Combinatorics, Probability and Computing (2022), **31**, pp. 1048–1069
doi:10.1017/S0963548322000098

PAPER

Community detection and percolation of information in a geometric setting

Ronen Eldan[†], Dan Mikulincer[*,‡] and Hester Pieters[§]

Weizmann Institute, Rehovot, Israel
[*]Corresponding author. Email: danmiku@gmail.com

(Received 12 July 2021; revised 21 January 2022; accepted 4 May 2022; first published online 31 May 2022)

Abstract

We make the first steps towards generalising the theory of stochastic block models, in the sparse regime, towards a model where the discrete community structure is replaced by an underlying geometry. We consider a geometric random graph over a homogeneous metric space where the probability of two vertices to be connected is an arbitrary function of the distance. We give sufficient conditions under which the locations can be recovered (up to an isomorphism of the space) in the sparse regime. Moreover, we define a geometric counterpart of the model of flow of information on trees, due to Mossel and Peres, in which one considers a branching random walk on a sphere and the goal is to recover the location of the root based on the locations of leaves. We give some sufficient conditions for percolation and for non-percolation of information in this model.

Keywords: random geometric graphs; community detection

2020 MSC Codes: Primary: 62G05, Secondary: 62E20, 60J80

1. Introduction

Community detection in large networks is a central task in data science. It is often the case that one gets to observe a large network, the links of which depends on some unknown, underlying community structure. A natural task in this case is to detect and recover this community structure to the best possible accuracy. Perhaps the most well-studied model in this topic is the *stochastic block model* (SBM) where a random graph whose vertex set is composed of several communities, $\{c_1, \ldots, c_k\}$ is generated in a way that every pair of nodes v, u which belong to communities $c(u), c(v)$, will be connected to each other with probability $p = p(c(v), c(u))$, hence with probability that only depends on the respective communities, and otherwise independently. The task is to recover the communities $c(\cdot)$ based on the graph. The (unknown) association of nodes with communities is usually assumed to be random and independent between different nodes. See [1] for an extensive review of this model. A natural extension of the SBM is the geometric random graph, where the discrete set of communities is replaced by a metric space. More formally, given a metric space (X, d), a function $f : V \to X$ from a vertex set V to the metric space and a function $\varphi : \mathbb{R}_+ \to [0, 1]$, a graph is formed by connecting each pair of vertices u, v independently, with probability

$$p(u, v) := \varphi\left(d(f(u), f(v))\right).$$

[†]Supported by a European Research Council Starting Grant (ERC StG) and by an Israel Science Foundation Grant no. 715/16.

[‡]Supported by an Azrieli foundation fellowship.

[§]Supported by an Israel Science Foundation Grant no. 715/16.

© The Author(s), 2022. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

https://doi.org/10.1017/S0963548322000098 Published online by Cambridge University Press

This model can sometimes mimic the behaviour of real-world networks more accurately than the SBM. For example, a user in a social network may be represented as a point in some linear space in a way that the coordinates correspond to attributes of her personality and her geographic location. The likelihood of two persons being associated with each other in the network will then depend on the proximity of several of these attributes. A flat community structure may therefore be two simplistic to reflect these underlying attributes. Therefore, a natural extension of the theory of SBMs would be to understand under what conditions the geometric representation can be recovered by looking at the graph. Our focus is on the case that the metric is defined over a symmetric space, such as the Euclidean sphere in d-dimensions. By symmetry, we mean that the probability of two vertices to be connected, given their locations, is invariant under a natural group acting on the space. We are interested in the sparse regime where the expected degrees of the vertices do not converge to infinity with the size of the graph. This is the (arguably) natural and most challenging regime for the SBM.

1.1 Inference in geometric random graphs

For the sake of simplicity, in what follows, we will assume that the metric space is the Euclidean sphere, and our main theorems will be formulated in this setting; It will be straightforward, yet technical, to generalise our results to more general symmetric space (see [6] for further discussion on this point).We begin by introducing a model for a random geometric graph. Let σ be the uniform probability measure on $\mathbb{S}^{d-1}$ and let $\varphi : \mathbb{S}^{d-1} \times \mathbb{S}^{d-1} \to \mathbb{R}_+$, be of the form $\varphi(x, y) = f(\langle x, y\rangle)$ for some $f : [-1, 1] \to \mathbb{R}_+$. Consider $\{X_i\}_{i=1}^n \sim \sigma$, a sequence of independently sampled vectors. Let $G\left(n, \frac{1}{n}\varphi(X_i, X_j)\right)$ be the inhomogeneous Erdös-Rényi graph model where edges are formed independently with probability $\frac{1}{n}\varphi(X_i, X_j)$ and let A be the adjacency matrix of a random graph drawn from $G\left(n, \frac{1}{n}\varphi(X_i, X_j)\right)$. We may freely assume that n is large enough so that, necessarily, $\frac{1}{n}\varphi(X_i, X_j) \leq 1$.We now define the task of recovering the latent positions of the sample $\{X_i\}_{i=1}^n$ from an observation of the random geometric graph.

Definition 1. We say that the model is $\varepsilon -$ *reconstructible* if, for all n large enough, there is an algorithm which returns an $n \times n$ matrix $\mathcal{X}$ such that

$$\frac{1}{n^2} \sum_{i,j} |\mathcal{X}_{i,j} - X_i \cdot X_j|^2 \leq \varepsilon.$$

Remark that, due the symmetry of the model, it is clear that the locations can only be reconstructed up to an orthogonal transformation, which is equivalent to reconstruction of the Gram matrix.

In order state the conditions under which we prove the possibility of reconstructions, we need some notation. For φ as above, define the integral operator $A_\varphi : L^2\left(\mathbb{S}^{d-1}\right) \to L^2\left(\mathbb{S}^{d-1}\right)$ by

$$A_\varphi(g)(x) = \int_{\mathbb{S}^{d-1}} \varphi(x, y)g(y)d\sigma(y).$$

It is standard to show that A_φ is a self-adjoint compact operator (see [10], for example) and so has a discrete spectrum, except at 0. By definition, φ is invariant to rotations and so A_φ commutes with the Laplacian. It follows that the eigenfunctions of A_φ are precisely the *spherical harmonics* which we denote by $\{\psi_i\}_{i=0}^\infty$. Thus, if $\lambda_i(A_\varphi)$ denotes the eigenvalue of φ corresponding to ψ_i we have the following identity,

$$\varphi = \sum_{i=0}^{\infty} \lambda_i \psi_i \otimes \psi_i. \tag{1}$$

https://doi.org/10.1017/S0963548322000098 Published online by Cambridge University Press

In particular, $\psi_0 = 1$ and for $i = 1, \ldots, d$, ψ_i are linear functionals such that, for $x, y \in \mathbb{S}^{d-1}$,

$$d \cdot \langle x, y \rangle = \sum_{l=1}^{d} \psi_l(x)\psi_l(y). \tag{2}$$

Note that in our notation the eigenvalues are indexed by the spherical harmonics, and are therefore not necessarily in decreasing order. By rotational invariance it must hold that

$$\lambda(\varphi) := \lambda_1 = \ldots = \lambda_d. \tag{3}$$

Define $\|\varphi\|_\infty = \sup_{x,y} \varphi(x, y)$. We make the following, arguably natural, assumptions on the function φ:

A1. There exist $\delta > 0$ such that $\min_{i \neq 1,\ldots,d} |\lambda(\varphi) - \lambda_i| > \delta$.
A2. Reordering the eigenvalues in decreasing order $\lambda_{l_0} \geq \lambda_{l_1} \geq \lambda_{l_2} \geq \ldots$ there exists $C > 0$ such that for every $i \geq 0$, $|\lambda_{l_i}| \leq \frac{C}{(i+1)^2}$.

Remark 2. For the spectral algorithm we propose, the eigenspace associated to $\lambda(\varphi)$ plays as a special role, since it may encode positions of latent vectors in $\mathbb{R}^d$. Therefore, to extract meaningful information from the algorithm, it is crucial that this eigenspace is d-dimensional; a property that is enforced by Assumption A1. However, it is not clear what would happen if, say, λ_{d+1} is very close to λ_d. Our analysis, which is based on stability of the eigenspaces to perturbations of the kernel, fails in this case, but it is plausible the algorithm (or another one) may still succeed.

Regarding Assumption A2. While, it is usually the case that, some integrability of eigenvalues is necessary for the problems we consider, we do not know whether the quadratic decay rate is tight. We have made no efforts to optimise this dependence and are willing to conjecture that there are possible improvements.

Theorem 3. *For every $\varepsilon > 0$ there exists a constant $C = C(\varepsilon, d)$, such that the model is ε-reconstructible whenever*

$$\min_{i \neq 0,\ldots,d} |\lambda_i - \lambda(\varphi)|^2 \geq C\|\varphi\|_\infty. \tag{4}$$

Remark 4. Observe that, since the left hand side of condition (4) is 2-homogeneous, whereas its right hand side is 1-homogeneous, we have that as long as the left hand side is nonzero, by multiplication of the function φ by a large enough constant, the condition can be made to hold true.

Theorem 3 can be compared to the known bounds for recovery in the SBM (see [1]). In particular, it is illuminating to compare the SBM with two communities to linear kernels in our model. In this case, both models are parameterised by two numbers. In the SBM these are the inter- and intra-communities probabilities and in our model, the coefficients of the linear function. In the SBM, denote the parameters as a and b, then recovery of the communities depends on the ratio $\frac{(a-b)^2}{a+b}$. The example below gives a similar result for linear kernels, with a dimensional affect, which typically makes reconstruction easier.

Example 5. Consider the linear kernel, $\varphi(x, y) = \gamma + \zeta \langle x, y \rangle$, with $|\zeta| \leq \gamma$. A calculation shows that

$$\lambda_0 = \gamma$$

$$\lambda(\varphi) = \frac{\zeta}{d}.$$

https://doi.org/10.1017/S0963548322000098 Published online by Cambridge University Press

Applying our theorem, we show that the model is reconstructible whenever

$$\left|\gamma - \frac{\zeta}{d}\right|^2 \geq C \cdot (\gamma + \zeta).$$

Methods and related works. Our reconstruction theorem is based on a spectral method, via the following steps:

1. We observe that by symmetry of our kernel, linear functions are among its eigenfunctions. We show that the kernel matrix (hence the matrix obtained by evaluating the kernel at pairs of the points (X_i)) will have respective eigenvalues and eigenvectors which approximate the ones of the continuous kernel.
2. Observing that the kernel matrix is the expectation of the adjacency matrix, we rely on a matrix concentration inequality due to Le-Levina-Vershynin [12] to show that the eigenvalues of the former are close to the ones of the latter.
3. We use the Davis-Kahan theorem to show that the corresponding eigenvectors are also close to each other.

