Geometry-Inspired Sampling Algorithms and Random Graphs

Elizabeth Yang

Electrical Engineering and Computer Sciences
University of California, Berkeley

Technical Report No. UCB/EECS-2023-87
http://www2.eecs.berkeley.edu/Pubs/TechRpts/2023/EECS-2023-87.html

May 10, 2023

Geometry-Inspired Sampling Algorithms and Random Graphs

by

Elizabeth Yang

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Computer Science

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Satish Rao, Chair
Professor Alistair Sinclair
Professor Nikhil Srivastava

Summer 2023

The dissertation of Elizabeth Yang, titled Geometry-Inspired Sampling Algorithms and Random Graphs, is approved:

Chair _____ Date _____

_____ Date _____

_____ Date _____

University of California, Berkeley

Geometry-Inspired Sampling Algorithms and Random Graphs

Abstract

Geometry-Inspired Sampling Algorithms and Random Graphs

by

Elizabeth Yang

Doctor of Philosophy in Computer Science

University of California, Berkeley

Professor Satish Rao, Chair

High-dimensional expansion, a generalization of graph expansion to higher-order edges, has recently garnered significant attention in the theoretical computer science community for the additional boost they give in applications like error-correction and approximate sampling. In this thesis, we explore two problems related to high-dimensional expansion, using tools from the geometry of polynomials as well as high-dimensional convex geometry.

First, we study approximate sampling from discrete distributions. The framework for sampling obtained from high-dimensional expansion provides both a natural set of random walks to use in MCMC algorithms, as well as a set of tools for their analysis. We show that the geometric properties (e.g. log-concavity) of a polynomial derived from the distribution allows us to speed up the implementations of these random walks.

Next, we study a random graph model called the "random geometric graph," with an eventual goal of understanding its modeling capabilities as well as its high-dimensional expansion properties. Along the way, we prove new results about distinguishing the random geometric graph model from the Erdös-Rényi model, and develop a new geometric toolkit for analyzing these graphs.

To my parents, for their unending support and many sacrifices,
and to Hansen, for being by my side as I chased a dream.

Contents

List of Figures

List of Tables

Acknowledgments

This document would not exist without the wisdom, guidance, and encouragement of many, many amazing people.

First and foremost, I want to thank my advisor, Satish Rao. Satish's door has, literally and figuratively, always been open for me. When I was a newcomer to theoretical computer science, Satish patiently helped me find my footing in the field. He always encouraged me to ask a lot of questions, and fostered a welcoming learning environment. Discussions with Satish are master classes in how to distill highly technical ideas down to their most intuitive core. And not only is Satish a great person to talk math with, he is also kind and supportive on a personal level. He has consistently expressed his confidence in my ability, and for that I'll always be grateful.

I'd also like to thank my thesis and quals committee: Alistair Sinclair, Nikhil Srivastava, and Prasad Raghavendra. I'd like to further thank Alistair for taking time during a hectic Spring 2019 to guide me through several HDX and MCMC papers. The earliest traces of this thesis came from our meetings that semester. I'd also like to thank Nikhil for welcoming me to his weekly reading group, where I found a lot of research inspiration and a wonderful community. Lastly, I am grateful to Umesh Vazirani for his support during my first year as my temporary advisor.

One of the greatest joys of my PhD was the chance learn from fantastic collaborators. Thank you to Nima Anari, Michał Dereziński, and June Vuong, for working with me during the pandemic on the project featured in Chapter 3, and to Siqi Liu, Sidhanth Mohanty, and Tselil Schramm, for a longtime collaboration that culminated in Chapters 5 and 6. Siqi and Sidhanth are two people who I had the great fortune to spend five years with at Berkeley. They are brilliant researchers, and even kinder friends. Some of my favorite research memories are the long afternoons we spent in Simons, drawing all over the whiteboards and doing the "OERS." I'd also like to thank Tselil for being an amazing research mentor, writing mentor, and even climbing mentor, and for the energy she brought to our discussions. I lastly thank June for our food outings, and for teaching me about the geometry of polynomials.

My internship experiences during my final years were very eye-opening and, more importantly, fun. For those opportunities, I am indebted to Ken Clarkson, Lior Horesh, and Shashanka Ubaru from IBM, as well as James Cook and Nina Mishra from AWS. Ken mentored me for two whole summers, and taught me everything I know about sketching. I am grateful to Ken for including me on so many team meetings, and for interesting technical discussion, often during walks around the Almaden office. I further thank James and Nina for an exciting chance to work on differential privacy. Though I met them at the end of my PhD, working with them has made tremendous impact on my grad school experience.

I owe much of my happiness at Berkeley to the theory group. Thanks to Ishaq Aden-Ali, Omar Alrabiah, James Bartusek, Zoë Bell, Thiago Bergamaschi, Adam Bouland, Jonah Brown-Cohen, Sarah Cannon, Ben Caulfield, Lynn Chua, Grace Dinh, Louis Golowich, Sam Gunn, Meghal Gupta, Sam Hopkins, Chris Ikeokwu, Fotis Iliopoulos, Malvika Joshi, Zeph Landau, Jingcheng Liu, Yunchao Liu, Xin Lyu, Urmila Mahadev, Pasin Manurangsi, Jarrod Millman, Peihan Miao, Ansh Nagda, Orr Paradise, Angelos Pelecanos, Vamsi Policharla, Bhaskar Roberts, Manuel Sabin, Aaron Schild, Jonathan Shafer, Abhishek Shetty, Akshayaram Srinivasan, Francisca Vasconcelos, Di Wang, David Wu, Kewen Wu, Hongxun Wu, Morris Yau, Henry Yuen, Fred Zhang, and Richard Zhang for making Soda so welcoming.

Additional thanks to those I have spent the most time with at Berkeley: Efe Aras, for the Mind Coffee and the constant laughs; Arun Ganesh, for karaoke, Survivor and relating to me on so many levels; Nathan Ju, for the SF carpool, and also for Survivor; Tarun Kathuria, for help and excellent K-drama recs during the trying times; Seri Khoury, for amazing home-cooked meals and for his generosity; Rachel Lawrence, for marina runs, and solidarity over courses and teaching; my incredible collaborators Siqi Liu and Sidhanth Mohanty; Chinmay Nirkhe, for the ice cream, office banter, and steadfast friendship, especially during the tough pandemic years; and Nick Spooner, for the shows at the Greek.

I would not be nearly as sane without support from other friends going through grad school at the same time: Jason Altschuler, Udit Gupta, Ghassen Jerfel, Swati Padmanabhan, Kiran Vodrahalli, and Angela Zhou. They consistently offered me good advice and encouragement, especially when I needed it most.

I am also grateful to those who made the Bay feel like home: Alana Bhatla, Andreea Bobu, Leah and Max Burkhardt, Kevin Chen, Reia Cho, Abrar Choudhury, Barak Gila, Chloe Hsu, David Hsu, Winter Ku, Harrison Liew, Young Lu, Joanna Pang, Stacey Park, Arjun Seshadri, Paul von Autenried, Serena Wang, Maggie White, Irvin Zhan, Sinan Zhang, Roy Zhao, and Quan Zhou. Lastly, I thank those who cheered me on from afar: Luyu Cheng, Priyanka Goyal, Sahana Jayaraman, Stacey Jou, Sushrut Karmalkar, Pallavi Koppol, Tori Lin, Kathryn Long, Sean Pan, Lucy Pei, Andrew Tsukamoto, Vicky Wang, Emily Yang, Ben Zhang, Charles Zhou, and Ben Zhu.

My partner, Hansen Qian, has been my rock for the length of two PhDs at this point. During my time at Berkeley, Hansen made sure I didn't stress too much and always reminded me to appreciate the small joys in life, like a tasty meal. More so than anyone else, he has witnessed the good, the bad, and the ugly of the past six years. He's comforted me through the rough patches, and celebrated all of my victories, no matter how small.

And last but certainly not least, I owe every opportunity I was ever given to my parents. This PhD was probably just as hard for them as it was for me, with our family split between opposite coasts. Words cannot fully describe how grateful I am for everything they've done for me during the last six years. Thank you, Mom and Dad.

Chapter 1

Introduction

But on a Wednesday in a cafe
I watched it begin again [Swi12]

An exciting trend in theoretical computer science is the incorporation of more and more tools from continuous mathematics to design algorithms and analyze combinatorial objects.

For instance, continuous optimization has inspired many breakthroughs in discrete optimization. One of the earliest such examples is the Goemans-Williamson approximation algorithm for max cut, based on semidefinite programming [GW95]. Linear programming relaxations also yield competitive and interpretable solutions to many combinatorial problems like job scheduling, Steiner tree, and facility location [WS11]. The recently discovered state-of-the-art algorithms for several classical graph problems, like max-flow and bipartite matching, also rely on interior point methods [Che+22; Bra+20].

If we look beyond continuous optimization, and look at continuous mathematics in general, we see even more examples. For one, Brownian motion is the basis of many rounding algorithms designed for discrepancy minimization [Abb+22]. The entire subfield of Boolean Fourier analysis is motivated by standard Fourier analysis for continuous functions [ODo14]. Barvinok's polynomial method for approximate counting is related to some of the geometric ideas surveyed in this thesis. In the field of approximate counting, Barvinok's method uses Taylor series expansion used to approximate partition functions such as the matching polynomial or the independence polynomial [Bar16]. Its analysis relies on studying the locations of the zeros of these polynomials.

In this thesis, we use geometric tools to tackle two problems related to "high-dimensional expansion." High-dimensional expanders have emerged in recent years as a promising construction for obtaining improvements in error correction, derandomization, approximate sampling, and beyond.

1.1 Why high-dimensional expansion?

Expander graphs (expanders, for short) are sparse graphs with good connectivity properties. "Connectivity" is usually quantified in two ways:

- **Combinatorial:** No matter how we split the vertices into two groups, the number of edges crossing the groups is large. This is also known as having high "conductance."
- **Spectral:** When we take a random walk on the graph, we converge quickly to the stationary distribution of the walk. This is equivalent to having a large gap between the top two eigenvalues of the (normalized) adjacency matrix of the graph.

These two notions of connectivity are very closely related, via Cheeger's inequality [Chu07]. For the remainder of this work, "expansion" will refer to *spectral* expansion.

At a first glance, it is not clear that expander graphs should even exist, let alone be easy to construct. Both combinatorial and spectral expansion encourage the graph to have as many edges as possible, while sparsity (i.e. constant degree) forces an upper bound on the number of edges in the graph. However, against these odds, we have discovered numerous expander constructions ranging from the algebraic [LPS88a] to the combinatorial [RVW00] to the probabilistic [Fri03]); expanders are now considered "commonplace" objects. For instance, Friedman's Theorem [Fri03] says that for constant $d \geq 3$, any d-regular graph is expanding with constant probability.

Because they simultaneously hold these conflicting, yet desirable properties, expanders unsurprisingly form the backbone of numerous advances in theoretical computer science. They have led to discoveries in error-correcting codes [SS96], pseudorandomness [INW94], and probabilistically checkable proofs [Din07], just to name a few examples. The expander "application area" that we will focus on for this thesis is the analysis of Markov chain Monte Carlo (MCMC) algorithms [Jer03].

MCMC refers to a family of algorithms for approximate sampling from distributions, whose supports are often exponential in the problem parameter, so techniques like rejection sampling are inefficient. Such algorithms typically consist of a low-degree Markov chain that mixes to the intended target distribution. In order for the algorithms to be efficient, we require *rapid mixing* of the Markov chain. Many techniques used to analyze MCMC algorithms, such as canonical paths [JS88] and Markov chain decomposition theorems [JSTV04], are directly developed from the study of both combinatorial and spectral expansion in graphs. In fact, any Markov chain's transition matrix can be viewed as the normalized adjacency matrix of a weighted graph, so if we can show that the graph is spectrally expanding, we obtain mixing time bounds for the Markov chain.

In a sense, MCMC is not a directly "application" of an expander graph, but more so an area that is deeply intertwined with the theory of expander graphs.

1.1.1 The HDX advantage

A slew of recent breakthroughs [ALGV19; DELLM22; PK22] suggest that substituting *high-dimensional expanders* (HDXes) in place of expander graphs gives us a "boost" in several application areas. There are a few different definitions of high-dimensional expansion, each of which generalize combinatorial and spectral expansion in different ways. In this thesis, we focus on high-dimensional spectral expansion, which is also referred to as *local spectral expansion*, and when we use the term "HDX," it will refer to a local spectral expander. For sake of this overview, we can think of HDXes as graphs that not only expand themselves, but whose "neighborhoods" also expand; the definition of neighborhood depends on the HDX dimension. We will precisely define everything in Chapter 2.

Graph expansion is a special case of high-dimensional expansion, for dimension 1. High-dimensional expansion (for dimension ≥ 1) is a strictly stronger condition than expansion, and an even more challenging one to satisfy from a mathematical standpoint due to the tension between connectivity and sparsity. Most expander graphs are not high-dimensional expanders; in fact, a random k-regular graph is 1-dimensionally expanding, but it is locally tree-like, so the neighborhoods in the graph are sets of isolated vertices. The neighborhoods are not even connected, let alone expanding, so random k-regular graphs are not high-dimensionally expanding.

The current known "true" HDX constructions (which have constant degree, rather than say, sub-linear degree) are primarily algebraic; there are a few combinatorial constructions for 2-dimensional expanders. The discovery of expander graphs has followed a similar trajectory, with the algebraic constructions [LPS88a] preceding the the combinatorial ones [RVW00] and probabilistic ones [Fri03]. It is still not known whether HDXes are as "commonplace"– whether their construction is fundamentally bottlenecked beyond the algebraic constructions, or if we will find additional combinatorial and probabilistid HDX constructions with time.

Another property of HDXes that set them apart from ordinary expander graphs is the *local-to-global* property. It is this local-to-global property that provides the "boost" over expander graphs in their various application areas. For instance, using expander graphs, we could get error-correcting codes with constant distance and constant rate. However, an HDX-like object helps us construct "c^3 codes," which satisy constant distance, constant rate, and constant *locality*. Constant locality means that we can query a constant number of bits in a string and decide whether or not it is a codeword with non-negligible probability. The local-to-global property is critical for achieving constant locality, and standard expander codes do not satisfy constant locality.

The local-to-global property also manifests in improvements to the design and (mainly) analysis of many MCMC algorithms. At a high level, the local-to-global property tells us that if we can encode a target distribution into an HDX, we obtain a natural Markov chain that we call a *higher-order random walk* for this HDX. We can then compute spectral gap and mixing time bounds for the higher-order random walk, which is a *global* object, solely from analyzing the *local* 1-dimensional expansion of the HDX's neighborhoods, which are often much simpler objects.

As a result, we obtain a unified framework for sampling from several interesting distributions where we reduce the analysis of a large Markov chain to analyzing several smaller, simpler chains. This framework has been successfully applied to several classic Markov chains, such as the basis exchange walk for sampling a uniform random spanning tree of a graph, or the Glauber dynamics for sampling independent sets in the graph.

1.2 Thesis overview

As mentioned before, there has been growing excitement and interest in incorporating tools from continuous mathematics and geometry to studying discrete objects. We now ask: would the study of expansion and high-dimensional expansion of graphs also benefit from additional tools in continuous mathematics?

The answer is already yes, at least for the advances in MCMC for discrete distributions using the HDX framework. Many recent results rely on deep connections between local spectral expansion in HDXes and the *geometry of polynomials*, the study of particular multivariate polynomials and properties like the locations of their zeros and their log-concavity. We elucidate this connection further in Section 2.4. Chapters 3 describes how to use a particular geometric property of a polynomial that encodes a discrete distribution to obtain speedups in approximate sampling from the distribution.

We also mentioned that currently, constructions of HDXes are lacking, in comparison to the plentiful constructions of expander graphs. Chapters 5, and 6 are motivated by whether a fundamentally geometric construction of a graph can yield high-dimensional expanders. Geometric graph constructions are also interesting as potential models for real-world networks. A big contribution of this work (primarily in Chapter 5, is a computer scientist's toolkit for the analyzing geometric graphs.

1.2.1 Chapter 3: Speedups of higher-order random walks

In Chapter 3 (with preliminaries in Chapter 2), we study the problem of approximately sampling from discrete distributions μ over size-k subsets of some domain $[n]$. There are many examples of distributions with such support, including the uniform distribution of spanning trees on a graph, k-determinantal point processes, and distributions that help count k-edge matchings in graphs. The prevailing algorithm used for this task is MCMC; in particular, we use the higher-order random walks of an HDX that encodes a such distribution $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$.

We can quantify how "expanding" this HDX is using a measure called $\frac{1}{\alpha}$-entropic independence [AJKPV21a; AJKPV21b], where $\alpha \in [\frac{1}{k}, 1]$, and a higher α corresponds to better local spectral expansion. For a distribution μ satisfying $\frac{1}{\alpha}$-entropic independence, there exists a higher-order random walk that generates samples from μ in time $O(n^{1/\alpha} \cdot \mathsf{poly}(k))$. While the dependence on k is acceptable, we can hope to improve the dependence on n if

we have access to the *marginals* of μ. Marginals here refers to estimates of $\Pr_{S\sim\mu}[i \in S]$ for individual ground set elements $i \in [n]$

We propose and analyze a meta-algorithm called *domain sparsification*, which uses the estimates of μ's marginals to reduce the size of the domain $[n]$ by a $O(n^\alpha)$ factor for $\frac{1}{\alpha}$-entropically independent distributions. In settings where we need to repeatedly generate samples from μ, we reduce the amortized cost of each sample by a factor of $\Omega(n)$. This extends the work of [AD20], which is specialized to the $\alpha = 1$ case.

The analysis of domain sparsification relies heavily on the connections between entropic independence and the "geometry of polynomials." Critically, we use the geometric characterization of entropically independent distributions from [AJKPV21b].

This chapter is based on joint work with Nima Anari, Michał Dereziński, and Thuy-Duong (June) Vuong, published in [ADVY22].

1.2.2 Results on random geometric graphs

Random geometric graphs are useful models for real-world data, as well as promising candidates for high-dimensional expanders. "Random geometric graph" can be defined in several different ways, with many kinds of underlying geometries; we generally represent vertices with vectors in $\mathbb{R}^d$, and place edges based on the distance between their vectors. The underlying geometry imposed on the graph may better capture observed phenomena like *triadic closure* than a geometry-agnostic network model, the Erdös-Rényi, can.

Graphs containing many triangles relative to its degree are promising candidates for high-dimensional expanders. It is then a natural question to understand the expansion properties of random geometric graphs. However, as we increase the underlying dimension d, random geometric graphs often "lose their geometry" and start behaving more like Erdös-Rényi graphs. It is known that Erdös-Rényi graphs are *not* good candidates for HDXes; by the time they exhibit sufficient connectivity properties, they are far too dense. Thus, if we hope for the random geometric graphs to be good HDXes, we need to choose their underlying dimension d carefully. This motivates our work in Chapter 5.

Chapter 5: Detecting latent geometry in random graphs

In Chapter 5 (with preliminaries in Chapter 4), we introduce and nearly resolve a conjecture regarding random geometric graphs.

We focus on a distribution we denote by $\mathsf{Geo}_d(n,p)$, where vertices are represented by unit vectors in $\mathbb{R}^d$ (which lie on the sphere $\mathbb{S}^{d-1}$) uniformly at random, and place edges between pairs of vectors whose dot product exceeds a threshold τ. We can choose τ so the edge probability is p. Formally, we answer a "Question 0" for $\mathsf{Geo}_d(n,p)$: for what dimension d is $\mathsf{Geo}_d(n,p)$ statistically indistinguishable from the Erdös-Rényi model $\mathsf{G}(n,p)$? In other words, at what threshold for d does $\mathsf{Geo}_d(n,p)$ lose its underlying geometry?

We improve upon the known indistinguishability threshold from [BBN20] for all choices of p. In the sparse regime when $p = O\left(\frac{1}{n}\right)$, we resolve the threshold question up to log factors,

settling an open problem from [BDER16]. A key contribution of our work is a toolkit for analyzing these graphs:

1. We use optimal transport maps to reduce questions about irregular distributions on the unit sphere to tractable computations on the uniform distribution over $\mathbb{S}^{d-1}$.

2. We apply the belief propagation algorithm to manually compute the distributions of RGG embedding vectors, conditioned on the vectors forming a particular graph.

We will discuss our use optimal transport in detail; we'll only provide a high-level outline of how the belief-propagation argument works.

This chapter is based on joint work with Siqi Liu, Sidhanth Mohanty, and Tselil Schramm, published in [LMSY22b].

Chapter 6: High-dimensional expansion of random geometric graphs

In Chapter 6 (with preliminaries in Chapters 2 and 4), we briefly outline a proof of the 2-dimensional expansion (with high probability) of graphs sampled from $\mathsf{Geo}_d(n,p)$ for $d = \Theta(\log n)$ and $p = 1/n^{1-\varepsilon}$ for any $\varepsilon \in (0,1)$. This distribution does not yield "true" 2-dimensional expansion because the expected degree of these graphs is n^{ε}, but selecting ε as a small constant does give us graphs with sublinear degree.

The choice of d is also influenced by our previous work on distinguishing $\mathsf{Geo}_d(n,p)$ and $\mathsf{G}(n,p)$. Graphs sampled from $\mathsf{G}(n,p)$ do not exhibit 2-dimensional expansion until $p \gg \frac{1}{\sqrt{n}}$, so we need to choose a dimension where $\mathsf{Geo}_d(n,p)$ and $\mathsf{G}(n,p)$ are distinguishable if we hope to achieve 2-dimensional expansion for smaller values of p.

The analysis of high-dimensional expansion is helped by a local-to-global result, namely Oppenheim's trickle-down theorem, which reduced the task of establishing 2-dimensional expansion to establishing spectral expansion with gap $> \frac{1}{2}$ for the neighborhoods of $\mathsf{Geo}_d(n,p)$. As a corollary of our proof of 2-dimensional expansion, we also show that most graphs sampled from $\mathsf{Geo}_d(n,p)$ yield examples of HDXes for which trickle-down is tight.

The key insight driving our analysis is the fact that the graph induced on the neighborhood of a vertex in $G \sim \mathsf{Geo}_d(n,p)$ is distributed like a random geometric graph on a sphere cap of $\mathbb{S}^{d-1}$. However, when d is high, most of the mass on the sphere cap is concentrated near its boundary. Thus, we can approximate the neighborhood of with a graph sampled from $\mathsf{Geo}_{d-1}(n,p')$, where $p' > p$ depends on the diameter of the sphere cap.

This chapter is based on joint work with Siqi Liu, Sidhanth Mohanty, and Tselil Schramm, published in [LMSY22a].

1.3 Notation

- We use $[n]$ to denote the set $\{1, 2, \ldots, n\}$. $\binom{[n]}{k}$ denotes all size-k subsets of $[n]$.

- **Asymptotics:** We use standard big-O notation, and we use $\tilde{O}$ and $\tilde{\Omega}$ to hide $\mathsf{polylog}(n)$ factors. The notation $f(x) \gg g(x)$ to denote that $\lim_{x\to\infty} \frac{g(x)}{f(x)} = 0$; the argument x will be clear from context. We use $\simeq$ to denote asymptotic equivalence.

- **Eigenvalues:** Given a matrix $M \in \mathbb{R}^{n\times n}$, we refer to its n eigenvalues by:

$$\lambda_1(M) \geq \lambda_2(M) \geq \ldots \geq \lambda_n(M)$$

- We use **boldface** for random variables (e.g. $\boldsymbol{v}$, $\boldsymbol{X}$). We'll use $\boldsymbol{X} \mid \mathcal{E}$ to denote the random variable $\boldsymbol{X}$ sampled from the conditional distribution of $\boldsymbol{X}$ conditioned on the event $\mathcal{E}$.

- **Binary entropy:** We use $\log x$ to denote the natural base logarithm, and for $x \in [0,1]$,

$$H(x) = x \ln \frac{1}{x} + (1-x) \ln \frac{1}{1-x}$$

  is the binary entropy function (with the understanding that $H(0) = H(1) = 0$).

- **Unit sphere:** We use $\mathbb{S}^{d-1}$ to denote $\{x \in \mathbb{R}^d : \|x\|_2 = 1\}$, the unit sphere in $\mathbb{R}^d$, which is inherently a $(d-1)$-dimensional surface.

- **p-norms of vectors and distributions:** For $v \in \mathbb{R}^d$ and $p \in (0,\infty)$, we use $\|\cdot\|_p$ to indicate the p-norm of v:

$$\|v\|_p = \left(\sum_{i\in[d]} |v_i|^p\right)^{1/p}$$

  The 0-norm is $\|v\|_0 = \#\{i \in [d] : v_i \neq 0\}$ and the infinity-norm is $\|v\|_\infty = \max_{i\in[d]} |v_i|$. In Chapter 5, we overload the same notation to denote the p-norm, for $p < \infty$, of a *function* $f : \mathbb{S}^{d-1} \to \mathbb{R}$ on the d-dimensional unit sphere as

$$\|f\|_p := (\mathbb{E}_{\boldsymbol{z}\sim\rho} |f(\boldsymbol{z})|^p)^{1/p}$$

  and for $p = \infty$ as $\|f\|_\infty := \sup_z$.

- **Probability densities on $\mathbb{S}^{d-1}$:** We use ρ to denote the uniform distribution on $\mathbb{S}^{d-1}$. Given a distribution ν on $\mathbb{S}^{d-1}$, we overload notation and use $\|\nu\|_p$ to denote $\|\frac{d\nu}{d\rho}\|_p$. We will also frequently use the symbol ν itself to denote its relative density to ρ.

  For a set A, we overload notation and use A to denote the uniform distribution on A when it is clear from context.

- For two distributions μ and ν over Ω, their *total variation distance* is

$$\mathrm{d_{TV}}(\mu,\nu) := \frac{1}{2}\|\mu - \nu\|_1$$

Their *Kullback-Leibler divergence (KL divergence)* or *relative entropy* is

$$\mathcal{D}_{\mathrm{KL}}(\mu\|\nu) := \mathbb{E}_{\omega\sim\mu}\left[\log\left(\frac{\mu(\omega)}{\nu(\omega)}\right)\right]$$

when ν is absolutely continuous with respect to μ. If ν is not, then their relative entropy is ∞.

A simple but useful observation about relative entropy is the following.

Observation 1.3.1. If for all $x \in \Omega$, $\frac{d\nu}{d\mu}(x) \leq C$: (1) $\mathcal{D}_{\mathrm{KL}}(\nu\|\mu) \leq \ln C$, and (2) for any event $\mathcal{E}$: $\nu(\mathcal{E}) \leq C \cdot \mu(\mathcal{E})$.

In Chapter 5, we will also use Pinsker's inequality to bound the total variation distance between two probability distributions in terms of the relative entropy.

Theorem 1.3.1 (Pinsker's inequality)**.** For distributions μ, ν over the same domain,

$$\mathrm{d}_{\mathrm{TV}}(\nu,\mu)^2 \leq \frac{1}{2}\mathcal{D}_{\mathrm{KL}}(\nu\|\mu)$$

Pinsker's inequality is a standard information theoretic tool. See e.g. Theorem 2.16 of [Mas07] for a proof.

Chapter 2

Higher-Order Random Walks

I'm only up when you're not down
Don't wanna fly if you're still on the ground [Swi06]

This chapter is intended as a friendly survey of the higher-order random walks that come from simplicial complexes, as well as the tools from the geometry of polynomials to establish their rapid mixing.

2.1 Markov chains and conventions

We refer to [Ber16] as a reference for the basics of Markov chains and their analysis.

A Markov chain is a stochastic process $\{X_t\}_{t\in\mathbb{N}}$ specified by parameters (Ω, P, μ_0). Ω is a state space, specifying the values each X_t may take. $P \in [0,1]^{\Omega\times\Omega}$ is a transition matrix that specifies the distribution of X_{t+1} given the value of X_t. We use the convention

$$P(i,j) = \Pr[X_{t+1} = j | X_t = i]$$

for $i, j \in \Omega$, so P is *row-stochastic* (i.e. every row sums to 1). μ_0 is a distribution that specifies where we start (X_0). When μ_0 is omitted, we assume a worst-case μ_0.

Given a weighted graph $G = (V, E, w)$, we can define a transition matrix using its:

$$\text{weighted adjacency matrix } A : A(i,j) = w(i,j)$$
$$\text{weighted diagonal matrix } D : D(i,i) = \sum_{j\in\Omega} w(i,j), \text{ otherwise } D(i,j) = 0$$

We call $D^{-1}A$ the *normalized adjacency matrix* of M; it is a transition matrix that describes a Markov chain that is a random walk on G, with edge transitions proportional to $w(\cdot)$. It is straightforward to verify that $D^{-1}A$ is row-stochastic.

Remark 2.1.1. We will sometimes abuse language and refer to the eigenvalues of G's normalized adjacency matrix as simply "the eigenvalues of G," even though G may have other canonical matrices like A with different eigenvalues. Almost all graphs we discuss in this work will be analyzed via their normalized adjacency matrices.

All of the Markov chains we consider with will be *irreducible* (we can reach j from i for all $i, j \in \Omega$) and *aperiodic* (we cannot find $i, j \in \Omega$ such that every path between i and j is a multiple of some $k \geq 2$). One property of an irreducible, aperiodic Markov chain is that $\lambda_1(P) = 1$, so in particular, there exists a row vector μ with $\|\mu\|_1 = 1$ such that $\mu = \mu P$. This particular eigenvector μ specifies a probability distribution over Ω, which we call the *stationary distribution* of our Markov chain. When the chain is irreducible and aperiodic, $\mu = \lim_{t\to\infty} vP^t$ for any $v \in \mathbb{R}^\Omega$, and moreover, μ is unique.

All of the Markov chains we consider will also be *time-reversible*, so in particular, they satisfy the *detailed balance* equations:

$$\mu_i \cdot P[i, j] = \mu_j \cdot P[j, i] \text{ for all } i, j \in \Omega$$

While we always converge to μ with infinite time, whether or not a Markov chain can help us efficiently sample from μ heavily depends on its *mixing time*, which lower bounds the rate that the chain converges to μ. For a distribution ν over Ω and $\varepsilon \in (0, 1)$,

$$\begin{aligned} t_{\text{mix}}(P, \nu, \epsilon) &= \min\{t \geq 0 \mid \mathrm{d}_{\text{TV}}\left(\nu P^t, \mu\right) \leq \varepsilon\} \\ t_{\text{mix}}(P, \varepsilon) &= \max_{x\in\Omega} \{t_{\text{mix}}(P, \delta_x, \varepsilon)\} \end{aligned}$$

where δ_x is the point mass distribution supported on x. When we don't specify ε, assume $\varepsilon = \frac{1}{4}$; any small constant suffices. This is due to the following well-known property of mixing times:

Theorem 2.1.1. If an irreducible aperiodic Markov chain with stationary distribution μ and transition matrix P satisfies $\mathrm{d}_{\text{TV}}\left(\nu P^t, \mu\right) \leq 1/4$ for all distributions ν over Ω and some $t \geq 1$, then for any $\varepsilon \in (0, 1/4]$,

$$t_{\text{mix}}(P, \epsilon) \leq t \log(1/\varepsilon).$$

The mixing time of a Markov chain can be analyzed by computing P's spectral gap, $\lambda_1(P) - \lambda_2(P) = 1 - \lambda_2(P)$.

Theorem 2.1.2. Let μ be the stationary distribution of a Markov chain given by (Ω, P), and let $\mu_{\min} = \min_{\omega\in\Omega} \mu(\omega)$.

$$t_{\text{mix}}\left(P, \frac{1}{4}\right) \leq \frac{1}{1 - \lambda_2(P)} \left(\frac{1}{2} \log \frac{1}{\mu_{\min}} + 1\right)$$

2.1.1 Mixing time via Poincaré and modified log-Sobolev

This section will not be critical for Chapter 3, Chapter 5, and Chapter 6, but may be useful for context to understand some of the exposition in Chapter 2; we keep it here for completeness.

Definition 2.1.2. Let μ be the stationary distribution of a Markov chain given by (Ω, P).

- For $f, g \in \mathbb{R}^\Omega$, the *Dirichlet form* of P is defined as
$$\mathcal{E}_P(f, g) = \langle f, (I - P)g\rangle_\mu = \langle (I - P)f, g\rangle_\mu.$$
- For a vector $v \in \mathbb{R}^\Omega$, and a distribution ν over Ω, we use $\mathbb{E}_\nu[v]$ to denote $\mathbb{E}_{i\sim\nu}[v_i]$. For $f \in \mathbb{R}^\Omega$, the *variance* of f is:
$$\mathrm{Var}_\mu[f] = \mathbb{E}_\mu\left[f^2\right] - \left(\mathbb{E}_\mu\left[f\right]\right)^2$$
For $f \in \mathbb{R}^\Omega_{\geq 0}$, its *entropy* is given by
$$\mathrm{Ent}_\mu[f] = \mathbb{E}_\mu\left[f \log f\right] - \mathbb{E}_\mu\left[f \log \mathbb{E}_\mu[f]\right]$$
- The *Poincaré constant* of P is defined to be
$$\lambda(P) = \inf\left\{\frac{\mathcal{E}_P(f, f)}{\mathrm{Var}_\mu[f]} : f \in \mathbb{R}^\Omega, \mathrm{Var}_\mu[f] \neq 0\right\}$$
- The *modified log-Sobolev constant* of P is defined to be
$$\rho(P) = \inf\left\{\frac{\mathcal{E}_P(f, \log f)}{\mathrm{Ent}_\mu[f]} : f \in \mathbb{R}^\Omega_{\geq 0}, \mathrm{Ent}_\mu[f] \neq 0\right\}$$
Again, it suffices to consider $f \in \mathbb{R}^\Omega_{\geq 0}$ satisfying $\mathrm{Ent}_\mu[f] \neq 0$ and $\mathbb{E}_\mu f = 1$.

Fact 2.1.3. The Poincaré constant is equal to P's spectral gap, $1 - \lambda_2(P)$.

Thus, we can infer an upper bound on mixing time bounds from a lower bound on the Poincaré constant, using Theorem 2.1.1.

Remark 2.1.4. For the rest of this work, we will almost always use the notation $1 - \lambda_2(P)$ to refer to the spectral gap, rather than use the (confusing) notation $\lambda(P)$ for the Poincaré constant.

The modified log-Sobolev constant provides an even stronger mixing time bound [BT03].

Theorem 2.1.3. Let μ be the stationary distribution of a Markov chain given by (Ω, P), and let $\mu_{\min} = \min_{\omega\in\Omega} \mu(\omega)$.
$$t_{\mathrm{mix}}\left(P, \frac{1}{4}\right) \leq \frac{1}{\rho(P)}\left(\log\log\frac{1}{\mu_{\min}} + 1\right)$$
Relatedly, it is usually more difficult to establish a lower bound on the modified log-Sobolev constant than it is to establish a lower bound on the spectral gap.

2.2 Basic HDX definitions

We borrow much of the notation for HDXes used in [ALGV19], though we define the weight function $w(\cdot)$ differently.

The object underlying a high-dimensional expander is called a simplicial complex.

Definition 2.2.1. A *simplicial complex* X over a ground set $[n]$ is a collection of downward-closed sets $S \subseteq [n]$. Downward-closed means that if $T \subseteq S$ and $S \in X$, then $T \in X$.

A simplicial complex is k-dimensional if its largest faces have size $(k+1)$. We assume that all simplicial complexes have maximal faces of equal size.

Definition 2.2.2. Given simplicial complex X, let $X(k')$ be the set of its k'-faces, given by $\{S \in X : |S| \leq k'+1\}$.

There is this "off-by-one" in the definition because a simplex on $(k+1)$ vertices inherently lives in k-dimensional space. For instance, a connected graph is a 1-dimensional complex X, where size 2 sets in X are edges (1-dimensional, as they are line segments), and size 1 sets are vertices (0-dimensional, as they are points).

Definition 2.2.3. The 1-skeleton of X is its "underlying graph," $X(1) \cup X(2)$.

We also have a high-dimensional analogue of "neighborhoods," called "links."

Definition 2.2.4. Let X be a simplicial complex, and $T \in X$ be one of its faces. The *link* of T, denoted by X_T, is the simplicial complex given by $\{S \setminus T : S \in X, S \supseteq T\}$.

In a 1-dimensional simplicial complex (i.e. graph), the link of a single vertex coincides with the usual notion of "neighborhood" in a graph.

Simplicial complexes naturally support weights as well. For a k-dimensional complex, let $w : X(k) \to \mathbb{R}_{\geq 0}$ be a weight function from the maximal (k-dimensional) faces to positive real numbers. The weights for the remainder of the faces propagate downwards inductively:

$$\text{For } T \in X(i),\ w(T) = \frac{1}{k-i} \cdot \sum_{S \in X(i+1): S \supset T} w(S)$$

It is not difficult to verify the following property of weights:

Proposition 2.2.5. Let X be a $(k-1)$-dimensional simplicial complex with weights specified by $w : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$. Then, for $T \in X(i)$ with $i \leq (k-1)$:

$$w(T) = \sum_{S \in X(k-1): S \supset T} w(S)$$

If $\sum_{S \subseteq [n]: |S| = k+1} w(S) = 1$, so w specifies a distribution μ over $X(k)$,

$$w(T) = \Pr_{S \sim \mu}[T \subseteq S]$$

2.2.1 Local spectral expansion and the trickle-down theorem

We remark that there are a few different notions of high-dimensional expansion. We focus on *local spectral expansion*, which generalizes spectral expansion in graphs. Other widely considered definitions of high-dimensional expansion include coboundary expansion [Gro10a; LMM16] and cosystolic expansion [EK16a], which generalize combinatorial expansion in graphs.

Definition 2.2.6. Let X be a simplicial complex, and let G denote its 1-skeleton, which can be represented by a weighted graph. Let $\lambda_2(G)$ denote the second largest eigenvalue of the normalized adjacency matrix of G.

The normalized adjacency matrix here refers to the transition matrix of the weighted random walk on G. The top eigenvalue of the normalized adjacency matrix, $\lambda_1(G)$ is always equal to 1, since there is an eigenvector corresponding to the stationary distribution.

In the expander graph literature, a family of (weighted) graphs $\{G_i\}_{i\in\mathbb{N}}$ is *spectrally expanding* if $\lambda_2(G_i)$ is bounded away from 1 by a constant for all $i \in \mathbb{N}$, i.e. $\lambda_2(G_i) \leq 1-\gamma$ for all $i \in \mathbb{N}$ and $\gamma > 0$. We also refer to such families of graphs as γ-spectral expanders. We can generalize this definition to simplicial complexes. We use the definition below when we talk about "HDXes:"

Definition 2.2.7. Let X be a weighted simplicial complex with maximal faces of size k, and recall that X_S is the link of $S \in X$. Let G_S denote the 1-skeletons X_S, respectively. X is a $(\gamma_0, \gamma_1, \ldots, \gamma_{k-2})$-*local spectral expander* if $\lambda_2(G_S) \leq \gamma_i$ for all $|S| = i$, and $0 \leq i \leq k-2$.

Note that in any simplicial complex, the only size-0 face is $\emptyset$. The link of $\emptyset$ is just X itself, so the condition that $\lambda_2(G_\emptyset) \leq \gamma_0$ is equivalent to the underlying graph of X itself being expanding. We also impose no conditions on G_S for S satisfying $|S| = k-1$ or $|S| = k$, since these links do not have any 1-faces (edges).

Example 2.2.8. In the case where $k = 2$ (which is the lowest-dimensional non-graph simplicial complex) and all of the 2-faces of X have uniform weight, X being a (γ_0, γ_1)-local-spectral-expander just means that

- The graph on the neighborhood of every vertex in X is a γ_1-expander.
- A weighted graph over all vertices of X is a γ_0-expander.

This is why we approximately described a "high-dimensional spectral expander" in Chapter 1 by a graph that is itself expanding, and whose neighborhoods are also expanding.

