Graph sampling with determinantal processes

Nicolas Tremblay*, Pierre-Olivier Amblard* and Simon Barthelmé*

*CNRS, GIPSA-lab, Grenoble-Alpes University, Grenoble, France

Abstract—We present a new random sampling strategy for k-bandlimited signals defined on graphs, based on determinantal point processes (DPP). For small graphs, *i.e.*, in cases where the spectrum of the graph is accessible, we exhibit a DPP sampling scheme that enables perfect recovery of bandlimited signals. For large graphs, *i.e.*, in cases where the graph's spectrum is not accessible, we investigate, both theoretically and empirically, a sub-optimal but much faster DPP based on loop-erased random walks on the graph. Preliminary experiments show promising results especially in cases where the number of measurements should stay as small as possible and for graphs that have a strong community structure. Our sampling scheme is efficient and can be applied to graphs with up to 10^6 nodes.

I. INTRODUCTION

Graphs are a central modelling tool for network-structured data. Data on a graph, called graph signals [1], such as individual hobbies in social networks, or blood flow of brain regions in neuronal networks, may be represented by a scalar per node. Studying such signals with respect to the particular graph topology on which it is defined, is the goal of graph signal processing (GSP). One of the top challenges of GSP is not only to adapt classical processing tools to arbitrary topologies defined by the underlying graph, but to do so with efficient algorithms that can handle the large datasets encountered today. A possible answer to this efficiency challenge is dimension reduction, and in particular sampling. Graph sampling [2]–[6] consists in measuring a graph signal on a reduced set of nodes carefully chosen to optimize a predefined objective, may it be the estimation of some statistics, signal compression via graph filterbanks, signal recovery, etc.

In an effort to generalize Shannon's sampling theorem of bandlimited signals, several authors [3]–[5] have proposed sampling schemes adapted to the recovery of *a priori* smooth graph signals, and in particular k-bandlimited signals, *i.e.*, linear combinations of only k graph Fourier modes. Classically, there are three main types of sampling schemes that one may try to adapt to graphs: periodic, irregular, and random.

Periodic graph sampling is ill-defined unless the graph is not bipartite (or multipartite) and has only been investigated in the design of graph filterbanks where the objective is to sample "one every two nodes" [7], [8].

Irregular graph sampling of k-bandlimited graph signals has been studied by Anis *et al.* [3] and Chen *et al.* [4] who, building upon Pensenson's work [2], look for optimal sampling sets of k-bandlimited graphs, in the sense that they are tight (*i.e.*, they contain only k nodes) and that they maximize noise-robustness of the reconstruction. To exhibit such an optimal set, one needs i) to compute the first k eigenvectors of the Laplacian which becomes prohibitive at large scale, and ii)

to optimize over all subsets of k nodes, which is a large combinatorial problem. Using greedy algorithms and spectral proxies [3] to sidestep some of the computational bottlenecks, these methods enable to efficiently perform optimal sampling on graphs of size up to a few thousands nodes.

Loosening the tightness constraint on sampling sets, random graph sampling [5] shows that if $O(k \log k)$ nodes are drawn with replacement according to a particular probability distribution p^* that depends on the first k eigenvectors of the Laplacian, then recovery is guaranteed with high probability. p^* was shown to be closely related to *leverage scores*, an important concept in randomized numerical linear algebra [9]. Moreover, an efficient estimation of p^*, bypassing expensive spectral computations, enables to sample very large graphs; as was shown in [10], where experiments were performed on graphs with $N \simeq 10^6$ and $k \simeq 200$.

In this paper, we detail another form of random graph sampling, motivated by a simple experimental context. Consider a graph composed of two almost disconnected communities of equal size. Its first Fourier mode is constant (as always), and its second one is positive in one community, and negative in the other, while its absolute value is more or less constant on the whole graph. Therefore, a 2-bandlimited graph signal will approximately be equal to a constant in one community, and to another constant in the other: any sampling set composed of one node drawn from each community will enable to recover the whole signal. Unfortunately, random graph sampling as proposed in [5] samples independently with replacement: it has a fair chance of sampling only in one community, thereby missing half of the information. To avoid such situations, we propose to take advantage of determinantal point processes (DPP), random processes that introduce negative correlations to force diversity within the sampling set.

Contributions. If the first k eigenvectors of the Laplacian are computable, we exhibit a DPP that always samples sets of size k perfectly embedding k-bandlimited graph signals. Moreover, in the case the graph is too large to compute its first k eigenvectors, and building upon recent results linking determinantal processes and random rooted spanning forests [11], we show to what extent a simple algorithm based on loop-erased random walks is an acceptable approximation.

