Abstract

A randomized algorithm for finding sparse cuts is given which is based on constructing a dual markov chain called multiscale rings process (MRP) and a new concept of entropy. It is shown how the time to absorption of the dual process measures the connectedness of the graph and mixing of the corresponding markov process which is then utilized to do clustering. The second algorithm uses the entropy which provides a new methodology and a set of tools to think about sparse cuts as well as sparsification of a graph.

1Email: noraveshrashad@gmail.com
1 Introduction

Partitioning a graph so that each cluster is densely connected with high expansion and clusters are sparsely connected to each other has always been an important problem in several areas such as computer science, big data, machine learning and bioinformatics. There are three major paradigms to deal with such a problem. The first approach comes from approximation algorithms and doing a combinatorial optimization and LP/SDP relaxations and roundings are the main classical tools specially in computer science. The second paradigm arises from the desire to generalize cheeger inequality and there are many important attempts such as (Louis et al. 2012),(Lee et al. 2014). The last paradigm which is relatively less appreciated is converting the problem to another problem that has the goal of finding all possible ways to reduce mixing time within each cluster. The absence of bottlenecks in the state space of a Markov chain would imply rapid mixing. Research on this paradigm is concentrated on some concepts as a proxy to mixing time. One approach is to use sparsification and electrical networks (Spielman & Srivastava 2008) and in the intrinsic level it uses hitting times. Another approach is dealing with mixing time of a markov chain (Andersen & Peres 2009),(Morris & Peres 2005). A very important article for showing how different concepts like mixing times, hitting times are related to each other is discovered in (Peres & Sousi 2011).

In the present paper the shortest path metric on a graph is denoted by $d_G$. The $\epsilon-$mixing time is defined as

$$\tau(\epsilon) := \min \{n : ||p^n(x, \cdot) - \pi|| \leq \epsilon \text{ for all } x \in V\} \quad (1.1)$$

(Morris & Peres 2005) is motivated by (László & Kannan 1999) who proved that

$$\tau(1/4) \leq 2000 \int_{\pi_u}^{3/4} \frac{du}{u\Phi^2(u)} \quad (1.2)$$

There are many approaches to bounding the mixing time. One of them is using relaxation time as a proxy for mixing time. In fact markov chains can be characterized by some important concepts such as hitting time,cover time,relaxation time and mixing time and they are all related to each other. For example (Oliveira & Peres 2019) proved an upper bound for hitting time:

$$t_{hit} \leq \frac{20d_{avg}}{d_{min}}n\sqrt{t_{rel}} + 1 \quad (1.3)$$

Another approach is to use coupling and there are many examples such as graph coloring and finding independent sets based on the idea of cou-
pling and more specifically path coupling. One of the most natural ideas to bound mixing time goes back to idea of (Aldous & Diaconis 1987) who defined strong stationary times and later (Diaconis & Fill 1990) constructed a dual Markov chain and analyzed the time to absorption of the dual process and later (Morris & Peres 2005) developed this idea and then used it for finding sparse cuts locally and naming this dual process the evolving set process (ESP). (Andersen & Peres 2009) further developed the idea to construct a non-vanishing process called volume biased evolving set process (VB-ESP). (Andersen & Peres 2009) proved that the cardinality of evolving set process is a martingale.

It can be easily observed that the probability of hitting 2k within 2n time steps for simple random walk is

\[ P(S_{2n} = 2k | S_0 = 0) = \binom{2n}{n+k} p^{n+k} q^{n-k}, \quad -n \leq k \leq n \] (1.4)

Hoeffding’s theorem states that, for all k > 0

\[ P(S_n - E(S_n) \geq k) \leq \exp\left(-\frac{2k^2}{\sum_{i=1}^{n} (b_i - a_i)^2}\right) \] (1.5)

