Sankhyā B : The Indian Journal of Statistics
https://doi.org/10.1007/s13571-021-00258-x

Convergence Details About k-DPP Monte-Carlo Sampling for Large Graphs

Diala Wehbe and Nicolas Wicker
University of Lille, Villeneuve D'Ascq, France
Diala Wehbe
Lebanese University, Tripoli, Lebanon

Abstract

This paper aims at making explicit the mixing time found by Anari et al. (2016) for k-DPP Monte-Carlo sampling when it is applied on large graphs. This yields a polynomial bound on the mixing time of the associated Markov chain under mild conditions on the eigenvalues of the Laplacian matrix when the number of edges grows.

AMS (2000) *subject classification.* Primary 05Cxx; Secondary 11Kxx.
Keywords and phrases. Determinantal point process, Kernel, Laplacian Kernel, Metropolis-Hastings, Markov chain, Mixing time

1 Introduction

Determinantal Point Processes, called DPPs, were identified by Macchi (1975) to have the ability to model the position of fermions, where nearby particles repel each other. They supply coherent probabilistic models in the presence of negative correlation. Recently, DPPs have developed a growing interest in statistics applications for machine learning such as finding the best translation of a sentence, computing the marginals of a Markov random field and modeling non-overlapping human poses in images or video (Kulesza and Taskar, 2012). A DPP is defined by a kernel L, basically a symmetric and positive semidefinite matrix. For an integer $k \leq n$, Kulesza and Taskar (2011) introduced the k-DPP, which requires a standard DPP on the event that the modeled set is of fixed size.

Kannan and Vempala (2009) & Deshpande and Rademacher (2010), and Kulesza and Taskar (2012) have presented algorithms for sampling from k-DPP. However, for a large dataset, these algorithms are inefficient to use since eigen-decomposition of a possibly huge matrix is needed. On the other hand, the Markov chain techniques are very attractive because of their plainness and effectiveness in generating k-DPP random samples. For instance, the Metropolis-Hastings algorithm was used by Kang (2013) who has considered a Markov chain Monte Carlo (MCMC) to sample from k-DPPs.

Anari et al. (2016) have given an accurate polynomial bound on the mixing time depending only the probability of the starting point which can deteriorate hugely the bound. Here, we make the bound precise in the specific case when k-DPPs are used to sample a diverse set of nodes in connected graphs. This has applications in video games to place randomly and with repulsion ressources or monsters. The kernel we use is the Moore-Penrose pseudo-inverse of the normalized Laplacian which is a classical kernel for graphs so that the bound we derive is intrinsically related to the Laplacian.

This paper includes the following sections: Section 2 gathers some basic facts about DPP, and Laplacian matrix properties. In Section 3, we define the k-DPP kernel and present the Markov Chain of Kang (2013) to efficiently generate random samples. Our main result is presented in Section 4 where we characterize the Laplacian eigenvalues and we focus on the mixing time of the Markov chain for the k-DPP.

2 Background

2.1. Determinantal Point Processes For a discrete set $\mathcal{X} = \{1, \cdots, n\}$ of n items, a DPP denoted $\mathcal{P}_L$, is a probability measure on the set $2^{[n]}$ of all subsets of $[n] = \{1, \cdots, n\}$. A DPP can be defined via an L-ensemble (Kulesza and Taskar, 2012) by using a $n \times n$ symmetric and positive semidefinite matrix L indexed by the elements of $\mathcal{X}$ such that if Y is a random subset drawn according to $\mathcal{P}_L$, we have

$$\mathcal{P}_L(Y) = \frac{\det(L_Y)}{\det(L+I)};$$

where for $i, j \in Y$, $(L_Y)_{ij} = L_{Y_i Y_j}$ and I is the $n \times n$ identity matrix. Note that the required normalization constant can be given in closed form for the reason that

$$\sum_{Y \subseteq \mathcal{X}} \det(L_Y) = \det(L+I).$$

Kulesza and Taskar (2011) propose an extension of DPPs that enables to model the content of a selected number of items. They obtained thus a k-DPP by conditioning a standard DPP so that the set Y has cardinality k.

The probability of Y is then denoted $\mathcal{P}_L^k$ and stands as follows:

$$\mathcal{P}_L^k(Y) = \frac{\det(L_Y)}{\sum_{|Y'|=k} \det(L_{Y'})}, \tag{2.1}$$

where $|Y| = k$ and L_Y is a $k \times k$ positive semidefinite matrix.

2.2. About the Graph Laplacian Let us consider an undirected unweighted graph $G = (V, E)$ where $V(G)$ and $E(G)$ are respectively the vertex and the edge set of G. Several kernels have effectively been used to capture the similarity between a pair of nodes induced by the local structure of the graph. One of them is based on the combinatorial Laplacian matrix of the graph which is defined as

$$L_G = D - A,$$

where the elements of the adjacency matrix A of the graph are: for all $i, j \in V(G)$

$$A_{ij} = \begin{cases} 1 & \text{if } i \sim j \\ 0 & \text{otherwise} \end{cases}$$

and $D = \operatorname{diag}(d_1, \cdots, d_n)$, with $d_i = \sum_j A_{ij}$.

