RANDOMNESS AS A TOOL FOR MODELING AND UNCOVERING STRUCTURE

A Dissertation
Presented to
The Academic Faculty

by

Samantha Petti

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in
Algorithms, Combinatorics, and Optimization

School of Mathematics
Georgia Institute of Technology
May 2020

RANDOMNESS AS A TOOL FOR MODELING AND UNCOVERING STRUCTURE

Approved by:

Dr. Santosh Vempala, Advisor
School of Computer Science
Georgia Institute of Technology

Dr. Will Perkins
Department of Mathematics, Statistics and Computer Science
University of Illinios at Chicago

Dr. Prasad Tetali
School of Mathematics
Georgia Institute of Technology

Dr. Eric Vigoda
School of Computer Science
Georgia Institute of Technology

Dr. Lutz Warnke
School of Mathematics
Georgia Institute of Technology

Date Approved: March TBD, 2020

To my grandmothers Barbara Ann Petti and Marcia Shafer for encouraging me as I pursued academic opportunities that they did not have.

ACKNOWLEDGEMENTS

I am incredibly grateful for the consistent support of my advisor Santosh Vempala. I am fortunate to have joined him on many long walks around campus during which he shared all sorts of exciting research ideas and also his extreme enthusiasm for research. I appreciate his patience while he helped bring me up to speed on the foundations of theoretical computer science during my first year. I would like to thank Christian Borgs and Jennifer Chayes for mentoring me during two wonderful summer internships at Microsoft Research. Thanks also to my other MSR collaborators: Souvik Dhara, Julia Gaudio, and Subhabrata Sen. Each has taught me a non-trivial amount of probability theory and has graciously listened to my sometimes incoherent ideas. Thanks to Alan Frieze for hosting me as a summer student, and to Will Perkins for sharing his excitement for spheres in high dimensions and many related research ideas. I am grateful to my committee, Will Perkins, Prasad Tetali, Santosh Vempala, Eric Vigoda, and Lutz Warnke, for taking the time to review this thesis.

I owe a special thanks to the other students in my cohort at Georgia Tech: Digvijay Boob, Matthew Fahrbach, Anna Kirkpatrick, Kevin Lai, Samira Samadi, and Saurabh Sawlani. Their willingness to work together on our ACO coursework and support throughout the past five years have made the experience of earning a Ph.D. substantially smoother. Thanks to the theory students in previous cohorts who helped show me the ropes: Prateek Bhakta, Sarah Cannon, Ben Cousins, and David Durfee. I would also like to thank my roommates Sarah Cannon, Alison Smith, and Sally Collins for making our home a great place to be. I am grateful for the continued support of my partner Travis Burrows.

Finally, I would like to thank my parents for providing me with many excellent educational opportunities and lots of encouragement through the years.

TABLE OF CONTENTS

LIST OF FIGURES

SUMMARY

This thesis contains four main research directions, united by the themes of using randomness to (i) construct structure and (ii) uncover structure. Randomness has long been used for these tasks. Random models are defined to mirror some system, and the probabilistic analysis of the model then provides insight into the properties and behavior of the system. Using random choices in algorithms yields faster results that are very accurate for most cases. The middle two research directions presented here focus on random graph models, in particular developing a theory of sparse graph limits. The first and final directions exemplify the themes in other contexts.

(1) Inspired by how computation may happen in the brain, we describe a method of building a threshold function by randomly connecting small primitive Boolean circuits. Our construction demonstrates the theme that a series of random choices can produce a structured object, in this case a Boolean circuit that computes a threshold function.

(2) We introduce the Random Overlapping Communities (ROC) model in order to model the local structure of sparse graphs. We show that the model can be tuned to produce graphs with a wide range of normalized closed walk counts, including the closed walk counts of the hypercube sequence. This direction also illustrates the first theme; the randomness in the model produces graphs with a specified set of closed walk counts.

(3) Fitting with the second theme, we explore the extent to which a small sample of a large graph can be used to deduce properties of the large graph. We introduce the Community Configuration model (CCM) to model graphs with overlapping community structure and varied degree distributions. We describe the sampling limit for sequences of CCMs and use the limit to draw parallels between CCMs and random graph models established in the literature. We also describe a hypothesis test to determine whether a graph was

produced from a CCM or a configuration model without community structure given access to a small random sample.

(4) The *hard sphere model* is a well-known statistical physics model of monatomic gases. We describe a Markov chain for sampling from the model that achieves rapid mixing for a wider range of fugacity parameters than previously known. Analyzing the sampling process governed by the Markov chain allows us deduce properties of the infinite volume limit of the model, thus demonstrating the second theme.

CHAPTER 1
INTRODUCTION

The work presented here is united by two related themes: (i) using randomness to construct an object with a desired structure, and (ii) using randomness to uncover structure. Graphs are a canonical example of an "object" in this context. To exemplify these themes, we begin with a brief overview of random graphs, property testing on graphs, and how these tasks are tied together by the beautiful theory of regularity (Section 1.1). Then we describe the results contained in this thesis and how they relate to these themes (Section 1.2).

1.1 Randomness in the study of graphs

We begin by discussing random graphs, the Szemerédi regularity lemma, graphons, property testing, and how together these ideas form an elegant theory of dense graph limits. A substantial part of this thesis is motivated by the goal of developing an analogous theory for sparse graphs (see Chapters 4 and 5). The following summary is brief and informal; see Chapter 2 for the formal statements of the definitions and theorems.

1.1.1 Random graph models

In 1959, Erdős and Rényi [ER59, ER60] introduced the first random graph model, and in doing so initiated a long line of research that uses randomness to model and understand the structure of graphs. Their model (which is also attributed to Gilbert [Gil59]), denoted $G_{n,p}$, constructs a random graph on n vertices by connecting every pair of vertices independently with probability p. Mathematicians have extensively studied properties of graphs generated from the Erdős and Rényi model and used the model to prove the existence of graphs with

certain properties. (See [FK15] for a survey.) In computer science, random graphs are often used to benchmark the runtime of graph algorithms. Moreover, the comparison of biological and social networks to random graph models is a popular technique for highlighting features of the network's structure [New03, New05, Son+05].

As Erdős and Rényi predicted, their work inspired more sophisticated models designed to mimic properties of real-world graphs. In their pre-internet article [ER59], they state

> This may be interesting not only from a purely mathematical point of view ... if one aims at describing such a real situation, one should replace the hypothesis of equiprobability of all connections by some more realistic hypothesis. It seems plausible that by considering the random growth of more complicated structures one could obtain fairly reasonable models of more complex real growth processes.

The Watts-Strogatz and Barabási-Albert models are two of many influential random graph models defined to produce graphs that have a particular structural feature observed in real world networks. The Watts-Strogatz *small world* model [WS98] generates graphs with small diameter and high clustering coefficient (the probability that two neighbors of a randomly selected vertex are adjacent). The Barabási-Albert *preferential attachment* model exhibits a power law in the distribution of vertex degrees. In this thesis, we will introduce two random graph models: the Random Overlapping Communities model and the Community Configuration model, both of which are designed to reflect community structure in sparse graphs.

1.1.2 The Szemerédi regularity lemma and graph limits

Informally, the Szemerédi regularity lemma states that a random graph model called the *Stochastic Block Model* can be used to succinctly summarize large dense graphs. The lemma guarantees the existence of a partition the vertices of a graph into a small number of blocks

such that the distribution of edges within a block and between most pairs of blocks resembles a random graph of prescribed density. The corresponding block model is a matrix that stores the densities between the partition classes. This representation encodes homorphism densities of the original graph (the chance of finding small subgraphs among randomly selected vertices) and the cut norm (which expresses the graph's cut structure). Moreover, sampling from the block model produces a smaller graph of desired size with similar properties as the original. This illustrates the first theme– using randomness to build a structured object.

The remarkable feature of the regularity lemma is that the size of the block model description is a constant, independent of the size of the original graph; the number of partition classes (and hence the total size of the description of the approximation) required to produce an approximation with absolute error $\varepsilon > 0$ in the cut norm is only a function of ε. Frieze and Kannan's weak regularity theorem [FK96] gives a weaker approximation (additive ε approximation in the cut norm), and the number of blocks needed is a much smaller function of ε. (Weak regularity requires $2^{2/\varepsilon}$ blocks, whereas the original form requires a tower function of height at least $1/\varepsilon^2$.) This work has lead to to efficient approximation algorithms for problems such as max-cut.

The closely related theory of graph limits shows that any sequence of graphs has a convergent subsequence, whose limit captures the limit of homomorphism densities and normalized cuts of the graphs, and is itself a probability distribution over the unit square (called a *graphon*). Moreover, if two graphs are close in the cut metric, then their homomorphism densities are also close. Qualitatively, these theorems give an essentially complete theory for the approximation of *dense* graphs, where the number of edges is $\Omega(n^2)$. For *sparse* graphs, where the number of edges is $o(n^2)$, the cut norm and homomorphism densities are trivially zero and so this theory is not insightful; the limit of any sequence of sparse graphs is the empty graphon.

Beyond approximating dense graphs, the regularity lemma has been fundamental in many

other areas of math including arithmetic combinatorics and extremal graph theory, see [KS96] for a survey. Next we will see that the regularity lemma is central in characterizing when property testing is possible.

1.1.3 Property testing on graphs

Property testing seeks to determine if an object satisfies a particular property or is far from any object satisfying the property. The notion was first introduced in [RS96] for functions with distance defined as the fraction of the domain where the functions disagree. In [GGR98] Goldreich et. al. formulate property testing for graphs. Given the ability to query whether or not a pair of vertices are adjacent in a graph G, the goal is determine with high probability if G satisfies property $\mathcal{P}$ or is ε-far from satisfying property $\mathcal{P}$, meaning no graph obtained by adding or deleting at εn^2 edges of G satisfies property $\mathcal{P}$. A property is said to be *testible* if there exists an algorithm that determines with accuracy at least $2/3$ whether G satisfies $\mathcal{P}$ or is ε-far from satisfying $\mathcal{P}$, and the total number of queries used is a function of ε (and therefore independent of the size of G). Randomized property testing algorithms exemplify the second theme of the thesis– using randomness to uncover some property of the system.

