A FEW COMBINATORIAL PROBLEMS

BY ROSS BERKOWITZ

A dissertation submitted to the
Graduate School—New Brunswick
Rutgers, The State University of New Jersey
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
Graduate Program in Mathematics
Written under the direction of
Swastik Kopparty
and approved by

_____

_____

_____

_____

New Brunswick, New Jersey
May, 2017

ABSTRACT OF THE DISSERTATION

A Few Combinatorial Problems

by ROSS BERKOWITZ

Dissertation Director: Swastik Kopparty

This thesis studies three problems in combinatorics.

Our first result is a quantitative local limit theorem for the distribution of the number of triangles in the Erdős-Renyi random graph $G(n,p)$, for a fixed $p \in (0,1)$. This proof is an extension of the previous work of Gilmer and Kopparty, who proved that the local limit theorem held asymptotically for triangles. Our work gives bounds on the ℓ^1 and ℓ^∞ distance of the triangle distribution from a suitable discrete normal.

In our second result we prove a stability version of a general result that bounds the permanent of a matrix in terms of its operator norm. More specifically, suppose A is an $n \times n$ matrix, and let $\mathcal{P}$ denote the set of $n \times n$ matrices that can be written as a permutation matrix times a unitary diagonal matrix. Then it is known that the permanent of A satisfies $|\mathrm{per}(A)| \leq \|A\|_2^n$ with equality iff $A/\|A\|_2 \in \mathcal{P}$ (where $\|A\|_2$ is the operator 2-norm of A). We show a stability version of this result asserting that unless A is very close (in a particular sense) to one of these extremal matrices, its permanent is exponentially smaller (as a function of n) than $\|A\|_2^n$. In particular, for any fixed $\alpha, \beta > 0$, we show that $|\mathrm{per}(A)|$ is exponentially smaller than $\|A\|_2^n$ unless all but at most αn rows contain entries of modulus at least $\|A\|_2(1-\beta)$.

Finally, we construct large sequences with the property that the contents of any small window determine the location of the window, *robustly*. Such objects have found

many applications in practical settings, from positioning of wireless devices to smart pens, and have recently gained some theoretical interest. In this context, we give the first explicit constructions of sequences with high rate and constant relative distance. Accompanying these efficient constructions, we also give efficient decoding algorithms, which can determine the position of the window given its contents, even if a constant fraction of the contents have been corrupted.

Acknowledgements

I want to thank my Mom for fostering my love of mathematics and my Dad for always having advice and for supporting me in general, even when I was unable to crack the fudge enigma. I also have to thank my sister, Lesley, for teaching me how to be a nerd.

Swastik Kopparty was an excellent adviser. He is always cheerful, helps me to find fruitful problems, is willing to listen to my crazy ideas, and to correct them when necessary. I also want to thank the other members of my committee, Doron Zeilberger and Jeff Kahn. Dr. Z was a teacher who always made his subject matter interesting, and taught me the value of attacking a problem with experimentation and not being afraid to try a new approach. Professor Kahn not only introduced me to interesting material, but he also supported me more than he was called upon to do.

I want to thank all of my fellow graduate students for fostering a fun, friendly and creative environment at the department. In particular I want to thank Pat Devlin for being the only other student crazy enough to work with me, Ed Karasiewicz for always being willing to talk math and inspiring me to work harder, Tom Sznigir and Jake Baron for inspiring me not to work too hard, and Justin Gilmer for lending me his problems.

I want to thank my cat, Garfield, for being cuddly.

Lastly, I want to thank Dana for bringing some much needed flowers and color into my life.

Dedication

I dedicate this to Charles Xavier Zak. He used his mental powers for good, ensuring that his Bris took place the day before my defense.

Table of Contents

Chapter 1

Introduction

Discrete mathematics and combinatorics form a growing and lively branch of mathematics. The field addresses topics which are varied and broad in scope, as are the tools needed to address them. In this thesis we will draw heavily from, among other tools, probability theory, Fourier analysis, and the properties of low degree polynomials.

The particular problems we will address are

- A local limit theorem for triangles in random graphs.
- A stability result relating the permanent to the matrix norm.
- A sequence with good error correcting properties.

Each of these topics will be outlined in the following section.

1.1 Brief Outline of Results

1.1.1 Local limit theorems for subgraph counts

The random graph $G(n,p)$, a model in which a graph on n vertices is selected by including each edge independently with probability p, is an important object of study in graph theory. We say that three vertices of G form a triangle if they are all pairwise connected by edges. Questions of the form, "how many triangles (or any fixed subgraph) do we see in a random graph" have been considered since the initial papers of Erdős and Renyi introducing the topic [13]. Erdős and Renyi showed in 1960 that the number of triangles in a graph is tightly concentrated about the expected value of $p^3\binom{n}{3}$. One question which received much attention in the '80s concerned pinning down the limiting distribution of subgraph counts in random graphs asymptotically as the number of

vertices goes to infinity. This line of study eventually lead to the work of Ruciński, which characterized when subgraph counts converged to a Gaussian [39].

In 2014, Gilmer and Kopparty showed a local limit theorem for the number of triangles in a random graph in the regime where p is a fixed constant [16]. They gave a qualitative result estimating the probability that a random graph in $G(n,p)$ has *exactly* k triangles. This should be contrasted with the previously shown central limit theorems, which only gave a coarse description of the number of triangles. I improve upon this result by proving a strong quantitative estimate for the distribution of triangles in random graph. A consequence of Theorem 2 is that for k near the the mean μ,

$$\Pr[\mathcal{T} = k] = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(k-\mu)^2}{2\sigma^2}\right)\left(1 + O(n^{-\frac{1}{2}-\epsilon})\right)$$

where $\mathcal{T}$ is the triangle counting function. Importantly, this strengthened result also allows us to prove that the distribution of triangle counts in random graphs is close to a discrete Gaussian in the ℓ^1 or total variation metric, a result which was previously out of reach. Further, the methods used extend to showing local limit theorems for paths of length 2 and providing bounds on the characteristic function of the random variables given by any subgraph count.

The ingredients of the proof involve a careful analysis of the Fourier transform of the subgraph count random variable, also known as its characteristic function. Characteristic functions of sums of dependent random variables, however, are typically difficult to analyze. To help with this difficulty, we decompose the triangle counting function using the orthogonal Walsh basis, a discrete analog of the traditional Fourier transform. This second transform helps reveal the underlying structure and low degree spectral concentration of the counting function, which we are then able to exploit to understand why these random variables are asymptotically normal. The results in this section were first posted here [6]

1.1.2 A stability result using the matrix norm to bound the permanent

The permanent of a square $n \times n$ matrix is defined to be

$$\mathrm{per}(A) := \sum_{\sigma \in S_n} \prod_{i=1}^{n} a_{i\sigma(i)}$$

The permanent is a classical object of study with connections to computational complexity and graph theory. Unlike the similarly defined determinant, the permanent is notoriously difficult to compute. Valiant showed [43] that computing the permanent of 0-1 matrices is $\#P$ complete, and even approximating the permanent is still a difficult problem.

In 2005 Leonid Gurvits [19] proved that the permanent of a matrix is upper bounded by its L_2 operator norm $\|A\|_2 := \sup_{\|x\|_2=1} \|Ax\|_2$. This bound is tight, with the extremal case of a permutation matrix P which has $\mathrm{per}(P) = \|P\|_2 = 1$. Motivated by an application to derandomizing permanent approximation, Aaronson and Hance asked whether it was possible to categorize when the permanent of a unitary matrix was large [2]. Aaronson and Nguyen later asked more formally if one could characterize $n \times n$ matrices such that $\|A\|_2 \leq 1$, but with permanent at least as large as n^{-C} for some fixed constant C [3].

In a joint work with Patrick Devlin we resolve these questions by showing that unless A is close to a permutation matrix, then $\mathrm{per}(A)$ is exponentially smaller than $\|A\|_2$. The definition needed to define closeness is

$$h_\infty(A) = \frac{1}{n} \sum_{i=1}^{n} \max_j(|a_{ij}|)$$

$h_\infty(A)$ returns the average of the largest entry in each row. Note that permutation matrices satisfy $h_\infty(A) = 1$. Conversely, matrices with $\|A\|_2 = 1$ and $h_\infty(A)$ close to 1 must be nearly permutation matrices in that almost every row and column is dominated by a single large entry. Formally stated, we proved that for a real $n \times n$ matrix A with

$\|A\|_2 = 1$ that

$$|\mathrm{per}(A)| \leq 2\exp\bigl(-n(1-h_\infty(A))^2/10^5\bigr)$$

For example, if we know that a matrix has every entry bounded above by $|a_{ij}| < .999$, then we find that $\mathrm{per}(A) \leq \exp(-\Omega(n))$.

The proof proceeds by expressing $\mathrm{per}(A)$ as the expectation of a random variable, then analyzing this random variable using various concentration inequalities including Talagrand's inequality and an extension of Khintchine's inequality. This chapter is a modified version of the draft [7], which is to appear in the Israel Journal of Mathematics.

1.1.3 Robust positioning patterns

Coding theory studies the error correcting properties of various forms of data subject to models of noisy channels. The field began with the pioneering work of Claude Shannon, and has since proven ripe for both application and theoretical study. One interesting model that has recently received attention is that of the error correcting sequence or torus. These are sequences of N symbols from some alphabet Σ with the property that all windows of n contiguous entries are significantly different from one another. Similarly, in the 2-dimensional setting, we study tori whose windows of $n \times n$ contiguous squares form a code of good distance. In the noiseless regime, these objects are the classical DeBruijn sequences and tori.

An illustrative application [1] of error correcting tori is simulating a touch screen with a device capable only of display, for example a monitor or special notepad. To do this, write an error correcting torus on the display and build a special stylus with a camera in its tip. When the stylus is pressed to the display its camera can read the window visible from its location on the screen and relay this information to the computer. The computer can then use this window, even if much of the data was corrupted by dirt or other text, to discover the location of the stylus.

The two important parameters measuring these objects are the rate R, which measures how large the sequence is relative to its theoretical maximum, and its relative distance $\delta(S)$, which measures the fractional number of errors the sequence can correct.

Kumar and Wei [29] showed that a random linear feedback shift register gives nearly optimal trade-off between rate and distance in the regime where $\delta(S) = O(n^{-1/2})$. In the 2-dimensional setting, Bruckstein, Etzion, Gordon, Holt, and Shuldiner [10] gave a construction of an error correcting torus, which was robust to a constant fraction errors, but had rate $o(1)$.

In a joint work with Swastik Kopparty we give an explicit construction of error correcting sequences with a near optimal trade-off between rate and distance. In particular, our construction yields sequences with constant rate and constant distance for both the large and small alphabet regimes. The tools used to construct this sequence included Reed-Solomon codes and a careful analysis of their structure, Gray codes, and concatenation.

Additionally, the sequence has a natural decoding algorithm based on the Guruswami-Sudan list decoding algorithm, which can recover position from a noisy window in polynomial time. The results in this section were first presented at SODA16 [8].

1.2 Overview of Techniques

1.2.1 Triangle LLT

The main result of the first chapter of this thesis is the following local limit theorem for the number of triangles in $G(n,p)$

Theorem 1. *For any $k \in \mathbb{N}$ we have that*

$$\Pr[\mathcal{T} = k] = \frac{1}{\sqrt{2\pi}\sigma_n} e^{-\frac{\left(k - p^3\binom{n}{2}\right)^2}{2\sigma_n^2}} + O(n^{-2.5+\epsilon})$$

where σ_n is the variance of the triangle counting random variable $\mathcal{T}$.

Our arguments are based on giving better bounds on the characteristic function of $\mathcal{T}$. The main improvements come from combining two different Fourier transforms, the first being the characteristic function, and the second being the finite Fourier transform of the triangle counting function. In particular, we will choose the p-biased Fourier basis over $\{0,1\}^{\binom{[n]}{2}}$ with basis functions denoted χ_S where $S \subset \binom{[n]}{2}$. That is, each

basis element is a function depending on some subset of the possible edges in our graph. The main mass of the triangle counting function will come from basis elements of the form χ_e, where e is an edge in $\binom{[n]}{2}$. In other words, we will find that $\mathcal{T}$ is highly concentrated on its weight 1 Fourier coefficients. This allows us to show that $\mathcal{T}$ may be reasonably well approximated as simply a linear function of the number of edges in the random graph. Informally, this follows the intuition that if one wanted to know how many triangles are in a fixed graph G, a reasonable estimator would be to simply ask how many edges are in the graph, and scale appropriately. This method is closely related to the method employed by Janson and Nowicki in [23]. It also bears similarity to the work of Friedgut [15] who gave an early and important example of the usefulness of the p-biased Fourier transform in studying random graphs.

The actual estimation will be performed in two steps. First we will normalize $\mathcal{T}$ to have mean 0 and variance 1 by defining $Z := \frac{\mathcal{T}-\mu}{\sigma}$. Then we will split Z up into two pieces $Z = X + Y$, where X contains the weight 1 fourier terms which dominate Z, and Y contains the higher order terms, which we will treat as error terms. We then use as blunt a tool as the mean value theorem to estimate the characteristic function by saying $\mathbb{E}[e^{itZ}] = \mathbb{E}[e^{itX} + O(|tY|)]$. Since X is a sum of i.i.d. random variables and Y is small this converges to the characteristic function of the normal distribution when t is small.

For slightly larger t we adapt this method slightly by first revealing some k-regular subgraph and then performing our estimates given this information. This will shrink the size of Y by a factor of $(k/n)^2$, but only shrink X by a factor of k/n. This gives us a better error term, at the cost of only slightly shrinking our main term. For this part of the argument we cannot give an exact main term for $|\varphi_Z(t)|$, as we could in the first method. However for t large, because the normal has very small characteristic function it suffices simply to show that $|\varphi_Z(t)|$ is very small as well. As part of this argument, we will need strong concentration of low degree functions (an application of hypercontractivity theorems) to ensure that even after we reveal a large portion of the edges, the randomness remaining in $\mathcal{T}$ is relatively well behaved.

1.2.2 Permanent Bounds

This section is focused on proving a bound on the permanent of matrices of operator norm at most 1. Recall that the L^2 operator norm is defined to be

$$\|A\|_2 := \sup_{\|x\|_2=1} \|Ax\|_2$$

Our first main result is (in an equivalent formulation of)

Theorem 2. *Let A be an $n \times n$ matrix over $\mathbb{C}$ and $\|A\|_2 \leq 1$. Then*

$$|\mathrm{per}(A)| \leq 2\exp[-n(1 - h_\infty(A))^2/10^5]$$

The "distance" operator we have used is

$$h_\infty(A) = \frac{1}{n}\sum_{i=1}^{n} \max_j(|a_{ij}|)$$

which can only be one in a matrix of operator norm 1 if A is essentially a permutation matrix. Our starting point in proving this result is to use a reformulation of the Permanent as the expectation of a particular random variable. The specific form we use is

$$\mathrm{per}(A) = \mathbb{E}[G_X(A)] = \mathop{\mathbb{E}}_{X\in\{-1,1\}^n}\left[\prod_{i=1}^{n} \overline{X}_i(AX)_i\right],$$

By the AM-GM inequality we can bound this above by $|\mathrm{per}(A)| \leq \mathbb{E}\left[\left(\frac{\|AX\|_1}{n}\right)^n\right]$. The rest of the work is dedicated to analyzing the random variable $\|AX\|_1$. Appealing to a result of König, Schütt, and Tomczak-Jaegermann[27] we are able to show that if A is far from a permutation matrix, then $\mathbb{E}\,\|AX\|_1/n$ is noticeably smaller than 1. The argument is then completed by bounding the probability of $\|AX\|_1$ deviating from its mean. This is done in two different settings. The most general setting is accomplished by using Talagrand's Inequality [30]. This approach yields the result given in Theorem 2.

Additionally, when $h_\infty(A)$ is very close to 1 and A is a real valued matrix, a tighter

concentration bound for $\|Ax\|_1$ can be achieved by appealing to hypercontractivity bounds. In particular, A is broken up into two submatrices: L the set of "little" rows which do not have any entries larger than $1-\lambda$ for an appropriately chosen value of λ, and B the set of "big" rows which are dominated by a single entry of size at least $1-\lambda$. For B the set of big rows, we will note that the sign of every row of Bx is with high probability the sign of the dominating entry of that row. Therefore

$$\|Bx\|_1 = \sum_{i=1}^{|B|} |Bx|_i \approx \sum_{i=1}^{|B|} (Bx)_i x_{ij_i^*}$$

where j_i^* is the index of the large entry in row i of B. The last term is simply a degree 2 polynomial with expectation $B_{ij_i^*}$ and variance at most 2λ (recall the rows of B are vectors of norm at most 1, and the entry $B_{ij_i^*}$ has size $\geq 1-\lambda$). Now an application of a hypercontractivity concentration result due to Bonami [9] for low degree polynomials showes that $\|Bx\|_1$ is well concentrated. For the set of little rows, L we again appeal to Talagrand to ensure that they do not deviate substantively from their mean. Combining these bounds we obtain our second main theorem

Theorem 3. *Let A be an $n \times n$ matrix over $\mathbb{R}$ and $\|A\|_2 \leq T \neq 0$. Then*

$$|\mathrm{per}(A)| \leq T^n (n+6) \exp\left[\frac{-\sqrt{n(1-h_\infty/T)}}{400}\right].$$

1.2.3 Robust Positioning Patterns

A robust positioning pattern S for windows of length n with alphabet Σ is a sequence in Σ^*. We say that the distance is the minimum over all distinct length n windows (contiguous subwords of length n) of $\Delta(w_1, w_2)$, where Δ is the usual Hamming Distance. The rate of the sequence is then defined to be $R := \log_{|\Sigma|}(|S|)$. These notions of rate and distance coincide exactly with the rate and distance of the code comprised of all length n windows of S (we allow the windows to wrap around). In this chapter we begin by describing a simple construction of a constant rate sequence over a large alphabet with constant fraction distance. This simple construction only leads to codes with rate $R \leq 1/3$, and having rate R approaching 1 seems to require some significantly

new ideas.

Let us also remark that there are several easy constructions of *low rate* sequences (with rate $< 1/2$) with constant relative window-n distance using "markers", but going to high rate introduces significant conceptual obstacles (in particular, one really needs to handle the overlap of the windows in the sequence).

Let C be the Reed-Solomon code of degree $\leq k$ polynomials over $\mathbb{F}_q$. We will take $n = q - 1$. Let us choose the sequence evaluation points for these polynomials to be $g, g^2, \ldots, g^{q-1}$, where g is a generator of $\mathbb{F}_q^*$. Thus the code is cyclic. That is every codeword can be rotated by i places by replacing $f(X)$ with $f(g^i X)$.

Partition C into the equivalence classes, where two codewords are equivalent if they are rotations of one another. Let $c_1, \ldots, c_M$ be a collection of codewords, one from each equivalence class. Let $\Sigma = \mathbb{F}_q$, let $N = M \cdot (q-1)$ and let $\sigma \in \Sigma^N$ be the sequence obtained by concatenating $c_1, c_2, \ldots, c_M$. Note that $N \approx q^{k+1}$.

We claim that σ has window-n distance at least $n - 3k$. Indeed, if we look at any length n window of σ, it looks like the concatenation of a suffix of c_i and a prefix of c_{i+1}. A moment's inspection, using the fact that every rotation of c_i is also a codeword of C shows that every n-window of this sequence looks like the splicing together of two codewords of C. Using this fact, it follows that the number of agreements between two distinct n windows is at most 3 times the maximum number of agreements between two codewords. Thus the distance between any two n-windows is at least $n - 3k$.

The above construction fails to do anything interesting if $k > n/3$. To go beyond, we will exploit our ability to carefully choose the ordering of $c_1, \ldots, c_M$. Our construction ensures that many of the windows that straddle c_i and c_{i+1} are (essentially) rotations of c_i (and in particular, they are essentially codewords of C). We do this using a Gray code. The analysis of the distance is somewhat mysterious, and takes advantage of the fact that windows now look like the gluing together of *overlapping* codewords. This leads to a bound of $(n - k)/3$ for the window-n distance, and yields a construction of efficiently decodable robust positioning patterns of any specified rate $R \in (0, 1)$, or any specified distance $\Delta \in (0, 1)$ (possibly with a large alphabet size).

Our binary construction in the one dimensional case is based on a new "augmented"

code concatenation scheme. This new scheme is based on two ideas: (1) using a low-autocorrelation sequence as a "marker", and (2) designing an inner code for the concatenation all of whose codewords are far away from all substrings of the marker.

Chapter 2

A Quantitiative Local Limit Theorem for Triangles in Random Graphs

2.1 Introduction

This chapter is concerned with the distribution of the number of triangles appearing in an Erdős-Renyi random graph $G(n,p)$ (a graph with n vertices where each edge is present independently with probability p). Recently, [16] showed a local limit theorem in this context which says that the distribution of the number of triangles approaches the discrete normal. Our main results show *quantitative* bounds, both pointwise and global, on how far the distribution of the number of triangles in a random graph can vary from a normal distribution. In particular, if $\mathcal{T}$ is the random variable corresponding to the number of triangles in $G(n,p)$ we show that for all $k \in \mathbb{Z}$ and $\epsilon > 0$,

$$\Pr[\mathcal{T} = k] = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(k-\mu)^2}{2\sigma^2}\right) + O(n^{-2.5+\epsilon})$$

where $\mu = \mathbb{E}[\mathcal{T}] = p^3\binom{n}{3}$ and $\sigma = Var(\mathcal{T})$. From this we are also able to obtain a quantitative bound on the ℓ^1 distance of $\mathcal{T}$ from a suitable discrete normal:

$$\sum_{k\in\mathbb{N}} \left| \Pr(\mathcal{T} = k) - \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\left(k - p^3\binom{n}{3}\right)^2}{2\sigma^2}\right) \right| = O(n^{-.5+\epsilon})$$

2.1.1 History

The study of subgraph counts dates back to the very beginning of random graphs, when Erdős and Renyi proved in 1960 [13] that certain subgraph counts behaved in expected ways by using the second moment method. In the 1980's there were several papers studying which subgraph counts obeyed a central limit theorem (see [26, 25, 35]). For

example, in this period a central limit theorem was shown for the triangle counting random variable $\mathcal{T}$, which stated that for any real numbers $a < b$

$$\Pr[\mathcal{T} \in [\mu + a\sigma_n, \mu + b\sigma_n]] = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-t^2/2} dt + o(1)$$

This line of questioning eventually found a complete solution in the work of Ruciński [39] who gave a characterization for when subgraph counts obeyed a central limit theorem. In 1989 there was progress made on showing central limit theorems with quantitative bounds in the work of Barbour, Karoński and Ruciński [5]. Slightly afterwards Janson and Nowicki [23] gave alternate arguments for central limit theorems using the language of U-statistics and a good basis for functions on the probability space of graphs.

If the edge probability $p \sim c/n$ for some constant c, then Erdős and Renyi [13] showed that the number of triangles in $G(n,p)$ converges to a Poisson distribution. This result was a local limit theorem, as it estimated the pointwise probabilities $\Pr[\mathcal{T} = k]$ for k constant. Further, Röllin and Ross [38] showed a local limit theorem when $p \sim cn^\alpha$ for $\alpha \in [-1, -\frac{1}{2}]$. In this regime they showed that the triangle counting distribution converges to a translated Poisson distribution (which is in turn close to a discrete Gaussian) in both the ℓ_∞ and total variation metrics.

In 2014, Gilmer and Kopparty [16] proved a local limit theorem for triangle counts for $G(n,p)$ in the regime where p is a fixed constant. In particular they proved that

$$\Pr[\mathcal{T} = k] = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(k-\mu)^2}{2\sigma_n^2}\right) \pm o(n^{-2})$$

It should be noted that this is largely a *qualitative* result, as the main term has size $\Theta(n^{-2})$ while the error term is $o(n^{-2})$. This type of result should also be contrasted with the central limit theorem given above. This theorem gives an estimate for the probability of having *exactly* k triangles or differing from the expected number of triangles by exactly 17. The central limit theorems estimate the probability of having a number of triangles in an interval of length proportional to the standard deviation.

The proof in [16] proceeded by using the characteristic function. The main step there was to show that $|\varphi(t) - \varphi_n(t)|$ is small for $t \in [-\pi\sigma_n,\ \pi\sigma_n]$, where φ represents the characteristic function of the standard normal distribution, and φ_n represents the characteristic function the triangle counting function $\mathcal{T}$.

2.1.2 Our Results

We improve the result of Gilmer and Kopparty by adding a quantitative estimate for the convergence of $\mathcal{T}$ to the normal. We strengthen their bound to give explicit distance bounds. The main result of this part of the thesis is the following local limit theorem for the number of triangles in $G(n,p)$

Theorem 2. *For any $k \in \mathbb{N}$ we have that*

$$\Pr[\mathcal{T} = k] = \frac{1}{\sqrt{2\pi}\sigma_n} e^{-\frac{\left(k - p^3\binom{n}{2}\right)^2}{2\sigma_n^2}} + O(n^{-2.5+\epsilon})$$

For $k = \mu_n + O(\sigma_n)$ this shows that $\Pr[\mathcal{T} = k]$ is within a $(1 + O(n^{-\frac{1}{2}+\epsilon}))$ multiplicative factor of $\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(k-\mu)^2}{2\sigma_n^2}\right)$, while the best known previous bound could only show a factor of $(1 + o(1))$. A polynomial factor is also the best possible bound, as even the binomial distribution of $\binom{n}{3}$ i.i.d. summands differs from the normal by a polynomial factor. As a consequence of Theorem 2 we also find a quantitative bound on the ℓ^1 distance between T and the normal.

Theorem 3.

$$\sum_{t\in\mathbb{N}} \left| \Pr(\mathcal{T} = t) - \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\left(t - p^3\binom{n}{3}\right)^2}{2\sigma^2}\right) \right| = O(n^{-0.5+\epsilon})$$

The results in [16] were not enough to imply ℓ^1 distance bounds, so this is the first result of this kind for triangle counts. Finally at the end of this chapter we highlight that these proof techniques continue to give characteristic function bounds for counting other subgraphs. While proving a local limit theorem for K_4's remains just out of reach, we can use these arguments to prove quantitative central limit theorems in this setting.

2.1.3 Organization of this Chapter

In section 2 we set up our notation and introduce some facts which will be necessary for the later sections. Section 3 contains the statements and proofs of our main results, modulo the main technical lemmas. In section 4 we examine the decomposition of $\mathcal{T}$ with respect to the p-biased Fourier basis, and in section 5 we exploit this decomposition to prove our main lemmas. Finally in section 6 we extend these arguments to a more general setting to capture larger subgraph counts.

2.2 Preliminaries and Notation

We will be working with a random variable which is defined as a graph function applied to an Erdős-Renyi random graph $G(n,p)$. We will be working in the regime where our probability p is a fixed constant, and $n \to \infty$. We will realize our probability space as drawing $\mathbf{x} \in \{0,1\}^{\binom{n}{2}}$ where each coordinate of $\mathbf{x}$ is labelled by an edge $e \in \binom{[n]}{2}$, and we have that for all edges, $\mathbf{x}_e$ is 0 with probability $1-p$ and 1 with probability p. $\binom{[n]}{2}$ refers equivalently to either the set of all pairs of distinct elements from $[n]$, or the set of possible edges of a graph with vertex set $[n]$.

Continuing our notation from the abstract, we use $\mathcal{T} : \{0,1\}^{\binom{n}{2}} \to \mathbb{N}$ to denote the triangle counting function, which returns the number of triangles in the graph with edge set given by the indicator vector $\{0,1\}^{\binom{n}{2}}$. One might note that the random variable $\mathcal{T}$ depends on both the probability p, and the size of the vertex set n in question. We will often supress the dependence on n and p, as we will be considering p to be fixed and our analysis will be done for a generic n, with limits only taken in the proof of the main theorem.