The probability that the random walk on \( \mathbb{Z} \) is at k after nth step is a direct consequence of Hoeffding’s inequality

\[ \sum_{|k| \geq d} q_n(k) \leq 2e^{-d^2/(2n)} \] (1.6)

(Carne & Varopoulos 1985) proved the following theorem

**Theorem 1.1.** Denote by \( P \) the \( L^2(\mu) \) operator associated to the transition kernel \( p \) and let \( |P| \) stand for its norm which is always less than or equal to one. Then:

\[ p^t(x, y) \leq 2\left(\frac{\mu(y)}{\mu(x)}\right)^{1/2} |P|^t \exp\left(-\frac{d(x, y)^2}{2t}\right) \] (1.7)

2 Construction of MRP

Subordination is an important tool to classify and analyze Markov processes such as Brownian motion (Kim et al. 2020) which is also extended to the discrete case in (Bendikov & Saloff-Coste 2012) and many different variations of it are defined in the literature for different purposes such as (Mimica 2015). Here the following definition is used in the present paper to define MRP at any scale.
Definition 1. A subordinated random walk is defined as
\[ Y_n := S_{\tau_n} \] (2.1)
where \( S_n = x + X_1 + X_2 + \ldots + X_n \) is a simple random walk starting at
\( x \in \mathbb{Z}^d \) and \( \tau_n = R_1 + R_2 + \ldots + R_n \) is a random walk on \( \mathbb{Z}_+ \) with independent
increments \( R_i \) having a probability distribution.

Definition 2. Multiscale Rings Process (MRP) at scale \( s \) is a subordinated
random walk defined in 2.1 while \( \tau_n \) is stopped at a fixed threshold time \( \tau_n^{(s)} \).
The scale \( s \) is induced by the distribution of \( R_i \) (increments on \( \mathbb{Z}_+ \)).

Definition 3. A ring at position \( r \) of scale \( s \) is defined by the set of points
inside the following zone
\[ \text{ring}_s(r) = \{ x : r \leq x \leq r + 2^s \} \] (2.2)

Figure 1: a random walk starting from the center and the rings \( \Omega_i \)
**Definition 4.** Solar system associated to a center node is the union of all rings up to a point whose shortest path distance to center node is bounded by $10 \times 2^s$ where $s$ is the fixed length scale of all rings at scale $s$.

A solar system is depicted in Figure 1. The algorithm starts by creating a metric embedding into $\ell^2$ using any favorite algorithm. Note that there may be a tradeoff between distortion of metric embedding and the time complexity of creating such an algorithm. After creating the metric embedding and representing it as a polar transformation, the solar system associated to a center is created by ten rings.

**Definition 5.** Any ring is divided into equal number of blocks where each block has only two neighbor blocks in that ring and has two other neighbors from previous and next rings. $\text{Ring}(r) = \bigcup_{i=1}^{n_b} B_i^{(r)}$

Note that there is no constraint on the graph to be planar and the blocks of a ring can be connected to each other by some edges in tangential direction but the shortest path distance in radial direction forces each block of a fixed ring to be connected to blocks of other rings by at most one block away and this can be done by choosing an appropriate scale of the ring. The algorithm then chooses a random set (call it $U_s(r)$) for all blocks of the solar system. The algorithm chooses randomly until the vertices inside each $U_s(r)$ produces a connected subgraph. Consider a block at ring$_s(r)$, and define the exterior boundary of it as follows:

$$B^{(r)}_\partial = \{ x \in V : d_G(x, B^{(r)}) = 1, x \notin B^{(r)} \}$$ (2.3)

The probability that a random walk on weighted graph starting from vertex $i$ hits $U_s(r)$ before it hits the boundary $B^{(r)}_\partial$ is denoted by $q(i, U_s(r), B^{(r)}_\partial) := P_i[T_U < T_{\partial B}]$. It is known that this type of function is a harmonic function and can be computed efficiently using a linear solver (Barlow 2017). Thus, imagine the function $q$ on graph to be the solution of the following linear set of equations

$$\sum_j (q(i) - q(j)) \frac{w_{ij}}{w_j} = 0 \text{ if } i \in B^{(r)} \setminus U_s(r)$$

$$q(j) = \begin{cases} 1, & \text{if } j \in U_s(r) \\ 0, & \text{if } l \in B^{(r)}_\partial \end{cases}$$ (2.4)