Using a variant of the regularity lemma, Alon and Shapira [AS08], showed that any monotone graph property is testible. Then Alon et. al. completely characterized testibility [Alo+09]; a graph property $\mathcal{P}$ testible if and only if knowing a regular partition of the graph is sufficient to determine whether G is close to satisfying $\mathcal{P}$.

1.2 Summary of contributions

This thesis contains four main research directions, united by the themes of using randomness to (i) construct structure and (ii) uncover structure. The middle two directions focus on developing a theory of sparse graph limits. The first and final directions exemplify these themes in other contexts. In the remainder of this Section, we introduce each research

directions (Sections 1.2.1 to 1.2.4). Then in Chapter 2, we give a formal introduction to graph limit theory (Chapter 2). Chapters 3 to 6 describe the above four research directions respectively.

1.2.1 Randomized construction of a threshold function

This research direction is joint work with Christos Papadimitriou and Santosh Vempala, published in [PPV16].

The motivation for this work is to construct a threshold function in neurally plausible way, meaning in a highly distributed manner using simple primitives.

We build a threshold function on n input bits[1] by constructing a circuit of AND/OR functions. A sequence of AND/OR functions can be represented as a tree in which the leaves are input bits and all other nodes are labeled with either "AND" or "OR". A node computes the function of its label applied to its children, and the tree returns the bit computed at the root. The basic algorithm for building a threshold function is as follows:

IterativeTree(L, m, C, X):

For each level j from 1 to L, apply the following iteration m times:
(level 0 consists of the input nodes X)

1. Choose a tree T according to C.
2. Choose nodes at random from the nodes on level $j-1$.
3. Build the tree T with these nodes as leaves. The root of T is a node on level $j+1$.

We show that a node at the highest level of the tree computes a t-threshold function whp. The distribution C over function trees determines the threshold t, and the width of the layers m and height L determine the accuracy. We show that it is possible to achieve any threshold t w.h.p. (with high probability) in this framework. The size of the function trees needed

[1] We say f is a t-threshold function if $f(X) = 1$ if and only if at least tn bits of X are one.

grows as t tends toward 0 or 1.

1.2.2 Approximating local structure with the random overlapping community (ROC) model

This Section describes joint work with Santosh Vempala, written in [PV18].

As discussed in Section 1.1.2, the regularity lemma and the subsequent theory of graphons provide an elegant framework for approximating dense graphs and describing the limit of dense graph sequences. Existing theories are limited in what they can achieve for families of graphs which are neither dense nor bounded-degree. In particular, they seem unable to answer the following representative question: *What is the limit of the sequence of hypercube graphs?*

We focus on approximating the simple cycle and closed walk counts of sparse graphs and graph sequences appropriately normalized. These counts encode information about the local structure of the graph and are related to its spectral properties; the number of closed k-walks in a graph is equal to the k^{th} moment of the graph's eigenspectrum.

We introduce a simple, easy to sample, random graph model that captures the limiting spectra of many sequences of interest, including the sequence of hypercube graphs. The Random Overlapping Communities (ROC) model is specified by a distribution on pairs (s, q), $s \in \mathbb{Z}_+$, $q \in (0, 1]$. A graph on n vertices is generated by repeatedly picking pairs (s, q) from the distribution, adding an Erdős-Rényi random graph of edge density q on a subset of vertices chosen by including each vertex with probability s/n, and repeating this process so that the expected degree is d. A variant of the model allows for bipartite communities. The selected subset of vertices is partitioned into two groups by assigning a vertex to either group with probability half, then a bipartite Erdős-Rényi random graph of edge density q is added. Our main results are as follows:

1. For almost all pairs of triangle-to-edge and four-cycle-to edge ratios, there exists a ROC construction with one community type that produces graphs with matching ratios.

2. We give a characterization based on the Stieltjes moment condition for when a vector of normalized closed walk counts can be realized by a family of ROC parameters.

3. We show that for any k, there exists parameters of the ROC model that produce graphs with the same normalized closed $3, 4 \ldots$ and k walk counts as the hypercube sequence. The core of the proof is showing that two Hankel matrices (consisting of the normalized closed walk counts of the hypercube) are positive semi-definite.

1.2.3 Sampling from sparse graphs with overlapping communities and heterogenous degrees

This section describes joint work with Christian Borgs, Jennifer Chayes, Souvik Dhara, and Subhabrata Sen.

Large graphs are ubiquitous in scientific and technological applications. However, computations involving the entire graph may be infeasible due to computational or privacy limitations. To this end, our goal is study small samples of large graphs. We study a specific notion of vertex sampling, referred to as p-sampling. A p-sample of a graph is obtained by selected each vertex with probability p and returning the induced subgraph on the selected vertices minus any isolated vertices (see Definition 5.1.1). This notion has its roots in the theory of sparse graph limits [VR16a, Bor+17a]. We demonstrate the power of p-samples in two ways: establishing connections among various graph models with overlapping communities proposed in the recent literature and detecting the presence of community structure.

We focus on the p-samples of graphs drawn from the Community Configuration model (CCM), a random graph model we designed to reflect the varied degree distributions and overlapping community structure characteristic of many real-world graphs. A CCM is determined by two parameters: a sequence of vectors containing colored half-edge counts and a matching that describes how to randomly pair the half-edges by color. Each color or pair of colors represent a community. Since vertices may have half-edges of many colors, the communities overlap.

First, we consider the limiting distributions of p-samples from increasingly large graphs drawn from CCMs and other popular models in the statistical physics literature including [HSM16, BKN11a, TMC16a]. By showing there exists parameterizations of the CCM and these models such that their limiting distributions of p-samples are the same, we establish a conceptual link between these models.

Second, we build on a large body of recent work that investigates the computational and information theoretic limits for community detection and clustering in stochastic block models (see [Abb17a]) by formulating analogous questions for CCMs. We describe a hypothesis test to determine if a graph came from a configuration model or a configuration model with community structure using only a p-sample. Under some mild assumptions, the test is accurate with probability $(1 - o(1))$ when p is $\Omega(\Delta^3/E(G)^2)$ where Δ denotes the maximum degree of G.

We note the CCM and ROC models are designed with different types of sparse graphs in mind. The p-samples cannot distinguish between approximately regular graphs with average degree $o(\sqrt{n})$ since the sample will almost always isolated edges for all such graphs. The p-sampling distribution is interesting when the degree distribution of the graph is highly varied, whereas a ROC graph cannot express graphs with a large range of degrees.

1.2.4 A Markov chain for the hard sphere model

This section describes joint work with Will Perkins and Tyler Helmuth.

For a fixed radius $r > 0$, the *hard sphere model* in a volume $\Lambda \subset \mathbb{R}^d$ at fugacity $\lambda \geq 0$ is a probability measure $\mu_{\Lambda,\lambda}$ on collections of non-overlapping spheres of radius r defined by conditioning a Poisson point process of intensity λ on Λ on the event that the points are at pairwise distance at least $2r$ and distance at least r from Λ^c. When the fugacity parameter λ is higher the probability measure favors configurations with more spheres. Conditioned on k the number of spheres, the distribution is uniform over all sphere packings of Λ with k

spheres.

It is an open mathematical problem to prove the existence of a phase transition in the hard sphere model. The *critical fugacity* $\lambda_c(d)$ is the supremum over λ such that the hard sphere model has a unique infinite volume limit in the sense of van Hove, i.e., such that the set of weak limit points of $\{\mu_{\Lambda,\lambda}\}_\Lambda$ is a singleton set. The *critical density* is the limiting expected packing density of the hard sphere model at the critical fugacity. We improve upon all the best known lower bounds on the critical fugacity and critical density of the hard sphere model in dimensions two and higher. As the dimension tends to infinity our improvements are by factors of 2 and 1.7, respectively.

We describe *single-center dynamics* and show that these dynamics are a rapidly mixing Markov chain for the hard sphere model at fugacity $\lambda < 2^{1-d}$. To prove this, we follow the approach of [Vig01], and apply path coupling with an optimized metric. Then using an equivalence between optimal spatial and temporal mixing for hard spheres, we show that 2^{1-d} is also an upper bound for the critical fugacity. Using lower bounds on the expected packing density of the hard sphere model, we translate this result to a lower bound on the critical density.

CHAPTER 2

GRAPH LIMIT BACKGROUND

In this Section, we formalize the definitions and theorems discussed in Section 1.1.2.

2.1 Szemerédi's Regularity Lemma

Definition 2.1.1. Let $G = (V, E)$ be a graph. For subsets $X, Y \subseteq V$, let $d_G(X, Y) = \frac{e_G(X,Y)}{|X||Y|}$ denote the density of edges between X and Y. A pair of disjoint subsets X and Y is ε-regular if

$$|d_G(X, Y) - d_G(A, B)| < \varepsilon$$

for all $A \subseteq X$ and $B \subseteq Y$ satisfying

$$|A| > \varepsilon|X| \quad \text{and} \quad |B| > \varepsilon|Y|.$$

Definition 2.1.2. A partition $\{V_0, V_1, \dots V_k\}$ of a vertex set V is ε-regular if the following three conditions hold:

(i) $|V_0| \leq \varepsilon|V|$,

(ii) $|V_0| = |V_1| = \cdots = |V_k|$,

(iii) all but at most εk^2 of the pairs V_i, V_j with $1 \leq i < j \leq k$ are ε-regular.

Theorem 2.1.3 (Szemerédi's Regularity Lemma, [Sze75, Sze76])**.** *For all $\varepsilon > 0$ and integers $m \geq 1$, there exists an integer $M = M(\varepsilon)$ such that every graph on at least m vertices has a ε-regular partition $\{V_0, V_1, \dots V_k\}$ with $m \leq k \leq M$.*

The proof of the lemma gives and upper bound on $M(\varepsilon)$ that is a tower function of height $1/\varepsilon$; similar lower bounds are given in [Gow97]. Frieze and Kannan developed weaker notion of regularity that requires fewer partition classes [FK96].