The idea in Steps 2 and 3 is not new, and rather standard (see [12] and references therein). Thus, the main technical contribution in proving our upper bound is in Step 1, where we prove a bound for the convergence of eigenvectors of kernel matrices. So far, similar results have only been obtained in the special case that the Kernel is positive definite, see for instance [4].The paper [20] considers kernels satisfying some Sobolev-type hypotheses similar to our assumptions on φ (but gives results on the spectrum rather than the eigenvectors). Reconstruction of the eigenspaces has been considered in [18], for positive definite kernels in the dense regime; in [17], for random dot products graphs; and in [2] in the dense and relatively sparse regimes, again for kernels satisfying some Sobolev-type hypotheses. Let us also mention other works which augmented the SBM with geometric information. The paper [7] considers the problem of discrete community recovery when presented with informative node covariates, inspired by the spiked covariance model. The authors derived a sharp threshold for recovery by introducing a spectral-like algorithm. However, the model is rather different than the one we propose in which the community structure is continuous. A model which is slightly closer to the one we consider appears in [9]. In this model, communities are still discrete but the edge connectivity depends continuously on the latent positions of nodes on some d dimensional sphere. In such a model, recovery of the communities may be reduced to more combinatorial arguments. Indeed, the number of common neighbours can serve as an indicator for checking whether two nodes come from the same community. A similar idea was previously explored in [5], where a triangle count was used to establish a threshold for detecting latent geometry in random geometric graphs.

1.2 Percolation of geometric information in trees

The above theorem gives an upper bound for the threshold for reconstruction. The question of finding respective lower bounds, in the SBM, is usually reduced to a related but somewhat simpler model of percolation of information on *trees*. The idea is that in the sparse regime, the neighbourhood of each node in the graph is usually a tree, and it can be shown that recovering the community of a specific node based on observation of the entire graph, is more difficult than the recovery of its location based on knowledge of the community association of the leaves of a tree rooted at this node. For a formal derivation of this reduction (in the case of the SBM), we refer to [1].This gives rise to the following model, first described in Mossel and Peres [15] (see also [14]): Consider a q-ary tree $T = (V, E)$ of depth k, rooted at $r \in V$. Suppose that each node

https://doi.org/10.1017/S0963548322000098 Published online by Cambridge University Press

in V is associated with a label $\ell : V \to \{1, .., k\}$ in the following way: The root r is assigned with some label and then, iteratively, each node is assigned with its direct ancestor's label with probability p and with a uniformly picked label with probability $1-p$ (independent between the nodes at each level). The goal is then to detect the assignment of the root based on observation of the leaves. Let us now suggest an extension of this model to the geometric setting. We fix a Markov kernel $\varphi(x,y) = f(\langle x, y\rangle)$, normalised such that $\int_{\mathbb{S}^{n-1}} \varphi(x,y)d\sigma(y) = 1$ for all $x \in \mathbb{S}^{d-1}$. We define $g : T \to \mathbb{S}^{d-1}$ in the following way. For the root r, $g(r)$ is picked according to the uniform measure. Iteratively, given that $g(v)$ is already set for all nodes v at the ℓ-th level, we pick the values $g(u)$ for nodes u at the $(\ell+1)$th level independently, so that if u is a direct descendant of v, the label $g(u)$ is distributed according to the law $\varphi(g(v), \cdot)d\sigma$.

Denote by $T_k \subset V$ the set of nodes at depth k, and define by μ_k the conditional distribution of $g(r)$ given $(g(v))_{v\in T_k}$. We say that the model has positive information flow if

$$\lim_{k\to\infty} \mathbb{E}\left[\mathrm{TV}(\mu_k, \sigma)\right] > 0.$$

Remark that by symmetry, we have

$$\mathbb{E}\left[\mathrm{TV}(\mu_k, \sigma)\right] = \mathbb{E}\left[\mathrm{TV}(\mu_k, \sigma)|g(r) = e_1\right]$$

where r is the root and e_1 is the north pole.

Our second objective in this work is to make the first steps towards understanding under which conditions the model has positive information flow, and in particular, our focus is on providing nontrivial sufficient conditions on q, φ for the above limit to be equal to zero.

Let us first outline a natural sufficient condition for the information flow to be positive which, as we later show, turns out to be sharp in the case of Gaussian kernels. Consider the following simple observable,

$$Z_k := \frac{1}{|T_k|} \sum_{v\in T_k} g(v).$$

By Bayes' rule, we clearly have that the model has positive information flow if (but not only if)

$$\liminf_{k\to\infty} \frac{\mathbb{E}[\langle Z_k, e_1\rangle|g(r) = e_1]}{\sqrt{\mathrm{Var}\left[\langle Z_k, e_1\rangle|\, g(r) = e_1\right]}} > 0. \tag{5}$$

This gives rise to the parameter

$$\lambda(\varphi) := \int_{\mathbb{S}^{d-1}} \langle x, e_1\rangle \varphi(e_1, x)d\sigma(x),$$

which is the eigenvalue corresponding to linear harmonics. As $\langle x, e_1\rangle$ is an eigenfunction of the Markov kernel, it is preserved under repeated applications. Coupling this observation with the linearity of expectation, we have

$$\mathbb{E}[\langle Z_k, e_1\rangle|g(r) = e_1] = \lambda(\varphi)^k. \tag{6}$$

Moreover, by symmetry if $j \neq 1$,

$$0 = \int_{\mathbb{S}^{d-1}} \langle x, e_j\rangle \varphi(e_1, x)d\sigma(x),$$

from which we may conclude for any $v \in \mathbb{S}^{d-1}$,

$$\lambda(\varphi)\langle v, e_1\rangle = \int_{\mathbb{S}^{d-1}} \langle x, v\rangle \varphi(e_1, x)d\sigma(x), \tag{7}$$

https://doi.org/10.1017/S0963548322000098 Published online by Cambridge University Press

For two nodes $u, v \in T$ define by $c(u, v)$ the deepest common ancestor of u, v and by $\ell(u, v)$ its level. A calculation gives

$$\begin{aligned}
\operatorname{Var}\left[\langle Z_k, e_1\rangle \,|\, g(r) = e_1\right] &= \frac{1}{q^{2k}} \sum_{u,v\in T_k} \mathbb{E}\left[g(v)_1 g(u)_1 | g(r) = e_1\right] - \lambda(\varphi)^{2k} \\
&= \frac{1}{q^{2k}} \sum_{u,v\in T_k} \mathbb{E}\left[\mathbb{E}[g(v)_1 | g(c(u,v))]\mathbb{E}[g(u)_1 | g(c(u,v))] | g(r) = e_1\right] - \lambda(\varphi)^{2k} \\
&= \frac{1}{q^{2k}} \sum_{u,v\in T_k} \mathbb{E}\left[g(c(u,v))_1^2 | g(r) = e_1\right] \lambda(\varphi)^{2(k-\ell(u,v))} - \lambda(\varphi)^{2k} \\
&\le \frac{1}{q^{2k}} \sum_{u,v\in T_k} \lambda(\varphi)^{2(k-\ell(u,v))} - \lambda(\varphi)^{2k} \\
&= \frac{\lambda(\varphi)^{2k}}{q^{2k}} \sum_{u,v\in T_k} \lambda(\varphi)^{-2\ell(u,v)} - \lambda(\varphi)^{2k} \\
&\le \frac{\lambda(\varphi)^{2k}}{q^{2k}} \sum_{\ell=0}^{k} q^{\ell} q^{2(k-\ell)} \lambda(\varphi)^{-2\ell} - \lambda(\varphi)^{2k} = \lambda(\varphi)^{2k} \sum_{\ell=1}^{k} \left(q\lambda(\varphi)^2\right)^{-\ell},
\end{aligned}$$

where in the third equality, we have applied (6) and (7) to $\mathbb{E}[g(v)_1 | g(c(u,v))]$ and $\mathbb{E}[g(u)_1 | g(c(u,v))]$. This gives a sufficient condition for (5) to hold true, concluding:

Claim 6. *The condition $q\lambda(\varphi)^2 > 1$ is sufficient for the model to have positive percolation of information.*

We will refer to this as the Kesten-Stigum (KS) bound.

We now turn to describe our lower bounds. For the Gaussian kernel, we give a lower bound which misses by a factor of 2 from giving a matching bound to the KS bound. To describe the Gaussian kernel, fix $\beta > 0$, let X be a normal random variable with law $\mathcal{N}(0, \beta)$ and suppose that $\varphi : \mathbb{S}^1 \times \mathbb{S}^1 \to \mathbb{R}$ is such that

$$\varphi(x, \cdot) \text{ is the density of } (x + X) \bmod 2\pi, \tag{8}$$

where we identify $\mathbb{S}^1$ with the interval $[0, 2\pi)$. We have the following result.

Theorem 7. *For the Gaussian kernel defined above, there is zero information flow whenever $q\lambda(\varphi) < 1$.*

Related results in the Gaussian setting. In the discrete setting, an analogous result was obtained in [13]. In fact, our method of proof is closely related. We use the inherent symmetries of the spherical kernels to decouple the values of $g(r)$ and $g(v)$, where v is some vertex which is sufficiently distant from r. This is same idea of Proposition 10 in [13] which uses a decomposition of the transition matrix to deduce a similar conclusion. Another related result appears in [16]. In the paper the authors consider a broadcast model on a binary tree with a Gaussian Markov random field and obtain a result in the same spirit as the one above. However, since the random field they consider is real valued, as opposed to our process which is constrained to the circle, the method of proof is quite different.