II. BACKGROUND

Consider an undirected weighted graph $\mathcal{G}$ of N nodes represented by its adjacency matrix W, where $W_{ij} = W_{ji} \geqslant 0$ is the weight of the connection between nodes i and j. Denote by $L = D - W$ its Laplacian matrix, where D is the diagonal matrix with entries $D_{ii} = \sum_j W_{ij}$. L is real and symmetrical,

therefore diagonalizable as $\mathsf{U}\Lambda\mathsf{U}^\mathsf{T}$ with $\mathsf{U} = (\boldsymbol{u}_1|\ldots|\boldsymbol{u}_N) \in \mathbb{R}^{N\times N}$ the graph Fourier basis and $\Lambda = \mathrm{diag}(\lambda_1,\ldots,\lambda_N)$ the diagonal matrix of the associated frequencies, that we sort: $0 = \lambda_1 \leqslant \lambda_2 \leqslant \ldots \leqslant \lambda_N$. For any function $h : \mathbb{R}^+ \to \mathbb{R}$, we write $h(\Lambda) = \mathrm{diag}(h(\lambda_1),\ldots,h(\lambda_N))$. The graph filter associated to the frequency response h reads $\mathsf{U}h(\Lambda)\mathsf{U}^\mathsf{T}$.

Given the frequency interpretation of the graph Fourier modes, one may define "smooth" signals as linear combinations of the first few low-frequency Fourier modes. Writing $\mathsf{U}_k = (\boldsymbol{u}_1|\ldots|\boldsymbol{u}_k) \in \mathbb{R}^{N\times k}$, we have the formal definition:

Definition II.1 (k-bandlimited signal on $\mathcal{G}$). *A signal $\boldsymbol{x} \in \mathbb{R}^N$ defined on the nodes of the graph $\mathcal{G}$ is k-bandlimited with $k \in \mathbb{N}^*$ if $\boldsymbol{x} \in \mathrm{span}(\mathsf{U}_k)$, i.e., $\exists\ \boldsymbol{\alpha} \in \mathbb{R}^k$ such that $\boldsymbol{x} = \mathsf{U}_k\boldsymbol{\alpha}$.*

Sampling consists in selecting a subset $\mathcal{A} = (\omega_1,\ldots,\omega_m)$ of m nodes of the graph and measuring the signal on it. To each possible sampling set, we associate a measurement matrix $\mathsf{M} = (\boldsymbol{\delta}_{\omega_1}|\boldsymbol{\delta}_{\omega_2}|\ldots|\boldsymbol{\delta}_{\omega_m})^\mathsf{T} \in \mathbb{R}^{m\times N}$ where $\boldsymbol{\delta}_{\omega_i}(j) = 1$ if $j = \omega_i$, and 0 otherwise. Given a k-bandlimited signal $\boldsymbol{x}$, its (possibly noisy) measurement on $\mathcal{A}$ reads:

$$\boldsymbol{y} = \mathsf{M}\boldsymbol{x} + \boldsymbol{n} \in \mathbb{R}^m, \tag{1}$$

where $\boldsymbol{n}$ models any kind of measurement noise and/or the fact that $\boldsymbol{x}$ may not exactly be in $\mathrm{span}(\mathsf{U}_k)$. In this latter case, $\boldsymbol{x} = \mathsf{U}_k\mathsf{U}_k^\mathsf{T}\boldsymbol{x} + \boldsymbol{e}$, where $\mathsf{U}_k\mathsf{U}_k^\mathsf{T}\boldsymbol{x}$ is the closest signal to $\boldsymbol{x}$ that is in $\mathrm{span}(\mathsf{U}_k)$; and $\mathsf{M}\boldsymbol{x} = \mathsf{M}\mathsf{U}_k\mathsf{U}_k^\mathsf{T}\boldsymbol{x} + \boldsymbol{n}$ with $\boldsymbol{n} = \mathsf{M}\boldsymbol{e}$.

A. If U_k is known

If U_k is known, the measurement on $\mathcal{A}$ reads $\boldsymbol{y} = \mathsf{M}\mathsf{U}_k\boldsymbol{\alpha} + \boldsymbol{n}$ and one recovers the signal by solving:

$$\boldsymbol{x}_{\mathrm{rec}} = \operatorname*{argmin}_{\boldsymbol{z}\in\mathrm{span}(\mathsf{U}_k)} \|\mathsf{M}\boldsymbol{z} - \boldsymbol{y}\|^2 = \mathsf{U}_k(\mathsf{M}\mathsf{U}_k)^\dagger\boldsymbol{y}, \tag{2}$$

where $(\mathsf{M}\mathsf{U}_k)^\dagger \in \mathbb{R}^{k\times m}$ is the Moore-Penrose pseudo-inverse of $\mathsf{M}\mathsf{U}_k \in \mathbb{R}^{m\times k}$ and $\|.\|$ the Euclidean norm. Of course, perfect reconstruction is impossible if $m < k$, as the system would be underdetermined. Hereafter, we thus suppose $m \geqslant k$. Denote by $\sigma_1 \leqslant \ldots \leqslant \sigma_k$ the singular values of $\mathsf{M}\mathsf{U}_k$.