Definition 2.1.4. A partition $V_1, V_2, \dots V_k$ is an ε-regular-FK partition if

$$\left| e_G(S,T) - \sum_{i,j=1}^{k} d_G(V_i, V_j)|S \cap V_i||T \cap V_j| \right| \leq \varepsilon |V|^2$$

for all $S, T \subseteq V$.

Theorem 2.1.5 ([FK96], in the notation of [FLZ19])**.** *Every graph has an ε-regular-FK partition with at most $2^{2/\varepsilon}$ parts.*

2.2 Graph limits: graphons

Next we review graphons and the associated theory of graph limits. All results in this section are as stated in [Lov12].

Definition 2.2.1. A bounded symmetric measurable function $W : [0,1]^2 \to [0,1]$ is called a graphon. We denote the set of graphons $\mathcal{W}$.

The empirical graphon is a graphon representation of a graph

Definition 2.2.2. Let G be a simple graph on a set of n vertices uniquely labeled with $[n]$. The empirical graphon $f^G : [0,1]^2 \to [0,1]$ is defined as follows:

$$f^G(x,y) = \begin{cases} 1 & (\lceil nx \rceil, \lceil ny \rceil) \text{ is an edge in } G \\ 0 & \text{otherwise.} \end{cases}$$

In order to view graphons as limit objects, one needs to construct a space of graphons and introduce a notion of distance. The cut metric will be used for this purpose. We begin

by defining the cut metric on graphs, and then we will extend the definition to graphons.

Definition 2.2.3. The cut distance between two graphs G and G' on the same vertex set is

$$d_{\square}(G, G') = \max_{U,W \subset V} \frac{|e_G(U,W) - e_{G'}(U,W)|}{|V|^2}.$$

Definition 2.2.4. The cut norm between two graphons $U, W \in \mathcal{W}$ is

$$d_{\square}(U, W) = \max_{S,T \subseteq [0,1]} \left| \int_{S \times T} U(x,y) - W(x,y) dx dy \right|.$$

Note that the above metric depends on the labeling of vertex set of the graph or the embedding into $[0,1]$. Next we define the cut distance, a pseudometric to compare the cut structure of unlabeled graphons. Let $S_{[0,1]}$ be the group of invertible measure preserving maps $[0,1] \to [0,1]$.

Definition 2.2.5. The cut distance between two graphons $U, W \in \mathcal{W}$ is

$$\delta_{\square}(U, W) = \inf_{\psi \in S_{[0,1]}} d_{\square}(U, W^{\psi}).$$

Define the relation $U \sim W$ if $\delta_{\square}(U, W) = 0$ and let $\tilde{\mathcal{W}} = \mathcal{W}/_{\sim}$ be the set of unlabeled graphons.

Theorem 2.2.6. *The space $(\tilde{\mathcal{W}}, \delta_{\square})$ is compact.*

The cut structure of a graph is a global property. It turns out that if two graphs are close in cut distance, they must also look similar locally in terms of how often particular subgraphs appear. For a finite graph $H = (V, E)$ and graphon W, define the homomorphism density

$$t(H, W) = \int_{[0,1]^V} \prod_{ij \in E} W(x_i, x_j) \prod_{i \in V} dx_i.$$

Lemma 2.2.7 (Counting Lemma)**.** *Let H be a simple graph and let $W, U \in \mathcal{W}_0$ be graphons. Then*

$$|t(H, W) - t(H, U)| \leq |e(H)|\delta_{\square}(W, U).$$

Lemma 2.2.8 (Inverse Counting Lemma)**.** *Let $k \in \mathbb{Z}^+$ and let $W, U \in \mathcal{W}_0$ be graphons. Assume that for every simple graph H on k nodes,*

$$|t(H, U) - t(H, W)| \leq 2^{-k^2}.$$

Then

$$\delta_{\square}(W, U) \leq \frac{50}{\sqrt{\log(k)}}.$$

Theorem 2.2.9. *Let (W_n) be a sequence of graphons in $\mathcal{W}_0$ and let $W \in \mathcal{W}_0$. The sequence $t(H, W_n)$ converges for all finite simple graphs H if and only if W_n is Cauchy in the $\delta_{\square}$ distance. Moreover, $t(H, W_n) \to t(H, W)$ for all finite simple graphs H if and only if $\delta_{\square}(W_n, W) \to 0$.*

CHAPTER 3
A RANDOMIZED CONSTRUCTION OF A THRESHOLD FUNCTION

This chapter is joint work with Christos Papadimitriou and Santosh Vempala, and appears as [PPV16].

3.1 Background and summary of results

We study a family of simple algorithmic processes motivated by neurally feasible computation. In particular, we focus on Boolean functions of an arbitrary number of input variables that can be realized by simple iterative constructions based on constant-size primitives. This restricted type of construction needs little global coordination or control and thus is a candidate for neurally feasible computation. We generalize Valiant's recursive construction of a majority function to realize any uniform threshold function in this neurally plausible manner. We study the rate of convergence, finding that while linear convergence to the correct function can be achieved for any threshold using a fixed set of primitives, for quadratic convergence, the size of the primitives must grow as the threshold approaches 0 or 1.

We also study finite realizations of this process and the learnability of the functions realized. We show that the constructions realized are accurate outside a small interval near the target threshold, where the size of the construction grows as the inverse square of the interval width. This phenomenon, that errors are higher closer to thresholds (and thresholds closer to the boundary are harder to represent), is a well-known cognitive finding.

Cortical computation. Among the many unexplained abilities of the cortex are learning complex patterns and invariants from relatively few examples. This is manifested in a range of cognitive functions including visual and auditory categorization, motor learning and language.

In spite of the highly varied perceptual and cognitive tasks accomplished, the substrate appears to be relatively uniform in the distribution and type of cells. How could these 80 billion cells organize themselves so effectively?

Cortical algorithms must therefore be highly distributed, require little *synchrony* (number of pairs of events that must happen in lock-step across neurons), little *global control* (longest chain of events that must happen in sequence) and be based on a small number of very simple primitives [PV15a]. Assuming that external stimuli are parsed as sets of binary sensory features, our central question is the following:

What functions can be represented and learned by cortical algorithms?

Perhaps the most natural primitives are the AND and OR functions on two input variables. These functions are arguably neurally plausible. They were studied as JOIN and LINK by [Val94, Val00, Val05, FV09], who showed how to implement them in the neuroidal model. An *item* is a collection of neurons (corresponding to a *neural assembly* in neuroscience) that represents some learned or sensed concept. Given two items A, B, the JOIN operation forms a new item $C = \text{JOIN}(A, B)$, which "fires" when both A and B fire, i.e., C represents $A \wedge B$. $\text{LINK}(A, B)$ captures association, and causes B to fire whenever A fires. By setting $\text{LINK}(A, C)$ and $\text{LINK}(B, C)$, we achieve that C is effectively $A \vee B$. While the precise implementation and neural correlates of JOIN and LINK are unclear, there is evidence that the brain routinely engages in hierarchical memory formation.

Monotone Boolean functions. Functions constructed by recursive processes based on AND/OR trees have been widely studied in the literature, motivated by the design of reliable circuits as in [MS56] and more recently, understanding the complexity-theoretic limitations of monotone Boolean functions. One line of work studies the set of functions that could be the limits of recursive processes, where at each step, the leaves of a tree are each replaced by constant-size functions. [MS56], showed that a simple recursive construction leads to

stable circuits and to threshold functions. [Val84] used their 4-variable primitive function $(A \vee B) \wedge (C \vee D)$ to derive a small depth and size threshold function that evaluates to 1 if at least $(2 - \phi) \approx 0.38$ fraction of the inputs are set to 1 and to zero otherwise. The depth and size were $O(\log n)$ and $O(n^{5.3})$ respectively. Calling it the *amplification method*, [Bop85] showed that Valiant's construction is optimal. [DZ92] extended the lower bound to classes of read-once formulae. [HMP06] gave smaller size Boolean *circuits* (where each gate can have fan out more than 1), of size $O(n^3)$ for the same threshold function. [LMS98] gave an alternative analysis of Valiant's construction along with applications to coding. The construction of a Boolean formula was extended by [Ser04] to monotone linear threshold functions, in that they can be approximated on most inputs by monotone Boolean formulae of polynomial size. [Fri86] gave more efficient constructions for threshold functions with small thresholds.

Savicky gives conditions under which the limit of such a process is the uniform distribution on all Boolean functions with n inputs [Sav87, Sav90] (see also [BP05, FGG09]). In a different application, [GKS93] showed how to use properties of these constructions to identify read-once formulae from their input-output behavior.

Our work. Unlike previous work, where a single constant-sized function is chosen and applied recursively, we will allow constructions that randomly choose one of two constant-sized functions. To be neurally plausible, our constructions are *bottom-up* rather than *top-down*, i.e., at each step, we apply a constant-size function to an existing set of outputs. In addition, the algorithm itself must be very simple — our goal is not to find ways to realize all Boolean functions or to optimize the size of such realizations. Here we address the following questions: What functions of n input items can be constructed by cortical algorithms in this iterative manner? Can arbitrary uniform threshold functions be realized? What size and depth of iterative constructions suffices to guarantee accurate computations? Can such functions and constructions be learned from examples, where the learning algorithm is also neurally

plausible?

Our rationale for uniform threshold functions is two-fold. First, uniform threshold functions are fundamental in computer science and likely also for cognition. Second, the restriction to JOIN and LINK as primitives ensures that any resulting function will be monotone since negation is not possible in this framework. Moreover, if we require the construction to be symmetric, it would seem that the only obtainable family of Boolean functions are uniform thresholds. However, as we will see, there is a surprise here, and in fact we can get *probabilistic thresholds*, i.e., functions whose output is 0 in an interval including 0, then equally likely to be 0 or 1 in an interval, then 1 in an interval including 1. To be able to describe our results precisely, we begin with a definition of iterative constructions.

