+2\sqrt{(c-1)(d-1)}+d-1} \right) \leq \frac{\tilde{\rho}}{\rho} \leq 1.$$

If $\lambda < \sqrt{c-1} + \sqrt{d-1}$ and both c and d are $n^{o(1)}$, then

$$\frac{\tilde{\rho}}{\rho} \leq \frac{cd}{2(c-1)(d-1)} + o(1).$$

Proof. We need to compare the eigenvalues of \tilde{P} to the eigenvalues of $P = D^{-1}A = \begin{bmatrix} 0 & \frac{1}{c}M \\ \frac{1}{d}M^T & 0 \end{bmatrix}$. Note that for λ an eigenvalue of A , we have

$$\begin{bmatrix} 0 & M \\ M^T & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \lambda \begin{bmatrix} x \\ y \end{bmatrix}$$

which implies $My = \lambda x$ and $M^T x = \lambda y$. Then observe

$$\begin{bmatrix} 0 & \frac{1}{c}M \\ \frac{1}{d}M^T & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{c}}x \\ \frac{1}{\sqrt{d}}y \end{bmatrix} = \begin{bmatrix} \frac{1}{c\sqrt{d}}My \\ \frac{1}{d\sqrt{c}}M^T x \end{bmatrix} = \frac{\lambda}{\sqrt{cd}} \begin{bmatrix} \frac{1}{\sqrt{c}}x \\ \frac{1}{\sqrt{d}}y \end{bmatrix},$$

so the eigenvalues of P are λ/\sqrt{cd} where λ ranges over the eigenvalues of A . Note that the largest eigenvalue of A is \sqrt{cd} .

Let μ equal the expression (12), and consider the following cases.

If $\sqrt{c-1} + \sqrt{d-1} \leq \lambda \leq \sqrt{cd}$, then μ is real. Direct computation verifies that, evaluating the expression (12) at $\lambda = \sqrt{cd}$ yields $\mu = 1 = \lambda/\sqrt{cd}$ and $\mu < \lambda/\sqrt{cd}$ for λ in this range. Therefore, in this case the eigenvalue of \tilde{P} always has smaller absolute value than the corresponding eigenvalue of P , implying $\tilde{\rho} \leq \rho$. The lower bound follows from (12) ignoring the square root inside. Thus the first case follows.

If $\lambda < \sqrt{c-1} + \sqrt{d-1}$, then μ is complex, and direct computation shows

$$|\mu|^2 = \frac{1}{\sqrt{(c-1)(d-1)}},$$

so

$$|\mu| = \frac{1}{((c-1)(d-1))^{1/4}}$$

A version of the Alon-Boppana Theorem exists for (c, d) -biregular graphs as well, proven by Feng and Li in [6] (see also [12]).

Theorem 7 ([6]). *Let G be a (c, d) -biregular graph, and let λ be the second largest eigenvalue of the adjacency matrix A of G . Then*

$$\lambda^2 \geq \left(\sqrt{c-1} + \sqrt{d-1}\right)^2 - \frac{2\sqrt{(c-1)(d-1)} - 1}{k}$$

where the diameter of G is greater than $2(k+1)$.

Observe that certainly the diameter is at least $\log_{cd} n$, so that the condition on the degrees and Theorem 7 imply that

$$\lambda^2 \geq 2\sqrt{(c-1)(d-1)}(1 - o(1)).$$

As $|\mu|^2 = \frac{1}{\sqrt{(c-1)(d-1)}}$, so this gives the result for the second case. \square

5 Conclusion

We have looked at non-backtracking random walks from the point of view of walking along directed edges. For the special cases of regular and biregular graphs, our weighted version of Ihara's Theorem (Theorem 4) has given us the complete spectrum of the transition probability matrix for the non-backtracking walk, allowing for easy comparison between the non-backtracking mixing rate, and the mixing rate of the usual random walk. Clearly, it would be desirable to extend these results to more general classes of graphs. The difficulty in applying Theorem 4 directly is with the term involving $\tilde{\tau}^2$. As seen in section 3, $\tilde{\tau}^2$ is a $2m \times 2m$ diagonal matrix with

$$\tilde{\tau}^2((u, v), (u, v)) = \frac{1}{(d_u - 1)(d_v - 1)}.$$

In the case of regular and biregular graphs, this expression is constant (we get $1/(d-1)$ and $1/(c-1)(d-1)$ for the d -regular and (c, d) -biregular cases respectively), making $\tilde{\tau}^2$ simply a multiple of the identity. This allows the difficulty to be handled relatively easily. Regular and biregular graphs are in fact the only graphs for which $\tilde{\tau}^2$ is a multiple of the identity, suggesting that these exact techniques will not work as nicely on more general classes of graphs. If a cleaner version of Theorem 4 could be proven, then, aside from being interesting in its own right, it could potentially be used to extend our results on non-backtracking random walks.

Acknowledgment. The author would like to thank Fan Chung for numerous helpful discussions throughout the process of writing this paper.

References

- [1] N. Alon, I. Benjamini, E. Lubetzky, and S. Sodin, Non-backtracking random walks mix faster, *Communications in Contemporary Mathematics*, **09**, (2007) 585.

- [2] O. Angel, J. Friedman, and S. Hoory, The non-backtracking spectrum of the universal cover of a graph, *Transactions of the American Mathematical Society*, **326**(6), (2015) 4287–4318.
- [3] H. Bass, The Ihara-Selberg zeta function of a tree lattice, *Internat. J. Math.*, **3** (1992), 717–797.
- [4] F. Chung, Laplacians and the Cheeger inequality for directed graphs, *Annals of Combinatorics*, **9** (2005), 1–19.
- [5] F. Chung, *Spectral Graph Theory*, AMS Publications, 1997.
- [6] K. Feng and W.-C. W. Li, Spectra of hypergraphs and applications, *Journal of Number Theory*, 60(1), (1996) 1–22.
- [7] R. Fitzner and R. van der Hofstad, Non-backtracking random walk, *J. Stat. Phys.*, **150**(2), (2013) 264–284.
- [8] K. Hishimoto, Artin-type L -functions and the density theorem for prime cycles on finite graphs, *Internat. J. Math.*, **3** (1992), 809–826.
- [9] Y. Ihara, On discrete subgroups of the two by two projective linear group over p -adic fields, *J. Math. Soc. Japan*, **18** (1966), 219–235.
- [10] M. Kotani and T. Sunada, Zeta functions of finite graphs, *J. Math. Sci. Univ. Tokyo*, **7** (2000), 7–25.
- [11] F. Krzakala, C. Moore, E. Mossel, J. Neeman, A. Sly, L. Zdeborova, and P. Zhang, Spectral redemption in clustering sparse networks, *Proceedings of the National Academy of Sciences*, **110**(52), (2013) 20935–20940.
- [12] W.-C. W. Li and P. Solé, Spectra of regular graphs and hypergraphs and orthogonal polynomials. *European Journal of Combinatorics*, 17(5), (1996) 461–477.
- [13] L. Lovász, Random walks on graphs: a survey, *Combinatorics, Paul Erdős is Eighty (Volume 2)*, Keszthely (Hungary) (1993) pp.1–46.
- [14] A. Nilli, On the second eigenvalue of a graph. *Discrete Math.*, 91 (1991), 207–210.
- [15] H. M. Stark and A. A. Terras, Zeta functions of finite graphs and coverings, *Advances in Mathematics*, **121**, (1996) 124–165.