As noted in Section 1.1.1, HDXes exhibit "local-to-global" behavior: that is, it suffices to only study links (and often, only the links of the largest faces) to deduce bounds on the expansion of objects like the down-up-walks that are defined over the whole complex. In Chapter 6, we leverage a particular local-to-global result known as Oppenheim's "trickle-down" theorem:

Theorem 2.2.1. (See [Opp18]) Let X be a $(k-1)$-dimensional simplicial complex satisfying $\lambda_2(G_S) < \lambda$ for all S of size $\leq (k-2)$. If G_S is connected for all S with $0 \leq |S| \leq k-2$, then G_S is a $(\frac{\lambda}{1-(k-2)\lambda}, \frac{\lambda}{1-(k-3)\lambda}, \ldots \frac{\lambda}{1-\lambda}, \lambda)$-spectral-expander.

In short, assuming only a mild constraint that the links of a simplicial complex are all connected, we can deduce the expansion of all links in the simplicial complex, using only a bound on $\lambda_2(G_S)$ for $|S| = k-2$. Thus, trickle-down helps us prove local spectral expansion to understanding the expansion of only the simplest of links in the complex.

Remark 2.2.9. The trickle-down theorem also has a key weakness: we only obtain meaningful results when $\lambda < \frac{1}{k}$, which is often too restrictive for many of the chains we see in practice. A recent work [ALG22] tries to remedy this by controlling all of the eigenvalues of G_S at once, rather than only bounding $\lambda_2(G_S)$ for $S \in X(k-3)$; they replace the condition $\lambda_2(G_2) \leq \lambda$ with a matrix inequality instead.

2.3 Higher-order random walks

Now that we have formalized how weights and local spectral expansion are defined for a simplicial complex X, we define the *higher-order random walks* associated with X.

To better understand the higher-order random walks, we'll first take a somewhat redundant view of the standard lazy random walk on a weighted graph, with a $\frac{1}{2}$ probability we stay at our current state in any given time step. In this walk, our state space is the vertex set V, and in each step, if we are not lazy and stay at the current state, then we transition to a neighbor of our current state, proportional to the weight of the edge to that neighbor. We can view this walk as the composition of two distinct processes: an *up* operator that starts at a vertex, and then selects an edge that contains it, and a *down* operator, where we are at an edge and select one of the vertices that contains it.

Higher-order random walks generalize the standard vertex→edge→vertex (i.e. 0-face→ 1-face→ 0-face) walk. They all can be decomposed into down and an up operators.

Definition 2.3.1. Let X be a simplicial complex with weights defined using $w : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$, and choose $\ell < k$. The *down operator*, denoted $D_{k\to\ell}$, is a transition matrix mapping $X(k-1)$ to $X(\ell-1)$. The probability we transition from $S \in X(k-1)$ to $T \in X(\ell-1)$ is:

$$D_{k\to\ell}(S,T) = \begin{cases} \frac{1}{\binom{k}{\ell}} & \text{if } T \subset S, \\ 0 & \text{otherwise.} \end{cases}$$

The *up operator*, denoted $U_{\ell\to k}$, is a transition matrix mapping $X(\ell-1)$ to $X(k-1)$. The probability we transition from $T \in X(\ell-1)$ to $S \in X(k-1)$ is:

$$U_{\ell\to k}(T,S) = \begin{cases} \frac{w(S)}{w(T)} & \text{if } T \subseteq S, \\ 0 & \text{otherwise.} \end{cases}$$

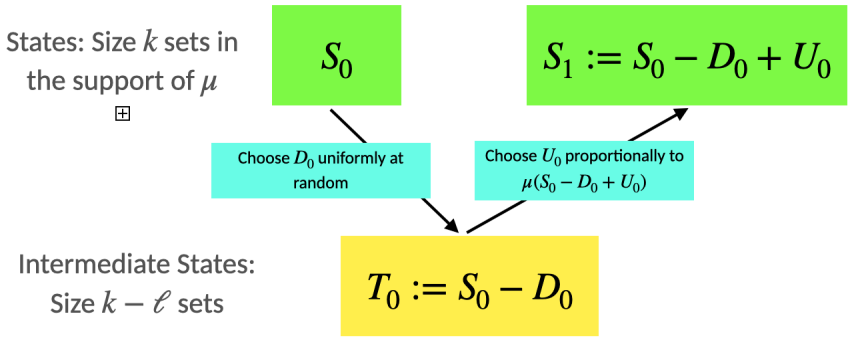

Figure 2.1: A diagram illustrating a single step of the down-up walk.

We note that in some other works (notably [ALGV19], [AL20]), the down and up operators are from k-faces to ℓ-faces, rather than size-k sets to size-ℓ sets.

Here, we are following the convention of [AD20], [AASV21], [AJKPV21b], and [ADVY22], which are closer related works to the content in Chapter 3.

Definition 2.3.2. The $(k \leftrightarrow \ell)$-*down-up walk* associated with μ is a Markov chain over sets in $X(k-1)$ with transition matrix $U_{\ell \to k} D_{k \to \ell}$.

Proposition 2.3.3. Let X be a simplicial complex with weights given using $w : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$, and choose $\ell < k$. The stationary distribution of the k-down-up walk is proportional to $w(\cdot)$.

This fact can be verified using detailed balance. In general, we can define down-up as well as up-down walks (where we apply the up operator before the down operator) over arbitrary $X(i)$. However, since a key problem of interest for this work is sampling from $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$, we will primarily be working with the $(k \leftrightarrow \ell)$-down-up walk.

This explains how we can design MCMC algorithms using simplicial complexes:

Remark 2.3.4. In typical approximate sampling problems, we only have oracle access to a function f that is proportional to μ. For instance, if we want to sample uniformly from the distribution of spanning trees in a graph, it is difficult to count the number of spanning trees a priori to specify μ exactly. However, it is easy to check whether or not a subset of the edges forms a spanning tree, so it is reasonable to assume we can access the weight function $w(T) = 1$ for a spanning tree $T \subseteq E$. Due to Proposition 2.3.3, having access to such a $w(\cdot)$ is enough to set up a $(k \leftrightarrow \ell)$-down-up-walk that mixes to μ.

When we implement a $(k \leftrightarrow \ell)$-down-up-walk, the down step takes time polynomial in k. However, the up step takes time polynomial in n, specifically $n^{O(\ell)}$ time, because we need to query $w(S)$ for each $S \supseteq T$.

Given a local spectral expander of dimension $(k-1)$, we can bound the spectral gap and the mixing time of the 1-down-1-up walk on $X(k-1)$.

Theorem 2.3.1. ([AL20], [AASV21]) Let X be a simplicial complex that represents a distribution $\mu : \binom{n}{k} \to \mathbb{R}_{\geq 0}$. If X is a $(\gamma_0, \gamma_1, \ldots, \gamma_{k-2})$-spectral expander, then:

$$1 - \lambda_2 \left(U_{\ell \to k} D_{k \to \ell}\right) \geq \frac{1}{k} \prod_{i=0}^{\ell-1} (1 - \gamma_i)$$

Remark 2.3.5. The case of $\ell = k-1$ recovers the original theorem stated in [AL20]. The version of the theorem produced above is not nearly as general as Theorem 47 of [AASV21]; we specialize it as follows:

- We use the same $\beta(S)$ for S of the same size, so the contraction factor is $\frac{1}{k}\prod_{i=0}^{\ell-1}\beta_i$. We assume the same for the contraction factors $\alpha(S)$, from their Definition 46.
- We replace f-divergence with "variance," which is equivalent to choosing $f(x) = x^2$.
- A local variance contraction of α at S is equivalent to a $\frac{1-\alpha}{\alpha}$ bound on the spectral gap of G_S if $\alpha \geq \frac{1}{2}$. See Fact 2.3.7 below.

Remark 2.3.6. Note that if we set $\ell = 0$, the spectral gap of the $(k \leftrightarrow 0)$-down-up-walk is $\frac{1}{k}$, so we always have $\mathsf{poly}(k)$ time mixing for this walk. However, we rarely implement this walk despite its fast mixing, because executing the up step takes time $n^{O(k)}$ (Remark 2.3.4). The real game here is to choose ℓ as close as possible to k while our spectral gap is still at least inverse polynomial in k.

We last describe how spectral gap relates to variance contraction between $\mu D_{k\to 2}$ and $\mu D_{k \to 1}$.

Fact 2.3.7. Let G be the 1-skeleton of a simplicial complex X corresponding to distribution $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$. Then, $1 - \lambda_2(G) \geq 2 \cdot (1-\alpha)$ if and only if μ satisfies a γ-variance contraction:

$$\mathrm{Var}_{\mu D_{k\to 1}}(\nu D_{2 \to 1}) \leq \alpha \cdot \mathrm{Var}_{\mu D_{k \to 2}}(\nu)$$

ν is an arbitrary function on $\binom{[n]}{2}$. It corresponds to the test function f in Definition 2.1.2.

Proof sketch. Let $P_2 = D_{2\to 1} U_{1 \to 2}$ and $P_1 = U_{1 \to 2} D_{2 \to 1}$. It is straightforward to show that $\lambda_2(P_2) = \lambda_2(P_1)$. In the proof of Lemma 14 in [CGM19], we have the following identity:

$$\mathrm{Var}_{\mu D_{k\to 1}}(\nu D_{2\to 1}) = \mathrm{Var}_{\mu D_{k \to 2}}(\nu) - \mathcal{E}_{P_2}(\nu, \nu)$$

If we incorporate the variance contraction inequality, we get $\mathcal{E}_{P_2}(\nu,\nu) \geq (1-\gamma)\mathrm{Var}_{\mu D_{k\to 2}}(\nu)$. This tells us $1 - \lambda_2(P_1) = 1 - \lambda_2(P_2) = 1 - \alpha$. We conclude by seeing that

$$1 - \lambda_2(P_1) = \frac{1}{2} \cdot (1 - \lambda_2(G))$$

since P_1 describes a lazy random walk on G with self-loop probability $\frac{1}{2}$. □

2.3.1 Spectral independence

We can also establish local spectral expansion, and thus, rapid mixing of the down-up walks, by proving something called *spectral independence* for distributions μ. Spectral independence (and the ensuing proof of local spectral expansion) has been at the crux of most of the recent breakthroughs in MCMC ([ALG21], [CGŠV21], [JPV21], [FGYZ22], [Bla+22]).

Though our work in Chapter 3 does not rely on spectral independences itself, it is heavily related to [AASV21], which employs spectral independence to establish rapid mixing of a wide family of down-up walks. We thus include a brief introduction to spectral independence for completeness.

Definition 2.3.8. For a distribution $\mu : \binom{n}{k} \to \mathbb{R}_{\geq 0}$, its *correlation matrix* ψ_μ is a matrix with rows and columns indexed by $[n]$ and entries specified by

$$\Psi_\mu(i,j) := \Pr_{S\sim\mu}[j \in S | i \in S] - \Pr_{S\sim\mu}[j \in S]$$

We drop the subscripts when the distribution is clear from the context.

Definition 2.3.9. A distribution $\mu : \binom{n}{k}$ is η-spectrally independent if

$$\lambda_1(\Psi_\mu) \leq \eta$$

μ is $(\eta_0, \eta_1, \ldots, \eta_{k-2})$-spectrally independent if for all sets S of size i for $0 \leq i \leq k$, the distribution μ conditioned on S (i.e. the distribution corresponding to X_S) is η_i-spectrally independent.

Spectral independence also directly implies local spectral expansion.

Theorem 2.3.2. Let $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$ be η-spectrally independent. Then, the 1-skeleton $G_\emptyset$ of its corresponding simplicial complex satisfies $\lambda_2(G_\emptyset) \leq \frac{\eta}{k-1}$.

Proof sketch. We will establish a Poincare inequality (Fact 2.1.3) for the random walk on $G_\emptyset$. It suffices to consider test vectors $f \in \mathbb{R}^{[n]}$ that satisfy $\mathbb{E}_{\mu D_{k\to 1}}[f] = 0$; we can shift f by a constant without affecting its Dirichlet form or variance. (Here, $\mu D_{k\to 1}$ describes the distribution proportional to the marginals $\Pr_{S\sim\mu}[i \in S]$.)

We first decompose the correlation matrix as follows:

$$\Psi_\mu = (k-1)P_\emptyset - M$$

Here, P is the transition matrix for $G_\emptyset$, while M is a matrix given by $M(i,j) = \Pr_{S\sim\mu}[j \in S]$. The factor of $(k-1)$ comes from normalization and the fact that

$$\sum_{j\neq i} \Pr_{S\sim\mu}[j \in S | i \in S] = k-1$$

Let Π be the diagonal matrix satisfying $\Pi(i,j)$. Since $\lambda_1(\Psi_\mu) \le \alpha$, we have:

$$(k-1)\cdot f^T\Pi Pf - f^T\Pi Mf \le \eta \cdot f^T\Pi f$$

Using $\mathbb{E}_\mu[f] = 0$, we can check that $f^T\Pi Mf = 0$. Rearranging the remaining terms, and adding $(k-1)\cdot f^T\Pi f$ to both sides:

$$(k-1-\eta)\cdot f^T\Pi f = (k-1)\cdot f^T\Pi(I-P)f$$

The quantity $f^T\Pi f$ is precisely $\mathrm{Var}_\mu(f)$, while $f^T\Pi Pf$ is $\mathcal{E}_P(f,f)$. □

Corollary 2.3.10. Let $\mu : \binom{[n]}{k} \to \mathbb{R}_{\ge 0}$ be $(\eta_0, \eta_1, \ldots, \eta_{k-2})$-spectrally independent. Then, its corresponding simplicial complex X is a $(\frac{\eta_0}{k-1}, \frac{\eta_1}{k-2}, \ldots, \eta)$-local spectral expander.

Corollary 2.3.11. If μ is $(\eta_0, \eta_1, \ldots, \eta_{k-2})$-spectrally independent, and $\eta_i \le O(1)$ for all $0 \le i \le k-1$, then the $(k \leftrightarrow k-1)$-down-up walk has spectral gap $\frac{1}{\mathsf{poly}(k)}$. Furthermore, if mixes in time $O(\mathsf{poly}(k)\cdot \log \mu_{\min}^{-1})$.

2.3.2 Entropic Independence

We can define *entropic independence* as a near-analogue of spectral independence: the quantities we previously saw that included variance are (almost) replaced with *entropy*. This yields modified log-Sobolev inequalities instead of Poincaré inequalities, which lead to much stronger upper bounds on mixing time.

We refer to [AJKPV21b] for a more complete treatment of entropic independence.

Definition 2.3.12. A probability distribution $\mu : \binom{[n]}{k} \to \mathbb{R}_{\ge 0}$ is $(1/\alpha)$-*entropically-independent* for $\alpha \in (0,1]$, if for all distributions ν on $\binom{[n]}{k}$,

$$\mathcal{D}_{\mathrm{KL}}(\nu D_{k\to 1} \| \mu D_{k\to 1}) \le \frac{1}{\alpha k}\cdot \mathcal{D}_{\mathrm{KL}}(\nu\|\mu)$$

Remark 2.3.13. We saw that spectral independence implied a variance contraction (Fact 2.3.7), but the corresponding entropy contraction

$$\mathrm{Ent}_{\mu D_{k\to 1}}(\nu D_{2\to 1}) \le c \cdot \mathrm{Ent}_{\mu D_{k\to 2}}(\nu)$$

for $c < 1$ is actually stronger than the entropy contraction implied in Definition 2.3.12. In this sense, entropic independence isn't a perfect analogue of spectral independence. Example 38 in the appendix of [AJKPV21b] has an example of a 2-entropically-independent distribution where the above entropy contraction does not hold.

Fact 2.3.14. Every distribution $\mu : \binom{[n]}{k} \to \mathbb{R}_{\ge 0}$ is k-entropically independent.

The fact is a direct consequence of the data processing inequality.

We will revisit entropic independence again and further elaborate on its relationship to spectral independence in Section 2.4.2.

2.4 Relationship to the geometry of polynomials

The mixing times of higher-order random walks are deeply connected to geometric aspects of a multivariate polynomial that encodes μ.

Definition 2.4.1. For a distribution $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$, its *generating polynomial* g_μ is

$$g_\mu(z_1, \ldots, z_n) := \sum_S \mu(S) \prod_{i \in S} z_i$$

In the expression for g_μ, we may also replace $\mu(\cdot)$ with $w(\cdot)$, a scalar multiple of $\mu(\cdot)$, and obtain the consistent definitions and the same results.

Definition 2.4.2. Let $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$ be a distribution with generating polynomial g_μ.

- We use $\mu|_T$ to denote μ restricted to sets $S \subseteq T$ rather than $S \subseteq [n]$. The generating polynomial of $\mu|_T$ equals g_μ evaluated at $z_i = 0$ for $i \notin T$.
- We use $\partial_T \mu$ to denote the stationary distribution over the $(k - |T| - 1)$-faces of the link of T, when considering μ as a simplicial complex. $\partial_T \mu$ is equivalent to the probability distribution specified by the polynomial

  $$\partial_{z_{i_1}} \cdots \partial_{z_{i_{|T|}}} g_\mu(z_1, \ldots, z_n) \text{ for } T = \{i_1, \ldots, i_{|T|}\}$$

  The order we take the partial derivatives does not matter.

2.4.1 Log-concavity

Definition 2.4.3. A distribution $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$ is *log-concave* if $\log g_\mu(z_1, \ldots, z_n)$ is concave over the positive orthant $\mathbb{R}^n_{\geq 0}$

Remark 2.4.4. We depart from the definition of *strongly log-concave* that is used in [ALGV19] and [CGM19]. Our Definition 2.4.3 also implies log-concavity of the derivatives. [AASV21] justifies this by using the fact that scaling preserves log-concavity.

$$\partial_i g_\mu(z_1, \ldots, z_{i-1}, z_{i+1}, \ldots, z_n) = \lim_{\lambda \to \infty} \frac{1}{\lambda} g_\mu(\lambda z_1, \ldots, \lambda z_{i-1}, z_i, \lambda z_{i+1}, \ldots, \lambda z_n)$$

Theorem 2.4.1. If $\mu : \binom{[n]}{k}$ is log-concave, then its corresponding simplicial complex X is a $(0, 0, \ldots, 0)$-local spectral expander.

Proof outline. A full proof is one of the central contributions of [ALGV19].

First consider $g_\mu(z_1, \ldots, z_n)$. We use the following chain of implications to relate log-concavity to local spectral expansion.

- $g_\mu(z_1, \ldots, z_n)$ being log-concave at $\vec{1}$ means that the Hessian of $\log g_\mu(z_1, \ldots, z_n)$ is negative semi-definite at $\vec{1}$.

- Using some tools from linear algebra, we can show that this implies that the Hessian of $g_\mu(z_1, \ldots, z_n)$ has at most one positive eigenvalue.
- If we normalize the rows of the Hessian of $g_\mu(z_1, \ldots, z_n)$ at $\vec{1}$, we obtain the (normalized) adjacency matrix of G, the 1-skeleton of X. The positive eigenvalue corresponds to $\lambda_1(G) = 1$; we can then conclude (with some more linear algebra) that $\lambda_2(G) \leq 0$.

Log-concavity of all of the derivatives of g_μ eventually translates to $\lambda_2(G_S) \leq 0$ for all S with size $\leq (k-2)$, where G_S is the 1-skeleton of X_S. □

Corollary 2.4.5. Let $\mu : \binom{[n]}{k} \to \mathbb{R}$ be log-concave. Then, its $(k \leftrightarrow k-1)$-down-up walk has spectral gap $\frac{1}{\mathsf{poly}(k)}$, and mixes in time $O(\mathsf{poly}(k) \cdot \log \mu_{\min}^{-1})$.

Next, we include a few examples of log-concave distributions.

Example 2.4.6. Let $\mathcal{M} = (E, \mathcal{I})$ be a matroid with bases $\mathcal{B}$. Then, if μ is the uniform distribution over $\mathcal{B}$, it is log-concave. This example also encapsulates the uniform distribution over spanning trees in a graph.

Example 2.4.7. Let L be a positive semidefinite matrix. The k-determinantal point process, or k-DPP, with kernel L [DM21] is a distribution $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$ defined by:

$$\mu(S) \propto \det(L_{S,S})$$

Here, $L_{S,S}$ is the principal submatrix of L whose rows and columns correspond to S.

If we have a collection of vectors $\{v_1, \ldots, v_n\}$ and let L be its Gram matrix, we are then sampling $S \subseteq [n]$ proportionally to the volume spanned by $\{v_i : i \in S\}$. Thus, we favor sets of vectors that are almost orthogonal to each other, and ignore vectors that only span a lower-dimensional space. The k-DPP is used in many applications in machine learning when we need *diverse* samples.

2.4.2 Fractional log-concavity

Definition 2.4.8. A distribution $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$ is *α-fractionally-log-concave* for parameter $\alpha \in (0, 1]$ if $\log g_\mu(z_1^\alpha, \ldots, z_n^\alpha)$ is concave over the positive orthant $(z_1, \ldots, z_n) \in \mathbb{R}^n_{\geq 0}$. When $\alpha = 1$, we recover the definition of log-concavity.

Remark 2.4.9. Remark 2.4.4 also holds for fractionally log-concave polynomials. As a result, we also get fractional log-concavity of the derivatives of g_μ.

Fractionally log-concave distributions also satisfy spectral independence.

Theorem 2.4.2 (See Remark 70 of [AASV21])**.** Let $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$ be α-fractionally-log-concave. Then, μ is $\frac{1}{\alpha}$-spectrally independent.

Corollary 2.4.10. If $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$ is α-fractionally-log-concave, then its $(k \leftrightarrow k - \lceil \alpha^{-1} \rceil)$-up-down walk has spectral gap $\frac{1}{\mathsf{poly}(k)}$, where the exponent of the polynomial depends on α^{-1}.

Proof sketch of Corollary 2.4.10. We apply Theorem 2.3.1 using $\ell = k - \lfloor \alpha^{-1} \rfloor$. Note that $\gamma_i \geq 1$ for $i \geq \ell$; treating α as a constant, we only obtain constant γ_i for $i \leq \ell - 1$. □

We next provide some examples of α-fractionally-log-concave polynomials and distributions for $\alpha < 1$. These examples are primarily from [AASV21]; they have proofs as well.

Example 2.4.11. If g is a degree-k homogeneous multi-affine polynomial, then it is $(1/k)$-log-concave. Every monomial $\prod_{i \in S} z_i^{1/k}$ is concave, since by Hölder's inequality

$$\prod_{i \in S} (\lambda z_i + (1-\lambda) y_i) \geq \left(\lambda \prod_{i \in S} z_i^{1/k} + (1-\lambda) \prod_{i \in S} y_i^{1/k} \right)^k$$

Now, $g(z_1^{1/k}, \ldots, z_n^{1/k})$ is concave, since it is a (positive-weighted) sum of concave functions $\prod_{i \in S} z_i^{1/k}$. Thus, $\log g(z_1^{1/k}, \ldots, z_n^{1/k})$ is also concave, as log is both monotone and concave.

Example 2.4.12. We present another toy class of α-fractionally-log-concave polynomials that provides some intuition despite not having many applications.

Let μ be an α-fractionally-log-concave polynomial over the variables $z_1, \ldots, z_n$. If we replace each z_i with the monomial $\prod_{j=1}^{m} z_i^{(j)}$, we obtain a degree mk, $\frac{\alpha}{m}$-fractionally-log-concave polynomial over the variables $\{z_i^{(j)} : i \in [n], j \in [m]\}$. For example, if the starting distribution μ is the uniform distribution over bases of a matroid, then $\alpha = 1$, and the resulting distribution will be $(1/m)$-fractionally log-concave.

Notice that if we treat this polynomial as a distribution, for any $i \in [n]$, the elements $i^{(1)}, \ldots, i^{(m)}$ are all perfectly correlated. On the other hand, if $i \neq j$, any two $i^{(m_i)}$ and $j^{(m_j)}$ inherit the correlations from the log-concave distribution.

Example 2.4.13. Let G be a graph and $k \in \mathbb{N}$. For each set $S \subseteq \binom{V}{2k}$, let $\mu(S)$ be proportional to the number of perfect matchings of the induced subgraph of G on S. Sampling from μ allows us to approximately count the number of k-matchings, i.e., matchings using k edges, in G. For any value of k, this distribution is fractionally log-concave with $\alpha \geq \frac{1}{4}$, and it is conjectured to be fractionally log-concave with $\alpha \geq \frac{1}{2}$.

Not all choices of G result in efficient sampling algorithms. The implementation of an oracle for μ (or a function w proportional to μ) involves counting perfect matchings over $S \subseteq V$, and we do not have a $\mathsf{poly}(k)$ time algorithm for counting matchings in general graphs. We thus only consider downward closed graph families with an FPRAS for counting perfect matchings, e.g., bipartite graphs [JSV04], planar graphs [Kas67], certain minor-free graphs [EV19], and small genus graphs [GL99].

Example 2.4.14. Let L be a nonsymmetric positive semidefinite matrix, i.e., an $n \times n$ matrix L that satisfies $L + L^\intercal \succeq 0$. Then, the nonsymmetric k-determinantal point process, or nonsymmetric k-DPP with kernel L [GBDK19; GHDGB20; AV21], is defined by

$$\mu(S) \propto \det(L_{S,S})$$

The nonsymmetric k-DPP is fractionally log-concave for $\alpha \geq \frac{1}{4}$.

The non-symmetric k-DPP is also conjectured to be fractionally log-concave with $\alpha \geq \frac{1}{2}$.

Example 2.4.15. Start with a measure μ_0 on $\binom{[n]}{k}$ that is Strongly Rayleigh (see [BBL09] for a definition), such as a (symmetric) determinant point process, or the uniform distribution over spanning trees of a graph. Suppose that we partition the ground set into a constant number $c = O(1)$ of parts: $[n] = A_1 \cup A_2 \cup \cdots \cup A_c$, and fix cardinalities $k_1, \ldots, k_c \in \mathbb{N}$, with $k_1 + \cdots + k_c = k$. Then the *partition-constrained* version of μ_0 can be defined as

$$\mu(S) \propto \mu_0(S) \cdot \mathbb{1}\left[|S \cap A_i| = k_i \text{ for } i = 1, \ldots, c\right]$$

As long as $c = O(1)$, this distribution μ will be $\Omega(1)$-fractionally-log-concave. For some discussion of partition-constrained Strongly Rayleigh measures, see [CDKSV16].

At a very high level, we can think of $\frac{1}{\alpha}$-fractionally-log-concave distributions as having some positive correlations among α^{-1}-sized subsets of $[n]$, and otherwise negative correlations. Example 2.4.12 especially captures this intuition.

We also remark that many of the results from [AASV21], which introduces fractional log-concavity, are actually stated and obtained for a property called *sector-stability*. Also, all of the examples above have α-fractional-log-concavity that is deduced from 2α-sector stability. Sector stability, which is stronger than fractional log-concavity, concerns the locations of the zeros of g_μ. It also generalizes *real-stability*, which is a stronger version of log-concavity. Given the focus of this work on the higher-order random walks, only working with (fractional) log-concavity suffices.

Connecting entropic independence and spectral independence

Finally, we formalize how α-fractional-log-concavity implies $\frac{1}{\alpha}$-entropic independence.

Theorem 2.4.3 (Theorem 3 of [AJKPV21b])**.** μ is $\frac{1}{\alpha}$-entropically independent if and only if for all $(z_1, \ldots, z_n) \in \mathbb{R}^n_{\geq 0}$,

$$g_\mu(z_1^\alpha, \ldots, z_n^\alpha) \leq \left(\sum_{i=1}^n p_i z_i\right)^{k\alpha} \tag{2.1}$$

where $p = (p_1, \ldots, p_n) := \mu D_{k \to 1}$ are the scaled marginals of μ.

Corollary 2.4.16. α-fractionally log concavity implies $(1/\alpha)$-entropic independence.

[AJKPV21b] also show that $\frac{1}{\alpha}$-entropic independence under external fields (a stronger condition than just entropic independence) is equivalent to fractional log-concavity:

Definition 2.4.17. For $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$ and $\lambda \in \mathbb{R}^n_{\geq 0}$, the distribution $\lambda * \mu$ is defined by:

$$\Pr_{\lambda * \mu}[S] = \mu(S) \cdot \prod_{i \in S} \lambda_i$$

Equivalently, $\lambda * \mu$ is the distribution given by polynomial $g_\mu(\lambda_1 z_1, \ldots, \lambda_n z_n)$.

Theorem 2.4.4 (Theorem 3 of [AJKPV21b])**.** If $\lambda * \mu$ is $\frac{1}{\alpha}$-entropically independent for all $\lambda \in \mathbb{R}^n_{\geq 0}$, then μ is α-fractionally log-concave.

Entropic independence of all conditionings (which is implied by fractional log-concavity) also results in a modified log-Sobolev inequality for some higher order random walks:

Theorem 2.4.5 (Theorem 5 of [AJKPV21b])**.** Let μ and all of its conditionings μ_S satisfy $\frac{1}{\alpha}$-entropic independence for $\frac{1}{\alpha} \in \mathbb{Z}$. Then, for all distributions ν over $\binom{[n]}{k}$

$$\begin{aligned}\mathcal{D}_{\mathrm{KL}}(\nu D_{k \to \ell} U_{\ell \to k} \| \mu D_{k \to \ell} U_{\ell \to k}) &\leq \mathcal{D}_{\mathrm{KL}}(\nu D_{k \to \ell} \| \mu D_{k \to \ell}) \\ &\leq \left(1 - \binom{k-\ell}{\alpha^{-1}} \cdot \binom{k}{\alpha^{-1}}^{-1}\right) \mathcal{D}_{\mathrm{KL}}(\nu \| \mu)\end{aligned}$$

We can then obtain mixing time bounds using the modified log-Sobolev inequalities.

Chapter 3

Domain Sparsification and Entropic Independence

Gave you too much but it wasn't enough
But I'll be all right, it's just a thousand cuts [Swi19]

As noted in Chapter 1, this chapter is based on joint work with Nima Anari, Michał Dereziński, and Thuy-Duong (June) Vuong, published in [ADVY22].

3.1 Problem statement and summary of results

Consider a distribution $\mu : \binom{[n]}{k}$ satisfying $\frac{1}{\alpha}$-entropic independence for all conditionings (which is implied by α-fractional log-concavity), and the problem of approximately generating samples from μ. As noted in Remark 2.3.4, we actually have access to some weight function w that is a scaling of μ, but we will only work with μ for the rest of this chapter; the same algorithms can be implemented with only knowledge of w. Using Corollary 2.4.10 and Remark 2.3.4, the run-time of the $k \leftrightarrow (k - \lceil \alpha^{-1} \rceil)$-down-up walk is $O(n^{\lceil \alpha^{-1} \rceil} \cdot \mathsf{poly}(k))$, where the time-consuming "up" step takes time $O(n^{\lceil 1/\alpha \rceil})$.

Can we speed up this sampling process if we have access to additional information about μ? More specifically, if we have estimates p_i of the *single element marginals* $\Pr_{S \sim \mu}[i \in S]$ for all $i \in [n]$, can we improve the runtime of the "up" step, by perhaps reducing the number of sets we need to consider?

We can think of this as a question of how to "sparsify" the space of transitions that we consider in the "up" step. This is not at all a novel concept. In the continuous setting, if we have n data points lying in $\mathbb{R}^d$, algorithms like the Johnson-Lindenstrauss transform and its variants ([Mat08], [KN14]) allow us to project the data vectors into a lower-dimensional subspace and gain runtime improvements for downstream tasks such as linear regression. In the

discrete world, cut sparsification and spectral sparsification algorithms ([BK96], [BSST13]) have been key subroutines in linear system solvers and other graph algorithms.

3.1.1 Approach and main result

Domain sparsification, or rather a simplification of it, is easiest to describe for the case of uniform marginals ($p_i = \frac{k}{n}$ for $i \in [n]$) and when the desired total variation distance from our sample to μ is a constant. We first sample $T \subseteq [n]$ uniformly at random, with $|T| \simeq n^{1-\alpha} \cdot \mathsf{poly}(k)$ elements from the ground set. Then, we sample S from the restricted distribution $\mu|_T$, and as noted in Theorem 3.1.1, we establish that this sample S is close enough in distribution to a sample we would have obtained from μ. We thus reduced the size of the ground set, or the domain $[n]$, by a factor of $n^{1/\alpha}$, before we begin sampling.

Remark 3.1.1. It suffices for μ to be $\frac{1}{\alpha}$-entropically log-concave for our choice of T to yield subsamples S that are close in distribution to μ. However, a stronger condition, like the fractional log-concavity of μ, helps guarantee that S can be sampled efficiently.

To handle non-uniform p_i, we first perform a transformation to μ, which algorithmically translates to selecting elements of T in an i.i.d. fashion with probabilities dictated by p_i (Section 3.2.1). Furthermore, we do not need $\Pr_{S\sim\mu}[i \in S]$ to be *exactly* the same for all i. Instead, it is enough to just have an *estimate* of these marginals.

Finally, in order to achieve better total variation distance to μ, we perform a Markov-chain-based variant of the mentioned process; in each step of this Markov chain, we simply *add* elements to the current set S_0 in order to get $S_0 \cup T_0$, and then subsample a new set $S_1 \subseteq (S_0 \cup T_0)$. From here, we can add elements to S_1 to get $S_1 \cup T_1$, and subsample S_2 from $S_1 \cup T_1$, and so on. For a high level overview, see Figure 3.1.1. For details see Section 3.2.2.

Theorem 3.1.1 (Informal)**.** Let $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$ be $(1/\alpha)$-entropically independent. Suppose that we have access to estimates $p_1, \ldots, p_n$ of the marginals and an oracle that can produce i.i.d. samples $i \in [n]$ with $\Pr i \propto p_i$; suppose that our estimates satisfy

$$p_1 + \cdots + p_n = k \text{ and } p_i \geq \Omega(\Pr_{S\sim\mu}[i \in S]) \text{ for all } i \in [n]$$

Then we can sample $T \subseteq [n]$ with $T \leq O\left(n^{1-\alpha} \cdot \mathsf{poly}(k)\right)$ nonzero entries, in time $n^{1-\alpha} \cdot \mathsf{poly}(k)$, such that a random sample S of $\mu|_T$ approximately follows μ.

Theorem 3.1.1 follows directly from Proposition 3.2.4, Proposition 3.2.5, and Lemma 3.2.6.

The main application of domain sparsification is accelerating the time it takes to produce *multiple* samples from a distribution $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$. Suppose that an algorithm $\mathcal{A}$ can produce (approximate) samples from μ and any distribution $\mu|_T$ obtained from it by restricting the domain $[n]$ to a subset T, in time $\mathcal{T}(m, k)$, which depends polynomially on $m = |T|$. Then after a preprocessing stage, where we use $\mathcal{A}$ on the unrestricted domain to estimate the marginals of μ, we can produce new samples in time $\mathcal{T}(n^{1-\alpha} \cdot \mathsf{poly}(k), k)$ per sample, which is polynomially smaller than $\mathcal{T}(n, k)$, as long as k is smaller than some $\mathsf{poly}(n)$ threshold.

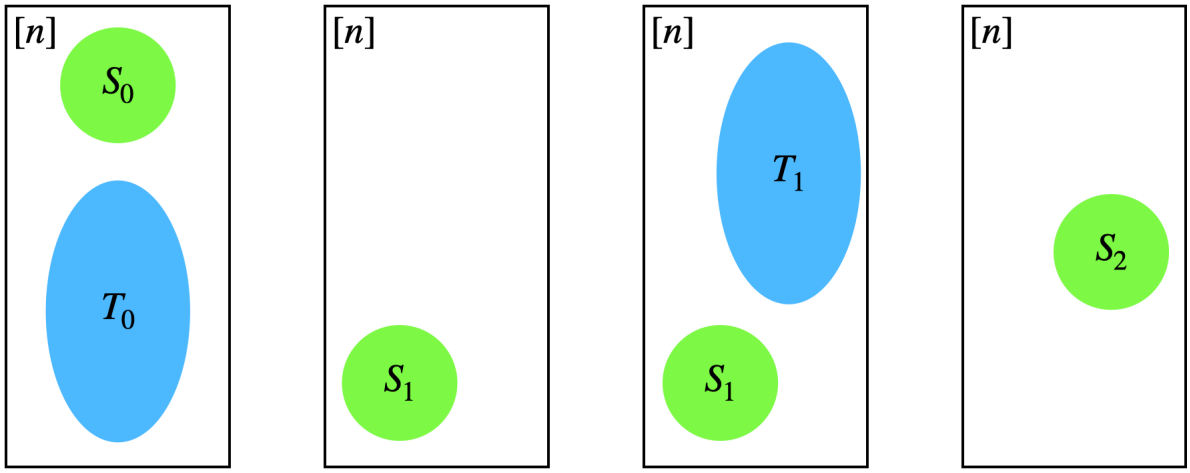

Figure 3.1: An overview of the domain sparsification scheme combined with Markov chain intermediate sampling. The first two panels demonstrate one step of the Markov chain to generate a sample S_1 approximately from μ, while the later two panels repeat the process to generate an additional sample S_2. Within each pair of panels, the blue sets T represent the sparsified domain (using marginals), which is much smaller than $[n]$.

The preprocessing step has to be done only once, and its cost gets amortized when we are interested in obtaining multiple samples from μ. A careful implementation, directly adapted from domain sparsification for log-concave polynomials by Anari and Dereziński [AD20], can bootstrap domain sparsification with estimation of marginals to complete the preprocessing step in roughly $\simeq \mathcal{T}(n,k) + n \cdot \mathsf{poly}(k, \log n) \cdot \mathcal{T}(n^{1-\alpha} \cdot \mathsf{poly}(k), k)$ time.

Corollary 3.1.2 (Informal)**.** Suppose that we have an algorithm $\mathcal{A}$ that can produce approximate samples from restriction of μ to a subset $T \subseteq [n]$ in time $\mathcal{T}(m,k)$, where $m = |T|$. Then we can produce the marginal estimates p_i in time

$$O\left(\mathcal{T}(n,k) + n \cdot \mathsf{poly}(k, \log n) \cdot \mathcal{T}(n^{1-\alpha} \cdot \mathsf{poly}(k), k)\right)$$

Furthermore, for any desired t, we can produce t i.i.d. approximate samples from μ in time

$$O\left(T(n,k) + \max\{t, n \cdot \mathsf{poly}(k, \log n)\} \cdot T(n^{1-\alpha} \cdot \mathsf{poly}(k), k)\right)$$

Sampling is also often used to solve the problem of approximate counting, or approximately computing the partition function when μ is specified by weights w:

$$\sum_{S \in \mathrm{supp}(\mu)} w(S).$$

To obtain an ε-relative error approximation, known reductions between counting and sampling [JVV86] introduce at least a multiplicative factor of $1/\varepsilon^2$ to the sampling time. Directly

Distribution	Reference	Time w/out D.S.	Amortized Time w/ D.S.
Matroid Bases	Ex. 2.4.6	$\tilde{O}(n \cdot \mathsf{poly}(k))$	$O(\mathsf{poly}(k) \cdot \mathsf{polylog}(n))$
k-DPP	Ex. 2.4.7	$\tilde{O}(n \cdot \mathsf{poly}(k))$	$O(\mathsf{poly}(k) \cdot \mathsf{polylog}(n))$
Matchings	Ex. 2.4.13	$\tilde{O}(n^4 \cdot \mathsf{poly}(k))$	$\tilde{O}(n^3 \cdot \mathsf{poly}(k))$
Nonsymmetric DPP	Ex. 2.4.14	$\tilde{O}(n^4 \cdot \mathsf{poly}(k))$	$\tilde{O}(n^3 \cdot \mathsf{poly}(k))$

Table 3.1: A summary of example distributions from Chapter 2 and the impact of domain sparsification on their amortized sampling time when we need to generate many samples. The column "Time w/out D.S." refers to the runtime of the down-up-walk used for sampling from $(1/\alpha)$-entropically independent distributions.

adapting the same technique for log-concave polynomials [AD20] and combining with our new domain sparsification result, we obtain an ε-relative error of the counts in time

$$\simeq \mathcal{T}(n,k) + \max\{n, 1/\varepsilon^2\} \cdot \mathsf{poly}(k, \log n) \cdot \mathcal{T}(n^{1-\alpha} \cdot \mathsf{poly}(k), k)$$

Notice that here $1/\varepsilon^2$ is multiplied by the term $\mathcal{T}(n^{1-\alpha} \cdot \mathsf{poly}(k), k)$ that can be substantially smaller than $\mathcal{T}(n,k)$; as a result, we can get a substantially improved running time for the high-precision regime where ε is inverse-polynomially small.