2.2.1 p-Biased Fourier Basis

To apply our analysis we use the p-biased Fourier basis for functions on this probability space. We define this as follows. For each edge $e \in \binom{[n]}{2}$ we define $\chi_e : \{0,1\}^{\binom{n}{2}} \to \mathbb{R}$ as

follows:

$$\chi_e := \chi_e(\mathbf{x}) := \frac{\mathbf{x}_e - p}{\sqrt{p(1-p)}} = \begin{cases} -\sqrt{\frac{p}{1-p}} & \text{if } \mathbf{x}_e = 0 \\ \sqrt{\frac{1-p}{p}} & \text{if } \mathbf{x}_e = 1 \end{cases}$$

This is just the transform of the bernoulli random variable $\mathbf{x}_e$ so that it has mean 0 and variance 1. Now for an arbitrary set $S \subset [n]$ we can define

$$\chi_S := \chi_S(\mathbf{x}) := \prod_{e \in S} \chi_e$$

We note that if we take our inner product of two functions $f, g : \{0,1\}^{\binom{n}{2}} \to \mathbb{R}$ to be defined by $\mathbb{E}[fg]$, then $\{\chi_S \mid S \subset [n]\}$ is an orthonormal basis (see [36] chapter 10 for more detail on this topic).

For any function $f : \{0,1\}^{\binom{n}{2}} \to \mathbb{R}$, if we define the Fourier transform $\hat{f} : \{0,1\}^{\binom{n}{2}} \to \mathbb{R}$ to be

$$\hat{f}(S) := \mathbb{E}[f(x)\chi_S(x)]$$

then by orthonormality we have that

$$f(\mathbf{x}) = \sum_{S \subset \binom{[n]}{2}} \hat{f}(S)\chi_S(\mathbf{x})$$

The Fourier Expansion will serve to reveal some of the underlying properties of the triangle counting random variable. For another work using the Fourier transform to reveal such structure, see [15].

2.2.2 Probability Terminology and Notation

In proving limit theorems, it is convenient to normalize the family of random variables to have mean 0 and variance 1, so throughout this chapter we will usually work with the related random variable $Z : \{0,1\}^{\binom{[n]}{2}} \to \mathbb{R}$

$$Z(\mathbf{x}) := Z_n(\mathbf{x}) := \frac{\mathcal{T} - \mu}{\sigma}$$

We will frequently refer to the characteristic function of Z as $\varphi_Z(t) := \mathbb{E}[e^{itZ}]$. Most of the work will be focused on studying φ_Z, and showing it is close to $e^{-t^2/2}$.

We will also throughout the chapter label the variance of $\mathcal{T}$ as $\sigma^2 := \sigma_n^2 := \mathbb{E}[\mathcal{T}^2] - \mathbb{E}[\mathcal{T}]^2$. A consequence of orthonormality gives us the following result, sometimes called Parseval's Theorem:

$$\sigma^2 := \mathbb{E}[\mathcal{T}^2] - \mathbb{E}[\mathcal{T}]^2 = \left(\sum_{S \subset \binom{[n]}{2}} \hat{\mathcal{T}}(S)^2 \right) - \hat{\mathcal{T}}(\varnothing)^2 = \sum_{S \neq \varnothing} \hat{\mathcal{T}}(S)^2 \tag{2.1}$$

2.2.3 Some Graph Notation

Let G be a graph with vertex set $[n]$ and edge set $E \subset \binom{[n]}{2}$. Given a triangle $\triangle$ with vertex set $\{v_1, v_2, v_3\} \subset [n]$ we will use the notation $e \in \triangle$ to denote that e is an edge in the triangle $\triangle$ i.e. $e \in \binom{\{v_1, v_2, v_3\}}{2}$. Additionally we will occasionally identify a triangle $\triangle$ with its edge set. That is, if we have $S \subset \binom{[n]}{2}$ and we write $S = \triangle$, that means S is the edge set of some triangle.

Also we will frequently need to refer to the case where e_1 and e_2 are two edges which are incident to a common vertex (i.e. $e_1 = (v_1, v_2)$ and $e_2 = (v_2, v_3)$). We will denote this as $e_1 \sim e_2$.

2.2.4 Notation for function restrictions

Often we will have a function $f : \{0,1\}^{\binom{n}{2}} \to \mathbb{R}$, and we will want to refer to the function obtained from f by restricting some input coordinates to have certain values. In particular assume that we have $H \subset \binom{[n]}{2}$ some fixed subset of input variables. Then for $\beta \in \{0,1\}^{H^c}$ we will define $f_\beta : \{0,1\}^H \to \mathbb{R}$ by

$$f_\beta(\alpha) = f(\alpha, \beta)$$

.

2.2.5 Ingredients for the Proof

In this section we cite some useful results from the literature. We will need the following Hypercontractivity result which bounds the probability that a low degree Boolean function deviates from its mean.

Theorem 4 ([36] Theorem 10.24)**.** *Let $f : \{0,1\}^n \to \mathbb{R}$ be a polynomial of degree k, and $\lambda := \min(p, 1-p)$. If $x \in \{0,1\}^n$ is chosen by setting each coordinate independently to be 1 with probability p and 0 with probability $1-p$ then for any $t \geq \sqrt{2e/\lambda}^k$,*

$$\Pr(|f(x)| \geq t\|f\|_2) \leq \lambda^k \exp\left(-\frac{k}{2e}\lambda t^{\frac{2}{k}}\right)$$

We will also use some of the existing bounds on the characteristic function of $\mathcal{T}$, which were derived in Gilmer-Kopparty. We slightly modify their result to have a different choice of numbers, but the proof remains unchanged.

Lemma 1 ([16] Theorem 5)**.** *Fix $\epsilon > 0$. If $\varphi_n(t)$ is the characteristic function of $Z = \frac{\mathcal{T} - p^3\binom{n}{3}}{\sigma}$, then for $|t| \in [n^{.5+\epsilon}, \pi\sigma_n]$ it holds that $|\varphi_n(t)| = O(|t|^{-50})$.*

We will frequently deal with Bernoulli random variables, and so the following bound on their characteristic function will be useful.

Lemma 2. *Let X be the mean 0 variance 1 random variable taking the values*

$$X := \begin{cases} -\sqrt{\frac{p}{1-p}} & \text{with probability } 1-p \\ \sqrt{\frac{1-p}{p}} & \text{with probability } p \end{cases}$$

Then for $|t| < \sqrt{p(1-p)}\pi$ we have that $|\mathbb{E}[e^{itX}]| < 1 - \frac{2t^2}{\pi^2}$.

Proof. Let Y be the random variable taking the value -1 with probability p and 1 with probability $1-p$. Y has variance $4p(1-p)$, and $X = \frac{Y - \mathbb{E}[Y]}{2\sqrt{p(1-p)}}$. Define $\tilde{t} := \frac{t}{2\sqrt{p(1-p)}}$. So we can compute that

$$\begin{aligned} |\mathbb{E}[e^{itX}]|^2 = \left|\mathbb{E}\left[e^{i\tilde{t}Y}\right]\right|^2 = |pe^{-i\tilde{t}} + (1-p)e^{i\tilde{t}}|^2 = \|(\cos(\tilde{t}),\ (1-2p)\sin(\tilde{t}))\|^2 \\ = 1 - 4p(1-p)\sin^2(\tilde{t}) \leq 1 - \frac{16p(1-p)}{\pi^2}\tilde{t}^2 \leq 1 - \frac{4t^2}{\pi^2} \end{aligned}$$

where we used the fact that $|\sin(\tilde{t})| \geq \frac{2|\tilde{t}|}{\pi}$ for $|\tilde{t}| \leq \frac{\pi}{2}$. Lastly noticing that $\sqrt{1-x} \leq 1 - \frac{x}{2}$ completes the proof. □

2.3 Main Results

Here we give the high level proof of our main results, deferring the proofs of the important lemmas to the next section. First, we need the following standard theorem from probability.

2.3.1 Local Limit Theorem for $\mathcal{T}$.

Theorem 5 (Fourier Inversion Formula for Lattices [14] XV.4 Theorem 4)**.** *Let X be a random variable supported in $b + h\mathbb{Z}$, and let $\varphi(t)$ be the characteristic function of X. Then for $x \in b + h\mathbb{Z}$*

$$\mathbb{P}(X = x) = \frac{h}{2\pi} \int_{-\frac{\pi}{h}}^{\frac{\pi}{h}} e^{-itx} \varphi_X(t) dt$$

As a consequence of this lemma we can turn characteristic function bounds for sequences of random variables into statements about their limiting distribution.

Lemma 3. *Let Y be the standard normal distribution which has density $\mathcal{N}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ and characteristic function $\varphi(t) = e^{-\frac{t^2}{2}}$. Let X_n be a sequence of random variables supported in the lattices $\mathcal{L}_n = b_n + h_n\mathbb{Z}$. Then*

$$|h_n \mathcal{N}(x) - \mathbb{P}(X_n = x)| \leq h_n \left(\int_{-\frac{\pi}{h_n}}^{\frac{\pi}{h_n}} |\varphi(t) - \varphi_n(t)| dt + \frac{1}{\sqrt{2\pi} t} e^{-\frac{t^2}{2}} \right)$$

Proof. By the general (that is, not the lattice version above) inversion principle for characteristic functions, we have $\mathcal{N}(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \varphi(t) dt$. By the above theorem we

have that $\mathbb{P}(X_n = x) = \frac{h_n}{2\pi}\int_{-\frac{\pi}{h_n}}^{\frac{\pi}{h_n}} e^{-itx}\varphi_n(t)dt$. So we have that

$$\begin{aligned}
|h_n\mathcal{N}(x) - \mathbb{P}(X_n = x)| &= \left|\frac{h_n}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\varphi(t)dt - \frac{h_n}{2\pi}\int_{-\frac{\pi}{h_n}}^{\frac{\pi}{h_n}} e^{-itx}\varphi_n(t)dt\right| \\
&\leq \left|\frac{h_n}{2\pi}\int_{-\frac{\pi}{h_n}}^{\frac{\pi}{h_n}} e^{-itx}(\varphi(t) - \varphi_n(t))dt\right| + \left|\frac{h_n}{2\pi}\int_{|t|>\frac{\pi}{h_n}} e^{-itx}\varphi(t)dt\right| \\
&\leq h_n\left(\int_{-\frac{\pi}{h_n}}^{\frac{\pi}{h_n}} |\varphi(t) - \varphi_n(t)|dt + \frac{1}{\sqrt{2\pi}t}e^{-\frac{t^2}{2}}\right) \qquad \square
\end{aligned}$$

The main calculation of this chapter is the following theorem, whose proof is given in Section 5.

Theorem 6. *Fix $\epsilon > 0$. Let $Z := \frac{\mathcal{T} - p^3\binom{n}{3}}{\sigma}$, and $\varphi_Z(t)$ be the characteristic function of Z. Then*

$$\int_{-\pi\sigma_n}^{\pi\sigma_n}\left|\varphi_Z(t) - e^{\frac{-t^2}{2}}\right| = O_\epsilon(n^{-.5+\epsilon})$$

We can now prove our main claim, Theorem 2, as it is elementarily equivalent to the following corollary.

Corollary 1. *Let $\mathcal{L}_n := \frac{1}{\sigma_n}(\mathbb{Z} - p^3\binom{n}{3})$. Then for any $x \in \mathcal{L}_n$ we have that*

$$\left|\mathbb{P}(Z_n = x) - \frac{\mathcal{N}(x)}{\sigma_n}\right| = O_\epsilon\left(\frac{1}{n^{2.5-\epsilon}}\right)$$

Proof. Apply Lemma 3 to Z, combined with the estimate for the characteristic function of Z given by Theorem 7. $\square$

2.3.2 Bounds on the Statistical Distance of $\mathcal{T}$ from Normal

We give a lemma which will allow us to turn the L^∞ bounds we obtain into bounds on the statistical difference of $\mathcal{T}$ from the normal.

Lemma 4. *Let $\mathcal{N}$ be the density of the standard normal and $\varphi(t)$ its characteristic function. Let X_n be a sequence of random variables supported in the lattice $\mathcal{L}_n := b_n + h_n\mathbb{Z}$, and with characteristic functions φ_n. Assume that the following hold:*

1. $\sup_{x\in\mathcal{L}_n} |\Pr(X_n = x) - h_n\mathcal{N}(x)| < \delta_n h_n$

2. $\Pr(|X_n| > A) \leq \epsilon_n$

Then $\sum_{x \in \mathcal{L}_n} |\Pr(X_n = x) - \mathcal{N}(x)| \leq 2A\delta_n + \epsilon_n + \frac{h_n}{\sqrt{2\pi}A} e^{\frac{-A^2}{2}}$.

Proof. We directly compute that:

$$\begin{aligned}
&\sum_{x \in \mathcal{L}_n} |\Pr(X_n = x) - h_n \mathcal{N}(x)| \\
&\qquad \leq \sum_{\substack{x \in \mathcal{L}_n \\ |x| < A}} |\Pr(X_n = x) - h_n \mathcal{N}(x)| + \sum_{\substack{x \in \mathcal{L}_n \\ |x| \geq A}} |\Pr(X_n = x) - h_n \mathcal{N}(x)| \\
&\qquad \leq \sum_{\substack{x \in \mathcal{L}_n \\ |x| < A}} |\Pr(X_n = x) - h_n \mathcal{N}(x)| + \Pr(X_n \geq A) + h_n \int_{|x| > A-1} \mathcal{N}(x) dx \\
&\qquad \leq \frac{2A}{h_n} \delta_n h_n + \epsilon_n + \frac{h_n}{\sqrt{2\pi}A} e^{\frac{-A^2}{2}}
\end{aligned}$$

□

We can now use this to give a proof that the statistical distance between triangle counts and discrete normal variable is asymptotically small. We will pick $A := \log^2(n)$. By an application of hypercontractivity (Theorem 4) we find that

$$\Pr(|Z_n| > \log^2(n)) \leq e^{-\Omega_p(\log^2(n))} = n^{-\Omega_p(\log(n))} = o\big(n^{-.5}\big)$$

This bounds the ϵ_n term in the above theorem. We also have from Corollary 1 that $\sup_{x \in \mathcal{L}_n} |\Pr(X_n = x) - h_n \mathcal{N}(x)| = O_\epsilon(n^{-2.5+\epsilon})$. Combining this with the calculation that $\sigma_n = \Theta(n^2)$ we obtain the following corollary, which is equivalent to Theorem 3:

Corollary 2. *Fix* $\epsilon > 0$. *Let* $\mathcal{L}_n := \frac{1}{\sigma_n}(\mathbb{Z} - p^3\binom{n}{3})$. *Then*

$$\sum_{x \in \mathcal{L}_n} \left| \Pr(Z = x) - \frac{1}{\sigma} \mathcal{N}(x) \right| = O_\epsilon(n^{-.5+\epsilon})$$

Proof. In the above Lemma for $X_n = \mathcal{T}_n$ we have that $h_n = \sigma_n$. We may take $\delta_n := n^{-.5+\frac{\epsilon}{2}}$ by Corollary 1, and we may fix $A = \log^2(n)$ as above. Then as argued above $\epsilon_n = O(n^{-.5})$ while $e^{-\frac{A^2}{2}}$ is minuscule. Plugging these choices into the bound given by Lemma 4 gives the desired estimate. □

2.4 Properties of the Triangle Counting Function

In this section we express the triangle counting function in the p-biased Fourier basis, and compute some basic properties.

Given a particular triangle $\triangle$ with vertex set v_1, v_2, v_3, we will use the notation $e \in \triangle$ to denote that e is an edge in the given triangle $\triangle$. The indicator function of this triangle's presence given the graph with edge indicator vector $\mathbf{x} \in \{0,1\}^{\binom{n}{2}}$ is given by

$$\begin{aligned} 1_{\triangle}(\mathbf{x}) &= \prod_{e \in \triangle} \mathbf{x}_e = \prod_{e \in \triangle} \left(\sqrt{p(1-p)}\chi_e(\mathbf{x}) + p \right) \\ &= p^3 + p^2\sqrt{p(1-p)} \sum_{e \in \triangle} \chi_e + p^2(1-p) \sum_{e_1 \neq e_2 \in \triangle} \chi_{\{e_1,e_2\}} + (p(1-p))^{\frac{3}{2}} \chi_{\{e_1,e_2,e_3\}} \end{aligned}$$

Given two edges every edge appears in $n-2$ triangles, and each pair of edges appear in exactly 1 triangle iff they are incident to a common vertex (an event which we denote by $e_1 \sim e_2$) we find by summing over all possible triangles that

$$\mathcal{T} = p^3 \binom{n}{3} + (n-2) \sum_{e \in \binom{[n]}{2}} p^2\sqrt{(p)(1-p)}\chi_e + \sum_{e_1 \sim e_2} p^2(1-p)\chi_{\{e_1,e_2\}} + \sum_{\triangle} p^{\frac{3}{2}}(1-p)^{\frac{3}{2}}\chi_{\triangle}$$

Restated we have found the Fourier Transform of $\mathcal{T}$ and it has the form

$$\hat{\mathcal{T}}(S) = \begin{cases} p^3\binom{n}{3} & \text{if } S = \varnothing \\ (n-2)p^2\sqrt{(p)(1-p)} & \text{if } |S| = 1 \\ p^2(1-p) & \text{if } S = \{e_1, e_2\},\ e_1 \sim e_2 \\ p^{\frac{3}{2}}(1-p)^{\frac{3}{2}} & \text{if } S = \triangle \\ 0 & \text{else} \end{cases} \tag{2.2}$$

We compute the variance of $\mathcal{T}$ using the orthonormality of our basis (or Parseval)

to be

$$\begin{aligned}\sigma^2 := \mathbb{E}[\mathcal{T}^2] - \mathbb{E}[\mathcal{T}]^2 &= \sum_{\substack{S \subset \binom{[n]}{2} \\ S \neq \varnothing}} \hat{T}^2(S) \\ &= \sum_{e \in \binom{[n]}{2}} \left((n-2)p^2\sqrt{(p)(1-p)}\right)^2 + \sum_{e_1 \sim e_2} \left(p^2(1-p)\right)^2 + \sum_{\triangle} \left(p^{\frac{3}{2}}(1-p)^{\frac{3}{2}}\right)^2 \\ &= \binom{n}{2}(n-2)^2 p^5(1-p) + 3\binom{n}{3}p^4(1-p)^2 + \binom{n}{3}p^3(1-p)^3 \\ &= \Theta(n^4)\end{aligned} \tag{2.3}$$

It should be noted that asymptotically we have $\sigma \sim \frac{p^{5/2}(1-p)^{1/2}}{2} n^2$. Also it is significant that the main term in the above expansion of σ^2 comes entirely from terms of the form χ_e, for a singleton set containing one edge e. This shows that $\mathcal{T}$ has Fourier spectrum highly concentrated on degree 1. In particular, if we define $W^1 := \sum_e \hat{\mathcal{T}}^2(e)$ then $\sigma^2 = W^1(1 + O(\frac{1}{n}))$.

Recall that we defined $Z := \frac{\mathcal{T}-\mu}{\sigma} = \frac{\mathcal{T}-p^3\binom{n}{3}}{\sigma}$. By construction Z has mean 0 and variance 1. The Fourier decomposition of Z is just a normalized version of $\mathcal{T}$. In particular $\hat{Z}(S) = \frac{\hat{\mathcal{T}}(S)}{\sigma}$ if $S \neq \varnothing$, and $\hat{Z}(\varnothing) = 0$.

2.5 Estimating the Characteristic Function of Z

2.5.1 Main Results of the Section

In this section we prove the following bound

Theorem 7. *Let $Z := \frac{\mathcal{T}-p^3\binom{n}{3}}{\sigma}$, and $\varphi_Z(t)$ be the characteristic function of Z. Then for any $\epsilon > 0$*

$$\int_{-\pi\sigma_n}^{\pi\sigma_n} \left| \varphi_Z(t) - e^{\frac{-t^2}{2}} \right| = O_\epsilon(n^{.5-\epsilon})$$

The work is done over 3 sections, each corresponding to different sizes of t. In Section 2.5.2 we prove the following bound which, while true for all t, is most useful for smaller values of t

Lemma 5.

$$\left|\varphi_Z(t) - e^{-\frac{t^2}{2}}\right| = O\left(\frac{t^3 e^{-\frac{t^2}{3}}}{n} + \frac{t}{\sqrt{n}}\right)$$

Subsequently in Section 2.5.3 we prove the result for "mid-sized" t that

Lemma 6. *Fix $0 < \epsilon < 1$. Then*

$$\int_{n^\epsilon}^{n^{\frac{1+\epsilon}{2}}} |\varphi_Z(t)|dt \leq O_\epsilon(n^{-.5+\epsilon})$$

Lastly for $|t| \geq n^{\frac{1+\epsilon}{2}}$ we simply cite Lemma 1. Combining all these results immediately gives Thoerem 7. For completeness we give the proof.

Proof of Theorem 7.

$$\begin{aligned}
\int_{-\pi\sigma_n}^{\pi\sigma_n}\left|\varphi_Z(t) - e^{\frac{-t^2}{2}}\right| &= \int_{|t|<n^{0.05}}\left|\varphi_Z(t) - e^{\frac{-t^2}{2}}\right| + \int_{n^{0.05}<|t|<n^{.5+\frac{\epsilon}{10}}}\left|\varphi_Z(t) - e^{\frac{-t^2}{2}}\right| \\
&\quad + \int_{n^{.5+\frac{\epsilon}{10}}<|t|<\pi\sigma_n}\left|\varphi_Z(t) - e^{\frac{-t^2}{2}}\right| \\
&\leq \int_{|t|<n^\epsilon} O\left(\frac{t^3 e^{-\frac{t^2}{3}}}{n} + \frac{t}{\sqrt{n}}\right)dt + O_{p,\epsilon}(n^{-.5+\epsilon}) + O(n^{-50}) \\
&\quad + 2\left|\int_{n^\epsilon}^{\pi\sigma_n} e^{-\frac{t^2}{2}}dt\right| \\
&= O_\epsilon(n^{-0.5+2\epsilon})
\end{aligned}$$

□

2.5.2 Bounds for small t

In this section we prove the following result

Lemma 5.

$$\left|\varphi_Z(t) - e^{-\frac{t^2}{2}}\right| = O\left(\frac{t^3 e^{-\frac{t^2}{3}}}{n} + \frac{t}{\sqrt{n}}\right)$$

This shows that the characteristic function of Z is very close to that of $N(0,1)$ for small t. In that regard this result is a generalization of a central limit theorem for T,

which is equivalent to the pointwise convergence of $\varphi_Z(t)$ to $e^{-t^2/2}$.

Proof. We can decompose Z into two parts, the dominant weight-1 part X, and a smaller term corresponding to Fourier coefficients of weight ≥ 2. In particular let $Q := \sqrt{\frac{1}{\binom{n}{2}}}$. Then we define

$$X := \sum_{e \in \binom{[n]}{2}} Q\chi_e \qquad\qquad Y := \sum_{e \in \binom{[n]}{2}} (\hat{Z}(e) - Q)\chi_e + \sum_{|S| \geq 2} \hat{Z}(S)\chi_S$$

First we examine X. It is the mean 0 variance 1 sum of independent random variables, and so by Berry-Esseen (see Petrov [37], Chapter V lemma 1) we know that if

$$L_n := \binom{n}{2} \mathbb{E}[|Q\chi_e|^3] = \frac{p^2 + (1-p)^2}{\sqrt{\binom{n}{2} p(1-p)}} = \Theta_p(1/n)$$

then for $t \leq \frac{1}{4L_n}$ we have that

$$\left| \mathbb{E}[e^{itX}] - e^{-\frac{t^2}{2}} \right| \leq 16 L_n |t|^3 e^{\frac{-t^2}{3}} \tag{2.4}$$

Now we turn our attention to Y. Y is best thought of as an error term. It is where the dependence of our random variable Z lives, and it will be always very small. In particular, using Cauchy-Schwarz and the orthogonality of our basis we obtain

$$\mathbb{E}\,|Y| \leq \sqrt{\mathbb{E}\,|Y|^2} = var(Y) = \sum_e (\hat{Z}(e) - Q)^2 + \sum_{|S| \geq 2} \hat{Z}^2(S)$$

We know from prior calculations that

$$\sum_{|S| \geq 2} \hat{Z}^2(S) = \frac{3\binom{n}{3} p^4 (1-p)^2 + \binom{n}{3} p^3 (1-p)^3}{\sigma^2} = O\left(\frac{1}{n}\right)$$

Further we can estimate

$$\binom{n}{2} \sigma^2 \hat{Z}^2(e) - \sigma^2 = \binom{n}{2} \hat{T}^2(e) - \sigma^2 = O(n^3) \implies \hat{Z}^2(e) - \frac{1}{\binom{n}{2}} = O(n^{-3})$$

Therefore using the fact that $(x - y) = (x^2 - y^2)/(x + y)$ coupled with the observation

that $\hat{Z}(e) + Q = \Theta\left(\frac{1}{n}\right)$, we find that

$$|\hat{Z}(e) - Q| \leq \left| \frac{\hat{Z}^2(e) - \frac{1}{\binom{n}{2}}}{\hat{Z}(e) + Q} \right| = O\left(\frac{1}{n^2}\right)$$

So as a result we can conclude that $var(Y) = O(1/n)$ and so $\mathbb{E}[|Y|] = O(\frac{1}{\sqrt{n}})$. Now we are ready for our characteristic function bound for Z. If $|t| \leq \frac{1}{4L_n} = \Theta_p(n)$ then combining the above with equation 2.10 yields.

$$\begin{aligned}
\left|\varphi_Z(t) - e^{-\frac{t^2}{2}}\right| &= \left|\mathbb{E}\left[e^{itZ}\right] - e^{-\frac{t^2}{2}}\right| = \left|\mathbb{E}\left[e^{it(X+Y)}\right] - e^{-\frac{t^2}{2}}\right| \\
&\leq \left|\mathbb{E}\left[e^{itX}\right] - e^{-\frac{t^2}{2}}\right| + \left|\mathbb{E}\left[e^{itX+Y}\right] - \mathbb{E}\, e^{itX}\right| \leq 16L_n|t|^3 e^{\frac{-t^2}{3}} + \mathbb{E}\,|tY| \\
&= O\left(\frac{t^3 e^{-\frac{t^2}{3}}}{n} + \frac{t}{\sqrt{n}}\right)
\end{aligned}$$

The last inequality comes from simply applying the mean value theorem to the function e^{itx}. The first term in the error is dominated for any choice of t, and so we can simplify the error to $|\varphi_Z(t) - e^{-t^2/2}| = O(tn^{-1/2})$. □

2.5.3 Bounds for slightly larger t

Here we perform the same operation as above, except we first reveal a fraction of the edges. The intuition behind this is that revealing a q fraction of the edges will reduce the number of edge variables over which we take our expectation by q, but it will reduce the influence of larger sets by even more, namely by $q^{|S|} \geq q^2$. Thus in the above estimate when we decompose Z into $X + Y$ we will find that Y will be significantly smaller, allowing us to get a better estimate.

For any natural number k, we can take H to be a k-regular bipartite graph on n vertices. Then it makes sense to talk about the restriction of Z to the variables in H. That is we are *revealing* the edges in H^c to be some vector $\beta \in \{0,1\}^{H^c}$, and consider the function $Z_\beta : \{0,1\}^H \to \mathbb{R}$ given by $Z_\beta(\alpha) = Z(\alpha, \beta)$. First we note that by the

law of total probability we have that

$$\mathbb{E}[e^{itZ}] = \mathop{\mathbb{E}}_{\beta\in\{0,1\}^{H^c}} \mathop{\mathbb{E}}_{\alpha\in\{0,1\}^{H}} [e^{itZ_\beta(\alpha)}]$$

So now we turn our attention to examining the form Z_β takes for a typical restriction β. First let us consider what happens to a generic basis function χ_S (for a general consideration of how restriction interacts with Fourier bases , particularly in the case of $p = \frac{1}{2}$, see [36] Chapter 3.3). If we split S as $S = S_H \cup S_{H^c}$ where $S_H \subset H$ and $S_H^c \subset H^c$ then

$$(\chi_S)_\beta(\mathbf{x}) = \chi_{S_{H^c}}(\beta)\chi_{S_H}(\mathbf{x})$$

So we can use this to compute the Fourier transform of $Z_\beta : \{0,1\}^H \to \mathbb{R}$. For an arbitrary $S \subset H$ we will have that

$$\widehat{Z_\beta}(S) = \sum_{T\subset H^c} \chi_T(\beta)\hat{Z}(S \cup T) \tag{2.5}$$

If we fix S, and think of β as an input, then $\widehat{Z_\beta}(S)$ can be viewed as a function of β, $\widehat{Z_\beta}(S) : \{0,1\}^{H^c} \to \mathbb{R}$.