In the general case, we were unable to give a corresponding bound, nevertheless, using the same ideas, we are able to give some nontrivial sufficient condition for zero flow of information

https://doi.org/10.1017/S0963548322000098 Published online by Cambridge University Press

for some $q > 1$, formulated in terms of the eigenvalues of the kernel. To our knowledge, this is the first known result in this direction. In order to formulate our result, we need some definitions.

We begin with a slightly generalised notion of a q-ary tree.

Definition 8. Let $q > 1$, we say that T is a tree of growth at most q if for every $k \in \mathbb{N}$,

$$|T_k| \leq \left\lceil q^k \right\rceil.$$

Now, recall that $\varphi(x, y) = f(\langle x, y\rangle)$. Our bound is proven under the following assumptions on the kernel.

- f is monotone.
- f is continuous.
- $\lambda(\varphi) > 0$ and for every $i \geq 1$, $|\lambda_i| \leq \lambda(\varphi)$.

We obtain the following result.

Theorem 9. *Let φ satisfy the assumptions above and let T be a tree of growth at most q. There exists a universal constant $c > 0$, such that if*

$$q \leq \left(1 - c\frac{\ln(\lambda(\varphi))(1-\lambda(\varphi))^2}{\ln\left(\frac{\lambda(\varphi)(1-\lambda(\varphi))}{f(1)}\right)}\right)^{-1}$$

then the model has zero percolation of information.

Possible connection to reconstruction on random geometric graphs. One might naively expect that the machinery developed in [15] might be used, in conjunction with Theorems 7 and 9, to deduce lower bounds for reconstructability in random geometric graphs. However, in contrast to the SBM, if v is a vertex and we condition on the latent positions of all vertices in the boundary of some ball around v, the position of v still retains some dependency with the position of vertices outside the ball.

Because of this fact, reducing a lower bound for the problem of recovering latent positions in random geometric graphs to the continuous broadcast model cannot be as straightforward as in the SBM. We leave the existence of a possible reduction as an interesting question for future research.

2. The upper bound: Proof of Theorem 3

Recall the following identity in $L^2(\mathbb{S}^{d-1})$,

$$\varphi = \sum_{k=0}^{\infty} \lambda_k \psi_k \otimes \psi_k,$$

with the eigenvalues λ_k indexed by the spherical harmonics. Define the random matrices M_n, Ψ_n by

$$(M_n)_{i,j} = \frac{1}{n}\varphi(X_i, X_j), \quad (\Psi_n)_{i,k} = \frac{1}{\sqrt{n}}\psi_k(X_i).$$

Note that M_n is an $n \times n$ matrix, while Ψ_n, has infinitely many columns. Furthermore, denote by Λ the diagonal matrix $\mathrm{diag}\{\lambda_i\}_{i=0}^{\infty}$. Then, σ-almost surely, for every $i, j \in [n]$,

$$(M_n)_{i,j} = \frac{1}{n}\sum_{k=0}^{\infty} \lambda_k \psi_k(X_i)\psi_k(X_j) = (\Psi_n \Lambda \Psi_n^T)_{i,j}.$$

https://doi.org/10.1017/S0963548322000098 Published online by Cambridge University Press

For $r \in \mathbb{N}$ we also denote

$$\varphi^r := \sum_{k=0}^{r} \lambda_k \psi_k \otimes \psi_k,$$

the finite rank approximation of φ, $\Lambda^r = \mathrm{diag}\{\lambda_k\}_{k=0}^r$, and Ψ_n^r the sub-matrix of Ψ_n composed of its first r columns. Finally, denote

$$M_n^r = \Psi_n^r \Lambda^r (\Psi_n^r)^T.$$

As before, let A be an adjacency matrix drawn from $G\left(n, \frac{1}{n}\varphi(X_i, X_j)\right)$ so that $\mathbb{E}\left[A|X_1, \ldots, X_n\right] = M_n$. Our goal is to recover Ψ_n^{d+1} from the observed A. The first step is to recover Ψ_n^{d+1} from M_n. We begin by showing that the columns of Ψ_n^r are, up to a small additive error, eigenvectors of M_n^r. To this end, denote

$$E_{n,r} := (\Psi_n^r)^T \Psi_n^r - \mathrm{Id}_r,$$

$C(n, r) = \left\|E_{n,r}\right\|_{op}^2$, and $K = \max_i \lambda_i$.

Lemma 10. *Let u_i be the i'th column of Ψ_n and let $\eta > 0$. Then*

$$\left\|M_n^r u_i - \lambda_i u_i\right\|_2^2 \leq K^2(\sqrt{C(n, r)} + 1)C(n, r).$$

Moreover, whenever $n \geq l(r) \log(2r/\eta)$, we have with probability larger than $1 - \eta$,

$$C(n, r) \leq \frac{4l(r) \log(2r/\eta)}{n}.$$

where $l(r)$ only depends on r and on the dimension.

Proof. Let $e_i \in \mathbb{R}^r$ be the i'th standard unit vector so that $u_i = \Psi_n^r e_i$. So,

$$(\Psi_n^r)^T u_i = (\Psi_n^r)^T \Psi_n^r e_i = (\mathrm{Id}_r + (\Psi_n^r)^T \Psi_n^r - \mathrm{Id}_r)e_i = e_i + E_{n,r} e_i.$$

We then have

$$\begin{aligned} M_n^r u_i &= \Psi_n^r \Lambda^r (\Psi_n^r)^T u_i = \Psi_n^r \Lambda^r e_i + \Psi_n^r \Lambda^r E_{n,r} e_i \\ &= \lambda_i \Psi_n^r e_i + \Psi_n^r \Lambda^r E_{n,r} e_i = \lambda_i u_i + \Psi_n^r \Lambda^r E_{n,r} e_i. \end{aligned}$$

To bound the error, we estimate $\left\|M_n^r u_i - \lambda_i u_i\right\|_2^2 = \left\|\Psi_n^r \Lambda^r E_{n,r} e_i\right\|_2^2$ as

$$\begin{aligned} \langle \Lambda^r E_{n,r} e_i, (\Psi_n^r)^T \Psi_n^r \Lambda^r E_{n,r} e_i \rangle &= \langle \Lambda^r E_{n,r} e_i, E_{n,r} \Lambda^r E_{n,r} e_i \rangle + \left\|\Lambda^r E_{n,r} e_i\right\|_2^2 \\ &\leq \left(\sqrt{C(n, r)} + 1\right) \left\|\Lambda^r E_{n,r} e_i\right\|_2^2 \leq K^2 \left(\sqrt{C(n, r)} + 1\right) C(n, r). \end{aligned}$$

It remains to bound $C(n, r)$. Let $X_i^r = (\psi_0(X_i), \ldots, \psi_{r-1}(X_i))$ stand for the i'th row of Ψ_n^r. Then, $E_{n,r} = \frac{1}{n}\sum_{i=1}^n \left((X_i^r)^T X_i^r - \mathrm{Id}_r\right)$, is a sum of independent, centred random matrices. We have

$$\begin{aligned} \sigma_{n,r}^2 &:= \left\|\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^n \left((X_i^r)^T X_i^r - \mathrm{Id}_r\right)\right)^2\right\|_{op} \\ &= \frac{1}{n}\left\|\mathbb{E}\left((X_1^r)^T X_1^r - \mathrm{Id}_r\right)^2\right\|_{op} \end{aligned}$$

https://doi.org/10.1017/S0963548322000098 Published online by Cambridge University Press

Furthermore, the norm of the matrices can be bounded by

$$\left\| \frac{1}{n} \left((X_1^r)^T X_1^r - \mathrm{Id}_r \right) \right\|_{op} = \frac{1}{n} \max \left(1, \left\| X_1^r \right\|_2^2 - 1\right)$$
$$\leq \frac{1}{n} \max \left(1, \left\| \sum_{i=0}^{r} \psi_i^2 \right\|_\infty - 1 \right).$$

Note that the right-hand side of the two last displays are of the form $\frac{1}{n} l(r)$ where $l(r)$ depends only on r and d (not on n). Applying matrix Bernstein ([19, Theorem 6.1]) then gives

$$\mathbb{P}\left(\left\| E_{n,r} \right\|_{op} \geq t \right) \leq 2r \exp\left(-\frac{n}{2l(r)} \frac{t^2}{1 + t/3} \right),$$

where $l(r)$ depends only on r and d. Choose now $t_0 = \frac{4l(r)\log(2r/\eta)}{n}$. As long as $n \geq l(r)\log(2r/\eta)$, $t_0 \leq 4$, and the above bound may be refined to

$$\mathbb{P}\left(\left\| E_{n,r} \right\|_{op} \geq t_0 \right) \leq 2r \exp\left(-\frac{n}{l(r)} \frac{t_0^2}{7} \right).$$

With the above conditions, it may now be verified that $2r \exp\left(-\frac{n}{l(r)} \frac{t_0^2}{7} \right) \leq \eta$, and the proof is complete. □

We now show that as r increases, the subset of eigenvectors of M_n^r, that encode the latent positions of $\{X_i\}_{i=1}^n$, converge to those of M_n. Order the eigenvalues in decreasing order $\lambda_{l_0} \geq \lambda_{l_1} \geq \lambda_{l_2} \geq \ldots$ and let $\Lambda_{>r} = \sum_{i=r}^{\infty} \lambda_{l_i}^2$. Note that it follows from assumption A2 that $\Lambda_{>r} = O(r^{-3})$. We will denote by $\lambda_i(M_n), \lambda_i(M_n^r)$ the respective eigenvalues of M_n and M_n^r, ordered in a decreasing way, and by $v_i(M_n), v_i(M_n^r)$ their corresponding unit eigenvectors. Suppose that s is such that

$$\lambda(\varphi) = \lambda_{l_{s+1}} = \ldots = \lambda_{l_{s+d}}. \tag{9}$$

Moreover, define

$$V_n := \operatorname{span}(v_{l_{s+1}}(M_n), \ldots, v_{l_{s+d}}(M_n)), \quad V_n^r := \operatorname{span}\left(v_{l_{s+1}}(M_n^r), \ldots, v_{l_{s+d}}(M_n^r) \right).$$

The next lemma shows that V_n is close to V_n^r whenever both n and r are large enough.