On occasion, it is also known as the Kirchhoff matrix or the information matrix. Similar to the combinatorial Laplacian, the normalized Laplacian yields eigenvalues which do not depend on the graph size, it is given by:

$$\mathcal{L}_G := D^{-1/2} L_G D^{-1/2} = I - D^{-1/2} A D^{-1/2}.$$

For a graph G we can see that, for all $i, j \in V(G)$

$$\mathcal{L}_{ij} = \begin{cases} 1 & \text{if } i = j \text{ and } d_j \neq 0 \\ -\frac{1}{\sqrt{d_i d_j}} & \text{if } i \sim j \\ 0 & \text{otherwise} \end{cases}$$

As mentioned by Chung (1997) & Cvetković et al. (2010), $\mathcal{L}_G = (\mathcal{L}_{ij})_{i,j \in V(G)}$ is a symmetric, positive semidefinite matrix, and its eigenvalues $\lambda_1\big(\mathcal{L}_G\big), \cdots,$ $\lambda_n\big(\mathcal{L}_G\big)$ satisfy

$$0 = \lambda_1\big(\mathcal{L}_G\big) \leq \lambda_2\big(\mathcal{L}_G\big) \leq \cdots \leq \lambda_n\big(\mathcal{L}_G\big) \leq 2.$$

The eigenvalues can be used to provide effective information about a graph. For instance, the number of connected components can be obtained from the multiplicity of the eigenvalue 0. If the graph G is connected then 0 is an eigenvalue with multiplicity one. The second smallest eigenvalue, discussed in next section, is an indicator of bottlenecks in the graph.

2.2.1. On the Second Smallest Normalized Laplacian Eigenvalue. Different types of studies were presented by Li et al. (2008) & Li and Li (2011) and Li et al. (2014) on the second smallest normalized Laplacian eigenvalue. Since $\mathcal{L}_G$ is symmetric and $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n \leq 2$, then its eigenvalues are all real and nonnegative. We recall some properties of the eigenvalues and eigenfunctions of $\mathcal{L}_G$ by using the variational characterization of those eigenvalues in terms of the Rayleigh quotient of $\mathcal{L}_G$.

Let G be a graph, $G = (V, E)$, and g be an eigenvector of $\mathcal{L}_G$. Then, g can be viewed as a function that assigns a real value $g(v)$ to each vertex v of G. By letting $g = D^{1/2}f$, we have

$$\frac{g^T \mathcal{L}_G g}{g^T g} = \frac{f^T D^{1/2} \mathcal{L}_G D^{1/2} f}{(D^{1/2}f)^T D^{1/2} f} = \frac{f^T L_G f}{f^T D f} = \frac{\sum_{u,v \in E(G)} (f(u) - f(v))^2}{\sum_{v \in V(G)} d(v)(f(v))^2}. \tag{2.2}$$

It is easy to deduce that 0 is an eigenvalue of $\mathcal{L}_G$ and $D^{1/2}e$ is an eigenfunction corresponding to 0 where e denotes the constant function which takes the value 1 on each vertex. Thus, by using Eq. (2.2) we can obtain the following formula corresponding to the second smallest normalized Laplacian eigenvalue of a graph G:

$$\lambda_2\big(\mathcal{L}_G\big) = \inf_{f \perp De} \frac{f^T L_G f}{f^T D f} = \inf_{f \perp De} \frac{\sum_{u,v \in E(G)} (f(u) - f(v))^2}{\sum_{v \in V(G)} d(v)(f(v))^2},$$

where f is called a harmonic eigenfunction associated with λ_2.

Moreover, $\lambda_2\big(\mathcal{L}_G\big)$ is closely related to the discrete Cheeger constant (Cheeger, 1970). For all non-empty subset $S \subset V(G)$ whose complement is $\overline{S} = V(G) - S$, we define

$$h_G(S) = \frac{\sum\limits_{i \in S, j \in \overline{S}} A_{ij}}{\min\Big(\sum\limits_{i \in S} d_i, \sum\limits_{j \in \overline{S}} d_j\Big)},$$

where A_{ij} are the elements of the adjacency matrix. The *Cheeger constant* h_G of a graph G is defined to be

$$h_G = \min_{S \subset V(G),\ S \neq \emptyset} h_G(S).$$

Here below, the Cheeger inequality makes the connection between $\lambda_2(\mathcal{L}_G)$ and the tightness of the bottleneck explicit.

Theorem 1. *(Cheeger Inequality* (Cheeger, 1970)*) If G is connected, then*

$$\frac{h_G^2}{2} \leq \lambda_2\big(\mathcal{L}_G\big) \leq 2h_G,$$

where h_G is the Cheeger constant of G.

2.2.2. The Moore-Penrose Pseudo-inverse of the Normalized Laplacian Matrix. The concept of pseudo-inverse generalizes the matrix inverse to matrices that are not of full rank or square. In fact, when the number of paths that link two nodes increases, the entries of the pseudo-inverse of the Laplacian matrix increase as well. This matrix represents the similarity between any pair of nodes. Then, it is motivating to use the pseudo-inverse of the normalized Laplacian matrix of the graph as a Gram matrix. This is a convenient way of defining a kernel on a graph (Lovász, 1993). The Moore-Penrose pseudo-inverse of the normalized Laplacian matrix $\mathcal{L}_G$ will be denoted $\mathcal{L}_G^\dagger$. Some of the important properties of $\mathcal{L}_G^\dagger$ are:

- $\mathcal{L}_G^\dagger$ is symmetric and positive-semidefinite.
- If $(V_i,\ \lambda_i \neq 0)$ are (eigenvectors, eigenvalues) of $\mathcal{L}_G$, then $(V_i,\ \lambda_i^{-1} \neq 0)$ are the corresponding (eigenvectors, eigenvalues) of $\mathcal{L}_G^\dagger$.
- If $(V_i,\ \lambda_i = 0)$ are (eigenvectors, eigenvalues) of $\mathcal{L}_G$, then are also (eigenvectors, eigenvalues) of $\mathcal{L}_G^\dagger$.