3.1.1 Iterative constructions

A sequence of AND/OR operations can be represented as a tree. Such a tree T with n leaves naturally computes a function $g_T : \{0,1\}^n \to \{0,1\}$. We can build larger trees in a neurally plausible way by using a set of small AND/OR trees as building blocks. Let C be a probability distribution on a finite set of trees. We define a *iterative tree for* C as follows.

IterativeTree(L, m, C, X)**:**

For each level j from 1 to L, apply the following iteration m times:
(level 0 consists of the input items X)

1. Choose a tree T according to C.
2. Choose items at random from the items on level $j-1$.
3. Build the tree T with these items as leaves.

The construction of small AND/OR trees is a decentralized process requiring a short sequence of steps, i.e., the synchrony and control parameters are small. Therefore, we consider them to be neurally plausible.

The above construction has a well-defined sequence of *levels*, with items from the next level having leaves only in the current level. A construction that needs even less coordination is the following: each item has a *refractory* period and the probability that it participates in future item creation decays exponentially with time. The weight of an item starts at 1 when it is created and decays by a factor of $e^{-\alpha}$ each time unit. We refer to such constructions as *exponential* iterative constructions. An extreme version of this, which we call *wild iterative* construction, is to have $\alpha = 0$, i.e, all items are equally likely to participate in the creation of new items.

3.1.2 Results

We are interested in the functions computed by iterative constructions. In the limit with respect to the width of the levels, the function computed by a high-level item of the iterative tree does not depend on the width of each level. However, as we discuss in Section 3.4, in a "bottom-up" construction in which the items at level $j-1$ are fixed before the items at level j are created, the width of the levels becomes important. The smaller the width of the levels the more likely the function computed deviates from expectation. The following theorems describe *in the limit* with respect to the width to the levels, the probability that a high level item of an iterative tree computes a threshold function.

To start, we restate Valiant's result [Val84] as a bottom-up construction. Here $\phi = (\sqrt{5}+1)/2$ is the golden ratio ($2-\phi \approx 0.38$).

Theorem 3.1.1. *Let R be the tree that computes $(A \vee B) \wedge (C \vee D)$. Then, in the limit, an item at level $\Omega(\log n + \log k)$ of an iteratively constructed tree for R computes a $(2-\phi)$-threshold function accurately with probability at least $1 - 2^{-k}$.*

In this construction, the iterative tree that computes the $2-\phi$ threshold function is built using only one small tree. We show that it is possible to achieve arbitrary threshold functions

if we allow our iterative tree to be built according to a probability distribution on two distinct smaller trees.

Theorem 3.1.2. *Let $0 < t < 1$ and let $R = \{\mathbb{P}(T_1) = t, \mathbb{P}(T_2) = 1 - t\}$ where T_1 is the tree that computes $(A \vee B) \wedge C$ and T_2 is the tree that computes $(A \wedge B) \vee C$. Then, in the limit, an item at level $\Omega(\log n + k)$ of an iteratively constructed tree for R computes a t-threshold function accurately with probability at least $1 - 2^{-k}$.*

The rate of convergence of this more general construction is linear rather than quadratic. While both are interesting, the latter allows us to guarantee a correct function on every input with depth only $O(\log n)$, since there are 2^n possible inputs.

Definition 3.1.3. A construction exhibits *linear convergence* if in expectation items at level $\Omega(\log n + k)$ accurately compute the threshold function with probability at least $1 - 2^{-k}$. A construction exhibits *quadratic convergence* if in expectation items at level $\Omega(\log n + \log k)$ accurately compute the threshold function with probability at least $1 - 2^{-k}$.

The next theorem gives constructions using slightly larger trees with 4 and 5 leaves respectively (illustrated in Figure 3.1) that converge quadratically to a t-threshold function for a range of values of t, with more leaves giving a larger range. Moreover, these ranges are tight, i.e. no construction on trees with 4 or 5 leaves yields quadratic convergence to a t-threshold function for t outside these ranges.

Theorem 3.1.4. *(A) Let $2 - \phi \leq t \leq \phi - 1$ and $\alpha(t) = \frac{1-t-t^2}{2t(t-1)}$. Define $R = \{\mathbb{P}(F_1) = \alpha(t), \mathbb{P}(F_2) = 1 - \alpha(t)\}$ be the probably distribution on trees in Figure 3.1. Then, in the limit, an item at level $\Omega(\log n + \log k)$ of an iteratively constructed tree for R computes a t-threshold function accurately with probability at least $1 - 2^{-k}$. Moreover, for t outside this range, there exists no such construction on trees with four leaves that converge quadratically to a t-threshold function.*

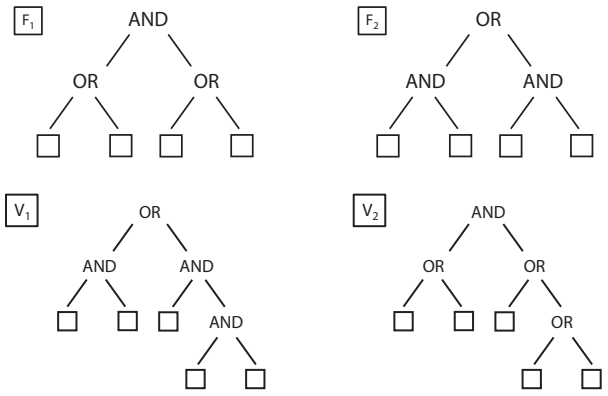

Figure 3.1: For $.38 \lesssim t \lesssim 0.62$, there exists a probability distribution on F_1 and F_2 that yields an iterative tree that converges quadratically to a t-threshold function. For $.26 \lesssim t \lesssim 0.74$, there exists a probability distribution on V_1 and V_2 that yields an iterative tree that converges quadratically to a t-threshold function.

(B) Let $\alpha(t) = \frac{-1+5t-4t^2+t^3}{5t(t-1)}$ *and let* t *be a value for which* $0 \leq \alpha(t) \leq 1$*, so* $0.26 \lesssim t \lesssim 0.74$*. Let* $R = \{\mathbb{P}(V_1) = \alpha(t), \mathbb{P}(V_2) = 1 - \alpha(t)\}$ *be the probably distribution on trees in Figure 3.1. Then, in the limit, an item at level* $\Omega(\log n + \log k)$ *of an iteratively constructed tree for* R *computes a* t*-threshold function accurately with probability at least* $1 - 2^{-k}$ *. Moreover, for* t *outside this range, there exists no such construction on trees with five leaves that converges quadratically to a* t*-threshold function.*

As the desired threshold t approaches 0 or 1, we show that an iterative tree that computes the t-threshold function must use increasingly large trees as building blocks.

Theorem 3.1.5. *Let* t *be a threshold,* $0 < t < 1$ *and let* $s = \min\{t, 1 - t\}$*. Then, the construction of an iterative tree whose level* $\Omega(\log n + \log k)$ *items compute a* t*-threshold function with probability at least* $1 - 2^{-k}$ *must be defined over a probability distribution on trees with at least* $\frac{1}{\sqrt{2s}}$ *leaves.*

This raises the question of whether it is possible to have quadratic convergence for any threshold. We can extend the constructions described in Theorem 3.1.4 by using analogous trees with six and seven leaves to obtain quadratic convergence for thresholds in the ranges

$0.15 \lesssim t \lesssim 0.85$ and $0.11 \lesssim t \lesssim 0.89$ respectively. However, it is not possible to generalize this construction beyond this point. Instead, we observe the emergence of probabilistic thresholds.

We define A_k as a tree on $2k$ leaves that computes $(x_1 \vee x_2 \vee \cdots \vee x_k) \wedge (x_{k+1} \vee x_{k+2} \vee \cdots \vee x_{2k})$ and B_k as a tree on $2k$ leaves that computes $(x_1 \wedge x_2 \wedge \cdots \wedge x_k) \vee (x_{k+1} \wedge x_{k+2} \wedge \cdots \wedge x_{2k})$.

Theorem 3.1.6. *Let $k \geq 4$. Consider an iterative construction in which A_k and B_k are each selected with probability $1/2$. Then there exists $\varepsilon > 0$ such that for inputs in which the fraction of items firing is in the range $[1/2 - \varepsilon, 1/2 + \varepsilon]$, high level items of the iterative construction fire with probability half.*

To achieve quadratic convergence for thresholds near the boundaries, we turn to the following construction, which asymptotically matches the lower bound of Theorem 3.1.5.

Theorem 3.1.7. *For any $0 < t \leq 2 - \phi$, there exists k and a probability distribution on A_k and A_{k+1} that yields an iterative tree with quadratic convergence to the uniform t-threshold function. Similarly for any $\phi - 1 \leq t < 1$, there exists k and a probability distribution on B_k and B_{k+1} that yields an iterative tree with quadratic convergence to the uniform t-threshold function.*

There is a trade-off between constructing iterative trees that converge faster and requiring minimal coordination in order to build the subtrees. Building a specified tree on a small number of leaves requires less coordination than building a specified tree on many leaves. Therefore, as t approaches 0 or 1, constructing an iterative tree with quadratic convergence becomes less neurally feasible because the construction of each subtree requires much coordination. These results are in line with behavioral findings [Ros+76, Ros78] and computational models [AV06, Arr+15] about categorization being easier when concepts are more robust.

Next we turn to finite realizations of iterative trees. The above theorems analyze the behavior of an iterative construction in the limit with respect to the width of the levels. We assumed that for any input the number of items turned on at given level of the tree is equal

to its expectation. This assumption holds when the width of each level is infinite; however, imagining a "bottom up" construction, we note that the chance that the number of items firing at a given level deviates from expectation is non-trivial. Such deviations percolate up the tree and effect the probability that high level items compute the threshold function accurately. The smaller the width of a level, the more likely that the number of items on at that level deviates significantly from expectation, rendering the tree less accurate. How large do the levels of an iteratively constructed tree need to be in order to ensure a reasonable degree of accuracy?