Proposition II.1 (Classical result). *If $\sigma_1 > 0$, then perfect reconstruction (up to the noise level) is possible.*

Proof. If $\sigma_1 > 0$, then $\mathsf{U}_k^\mathsf{T}\mathsf{M}^\mathsf{T}\mathsf{M}\mathsf{U}_k$ is invertible; and since $\boldsymbol{x} = \mathsf{U}_k\boldsymbol{\alpha}$, we have:

$$\boldsymbol{x}_{\mathrm{rec}} = \mathsf{U}_k(\mathsf{M}\mathsf{U}_k)^\dagger\boldsymbol{y} = \mathsf{U}_k(\mathsf{U}_k^\mathsf{T}\mathsf{M}^\mathsf{T}\mathsf{M}\mathsf{U}_k)^{-1}\mathsf{U}_k^\mathsf{T}\mathsf{M}^\mathsf{T}(\mathsf{M}\boldsymbol{x} + \boldsymbol{n})$$
$$= \boldsymbol{x} + \mathsf{U}_k(\mathsf{U}_k^\mathsf{T}\mathsf{M}^\mathsf{T}\mathsf{M}\mathsf{U}_k)^{-1}\mathsf{U}_k^\mathsf{T}\mathsf{M}^\mathsf{T}\boldsymbol{n}. \qquad \square$$

In the following, in the cases where U_k is known, we fix $m = k$. Chen et al. [4] showed that a sample of size k always exists s.t. $\sigma_1 > 0$. In fact, in general, many possible such subsets exist. Finding the optimal one is a matter of definition. Authors in [3], [4] propose two optimality definitions via noise robustness. A first option is to minimize the worst-case error, which translates to finding the subset that maximizes σ_1^2:

$$\mathcal{A}^{\mathrm{WCE}} = \operatorname*{arg\,max}_{\mathcal{A}\ \mathrm{s.t.}\ |\mathcal{A}|=k} \sigma_1^2. \tag{3}$$

Algorithm 1 Greedy algorithm to approximate (3), (4), or (5)

Input: U_k, `objective` = WCE, MSE or MV.
$\mathcal{Y} \leftarrow \emptyset$
while $|\mathcal{Y}| \neq k$ **do**:
    for $n = 1, 2, \ldots, N$ **do**:
        · $\mathcal{Y}_n \leftarrow \mathcal{Y} \cup \{n\}$
        · Compute $\{\sigma_i\}_{i=1,\ldots,|\mathcal{Y}_n|}$ the singular values of the restriction of U_k to the lines indexed by $\mathcal{Y}_n$.
        · Store the gain on the `objective` of adding n
    end for
    Add to $\mathcal{Y}$ the node for which the gain is maximal.
end while
Output: $\mathcal{Y}$.

A second option is to find the subset that minimizes the mean square error (assuming $\mathbb{E}(\boldsymbol{n}\boldsymbol{n}^\mathsf{T})$ is the identity) :

$$\mathcal{A}^{\mathrm{MSE}} = \operatorname*{arg\,min}_{\mathcal{A}\ \mathrm{s.t.}\ |\mathcal{A}|=k} \mathrm{tr}\left[(\mathsf{U}_k^\mathsf{T}\mathsf{M}^\mathsf{T}\mathsf{M}\mathsf{U}_k)^{-1}\right] = \operatorname*{arg\,min}_{\mathcal{A}\ \mathrm{s.t.}\ |\mathcal{A}|=k} \sum_{i=1}^{k} \frac{1}{\sigma_i^2}, \tag{4}$$

where tr is the trace operator.

We argue that another possible definition of optimal subset is such that the determinant of $\mathsf{U}_k^\mathsf{T}\mathsf{M}^\mathsf{T}\mathsf{M}\mathsf{U}_k$ is maximal, *i.e.*:

$$\mathcal{A}^{\mathrm{MV}} = \operatorname*{arg\,max}_{\mathcal{A}\ \mathrm{s.t.}\ |\mathcal{A}|=k} \prod_{i=1}^{k} \sigma_i^2, \tag{5}$$

where MV stands for Maximum Volume, as the determinant is proportional to the volume spanned by the k sampled lines of U_k. In other words, this definition of optimality maximizes the determinant of the inverse covariance of the error (here supposed to be the identity). It is related to the previous definitions : in particular, increasing σ_1 necessarily increases the determinant. Also, such volume sampling as it is sometimes called has been largely studied and successfully applied in diverse contexts [12].

In all three cases, finding the optimal subset is a very large combinatorial problem (it implies a search over all combinations of k nodes): in practice, greedy algorithms such as Alg. 1 are used to find approximate solutions. In the following, we refer to $\hat{\mathcal{A}}^{\mathrm{WCE}}$, $\hat{\mathcal{A}}^{\mathrm{MSE}}$ and $\hat{\mathcal{A}}^{\mathrm{MV}}$ the (deterministic) solutions obtained by greedy optimization of the above 3 objectives. We refer to $\hat{\mathcal{A}}^{\mathrm{maxvol}}$ the solution to the `maxvol` algorithm proposed in [13], which is another approximation of $\mathcal{A}^{\mathrm{MV}}$. Note that greedy approaches have been proved near-optimal in this context [14], [15].