Corollary 3.1.3 (Informal, adapted from [AD20])**.** Suppose that we have an algorithm $\mathcal{A}$ that can produce approximate samples from any restriction $\mu|_T$ for $T \subseteq [n]$ in time $\mathcal{T}(m,k)$, where $m = |T|$. Then we can compute an ε relative error approximation of $\sum_S \mu(S)$ in time

$$O\left(T(n,k) + \max\{n, 1/\varepsilon^2\} \cdot \mathsf{poly}(k, \log n) \cdot \mathcal{T}(n^{1-\alpha} \cdot \mathsf{poly}(k), k)\right)$$

3.1.2 Applications

Our main results imply that as long as we estimate the probability of every vertex being part of a random k-matching, we can reduce the task of sampling k-matchings on an n vertex graph to graphs with only $n^{3/4} \cdot \mathsf{poly}(k)$ many vertices. Recall the examples of α-fractionally log-concave distributions from Section 2.4.2. We summarize the *amortized* speedups obtained after applying domain sparsification in Table 3.1.2 Note that domain sparsification for $\alpha = 1$ was already established in [AD20]; we include the results in our table for completeness.

3.1.3 Related work

A class of domain sparsification algorithms, related to the algorithms we used here, called intermediate sampling was first proposed by [DWH18; Der19] in the context of sampling from Determinantal Point Processes (DPPs, [DM21]), also known as Volume Sampling [DRVW06; DR10; GS12]. DPPs are a family of distributions (a small, but important, subset of distributions with log-concave generating polynomials) which arise for instance when sampling

random spanning trees [Gue83], as well as in randomized linear algebra [DW17; DCMW19], machine learning [KT11; KT+12; DKM20], optimization [NST22; DBPM20; MDK20], and other areas [Mac75; HKPV06; BLMV17].

The complexity of intermediate sampling for DPPs was further improved by [DCV19; CDV20], and the approach was extended to DPPs over continuous domains by [DWH22]. Crucially, these algorithms take advantage of the additional structure in DPPs, to enable *distortion-free* intermediate sampling: instead of using a Markov chain, this uses rejection sampling to draw *exactly* from the target distribution. This approach is not possible more generally, since μ typically does not have a tractable partition function. However, [AD20] showed that the original analysis of distortion-free intermediate sampling can largely be retained for distributions with log-concave generating polynomials, as long as we switch to a Markov chain implementation.

On the other hand, in this work, we largely abandon the original analysis in favor of a new one which is specific to the Markov chain and requires less precision in marginal estimates. As a result, we show that the preprocessing cost for Markov chain intermediate sampling is substantially smaller than for distortion-free intermediate sampling. This leads to significant improvements in time complexity even for DPPs, e.g., by reducing the preprocessing cost in [DCV19] from $\tilde{O}(n \cdot k^6 + k^9)$ to $\tilde{O}(n \cdot k^2 + k^3)$, where $\tilde{O}$ hides polylogarithmic terms.

3.2 The intermediate sampling algorithm

3.2.1 Isotropic transformation

We say a distribution μ is isotropic if for all $i \in [n]$, the marginal probability $\Pr_{S\sim\mu}[i \in S]$ is $\frac{k}{n}$, as done in [AD20]. We remark that this is only similar in name and spirit, but different in nature, to the analogous notion of isotropy for continuous distributions; the latter is defined based on the covariance matrix of the distribution, while the former is defined based on marginals. In this paper, isotropy captures "uniformity" over the elements of $[n]$ in their marginal probabilities. Below, we discuss a subdivision process [AD20] that transforms an arbitrary distribution μ over $\binom{[n]}{k}$ into a distribution μ' that is nearly-isotropic.

Definition 3.2.1. Let $\mu : \binom{n}{k} \to \mathbb{R}_{\geq 0}$ be an arbitrary probability distribution, and assume that we have estimates $p_1, \ldots, p_n$ of the marginals with

$$p_1 + \ldots + p_n = k \text{ and } p_i \geq \Omega\left(\Pr_{S\sim\mu}[i \in S]\right) \text{ for all } i \in [n]$$

Let $t_i := \lceil \frac{n}{k} \cdot p_i \rceil$. We will create a new distribution out of μ and these t_i: for each $i \in [n]$, create t_i copies of the element i and let the collection of all these copies be the new ground set: $U = \bigcup_{i=1}^{n} \{i^{(1)}, \ldots, i^{(t_i)}\}$. Define the following distribution $\mu' : \binom{U}{k} \to \mathbb{R}_{\geq 0}$ from μ:

$$\mu'\left(\left\{i_1^{(j_1)}, \ldots, i_k^{(j_k)}\right\}\right) := \frac{\mu(\{i_1, \ldots, i_k\})}{t_1 \cdots t_k}.$$

We call μ' the *isotropic transformation* of μ. Another way we can define μ' is that to produce a sample from it, we first generate a sample $\{i_1, \ldots, i_k\}$ from μ, and then choose a copy $i_m^{(j_m)}$ for each element i_m uniformly at random.

Remark 3.2.2. Subdivision is mostly a tool for analysis. In practice, we do not have to formally perform subdivision; it suffices to sample elements proportionally to their marginals or marginal estimates.

Remark 3.2.3. To obtain the estimates $\{p_i\}$ for all i, we can apply the proof of Lemma 23 in [AD20], with ε constant, rather than $\varepsilon = O\left(\frac{1}{k}\right)$. This provides a runtime reduction for our preprocessing step, even in the case of log-concave polynomials.

There are three desirable properties of μ' we need to establish for subdivision to be an effective preprocessing step.

- Subdivision preserves $(1/\alpha)$-entropic independence (see Proposition 19 of [ADVY22]).
- The marginals $\Pr_{S\sim\mu'}[i^{(j)} \in S]$ are all close to $\frac{k}{|U|}$ for all $i^{(j)} \in U$; in other words, μ' is actually close to isotropic.
- $|U| \leq O(n)$, so if we ran a sampling algorithm on μ', the increased size of our ground set does not accidentally inflate our desired asymptotic running times.

Formally, we show:

Proposition 3.2.4. Let $\mu : \binom{n}{k} \to \mathbb{R}_{\geq 0}$, and let $\mu' : \binom{U}{k} \to \mathbb{R}_{\geq 0}$ be the subdivided distribution from Definition 3.2.1. The following hold for μ':

1. Near-isotropy: For all $i^{(j)} \in U$, the marginal $\Pr_{S\sim\mu'}[i^{(j)} \in S] \leq O\left(\frac{k}{|U|}\right)$.
2. Linear ground set size: The number of elements $|U| \leq O(n)$.

Proof. First, we verify that $|U|$ is at most $O(n)$:

$$|U| = \sum_{i=1}^{n} t_i \leq \sum_{i=1}^{n} \left(1 + \frac{n}{k} \cdot p_i\right) = n + \frac{n}{k} \sum_{i=1}^{n} p_i = 2n.$$

Next, we check that for any $i^{(j)}$, the marginal probabilities $\Pr_{S\sim\mu'}[i^{(j)} \in S]$ are at most $O\left(\frac{k}{|U|}\right)$. Here, we interpret the sampling from μ' as first sampling from μ, and then choosing a copy for each element.

$$\begin{aligned}\Pr_{S\sim\mu'}[i^{(j)} \in S] &= \sum_{S\ni i} \Pr[\text{we chose copy } j \mid \text{we sampled } S \text{ from } \mu] \cdot \Pr[\text{we sampled } S \text{ from } \mu] \\ &= \sum_{S\ni i} \frac{1}{t_i} \cdot \mu(S) = \frac{1}{t_i} \sum_{S\ni i} \mu(S) = \frac{1}{t_i} \cdot \Pr_{S\sim\mu}[i \in S]\end{aligned}$$

Since $t_i \geq \frac{n}{k} \cdot p_i \geq \frac{n}{k} \cdot \Omega\left(\Pr_{S\sim\mu}[i \in S]\right)$, we get that

$$\Pr_{S\sim\mu'}[i^{(j)} \in S] \leq O\left(\frac{\Pr_{S\sim\mu}[i \in S]}{\frac{n}{k} \cdot \Pr_{S\sim\mu}[i \in S]}\right) = O\left(\frac{k}{n}\right) \leq O\left(\frac{k}{|U|}\right). \qquad \square$$

3.2.2 Domain sparsification via Markov chain intermediate sampling

Here, we first describe, for any general distributions μ, a Markov chain based on generating intermediate samples $T \subseteq [n]$, that mixes to μ. Then, in Lemma 3.2.6 and Proposition 3.2.7, we state our main result that for distributions μ which are $(1/\alpha)$-entropically independent and nearly-isotropic, the size of T only needs to be $n^{1-\alpha} \cdot \mathsf{poly}(k)$ for the mixing to occur in one step.

For distribution $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$, consider the Markov chain M_μ^t defined for any positive integer t, with the state space $\mathrm{supp}(\mu)$. Starting from $S_0 \in \mathrm{supp}(\mu)$, one step of M_μ^t is:

1. Sample $T \sim \binom{[n]\setminus S_0}{t-k}$.
2. Downsample $S_1 \sim \mu|_{S_0 \cup T}$, and update S_0 to be S_1.

We note that the requirement $S_0 \in \mathrm{supp}(\mu)$ is not strictly necessary for this step to be defined.

Proposition 3.2.5. For any distribution $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$, the chain M_μ^t for $t \geq 2k$ is irreducible, aperiodic, and has stationary distribution μ.

Proof. Let P denote the transition probability matrix of M_μ^t. Since $t \geq 2k$, for any $S, S' \in \mathrm{supp}(\mu)$, there is a positive probability that we sample $T \supseteq (S \cup S')$. Thus, we have $P(S, S') > 0$, and P is both irreducible and aperiodic.

To derive the stationary distribution, suppose that we perform one step of the chain starting from $S_0 \sim \mu$. We first derive the distribution of the intermediate set $R := S_0 \cup T$.

For any $\tilde{R} \in \binom{[n]}{t}$, the probability of sampling $\tilde{R}$ for the intermediate set R is

$$\Pr[R = \tilde{R}] = \sum_{S_0 \in \binom{\tilde{R}}{k}} \mu(S_0) \cdot \Pr[T = \tilde{R} \setminus S_0] = \frac{1}{\binom{n-k}{t-k}} \cdot \mu(\tilde{R}).$$

Here, we are abusing notation and letting $\mu(\tilde{R}) = \sum_{S \subseteq \tilde{R}} \mu(S)$.

For any $\tilde{S}_1 \in \mathrm{supp}(\mu)$, the probability of sampling $\tilde{S}_1$ is

$$\begin{aligned}\Pr[S_1 = \tilde{S}_1] &= \sum_{\tilde{R}\in\binom{[n]}{r}} \Pr[S_1 = \tilde{S}_1 \mid R = \tilde{R}] \Pr[R = \tilde{R}] \\ &= \sum_{\tilde{R}\in\binom{[n]}{t}:\tilde{R}\supseteq\tilde{S}_1} \frac{\mu(\tilde{S}_1)}{\mu(\tilde{R})} \cdot \frac{1}{\binom{n-k}{t-k}} \cdot \mu(\tilde{R}) \\ &= \mu(\tilde{S}_1) \sum_{(\tilde{R}\setminus\tilde{S}_1)\in\binom{[n]\setminus\tilde{S}_1}{t-k}} \frac{1}{\binom{n-k}{t-k}} = \mu(\tilde{S}_1)\end{aligned}$$

Above, we summed over all $\tilde{R}$ that contain the target set $\tilde{S}_1$. □

The following lemma is the key to analyzing the sampling algorithm, since it quantifies the decrease in total variation distance after running one step of M_μ^t. It will be proven in Section 3.2.3.

Lemma 3.2.6. Let $\mu : \binom{[n]}{k}$ be a $(1/\alpha)$-entropically independent distribution. Suppose that for all $i \in [n]$, we have $\Pr_{S\sim\mu}[i \in S] \le \frac{Ck}{n}$. Then, for any constant $\varepsilon \in (0, \frac{1}{4}]$, and $t = \Omega\left(n^{1-\alpha}(Ck^2 \log \frac{1}{1-\varepsilon})^\alpha\right)$, the output S_1 of a single step of the Markov chain M_μ^t starting from S_0 satisfies

$$\forall S \in \binom{[n]}{k} : \Pr[S_1 = S] \ge \mu(S) \cdot (1 - \varepsilon)$$

Recall that if we used the marginal estimates required by Theorem 3.1.1, then by applying Proposition 3.2.4, we get an equivalent distribution μ' on a ground set of size $O(n)$ that satisfies the above assumption of $\Pr_{S\sim\mu'}[i \in S] \le \frac{Ck}{n}$ for some $C = O(1)$ (see Lemma 3.2.6).

Proposition 3.2.7. Let $\mu : \binom{[n]}{k}$ be a $(1/\alpha)$-entropically independent distribution, and let $\varepsilon \in (0, \frac{1}{4}]$ be a constant. Suppose $\Pr_{S\sim\mu}[i \in S] \le \frac{Ck}{n}$ for all i. Choose the intermediate sample size t according to Lemma 3.2.6. Then

$$\mathrm{d_{TV}}\left(P(S_0, \cdot), \mu\right) \le \varepsilon$$

and M_μ^t mixes to a distribution that has TV distance $\varepsilon' < \varepsilon$ from μ in $O\left(\log\left(\frac{1}{\varepsilon'}\right)\right)$ steps.

Proof. The bound on total variation distance follows via:

$$\begin{aligned}\mathrm{d_{TV}}\left(P(S_0, \cdot), \mu\right) &= \sum_{S\in\binom{[n]}{k}:\Pr[S_1=S]<\mu(S)} (\mu(S) - \Pr[S_1 = S]) \\ &\le \varepsilon \sum_{S\in\binom{[n]}{k}:\Pr[S_1=S]<\mu(S)} \mu(S) \le \varepsilon\end{aligned}$$

The mixing time bound follows from Theorem 2.1.1. □

We have shown that M_μ^t is fast mixing (in fact, mixing in one step for appropriately large t). Next, we show that for a wide class of distributions, namely, the class of α-fractionally-log-concave distributions with $\alpha = \Omega(1)$ (see [AASV21] for examples), each step of M_μ^t can be implemented in $\mathsf{poly}(n,k)$ time via a local Markov chain, i.e., the (muti-step) down-up random walk (see Definition 1 of [AASV21]).

Lemma 3.2.8 (Runtime analysis)**.** Suppose μ is α-fractionally-log-concave, and we start with $S_0^{(0)}$ such that $\mu(S_0^{(0)}) \geq 2^{-n^c}$ for some constant $c > 1$ and we run the chain for τ steps. The down-sampling takes time

$$O\left((t-k)^{\lceil 1/\alpha \rceil} k^{1/\alpha}\left(c\log n + \log\tau + \log\log\frac{1}{1-\varepsilon}\right)\right)$$

and the total runtime is

$$O\left(\tau(t-k)^{\lceil 1/\alpha \rceil} k^{1/\alpha}\left(c\log n + \log\tau + \log\log\frac{1}{1-\varepsilon}\right)\right).$$

Proof. We first show that with probability $\geq 1-\tau 2^{-n}$, for all $0 \leq i \leq \tau$, the i^{th}-step starting point, denoted by $S_0^{(i)}$, satisfies $\mu(S_0^{(i)}) \geq 2^{-(n^c+2ni)}$. This can be proven via induction on i. Conditioned on $\mu(S_0^{(i)}) \geq 2^{-n^c-2ni}$, we have

$$\begin{aligned}\Pr[\mu(S_0^{(i+1)}) \leq 2^{-(n^c+2(i+1)n)}] &= \mu(S_0^{(i)} \cup T)^{-1} \sum_{S \subseteq (S_0^{(i)} \cup T): \mu(S) \leq 2^{-(n^c+2(i+1)n)}} \mu(S) \\ &\leq_{(1)} \frac{2^{-(n^c+2(i+1)n)} \cdot 2^n}{\mu(S_0^{(i)})} \leq 2^{-n}.\end{aligned}$$

where in (1) we use the crude bound $\left|\left\{S \subseteq (S_0^{(i)} \cup T) : \mu(S) \leq 2^{-(n^c+2(i+1)n)}\right\}\right| \leq 2^n$.

Suppose that this good event happens, i.e.

$$\text{for all } i \in [0,\tau] : \mu(S_0^{(i)}) \geq 2^{-n^c-2ni}.$$

We observe that α-fractional-log-concavity is preserved by subdividing and restricting to a subset of the ground set ([AASV21]). In the down-sampling step, we run the (multi-step) down-up walk starting at $S_0^{(i)}$, and use Theorem 4 in [AJKPV21b] to bound the runtime and

obtain the desired result. To this end, we need to bound

$$\begin{aligned}\mathbb{E}_{T\sim\binom{[n]\setminus S_0^{(i)}}{t-k}}\left[\log\left(1+\log\frac{\mu(S_0^{(i)}\cup T)}{\mu(S_0^{(i)})}\right)\right] &\leq_{(1)} \log\left(1+\log\mathbb{E}_{T\sim\binom{[n]\setminus S_0^{(i)}}{t-k}}\left[\frac{\mu(S_0^{(i)}\cup T)}{\mu(S_0^{(i)})}\right]\right)\\ &= \log\left(1+\log\frac{1}{\Pr[S_1=S_0^{(i)}]}\right)\\ &\leq_{(2)} \log\left(1+\log\frac{1}{\mu(S_0^{(i)})(1-\varepsilon)}\right)\\ &\leq_{(3)} c\log n+\log\tau+\log\log\frac{1}{1-\varepsilon}.\end{aligned}$$

where (1) follows from Jensen's inequality for the concave function $f(x)=\log(1+\log(x))$ on $[1,\infty)$, (2) from Lemma 3.2.6 and (3) from lower bound on $\mu(S_0^{(i)})$. □

Remark 3.2.9. As a slight optimization, we can replace $(t-k)^{\lceil 1/\alpha\rceil}k^{1/\alpha}$ with $k^{\lceil 1/\alpha\rceil}(t-k)^{1/\alpha}$ when both μ and its complement μ^{comp} are α-fractionally-log concave. We can down-sample from $\mu^{\text{comp}}|_{S_0\cup T}$, then output the complement as S_1, where $\mu^{\text{comp}}:\binom{[n]}{n-k}\to\mathbb{R}_{\geq 0}$ is the complement of μ, defined by

$$\mu^{\text{comp}}([n]\setminus S):=\mu(S)\text{ for all }S\in\binom{[n]}{k}$$

In all important instances of α-fractionally-log-concavity, $\frac{1}{\alpha}\in\mathbb{N}$ and this optimization is unnecessary. The bound on total runtime can be simplified into $O(n^{1/\alpha-1}\mathsf{poly}(k,\log\frac{1}{\varepsilon}))$.

Advantage over rejection sampling

While we use a similar intermediate sampling framework as [AD20], our novel analysis of Markov chain intermediate sampling improves the runtime and applies to wider families of distributions. In order to fully understand the advantages realized by our intermediate sampling framework, we first need an overview of a rejection sampling-based implementation of intermediate sampling [Der19], which inspired the analysis of [AD20]. We then provide an example of $\frac{1}{2}$-log-concave distributions where Markov chain intermediate sampling succeeds using a smaller intermediate sample size than what is required for rejection sampling.

Let $S_0\in\text{supp}(\mu)$. One step of rejection sampling is given by:

1. Sample $T\sim\binom{[n]\setminus S_0}{t-k}$.
2. Accept the set $S_0\cup T$ with probability

$$\frac{\mu(S_0\cup T)}{\max_{T'\in\binom{[n]\setminus S_0}{t-k}}\mu(S_0\cup T')}.$$

3. Downsample $S_1 \sim \mu_{S_0 \cup T}$.

The key difference between rejection sampling and our Markov chain intermediate sampling algorithm is the *rejection step*, which is necessary if we want our chain to mix to the correct stationary distribution μ. In order to have a sufficiently large acceptance probability, and assuming μ is isotropic, we require that for all T,

$$\mu(T) \leq \left(\frac{t}{n}\right)^k \cdot (1+\varepsilon)^k.$$

Here, ε is a parameter related to the guarantee on $\mathrm{d_{TV}}\left(P(S_0, \cdot), \mu\right)$. Using this bound, we can ensure that the expected acceptance probability is $1 - O(\varepsilon k)$.

This inequality describes a "worst-case" condition on T. This "worst-case" type analysis originated from earlier works that introduced intermediate sampling for Determinantal Point Processes [Der19]. The proof of our worst-case inequality on $\mu(T)$ relies heavily on the fact that the KL divergence between a log-concave distribution μ and an arbitrary distribution ν contracts by a precise amount when applying the down operator $D_{k \to m}$:

$$\mathcal{D}_{\mathrm{KL}}(\nu D_{k\to\ell} \| \mu D_{k\to\ell}) \leq \frac{\ell}{k} \cdot \mathcal{D}_{\mathrm{KL}}(\nu \| \mu)$$

This contraction is well-known for log-concave distributions [CGM19], but does not hold with the factor ℓ/k for α-fractionally-log-concave distributions. On the other hand, the inequality we need to show (from the proof of Lemma 3.2.6) is "average-case" in nature, and when μ is isotropic, it takes the form:

$$\mathbb{E}_{T \sim \binom{[n]}{t}}[\mu(T)] \leq \left(\frac{t}{n}\right)^k \cdot \frac{1}{1-\varepsilon}$$

To concretely illustrate the advantage of Markov chain intermediate sampling, let us consider an example where the worst-case inequality fails to hold. Suppose that $k = 2$, n is even, and μ samples a set from $\left\{1, \frac{n}{2}+1\right\}, \left\{2, \frac{n}{2}+2\right\}, \ldots$ uniformly at random, so $\mu\left(\left\{i, \frac{n}{2}+i\right\}\right) = \frac{2}{n}$. This distribution is isotropic, $\frac{1}{2}$-sector stable [AASV21], and $\frac{1}{2}$-fractionally-log-concave, and yet, according to the worst-case analysis, it does not yield enough acceptance probability when $t = o(n)$. For any set T, we have

$$\mu(T) \leq \frac{t}{2} \cdot \frac{2}{n} = \frac{t}{n}.$$

Equality is achieved by selecting a subset T that contains as many pairs of the form $\{i, \frac{n}{2}+i\}$ as possible, i.e., at least $(t-1)/2$. Thus, the worst-case analysis would suggest that no non-trivial intermediate sampling is possible for the distribution μ; this is because $\frac{t}{n} \gg (t/n)^2$ for small values of t.

However, our relaxed average-case analysis captures the fact that realistically, not every element of T will be paired up. In fact, we expect only a constant number of pairs when $t = O(\sqrt{n})$, so for this example, we have:

$$\mathbb{E}_{T \sim \binom{[n]}{t}}[\mu(T)] \leq C \cdot \frac{2}{n} \leq O\left(\frac{1}{n}\right) = O\left(\frac{t^2}{n^2}\right).$$

3.2.3 Proof of Lemma 3.2.6

In this section, we will prove Lemma 3.2.6.

Lemma 3.2.10. Let U, V be a sets of size $u, v \leq k$ respectively with $U \cap V = \emptyset$. We have:

$$\left(\frac{t-(u+v)}{n-(u+v)}\right)^u \leq \Pr_{T\in\binom{[n]\setminus V}{t-v}}[U \subseteq T] \leq \left(\frac{t-v}{n-v}\right)^u.$$

Proof. Since we are sampling the elements of T uniformly at random from $[n]$,

$$\Pr_{T\in\binom{[n]\setminus V}{t-v}}[U \subseteq T] = \frac{\binom{n-v-u}{t-v-u}}{\binom{n-v}{t-v}} = \frac{(t-v)(t-v-1)\cdots(t-v-u+1)}{(n-v)(n-v-1)\cdots(n-v-u+1)} \leq \left(\frac{t-v}{n-v}\right)^u.$$

Similarly, we also have:

$$\Pr_{T\in\binom{[n]\setminus V}{t-v}}[U \subseteq T] \geq \left(\frac{t-v-u+1}{n-v-u+1}\right)^u \geq \left(\frac{t-(u+v)}{n-(u+v)}\right)^u.$$

□

Proof of Lemma 3.2.6. Let $R = S_0 \cup S$, and let $r = |S_0 \cup S|$. Note that $|S \setminus S_0| = r - k$ and

$$\begin{aligned}
\Pr[S_1 = S] &= \mathbb{E}_{T\sim\binom{[n]\setminus S_0}{t-k}}\left[\frac{\mu(S)}{\sum_{S'\subseteq(S_0\cup T)}\mu(S')} \mid S \subseteq T\right] \cdot \Pr_{T\sim\binom{[n]\setminus S_0}{t-k}}[(S\setminus S_0)\subseteq T] \\
&= \mathbb{E}_{T'\sim\binom{[n]\setminus R}{t-r}}\left[\frac{\mu(S)}{\sum_{S'\subseteq(R\cup T')}\mu(S')}\right] \cdot \Pr_{T\sim\binom{[n]\setminus S_0}{t-k}}[(S\setminus S_0)\subseteq T] \\
&\geq_{(1)} \frac{\mu(S)}{\mathbb{E}_{T'\sim\binom{[n]\setminus R}{t-r}}\left[\sum_{S'\subseteq(R\cup T')}\mu(S')\right]} \cdot \Pr_{T\sim\binom{[n]\setminus S_0}{t-k}}[(S\setminus S_0)\subseteq T] \\
&\geq_{(2)} \frac{\mu(S)}{\mathbb{E}_{T'\sim\binom{[n]\setminus R}{t-r}}\left[\sum_{S'\subseteq(R\cup T')}\mu(S')\right]} \cdot \left(\frac{t-r}{n-r}\right)^{r-k}.
\end{aligned}$$

Inequality (1) is an application of Jensen's inequality to the function $f(x) = \frac{c}{x}$, which is convex when $x > 0$. Inequality (2) use Lemma 3.2.10 with $U = (S \setminus S_0)$ and $V = S_0$.

Now if we bound $\mathbb{E}_{T'\sim\binom{[n]\setminus R}{t-r}}\left[\sum_{S'\subseteq(R\cup T')}\mu(S')\right]$ by $(\frac{t-r}{n-r})^{r-k}\cdot\frac{1}{1-\varepsilon}$, then we are done.

$$\begin{aligned}
\mathbb{E}_{T'\sim\binom{[n]\setminus R}{t-r}}\left[\sum_{S'\subset(T'\cup R)}\mu(S')\right] &= \sum_{S'\in\binom{[n]}{k}}\mu(S')\cdot\Pr_{T'\sim\binom{[n]\setminus R}{t-r}}[(S'\setminus R)\subseteq T'] \\
&\leq \sum_{S'\in\binom{[n]}{k}}\mu(S')\cdot\left(\frac{t-r}{n-r}\right)^{|S'\setminus R|}.
\end{aligned}$$

In the very last line, we applied Lemma 3.2.10 with $U = S' \setminus R$ and $V = R$. If we set

$$z_i = \begin{cases} \left(\frac{n-r}{t-r}\right)^{1/\alpha} & \text{if } i \in (S_0 \cup S) \\ 1 & \text{otherwise} \end{cases}$$

then we can rewrite

$$\begin{aligned} \sum_{S' \in \binom{[n]}{k}} \mu(S') \cdot \left(\frac{t-r}{n-r}\right)^{|S' \setminus (S_0 \cup S)|} &= \left(\frac{t-r}{n-r}\right)^{k} \sum_{S' \in \binom{[n]}{k}} \mu(S') \cdot \left(\frac{n-r}{t-r}\right)^{|S' \cap (S_0 \cup S)|} \\ &= \left(\frac{t-r}{n-r}\right)^{k} \cdot g_\mu\left(z_1^\alpha, \ldots, z_n^\alpha\right) \end{aligned}$$

Applying Equation 2.1 to $g_\mu(z_1^\alpha, \ldots, z_n^\alpha)$ and noting that $p_i = \frac{\Pr_{S \sim \mu}[i \in S]}{k}$, we obtain

$$\begin{aligned} g_\mu\left(z_1^\alpha, \ldots, z_n^\alpha\right) &\leq \left(\sum_{i=1}^{n} \frac{\Pr_{S \sim \mu}[i \in S]}{k} \cdot z_i\right)^{k\alpha} \\ \log g_\mu\left(z_1^\alpha, \ldots, z_n^\alpha\right) &\leq k\alpha \log\left(\sum_{i=1}^{n} \frac{\Pr_{S \sim \mu}[i \in S]}{k} \cdot z_i\right) \\ &\leq_{(1)} k\alpha \left(-1 + \sum_{i=1}^{n} \frac{\Pr_{S \sim \mu}[i \in S]}{k} \cdot z_i\right) \\ &=_{(2)} k\alpha \left(\sum_{i=1}^{n} \frac{\Pr_{S \sim \mu}[i \in S]}{k} \cdot (z_i - 1)\right) \\ &= \alpha \sum_{i=1}^{n} \Pr_{S \sim \mu}[i \in S] \cdot (z_i - 1), \end{aligned}$$

where in (1), we use $\log x \leq x - 1$ for $x \in (0, \infty)$ and in (2) we use $\sum_{i=1}^{n} \frac{\Pr_{S \sim \mu}[i \in S]}{k} = 1$. Substituting z_i as specified above into the final inequality, we get

$$\begin{aligned} \log g_\mu\left(z_1^\alpha, \ldots, z_n^\alpha\right) &\leq \alpha \sum_{i \in (S_0 \cup S)} \frac{Ck}{n} \cdot \left(\frac{n-r}{t-r}\right)^{1/\alpha} \\ &= \frac{C\alpha k r}{n} \left(\frac{n-r}{t-r}\right)^{1/\alpha} \\ &\leq \frac{2Ck^2}{n} \left(\frac{n-r}{t-r}\right)^{1/\alpha} \end{aligned}$$

□

3.3 Lower bound on intermediate sampling

We first show that the dependence of our domain sparsification analysis on n is optimal. Consider the uniform distribution μ_0 over singletons of a ground set of $\frac{n}{k}$ elements. Any distribution on singletons ($k = 1$) is log-concave as the generating polynomial is linear, and thus, log-concave.

Now apply the construction of Example 2.4.12 with $m = k$ to μ_0 to obtain a new distribution μ on $\binom{[n]}{k}$. This distribution is uniform over the parts of a particular partition of the ground set $[n]$ into $\frac{n}{k}$ sets $S_1, \ldots, S_{n/k}$. Note that this distribution is also isotropic and k-entropically independent.

If we sample a uniformly random set T of size t, then the chance that S_i is contained in T can be upperbounded by

$$\binom{n-k}{t-k} \cdot \binom{n}{t}^{-1} \simeq \left(\frac{t}{n}\right)^k.$$

Thus, the chance that any of the S_i are contained in T can be upper bounded (via a union bound) by roughly

$$n \cdot \left(\frac{t}{n}\right)^k.$$

Thus, as long as $t \ll n^{1-1/k}$, the above is negligible. Without having any S_i in the support with high probability, we cannot faithfully produce a sample of μ from subsets of T.

Next we construct an example showing that even higher-order marginals cannot remove this dependence on n for entropically independent distributions (in sharp contrast with Conjecture 7.1.1). Our construction is based on Reed-Solomon codes.

Lemma 3.3.1. Let q be a prime number and $\mathbb{F}_q$ the finite field of size q. Fix k points $\{x_1, \ldots, x_k\} \subseteq \mathbb{F}_q$ where k is a constant and choose a set of k random permutations from $\mathbb{F}_q$ to $\mathbb{F}_q$ and call them $\pi_1, \ldots, \pi_k$. Let $\mu : \binom{\Omega}{k} \to \mathbb{R}_{\geq 0}$ be the uniform distribution over sets $\{(x_i, y_i) : i \in [k]\}$ such that $p(x_i) = \pi_i(y_i)$ for some polynomial p of $\deg(p) \leq d < k$. The ground set Ω is $\{x_1, \ldots, x_k\} \times \mathbb{F}_q$. Then

1. μ satisfies $(1/\alpha)$-entropic independence, with $\alpha = \frac{d+1}{k}$.
2. Any domain sparsification scheme to sample from μ requires $t = \tilde{\Omega}(n^{1-\alpha})$, even when we are allowed to sample higher order marginals.

Proof. The distribution $\mu D_{k\to(d+1)}$ is uniform over $\{(x_j, y_j) : j \in J \subseteq [k], |J| = (d+1)\}$, because for any such set, there exists a unique polynomial p of degree at most d such that $p(x_j) = \pi_j(y_j)$ for all $j \in J$. Thus, high-order marginals, up to order $(d+1)$, are independent of the choice of permutations $\pi_1, \ldots, \pi_k$.

The support of $\mu D_{k\to(d+1)}$ forms the basis of a *partition matroid*. For each $x = x_i$, we have a block consisting of all points $\{(x, y) : y \in \mathbb{F}_q\}$, and for each set in the support of

$\mu D_{k\to(d+1)}$, we have at most one element per block. Since we have a uniform distribution over matroid bases, $\mu D_{k\to(d+1)}$ is log-concave, and thus it satisfies 1-entropic independence.

We use this to upper bound $\mathcal{D}_{\mathrm{KL}}(\nu D_{k\to 1}\|\mu D_{k\to 1})$, and from here, conclude $\frac{d+1}{k}$-entropic independence of μ:

$$\begin{aligned}\mathcal{D}_{\mathrm{KL}}(\nu D_{k\to 1}\|\mu D_{k\to 1}) &= \mathcal{D}_{\mathrm{KL}}\left(\nu D_{k\to(d+1)})D_{(d+1)\to 1}\|(\mu D_{k\to(d+1)})D_{(d+1)\to 1}\right)\\ &\le \frac{1}{d+1}\cdot\mathcal{D}_{\mathrm{KL}}\left(\nu D_{k\to(d+1)}\|\mu D_{k\to(d+1)}\right)\\ &\le \frac{1}{d+1}\cdot\mathcal{D}_{\mathrm{KL}}(\nu\|\mu) = \frac{1}{\frac{d+1}{k}\cdot k}\cdot\mathcal{D}_{\mathrm{KL}}(\nu\|\mu).\end{aligned}$$

The second line follows from $\mu D_{k\to(d+1)}$ satisfying 1-entropic independence, and the third line comes from the data processing inequality.

We now prove that for all $t \le o\left(n^{1-\alpha}\right)$, no domain sparsification scheme exists, even with access to higher order marginals. For $d' \le (d+1)$, the distribution $\mu D_{k\to d'}$ is uniform over the size d' independent sets of the partition matroid defined above. One consequence of the independence of high-order marginals from the choice of permutations $\pi_1,\ldots,\pi_k$ is that the higher order marginals do not provide any information about the identity of μ.

Suppose that we choose our sample in domain sparsification from a distribution whose ground set is the sparse subset T. We want an upper bound on the probability (over the choice of permutations) that T contains some $S \in \mathrm{supp}(\mu)$.

In order for T to contain a valid S, there must be some subset in $S \in \binom{T}{k}$ associated to a degree $\le d$ polynomial p satisfying $p(x_i) = \pi_i(y_i)$. However it is easy to see that for any particular set S the probability (over the choice of permutations) that S is in the support of μ is $\simeq 1/q^{k-d-1}$. We can upper bound $\Pr[S \subseteq T \text{ for some } S \in \mathrm{supp}(\mu)]$ as follows:

$$\binom{t}{k}\cdot\frac{1}{q^{k-d-1}} \le \frac{t^k}{q^{k-d-1}}.$$

This implies that for any $t \le o\left(q^{(k-d-1)/k}\right) \le o\left(n^{(k-d-1)/k}\right)$, the probability of containing a set in the support is negligible. Note that we have $\alpha = \frac{d+1}{k}$, so $1-\alpha = \frac{k-d-1}{k}$, which completes the lower bound. □

Chapter 4

Sphere Caps and Random Graphs

As if the street lights pointed in an arrowhead
Leading us home [Swi19]

In this chapter, we introduce some of the preliminaries required for Chapters 5 and 6. In particular, we have amassed an extensive collection of concentration inequalities about sphere caps in high dimensions that may have use cases outside of our work.

4.1 Sphere caps and dot products of unit vectors

Here, we introduce some useful bounds on the measure of sphere caps, and a concentration bound on the dot products of two unit vectors.

Definition 4.1.1. Fix $x \in \mathbb{S}^{d-1}$. For $p \in [0, \frac{1}{2}]$, we let $\tau(p, d)$ be the threshold τ at which

$$\Pr_{\boldsymbol{y} \sim \rho}[\langle x, \boldsymbol{y} \rangle \geq \tau] = p$$

Definition 4.1.2 (p-cap)**.** For a vector $v \in \mathbb{S}^{d-1}$, its *p-cap* is

$$\mathsf{cap}_p(v) := \{x \in \mathbb{S}^{d-1} : \langle v, x \rangle \geq \tau(p, d)\}$$

Similarly, we define its *p-anticap* as

$$\overline{\mathsf{cap}}_p(v) := \{x \in \mathbb{S}^{d-1} : \langle v, x \rangle < \tau(p, d)\}$$

We drop the p in the subscript when its value is clear from context.

4.1.1 Fine-grained estimates of the inner product distribution

For a refresher on some of the notation related to $\mathbb{S}^{d-1}$ and the distribution ρ, see Section 1.3. In order to study sphere caps, it is natural that we need a good handle on the density ψ_d, which describes the distribution of $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ when $\boldsymbol{x}, \boldsymbol{y} \sim \rho$.

$$\psi_d(x) = \frac{\Gamma\left(\frac{d}{2}\right)}{\Gamma\left(\frac{d-1}{2}\right)\sqrt{\pi}} \cdot \left(1 - x^2\right)^{(d-3)/2}.$$

ψ_d is supported on $[-1, 1]$. We'll overload notation a bit and also use ψ_d when we reference the distribution described by the density.

Remark 4.1.3. $\psi_d(x)$ also describes the distribution of $\langle x, \boldsymbol{y} \rangle$ for fixed $x \in \mathbb{S}^{d-1}$ when $\boldsymbol{y} \sim \rho$. It is invariant under choice of x.

We use Z_d to denote the normalizing constant $\frac{\Gamma\left(\frac{d}{2}\right)}{\Gamma\left(\frac{d-1}{2}\right)\sqrt{\pi}}$. The quantity Z_d shows up in [LMSY22a]; though it won't be needed in this thesis, we include Fact 4.1.4 for completeness.

Fact 4.1.4. $Z_d \leq O(\sqrt{d})$.

Using elementary techniques, we can get a sharp estimate of the tail of ψ_d.