**Definition 6.** Relative absorption to a set $U$ is a measure of absorption of a random walk to $U$ as follows

$$RA = \sum_i \frac{q(i)}{d(i, B^{(r)})}$$ (2.5)
Since $RA$ is calculated for each block, a fixed threshold for all blocks can be set to check if $RA$ is less than the allowed threshold $RA_{al}$. The algorithm checks if relative absorption to $U_s(r)$ is smaller than $RA_{al}$ and in the cases that it is so we call these $U_s(r)$, safe sets. One can think that the whole process is like a percolation of light from the center of solar system until it reaches the farthest ring in the solar system. So the probability that light passes through some rings and reaches the latest ring can be approximated by a geometric distribution. The approximation is exact if $RA$ is the same for all rings. The sets $U_s(r)$ are merged to each other if their mutual conductances is higher than a threshold $\phi_{al}$ which controls the sparsity among these sets. After merging many sets together, some giant components appears which we call them long radial constellations and can be considered as a cluster for the output of the algorithm. Apart from these long radial constellations, some clusters are formed by merging safe sets. Since each solar system covers only a subset of vertices of the graph, the algorithm repeats for many different solar centers and can be done efficiently in parallel.

**Definition 7.** galaxy is the sum of disjoint solar systems.

The algorithm then outputs the set of clusters by union of clusters from different solar systems.

### 3 MRP as a dual process

**Definition 8.** The green function upto ring $r$ is the green function of the random walk which is killed out of set $A_r$ and is denoted by

$$G_{A_r}(x, y) = \sum_{n=0}^{\infty} p_n^{A_r}(x, y)$$

(3.1)

where $A_r = \bigcup_{i=1}^{r} \text{Ring}(i)$

**Lemma 3.1.** The green function upto ring $r$ for a recurrent solar system for center $c$ satisfies

$$G_{A_r}(c, c) = \frac{1}{\omega(c) P_c(\tau_{A_r} \leq T_c^+)}$$

(3.2)

**Proof.** Lets define the local time of the Markov chain on $c$ to be

$$L_c^e_n = \sum_{r=0}^{n-1} 1_{\{X_r = c\}}$$

(3.3)
Observe that
\[ E^x L^c_{\tau_{A_r}} = \sum_{n=0}^{\infty} P^x(X_n = c, \tau_{A_r} > n) = g_{A_r}(x,c) \omega_c \quad (3.4) \]

Strong Markov property implies that local time on \( c \) has a geometric distribution, Thus:
\[ \omega_c g_{A_r}(c,c) = E^c L^c_{\tau_{A_r}} = \sum_{k=1}^{\infty} k(1 - P^c(T^+_c \leq \tau_{A_r}))^{k-1} \]
\[ = \frac{1}{1 - P^c(T^+_c \leq \tau_{A_r})} = \frac{1}{P^c(\tau_{A_r} \leq T^+_c)} \quad (3.5) \]

The proof of the following two lemma can be found for example in [Barlow 2017]

**Lemma 3.2.** Suppose \( X_j \) is the random walk on the solar system, \( A_i = \text{Ring}(i), B_i = \text{Ring}(i) \cup \text{Ring}(i + 1) \), \( \tau_i = \inf\{j \geq 1 : X_j \in A_i\} \) and \( \bar{\tau}_i = \inf\{j \geq 0 : X_j \notin B_i\} \), then
\[ P^x\{\bar{\tau}_i \leq \eta_i\} = \sum_{y \in A_i} G_{B_i}(x,y) P^y\{\tau_i > \eta_i\} \quad (3.6) \]

holds for all rings \( i = 1, \ldots, n_r \)