B. If U_k is unknown

If U_k is unknown, one may use a proxy that penalizes high frequencies in the solution, and thus recover an approximation of the original signal by solving the regularized problem:

$$\boldsymbol{x}_{\mathrm{rec}} = \operatorname*{argmin}_{\boldsymbol{z}\in\mathbb{R}^N} \|\mathsf{M}\boldsymbol{z} - \boldsymbol{y}\|^2 + \gamma\boldsymbol{z}^\mathsf{T}\mathsf{L}^r\boldsymbol{z}, \tag{6}$$

where $\gamma > 0$ is the regularization parameter and r the power of the Laplacian that controls the strength of the high frequency penalization. By differentiating with respect to $\boldsymbol{z}$, one has:

$$(\mathsf{M}^\mathsf{T}\mathsf{M} + \gamma \mathsf{L}^r)\boldsymbol{x}_{\mathrm{rec}} = \mathsf{M}^\mathsf{T}\boldsymbol{y}, \tag{7}$$

which may be solved by direct inversion if N is not too large, or –thanks to the objective's convexity– by iterative methods such as gradient descent.

C. Reweighting in the case of random sampling

In random sampling, the sample $\mathcal{A} = (\omega_1, \ldots, \omega_m)$ is a random variable, and therefore so is the measurement $\boldsymbol{y} = \mathsf{M}\boldsymbol{x}$. Consider two signals $\boldsymbol{x}_1$ and $\boldsymbol{x}_2 \neq \boldsymbol{x}_1$, and their measurements $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$. If $\boldsymbol{y}_1 = \boldsymbol{y}_2$, perfect reconstruction is impossible. To prevent this, one possible solution is to reweight the measurement such that the expected norm of the reweighted measurement equals the signal's norm. For instance, in [5], m nodes are drawn independently with replacement. At each draw, the probability to sample node i is $p_i^* = \|\mathsf{U}_k^\mathsf{T}\boldsymbol{\delta}_i\|^2 / k$. One has $\mathbb{E}\left(\|\mathsf{P}^{-1/2}\mathsf{M}\boldsymbol{x}\|^2\right) = \|\boldsymbol{x}\|^2$ if

$$\mathsf{P} = \mathrm{diag}(mp_{\omega_1}^* | \ldots | mp_{\omega_m}^*). \tag{8}$$

This guarantees that if $\|\boldsymbol{x}_1 - \boldsymbol{x}_2\| > 0$, then, for a large enough m, the reweighted measures will necessarily be distinct in the measurement space, i.e., $\|\mathsf{P}^{-1/2}(\boldsymbol{y}_1 - \boldsymbol{y}_2)\| > 0$, ensuring that $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ have a chance of being recovered.

Recovery with known U_k (Eq. (2)) thus transforms into

$$\boldsymbol{x}_{\mathrm{rec}} = \underset{\boldsymbol{z}\in\mathrm{span}(\mathsf{U}_k)}{\mathrm{argmin}} \left\|\mathsf{P}^{-1/2}\left(\mathsf{M}\boldsymbol{z} - \boldsymbol{y}\right)\right\|^2; \tag{9}$$

and recovery with unknown U_k (Eq. (6)) transforms into

$$\boldsymbol{x}_{\mathrm{rec}} = \underset{\boldsymbol{z}\in\mathbb{R}^N}{\mathrm{argmin}} \left\|\mathsf{P}^{-1/2}\left(\mathsf{M}\boldsymbol{z} - \boldsymbol{y}\right)\right\|^2 + \gamma\boldsymbol{z}^\mathsf{T}\mathsf{L}^r\boldsymbol{z}. \tag{10}$$

III. DETERMINANTAL PROCESSES

Denote by $[N]$ the set of all subsets of $\{1, 2, \ldots, N\}$.

Definition III.1 (Determinantal Point Process [12])**.** *Consider a point process, i.e., a process that randomly draws an element* $\mathcal{A} \in [N]$. *It is determinantal if, for every* $\mathcal{S} \subseteq \mathcal{A}$,

$$\mathbb{P}(\mathcal{S} \subseteq \mathcal{A}) = det(\mathsf{K}_\mathcal{S}),$$

where $\mathsf{K} \in \mathbb{R}^{N\times N}$, *a semi-definite positive matrix s.t.* $0 \preceq \mathsf{K} \preceq 1$, *is called the marginal kernel; and* $\mathsf{K}_\mathcal{S}$ *is the restriction of* K *to the rows and columns indexed by the elements of* $\mathcal{S}$.

A. Sampling from a DPP and signal recovery

Given the constraints on K, its eigendecomposition $\mathsf{K} = \sum_{i=1}^N \mu_i\boldsymbol{v}_i\boldsymbol{v}_i^\mathsf{T}$ always exists, and Alg. 2 provides a sample from the associated DPP [12]. Note that $|\mathcal{A}|$, the number of elements in $\mathcal{A}$, is distributed as the sum of N Bernoulli trials of probability μ_i (see the "for loop" of Alg. 2). In particular:

$$\mathbb{E}(|\mathcal{A}|) = \mathrm{tr}(\mathsf{K}) = \sum_{i=1}^N \mu_i \text{ and } \mathrm{Var}(|\mathcal{A}|) = \sum_{i=1}^N \mu_i(1 - \mu_i). \tag{11}$$

Algorithm 2 Sampling a DPP with marginal kernel K [12]