Lemma 11. *Let δ and C be the constants from Assumption A1 and Assumption A2. For all n, r, let $P_{n,r}$ be the orthogonal projection onto V_n^r. Then, for all $\eta > 0$ there exist constants n_0, r_0, which may depend on δ, C, and η, such that for all $n > n_0$ and $r > r_0$, we have with probability at least $1 - \eta$ that, for all $w \in V_n$,*

$$\left\| w - P_{n,r} w \right\|_2 \leq \frac{4C}{\eta \delta^2 r^3}.$$

https://doi.org/10.1017/S0963548322000098 Published online by Cambridge University Press

Proof. We have

$$\begin{aligned}\mathbb{E}\left\|M_n - M_n^r\right\|_F^2 &= \sum_{i,j} \mathbb{E}(M_n - M_n^r)_{i,j}^2 \\ &= \sum_{i,j} \frac{1}{n^2}\mathbb{E}\left(\sum_{k=r}^{\infty} \lambda_k \psi_k(X_i)\psi_k(X_j)\right)^2 \\ &= \mathbb{E}_{x,y\sim\sigma}\left(\sum_{k=r}^{\infty} \lambda_k \psi_k(x)\psi_k(y)\right)^2 \\ &= \sum_{k=r}^{\infty} \lambda_k^2 = \Lambda_{>r}.\end{aligned}$$

Applying Markov's inequality gives that with probability $1-\eta$

$$\left\|M_n - M_n^r\right\|_F^2 \le \frac{\Lambda_{>r}}{\eta} \le \frac{C}{\eta r^3}. \tag{10}$$

Theorem 1 in [20] shows that there exists n large enough such that with probability larger than $1-\eta$, one has

$$|\lambda_i(M_n) - \lambda_{l_{s+i}}| \le \delta/4,$$

with δ being the constant from Assumption A1. It follows that

$$\lambda_{l_{s+1}}(M_n), \ldots, \lambda_{l_{s+d}}(M_n) \in \left[\lambda(\varphi) - \frac{\delta}{4}, \lambda(\varphi) + \frac{\delta}{4}\right], \tag{11}$$

while by (10) and Weyl's perturbation theorem (e.g., [3, Corollary III.2.6]), for r large enough with probability $1-\eta$,

$$\lambda_i(M_n^r) \notin \left[\lambda(\varphi) - \frac{3\delta}{4}, \lambda(\varphi) + \frac{3\delta}{4}\right], \text{ for } i \ne l_{s+1}, \ldots l_{s+d}. \tag{12}$$

Combining (10), (11) and (12) it follows from the classical Davis-Kahan theorem (see e.g. [3, Section VII.3]) that with probability at-least $1-2\eta$, for every $w \in V_n$,

$$\left\|w - P_{n,r}w\right\|_2^2 \le \frac{4C}{\eta\delta^2 r^3}.$$

□

Denote

$$G_n := \frac{1}{d}\sum_{k=1}^{d} v_{l_{s+k}}(M_n)v_{l_{s+k}}(M_n)^T, \quad (G_n')_{i,j} = \frac{1}{n}\langle X_i, X_j\rangle.$$

A combination of the last two lemmas produces the following:

Theorem 12. *One has*

$$\|G_n - G_n'\|_F \to 0$$

in probability, as $n \to \infty$.

https://doi.org/10.1017/S0963548322000098 Published online by Cambridge University Press

Proof. Denote

$$G_n^r := \frac{1}{d}\sum_{k=1}^{d} v_{l_{s+k}}(M_n^r)(v_{l_{s+k}}(M_n^r))^T.$$

Then

$$\|G_n - G_n'\|_F \le \|G_n^r - G_n'\|_F + \|G_n - G_n^r\|_F$$

We will show that the two terms on the right hand side converge to zero. Let $r(n)$ be a function converging to infinity slowly enough so that $C(n,r) \to 0$, for the constant $C(n,r)$ defined in Lemma 10. Taking $\eta = \eta(n)$ to converge to zero slowly enough and applying Lemma 10, gives for all $1 \le i \le d$,

$$\|(M_n^r - \lambda(\varphi)\mathrm{Id}_n)u_i\|_2^2 \le \varepsilon_n$$

with u_i the i'th column of Ψ_n^r and where $\varepsilon_n \to 0$ as $n \to \infty$. Now, if we write

$$u_i = \sum_{j=0}^{\infty} \alpha_{i,j} v_j(M_n^r),$$

the last inequality becomes

$$\sum_j |\lambda_j(M_n^r) - \lambda(\varphi)|^2 \alpha_{i,j}^2 = \sum_j |(M_n^r - \lambda(\varphi)\mathrm{Id}_n)v_j(M_n^r)|^2 \alpha_{i,j}^2 \le \varepsilon_n, \ \ \forall 1 \le i \le d.$$

Recall that, by (12), for $j \notin \{l_{s+1}, .., l_{s+d}\}$, we have $\frac{3\delta}{4} \le |\lambda_j(M_n^r) - \lambda(\varphi)|$. Hence.

$$\sum_{j \notin \{l_{s+1},..,l_{s+d}\}} \alpha_{i,j}^2 \le \frac{16\varepsilon_n}{9\delta^2} \to 0, \tag{13}$$

and thus

$$\left\| u_i - \sum_{k=1}^{d} \alpha_{i,l_{s+k}} v_{l_{s+k}}(M_n^r) \right\|_2^2 \to 0, \ \ \forall 1 \le i \le d.$$

Define a $d \times d$-matrix B by $B_{i,j} = \alpha_{i,l_{s+j}}$. Then we can rewrite the above as

$$\left\|(u_1, \ldots, u_d) - (v_{l_{s+1}}(M_n^r), \ldots, v_{l_{s+d}}(M_n^r)) \cdot B\right\|_F^2 \to 0.$$

Now, since for two $n \times d$ matrices R, S we have $\|RR^T - SS^T\|_F \le (\|R\|_{op} + \|S\|_{op})\|R - S\|_F$. It follows that

$$\|G_n' - (v_{l_{s+1}}(M_n^r), \ldots, v_{l_{s+d}}(M_n^r))BB^T(v_{l_{s+1}}(M_n^r), \ldots, v_{l_{s+d}}(M_n^r))^T\|_F \to 0. \tag{14}$$

Observe that

$$B_{i,j} = \langle u_i, u_j \rangle - \sum_{k \notin \{l_{s+1},..,l_{s+d}\}} \alpha_{i,k}\alpha_{j,k},$$

implying that

$$|(BB^T)_{i,j} - (E_{i,j} + \mathrm{Id}_d)| \le \sqrt{\sum_{k \notin \{l_{s+1},..,l_{s+d}\}} \alpha_{i,k}^2 \sum_{k \notin \{l_{s+1},..,l_{s+d}\}} \alpha_{j,k}^2} \xrightarrow{13} 0.$$

where $E = E_{n,r}$. Consequently we have

$$\|BB^T - \mathrm{Id}_d\|_{op} \le \sqrt{\sum_{k \notin \{l_{s+1},..,l_{s+d}\}} \alpha_{i,k}^2 \sum_{k \notin \{l_{s+1},..,l_{s+d}\}} \alpha_{j,k}^2} + \sqrt{C(n,r)} \to 0,$$

https://doi.org/10.1017/S0963548322000098 Published online by Cambridge University Press

which implies that

$$\|v_{l_{s+1}}(M_n^r), \ldots, v_{l_{s+d}}(M_n^r))(BB^T - \mathrm{Id}_d)(v_{l_{s+1}}(M_n^r), \ldots, v_{l_{s+d}}(M_n^r))^T\|_F^2 \to 0.$$

Combining with (14) finally yields

$$\|G_n' - G_n^r\|_F \to 0.$$

in probability, as $n \to \infty$. If P is the orthogonal projection onto $V_n^r = \mathrm{span}(v_{l_{s+1}}(M_n^r), \ldots, v_{l_{s+d}}(M_n^r))$, and Q is the orthogonal projection onto $\mathrm{span}(v_{l_{s+1}}(M_n), \ldots, v_{l_{s+d}}(M_n))$, then Lemma 11 shows that for all $\eta > 0$, with probability at least $1 - \eta$, as $n \to \infty$ (and $r = n^{\frac{1}{2d}} \to \infty$), we have for every unit vector v

$$|(P - \mathrm{Id}_n)Qv| \le \varepsilon_n \tag{15}$$

with some $\varepsilon_n \to 0$. By symmetry, we also have for every unit vector v that

$$|(Q - \mathrm{Id}_n)Pv| \le \varepsilon_n$$

(this uses that fact that both P and Q are projections into subspaces of the same dimension). The last two inequalities easily yield that $\|P - Q\|_{op} \le \varepsilon_n$. Since this is true for every $\eta > 0$, it follows that

$$\|G_n - G_n^r\|_F \to 0,$$

in probability, as $n \to \infty$. □

Now, after establishing that M_n is close to Ψ_n^{d+1}, the second step is to recover M_n (and therefore Ψ_n^{d+1}), from the observed A. For the proof we will need the following instance of the Davis-Kahan theorem.

Theorem 13. ([21, Theorem 2]). *Let X, Y be symmetric matrices with eigenvalues $\lambda_0 \ge \ldots \ge \lambda_p$ resp. $\hat{\lambda}_0 \ge \ldots \ge \hat{\lambda}_p$ with corresponding orthonormal eigenvectors $v_0, \ldots, v_p$ resp. $\hat{v}_0, \ldots, \hat{v}_p$. Let $V = (v_{s+1}, \ldots, v_{s+d})$ and $\hat{V} = (\hat{v}_{s+1}, \ldots, \hat{v}_{s+d})$. Then there exists an orthogonal $d \times d$ matrix R such that*

$$\|\hat{V}R - V\|_F \le \frac{2^{3/2} \min\left(d^{1/2}\|Y - X\|_{op}, \|Y - X\|_F\right)}{\min\left(\lambda_s - \lambda_{s+1}, \lambda_{s+d} - \lambda_{s+d+1}\right)}.$$

Our main tool to pass from the expectation of the adjacency matrix to the matrix itself is the following result regarding concentration of random matrices, which follows from [11, Theorem 1.1] (see also [12, Theorem 5.1]).