Claim 1. *Let A be the event (over the space of revelations $\beta \in \{0,1\}^{H^c}$) that for* every *edge $e \in H$ we have that*

$$|\widehat{Z_\beta}(e) - \hat{Z}(e)| < \frac{\sqrt{3}n^{0.6}}{\sigma}$$

Let $\lambda := \min(p, 1-p)$. Then $\Pr(A) \geq 1 - n^2\lambda^2 e^{-\lambda n^{.01}}$.

Claim 2. *Assume $\beta \in A$. Then for $t \leq \sigma\pi\sqrt{p(1-p)}/2n = \Theta_p(n)$*

$$|\mathop{\mathbb{E}}_{\alpha\subset H}[e^{itZ_\beta(\alpha)}]| \leq \exp\left(-\frac{kt^2n^3}{4\pi^2\sigma^2}\right) + \frac{4|t|n\binom{k}{2}}{\sigma^2}$$

Assuming these two claims we can prove the main result for this subsection.

Lemma 6. *Fix $0 < \epsilon < 1$. Then*

$$\int_{n^\epsilon}^{n^{\frac{1+\epsilon}{2}}} |\varphi_Z(t)|dt \leq O_\epsilon(n^{-.5+\epsilon})$$

Proof. Let A, as in Claim 4, be the event that for all $e \in H$ we have that $|\widehat{Z_\beta}(e) - \hat{Z}(e)| < \frac{\sqrt{3}n^{0.6}}{\sigma}$. We can break up $\{0,1\}^{H^c}$ into A and A^c and estimate

$$|\varphi_Z(t)| := \mathop{\mathbb{E}}_{(\alpha,\beta)\in 2^{\binom{n}{2}}}[e^{itZ(\alpha,\beta)}] \leq \mathop{\mathbb{E}}_{\beta\subset H^c}|\mathop{\mathbb{E}}_{\alpha\subset H}[e^{itZ_\beta(\alpha)}]| \leq \Pr(A) + (1-\Pr(A))\mathop{\mathbb{E}}_{\beta\in A^c}\left|\mathop{\mathbb{E}}_{\alpha}[e^{itZ_\beta}]\right|$$

Now combining Claims 4 and 2 we find that

$$\Pr(A^c) + (Pr(A))\mathop{\mathbb{E}}_{\beta\in A}\left|\mathop{\mathbb{E}}_{\alpha}[e^{itZ_\beta}]\right| \leq \lambda^2 n^2 e^{-\frac{\lambda}{e}n^{0.1}} + \exp\left(-\frac{kt^2n^3}{4\pi^2\sigma^2}\right) + \frac{2k|t|\sqrt{n}}{\sigma}$$

We may choose k to be an integer of size $n\lceil|t|^{-2+\epsilon}\rceil$ (which may be done for $0 < |t| \leq n^{\frac{1+\epsilon}{2}}$). Recalling that $\sigma = \Theta(n^2)$ we find that

$$|\varphi_Z(t)| = O\left(n^2 e^{-\frac{\lambda}{e}n^{0.1}} + \exp\left(-\frac{-t^\epsilon n^4}{4\pi^2\sigma^2}\right) + \frac{1}{|t|^{1-\epsilon}\sqrt{n}}\right)$$

Using this we may make the following estimate

$$\int_{n^\epsilon}^{n^{\frac{1+\epsilon}{2}}} |\varphi_Z(t)|dt \leq O\left(n^{2+\frac{1+\epsilon}{2}}e^{-\frac{\lambda}{e}n^{0.1}} + n^{\frac{1+\epsilon}{2}}\exp(-n^\epsilon) + \left[\frac{1}{\epsilon}t^\epsilon n^{-.5}\right]_{n^\epsilon}^{n^{\frac{1+\epsilon}{2}}}\right) = O_\epsilon\left(n^{-.5+\epsilon}\right)$$

□

Proof Of Claims 4 and 2

Claim 4. *Let A be the event (over the space of revelations $\beta \in \{0,1\}^{H^c}$) that for* every *edge $e \in H$ we have that*

$$|\widehat{Z_\beta}(e) - \hat{Z}(e)| < \frac{\sqrt{3}n^{0.6}}{\sigma}$$

Let $\lambda := \min(p, 1-p)$. Then $\Pr(A) \geq 1 - n^2\lambda^2 e^{-\lambda n^{.01}}$.

We prove Claim 4 by noting that the formula for $\widehat{Z_\beta}(S)$ (a coefficient in the polynomial Z_β) is *itself* a low degree polynomial, and therefore may be shown to have tight concentration by Theorem 4.

Proof Of Claim 4. Recall equation 2.5 which states that

$$\widehat{Z_\beta}(e) = \sum_{T \subset H^c} \hat{Z}(e \cup T)\chi_T(\beta)$$

$\widehat{Z_\beta}(e) : \{0,1\}^{H^c} \to \mathbb{R}$ is a polynomial (in the functions χ_e), and we can began by estimating its coefficients. First we see that

$$\mathbb{E}[\widehat{Z_\beta}(e)] = \widehat{\widehat{Z_\beta}(e)}(\varnothing) = \hat{Z}(e)$$

Also for any $T \subset \{0,1\}^{H^c}$ we know that $\hat{Z}(e \cup T) \neq 0$ iff e and T are in a common triangle. There are at most $3(n-2)$ choices of T (corresponding to completing the $n-2$ triangles containing the edge e). Therefore Combining this with the fact that $\hat{Z}(S') \leq \sigma^{-1}$ for all sets of size $|S'| \geq 2$ we find that

$$var_\beta(\widehat{Z_\beta}(e)) = \sum_{\substack{T \subset H^c \\ T \neq \varnothing}} \hat{Z}(e \cup T)^2 \leq \frac{3(n-2)}{\sigma^2}$$

Since $\widehat{Z_\beta}(e)$ has degree 2, an application of Theorem 4 gives us that for any $e \in H$ if we set $\lambda = \min(p, 1-p)$ then

$$\Pr\left[\left|\widehat{Z_\beta}(e) - \hat{Z}(e)\right| \geq \frac{\sqrt{3}n^{0.6}}{\sigma}\right] < \lambda^2 \exp\left(-\frac{\lambda n^{0.1}}{e}\right)$$

Applying a union bound over all edges in H completes the proof. □

Claim 2 is concerned with estimating $|\,\mathbb{E}[e^{itZ_\beta}]|$, given that β is a typical, well behaved revelation. When β is well behaved Z_β will be dominated by a sum of independent monomials, and so the proof proceeds in a manner very similar to the arguments in Section 2.5.2.

Claim 2. *Recall A is the event in 2^{H^c} such that for all edges $e \in H$ we have $\left|\widehat{Z_\beta}(e) - \hat{Z}(e)\right| \leq \sqrt{3}n^{\cdot 6}$ (that is, the set of all revelations of the edges of H^c which are well behaved).*

Assume $\beta \in A$. Then for $t \leq \sigma\pi\sqrt{p(1-p)}/2n = \Theta_p(n)$

$$|\mathop{\mathbb{E}}_{\alpha \subset H}[e^{itZ_\beta(\alpha)}]| \leq \exp\left(-\frac{kt^2n^3}{4\pi^2\sigma^2}\right) + \frac{2k|t|\sqrt{n}}{\sigma}$$

Proof of Claim 2. Assume that $\beta \in A$. Let X and Y be

$$X := \sum_{e \in \binom{[n]}{2}} \widehat{Z_\beta}(e)\chi_e \qquad\qquad Y := \sum_{|S| \geq 2} \widehat{Z_\beta}(S)\chi_S$$

then $Z_\beta = X + Y$, and we will be able to obtain bounds similar to our previous ones. In particular X is the sum of indpendent random variables so if for each e we define $Q_e := \widehat{Z_\beta}(e)$ then we will have because of our assumptions that $\frac{n}{2\sigma} \leq \hat{Z}(e) - \frac{\sqrt{3}n^{\cdot 6}}{\sigma} \leq Q_e \leq \frac{2n}{\sigma}$

because $\widehat{Z_\beta}(e) \leq \frac{2n}{\sigma}$ and $t \leq \frac{\sigma\pi\sqrt{p(1-p)}}{2n}$ we can use Lemma 2 to show that

$$|\,\mathbb{E}[e^{it\widehat{Z_\beta}(e)\chi_e}]| \leq 1 - \frac{t^2n^2}{2\pi^2\sigma^2} \leq \exp\left(-\frac{t^2n^2}{2\pi^2\sigma^2}\right)$$

So now we find that

$$\begin{aligned}\mathbb{E}[e^{itX}] &= \prod_{e \in H} \mathbb{E}[\exp\left(it\widehat{Z_\beta}(e)\chi_e\right)] \leq \exp\left(-\sum_{e \in H}(\widehat{tZ_\beta(e)})^2\right) \leq \exp\left(\sum_{e \in H} -\frac{t^2n^2}{\pi^2\sigma^2}\right) \\ &= \exp\left(-\frac{kt^2n^3}{4\pi^2\sigma^2}\right)\end{aligned}$$

Now we turn our attention to Y. If $|S| = 2$ with $S = \{e_1, e_2\}$ then $\hat{Z}(S)$ is 0 unless $e_1 \sim e_2$, and therefore e_1, e_2 lie in a common triangle $\triangle = \{e_1, e_2, e_3\}$. However this is the only triangle containing S, and so we can quickly compute using equation 2.5, and the fact for $|S| \geq 2$ we have $|\hat{Z}(S)| \leq \frac{1}{\sigma}$ (see equation 2.2 and normalize to obtain Z) that

$$\widehat{Z_\beta}(S) = \sum_{T \subset H^c} \chi_T(\beta)\hat{Z}(S \cup T) = \chi_\varnothing(\beta)\hat{Z}(S) + \chi_{e_3}(\beta)\hat{Z}(\triangle) \leq \frac{2}{\sigma}$$

So we can compute, again using Cauchy Schwartz and the fact that H is k-regular that

$$\mathbb{E}[|Y|]^2 \leq \mathbb{E}[|Y|^2] = \sum_{\substack{e_1 \sim e_2 \\ e_1, e_2 \in H}} \widehat{Z_\beta}^2(S) \leq n\binom{k}{2}\frac{4}{\sigma^2}$$

Combining this information, we compute that

$$\begin{aligned}\left|\mathop{\mathbb{E}}_{\alpha \in 2^H}[e^{itZ_\beta(\alpha)}]\right| = \left|\mathop{\mathbb{E}}_{\alpha}[e^{it(X+Y)}]\right| &\leq \left|\mathbb{E}[e^{itX} + |tY|]\right| \\ &\leq \exp\left(-\frac{kt^2n^3}{2\pi^2\sigma^2}\right) + |t|\sqrt{\binom{k}{2}n\frac{(2)^2}{\sigma^2}}\end{aligned}$$

□

2.6 General Subgraph Counts in $G(n,p)$

In this section we take the arguments we have used so far in this chapter and extend them to counting subgraphs other than triangles. We will be able to give good characteristic function bounds for the corresponding random variables; however these results as of yet do not yield any local limit theorems for any graphs on more than 3 vertices. We will, however, be able to give a new proof of quantitative central limit theorems for subgraph counts in $G(n,p)$. Section 2.6.1 will introduce necessary notation and definitions. Section 2.6.2 will contain the main results of this section. The remaining sections will cover the properties of graph statistics and then the proofs of the theorems.

2.6.1 Definitions and Graph Statistics

Falling factorials will frequently appear in our analysis, and we will use the following notation:

Definition 1. Let $n, k \in \mathbb{N}$. We define $(n)\downarrow_k := \prod_{i=0}^{k-1}(n-i)$. For the case $k = 0$ we set $(n)\downarrow_0 = 1$.

Throughout this section we will be working with functions defined on graphs. To capture subgraph counts we will need two graphs: our random graph G on a large

growing vertex set of size n, and a second graph Γ on vertex sets of a fixed size k that will define the subgraphs we are interested in counting.

Definition 2. Let $\mathcal{S}_G := \mathcal{S}_G(n,k)$ denote the set of all labeled (with vertices distinguishable by their origin in $[n]$, and given a labeling in $[k]$) induced subgraphs of the graph G with k vertices. It will also be useful to denote this as the set of injections of $\psi : [k] \hookrightarrow [n]$, with the map extended to edges in the obvious way.

Let's denote the edge set in the big graph to be $E = \binom{[n]}{2}$ and the edge set in the small graph to be $D := \binom{[k]}{2}$. Here we will give a standard notation to a slight generalization of subgraph counts, which we will call graph statistics, and the rest of this section will be concerned with analyzing such functions

Definition 3. Fix a graph function $f : 2^{\binom{[k]}{2}} \to \mathbb{R}$. For any $n \geq k$ we can define the *graph statistic* $F_f : 2^{\binom{[n]}{2}} \to \mathbb{R}$ (typically denoted simply as F) for f to be

$$F(G) := F_f(G) := \sum_{\Gamma \in \mathcal{S}_G} f(\Gamma)$$

A function $F(G)$ defined this way sums f as applied to all ordered subgraphs of size k in G. In particular if f is the indicator of a fixed graph H (induced or otherwise), then the graph statistic $F(G)$ counts the number of copies of this graph inside G. To help our study of the properties of F, it will be useful to have some notation aggregating information about the base function f.

Definition 4. For a set $T \subset E$ let

$$h_T := \sum_{\phi : \mathrm{supp}(T) \hookrightarrow D} \hat{f}(\psi(T))$$

where the summation is over all injections of $\mathrm{supp}(T)$ into $D = \binom{[k]}{2}$.

Note the arrow here is reversed from the maps in the definition of $\mathcal{S}_G$. Also, h_T depends only on the isomorphism class of T, and importantly does not change as the

parameter n changes. It will also be handy to define the largest such coefficient to be

$$h_* := \max_T |h_T| \tag{2.6}$$

When analyzing the low weight spectral concentration of F, a better measure for estimating $\hat{F}(S)$ than simply $|S|$, will be the number of vertices incident to edges in S. We call this set of vertices the support of S.

Definition 5. Given a set of edges S, define $\text{supp}(S) := \cup_{e \in S} e$, the set of all vertices incident to edges in S.

Our main theorems in the next section will be aimed at bounds on the characteristics of subgraph counting random variables. However, our arguments will work in the slightly more general setting of graph statistics which are *edge dominated*.

Definition 6. If f has the property that $h_e = \sum_{e \in \binom{[k]}{2}} \hat{f}(e) \neq 0$, then we say that F is *edge dominated*.

A few examples to illustrate these definitions are in order.

Example 2.6.1. Consider $|\Gamma| = 3$, and so $f : 2^{\binom{[3]}{2}} \to \mathbb{R}$ defined by

$$f(\mathbf{x}) = \mathbf{x}_{12}\mathbf{x}_{23} = \Big(\sqrt{p(1-p)}\chi_{(12)}(\mathbf{x}) + p\Big)\Big(\sqrt{p(1-p)}\chi_{(23)}(\mathbf{x}) + p\Big)$$

Then f is the indicator of whether the input graph Γ contains the length 2 path from 1 to 3, but puts no condition on the edge between vertices 2 and 3. F_f will count all ordered paths of length 2 in the graph and will be edge dominated for any p (as can be seen by expanding out the above product).

Example 2.6.2. Again take $|\Gamma| = 3$, and so $f : 2^{\binom{[3]}{2}} \to \mathbb{R}$ defined by

$$\begin{aligned} f(\mathbf{x}) &= \mathbf{x}_{12}\mathbf{x}_{23}(1 - x_{13}) \\ &= \Big(\sqrt{p(1-p)}\chi_{(12)}(\mathbf{x}) + p\Big)\Big(\sqrt{p(1-p)}\chi_{(23)}(\mathbf{x}) + p\Big) \cdot \Big(\sqrt{p(1-p)}\chi_{(13)}(\mathbf{x}) + p - 1\Big) \end{aligned}$$

Then f is the indicator of whether the input graph Γ is exactly the length 2 path from 1 to 3, with edge (23) excluded. F_f will count all induced copies of P_2 in the graph.

We can compute h_e to be

$$h_e = 2\Big(p(p-1)\sqrt{p(1-p)}\Big) + \sqrt{p(1-p)}p^2 = p^{3/2}(1-p)^{1/2}(3p-2)$$

So $h_e \neq 0$ and F is edge dominated so long as $p \neq \frac{2}{3}$. Note this condition is quite logical, as $\frac{2}{3}$ is the edge density of P_2, and intuitively it is at this point that observing an edge in our random graph gives us the least information about how many induced copies of P_2 we should expect.

In general, these above examples extend to the case of all homomorphic or induced subgraph counts. In particular, if f checks for noninduced copies of a fixed grah H, then F_f will always be edge dominated and obey the characteristic function bounds of Theorems 8, 9, and 10 and the Central Limit Theorem of Theorem 11. Meanwhile if f counts induced copies of H, then F_f will still be edge dominated *so long as* $p \neq \frac{|E(H)|}{\binom{k}{2}}$, that is p is not exactly the edge density of H.

2.6.2 Characteristic Function Bounds for Subgraph Counts and an Application

Our first main result will be showing that the characteristic function of a function/random variable defined by applying an edge dominated graph statistic F to $G(n,p)$ is close to that of the Gaussian.

Theorem 8. *Let* $F : \binom{[n]}{2} \to \mathbb{R}$ *be an edge dominated graph statistic defined from* $f : \binom{[k]}{2} \to \mathbb{R}$ *be as in definition 3. Let* $Z := \frac{F - \mathbb{E}F}{\sigma}$*, then*

$$\left|\varphi_Z(t) - e^{-\frac{t^2}{2}}\right| = O\left(\frac{t^3 e^{-\frac{t^2}{3}}}{n} + \frac{t}{\sqrt{n}}\right)$$

This result is always true, but useless for $t >> \sqrt{n}$. To address the situation as t grows larger we prove the following result.

Theorem 9. *Fix $\epsilon > 0$. For $n^{\epsilon} < t \leq n^{\frac{1}{2}+\frac{\epsilon}{4}}$*

$$|\varphi_Z(t)| \leq O\left(\frac{1}{\sqrt{n}t^{1-\epsilon}}\right)$$

Lastly we have one more case which covers yet more values of t.

Theorem 10. *Fix $\epsilon > 0$. For $n^{\frac{1}{2}+\epsilon} \leq t \leq n^{1-\epsilon}$ we have that*

$$|\varphi_Z(t)| \leq O\left(\frac{1}{tn^{1-\epsilon}}\right)$$

We then show an application of all of these characteristic function bounds in the form of a quantitative central limit theorem for subgraph counts by the use of the Esseen Smoothing Lemma. We restate an appropriate version of the smoothing result (Lemma 2 of Chapter 16 in Feller [14] following a result of A.C. Berry).

Lemma 7. *Assume Z has $\mathbb{E}[Z] = 0$ and characteristic function $\varphi_Z(t)$. Then if we let $\mathcal{N}(x) := \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$, the density of the normal, and $\varphi := e^{-t^2/2}$ be the characteristic function of the normal. Finally let $\mathcal{Z}$ be the cumulative distribution function of Z and $\mathfrak{N}$ the c.d.f. of the standard unit normal. Then for any x and $T > 0$*

$$|\mathcal{Z}(x) - \mathfrak{N}(x)| \leq \frac{1}{\pi}\int_{-T}^{T}\left|\frac{\varphi_Z(t) - \varphi(t)}{t}\right|dt + \frac{24}{\pi\sqrt{2\pi}T} \tag{2.7}$$

We can now easily obtain the following quantitative central limit theorem for subgraph count like functions.

Theorem 11. *Assume $F : \binom{[n]}{2} \to \mathbb{R}$, a graph statistic defined from $f : \binom{[k]}{2} \to \mathbb{R}$, is edge dominated and $Z = \frac{F-\mu}{\sigma}$. Then we have that for any $a < b$ fixed constants and $\epsilon > 0$*

$$\Pr(Z \in (a, b)) = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-x^2/2}dx + O_{\epsilon}\left(\frac{1}{n^{\frac{1}{2}-\epsilon}}\right)$$

Proof. Fix $T = \sqrt{n}$. For all $t \leq \sqrt{n}$ we can apply either Theorem 8 or 9 to bound

$$\begin{aligned}\frac{1}{\pi}\int_{-T}^{T}\left|\frac{\varphi_Z(t)-\varphi(t)}{t}\right|dt &\leq 2\int_0^{n^\epsilon}\frac{1}{t}O\left(\frac{t^3e^{-\frac{t^2}{3}}}{n}+\frac{t}{\sqrt{n}}\right)dt + 2\int_{n^\epsilon}^{\sqrt{n}}\frac{1}{t}O\left(\frac{1}{\sqrt{n}t^{1-\epsilon}}\right)dt \\ &= O_\epsilon\left(\frac{1}{n^{\frac{1}{2}-\epsilon}}\right)\end{aligned}$$

The result now follows immediately from Lemma 7 □

2.6.3 Properties of Graph Statistics

In this subsection we compute the Fourier Coefficients, variance and spectral concentration of F, where F is a graph statistic defined from f as in Definition 3. Fix a set $T \subset \binom{[n]}{2}$. Note that a map $\psi \in \mathcal{S}_G$ such that $\text{supp}(T) \subset \psi(D)$ can be determined as follows: Pick an injection $\phi : \text{supp}(T) \hookrightarrow D$, and for $v \in \phi(\text{supp}(T))$ set $\psi(v) = \phi^{-1}(v)$. We can then extend ψ to a map on all of $[k]$ it by specifying the image of ψ on $\phi(T)^c$ arbitrarily. An extension can be picked in $(n-|\text{supp}(T)|)\downarrow_{k-|\text{supp}(T)|}$ ways. So we have that

$$\begin{aligned}\hat{F}(T) = \sum_{\substack{\psi\in\mathcal{S}_G \\ \text{supp}(T)\subset\psi(D)}}\hat{f}(\psi^{-1}(T)) &= \sum_{\phi:\text{supp}(T)\hookrightarrow D}\ \sum_{\psi:[k]-\varphi(T)\hookrightarrow[n]}\hat{f}(\psi^{-1}(T)) \\ &= (n-|\text{supp}(T)|)\downarrow_{k-|\text{supp}(T)|}\sum_{\varphi:\text{supp}(T)\hookrightarrow D}\hat{f}(\psi^{-1}(T)) \\ &= (n-|\text{supp}(T)|)\downarrow_{k-|\text{supp}(T)|} h_T\end{aligned} \tag{2.8}$$

Furthermore, this shows us that $\hat{F}(T) = \Theta(n^{k-|\text{supp}(T)|})$, so long as we have that $h_T \neq 0$. It is of particular importance whether $h_e = 0$ were e is a set consisting of a single edge. Using these estimates and Parseval's 2.1 we can compute the variance of F to be

$$\begin{aligned}\sigma^2 := Var(F) &= \binom{n}{2}h_e^2((n-2)\downarrow_{k-2})^2 + \sum_{i=3}^{k}((n-i)\downarrow_{k-i})^2\sum_{|\text{supp}(T)|=i}h_T^2 \\ &= \binom{n}{2}h_e^2n^{2k-2} + O\left(\sum_{i=3}^{k}((n-i)\downarrow_{k-i})^2\left[\binom{n}{i}\max_{|\text{supp}(T)=i|}h_T^2\right]\right) \\ &= \sum\binom{n}{2}h_e^2n^{2k-2} + O(n^{2k-3})\end{aligned} \tag{2.9}$$

So we see that if $h_e \neq 0$, that is f is edge dominated, then $\sigma^2 - W^1(F) = O(n^{2k-3})$. In fact we have shown that more is true, and that for any $j \geq 1$ we have that $W^j(F)/\sigma^2 = O(n^{-j+1})$

2.6.4 Small values of t

The goal of this subsection is to prove Theorem 8, which we restate.

Theorem 8. *Let $F : \binom{[n]}{2} \to \mathbb{R}$ be a graph statistic defined from $f : \binom{[k]}{2} \to \mathbb{R}$ be as in Definition 3. Assume F is edge dominated, that is $h_e = \sum_{e \in \binom{[k]}{2}} \hat{f}(e) \neq 0$. Let $Z := \frac{F - \mathbb{E}F}{\sigma}$, then*

$$\left| \varphi_Z(t) - e^{-\frac{t^2}{2}} \right| = O\left(\frac{t^3 e^{-\frac{t^2}{3}}}{n} + \frac{t}{\sqrt{n}} \right)$$

Theorem 8 concerns Z, a normalized form of F with mean 0 and variance 1. We can decompose Z into two parts, as we did in the triangle case, the dominant weight one part X, and a smaller term Y corresponding to Fourier coefficients of weight ≥ 2. Let $Q := \sqrt{\frac{1}{\binom{n}{2}}}$ and

$$X := \sum_{e \in \binom{[n]}{2}} Q\chi_e \qquad\qquad Y := \sum_{e \in \binom{[n]}{2}} (\hat{Z}(e) - Q)\chi_e + \sum_{|S| \geq 2} \hat{Z}(S)\chi_S$$

First we examine X. It is the mean 0 variance 1 sum of independent random variables, and so by Berry-Esseen (see Petrov [37], Chapter V lemma 1) we know that if

$$L_n := \binom{n}{2} \mathbb{E}[|Q\chi_e|^3] = \frac{p^2 + (1-p)^2}{\sqrt{\binom{n}{2} p(1-p)}} = \Theta_p(1/n)$$

then for $t \leq \frac{1}{4L_n}$ we have that

$$\left| \mathbb{E}[e^{itX}] - e^{-\frac{t^2}{2}} \right| \leq 16 L_n |t|^3 e^{\frac{-t^2}{3}} \tag{2.10}$$

Next we examine Y. It is best considered as an error term, and we will show that $\mathbb{E}\,|Y|$

is small. We know from prior calculations 2.8 and 2.9

$$\sum_{|S|\geq 2} \hat{Z}^2(S) = O\left(\frac{kn^{2k-3}}{\sigma^2}\right) = O_k\left(\frac{1}{n}\right)$$

Further we can estimate

$$\binom{n}{2}\sigma^2\hat{Z}^2(e) - \sigma^2 = \binom{n}{2}\hat{F}^2(e) - \sigma^2 = O(n^{2k-3}) \implies \hat{Z}^2(e) - \frac{1}{\binom{n}{2}} = O(n^{-3})$$

Therefore using the fact that $(x-y) = (x^2-y^2)/(x+y)$ coupled with the observation that $\hat{Z}(e) + Q = \Theta\left(\frac{1}{n}\right)$, we find that

$$|\hat{Z}(e) - Q| \leq \left|\frac{\hat{Z}^2(e) - \frac{1}{\binom{n}{2}}}{\hat{Z}(e) + Q}\right| = O\left(\frac{1}{n^2}\right)$$

So as a result we can conclude that $var(Y) = O(1/n)$ and so $\mathbb{E}[|Y|] = O(\frac{1}{\sqrt{n}})$. Now we are ready for our characteristic function bound for Z. If $|t| \leq \frac{1}{4L_n} = \Theta_p(n)$ then combining the above with equation 2.10 yields.

$$\begin{aligned}
\left|\varphi_Z(t) - e^{-\frac{t^2}{2}}\right| &= \left|\mathbb{E}\left[e^{itZ}\right] - e^{-\frac{t^2}{2}}\right| = \left|\mathbb{E}\left[e^{it(X+Y)}\right] - e^{-\frac{t^2}{2}}\right| \\
&\leq \left|\mathbb{E}\left[e^{itX}\right] - e^{-\frac{t^2}{2}}\right| + \left|\mathbb{E}\left[e^{itX+Y}\right] - \mathbb{E}\, e^{itX}\right| \leq 16L_n|t|^3 e^{\frac{-t^2}{3}} + \mathbb{E}\,|tY| \\
&= O\left(\frac{t^3 e^{-\frac{t^2}{3}}}{n} + \frac{t}{\sqrt{n}}\right)
\end{aligned}$$

But this is exactly the statement of Theorem 8

2.6.5 Bounds for slightly larger t

The goal for this subsection is to prove

Theorem 9. *Fix $\epsilon > 0$. For $n^\epsilon < t \leq n^{\frac{1}{2}+\frac{\epsilon}{4}}$*

$$|\varphi_Z(t)| \leq O\left(\frac{1}{\sqrt{n}t^{-1+\epsilon}}\right)$$

To prove this statement we will first need the following claims:

Claim 3. *Fix $\epsilon > 0$. For all sufficiently large n, we have that for any $\alpha \in (n^{-1+\epsilon}, 1)$ there exists a set of edges $H \subset \binom{[n]}{2}$ with $|H| \geq \frac{\alpha n}{2}$ such that*

$$\sum_{\substack{S \in H \\ |S| \geq 2}} n^{2k-2|\mathrm{supp}(S)|} \leq C\alpha^2 n^{2k-3}$$

Where C is a fixed constant depending only on f.