**Proof.** let \( \sigma_i = \sup\{j : X_j \in A_i, j \leq \eta_i\} \) then
\[ P^x\{\bar{\tau}_i \leq \eta_i\} = \sum_{j=0}^{\infty} P^x\{\sigma = j\} \]
\[ = \sum_{y \in A_i} \sum_{j=0}^{\infty} P^x\{\sigma = j, X_j = y\} \quad (3.7) \]

Simple markov property produces
\[ = \sum_{y \in A_i} \sum_{j=0}^{\infty} P^x\{X_j = y; j \leq \eta_i; X_k \notin A_i; j < k \leq \eta_i\} \]
\[ = \sum_{y \in A_i} \sum_{j=0}^{\infty} P^x\{X_j = y; j \leq \eta_i\} P^y\{\tau_i > \eta_i\} \quad (3.8) \]
A simple rearrangement shows that

\[
P_x \{ \tilde{\tau}_i \leq \eta_i \} = \sum_{y \in A_i} P_y \{ \tau_i > \eta_i \} \sum_{j=0}^{\infty} P_x \{ X_j = y; j \leq \eta \}
\]

\[
= \sum_{y \in A_i} P_y \{ \tau_i > \eta_i \} G_{B_i}(x, y)
\]

Lemma 3.3. Let \( S_n \) be a simple symmetric random walk on \( \mathbb{Z} \) with \( S_0 = 0 \) then

\[
\sum_{r \in \mathbb{Z}} \lambda^r \mathbb{P}(S_n = r) = 2^{-n} \sum_{r=0}^{n} \lambda^{2r-n} \binom{n}{r} - 1 \leq t \leq 1
\]

Multiplying chebyshev polynomials \( H_k(t) = \frac{1}{2}(t + i(1 - t^2)^\frac{1}{2})^k + \frac{1}{2}(t - i(1 - t^2)^\frac{1}{2})^k \) by probability that random walk hits \( k \) and summing over all \( k \) has the following equality

Lemma 3.4. For each \( k \), \( H_k(t) \) is a real polynomial in \( t \) of degree \( k \) and \( |H_k(t)| \leq 1 \), Further, for each \( n \geq 0 \)

\[
t^n = \sum_{k \in \mathbb{Z}} \mathbb{P}(S_n = k) H_{|k|}(t)
\]

The importance of Lemma 3.4 is that the scalar can be substituted by the transition probability matrix of the random walk on weighted graph:

\[
P^n = \sum_{k \in \mathbb{Z}} \mathbb{P}(S_n = k) H_{|k|}(P)
\]

\[3.12\] shows the duality between the primary process and the random walk on \( \mathbb{Z} \) which is also used for proof of Carne–Varopoulos bound in Theorem 1.1.

Now \( p_n(x, y) \) can be written as

\[
p_n(x, y) = \langle f_1, P^n f_2 \rangle
\]

\[
= \sum_{k \in \mathbb{Z}} \mathbb{P}(S_n = k) \langle f_1, H_{|k|}(P) f_2 \rangle
\]

where \( f_i(x) = 1_{x_i \mu x_i^2} \) for \( i = 1, 2 \) This duality is too strong and it is better to work with some expectations(such as characteristics for real random variable and generating functions for discrete random variables) rather than
transition probabilities. In fact in the present paper, convergence in distribution (the weakest form of convergence) is the main tool. It is well known in the theory of discrete subordination of random walk, any Bernstein function has a Lévy–Khintchine representation. (Schilling et al. 2012)

\[
\phi(x) = a + bx + \int_0^\infty (1 - e^{xy}) \nu(dy), \ x \geq 0
\]

(3.14)

where \( \nu \) is the radon measure.