Interestingly, the pseudo-inverse of the Laplacian matrix can be used to determine the average commute time, which is defined as the average number of steps taken by a random walker when starting from node i to reach node j and returning to node i. Let $\mathrm{vol}(G)$ denotes the volume of the graph, the average commute time can be computed as follows: Gobel and Jagers (1974)

$$n(i,j) = \mathrm{vol}(G) \cdot \Big(\big(L_G^\dagger\big)_{ii} + \big(L_G^\dagger\big)_{jj} - 2\big(L_G^\dagger\big)_{ij} \Big)$$

with $\mathrm{vol}(G) = \sum_{l=1}^{n} d_l$.

Finally, the following interlacing eigenvalues lemma is well-known in matrix analysis. It will be applied in this paper on $\mathcal{L}_G^\dagger$. First, let us denote $\lambda_k(A)$ as the k-th smallest eigenvalue of A.

Lemma 1. *Horn and Johnson (1985) Let M be a real symmetric matrix and M_r denotes any r-by-r principal submatrix of M. For any integer k such that $1 \leq k \leq r$ we have*

$$\lambda_k(M) \leq \lambda_k(M_r) \leq \lambda_{k+n-r}(M).$$

3. Sampling via k-DPP

Henceforth, we will use the Moore-Penrose pseudo-inverse of the normalized Laplacian matrix as kernel which is given by:

$$L(i,j) = \big(\mathcal{L}_G^\dagger\big)_{ij}. \tag{3.1}$$

Kang (2013) presented an efficient method to construct a Markov chain with the stationary distribution $\mathcal{P}_L^k$ generated by the Metropolis-Hastings algorithm (Chib and Greenberg, 1995). The main idea behind this method is to suggest a new configuration by choosing two elements x and y: x to be removed from the current set X of size $|X| = k$, and y to be added. Hence, for $X = Y \cup \{x\}$ and $y \notin X$, the acceptance probability of removing x from X and replacing it with y is computed as follows:

$$\frac{1}{2}\min\left\{1, \frac{\det L_{Y\cup\{y\}} \cdot q\big(Y \cup \{y\}, Y \cup \{x\}\big)}{\det L_{Y\cup\{x\}} \cdot q\big(Y \cup \{x\}, Y \cup \{y\}\big)}\right\} = \frac{1}{2}\min\left\{1, \frac{\det L_{Y\cup\{y\}}}{\det L_{Y\cup\{x\}}}\right\}. \tag{3.2}$$

with $q\big(Y \cup \{y\}, Y \cup \{x\}\big) = \frac{1}{k} \cdot \frac{1}{n-k}$ representing the proposal distribution.

Consequently, $P(Y\cup\{x\}, Y\cup\{y\}) = \begin{cases} q\big(Y \cup \{x\}, Y \cup \{y\}\big) \cdot \frac{1}{2}\min\left\{1, \frac{\det L_{Y\cup\{y\}}}{\det L_{Y\cup\{x\}}}\right\} & \text{if } y \neq x \\ 1 - \sum\limits_{z \neq x} q\big(Y \cup \{x\}, Y \cup \{z\}\big) \cdot \frac{1}{2}\min\left\{1, \frac{\det L_{Y\cup\{z\}}}{\det L_{Y\cup\{x\}}}\right\} & \text{otherwise.} \end{cases}$

Since $P(X, X) \geq \frac{1}{2}$ for all $X \subseteq [n]$, the Markov chain described above is said to be a lazy chain (Greenhill, 2013).

The lazy MCMC produced by the Metropolis-Hastings algorithm is introduced to acquire a sample from k-DPP as follows:

Algorithm 1 Markov chain for sampling from $\mathcal{P}_L^k$ (Anari et al., 2016).

```
Require: Itemset $S = [n]$, similarity matrix $L \succ 0$
  Randomly initialize state $X \subseteq S$, s.t. $|X| = k$
  Sample $r \sim \mathrm{Unif}(0, 1)$
  if $r < \frac{1}{2}$ then
      $X \leftarrow X$
  else
      while Not mixed do
          Sample $x \in X$, and $y \notin X$ u.a.r.
          Letting $Y = X \setminus \{x\}$, set
          $p \leftarrow \min\left\{1, \frac{\det L_{Y\cup\{y\}}}{\det L_{Y\cup\{x\}}}\right\}$.
          $X \leftarrow Y \cup \{y\}$ with prob. $p$
          $X \leftarrow X$ with prob. $1 - p$
      end while
  end ifreturn X
```

The essential idea of this algorithm is to obtain a rapidly-mixing Markov chain whose stationary distribution is the k-DPP $\mathcal{P}_L^k$.

4. Mixing Time

In this section, we will study the convergence of the distribution of a Markov chain $\mathcal{M}$ to the stationary distribution $\mathcal{P}_L^k$.

Definition 1. *The total variation distance between two probability distributions ν and μ on a finite space Ω is:*

$$||\mu - \nu||_{TV} = \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \nu(x)|.$$

If Y is a random variable samples according to μ and $||\mu - \nu||_{TV} \leq \epsilon$, then we say Y is an ϵ-approximate sample of ν.

Definition 2. *(Mixing Time). For a chain started at $x \in \Omega$ with transition probability matrix P and stationary distribution π. For $\epsilon > 0$, the total variation mixing time of the Markov chain is defined as follows:*

$$\tau_x(\epsilon) := \min\{t : ||P^t(x, \cdot) - \pi||_{TV} \leq \epsilon\},$$

where $P^t(x, \cdot)$ is the distribution of the chain started at x at time t.