Claim 4. *Let A be the event (over the space of revelations $\beta \in \{0,1\}^{H^c}$) that for* every *edge $e \in H$ we have that*

$$|\widehat{Z_\beta}(e) - \hat{Z}(e)| < \frac{1}{n^{1.4}}$$

Let $\lambda := \min(p, 1-p)$. Then $\Pr(A) \geq 1 - n^2\lambda^2 e^{-\Omega\left(\lambda n^{\frac{0.1}{k^2}}\right)}$.

Claim 5. *Let B be the event (over the space of revelations $\beta \in \{0,1\}^{H^c}$) that for* every *set $S \subset E$ with $|S| \geq 2$*

$$|\widehat{Z_\beta}(S)| \leq Cn^{k-s}$$

where C is the fixed constant $C := h_* 2^{\binom{k}{2}} + 1$ and $s = |\mathrm{supp}(S)|$. Let $\lambda := \min(p, 1-p)$. Then $\Pr(B) \geq 1 - O\left(n^k e^{-\Omega\left(n^{\frac{2}{k^2}}\right)}\right)$

Claim 6. *Assume $\beta \in A \cap B$. Then for any $\alpha \in (n^{-1+\epsilon}, 1)$ and $t = o(n)$*

$$\mathop{\mathbb{E}}_{x \in 2^H}[e^{itZ_\beta}] \leq \exp\left(-\frac{\alpha t^2}{8\pi^2}\right) + O\left(|t|\alpha n^{k-\frac{3}{2}}\right)$$

Claim 7. *For $\alpha \in (n^{-1+\epsilon}, 1)$we have that*

$$|\varphi_Z(t)| < \exp\left(-\frac{\alpha t^2}{8\pi^2}\right) + O\left(\alpha|t|n^{k-\frac{3}{2}} + n^k e^{-\Omega\left(n^{\frac{2}{k^2}}\right)} + n^2 e^{-\Omega\left(\lambda n^{\frac{0.1}{k^2}}\right)}\right)$$

Theorem 9 now follows simply by making a good choice of α.

Proof of Theorem 9. We can now fix α to be of size $t^{-2+\epsilon}$, which is feasible for the hypothesis of Claim 3 so long as we assure that $n^\epsilon < t < n^{\frac{1}{2}+\frac{\epsilon}{4}}$, and so $\alpha > n^{-1+\epsilon/2}$. Plugging this choice of α into Claim 7 completes the proof. □

Proof of Claims

Proof Of Claim 3. Fix $\ell = \lfloor \alpha n \rfloor$. Let H be the subgraph given by taking the union of $\lfloor \frac{n}{\ell} \rfloor$ disjoint cliques of size ℓ, and the remaining vertices with no edges. The number of edges in H is

$$\binom{\ell}{2} \lfloor \frac{n}{\ell} \rfloor \geq \frac{n(\ell-1)}{2} - \binom{\ell}{2} \geq \frac{\alpha n^2}{2} - \ell^2 - \frac{n}{2}$$

Meanwhile the number of subgraphs of H with support of size $|\text{supp}(S)| = i$ is upper bounded by

$$\lfloor \frac{n}{\ell} \rfloor \left(\binom{\ell}{i} 2^{\binom{i}{2}} \right) \leq n \ell^{i-1} 2^{i^2} = \alpha^{i-1} n^i (2^{i^2} + O(\frac{1}{\ell}))$$

So we can see that the number of edges is at least $\frac{\alpha n}{2}$ for n sufficiently large. Further we can compute that

$$\begin{aligned} \sum_{\substack{S \in H \\ |S| \geq 2}} n^{2k-2|\text{supp}(S)|} &\leq \sum_{i=3}^{k} \sum_{\substack{S \in H \\ |\text{supp}(S)|=i}} n^{2k-2i} \leq \sum_{i=3}^{k} (2^{i^2} + O(\frac{1}{\ell})) \alpha^{i-1} n^i n^{2k-2i} \\ &\leq (k + O(\frac{1}{\ell})) \alpha^2 n^{2k-i} = O(\alpha^2 n^{2k-3}) \end{aligned}$$

Where the last inequality is justified by the assumption that $\ell \geq h(n)$ where $n \to \infty$. $\square$

We prove Claim 4 by noting that the formula for $\widehat{Z_\beta}(S)$ (a coefficient in the polynomial Z_β) is *itself* a low degree polynomial, and therefore may be shown to have tight concentration by hypercontractivity.

Proof Of Claim 4. Recall that

$$\widehat{Z_\beta}(e) = \sum_{T \subset H^c} \hat{Z}(e \cup T) \chi_T(\beta)$$

So $\widehat{Z_\beta}(e) : \{0,1\}^{H^c} \to \mathbb{R}$ is a polynomial (in the functions χ_e), and we can began by estimating its coefficients. First we see that

$$\mathbb{E}[\widehat{Z_\beta}(e)] = \widehat{\widehat{Z_\beta}(e)}(\varnothing) = \hat{Z}(e)$$

Also for any $T \subset \{0,1\}^{H^c}$ we know that $\hat{Z}(e \cup T) \neq 0$ only if $|\text{supp}(e \cup T)| \leq k$. So we can compute:

$$\begin{aligned} Var_\beta(\widehat{Z_\beta}(e)) &= \sum_{\substack{T \subset H^c \\ T \neq \varnothing}} \hat{Z}(e \cup T)^2 = \sum_{i=3}^{k} \sum_{\substack{T \subset H^c \\ |\text{supp}(T \cup e)|=i}} \hat{Z}(e \cup T)^2 \\ &\leq \sum_{i=3}^{k} \sum_{|\text{supp}(T \cup e)|=i} \hat{Z}(e \cup T)^2 \leq \sum_{i=3}^{k} \binom{n-2}{i-2} \frac{h_*^2 n^{2(k-i)}}{\sigma^2} \leq k h_*^2 \frac{n^{2k-5}}{\sigma^2} \\ &= O\left(\frac{1}{n^3}\right) \end{aligned}$$

Since $\widehat{Z_\beta}(e)$ has degree less than $\binom{k}{2}$, an application of Theorem 4 gives us that for any $e \in H$ if we set $\lambda = \min(p, 1-p)$ then

$$\Pr\left[\left|\widehat{Z_\beta}(e) - \hat{Z}(e)\right| \geq \frac{1}{n^{1.4}}\right] < \lambda^2 \exp\left(-\Omega\left(\frac{\lambda n^{\frac{0.1}{k^2}}}{e}\right)\right)$$

Applying a union bound over all edges in H completes the proof. □

Proof of Claim 5. Again we use the decomposition

$$\widehat{Z_\beta}(S) = \sum_{T \subset H^c} \hat{Z}(S \cup T) \chi_T(\beta)$$

So $\widehat{Z_\beta}(S) : \{0,1\}^{H^c} \to \mathbb{R}$ is a polynomial (in the functions χ_e), and we can began by estimating its coefficients. First we see that

$$\mathbb{E}[\widehat{Z_\beta}(S)] = \widehat{\widehat{Z_\beta}(S)}(\varnothing) = \hat{Z}(S)$$

Assume $|\text{supp}(S)| = s$. For any $T \subset \{0,1\}^{H^c}$ we know that $\hat{Z}(S \cup T) \neq 0$ iff $|\text{supp}(S \cup T)| \leq k$. There are at most $2^{\binom{\ell}{2}}(n-s)\downarrow_{\ell-s} \leq 2^{k^2} n^{\ell-s}$ choices of T such that $|\text{supp}(S \cup T)| = \ell$. And further for each of these choices we know that $\hat{Z}(S \cup T) \leq h_* n^{k-\ell}$.

Let $g := \sum_{\substack{T \subset H^c \\ |\mathrm{supp}(S \cup T)| > s}} \hat{Z}(S \cup T)\chi_T(\beta)$ So we can compute that

$$\begin{aligned} Var(g) &\leq \sum_{\ell=s+1}^{k} \sum_{|\mathrm{supp}(S\cup T)=\ell} \left(\hat{Z}(S \cup T)\right)^2 \leq \sum_{\ell=s+1}^{k} 2^{k^2} n^{\ell-s} (h_*)^2 n^{2k-2\ell} \\ &\leq k 2^{k^2} (h_*)^2 n^{2k-2s-1} \end{aligned}$$

Further we can see that g is a polynomial of degree at most $2^{\binom{k}{2}}$, and so by Hypercontractivity 4 we see that

$$\begin{aligned} \Pr\Big[|g| \geq n^{k-s}\Big] &= \Pr\left[g \geq \frac{1}{\sqrt{k 2^{k^2} (h_*)^2}} \sqrt{n} \|g\|_2\right] \leq \lambda^{\binom{k}{2}} \exp\left(-\frac{\binom{k}{2}}{2e} \lambda \left(\frac{t}{k 2^{k^2} (h_*)^2}\right)^{\frac{2}{\binom{k}{2}}}\right) \\ &= O(e^{-\Omega(n^{\frac{2}{k^2}})}) \end{aligned}$$

If $|g| < n^{k-s}$ then we can conclude that

$$\hat{Z}(S) = \sum_{|\mathrm{supp}(S\cup T)|=s} \hat{Z}(S \cup T)\chi_T(\beta) + g(\beta) \leq 2^{\binom{s}{2}} h_* n^{k-s} + n^{k-s}$$

So for any $S \subset H$ we find that $|\widehat{Z_\beta}(S)| \leq C n^{2k-2s}$ with probability at least $1 - O(e^{-\Omega(n^{\frac{2}{k^2}})})$. Taking a union bound over all such S finishes the proof. □

Proof of Claim 6. Assume that $\beta \in A \cap B$. Let X and Y be

$$X := \sum_{e \in \binom{[n]}{2}} \widehat{Z_\beta}(e)\chi_e \qquad\qquad Y := \sum_{|S| \geq 2} \widehat{Z_\beta}(S)\chi_S$$

then $Z_\beta = X + Y$, where X is an independent sum, and Y is likely small, so we will be able to obtain bounds similar to our previous ones. Let $Q = \sqrt{\frac{1}{\binom{n}{2}}} = (1 + O(\frac{1}{n}))\hat{Z}(e)$, and that $Q \approx \frac{\sqrt{2}}{n}$. Because of our assumption that $\beta \in A$ we have that

$$\frac{Q}{2} \leq \hat{Z}(e) - n^{-1.4} \leq \widehat{Z_\beta}(e) \leq \hat{Z}(e) + n^{-1.4} \leq \frac{3Q}{2}$$

Using our bound on $\widehat{Z_\beta}(e)$ and the fact that $t = o(1/Q)$ we may apply Lemma 2 to

say that

$$|\mathbb{E}[e^{it\widehat{Z_\beta}(e)\chi_e}]| = 1 - \frac{t^2Q^2}{2\pi^2} \leq \exp\left(-\frac{t^2}{2n^2\pi^2}\right)$$

So now we find that

$$\begin{aligned}\mathbb{E}[e^{itX}] &= \prod_{e\in H}\mathbb{E}[\exp\left(it\widehat{Z_\beta}(e)\chi_e\right)] \leq \exp\left(-\sum_{e\in H}(t\widehat{Z_\beta(e)})^2\right) \leq \exp\left(\sum_{e\in H}-\frac{t^2}{2\pi^2n^2}\right)\\ &= \exp\left(-\frac{\alpha n^2}{4}\cdot\frac{t^2}{2\pi^2n^2}\right)\end{aligned}$$

Next we turn our attention to Y. We can use Cauchy Schwartz, the assumption that $\beta \in B$ and the fact that H satisfies the conditions of Claim 3 to bound

$$\mathbb{E}[|Y|]^2 \leq \mathbb{E}[|Y|^2] = \sum_{\substack{S\subset H\\|S|\geq 2}}\widehat{Z_\beta}^2(S) \leq \sum_{\substack{S\subset H\\|S|\geq 2}} C^2n^{2k-2|\mathrm{supp}(S)|} \leq O(\alpha^2n^{2k-3})$$

Finally we combine all of these estimates to bound $\mathbb{E}_H[e^{itZ_\beta}]$ and finish the proof of Claim 6

$$\begin{aligned}\left|\mathop{\mathbb{E}}_{\alpha\in 2^H}[e^{itZ_\beta(\alpha)}]\right| &= \left|\mathop{\mathbb{E}}_{\alpha}[e^{it(X+Y)}]\right| \leq \left|\mathbb{E}[e^{itX}+|tY|]\right|\\ &\leq \exp\left(-\frac{\alpha t^2}{8\pi^2}\right) + O\left(|t|\alpha n^{k-\frac{3}{2}}\right)\end{aligned}$$

□

Proof of Claim 7. Let A, and B be as defined in Claims 4 and 5. We can break up $\{0,1\}^{H^c}$ into $A\cap B$ and $(A\cap B)^c$ and estimate

$$\begin{aligned}|\varphi_Z(t)| := \mathop{\mathbb{E}}_{(\alpha,\beta)\in 2^{\binom{n}{2}}}[e^{itZ(\alpha,\beta)}] &\leq \mathop{\mathbb{E}}_{\beta\subset H^c}|\mathop{\mathbb{E}}_{\alpha\subset H}[e^{itZ_\beta(\alpha)}]|\\ &\leq \Pr[(A\cap B)^c] + \Pr[A\cap B]\mathop{\mathbb{E}}_{\beta\in(A\cap B)^c}\left|\mathop{\mathbb{E}}_{\alpha}[e^{itZ_\beta}]\right|\end{aligned}$$

Now combining Claims 4 and 5 we find that

$$\Pr[(A \cap B)^c] + Pr[A \cap B] \mathop{\mathbb{E}}_{\beta \in A} \left| \mathop{\mathbb{E}}_{\alpha}[e^{itZ_\beta}] \right| \leq$$
$$\exp\left(-\frac{\alpha t^2}{8\pi^2}\right) + O\left(\alpha|t|n^{k-\frac{3}{2}} + n^k e^{-\Omega\left(n^{\frac{2}{k^2}}\right)} + n^2 e^{-\Omega\left(\lambda n^{\frac{0.1}{k^2}}\right)}\right)$$

□

2.6.6 Middle values of t

This subsection does not have a direct analog in the triangle case, as the tighter Cauchy-Schwarz bound given in [16] may be used in that case. The goal of this subsection is to prove

Theorem 10. *Fix $\epsilon > 0$. For $n^{\frac{1}{2}+\epsilon} \leq t \leq n^{1-\epsilon}$ we have that*

$$|\varphi_Z(t)| \leq O\left(\frac{1}{tn^{1-\epsilon}}\right)$$

For $t \geq n^{\frac{1}{2}+\epsilon}$ we use a different choice of H, the subgraph whose complement we reveal, from in the previous arguments. Here we take H to be a matching of size ℓ. Again let $\beta \in 2^{H^c}$ be a revelation of all of the edges in H^c and look at

$$Z_\beta = Z(\mathbf{x}_H, \beta) = X_\beta + Y_\beta$$

where

$$X_\beta := \sum_{e \in H} \widehat{Z_\beta}(e)\chi_e \qquad\qquad Y_\beta := \sum_{\substack{S \subset H \\ |S| \geq 2}} \widehat{Z_\beta}(S)\chi_S$$

Because H is a matching, any set $S \subset H$ has support of size $|\text{supp}(S)| = 2|S|$. So assuming that we are again in the event $|A \cap B|$ (i.e. all of the Fourier coefficients are behaved nicely where A and B are as defined in Claims 4 and 5 respectively) we can

compute that

$$\mathbb{E}[|Y_\beta|]^2 \leq \mathbb{E}[|Y_\beta|^2] = \sum_{\substack{S\subset H\\|S|\geq 2}} \widehat{Z_\beta}^2(S) \leq \sum_{i=2}^{2\ell} \sum_{\substack{S\subset H\\|S|=i}} \frac{C^2}{\sigma^2} n^{2k-4i}$$
$$\leq \sum_{i=2}^{2\ell} \binom{\ell}{i} \frac{C^2}{\sigma^2} n^{2k-4i} \leq 2\ell^2 \frac{C^2}{\sigma^2} n^{2k-8}$$

So we have $\mathbb{E}[|Y_\beta|] = O(\ell n^{-3})$. Meanwhile so long as $t\widehat{Z_\beta}(e) < \sqrt{p(1-p)}\pi$ we can use Lemma 2 to compute

$$\mathbb{E}[e^{itX_\beta}] = \prod_{e\in H} \mathbb{E}[\exp\Big(it\widehat{Z_\beta}(e)\chi_e\Big)] \leq \exp\left(-\sum_{e\in H}(t\widehat{Z_\beta(e)})^2\right)$$
$$\leq \exp\left(\ell\frac{t^2}{2n^2}\right)$$

So for $n^{\frac{1}{2}+\epsilon} < t < n^{1-\epsilon}$ we can choose $\ell = \lfloor \frac{n^{2+\epsilon}}{t^2} \rfloor \in [n/2]$ (the size bound verifying that there does exist a matching of size ℓ) and then we find $\mathbb{E}[e^{itX_\beta}] \leq \exp(-n^\epsilon)$. So then in total we have that if $\beta \in A \cap B$ then

$$\mathop{\mathbb{E}}_{H}[e^{itZ_\beta}] = \mathbb{E}[e^{it(X_\beta+Y_\beta)}] \leq \mathbb{E}[e^{itX_\beta} + t|Y_\beta|] \leq e^{-n^\epsilon} + O\left(\frac{t\ell}{n^3}\right) = e^{-n^\epsilon} + O\left(\frac{1}{tn^{1-\epsilon}}\right)$$

Also, arguing as we did in Claim 7, we can use Claims 4 and 5 to show that $\Pr[\beta \in (A\cap B)^c] \leq O\big(\frac{1}{tn^{1-\epsilon}}\big)$. So we can now conclude that

$$\mathbb{E}[e^{itZ}] = \mathop{\mathbb{E}}_{\beta}\mathop{\mathbb{E}}_{H} e^{itZ} \leq O\left(\frac{1}{tn^{1-\epsilon}}\right)$$

Concluding the proof.

Chapter 3

A Stability Result Using the Matrix Norm to Bound the Permanent

3.1 Introduction

The *permanent* of an $n \times n$ matrix, A, has long been an important quantity in combinatorics and computer science, and more recently it has also had applications to physics and linear-optical quantum computing. It is defined as

$$\mathrm{per}(A) := \sum_{\sigma \in S_n} \prod_{i=1}^{n} a_{i,\sigma(i)},$$

where S_n denotes the set of permutations of $[n] = \{1, 2, \ldots, n\}$. For instance, if A only has entries in $\{0, 1\} \subseteq \mathbb{R}$, then the permanent counts the number of perfect matchings in the bipartite graph whose bipartite adjacency matrix is A.

The definition of the permanent is of course reminiscent of that for the determinant; however, whereas the determinant is rich in algebraic and geometric meaning, the more combinatorial permanent is notoriously difficult to understand. For example, computing $\mathrm{per}(A)$ even for $\{0, 1\}$-matrices is the prototypical #P-complete problem (Valiant [43]).

On the other hand, the *operator 2-norm* (also called the *operator norm*) of a matrix is a particularly nice parameter. For an $n \times n$ matrix A with entries in $\mathbb{C}$, it is defined as

$$\|A\|_2 = \sup_{\|\vec{x}\|_2 \leq 1,\ \vec{x} \in \mathbb{C}^n} \|A\vec{x}\|_2,$$

where $\|\vec{v}\|_p$ is the usual l_p norm (i.e., $\|\vec{v}\|_p^p = \sum_i |v_i|^p$ for $p \in (0, \infty)$, and $\|\vec{v}\|_\infty = \max |v_i|$). The operator norm of a matrix has the advantages of being both algebraically and analytically well-behaved as well as computationally easy to determine (as this amounts to finding the largest singular value of A).

Considering how differently behaved the permanent and operator norm are, it is perhaps strange to think that there would be much of a connection between them. Nonetheless, they are related by the following extremal result, which is due to Gurvits [19] (see also [2, 3]).

Theorem 1. *Suppose A is an $n \times n$ matrix over $\mathbb{C}$ (resp. $\mathbb{R}$), and let $\mathcal{P}$ denote the set of $n \times n$ matrices over $\mathbb{C}$ (resp. $\mathbb{R}$) that can be written as a permutation matrix times a unitary diagonal matrix. Then $|\mathrm{per}(A)| \leq \|A\|_2^n$ with equality iff A is a scalar multiple of a matrix in $\mathcal{P}$.*

Note that this extremal set $\mathcal{P}$ is simply the set of matrices with exactly n non-zero entries, each having modulus 1, and no two of which are in the same row or column. Such a matrix $P \in \mathcal{P}$ has $\|P\|_2 = |\mathrm{per}(P)| = 1$ and satisfies

$$\|AP\|_2 = \|PA\|_2 = \|A\|_2, \qquad \text{and} \qquad |\mathrm{per}(AP)| = |\mathrm{per}(PA)| = |\mathrm{per}(A)|$$

for all matrices A (which is equivalent to membership in $\mathcal{P}$). Moreover, $\mathcal{P}$ is a subgroup of the group of unitary matrices, and as a set, it has a very tractable topological structure.

Motivated by algorithmic questions related to approximating the permanent, Aaronson and Hance [2] asked whether one could prove a stability version of Theorem 1:

Question A:

If $|\mathrm{per}(A)|$ is close to $\|A\|_2^n$, must $A/\|A\|_2$ be 'close' to a matrix in $\mathcal{P}$?

A somewhat more concrete version was suggested by Aaronson and Nguyen [3]:

Question B:

Characterize $n \times n$ matrices A such that $\|A\|_2 \leq 1$ and there exists a constant $C > 0$ such that $|\mathrm{per}(A)| \geq n^{-C}$.

Using techniques of inverse Littlewood-Offord theory, Aaronson and Nguyen gave a substantial answer to an analogous question under the (stronger) assumptions that A is orthogonal and that the intersection of the hypercube $\{\pm 1\}^n$ with its image under A is large. They also proved something like (actually slightly stronger than) our results below for stochastic matrices. Further results in the direction of Question B were given by Nguyen [34].

The two main results of the present paper are Theorems 2 and 3 below. The first provides a positive answer to Question A for matrices over $\mathbb{C}$ (or $\mathbb{R}$), and the second is a more refined result that (depending on your philosophical views) at least partially addresses Question B for matrices over $\mathbb{R}$. More specifically, we bound $\mathrm{per}(A)$ in terms of the following easily computed parameters.

Definition:

Let A be a matrix with rows $r_1, r_2, \ldots, r_n$, and $p \in \mathbb{R} \cup \{\infty\}$. Then the parameter $h_p(A)$ is defined as $h_p(A) = h_p = \frac{1}{n} \sum_i \|r_i\|_p$.

We will only consider h_∞ and h_2. First note $0 \leq h_\infty(A) \leq h_2(A) \leq \|A\|_2$. Moreover, it is easy to show $h_2(A) = \|A\|_2$ iff $A/\|A\|_2$ is a unitary matrix, and $h_\infty(A) = \|A\|_2$ iff $A/\|A\|_2$ is in $\mathcal{P}$. Thus, in some sense, the quantity $1 - h_2(A)/\|A\|_2 \in [0,1]$ measures how close $A/\|A\|_2$ is to being unitary, and $1 - h_\infty(A)/\|A\|_2 \in [0,1]$ measures how close $A/\|A\|_2$ is to being in $\mathcal{P}$. Broadly speaking, $h_\infty/\|A\|_2$ is close to 1 precisely when most of the rows of A each have one entry of modulus close to $\|A\|_2$ and all the other entries in that row are close to 0.

Before stating the first of our main results, notice that in addressing either of the above questions, we lose no generality in assuming $\|A\|_2 \leq 1$, since Question A is invariant under scaling. However, to facilitate any application of our results, we state them in the "more general" case that $\|A\|_2 \leq T$.

Theorem 2. *Let A be an $n \times n$ matrix over $\mathbb{C}$ and $\|A\|_2 \leq T \neq 0$. Then*

(i) $|\mathrm{per}(A)| \leq T^n \exp\left[-3n\Big(1 - \frac{\sqrt{\pi}}{2}h_2/T - \Big(1 - \frac{\sqrt{\pi}}{2}\Big)h_\infty/T\Big)^2/100\right]$,

(ii) $|\mathrm{per}(A)| \leq T^n \exp[-n(1 - h_\infty/T)^2/10^5]$.

As discussed above, this provides a positive answer to Question A by viewing h_∞ (and to a lesser extent h_2) as a proxy for 'closeness' of a matrix A to those in $\mathcal{P}$. As an easy corollary, if $\alpha, \beta \geq 0$ satisfy $|\mathrm{per}(A)| \geq 2T^n \exp[-n\alpha^2\beta^2/10^5]$, then all but at most αn of the rows of A contain an entry whose modulus is at least $T(1-\beta)$. And since the l_2 norm of any row of A is at most $\|A\|_2$, no entry of A can have modulus larger than T. Thus, entries of modulus $T(1-\beta)$ are nearly as large as possible. Moreover, if a row (or column) has an entry with very large modulus, then the remaining entries must have very small moduli (again since its l_2 norm is at most $\|A\|_2$). Thus, this theorem also provides a *qualitative* stability result stating that matrices with large permanent must have many very large entries, and a row (or column) containing a large entry must have all its other entries small.

Note that Theorem 2 is only useful for values of h_∞/T that are not very close to 1—namely when $1 - h_\infty/T \gg n^{-1/2}$. Although this does well in many cases, we believe that for large values of h_∞/T, it is not optimal. For comparison, if A is δ times the identity matrix, and $\delta \approx 1$, then $|\mathrm{per}(A)| \approx e^{-n(1-\delta)} = e^{-n(1-h_\infty)}$, and we conjecture that this is essentially tight.

Conjecture 1. *There is some constant $C > 0$ and some polynomial $f(n)$ such that the following holds. If A is an $n \times n$ matrix with complex entries and $\|A\|_2 \leq 1$, then* $|\mathrm{per}(A)| \leq f(n)e^{-Cn(1-h_\infty)}$.

As a step in this direction, we are able to prove the following, which better addresses Question B for matrices over $\mathbb{R}$.

Theorem 3. *Let A be an $n \times n$ matrix over $\mathbb{R}$ and $\|A\|_2 \leq T \neq 0$. Then*

$$|\mathrm{per}(A)| \leq T^n(n+6)\exp\left[\frac{-\sqrt{n(1-h_\infty/T)}}{400}\right].$$

As with Theorem 2, a result like Theorem 3 that involves h_2 is also possible, and it essentially falls out of our proof directly. Theorem 3 is an improvement over Theorem 2 when $n^{-1/3} \gg 1-h_\infty/T$ and gives a meaningful bound provided $1-h_\infty/T \gg \log(n)^2/n$. Although this yields a quantitatively better understanding for matrices over $\mathbb{R}$, we cannot shake the belief that neither of our main results (i.e., Theorems 2 and 3) is best possible, and we discuss this further in Section 5.2.

Structure of paper

The paper is devoted to proving Theorems 2 and 3, which goes roughly as follows. First, we appeal to a result of Glynn [17] that allows us to convert the problem of estimating the permanent into a problem about estimating the expected value of a certain random variable (Section 3.2). We then use standard probabilistic tools to show certain concentration results for the random variable of interest, which in turn yield the estimates needed for our results. This is done for the complex-valued case in Section 3.3, which proves Theorem 2. In Section 3.4, we consider the real-valued case, where we analyze the corresponding random variable more carefully to obtain Theorem 3. We conclude in Section 5.2 with several open questions and conjectures, as well as a discussion of Question B.