Theorem 3.1.8. *Consider a construction of a t-threshold function with quadratic convergence described in Theorem 3.1.4 or Theorem 3.1.7 in which each level ℓ has m_ℓ items and the fraction of input items firing is at least ε from the threshold t. Then, with probability at least $1-\gamma$, items at level $\Omega(\log \frac{1}{\gamma} + \log \frac{1}{\varepsilon}))$ will accurately compute the threshold function for $m_1 = \Omega\left(\frac{\ln(1/\gamma)}{\varepsilon^2}\right)$ and $\sum_\ell m_\ell = O(m_1)$.*

As a direct corollary, by setting $\varepsilon = O(1/n)$ and $\gamma = 2^{-n-1}$, we realize a t-threshold construction of size $O(n^3)$ for any t, matching the best-known construction which was for a specific threshold [HMP06]. The finite-width version of Theorem 3.1.2 is given in Section 3.4. The exponential iterative construction also converges to a t-threshold function. We give the statement here for the wild iterative construction (with no weight decay).

Theorem 3.1.9. *Consider a wild iterative construction on n inputs corresponding to a polynomial f that converges to a t-threshold function. Then after $O\left(\frac{n^2}{\varepsilon}\right)$ items are created in the wild iterative construction, the next item will accurately compute the threshold function with probability at least $1-\varepsilon$.*

Finally, in Section 3.5, we give a simple cortical algorithm to learn a uniform threshold function from a single example. We also discuss how a more complex setting of a noisy threshold function with monotone noise can be learned from multiple examples.

Theorem 3.1.10. *Let $X \in \{0,1\}^n$ such that $||X||_1 = tn$, $L = \Omega\left(\log \frac{1}{\gamma} + \log \frac{1}{\varepsilon}\right)$, and $\varepsilon = \Omega\left(\sqrt{\frac{\ln(1/\gamma)}{m}}\right)$. Then, on any input in which the fraction of input items firing is outside $[t-\varepsilon, t+\varepsilon]$, items at level L of an iterative tree produced by LearnThreshold(L, m, X) will compute a t-threshold function with probability at least $1-\gamma$.*

The next section provides the groundwork for these theorems, and the proofs are in Sections 3.3, 3.4, 3.5 and 3.6. We discuss several open questions and directions for future research in Section 3.7.

3.2 Polynomials of AND/OR Trees

Let $g_T : \{0,1\}^n \to \{0,1\}$ be the Boolean function computed by an AND/OR tree T with n leaves. We define f_T as the probability that T evaluates to 1 if a p-fraction of input items are set to 1.

$$f_T(p) = \mathbb{P}\left(g_T(X) = 1 \,|\, \|X\|_1 = pn\right).$$

We analogously define $f_C(p)$ for probability distributions on trees; let f_C be the probability that a tree chosen according to C evaluates to 1 if a p-fraction of input items are set to 1. Let λ_T be the probability of T in distribution C. We have

$$f_C(p) = \sum_{T \in C} \lambda_T f_T(p).$$

In an iterative construction for the probability distribution C, an item at level k evaluates to 1 with probability $f_C(p_{k-1})$ where p_{k-1} is the probability that an item at level $k-1$ evaluates to 1. Thus, in expectation the probability that items at level k evaluate to 1 is $f_C^{(k)}(p)$ where p is the probability an input is set to 1. This follows directly from the recurrence relation:

$$f_C^{(k)}(p) = f_C(f_C^{(k-1)}(p)).$$

In the remainder of this section, we collect properties of polynomials of AND/OR trees to be used in the analysis of iterative trees. The proofs are provided in Section 3.6.1.

We call a polynomial *achievable* if it can be written as f_T for some AND/OR tree T. We call a polynomial *achievable through convex combinations* if it can be written as f_C for some probability distribution on AND/OR trees C. Table 3.1 lists all achievable polynomials with degree at most five. Note that $\mathcal{A}$ is closed under the AND and OR operations. If $a, b \in \mathcal{A}$, then $a \cdot b \in \mathcal{A}$ and $a + b - a \cdot b \in \mathcal{A}$. The set of polynomials achievable through convex combinations is the convex hull of $\mathcal{A}$.

Degree	Polynomials in $\mathcal{A}$
1	(0,1)
2	(0, 0, 1) (0, 2, -1)
3	(0, 0, 0, 1) (0, 1, 1, -1) (0, 0, 2, -1) (0, 3, -3, 1))
4	((0, 0, 0, 0, 1) (0, 1, 0, 1, -1) (0, 0, 1, 1, -1) (0, 2, 0, -2, 1) (0, 0, 0, 2, -1) (0, 1, 2, -3, 1) (0, 0, 3, -3, 1) (0, 4, -6, 4, -1) (0, 0, 2, 0, -1) (0, 0, 4, -4, 1))

Degree	Polynomials in $\mathcal{A}$
5	(0, 0, 0, 0, 0, 1) (0, 1, 0, 0, 1, -1) (0, 0, 1, 0, 1, -1) (0, 2, -1, 1, -2, 1) (0, 0, 0, 1, 1, -1) (0, 1, 1, 0, -2, 1) (0, 0, 2, 0, -2, 1) (0, 3, -2, -2, 3, -1) (0, 0, 0, 0, 2, -1) (0, 1, 0, 2, -3, 1) (0, 0, 1, 2, -3, 1) (0, 2, 1, -5, 4, -1) (0, 0, 0, 3, -3, 1) (0, 1, 3, -6, 4, -1) (0, 0, 4, -6, 4, -1) (0, 5, -10, 10, -5, 1) (0, 0, 0, 2, 0, -1) (0, 1, 2, -2, -1, 1) (0, 0, 0, 4, -4, 1) (0, 1, 4, -8, 5, -1) (0, 0, 1, 1, 0, -1) (0, 0, 3, -1, -2, 1) (0, 0, 2, 1, -3, 1) (0, 0, 6, -9, 5, -1)

Table 3.1: Achievable polynomials for AND/OR trees. The polynomial $a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 + a_5x^5$ is denoted by $(a_0, a_1, a_2, a_3, a_4, a_5)$.

Lemma 3.2.1. *Let $\mathcal{A}$ be the set of achievable polynomials. Let $A(x) = a_0 + a_1x_1 + \cdots + a_nx^n$ be a polynomial in $\mathcal{A}$. Then,*

1. $a_0 = 0$
2. $a_n = -1$ *or* 1
3. $\sum_{i=0}^{n} a_i = 1$
4. *If $A(x)$ has degree d, then $A(x)$ is the polynomial for a tree on d leaves.*

Next we prove bounds on the values of the coefficients of achievable polynomials.

Lemma 3.2.2. *Let $f \in \mathcal{A}$ be an achievable polynomial of degree d, $f = a_0 + a_1x + a_2x^2 + \ldots a_dx^d$. Then $|a_\ell| \leq d^\ell$.*

We observe a relationship between the polynomial of a tree and the polynomial of its complement. We define the complement of the AND/OR tree T to be the tree obtained from T by switching the operation at each node.

Lemma 3.2.3. *Let A and B be complementary AND/OR trees and let f_A and f_B be the corresponding polynomials. Then $f_B(1-p) = 1 - f_A(p)$ for all $0 < p < 1$.*

Let f_A be a polynomial achievable through convex combinations, $f_A = \sum_{i=1}^{n} \lambda_i f_{A_i}$. Let A_i and B_i be complementary AND/OR trees. Let $f_B = \sum_{i=1}^{n} \lambda_i f_{B_i}$. We say that f_A and f_B are *complementary polynomials*.

Corollary 3.2.4. *Let f_A and f_B be complementary polynomials. Then*

1. *For all $0 < p < 1$, $f_B(1-p) = 1 - f_A(p)$*
2. *If p is a fixed point of f_A then $1-p$ is a fixed point of f_B*
3. *For all $0 < p < 1$, $f_B^{(k)}(1-p) = 1 - f_A^{(k)}(p)$.*

Finally, we make some observations about the polynomials associated with the specific family of trees we use in many of our constructions.

Definition 3.2.5. Let A_k be a tree on $2k$ leaves that computes $(x_1 \vee x_2 \vee \cdots \vee x_k) \wedge (x_{k+1} \vee x_{k+2} \vee \cdots \vee x_{2k})$. Let B_k be a tree on $2k$ leaves that computes $(x_1 \wedge x_2 \wedge \cdots \wedge x_k) \vee (x_{k+1} \wedge x_{k+2} \wedge \cdots \wedge x_{2k})$.

Lemma 3.2.6. *Let f_{A_k} and f_{B_k} be the polynomials corresponding to A_k and B_k respectively. Then f_{A_k} has a unique fixed point in the interval $\left(\frac{1}{k^2}, \frac{1}{k(k-1)}\right)$ and f_{B_k} has a fixed point in the interval $\left(1 - \frac{1}{k(k-1)}, 1 - \frac{1}{k^2}\right)$.*

Lemma 3.2.7. *Let $0 \leq \alpha \leq 1$ and $f = \alpha f_{B_k} + (1-\alpha) f_{B_{k+1}}$ where f_{B_k} is the polynomial corresponding to B_k. Let t be the fixed point of f in $(0,1)$. Then $g(p) = \frac{f(p)-p}{p(1-p)(p-t)} \geq \frac{1}{t}$ for all $p \in [0,1]$.*

Lemma 3.2.8. *Let f_{A_k} and f_{B_k} be the polynomials corresponding to A_k and B_k respectively. For $t \leq 2 - \phi$, there exists some k and α such that $f_A = \alpha f_{A_k} + (1-\alpha) f_{A_{k+1}}$ has fixed point t. Moreover, $\frac{t - f_A(p)}{t-p} \geq \left(1 + \frac{p(1-p)}{t}\right)$. Similarly, for $t \geq \phi - 1$, there exists some k and α such that $f_B = \alpha f_{B_k} + (1-\alpha) f_{B_{k+1}}$ has fixed point t. Moreover, $\frac{t - f_B(p)}{t-p} \geq \left(1 + \frac{p(1-p)}{t}\right)$.*

3.3 Convergence of iterative trees to threshold functions

In the previous section, we showed that in the limit with respect to the width of the levels items at level k of an iterative tree evaluate to 1 with probability $f_C^{(k)}(p)$ when the inputs are set according to a Bernoulli distribution with probability p. In this section, we demonstrate ways of selecting C so that $f_C^{(k)}(p)$ converges to a t-threshold function.