4 Many sparse cuts

As mentioned in the introduction, algorithm 1 is a randomized algorithm that might be practical for cases that involves big data that a parallel algorithm is essential. This algorithm can be combined with evolutionary algorithms to find semi-optimum clusters since each run of algorithm produces a measure such as conductance for the quality of clustering.

Algorithm 1 outputs many clusters

Input: A weighted undirected graph with \( n \) vertices and \( m \) edges

loop over \( n_c \) centers:

loop for different randomizations:

1: create a metric embedding into \( \ell^2 \) and save records in polar coordinates inside the database
2: create the associated solar system
3: solve the linear equations associated to blocks of each ring in parallel
4: check if relative absorption to random sets in step 5 is smaller than the threshold \( RA_{al} \), if it is smaller, label the random set as safe set.
5: merge safe sets if the conductance between them is higher than a threshold \( \phi_{al} \)

Output: giant components and merged safe sets of all solar systems

5 Entropy

The concept of optimal transport time or sometimes called access time from \( \mu \) to \( \nu \) is defined by (László & Winkler 1995)

\[
H(\mu, \nu) := \inf_{\tau \in \Gamma} \mathbb{E}[\tau]
\]

(5.1)
where $\Gamma$ is the set of stopping times that transport $\mu$ to $\nu$. This problem has a close relation with skorokhod embedding if one replaces brownian motion with a general discrete markov chain. (Beiglboeck et al. [2017]) and a practical tool to solve these problems are by PDE and free boundary problem approaches which can be efficiently solved by computers.(Ghoussoub et al. [2019])

Now consider the following function

$$V(x) = \sup_{\tau} \mathbb{E}_x(M(X_{\tau})) + \int_0^\tau L(X_t)dt + \sup_{0 \leq t \leq \tau} K(X_t) + F(x)$$ \tag{5.2}$$

Important quantities such as (5.1) are just a special case of (5.2) when the functions $L, K, F$ are zero and $M = -\tau$. Another important special case is average value of a process $Y_t = L(X_t)$ over time and is related to large deviation theory for markov chains and ergodic theory:

$$A_t = \frac{1}{T} \sum_{i=1}^T Y_i$$ \tag{5.3}$$

where $Y_i$ measures how far the markov chain is from a fixed reference vertex $x$ and the corresponding event can be approximated by rate function $I(a)$ as follows:

$$\mathbb{P}(A_t = a) \approx \exp^{-TI(a)}$$ \tag{5.4}$$

One example that is also related to American put option is when (5.2) takes the following special form

$$V(x) = \sup_{\tau} \mathbb{E}_x e^{-r\tau}(K - X_{\tau})^+$$ \tag{5.5}$$

where $K$ is the strike price, $r$ is the interest rate, and it can easily be reduced to free boundary value problem and the corresponding optimal stopping time turns out to exist and is

$$\tau_* = \inf\{0 \leq t \leq T : X_t \leq b_*(t)\}$$ \tag{5.6}$$

Note that function $F(x)$ in (5.2) is independent of time and stopping time. A very important example is when $F(x)$ is a probability that the the discrepancy between shortest path distance between a fixed vertex and a subset $\Omega$ and the time it takes for the random walk to hit $\Omega$, is $C$ times greater than shortest path distance as follows

$$F(x) = \mathbb{P}(T(x, \Omega) - d(x, \Omega) > Cd(x, \Omega))$$ \tag{5.7}$$
The methodology of algorithm 2 for clustering is based on including the vertices as much as possible to increase entropy within each cluster. Two concepts of entropy namely $\alpha$-entropy and $\beta$-entropy is defined. The former is difficult to compute in practice while the latter is easy to calculate, it has no limit, and is used in algorithm 2. The relation between these two entropies is an open question. Note that at each moment of time, the random walk on the weighted graph belongs to one or more subsets of the state space of the markov chain. Thus an example of trajectory of the dynamics of a particle going by the random walk could be like $\{\ldots, \Omega_3, \Omega_2, \Omega_2, \Omega_4, \Omega_1, \Omega_2, \ldots\}$. So at each time the random walk could belong to any of these subsets. For simplicity and without lack of generality assume these subsets are balls of some fixed radius in the shortest path metric on the normalized weighted graph. Thus, $\Omega$ is shown in figure 2 and is decomposed as follows:

$$\Omega = \bigcup_{i=1}^{n} \Omega_i$$

(5.8)

Figure 2: a random walk on weighted graph and the covering $\Omega_i$
Observer that as $n$ is increased in (5.8), the number of balls are increased to cover all vertices and balls have smaller radius.