Lemma 4.1.5. Let $\Phi_{\psi_d}(t) := \Pr_{X \sim \psi_d}[X \leq t]$ be the CDF of ψ_d, and let $\overline{\Phi}_{\psi_d}(t) = 1 - \Phi_{\psi_d}(t)$. Then, for $t \geq 0$:

$$\frac{Z_d}{t(d-1)} \cdot \left(1 - t^2\right)^{(d-1)/2} \cdot \left(1 - \frac{4\log\left(1 + d \cdot t^2\right)}{d \cdot t^2}\right) \leq \overline{\Phi}_{\psi_d}(t) \leq \frac{Z_d}{t(d-1)} \cdot \left(1 - t^2\right)^{(d-1)/2}$$

Proof. It suffices to upper and lower bound $\int_t^1 (1 - x^2)^{(d-3)/2}$. We first upper bound:

$$\begin{aligned}
\int_t^1 \left(1 - x^2\right)^{(d-3)/2} dx &= \frac{1}{t} \int_t^1 t \left(1 - x^2\right)^{(d-3)/2} dx \\
&\leq \frac{1}{t} \int_t^1 x \left(1 - x^2\right)^{(d-3)/2} dx \\
&= -\frac{1}{t(d-1)} \cdot \left(1 - x^2\right)^{(d-1)/2} \Bigg|_t^1 \\
&= \frac{1}{t(d-1)} \cdot \left(1 - t^2\right)^{(d-1)/2}
\end{aligned}$$

Now we prove the lower bound. For any $\varepsilon > 0$ such that $t \cdot \sqrt{1 - \varepsilon + \frac{\varepsilon}{t^2}} \leq 1$, we define $\delta := \frac{\varepsilon}{t^2} - \varepsilon$. Then:

$$\begin{aligned}
\int_t^1 (1-x^2)^{(d-3)/2} dx &\geq \frac{1}{t\sqrt{1+\delta}} \int_t^{t\sqrt{1+\delta}} \left(t\sqrt{1+\delta}\right) \left(1-x^2\right)^{(d-3)/2} dx \\
&\geq \frac{1-\delta}{t} \int_t^{t\sqrt{1+\delta}} x \left(1-x^2\right)^{(d-3)/2} dx \\
&= -\frac{1-\delta}{t(d-1)} \cdot \left(1-x^2\right)^{(d-1)/2} \Bigg|_t^{t\sqrt{1+\delta}} \\
&= \frac{1-\delta}{t(d-1)} \cdot \left(1-t^2\right)^{(d-1)/2} \cdot \left(1-(1-\varepsilon)^{(d-1)/2}\right)
\end{aligned}$$

where the second line uses $\frac{1}{\sqrt{1+\delta}} \geq 1-\delta$ and the last equality uses $1-t^2(1+\delta) = (1-t^2)(1-\varepsilon)$. Choosing $\varepsilon = \frac{2\log\left(1+dt^2\right)}{d-1}$ yields:

$$\int_t^1 \left(1-x^2\right)^{(d-3)/2} \geq \frac{1}{t(d-1)} \cdot \left(1-t^2\right)^{(d-1)/2} \cdot \left(1 - \frac{4\log\left(1+dt^2\right)}{dt^2}\right)$$

□

We also have the following constant factor approximation for the measure of a sphere cap in terms of the threshold $\tau(p,d)$, when $\tau(p,d) \geq \sqrt{\frac{2}{d}}$ given by Lemma 2.1(b) of [Bri+01]:

Theorem 4.1.1. Consider a p-cap where $\tau(p,d) \geq \sqrt{\frac{2}{d}}$. Then:

$$\frac{1}{6\tau(p,d)\sqrt{d}} \left(1-\tau(p,d)^2\right)^{(d-1)/2} \leq p \leq \frac{1}{2\tau(p,d)\sqrt{d}} \left(1-\tau(p,d)^2\right)^{(d-1)/2}$$

We lastly present a lemma from [BBN20] that contains additional facts about ψ_d.

Lemma 4.1.6 (Lemma 5.1 of [BBN20])**.** Let $\tau(p,d)$ and ψ_d be defined as above. Then:

1. For $0 \leq \tau \leq \frac{1}{2}$ and $\delta > 0$, we have:

$$\frac{\psi_d(\tau-\delta)}{\psi_d(\tau)} \leq \exp(2\tau d\delta).$$

2. $\psi_d(\tau) \leq C_{4.1.6} \cdot p \cdot \max\{\sqrt{d}, d\tau\}$ for a universal constant $C_{4.1.6}$.

4.1.2 Simpler estimates for sphere cap measures

In this section, we present some cruder estimates of sphere cap measures (alternatively, tail bounds for ψ_d) that are often good enough for many of our results.

Recall Lévy's theorem for concentration of measure on the unit sphere (Theorem 14.1.1 of [Mat13]). We use it to upper bound the measure of a sphere cap with threshold τ.

Lemma 4.1.7. Let $y \in \mathbb{S}^{d-1}$ be any vector. Then,

$$\Pr_{\boldsymbol{w}\sim\rho}[|\langle \boldsymbol{w}, y\rangle| \geq \tau] \leq 4\exp(-\tau^2 d/2)$$

This is also equivalent to a tail bound on ψ_d:

$$\overline{\Phi}_{\psi_d}(\tau) \leq 4\exp(-\tau^2 d/2)$$

While this upper bound is off from the upper bound in Lemma 4.1.5 by a factor of $\sqrt{d}$, we see the same sub-Gaussian tail present in both bounds.

We now present a convenient upper bound on the dot product threshold $\tau(p, d)$ of a p-cap.

Lemma 4.1.8. For any $p \leq \frac{1}{2}$, we have $\tau(p, d) \leq \sqrt{\frac{2\log(1/p)}{d}}$.

Proof. We case on whether $\tau(p, d)$ is smaller or larger than $\sqrt{\frac{2}{d}}$. In the first case observe that $\log(1/p) \geq 1$ by our bound on p, so $\tau(p, d) \leq \sqrt{\frac{2}{d}} \leq \sqrt{\frac{2\log(1/p)}{d}}$.

When $\tau(p, d) \geq \sqrt{\frac{2}{d}}$, we use the upper bound in Theorem 4.1.1. Let $\tau' = \sqrt{\frac{2\log(1/p)}{d}}$, and let p' denote its corresponding tail probability $\Pr_{x,y\in\mathbb{S}^{d-1}}[\langle x, y\rangle \geq \tau']$. If we can show $p' \leq p$, then we know $\tau(p, d) \leq \tau'$.

$$\begin{aligned} p' &\leq \frac{1}{2\tau'\sqrt{d}} \cdot \left(1 - t^2\right)^{(d-1)/2} = \frac{1}{2\sqrt{\log(1/p)}} \cdot \left(1 - \frac{2\log(1/p)}{d}\right)^{(d-1)/2} \\ &\leq \frac{1}{2\sqrt{\log(1/p)}} \cdot p \leq p \end{aligned}$$

Since $p' \leq p$, this tells us $\tau \leq \tau'$ as well, giving us the desired inequality. □

The next lemma helps us understand the deviations in cap volume p when we make small adjustments to its dot product threshold $\tau(p, d)$.

Lemma 4.1.9. Fix $x \in \mathbb{S}^{d-1}$. Let $\tau \geq 0$ and $p := \Pr_{\boldsymbol{z}\sim\rho}[\langle x, \boldsymbol{z}\rangle \geq \tau]$. For any $\varepsilon \geq 0$, there is a universal constant $C_{4.1.9}$ such that:

$$\Pr_{\boldsymbol{z}\sim\rho}[\tau - \varepsilon \leq \langle x, \boldsymbol{z}\rangle \leq \tau + \varepsilon] \leq p \cdot \left(C_{4.1.9} \cdot \varepsilon \cdot \exp(2d\tau\varepsilon) \cdot \sqrt{d\log(1/p)}\right)$$

Understanding the joint distribution of two independently chosen unit vectors in $\mathbb{S}^{d-1}$ helps complete the proof of this lemma.

Proof. When $t \geq 0$, the density $\psi_d(t)$ is a decreasing function in t. Thus:

$$\begin{aligned}\Pr_{\boldsymbol{z}\sim\rho}[\tau-\varepsilon \leq \langle x, \boldsymbol{z}\rangle \leq \tau+\varepsilon] &= \int_{\tau-\varepsilon}^{\tau+\varepsilon} \psi_d(t)dt \\ &\leq (2\varepsilon)\cdot[\psi_d(\tau)\cdot\exp(2d\tau\varepsilon)]\end{aligned}$$

In the last line, we used Part 1 of Lemma 4.1.6, and noted that this is an upper bound even when $\varepsilon > \tau$. Using Part 2 of Lemma 4.1.6, and Lemma 4.1.8, we can upper bound $\psi_d(\tau)$:

$$\psi_d(\tau) \leq C_{4.1.6}p\cdot\max\{\sqrt{d}, d\tau\} \leq p\cdot\left(C_{4.1.6}\sqrt{2}\cdot\sqrt{d}\cdot\sqrt{\log(1/p)}\right)$$

from which the desired bound follows. □

4.2 Basic results about random graphs

The core problem solved in Chapter 5 is to distinguish between two random graph models: the widely studied Erdös-Rényi model, and the random geometric graph. We first provide some basic facts about both of these distributions.

- For a graph $G = (V, E)$, and a subset of vertices $S \subseteq V$, we use $G[S]$ to denote the induced subgraph of G on S.
- $B_G(v, \ell)$ to denotes the ball of radius-ℓ around a vertex v in graph G (i.e. all vertices of distance $\leq \ell$ from v).
- $N_G(i)$ denotes the neighbors of vertex i in graph G.

Definition 4.2.1 ($\mathsf{G}(n,p)$)**.** To sample a graph $G = (V, E)$ from the *Erdös-Rényi* distribution, denoted $\mathsf{G}(n,p)$, we let $V = [n]$ and for each pair of vertices (i, j), we place $(i, j) \in E$ with probability p.

Definition 4.2.2 ($\mathsf{Geo}_d(n,p)$)**.** Given a set of vectors $V = (v_1, \ldots, v_n) \in \left(\mathbb{S}^{d-1}\right)^n$ the associated *geometric graph* $\mathsf{gg}(V,p) = ([n], E)$ is given by choosing the edge set as all (i, j) where $\langle v_i, v_j\rangle \geq \tau(p, d)$. In particular, when $\boldsymbol{V} \sim \rho^{\otimes n}$, $\mathsf{gg}(\boldsymbol{V}, p)$ is distributed as $\mathsf{Geo}_d(n,p)$.

For an n-vertex graph G, ρ^G denotes the conditional distribution $\rho^G = \rho^{\otimes n} \mid \mathsf{gg}(V,p) = G$.

First, we state a high-probability upper bound of $O\left(\frac{\log n}{\log\log n}\right)$ on the maximum degree in both random graph models when $p = \frac{\alpha}{n}$ for constant α. Since the degree of each vertex is distributed like $\mathsf{Binom}(n,p)$, by applying the standard tail bound for a Binomial random variable and taking a union bound over all vertices we get the following.

Lemma 4.2.3. Let $\Delta(G)$ denote the maximum degree of a graph G. If $p = \frac{\alpha}{n}$, then for both $\boldsymbol{G} \sim \mathsf{G}(n,p)$ and $\boldsymbol{G} \sim \mathsf{Geo}_d(n,p)$ and for all d:

$$\Pr[\Delta(\boldsymbol{G}) \geq k] \leq n \cdot \left(\frac{k}{e\alpha}\right)^{-k}.$$

Remark 4.2.4. This allows us to ignore graphs where the max degree is too high when we upper bound $\mathrm{d_{TV}}\left(\mathsf{G}(n,p), \mathsf{Geo}_d(n,p)\right)$, since such graphs comprise a negligible fraction of each distribution.

More generally, we know the following bound on the number of vertices in an Erdös-Rényi graph at distance at most ℓ from any given vertex, as well as a high probability statement about their structure.

Lemma 4.2.5 (Lemma 29 of [BLM15])**.** Let $\boldsymbol{G} \sim \mathsf{G}(n,p)$ for $p = \frac{\alpha}{n}$ for constant α. Define $\boldsymbol{h}_t(v)$ as the number of vertices with distance exactly t from v in $\boldsymbol{G}$. Then for any vertex v, there are constants c, C such that:

$$\Pr\left[\text{there exists } t \geq 0 : \boldsymbol{h}_t(v) > s\alpha^t\right] \leq C\exp(-cs).$$

Lemma 4.2.6 (Lemma 30 of [BLM15])**.** Let $\boldsymbol{G} \sim \mathsf{G}(n,p)$ for $p = \frac{\alpha}{n}$ for constant α, and let $B_{\boldsymbol{G}}(v,t)$ be the set of all vertices with distance $\leq t$ from vertex v in $\boldsymbol{G}$. Then, for any vertex v, there is a constant c' such that:

$$\Pr[B_{\boldsymbol{G}}(v,t) \text{ is not a tree}] \leq \frac{c'\alpha^t}{n}$$

Given an n-vertex graph G, we are interested in $\boldsymbol{V} \sim \rho^{\otimes n} | \mathsf{gg}(\boldsymbol{V}, p) = G$, whose distribution we shorten to ρ^G. A simple but crucial observation for us is the following.

Observation 4.2.7. Let $f : \left(\mathbb{S}^{d-1}\right)^n \to \{0,1\}$ be a Boolean function. If $\mathbb{E}_{\boldsymbol{V}\sim\rho^{\otimes n}} f(\boldsymbol{V}) \leq \delta$, then:

$$\Pr_{\boldsymbol{G}\sim\mathsf{Geo}_d(n,p)}\left[\mathbb{E}_{\boldsymbol{V}\sim\rho^{\boldsymbol{G}}} f(\boldsymbol{V}) \geq \sqrt{\delta}\right] \leq \sqrt{\delta}.$$

Proof. We can write:

$$\mathbb{E}_{\boldsymbol{V}\sim\rho^{\otimes n}} f(\boldsymbol{V}) = \mathbb{E}_{\boldsymbol{G}\sim\mathsf{Geo}_d(n,p)} \mathbb{E}_{\boldsymbol{V}\sim\rho^{\boldsymbol{G}}}\left[f(\boldsymbol{V})\right] \leq \delta$$

The statement then follows from Markov's inequality. □

To conclude these preliminaries, we state one more result that we can find a coupling that "sandwiches" a random geometric graph with edge probability p between two Erdös-Rényi graphs with edge probabilities slightly above and slightly below p. The full proof of this result is in Section 6 of [LMSY22b].

Proposition 4.2.8. Let n, d be positive integers, and let $p \in \mathbb{R}_{\geq 0}$, satisfying $\frac{\alpha}{n} \ll p \leq \frac{1}{2}$ for α a fixed constant, $(nH(p) + \ln^2 n)^3 \ln n \ll d \leq n^{100}$. Let

$$\varepsilon > C \cdot \sqrt{\frac{n \cdot H(p) + \log^2 n}{d}}$$

for $C > 0$ a universal constant. Then there exists a three-way "coupling" $\mathcal{C}$ of

$$\mathsf{G}(n, (1-\varepsilon)p), \mathsf{Geo}_d(n, p), \mathsf{G}(n, (1+\varepsilon)p)$$

so that with probability at least $1 - n^{-10}$ over $(\boldsymbol{G}_-, \boldsymbol{G}, \boldsymbol{G}_+) \sim \mathcal{C}$, with $\boldsymbol{G}_- \sim \mathsf{G}(n, (1-\varepsilon)p)$, $\boldsymbol{G} \sim \mathsf{Geo}_d(n, p)$, and $\boldsymbol{G}_+ \sim \mathsf{G}(n, (1+\varepsilon)p)$, it holds that

$$\boldsymbol{G}_- \subseteq \boldsymbol{G} \subseteq \boldsymbol{G}_+$$

Chapter 5

Testing for Geometry in Graphs

> Well, I thought you might be different than the rest
> I guess you're all the same [Swi21]

This chapter is based on joint work with Siqi Liu, Sidhanth Mohanty, and Tselil Schramm, published in [LMSY22b].

5.1 Problem statement and summary of results

The Erdös-Rényi graph model (Definition 4.2.1, and reproduced below) has arisen as a valuable tool for combinatorialists and theoretical computer scientists alike.

Definition 5.1.1 (Restatement of Definition 4.2.1)**.** To sample a graph $G = (V, E)$ from the *Erdös-Rényi* distribution, denoted $\mathsf{G}(n, p)$, we let $V = [n]$ and for each pair of vertices (i, j), we place $(i, j) \in E$ with probability p.

Due to the independence between edges, it is a very tractable model to study for establishing theoretical results. Graphs sampled from the Erdös-Rényi model have been employed in conjunction with the probabilistic method, most notably to to prove the existence of graphs with simultaneously high girth and high chromatic number (see [Erd59; AS04]), and have also been used in the average-case analysis of graph algorithms (see [FK01; BCCFV10]).

However, the assumption that edges are placed independently makes Erdös-Rényi graph unattractive for modeling many real-world networks. For instance, they do not capture networks where the vertex degree distribution exhibits a power law (such as the web graph); the Barabasi-Albert model ([AB02]), which implements a preferential attachment mechanism, is instead able to capture this empirically observed behavior. Erdös-Rényi graphs also have a low *clustering coefficient*, which is counter to what we see in social networks; the Watts-Strogatz model ([WS98]) attempts to remedy this.

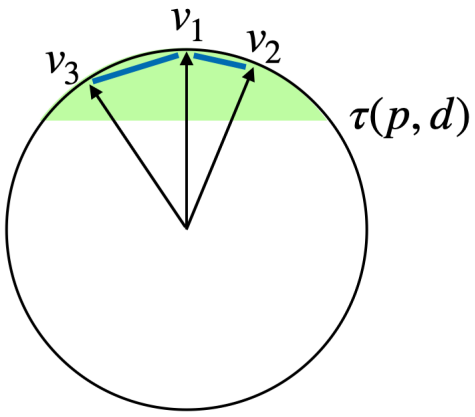

Figure 5.1: A demonstration of triadic closure with d is small: The threshold $\tau(p, d)$ is high, and v_1 being close to v_2, v_3 also forces v_2, v_3 to be closer.

It is also natural to model random networks with an underlying geometry in mind. Almost all of the data we see today, including network data, can be embedded into some high-dimensional space, with the coordinates representing different features about the data. It is then natural to place edges between nodes based on the proximity of their embeddings. While such embedding-inspired graphs may be difficult to analyze in full generality, as a first step towards the theory of random graphs, we assume all of the vectors are embedded on a high-dimensional unit sphere (Definition 4.2.2).

Definition 5.1.2 (Restatement of Definition 4.2.2)**.** Given a set of vectors $V = (v_1, \dots, v_n) \in \left(\mathbb{S}^{d-1}\right)^n$ the associated *geometric graph* $\mathsf{gg}(V, p) = ([n], E)$ is given by choosing the edge set as all (i, j) where $\langle v_i, v_j \rangle \geq \tau(p, d)$. In particular, when $\boldsymbol{V} \sim \rho^{\otimes n}$, $\mathsf{gg}(\boldsymbol{V}, p)$ is distributed as $\mathsf{Geo}_d(n, p)$.

The random geometric graph distribution on the unit sphere, $\mathsf{Geo}_d(n, p)$, behaves very differently depending on the value of d relative to n and p. When d is very small, graphs sampled from $\mathsf{Geo}_d(n, p)$ behave much like discretizations of the unit sphere. In such cases, we observe *triadic closure*, which is a tendency towards seeing triangles conditioned on two of the edges already being present.

In other words, if we have edges (i, j) and (i, k), the dot products between their associated vectors $\langle v_i, v_j \rangle$ and $\langle v_i, v_k \rangle$ are also high; this makes it likelier that $\langle v_j, v_k \rangle$ is high as well.

When d is very large, we lose this dependence between edges, and $\mathsf{Geo}_d(n, p)$ looks more and more like $\mathsf{G}(n, p)$. This begs a "Question 0" about $\mathsf{Geo}_d(n, p)$: for what dimension d is $\mathsf{Geo}_d(n, p)$ statistically indistinguishable from $G(n, p)$? In other words,

At what threshold for d (in terms of n, p) does $\mathsf{Geo}_d(n, p)$ lose its underlying geometry?

In a sense, this is also asking "at what dimension d do these random geometric graphs lose their modeling power over the Erdös Rényi model?

In more formal mathematical terms, we are asking:

For what d (in terms of n, p) does $\lim_{n\to\infty} \mathrm{d_{TV}}\left(\mathsf{G}(n,p), \mathsf{Geo}_d(n,p)\right) = 0$?

5.1.1 Related work

The first work that initiates study in this question is [BDER16], who show that a variant of the triangle counts (called "signed triangle counts") are sufficient for distinguishing between $\mathsf{G}(n,p)$ and $\mathsf{Geo}_d(n,p)$ when $d \ll n^3$ for $p = \Theta(1)$, and when $d \ll \log^3(n)$ for $p = \Theta\left(\frac{1}{n}\right)$. In [LMSY22b], we extend their results and also show that their signed triangle statistic is a good distinguisher when $d \ll n^3 H(p)^3$ for general p.

[BDER16] also show that their distinguisher is optimal in the dense regime $p = \Theta(1)$, using tools from random matrix theory:

Theorem 5.1.1 (Theorem 1(c) of [BDER16])**.** For $p = \Theta(1)$, if $d \gg n^3$, then

$$\lim_{n\to\infty} \mathrm{d_{TV}}\left(\mathsf{G}(n,p), \mathsf{Geo}_d(n,p)\right) = 0$$

They also conjecture that the signed triangle counts are the optimal distinguisher (up to $\mathsf{polylog}(n)$ factors) for $p = \Theta\left(\frac{1}{n}\right)$; we extend their conjecture to all p.

Conjecture 5.1.3. If $d \gg n^3 H(p)^3$, then

$$\lim_{n\to\infty} \mathrm{d_{TV}}\left(\mathsf{G}(n,p), \mathsf{Geo}_d(n,p)\right) = 0$$

In particular, when $p = \Theta\left(\frac{1}{n}\right)$, the above statement holds for $d \gg \log^3(n)$.

More recently, the work of [BBN20] improve upon the indistinguishability threshold for general p.

Theorem 5.1.2. If $d \gg \min\{pn^3 \log\frac{1}{p}, p^2 n^{7/2}\mathsf{polylog}(n)\}$, then

$$\lim_{n\to\infty} \mathrm{d_{TV}}\left(\mathsf{G}(n,p), \mathsf{Geo}_d(n,p)\right) = 0$$

In particular, if $p = \Theta\left(\frac{1}{n}\right)$, the above holds for $d \gg n^{3/2}\mathsf{polylog}(n)$.

5.1.2 Main results and technical overview

Our main results are twofold:

- First, we have an improved indistinguishability bound for general $p = \Omega\left(\frac{1}{n}\right)$.

Theorem 5.1.3. For any fixed constant $\alpha > 0$, if $\frac{\alpha}{n} < p < \frac{1}{2}$ and $d = \tilde{\Omega}(p^2 n^3)$,

$$\lim_{n\to\infty} \mathrm{d_{TV}}\left(\mathsf{Geo}_d\left(n,p\right), \mathsf{G}\left(n,p\right)\right) = 0.$$

This improves, by polynomial factors in p and n, on the previous best-known bound of [BBN20]. However, this result is not tight (at least for small p); in particular it does not recover Theorem 5.1.4.

- We also present an indistinguishability result for sparse random geometric graphs and Erdös-Rényi graphs when the dimension d exceeds $\mathsf{polylog}(n)$.

  Theorem 5.1.4. For any fixed constant $\alpha \geq 1$, if $d = \Omega(\log^{36} n)$, then

  $$\lim_{n\to\infty} \mathrm{d_{TV}}\left(\mathsf{Geo}_d\left(n, \tfrac{\alpha}{n}\right), \mathsf{G}\left(n, \tfrac{\alpha}{n}\right)\right) = 0.$$

  Our result settles the conjecture of [BDER16] up to logarithmic factors, and gave an exponential improvement over the previous bound of [BBN20], which required $d \gg n^{3/2}$. We remark that we have not made an effort to optimize the logarithmic factors; it is possible that our current proofs in combination with chaining-style arguments will yield $\log^3 n$, matching their conjecture.

In this thesis, we will prove Theorem 5.1.3, with an emphasis on the geometry-inspired and optimal transport-inspired toolkit for analyzing sphere caps. We will not fully prove Theorem 5.1.4, but rather just highlight one portion of the belief propagation argument that also heavily relies on the geometry of sphere caps.

We now provide technical outlines of the proofs of Theorems 5.1.3 and 5.1.4.

Relative entropy tensorization

Our goal is to determine the dimension d at which the total variation distance

$$\mathrm{d_{TV}}\left(\mathsf{G}(n,p), \mathsf{Geo}_d(n,p)\right)$$

goes to 0 as $n \to \infty$. Like the authors of [BBN20], we relate the TV distance between these two distributions to their relative entropy $\mathcal{D}_{\mathrm{KL}}(\mathsf{Geo}_d(n,p) \| \mathsf{G}(n,p))$ via Pinsker's inequality (Theorem 1.3.1), and then apply the tensorization of the relative entropy (Claim 5.4.1). Roughly, the tensorization says that given a decomposition of $\mathsf{G}(n,p)$ as a product distribution, we can reduce the problem of bounding $\mathcal{D}_{\mathrm{KL}}(\mathsf{Geo}_d(n,p) \| \mathsf{G}(n,p))$ to bounding the relative entropy over (potentially simpler) distributions with smaller support.

$\mathsf{G}(n,p)$ is conveniently a product distribution over edges. However, unlike [BBN20], we do not use this straightforward decomposition of $\mathsf{G}(n,p)$ by edge. Instead, let μ_t be the distribution of vertex t's edges to $[t-1]$ under $\mathsf{G}(n,p)$. Similarly, let ν_t be the marginal

distribution of vertex t's edges to $[t-1]$ over the graph being sampled from $\mathsf{Geo}_d(n,p)$. Our bound via tensorization now becomes

$$\begin{aligned}\mathcal{D}_{\mathrm{KL}}(\mathsf{Geo}_d(n,p)\|\mathsf{G}(n,p)) &= \sum_{t=1}^{n} \mathbb{E}_{\boldsymbol{G}_{t-1}\sim\mathsf{Geo}_d(t-1,p)}\left[\mathcal{D}_{\mathrm{KL}}\left(\nu_t(\cdot \mid \boldsymbol{G}_{t-1}) \,\|\, \mu_t\right)\right] \\ &\le n \cdot \mathbb{E}_{\boldsymbol{G}_{n-1}\sim\mathsf{Geo}_d(n-1,p)}\left[\mathcal{D}_{\mathrm{KL}}\left(\nu_n(\cdot \mid \boldsymbol{G}_{n-1}) \,\|\, \mu_n\right)\right]\end{aligned}$$

where the final inequality follows from the chain rule for relative entropy (Claim 5.4.2).

The coupling view. The tensorization inequality reduces bounding the TV distance to comparing the probability distribution of the neighborhood of the "final" vertex in $\mathsf{G}(n,p)$ and $\mathsf{Geo}_d(n,p)$. Specifically, we study $\mathbb{E}_{\boldsymbol{G}_{n-1}\sim\mathsf{Geo}_d(n-1,p)}\left[\mathcal{D}_{\mathrm{KL}}\left(\nu_n(\cdot \mid \boldsymbol{G}_{n-1}) \,\|\, \mu_n\right)\right]$ by considering the following scenario: we already have a graph $\boldsymbol{G}_{n-1}$ sampled on $n-1$ vertices, and we want to incorporate vertex n into our graph.

By the definition of $\mathsf{G}(n,p)$, μ_n will sample the neighbor set $S \subseteq [n-1]$ with probability $p^{|S|}(1-p)^{n-1-|S|}$. For a random geometric graph, we can sample a vector $\boldsymbol{v}_n \sim \mathbb{S}^{d-1}$, and take its dot products to vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{n-1}$ sampled uniformly from $\mathbb{S}^{d-1}$ conditioned on producing $\boldsymbol{G}_{n-1}$, to determine the neighbors of n in $\boldsymbol{G}$ (which we denote by $N_{\boldsymbol{G}}(n)$).

Our goal now is to compare, for $S \subseteq [n-1]$,

$$\Pr_{\boldsymbol{G}\sim\mathsf{G}(n,p)}[N_{\boldsymbol{G}}(n) = S] \text{ and } \Pr_{\boldsymbol{G}\sim\mathsf{Geo}_d(n,p)}[N_{\boldsymbol{G}}(n) = S]$$

A geometric interpretation of neighborhood probability

For $\boldsymbol{G} \sim \mathsf{Geo}_d(n,p)$, if vertex i is associated to a (random) vector $\boldsymbol{v}_i$, and (i,j) is an edge, we consequently know that $\langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle \ge \tau$. On the sphere $\mathbb{S}^{d-1}$, the locus of points where $\boldsymbol{v}_j$ can be, conditioned on (i,j) being an edge, is a sphere cap centered at $\boldsymbol{v}_i$ with a p fraction of the sphere's surface area, which we denote by $\mathsf{cap}(\boldsymbol{v}_i)$. Similarly, if we know that i and j are *not* adjacent, the locus of points where $\boldsymbol{v}_j$ can fall is the complement of a sphere cap, which we call an "anti-cap," with measure $1-p$.

Equipped with this geometric picture, we can view the probability that vertex n's neighborhood is exactly equal to $S \subseteq [n-1]$ as the measure $\rho(\boldsymbol{L}_S)$, where ρ is the uniform distribution over $\mathbb{S}^{d-1}$ and $\boldsymbol{L}_S \subseteq \mathbb{S}^{d-1}$ is a random set defined as

$$\boldsymbol{L}_S := \left(\bigcap_{i\in S} \mathsf{cap}(\boldsymbol{v}_i)\right) \cap \left(\bigcap_{i\notin S} \overline{\mathsf{cap}(\boldsymbol{v}_i)}\right)$$

To show that the distance between $\mathsf{Geo}_d(n,p)$ and $\mathsf{G}(n,p)$ is small, we must show that $\rho(\boldsymbol{L}_S)$ concentrates around $p^{|S|}(1-p)^{n-1-|S|}$, which is the probability that n's neighborhood is equal to S under the Erdös-Rényi model.

Optimal transport. The backbone of our result is a (to our knowledge) novel application of optimal transport. In Section 5.2, we prove for a generic distribution ν supported on $\mathbb{S}^{d-1}$, and $\boldsymbol{z}$ independently sampled from ρ that

$$\Pr_{\boldsymbol{x}\sim\nu}[\langle \boldsymbol{x}, \boldsymbol{z}\rangle \geq \tau] \in (1 \pm \varepsilon) \cdot p \text{ for } \varepsilon \leq \tilde{O}\left(\sqrt{\frac{\ln \|\nu\|_\infty}{d}}\right)$$

with high probability over $\boldsymbol{z}$.

In other words, how tightly the random variable $X_\nu(\boldsymbol{z}) = \Pr_{\boldsymbol{x}\sim\nu}[\langle \boldsymbol{x}, \boldsymbol{z}\rangle \geq \tau]$ concentrates is directly related to the maximum value of its relative density, $\|\nu\|_\infty = \max_{x\in\mathbb{S}^{d-1}} \frac{d\nu(x)}{d\rho(x)}$.

To give some intuition for this result, first consider the case when $\nu = \rho$, the uniform distribution over $\mathbb{S}^{d-1}$. Then, the variable $X_\rho(\boldsymbol{z}) = \Pr_{\boldsymbol{y}\sim\rho}[\langle \boldsymbol{y}, \boldsymbol{z}\rangle \geq \tau] = p$ deterministically. Now, when $\nu \neq \rho$, we can work with a transport map $\mathcal{D}$ between ρ and ν, and we can couple $\boldsymbol{x} \sim \nu$ and $\boldsymbol{y} \sim \rho$ according to $\mathcal{D}$, so that

$$X_\nu(z) = \Pr_{\boldsymbol{x}\sim\nu}[\langle \boldsymbol{x}, z\rangle \geq \tau] = \Pr_{\substack{(\boldsymbol{x},\boldsymbol{y})\sim\mathcal{D}(\nu,\rho)\\ \boldsymbol{e}=\boldsymbol{x}-\boldsymbol{y}}}[\langle \boldsymbol{y}, z\rangle \geq \tau - \langle \boldsymbol{e}, z\rangle].$$

The smaller $\|\nu\|_\infty$ is, the smaller the average of the transport distance $\|\boldsymbol{e}\| = \|\boldsymbol{x}-\boldsymbol{y}\|$; further when $\boldsymbol{z} \sim \rho$ the quantity $\langle \boldsymbol{e}, \boldsymbol{z}\rangle$ concentrates tightly in a $\pm\frac{1}{\sqrt{d}}\|\boldsymbol{e}\|$ window around 0. In this way, we translate the concentration of transport distance into tail bounds on $|X_\rho(\boldsymbol{z}) - X_\nu(\boldsymbol{z})|$.

To analyze $\rho(\boldsymbol{L})$ for $\boldsymbol{L}$ the intersection of caps and anti-caps defined above, we will apply the above in sequence inside a martingale concentration argument, building up $\boldsymbol{L}$ one cap at a time (Lemma 5.3.1, Corollary 5.3.10). Using this approach, our transport result alone is enough to conclude Theorem 5.1.3. (The proof is assembled in Section 5.4.1.)

The need to resample vectors. In the general p setting, we can think of our analysis of $\nu_n(\cdot|\boldsymbol{G}_{n-1})$ as considering a fixed vector embedding $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{n-1}$ of $\boldsymbol{G}_{n-1}$, and then analyzing the probability that n connects to some $S \subseteq [n-1]$. When $p = \frac{\alpha}{n}$, this does not yield tight results; moreover, one can show that this is not due to loose tail bounds on $\rho(\boldsymbol{L})$, as our concentration results have matching anti-concentration results.

Hence, in order to prove Theorem 5.1.4, we must additionally consider the concentration of $\rho(\boldsymbol{L}_S)$ *on average over vector embeddings of* $\boldsymbol{G}_{n-1}$ as well.

We will first sample $\boldsymbol{G}_{n-1}$, and then for each set S, we bound the deviation in the random variable $\rho(\boldsymbol{L}_S) = \rho(\bigcap_{i\in S} \mathsf{cap}(\boldsymbol{u}_i) \cap \bigcap_{j\notin S} \overline{\mathsf{cap}}(\boldsymbol{u}_j))$ conditioned on $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{n-1}$ producing $\boldsymbol{G}_{n-1}$. To do this, we will use a "cavity-method" style argument: we will view all vectors at distance $> \ell = \frac{\log n}{\log\log n}$ from S as fixed and arbitrary, and then exactly compute the marginal distributions over $\boldsymbol{u}_i$ for i at distance $\leq \ell$ from S, conditional on forming $\boldsymbol{G}_{n-1}$.

Neighborhood containment as a constraint satisfaction problem.

We first reduce the need for high-probability estimates for $\Pr_{\boldsymbol{G}\sim\mathsf{Geo}_d(n,p)}[N_{\boldsymbol{G}}(n) = S]$ where $S \subseteq [n-1]$ to obtaining estimates for $\Pr_{\boldsymbol{G}\sim\mathsf{Geo}_d(n,p)}[N_{\boldsymbol{G}}(n) \supseteq S]$ instead. This simplification

is possible because the measure of anti-cap intersections concentrates dramatically better than the measure of cap intersections. With this step, we eliminate the need to study anti-correlations between $\boldsymbol{v}_i, \boldsymbol{v}_j$ that do not have an edge between them.

Given S and $\boldsymbol{G}_{n-1}$ (and its corresponding vectors), we fix all vectors except those corresponding to the depth-$\frac{\log n}{\log \log n}$ neighborhood of S in $\boldsymbol{G}_{n-1}$, which is with high probability a union of trees. To formally analyze the distribution of the unfixed vectors upon resampling them, we set up a 2-CSP (constraint satisfaction problem) instance over a continuous alphabet that encodes the edges of $\boldsymbol{G}_{n-1}$ within the trees around S: each node has a vector-valued variable in $\mathbb{S}^{d-1}$, and the constraints are that nodes joined by an edge must have vectors with inner product at least τ.

Belief propagation. Since our 2-CSPs are over trees, the *belief propagation (BP) algorithm* exactly computes the marginal distribution of each variable vector (see Section 5.5 for the definition of BP).

Using our results on the concentration of $\Pr_{\boldsymbol{x}\sim\nu}[\langle \boldsymbol{x}, \boldsymbol{z}\rangle \geq \tau]$ over $\boldsymbol{z} \sim \rho$, we can quantify the TV distance between the marginal distributions of our resampled vectors and the uniform distribution over $\mathbb{S}^{d-1}$. At a high level, the farther some $\boldsymbol{v}_i$ is from a fixed vector in our 2-CSP, the closer its distribution is to uniform.

The key insight is that the message from i to its neighbor j in our belief propagation algorithm correspond to a convolution of the marginal distribution of $\boldsymbol{v}_i$ with a spherical cap. We can then use our concentration of measure for spherical caps from Section 5.2 to show that convolutions of spherical caps mix to uniform rapidly, causing the correlations between far away vertices to decay.

This can also be seen as a form of the "decay of correlations" phenomenon. This analysis gives us the finer-grained control over $\Pr[N_{\boldsymbol{G}}(n) = S]$ needed to conclude Theorem 5.1.4. [1]

5.2 Concentration via optimal transport

In this section we establish that for a probability distribution ν over $\mathbb{S}^{d-1}$, a random p-cap on the sphere contains a p-fraction of ν's measure with high probability, where the strength of the concentration depends on $\|\nu\|_\infty$. We do so by analyzing the optimal transport mapping $\mathcal{D}$ between ν and the uniform measure ρ.

5.2.1 Optimal transport and the Wasserstein metric

The Wasserstein metric quantifies the "physical distance" between a pair of probability distributions. We say $\pi(x, y)$ is a transport coupling between distributions μ and ν if

$$\pi(x, \cdot) = \mu(x) \text{ and } \pi(\cdot, y) = \nu(y)$$

[1] Unfortunately, this analysis only works for $p = \frac{\alpha}{n}$; otherwise the neighborhoods around vertices in S are only trees at a depth which is too shallow for the correlations to decay sufficiently, so the resampled vectors' distributions are not close enough to uniform.

Definition 5.2.1 (Wasserstein Distance)**.** Let μ and ν be two probability distributions over Ω, and Π be the set of all transport couplings $\pi(x, y)$ between μ and ν. Then, the *Wasserstein-2 distance* between μ and ν is

$$W_2(\mu, \nu) := \sqrt{\inf_{\pi \in \Pi} \int_{\Omega \times \Omega} \|x - y\|_2^2 d\pi(x, y)}$$

It is straightforward to verify that $W_2(\cdot, \cdot)$ is in fact a metric.

In other words, the square of the Wasserstein distance is the average squared Euclidean distance $\|\boldsymbol{x} - \boldsymbol{y}\|^2$ between $(\boldsymbol{x}, \boldsymbol{y}) \sim \pi$ for the most efficient "transport coupling" π. For intuition, we may think of μ's density of a pile of sand over Ω, and of π as a map that shifts grains of sand from μ to form the shape of ν in such a way that minimizes the average distance traveled.

Fact 5.2.2 (Proposition 9.1.2 of [BGL13])**.** Given two probability distributions μ and ν over Ω, there exists a coupling π such that:

$$\int_{\Omega \times \Omega} \|x - y\|_2^2 d\pi(x, y) = W_2(\mu, \nu)^2.$$

We call any such π an *optimal coupling*.

We specifically will need bounds on the Wasserstein-2 distance between an arbitrary distribution ν on the unit sphere and the uniform distribution on the unit sphere ρ. We obtain a handle on the distances we need via an *entropy-transport inequality*, which bounds the Wasserstein-2 distance in terms of the relative entropy. The following is a direct corollary of Theorem 22.17(i) of [Vil08] and Corollary 2 of [DEKL14].

Lemma 5.2.3 (Talagrand's T_2 inequality on the sphere)**.** For any distribution ν on $\mathbb{S}^{d-1}$, and ρ the uniform measure on $\mathbb{S}^{d-1}$,

$$W_2(\nu, \rho) \leq \sqrt{\frac{2}{d-1} \cdot \mathcal{D}_{\mathrm{KL}}(\nu \| \rho)}.$$

For a reader interested in the proof of Lemma 5.2.3, we recommend the proof of the analogous statement in Gaussian space by [Tal96] due to its relative simplicity. In fact, it is possible to derive a slightly weaker bound of $\sqrt{\frac{2}{d} \cdot \mathcal{D}_{\mathrm{KL}}(\nu \| \rho)} + \frac{2}{\sqrt{d}}$ from the Gaussian case via an elementary proof. We also find it worthwhile to point to Theorem 9.2.1 of [BGL13] for a comprehensive exposition.

5.2.2 The concentration of sphere cap measure

Let ρ be the uniform measure over $\mathbb{S}^{d-1}$, let ν be a probability measure over $\mathbb{S}^{d-1}$, and let $\mathcal{D}$ be a coupling between them. For $(\boldsymbol{x},\boldsymbol{y}) \sim \mathcal{D}$ we use the convention that $\boldsymbol{x}$ is distributed according to ν and $\boldsymbol{y}$ is distributed according to ρ. We first prove quantitative bounds on the concentration of $\|\boldsymbol{x}-\boldsymbol{y}\|$ when $(\boldsymbol{x},\boldsymbol{y}) \sim \mathcal{D}$.