By an abuse of notation, we say that $f(p)$ converges to a t-threshold function if

$$\lim_{k \to \infty} f^{(k)}(p) = \begin{cases} 1 & 0 \leq p < t \\ 0 & t < p \leq 1 \\ p & p = r. \end{cases}$$

Moreover, we say that f converges quadratically to a t-threshold function if the corresponding iterative construction exhibits quadratic convergence. The function depicted in Figure 3.2 converges to a t-threshold function.

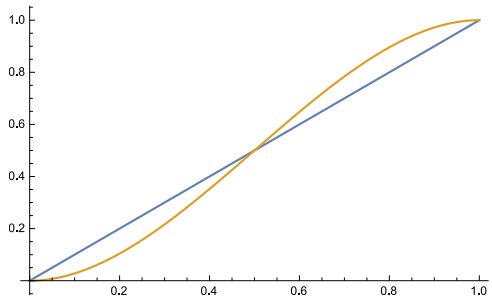

Figure 3.2: A function that converges to a 1/2-threshold function.

Lemma 3.3.1. *The fixed points of $f_C(p)$ are the roots of the polynomial $f_C(p) - p$.*

We now prove that the construction described in Theorem 3.1.2 converges to a t-threshold function.

of Thm. 3.1.2. Let f_R be the polynomial that describes the iterative construction in which T_1 and T_2 are selected with probability t and $1-t$ respectively. Since, $f_{T_1}(p) = 2p^2 - p^3$ and $f_{T_2}(p) = p + p^2 - p^3$,

$$f_R(p) = tf_{T_1}(p) + (1-t)f_{T_2}(p) = (1-t)p + (1+t)p^2 - p^3.$$

Since $f_R(p) - p = p(1-p)(p-t)$, the fixed points of f_R are $0, t$, and 1. We claim that f_R exhibits linear convergence to a t-threshold function.

Let p be the probability that an input item fires. It suffices to consider the case when $p \leq t - 1/n$. By Corollary 3.2.4, convergence to 1 for $p \geq t + \frac{1}{n}$ follows from the complementary construction.

First we show that the probability an item at level $\Omega(\log n)$ fires is less than $\frac{t}{2}$. By definition $p - f(p) = p(1-p)(t-p)$. Observe that for $t/2 < p \leq t - 1/n$

$$\frac{t - f(p)}{t - p} = 1 + \frac{p - f(p)}{t - p} = 1 + p(1-t) \geq 1 + \frac{t(1-t)}{2}.$$

It follows that

$$t - f^{(l)}(p) \geq \left(1 + \frac{t(1-t)}{2}\right)^l (t - p) \geq \left(1 + \frac{t(1-t)}{2}\right)^l \frac{1}{n}.$$

For $l = \log_{1+\frac{t(1-t)}{2}} \frac{tn}{2}$, $f^l(p) < \frac{t}{2}$.

Next, we show that at $\Omega(k)$ additional levels, the probability an items fires is less than 2^{-k}. For $p < \frac{t}{2}$,

$$f(p) = p(1-p)(p-t) + p = p(1 - (1-p)(t-p)) \leq p\left(1 - \left(1 - \frac{t}{2}\right)\frac{t}{2}\right).$$

It follows

$$f^{(l)}(p) < \left(1 - \left(1 - \frac{t}{2}\right)\frac{t}{2}\right)^l p < \left(1 - \left(1 - \frac{t}{2}\right)\frac{t}{2}\right)^l \frac{t}{2}.$$

Thus, for $l = \log_{\left(1-\left(1-\frac{t}{2}\right)\frac{t}{2}\right)} \frac{1}{t2^{k-1}}$, $f^l(p) < 2^{-k}$. We have shown that when the input items fire with probability $p \leq t - 1/n$, items level $\Omega(k + \log n)$ will evaluate to 1 with probability less than 2^{-k}. □

3.3.1 Quadratic convergence from iterative trees with small building blocks

In this section we show that using trees with four or five leaves as building blocks, we can construct an iterative tree that converges quadratically to a t-threshold function for restricted values of t. We begin with a lemma that provides sufficient conditions for quadratic convergence.

Lemma 3.3.2. *Let f be a function corresponding to an iterative construction on n inputs that satisfies the following conditions:*

- *On the interval $[0,1]$, f has precisely three fixed points: $0, t,$ and 1.*
- *(Linear Divergence) There exists constants u, v satisfying $0 < u < t$ and $t < v < 1$ and constants $c_1, c_2 > 1$ such that*
  1. *$t - f(p) \geq c_1(t-p)$ for $p \in [u, t - \frac{1}{n}]$, and*
  2. *$f(p) - t \geq c_2(p-t)$ for $p \in [t + \frac{1}{n}, v]$.*
- *(Quadratic Convergence) For the constants u, v as above, there exists constants c_3, c_4 such that $c_3 u < 1$ and $c_4(1-v) < 1$ and*
  1. *$f(p) < c_3 p^2$ for $p \in (0, u)$, and*
  2. *$1 - f(p) < c_4(1-p)^2$ for $p \in (v, 1)$.*

Then f exhibits quadratic convergence to a t-threshold function, meaning that in expectation items at level $\Omega(\log n + \log k)$ of the corresponding iterative construction compute a t-threshold function with probability at least $1 - 2^{-k}$.

Proof. Let p be the probability an input item fires. First we consider the case when $p \leq t - \frac{1}{n}$. By the linear divergence assumption, $t - f(p) \geq c_1(t-p)$ for $p \in [u, t - \frac{1}{n}]$. It follows that

$$t - f^{(\ell)}(p) \geq c_1^{\ell}(t-p) \geq c_1^{\ell}(1/n).$$

Thus for $\ell = \log_{c_1} n(t-u)$, $f^{(\ell)}(p) \leq u$. Therefore, level $\Omega(\log n)$ items fire with probability at most u. Next we show that given a level in which items fire with probability at most u, the items at $\Omega(\log k)$ levels higher in the iterative tree fire with probability at most 2^{-k}. Let p' be the probability an item fires at the first level for which the probability an item

fires is below u. By the quadratic convergence assumption, $f(p') < c_3(p')^2$ for $p' \in (0, u)$. It follows that for $\ell > \log_2(\log_{c_3}(1/2)) + \log_2(k) - \log_2(1 + \log_{c_3}(t - 1/n)) + 1$, $f^{(\ell)}(p') \leq c^{2^l-1}(p')^{2^l} \leq c^{2^l-1}u^{2^l} < 2^{-k}$. We have shown that in expectation items at $\Omega(\log n + \log k)$ fire with probability less than 2^{-k} when $p \leq t - \frac{1}{n}$.

Next we consider the case when $p \geq t + \frac{1}{n}$. By the linear divergence assumption, $f(p) - t \geq c_2(p - t)$ for $p \in [t + \frac{1}{n}, v]$. It follows that

$$f^{(\ell)}(p) - t \geq c_2^{\ell}(p - t) \geq c_2^{\ell}(1/n).$$

Thus for $\ell = \log_{c_2} n(v - t)$, $f^{(\ell)}(p) \geq v$. Therefore, level $\Omega(\log n)$ items fire with probability at least v. Next we show that given a level in which items fire with probability at least v, the items at $\Omega(\log k)$ levels higher in the iterative tree fire with probability at most 2^{-k} in expectation. Let p' be the probability an item fires at the first level for which the probability an item fires is at least v. By the quadratic convergence assumption, $1 - f(p') < c_3(1 - p')^2$ for $p' \in (v, 1)$. It follows that for $\ell > \log_2(\log_{c_4}(1/2)) + \log_2(k) - \log_2(1 + \log_{c_4}(1 - (t + 1/n))) + 1$, $1 - f^{(\ell)}(p') \leq c^{2^\ell-1}(1 - p')^{2^\ell} \leq c^{2^\ell-1}(1 - v)^{2^\ell} < 2^{-k}$. We have shown that in expectation items at $\Omega(\log n + \log k)$ fire with probability at least $1 - 2^{-k}$ when $p \geq t + \frac{1}{n}$. □

Remark 3.3.3. Let f be a function corresponding to an iterative construction with fixed point t. Then there exists u and v for which the quadratic convergence condition of Lemma 3.3.2 holds if and only if $f'(0) = 0$ and $f'(1) = 0$.

Proof. Quadratic convergence to 0 is observed if and only if there exists some positive constant u sufficiently close to 0 for which all $x < u$, $f(x) = O(x^2)$. Writing $f(x)$ according to its Taylor series expansion about 0 implies that such behavior occurs if and only if $f'(0) = 0$. Similarly, the observing the Taylor series expansion about 1 allows us to conclude that quadratic convergence to 1 is observed if and only if $f'(1) = 0$. □

Next, we prove that the construction given in Theorem 3.1.4A converges quadratically to

t- threshold functions. The proof of Theorem 3.1.4B is provided in Section 3.6.2.

Proof. (of Theorem 3.1.4A)

Since $2-\phi \leq t \leq \phi - 1$, $0 \leq \alpha(t) \leq 1$ and the probability distribution R is well-defined. By construction, $f_{F_1}(p) = 4p^2 - 4p^3 + p^4$ and $f_{F_2}(p) = 2p^2 - p^4$, so

$$f_R(p) = \frac{1+t-3t^2}{t(1-t)}p^2 + \frac{-2+2t+2t^2}{t(1-t)}p^3 + \frac{1-2t}{t(1-t)}p^4.$$

We apply Lemma 3.3.2. First note that $0, t$, and 1 are fixed points. Let p be the fraction of input items firing. It suffices to show convergence to 0 when $p \leq t - \frac{1}{n}$. By Corollary 3.2.4, convergence to 1 for $p \geq t + \frac{1}{n}$ follows from the complementary construction.