3.2 Definitions and set-up with random variables

We will work over the field $\mathbb{K}$, which will either be $\mathbb{R}$ or $\mathbb{C}$. Given an $n \times n$ matrix A over $\mathbb{K}$ and $x \in \mathbb{K}^n$, set $y = Ax$, and define

$$G_x(A) = \prod_{i=1}^{n} \overline{x}_i \times \prod_{i=1}^{n} y_i,$$

where $\overline{z}$ denotes the complex conjugate of z. Let $X \in \mathbb{K}^n$ be the random variable whose coordinates are independently selected uniformly on $|z| = 1$, and let $Y = AX$ (note: if $\mathbb{K} = \mathbb{C}$, then each coordinate of X is distributed continuously over the unit circle, whereas if $\mathbb{K} = \mathbb{R}$, then X is chosen uniformly from the discrete set $\{-1, 1\}^n$). Then

$$\mathrm{per}(A) = \mathbb{E}[G_X(A)] = \mathbb{E}\left[\prod_{i=1}^{n} \overline{X}_i Y_i\right],$$

obtained simply by expanding out the product in $G_X(A)$ and using the fact that the X_i are independent with mean 0 and variance 1 (for proofs of this fact, see [19, 17, 2, 3]). Therefore, by convexity (which we are about to use twice), we have

$$|\mathrm{per}(A)| \leq \mathbb{E}\left[\prod_{i=1}^{n} |\overline{X}_i Y_i|\right] = \mathbb{E}\left[\prod_{i=1}^{n} |Y_i|\right] \leq \mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n} |Y_i|\right)^n\right] = \mathbb{E}\left[\left(\frac{\|AX\|_1}{n}\right)^n\right].$$

Note that from here, we could say (by Cauchy-Schwartz)

$$\frac{\|AX\|_1}{n} \leq \frac{\|AX\|_2}{\sqrt{n}} = \frac{\|AX\|_2}{\|X\|_2} \leq \|A\|_2,$$

thus obtaining the inequality $|\mathrm{per}(A)| \leq \|A\|_2^n$ of Theorem 1 (the equality case follows by considering equality in the above estimates).

Specializing to norm at most 1

Note that to prove our results, it suffices to prove them for the case $\|A\|_2 \leq 1$. This is because otherwise, we could simply scale the matrix by some α to have norm at most 1, and because $\mathrm{per}(A) = \alpha^n \mathrm{per}(A/\alpha)$, our results would follow. As such, we will

henceforth assume $\|A\|_2 \leq 1$ (explicitly making note of when we do), but this choice is simply for notational ease. We remark that the set-up thus far has also been employed in several other papers [19, 2, 3]; however, the remainder of this paper deviates from the previous literature.

3.3 Proof of Theorem 2 ($\mathbb{K} = \mathbb{C}$)

In the setting where $\|A\|_2 \leq 1$, the permanent is always bounded above by 1 (as shown above), and we want to conclude that under certain conditions, it must be (exponentially) small. We know (since $0 \leq \|AX\|_1/n \leq \|A\|_2 \leq 1$) that for all $\varepsilon \geq 0$ and all $\tilde{\mu} \geq 0$,

$$|\mathrm{per}(A)| \leq \mathbb{E}\left[\left(\frac{\|AX\|_1}{n}\right)^n\right] \leq (\tilde{\mu}/n + \varepsilon)^n + \mathbb{P}(\|AX\|_1 \geq \tilde{\mu} + \varepsilon n).$$

We will pick $\tilde{\mu}$ suitably small with $\tilde{\mu} \geq \mathbb{E}[\|AX\|_1]$ and then argue that $\|AX\|_1$ is tightly concentrated about its mean, which will complete the proof.

The mean of $\|AX\|_1$

We appeal to a theorem of König, Schütt, and Tomczak-Jaegermann [27], which is a variant of Khintchine's inequality conveniently well-suited for our situation (in fact, X was chosen in part so that we could apply this result directly).

Theorem 4 (König et al. [27], 1999)**.** *Let $\mathbb{K}$ be $\mathbb{R}$ or $\mathbb{C}$. Suppose $\vec{a} = (a_1, \ldots, a_n) \in \mathbb{K}^n$ is fixed, and suppose each coordinate of $\xi \in \mathbb{K}^n$ is independently distributed uniformly on $|z| = 1$. Then*

$$\left|\mathbb{E}\left[\left|\sum_i a_i \xi_i\right|\right] - \Lambda_{\mathbb{K}}\|\vec{a}\|_2\right| \leq (1 - \Lambda_{\mathbb{K}})\|\vec{a}\|_\infty,$$

where $\Lambda_{\mathbb{R}} = \sqrt{2/\pi}$ and $\Lambda_{\mathbb{C}} = \sqrt{\pi}/2$.

Applying this to each row of A (and using linearity of expectation) gives

Proposition 5. *With A and $X \in \mathbb{C}^n$ as in Section 3.2, we have*

$$\mathbb{E}[\|AX\|_1/n] \leq \frac{1}{n}\sum_{i=1}^{n}\big[\sqrt{\pi}/2\|r_i\|_2 + \big(1-\sqrt{\pi}/2\big)\|r_i\|_\infty\big] = \frac{\sqrt{\pi}}{2}h_2(A) + \left(1-\frac{\sqrt{\pi}}{2}\right)h_\infty(A).$$

Concentration about mean

To show concentration of $\|AX\|_1$ about its mean, we use a very general and useful result of Talagrand (a form of "Talagrand's inequality"), which can be found in chapter 1 of his book [30].

Theorem 6 (Talagrand [30], 1991)**.** *Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is such that $|f(x) - f(y)| \leq \sigma\|x-y\|_2$ for all $x, y \in \mathbb{R}^n$, and define the random variable $F = f(\xi_1, \xi_2, \ldots, \xi_n)$, where the ξ_i are independent standard normal random variables. Then for all $t \geq 0$,*

$$\mathbb{P}(F > \mathbb{E}[F] + t) \leq e^{-2t^2/(\pi\sigma)^2}.$$

We apply this result to our setting by way of a now standard trick that expresses our random variable of interest as a function of standard Gaussians. In fact, this trick is even discussed in [30], so we could have saved a few lines of the following argument by simply citing a "more applicable" version of Theorem 6 (i.e., one for which this trick has already been incorporated); however, the trick so nicely captures the usefulness of Theorem 6, that we thought it worth recalling here.

Proposition 7. *Suppose $\|A\|_2 \leq 1$, and let $X \in \mathbb{C}^n$ be as in Section 3.2. Then for all $t \geq 0$,*

$$\mathbb{P}(\|AX\|_1 > \mathbb{E}[\|AX\|_1] + tn) \leq e^{-nt^2/\pi^3}.$$

Proof. To make use of Theorem 6, we need to define a suitable $f : \mathbb{R}^n \to \mathbb{R}$, which we do in pieces. First define $\Phi : \mathbb{R} \to \mathbb{R}$ via

$$\Phi(u) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{u} e^{-x^2/2}\,dx,$$

which is the probability that a standard Gaussian is at most u. Then define $g : \mathbb{R}^n \to \mathbb{C}^n$

as

$$g(x_1, \ldots, x_n) = \begin{pmatrix} e^{2\pi i \Phi(x_1)} \\ e^{2\pi i \Phi(x_2)} \\ \vdots \\ e^{2\pi i \Phi(x_n)} \end{pmatrix},$$

and, finally, set $f(x) = \|Ag(x)\|_1$.

Notice that if $\xi_1, \xi_2, \ldots, \xi_n$ are independently sampled from the standard normal distribution, then each $\Phi(\xi_i)$ is distributed uniformly on $[0, 1]$. Therefore $g(\xi_1, \ldots, \xi_n)$ has the same distribution as X, and so $F := f(\xi_1, \ldots, \xi_n)$ has the same distribution as $\|AX\|_1$.

Now let $x, y \in \mathbb{R}^n$ be arbitrary. Then we have

$$\begin{aligned} |f(x) - f(y)| &= \Big| \|Ag(x)\|_1 - \|Ag(y)\|_1 \Big| \leq \|Ag(x) - Ag(y)\|_1 \leq \sqrt{n}\|A(g(x) - g(y))\|_2 \\ &\leq \sqrt{n}\|A\|_2 \|g(x) - g(y)\|_2 \leq \sqrt{n}\|g(x) - g(y)\|_2. \end{aligned}$$

Using the fact that $|e^{i\alpha} - 1| \leq |\alpha|$ for all $\alpha \in \mathbb{R}$, we further bound the above by

$$\begin{aligned} \|g(x) - g(y)\|_2^2 &= \sum_{j=1}^n |e^{2\pi i \Phi(x_j)} - e^{2\pi i \Phi(y_j)}|^2 = \sum_{j=1}^n |e^{2\pi i (\Phi(x_j) - \Phi(y_j))} - 1|^2 \\ &\leq (2\pi)^2 \sum_{j=1}^n |\Phi(x_j) - \Phi(y_j)|^2 \leq 2\pi \sum_{j=1}^n |x_j - y_j|^2 = 2\pi \|x - y\|_2^2. \end{aligned}$$

Thus, $|f(x) - f(y)| \leq \sqrt{2\pi n}\|x - y\|_2$, and appealing to Theorem 6 with $\sigma = \sqrt{2\pi n}$ yields

$$\mathbb{P}(\|AX\|_1 > \mathbb{E}[\|AX\|_1] + tn) = \mathbb{P}(F > \mathbb{E}[F] + tn) \leq e^{-2(nt)^2/(\pi\sqrt{2\pi n})^2} = e^{-nt^2/\pi^3}. \quad \square$$

Finishing the proof for $\mathbb{K} = \mathbb{C}$

Proposition 8. *Let $\|A\|_2 \leq 1$ and $X \in \mathbb{C}^n$ be as in Section 3.2. If $\mathbb{E}[\|AX\|_1/n] = \mu$, then*

$$\mathbb{E}[(\|AX\|_1/n)^n] \leq 2\exp[-3n(1-\mu)^2/100].$$

Proof. Let $L = t\mu + (1-t)$ with $t \in [0,1]$ to be determined. Since $0 \leq \|AX\|_1/n \leq 1$, we have (appealing to Proposition 7 for the last inequality)

$$\begin{aligned}
\mathbb{E}[(\|AX\|_1/n)^n] &\leq L^n + \mathbb{P}(\|AX\|_1/n > L) \\
&\leq \exp[-n(1-L)] + \mathbb{P}(\|AX\|_1/n - \mu > (1-t)(1-\mu)) \\
&\leq \exp[-nt(1-\mu)] + \exp[-n(1-t)^2(1-\mu)^2/\pi^3],
\end{aligned}$$

We now take $2t(1-\mu) = \pi^3 + 2 - 2\mu - \pi^{3/2}\sqrt{\pi^3 + 4 - 4\mu}$ (for which t does lie in the interval $[0,1]$), so as to make the exponents equal. For this t, we obtain

$$\mathbb{E}[(\|AX\|_1/n)^n] \leq 2\exp\Big[-n(2\mu + \pi^{3/2}\sqrt{\pi^3 + 4 - 4\mu} - \pi^3 - 2)/2\Big].$$

Then appealing to the Taylor series at $\mu = 1$, we see that for all $\mu \in [0,1]$,

$$\frac{2\mu + \pi^{3/2}\sqrt{\pi^3 + 4 - 4\mu} - \pi^3 - 2}{2} \geq \frac{(1-\mu)^2}{\pi^3} - \frac{2(1-\mu)^3}{\pi^6} \geq (1-\mu)^2\left(\frac{1}{\pi^3} - \frac{2}{\pi^6}\right) \geq \frac{3(1-\mu)^2}{100}.$$

□

We then readily obtain Theorem 2 simply by combining Propositions 5 and 8 and using the fact that if $\|A\|_2 \leq 1$, then $0 \leq h_\infty(A) \leq h_2(A) \leq 1$. A leading factor of 2 coming from Proposition 8 can be removed. More generally, any bound of the form $|\mathrm{per}(A)| \leq \|A\|^n \exp[-nF(h_\infty) + o(n)]$ can immediately be improved to $|\mathrm{per}(A)| \leq \|A\|^n \exp[-nF(h_\infty)]$ by a simple amplification trick (considering a block diagonal matrix consisting of many copies of the matrix A).

3.4 Proof of Theorem 3 (better results for $\mathbb{K} = \mathbb{R}$)

For matrices over $\mathbb{R}$, our general strategy is the same as before, but we first partition the rows of A into those that contain 'big' entries and those that do not. We show that the contribution due to rows with large entries has small variance, and although the rows without large entries may each contribute something of high variance, we benefit from the fact that there simply aren't that many such rows. In this way, we are able to obtain better concentration of $\|AX\|_1$ about its mean, which in turn gives a better bound on $\mathrm{per}(A)$.

We are not sure exactly how to adapt this argument when $\mathbb{K} = \mathbb{C}$, although we admittedly didn't try very hard to do so. We feel confident (especially in light of Theorem 3) that Theorem 2 can be improved, but we do not think that Theorem 3 is best possible either (which is why we haven't worried so much about extending it to $\mathbb{K} = \mathbb{C}$). See Section 5.2 for a discussion of several related conjectures (some perhaps more true than others) and open problems.

Set-up for the real-valued case

As in Section 3.2, we let A be an $n \times n$ matrix over $\mathbb{R}$ with $\|A\|_2 \leq 1$. Define $t = 1 - h_\infty(A)$. Then to prove Theorem 3, our goal is to show

$$|\mathrm{per}(A)| \leq (n+6)\exp[-\sqrt{nt}/400].$$

Let $\varepsilon > 0$ and $1/10 > \lambda > 0$ be parameters to be determined (we will end up choosing $\varepsilon = t/10$ and $\lambda = 64/\sqrt{nt}$). We now partition the rows of A into "big rows" (those containing an element of absolute value at least $1 - \lambda$) and "small rows" (the rest). Suppose there are b big rows and $l = n - b$ small rows. Recall that because $\|A\|_2 \leq 1$, each row and column of A has l_2-norm at most 1. Thus, 'large' entries (those of absolute

value at least $1-\lambda$) must appear in different rows and columns. By multiplying A by appropriate permutation matrices and the appropriate ± 1-diagonal matrix (which changes neither the norm, nor the absolute value of the permanent, nor the values of t, b, or l), we can assume A is of the form:

$$A = \begin{pmatrix} B \\ L \end{pmatrix},$$

where B is a $b \times n$ matrix, the (i,i)-entries of B are all positive with size at least $1-\lambda$, and all the rest of the entries in A have absolute value less than $1-\lambda$. For convenience, we will assume $b > 0$ and $l > 0$, for if not, our same argument would apply with only superficial alterations.

We recall our earlier set-up as in the complex-case (but with $X \in \mathbb{R}^n$ now uniformly distributed over $\{-1,1\}^n$). Then for all $\tilde{\mu}_B, \tilde{\mu}_L \geq 0$, we have

$$\begin{aligned} |\mathrm{per}(A)| &\leq \mathop{\mathbb{E}}_{X}\left[\left(\frac{\|AX\|_1}{n}\right)^n\right] = \mathop{\mathbb{E}}_{X}\left[\left(\frac{\|LX\|_1 + \|BX\|_1}{n}\right)^n\right] \\ &\leq \left(\frac{\tilde{\mu}_L + \tilde{\mu}_B}{n} + 2\varepsilon\right)^n + \mathbb{P}(\|LX\|_1 \geq \tilde{\mu}_L + \varepsilon n) + \mathbb{P}(\|BX\|_1 \geq \tilde{\mu}_B + \varepsilon n), \end{aligned} \tag{3.1}$$

where (as before) the last inequality is justified by the fact that the random variable within the expected value is bounded above by 1.

We choose

$$\begin{aligned} \tilde{\mu}_B &= \sum_{i=1}^{b}\left[\sqrt{\frac{2}{\pi}} + \left(1-\sqrt{\frac{2}{\pi}}\right)\|r_i\|_\infty\right] = \sum_{i=1}^{b}\left[1 - \left(1-\sqrt{\frac{2}{\pi}}\right)(1-\|r_i\|_\infty)\right], \qquad \text{and} \\ \tilde{\mu}_L &= \sum_{i>b}^{n}\left[\sqrt{\frac{2}{\pi}} + \left(1-\sqrt{\frac{2}{\pi}}\right)\|r_i\|_\infty\right] = \sum_{i>b}^{n}\left[1 - \left(1-\sqrt{\frac{2}{\pi}}\right)(1-\|r_i\|_\infty)\right], \end{aligned}$$

where (again) r_i is the i^{th} row of A (note, $\|r_i\|_\infty = b_{i,i}$ for all $i \leq b$). Then by Theorem 4 (this time with $\mathbb{K} = \mathbb{R}$), we have $\tilde{\mu}_L \geq \mathbb{E}[\|LX\|_1]$ and $\tilde{\mu}_B \geq \mathbb{E}[\|BX\|_1]$, and by the definitions

$$\frac{\tilde{\mu}_L + \tilde{\mu}_B}{n} = 1 - \left(1 - \sqrt{\frac{2}{\pi}}\right)\frac{1}{n}\sum_{i=1}^{n}\left(1 - \|r_i\|_\infty\right) = 1 - \left(1 - \sqrt{\frac{2}{\pi}}\right)t. \tag{3.2}$$

To take advantage of (3.1), we need only exhibit concentration bounds for $\|LX\|_1$ and $\|BX\|_1$.

Concentration of $\|LX\|_1$

To show concentration of $\|LX\|_1$ about its mean, we will again apply a version of Talagrand's inequality (but this time suited for the discrete distribution over $\{-1, 1\}^n$). Instead of showing the derivation of this from the corresponding general result in [30] (as we did before), we will simply cite [4], in which the following statement appears as Theorem 3.3.

Theorem 9. *Suppose M is a $k \times n$ real-valued matrix such that $\|M\vec{x}\|_1 \leq \sigma\|\vec{x}\|_2$ for all $\vec{x} \in \mathbb{R}^n$. Let $\xi \in \mathbb{R}^n$ be chosen uniformly from $\{-1, 1\}^n$, and let m be a median of $\|M\xi\|_1$. Then for all $\gamma \geq 0$, we have $\mathbb{P}(|\|M\xi\|_1 - m| > \gamma) \leq 4e^{-\gamma^2/(8\sigma^2)}$.*

Lemma 10. *With notation as before, if $\varepsilon n \geq 16\sqrt{nt\log(n)/\lambda}$, then*

$$\mathbb{P}(\|LX\|_1 \geq \tilde{\mu}_L + \varepsilon n) \leq 4\exp\left[\frac{-\varepsilon^2 n\lambda}{32t}\right].$$

Proof. Note that for all $\vec{x} \in \mathbb{R}^n$, we have $\|L\vec{x}\|_1 \leq \sqrt{l}\|L\vec{x}\|_2 \leq \sqrt{l}\|A\vec{x}\|_2 \leq \sqrt{l}\|\vec{x}\|_2$. Thus, if m is a median of $\|LX\|_1$, then by Theorem 9, we have

$$\mathbb{P}(|\|LX\|_1 - m| > \gamma) \leq 4e^{-\gamma^2/(8l)}. \tag{3.3}$$

From this, we see that $\|LX\|_1$ is tightly concentrated about its *median*. However, this

also implies

$$m \leq \mathbb{E}[\|LX\|_1] + 8\sqrt{l \log n}, \tag{3.4}$$

since otherwise, we would have

$$\begin{aligned}
\mathbb{E}[\|LX\|_1] &\geq \Big(\mathbb{E}[\|LX\|_1] + 4\sqrt{l \log n}\Big) \cdot \mathbb{P}\Big(|\|LX\|_1 - m| \leq 4\sqrt{l \log n}\Big) \\
&\geq \Big(\mathbb{E}[\|LX\|_1] + 4\sqrt{l \log n}\Big) \cdot (1 - 4/n^2) \\
&= \mathbb{E}[\|LX\|_1] + 4\sqrt{l \log n} - \Big(\mathbb{E}[\|LX\|_1] + 4\sqrt{l \log n}\Big) \cdot 4/n^2.
\end{aligned}$$

And subtracting $\mathbb{E}[\|LX\|_1]$ from both sides and rearranging, we would obtain

$$n^2 \leq 4 + \frac{\mathbb{E}[\|LX\|_1]}{\sqrt{l \log n}} \leq 4 + \frac{n}{\sqrt{\log n}},$$

which is a contradiction if $n > 2$ (whereas for $n \leq 2$, the desired bound on m is implied by $m \leq n$ [not that it matters]). Therefore, appealing to (3.4), we have

$$\mathbb{P}(\|LX\|_1 \geq \tilde{\mu}_L + \varepsilon n) \leq \mathbb{P}(\|LX\|_1 \geq \mathbb{E}[\|LX\|_1] + \varepsilon n) \leq \mathbb{P}\Big(\|LX\|_1 \geq m + \varepsilon n - 8\sqrt{l \log n}\Big).$$

Furthermore, if $\varepsilon n \geq 16\sqrt{l \log n}$, then we can combine this with (3.3) to obtain

$$\text{if } \varepsilon n \geq 16\sqrt{l \log n}, \text{ then} \qquad \mathbb{P}(\|LX\|_1 \geq \tilde{\mu}_L + \varepsilon n) \leq 4 \exp\left[\frac{-\varepsilon^2 n^2}{32l}\right]. \tag{3.5}$$

Finally, since $nt \geq \sum_{i=b+1}^{n}(1 - \|r_i\|_\infty) \geq l\lambda$, we know $l \leq nt/\lambda$, completing the proof by (3.5). □

Concentration of $\|BX\|_1$

We now focus on getting an upper bound on $\mathbb{P}(\|BX\|_1 \geq \tilde{\mu}_B + \varepsilon n)$. We first recall the following classical concentration result.

Proposition 11 (Hoeffding's inequality)**.** *Let $a_1, \ldots, a_k$ be real numbers (not all of which are 0), and let $\xi_1, \xi_2, \ldots, \xi_k$ be independent each distributed uniformly on $\{-1, 1\}$.*

Then for all $\gamma \geq 0$,

$$\mathbb{P}\left(\sum_{i=1}^{k} a_i \xi_i \geq \gamma\right) \leq \exp\left[\frac{-\gamma^2}{2\sum_{i=1}^{k} a_i^2}\right].$$

Let $\tilde{B} = \begin{pmatrix} B \\ 0 \end{pmatrix}$ be the $n \times n$ matrix whose first b rows are given by B and the rest are 0. Our key step here is replacing $\|BX\|_1$ with $\langle X, \tilde{B}X\rangle$, via the following lemma[1].

Lemma 12. *With notation as before, if* $\lambda < 0.1$ *then*

$$\mathbb{P}(\|BX\|_1 \geq \tilde{\mu}_B + \varepsilon n) \leq \mathbb{P}(\langle X, \tilde{B}X\rangle \geq \tilde{\mu}_B + \varepsilon n) + ne^{-1/(5\lambda)}.$$

Proof. It suffices to show $\mathbb{P}(\|BX\|_1 \neq \langle X, \tilde{B}X\rangle) \leq ne^{-1/(5\lambda)}$. The idea is that since each row of B is dominated by a single large entry (namely $b_{i,i}$), each entry of BX is a random sum dominated by a single large term (namely $X_i b_{i,i}$). Thus, it is very unlikely that any entry of BX would have a different sign than $X_i b_{i,i}$. This is made rigorous as follows.

Recall that we ordered the columns of B so that the (i,i)-entry is the largest in its row, and that $b_{i,i} \geq 1 - \lambda$. Letting Y_i be the i^{th} coordinate of BX, we have, by a simple union bound,

$$\mathbb{P}(\|BX\|_1 \neq \langle X, \tilde{B}X\rangle) \leq \sum_{i=1}^{b} \mathbb{P}(|Y_i| \neq X_i Y_i) = \sum_{i=1}^{b} \mathbb{P}(X_i Y_i < 0) = \sum_{i=1}^{b} \mathbb{P}\left(\sum_{j=1}^{n} X_i X_j b_{i,j} < 0\right)$$

Using the fact that for any given i, the random vector $(X_i X_j)_{j \neq i}$ has the same joint distribution as $(X_j)_{j \neq i}$ (and that $X_i^2 = 1$), we obtain by Proposition 11

$$\sum_{i=1}^{b} \mathbb{P}\left(\sum_{j=1}^{n} X_i X_j b_{i,j} < 0\right) = \sum_{i=1}^{b} \mathbb{P}\left(b_{i,i} < \sum_{j \neq i}^{n} X_j b_{i,j}\right) \leq \sum_{i=1}^{b} \exp\left[\frac{-b_{i,i}^2}{2\sum_{i \neq j} b_{i,j}^2}\right].$$

[1]Extending this step is the main obstacle to applying the present argument when $\mathbb{K} = \mathbb{C}$.

Since $b_{i,i} \geq 1 - \lambda$ and $\sum_j b_{i,j}^2 \leq 1$, this in turn is bounded by

$$\sum_{i=1}^{b} \exp\left[\frac{-b_{i,i}^2}{2\sum_{i\neq j} b_{i,j}^2}\right] \leq n \exp\left[\frac{-(1-\lambda)^2}{2(1-(1-\lambda)^2)}\right] \leq ne^{-1/(5\lambda)},$$

where the last inequality is justified because $0 < \lambda < 0.1$. □

We can now exploit the fact that $\langle X, \tilde{B}X\rangle$ is a degree two polynomial over $\{-1,1\}^n$, allowing us to use any of a variety of concentration inequalities. We will use an inequality of Bonami [9], which was the first *hypercontractivity inequality* of its type. A detailed exposition of such results can be found in chapter 9 of O'Donnell's book [36], and a comparison of this to more recent polynomial concentration inequalities can be found in [41].

Theorem 13 (Bonami [9], 1970)**.** *Let $F : \mathbb{R}^n \to \mathbb{R}$ be a degree k polynomial, and consider the random variable $Z = F(\xi_1, \xi_2, \ldots, \xi_n)$, where the ξ_i are independent with each distributed uniformly over $\{-1, 1\}$. Then for all $q \geq 2$, we have $\mathbb{E}[|Z|^q] \leq \left((q-1)^k \mathbb{E}[Z^2]\right)^{q/2}$.*

Lemma 14. *With notation as before, if $\varepsilon n \geq 4e\sqrt{nt}$, then*

$$\mathbb{P}(\langle X, \tilde{B}X\rangle \geq \tilde{\mu}_B + \varepsilon n) \leq \exp\left(\frac{-\varepsilon n}{2e\sqrt{nt}}\right).$$

Proof. For $\vec{x} \in \mathbb{R}^n$, define $F(x_1, x_2, \ldots, x_n) = \langle \vec{x}, \tilde{B}\vec{x}\rangle - \sum_{i=1}^{b} b_{i,i}$, and define the random variable $Z = F(X_1, \ldots, X_n)$. Then $\mathbb{P}(\langle X, \tilde{B}X\rangle \geq \tilde{\mu}_B + \varepsilon n) \leq \mathbb{P}(Z \geq \varepsilon n)$, since[2] $\tilde{\mu}_B \geq \sum_{i\leq b} b_{i,i}$. Now $F(x_1, x_2, \ldots, x_n)$ is a degree 2 polynomial, and moreover, by expanding out the sums and using the fact that terms such as $\mathbb{E}[X_i X_j]$ vanish when

[2]In fact, we could have simply taken $\tilde{\mu}_B = \sum_{i\leq b} b_{i,i}$, but we chose instead to define it similarly to $\tilde{\mu}_L$, a change which only affects the constants in our end result.