Lemma 5.2.4. Let ν be a distribution over $\mathbb{S}^{d-1}$ and let $\mathcal{D}$ be the optimal transport coupling of ν and ρ. Then for all $t > 0$,

$$\Pr_{(\boldsymbol{x},\boldsymbol{y})\sim\mathcal{D}}\left[\|\boldsymbol{x}-\boldsymbol{y}\| \geq t + \sqrt{\tfrac{2}{d-1}\ln\|\nu\|_\infty}\right] \leq \exp\left(-\frac{d-1}{8}\cdot t^2\right).$$

Proof. Let $(\boldsymbol{x},\boldsymbol{y}) \sim \mathcal{D}$ with $\boldsymbol{x} \sim \nu$ and $\boldsymbol{y} \sim \rho$. Let $\mathcal{E}_s$ be the event that $\|\boldsymbol{x}-\boldsymbol{y}\| \geq s$, and let $p_s := \Pr_{\mathcal{D}}[\mathcal{E}_s]$. Let $\mathcal{D}^s = \mathcal{D}|\mathcal{E}_s$, and let $\mathcal{D}^s_\nu, \mathcal{D}^s_\rho$ be the marginal distributions of $\mathcal{D}^s$ on $\boldsymbol{x}$ and $\boldsymbol{y}$ respectively. We claim

$$W_2(\mathcal{D}^s_\nu, \mathcal{D}^s_\rho) \geq s.$$

Indeed, suppose not; then, one could obtain a coupling which further decreases the transport distance between ν and ρ, contradicting the optimality of $\mathcal{D}$. Now, since W_2 is a metric:

$$\begin{aligned} s &\leq W_2(\mathcal{D}^s_\nu, \mathcal{D}^s_\rho) \leq W_2(\mathcal{D}^s_\rho, \rho) + W_2(\mathcal{D}^s_\nu, \rho) \\ &\leq \sqrt{\frac{2}{d-1}\mathcal{D}_{\mathrm{KL}}(\mathcal{D}^s_\rho\|\rho)} + \sqrt{\frac{2}{d-1}\mathcal{D}_{\mathrm{KL}}(\mathcal{D}^s_\nu\|\rho)}, \end{aligned} \tag{5.1}$$

where we have used the triangle inequality in conjunction with Talagrand's T_2 inequality (Lemma 5.2.3). Finally, since $\mathcal{D}^s_\rho(x) = \rho(x|\mathcal{E}_s) \leq \frac{1}{p_s}\rho(x)$,

$$\mathcal{D}_{\mathrm{KL}}(\mathcal{D}^s_\rho\|\rho) = \int_{\mathbb{S}^{d-1}} \mathcal{D}^s_\rho(x)\cdot\ln\frac{\mathcal{D}^s_\rho(x)}{\rho(x)}dx \leq \int_{\mathbb{S}^{d-1}} \mathcal{D}^s_\rho(x)\cdot\ln\frac{1}{p_s}dx = \ln\frac{1}{p_s},$$

and similarly, because $\mathcal{D}^s_\nu(x) = \nu(x|\mathcal{E}_s) \leq \frac{1}{p_s}\nu(x) \leq \frac{\|\nu\|_\infty}{p_s}\rho(x)$,

$$\mathcal{D}_{\mathrm{KL}}(\mathcal{D}^s_\nu\|\rho) = \int_{\mathbb{S}^{d-1}} \mathcal{D}^s_\nu(x)\cdot\ln\frac{\mathcal{D}^s_\nu(x)}{\rho(x)}dx \leq \int_{\mathbb{S}^{d-1}} \mathcal{D}^s_\nu(x)\cdot\ln\frac{\|\nu\|_\infty}{p_s}dx = \ln\frac{1}{p_s} + \ln\|\nu\|_\infty.$$

Putting these together with Equation 5.1 and using $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$, we have

$$s \leq 2\sqrt{\frac{2}{d-1}\ln\frac{1}{p_s}} + \sqrt{\frac{2}{d-1}\ln\|\nu\|_\infty},$$

and then re-arranging we have

$$p_s \leq \exp\left(-\frac{d-1}{8}\left(s - \sqrt{\frac{2}{d-1}\ln\|\nu\|_\infty}\right)^2\right),$$

when $s \geq \sqrt{\frac{2}{d-1}\ln\|\nu\|_\infty}$. Applying a change of variables completes the proof. □

Having established in Lemma 5.2.4 that the optimal transport map $\pi(\nu, \rho)$ between $\boldsymbol{x} \sim \nu$ and $\boldsymbol{y} \sim \rho$ has bounded length with high probability, we can translate this into a tail bound for the inner product $\langle \boldsymbol{z}, \boldsymbol{x} - \boldsymbol{y} \rangle$ for a random vector $\boldsymbol{z} \sim \rho$.

Lemma 5.2.5. Let ν be a distribution on $\mathbb{S}^{d-1}$, and let $\mathcal{D}$ be the optimal transport coupling of ν and ρ. For $z \in \mathbb{S}^{d-1}, t \in \mathbb{R}_{\geq 0}$ and any $\kappa > 0$, define $X(z,t)$ as:

$$X(z,t) := \Pr_{(\boldsymbol{x},\boldsymbol{y})\sim\mathcal{D}}\left[|\langle z, \boldsymbol{x} - \boldsymbol{y}\rangle| \geq \left(\sqrt{\frac{2}{d-1}\ln\|\nu\|_\infty} + \sqrt{\frac{8\kappa}{d-1}}\right)\cdot t\right].$$

Then:

$$\Pr_{\boldsymbol{z}\sim\rho}\left[X(\boldsymbol{z},t) \geq 2\exp(-dt^2/4) + \exp(-\kappa)\right] \leq 2\exp(-dt^2/4).$$

Remark 5.2.6. One should think of $X(z,t)$ as a measure of how often a randomly chosen transport vector $\boldsymbol{x} - \boldsymbol{y}$, with $(\boldsymbol{x}, \boldsymbol{y}) \sim \mathcal{D}$, has a large projection in the z direction.

Proof of Lemma 5.2.5. For any $z \in \mathbb{S}^{d-1}$ and $(\boldsymbol{x}, \boldsymbol{y}) \sim \mathcal{D}$, suppose

$$|\langle z, \boldsymbol{x} - \boldsymbol{y}\rangle| \geq \left(\sqrt{\frac{2}{d-1}\ln\|\nu\|_\infty} + \sqrt{\frac{8\kappa}{d-1}}\right)\cdot t$$

Then $|\langle z, \boldsymbol{x} - \boldsymbol{y}\rangle| \geq \|\boldsymbol{x} - \boldsymbol{y}\| \cdot t$ or $\|\boldsymbol{x} - \boldsymbol{y}\| \geq \sqrt{\frac{2}{d-1}\ln\|\nu\|_\infty} + \sqrt{\frac{8\kappa}{d-1}}$. Defining

$$Y(z,t) := \Pr_{(\boldsymbol{x},\boldsymbol{y})\sim\mathcal{D}}\left[|\langle z, \boldsymbol{x} - \boldsymbol{y}\rangle| \geq \|\boldsymbol{x} - \boldsymbol{y}\| \cdot t\right],$$

we can write

$$X(z,t) \leq Y(z,t) + \Pr_{(\boldsymbol{x},\boldsymbol{y})\sim\mathcal{D}}\left[\|\boldsymbol{x} - \boldsymbol{y}\| \geq \sqrt{\frac{8\kappa}{d-1}} + \sqrt{\frac{2}{d-1}\ln\|\nu\|_\infty}\right] \leq Y(z,t) + \exp(-\kappa) \tag{5.2}$$

where to obtain the upper bound we have applied Lemma 5.2.4. Next, we prove that

$$\Pr_{\boldsymbol{z}\sim\rho}\left[Y(\boldsymbol{z},t) \geq 2\exp(-dt^2/4)\right] \leq 2\exp(-dt^2/4), \tag{5.3}$$

by showing $\mathbb{E}\left[Y(\boldsymbol{z},t)\right] \leq 4\exp(-dt^2/2)$, which implies Equation 5.3 via Markov's inequality.

$$\begin{aligned}\mathbb{E}_{\boldsymbol{z}\sim\rho}\left[Y(\boldsymbol{z},t)\right] &= \Pr_{\substack{\boldsymbol{z}\sim\rho\\(\boldsymbol{x},\boldsymbol{y})\sim\mathcal{D}}}\left[|\langle \boldsymbol{z}, \boldsymbol{x} - \boldsymbol{y}\rangle| \geq \|\boldsymbol{x} - \boldsymbol{y}\| \cdot t\right] = \Pr_{\substack{\boldsymbol{z}\sim\rho\\(\boldsymbol{x},\boldsymbol{y})\sim\mathcal{D}}}\left[|\langle \boldsymbol{z}, \frac{\boldsymbol{x} - \boldsymbol{y}}{\|\boldsymbol{x} - \boldsymbol{y}\|}\rangle| \geq t\right] \\ &\leq 4\exp(-dt^2/2) \text{ by Lemma 4.1.7}\end{aligned}$$

We can then complete the proof starting at Equation 5.3 as follows:

$$\begin{aligned}\Pr_{\boldsymbol{z}\sim\rho}\left[Y(\boldsymbol{z},t) + \exp(-\kappa) \geq 2\exp(-dt^2/4) + \exp(-\kappa)\right] &\leq 2\exp(-dt^2/4) \\ \Pr_{\boldsymbol{z}\sim\rho}\left[X(\boldsymbol{z},t) \geq 2\exp(-dt^2/4) + \exp(-\kappa)\right] &\leq 2\exp(-dt^2/4). \quad \text{(by Equation 5.2)}\end{aligned}$$

This yields the desired conclusion. □

Since a p-cap around x is given by the set of vectors z with inner product $\langle x, z\rangle \geq \tau(p,d)$, and $\langle x, z\rangle = \langle y, z\rangle \pm |\langle x-y, z\rangle|$, we can finally use Lemma 5.2.5 to relate

$$X(z) = \Pr_{\boldsymbol{x}\sim\nu}[\langle z, \boldsymbol{x}\rangle > \tau(p,d)]$$

the measure of ν that falls into the p-cap of z, to

$$p = \Pr_{\boldsymbol{y}\sim\rho}[\langle z, \boldsymbol{y}\rangle > \tau(p,d)]$$

That is, we can now show that $X(\boldsymbol{z})$ for $\boldsymbol{z}\sim\rho$ concentrates tightly around p, so that most vectors in $\mathbb{S}^{d-1}$ contain very close to a p-fraction of ν's mass in their p-caps.

Lemma 5.2.7. Let ν be a distribution on $\mathbb{S}^{d-1}$. For $z\in\mathbb{S}^{d-1}$, let $X(z) := \Pr_{\boldsymbol{x}\sim\nu}[\langle \boldsymbol{x}, z\rangle > \tau(p,d)]$, and for any $\kappa>0$, let $u(t) := \left(\sqrt{\frac{8}{d-1}\ln\|\nu\|_\infty} + \sqrt{\frac{8\kappa}{d-1}}\right)\cdot t$. Then for any $t\geq 0$:

$$\Pr_{\boldsymbol{z}\sim\rho}\left[|X(\boldsymbol{z}) - p| > p\cdot\varepsilon(t)\right] \leq 2\exp(-dt^2/4)$$

where $\varepsilon(t) := C_{4.1.9}\cdot u(t)\cdot\exp\left(2d\cdot\tau(p,d)\cdot u(t)\right)\cdot\sqrt{d\log\frac{1}{p}} + \frac{2\exp(-dt^2/4)+\exp(-\kappa)}{p}$.

Proof. Let $\mathcal{D}$ be the optimal coupling between ν and ρ. For any $z\in\mathbb{S}^{d-1}$ and $t\geq 0$:

$$\begin{aligned}
X(z) &= \Pr_{(\boldsymbol{x},\boldsymbol{y})\sim\mathcal{D}}[\langle \boldsymbol{y}, z\rangle > \tau(p,d) - \langle z, \boldsymbol{x}-\boldsymbol{y}\rangle] \\
&\leq \Pr_{(\boldsymbol{x},\boldsymbol{y})\sim\mathcal{D}}[\langle \boldsymbol{y}, z\rangle > \tau(p,d) - \max\{\langle z, \boldsymbol{x}-\boldsymbol{y}\rangle, u(t)\}] \\
&\leq \Pr_{\boldsymbol{y}\sim\rho}[\langle \boldsymbol{y}, z\rangle > \tau(p,d) - u(t)] + \Pr_{(\boldsymbol{x},\boldsymbol{y})\sim\mathcal{D}}[|\langle z, \boldsymbol{x}-\boldsymbol{y}\rangle| > u(t)] \\
&\leq p\left(1 + C_{4.1.9}\cdot u(t)\cdot\exp\left(2d\tau(p,d)u(t)\right)\cdot\sqrt{d\log\frac{1}{p}}\right) + \Pr_{(\boldsymbol{x},\boldsymbol{y})\sim\mathcal{D}}[|\langle z, \boldsymbol{x}-\boldsymbol{y}\rangle| > u(t)]
\end{aligned} \tag{5.4}$$

where the last step of the chain of inequalities follows from Lemma 4.1.9. Identically,

$$X(z) \geq p\left(1 - C_{4.1.9}\cdot u(t)\cdot\exp\left(2d\tau(p,d)u(t)\right)\cdot\sqrt{d\log\frac{1}{p}}\right) - \Pr_{(\boldsymbol{x},\boldsymbol{y})\sim\mathcal{D}}[|\langle z, \boldsymbol{x}-\boldsymbol{y}\rangle| > u(t)] \tag{5.5}$$

Then by Lemma 5.2.5, when $\boldsymbol{z}\sim\rho$, we can obtain an upper bound of $2\exp(-dt^2/4)+\exp(-\kappa)$ on the second term in the right hand side of Equation 5.4 and Equation 5.5 that holds except with probability $2\exp(-dt^2/4)$, which implies

$$\begin{aligned}
\Pr_{\boldsymbol{z}\sim\rho}\Bigg[|X(\boldsymbol{z}) - p| > p\cdot C_{4.1.9}\cdot u(t)\cdot\exp\left(2d\tau(p,d)u(t)\right)\cdot\sqrt{d\log\frac{1}{p}} + 2\exp(-dt^2/4) + \exp(-\kappa)\Bigg] \\
\leq 2\exp(-dt^2/4).
\end{aligned}$$

thus completing the proof. □

5.2.3 Different parameterizations of the sphere cap concentration

We'll now derive a few useful corollaries of Lemma 5.2.7. First, for intuition, consider the following immediate consequence regarding the intersection of a set in $\mathbb{S}^{d-1}$ with a random sphere cap.

Corollary 5.2.8. Let $d \geq \log^{10} n$ and let $Q \subseteq \mathbb{S}^{d-1}$ be a set such that $\rho(Q) \geq \frac{1}{n^{\log^3 n}}$. Then for $\boldsymbol{z} \sim \rho$:

$$\frac{\rho\left(\mathsf{cap}(\boldsymbol{z}) \cap Q\right)}{\rho\left(\mathsf{cap}(\boldsymbol{z})\right) \cdot \rho(Q)} \notin \left(1 \pm \frac{\log^5 n}{\sqrt{d}} \pm n^{-\log^2 n}\right)$$

with probability at most $n^{-\Omega(\log^3 n)}$.

We will make use of the following convenient specialization of Lemma 5.2.7.

Corollary 5.2.9. Let ν be a distribution on $\mathbb{S}^{d-1}$ and for $z \in \mathbb{S}^{d-1}$, let

$$X(z) := \Pr_{\boldsymbol{x} \sim \nu}\left[\langle \boldsymbol{x}, z\rangle \geq \tau(p,d)\right]$$

Then for $s \leq 1$ and for some constant $C_{5.2.9}$,

$$\Pr_{\boldsymbol{z}\sim\rho}\left[|X(\boldsymbol{z}) - p| > ps\right] \leq 2\exp\left(-\frac{ds^2}{C_{5.2.9}\left(\sqrt{\ln\|\nu\|_\infty} + \sqrt{\ln\frac{d}{p}}\right)^2 \cdot \log\frac{1}{p} \cdot \log\frac{d}{p}}\right)$$

Proof. The idea is to apply Lemma 5.2.7 by setting $\kappa = 4\ln\frac{d}{p}$ and using the parameterization

$$t = \frac{s}{2\left(\sqrt{8\ln\|\nu\|_\infty} + \sqrt{32\ln\frac{1}{p}}\right)\sqrt{\ln\frac{1}{p}}}.$$

Clearly, the statement is true when

$$s \leq \sqrt{\frac{C_{5.2.9}}{2d}} \cdot \left(\sqrt{\ln\|\nu\|_\infty} + \sqrt{\ln\frac{d}{p}}\right) \cdot \sqrt{\log\frac{1}{p} \cdot \log\frac{d}{p}}$$

since $\Pr\left[|X(z) - p| > ps\right] \leq 1 < 2\exp\left(-\frac{1}{2}\right)$. Hence, we restrict our attention to when

$$s \in H := \left[\sqrt{\frac{C_{5.2.9}}{2d}} \cdot \left(\sqrt{\ln\|\nu\|_\infty} + \sqrt{\ln\frac{d}{p}}\right) \cdot \sqrt{\log\frac{1}{p} \cdot \log\frac{d}{p}}, 1\right]$$

Let $\varepsilon(t)$ and $u(t)$ be as in the statement of Lemma 5.2.7. Using $s \leq 1$ and Lemma 4.1.8, we know $\exp(2d\tau(p,d)u(t)) \leq O(1)$. This tells us that for some constant C,

$$\varepsilon(t) \leq C\left(\sqrt{\frac{8}{d-1}\ln\|\nu\|_\infty} + \sqrt{\frac{32\ln\frac{1}{p}}{d-1}}\right)\sqrt{d\log\frac{1}{p}}t + \frac{2\exp(-dt^2/4) + \frac{p^4}{d^4}}{p}.$$

We can choose our constant $C_{5.2.9}$ to be a large enough so that when $s \in H$, then $t \geq 4\sqrt{\frac{\log \frac{d}{p}}{d}}$. Observe that once $t \geq 4\sqrt{\frac{\log \frac{d}{p}}{d}}$, for large enough d:

$$\begin{aligned}\varepsilon(t) &\leq 2C\left(\sqrt{\frac{8}{d-1}\ln\|\nu\|_\infty} + \sqrt{\frac{32\ln\frac{1}{p}}{d-1}}\right)\sqrt{d\log\frac{1}{p}}t \\ &= C'\cdot 2\left(\sqrt{8\ln\|\nu\|_\infty} + \sqrt{32\ln\frac{1}{p}}\right)\cdot\sqrt{\ln\frac{1}{p}}\cdot t \\ &= C's\end{aligned}$$

for some other constant C'. Then by Lemma 5.2.7, when $s \in H$:

$$\begin{aligned}\Pr\left[|X(z)-p| > ps\right] &\leq 2\exp\left(-\frac{ds^2}{4C'^2\left(\sqrt{8\ln\|\nu\|_\infty}+\sqrt{32\ln\frac{1}{p}}\right)^2\ln\frac{1}{p}}\right) \\ &\leq 2\exp\left(-\frac{ds^2}{C_{5.2.9}\left(\sqrt{\ln\|\nu\|_\infty}+\sqrt{\ln\frac{d}{p}}\right)^2\cdot\log\frac{1}{p}\cdot\log\frac{d}{p}}\right)\end{aligned}$$

where the second inequality arises from choosing $C_{5.2.9}$ to be large enough, which completes the proof. □

Corollary 5.2.9 can be extended to the case when $\boldsymbol{z} \sim \mu$ with a worse quantitative upper bound depending on μ from Observation 1.3.1.

Corollary 5.2.10. Let ν and μ be distributions on $\mathbb{S}^{d-1}$ and for $z \in \mathbb{S}^{d-1}$, let $X(z) := \Pr_{\boldsymbol{x}\sim\nu}[\langle \boldsymbol{x}, z\rangle > \tau(p,d)]$. Then for $s \leq 1$ we have:

$$\Pr_{\boldsymbol{z}\sim\mu}\left[|X(\boldsymbol{z})-p| > ps\right] \leq 2\exp\left(-\frac{ds^2}{C_{5.2.9}\left(\sqrt{\ln\|\nu\|_\infty}+\sqrt{\ln\frac{d}{p}}\right)^2\cdot\log\frac{1}{p}\cdot\log\frac{d}{p}}\right)\cdot\|\mu\|_\infty$$

5.3 Intersections of caps and anti-caps

In this section, we prove concentration of measure for the intersection of random p-caps and p-anticaps with any fixed set $L \subseteq \mathbb{S}^{d-1}$.

Lemma 5.3.1 (Concentration for the intersection of j caps and $k-j$ anti-caps)**.** Let ρ be the uniform measure over $\mathbb{S}^{d-1}$, and let $L \subset \mathbb{S}^{d-1}$, and let $k > 0$ be an integer. For $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k) \sim \rho^{\otimes k}$, let

$$\boldsymbol{S}_i := \begin{cases}\mathsf{cap}(\boldsymbol{v}_i) & \text{if } i \leq j \\ \overline{\mathsf{cap}}(\boldsymbol{v}_i) & \text{if } i > j,\end{cases}$$

and let $\boldsymbol{L}_t = L \cap \bigcap_{i=1}^t \boldsymbol{S}_i$. Then the ratio $\boldsymbol{R} := \frac{\rho(\boldsymbol{L}_k)}{\rho(L)\cdot p^j\cdot(1-p)^{k-j}}$ is concentrated as follows:

$$\Pr\left[|\boldsymbol{R}-1| > s\right] \leq \varepsilon_1(s) + k\varepsilon_1(.5) + \varepsilon_2$$

where

$$\varepsilon_1(s) := 2\exp\left(-\frac{ds^2}{C'(j+(k-j)p^2)F(j)}\right)$$
$$\varepsilon_2 := \frac{4k}{p^2}\cdot\exp\left(-\frac{d}{CF(j)}\right)$$

$F(j) := \left(\sqrt{\ln\frac{1}{\rho(L)} + j\ln\frac{1}{p} + (k-j)\ln\frac{1}{1-p}} + \sqrt{\ln\frac{d}{p}}\right)^2 \log\frac{1}{p}\log\frac{d}{p}$, and $C, C' > 0$ are universal constants.

Before proving the lemma, we will articulate one useful corollary: extremely good concentration for intersections anti-caps. For example, in the $p = \frac{\alpha}{n}$ regime, an intersection of $m = \Theta(n)$ anti-caps will have measure $(1-p)^m(1\pm\varepsilon)$ with typical deviations $\varepsilon = o(\sqrt{1/n})$ if $d > \mathsf{polylog}(n)$ for a sufficiently large power of $\log n$. (See also Corollary 5.3.10 for another application of Lemma 5.3.1).

Corollary 5.3.2 (Intersection of anti-caps)**.** Suppose $m, n, d \in \mathbb{Z}_{\geq 0}$, $a \geq 1$, and $p \in \mathbb{R}_{\geq 0}$ satisfy $m \leq n$, $d \leq n^{100}$, $\frac{1}{n^2} \ll p \leq \frac{1}{2}$, and $mp \leq \log^a n$ for n sufficiently large. Let $L \subset \mathbb{S}^{d-1}$ with $\rho(L) \geq e^{-\log^a n}$. Let $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m \sim \mathbb{S}^{d-1}$ uniformly at random, and let

$$\boldsymbol{A} = \{w \in \mathbb{S}^{d-1} \mid \langle w, \boldsymbol{v}_i\rangle \leq \tau(p,d)\ \forall i \in [m]\}$$

be the intersection of anti-caps of the $\boldsymbol{v}_i$'s. Then there exist universal constants C, C' such that for all n sufficiently large, for all $t \geq 0$,

$$\Pr_{\boldsymbol{A}}\left[\left|\frac{\rho(\boldsymbol{A}\cap L)}{(1-p)^m\rho(L)} - 1\right| > t\cdot\sqrt{mp^2}\right] \leq \exp\left(-C\min\left\{\frac{t^2\cdot d}{(\ln n)^{a+4}}, \frac{d - C'(\ln n)^{2a+5}}{(\ln n)^{2a+4}}\right\}\right).$$

Proof. Using that $\ln(1+x) \leq x$ for $x > 0$,

$$\begin{aligned}\frac{1}{\log\frac{1}{p}\log\frac{d}{p}}\sqrt{F(j)} &= \sqrt{\ln\tfrac{1}{\rho(L)} + m\ln(1+\tfrac{p}{1-p})} + \sqrt{\ln\tfrac{d}{p}} \\ &\leq \sqrt{\ln\tfrac{1}{\rho(L)}} + \sqrt{m\frac{p}{1-p}} + \sqrt{\ln\tfrac{d}{p}} \\ &\leq c\cdot\ln^{a/2} n\end{aligned}$$

for $c > 0$ a universal constant, where in the final inequality we have applied the assumption $p \leq \frac{1}{2}$ and the assumption $p \gg n^{-2}$. Re-arranging, we have $F(j) \leq c'(\ln n)^{a+4}$ for some

constant c'. Hence, Lemma 5.3.1 implies that there exist constants c'', c''' so that for n sufficiently large,

$$\Pr_{\boldsymbol{A}}\left[\left|\frac{\rho(\boldsymbol{A}\cap L)}{(1-p)^m\rho(L)}-1\right|>s\right]\leq 2\exp\left(-c''\frac{ds^2}{mp^2(\ln n)^{a+4}}\right) + \left(2k+\frac{4m}{p^2}\right)\cdot\exp\left(-c'''\frac{d}{(\ln n)^{2a+4}}\right),$$

where we used our bounds on p, k, m to combine the ε_2 and $k\varepsilon_1(\frac{1}{2})$ terms from Lemma 5.3.1 into our second term. The conclusion now follows by substituting $s = t\sqrt{mp^2}$ and applying asymptotic simplifications. □

Our proof proceeds by a martingale argument. First, we notice that the (rescaled) area of the intersection of sets is a martingale.

Observation 5.3.3 (Scaled intersection is a martingale)**.** Let

$$\boldsymbol{L}_t = L\cap\bigcap_{i=1}^{t}\boldsymbol{S}_i \text{ and } \boldsymbol{R}_t := \frac{\rho(\boldsymbol{L}_t)}{\rho(L)\prod_{i=1}^{t}\rho(\boldsymbol{S}_i)}$$

as introduced in Lemma 5.3.1. Then $(\boldsymbol{R}_t)_{t\in[k]}$ is a martingale with respect to the filtration $(\mathcal{V}_t)_{t\in[k]}$ induced by $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots$, with

$$\mathbb{E}[\boldsymbol{R}_t \mid \mathcal{V}_{t-1}] = \boldsymbol{R}_{t-1} \text{ and } |\boldsymbol{R}_t| \leq \frac{1}{p^j(1-p)^{k-j}}$$

Proof. The quantities $\rho(\boldsymbol{S}_i)$ are fixed for all i. $\rho(\boldsymbol{L}_t) \leq \rho(L)$ implies $\boldsymbol{R}_t \leq \frac{1}{p^j(1-p)^{k-j}}$.

By definition $\boldsymbol{R}_t = \frac{\rho(\boldsymbol{S}_t\cap\boldsymbol{L}_{t-1})}{\rho(L)\cdot\prod_{i=1}^{t}\rho(\boldsymbol{S}_i)}$. Since $\boldsymbol{v}_t \sim \mathbb{S}^{d-1}$ independently of $\mathcal{V}_{t-1}$, $\mathbb{E}_{\boldsymbol{v}_t\sim\mathbb{S}^{d-1}}[\rho(\boldsymbol{S}_t\cap \boldsymbol{L}_{t-1})] = \rho(\boldsymbol{S}_t)\cdot\rho(\boldsymbol{L}_{t-1})$. The conclusion now follows, since $\mathbb{E}_{\boldsymbol{v}_t\sim\mathbb{S}^{d-1}}[\boldsymbol{R}_t] = \boldsymbol{R}_{t-1}$. □

We will next need concentration inequalities for martingales that arise as sums of sub-Gaussian random variables.

Definition 5.3.4. The *sub-Gaussian norm* of a real-valued random variable $\boldsymbol{X}$ is

$$\|\boldsymbol{X}\|_{\mathsf{gauss}} := \inf\left\{K>0 : \mathbb{E}\left[\exp(\boldsymbol{X}^2/K^2)\right]\leq 2\right\}.$$

We say $\boldsymbol{X}$ is a *sub-Gaussian random variable* if $\|\boldsymbol{X}\|_{\mathsf{gauss}} < \infty$.

We will need a version of Azuma's inequality for martingales with centered sub-Gaussian increments. A proof may be found in [Van14], Lemma 3.7.

Lemma 5.3.5 (sub-Gaussian martingale concentration)**.** There exists a constant $C_{5.3.5} > 0$ such that if $\boldsymbol{X}_0, \boldsymbol{X}_1, \ldots, \boldsymbol{X}_m$ is a martingale sequence with respect to a filtration $(\mathcal{V}_t)_{t\in[m]}$ and $K_i := \sup_{\mathcal{V}_{i-1}}\|\boldsymbol{X}_i - \boldsymbol{X}_{i-1}|\mathcal{V}_{i-1}\|_{\mathsf{gauss}} < \infty$ for all $i \in [m]$, then:

$$\Pr\left[|\boldsymbol{X}_m - \boldsymbol{X}_0| \geq t\right] \leq 2\exp\left(-C_{5.3.5}\cdot\frac{t^2}{\sum_{i=1}^{m}K_i^2}\right).$$

Also, we will make use of the following statement to bound the sub-Gaussian norm of a random variable in terms of its tail probabilities.

Lemma 5.3.6 (Proposition 2.5.2 of [Ver18])**.** There exists constant $C_{5.3.6} > 0$ such that if $\boldsymbol{X}$ is a random variable satisfying $\Pr[|\boldsymbol{X}| > t] \leq 2\exp\left(-\frac{t^2}{K^2}\right)$ for all $t \geq 0$, then

$$\|\boldsymbol{X}\|_{\mathsf{gauss}} \leq C_{5.3.6} \cdot K$$

Now, we will sketch the proof of concentration for the martingale introduced in Observation 5.3.3, using the concentration inequality for martingales with sub-Gaussian increments, Lemma 5.3.5. The $(\boldsymbol{R}_t)_{t\in[k]}$ do not quite have sub-Gaussian increments as described. Thus we must "tame" them by making some minor technical modifications.

Proof sketch for Lemma 5.3.1. Recall the martingale sequence $\boldsymbol{R}_t$ defined in Observation 5.3.3. We are then interested in deviation bounds for $|\boldsymbol{R}_k - 1|$. Our proof strategy is to couple $\boldsymbol{R}_t$ with a more well-behaved martingale sequence and use the sub-Gaussian martingale concentration inequality on the more well-behaved sequence. Let $\boldsymbol{T}$ be the first time at which $\boldsymbol{R}_{\boldsymbol{T}} > 2$ or $\boldsymbol{R}_{\boldsymbol{T}} < \frac{1}{2}$. Define the process $(\boldsymbol{Q}_t)_{t\geq 1}$ so that

$$\boldsymbol{Q}_i = \begin{cases} \boldsymbol{R}_i & i \leq \boldsymbol{T} \\ \boldsymbol{R}_{\boldsymbol{T}} & \text{otherwise.} \end{cases}$$

Note that the sequence $(\boldsymbol{Q}_t)_{t\geq 1}$ is a martingale. This is because when $t \leq \boldsymbol{T}$, $\boldsymbol{Q}_t - \boldsymbol{Q}_{t-1} \mid \mathcal{V}_{t-1}$ has the same distribution as $\boldsymbol{R}_t - \boldsymbol{R}_{t-1} \mid \mathcal{V}_{t-1}$ and hence $\mathbb{E}[\boldsymbol{Q}_t - \boldsymbol{Q}_{t-1} \mid \mathcal{V}_{t-1}] = 0$, and when $t \geq \boldsymbol{T}$, $\boldsymbol{Q}_t - \boldsymbol{Q}_{t-1} \mid \mathcal{V}_{t-1}$ is identically zero. We will need to make an additional modification on top of $\boldsymbol{Q}_i$ to make the random variable well-behaved, since Corollary 5.2.9 only guarantees that $\Pr[|\boldsymbol{Q}_i - \boldsymbol{Q}_{i-1}| > s]$ remains sub-Gaussian up to $s = 1$. We will truncate $\boldsymbol{Q}_i$ as described by the following definition:

Definition 5.3.7 ((α,β)-truncation)**.** Given a centered random variable $\boldsymbol{X}$, we define a random variable $\boldsymbol{X}_{\alpha,\beta}$ for $\alpha, \beta \in \mathbb{R}_{\geq 0}$. First, let $\theta = \min\left(\frac{\Pr[|\boldsymbol{X}|>\alpha]}{\beta}, 1\right)$. Now, define

$$\boldsymbol{X}_{\alpha,\beta} := \begin{cases} \boldsymbol{X} \mid |\boldsymbol{X}| \leq \alpha & \text{with probability } (1-\theta)\Pr[|\boldsymbol{X}| \leq \alpha] \\ \mathbb{E}\left[\boldsymbol{X} \mid |\boldsymbol{X}| > \alpha\right] \cdot \beta & \text{with probability } (1-\theta)\cdot\theta \\ 0 & \text{otherwise.} \end{cases}$$

$\boldsymbol{X}_{\alpha,\beta}$ is well-defined; in the case when $\theta = 1$ it takes value 0 deterministically, and otherwise if $\theta < 1$ the probabilities sum to 1.

We make two useful observations about (α,β)-truncations.

Observation 5.3.8. For a random variable $\boldsymbol{X}$ and parameters α, β, $\mathbb{E}\boldsymbol{X}_{\alpha,\beta} = (1-\theta)\cdot\mathbb{E}\boldsymbol{X}$.

Observation 5.3.9. For a random variable $\boldsymbol{X}$ and parameters α, β satisfying $\beta \leq 1$,

$$\mathrm{d}_{\mathrm{TV}}\left(\boldsymbol{X}, \boldsymbol{X}_{\alpha,\beta}\right) \leq 2\theta$$

For now, we take these observations without proof and proceed with the proof of Lemma 5.3.1. Define $\boldsymbol{\Delta}_i := \boldsymbol{Q}_i - \boldsymbol{Q}_{i-1}$, and define $\tilde{\boldsymbol{\Delta}}_i \mid \mathcal{V}_{i-1} := (\boldsymbol{\Delta}_i \mid \mathcal{V}_{i-1})_{\alpha,\beta}$ where

$$\alpha = \begin{cases} 1 & \text{if } i \leq j \\ \frac{p}{1-p} & \text{if } i > j. \end{cases}$$

and $\beta = p^2\alpha$. We have chosen α such that Corollary 5.2.9 guarantees that we can control the tail probability $\Pr[|\boldsymbol{\Delta}_i| > t]$ for $t < \alpha$. Since $|\boldsymbol{\Delta}_i| \leq \frac{1}{p}$, $|\mathbb{E}\left[\boldsymbol{\Delta}_i \mid |\boldsymbol{\Delta}_i| > \alpha\right]| \cdot \beta \leq p\alpha$, and consequently $|\tilde{\boldsymbol{\Delta}}_i| \leq \alpha$. Let $(\tilde{\boldsymbol{Q}}_t)_{t\geq 1}$ be the random process obtained by setting $\tilde{\boldsymbol{Q}}_1 = \boldsymbol{Q}_1$ and $\tilde{\boldsymbol{Q}}_t = \tilde{\boldsymbol{Q}}_{t-1} + \tilde{\boldsymbol{\Delta}}_t$ for $t \geq 2$. By Observation 5.3.8, $(\tilde{\boldsymbol{Q}}_t)_{t\geq 1}$ is a martingale.

Using Corollary 5.2.9 and Lemma 5.3.6, we can bound the sub-Gaussian norms of $\tilde{\boldsymbol{\Delta}}_i \mid \mathcal{V}_i$. Once we have the sub-Gaussian norms, we can use Lemma 5.3.5 to obtain high-probability upper bounds on $|\tilde{\boldsymbol{Q}}_i - 1|$. We lastly use Observation 5.3.9 to relate $\tilde{\boldsymbol{Q}}$ back to $\boldsymbol{R}$. □

We refer to [LMSY22b] for more proof details. The proofs of our observations regarding truncated variables can also be found in [LMSY22b]; we will omit them in this work.

Lastly, we present a convenient corollary of Lemma 5.3.1.

Corollary 5.3.10 (Area of $\approx pk$ caps and $\approx (1-p)k$ anti-caps)**.** Let $n, j, k, d \in \mathbb{Z}_{\geq 0}$, and let $p \in \mathbb{R}_{\geq 0}$, satisfying $\frac{1}{n^2} \ll p \leq \frac{1}{2}$, $1 \leq d \leq n^{100}$, $j \leq k \leq n$, with $j = pk + \Delta$. Sample $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$ uniformly from $\mathbb{S}^{d-1}$ and let $\boldsymbol{L} = \left(\bigcap_{i=1}^{j} \mathsf{cap}(\boldsymbol{v}_i)\right) \cap \left(\bigcap_{i=j+1}^{k} \overline{\mathsf{cap}}(\boldsymbol{v}_i)\right)$. Then there exist constants $C_{5.3.10}, C'_{5.3.10} > 0$ such that for all n sufficiently large,

$$\Pr\left[\left|\frac{\rho(\boldsymbol{L})}{e^{-kH(p)}\left(\frac{p}{1-p}\right)^{\Delta}} - 1\right| \geq t\right] \leq \exp\left(-C_{5.3.10} \cdot \min\left(\frac{dt^2}{M(k,p,\Delta)^2 \ln n}, \frac{d}{M(k,p,\Delta)^2 \ln n} - C'_{5.3.10} \log n\right)\right)$$

where $M(k,p,\Delta) = \max(kH(p), |\Delta| \ln \frac{1}{p}, \ln n)$.

We omit the proof and refer to [LMSY22b] for the details.

5.4 Total variation bound

In this section, we prove Theorem 5.1.3 and Theorem 5.1.4. As was done in the prior work by Brennan, Bresler, and Nagaraj [BBN20], we use an analogue of the tensorization of the relative entropy for non-product measures:

Claim 5.4.1 (Relative entropy tensorization, similar to Lemma 2.1 of [BBN20])**.** Suppose $\mu = \mu_1 \otimes \cdots \otimes \mu_n$ is a product measure and ν is a measure over the same domain. Let ν_t denote the marginal of ν on the t-th coordinate x_t, and let $x_{a:b}$ denote coordinates a through b of x. Then

$$\mathcal{D}_{\mathrm{KL}}(\nu \,\|\, \mu) = \sum_{t=1}^{n} \mathbb{E}_{x_{1:t-1}\sim\nu} \left[\mathcal{D}_{\mathrm{KL}}\left(\nu_t(x_t \mid x_{1:t-1}) \,\|\, \mu_t\right)\right].$$

Proof. By the chain rule for relative entropy,

$$\begin{aligned}\mathbb{E}_{\boldsymbol{x}\sim\nu} \log \frac{\nu(\boldsymbol{x})}{\mu(\boldsymbol{x})} &= \sum_{t=1}^{n} \mathbb{E}_{\boldsymbol{x}\sim\nu} \log \frac{\nu_t(\boldsymbol{x}_t \mid \boldsymbol{x}_{1:t-1})}{\mu_t(\boldsymbol{x}_t)} \\ &= \sum_{t=1}^{n} \mathbb{E}_{\boldsymbol{x}_{1:t-1}\sim\nu} \left(\mathbb{E}_{\boldsymbol{x}_t\sim\nu_t|\boldsymbol{x}_{1:t-1}} \log \frac{\nu_t(\boldsymbol{x}_t \mid \boldsymbol{x}_{1:t-1})}{\mu_t(\boldsymbol{x}_t)}\right),\end{aligned}$$

by linearity of expectation and by definition of the marginal distribution. The expression on the right simplifies using the definition of the relative entropy, completing the proof. □

In combination with Pinsker's inequality, this lemma reduces bounding the TV distance between a product measure μ and a general measure ν, to bounding the relative entropy $\mathcal{D}_{\mathrm{KL}}\left(\nu_t(x_t \mid x_{1:t-1}) \,\|\, \mu_t\right)$.