First we show linear divergence from t. Let $g(p) = \frac{f(p)-p}{p(1-p)(p-t)} = \frac{1}{t} + \frac{p(2t-1)}{(1-t)t}$. We claim $g(p) \geq 1$ for $0 \leq p \leq 1$. If $t \geq \frac{1}{2}$, then $g(p) \geq \frac{1}{t} > 1$. If $t < \frac{1}{2}$, then $g(p) \geq \frac{1}{t} + \frac{2t-1}{(1-t)t} = \frac{1}{1-t} \geq 1$. Observe that for any constant $0 < u < t$ and $u < p \leq t - 1/n$

$$t - f(p) = t - (p + p(1-p)(p-t)g(p)) \geq (t-p)\left(1 + p(1-p)\right) \geq (t-p)\left(1 + u(1-t)\right).$$

Thus $c_1 = 1 + u(1-t)$ satisfies the first linear divergence condition.

Next, we show quadratic convergence. Let $u = 1/5$. Observe that

$$f(p) \leq 4p^2 - 4p^3 + p^4 < 4p^2.$$

Since $(1/5)4 < 1$, taking $c_3 = 4$ satisfies the first condition of quadratic convergence. Thus, we may apply Lemma 3.3.2 to conclude that items at level $\Omega(\log n + \log k)$ in the limit of the iterative construction compute a t-threshold function with probability at least $1 - 2^{-k}$.

It remains to show that no construction using trees with four leaves will yield quadratic convergence to a t-threshold function for t outside the range $2 - \phi \leq t \leq \phi - 1$. A t-threshold

function with quadratic convergence must satisfy the following five constraints: (i) $f(0) = 0$, (ii) $f(1) = 1$, (iii) $f(t) = t$, (iv) $f'(0) = 0$, (v) $f'(1) = 0$. Solving these equations gives the function

$$f(p) = \frac{1 + t - 3t^2}{t(1-t)}p^2 + \frac{-2 + 2t + 2t^2}{t(1-t)}p^3 + \frac{1 - 2t}{t(1-t)}p^4.$$

Suppose that f can realized by a convex combination of degree four polynomials. Then the leading coefficient of f must be between -1 and 1 since all achievable polynomials have leading coefficient -1 or 1. Thus, $0 \leq \frac{1-2t}{t(1-t)} \leq 1$, which implies that $2 - \phi \leq t \leq \phi - 1$. $\square$

Using a similar technique as in the proof above, it is possible to show that the analogous constructions on six and seven leaves yield iterative constructions that converge quadratically to threshold functions for thresholds in the ranges $0.15 \lesssim t \lesssim 0.85$ and $0.11 \lesssim t \lesssim 0.89$ respectively. However, it is not possible to generalize such a construction beyond this point. Instead, we observe the emergence of probabilistic thresholds.

3.3.2 The emergence of probabilistic thresholds

For $k \geq 4$, the function $h = (f_{A_k} + f_{B_k})/2$ does not exhibit quadratic convergence to a threshold function. The function h has a fixed point $s \in (0, 1/2)$, a fixed point $t \in (1/2, 1)$, and a fixed point at $1/2$. Figure 3.3 illustrates $h_6(p)$. There is no linear divergence away from $1/2$; instead $1/2$ is an attractive fixed point. Therefore with high probability, high level items of such an iterative construction return 0 or 1 with equal probability for inputs in the interval (s, t), return 0 for inputs in the interval $[0, s)$, and return 1 for inputs in the interval $(t, 1]$.

of Theorem 3.1.6. Let $h = (f_{A_k} + f_{B_k})/2$. It suffices to show that for $k \geq 4$, $1/2$ is an attractive fixed point of h. First, note $1/2$ is a fixed point of h since $h(1/2) = 1/2$. We show that $1/2$ is an attractive fixed point by proving that there is some ε neighborhood of $\frac{1}{2}$ such that $h(p) - p > 0$ for $p > \frac{1}{2} - \varepsilon$ and $h(p) - p < 0$ for $p < \frac{1}{2} + \varepsilon$. Let $d(p) = h(p) - p$. By

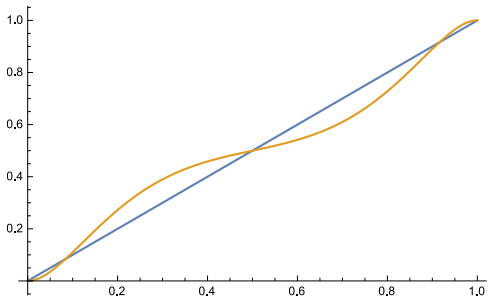

Figure 3.3: For $k \geq 4$, the function $(f_{A_k} + f_{B_k})/2$ has five fixed points on the interval $[0, 1]$ and $1/2$ is an attractive fixed point.

definition

$$d(p) = \frac{2p^k - p^{2k} + 1 - 2(1-p)^k + (1-p)^{2k}}{2} - p.$$

We compute the derivatives,

$$\begin{aligned}
d'(p) &= k\left(p^{k-1} + (1-p)^{k-1}\right) - k\left(p^{2k-1} + (1-p)^{2k-2}\right) - 1 \\
d''(p) &= k(k-1)\left(p^{k-2} - (1-p)^{k-2}\right) - k(2k-1)\left(p^{2k-1} - (1-p)^{2k-2}\right) \\
d'''(p) &= k(k-1)(k-2)\left(p^{k-3} + (1-p)^{k-3}\right) - k(2k-1)(2k-2)\left(p^{2k-3} + (1-p)^{2k-3}\right).
\end{aligned}$$

Evaluating at $p = 1/2$ we obtain

$$\begin{aligned}
d''(1/2) &= 0 \\
d'''(1/2) &= \frac{k}{2^{k-2}}\left((k-1)(k-2) - \frac{(2k-1)(2k-2)}{2^k}\right).
\end{aligned}$$

Note that $1/2$ is both a zero and an inflection point of d. For $k \geq 4$, $d'''(1/2) > 0$ meaning $d(p)$ changes from concave down to concave up at $1/2$. It follows that for some ε, $d(p) > 0$ for $1/2 - \varepsilon < p < 1/2$ and $d(p) < 0$ for $1/2 < p < 1/2 + \varepsilon$. □

Since iterative constructions achieved by averaging A_k and B_k do not yield threshold functions for $k \geq 4$, we must employ a new strategy to achieve threshold functions near 0 or

1.

3.3.3 Quadratic convergence for arbitrary thresholds.

In this section we show that as t approaches 0 or 1, increasingly large building blocks trees are needed to construct an iterative tree that converges quadratically to a t- threshold function. Further, we give a construction that exhibits quadratic convergence for arbitrary thresholds near 0 and 1. We begin by proving Theorem 3.1.5, which can also be restated as follows: Let f be an achievable polynomial with fixed points 0, t, and 1 that exhibits quadratic convergence to a t-threshold function. Then, f has degree at least $\frac{1}{\sqrt{2s}}$ where $s = \min\{t, 1-t\}$.

Proof. (of Thm. 3.1.5.) Let f be an achievable polynomial with fixed points 0, t, and 1 that exhibits quadratic convergence. Then for ε sufficiently small, $f(\varepsilon) = O(\varepsilon^2)$, which implies $a_1 = 0$. For $x < \frac{1}{2d}$, we have

$$f(x) = a_2x^2 + a_3x^2 + \cdots + a_dx^d \leq d^2x^2 + d^3x^3 + \ldots d^dx^d < d^2x^2\left(\frac{1}{1-dx}\right) < 2d^2x^2.$$

Since t is a fixed point of f, $f(t) = t$. Thus, $t < 2d^2t^2$. It follows that $d > \frac{1}{\sqrt{2t}}$. By Lemma 3.2.3, if there exists an achievable polynomial with fixed point t, then there also exists a complementary achievable polynomial with fixed point $1-t$. Thus, $d > \frac{1}{\sqrt{2(1-t)}}$. □

We now prove that a nearly matching iterative construction exists. To achieve quadratic convergence to thresholds near 0 or 1, we average trees of the form A_k and A_{k+1} or B_k and B_{k+1} respectively.

Proof. (of Thm. 3.1.7.) By Corollary 3.2.4, it suffices to prove the theorem for $1-\phi \leq t < 1$. The complement of a construction that achieves quadratic convergence to a t-threshold function

yields quadratic convergence for to a $(1-t)$-threshold function. By Lemma 3.2.8, there exists k and α such that $f = \alpha f_{B_k} + (1-\alpha) f_{B_{k+1}}$has fixed point t. Moreover, $\frac{t-f(p)}{t-p} \geq \left(1 + \frac{p(1-p)}{t}\right)$.

We apply Lemma 3.3.2 to prove that f converges to a t-threshold function. Let p be the probability an input item is on. First suppose that $p \leq t - \frac{1}{n}$. We show linear divergence away from t. For any constant $0 < u < t$, and $u \leq p \leq t - \frac{1}{n}$ by Lemma 3.2.8 we have

$$t - f(p) \geq (t-p)\left(1 + \frac{p(1-p)}{t}\right) \geq (t-p)\left(1 + \frac{u(1-t)}{t}\right).$$

Thus, $c_1 = 1 + \frac{u(1-t)}{t}$ is a valid choice for c_1 in Lemma 3.3.2.

Next, we claim that $u = 1 - \frac{1}{k-1}$ is a valid starting point for quadratic convergence towards 0. We write $f(p) = p^2(\alpha d_k(p) + (1-\alpha) d_{k+1}(p))$ where $d_k(p) = 2p^{k-2} - p^{2k-2}$. Let $d(p) = \alpha d_k(p) + (1-\alpha) d_{k+1}(p)$. Note that $d(p)$ is increasing on the interval $(0, u)$ since each d_k increases on this interval. For $p < u$,

$$\frac{2k-4}{2k-2} = u > u^k > p^k.$$

It follows that $d_k'(p) = p^{k-3}((2k-4) - (2k-2)p^k) > 0$. Thus, d_k is increasing on the interval $(0, u)$. Thus, $c_3 = d(u)$ is a valid choice for c_3 in Lemma 3.3.2.