Input: Eigendecomposition of K : $\{\mu_i, \mathsf{V} = (\boldsymbol{v}_1| \ldots |\boldsymbol{v}_N)\}$.
$\mathcal{J} \leftarrow \emptyset$
for $n = 1, 2, \ldots, N$ **do**:
    $\mathcal{J} \leftarrow \mathcal{J} \cup \{n\}$ with probability μ_n
end for
$V \leftarrow \{\boldsymbol{v}_n\}_{n\in\mathcal{J}}$
$\mathcal{Y} \leftarrow \emptyset$
while $|V| > 0$ **do**:
    · Select a new node i with $\mathrm{Pr}(i) = \frac{1}{|V|}\sum_{\boldsymbol{v}\in V}(\boldsymbol{v}^\mathsf{T}\boldsymbol{e}_i)^2$
    · $\mathcal{Y} \leftarrow \mathcal{Y} \cup \{i\}$
    · $V \leftarrow V_\perp$, an orthonormal basis for the subspace of V
      $\perp$ to $\boldsymbol{e}_i$.
end while
Output: $\mathcal{Y}$.

Consider $\mathcal{A}$, the random variable of a DPP with kernel K; and M its associated measurement matrix. Denote by $\pi_i = \mathsf{K}_{ii}$ the marginal probability that $\{i\} \subseteq \mathcal{A}$ and write:

$$\mathsf{P} = \mathrm{diag}(\pi_{\omega_1}| \ldots |\pi_{\omega_m}) \tag{12}$$

Proposition III.1. $\forall \boldsymbol{x} \in \mathbb{R}^N$ $\mathbb{E}_\mathcal{A}\left(\|\mathsf{P}^{-1/2}\mathsf{M}\boldsymbol{x}\|^2\right) = \|\boldsymbol{x}\|^2$.

Proof. Let $\boldsymbol{x} \in \mathbb{R}^N$. In the following, we write $\mathbb{P}(\mathcal{A})$ the probability that $\mathcal{A}$ is sampled. Note that: $\forall i$ $\sum_{\mathcal{A}\supset i}\mathbb{P}(\mathcal{A}) = \mathbb{P}(\{i\}$ is sampled$) = \pi_i$. Thus:

$$\mathbb{E}_\mathcal{A}\left(\|\mathsf{P}^{-1/2}\mathsf{M}\boldsymbol{x}\|^2\right) = \mathbb{E}_\mathcal{A}\left(\sum_{i\in\mathcal{A}}\frac{x_i^2}{\pi_i}\right) = \sum_\mathcal{A}\mathbb{P}(\mathcal{A})\sum_{i\in\mathcal{A}}\frac{x_i^2}{\pi_i}$$
$$= \sum_{i=1}^N x_i^2\sum_{\mathcal{A}\supset i}\frac{\mathbb{P}(\mathcal{A})}{\pi_i} = \sum_{i=1}^N x_i^2 = \|\boldsymbol{x}\|^2 \quad\square$$

This result is the first step to prove a restricted isometry property (RIP) for $\mathsf{P}^{-1/2}\mathsf{M}$ – currently work in progress. In this paper's context, it shows how to reweight the measurement in the case of sampling from a DPP: recovery should be performed via optimization of (9) or (10) with P as in (12).

B. If U_k is known: the ideal low-pass marginal kernel

Consider the DPP defined by the following marginal kernel:

$$\mathsf{K}_k = \mathsf{U}_k\mathsf{U}_k^\mathsf{T} = \mathsf{U}\,h_{\lambda_k}(\Lambda)\,\mathsf{U}^\mathsf{T}, \tag{13}$$

with h_{λ_k} s.t. $h_{\lambda_k}(\lambda) = 1$ if $\lambda \leqslant \lambda_k$ and 0 otherwise. In GSP words, K_k is the ideal low-pass with cutting frequency λ_k.

Proposition III.2. *A sample from a DPP with kernel* K_k *is of size k and* $\forall\boldsymbol{x} \in \mathrm{span}(\mathsf{U}_k)$, *the measurement* $\boldsymbol{y} = \mathsf{M}\boldsymbol{x} + \boldsymbol{n} \in \mathbb{R}^k$ *enables perfect reconstruction up to the noise level.*

Proof. $\mathsf{K}_k = \sum_{i=1}^N \mu_i\boldsymbol{u}_i\boldsymbol{u}_i^\mathsf{T}$, with $\mu_1 = \ldots = \mu_k = 1$ and $\mu_{k+1} = \ldots = \mu_N = 0$. Therefore, Eq. (11) implies that any sample $\mathcal{A}$ has size k. Also, by definition, a sample $\mathcal{A}$ verifies: $det(\mathsf{K}_{k\mathcal{A}}) > 0$, i.e., $det(\mathsf{M}\mathsf{U}_k\mathsf{U}_k^\mathsf{T}\mathsf{M}^\mathsf{T}) = \prod_{i=1}^k \sigma_i^2 > 0$, implying $\sigma_1 > 0$. Using Prop. II.1 completes the proof. $\quad\square$

Algorithm 3 A variation of Wilson's algorithm (see Sec. III-C)

Input: Adjacency matrix W of the graph and $q > 0$
$\mathcal{Y} \leftarrow \emptyset, \mathcal{V} \leftarrow \emptyset$
while $\mathcal{V} \neq \{1, \ldots, N\}$ **do**:
  · Start a random walk from any node $i \in \{1, \ldots, N\} \backslash \mathcal{V}$
    until it reaches either Δ, or a node in $\mathcal{V}$.
  · Erase all the loops of the trajectory and denote by $\mathcal{S}$
    the set of all new visited nodes (after erasure).
  · $\mathcal{V} \leftarrow \mathcal{V} \cup \mathcal{S}$
  if the last node of the trajectory is Δ **do**:
    · Denote by l the last visited node before Δ
    · $\mathcal{Y} \leftarrow \mathcal{Y} \cup \{l\}$
Output: $\mathcal{Y}$.