After recalling the result of Anari et al. (2016), we will make it explicit in the case of sampling on large graphs using the kernel given before. In what follows, $P_{\mathcal{P}_L^k}$ denotes the transition probability matrix of $\mathcal{M}$.

Theorem 2. *Anari et al. (2016) For any k-DPP $\mathcal{P}_L^k : 2^{[n]} \to \mathbb{R}_+,\ X \in S$ where $S = supp\{\mathcal{P}_L^k\}$ and $|X| = k$. For $\epsilon > 0$,*

$$\tau_X(\epsilon) \leq \frac{1}{C_{\mathcal{P}_L^k}} \cdot \log\left(\frac{1}{\epsilon \cdot \mathcal{P}_L^k(X)}\right),$$

where

$$C_{\mathcal{P}_L^k} = \min_{X, X' \in S} \max\left(P_{\mathcal{P}_L^k}(X, X'), P_{\mathcal{P}_L^k}(X', X)\right)$$

and is at least $\frac{1}{2kn}$ by construction.

Our contribution is to provide a lower bound for $\mathcal{P}_L^k(X)$.

4.1. Convergence Theorem Since the acceptance probability (Eq. (3.2)) is bounded below by the ratio of the determinants of two matrices, we must bound the spectrum of L_X, which denotes the restriction of L to the entries indexed by elements of X. Since in a connected graph 0 is a single eigenvalue of $\mathcal{L}_G$, then it is also a single eigenvalue of $\mathcal{L}_G^\dagger$. This suggests that under some conditions, the eigenvalues of L_X can be bounded below by a strictly positive constant. This is achieved by using the fact that $X \subset V(G)$ as it does not make sense to take $k = n$ in most applications.

As the eigenvalues of $\mathcal{L}_G^\dagger$ are the inverse of those of $\mathcal{L}_G$, to bound the largest eigenvalue of $\underline{\mathcal{L}}_G^\dagger$, the lower bound of the smallest eigenvalue of $\underline{\mathcal{L}}_G$ is needed. In the following lemma we provide a lower bound for $\lambda_1(\underline{\mathcal{L}}_G)$.

Lemma 2. *Noting m the number of edges and considering for all $j \in 1, \dots, n$ $\mu_j = \sqrt{\sum_{i=1}^{n-1} \frac{\delta_{ij}}{d_i d_j}}$ where with $d_1, \dots, d_n$ stand for the nodes degrees and $\delta_{ij} = \begin{cases} 1 & \textit{if } i \textit{ is connected to } j \\ 0 & \textit{otherwise} \end{cases}$ we have the following bound: $\lambda_1(\underline{\mathcal{L}}_G) \geq$ $\max\{\lambda^*, 1\}$ with:*

$$\lambda^* = \max\left\{0, \min\left\{ f\left(\sqrt{\frac{1}{2} + \frac{1}{2}\sqrt{1 - \frac{4B_j^2}{4B_j^2 + C_j^2}}}\right), f\left(\sqrt{\frac{1}{2} - \frac{1}{2}\sqrt{1 - \frac{4B_j^2}{4B_j^2 + C_j^2}}}\right)\right\}\right\}$$

where $A_j = \lambda_2(\mathcal{L}_G) - \mu_j$, $B_j = 2\sqrt{\frac{d_j}{2m}(1 + \mu_j^2)}$ and $C_j = \left(-2A_j + 2\frac{d_j}{m}\right)$.

Proof. Without loss of generality we consider that $j = n$. First we remark that the smallest eigenvalue of $\underline{\mathcal{L}}_G$ which is larger or equal to the smallest eigenvalue of the $n \times n$ matrix $\begin{pmatrix} \underline{\mathcal{L}}_G & 0 \\ 0 & 1 \end{pmatrix}$. Then, it can be rewritten as:

$$\lambda_1(\underline{\mathcal{L}}_G) \geq \lambda_1 \begin{pmatrix} \underline{\mathcal{L}}_G & 0 \\ 0 & 1 \end{pmatrix} = \lambda_1 \left(\mathcal{L}_G + \begin{pmatrix} 0 & \dots & 0 & l_1 \\ \vdots & & \vdots & \vdots \\ 0 & \dots & 0 & l_{n-1} \\ l_1 & \dots & l_{n-1} & 0 \end{pmatrix} \right),$$

where for all $i \in \{1, \cdots, n-1\}$, $l_i = \frac{\delta_{in}}{\sqrt{d_i d_n}}$. The matrix $\begin{pmatrix} 0 & \dots & 0 & l_1 \\ \vdots & & \vdots & \vdots \\ 0 & \dots & 0 & l_{n-1} \\ l_1 & \dots & l_{n-1} & 0 \end{pmatrix}$ has only 0 and some eigenvalues that we will denote by μ^*. In order to find μ^*, the following system of linear equations is used:

$$\begin{cases} \mu^* x_1 & = & l_1 x_n \\ \vdots & & \vdots \\ \mu^* x_{n-1} & = & l_{n-1} x_n \\ \mu^* x_n & = & \sum_{i=1}^{n-1} l_i x_i \end{cases}$$

thereby $\mu^* x_n = \sum_{i=1}^{n-1} l_i^2 \frac{x_n}{\mu^*}$ hence $\mu^* = \pm\sqrt{\sum_{i=1}^{n-1} l_i^2}$ as obviously $x_n \neq 0$, that is $\mu^* \in \{\mu, -\mu\}$. Next, the eigenvectors associated with μ and $-\mu$ are respectively

$$u_1' = \Big(l_1, \dots, l_{n-1}, \sqrt{\sum_{i=1}^{n-1} l_i^2}\Big) \text{ and } u_2' = \Big(l_1, \dots, l_{n-1}, -\sqrt{\sum_{i=1}^{n-1} l_i^2}\Big).$$