Claim 5.4.2. Let μ_t be the distribution of the neighborhood of t to vertices $[t-1]$ under $\mathsf{G}(n,p)$, and let $\nu_t(\cdot \mid G_{t-1})$ be the distribution of the neighborhood of t under $\mathsf{Geo}_d(n,p)$, conditioned on the subgraph $\boldsymbol{G}_{t-1} \sim \mathsf{Geo}_d(t-1,p)$ on the vertices $[t-1]$. Then,

$$2\mathrm{d}_{\mathrm{TV}}\left(\mathsf{Geo}_d(n,p), \mathsf{G}(n,p)\right)^2 \le n \cdot \mathbb{E}_{\boldsymbol{G}_{n-1}\sim\mathsf{Geo}_d(n-1,p)} \left[\mathcal{D}_{\mathrm{KL}}\left(\nu_n(\cdot \mid \boldsymbol{G}_{n-1}) \,\|\, \mu_t\right)\right]$$

Proof. Applying Pinsker's inequality (Theorem 1.3.1), $2 \cdot \mathrm{d}_{\mathrm{TV}}\left(\nu, \mu\right)^2 \le \mathcal{D}_{\mathrm{KL}}(\nu \,\|\, \mu)$. We then apply Claim 5.4.1 with $\mu = \mathsf{G}(n,p)$ and $\nu = \mathsf{Geo}_d(n,p)$, and μ_t, ν_t as defined in the claim:

$$2\mathrm{d}_{\mathrm{TV}}\left(\mathsf{Geo}_d(n,p), \mathsf{G}(n,p)\right)^2 \le \sum_{t=1}^{n} \mathbb{E}_{\boldsymbol{G}_{t-1}\sim\mathsf{Geo}_d(t-1,p)} \left[\mathcal{D}_{\mathrm{KL}}\left(\nu_t(\cdot \mid \boldsymbol{G}_{t-1}) \,\|\, \mu_t\right)\right]$$

Let G_S denote a graph over vertices S, and let ν_t^S, μ_t^S refer to the distribution of vertex t's neighbors in S under $\mathsf{Geo}_d(n,p)$ and $\mathsf{G}(n,p)$ respectively. By symmetry,

$$\mathbb{E}_{x_{[n]\setminus t}\sim\nu} \left[\mathcal{D}_{\mathrm{KL}}\left(\nu_t^{[n]\setminus t}(\cdot \mid x_{[n]\setminus t}) \,\|\, \mu_n\right)\right]$$

is the same for all $t \in [n]$. Via the chain rule for relative entropy, and the non-negativity of relative entropy,

$$\begin{aligned}\mathcal{D}_{\mathrm{KL}}\left(\nu_t^{[n]\setminus t}(\cdot \mid G_{[n]\setminus t}) \,\|\, \mu_n\right) &\ge \mathcal{D}_{\mathrm{KL}}\left(\nu_t^{[t-1]}(\cdot \mid G_{[n]\setminus t}) \,\|\, \mu_t^{[t-1]}\right) \\ &= \mathcal{D}_{\mathrm{KL}}\left(\nu_t^{[t-1]}(\cdot \mid G_{t-1}) \,\|\, \mu_t^{[t-1]}\right)\end{aligned}$$

The final equality comes from the fact that t's neighbors in $[t-1]$ only depends on $\boldsymbol{G}_{t-1}$.

Upper bounding each $\mathcal{D}_{\mathrm{KL}}\left(\nu_t(\cdot \mid \boldsymbol{G}_{t-1}) \,\|\, \mu_t\right)$ by $\mathcal{D}_{\mathrm{KL}}\left(\nu_n(\cdot \mid \boldsymbol{G}_{n-1}) \,\|\, \mu_n\right)$ completes the proof. □

5.4.1 Bounding neighborhood relative entropy

Via Claim 5.4.2, our goal now is to upper bound:

$$\mathbb{E}_{\boldsymbol{G}_{n-1}\sim\mathsf{Geo}_d(n-1,p)}\left[\mathcal{D}_{\mathrm{KL}}\left(\nu_n(\cdot \mid \boldsymbol{G}_{n-1}) \,\|\, \mu_t\right)\right] = \mathbb{E}_{\boldsymbol{G}_{n-1}\sim\nu_{[n-1]}}\mathbb{E}_{\boldsymbol{S}\sim\nu_n(\cdot|G_{n-1})} \ln \frac{\nu_n(\boldsymbol{S} \mid G_{n-1})}{\mu_n(\boldsymbol{S})}$$
$$\leq o\left(\frac{1}{n}\right)$$

For most events under $\nu_{[n-1]}$ and $\nu_n(\cdot|G_{n-1})$, we will upper bound the relative entropy via a Chi-square-like quantity: we use the Chi-squared distance, but we allow the omission of a low-probability event $\mathcal{E}$ (to allow the removal of events which cause the Chi-square distance to blow up).

We then specialize the resulting Chi-square-like bound (Lemma 5.4.3) by removing different low-probability events for the general p case (Section 5.4.1) and the sparse case, and separately conclude Theorem 5.1.3 and Theorem 5.1.4. We now formally state the Chi-square bound:

Lemma 5.4.3. Let $\mathcal{E}$ be an event satisfying both

$$\Pr_{\boldsymbol{G}_{n-1}\sim\mathsf{Geo}_d(n-1,p),\boldsymbol{S}\sim\nu_n(\cdot|\boldsymbol{G}_{n-1})}[\mathcal{E}] \leq o\left(\frac{1}{n^2\ln n}\right) \text{ and } \Pr_{\boldsymbol{G}_{n-1}\sim\mathsf{Geo}_d(n-1,p),\boldsymbol{S}\sim\mu_n}[\mathcal{E}] \leq o\left(\frac{1}{n^2\ln n}\right)$$

and for $S \subseteq [n-1]$, define $\Delta_{G_{n-1}}(S) = \frac{\nu_n(S|G_{n-1})}{\mu_n(S)} - 1$. Then,

$$\mathbb{E}_{\boldsymbol{G}_{n-1}\sim\nu_{[n-1]}}\mathbb{E}_{\boldsymbol{S}\sim\nu_n(\cdot|\boldsymbol{G}_{n-1})} \ln \frac{\nu_n(\boldsymbol{S} \mid \boldsymbol{G}_{n-1})}{\mu_n(\boldsymbol{S})} \leq \mathbb{E}_{\boldsymbol{G}_{n-1}\sim\nu_{[n-1]}}\mathbb{E}_{\boldsymbol{S}\sim\mu_n}\Delta_{\boldsymbol{G}_{n-1}}(\boldsymbol{S})^2 \cdot \mathbb{1}(\overline{\mathcal{E}}) + o\left(\frac{1}{n}\right)$$

Proof. Before we introduce the Chi-square bound, we first perform some conditioning on $\overline{\mathcal{E}}$.

$$\mathbb{E}_{\boldsymbol{G}_{n-1}\sim\nu_{[n-1]}}\mathbb{E}_{\boldsymbol{S}\sim\nu_n(\cdot|\boldsymbol{G}_{n-1})} \ln \frac{\nu_n(\boldsymbol{S} \mid \boldsymbol{G}_{n-1})}{\mu_n(\boldsymbol{S})} = \mathbb{E}_{\boldsymbol{G}_{n-1}\sim\nu_{[n-1]}}\mathbb{E}_{\boldsymbol{S}\sim\nu_n(\cdot|\boldsymbol{G}_{n-1})} \ln\left(\Delta_{G_{n-1}}(\boldsymbol{S})+1\right)$$
$$= \mathbb{E}_{\boldsymbol{G}_{n-1}\sim\nu_{[n-1]}}\mathbb{E}_{\boldsymbol{S}\sim\nu_n(\cdot|\boldsymbol{G}_{n-1})} \ln\left(\Delta_{\boldsymbol{G}_{n-1}}(\boldsymbol{S})+1\right) \cdot \mathbb{1}(\overline{\mathcal{E}})$$
$$+ \max_S\left(\ln \frac{\nu_n(S \mid \boldsymbol{G}_{n-1})}{\mu_n(S)}\right) \cdot \Pr_{\mathsf{Geo}_d(n,p)}[\mathcal{E}]$$

The maximum of $\ln \frac{\nu_n(S|G_{n-1})}{\mu_n(S)}$ is upper bounded by $n \ln \frac{1}{p}$, which comes from bounding

$$\nu_n(S \mid G_{n-1}) \leq 1$$

achieving the smallest possible $\mu_n(S)$ when $|S| = n-1$. Applying our assumption on $\Pr[\mathcal{E}]$, the term $\max\left(\ln \frac{\nu_n(S|G_{n-1})}{\mu_n(S)}\right) \cdot \Pr[\mathcal{E}]$ is at most $o\left(\frac{1}{n}\right)$.

We now turn our attention to the remaining expectation term:

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{G}_{n-1}\sim\nu_{[n-1]}}\mathbb{E}_{\boldsymbol{S}\sim\nu_n(\cdot|G_{n-1})} \ln\left(\Delta_{G_{n-1}}(\boldsymbol{S})+1\right)\cdot\mathbb{1}(\overline{\mathcal{E}}) \\
&\qquad\le \mathbb{E}_{\boldsymbol{G}_{n-1}\sim\nu_{[n-1]}}\mathbb{E}_{\boldsymbol{S}\sim\nu_n(\cdot|G_{n-1})}\Delta_{G_{n-1}}(\boldsymbol{S})\cdot\mathbb{1}(\overline{\mathcal{E}}) \\
&\qquad= \mathbb{E}_{\boldsymbol{G}_{n-1}\sim\nu_{[n-1]}}\mathbb{E}_{\boldsymbol{S}\sim\mu_n}(1+\Delta_{G_{n-1}}(\boldsymbol{S}))\Delta_{G_{n-1}}(\boldsymbol{S})\cdot\mathbb{1}(\overline{\mathcal{E}}) \\
&\qquad= \mathbb{E}_{\boldsymbol{G}_{n-1}\sim\nu_{[n-1]}}\mathbb{E}_{\boldsymbol{S}\sim\mu_n}\Delta_{G_{n-1}}(\boldsymbol{S})\cdot\mathbb{1}(\overline{\mathcal{E}}) + \mathbb{E}_{\boldsymbol{G}_{n-1}\sim\nu_{[n-1]}}\mathbb{E}_{\boldsymbol{S}\sim\mu_n}\Delta_{G_{n-1}}(\boldsymbol{S})^2\cdot\mathbb{1}(\overline{\mathcal{E}})
\end{aligned}$$

The inequality follows because $1+x\le e^x$ for all x, and the second equality follows from a change in the randomness of S, and because $(1+\Delta_{G_{n-1}}(S)) = \frac{\nu_n(S|G_{n-1})}{\mu_n(S)}$.

By the definition of $\Delta_{G_{n-1}}$, we can further simplify the expectation of $\Delta_{G_{n-1}}(S)\cdot\mathbb{1}(\overline{\mathcal{E}})$:

$$\begin{aligned}
\left|\mathbb{E}_{\boldsymbol{G}_{n-1}\sim\nu_{[n-1]}}\mathbb{E}_{\boldsymbol{S}\sim\mu_n}\Delta_{G_{n-1}}(\boldsymbol{S})\cdot\mathbb{1}(\overline{\mathcal{E}})\right| &= \left|\Pr_{\boldsymbol{G}_{n-1}\sim\mathsf{Geo}_d(n-1,p),\boldsymbol{S}\sim\mu_n}[\overline{\mathcal{E}}] - \Pr_{\boldsymbol{G}_{n-1}\sim\mathsf{Geo}_d(n-1,p),\boldsymbol{S}\sim\nu_n(\cdot|\boldsymbol{G}_{n-1})}[\mathcal{E}]\right| \\
&\le \Pr_{\boldsymbol{G}_{n-1}\sim\mathsf{Geo}_d(n-1,p),\boldsymbol{S}\sim\mu_n}[\mathcal{E}] + \Pr_{\boldsymbol{G}_{n-1}\sim\mathsf{Geo}_d(n-1,p),\boldsymbol{S}\sim\nu_n(\cdot|\boldsymbol{G}_{n-1})}[\mathcal{E}] \\
&\le o\left(\frac{1}{n^2\ln n}\right)
\end{aligned}$$

□

Remark 5.4.4. We will ultimately choose the event $\mathcal{E}$ based on when the Chi-square estimate is too loose of an upper bound on the relative entropy. If the conditions of Lemma 5.4.3 hold, the overall TV bound we want would follow from

$$\mathbb{E}_{\boldsymbol{G}_{n-1}\sim\mathsf{Geo}_d(n-1,p)}\mathbb{E}_{\boldsymbol{S}\sim\mathsf{Binom}(n-1,p)}\left[\left(\frac{\nu_n(\boldsymbol{S}\mid\boldsymbol{G}_{n-1})}{p^{|\boldsymbol{S}|}(1-p)^{n-1-|\boldsymbol{S}|}}-1\right)^2\cdot\mathbb{1}(\overline{\mathcal{E}})\right] = o\left(\frac{1}{n}\right). \tag{5.6}$$

The general p case

The goal of this section is to prove Theorem 5.1.3.

Proof. We apply Lemma 5.4.3, when $\mathcal{E}$ is the failure of degree concentration: $|S|\ge pn+\Delta$, with $\Delta = 10\max(\log n, pn)$. Via Lemma 4.2.3, $\mathcal{E}$ has probability $O(n^{-3})$, regardless of whether $S\sim\mu_n$ or $S\sim\nu_n(\cdot|\boldsymbol{G}_{n-1})$ and $\boldsymbol{G}_{n-1}\sim\mathsf{Geo}_d(n-1,p)$. (The latter distribution is equivalent to $\boldsymbol{G}_n\sim\mathsf{Geo}$.) We thus satisfy:

$$\Pr_{\boldsymbol{G}_{n-1}\sim\mathsf{Geo}_d(n-1,p),S\sim\mu_n}[\mathcal{E}], \Pr_{\boldsymbol{G}_{n-1}\sim\mathsf{Geo}_d(n-1,p),S\sim\nu_n(\cdot|\boldsymbol{G}_{n-1})}[\mathcal{E}] \le o\left(\frac{1}{n^2\ln n}\right)$$

After the reduction to Equation 5.6 in Remark 5.4.4, we complete the proof by bounding:

$$\mathbb{E}_{\boldsymbol{G}_{n-1}\sim\nu_{[n-1]}}\mathbb{E}_{\boldsymbol{S}\sim\mathsf{Binom}(n-1,p)}\left[\left(\frac{\nu_n(S\mid\boldsymbol{G}_{n-1})}{p^{|S|}(1-p)^{n-1-|S|}}-1\right)^2\cdot\mathbb{1}(\overline{\mathcal{E}})\right]$$
$$\leq\int_0^\infty\Pr_{\boldsymbol{G}_{n-1}\sim\nu_{[n-1]},\boldsymbol{S}\sim\mathsf{Binom}(n-1,p)}\left[\left(\frac{\nu_n(S\mid\boldsymbol{G}_{n-1})}{p^{|S|}(1-p)^{n-1-|S|}}-1\right)^2>t\mid\overline{\mathcal{E}}\right]dt$$

We now apply Corollary 5.3.10 to control these tail probabilities. By conditioning on $\overline{\mathcal{E}}$, we may assume $\Delta=10\max(pn,\log n)$ in the tail bound, as this choice of Δ leads to the worst case tail probability for all $|S|\leq pn+\Delta$. Recall that $M(n,p,\Delta)=\max(nH(p),|\Delta|\ln\frac{1}{p},\ln n)$ in Corollary 5.3.10, and note that $p^{|S|}(1-p)^{n-1-|S|}=e^{-nH(p)}\cdot\left(\frac{p}{1-p}\right)^\Delta$.

$$\int_0^\infty\Pr\left[\left(\frac{\nu_n(S\mid\boldsymbol{G}_{n-1})}{p^{|S|}(1-p)^{n-1-|S|}}-1\right)^2>t\mid\overline{\mathcal{E}}\right]dt$$
$$\leq\int_0^{\frac{1}{n\log n}}1\cdot dt+\int_{\frac{1}{n\log n}}^1\exp\left(\frac{-C_{5.3.10}d\cdot t}{M(n,p,\Delta)^2\ln n}\right)dt$$
$$+\int_1^{e^{2nH(p)}\cdot\left(\frac{p}{1-p}\right)^\Delta}\exp\left(\frac{-C_{5.3.10}d}{M(n,p,\Delta)^2\ln n}+C'_{5.3.10}\log n\right)dt$$
$$\leq\frac{1}{n\log n}-\frac{M(n,p,\Delta)^2\ln n}{C_{5.3.10}d}\exp\left(\frac{-C_{5.3.10}d\cdot t}{M(n,p,\Delta)^2\ln n}\right)\Big|_{\frac{1}{n\log n}}^1$$
$$+e^{2nH(p)}\cdot\left(\frac{p}{1-p}\right)^\Delta\cdot\exp\left(\frac{-C_{5.3.10}d}{M(n,p,\Delta)^2\ln n}+C'_{5.3.10}\log n\right)$$

In the first line, we split the integral based on the appropriate tail bound expression in Corollary 5.3.10, and also remark that it suffices to consider $t\leq e^{2nH(p)}\cdot\left(\frac{p}{1-p}\right)^\Delta$, as $\frac{\nu_n(S|\boldsymbol{G}_{n-1})}{p^{|S|}(1-p)^{n-1-|S|}}$ is maximized at that value when $|S|\leq np+\Delta$. If we choose $d\geq n\cdot M(n,p,\Delta)^2\cdot\ln^3 n$, we can bound each summation term individually to obtain the desired bound on $\mathbb{E}_{\boldsymbol{G}_{n-1}\sim\nu_{[n-1]}}\mathbb{E}_{\boldsymbol{S}\sim\mathsf{Binom}(n-1,p)}\left[\left(\frac{\nu_n(S|\boldsymbol{G}_{n-1})}{p^{|S|}(1-p)^{n-1-|S|}}-1\right)^2\cdot\mathbb{1}(\overline{\mathcal{E}})\right]$.

$$\int_0^\infty\Pr\left[\left(\frac{\nu_n(S\mid\boldsymbol{G}_{n-1})}{p^{|S|}(1-p)^{n-1-|S|}}-1\right)^2>t\mid\overline{\mathcal{E}}\right]dt\leq\frac{1}{n\log n}+\frac{1}{n}\cdot e^{-\log n}+e^{2nH(p)}\cdot e^{-Cn\ln^2 n}$$
$$\leq o\left(\frac{1}{n}\right)$$

We note that $M(n,p,\Delta)\leq\max(np\ln n,\ln^2 n)$ for $\frac{1}{n}\leq p\leq\frac{1}{2}$, so we can restate our requirement on d as $d=\tilde{\Omega}(n^3p^2)$. □

5.5 Belief propagation over a continuous alphabet

In this section, we sketch some of the techniques used to prove Theorem 5.1.4. Full proofs are in [LMSY22b]. As mentioned in Section 5.1.2, to obtain stronger bounds on $\Pr_{\boldsymbol{G}\sim\mathsf{Geo}_d(n,p)}[N_{\boldsymbol{G}}(n) = S]$, we need to *average* over different vector embeddings that still form $\boldsymbol{G}$.

Let's make this a bit more concrete. Say we instead want to compare

$$\Pr_{\boldsymbol{G}\sim\mathsf{Geo}_d(n,p)}[N_{\boldsymbol{G}}(n) \supseteq S] \text{ to } \Pr_{\boldsymbol{G}\sim\mathsf{G}(n,p)}[N_{\boldsymbol{G}}(n) \supseteq S] = p^{|S|}$$

Without loss of generality, let $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{|S|}\}$ be the vectors representing S. If somehow, conditioned on $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{n-1}\}$ forming $\boldsymbol{G}_{n-1}$, the vectors $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{|S|}\}$ were independent and each distributed according to ρ, then we would conclude that $\Pr_{\boldsymbol{G}\sim\mathsf{Geo}_d(n,p)}[N_{\boldsymbol{G}}(n) \supseteq S] = p^{|S|}$ as well. Sadly, this is incorrect: $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{n-1}\}$ are *not* independent so long as $\boldsymbol{G}_{n-1}$ is connected, and the dependences between them are potentially complex, based on how they are connected within $\boldsymbol{G}_{n-1}$.

However, if there are no edges among them (which is the case when $p = \Theta\left(\frac{\alpha}{n}\right)$, it *is* the case that $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{n-1}\}$ are independent conditioned on fixing $\{\boldsymbol{v}_{|S|+1}, \ldots, \boldsymbol{v}_{n-1}\}$. This is good news, but conditioning on the placements of the entire set $\{\boldsymbol{v}_{|S|+1}, \ldots, \boldsymbol{v}_{n-1}\}$ is prohibitive. The marginals of each vector in $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{|S|}\}$ will not resemble ρ.

The game here is to condition on enough vectors to treat $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{n-1}\}$ independently, but also have these vectors correspond to vertices far away from S, so the marginals of each vector in $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{n-1}\}$ can get as close as possible to ρ. We found that when $p = \Theta\left(\frac{1}{n}\right)$, the $\frac{\log n}{\log\log n}$-neighborhoods around each $\boldsymbol{v}_i$ for $i \in S$ will be disconnected, and thus independent. To estimate the *marginal distributions* of each $\boldsymbol{v}_i$ for $i \in S$, we use the belief propagation algorithm, detailed below.

5.5.1 Basics of CSPs and belief propagation

Definition 5.5.1. A *constraint satisfaction problem instance* (CSP instance) $\mathcal{I}$ consists of a *variable set* V and a *constraint set* E:

- The variables $v \in V$ each belong to an alphabet Σ.
- The constraints $f \in E$ consist of a k-tuple of variables ∂f and a function

$$\psi_f : \Sigma^k \to \{0, 1\}$$

  Here, a value of 1 corresponds to a constraint being *satisfied*.

An assignment to variables $c : V \to \Sigma$ is *satisfying* if for each constraint $f \in F$,

$$\psi_f(c(\partial f)) = 1$$

Definition 5.5.2. We can represent any CSP instance $\mathcal{I}$ as a bipartite graph F, which we call a *factor graph*. The two sides of the bipartition are V and E, and we place an edge between $v \in V$ and $f \in E$ if the variable v participates in constraint f. We use ∂v and ∂f to denote the neighborhoods of variables v and clauses f, respectively, in this graph.

When the factor graph F does not contain cycles, the marginal on a variable v can be computed exactly from the fixed point of the belief propagation algorithm.

Definition 5.5.3. A *belief propagation fixed point* for a factor graph F is a collection of *messages*

$$\left\{m^{v\to f}, m^{f\to v}\right\}_{v\in V, f\in E}$$

for all pairs v, f such that f is a neighbor of v, where each message is a *probability distribution* on Σ, such that

$$m^{f\to v}(x) \propto \int_{c|_{\partial f}: c(v)=x} \psi_f(c|_{\partial f}) \prod_{v'\in\partial f\setminus v} m^{v'\to f}(c(v')) \tag{5.7}$$

$$m^{v\to f}(x) \propto \prod_{f'\in\partial v\setminus f} m^{f'\to v}(x) \tag{5.8}$$

Theorem 5.5.1 (Theorem 14.1 of [MM09])**.** Suppose F is a forest factor graph corresponding to a CSP instance $\mathcal{I}$ where every vertex is attached to a unary constraint. Then there is a unique belief propagation fixed point and the marginal distribution ν on variable v over the uniform distribution over satisfying assignments to i is given via the following formula.

$$\nu \propto \prod_{f\in\partial v} m^{f\to v}.$$

5.5.2 Belief propagation for sparse random geometric graphs

We want to set up a CSP that captures configurations of vectors $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{n-1}\}$ that form a particular graph $\boldsymbol{G}_{n-1}$. Recall that we want to understand the marginals of vectors that are within a $\frac{\log n}{\log\log n}$-depth neighborhood around S. Let K denote this set of vectors.

For vertices outside of K, we fix their vectors (i.e. they follow a Dirac delta distribution). Let $R = [n-1] \setminus K$ represent these vectors.

- The alphabet Σ is the unit sphere $\mathbb{S}^{d-1}$, and the variables are vectors $\{\boldsymbol{v}_i\}$ for $i \in K$.
- **Unary constraints:** We require each v_i to respect their non-edges (and sometimes, edges) to the vertices in R. Define

$$\boldsymbol{L}_i \coloneqq \left(\bigcap_{j\in R\setminus N(i)} \mathsf{cap}(\boldsymbol{v}_j)\right) \cap \left(\bigcap_{j\in R\cap N(i)} \overline{\mathsf{cap}}(\boldsymbol{v}_j)\right)$$

  Thus, we have constraints of the form $f(\boldsymbol{v}_i) = 1$ if $\boldsymbol{v}_i \in \boldsymbol{L}_i$ and $f(\boldsymbol{v}_i) = 0$ otherwise. Such f are *unary*; they do not depend on multiple variables.

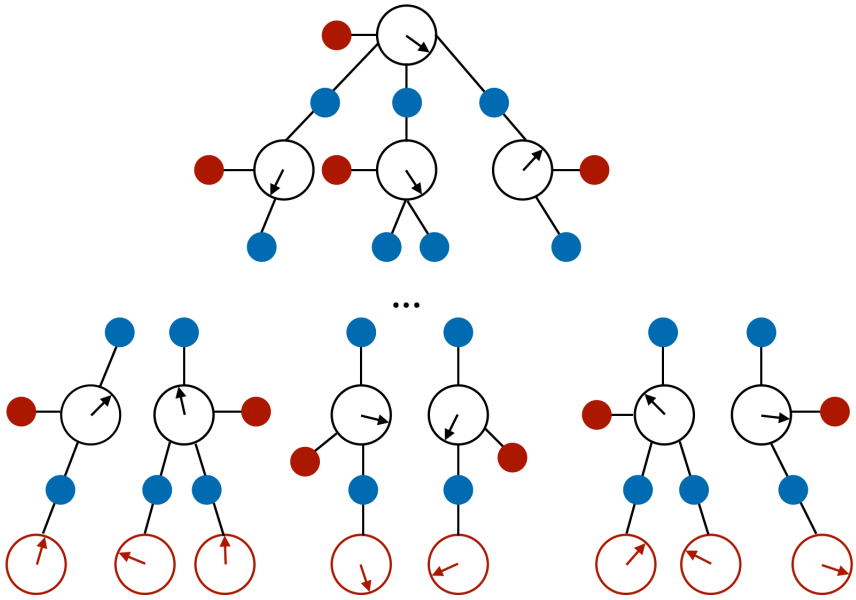

Figure 5.2: The factor graph without non-edge constraints. Here, all circles with arrows represent variable nodes, and solid circles represnet constraint nodes. The red variables represent *fixed* vectors. The red solid circles are the unary constraints (which are non-edge constraints to fixed vectors) while the blue solid circles are the edge constraints.

- **Edge constraints:** For each edge (i, j) in $\boldsymbol{G}_{n-1}$, we define a constraint f such that $f(\boldsymbol{v}_i, \boldsymbol{v}_j) = 1$ when $\langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle \geq \tau(p, d)$ and $f(\boldsymbol{v}_i, \boldsymbol{v}_j) = 0$ otherwise.
- **Non-edge constraints:** For each non-edge (i, j), we have a constraint f that equals 1 when $\langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle < \tau(p, d)$ and 0 otherwise (i.e. the opposite of the edge constraints).

For the rest of the chapter, we will assume the following simplifications:

Assumption 5.5.4. It suffices to consider the above CSP without non-edge contraints.

When $p = \Theta\left(\frac{1}{n}\right)$, the likelihood that two random vectors do not have an edge is very close to 1, and the non-edge constraints actually do not affect the solution space in a noticeable way. Specifically, a random solution for the CSP *without* the non-edge constraints is also a solution for the CPS *with* non-edge constraints with high probability. (We formally prove this in Section 7.4 of [LMSY22b].)

Under these assumptions, and conditioned on $G_{n-1}[K]$ being a forest, our factor graph will also be a forest, depicted in Figure 5.2.

In our setting, the BP equations are the following:

Variable-to-constraint messages:

$$m^{v\to f} = \frac{\prod_{e\in\partial v\setminus f} m^{e\to v}}{\int \prod_{e\in\partial v\setminus f} m^{e\to v}(x)d\rho(x)} \tag{5.9}$$

Unary constraint-to-variable messages:

$$m^{f\to v} = \frac{f}{\int f(x)d\rho(x)} \tag{5.10}$$

Edge constraint-to-variable messages: Let $\partial f = \{v, w\}$, then:

$$m^{f\to v}(x_v) = \frac{\int f(x_v, x_w)\cdot m^{w\to f}(x_w)d\rho(x_w)}{\int\int f(x_v, x_w)\cdot m^{w\to f}(x_w)d\rho(x_w)d\rho(x_v)} \tag{5.11}$$

Perhaps the unary constraint-to-variable messages are simplest to interpret; they are simply the uniform distribution over $\boldsymbol{L}_i$, which depends on fixed (in the sense they are not BP variables) but randomly chosen vectors corresponding to R.

Edge constraint-to-vertex messages as convolutions

The denominator in the case of edge constraint-to-vertex messages further simplifies:

$$\begin{aligned}\int\int f(x_v, x_w)\cdot m_t^{w\to f}(x_w)d\rho(x_v)d\rho(x_w) &= \int m_t^{w\to f}(x_w)\int \mathbb{1}[\langle x_v, x_w\rangle \geq \tau(p,d)]d\rho(x_v)d\rho(x_w)\\ &= p\int m_t^{w\to f}(x_w)d\rho(x_w)\\ &= p\end{aligned}$$

The final equality comes from the fact that each message $m_t^{w\to f}$ is a distribution. For binary constraints f such that $\partial f = \{v, w\}$, we can rewrite the constraint-to-vertex messages as:

$$m_{t+1}^{f\to v}(x_v) = \int \frac{f(x_v, x_w)}{p} m_t^{w\to f}(x_w)d\rho(x_w),$$

which, in particular, can be written as $Pm^{w\to f}$ for a linear operator P, defined as follows:

Definition 5.5.5. Let P be the linear operator defined so that for any function $h : \mathbb{S}^{d-1}\to\mathbb{R}$,

$$Ph(x) = \frac{1}{p}\int_{\mathsf{cap}_p(x)} h(y)\, d\rho(y),$$

which we alternately denote $\frac{1}{p}\cdot h(\mathsf{cap}_p(x))$.

In words, the operator P convolves its input with the uniform distribution over a spherical cap. Since P is a convolution operator, it preserves the ℓ_1-norm of nonnegative functions.

Observation 5.5.6. Suppose ν is a nonnegative function, then $\|P\nu\|_1 = \|\nu\|_1$. Additionally, for an arbitrary function ν, $\|P\nu\|_1 \le \|\nu\|_1$.

When ν is a distribution in particular, P can be construed as the transition operator of the Markov chain on $\mathbb{S}^{d-1}$ where a single step entails walking from v to a uniformly random point in $\mathsf{cap}(v)$. The following useful observation is immediate from how we define P.

Observation 5.5.7. For any function ν, $\|P\nu\|_\infty \le \frac{\|\nu\|_1}{p}$. This is because $\nu(\mathsf{cap}(z)) \le \|\nu\|_1$ for any $z \in \mathbb{S}^{d-1}$. In particular, if ν is a distribution, $\|P\nu\|_\infty \le \frac{1}{p}$.

We will to show that P "flattens" the distribution; to quantify this "flatness," we study the *spread* of the distribution.

Definition 5.5.8. Given a function ν on $\mathbb{S}^{d-1}$, its *deviation profile* is

$$\mathrm{Dev}_\nu(\varepsilon) := \Pr_{\boldsymbol{z}\sim\rho}[\nu(\boldsymbol{z}) \notin \mathbb{E}_{\boldsymbol{y}\sim\rho}[\nu(\boldsymbol{y})] \pm \varepsilon \cdot \|\nu\|_1]$$

The *spread profile* of ν, denoted $\mathrm{Spr}_\nu(\delta)$, is

$$\mathrm{Spr}_\nu(\delta) := \inf\{\varepsilon \in \mathbb{R}_{\ge 0} : \mathrm{Dev}_\nu(\varepsilon) \le \delta\}.$$

Smaller spread implies tighter concentration of $\nu(\boldsymbol{z})$ around its mean. Below, we see that applying P reduces the spread of any function ν on the unit sphere by a factor of $\tilde{O}\left(\frac{1}{\sqrt{d}}\right)$.

Lemma 5.5.9. Let ν be any function on $\mathbb{S}^{d-1}$ with $\mathbb{E}_{\boldsymbol{y}\sim\rho}[\nu(\boldsymbol{y})] = 0$. If

$$d \ge \log^{10} n \cdot \left(1 + \log^3 \frac{\|\nu\|_\infty}{\|\nu\|_1}\right)$$

Then:

$$\mathrm{Dev}_{P\nu}\left(\varepsilon \cdot \frac{\|\nu\|_1}{\|P\nu\|_1}\right) \le n^{-\log^4 n}$$

where $\varepsilon := \sqrt{\frac{C\log^{11} n}{d}} \cdot \left(2\ln\frac{\|\nu\|_\infty}{\|\nu\|_1} + 8\ln\frac{d}{p}\right)$ for C an absolute constant. Thus,

$$\mathrm{Spr}_{P\nu}(n^{-\log^4 n}) \le \varepsilon \cdot \frac{\|\nu\|_1}{\|P\nu\|_1}$$

Now, let $\boldsymbol{L} := \bigcap_{j\in R} \overline{\mathsf{cap}}(\boldsymbol{v}_j)$. Recall we will overload notation and also use $\boldsymbol{L}$ to denote the uniform distribution over $\boldsymbol{L}$.

Many of the unary constraints–in fact, all of them above the roots of the BP factor tree–involve the uniform distribution over $\boldsymbol{L}$; when we perform the variable-to-constraint updates, we will be continually taking restrictions of distributions over $\boldsymbol{L}$. We thus establish that P gives us a "spread contraction" for distributions close to the uniform distribution over $\boldsymbol{L}$.

Lemma 5.5.10. Let ν be a distribution over $\mathbb{S}^{d-1}$ such that $\|\nu\|_\infty \le n^{\log^5 n}$, and let

$$\delta = \mathrm{d_{TV}}(\nu, \boldsymbol{L})$$

Then, for a universal constant C,

$$\mathrm{Spr}_{P\nu}\left(2n^{-\log^4 n}\right) \le \max\left\{2\sqrt{\frac{\log^{11} n}{nd}}, C\cdot\delta\cdot\sqrt{\frac{\log^{23} n}{d}}\right\}$$

Our results from Section 5.2 help us prove these lemmas; the spread of $P\nu$ at any point is precisely controlled by the concentration of the measure of a random sphere cap under ν. In Section 5.5.3, we provide some additional analysis of the P operator.

Analyzing vertex-to-constraint messages.

Recall that our vertex-to-constraint messages are given as a product of the vertex's incoming messages. Each vertex's incoming messages are themselves measures over $\mathbb{S}^{d-1}$ that have bounded spread; we show that as long as a vertex's degree is bounded, the spread of their product is not too large.

Lemma 5.5.11. Let $\nu_1, \dots, \nu_j$ be distributions over $\mathbb{S}^{d-1}$ with intersecting support. For each $i \in [j]$, suppose $\mathrm{Spr}_{\nu_i}(\varepsilon) \le \eta$ and $\|\nu_i\|_\infty \le \frac{1}{p}$. Further, assume $\eta(j+1) < \frac{1}{2}$ and $\frac{j\varepsilon}{p^j} \le \eta$. Let ν be the distribution whose density is $\nu(x) \propto \prod_{i=1}^{j} \nu_i(x)$. Then,

$$\mathrm{Spr}_\nu(j\cdot\varepsilon) \le 8j\eta \qquad \text{and} \qquad \|\nu\|_\infty \le \frac{2}{p^j}.$$

As a consequence of Observation 5.5.17:

$$\mathrm{d_{TV}}(\nu, \rho) \le 10j\eta.$$

Tracking "spread" up the BP tree

Starting from the leaves of the factor graph (Figure [**fig:factor-graph**]), we will track both the *infinity norm* of the messages as well as their spread.

Definition 5.5.12. For a variable vertex $v \in F$, we use $\mathrm{Ch}(v)$ to denote the set of constraint vertices corresponding to edges $\{v, w\}$ for w that are children of v in T.

Definition 5.5.13. Given a vertex $i \in V(T)$ we say its *tier* is the distance to its closest neighbor in R. In particular,

$$\mathrm{Tier}(v) = \min\{\beta : v \text{ has a distance } \beta \text{ to } R\}.$$

We use the convention that v is a descendant of itself and hence $\mathrm{Tier}(v) = 0$ when $v \in R$.

We make the following simplification, which holds with high probability when $p = \Theta\left(\frac{1}{n}\right)$.

Assumption 5.5.14. We assume that the degree of every vertex in T is bounded by $\log^2 n$, and that for every $i \in V(T)$, the $\rho(L_i) \geq n^{-\log^3 n}$.

In Lemma 7.22 of [LMSY22b], we prove:

Lemma 5.5.15. For a variable vertex i in F, let ν_i be the distribution

$$\nu_i \propto \prod_{a \in \mathrm{Ch}(i)} m^{a \to i}.$$

Then, for a universal constant C and $\varepsilon := C\sqrt{\frac{\log^{27} n}{d}}$:

$$\mathrm{Spr}_{\nu_i}\left(n^{-\log^4 n/2}\right) \leq \min\left\{\varepsilon, \max\left\{\varepsilon^{\mathrm{Tier}(i)}, C\sqrt{\frac{\log^{15} n}{nd}}\right\}\right\} \qquad \text{and} \qquad \|\nu_i\|_\infty \leq \frac{2}{p^{\log^2 n}}.$$

The thrust of the proof is sketched below:

Proof sketch. We proceed up the BP factor tree level by level. The outgoing messages from any "edge constraint" node experience a $\tilde{O}\left(\frac{1}{\sqrt{d}}\right)$ factor decrease in spread (Lemma 5.5.10), and the outgoing messages also have ∞-norm upper bounded by $\frac{1}{p}$ (Observation 5.5.7).

Then, when we consider a "variable" node ν_i, the bound $\|\nu_i\|_\infty \leq \frac{2}{p^{\log^2 n}}$ is immediate after combining Lemma 5.5.11 with Assumption 5.5.14. We also obtain the correct contraction in spread, using $j \leq \log^2 n$; Lemma 5.5.11 also tells us that when we combine edge constraints to a variable node, the spread only takes a $\mathsf{polylog}(n)$ hit. □

5.5.3 A potential function for sphere cap convolutions

Recall the definition of *deviation* and *spread*, reproduced below.

Definition 5.5.16. Given a function ν on $\mathbb{S}^{d-1}$, its *deviation profile* is

$$\mathrm{Dev}_\nu(\varepsilon) := \Pr_{\boldsymbol{z} \sim \rho}[\nu(\boldsymbol{z}) \notin \mathbb{E}_{\boldsymbol{y} \sim \rho}[\nu(\boldsymbol{y})] \pm \varepsilon \cdot \|\nu\|_1]$$

The *spread profile* of ν, denoted $\mathrm{Spr}_\nu(\delta)$, is

$$\mathrm{Spr}_\nu(\delta) := \inf\left\{\varepsilon \in \mathbb{R}_{\geq 0} : \mathrm{Dev}_\nu(\varepsilon) \leq \delta\right\}.$$

In the special case where ν is a relative density of a distribution with respect to ρ,

$$\mathrm{Dev}_\nu(\varepsilon) = \Pr_{\boldsymbol{z} \sim \rho}[|\nu(\boldsymbol{z}) - 1| > \varepsilon].$$

We comment that we can think of the spread profile as an "inverse" to the deviation profile, since it takes in a tail probability and returns the corresponding deviation ε.