$i \neq j$, we obtain

$$\begin{aligned}
\mathbb{E}[Z^2] &= \mathbb{E}\left[\left(\sum_{i=1}^{b}\left[-b_{i,i} + \sum_{j=1}^{b} X_i X_j b_{i,j}\right] + \sum_{i=1}^{b}\sum_{j=b+1}^{n} X_i X_j b_{i,j}\right)^2\right] \\
&= \mathbb{E}\left[\left(\sum_{i=1}^{b}\left[-b_{i,i} + \sum_{j=1}^{b} X_i X_j b_{i,j}\right]\right)^2\right] + \mathbb{E}\left[\left(\sum_{i=1}^{b}\sum_{j=b+1}^{n} X_i X_j b_{i,j}\right)^2\right] \\
&= \sum_{i=1}^{b}\sum_{j<i}(b_{i,j}+b_{j,i})^2 + \sum_{i=1}^{b}\sum_{j=b+1}^{n} b_{i,j}^2 \leq 2\sum_{i=1}^{b}\sum_{j<i}(b_{i,j}^2 + b_{j,i}^2) + 2\sum_{i=1}^{b}\sum_{j=b+1}^{n} b_{i,j}^2 \\
&= 2\sum_{i=1}^{b}\left(-b_{i,i}^2 + \sum_{j=1}^{n} b_{i,j}^2\right) \leq 2\sum_{i=1}^{b}(1-b_{i,i}^2) \leq 4\sum_{i=1}^{b}(1-b_{i,i}) \leq 4nt.
\end{aligned}$$

Applying Theorem 13 with $q = \varepsilon n/(2e\sqrt{nt})$—which is valid since by hypothesis this ratio is at least 2—together with Markov's inequality, we obtain

$$\mathbb{P}(Z \geq \varepsilon n) \leq \mathbb{P}(|Z|^q \geq (\varepsilon n)^q) \leq \frac{\mathbb{E}[|Z|^q]}{(\varepsilon n)^q} \leq \left(\frac{(q-1)2\sqrt{nt}}{\varepsilon n}\right)^q \leq \exp\left(\frac{-\varepsilon n}{2e\sqrt{nt}}\right). \quad \square$$

Finishing the proof for $\mathbb{K} = \mathbb{R}$

We now need to pick ε and λ to optimize the tradeoffs between our various upper bounds. We need the assumptions of Lemmas 10, 12, and 14—namely (i) $\varepsilon n \geq 16\sqrt{nt\log(n)/\lambda}$, (ii) $\lambda < 0.1$, and (iii) $\varepsilon n \geq 4e\sqrt{nt}$—in which case we can combine these lemmas with (3.1) and (3.2) to obtain

$$\begin{aligned}
|\mathrm{per}(A)| &\leq \left(2\varepsilon + \frac{\tilde{\mu}_L + \tilde{\mu}_B}{n}\right)^n + \mathbb{P}(\|LX\|_1 \geq \tilde{\mu}_L + \varepsilon n) + \mathbb{P}(\|BX\|_1 \geq \tilde{\mu}_B + \varepsilon n) \\
&\leq \left(2\varepsilon + 1 - \left(1 - \sqrt{\frac{2}{\pi}}\right)t\right)^n + 4\exp\left[\frac{-\varepsilon^2 n\lambda}{32t}\right] + ne^{-1/(5\lambda)} + \exp\left(\frac{-\varepsilon n}{2e\sqrt{nt}}\right).
\end{aligned}$$

We will take $\varepsilon = t/10$ and $\lambda = 64/\sqrt{nt}$, for which we claim that conditions (i), (ii), and (iii) are satisfied. Note that since our goal is to show $|\mathrm{per}(A)| \leq (n+6)\exp[-\sqrt{nt}/400]$, we may assume $\sqrt{nt}/\log(n+6) \geq 400$ (or the bound we are trying for is worse than the trivial bound of 1) (of course, in any case we are really more interested in large n).

Notice that with ε and λ as above:

(i) $\varepsilon n \geq 16\sqrt{nt\log(n)/\lambda}$ is equivalent to $\sqrt{nt} \geq 400\log n$;

(ii) $\lambda < 0.1$ is equivalent to $\sqrt{nt} > 640$; and

(iii) $\varepsilon n \geq 4e\sqrt{nt}$ is equivalent to $\sqrt{nt} \geq 40e$.

Thus, these choices of λ and ε allow us to appeal to the aforementioned results, obtaining

$$\begin{aligned}
|\mathrm{per}(A)| &\leq \left(2\varepsilon + 1 - \left(1 - \sqrt{\frac{2}{\pi}}\right)t\right)^n + 4\exp\left[\frac{-\varepsilon^2 n\lambda}{32t}\right] + ne^{-1/(5\lambda)} + \exp\left(\frac{-\varepsilon n}{2e\sqrt{nt}}\right) \\
&\leq \exp\left[-nt\left(1 - \sqrt{2/\pi} - 0.2\right)\right] + 4\exp\left[\frac{-\sqrt{nt}}{50}\right] + n\exp\left[-\frac{\sqrt{nt}}{320}\right] + \exp\left[\frac{-\sqrt{nt}}{20e}\right] \\
&\leq (n+6)\exp\left[\frac{-\sqrt{nt}}{400}\right],
\end{aligned}$$

which completes the proof of Theorem 3.

3.5 Conclusion

Our most natural open question concerns the optimality of our main results. Namely, a proof of Conjecture 1 as stated in Section 3.1 would be very interesting. The main barrier preventing us from proving this conjecture is our reliance on Talagrand's inequality. For $\mathbb{K} = \mathbb{R}$, we partially mitigated the cost of using this inequality via Lemma 10, but the application of Theorem 9 was still a crucial (though not the only) bottleneck. Our argument could conceivably be pushed further either by a more careful analysis that better uses (3.5) or by a more nuanced argument that splits the matrix A into more than two pieces.

One could also try to avoid using Talagrand's inequality altogether. It is possible that some stronger inequality could replace it (by taking advantage of some aspects particular to our situation), but a more likely "quick fix" of this sort would be a more direct estimate of $\mathbb{E}[(\|AX\|_1/n)^n]$ (in the real case, AX is simply a vector-valued Rademacher

sum, which is a well-studied random variable). On the other hand, it could be that the convexity bounds on $G_X(A)$ already give away too much to recover anything stronger than what we have.

An entirely different approach would be to determine among matrices with given norm and h_∞, which ones maximize $|\mathrm{per}(A)|$ (it does not seem impossible that this maximum is always attained by a circulant matrix with all real entries). A characterization of these extremal matrices would certainly be very appealing, and one might hope that thinking along these lines would suggest a more combinatorial approach.

As far as Question B is concerned, we feel that there is still more to be said beyond the present results. Namely, our results only provide a necessary condition for a matrix to have a large permanent (i.e., h_∞ must be large). But there is no clean converse to this statement; consider for example a diagonal matrix with most of its diagonal entries equal to 1 except for one of them equal to 0 (this has large h_∞ and permanent 0). To continue the spirit of the question, we state the following variation of Question B (essentially echoing a question of [2]):

Problem B′:

Find a (deterministic) polynomial-time algorithm that takes an $n \times n$ matrix A of norm 1 and decides whether $|\mathrm{per}(A)| < n^{-100}$ or $|\mathrm{per}(A)| > n^{-10}$ (with the promise that the input matrix will satisfy one of these inequalities).

We attempted this along the following lines: "if the matrix has large permanent, it must have many rows each of which is dominated by a single large entry. If the matrix is of this form, then [heuristic] hopefully that means the permanent is dominated by terms that use at least most of these large entries. Since there are so many large entries, we

can efficiently compute the exact contribution of these dominant terms." However, our current results do not allow us to conclude that there are enough rows with large entries (we would like all but about $\log n$ of the rows but are limited to all but about $\log^2 n$ when $\mathbb{K} = \mathbb{R}$ and $\sqrt{n \log n}$ when $\mathbb{K} = \mathbb{C}$). And in fact, even if we could improve our result to the conjectured (and best possible) bound mentioned above, we still do not quite see how to make this heuristic argument yield a polynomial-time algorithm. We should note that Gurvits [19] found a *randomized* algorithm accomplishing the goal of Problem B′, and in the deterministic setting, progress towards Problem B′ was made in [2] which gives an algorithm in the case that the entries of A are non-negative.

Further remarks

- We note that there is a lot of freedom in choosing the random variable $X \in \mathbb{K}^n$ for $G_X(A)$ (X just needs to have independent components each satisfying $\mathbb{E}[X_i] = 0$ and $\mathbb{E}[|X_i|^2] = 1$). For example, when $\mathbb{K} = \mathbb{R}$, it is tempting to replace $X \in \mathbb{R}^n$ with an n-dimensional Gaussian and bound the permanent by something like

$$|\mathrm{per}(A)| = \left|\mathbb{E}\left[\prod_i X_i Y_i\right]\right| \leq \mathbb{E}\left[\prod_i |X_i Y_i|\right] \leq \mathbb{E}\left[\left(\frac{1}{n}\sum_i |X_i Y_i|\right)^n\right].$$

  But even if A is the identity matrix this is already (exponentially) larger than 1, which illustrates the difficulty with this approach.

- Via an entirely different method, we were also able to get an upper bound on the permanent for matrices having only non-negative real entries by appealing to the results of [20]. Unfortunately, the bound we obtained is strictly weaker than the results of the present paper, so it is omitted.

Acknowledgement: We thank Hoi Nguyen for introducing us to this problem and sharing [34].

Chapter 4

Robust Positioning Patterns

4.1 Introduction

A 1-dimensional "positioning pattern" is a sequence of N symbols from some alphabet, with the property that any window of n consecutive elements from the sequence uniquely determines the position of the window. Similarly, a 2-dimensional "positioning pattern" is an $N \times N$ matrix of symbols from some alphabet, with the property that any $n \times n$ (contiguous) window of elements uniquely determines the position of the window. Positioning patterns have been classically studied in combinatorics under various names: de Bruijn sequences, perfect maps, pseudorandom sequences and arrays, etc. In recent years, these objects have found a number of useful real-world applications, such as robot localization [40], camera localization [42], the Echo Smartpen, and smart stylus' [1].

To see the utility of positioning patterns, let us briefly describe the application from [1]. We are given a display device (such as a monitor or a laptop screen) whose sole capability is display (in particular, it cannot detect touch or the presence of a stylus/pen). The smart stylus from [1] is based on a combination of software and hardware, and converts any such display into one which can take input from the stylus. The hardware component is a pen with a small camera at its nib, which when brought near the screen of the display device can view a small $n \times n$ window of the screen. The software component sets the lower order bits of the color attribute for each pixel on the screen according to a positioning pattern. This ensures that the lower order bits for any $n \times n$ window of the screen uniquely determines the position of the window: thus one can use the image from the pen camera to determine the location of the pen, and this is just as good as having a display that can detect the location of an associated stylus.

A "robust positioning pattern" is a sequence/matrix of symbols, which allows such position determination by reading a small window from the pattern *even if some errors occur while reading the small window.* Concretely, the sequence/matrix has the property that the contents of the different windows should be far apart from each other in Hamming distance. Algorithmically, we would like to be able to *efficiently decode* the position of the window, given the corrupted contents of a window.

We are interested in constructing such robust positioning patterns and designing associated decoding algorithms for them. Our motivation comes from both practice and theory. Firstly, these problems are naturally motivated by the applications of positioning patterns given above, which rely on physical devices and are thus prone to error. Secondly, this topic presents interesting combinatorial and algorithmic challenges at the confluence of error-correcting codes and combinatorial sequence design, both of which are extensively studied and have highly developed theories.

Our main results give explicit constructions of robust positioning patterns, along with associated decoding algorithms. These constructions are the first to achieve constant rate while being robust to a constant fraction of errors, and are also the first to achieve robustness to a constant number of errors with redundancy within a constant factor of optimal.

4.1.1 Results

We begin with the 1 dimensional setting.

Let $\sigma \in \Sigma^N$ be a string. We let $\sigma[i, j)$ denote the substring $\sigma_i \sigma_{i+1} \dots \sigma_{j-1}$. We will be interested in substrings of the form $\sigma[i, i+n)$, which we will also call the "windows of length n". We define the window-n distance of σ to equal the minimum, over distinct $i, j \in [N - n + 1]$ of

$$\Delta(\sigma[i, i+n), \sigma[j, j+n)),$$

where Δ denotes the Hamming distance.

The basic combinatorial problem here is to determine the length of the longest string with window-n distance at least d. The basic algorithmic problems here are: (1)

Encoding: to explicitly construct a long string with window-n distance at least d, and (2) **Decoding:** for this sequence, given a "received string" $r \in \Sigma^n$ which is within distance e of some window $\sigma[i, i+n-1]$, to find i.

It is sometimes convenient to use the following terminology. Define the window-n rate of σ to equal $\frac{\log N}{n \log |\Sigma|}$. Define the window-$n$ relative distance to be the window-n distance divided by n.

It is clear that the length N of any sequence with window-n distance d cannot be more than the size of the largest error-correcting code $C \subseteq \Sigma^n$ with minimum distance d (since the n-windows of the sequence form such an error-correcting code). Thus we have the following rough upper bounds on the length of such a sequence:

1. for $d = \delta n$ (with $\delta > 0$ a constant), we have $N \leq |\Sigma|^{n(1-f(\delta))}$, for some function $f(\delta)$ that goes to 0 as δ goes to 0,
2. for $d = O(1)$, $|\Sigma|$ large, we have $N \leq \frac{|\Sigma|^n}{|\Sigma|^{\Omega(d)}}$,
3. for $d = O(1)$, $|\Sigma| = 2$, we have $N \leq \frac{2^n}{n^{\Omega(d)}}$.

A simple application of the Lovász Local Lemma (suggested to us by Nathaniel Shar) shows that the above upper bounds on N are essentially tight (nonconstructively); there exist strings in Σ^N matching the above bounds. A very nice result of Kumar and Wei [29] shows that a random irreducible Linear Feedback Shift Register Sequence matches the third of the above upper bounds with high probability (this result holds for all $d \leq \sqrt{n}$). It is natural to ask if we can match these bounds with explicit constructions and efficient decoding algorithms.

Our main results for 1-dimensional sequences give explicit constructions and efficient decoding algorithms for sequences, essentially matching the above parameters[1].

Theorem 12 (1-Dimension, Large Σ, constant δ)**.** *There exists an infinite sequence of n and alphabets Σ_n (with $|\Sigma_n| \leq O(n)$), such that for every $R \in (0,1)$, there is a sequence $\sigma \in \Sigma^{N_n}$ with:*

[1] We require a widely believed number theoretic conjecture to attain the third set of parameters.

1. *the rate of* σ *is at least* R,

2. *the window-*n *relative distance of* σ *is at least* $\max(1-3R,(1-R)/3)-o(1)$,

3. *the* i*'th coordinate of* σ *can be computed in time* $\mathrm{poly}(n)$,

4. n*-windows of* σ *can be decoded from a constant fraction of errors in* $\mathrm{poly}(n)$ *time.*

This theorem follows from Theorem 16.

Theorem 13 (1-Dimension, $|\Sigma|=2$, constant δ)**.** *There exists an infinite sequence of* n *such that for every* $R\in(0,1)$*, there is a sequence* $\sigma\in\{0,1\}^{N_n}$ *with:*

1. *the rate of* σ *is at least* R,

2. *the window-*n *relative distance of* σ *is at least* $h(R)-o(1)$*, (where* $h(R)>0$ *),*

3. *the* i*'th coordinate of* σ *can be computed in time* $\mathrm{poly}(n)$,

4. n*-windows of* σ *can be decoded from a constant fraction of errors in* $\mathrm{poly}(n)$ *time.*

This theorem follows from Corollary 5.

Theorem 14 (1-D, Large Σ, constant distance)**.** *There exists an infinite sequence of* n *and alphabets* Σ_n *(with* $|\Sigma_n|=O(n)$*), such that for every constant* d*, there is a sequence* $\sigma\in\Sigma_n^{N_n}$ *with:*

1. $N_n\geq\frac{|\Sigma_n|^n}{|\Sigma_n|^{O(d)}}$,

2. *the window-*n *distance of* σ *is at least* d,

3. *the* i*'th coordinate of* σ *can be computed in time* $\mathrm{poly}(n)$,

4. n*-windows of* σ *can be decoded from* $\Omega(d)$ *errors in* $\mathrm{poly}(n)$ *time.*

This theorem follows from Theorem 16.

Our result for constant distance binary codes depends on the existence of suitable Mersenne-like primes. Such primes are widely believed to exist based on standard number theoretic heuristics.

Conjecture C: There exists a constant c and infinitely many n such that there exists a prime between $2^n - c \cdot n$ and $2^n - 1$.

Note that this conjecture would be implied by the existence of infinitely many Mersenne primes.

Theorem 15 (1-D, $|\Sigma| = 2$, constant distance)**.** *Assume conjecture C. There exists an infinite sequence of n such that for every constant d, there is a sequence $\sigma \in \{0,1\}^{N_n}$ with:*

1. $N_n \geq \frac{2^n}{n^{O(d)}}$,
2. *the window-n distance of σ is at least d,*
3. *the i'th coordinate of σ can be computed in time* $\mathrm{poly}(n)$,
4. *n-windows of σ can be decoded from $\Omega(d)$ errors in* $\mathrm{poly}(n)$ *time.*

Our large alphabet constructions all use properties of polynomial-based error-correcting codes (especially using their cyclicity when the evaluation set is special), in conjunction with Gray codes.

Our binary constructions are based on a new "augmented" code concatenation scheme. This new scheme is based on two ideas: (1) using a low-autocorrelation sequence as a "marker", and (2) designing an inner code for the concatenation all of whose codewords are far away from all substrings of the marker.

4.1.2 Related work

The classical notions of de Bruijn sequences and M sequences are the basic examples of positioning patterns. The two-dimensional "de Bruijn torus" is the natural generalization to two dimensions, and were first constructed by [31]. These found applications in various practical settings for localization / positioning [42, 40, 1].

Efficiently decodable de Bruijn sequences and tori, which are extremely natural for the positioning applications, were given by [33, 32, 11, 12].

The requirement for robustness in positioning patterns is very natural for real-world applications where the positioning pattern is "measured" by a physical device. Indeed,

several applied works encountered these problems (in applications such as wireless device localization, and markers for "augmented reality") [28, 24, 22], and proposed ad hoc solutions.

On the theoretical side, there were some important papers on robust positioning such as [29, 10, 21]. [29] showed that a random linear feedback shift register sequence provides a nearly optimal tradeoff between the window-n distance and the length of the sequence (in the regime where the number of errors is less than $\sqrt{n}$). [10] gave constructions of $N \times N$ 2-dimensional robust positioning patterns (for $n \times n$ windows) with $N = 2^{O(n)}$ (while there exist such patterns with $N = 2^{O(n^2)}$).

4.2 Preliminaries and Notation for 1 Dimensional Robust Positioning Sequences

Some basic preliminaries. Throughout this chapter we will use $[n]$ to refer to the first n natural numbers with 0 *included*. That is

$$[n] := \{0, 1, 2, \ldots, n-1\}$$

We need some notation for expressing and accessing values of sequences.

Definition 7. Given a sequence $S := (s_1, \ldots, s_N)$ we define $S[i]$ to be the i^{th} entry of S, i.e. $S[i] = s_i$. Further if $I := (i_0, i_2, \ldots, i_n)$ then we define

$$S[I] := (S[i_0],\ S[i_1], \ldots,\ S[i_n])$$

Furthermore for our robust positioning patterns we will denote them as a sequence of length N, however we will frequently wish to consider the coordinates cyclically. To that extent for a sequence S of length N we will say that for any integer m even if m is negative or $> N$ we have that $S(m) := S(m \mod N)$.

Also, we will frequently need to refer to intervals of integers, and so we use the notation $[m_1, m_2]$.

Definition 8. Given $m_1 < m_2$ define $[m_1, m_2] := (m_1, m_1 + 1 \ldots, m_2)$. We will use

square brackets for inclusive and open brackets for open boundaries much like intervals in $\mathbb{R}$. For example $(m_1, m_2] := (m_1 + 1, \ldots, m_2)$. Sometimes when more compact notation is needed, we will use $\langle m \rangle_n$ to denote $[m, m + n)$.

For sequences $S_1, S_2 \in \Sigma^n$, we denote their Hamming distance by $\Delta(S_1, S_2)$, and denote their *agreement* by $\text{agree}(S_1, S_2)$. Thus $\Delta(S_1, S_2) + \text{agree}(S_1, S_2) = n$.

For sequences $S_1, S_2, \ldots S_n$, we will denote their concatenation by $(S_1, S_2, \ldots, S_n)$.

We will frequently want to rotate sequences, cyclically permuting their entries. We give special notation to this operation. We define $\rho : \Sigma^n \mapsto \Sigma^n$ be the coordinate rotation map $\rho((x_1, \ldots, x_n)) = (x_2, x_3, \ldots, x_n, x_1)$.

The following definitions caption the relationship of two sequences being almost the same (i.e. differ in only one position) after a rotation.

Definition 9. Given two sequences $S_1, S_2 \in \Sigma^n$ we write $S_1 \sim S_2$ if there exists some rotation ρ^j such that $\dot{(}\rho^j S_1, S_2) \leq 1$. Similarly if $T_1 \in \Sigma^m$ and $T_2 \in \Sigma^n$ where $m \leq n$ we write $T_1 \lesssim T_2$ if there is some i such that $T_1 \sim T_2[i, i + m - 1]$.

Finally we will need a way to quantify the error correcting properties of the sequences we create. We borrow the terms rate, distance and relative distance from Coding Theory as follows:

Definition 10. Given a q-ary sequence S of length N and an integer n (the window length), we say that the rate of S is

$$R(S) := R := \frac{\log_q(N)}{n}.$$

We define the distance of S to be $\min_{0 \leq i \neq j \leq N} \dot{(}S[i, i+n), S[j, j+n))$. Finally we define the relative distance of S to be

$$\delta_S := \min_{0 \leq i \neq j \leq N} \frac{\dot{(}S[i, i + n), S[j, j + n))}{n}$$

4.3 Robust Positioning Sequences Over Large Alphabets

4.3.1 Overview

In this section we'll show an explicit construction of a robust positioning pattern over large alphabets which is "good" in the sense that it will acheive constant fraction distance and constant rate. Further the construction is capable of acheiving any rate between 0 and 1, and any relative distance between 0 and 1 as well.

The positioning pattern itself is acheived by listing consecutively the entries of a Reed-Solomon code, which has been suitably pruned so that no two codewords are rotations of one another. Further we need to list the remaining codewords in a specific order, namely in such a way that their prefixes form a q-ary gray code of length $\deg(p)$ (See Figure 4.1). This ordering will ensure that windows which are slightly misaligned and which see the end of some codeword C_i and the gray code prefix of its successor codeword C_{i+1} are tricked into believing they instead see a rotation of the first codeword $\rho^j(C_i)$ (see Remark 1 and the accompanying diagram Figure 4.2). But since we ensured that our codewords were not rotations of eachother, any such rotated codewords will be unique and distant and will allow us to recover the codeword C_i, while the rotation will tell us exactly the location of the current window.

One more small remark is that ocasionally our window will not only see the gray code bits of the subsequent Reed-Solomon codeword, but in such cases we may break the window down into two subwindows which are small windows of rotated codewords, and at a cost of losing some distance from the original Reed-Solomon code, we will be able to decode one of these shortened subwindows.

4.3.2 Definitions and Construction

First in order to explicitly write down a Reed-Solomon codeword, we need to fix an ordering of our underlying base field $\mathbb{F}$. So to that aim fix g a generator of $\mathbb{F}_q^\times$, and we will order the elments of $\mathbb{F}^\times$ as subsequent powers of g.

Definition 11. Fix $n := |\mathbb{F}^\times| = q - 1$. Given the function $f : [\mathbb{F}_q] \to [\mathbb{F}_q]$ we define the

word $C^f := C(f) \in \mathbb{F}_q^n$ by setting

$$C(f) := (f(g^0),\ f(g^1),\ \ldots,\ f(g^{n-1}))$$

Let $\mathcal{C} := \{C(f) \text{ s.t. } f \in \mathcal{F}\}$.

Then let Σ be a q-ary gray code of length k. Our robust positioning pattern will be built out of blocks consisting of encodings of a certain family of polynomials $\mathcal{F}$ given by interpolating a polynomial with of degree k with prefix given by some $\sigma \in \Sigma$, and with constant term 0, and coefficient of X fixed to be 1 (these last two properties will ensure that the family of polynomials define words which are not rotations of one another).

Definition 12. Given $\sigma \in \mathbb{F}^k$ let $f^\sigma(X) \in \mathbb{F}[X]$ be the unique interpolating polynomial of degree $k+1$ so that:

- $\text{coeff}_X(f^\sigma) = 1$
- $\text{coeff}_1(f^\sigma) = 0$
- for each $i \in [0, k)$, $f^\sigma(g^i) = \sigma_i$

Further, define $\mathcal{F} := \{f^\sigma \text{ s.t. } \sigma \in \mathbb{F}^k\}$.

The first two conditions above are equivalent to saying $f^\sigma(X) := Xh^\sigma(X)$ where

- $h^\sigma(0) = 1$
- for each $i \in [0, k-1]$, $h^\sigma(g^i) = \sigma_i g^{-i}$

Given a polynomial we will encode it in the following manner.

Definition 13. Given the function $f : [\mathbb{F}_q] \to [\mathbb{F}_q]$ we define the word $C^f := C(f) \in \mathbb{F}_q^n$ by setting

$$C(f) := (f(g^0),\ f(g^1),\ \ldots,\ f(g^{q-2}))$$

Let $\mathcal{C} := \{C(f) \text{ s.t. } f \in \mathcal{F}\}$.

We now are ready to present the definition of the robust positioning pattern we will study:

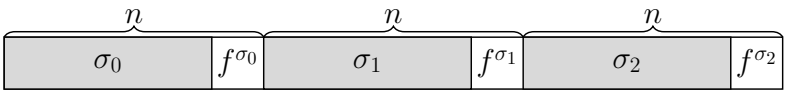

Figure 4.1: A view of the beginning of the robust positioning pattern S_Σ constructed in Definition 14. σ_i represents the i^{th} word in the Gray code Σ, and f^{σ_i} is the rest of the appropriate interpolated polynomial word so that the codeword $C(f^{\sigma_i})$ has σ_i as a prefix.

Definition 14. Let $\Sigma = \sigma^0, \sigma^1, \ldots, \sigma^{q^k-1}$ be a q-ary gray code of window length k. For convenience of notation we will often write f^a for f^{σ_a} and C^a for $C^{f^{\sigma_a}}$. Then define the sequence S to be

$$\begin{aligned} S := S_\Sigma &:= \left(C^{f^{\sigma_1}}, C^{f^{\sigma_2}}, \ldots, C^{f^{\sigma_{q^k}}}\right) \\ &:= \left(C^0, C^1, \ldots, C^{q^k-1}\right) \end{aligned}$$

See Figure 4.1 for a depiction of part of this construction.

4.3.3 Proof of Distance of S_Σ

The goal of this section is to prove the following distance result for S.

Theorem 16. *The sequence $S := S_\Sigma = [C^{f^1}, C^{f^2}, \ldots, C^{f^{q^k}}]$ defined in Definition 14 is a q-ary sequence of rate $\frac{k+1}{q}$ and distance $\max\left(\frac{q-k}{3} - 3, q - 3k - 9\right)$ with window size $n := q - 1$.*

When considering a window $w = S[m] := S[m, m+n)$ often the most important identifying feature is $\bar{m} = m \mod n$ where $0 \leq \bar{m} < n$. This tells us which symbols correspond to Gray code entries, and which are values of the interpolated polynomial f^σ. Larger values of $\bar{m}$ indicate that the Gray code has been pushed leftward (wrapping around) in our window.

Our first observation is that when $\bar{m}$ is small, then we see almost exactly a rotation of a copy of some codeword C^a.

Observation 1. *Let w be a length n window of S_Σ, $w = S_\Sigma\langle m\rangle_n := (S(m),\ S(m+1), \ldots,\ S(m+n-1))$ where $m = an + \bar{m}$ and $0 \leq \bar{m} < k$. Then $\rho^{\bar{m}} C^a \sim w$,(i.e.*

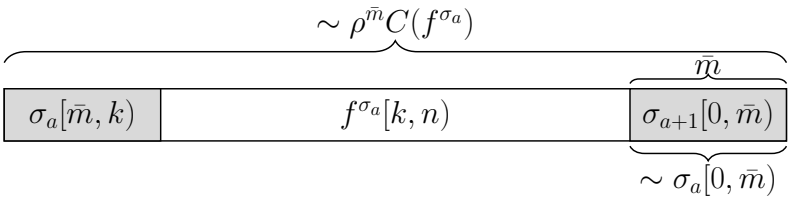

Figure 4.2: Accompanying diagram for Observation 1. Note how the size of $\bar{m}$ affects the position of the Gray code bits, and the fact that $\bar{m} < k$ is important to ensuring that the Gray code bits are a suffix.

$\Delta(w, \rho^{\bar{m}} C^a) \leq 1$*).*

Proof. For a depiction of the argument see Figure 4.2. If $m = an + \bar{m}$ then we see that

$$\begin{aligned} w = S_\Sigma[aq + \bar{m}] &= \big(C^a[\bar{m}, n),\ C^{a+1}[0, \bar{m})\big) \\ &= \big(\sigma_a[\bar{m}, k),\ C^a[k, n),\ \sigma_{a+1}[0, \bar{m})\big) \end{aligned}$$

Therefore

$$w = \rho^{\bar{m}}\big(\sigma_{a+1}[0, \bar{m}), \sigma_a[\bar{m}, k),\ C^a[k, n)\big)$$

and from the definition

$$\rho^{\bar{m}} C^a = \rho^{\bar{m}}\big(\sigma_a[0, \bar{m}), \sigma_a[\bar{m}, k),\ C^a[k, n)\big)$$

As Σ is a Gray code we have that $\Delta(\sigma_a, \sigma_{a+1}) = 1$ so comparing the above two expressions it follows immediately that $\Delta(w, \rho^{\bar{m}} C^a) \leq \Delta(\sigma_a, \sigma_{a+1}) = 1$

□

Second, we observe that when $\bar{m}$ is larger, the situation isn't as nice, but we can split the window up into two overlapping parts which do look like subwindows of codewords.