The most probable sample from this DPP is the sample $\mathcal{A}$ of size k that maximizes $\det(\mathsf{M}\mathsf{U}_k\mathsf{U}_k^\mathsf{T}\mathsf{M}^\mathsf{T})$ *i.e.*, $\mathcal{A}^{\mathrm{MV}}$, the solution to the maximal volume optimisation problem (5).

C. If U_k is unknown: Wilson's marginal kernel

Avena and Gaudillière [11] slightly change the so-called Wilson's algorithm [16] originally introduced in the context of uniform spanning tree sampling, in order to output a random forest, the roots of which we will consider our sampling set. Basically, the idea is to first add a node, Δ, to the graph, that we will call "absorbing state" and that we connect to all existing N nodes with a weight q. We will propagate loop-erased random walks on this extended graph. Note that the probability to jump from any node $i \neq \Delta$:

- to any other node $j \neq \Delta$ is equal to $\mathsf{W}_{ij}/(\mathsf{D}_{ii} + q)$;
- to Δ is $q/(\mathsf{D}_{ii} + q)$.

Once the random walk reaches Δ, it cannot escape from it. Algorithm 3 sums up the sampling procedure.

Proposition III.3 (Corollary of Thm 3.4 in [11]). *$\mathcal{Y}$ is a sample from a DPP with marginal kernel $\mathsf{K}_q = \mathsf{U}g_q(\Lambda)\mathsf{U}^\mathsf{T}$, with $g_q(\lambda) = \frac{q}{q+\lambda}$.*

Proof. As q is constant over all nodes, Thm 3.4 and Lemma 3.3 in [11] state that $\mathcal{Y}$ is distributed as a DPP with kernel $(L + q\mathsf{I})^{-1}q\mathsf{I}$ where I is the identity matrix (note that L in [11] refers to $-\mathsf{L}$ in our paper). Writing L as $\mathsf{U}\Lambda\mathsf{U}^\mathsf{T}$ ends the proof. $\square$

Prop. III.2 shows that sampling with K_k is optimal in the sense it enables perfect reconstruction with only $m = k$ measurements. Unfortunately, as N or k increase, computing K_k becomes prohibitive. Considering g_q as an approximation of h_{λ_k}, one may see sampling with K_q as an approximation of optimal sampling with K_k. Prop. III.3 has a beautiful consequence: one does not need to explicitly compute K_q to sample from it. In fact, Alg. 3 efficiently approximates sampling from K_k without any spectral computation of L.

IV. EXPERIMENTS

The Stochastic Block Model (SBM). We consider random community-structured graphs drawn from the SBM. We specifically look at graphs with k communities of same size

---

Algorithm 4 Estimation of $\pi_i = \sum_{j=1}^{N} g_q(\lambda_j)\boldsymbol{u}_j(i)^2$

Input: Laplacian L, $g_q : \mathbb{R} \to \mathbb{R}^+$, parameters d and n
· Compute λ_N the largest eigenvalue of L
· Approximate $s = \sqrt{g_q}$ on $[0, \lambda_N]$ with a polynomial of order d *i.e.*, $\mathsf{S} = \mathsf{U}s(\Lambda)\mathsf{U}^\mathsf{T} \simeq \mathsf{U}\sum_{l=1}^{d}\beta_l\Lambda^l\mathsf{U}^\mathsf{T} = \sum_{l=1}^{d}\beta_l\mathsf{L}^l$
· Generate $\mathsf{R} \in \mathbb{R}^{N \times n}$ with $\forall(i,j)$ $\mathsf{R}_{ij} = \mathcal{N}(0, 1/n)$
· Compute $\mathsf{SR} \simeq \sum_{l=1}^{d}\beta_l\mathsf{L}^l\mathsf{R}$ recursively
Output: $\forall i$ $\hat{\pi}_i = \|\boldsymbol{\delta}_i^\mathsf{T}\mathsf{SR}\|^2$

N/k. In the SBM, the probability of connection between any two nodes i and j is q_1 if they are in the same community, and q_2 otherwise. One can show that the average degree reads $c = q_1\left(\frac{N}{k} - 1\right) + q_2\left(N - \frac{N}{k}\right)$. Thus, instead of providing the probabilities (q_1, q_2), one may characterize a SBM by considering $(\epsilon = \frac{q_2}{q_1}, c)$. The larger ϵ, the fuzzier the community structure. In fact, authors in [17] show that above the critical value $\epsilon_c = (c - \sqrt{c})/(c + \sqrt{c}(k-1))$, community structure becomes undetectable in the large N limit. In the following, $N = 100$, $k = 2$, $c = 16$ and we let ϵ vary.