Now we can proceed with bounding $\lambda_1(\underline{\mathcal{L}}_G)$. Let us note $V = \sum_{i=1}^n a_i V_i$, where the eigenvectors $V_1, \dots, V_n$ of $\mathcal{L}_G$ are ordered with increasing $\lambda_i(\mathcal{L}_G)$ and form an orthonormal basis. As we are interested in the eigenvalues, we consider hence without loss of generality that $a_1 \geq 0$ and $\|V\| = 1$, that is $\sum_{i=1}^n a_i^2 = 1$. Now we can minimize $a'\underline{\mathcal{L}}_G a$ to bound $\lambda_1\big(\underline{\mathcal{L}}_G\big)$. This yields,

$$\begin{aligned}
V' \begin{pmatrix} \underline{\mathcal{L}_G} & 0 \\ 0 & 1 \end{pmatrix} V \quad &= \quad \left(\sum_{i=1}^n a_i V_i\right)' \left(\mathcal{L}_G + \mu \frac{u_1 u_1'}{||u_1||^2} - \mu \frac{u_2 u_2'}{||u_2||^2}\right) \left(\sum_{i=1}^n a_i V_i\right) \\
&= \quad \left(\sum_{i=2}^n a_i V_i\right)' \left(\mathcal{L}_G + \mu \left(\frac{u_1 u_1'}{||u_1||^2} - \frac{u_2 u_2'}{||u_2||^2}\right)\right) \left(\sum_{i=2}^n a_i V_i\right) \\
&\quad + \frac{\mu}{||u_1||^2} a_1^2 \big((V_1' u_1)^2 - (V_1' u_2)^2\big) + 2(a_1 V_1') \Bigg[\mu \frac{u_1 u_1'}{||u_1||^2} \\
&\quad - \mu \frac{u_2 u_2'}{||u_2||^2}\Bigg] \left(\sum_{i=2}^n a_i V_i\right) \text{ as } \lambda_1(\mathcal{L}_G) = 0 \\
&\geq \quad (\lambda_2(\mathcal{L}_G) - \mu) \left\|\sum_{i=2}^n a_i V_i\right\|^2 + \frac{1}{2\mu} a_1^2 \left[\left(\sum_{i=1}^n V_{1i} u_{1i}\right)^2 - \left(\sum_{i=1}^n V_{1i} u_{2i}\right)^2\right] \\
&\quad + \quad 2 \frac{a_1 V_1'}{2\mu} [u_1 u_1' - u_2 u_2'] \left(\sum_{i=2}^n a_i V_i\right).
\end{aligned}$$

Recalling that the matrix $u_1 u_1' - u_2 u_2'$ is as follows:

$$u_1 u_1' - u_2 u_2' = 2\mu \begin{pmatrix} 0 & \dots & 0 & l_1 \\ \vdots & & \vdots & \vdots \\ 0 & \dots & 0 & l_{n-1} \\ l_1 & \dots & l_{n-1} & 0 \end{pmatrix}.$$

We set $w' := V_1'(u_1u_1' - u_2u_2') = 2\mu\left(V_{1n}l_1, \cdots, V_{1n}l_{n-1}, \sum_{i=1}^{n-1} l_iV_{1i}\right)$ and by using the fact that $\mathcal{L}_G$ has an orthonormal basis of eigenvectors, we get

$$\begin{aligned}
a'\underline{\mathcal{L}}_G a &\geq (\lambda_2(\mathcal{L}_G) - \mu)\sum_{i=2}^n a_i^2 + \frac{a_1^2}{2\mu}\left[\left(\sum_{i=1}^n V_{1i}u_{1i} - \sum_{i=1}^n V_{1i}u_{2i}\right)\left(\sum_{i=1}^n V_{1i}u_{1i} + \sum_{i=1}^n V_{1i}u_{2i}\right)\right] \\
&\quad + \frac{a_1}{\mu}2\mu\left(V_{1n}l_1, \cdots, V_{1n}l_{n-1}, \sum_{i=1}^{n-1} l_iV_{1i}\right)\left(\sum_{i=2}^n a_iV_i\right) \\
&\geq \big(\lambda_2(\mathcal{L}_G) - \mu\big)\big(1 - a_1^2\big) + \frac{a_1^2}{2\mu}\left[\left(\sum_{i=1}^n V_{1i}(u_{1i} - u_{2i})\right)\left(\sum_{i=1}^n V_{1i}(u_{1i} + u_{2i})\right)\right] \\
&\quad + \frac{a_1}{\mu}\left[w'\left(\sum_{i=2}^n a_iV_i\right)\right] \\
&\geq \big(\lambda_2(\mathcal{L}_G) - \mu\big)\big(1 - a_1^2\big) + \frac{a_1^2}{2\mu}\left[\left(V_{1n}\left(2\sqrt{\sum_{i=1}^{n-1} l_i^2}\right)\right)\left(\sum_{i=1}^{n-1} V_{1i}(2l_i)\right)\right] \\
&\quad + \frac{a_1}{\mu}w'\left(\sum_{i=2}^n a_iV_i\right) \\
&\geq \big(\lambda_2(\mathcal{L}_G) - \mu\big)\big(1 - a_1^2\big) + 2a_1^2V_{1n}\left(\sum_{i=1}^{n-1} V_{1i}l_i\right) - \frac{a_1}{\mu}\|w\|\left\|\sum_{i=2}^n a_iV_i\right\| \text{ by Cauchy-Schwarz.}
\end{aligned}$$

Recalling that V_1 is the eigenvector of $\mathcal{L}_G$ corresponding to eigenvalue 0 as mentioned in Section 2.2.1 we get $V_1 = \frac{1}{\sqrt{2m}}\begin{pmatrix}\sqrt{d_1}\\ \vdots \\ \sqrt{d_n}\end{pmatrix}$, with $\sum_{i=1}^n d_i = 2m$.