Theorem 14. *Let A be the adjacency matrix of a random graph drawn from $G\left(n, \frac{1}{n}\varphi(X_i, X_j)\right)$. For every vertex of degree larger than $\max_{ij} \varphi(X_i, X_j)$, remove all incident edges and let A' denote the new adjacency matrix. Then,*

$$\|A' - M_n\| \le C\sqrt{\|\varphi\|_\infty}.$$

We can now prove the main reconstruction theorem.

Proof of Theorem 3. Let A be the adjacency matrix of a random graph drawn from the model $G\left(n, \frac{1}{n}\varphi(X_i, X_j)\right)$ and let A' be the adjacency matrix defined in Theorem 14.

Denote by $\lambda_0' \ge \lambda_1' \ge \ldots$ the eigenvalues of A' and by $v_0', v_1', \ldots$ the corresponding orthonormal eigenvectors. Let $Y = (v_{l_{s+1}}', \ldots, v_{l_{s+d}}')$. By Theorem 13 there exists an $R \in \mathcal{O}(d)$ such that

$$\left\|\left(v_{l_{s+1}}(M_n), \ldots, v_{l_{s+d}}(M_n)\right) - YR\right\|_F \le \frac{2^{3/2} d^{1/2} \|M_n - A'\|_{op}}{\min_{i \ne 1, \ldots, d} |\lambda_i - \lambda(\varphi)|}.$$

https://doi.org/10.1017/S0963548322000098 Published online by Cambridge University Press

Hence by Theorem 14 we have

$$\left\| \left(v_{l_{s+1}}(M_n), \ldots, v_{l_{s+d}}(M_n)\right) - YR \right\|_F \leq C\sqrt{d} \cdot \frac{\sqrt{\|\varphi\|_\infty}}{\min_{i\neq 1,\ldots,d} |\lambda_i - \lambda(\varphi)|},$$

with probability $1 - n^{-1}$. It follows that

$$\left\| G_n - YY^T \right\|_F \leq C\sqrt{d} \cdot \frac{\sqrt{\|\varphi\|_\infty}}{\min_{i\neq 1,\ldots,d} |\lambda_i - \lambda(\varphi)|}$$

Combining this with Theorem 12 yields

$$\left\| G_n' - YY^T \right\|_F \leq C\sqrt{d} \cdot \frac{\sqrt{\|\varphi\|_\infty}}{\min_{i\neq 1,\ldots,d} |\lambda_i - \lambda(\varphi)|},$$

So,

$$\frac{1}{n^2} \sum_{i,j} \left| X_i \cdot X_j - (nYY^T)_{i,j} \right|^2 \leq Cd \frac{\|\varphi\|_\infty}{\min_{i\neq 1,\ldots,d} |\lambda_i - \lambda(\varphi)|^2}$$

which gives the desired reconstruction bound. □

3. Lower bounds

Our approach to proving lower bounds will be to exploit some symmetries which are inherent to well behaved kernel functions. We thus make the following definition:

Definition 15. *(DPS property).* Let μ be a probability measure on $\mathbb{S}^{d-1}$, and let $w \in \mathbb{S}^{d-1}$. We say that μ has the diminishing post-translation symmetry (DPS) around w property with constant ε, if there exists a decomposition,

$$\mu = (1-\varepsilon)\mu_w + \varepsilon \mu_w^s.$$

Here μ_w^s is a probability measure, invariant to reflections with respect to $w^\perp$. In other words, if $R = \mathrm{Id}_d - 2ww^T$, $R_* \mu_w^s = \mu_w^s$. For such a measure we denote $\mu \in \mathrm{DPS}_w(\varepsilon)$.

If instead μ is a measure on $(\mathbb{S}^{d-1})^{|T_k|}$, we say that $\mu \in \mathrm{DPS}_w^k(\varepsilon)$ if a similar decomposition exists but now the reflections should be applied to each coordinate separately.

We now explain the connection between the DPS property and percolation of information in trees. For this, let us recall the random function $g{:}T \to \mathbb{S}^{d-1}$, introduced in Section 1.2, which assigns to the root, r, a uniformly random value and for any other $u \in T$, the label $g(u)$ is distributed according to $\varphi(g(\mathrm{parent}(u)), \cdot)d\sigma$.

Lemma 16. *Suppose that there exist a sequence $(p_k)_k$ with $\lim_{k\to\infty} p_k = 1$ such that for every $w, x_0 \in \mathbb{S}^{d-1}$ and every $k > 0$,*

$$\mathrm{Law}((g(v))_{v\in T_k} | g(r) = x_0) \in \mathrm{DPS}_w^k(p_k).$$

Then there is zero percolation of information along T.

Proof. Denote $X = g(r)$ and $Y = g(v)_{v\in T_k}$ and let $\rho_{X|Y}$ be the density of $X|Y$ with respect to σ. Our aim is to show that $\mathbb{E}_Y\left[\mathrm{TV}(X|Y, \sigma)\right] = o(1)$. We first claim that it is enough to show that for all $x, x' \in \mathbb{S}^{d-1}$ and all $\delta > 0$ one has,

$$\mathbb{P}\left(\frac{\rho_{X|Y}(x)}{\rho_{X|Y}(x')} - 1 \leq \delta \,\middle|\, X = x \right) = 1 - o(1). \tag{16}$$

https://doi.org/10.1017/S0963548322000098 Published online by Cambridge University Press

Indeed, let $H = \left\{ \frac{\rho_{X|Y}(x)}{\rho_{X|Y}(x')} - 1 \le \delta \right\}$. Note that, by Bayes' rule,

$$\frac{\rho_{X|Y}(x)}{\rho_{X|Y}(x')} = \frac{\rho_{Y|X=x}(Y)}{\rho_{Y|X=x'}(Y)}.$$

Let $x \in \mathbb{S}^{d-1}$ be some fixed point and let x' be uniformly distributed is $\mathbb{S}^{d-1}$ and independent from Y. We will use $\mathbb{E}_{x'}$ to denote integration with respect to x', which has density ρ_X. Similarly, $\mathbb{E}_Y$ stands for integration with respect to the density ρ_Y of Y and $\mathbb{E}_{x',Y}$ is with respect to the product $\rho_X \cdot \rho_Y$. The symbol $\mathbb{P}$ will always mean probability over x' and Y.

As a consequence of the assumption in (16), as long as k is large enough, we have,

$$\mathbb{P}\left(H|X=x\right) \ge 1-\delta.$$

For any Borel measurable set A, we denote the projection:

$$\Pi_Y A = \{y \in (\mathbb{S}^{d-1})^{|T_k|} : (\tilde{x}, y) \in A \text{ for some } \tilde{x} \in \mathbb{S}^{d-1}\}.$$

Define now the set

$$H' := \{(\tilde{x}, y) | (\tilde{x}, y) \in H \text{ and } \mathbb{E}_{x'}[\mathbf{1}_H(\cdot, y)] \ge 1 - 4\delta\}.$$

In particular, for every $y \in \Pi_Y H'$,

$$\mathbb{E}_{x'}[\mathbf{1}_H(\cdot, y)] \ge 1 - 4\delta,$$

and by Fubini's theorem, $\mathbb{P}\left(H'|X=x\right) \ge 1 - 4\delta$.

Consider the random variables $\alpha := \alpha(Y) = \frac{\rho_{Y|X=x}(Y)}{\rho_Y(Y)}$ and $\beta := \beta(x', Y) = \frac{\rho_{Y|X=x'}(Y)}{\rho_Y(Y)}$. By definition of H, we have

$$(1-\delta)\mathbf{1}_H \alpha \le \mathbf{1}_H \beta. \tag{17}$$

Moreover, for almost every Y,

$$\mathbb{E}_{x'}\left[\beta\right] = \frac{1}{\rho_Y(Y)} \int \rho_{Y|X=x'}(Y)\rho_X(x') = \frac{1}{\rho_Y(Y)}\rho_Y(Y) = 1.$$

So,

$$(1-\delta)(1-4\delta)\mathbb{E}_Y\left[\alpha \mathbf{1}_{\Pi_Y H'}\right] \le (1-\delta)\mathbb{E}_Y\left[\alpha \mathbb{E}_{x'}\left[\mathbf{1}_{H'}\right]\right] \le \mathbb{E}_{x',Y}\left[\beta \mathbf{1}_{H'}\right] \le \mathbb{E}_{x',Y}\left[\beta\right] = 1.$$

Observe that $\mathbb{E}_Y\left[\alpha \mathbf{1}_{\Pi_Y H'}\right] = \mathbb{P}\left(\Pi_Y H' | X = x\right) = 1 - o(1)$, where the second equality is a consequence of Fubini's theorem. Hence, let us write $h_1(\delta)$, so that

$$1 - h_1(\delta) = (1-\delta)(1-4\delta)\mathbb{E}_Y\left[\alpha \mathbf{1}_{\Pi_Y H'}\right] \le (1-\delta)\mathbb{E}_{x',Y}\left[\alpha \mathbf{1}_{H'}\right].$$

Markov's inequality then implies,

$$\mathbb{P}\left(\beta \mathbf{1}_{H'} \ge (1-\delta)\alpha \mathbf{1}_{H'} + \sqrt{h_1(\delta)}\right) \le \frac{\mathbb{E}_{x',Y}\left[\beta \mathbf{1}_{H'}\right] - (1-\delta)\mathbb{E}_{x',Y}\left[\alpha \mathbf{1}_{H'}\right]}{\sqrt{h_1(\delta)}} \le \sqrt{h_1(\delta)}. \tag{18}$$

Now, we integrate over x' to obtain,

$$(1-\delta)(1-4\delta)\alpha \mathbf{1}_{\Pi_Y H'} \le (1-\delta)\alpha \mathbb{E}_{x'}\left[\mathbf{1}_{H'}\right] \le \mathbb{E}_{x'}\left[\beta \mathbf{1}_{H'}\right] \le 1,$$

which implies

$$\alpha \mathbf{1}_{H'} \le \frac{1}{(1-\delta)(1-4\delta)}. \tag{19}$$

https://doi.org/10.1017/S0963548322000098 Published online by Cambridge University Press

Keeping in mind the previous displays, we may choose $h_2(\delta)$, which satisfies $\lim_{\delta\to 0} h_2(\delta) = 0$, $\alpha \mathbf{1}_{H'} \leq 1 + h_2(\delta)$ and $\mathbb{E}_{x',Y}\left[\alpha \mathbf{1}_{H'}\right] \geq 1 - h_2(\delta)$.