Definition 9. $\alpha$-entropy is defined as follows

$$H(\alpha) = \lim_{n \to \infty} -\frac{1}{n} \sum_{i_1 \ldots i_n} \mathbb{P}(\Omega_{i_1}, \Omega_{i_2}, \ldots, \Omega_{i_n}) \log \mathbb{P}(\Omega_{i_1}, \Omega_{i_2}, \ldots, \Omega_{i_n})$$

(5.9)

Definition 10. $\beta$-entropy is defined as follows

$$H(\beta) = \frac{1}{n} \sum_{i=1}^{n} V_i(x) \log(V_i(x))$$

(5.10)

while functions $V_i$ corresponds to (5.2) where function $M = d(x, X_r)$, $K$ is zero, and $L_i$ is associated with $V_i$ and is defined as follows

$$L_i(X_t) = 1(X_t \in \Omega_i), i = 1, \ldots, n$$

(5.11)

Define the following stochastic process:

$$B_N^n = M_N$$

$$B_N^n = \max\{M_n, \mathbb{E}(B_N^{n+1} | \mathcal{F}_n)\}, \quad n = N - 1, N - 2, \ldots$$

(5.12)

The importance of (5.12) is that it is the smallest supermartingale that majorizes $M_k$ for $n \leq k \leq N$. The following class of stopping times is used in the next lemma.

$$\tau_N^n = \inf\{k \in \{n, n+1, \ldots, N\} : B_N^k = M_k\}$$

(5.13)

Lemma 5.1. Suppose $M = (M_k)_{k=n}^{N}$ satisfies $\mathbb{E}\max_{n \leq k \leq N} M_k < \infty$, then the stopping time defined in (5.13) is optimal in (5.2).

Proof. First the proof sketch is described. Observe that it suffices to show that $B_N^n$ is the smallest supermartingale that majorizes $M$ and $B_N^n$ is martingale for $n \leq k \leq N$ and then Doob’s optional stopping theorem applied to this martingale shows that

$$B_n^n \geq \mathbb{E}(M_r | \mathcal{F}_n)$$

$$B_n^n = \mathbb{E}(M_{\tau_N^n} | \mathcal{F}_n)$$

(5.14)

Taking expectations of both sides of (5.14) gives

$$\mathbb{E}M_r \leq \mathbb{E}M_n^n = \mathbb{E}M_{\tau_N^n}$$

(5.15)
which shows that $\tau_n^N$ is the optimal stopping time. Now that the proof sketch is given, the pieces of it is proven as follows: The following relation proves that if $B$ is another supermartingale that majorizes $M$, it is not the smallest supermartingale.

$$\bar{B}^N_{k-1} \geq \max\{M_{k-1}, \mathbb{E}B^N_k | \mathcal{F}_{k-1}\} \geq \max\{M_{k-1}, \mathbb{E}B^N_k | \mathcal{F}_{k-1}\} = B^N_{k-1} \quad (5.16)$$

The next relationship proves that $B^N_{\tau^N_n \wedge k}$ is martingale.

$$\mathbb{E}[B^N_{\tau^N_n \wedge (k+1)} | \mathcal{F}_k] = B^N_{\tau^N_n \wedge (k)} 1_{\{\tau^N_n \leq k\}} + B^N_k 1_{\{\tau^N_n > k\}} = B^N_{\tau^N_n \wedge k} \quad (5.17)$$