Observation 2. *Let $w = S\langle m \rangle_n$ where $m = an + \bar{m}$ and $k < \bar{m} \leq n$. If we let $x_1 = n - \bar{m}$ then $w[0, x_1 + k) \lesssim C(f^a)$ and $w[x_1, n) \lesssim C(f^{a+1})$.*

Proof. $w[0, x_1 + k - 1] \subset S\langle m - \bar{m} + k \rangle_n \sim C(f^a)$ by Observation 1. We also have that $w\langle x_1, q - 1 \rangle_n \subset S\langle m + (q - \bar{m}) \rangle_n = C(f^{a+1})$. □

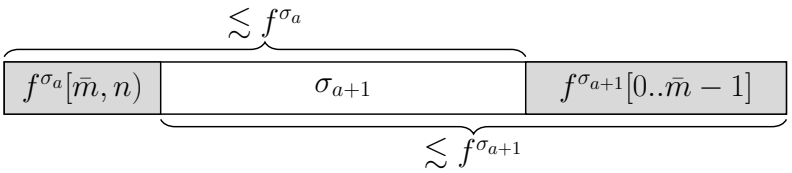

Figure 4.3: Accompanying figure for Observation 2

We combine these two observations into a single corollary.

Corollary 3. *Let $w_1 \neq w_2$ be windows of S_Σ of length $\ell \leq n$. Assume that $w_1 = S\langle m_1 \rangle_n$ and $w_2 = S\langle m_2 \rangle_n$ where for all $i = 1, 2$ we have either ($k \leq \bar{m}_i$ and $\bar{m}_i + \ell < n + k$) or ($0 \leq \bar{m}_i < k$). Then agree$(w_1, w_2) \leq \min(\ell, k + 3)$.*

Proof. Note that the congruence conditions are exactly the conditions we need to apply the above observations. If $k \leq \bar{m}_i$ and $\ell < n - \bar{m}_i + k$ then by Observation 2 $w_i \lesssim C^{a_i}$ for some a_i. In the second case if $0 \leq \bar{m}_i \leq k$ then $w_i \subset S[m_i, m_i{+}n) \lesssim \rho^{\bar{m}_i} C^{a_i}$ for some a_i by Observation 1. So by using triangle inequality, the fact that $\rho^{\bar{m}_1} C^{a_1} \neq \rho^{\bar{m}_2} C^{a_2}$, and Lemma 15 we obtain that

$$\text{agree}(w_1, w_2) \leq \min(\ell, \text{agree}(C^1, C^2) + 3) \leq \min(\ell, k + 3)$$

□

Now we are ready to begin the proof of our main theorem. The basic strategy will be as follows: Corollary 3 will allow us to break each window into two pieces, each of which is a rotation of a subwindow of a codeword from $\mathcal{C}$. Then we will break our windows up into pieces based on these subwindows (a process which will require several cases), analyze what distance and agreement bounds we can get on each piece, and then recombine our answers for the final estimate.

THEOREM 16 *Let $w_1 \neq w_2$ be windows of size q of S_Σ. Then $\Delta(w_1, w_2) \geq \max(q - 3k - 9,\ \frac{q-k}{3} - 3)$.*

Proof. Assume $w_1 = S\langle m_1 \rangle_n$ and $w_2 = S\langle m_2 \rangle_n$. Let $m_1 \equiv \bar{m}_1,\ m_2 \equiv \bar{m}_2$ where

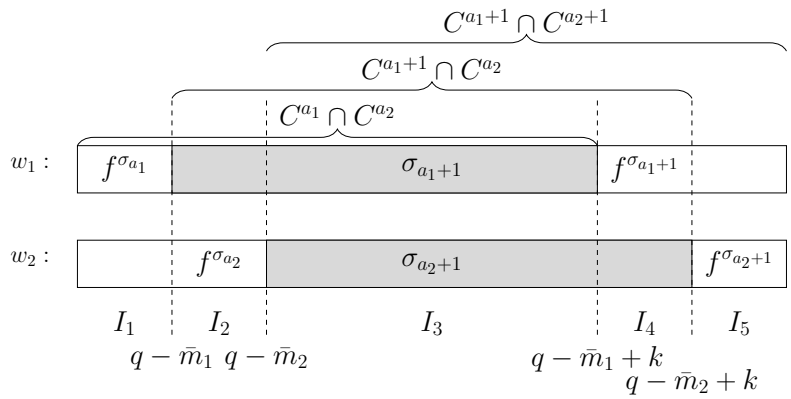

Figure 4.4: The partition in Case 1. The decomposition of w_1, w_2 into the intersections of rotations of codewords from $\mathcal{C}$ (shown by the curly brackets) is used to provide our distance bounds.

$0 \leq \bar{m}_1, \bar{m}_2 < n$. Assume without loss of generality that $\bar{m}_2 \leq \bar{m}_1$. We proceed by cases.

Case 1 First assume that $\bar{m}_1 - \bar{m}_2 < k$ and $k < \bar{m}_2$. As a result we will have that

$$0 < n - \bar{m}_1 \leq n - \bar{m}_2 < n - (\bar{m}_1 - k) \leq n - \bar{m}_2 + k < n$$

Therefore we can partition the interval window $[0, n)$ into 5 pieces by letting (see Figure 4.4)

$$I_1 := [0, n - \bar{m}_1)$$

$$I_2 := [n - \bar{m}_1, n - \bar{m}_2)$$

$$I_3 := [n - \bar{m}_2, n - \bar{m}_1 + k)$$

$$I_4 := [n - (\bar{m}_1 - k), \bar{m}_2 + k)$$

$$I_5 := [n - (\bar{m}_2 - k), n)$$

Note that it is possible that some of these intervals are empty (i.e. if $\bar{m}_1 = \bar{m}_2$) but this will not affect our argument.

For each j let $\text{agree}_j := \text{agree}(w_1[I_j], w_2[I_j])$. By Observation 2 for some $a_1,\ a_2$ we have that $w_1[I_1, I_2, I_3] \lesssim C^{a_1}$ and $w_1[I_3, I_4, I_5] \lesssim C^{a_1+1}$. Similarly we also have that

$w_2[I_1, I_2, I_3, I_4] \lesssim C^{a_2}$ and $w_2[I_3, I_4, I_5] \sim C^{a_2+1}$. Therefore by Corollary 3

$$\begin{aligned}
&\text{agree}_1 + \text{agree}_2 + \text{agree}_3 \\
&\qquad = \text{agree}(w_1[I_1, I_2, I_3], w_2[I_1, I_2, I_3]) \leq k + 3 \\
&\text{agree}_2 + \text{agree}_3 + \text{agree}_4 \\
&\qquad = \text{agree}(w_1[I_2, I_3, I_4], w_2[I_2, I_3, I_4]) \leq k + 3 \\
&\text{agree}_3 + \text{agree}_4 + \text{agree}_5 \\
&\qquad = \text{agree}(w_1[I_3, I_4, I_5], w_2[I_3, I_4, I_5]) \leq k + 3
\end{aligned}$$

Simply by noting that $|I_2| + |I_3| = |I_3| + |I_4| = k$ we find that

$$\begin{aligned}
\text{agree}_1 + \text{agree}_2 + \text{agree}_5 &\leq n - |I_2| - |I_3| = n - k \\
\text{agree}_1 + \text{agree}_4 + \text{agree}_5 &\leq n - |I_3| - |I_4| = n - k
\end{aligned}$$

Summing these five inequalities yields $\text{agree}(w_1, w_2) = \sum \text{agree}_m \leq \frac{2n+k}{3} + 3$. Summing only the first and third inequalities we find that

$$\begin{aligned}
\text{agree}(w_1, w_2) &= \sum \text{agree}_j \\
&\leq \text{agree}_1 + 2\text{agree}_2 + 2\text{agree}_3 + 2\text{agree}_4 + \text{agree}_5 \\
&\leq 2k + 6
\end{aligned}$$

Therefore in this case we find that $\text{agree}(w_1, w_2) \leq \min(2k + 6, \frac{2n+k}{3} + 3)$.

Case 2 Here assume again that $\bar{m}_2 > k$ but now $\bar{m}_1 - \bar{m}_2 \geq k$. Here we have to partition slightly differently, as the Gray code bits will not overlap. Note that we have

$$0 < n - \bar{m}_1 \leq n - \bar{m}_1 + k \leq n - \bar{m}_2 \leq n - \bar{m}_2 + k \leq n$$

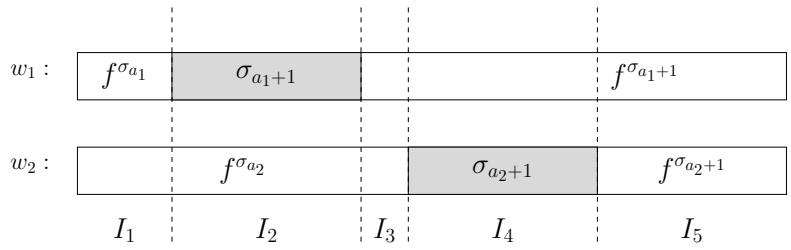

Figure 4.5: The partition in Case 2

So therefore we can partition $[0, n)$ as follows:

$$I_1 := [0, n - \bar{m}_1)$$
$$I_2 := [n - \bar{m}_1, n - \bar{m}_1 + k)$$
$$I_3 := [n - \bar{m}_1 + k, n - \bar{m}_2)$$
$$I_4 := [n - \bar{m}_2, n - \bar{m}_2 + k)$$
$$I_5 := [n - \bar{m}_2 + k, n)$$

Here we will have by Observations 1 and 2 that for some a_1, a_2 that $w_1[I_1, I_2] \lesssim C^{a_1}$ and $w_1[I_2, I_3, I_4, I_5] \lesssim C^{a_1+1}$, while $w_2[I_1, I_2, I_3, I_4] \lesssim C^{a_2}$ and $w_2[I_4, I_5] \lesssim C^{a_2+1}$. So again defining $\text{agree}_j := \text{agree}(w_1[I_j], w_2[I_j])$ we can use Corollary 3 to compute that

$$\text{agree}_1 + \text{agree}_2 = \text{agree}(w_1[I_1, I_2], w_2[I_1, I_2]) \leq k + 3$$

And again by similar reasoning applied on each pair of overlapping subwords

$$\text{agree}_1 + \text{agree}_2 \leq k + 3$$
$$\text{agree}_2 + \text{agree}_3 + \text{agree}_4 \leq k + 3$$
$$\text{agree}_4 + \text{agree}_5 \leq k + 3$$

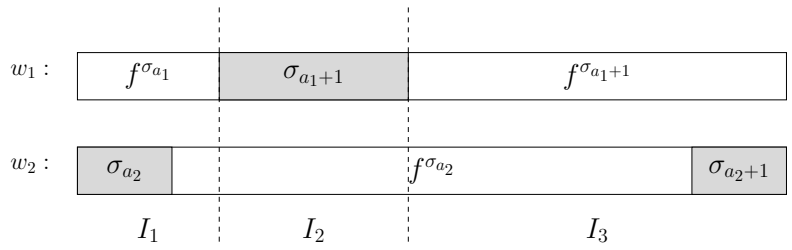

Figure 4.6: The partition in Case 4

Also, simply by noting that $|I_2| = |I_4| = k$ we find that

$$\begin{aligned}\text{agree}_1 + \text{agree}_2 + \text{agree}_3 + \text{agree}_5 &\leq q - |I_2| = n - k \\ \text{agree}_1 + \text{agree}_3 + \text{agree}_4 + \text{agree}_5 &\leq q - |I_4| = n - k\end{aligned}$$

So summing all five inequalities we find that $3\sum \text{agree}_m \leq 2q + k + 9$. And so $\text{agree}(w_1, w_2) \leq \frac{2q+k}{3} + 3$. Also summing over only the first 3 inequalities we find that

$$\begin{aligned}\text{agree}(w_1, w_2) &= \sum \text{agree}_j \\ &\leq \text{agree}_1 + 2\text{agree}_2 + \text{agree}_3 + 2\text{agree}_4 + \text{agree}_5 \\ &\leq 3k + 9\end{aligned}$$

Therefore in this case we find that $\text{agree}(w_1, w_2) \leq \min(3k + 9, \frac{2q+k}{3} + 3)$.

Case 3 In this case we assume that $0 \leq \bar{m}_2 \leq \bar{m}_1 \leq k$. By Lemma 1 we find that for some a_i $w_i \sim C^{a_i}$. So in this case we have by Corollary 3 that $\text{agree}(w_1, w_2) \leq k+3$.

Case 4 The last case is when $\bar{m}_2 \leq k$ but $\bar{m}_1 > k$. In this case if we let

$$\begin{aligned}I_1 &:= [0, n - \bar{m}_1 - 1] \\ I_2 &= [n - \bar{m}_1, q - \bar{m}_1 + d - 1] \\ I_3 &:= [n - \bar{m}_1 + d, q - 1]\end{aligned}$$

Then we have again by Observation 1 that $w_2 \sim C^{a_2}$ for some a_2. By Observation 2 that for some a_1, $w_1[I_1, I_2] \lesssim C^{a_1}$ and $w_1[I_2, I_3] \lesssim C^{a_1+1}$. If we define $\text{agree}_j := \text{agree}(w_1[I_j],\ w_2(w_2[I_j])$ then we will have that

$$\text{agree}_1 + \text{agree}_2 = \text{agree}(w_1[I_1, I_2], w_2[I_1, I_2]) \leq k + 3$$

And due to similar reasoning to the above we will have that

$$\text{agree}_1 + \text{agree}_2 \leq k + 3$$
$$\text{agree}_2 + \text{agree}_3 \leq k + 3$$

Also, simply by noting that $|I_2| = k$ we find that

$$\text{agree}_1 + \text{agree}_3 \leq n - k$$

So aggregating these inequalities we find that both $2\sum \text{agree}_m \leq n + k + 6$ and $\sum \text{agree}_m \leq 2k + 6$. As a result in this final case we get

$$\begin{aligned}\text{agree}(w_1, w_2) = \sum \text{agree}_j &\leq \min\left(\frac{n+k}{2} + 3, 2k + 6\right) \\ &\leq \min\left(\frac{2n+k}{3} + 3, 3k + 9\right)\end{aligned}$$

□

Corollary 4. *For any $0 < R < 1$ and $\delta < \max(\frac{1-R}{3}, 1 - 3R)$, for large enough q there exists a q-ary sequence of window length q, rate R and relative distance δ.*

Proof. We can compute the rate of S_Σ is

$$R = \frac{\log_n(n \cdot q^k)}{n} \geq \frac{\log_q(nq^k)}{n} = \frac{k+1}{n} - o_n(1)$$

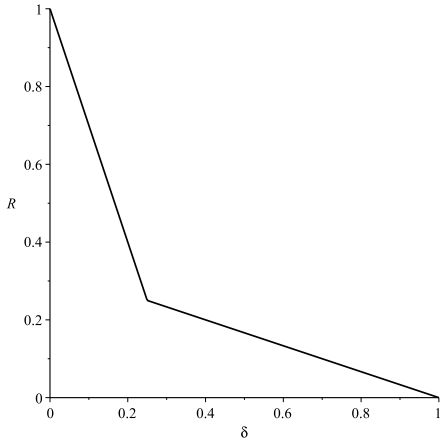

Figure 4.7: The Rate vs. Distance tradeoff of our construction as $q \to \infty$

And by Theorem 16 we have that the relative distance is

$$\begin{aligned}\delta &= \frac{\max\left(n-3k,\ \frac{n-k}{3}\right)}{n} - o_n(1) \\ &\geq \max\left(\frac{1-R}{3}, 1-3R\right) - o_n(1)\end{aligned}$$

□

Here we prove the useful fact that not only are any two codewords in $\mathcal{C}$ far apart, but also any two rotations of codewords of $\mathcal{C}$ are also distant as well. This is crucial for our analysis which routinely uses the fact that misaligned windows of the robust positioning pattern S_Σ still look like rotated codewords from $\mathcal{C}$.

Lemma 15. *For any $i_1, i_2 \in [n]$ and any $C^{a_1}, C^{a_2} \in \mathcal{C}$, so long as $(a_1, i_1) \not\equiv (a_2, i_2)$, then $\Delta(\rho^{i_1}(C^{a_1}), \rho^{i_2}(C^{a_2})) \geq q-k-1$ and therefore also $agree(\rho^{i_1}(C^{a_1}), \rho^{i_2}(C^{a_2})) \leq k+1$.*

Proof. First, we note that

$$\rho^{i_\ell}(C^{a_\ell}) = (f^{a_\ell}(g^{i_\ell}), f^{a_\ell}(g^{i_\ell+1}), \ldots, f^{a_\ell}(g^{i_\ell-1}))$$

But this is exactly the encoding $C(p_\ell)$ of the degree $k+1$ polynomial $p_\ell(X) = f^{i_m}(g^{i_\ell} X)$. Furthermore, we have that $p_1(0) = p_2(0) = 0$.

Therefore we will have that $\Delta(\rho^{j_2}(C^{i_2}), \rho^{j_2}(C^{i_2})) \geq q - deg(p_2 - p_1) \geq q - k - 1$ if we

can show that $p_2 - p_1 \neq 0$.

But note that

$$\text{coeff}_X(p_1) = \text{coeff}_X(f^{a_1}(g^{i_1}X)) = g^{i_1}$$
$$\text{coeff}_X(p_2) = \text{coeff}_X(f^{a_2}(g^{i_2}X)) = g^{i_2}$$

Therefore $p_1 = p_2$ only if $g^{i_1} = g^{i_2}$, which occurs only if $i_1 \equiv i_2 \mod n$.

If that is the case then $p_1 = p_2$ directly implies that $f^{a_1} = f^{a_2}$, contradicting our assumption that $(a_1, i_1) \not\equiv (a_2, i_2)$. □

4.4 Binary Positioning Sequences

4.4.1 Preliminaries

To concatenate down to binary we first need a marker to let us know the boundaries between words. To this end we construct a suitable binary word Ψ so that any two rotations of Ψ agree and differ in almost exactly half the coordinates.

Lemma 8. *For any $t \in \mathbb{N}$ there exists a binary word Ψ of length $2^t - 1$ so that for any two i, j, $0 \leq i \neq j \leq 2^t - 1$, the rotations $\rho^i(\Psi), \rho^j(\Psi)$ have the property $\left|agree(\rho^i(\Psi), \rho^j(\Psi)) - \frac{\ell}{2}\right| \leq 2^{\frac{t}{2}-1} \cdot O(t)$.*

We sketch the construction. Let g be a generator of $\mathbb{F}_{2^t}^{\times}$. Order the elements of $\mathbb{F}_{2^t}^{\times}$ by $x_i := g^i$, and take $\psi : \mathbb{F} \to \{\pm 1\}$ to be a nontrivial additive character of $\mathbb{F}_{2^t}$. Now we can define $\tilde{\Psi} := [\psi(g^0), \psi(g^1), \ldots, \psi(g^{2^t-2})]$. We will then let our codeword be the binary verison of this string by replacing -1 with 0.

Next, in order to use this marker appopriately in our concatenation, we must have a binary outer code which is also far from any window of Ψ.

Lemma 9. *Given $\delta < \frac{1}{2}$ and $R < 1 - H_2(\delta)$ For sufficiently large n, and any word W of length n, there exists a binary code $\mathcal{C}$ of block length n, relative distance δ and rate R so that all codewords of $\mathcal{C}$ have distance at least $\frac{\delta n}{2}$ from any rotation of W.*

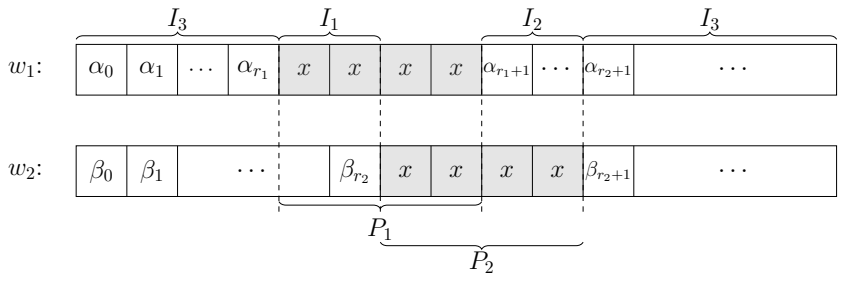

Figure 4.8: An illustration of the argument in Lemma 10. $w_i[I_3]$ is a subset of a window of S of length $n-|P_1 \cup P_2|$, and so contributes at least $d-|P_1 \cup P_2|$ in distance. Meanwhile there is additional contribution of $|P_1| + |P_2| - 2|P_1 \cap P_2|$ to the distance from I_1 and I_2 where copies of x are compared to elements of Σ.

Proof. By the Gilbert-Varshamov bound, for sufficiently large n there exists a code $\mathcal{C}_0$ with block length n relative distance δ and rate $R > 1 - H_2(\delta)$. Now we can define $\mathcal{C}$ from $\mathcal{C}_0$ by simply removing any codeword which has distance less than $\frac{\delta n}{2}$ from any rotation of W. Since $\mathcal{C}$ has distance greater than δn there can be at most one codeword removed per rotation of W. Therefore $|\mathcal{C}| \geq |\mathcal{C}| - n$ and so the rate is asymptotically unchanged. As the distance of $\mathcal{C}$ as at least the distance of $\mathcal{C}_0$, $\mathcal{C}$ satisfies the conditions we need. □

4.4.2 Augmented Sequences

First to build our binary sequence we will need to define a method of augmenting large alphabet sequences to include marker symbols which will help us with alignment issues.

Definition 15. Let S be a positioning sequence over Σ with window length n. Fix any $x \notin \Sigma$ and define the s-augmented sequence $A := A_s(S)$ over the alphabet $\Sigma \cup x$ of window length $n + s$ by

$$A[a(n+s)+b] := \begin{cases} x & \text{if } 0 \leq b < s \\ S[an+b-s] & \text{if } s \leq b < n+s \end{cases}$$

Lemma 10. *Let S be a positioning sequence and A the s-augmented sequence. If S has distance d then A has distance at least* $\min(d, 2s)$.

Proof. Take $w_1 \neq w_2$ to be arbitrary distinct length $n+s$ windows of U where $w_i := U\langle m_i \rangle_{n+s}$. Now for each window we define P_i to be the places corresponding to the copies of x. That is

$$P_i := \{j \in [n+s] \text{ s.t. } m_i + j \mod (n+s) < s\}$$

Also define

$$I_1 := P_1 \cap P_2^c \qquad I_2 := P_1^c \cap P_2 \qquad I_3 := (P_1 \cup P_2)^c$$

Because $w_1[I_1]$ is a string of only the character x and $w_2[I_1]$ contains no copies of x at all, we have that $\Delta(w_1[I_1], w_2[I_1]) = |I_1|$. Similarly we find that $\Delta(w_1[I_2], w_2[I_2]) = |I_2|$

There are two cases to consider. First if P_1 and P_2 are disjoint then $\Delta(w_1, w_2) \geq 2s$ as $|I_1| = |I_2| = s$.

In the second case we have $P_1 \cap P_2$ is nonempty. Therefore as the union of two contiguous (on the circle) intervals of length s, $P_1 \cup P_2$ is also such an interval but of length $\ell := |P_1 \cup P_2|$. There are two more subcases to consider. First if $P_1 \cup P_2 = [a, a+\ell)$ for some $a \leq n+s-\ell$ or second if $P_1 \cup P_2 = [a, n+s) \cup [0, \ell - n - s + a)$ for some $a \geq n+s-\ell$. Define $\tilde{w}_i := w_i[P_i^c]$, and note that by the construction of A, $\tilde{w}_i$ is a length n window of S and $\tilde{w}_1 \neq \tilde{w}_2$ so $\Delta(\tilde{w}_1, \tilde{w}_2) \geq d$. Here we will have that $I_3 := (P_1 \cup P_2)^c = [0, a) \cup [a+\ell, n+s)$ and so for $i = 1, 2$ it can be seen that $w_i[I_3] = \tilde{w}_i[I]$ where $I = [0, a) \cup [a + \ell - s, n)$. Therefore as $|I| = n + \ell - s$ it follows that

$$\begin{aligned}
\Delta(w_1, w_2) &\geq \Delta(w_1[I_1], w_2[I_1]) + \Delta(w_1[I_2], w_2[I_2]) \\
&\quad + \Delta(w_1[I_3], w_2[I_3]) \\
&\geq |I_1| + |I_2| + \Delta(\tilde{w}_1, \tilde{w}_2) - (\ell - s) \\
&= 3s - 2|P_1 \cap P_2| + d - |P_1 \cup P_2| \\
&\geq d
\end{aligned}$$

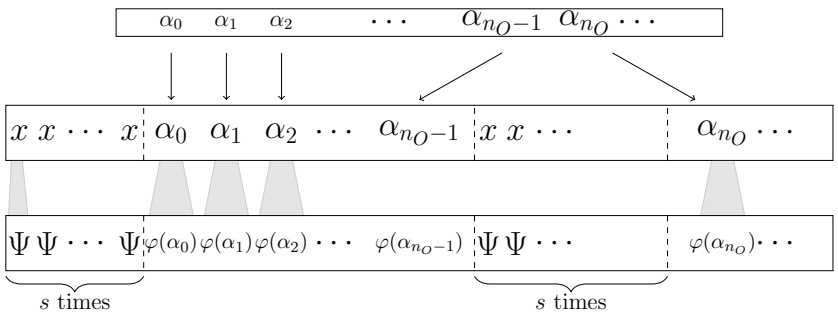

Figure 4.9: A view of the binary Robust Self Location Pattern T constructed in Section 4.4.3. Note that the sequence consists of a concatenation of the large alphabet sequence S to binary, intermixed with a locator word Ψ, which will aid in making detecting where a window lies in T modulo n_I.

In the second subcase the argument is exactly the same, but here we will have $I_3 := (P_1 \cup P_2)^c = [\ell - n - s + a, a)$ a contiguous interval of length $n + s - \ell$, and so $w_1[I_3]$ and $w_2[I_3]$ are subwindows of S of length $n + s - \ell$ and so the proof follows through using the same estimates as above.

□

4.4.3 Construction of the Binary Robust Positioning Sequence

Now we are ready to proceed with the construction of our binary robust positioning pattern.

Fix some $t \in \mathbb{N}$ and let $n := 2^t - 1$. Then take S to be a q-ary robust positioning sequence with window length n_O, rate R_O and relative distance δ_O. Let Ψ be a word of the form promised by Lemma 8 with length $n = 2^t - 1$ and let $\mathcal{C}$ be a binary code with q messages and block length n (and therefore rate $R_I := \frac{\log q}{n}$), relative distance $\delta_I > H_2^{-1}(1 - R_I)$ chosen as promised by Lemma 9 to have all windows distant from Ψ.

Take A to be an s-augmentation of the q-ary sequence S, then let $\varphi : \mathbb{F} \cup \{x\} \to \mathcal{C}$ to be an encoding of $\mathbb{F}_q$ to codewords of $\mathcal{C}$ with $\varphi(x) := \Psi$. Then we can define our binary positioning pattern to be given by

$$T := [\varphi(A[1]),\ \varphi(A[2]),\ \varphi(A[3]), \ldots]$$

$n_I := \lvert\varphi(\alpha)\rvert + \lvert\Psi\rvert = \tilde{n}_I + u$	$R_I := \frac{\log_2 q}{n}$
$\delta_I := H_2^{-1}(1-r) - \eta$	$\lvert\Psi\rvert := 2^t - 1 := n$
$N := n(n_O + s)$	$\bar{m} := m \mod \star$

Figure 4.10: Summary of Notation

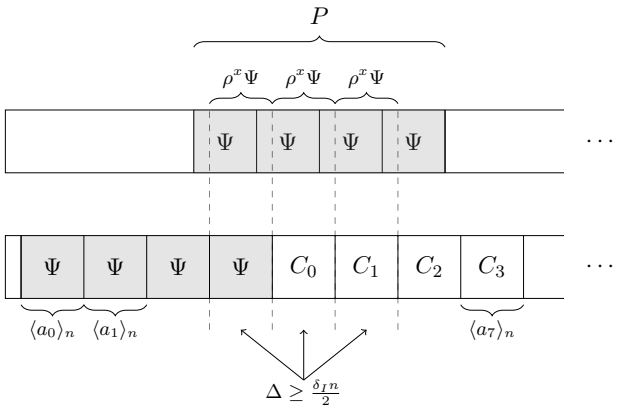

Figure 4.11: An illustration of the argument in Lemma 11. The distance in the non-aligned case comes from comparing rotations of the marker word Ψ to codewords of $\mathcal{C}$ and other copies of Ψ.