It now follows from averaging arguments that the deviation profile and spread profile of a distribution ν give us useful upper and lower bounds on $\mathrm{d_{TV}}(\nu, \rho)$.

Observation 5.5.17. For any $\varepsilon > 0$ and distribution ν,

$$\mathrm{Dev}_\nu(\varepsilon) \cdot \varepsilon \le \mathrm{d_{TV}}(\nu, \rho) \le \varepsilon + \mathrm{Dev}_\nu(\varepsilon) \cdot \|\nu\|_\infty$$

Similarly, for any $\delta > 0$ and distribution ν,

$$\delta \cdot \mathrm{Spr}_\nu(\delta) \le \mathrm{d_{TV}}(\nu, \rho) \le \mathrm{Spr}_\nu(\delta) + \delta \cdot \|\nu\|_\infty$$

We list some properties of the deviation and spread profiles that will be useful in our analysis.

Observation 5.5.18. Both the deviation profile and spread profile are non-increasing.

Observation 5.5.19. We have $\mathrm{Dev}_\rho(\varepsilon) = 0$ for all $\varepsilon > 0$ and $\mathrm{Spr}_\rho(\delta) = 0$ for all $\delta > 0$.

Observation 5.5.20. Both the deviation and spread profiles are invariant under a constant factor multiplication to ν, i.e. for $\alpha \neq 0$:

$$\mathrm{Dev}_{\alpha\nu}(\varepsilon) = \mathrm{Dev}_\nu(\varepsilon) \text{ and } \mathrm{Spr}_{\alpha\nu}(\delta) = \mathrm{Spr}_\nu(\delta)$$

Observation 5.5.21. For $\nu = \nu_1 + \nu_2$, and $\varepsilon > 0$, the deviation profile satisfies:

$$\mathrm{Dev}_\nu\left(\varepsilon \cdot \frac{\|\nu_1\|_1 + \|\nu_2\|_1}{\|\nu\|_1}\right) \le \mathrm{Dev}_{\nu_1}(\varepsilon) + \mathrm{Dev}_{\nu_2}(\varepsilon)$$

The "triangle inequality" follows from the following containment of events:

$$\begin{aligned}\{\nu(\boldsymbol{z}) \notin \mathbb{E}_{\boldsymbol{y}\sim\rho}[\nu(y)] \pm \varepsilon \cdot (\|\nu_1\|_1 + \|\nu_2\|_1)\} \\ \subseteq \{\nu_1(\boldsymbol{z}) \notin \mathbb{E}_{\boldsymbol{y}\sim\rho}[\nu_1(y)] \pm \varepsilon \cdot \|\nu_1\|_1\} \cup \{\nu_2(\boldsymbol{z}) \notin \mathbb{E}_{\boldsymbol{y}\sim\rho}[\nu_2(y)] \pm \varepsilon \cdot \|\nu_2\|_1\}.\end{aligned}$$

Similarly, we also have a "triangle inequality" for the spread profile. For $\nu = \nu_1 + \nu_2$ and $\delta_1, \delta_2 > 0$:

$$\|\nu\|_1 \cdot \mathrm{Spr}_\nu(\delta_1 + \delta_2) \le \|\nu_1\|_1 \cdot \mathrm{Spr}_{\nu_1}(\delta_1) + \|\nu_2\|_1 \cdot \mathrm{Spr}_{\nu_2}(\delta_2)$$

We can think of the left hand side as the $(1 - \delta_1 - \delta_2)$-confidence interval of ν, and the right hand terms as the $(1 - \delta_1)$- and $(1 - \delta_2)$-confidence intervals of ν_1 and ν_2, respectively.

Proof of Lemma 5.5.9

The key property we use to establish Lemma 5.5.9 is that the value of $\nu(\mathsf{cap}(\boldsymbol{z}))$ concentrates when $\boldsymbol{z}$ is chosen uniformly at random from $\mathbb{S}^{d-1}$ and $\|\nu\|_\infty$ is reasonably bounded. This was already shown in Section 5.2; it is a nice coincidence that our prior results found additional application here.

Proof of Lemma 5.5.9. First, let us write $\nu = \nu^+ - \nu^-$ where

$$\nu^+ = \max(\nu, 0) \text{ and } \nu^- = -\min(\nu, 0)$$

Here, both ν^+ and ν^- are nonnegative functions, and $\|\nu\| = 2\|\nu^+\| = 2\|\nu^-\| = \|\nu^+\| + \|\nu^-\|$ for $\|\cdot\|$ any ℓ_p norm. Since $P\nu = P\nu^+ - P\nu^-$, by Observation 5.5.21 and Observation 5.5.20:

$$\mathrm{Dev}_{P\nu}\left(\varepsilon \cdot \frac{\|P\nu^+\|_1 + \|P\nu^-\|_1}{\|P\nu\|_1}\right) \leq \mathrm{Dev}_{P\nu^+}(\varepsilon) + \mathrm{Dev}_{P\nu^-}(\varepsilon) = \mathrm{Dev}_{P\frac{\nu^+}{\|\nu^+\|}}(\varepsilon) + \mathrm{Dev}_{P\frac{\nu^-}{\|\nu^-\|}}(\varepsilon).$$

Note that $\frac{1}{\|\nu^+\|_1}\nu^+$ and $\frac{1}{\|\nu^-\|_1}\nu^-$ are both probability measures. By Definition 5.5.5, for any $x \in \mathbb{S}^{d-1}$:

$$\frac{P\nu^+(x)}{\|\nu^+\|_1} = \frac{1}{p} \cdot \frac{\nu^+(\mathsf{cap}(x))}{\|\nu^+\|_1} \qquad \text{and} \qquad \frac{P\nu^-(x)}{\|\nu^-\|_1} = \frac{1}{p} \cdot \frac{\nu^-(\mathsf{cap}(x))}{\|\nu^-\|_1}$$

By Corollary 5.2.9 with s set as $\frac{1}{d} \cdot C \log^5 n \cdot \left(2 \ln \frac{\|\nu^+\|_\infty}{\|\nu^+\|_1} + 8 \ln \frac{d}{p}\right)$ for some constant $C > 0$:

$$\begin{aligned}\Pr_{\boldsymbol{x}\sim\rho}\left[\left|\frac{1}{p} \cdot \frac{\nu^+(\mathsf{cap}(\boldsymbol{x}))}{\|\nu^+\|_1} - 1\right| > C \cdot C_{5.2.9}\sqrt{\frac{1}{d} \cdot \log\left(\frac{1}{p}\right)} \cdot \log^5(n) \cdot \left(2 \ln \frac{\|\nu^+\|_\infty}{\|\nu^+\|_1} + 8 \ln \frac{d}{p}\right)\right] \\ \leq \frac{1}{2} \cdot n^{-\log^4 n}.\end{aligned}$$

Using $p \geq \frac{1}{n}$, $\|\nu^+\|_\infty \leq \|\nu\|_\infty$ and $\|\nu^+\|_1 = \frac{1}{2}\|\nu\|$ we can conclude:

$$\mathrm{Dev}_{\frac{P\nu^+}{\|\nu^+\|}}(\varepsilon) \leq \frac{1}{2} n^{-\log^4 n}.$$

Identically, $\mathrm{Dev}_{\frac{P\nu^-}{\|\nu^-\|}}(\varepsilon) \leq \frac{1}{2} n^{-\log^4 n}$, and thus:

$$\mathrm{Dev}_{P\nu}\left(\varepsilon \cdot \frac{\|P\nu^+\|_1 + \|P\nu^-\|_1}{\|P\nu\|_1}\right) \leq n^{-\log^4 n}.$$

Since ν^+ and ν^- are nonnegative functions, $\|P\nu^+\| = \|\nu^+\|$ and $\|P\nu^-\| = \|\nu^-\|$, and the desired statement then immediately follows. □

Proof of Lemma 5.5.10

Lemma 5.5.22. Define

$$\begin{aligned}g\left(\boldsymbol{G}_{n-1}\right) &:= \Pr_{\boldsymbol{w}_1,\ldots,\boldsymbol{w}_{n-1}\sim\rho^{\boldsymbol{G}_{n-1}}}\left[\mathrm{Dev}_{P\boldsymbol{L}}(\varepsilon) > n^{-\log^4 n}\right] \\ &= \Pr_{\boldsymbol{w}_1,\ldots,\boldsymbol{w}_{n-1}\sim\rho^{\boldsymbol{G}_{n-1}}}\left[\mathrm{Spr}_{P\boldsymbol{L}}\left(n^{-\log^4 n}\right) > \varepsilon\right]\end{aligned}$$

for $\varepsilon = \sqrt{\frac{\log^{11} n}{nd}}$. Then, $g\left(\boldsymbol{G}_{n-1}\right)$ is at most $O\left(n^{-\log^2 n}\right)$ except with probability $O\left(n^{-\log^2 n}\right)$.

Proof of Lemma 5.5.10. We can express $\nu = \boldsymbol{L} + \Delta$. Then, the triangle inequality for $\mathrm{Spr}_{P\nu}$ as articulated in Observation 5.5.21 implies:

$$\mathrm{Spr}_{P\nu}(2n^{-\log^4 n}) \leq \mathrm{Spr}_{P\boldsymbol{L}}(n^{-\log^4 n}) + \|P\Delta\|_1 \cdot \mathrm{Spr}_{P\Delta}(n^{-\log^4 n}) \tag{5.12}$$

By Lemma 5.5.22, $\mathrm{Spr}_{P\boldsymbol{L}}(n^{-\log^4 n}) \leq \sqrt{\frac{\log^{11} n}{nd}}$.

We now turn our attention to $P\Delta$. We would like to understand $P\Delta$ through Lemma 5.5.9, but the lower bound on d of $\log^{10} n \cdot \left(1 + \log^3 \frac{\|\nu\|_\infty}{\|\nu\|_1}\right)$ in the hypothesis of its statement prevents us from applying it when δ is too small, and hence we case on the value of δ. Suppose $\delta \leq n^{-2\log^4 n}$, then by Observation 5.5.6, $\|P\Delta\|_1 \leq n^{-2\log^4 n}$ and thus by Markov's inequality $\mathrm{Spr}_{P\Delta}(n^{-\log^4 n}) \leq n^{\log^4 n}$. Plugging this into Equation 5.12 gives:

$$\mathrm{Spr}_{P\nu}(2n^{-\log^4 n}) \leq \sqrt{\frac{\log^{11} n}{nd}} + n^{-\log^4 n} \leq 2\sqrt{\frac{\log^{11} n}{nd}}.$$

When $\delta > 2n^{-\log^4 n}$, by Lemma 5.5.9, we have for some constant C:

$$\begin{aligned}
\mathrm{Spr}_{P\Delta}(n^{-\log^4 n}) &\leq \sqrt{\frac{1}{d}} \cdot C\log^{5.5} n \cdot \left(2\ln \frac{n^{\log^5 n}}{\|\Delta\|_1} + 8\ln\frac{d}{p}\right) \cdot \frac{\|\Delta\|_1}{\|P\Delta\|_1} \\
&\leq \sqrt{\frac{1}{d}} \cdot C\log^{5.5} n \cdot \left(4\log^6 n + 8\ln\frac{d}{p}\right) \cdot \frac{\|\Delta\|_1}{\|P\Delta\|_1}
\end{aligned}$$

where the second inequality uses $\|\Delta\|_1 = \delta > n^{-2\log^4 n}$. Consequently, plugging into Equation 5.12,

$$\mathrm{Spr}_{P\nu}\left(n^{-\log^4 n}\right) \leq \sqrt{\frac{\log^{11} n}{nd}} + C\sqrt{\frac{\log^{23} n}{d}} \cdot \delta$$

which completes the proof. □

Chapter 6

Random Geometric Graphs as HDXes

I would be complex
I would be cool [Swi19]

This chapter is based on joint work with Siqi Liu, Sidhanth Mohanty, and Tselil Schramm, published in [LMSY22a]. We do not include full proofs; we only state some relevant results.

6.1 Problem statement and summary of results

While we have algebraic and combinatorial constructions of HDXes (for dimension 2), we are quite far from a result like Friedman's Theorem [Fri03] for expander graphs; we do not know of a natural distribution over simplicial complexes from which to sample high-dimensional expanders. Whether high-dimensional expanders are highly contrived objects or relatively commonplace is still not known. See the upcoming Section 6.1.2 for a more thorough survey of existing HDX constructions.

In this chapter, we study a natural distribution over 2-dimensional simplicial complexes, inspired by the $\mathsf{Geo}_d(n,p)$ distribution, and prove that (with high probability), complexes sampled from this distribution have local spectral expansion bounded away from $\frac{1}{2}$. Ideally, these complexes would have constant degree, but we settle for average degree $\Theta(n^\varepsilon)$ for $\varepsilon < 1$.

To prove that these 2-dimensional simplicial complexes are local spectral expanders, it suffices to just understand the expansions of the vertex neighborhoods (i.e. the *links* of the vertices), because of Oppenheim's trickle-down theorem:

Theorem 6.1.1 (Restatement of Theorem 2.2.1)**.** Let X be a 2-dimensional simplicial complex. If its 1-skeleton is connected, and the second eigenvalue of every link's random walk matrix is at most λ, then the second absolute eigenvalue of the random walk matrix of the 1-skeleton of X is at most $\frac{\lambda}{1-\lambda}$.

This theorem explains the significance of $\lambda = \frac{1}{2}$, since when $\lambda < \frac{1}{2}$, local expansion "trickles down" to imply global expansion. We will show that random geometric graphs, in a carefully-chosen parameter regime, have sufficient link expansion.

6.1.1 Our results

We first describe how we use $\mathsf{Geo}_d(n,p)$ to sample a 2-dimensional simplicial complex.

Definition 6.1.1 (Random geometric complex)**.** The *random geometric k-complex* $\mathsf{Geo}_d^{(k)}(n,p)$ is the distribution defined by sampling $\boldsymbol{G} \sim \mathsf{Geo}_d(n,p)$ and taking the downward-closure of the complex whose k-faces are the cliques of size $(k+1)$ in $\boldsymbol{G}$.

Our main result proves that under mild conditions, random geometric 2-complexes of degree n^ε are high-dimensional expanders enjoying the trickling-down phenomenon:

Theorem 6.1.2. For every $0 < \varepsilon < 1$, there exist constants C_ε and $\delta = \exp(-O(1/\varepsilon))$ such that when $\boldsymbol{H} \sim \mathsf{Geo}_d^{(2)}(n, n^{-1+\varepsilon})$ for $d = C_\varepsilon \log n$, with high probability every vertex link of $\boldsymbol{H}$ is a $(\frac{1}{2} - \delta)$-expander, and hence its 1-skeleton is a $\left(1 - \frac{4\delta}{1+2\delta}\right)$-expander.

Remark 6.1.2. The complexes arising from Theorem 6.1.2 have degree bounded by $O(n^{2\varepsilon})$ with high probability, as the number of triangle a vertex participates in is the square of its degree in the 1-skeleton.

Along the way to proving Theorem 6.1.2, we also analyze the spectrum of $\boldsymbol{G} \sim \mathsf{Geo}_d(n,p)$ directly and obtain sharper control of its second eigenvalue in a more general setting, bounding the spectral gap of random geometric graphs in the full high-dimensional ($d \to_n \infty$) regime. To our knowledge, previous results in this vein are only for $d \sim n^{1/k}$ for fixed integers k [El 10; CS13; DV13; Bor13; FM19; LY22].

Theorem 6.1.3. Let $\boldsymbol{G} \sim \mathsf{Geo}_d(n,p)$ and $\tau := \tau(p,d)$. Then $\boldsymbol{G}$ is a μ-expander w.h.p., with

$$\mu := (1 + o(1)) \cdot \max\left\{ (1 + o_{d\tau^2}(1)) \cdot \tau, \frac{\log^4 n}{\sqrt{pn}} \right\}$$

and $o_{d\tau^2}(1)$ denotes a function that goes to 0 as $d \cdot \tau(p,d)^2 \to \infty$.

Our family of random geometric 2-complexes also exhibits an example where Oppenheim's trickle-down theorem is tight.

Proposition 6.1.3 (Trickling-down theorem is tight)**.** For each $\lambda \in (0, \frac{1}{2}]$ and $\eta > 0$ there exists a 2-dimensional expander in which all vertex link eigenvalues are at most λ for which the 1-skeleton is connected with eigenvalue at least $\frac{\lambda}{1-\lambda} - \eta$.

Spectra of random restrictions

Theorem 6.1.3 (and morally Theorem 6.1.2) is a consequence of a more general theorem that we prove concerning the spectral properties of *random restrictions* of graphs. We describe this result here, both because it may be of independent interest, and because it may help demystify Theorem 6.1.2.

Random restriction is a procedure for approximating a large graph X by a smaller graph $\boldsymbol{G}$: one selects a random subset of vertices $\boldsymbol{S}$, and then takes $\boldsymbol{G}$ to be the induced graph $X[\boldsymbol{S}]$. The random restriction $\boldsymbol{G}$ is now a smaller (and often sparser) approximation to X. This idea has been useful in a number of contexts in theoretical computer science (e.g. [GGR98; AVKK03; BHHS11; LRS15; Hop+17]). The core question here is: to what extent do random restrictions actually inherit properties of the original graph? We will show that if random walks on X mix rapidly enough, then random restrictions inherit the *spectral* properties of the original graph.

To see the relevance of this result in our context, notice that a random geometric graph on the sphere is a random restriction of the (infinite) graph with vertex set $\mathbb{S}^{d-1}$ and edge set $\{(u, v) \mid \langle u, v\rangle \geq \tau\}$. Theorem 6.1.2 is then a consequence of the fact that the sphere is *itself* a 2-dimensional expander. We state the theorem precisely below.

Definition 6.1.4 (Random restriction)**.** Suppose X is a (possibly infinite) graph, and that the simple random walk on X has unique stationary distribution ρ. We define an *n-vertex random restriction* of X to be a graph $\boldsymbol{G} \sim \mathsf{RR}_n(X)$ sampled by sampling n vertices independently according to ρ, $\boldsymbol{S} \sim \rho^{\otimes n}$, then taking $\boldsymbol{G} = X[\boldsymbol{S}]$ to be the graph induced on those vertices.

We show that if the average degree in $\boldsymbol{G}$ is not too small, $\lambda_2(\boldsymbol{G})$ reflects the rapid mixing of the random walk on X.

Theorem 6.1.4. Let X be a (possibly infinite) vertex-transitive graph on which the associated simple random walk has a unique stationary distribution ρ, and let

$$p = \Pr_{\boldsymbol{G} \sim \mathsf{RR}_n(X)}[(i, j) \in E(\boldsymbol{G})]$$

be the marginal edge probability of a n-vertex random restriction of X. Suppose there exist $C \geq 1$ and $\lambda \in [(np)^{-1/2}, 1]$ such that for any $k \in \mathbb{N}$, k-step walks on X satisfy the following mixing property: for any distribution α over $V(X)$,

$$\mathrm{d}_{\mathrm{TV}}\left(X^k \alpha, \rho\right) \leq C \cdot \lambda^k,$$

where X^k denotes the k-step random walk operator on X. Furthermore, suppose $pn \gg C^6 \log^4 n$. Then for any constant $\gamma > 0$,

$$\Pr_{\boldsymbol{G} \sim \mathsf{RR}_n(X)}\left[\left|\lambda_2(\widehat{A}_{\boldsymbol{G}})\right|, \left|\lambda_n(\widehat{A}_{\boldsymbol{G}})\right| \leq (1 + o(1)) \cdot \max\left(\lambda, \frac{\log^4 n}{\sqrt{pn}}\right)\right] \geq 1 - n^{-\gamma},$$

where $\widehat{A}_{\boldsymbol{G}}$ is the (normalized) adjacency matrix of $\boldsymbol{G}$.

Remark 6.1.5. It is likely that some of the conditions of Theorem 6.1.4 could be weakened. The decay of total variation could plausibly be replaced with a (much weaker) assumption about the spectral gap of X; this would not impact our results for $\mathbb{S}^{d-1}$, but may be useful in other applications. Transitivity is assumed mostly to make the proof of Theorem 6.1.4 go through at this level of generality; to prove Theorem 6.1.2 we re-prove a version of Theorem 6.1.4 for the specific non-transitive case where X is a link of a vector in the sphere (i.e. a random geometric graph restricted to a spherical cap).

6.1.2 Related work

While so far we have focused on a spectral notion of high-dimensional expanders (HDX), there are two additional notions: coboundary and cosystolic expansion. These are meant to generalize the Cheeger constant, a cut-based measure of graph expansion.

Distributions over high-dimensional expanders

The existence of natural distributions over sparse HDXs has been a question of interest since sparse HDX were first shown to exist (and this was highlighted as an important open problem in e.g. [Lub18; Lub22]).

The early work of Linial and Meshulam [LM06] considered the distribution over 2-dimensional complexes in which all edges $\binom{[n]}{2}$ are included, and each triangle is included independently with probability p; they identified the phase transition at p for coboundary connectivity for this distribution (see also the follow-ups [BHK11; MW09; LP16]). This distribution has the drawback that the 1-skeleton of these complexes is K_n, and so the resulting complex is far from sparse.

In [FGLNP12], the authors show that a union of d random partitions of $[n]$ into sets of size $k+1$ with high probability produces a *geometric* expander [Gro10b], which is a notion of expansion which measures how much the faces must intersect when the complex is embedded into $\mathbb{R}^k$. The resulting complexes have disconnected links when $d \ll \sqrt{n}$, and so they fail to be spectral HDXs.

The work of [LMY20] introduces a distribution over spectral expanders with expansion exactly $\frac{1}{2}$ by taking a tensor product of a random graph and a HDX; the authors show that down-up walks on these expanders mix rapidly, and [Gol21] introduces a reweighing of these complexes which yields improved mixing time bounds. However, the links in these complexes fail to satisfy $\lambda < \frac{1}{2}$, and so fall outside of the range of the trickling-down theorem. The same drawback applies to [Con19; CTZ20]: they show that up-down walks mix on random polylogarithmic-degree graphs given by subsampling a random set of generators of a Cayley graph. However, these graphs do not satisfy the conditions of the trickling-down theorem.

Explicit constructions of HDXes

One of the first constructions of sparse high-dimensional spectral expanders was the Ramanujan complex of [CSŻ03; Li04; LSV05a; LSV05b], which generalize the Ramanujan expander graphs of [LPS88b]. Not only are these spectral expanders, but [KKL14; EK16b] also show that they are co-systolic expanders. These Ramanujan complexes are algebraic by nature, constructed from the Cayley graphs of $\mathrm{PSL}_d(\mathbb{F}_q)$. Other algebraic constructions include that of [KO18]; the authors analyze the expansion properties of coset complexes for various matrix groups. They achieve sparse spectral expanders, with local expansion arbitrarily close to 0. More recently, [OP22] extend the coset complex construction to the more general family of Chevalley groups.

A few combinatorial constructions for HDX are also known. [CLP20] prove that objects called (a, b)-expanders are two-dimensional spectral expanders; they give a graph-product-inspired construction of a family of such expanders, and show that other known complexes [CSŻ03; Li04; LSV05a; LSV05b; KO18] are also (a, b)-expanders. Their work is extended by [FI20] to higher dimensions.

Random geometric graphs and random kernel matrices

Random restrictions of metric spaces such as $\mathbb{S}^{d-1}$ and $[-1, 1]^d$ are well-studied in the fixed-dimensional regime, where $d = O(1)$ and $n \to \infty$ (see the survey of Penrose [Pen03]). In our work, we are interested in the high-dimensional setting, where $d \to \infty$ with n. The high-dimensional setting was first studied only recently, initiated by [DGLU11; BDER16], and many mysteries remain in this young area of study.

Our Theorem 6.1.3 is related to the study of *kernel random matrices*: random $n \times n$ matrices whose (i, j)-th entry is given by $f_d(\langle \boldsymbol{u}_i, \boldsymbol{u}_j \rangle)$, for $f_d : \mathbb{R} \to \mathbb{R}$ and $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$ sampled independently from some distribution over $\mathbb{R}^d$. The special case of $\boldsymbol{u}_i \sim \mathrm{Unif}(\mathbb{S}^{d-1})$ and $f_d(x) = \mathbb{1}[x \geq \tau(p, d)]$ yields the adjacency matrix of $\mathsf{Geo}_d(n, p)$. A line of work initiated by [KG00] studies the spectrum of kernel random matrices [El 10; CS13; DV13; Bor13; FM19], and the most recent work [LY22] characterizes the limiting empirical spectral distribution when $d = \Theta(n^{1/k})$ for k a fixed constant and f can be "reasonably" approximated by polynomials (in a sense that is flexible enough to capture the indicator $f_d(x) = \mathbb{1}[x \geq \tau(p, d)]$). In comparison with our results, they characterize the entire empirical spectral distribution, but we do not need to restrict $d \sim n^{1/k}$ for integer k, which is crucial for our applications.

6.1.3 Technical Overview

We now explain how we prove Theorem 6.1.2, which states that for a complex sampled from $\boldsymbol{H} \sim \mathsf{Geo}_d^{(2)}(n, p)$ for $p = n^{-1+\varepsilon}$ with $0 < \varepsilon < 1$ and $d = C_\varepsilon \log n$, with high probability every link of $\boldsymbol{H}$ is a $\left(\frac{1}{2} - \delta\right)$-expander for some $\delta = \exp(-O(1/\varepsilon))$, and its 1-skeleton is a $\left(1 - \frac{4\delta}{1+2\delta}\right)$-expander. By the trickling-down theorem, it suffices for us to prove:

1. All n vertices' corresponding links in $\boldsymbol{H}$ are $\left(\frac{1}{2}-\delta\right)$-expanders with high probability.
2. The 1-skeleton of $\boldsymbol{H}$ is connected with high probability.

To show Item 2, it is enough to show that some reweighting of the 1-skeleton expands; Item 1 implies that every edge (i,j) must participate in at least one triangle (otherwise the link would contain isolated vertices), so the unweighted 1-skeleton is just the adjacency matrix of an unweighted graph from $\mathsf{Geo}_d(n,p)$. En route to proving Item 1 we'll prove that unweighted random geometric graphs expand, by this logic yielding Item 2 a consequence.

Analyzing link expansion

We establish Item 1 by showing that that each of the n links is a $\left(\frac{1}{2}-\delta\right)$-expander with probability $1-o(1/n)$, then applying a union bound.

We can think of sampling the link of vertex i_w in $\boldsymbol{H}$ by first choosing the number of neighbors $\boldsymbol{r} \sim \mathsf{Binom}(n-1,p)$, then sampling $\boldsymbol{r}$ points $\boldsymbol{v}_1,\dots,\boldsymbol{v}_{\boldsymbol{r}}$ independently and uniformly from a measure-p cap in $\mathbb{S}^{d-1}$ centered at some point w (corresponding to the vector of the link vertex i_w), placing an edge between every i,j such that $\langle \boldsymbol{v}_i, \boldsymbol{v}_j\rangle \geq \tau(p,d)$. Finally, we remove any isolated vertices; here, we'll show that the graph expands with high probability before removing these isolated vertices, which implies that no isolated vertices have to be removed. For the remainder of the overview, let $\tau = \tau(p,d)$. We'll show that:

Theorem 6.1.5 (Informal)**.** Let $\boldsymbol{G}$ be the link of some point $w \sim \mathbb{S}^{d-1}$ induced by $\boldsymbol{v}_1,\dots,\boldsymbol{v}_m \sim \mathsf{cap}_p(w)$. Then with high probability $\boldsymbol{G}$ is a μ-expander where

$$\mu := (1+o(1)) \cdot \max\left\{\frac{\tau}{\tau+1}, \frac{\log^4 m}{\sqrt{qm}}\right\} + o_d(1).$$

Here $q = \Pr_{u,v\sim\mathbb{S}^{d-2}}\left[\langle u,v\rangle \geq \frac{\tau}{\tau+1}\right]$.

Links are essentially random geometric graphs in one lower dimension

Since most of the measure of the cap lies close to its boundary, intuitively the link is distributed *almost* like a random geometric graph with points drawn independently from the cap boundary, i.e. $\mathsf{shell}_p(w) := \{x : \langle x,w\rangle = \tau\}$. Our proof of Theorem 6.1.5 must pay attention to the fluctuations in $\langle \boldsymbol{v}_i, w\rangle - \tau$, but to simplify our current discussion, we assume each link is in fact a random geometric graph on $\mathsf{shell}_p(w)$.

Observe that a uniformly random $\boldsymbol{v}$ from $\mathsf{shell}_p(w)$ is distributed as

$$\tau \cdot w + \sqrt{1-\tau^2}\cdot \boldsymbol{u}$$

where $\boldsymbol{u}$ is a uniformly random unit vector orthogonal to w. Using this decomposition,

$$\langle \boldsymbol{v}_i, \boldsymbol{v}_j\rangle \geq \tau \text{ if and only if } \langle \boldsymbol{u}_i, \boldsymbol{u}_j\rangle \geq \frac{\tau}{1+\tau}$$

Thus, under our simplifying assumption, the link is distributed exactly like a random geometric graph on $\mathbb{S}^{d-2}$ with inner product threshold $\frac{\tau}{1+\tau}$. Hence (up to the difference between $\mathsf{cap}_p(w)$ and $\mathsf{shell}_p(w)$), to understand link expansion we can study the second eigenvalue of a random geometric graph on the sphere.

Remark 6.1.6 (Requiring $d = \Theta(\log n)$)**.** In light of Theorem 6.1.5 (and even the heuristic discussion above), it turns out that $d = \Theta(\log n)$ is the only regime for which the links can be connected while the 1-skeleton has average degree $\ll \sqrt{n}$. To see this, we consider the relationship between p, τ, and d:

$$p = \Pr_{\boldsymbol{v},\boldsymbol{v}' \sim \mathbb{S}^{d-1}} [\langle \boldsymbol{v}, \boldsymbol{v}' \rangle \geq \tau] = \Theta\left(\tfrac{1}{\tau d}\right) \cdot \left(1 - \tau^2\right)^{\frac{d-1}{2}} \approx \exp(-d\tau^2/2). \tag{6.1}$$

See Lemma 4.1.5 for a formal argument. The arguments above, in conjunction with Equation 6.1 imply that the probability that two vertices within a link are connected is also roughly

$$q = \Pr_{\boldsymbol{u},\boldsymbol{u}' \sim \mathbb{S}^{d-2}} \left[\langle \boldsymbol{u}, \boldsymbol{u}' \rangle \geq \tfrac{\tau}{1+\tau}\right] = \Theta\left(\tfrac{1}{\tau d}\right) \cdot \left(1 - \tfrac{\tau^2}{(1+\tau)^2}\right)^{\frac{d-2}{2}},$$

since the link is like a random geometric graph on $\mathsf{shell}_p(w)$.

Connectivity within the links in conjunction with sparsity requires us to have $d \in \Theta(\log n)$: the number of vertices inside each link concentrates around $m = np$, so the average degree inside the link is $qm \approx qpn$. We must have the average link degree $qpn \geq 1$, otherwise the link is likely disconnected.

Now, if $\tau = o(1)$, then $\tau \approx \frac{\tau}{1+\tau}$ and $p \approx q$, so

$$qpn \geq 1 \Rightarrow p^2 n \gtrapprox 1 \Rightarrow p \gtrapprox n^{-1/2}$$

ruling out a 1-skeleton with average degree $\ll \sqrt{n}$. Hence we need $\tau = \Omega(1)$. Given that $\tau = \Omega(1)$, Equation 6.1 implies that to have the average 1-skeleton degree $\sqrt{n} \geq pn \geq 1$ we need $d \in \Theta(\log n)$.

Spectral expansion in random geometric graphs

We now explain how to prove near-sharp second eigenvalue bounds for random geometric graphs (Theorem 6.1.3).

As mentioned above, Theorem 6.1.3 is a consequence of the more general Theorem 6.1.4 about the second eigenvalue of random restrictions of vertex-transitive graphs, and the inner product threshold $\tau = \tau(p, d)$ appears as the mixing rate of the random walk on $\mathbb{S}^{d-1}$ where a step originating at v walks to a random vector in $\mathsf{cap}_p(v)$. Via standard concentration arguments applied to the vertex degrees, to prove the above it suffices to bound $\|A_{\boldsymbol{G}} - \mathbb{E}A_{\boldsymbol{G}}\| \leq \mu \cdot pn$, where $A_{\boldsymbol{G}}$ is the (unnormalized) adjacency matrix of $\boldsymbol{G}$. We'll focus on the regime where $pn \gg \mathsf{polylog} n$, so that $\mu \approx \tau$.

Trace method for random geometric graphs

To bound $\|A_{\boldsymbol{G}} - \mathbb{E}A_{\boldsymbol{G}}\|$, we employ the trace method, bounding the expected trace of a power of $A_{\boldsymbol{G}} - \mathbb{E}A_{\boldsymbol{G}}$. This is sufficient for the following reason: for convenience, let $\overline{A}_{\boldsymbol{G}} = A_{\boldsymbol{G}} - \mathbb{E}A_{\boldsymbol{G}}$, and let ℓ be any non-negative, even integer. Since ℓ is even,

$$\left\|\overline{A}_{\boldsymbol{G}}\right\|^{\ell} = \left\|\overline{A}_{\boldsymbol{G}}^{\ell}\right\| \leq \operatorname{tr}\left(\overline{A}_{\boldsymbol{G}}^{\ell}\right)$$

And so applying Markov's inequality,

$$\Pr\left(\left\|\overline{A}_{\boldsymbol{G}}\right\| \geq e^{\varepsilon}\left(\mathbb{E}\operatorname{tr}\left(\overline{A}_{\boldsymbol{G}}^{\ell}\right)\right)^{1/\ell}\right) = \Pr\left(\left\|\overline{A}_{\boldsymbol{G}}\right\|^{\ell} \geq e^{\varepsilon\ell}\mathbb{E}\operatorname{tr}\left(\overline{A}_{\boldsymbol{G}}^{\ell}\right)\right) \leq \exp(-\varepsilon\ell).$$

Thus, our goal reduces to bounding the expectation of $\operatorname{tr}\left(\overline{A}_{\boldsymbol{G}}^{\ell}\right)$ for a sufficiently large even ℓ; in particular, if we choose $\ell \gg \log n$, then since $\overline{A}_{\boldsymbol{G}}$ has n eigenvalues, $\operatorname{tr}\left(\overline{A}_{\boldsymbol{G}}^{\ell}\right)^{1/\ell}$ is a good "soft-max" proxy for $\left\|\overline{A}_{\boldsymbol{G}}\right\|$, and we will obtain high-probability bounds.

We now explain why properties of random walks on $\mathbb{S}^{d-1}$ naturally arise when applying the trace method. Concretely, $\operatorname{tr}\left(\overline{A}_{\boldsymbol{G}}^{\ell}\right)$ is a sum over products of entries of $\overline{A}_{\boldsymbol{G}}$ corresponding to closed walks of length ℓ in the complete graph K_n on n vertices:

$$\operatorname{tr}\left(\overline{A}_{\boldsymbol{G}}^{\ell}\right) = \sum_{i_0,\ldots,i_{\ell-1}\in[n]} \prod_{t=0}^{\ell-1} (\overline{A}_{\boldsymbol{G}})_{i_t i_{t+1 \bmod \ell}},$$

The walk $i_0, i_2, \ldots, i_{\ell-1}, i_0$ can be represented as a directed graph. When we take the expectation, the symmetry of the distribution means that all sequences $i_0, \ldots, i_{\ell-1}$ which result in the same graph (up to relabeling) give the same value. That is, letting $\mathcal{W}_\ell$ be the set of all such graphs, and for each $W \in \mathcal{W}_\ell$ letting N_W be the number of ways it can arise in the sum above,

$$\mathbb{E}\operatorname{tr}\left(\overline{A}_{\boldsymbol{G}}^{\ell}\right) = \sum_{W\in\mathcal{W}_\ell} N_W \cdot \mathbb{E}\prod_{(i,j)\in W} (\overline{A}_{\boldsymbol{G}})_{ij}. \tag{6.2}$$

To bound this sum, we must bound the expectation contributed by each $W \in \mathcal{W}_\ell$. For the sake of this overview we will consider only the case when $W = C_\ell$, the cycle on ℓ vertices, as it requires less accounting than the other cases; however it is reasonable to restrict our attention to this case for now, as bounding it already demonstrates our main ideas, and because this term roughly dominates the sum with $N_{C_\ell} \gg N_{W'}$ for all other $W' \in \mathcal{W}_\ell$ at $\ell = \mathsf{polylog}n$ and $pn \gg \mathsf{polylog}n$.[1]

We now bound the expectation for the case $W = C_\ell$; readers uninterested in the finer details may skip to the conclusion in Equation 6.4. We expand the product using that

[1] Briefly, this is because whenever $i_0, \ldots, i_{\ell-1}$ are all distinct elements of $[n]$, the resulting walk's graph is a cycle, and when $\ell = \mathsf{polylog}n$, ℓ indices sampled at random from $[n]$ are all distinct with high probability.

$(\overline{A}_{\boldsymbol{G}})_{ij} = \boldsymbol{A}_{ij} - p$ (since $\mathbb{E}[\boldsymbol{A}_{ij}] = p$):

$$\begin{aligned}\mathbb{E}\prod_{i=1}^{\ell}(\boldsymbol{A}_{i,i+1} - p) &= \sum_{T\subseteq[\ell]}(-p)^{\ell-|T|}\mathbb{E}\prod_{i\in T}\boldsymbol{A}_{i,i+1} \\ &= \sum_{T\subseteq[\ell]}(-p)^{\ell-|T|}\Pr[\{(i,i+1) : i\in T\} \text{ is subgraph of } \boldsymbol{G}]. \qquad (6.3)\end{aligned}$$

and thus our focus is to understand subgraph probabilities in a random geometric graph. It is not too hard to see that when the edges specified by T form a forest, its subgraph probability is $p^{|T|}$, identical to its counterpart in an Erdös-Rényi graph; the nontrivial correlations introduced by the geometry only play a role when T has cycles. Hence, the sum 6.3 simplifies,

$$\begin{aligned}\mathbb{E}\prod_{i=1}^{\ell}(\overline{A}_{\boldsymbol{G}})_{i,i+1} &= \sum_{T\subsetneq[\ell]}(-p)^{\ell-|T|}p^{|T|} + \Pr[C_\ell \text{ is subgraph of } \boldsymbol{G}] \\ &= \Pr[C_\ell \text{ is subgraph of } \boldsymbol{G}] - p^\ell, \qquad (6.4)\end{aligned}$$

where we used that the binomial sum is equal to $(p-p)^\ell = 0$.

Hence it remains to estimate the subgraph probability of a length-ℓ cycle. We will now see how subgraph probabilities are related to the mixing rate of a random walk on $\mathbb{S}^{d-1}$.

Subgraph probability of a cycle in a random geometric graph

For the cycle $C_\ell = 0, 1, \ldots, \ell-1, 0$, by Bayes' rule:

$$\begin{aligned}\Pr[C_\ell \in \boldsymbol{G}] &= \prod_{i=0}^{\ell-1}\Pr[(i,i+1)\in\boldsymbol{G} \mid \forall j < i, (j,j+1)\in\boldsymbol{G}] \\ &= p^{\ell-1}\cdot\Pr[(\ell-1,0)\in\boldsymbol{G} \mid 0,1,\ldots\ell-1\in\boldsymbol{G}],\end{aligned}$$

since in all but the step $i+1=\ell$, the graph in question is a forest. Identifying each i with a point $\boldsymbol{x}_i$ on $\mathbb{S}^{d-1}$, for any choice of $\boldsymbol{x}_0$ the above probability can equivalently be written as

$$p^{\ell-1}\cdot\Pr\left[\langle\boldsymbol{x}_{\ell-1},\boldsymbol{x}_0\rangle \ge \tau \mid \langle\boldsymbol{x}_i,\boldsymbol{x}_{i+1}\rangle\ge\tau : 0\le i\le\ell-2\right]$$

Denoting with P the transition kernel of the random walk we alluded to earlier, where in one step we walk from a point x to a uniformly random point in $\mathsf{cap}_p(x)$, we can write the distribution of $\boldsymbol{x}_\ell \mid \{\boldsymbol{x}_0, \langle\boldsymbol{x}_i,\boldsymbol{x}_{i+1}\rangle\ge\tau : 0\le i\le\ell-2\}$ as $P^{\ell-1}\delta_{\boldsymbol{x}_0}$ where $\delta_{\boldsymbol{x}_0}$ refers to the point mass probability distribution supported at $\boldsymbol{x}_0$. In turn, we can write the subgraph probability as:

$$p^{\ell-1}\cdot\Pr_{\boldsymbol{x}_{\ell-1}\sim P^{\ell-1}\delta_{\boldsymbol{x}_0}}\left[\boldsymbol{x}_{\ell-1}\in\mathsf{cap}_p(\boldsymbol{x}_0)\right]$$

If $\boldsymbol{x}_{\ell-1}$ were sampled from the uniform distribution ρ on $\mathbb{S}^{d-1}$ then the probability of landing in $\mathsf{cap}_p(\boldsymbol{x}_0)$ would be p, which lets us upper bound the subgraph probability by:

$$p^{\ell-1} \cdot \left(p + \mathrm{d_{TV}}\left(P^{\ell-1}\delta_{\boldsymbol{x}_0}, \rho\right)\right).$$

The terms for more complicated subgraphs $W' \in \mathcal{W}_\ell$ also similarly depend on the mixing properties of P via subgraph probabilities. Our next goal then is to understand the mixing properties of P.