To generate a k-bandlimited signal, we draw k realisations of a normal distribution with mean 0 and variance 1, to build the vector $\boldsymbol{\alpha} \in \mathbb{R}^k$. Then, renormalize to make sure that $\|\boldsymbol{\alpha}\|^2 = 1$, and write $\boldsymbol{x} = \mathsf{U}_k\boldsymbol{\alpha}$. In our experiments, the measurement noise is Gaussian with mean 0 and variance $\sigma_n^2 = 10^{-4}$, corresponding to an SNR equal to 20 dB.

First series: U_k known. We compare DPP sampling from K_k followed by recovery using (9); to deterministic sampling with $\hat{\mathcal{A}}^{\mathrm{WCE}}$, $\hat{\mathcal{A}}^{\mathrm{MSE}}$, $\hat{\mathcal{A}}^{\mathrm{MV}}$ and $\hat{\mathcal{A}}^{\mathrm{maxvol}}$ presented in Sec. II-A followed by recovery using (2). We show in Fig. 1a) the recovery performance with respect to ϵ/ϵ_c: all greedy methods perform similarly, and on average slightly outperform the DPP-based method, especially as the graphs become less structured.

Second series: U_k unknown. We compare negatively correlated sampling from K_q (using Wilson's algorithm) followed by recovery using (10) with P as in (12); to the uncorrelated random sampling [5] where nodes are sampled independently with replacement from p^* with $p_i^* = \|\mathsf{U}_k^\mathsf{T}\boldsymbol{\delta}_i\|^2/k$, followed by recovery using (10) with P as in (8). Several comments are in order. i) the number of samples from Wilson's algorithm is not known in advance, but we know its expected value $\mathbb{E}(|\mathcal{A}|) = \sum_i q/(q + \lambda_i)$. In order to explore the behavior around the critical number of measurements $m = k$, one needs to choose q s.t. $\sum_i q/(q + \lambda_i) \simeq k$, but without computing the λ_i! To do so, we follow the proposition of [11] (see discussion around Eq. (4.14)) based on a few runs of Alg. 3. Once a sample of size approximately equal to k is exhibited, and for fair comparison, one uses the same number of nodes for the uncorrelated sampling. ii) exact computation of p^* for the uncorrelated sampling requires U_k. We follow the efficient Alg. 1 of [5] to approximate p^*. iii) recovery using P as in (12) requires to know $\pi_i = \sum_{j=1}^{N} g_q(\lambda_j)\boldsymbol{u}_j(i)^2$ for each node i. To estimate π_i without spectral decomposition of L, an elegant solution is via fast graph filtering of random signals, as in Alg. 4. Building upon the Johnson-Lindenstrauss lemma, one
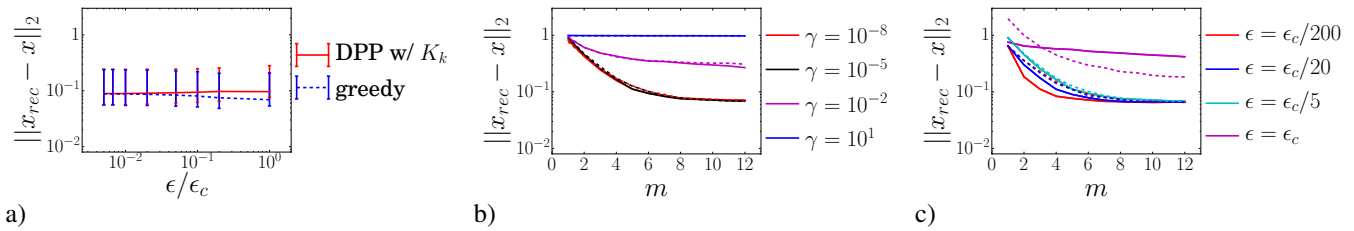
Fig. 1: a) Comparison of the recovery performance after noisy measurements of k-bandlimited signals between DPP sampling with K_k, and the 4 sampling methods presented in Sec. II-A (grouped here as "greedy": they perform equivalently). The intervals represent the 10- and 90-percentiles; the lines the averages, computed over 10^4 different signals defined on 100 different SBM realisations for each value of ϵ. b) Reconstruction performance versus the number of measurements m for different values of the regularization parameter γ, for $\epsilon = \epsilon_c/5$. c) Performance versus m for different values of ϵ. In b) and c), the full (resp. dashed) lines represent results obtained after DPP sampling via Wilson's algorithm (resp. after uncorrelated sampling such as in [5]). For each value of m, we average over 3500 bandlimited signals defined on 100 different SBM realisations.

can show that $\hat{\pi}_i$ concentrates around π_i for $n = O(\log N)$ (see a similar proof in [18]). We fix $n = 20 \log N$ and $d = 30$.