So, an application of the reverse Markov's inequality for bounded and positive random variables shows,

$$\mathbb{P}\left(\alpha \mathbf{1}_{H'} \geq 1 - \sqrt{h_2(\delta)}\right) \geq \frac{\mathbb{E}_{x',Y}\left[\alpha \mathbf{1}_{H'}\right] - (1 - \sqrt{h_2(\delta)})}{1 + h_2(\delta) - (1 - \sqrt{h_2(\delta)})} \geq \frac{1 - h_2(\delta) - (1 - \sqrt{h_2(\delta)})}{1 + h_2(\delta) - (1 - \sqrt{h_2(\delta)})}$$
$$= \frac{\sqrt{h_2(\delta)} - h_2(\delta)}{\sqrt{h_2(\delta)} + h_2(\delta)}. \tag{20}$$

Note that as $\delta \to 0$ the RHS goes to 1. Thus, by combining the above displays, there exists a function h, which satisfies $\lim_{\delta\to 0} h(\delta) = 0$ and some $H'' \subset H'$, with $\mathbb{P}\left(H'\right) \geq 1 - h(\delta)$, such that, by (19) and (20),

$$\mathbf{1}_{H''}|\alpha - 1| \leq h(\delta),$$

which implies, together with (17) and (18)

$$\mathbf{1}_{H''}|\alpha - \beta| \leq h(\delta).$$

This then gives

$$\mathbf{1}_{H''}|1 - \beta| \leq 2h(\delta).$$

We can thus conclude,

$$\begin{aligned}
\mathbb{E}_Y \mathrm{TV}\,(X|Y, X) &= \mathbb{E}_{Y,x'}\left[|\beta - 1|\right] = 2\mathbb{E}_{Y,x'}\left[(1-\beta)\mathbf{1}_{\beta\leq 1}\right] \\
&\leq 2\mathbb{E}_{Y,x'}\left[(1-\beta)\mathbf{1}_{\beta\leq 1}\mathbf{1}_{H''}\right] + 1 - \mathbb{P}\left(H''\right) \\
&= 2\mathbb{E}_{Y,x'}\left[|1-\beta|\mathbf{1}_{H''}\right] + h(\delta) \\
&\leq 5h(\delta).
\end{aligned}$$

Take now $\delta \to 0$ to get $\mathbb{E}_Y \mathrm{TV}\,(X|Y, X) \to 0$.

Thus, we may assume towards a contradiction that there exist $x, x' \in \mathbb{S}^{d-1}$ and a set $F \subset (\mathbb{S}^{d-1})^{|T_k|}$, such that

$$\mathbb{P}\left(Y \in F | X = x\right) \geq \delta, \tag{21}$$

and under $\{Y \in F\}$,

$$\frac{\rho_{X|Y}(x)}{\rho_{X|Y}(x')} \geq 1 + \delta, \tag{22}$$

for some constant $\delta > 0$.

Let $w \in \mathbb{S}^{d-1}$ be such that the reflection $R := \mathrm{Id}_d - 2ww^T$ satisfies $Rx = x'$. Under our assumption, there exists an event A_k, which satisfies

$$\mathbb{P}(A_k | X = x) = 1 - o(1)$$

and such that $Y|(X = x, A_k)$ is R-invariant. By (21), we also have

$$\mathbb{P}(A_k | X = x, Y \in F) = 1 - o(1),$$

and thus there exists $y \in F$ such that

$$\mathbb{P}(A_k | X = x, Y = y) = 1 - o(1). \tag{23}$$

https://doi.org/10.1017/S0963548322000098 Published online by Cambridge University Press

By continuity, we can make sense of conditioning on the zero probability event $E := \{X = x,\ Y \in \{y, Ry\}\}$, in the sense of regular probabilities. Note that we have by symmetry and since $y \in F$,

$$Y = y \ \Rightarrow\ 1 + \delta \leq \frac{\rho_{X|Y}(x)}{\rho_{X|Y}(x')} = \frac{\mathbb{P}(Y = y|E)}{\mathbb{P}(Y = Ry|E)}. \tag{24}$$

On the other hand, we have by definition of A_k,

$$\mathbb{P}(Y = Ry|E, A_k) = \mathbb{P}(Y = y|E, A_k),$$

which implies that

$$\begin{aligned}\mathbb{P}(Y = Ry|E) &\geq \mathbb{P}\big(\{Y = Ry\} \cap A_k|E\big)\\ &= \mathbb{P}\big(\{Y = y\} \cap A_k|E\big)\\ &\geq \mathbb{P}(Y = y|E)(1 - o(1))\end{aligned}$$

which contradicts (24). The proof is complete. □

3.1 The Gaussian case

Our aim is to show that certain classes of kernel functions satisfy the DPS condition. We begin by considering the case where the kernel φ is Gaussian, as in (8). In this case, the function g may be defined as follows. Let T be a q-ary tree. To each edge $e \in E(T)$ we associate a Brownian motion $(B_e(t))_{t\in(0,1)}$ of rate β such that for every node $v \in V$ we have

$$g(v) = \sum_{e\in P(v)} B_e(1) \bmod 2\pi$$

where $P(v)$ denotes the shortest path from the root to v.

For every node $v \in T_k$ let us now consider the Brownian motion $(B_v(t))_{t=0}^k$ defined by concatenating the Brownian motions B_e along the edges $e \in P(v)$. Define by E_v the event that the image of $B_v \bmod \pi$ contains the entire interval $[0, \pi)$, and define $E_k = \bigcap_{v\in T_k} E_v$. Our lower bound relies on the following observation.

Claim 17. *Fix $v \in T_k$ and set $p_k := \mathbb{P}(E_k)$. Then, for every $\theta, x_0 \in \mathbb{S}^1$, $\mathrm{Law}(g(v)|g(r) = x_0) \in \mathrm{DPS}_\theta^k(p_k)$.*

Proof. Fix $\theta \in [0, \pi)$. Given the event E_v, we have almost surely that there exists a time $t_v \leq k$ such that $B_v(t_v) \in \{\theta - \pi, \theta\}$. By symmetry and by the Markov property of Brownian motion, we have that the distribution of $g(v)$ conditioned on the event $\{B_v(t_v) = \theta,\ \forall v\}$ is symmetric around θ. Thus, by considering $(B_v(t))_{v\in T_k}$, under the event $\{t_v \leq k | v \in T_k\}$ we get that for any x_0, $\mathrm{Law}((g(v))_{v\in T_k}|g(r) = x_0)$ is symmetric around θ. So,

$$\mathrm{Law}((g(v))_{v\in T_k}|g(r) = x_0) \in \mathrm{DPS}_\theta^k(p_k).$$

□

We will also need the following bound, shown for example in [8].

Lemma 18. *Let $B(t)$ be a Brownian motion of rate β on the unit circle. Then,*

$$\mathbb{P}(\mathrm{Image}(B(s) \bmod \pi)_{0\leq s\leq t} = [0, \pi)) \geq 1 - Ce^{-t\beta/2}.$$

We are now in a position to prove Theorem 7.

Proof of Theorem 7. Lemma 18 immediately implies that for all $v \in T_k$,

$$\mathbb{P}(E_v) \geq 1 - Ce^{-k\beta/2}.$$

https://doi.org/10.1017/S0963548322000098 Published online by Cambridge University Press

On the other hand, a calculation gives $\lambda(\varphi) = \mathbb{E}[\cos(B_1)] = e^{-\beta/2}$. Thus, applying a union bound implies that for some constant $C > 0$, $\mathbb{P}(E_k) \geq 1 - Cq^k\lambda(\varphi)^k$. Hence, by Claim 17,

$$\mathrm{Law}((g(v))_{v\in T_k}|g(r) = x_0) \in \mathrm{DPS}_w(1 - Cq^k\lambda(\varphi)^k).$$

The result is now a direct consequence of Lemma 16. □

In the next section, we generalise the above ideas and obtain a corresponding bound which holds for distributions other than the Gaussian one.

3.2 The general case

3.2.1 On symmetric functions

We begin this section with simple criterion to determine whether a measure belongs to some DPS class. In the sequel, for $w \in \mathbb{S}^{d-1}$, we denote,

$$H_w^+ = \{x \in \mathbb{S}^{d-1} | \langle x, w\rangle \geq 0\} \text{ and } H_w^- = \{x \in \mathbb{S}^{d-1} | \langle x, w\rangle < 0\}.$$

Lemma 19. *Let $f : [-1, 1] \to \mathbb{R}^+$ satisfy the assumptions of Theorem 9 and let $y \in \mathbb{S}^{d-1}$. If $\mu = f(\langle\cdot, y\rangle)d\sigma$, then for any $w \in \mathbb{S}^{d-1}$, $\mu \in \mathrm{DPS}_w(2 \cdot p_w)$, where $p_w = \min\left(\mu(H_w^+), \mu(H_w^-)\right)$.*

Proof. Without loss of generality let us assume that $y \in H_w^+$. Monotonicity of f implies $\mu(H_w^+) \geq \mu(H_w^-)$. Now, if $R = \mathrm{Id}_d - 2ww^T$ is the reflection matrix with respect to $w^\perp$, then we have for any $x \in H_w^-$,

$$f(\langle x, y\rangle) \leq f(\langle Rx, y\rangle).$$

This follows since $\langle x, y\rangle \leq \langle Rx, y\rangle$.