Consequently,

$$\begin{aligned}
a'\underline{\mathcal{L}}_G a &\geq \big(\lambda_2(\mathcal{L}_G) - \mu\big)\big(1 - a_1^2\big) + 2a_1^2\frac{\sqrt{d_n}}{2m}\left(\sum_{i=1}^{n-1}\sqrt{d_i}\frac{\delta_{in}}{\sqrt{d_id_n}}\right) \\
&\quad - \frac{a_1}{\mu}\|w\|\sqrt{1 - a_1^2} \text{ by Cauchy-Schwarz} \\
&\geq \big(\lambda_2(\mathcal{L}_G) - \mu\big)\big(1 - a_1^2\big) + \frac{a_1^2}{m}\sum_{i=1}^{n-1}\delta_{in} - \frac{a_1}{\mu}\|w\|\sqrt{1 - a_1^2} \\
&\geq \big(\lambda_2(\mathcal{L}_G) - \mu\big)\big(1 - a_1^2\big) + \frac{d_n}{m}a_1^2 - \frac{a_1}{\mu}\|w\|\sqrt{1 - a_1^2}.
\end{aligned}$$

Now, we need an upper bound on $\|w\|$:

$$\|w\|^2 \leq \left\|2\mu\left(V_{1n}l_1, \cdots, V_{1n}l_{n-1}, \sum_{i=1}^{n-1} l_iV_{1i}\right)\right\|^2$$

$$\begin{aligned}
&\leq \quad 4\mu^2\left[(V_{1n}l_1)^2+\cdots+(V_{1n}l_{n-1})^2+\left(\sum_{i=1}^{n-1}l_iV_{1i}\right)^2\right]\\
&\leq \quad 4\mu^2\left[\frac{d_n}{2m}\left(\sum_{i=1}^{n-1}\frac{\delta_{in}}{d_id_n}\right)+\left(\sum_{i=1}^{n-1}\frac{\delta_{in}}{\sqrt{d_id_n}}\frac{\sqrt{d_i}}{\sqrt{2m}}\right)^2\right]\\
&\leq \quad 4\mu^2\left[\frac{d_n}{2m}\mu^2+\left(\frac{1}{\sqrt{2m}\sqrt{d_n}}\sum_{i=1}^{n-1}\delta_{in}\right)^2\right]\\
&\leq \quad \frac{4\mu^2 d_n}{2m}\left(1+\mu^2\right),
\end{aligned}$$

This implies that

$$a'\underline{\mathcal{L}}_G a \geq \big(\lambda_2(\mathcal{L}_G)-\mu\big)\big(1-a_1^2\big)+\frac{d_n}{m}a_1^2-2a_1\sqrt{\frac{d_n}{2m}(1+\mu^2)}\sqrt{1-a_1^2}. \quad (4.1)$$

To shorten notation, we introduce A, B, C and function f as follows: $A=\lambda_2(\mathcal{L}_G)-\mu$, $B=2\sqrt{\frac{d_n}{2m}(1+\mu^2)}$, $C=\left(-2A+2\frac{d_n}{m}\right)$ and $f(a_1)=A\big(1-a_1^2\big)+\frac{d_n}{m}a_1^2-Ba_1\sqrt{1-a_1^2}$. Then, the derivative of the function f is given by

$$f'(a_1)=\Big(-2A+2\frac{d_n}{m}\Big)a_1-B\left(\sqrt{1-a_1^2}-\frac{a_1^2}{\sqrt{1-a_1^2}}\right)=Ca_1-B\left(\frac{1-2a_1^2}{\sqrt{1-a_1^2}}\right).$$

Therefore, $f'(a_1)=0$ implies:

$$C^2a_1^2=B^2\frac{(1-2a_1^2)^2}{1-a_1^2}\Longrightarrow C^2a_1^2-C^2a_1^4=B^2+4B^2a_1^4-4B^2a_1^2. \quad (4.2)$$

Re-noting $x=a_1^2$, Eq. (4.2) can be written as a second order equation: $(4B^2+C^2)x^2-(4B^2+C^2)x+B^2=0$. Then,

$$x=\frac{(4B^2+C^2)\pm(4B^2+C^2)\sqrt{1-\frac{4B^2}{4B^2+C^2}}}{2(4B^2+C^2)}=\frac{1}{2}\pm\frac{1}{2}\sqrt{1-\frac{4B^2}{4B^2+C^2}}\in[0;1]. \quad (4.3)$$

Moreover, as $f'(0)=-B<0$ and $f'(x)\underset{x\to 1}{\longrightarrow}+\infty$, the minimum of the above function is reached for either:

$$\sqrt{\frac{1}{2}+\frac{1}{2}\sqrt{1-\frac{4B^2}{4B^2+C^2}}}\text{ or }\sqrt{\frac{1}{2}-\frac{1}{2}\sqrt{1-\frac{4B^2}{4B^2+C^2}}}.$$

Therefore, according to Eq. (4.1), we get

$$a'\underline{\mathcal{L}}_G a \geq \min\left\{ f\left(\sqrt{\frac{1}{2}+\frac{1}{2}\sqrt{1-\frac{4B^2}{4B^2+C^2}}}\right), f\left(\sqrt{\frac{1}{2}-\frac{1}{2}\sqrt{1-\frac{4B^2}{4B^2+C^2}}}\right)\right\}.$$

This result has been obtained without loss of generality for vertex, so that we can conclude for any vertex j.