Algorithm 2 outputs clusters by entropy maximization

Input: A weighted undirected graph with $n$ vertices and $m$ edges

loop:
1: generate random sets $K_l$ of vertices of the graph each of fixed cardinality
2: calculate $\beta$-entropy for all $K_l$
3: sort entropies in step 2
Output: top high entropy sets in step 3

6 Conclusion

Since the second largest eigenvalue of transition matrix and cheeger inequality characterize the markov chain on graph only globally, the present algorithm is designed for big data problems where computing the eigenvectors and eigenvalues of a big graph is time consuming and not robust to variations inside the graph.

References

Aldous, D. & Diaconis, P. (1987), ‘Strong uniform times and finite random walks’, Advances in Applied Mathematics 8(1), 69-97.

Andersen, R. & Peres, Y. (2009), ‘Finding sparse cuts locally using evolving sets’, ArXiv abs/0811.3779.

Barlow, M. T. (2017), Random Walks and Heat Kernels on Graphs, London Mathematical Society Lecture Note Series, Cambridge University Press.

Beiglboeck, M., Huesmann, M. & Cox, A. (2017), ‘Optimal transport and skorokhod embedding’, Inventiones mathematicae 208.
Bendikov, A. & Saloff-Coste, L. (2012), ‘Random walks on groups and discrete subordination’, *Mathematische Nachrichten* 285.

Carne, T. K. & Varopoulos, N. T. (1985), ‘A transmutation formula for markov chains’, *Bulletin Des Sciences Mathematiques* 109, 399–405.

Diaconis, P. & Fill, J. A. (1990), ‘Strong Stationary Times Via a New Form of Duality’, *The Annals of Probability* 18(4), 1483 – 1522.

Ghoussoub, N., Kim, Y.-H. & Palmer, A. (2019), ‘Pde methods for optimal skorokhod embeddings’, *Calculus of Variations and Partial Differential Equations* 58, 113.

Kim, P., Song, R. & Vondraček, Z. (2020), ‘On the boundary theory of subordinate killed lévy processes’, *Potential Analysis* 53.

Lee, J. R., Gharan, S. O. & Trevisan, L. (2014), ‘Multiway spectral partitioning and higher-order cheeger inequalities’, *J. ACM* 61(6).

Louis, A., Raghavendra, P., Tetali, P. & Vempala, S. (2012), Many sparse cuts via higher eigenvalues, in ‘Proceedings of the Forty-Fourth Annual ACM Symposium on Theory of Computing’, STOC ’12, Association for Computing Machinery, p. 1131–1140.

László, L. & Kannan, R. (1999), ‘Faster mixing via average conductance’.

László, L. & Winkler, P. (1995), ‘Fast mixing in a markov chain’.

Mimica, A. (2015), ‘On subordinate random walks’, *Forum Mathematicum* 29.

Morris, B. & Peres, Y. (2005), ‘Evolving sets, mixing and heat kernel bounds’, *Probability Theory and Related Fields* 133, 245–266.

Oliveira, R. I. & Peres, Y. (2019), Random walks on graphs: new bounds on hitting, meeting, coalescing and returning, in M. Mishna & J. I. Munro, eds, ‘Proceedings of the Sixteenth Workshop on Analytic Algorithmics and Combinatorics, ANALCO 2019, San Diego, CA, USA, January 6, 2019’, SIAM, pp. 119–126.

Peres, Y. & Sousi, P. (2011), ‘Mixing times are hitting times of large sets’, *Journal of Theoretical Probability* 28.

Schilling, R. L., Song, R. & Vondracek, Z. (2012), *Bernstein Functions: Theory and Applications*, De Gruyter.

Spielman, D. A. & Srivastava, N. (2008), Graph sparsification by effective resistances, STOC ’08, Association for Computing Machinery, New York, NY, USA, p. 563–568.