In particular we can see that if $m = aN + bn + c$ with $0 \leq b < N$ and $0 \leq s < n$ then

$$T[m] = T[aN + bn + c] := \begin{cases} \Psi[c] & \text{if } b < s \\ \phi(T[an_O + b])[c] & \text{if } s \leq b \end{cases}$$

4.4.4 Proof of Distance

We state our main result about the distance of T.

Theorem 17. *T is a binary robust positoning pattern of block length $(n_O + s)n_I$ and distance at least*

$$\min\left(\frac{(\min(n_O\delta_O - 1, 2s)\delta_I n_I}{2},\ (s-2)\frac{\delta_I n}{2}\right)$$

and rate at least $R_O R_I \frac{n_O}{n_O+s}$

Lemma 11. *Let $w_1 \neq w_2$ be any two windows of T of the form $w_i := T\langle m_i\rangle_N$. If $m_1 \neq m_2 \mod n$ then $\Delta(w_1, w_2) \geq (s-2)\frac{\delta n}{2}$.*

Proof. Define

$$P_1 := \{i \in [N] \text{ s.t. } m_1 + i = aN + bn + c \quad \text{where } 0 \leq c < n \text{ and } 0 \leq b < s\}$$

That is to say that P_1 is the set of indices corresponding to entries in w_1 coming from copies of Ψ. In particular if $i \in P_1$ and $i \equiv \bar{i} \mod n$ then $w_1[i] = \Psi[\bar{i}]$.

Now let

$$A_2 := \{i \in [N] \text{ s.t. } m_2 + i \equiv 0 \mod n\}$$

Assume that $m_2 \equiv \bar{m}_2 \mod n$. If we consider the sequence of subwindows $\langle in + n - m_2 \rangle_n$.

So A_2 is the set of beginnings of length n windows corresponding to either a copy of Ψ or a codeword of $\mathcal{C}$. Because P_1 is a set of size sn consisting of at most 2 runs of consecutive integers, it must contain at least $s-2$ length n windows of the form $\langle a \rangle_n$ where $a \in A_2$. But for such a window we will have $w_2\langle a \rangle_n$ is either a codeword of $\mathcal{C}$ or a copy of Ψ, while if $x := m_2 - m_1 \not\equiv 0 \mod n$ we will have that $w_1\langle a \rangle_n = \rho^x \Psi$. So either by the construction of Ψ to have low autocorrelation (Lemma 8) or by the distance of all codewords of $\mathcal{C}$ from all rotations of Ψ (Lemma 9) we will have that

$$\Delta(w_1[\langle a \rangle_n], w_2[\langle a \rangle_n]) \geq \min\left(\frac{\delta_I n}{2}, (1 - o(1))\frac{n}{2}\right) = \frac{\delta_I n}{2}$$

Because P_1 contains $s-2$ disjoint such windows, the result follows. □

Next we cover the case when the codewords are aligned modulo n. Here we will have that binary codewords and copies of Ψ will be compared to eachother and will get our distance from the distance of the concatenation code combined with the distance of the large alphabet positioning sequence.

Lemma 12. *Let $w_1 \neq w_2$ be any two windows of T of the form $w_i := T\langle m_i \rangle_N$. If $m_1 \equiv m_2 \mod n$ then $\Delta(w_1, w_2) \geq \frac{(\min(n_O \delta_O - 1, 2s)\delta_I n_I}{2}$.*

Proof. Let $0 \leq \bar{m} < n$ such that $m_1 \equiv m_2 \equiv \bar{m} \mod n$. Now for $j \in [n_O + s - 1]$

define $I_j := \langle nj + n - \bar{m} \rangle_n$. Then for any i, j $w_i[I_j]$ corresponds to a codeword in $\mathcal{C}$ or a copy of Ψ, and in fact it must be that $u_1 := \{\varphi^{-1}(w_1[I_j])\}_{j \in [n_O+s-1]}$ and $u_2 := \{\varphi^{-1}(w_2[I_j])\}_{j \in [n_O+s-1]}$ are distinct windows of A of length $n_O + s - 1$. By Lemma 10 we know that these windows differ in at least $\min(n_O\delta_O, 2s) - 1$ positions. Since $\mathcal{C} \cup \{\Psi\}$ forms a code of distance $\frac{n\delta_I}{2}$ our result follows. □

We are now ready to restate and prove our main result of this section:

Theorem 18. *T is a binary robust positoning pattern of block length $(n_O + s)n_I$ and distance at least*

$$\min\left(\frac{(\min(n_O\delta_O - 1, 2s)\delta_I n_I}{2},\ (s-2)\frac{\delta_I n}{2}\right)$$

and rate at least $R_O R_I \frac{n_O}{n_O+s}$

Proof. The statement of distance is a combination of Lemmas 12 and 11. Meanwhile the rate statement follows from the standard calculation

$$R = \frac{\log_2 |T|}{(n_O + s)n_I} \geq \frac{\log_2 |q| \log_2 |S|}{(n_O + s)n_I} = R_O R_I \frac{n_O}{n_O + s}$$

□

We note that choosing s, R_O and R_I properly we can obtain the following corollary.

Corollary 5. *For any $0 < R < 1$ there is some $\delta(R)$ such that the above construction yields binary positioning patterns of arbitrarily long block length, rate R and relative distance $\delta(R)$.*

4.5 Encoding/Decoding Over Large Alphabets

The following is an algorithm for encoding a window of $S := S_\Sigma$ as defined in Definition 14. Assume we wish to compute $S[m]$ where $0 \leq m \leq q^k - 1$. Assume we are using the Gray Code Σ given by the standard inductive construction, for which an efficient method of encoding and decoding exists.

Algorithm 1 Algorithm for Encoding our large alphabet pattern 14

Given m to compute $S\langle m\rangle_n$ do:

1. Find $0 \leq \bar{m} < q$ and a so that $m = aq + \bar{m}$
2. Find σ_a and σ_{a+1} (the a^{th} and $(a+1)^{st}$ entries in the q-ary gray code Σ.
3. Interpolate the polynomials f^a and f^{a+1} so that for $j = 0, 1$
   (a) $f^{a+j}(0) = 0$
   (b) $(f^{a+j})'(0) = 1$
   (c) $f^{a+j}(g^i) = \sigma_i^{a+j}$ for $0 \leq i \leq k-1$
4. Output $[(f^a(g^j))_{j=\bar{m}}^{q-1},\ (f^a(g^j))_{j=0}^{\bar{m}-1}]$

Theorem 19. *Algorithm 2 for decoding received words of the window sequence S can correct* $\min(\delta n, \frac{q+k}{2} - 1 - \sqrt{q(k+1)})$ *errors (where* $\delta n = \max(q - 3k - 9,\ \frac{q-k}{3} - 3)$*). Furthermore the algorithm runs in time $O(poly(q))$.*

Algorithm 2 Algorithm for Decoding our large alphabet pattern 14

Assume we receive a window w of length q. To decode do:

1. Run the Guruswami-Sudan list decoding algorithm [18] for Reed Solomon codes on w, reuturning the list of degree $k+1$ polynomials $L := \{p \text{ s.t. } \Delta(C(p), w) \leq q - \sqrt{q(k+1)}\}$.
2. For each polynomial $p \in L$ do the following:
   (a) For each $i \in [q]$ find the index a_i such that

$$\begin{aligned}\rho^i(C^p)[0, k-1] \\ &= (p(g^i), p(g^{i+1}), \ldots, p(g^{i+k-1})) \\ &= \Sigma[a_i, a_i + k - 1)\end{aligned}$$

   where Σ is the q-ary Gray code in use.
   (b) Make the guesses $\mu_p^1 = a_i q + (q - i)$ and $\mu_p^0 = (a_i - 1)q + (q - i)$
   (c) For each guess μ_p^i let $w_p^i = S\langle \mu_p^i \rangle_n$. If $\dot{(}w, w_p^i) < \frac{\delta q}{2} = \frac{1}{2}\max(q - 3k - 6,\ \frac{q-k}{3} - 2)$ then return w_p^i and its index μ_p^i

Proof. Assume that the sent window was $u = S[m]$ with $m = aq + \bar{m}$, so that $\dot{(}u, w) \leq \frac{q+k}{2} - 1 - \sqrt{q(k+1)}$. Let $x = q - \bar{m}$, and $I_0 = [0, x + k - 1]$, $I_1 = [x, q - 1]$. If $\bar{m} < \frac{q-k}{2}$

then by either Observation 1 or 2 we will have that $u[I_0] \sim \rho^{\bar{m}} C^a[I_0]$, and therefore

$$\begin{aligned}\Delta(u, \rho^{\bar{m}} C^a) &\leq (q - |I_0|) + (\dot{u}[I_0], \rho^{\bar{m}} C^a[I_0]) \\ &\leq (q - |I_0|) + 1\end{aligned}$$

Similarly if $\bar{m} \geq \frac{q-k}{2}$ then $u[I_1] \sim C^{a+1}[0, k + \bar{m} - 1] = \rho^{\bar{m}} C^{a+1}[x, q-1]$. So

$$\begin{aligned}(\dot{\rho^x} u, C^{a+1}) &= (\dot{u}, \rho^{\bar{m}} C^{a+1}) \\ &\leq (q - |I_1|) + (\dot{u}[I_1], \rho^{\bar{m}} C^{a+1}[I_1]) \\ &\leq (q - |I_1|) + 1\end{aligned}$$

Because $|I_0| + |I_1| = q + k$ for some j we must have $|I_j| \geq \frac{q+k}{2}$.

Therefore we can compute that

$$\begin{aligned}(\dot{w}, \rho^{\bar{m}} C^{a+j}) &\leq (\dot{w}, u) + (\dot{u}, \rho^{\bar{m}} C^{a+j}) \\ &\leq \frac{q+k}{2} - 1 - \sqrt{q(k+1)} + q - |I_j| + 1 \\ &\leq q - \sqrt{q(k+1)}\end{aligned}$$

Therefore we see that the Guruswami-Sudan list decoding algorithm will place either $f^{a+j}(x + \bar{m})$ in its list L.

Therefore when step $2a$ tries $p = f^{a+j}(x + \bar{m})$ and $i = q - \bar{m}$ we will have $\rho^i \rho^{\bar{m}} C^{a+j}[0, k-1] = C^{a+j}[0, k-1] = \sigma^{a+j}$, and conseuqently the guesses $\mu_p^1 = (a+j)q + \bar{m}$ and $\mu_p^0 = (a+j-1)q + \bar{m}$ will be made. If $j = 0$ then the former will be correct, and if $j = 1$ then the latter will be. Either way step c will check $\mu = aq + \bar{m} = m$, and because $(\dot{w}, S\langle m \rangle_n) < \frac{\delta n}{2}$ the algorithm will return m and $S\langle m \rangle_n$.

The only thing left to check is that the algorithm returns no false positives. But, by Theorem **??** we know that all windows of S have distance at least $\delta n := \max(q - 3k - 9, \ \frac{q-k}{3} - 3)$ from eachother. Therefore it follows that there is always at most one window $S\langle \mu \rangle_n$ so that $(\dot{S}\langle \mu \rangle_n, w) \leq \frac{\delta n}{2}$, for any window w, and therefore only the correct window could ever be returned.

To check the runtime claim we note that the Algorithm in step 1 runs in time $poly(q)$ and returns a list of size $|L| \leq q^2$. Furthermore each of the operation in steps 2a takes time $poly(q)$ by the decodability of Σ. The operations in step 2b take time $O(1)$, and each step in 2c runs in time $poly(q)$ by the encoding algorithm of S given $\square$

4.5.1 Encoding/Decoding in Binary

Let T be a binary robust positioning sequence as constructed section 4.4.3. with window length $N = (n_O + s)n$ and distance d.

First we comment on construction of the sequence. The only point of interest here is to find the locator word Ψ its accompanying code $\mathcal{C}$. Ψ is just a character over $\mathbb{F}_{2^n}$ and so can be any multilinear polynomial. over $\mathbb{F}_2[X_1, \ldots, X_n]$. For $\mathcal{C}$ we may pick any efficiently encodable and decodable good distance binary code of rate R, of which numerous constructions exist. Then to pair it with Ψ we only have to remove the codeword in $\mathcal{C}$ of distance less than $\frac{d}{2}$ from each rotation of Ψ, a process which takes at most n calls to the decoding algorithm of $\mathcal{C}$.

Now we discuss the decoding algorithm. Here the process proceeds in two steps. First we find where in the window are the s copies of Ψ. Once we know that, we know exactly which blocks of length n correspond to concatenated codewords, and can apply usual decoding methods for concatenated words.

Assume that S has a decoder algorithm D_O which given a window $w \in \Sigma_O^{n_O}$ will determine (if possible) the unique window $S\langle m \rangle_{n_O}$ so that $\underset{\cdot}{(}S\langle m \rangle_{n_O}, w) < d$ in time $poly(q)$. Assume also that we also have for the inner alphabet a decoding algorithm D_I so that for any received word $w \in [2]^n$ D_I returns the unique letter $\alpha \in [q]$ so that $\underset{\cdot}{(}C_I(\alpha), w) < \frac{\delta_I n}{2}$ in time $poly(q)$.

Theorem 20. *Algorithm 3 runs in time* $\mathrm{poly}(N)$ *and given a window* $w \subset [2]^N$ *returns (if it exists) the unique window* $u := T\langle m \rangle_N$ *such that* $\underset{\cdot}{(}u, w) < \delta_O \delta_I \frac{(n-1)n_O}{4}$ *in time* $poly(N)$.

Algorithm 3 Algorithm for decoding our binary pattern 4.4.3

For each i from 0 to N do

1. For each $0 \leq j < n_O$ decode (if possible) the length n window $(\rho^i w)\langle jn \rangle_n$ to α_j using the decoder of $\mathcal{C}$.
2. Let $\tilde{w}_i$ be the q-ary string $(\alpha_0, \ldots, \alpha_{n_O - 1})$.
3. Run the decoder D_O over large alphabets on $\tilde{w}_i$ (padded with an extra bit if necessary) and return its index $\tilde{\mu}_i$
4. Let $\mu = n_O \tilde{\mu} - i$.
5. If $\underset{\cdot}{(}T[\mu], w) < \frac{d}{2}$ then return $T[\mu], \mu$.

The argument that this decoding works is very similar to decoding an ordinary concatenated code. We can brute force through every rotation $\rho^i w$ for $i \in [N]$ of the received window w, and try decoding $\rho^i w$ as we would any concatenated code of length $n_O n$ (the algorithm will return when the unchecked sn entries correspond to the s copies of Ψ). Since one of the rotations will correspond to us having a word which is $n - 1$ blocks of concatenated codewords (and possibly 1 junk block from the beginning and end of the window), we will be able to decode at least a $\delta_O N_O$ fraction of these blocks correctly, and the decoder for the large alphabet robust positioning sequence handles decoding the resulting large alphabet sequence. In step 5 we use the fact that T has good distance to eliminate any possible false positives.

4.6 Positioning Sequences with Constant Distance

In this section, we give a brief description of our construction of positioning sequences with constant distance d.

Over large alphabets Σ, it follows by inspecting the parameters in Theorem 16 that the construction there leads to sequences of length $\frac{|\Sigma|^n}{|\Sigma|^{O(d)}}$, which is essentially optimal (upto the constant in the $O(d)$) by the Singleton bound.

Over the binary alphabet, we have to do something different. Here we are aiming to get a sequence of length $\frac{2^n}{n^{O(d)}}$. The concatenation scheme described for the case of constant relative distance codes is insufficient, since any nontrivial concatenation map

leads to a drastic reduction in the length of the sequence. Instead, we will use a trivial concatenation map, along with a simpler marker, at the cost of having to rely on an unproven conjecture (Conjecture C from the introduction).

Assuming Conjecture C, for infinitely many r we can choose a prime q between $2^r - cr$ and $2^r - 1$. We start with a large alphabet positioning sequence over the alphabet $\Sigma = \mathbb{F}_q$ with distance d. Now choose a one-to-one map $\phi : \mathbb{F}_q \to \{0,1\}^r$ whose image avoids the string 0^r: this is possible since $q \leq 2^r - 1$. We will be using the map ϕ to encode large alphabet symbols into sequences of binary symbols. The final binary sequence is then obtained by taking the ϕ-encoding of each symbol of the large alphabet sequence, along with the marker sequence $(0^{2r}1^r)^{3d}$. The goal of this marker sequence, as in the case of the constant relative distance codes, is to ensure alignment. The fact that the image of ϕ avoids 0^r is what ensures that this marker sequence cannot have too much agreement with any symbols outside the marker sequence. Finally, the fact that $q > 2^r - cr$ ensures that the length of the sequence so constructed is as long as $\frac{2^n}{n^{O(d)}}$: encoding elements of Σ by ϕ did not make us lose too much in the rate.

Chapter 5

Conclusion

We will conclude by highlighting some of the most important open questions left from the work in this thesis.

5.1 Open Questions for Chapter 2

In Chapter 2 we proved a local limit theorem for triangle counts in the random graph for the regime where p is a fixed constant. The most central and broadest open question is

Question 1. *Characterize when combinatorial random variables which obey central limit theorems also obey local limit theorems.*

This question is open ended, and I do not yet have a solid guess for what the right criteria are. I suspect that there is much to be done here, and that that there may be many subtle conditions which need to be considered beyond the low degree polynomials and spectral concentration arguments in this chapter of the thesis. For example the simple function E^2 which counts the *square* of the number of edges in $G(n,p)$ obeys a central limit theorem for any fixed $p \in (0,1)$, and has very good weight 1 spectral concentration (leading to good characteristic function bounds for small values of t). However it is readily seen not to obey a local limit theorem as it is only supported on square numbers. In the interim a more concrete question, which I believe to be significantly more approachable is the following

Question 2. *For which k do we have that for all $p \in (0,1)$ the number of complete graphs of order k in $G(n,p)$ obeys a local limit theorem?*

5.2 Open Questions for Chapter 3

Our most natural open question of this chapter concerns the optimality of our main results. Namely, a proof of Conjecture 1 as stated in Section 3.1 would be very interesting. We restate this in the form of a specific question:

Question 3. *Does there is some constant $C > 0$ such that the following holds. If A is an $n \times n$ matrix with complex entries and $\|A\|_2 \leq 1$, then $|\mathrm{per}(A)| \leq e^{-Cn(1-h_\infty)}$.*

This would be a tight result, as it is acheived (wastefully) by the matrix αI, where I is the identity matrix and α is an arbitrary constant less than 1.

We are also very interested in whether the results of this chapter can lead to a polynomial time deterministic algorithm for approximating the permanent. We again echo the question from [2] and ask:

Question 4. *Find a (deterministic) polynomial-time algorithm that takes an $n \times n$ matrix A of norm 1 and decides whether $|\mathrm{per}(A)| < n^{-100}$ or $|\mathrm{per}(A)| > n^{-10}$ (with the understanding that the input matrix will satisfy one of these inequalities).*

5.3 Open Questions for Chapter 4

The most natural question would be to determine the true rate-distance trade-off for robust positioning patterns. Currently we have our construction, and an omitted argument using the Lovasz Local Lemma as lower bounds, but the only upper bounds are those taken directly from standard coding theory. We state this as a question:

Question 5. *For any fixed $\delta^* \in (0, 1)$ what is the largest rate R^* such that there exist an infinite family of sequences $\{S_i\}$ all having rate at least R^* and also relative distance at least δ^*?*

Also, it is worth noting that while we have a very natural construction of 2-dimensional robust positioning patterns, the proof of distance for these constructions remains somewhat thorny, and does not immediately extend to dimensions larger than 2. It would be interesting if a slick proof could show that our 2 dimensional construction and its natural extensions to higher dimensions all have constant fraction distance.

Bibliography

[1] Microsoft mulls a stylus for any screen. http://www.technologyreview.com/news/428521/microsoft-mulls-a-stylus-for-any-screen/ .

[2] Scott Aaronson and Travis Hance. Generalizing and derandomizing Gurvits's approximation algorithm for the permanent. *Quantum Inf. Comput.*, 14(7-8):541–559, 2014.

[3] Scott Aaronson and Hoi Nguyen. Near invariance of the hypercube. *Israel Journal of Mathematics*, 2016.

[4] Nir Ailon and Edo Liberty. Fast dimension reduction using Rademacher series on dual BCH codes. *Discrete Comput. Geom.*, 42(4):615–630, 2009.

[5] A. D. Barbour, Michał Karoński, and Andrzej Ruciński. A central limit theorem for decomposable random variables with applications to random graphs. *J. Combin. Theory Ser. B*, 47(2):125–145, 1989.

[6] Ross Berkowitz. A quantitative local limit theorem for triangles in random graphs, 2016.

[7] Ross Berkowitz and Pat Devlin. A stability result using the matrix norm to bound the permanent, 2016.

[8] Ross Berkowitz and Swastik Kopparty. Robust positioning patterns. In *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1937–1951.

[9] Aline Bonami. Étude des coefficients de Fourier des fonctions de $L^p(G)$. *Ann. Inst. Fourier (Grenoble)*, 20(fasc. 2):335–402 (1971), 1970.

[10] A.M. Bruckstein, T. Etzion, R. Giryes, N. Gordon, R.J. Holt, and D. Shuldiner. Simple and robust binary self-location patterns. *Information Theory, IEEE Transactions on*, 58(7):4884–4889, July 2012.

[11] John Burns and Chris J Mitchell. Coding schemes for two-dimensional position sensing. In *Institute of Mathematics and Its Applications Conference Series*, volume 45, pages 31–31. Oxford University Press, 1993.

[12] ZD Dai, KM Martin, MJB Robshaw, and PR Wild. Orientable sequences. In *INSTITUTE OF MATHEMATICS AND ITS APPLICATIONS CONFERENCE SERIES*, volume 45, pages 97–97. OXFORD UNIVERSITY PRESS, 1993.

[13] P. Erdős and A. Rényi. On the evolution of random graphs. *Bull. Inst. Internat. Statist.*, 38:343–347, 1961.

[14] William Feller. *An introduction to probability theory and its applications. Vol. II.* Second edition. John Wiley & Sons, Inc., New York-London-Sydney, 1971.

[15] Ehud Friedgut. Sharp thresholds of graph properties, and the k-sat problem. *J. Amer. Math. Soc.*, 12(4):1017–1054, 1999. With an appendix by Jean Bourgain.

[16] Justin Gilmer and Swastik Kopparty. A local central limit theorem for the number of triangles in a random graph. *ArXiv e-prints*, November 2014.

[17] David G. Glynn. The permanent of a square matrix. *European J. Combin.*, 31(7):1887–1891, 2010.

[18] Venkatesan Guruswami and Madhu Sudan. Improved decoding of Reed-Solomon and algebraic-geometric codes. *IEEE Transactions on Information Theory*, 45:1757–1767, 1999.

[19] Leonid Gurvits. On the complexity of mixed discriminants and related problems. In *Mathematical foundations of computer science 2005*, volume 3618 of *Lecture Notes in Comput. Sci.*, pages 447–458. Springer, Berlin, 2005.

[20] Leonid Gurvits and Alex Samorodnitsky. Bounds on the permanent and some applications. In *2014 IEEE 55th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 90–99. IEEE, 2014.

[21] Mariko Hagita, Makoto Matsumoto, Fumio Natsu, and Yuki Ohtsuka. Error correcting sequence and projective de Bruijn graph. *Graphs and Combinatorics*, 24(3):185–194, 2008.

[22] Zsolt Horváth, Adam Herout, István Szentandrási, and Michal Zachariáš. Design and detection of local geometric features for deformable marker fields. In *Proceedings of the 29th Spring Conference on Computer Graphics*, SCCG '13, pages 073:73–073:80, New York, NY, USA, 2013. ACM.

[23] Svante Janson and Krzysztof Nowicki. The asymptotic distributions of generalized U-statistics with applications to random graphs. *Probab. Theory Related Fields*, 90(3):341–375, 1991.

[24] Lode Jorissen, Steven Maesen, Ashish Doshi, and Philippe Bekaert. Robust global tracking using a seamless structured pattern of dots. In Lucio Tommaso De Paolis and Antonio Mongelli, editors, *Augmented and Virtual Reality*, volume 8853 of *Lecture Notes in Computer Science*, pages 210–231. Springer International Publishing, 2014.

[25] Michał Karoński. *Balanced subgraphs of large random graphs*, volume 7 of *Seria Matematyka [Mathematics Series]*. Uniwersytet im. Adama Mickiewicza w Poznaniu, Poznań, 1984. With a Polish summary.

[26] Michał Karoński and Andrzej Ruciński. On the number of strictly balanced subgraphs of a random graph. In *Graph theory (Łagów, 1981)*, volume 1018 of *Lecture Notes in Math.*, pages 79–83. Springer, Berlin, 1983.

[27] Hermann König, Carsten Schütt, and Nicole Tomczak-Jaegermann. Projection constants of symmetric spaces and variants of Khintchine's inequality. *J. Reine Angew. Math.*, 511:1–42, 1999.

[28] Bhaskar Krishnamachari and Kiran Yedavalli. Secure sequence-based localization for wireless networks. In Radha Poovendran, Sumit Roy, and Cliff Wang, editors, *Secure Localization and Time Synchronization for Wireless Sensor and Ad Hoc Networks*, volume 30 of *Advances in Information Security*, pages 237–247. Springer US, 2007.

[29] P.Y. Kumar and V.K. Wei. Minimum distance of logarithmic and fractional partial m-sequences. *Information Theory, IEEE Transactions on*, 38(5):1474–1482, Sep 1992.

[30] Michel Ledoux and Michel Talagrand. *Probability in Banach spaces*. Classics in Mathematics. Springer-Verlag, Berlin, 2011. Isoperimetry and processes, Reprint of the 1991 edition.

[31] F.J. MacWilliams and N.J.A. Sloane. Pseudo-random sequences and arrays. *Proceedings of the IEEE*, 64(12):1715–1729, Dec 1976.

[32] ChrisJ. Mitchell and KennethG. Paterson. Decoding perfect maps. *Designs, Codes and Cryptography*, 4(1):11–30, 1994.

[33] C.J. Mitchell, T. Etzion, and K.G. Paterson. A method for constructing decodable de Bruijn sequences. *Information Theory, IEEE Transactions on*, 42(5):1472–1478, Sep 1996.

[34] Hoi Nguyen. On matrices of large permanent. Private communication, 2016.

[35] Krzysztof Nowicki and John C. Wierman. Subgraph counts in random graphs using incomplete U-statistics methods. In *Proceedings of the First Japan Conference on Graph Theory and Applications (Hakone, 1986)*, volume 72, pages 299–310, 1988.

[36] Ryan O'Donnell. *Analysis of Boolean functions*. Cambridge University Press, 2014.

[37] V. V. Petrov. *Sums of independent random variables*. Springer-Verlag, New York-Heidelberg, 1975. Translated from the Russian by A. A. Brown, Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 82.

[38] Adrian Röllin and Nathan Ross. Local limit theorems via Landau-Kolmogorov inequalities. *Bernoulli*, 21(2):851–880, 2015.

[39] Andrzej Ruciński. When are small subgraphs of a random graph normally distributed? *Probab. Theory Related Fields*, 78(1):1–10, 1988.

[40] E.R. Scheinerman. Determining planar location via complement-free de brujin sequences using discrete optical sensors. *Robotics and Automation, IEEE Transactions on*, 17(6):883–889, Dec 2001.

[41] Warren Schudy and Maxim Sviridenko. Concentration and moment inequalities for polynomials of independent random variables. In *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 437–446. ACM, New York, 2012.

[42] I. Szentandrasi, M. Zacharias, J. Havel, A. Herout, M. Dubska, and R. Kajan. Uniform marker fields: Camera localization by orientable de Bruijn tori. In *Mixed and Augmented Reality (ISMAR), 2012 IEEE International Symposium on*, pages 319–320, Nov 2012.

[43] L. G. Valiant. The complexity of computing the permanent. *Theoret. Comput. Sci.*, 8(2):189–201, 1979.