Using previous result, we can make the polynomial bounds on the mixing time more precise in the following theorem.

Theorem 3. *Let $\mathcal{P}_L^k : 2^{[n]} \to \mathbb{R}^+$ where L is the matrix defined by Eq. (3.1). For any $\epsilon > 0$ the lazy Markov chain defined in Algorithm 1 generates an ϵ-approximate sample of $\mathcal{P}_L^k$ in time*

$$\tau_\epsilon \leq 2k^2 n \cdot \log\left[\frac{2en}{\lambda^* k \epsilon^{\frac{1}{k}}}\right]$$

where λ^* is the bound provided by lemma 2.

Proof. The proof is based on the main theorem of Anari et al. (2016) by considering L the matrix defined by Eq. (3.1) and noting that if $k \leq n-1$, then the determinant defining $\mathcal{P}_L^k$ can be seen as a principal submatrix of a matrix of size $n-1$ enabling us to use the bound of lemma 2. By applying Lemma 1, the eigenvalues of L_X are as follows:

$$0 < \lambda_i(\underline{\mathcal{L}}_G^\dagger) \leq \lambda_i(L_X) \leq \lambda_{n-1-k+i}(\underline{\mathcal{L}}_G^\dagger). \tag{4.4}$$

By using Eq. (4.4), lower and upper bounds of the determinant of L_X can be obtained as:

$$\det\left(L_X\right) \quad = \quad \det\left(\mathcal{L}_X^\dagger\right) \geq \lambda_1(\underline{\mathcal{L}}_G^\dagger) \times \cdots \times \lambda_k(\underline{\mathcal{L}}_G^\dagger) \geq \left(\lambda_1(\underline{\mathcal{L}}_G^\dagger)\right)^k, \tag{4.5}$$

$$\det\left(L_X\right) \quad = \quad \det\left(\mathcal{L}_X^\dagger\right) \leq \lambda_{n-1-k+1}(\underline{\mathcal{L}}_G^\dagger) \times \cdots \times \lambda_{n-1-k+k}(\underline{\mathcal{L}}_G^\dagger) \leq \left(\lambda_{n-1}(\underline{\mathcal{L}}_G^\dagger)\right)^k \cdot \tag{4.6}$$

Recalling that the eigenvalues of $\mathcal{L}_G$ satisfy:

$$0 = \lambda_1(\mathcal{L}_G) \leq \cdots \leq \lambda_n(\mathcal{L}_G) \leq 2,$$

then by deleting its i^{th} row and i^{th} column and by applying Lemma 1, the eigenvalues of $\underline{\mathcal{L}}_G$ stand as follows:

$$\lambda^* \leq \lambda_1(\underline{\mathcal{L}}_G) \leq \lambda_2(\mathcal{L}_G) \leq \cdots \leq \lambda_{n-1}(\underline{\mathcal{L}}_G) \leq \lambda_n(\mathcal{L}_G) \leq 2. \tag{4.7}$$

Thus, Eq. (4.7) yields:

$$\frac{1}{2} \leq \frac{1}{\lambda_n(\underline{\mathcal{L}}_G)} \leq \frac{1}{\lambda_{n-1}(\underline{\mathcal{L}}_G)} = \lambda_1(\underline{\mathcal{L}}_G^{\dagger}) \leq \cdots \leq \frac{1}{\lambda_2(\underline{\mathcal{L}}_G)} \leq \frac{1}{\lambda_1(\underline{\mathcal{L}}_G)} = \lambda_{n-1}(\underline{\mathcal{L}}_G^{\dagger}).$$

Hence, for any set X of cardinality k and by using Eq. (4.5), we have

$$\det(L_X) \geq \left(\lambda_1(\underline{\mathcal{L}}_G^{\dagger})\right)^k \geq \left(\frac{1}{2}\right)^k.$$

Indeed, by using Eq. (4.6) we can bound the determinant of L_X as follows:

$$\det L_X \quad \leq \left(\lambda_{n-1}(\underline{\mathcal{L}}_G^{\dagger})\right)^k = \left(\frac{1}{\lambda_1(\underline{\mathcal{L}}_G)}\right)^k \leq \left(\frac{1}{\lambda^*}\right)^k$$

Consequently,

$$\mathcal{P}_L^k(X) = \frac{\det(L_X)}{\sum_{|X'|=k} \det(L_{X'})} \geq \frac{\left(\frac{1}{2}\right)^k}{\begin{pmatrix} n \\ k \end{pmatrix} \left(\frac{1}{\lambda^*}\right)^k} \geq \frac{(\lambda^*)^k k!}{n^k 2^k} \geq \left(\frac{\lambda^* k}{2en}\right)^k \text{ using Stirling inequality.}$$

Now, by using Theorem 2 we can directly upper bound the mixing time in total variation distance as follows:

$$\tau_X(\epsilon) \leq \frac{1}{C_{\mathcal{P}_L^k}} \cdot \log\left(\frac{1}{\epsilon \cdot \mathcal{P}_L^k}\right) \leq 2kn \cdot \log\left[\left(\frac{2en}{\lambda^* k}\right)^k \frac{1}{\epsilon}\right] \leq 2k^2 n \cdot \log\left[\frac{2en}{\lambda^* k \epsilon^{\frac{1}{k}}}\right].$$

This proves the result.