Let us now define the measure $\tilde{\mu}_w^s$ such that

$$\frac{d\tilde{\mu}_w^s}{d\sigma}(x) = \begin{cases} f(\langle x, y\rangle) & \text{if } x \in H_w^- \\ f(\langle Rx, y\rangle) & \text{if } x \in H_w^+. \end{cases}$$

$\tilde{\mu}_w^s$ is clearly R-invariant and the above observation shows that $\tilde{\mu}_w^s(\mathbb{S}^{d-1}) \leq 1$. We can thus define $\tilde{\mu}_w = \mu - \tilde{\mu}_w^s$. To obtain a decomposition, define $\mu_w^s = \frac{\tilde{\mu}_w^s}{\tilde{\mu}_w^s(\mathbb{S}^{d-1})}$ and $\mu_w = \frac{\tilde{\mu}_w}{\tilde{\mu}_w(\mathbb{S}^{d-1})}$, for which

$$\mu = (1 - \tilde{\mu}_w^s(\mathbb{S}^{d-1}))\mu_w + \tilde{\mu}_w^s(\mathbb{S}^{d-1})\mu_w^s.$$

The proof is concluded by noting $\tilde{\mu}_w^s(\mathbb{S}^{d-1}) = 2\mu(H_w^-)$. □

Our main object of interest will be the measure $\mu_\varphi(y)$ which, for a fixed y, is defined by

$$\mu_\varphi(y) := \varphi(x, y)d\sigma(x) = f(\langle x, y\rangle)d\sigma(x). \tag{25}$$

Let us now denote,

$$\beta_d(t) := \frac{\Gamma(\frac{d}{2})}{\Gamma(\frac{d-1}{2})}(1 - t^2)^{(d-3)/2},$$

which is the marginal of the uniform distribution on the sphere. We now show that the spectral gap of φ may determine the DPS properties of $\mu_\varphi(y)$.

Lemma 20. *Let $w \in \mathbb{S}^{d-1}$ and suppose that f is monotone. If $|\langle w, y\rangle| \leq \frac{1-\lambda(\varphi)}{16\sqrt{d}}$, then*

$$\mu_\varphi(y) \in \mathrm{DPS}_w\left(\frac{1 - \lambda(\varphi)}{35}\right).$$

https://doi.org/10.1017/S0963548322000098 Published online by Cambridge University Press

Proof. Assume W.L.O.G. that $\langle y, w\rangle > 0$. By Lemma 19 it will be enough to bound $\int_{H_w^-} \mu_\varphi(y)$ from below. Let $X \sim \mu_\varphi(y)$ and define $Z = \langle X, y\rangle$. We have $\mathbb{E}[Z] = \lambda(\varphi)$ and by Markov's inequality,

$$\mathbb{P}\left(Z \le \frac{1+\lambda(\varphi)}{2}\right) = \mathbb{P}\left(Z+1 \le \frac{1+\lambda(\varphi)}{2}+1\right) \ge 1 - \frac{2(\lambda(\varphi)+1)}{3+\lambda(\varphi)} \ge \frac{1-\lambda(\varphi)}{4}.$$

For $t \in [-1, 1]$, set $S_t = \{x \in \mathbb{S}^{d-1} | \langle x, y\rangle = t\}$ and let $\mathcal{H}^{d-2}$ stand for $d-2$-dimensional Hausdorff measure. To bound $\int_{H_w^-} \mu_\varphi(y)$ we would first like to estimate $\frac{\mathcal{H}^{d-2}(S_t \cap H_w^-)}{\mathcal{H}^{d-2}(S_t)}$.

We know that $0 \le \langle w, y\rangle \le \frac{1-\lambda(\varphi)}{16\sqrt{d}}$. Denote $t_y := \langle w, y\rangle$ and fix $t \le t_0 := \frac{1+\lambda(\varphi)}{2}$. With no loss of generality, let us write $w = e_1$ and $y = t_y e_1 + \sqrt{1-t_y^2} e_2$. Define now $z = -\sqrt{1-t_y^2} e_1 + t_y e_2$. We claim that

$$\left\{ v \in S_t \middle| \frac{\langle v, z\rangle}{\sqrt{1-t^2}} \ge \frac{1}{2\sqrt{d}} \right\} \subseteq S_t \cap H_w^-. \tag{26}$$

If $v \in S_t$, its projection onto the plane span(y, w), can be written as $t \cdot y + \sqrt{1-t^2} c \cdot z$, for some $c \in [-1, 1]$. So,

$$\langle v, w\rangle = t \cdot t_y - \sqrt{1-t^2}\sqrt{1-t_y^2}\, c.$$

Now, whenever

$$c > \frac{t \cdot t_y}{\sqrt{1-t^2}\sqrt{1-t_y^2}},$$

we get $\langle v, w\rangle < 0$. Also,

$$\frac{t \cdot t_y}{\sqrt{1-t^2}\sqrt{1-t_y^2}} \le \frac{1}{\sqrt{3}} \frac{t_y}{\sqrt{1-t_y^2}} \le \frac{1}{2\sqrt{d}},$$

where we have used $t \le \frac{1}{2}$ for the first inequality and $t_y \le \frac{1}{2\sqrt{d}}$ in the second inequality. By combining the above displays with $\frac{\langle v,z\rangle}{\sqrt{1-t^2}} = c$, (26) is established. Thus, by taking the marginal of S_t in the direction of $-z$, we see

$$\frac{\mathcal{H}^{d-2}(S_t \cap H_w^-)}{\mathcal{H}^{d-2}(S_t)} \ge \int_{-1}^{-\frac{1}{2\sqrt{d}}} \beta_{d-1}(s)ds \ge \int_{-\frac{1}{\sqrt{d}}}^{-\frac{1}{2\sqrt{d}}} \beta_{d-1}(s)ds \ge \frac{1}{2\sqrt{d}} \beta_{d-1}\left(\frac{1}{\sqrt{d}}\right)$$

$$\ge \frac{1}{2\sqrt{d}} \frac{\Gamma(\frac{d-1}{2})}{\Gamma(\frac{d-2}{2})} \left(1 - \frac{1}{d}\right)^{(d-4)/2} \ge \frac{1}{10\sqrt{e}},$$

https://doi.org/10.1017/S0963548322000098 Published online by Cambridge University Press

where we used $\frac{\Gamma(\frac{d-1}{2})}{\Gamma(\frac{d-2}{2})} \geq \frac{\sqrt{d}}{5}$, valid for any $d \geq 3$. We use the above estimates with Fubini's theorem to obtain:

$$\mathbb{P}\left(X \in H_w^-\right) = \int_{-1}^{1} f(t)\mathcal{H}^{d-2}(S_t \cap H_w^-)dt \geq \int_{-1}^{t_0} f(t)\mathcal{H}^{d-2}(S_t \cap H_w^-)dt$$

$$\geq \frac{1}{10\sqrt{e}} \int_{-1}^{t_0} f(t)\mathcal{H}^{d-2}(S_t)dt = \frac{1}{10\sqrt{e}}\mathbb{P}\left(Z \leq \frac{1+\lambda(\varphi)}{2}\right) \geq \frac{1-\lambda(\varphi)}{70}.$$

□

3.2.2 Mixing

Recall the random function $g{:}T \to \mathbb{S}^{d-1}$, introduced in Section 1.2, which assigns to the root, r, a uniformly random value and for any other $u \in T$, the label $g(u)$ is distributed according to $\varphi(g(\text{parent}(u)), \cdot)d\sigma =: \mu_\varphi(\text{parent}(u))$. Suppose that $v \in T_k$ and let $\{v_i\}_{i=0}^k$ denote the simple path from r to v in T. Fix $x_0 \in \mathbb{S}^{d-1}$, for $i = 0, \ldots, k$, we now regard,

$$X_i := g(v_i)|g(r) = x_0,$$

as a random walk on $\mathbb{S}^{d-1}$. Observe that given X_{i-1}, $X_i \sim \mu_\varphi(X_{i-1})$. The following lemma shows that this random walk is rapidly mixing.

Lemma 21. *For $w \in \mathbb{S}^{d-1}$, let*

$$S(w) = \left\{u \in \mathbb{S}^{d-1} : |\langle u, w\rangle| \leq \frac{1-\lambda(\varphi)}{16\sqrt{d}}\right\},$$

and set $k_0 = \frac{\ln\left(\frac{\lambda(\varphi)(1-\lambda(\varphi))}{32f(1)}\right)}{\ln(\lambda(\varphi))}$. Then, if f satisfies the assumptions of Theorem 9,

$$\mathbb{P}(X_{k_0} \in S(w)) \geq \frac{1-\lambda(\varphi)}{32}.$$

Proof. Note that if $U \sim \text{Uniform}(\mathbb{S}^{d-1})$, then

$$\mathbb{P}(U \in S(w)) = \int_{|t| \leq \frac{1-\lambda(\varphi)}{16\sqrt{d}}} \beta_d(t)dt \geq 2\frac{\Gamma(\frac{d}{2})}{\Gamma\left(\frac{d-1}{2}\right)}\beta_d\left(\frac{1}{16\sqrt{d}}\right)\frac{1-\lambda(\varphi)}{16\sqrt{d}} \geq \frac{1-\lambda(\varphi)}{16}. \tag{27}$$

It will then suffice to show that $\mathbb{P}(X_{k_0} \in S(w))$ can be well approximated by $\mathbb{P}(U \in S(w))$. Since X_{k_0} has density $A_\varphi^{k_0-1}f(\langle x, x_0\rangle)$, the following holds true,

$$\left(\mathbb{P}(U \in S(w)) - \mathbb{P}(X_{k_0} \in S(w))\right)^2 = \left(\int_{S(w)} \left(A_\varphi^{k_0-1}f(\langle x, x_0\rangle) - 1\right) d\sigma(x)\right)^2$$

$$\leq \int_{S(w)} \left(A_\varphi^{k_0-1}f(\langle x, x_0\rangle) - 1\right)^2 d\sigma(x).$$

https://doi.org/10.1017/S0963548322000098 Published online by Cambridge University Press

We now decompose the density as $f(\langle x, x_0\rangle) = \sum_{i=0}^{\infty} \lambda_i \psi_i(x)\psi_i(x_0)$. So that

$$A_\varphi^{k_0-1} f(\langle x, x_0\rangle) = \sum_{i=0}^{\infty} \lambda_i^{k_0} \psi_i(x)\psi_i(x_0).$$

We know that $\psi_0 \equiv 1$ with eigenvalue $\lambda_0 = 1$, and we've assumed that $|\lambda_i| \le \lambda_1 = \lambda(\varphi)$ for every $i \ge 1$. Thus,

$$\begin{aligned}\int_{S(w)} \left(A_\varphi^{k_0-1} f(\langle x, x_0\rangle) - 1\right)^2 d\sigma(x) &= \int_{S(w)} \left(\sum_{i=1}^{\infty} \lambda_i^{k_0} \psi_i(x)\psi_i(x_0)\right)^2 d\sigma(x) \\ &\le (\lambda(\varphi))^{2k_0-2} \int_{S(w)} \sum_{i=1}^{\infty} (\lambda_i \psi_i(x))^2 \sum_{i=1}^{\infty} (\lambda_i \psi_i(x_0))^2 d\sigma(x) \\ &\le \lambda(\varphi)^{2k_0-2} f(1)^2.\end{aligned}$$

where in the last inequality we have used $f(1) = \sum \lambda_i \psi_i(y)\psi_i(y)$, which is valid for any $y \in \mathbb{S}^{d-1}$. Thus, since $k_0 = \frac{\ln\left(\frac{\lambda(\varphi)(1-\lambda(\varphi))}{32 f(1)}\right)}{\ln(\lambda(\varphi))}$, by (27), we get,

$$\mathbb{P}(X_{k_0} \in S(w)) \ge \mathbb{P}(U \in S(w)) - \frac{1-\lambda(\varphi)}{32} \ge \frac{1-\lambda(\varphi)}{32}.$$

□

Since the random walk X_k mixes well, we now use Lemma 20 to show that after enough steps, X_k will be approximately invariant to a given reflection.