The reconstruction's parameters are γ and r. Following [5], r is fixed to 4. Fig. 1b) shows that the reconstruction performance saturates at $\gamma = 10^{-5}$. With these values of r and γ, we compare DPP sampling with K_q vs uncorrelated sampling in Fig. 1c) with respect to the number of measurements, for different values of ϵ. The lower ϵ, *i.e.* the stronger the community structure, the better DPP sampling is compared to uncorrelated sampling. In terms of computation time, a precise comparison of both methods versus the different parameters is out of this paper's scope. Nevertheless, to give an idea of the method's scalability, for a SBM with $N = 10^5$ (resp. 10^6), $k = 2$, $c = 16$ and $\epsilon = \epsilon_c/5$, and with $q = 5 \cdot 10^{-4}$, Wilson's algorithm outputs in average 5 (resp. 36) samples in a mean time of 7 (resp. 90) seconds, using Python on a laptop.

V. CONCLUSION

We show that sampling from a DPP with marginal kernel $\mathsf{K}_k = \mathsf{U}_k \mathsf{U}_k^\mathsf{T}$ outputs a subset of size k that enables perfect reconstruction, with a robustness to noise comparable to state-of-the-art greedy sampling algorithms. Moreover, in the important and fairly common case where the first k eigenvectors of the Laplacian are not computable, we show that Wilson's algorithm (Alg. 3) may be leveraged as an efficient graph sampling scheme that approximates sampling from K_k. Preliminary experiments on the SBM suggest that in the cases where the number of samples m needs to stay close to k and where the community structure of the graph is strong, sampling with Wilson's algorithm enables a better reconstruction than the state-of-the-art uncorrelated random sampling. Further investigation, both theoretical and experimental, is necessary to better specify which families of graphs are favorable to DPP sampling, and which are not.

REFERENCES

[1] D. Shuman, S. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *Signal Processing Magazine, IEEE*, vol. 30, no. 3, pp. 83–98, May 2013.

[2] I. Pesenson, "Sampling in Paley-Wiener spaces on combinatorial graphs," *Transactions of the AMS*, vol. 360, no. 10, 2008.

[3] A. Anis, A. Gadde, and A. Ortega, "Efficient Sampling Set Selection for Bandlimited Graph Signals Using Graph Spectral Proxies," *IEEE Transactions on Signal Processing*, vol. 64, no. 14, pp. 3775–3789, 2016.

[4] S. Chen, R. Varma, A. Sandryhaila, and J. Kovačević, "Discrete signal processing on graphs: Sampling theory," *IEEE Transactions on Signal Processing*, vol. 63, no. 24, pp. 6510–6523, 2015.

[5] G. Puy, N. Tremblay, R. Gribonval, and P. Vandergheynst, "Random sampling of bandlimited signals on graphs," *Applied and Computational Harmonic Analysis*, 2016, in press.

[6] M. Tsitsvero, S. Barbarossa, and P. Di Lorenzo, "Signals on Graphs: Uncertainty Principle and Sampling," *IEEE Transactions on Signal Processing*, vol. 64, no. 18, pp. 4845–4860, Sep. 2016.

[7] D. I. Shuman, M. J. Faraji, and P. Vandergheynst, "A Multiscale Pyramid Transform for Graph Signals," *IEEE Transactions on Signal Processing*, vol. 64, no. 8, pp. 2119–2134, Apr. 2016.

[8] S. Narang and A. Ortega, "Compact Support Biorthogonal Wavelet Filterbanks for Arbitrary Undirected Graphs," *Signal Processing, IEEE Transactions on*, vol. 61, no. 19, pp. 4673–4685, Oct. 2013.

[9] P. Drineas and M. W. Mahoney, "RandNLA: randomized numerical linear algebra," *Communications of the ACM*, vol. 59, no. 6, 2016.

[10] N. Tremblay, G. Puy, R. Gribonval, and P. Vandergheynst, "Compressive spectral clustering," in *Proceedings of the 33 rd International Conference on Machine Learning (ICML)*, vol. 48, 2016, pp. 1002–1011.

[11] L. Avena and A. Gaudillière, "On some random forests with determinantal roots," *arXiv preprint arXiv:1310.1723*, 2013.

[12] A. Kulesza and B. Taskar, "Determinantal Point Processes for Machine Learning," *Found. and Trends in Mach. Learn.*, vol. 5, no. 2–3, 2012.

[13] S. A. Goreinov, I. V. Oseledets, D. V. Savostyanov, E. E. Tyrtyshnikov, and N. L. Zamarashkin, "How to find a good submatrix," *Research Report 08-10, ICM HKBU, Kowloon Tong, Hong Kong*, pp. 08–10, 2008.

[14] S. P. Chepuri and G. Leus, "Subsampling for graph power spectrum estimation," in *SAM SP Workshop*. IEEE, 2016.

[15] L. F. O. Chamon and A. Ribeiro, "Greedy sampling of graph signals," *arXiv preprint arXiv:1704.01223*, 2017.

[16] D. B. Wilson, "Generating random spanning trees more quickly than the cover time," in *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*. ACM, 1996, pp. 296–303.

[17] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová, "Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications," *Phys. Rev. E*, vol. 84, no. 6, 2011.

[18] N. Tremblay, G. Puy, P. Borgnat, R. Gribonval, and P. Vandergheynst, "Accelerated Spectral Clustering Using Graph Filtering Of Random Signals," in *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016, pp. 4094–4098.