5 Discussion and Conclusion

The bound provided by Lemma 2, as mentionned in Theorem 3 is only useful if $\lambda^* > 0$. This is unfortunately not always the case but we can report here below in Table 1 some real graphs where this happens. These three datasets, hens, eies_relations and bktecc all stem from the R networkdata package.

Table 1: Application to three small graphs

	Hens	eies_relations	bktecc
Number of vertices	32	32	34
Number of edges	496	435	561
λ^*	0.023	0.0.07	0.017

Table 2: Application to Erdös-Renyi graphs with varying size and parameter p which is the probability for each edge to exist

number of vertices	1000	1000	1000	1000	10000	10000	10000	10000
p	0.01	0.05	0.1	0.2	0.01	0.05	0.1	0.2
λ^*	$-4.7 * 10^{-2}$	$1.5 * 10^{-4}$	$4.2 * 10^{-4}$	$6.1 * 10^{-4}$	$3.7 * 10^{-5}$	$7.1 * 10^{-5}$	$8.0 * 10^{-5}$	$8.6 * 10^{-5}$

We have tested for dozens of other graphs where the lower bound appeared useless being negative, some explanation for it was that many times the smallest vertex degree was quite small as well as the graph size. That is why we have also applied the bound to Erdös-Renyi graphs where the bound appears much more useful as can be seen in Table 2. Then it also seems that the bigger the graph, the easier it is to have a positive bound.

To conclude, this paper makes the bound of Anari et al. (2016) more precise in the particular case where we sample k-DPP on graphs using as kernel the Moore-Penrose pseudo-inverse of the normalized Laplacian. The proof involving some linear algebra observations is simply an application of the main theorem of Anari et al. (2016) in a practical case. In the future we would like to provide lower bounds of a more general use for graph sampling looking at for example what happens when more than one vertex is removed from the whole graph.

References

ANARI, N., GHARAN, S.O. and REZAEI, A. (2016). Monte Carlo Markov chain algorithms for sampling strongly Rayleigh distributions and determinantal point processes. In: COLT.

BORCEA, J., BRANDEN, P. and LIGGETT, T.M. (2009). Negative dependence and the geometry of polynomials. *J. Am. Math. Soc.* **22**, 521–567.

CVETKOVIĆ, D., ROWLINSON, P. and SIMIĆ, S. (2010). An Introduction to the Theory of Graph Spectra. Cambridge University Press.

CHEEGER, J. (1970). A lower bound for the smallest eigenvalue of the Laplacian. Problems in Analysis, (R. C. Gunning, ed.) Princeton University Press, pp. 195–199.

CHIB, S. and GREENBERG, E. (1995). Understanding the Metropolis-Hastings algorithm. *Am. Stat.* **49**, 327–335.

CHUNG, F.R.K. (1997). *Spectral Graph Theory CBMS Regional Conference Series in Mathematics, 92.* American Mathematical Society, Providence.

DESHPANDE, A. and RADEMACHER, L. (2010). Efficient volume sampling for row/column subset selection. FOCS Number 1-3-4 329–338.

GREENHILL, C. (2013). Making Markov chains less lazy arXiv:1203.6668.

GOBEL, F. and JAGERS, A. (1974). Random walks on graphs. *Stoch. Process. Appl.* **2**, 311–336.

HORN, R.A. and JOHNSON, C.R. (1985). Matrix analysis. Cambridge University Press.

KANNAN, R. and VEMPALA, S. (2009). Spectral algorithms. *Found. Trends Theor. Comput. Sci.* **4**, 157–288.

KANG, B. (2013). Fast determinantal point process sampling with application to clustering. NIPS, pp. 2319–2327.

KULESZA, A. and TASKAR, B. (2012). Determinantal point processes for machine learning. *Found. Trends Mach. Learn.* **5**, 123–286.

KULESZA, A. and TASKAR, B. (2011). kDPPs: fixed size determinantal point processes. In: Proceedings of the 28th International Conference on Machine Learning. pp 1193–1200.

LEVIN, A.D., PERES, Y. and WILMER, E.L. (2006). Markov chains and mixing times american. Mathematical Society.

LI, H.H., LI, J.S. and FAN, Y.-Z. (2008). The effect on the second smallest eigenvalue of the normalized Laplacian of a graph by grafting edges, vol. 56.

LI, H.H. and LI, J.S. (2011). A note on the normalized Laplacian spectra, vol. 15.

LI, J., GUO, J.-M., SHIU, W.C. and CHANG, A. (2014). An edge-separating theorem on the second smallest normalized Laplacian eigenvalue of a graph and its applications, vol. 171.

LOVÁSZ, L. (1993). Random walks on graphs: A survey. *Combinatorics, Paul erdös is Eighty* **2**, 1–46.

MACCHI, O. (1975). The Coincidence approach to stochastic point processes. *Adv. Appl. Probab.* **7**, 83–122.

Publisher's Note. Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Diala Wehbe
Nicolas Wicker
Paul Painlevé Laboratory, University of Lille, 59650, Villeneuve D'Ascq, France
E-mail: diala.wehbe@net.usj.edu.lb

Diala Wehbe
EDST, Lebanese University, Tripoli, Lebanon

Paper received: 30 September 2019; accepted 5 March 2021.