Lemma 22. *Let $w, x_0 \in \mathbb{S}^{d-1}$. Then,*

$$\mathrm{Law}((g(v))_{v \in T_k} | g(r) = x_0) \in \mathrm{DPS}_w^k \left(1 - q^k p^k\right),$$

where $p = \left(1 - \frac{\ln(\lambda(\varphi))}{\ln\left(\frac{\lambda(\varphi)(1-\lambda(\varphi))}{32 f(1)}\right)} \frac{(1-\lambda(\varphi))^2}{600}\right)$.

Proof. Let $R = \mathrm{Id}_d - 2ww^T$ denote the linear reflection with respect to $w^\perp$. Then, the claim is equivalent to the decomposition,

$$X_k = P\tilde{X}_k + (1-P)X_k^R,$$

where X_k^R is invariant to reflections by R and $P \sim \mathrm{Bernoulli}(s_k)$ is independent from $\{\tilde{X}_k, X_k^R\}$ with $s_k \le \left(1 - \frac{\ln(\lambda(\varphi))}{\ln\left(\frac{\lambda(\varphi)(1-\lambda(\varphi))}{32 f(1)}\right)} \frac{(1-\lambda(\varphi))^2}{600}\right)^k$.

Consider the case that for some $i = 0, \ldots, k$, $|\langle X_i, w\rangle| \le \frac{1-\lambda(\varphi)}{16\sqrt{d}}$. In this case, from Lemma 20, given X_i, we have the decomposition,

$$\mu_\varphi(X_i) = \left(1 - \frac{(1-\lambda(\varphi))}{35}\right) \mu_{\varphi,w} + \frac{(1-\lambda(\varphi))}{35} \mu_{\varphi,w}^s(X_i),$$

where $\mu_{\varphi,w}^s(X_i)$ is R-invariant.

https://doi.org/10.1017/S0963548322000098 Published online by Cambridge University Press

We now generate the random walk in the following way. For $i = 0, \ldots, k$, let

$$P_i \sim \text{Bernoulli}\left(\frac{(1-\lambda(\varphi))}{35}\right), \tag{28}$$

be an *i.i.d* sequence. Given X_i, if $|\langle X_i, w\rangle| > \frac{1-\lambda(\varphi)}{16\sqrt{d}}$ then $X_{i+1} \sim \mu_\varphi(X_i)$. Otherwise, $|\langle X_i, w\rangle| \leq \frac{1-\lambda(\varphi)}{16\sqrt{d}}$. To decide on the position of X_{i+1} we consider P_i. If $P_i = 0$ then $X_{i+1} \sim \mu_{\varphi,w}(X_i)$. If $P_i = 1$, we generate $X_{i+1} \sim \mu^s_{\varphi,w}(X_i)$. We denote the latter event by A_i and $A = \cup_{i=0}^{k-1} A_i$.

It is clear that, conditional on A, $RX_k \overset{law}{=} X_k$. Thus, to finish the proof, if $\bar{A}$ is the complement of A, we will need to show

$$\mathbb{P}(\bar{A}) \leq \left(1 - \frac{\ln(\lambda(\varphi))}{\ln\left(\frac{\lambda(\varphi)(1-\lambda(\varphi))}{32f(1)}\right)} \frac{(1-\lambda(\varphi))^2}{600}\right)^k.$$

Towards this, let $S(w)$ and k_0 be as in Lemma 21. Coupled with (28), the lemma tells us that

$$\mathbb{P}\left(A_{k_0}\right) \geq \frac{(1-\lambda(\varphi))^2}{600}.$$

Now, by restarting the random walk from X_{k_0} if needed, we may show,

$$\mathbb{P}\left(\bar{A}\right) \leq \sum_{m \leq \frac{k}{k_0}} \mathbb{P}(\bar{A}_{m \cdot k_0}) \leq \left(1 - \frac{(1-\lambda(\varphi))^2}{600}\right)^{\frac{k}{k_0}} \leq \left(1 - \frac{(1-\lambda(\varphi))^2}{600 k_0}\right)^k.$$

The claim now follows by taking a union bound over all paths. Indeed, for each $v \in T_k$ define the corresponding event A_v. Then, with a union bound,

$$\mathbb{P}\left(\bigcup_{v \in T_k} \bar{A}_v\right) \leq |T_k| p^k \leq q^k p^k.$$

By definition, conditioned on $\bigcap_{v \in T_k} A_v$, $\{X_v\}_{v \in T_k}$ is symmetric around w, which completes the lemma. □

3.2.3 Proving Theorem 9

Proof. By Lemma 22, for every $w, x_0 \in \mathbb{S}^{d-1}$.

$$\text{law}(g(v)_{v \in T_k} | g(r) = x_0) \in \text{DPS}^k_w(1 - q^k p^k),$$

where $p = \left(1 - \frac{\ln(\lambda(\varphi))}{\ln\left(\frac{\lambda(\varphi)(1-\lambda(\varphi))}{32f(1)}\right)} \frac{(1-\lambda(\varphi))^2}{600}\right)$. By assumption

$$q \leq p^{-1},$$

and Lemma 16 gives the result. □

Acknowledgments

We are grateful to two anonymous referees for carefully reading the paper and providing many insightful comments.

References

[1] Abbe, E. (2018) Community detection and stochastic block models. *Found. Trends® Commun. Inform. Theory* **14**(1-2) 1–162.

[2] Valdivia, E. A and De Castro, Y. (2019) Latent distance estimation for random geometric graphs, *Advances in Neural Information Processing Systems 32*. Curran Associates, Inc, 8724–8734.

[3] Bhatia, R. (1997) *Matrix Analysis*. Springer-Verlag, New York.

[4] Braun, M. L. (Dec. 2006) Accurate error bounds for the eigenvalues of the kernel matrix. *J. Mach. Learn. Res.* **7** 2303–2328.

[5] Bubeck, S., Ding, J., Eldan, R. and Rácz, M. Z. (2016) Testing for high-dimensional geometry in random graphs. *Random Struct. Algorithms* **49**(3) 503–532.

[6] De Castro, Y., Lacour, C. and Ngoc, T. M. P. (2017) Adaptive estimation of nonparametric geometric graphs, arXiv preprint arXiv: 1708. 02107.

[7] Deshpande, Y., Sen, S., Montanari, A. and Mossel, E. (2018) Contextual stochastic block models, *Advances in Neural Information Processing Systems 31*, Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N. and Garnett, R., Curran Associates, Inc, 8581–8593.

[8] Ernst, P. and Shepp, L. (2016) On the time for Brownian motion to visit every point on a circle. *J. Statist. Plann. Inference* **171** 130–134.

[9] Galhotra, S., Mazumdar, A., Pal, S. and Saha, B.. The geometric block model. arXiv preprint arXiv: 1709. 05510, 2017.

[10] Hirsch, F. and Lacombe, G. (2012) *Elements of functional analysis*. Springer Science & Business Media, **192**

[11] Le, C. M., Levina, E. and Vershynin, R. (2017) Concentration and regularization of random graphs. *Random Struct. Algorithms* **51**(3) 538–561.

[12] Le, C. M., Levina, E. and Vershynin, R. (2018) Concentration of random graphs and application to community detection. *Proc. Int. Cong. Math.* **3** 2913–2928.

[13] Mossel, E. (2001) Reconstruction on trees: beating the second eigenvalue. *Ann. Appl. Probab.* **11**(1) 285–300.

[14] Mossel, E. (2004) *Survey: Information Flow on Trees*. American Mathematical Society.

[15] Mossel, E. and Peres, Y. (08 2003) Information flow on trees. *Ann. Appl. Probab.* **13**(3) 817–844.

[16] Mossel, E., Roch, S. and Sly, A. (2013) Robust estimation of latent tree graphical models: inferring hidden states with inexact parameters. *IEEE Trans. Inform. Theory* **59**(7) 4357–4373.

[17] Sussman, D. L., Tang, M. and Priebe, C. E. (Jan 2014) Consistent latent position estimation and vertex classification for random dot product graphs. *IEEE Trans. Pattern Anal. Mach. Intell.* **36**(1) 48–57.

[18] Tang, M., Sussman, D. L. and Priebe, C. E. (2013) Universally consistent vertex classification for latent positions graphs. *Ann. Stat.* **41**(3) 1406–1430.

[19] Tropp, J. (2012) User-friendly tail bounds for sums of random matrices. *Found. Comput. Math.* **12**(4) 389–434.

[20] Valdivia, E. A. (2018) Relative concentration bounds for the kernel matrix spectrum, arXiv preprint arXiv: 1812. 02108.

[21] Yu, Y., Wang, T. and Samworth, R. J. (2014) A useful variant of the Davis–Kahan theorem for statisticians. *Biometrika* **102**(2) 315–323.

Cite this article: Eldan R, Mikulincer D, and Pieters H (2022). Community detection and percolation of information in a geometric setting. *Combinatorics, Probability and Computing* **31**, 1048–1069. https://doi.org/10.1017/S0963548322000098

https://doi.org/10.1017/S0963548322000098 Published online by Cambridge University Press