Remark 6.1.7. To prove 6.1.4 about random restrictions, the same strategy is used to relate subgraph probabilities with mixing rate of the random walk on the original graph we start with.

Mixing properties of P

We show that the walk over $\mathbb{S}^{d-1}$ with transition kernel P contracts the TV distance by coupling this discrete walk with the continuous Brownian motion U_t over $\mathbb{S}^{d-1}$. Then via a known log-Sobolev inequality for Brownian motion on spheres, we can prove the following contraction property for P.

Theorem 6.1.6 (Informal)**.** For any probability measure α over $\mathbb{S}^{d-1}$ and integer $k \geq 0$,

$$\mathrm{d_{TV}}\left(P_p^k\alpha, \rho\right) \leq \left((1 + o_{d\tau^2}(1)) \cdot \tau\right)^{k-1} \cdot \sqrt{\frac{1}{2}\log\frac{1}{p}},$$

where P_p denotes the transition kernel in which every $x \in \mathbb{S}^{d-1}$ walks to a uniformly random point in the measure-p cap around it and $o_{d\tau^2}(1)$ denotes a function that goes to 0 as $d\tau^2 \to \infty$.

In brief, the reason we are able to execute this coupling is that the probability mass in $P\delta_{\boldsymbol{x}_0}$ concentrates around $\mathsf{shell}_{=\tau}(\boldsymbol{x}_0)$, and most of the $(\frac{1}{d-1}\log\frac{1}{\tau})$-step Brownian motion starting from $\boldsymbol{x}_0$ concentrates at $\mathsf{shell}_{=\tau}(\boldsymbol{x}_0)$, so when $t = \frac{1}{d-1}\log\frac{1}{\tau}$ the operators P and U_t have similar action.

We can now apply Theorem 6.1.6 to bound $\mathrm{d_{TV}}\left(P^{\ell-1}\delta_{\boldsymbol{x}_0}, \rho\right)$ with $\alpha = \delta_{\boldsymbol{x}_0}$ and $k = \ell - 1$:

$$\mathrm{d_{TV}}\left(P^{\ell-1}\delta_{\boldsymbol{x}_0}, \rho\right) \leq \left((1 + o(1))\tau\right)^{\ell-2}\sqrt{\tfrac{1}{2}\log\tfrac{1}{p}}.$$

Spectral norm of random geometric graph

We now return to bounding the expected trace of $\overline{A}_{\boldsymbol{G}}^{\ell}$; putting together the above, we have the bound

$$\begin{aligned}
\mathbb{E}\prod_{(i,j)\in C_\ell}(\overline{A}_{\boldsymbol{G}})_{ij} &\leq \Pr[C_\ell \in \boldsymbol{G}] - p^\ell \\
&\leq p^{\ell-1}\left(p + \mathrm{d_{TV}}\left(P^{\ell-1}\delta_{\boldsymbol{x}_0}, \rho\right)\right) - p^\ell \\
&\leq p^{\ell-1}\left((1 + o(1))\tau\right)^{\ell-2}\sqrt{\tfrac{1}{2}\log\tfrac{1}{p}}.
\end{aligned}$$

The coefficient N_{C_ℓ} in front of the $W = C_\ell$ term in Equation 6.2 is the number of sequences $i_1, \dots, i_\ell \in [n]$ which yield an ℓ-cycle graph; this happens if and only if all of the indices are distinct, so $N_{C_\ell} = \ell! \cdot \binom{n}{\ell} \leq n^\ell$. Hence the contribution of the ℓ-cycle to the sum is at most $((1+o(1))np\tau)^{\ell-2} \cdot \mathsf{poly}(n)$ when $p > 1/n$. By a careful accounting similar to the above for all graphs $W \in \mathcal{W}_\ell$, one can show that in the parameter regime $pn \gg \mathsf{polylog}(n)$ and $\ell = \mathsf{polylog} n$, the term $W = C_\ell$ contains $(1 - o(1))$ of the total value of this sum, so we obtain the bound

$$\left[\mathbb{E}\mathrm{tr}(\overline{A}_{\boldsymbol{G}}^\ell)\right]^{1/\ell} \leq \left((1+o(1)) \cdot ((1+o(1))np\tau)^{\ell-2} \cdot \mathsf{poly}(n)\right)^{1/\ell} = (1+o(1))np\tau,$$

when we choose $\ell = \omega(\log n)$. Applying Markov's inequality we conclude that

$$\|\overline{A}_{\boldsymbol{G}}\| \leq (1+o(1))np\tau$$

with high probability, and normalizing by the degrees (which concentrate well around np) we conclude our upper bound of τ in Theorem 6.1.3.

Chapter 7

Conclusion and Open Problems

So many things that you wish I knew
But the story of us might be ending soon [Swi10]

Throughout this thesis, we have seen several instances of the interplay between geometric ideas and techniques with the fundamentally discrete study of higher-order random walks and high-dimensional expansion.

We now survey some open problems in several distinct directions, building off of discrete sampling (Chapters 2 and 3), the toolkit for analyzing random geometric graphs (Chapters 4 and 5), and constructions of high-dimensional expanders (Chapter 6).

7.1 Additional questions about domain sparsification

The question of higher-order marginals. We saw in Section 3.3 that our analysis of our domain sparsification scheme is tight for distributions satisfying $\frac{1}{\alpha}$-entropic independence. However, our analysis is not necessarily tight if we instead impose the stronger condition of fractional log-concavity. In this case, if we use the higher-order marginals, could we sparsify domains to sizes as small as $O(\mathsf{poly}(k) \cdot \mathsf{polylog}(n))$? We make the following conjecture.

Conjecture 7.1.1 (Informal)**.** Let μ be an α-fractionally-log-concave distribution for some $\alpha = \Omega(1)$. Given access to estimates for high-order marginals of the form $\Pr_{S\sim\mu}[T \subseteq S]$ for all T of size $\ell \simeq 1/\alpha$, and an oracle that produces i.i.d. samples from these marginals, there is a domain sparsification scheme for μ which reduces the domain size to $O(\mathsf{poly}(k)\cdot\mathsf{polylog}(n))$.

We suspect that obtaining these domain sparsification schemes likely require entirely new ideas. In fact, the original problem we set out to solve was: given pairwise marginals and a $\frac{1}{2}$-fractionally log-concave distribution, can we sparsify the domain to the desired $O(\mathsf{poly}(k) \cdot \mathsf{polylog}(n))$ size? If we wanted to generalize the strategy of [AD20] to pairwise marginals, we would need to count matchings in non-bipartite graphs, which still does not

admit a FPRAS. Perhaps a reasonable starter question, specifically for $\alpha = \frac{1}{2}$ is whether we can find such a domain sparsification scheme for the matching distribution or the asymmetric DPP, which each have additional structure that might make the problem more tractable.

Extension to continuous domains. To our knowledge, the framework of higher order random walks has not been formally extended to distributions over size-k subsets of a continuous domain; for instance, there is no explicit local-to-global theorem for continuous domains. However, the work of [GR18] suggests that this is feasible. Their "Gibbs sampler" implements a continuous version of a down-up walk, which is a higher-order walk known to rapidly mix on HDXes.

One distribution of interest where the domain is continuous is the uniform distribution over all vector embeddings $\{v_1, \ldots, v_k\}$ that realize a particular k-vertex graph G. From our work on random geometric graphs, this problem is likely easier when the underlying dimension d is high since the intersections of sphere caps exhibit tighter concentration. Unfortunately, this is also the less-interesting setting where random geometric graphs are statistically indistinguishable from Erdős-Rényi graphs anyways. It would be interesting to find a regime where d is small, *and* the higher-order random walks allow us to sample from this distribution.

If we could understand higher-order random walks in continuous domains, we can also ask if we can implement domain sparsification. Does some version of subsampling to a discrete subdomain suffice? Or could we be even use an inherently continuous method, like random subspace projection?

7.2 Geometric graphs for modeling and data science

Alternative geometric graph models. One particular family of random geometric graphs that could model communities and clustering is the geometric generalization of the stochastic block model. There has been some prior work in the area ([GMPS18b]) that defines the *geometric block model* by first sampling unit vectors uniformly from $\mathbb{R}^d$, choosing an arbitrary partition of the vectors into two "communities," and then connecting vectors within each community using a lower dot product threshold than vectors between the two communities. The techniques we developed in [LMSY22b] could potentially be applied to understand when we can distinguish the geometric block model from the stochastic block model.

We could impose even more structure to the block model: for instance, we can form the block model's communities based on geometric criteria rather than choosing them arbitrarily. As an example, we could use a random hyperplane to divide the vectors. This may be a more realistic model for the communities we see in practice, where feature vectors for members in the same community tend to be closer. In this setting, we would hope that information about the underlying geometry of the communities will give us more effective community recovery algorithms.

Other interesting models of random geometric graphs include those where edges are placed between vectors probabilistically ([LR21b; LR21a]), or where edges are placed between vectors if the Euclidean distance between them falls into a range rather than exceeds a threshold ([GMPS18a]). Recently, [MLS22] also study the problem of distinguishing an Erdös-Rényi graph from a random geometric graph living in hyperbolic (negatively-curved) space, as opposed to spherical (positively-curved) space. Fascinatingly, pure triangle counts, which were effective for detecting spherical geometry ([BDER16]) are not effective tests for underlying hyperbolic geometry. The question of indistinguishability in hyperbolic space also remains open, and could be an interesting avenue of exploration as well.

Higher-order network analysis. Higher-order network analysis is an emerging research area, motivated by the observation that some network dynamics cannot be modeled solely using pairwise interactions (i.e. edges). In fact, there are networks we encounter in ecology and neuroscience where models that include "higher-order" interactions, i.e. hypergraphs networks, are suspected to provide better insights on the network behavior ([BGH21]).

Several natural theory questions ensue. What kinds of structural assumptions do we need to place, so that we can get theoretical guarantees on the hypergraph algorithms used in practice, like higher-order PageRank and clustering ([GLY15; YBLG17])? Also, does domain sparsification admit a reasonable heuristic for speeding up such algorithms in practice?

7.3 New constructions of HDXes

As noted in Chapter 1, the study of high-dimensional expansion is still a emerging field, and we do not fully understand their power over expander graphs in many application areas. Finding additional application areas where HDXes have a demonstrated advantage is itself an ongoing open problem.

We also remarked earlier that we don't know many constructions of HDXes, to either use in applications or, at a more basic level, to gain understanding of their inner workings. We took a step towards this by proving local spectral expansion of random geometric graphs, but they left something to be desired, as the parameters we required for local spectral expansion forced polynomial average degree.

Another shortcoming of our construction is that we only obtain 2-dimensional expanders, rather than k-dimensional ones for any constant k. However, we attribute this to a bottleneck in analysis rather than a lack of confidence in the expansion of the k-faces' links in $\mathsf{Geo}_d(n,p)$. To analyze expansion of links of faces with more than one vertex, we would need to analyze the random walk of the intersection of multiple sphere caps, which has a much more irregular structure than a single sphere cap.

The random restriction route. We may hope to obtain HDXes with better parameters if we consider different underlying geometries, rather than just the unit sphere. As noted in Section 6.1.1, the random geometric complex fits in the broader framework of *random*

restrictions of simplicial complexes: starting with a dense high-dimensional expander X, we sample a subset of vertices $\boldsymbol{S}$ of X to produce the sparser induced complex $X[\boldsymbol{S}]$.

We have shown in Section 6.1.4 that $X[\boldsymbol{S}]$ (to some extent) inherits the spectral properties of X itself, and we've leveraged this to show that for any polynomial average degree, one can produce a 2-dimensional expander by taking a random restriction of X the sphere in a particular dimension and with a particular connectivity distance. We hope that Section 6.1.4 (or a strengthening thereof, see Remark 6.1.5) might help us identify additional natural distributions over sparser and/or higher-dimensional complexes. More specifically,

> Is there a simplicial complex X whose random restrictions yield high-dimensional expanders, whose links have eigenvalue $< \frac{1}{2}$, of sub-polynomial or polylogarithmic degree?

We note that constant average degree would likely require additional work; this is not just because of the polylogarithmic factors appearing in the statement of Theorem 6.1.4, but because in a random restriction, the degree distribution of each vertex is $\mathsf{Binom}(n, p)$ and so when $p = \Theta(1/n)$, we will have isolated vertices. This is the same as the phenomenon wherein Erdös-Rényi graphs of degree $O(1)$ are not expanders until one restricts to the giant component. As a starting point, we remark that geometric graphs on the unit sphere work because the corresponding X itself has link expansion better than $\frac{1}{2}$, witnessing that $\mathbb{S}^{d-1}$ itself is an expander.

Bibliography

[AASV21] Yeganeh Alimohammadi et al. "Fractionally log-concave and sector-stable polynomials: counting planar matchings and more". In: *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*. 2021, pp. 433–446.

[AB02] Réka Albert and Albert-László Barabási. "Statistical mechanics of complex networks". In: *Reviews of modern physics* 74.1 (2002), p. 47.

[Abb+22] Sepehr Abbasi-Zadeh et al. "Sticky brownian rounding and its applications to constraint satisfaction problems". In: *ACM Transactions on Algorithms (TALG)* 18.4 (2022), pp. 1–50.

[AD20] Nima Anari and Michał Dereziński. "Isotropy and Log-Concave Polynomials: Accelerated Sampling and High-Precision Counting of Matroid Bases". In: *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE. 2020, pp. 1331–1344.

[ADVY22] Nima Anari et al. "Domain Sparsification of Discrete Distributions Using Entropic Independence". In: *13th Innovations in Theoretical Computer Science Conference (ITCS 2022)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik. 2022.

[AJKPV21a] Nima Anari et al. "Entropic Independence II: Optimal Sampling and Concentration via Restricted Modified Log-Sobolev Inequalities". In: *arXiv preprint arXiv:2111.03247* (2021).

[AJKPV21b] Nima Anari et al. "Entropic Independence in High-Dimensional Expanders: Modified Log-Sobolev Inequalities for Fractionally Log-Concave Polynomials and the Ising Model". In: *arXiv preprint arXiv:2106.04105* (2021).

[AL20] Vedat Levi Alev and Lap Chi Lau. "Improved Analysis of Higher Order Random Walks and Applications". In: *arXiv e-prints* (2020), arXiv–2001.

[ALG21] Nima Anari, Kuikui Liu, and Shayan Oveis Gharan. "Spectral independence in high-dimensional expanders and applications to the hardcore model". In: *SIAM Journal on Computing* 0 (2021), FOCS20–1.

[ALG22] Dorna Abdolazimi, Kuikui Liu, and Shayan Oveis Gharan. "A matrix trickle-down theorem on simplicial complexes and applications to sampling colorings". In: *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE. 2022, pp. 161–172.

[ALGV19] Nima Anari et al. "Log-Concave Polynomials II: High-Dimensional Walks and an FPRAS for Counting Bases of a Matroid". In: *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*. 2019, pp. 1–12.

[AS04] Noga Alon and Joel H Spencer. *The probabilistic method*. John Wiley & Sons, 2004.

[AV21] Nima Anari and Thuy-Duong Vuong. "Simple and near-optimal map inference for nonsymmetric dpps". In: *arXiv preprint arXiv:2102.05347* (2021).

[AVKK03] Noga Alon et al. "Random sampling and approximation of MAX-CSPs". In: *Journal of computer and system sciences* 67.2 (2003), pp. 212–243.

[Bar16] Alexander Barvinok. *Combinatorics and complexity of partition functions*. Vol. 30. Springer, 2016.

[BBL09] Julius Borcea, Petter Brändén, and Thomas Liggett. "Negative dependence and the geometry of polynomials". In: *Journal of the American Mathematical Society* 22.2 (2009), pp. 521–567.

[BBN20] Matthew Brennan, Guy Bresler, and Dheeraj Nagaraj. "Phase transitions for detecting latent geometry in random graphs". In: *Probability Theory and Related Fields* 178.3 (2020), pp. 1215–1289.

[BCCFV10] Aditya Bhaskara et al. "Detecting high log-densities: an $O(n^{1/4})$ approximation for densest k-subgraph". In: *Proceedings of the forty-second ACM symposium on Theory of computing*. 2010, pp. 201–210.

[BDER16] Sébastien Bubeck et al. "Testing for high-dimensional geometry in random graphs". In: *Random Structures & Algorithms* 49.3 (2016), pp. 503–532.

[Ber16] Nathanaël Berestycki. "Mixing times of markov chains: Techniques and examples". In: *Alea-Latin American Journal of Probability and Mathematical Statistics* (2016).

[BGH21] Austin R Benson, David F Gleich, and Desmond J Higham. "Higher-order network analysis takes off, fueled by classical ideas and new data". In: *arXiv preprint arXiv:2103.05031* (2021).

[BGL13] Dominique Bakry, Ivan Gentil, and Michel Ledoux. *Analysis and geometry of Markov diffusion operators*. Vol. 348. Springer Science & Business Media, 2013.

[BHHS11] Boaz Barak et al. "Subsampling mathematical relaxations and average-case complexity". In: *Proceedings of the twenty-second annual ACM-SIAM Symposium on Discrete Algorithms*. SIAM. 2011, pp. 512–531.

[BHK11] Eric Babson, Christopher Hoffman, and Matthew Kahle. "The fundamental group of random 2-complexes". In: *Journal of the American Mathematical Society* 24.1 (2011), pp. 1–28.

[BK96] András A Benczúr and David R Karger. "Approximating st minimum cuts in $\tilde{O}$ (n 2) time". In: *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*. 1996, pp. 47–55.

[Bla+22] Antonio Blanca et al. "On Mixing of Markov Chains: Coupling, Spectral Independence, and Entropy Factorization". In: *Proceedings of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. SIAM. 2022, pp. 3670–3692.

[BLM15] Charles Bordenave, Marc Lelarge, and Laurent Massoulié. "Non-backtracking spectrum of random graphs: community detection and non-regular ramanujan graphs". In: *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*. IEEE. 2015, pp. 1347–1357.

[BLMV17] Rémi Bardenet et al. "On a few statistical applications of determinantal point processes". In: *ESAIM: Proceedings and Surveys* 60 (2017), pp. 180–202.

[Bor13] Charles Bordenave. "On Euclidean random matrices in high dimension". In: *Electronic Communications in Probability* 18 (2013), pp. 1–8.

[Bra+20] Jan van den Brand et al. "Bipartite matching in nearly-linear time on moderately dense graphs". In: *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE. 2020, pp. 919–930.

[Bri+01] Andreas Brieden et al. "Deterministic and randomized polynomial-time approximation of radii". In: *Mathematika* 48.1-2 (2001), pp. 63–105.

[BSST13] Joshua Batson et al. "Spectral sparsification of graphs: theory and algorithms". In: *Communications of the ACM* 56.8 (2013), pp. 87–94.

[BT03] Sergey Bobkov and Prasad Tetali. "Modified log-Sobolev inequalities, mixing and hypercontractivity". In: *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*. 2003, pp. 287–296.

[CDKSV16] L Elisa Celis et al. "On the complexity of constrained determinantal point processes". In: *arXiv preprint arXiv:1608.00554* (2016).

[CDV20] Daniele Calandriello, Michal Derezinski, and Michal Valko. "Sampling from a k-DPP without looking at all items". In: *Advances in Neural Information Processing Systems* 33 (2020), pp. 6889–6899.

[CGM19] Mary Cryan, Heng Guo, and Giorgos Mousa. "Modified log-Sobolev inequalities for strongly log-concave distributions". In: *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE. 2019, pp. 1358–1370.

[CGŠV21] Zongchen Chen et al. "Rapid mixing for colorings via spectral independence". In: *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*. SIAM. 2021, pp. 1548–1557.

[Che+22] Li Chen et al. "Maximum flow and minimum-cost flow in almost-linear time". In: *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE. 2022, pp. 612–623.

[Chu07] Fan Chung. "Four proofs for the Cheeger inequality and graph partition algorithms". In: *Proceedings of ICCM*. Vol. 2. Citeseer. 2007, p. 378.

[CLP20] Michael Chapman, Nati Linial, and Yuval Peled. "Expander graphs – both local and global". In: *Combinatorica* 40.4 (2020), pp. 473–509.

[Con19] David Conlon. "Hypergraph expanders from Cayley graphs". In: *Israel Journal of Mathematics* 233.1 (2019), pp. 49–65.

[CS13] Xiuyuan Cheng and Amit Singer. "The spectrum of random inner-product kernel matrices". In: *Random Matrices: Theory and Applications* 2.04 (2013), p. 1350010.

[CSŻ03] Donald I Cartwright, Patrick Solé, and Andrzej Żuk. "Ramanujan geometries of type $\tilde{A}_n$". In: *Discrete mathematics* 269.1-3 (2003), pp. 35–43.

[CTZ20] David Conlon, Jonathan Tidor, and Yufei Zhao. "Hypergraph expanders of all uniformities from Cayley graphs". In: *Proceedings of the London Mathematical Society* 121.5 (2020), pp. 1311–1336.

[DBPM20] Michal Derezinski et al. "Debiasing distributed second order optimization with surrogate sketching and scaled regularization". In: *Advances in Neural Information Processing Systems* 33 (2020), pp. 6684–6695.

[DCMW19] Michał Dereziński et al. "Minimax experimental design: Bridging the gap between statistical and worst-case approaches to least squares regression". In: *Conference on Learning Theory*. PMLR. 2019, pp. 1050–1069.

[DCV19] Michal Derezinski, Daniele Calandriello, and Michal Valko. "Exact sampling of determinantal point processes with sublinear time preprocessing". In: *Advances in neural information processing systems* 32 (2019).

[DEKL14] Jean Dolbeault et al. "Sharp interpolation inequalities on the sphere: new methods and consequences". In: *Partial Differential Equations: Theory, Control and Approximation*. Springer, 2014, pp. 225–242.

[DELLM22] Irit Dinur et al. "Locally Testable Codes with Constant Rate, Distance, and Locality". In: *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*. 2022, pp. 357–374.

[Der19] Michał Dereziński. "Fast determinantal point processes via distortion-free intermediate sampling". In: *Conference on Learning Theory*. PMLR. 2019, pp. 1029–1049.

[DGLU11] Luc Devroye et al. "High-dimensional random geometric graphs and their clique number". In: *Electronic Journal of Probability* 16 (2011), pp. 2481–2508.

[Din07] Irit Dinur. "The PCP Theorem by Gap Amplification". In: *Journal of the ACM (JACM)* 54.3 (2007), 12–es.

[DKM20] Michal Derezinski, Rajiv Khanna, and Michael W Mahoney. "Improved guarantees and a multiple-descent curve for Column Subset Selection and the Nystrom method". In: *Advances in Neural Information Processing Systems* 33 (2020), pp. 4953–4964.

[DM21] Michał Derezinski and Michael W Mahoney. "Determinantal point processes in randomized numerical linear algebra". In: *Notices of the American Mathematical Society* 68.1 (2021), pp. 34–45.

[DR10] Amit Deshpande and Luis Rademacher. "Efficient volume sampling for row/column subset selection". In: *2010 ieee 51st annual symposium on foundations of computer science*. IEEE. 2010, pp. 329–338.

[DRVW06] Amit Deshpande et al. "Matrix approximation and projective clustering via volume sampling". In: *Theory of Computing* 2.1 (2006), pp. 225–247.

[DV13] Yen Do and Van Vu. "The spectrum of random kernel matrices: universality results for rough and varying kernels". In: *Random Matrices: Theory and Applications* 2.03 (2013), p. 1350005.

[DW17] Michal Derezinski and Manfred KK Warmuth. "Unbiased estimates for linear regression via volume sampling". In: *Advances in Neural Information Processing Systems* 30 (2017).

[DWH18] Michal Derezinski, Manfred KK Warmuth, and Daniel J Hsu. "Leveraged volume sampling for linear regression". In: *Advances in Neural Information Processing Systems* 31 (2018).

[DWH22] Michał Dereziński, Manfred K Warmuth, and Daniel Hsu. "Unbiased estimators for random design regression". In: *The Journal of Machine Learning Research* 23.1 (2022), pp. 7539–7584.

[EK16a] Shai Evra and Tali Kaufman. "Bounded degree cosystolic expanders of every dimension". In: *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*. 2016, pp. 36–48.

[EK16b] Shai Evra and Tali Kaufman. "Bounded degree cosystolic expanders of every dimension". In: *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*. 2016, pp. 36–48.

[El 10] Noureddine El Karoui. "The spectrum of kernel random matrices". In: *The Annals of Statistics* 38.1 (2010), pp. 1–50.

[Erd59] Paul Erdös. "Graph theory and probability". In: *Canadian Journal of Mathematics* 11 (1959), pp. 34–38.

[EV19] David Eppstein and Vijay V Vazirani. "NC algorithms for computing a perfect matching, the number of perfect matchings, and a maximum flow in one-crossing-minor-free graphs". In: *The 31st ACM Symposium on Parallelism in Algorithms and Architectures*. 2019, pp. 23–30.

[FGLNP12] Jacob Fox et al. "Overlap properties of geometric expanders". In: *Journal für die reine und angewandte Mathematik (Crelles Journal)* 2012.671 (2012), pp. 49–83.

[FGYZ22] Weiming Feng et al. "Rapid mixing from spectral independence beyond the Boolean domain". In: *ACM Transactions on Algorithms (TALG)* 18.3 (2022), pp. 1–32.

[FI20] Ehud Friedgut and Yonatan Iluz. "Hyper-regular graphs and high dimensional expanders". In: *arXiv preprint arXiv:2010.03829* (2020).

[FK01] Uriel Feige and Joe Kilian. "Heuristics for semirandom graph problems". In: *Journal of Computer and System Sciences* 63.4 (2001), pp. 639–671.

[FM19] Zhou Fan and Andrea Montanari. "The spectral norm of random inner-product kernel matrices". In: *Probability Theory and Related Fields* 173.1 (2019), pp. 27–85.

[Fri03] Joel Friedman. "A Proof of Alon's Second Eigenvalue Conjecture". In: *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*. 2003, pp. 720–724.

[GBDK19] Mike Gartrell et al. "Learning nonsymmetric determinantal point processes". In: *Advances in Neural Information Processing Systems* 32 (2019).

[GGR98] Oded Goldreich, Shafi Goldwasser, and Dana Ron. "Property testing and its connection to learning and approximation". In: *Journal of the ACM (JACM)* 45.4 (1998), pp. 653–750.

[GHDGB20] Mike Gartrell et al. "Scalable learning and MAP inference for nonsymmetric determinantal point processes". In: *arXiv preprint arXiv:2006.09862* (2020).

[GL99] Anna Galluccio and Martin Loebl. "On the theory of Pfaffian orientations. I. Perfect matchings and permanents". In: *the electronic journal of combinatorics* (1999), R6–R6.

[GLY15] David F Gleich, Lek-Heng Lim, and Yongyang Yu. "Multilinear pagerank". In: *SIAM Journal on Matrix Analysis and Applications* 36.4 (2015), pp. 1507–1541.

[GMPS18a] Sainyam Galhotra et al. "Connectivity in Random Annulus Graphs and the Geometric Block Model". In: *arXiv preprint arXiv:1804.05013* (2018).

[GMPS18b] Sainyam Galhotra et al. "The Geometric Block Model". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 32. 1. 2018.

[Gol21] Louis Golowich. "Improved Product-Based High-Dimensional Expanders". In: *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2021)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik. 2021.

[GR18] Shayan Oveis Gharan and Alireza Rezaei. "A Polynomial Time MCMC Method for Sampling from Continuous DPPs". In: *arXiv preprint arXiv:1810.08867* (2018).

[Gro10a] Mikhail Gromov. "Singularities, expanders and topology of maps. Part 2: From combinatorics to topology via algebraic isoperimetry". In: *Geometric and Functional Analysis* 20.2 (2010), pp. 416–526.

[Gro10b] Mikhail Gromov. "Singularities, expanders and topology of maps. Part 2: From combinatorics to topology via algebraic isoperimetry". In: *Geometric and Functional Analysis* 20.2 (2010), pp. 416–526.

[GS12] Venkatesan Guruswami and Ali Kemal Sinop. "Optimal column-based low-rank matrix reconstruction". In: *Proceedings of the twenty-third annual ACM-SIAM symposium on Discrete Algorithms*. SIAM. 2012, pp. 1207–1214.

[Gue83] Alain Guenoche. "Random spanning tree". In: *Journal of Algorithms* 4.3 (1983), pp. 214–220.

[GW95] Michel X Goemans and David P Williamson. "Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming". In: *Journal of the ACM (JACM)* 42.6 (1995), pp. 1115–1145.

[HKPV06] J Ben Hough et al. "Determinantal processes and independence". In: (2006).

[Hop+17] Samuel B Hopkins et al. "The power of sum-of-squares for detecting hidden structures". In: *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE. 2017, pp. 720–731.

[INW94] Russell Impagliazzo, Noam Nisan, and Avi Wigderson. "Pseudorandomness for Network Algorithms". In: *Proceedings of the twenty-sixth annual ACM symposium on Theory of computing*. 1994, pp. 356–364.

[Jer03] Mark Jerrum. *Counting, Sampling and Integrating: Algorithms and Complexity*. Springer Science & Business Media, 2003.

[JPV21] Vishesh Jain, Huy Tuan Pham, and Thuy Duong Vuong. "Spectral independence, coupling with the stationary distribution, and the spectral gap of the Glauber dynamics". In: *arXiv preprint arXiv:2105.01201* (2021).

[JS88] Mark Jerrum and Alistair Sinclair. "Conductance and the rapid mixing property for Markov chains: the approximation of permanent resolved". In: *Proceedings of the twentieth annual ACM symposium on Theory of computing*. 1988, pp. 235–244.

[JSTV04] Mark Jerrum et al. "Elementary Bounds on Poincaré and Log-Sobolev constants for Decomposable Markov Chains". In: *The Annals of Applied Probability* 14.4 (2004), pp. 1741–1765.

[JSV04] Mark Jerrum, Alistair Sinclair, and Eric Vigoda. "A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries". In: *Journal of the ACM (JACM)* 51.4 (2004), pp. 671–697.

[JVV86] Mark R Jerrum, Leslie G Valiant, and Vijay V Vazirani. "Random generation of combinatorial structures from a uniform distribution". In: *Theoretical computer science* 43 (1986), pp. 169–188.

[Kas67] Pieter Kasteleyn. "Graph theory and crystal physics". In: *Graph theory and theoretical physics* (1967), pp. 43–110.

[KG00] Vladimir Koltchinskii and Evarist Giné. "Random matrix approximation of spectra of integral operators". In: *Bernoulli* (2000), pp. 113–167.

[KKL14] Tali Kaufman, David Kazhdan, and Alexander Lubotzky. "Ramanujan complexes and bounded degree topological expanders". In: *2014 IEEE 55th Annual Symposium on Foundations of Computer Science*. IEEE. 2014, pp. 484–493.

[KN14] Daniel M Kane and Jelani Nelson. "Sparser johnson-lindenstrauss transforms". In: *Journal of the ACM (JACM)* 61.1 (2014), pp. 1–23.

[KO18] Tali Kaufman and Izhar Oppenheim. "Construction of new local spectral high dimensional expanders". In: *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*. 2018, pp. 773–786.

[KT+12] Alex Kulesza, Ben Taskar, et al. "Determinantal point processes for machine learning". In: *Foundations and Trends® in Machine Learning* 5.2–3 (2012), pp. 123–286.

[KT11] Alex Kulesza and Ben Taskar. "k-DPPs: Fixed-size determinantal point processes". In: *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*. 2011, pp. 1193–1200.

[Li04] Wen-Ching Winnie Li. "Ramanujan hypergraphs". In: *Geometric & Functional Analysis GAFA* 14.2 (2004), pp. 380–399.

[LM06] Nathan Linial and Roy Meshulam. "Homological connectivity of random 2-complexes". In: *Combinatorica* 26.4 (2006), pp. 475–487.

[LMM16] Alexander Lubotzky, Roy Meshulam, and Shahar Mozes. "Expansion of building-like complexes". In: *Groups, Geometry, and Dynamics* 10.1 (2016), pp. 155–175.

[LMSY22a] Siqi Liu et al. "Local and Global Expansion in Random Geometric Graphs". In: *arXiv preprint arXiv:2210.00158* (2022).

[LMSY22b] Siqi Liu et al. "Testing Thresholds for High-Dimensional Sparse Random Geometric Graphs". In: *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*. 2022, pp. 672–677.

[LMY20] Siqi Liu, Sidhanth Mohanty, and Elizabeth Yang. "High-Dimensional Expanders from Expanders". In: *11th Innovations in Theoretical Computer Science Conference (ITCS 2020)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik. 2020.

[LP16] Nathan Linial and Yuval Peled. "On the phase transition in random simplicial complexes". In: *Annals of mathematics* (2016), pp. 745–773.

[LPS88a] Alexander Lubotzky, Ralph Phillips, and Peter Sarnak. "Ramanujan Graphs". In: *Combinatorica* 8.3 (1988), pp. 261–277.

[LPS88b] Alexander Lubotzky, Ralph Phillips, and Peter Sarnak. "Ramanujan graphs". In: *Combinatorica* 8.3 (1988), pp. 261–277.

[LR21a] Suqi Liu and Miklos Z Racz. "A Probabilistic View of Latent Space Graphs and Phase Transitions". In: *arXiv preprint arXiv:2110.15886* (2021).

[LR21b] Suqi Liu and Miklós Z Rácz. "Phase Transition in Noisy High-Dimensional Random Geometric Graphs". In: *arXiv preprint arXiv:2103.15249* (2021).

[LRS15] James R Lee, Prasad Raghavendra, and David Steurer. "Lower bounds on the size of semidefinite programming relaxations". In: *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*. 2015, pp. 567–576.

[LSV05a] Alexander Lubotzky, Beth Samuels, and Uzi Vishne. "Explicit constructions of Ramanujan complexes of type Ad". In: *European Journal of Combinatorics* 26.6 (2005), pp. 965–993.

[LSV05b] Alexander Lubotzky, Beth Samuels, and Uzi Vishne. "Ramanujan complexes of typeÃ d". In: *Israel journal of Mathematics* 149.1 (2005), pp. 267–299.

[Lub18] Alexander Lubotzky. "High dimensional expanders". In: *Proceedings of the International Congress of Mathematicians: Rio de Janeiro 2018*. World Scientific. 2018, pp. 705–730.

[Lub22] Alex Lubotzky. *High-Dimensional Expanders and Property Testing*. STOC workshop on advances in coding theory, `https://youtu.be/P5vs4ARRQjU?t=2967`. 2022.

[LY22] Yue M Lu and Horng-Tzer Yau. "An Equivalence Principle for the Spectrum of Random Inner-Product Kernel Matrices". In: *arXiv preprint arXiv:2205.06308* (2022).

[Mac75] Odile Macchi. "The coincidence approach to stochastic point processes". In: *Advances in Applied Probability* 7.1 (1975), pp. 83–122.

[Mas07] Pascal Massart. "Concentration inequalities and model selection". In: (2007).

[Mat08] Jiři Matoušek. "On variants of the Johnson–Lindenstrauss lemma". In: *Random Structures & Algorithms* 33.2 (2008), pp. 142–156.

[Mat13] Jiri Matousek. *Lectures on discrete geometry*. Vol. 212. Springer Science & Business Media, 2013.

[MDK20] Mojmir Mutny, Michal Derezinski, and Andreas Krause. "Convergence analysis of block coordinate algorithms with determinantal sampling". In: *International Conference on Artificial Intelligence and Statistics*. PMLR. 2020, pp. 3110–3120.

[MLS22] Riccardo Michielan, Nelly Litvak, and Clara Stegehuis. "Detecting Hyperbolic Geometry in Networks: Why Triangles Are Not Enough". In: *Physical Review E* 106.5 (2022), p. 054303.

[MM09] Marc Mézard and Andrea Montanari. *Information, physics, and computation*. Oxford University Press, 2009.

[MW09] Roy Meshulam and Nathan Wallach. "Homological connectivity of random k-dimensional complexes". In: *Random Structures & Algorithms* 34.3 (2009), pp. 408–417.

[NST22] Aleksandar Nikolov, Mohit Singh, and Uthaipon Tantipongpipat. "Proportional volume sampling and approximation algorithms for A-optimal design". In: *Mathematics of Operations Research* 47.2 (2022), pp. 847–877.

[ODo14] Ryan O'Donnell. *Analysis of boolean functions*. Cambridge University Press, 2014.

[OP22] Ryan O'Donnell and Kevin Pratt. "High-Dimensional Expanders from Chevalley Groups". In: *arXiv preprint arXiv:2203.03705* (2022).

[Opp18] Izhar Oppenheim. "Local Spectral Expansion Approach to High Dimensional Expanders Part I: Descent of Spectral Gaps". In: *Discrete & Computational Geometry* 59.2 (2018), pp. 293–330.

[Pen03] Mathew Penrose. *Random geometric graphs*. Vol. 5. OUP Oxford, 2003.

[PK22] Pavel Panteleev and Gleb Kalachev. "Asymptotically Good Quantum and Locally Testable Classical LDPC codes". In: *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*. 2022, pp. 375–388.

[RVW00] Omer Reingold, Salil Vadhan, and Avi Wigderson. "Entropy Waves, the Zig-Zag Graph Product, and New Constant-Degree Expanders and Extractors". In: *Proceedings 41st Annual Symposium on Foundations of Computer Science*. IEEE. 2000, pp. 3–13.

[SS96] Michael Sipser and Daniel A Spielman. "Expander Codes". In: *IEEE Transactions on Information Theory* 42.6 (1996), pp. 1710–1722.

[Swi06] Taylor Swift. *Taylor Swift*. 2006.

[Swi10] Taylor Swift. *Speak Now*. 2010.

[Swi12] Taylor Swift. *Red*. 2012.

[Swi19] Taylor Swift. *Lover*. 2019.

[Swi21] Taylor Swift. *Fearless (Taylor's Version)*. 2021.

[Tal96] Michel Talagrand. "Transportation cost for Gaussian and other product measures". In: *Geometric & Functional Analysis GAFA* 6.3 (1996), pp. 587–600.

[Van14] Ramon Van Handel. *Probability in high dimension*. Tech. rep. Princeton University, 2014.

[Ver18] Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*. Vol. 47. Cambridge university press, 2018.

[Vil08] Cédric Villani. *Optimal transport: old and new*. Vol. 338. Springer Science & Business Media, 2008.

[WS11] David P Williamson and David B Shmoys. *The design of approximation algorithms*. Cambridge university press, 2011.

[WS98] Duncan J Watts and Steven H Strogatz. "Collective dynamics of 'small-world'networks". In: *nature* 393.6684 (1998), pp. 440–442.

[YBLG17] Hao Yin et al. "Local higher-order graph clustering". In: *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*. 2017, pp. 555–564.