It remains to show that for $p \geq t + \frac{1}{n}$ we observe linear divergence from t then quadratic convergence to 1. We show linear divergence away from t. For any constant $t < v < 1$, and $t + \frac{1}{n} \leq p \leq 1$ by Lemma 3.2.8 we have

$$f(p) - t \geq (p-t)\left(1 + \frac{p(1-p)}{t}\right) \geq (p-t)\left(1 + \frac{t(1-v)}{t}\right).$$

Thus, $c_2 = 1 + \frac{t(1-v)}{t}$ is a valid choice for c_2 in Lemma 3.3.2.

We claim that $v > 1 - \frac{1}{8(k+1)^2}$ is a valid starting point for quadratic convergence to 1. By

Corollary 3.2.4, $f_{A_k}(1-p) = 1 - f_{B_k}(p)$. It follows

$$1 - f(p) = \alpha - \alpha f_{B_k}(p) + (1-\alpha) - (1-\alpha) f_{B_{k+1}}(p) = \alpha f_{A_k}(1-p) + (1-\alpha) f_{A_{k+1}}(1-p).$$

Recall from the proof of Theorem 3.1.5, $f(x) < 2dx^2$ where d is the degree of x. Therefore,

$$f_{A_k}(1-p) < 8k^2(1-p)^2 < 8(k+1)^2(1-p)^2 \text{ and } f_{A_{k+1}}(1-p) < 8(k+1)^2(1-p)^2.$$

Since $(1-v)8(k+1)^2 < 1$, $c_4 = 8(k+1)^2$ is a valid choice for c_4 in Lemma 3.3.2. □

3.4 Finite iterative constructions of threshold trees

In the above section, we analyzed the behavior of iterative trees in the limit with respect to level width. We assumed that for any input the number of items turned on at level l of the tree is equal to its expectation, $mf^{(l)}(p)$ where m is the width of level l and p is the fraction of the inputs turned on. In a "bottom up" construction in which the items of one level are fixed before the next level is built, we note that the chance that the number of items that fire at a given level deviates from expectation is non-trivial. In this section, we give a bound on the width of the levels required to achieve a desired degree of accuracy for a finite realization of iterative constructions.

Remark 3.4.1. We can use a transition matrix to directly compute the probability that a high level item of an iterative construction fires given the width of the levels. Let f be the function corresponding to the construction, p be the fraction of input items firing, and m the width of the levels. Define $s \in \mathbb{R}^{1\times(m+1)}$, $A \in \mathbb{R}^{(m+1)\times(m+1)}$, and $t \in \mathbb{R}^{1\times(m+1)}$

$$s_i = \binom{m}{i} f(p)^i \left(1 - f(p)\right)^{m-i}, \quad A_{i,j} = \binom{m}{j} f\left(\frac{i}{m}\right)^j \left(1 - f\left(\frac{i}{m}\right)\right)^{m-j}, \quad t_i = i$$

for $i, j = 0, 1, \ldots m$. Then the probability that an item at level L fires is $sA^{L-1}t^T$.

We will use the following concentration inequalities.

Lemma 3.4.2. *(Chernoff) Let $Y_1, Y_2, \ldots Y_m$ be independent with $0 \leq Y_i \leq 1$ and $Y = \sum_{i=1}^{n} Y_i$. Then, for any $\delta > 0$,*

$$\mathbb{P}(Y - E(Y) \geq \delta E(Y)) \leq \exp\left(\frac{-\delta^2 m E(Y)}{2+\delta}\right).$$

Lemma 3.4.3. *Let X be a sum of n binomial random variables with mean μ. Then, for $k \geq n\mu$,*

$$\mathbb{P}(X \geq k) = \sum_{i=k}^{n} \binom{n}{i} \mu^i (1-\mu)^{n-i} < \exp\left(-nH(\mu, k/n)\right)$$

where $H(p,q) = q\log(q/p) + (1-q)log((1-q)/(1-p))$.

The following lemma describes linear divergence for finite width constructions.

Lemma 3.4.4. *Consider the construction of a t-threshold function in which each level ℓ has m_ℓ items and the fraction of input items firing is at least ε below the threshold t. Let d be the minimum value of $\frac{f(p)-p}{p(1-p)(p-t)}$ on the interval $[0,1]$. Then, with probability at least $1-\gamma$, the fraction of inputs firing at level $\Omega(\frac{1}{\varepsilon})$ will be less than any fixed constant u when*

$$m_\ell = \frac{8\ln(\frac{1}{u(1-t)\gamma})}{d^2 u(1-t)^2\left(1+\frac{c_1}{2}\right)^{\ell-1}\varepsilon^2}$$

where c_1 is the linear divergence constant.

Proof. Let X_i be the fraction of items firing at level X_i. Then $\mathbb{E}(X_i) = f(X_{i-1})$. In expectation, the sequence $X_1, X_2, X_3, \ldots$ convergences to 0. We will show that with

probability at least $1-\gamma$, the sequence obeys the *half-progress* relation $X_{i+1} \leq \frac{X_i+f(X_i)}{2}$ and therefore $X_L < u$ for $L = \Omega(\frac{1}{\varepsilon})$.

Write $f(p) - p = p(1-p)(p-t)g(p)$ where g is a polynomial in p. Let d be the minimum value obtained by g on the interval $[0,1]$. First we compute probability that $X_{i+1} > \frac{X_i+f(X_i)}{2}$ by applying Lemma 3.4.2. Observe

$$\begin{aligned}
\mathbb{P}\left(X_{i+1} > \frac{X_i + f(X_i)}{2}\right) &= \mathbb{P}\left(X_{i+1} - \mathbb{E}(X_{i+1}) > \frac{X_i - f(X_i)}{2}\right) \\
&\leq \mathsf{exp}\left(\frac{-\left(\frac{X_i - f(X_i)}{2f(X_i)}\right)^2 m f(X_i)}{2 + \frac{X_i - f(X_i)}{2f(X_i)}}\right) \\
&= \mathsf{exp}\left(-\frac{(X_i(1-X_i)(X_i-t)g(X_i))^2 m}{2(X_i + 3(X_i + X_i(1-X_i)(X_i-t)g(X_i)))}\right) \\
&\leq \mathsf{exp}\left(-\frac{X_i(1-X_i)^2(t-X_i)^2 d^2 m}{8}\right)
\end{aligned}$$

Let $\varepsilon_i = t - X_i$ and $\alpha = \frac{u(1-t)^2 d^2}{8}$. Then for $u \leq X_i \leq t - \varepsilon$,

$$\mathbb{P}\left(X_{i+1} > \frac{X_i + f(X_i)}{2}\right) < \mathsf{exp}\left(-\alpha m \varepsilon_i^2\right).$$

Next we compute the probability that i is the first value for which the half-progress relation is not satisfied given $X_i > u$. If the half-progress relation is satisfied meaning $X_{i+1} > \frac{X_i+f(X_i)}{2}$, then $\varepsilon_{i+1} \geq \varepsilon_i \beta$ where $\beta = 1 + \frac{u}{2}(1-t)$. It follows that if the half-progress relation is satisfied for all $j < i$, then $\varepsilon_{i+1} \geq \varepsilon \beta^i$. Thus,

$$\mathbb{P}\left(i \text{ is the first value for which } X_{i+1} > \frac{X_i + f(X_i)}{2}\right) \leq \mathsf{exp}\left(-\alpha m \varepsilon^2 \beta^{2i}\right).$$

By linear divergence, there exists $L = \Omega(\log(\frac{1}{\varepsilon}))$ such that if the sequence satisfies the half-progress relation for all $i < L$, then $X_L < u$. We bound the probability that this does

not happen. Let $m_\ell = \frac{8\ln(\frac{1}{u(1-t)\gamma})}{d^2u(1-t)^2\beta^i\varepsilon^2}$. For ease of notation, let $c = \ln\frac{1}{u(1-t)\gamma} < 1$. Observe

$$\begin{aligned}
\mathbb{P}(X_L > u) &\leq \sum_{i=0}^{L} \exp\left(-\alpha m_\ell \varepsilon^2 \beta^{2i}\right) \\
&= \sum_{i=0}^{L} \exp\left(-c\beta^i\right) \\
&\leq \sum_{i=0}^{L} \exp\left(-c(1+iu(1-t))\right) \\
&< \exp\left(-c\right) \sum_{i=0}^{L} e^{-iu(1-t))} \\
&< \frac{\exp\left(-c\right)}{1-\exp\left(-u(1-t)\right)} \\
&< \frac{\exp\left(-c\right)}{u(1-t)} = \gamma.
\end{aligned}$$

□

Theorem 3.4.5. *Consider the construction of a t-threshold function with linear convergence given in Theorem 3.1.2 in which each level ℓ has m_ℓ items and the fraction of input items firing is at least ε from the threshold t. Then, with probability at least $1-\gamma$, items at level $\Omega(\log\frac{1}{\gamma} + \log\frac{1}{\varepsilon}))$ will accurately compute the threshold function for $m = \Omega\left(\ln(\frac{1}{\gamma})(\frac{1}{\gamma} + \frac{1}{\varepsilon^2})\right)$.*

Proof. Let X_i be the fraction of items firing at level X_i. Then $E(X_i) = f(X_{i-1})$. By Corollory 3.2.4, it suffices to consider the case when the fraction of inputs firing is less that $t-\varepsilon$. As proved in Theorem 3.1.2, in expectation, the sequence $X_1, X_2, X_3, \ldots$ convergences to 0. We will show that with probability at least $1-\frac{\gamma}{2}$, the sequence drops below $\frac{\gamma}{2}$. First we apply Lemma 3.4.4. Recall that the polynomial corresponding to this construction is $f(p) = p + p(1-p)(p-t)$ and therefore d in the statement of Lemma 3.4.4 is 1. Let u be a constant $0 < u < t$, $m \geq \frac{8\ln(\frac{4}{u(1-t)\gamma})}{u(1-t)^2\varepsilon^2}$ and $L = \Omega(\frac{1}{\varepsilon})$. Thus, $X_L < u$ with probability at least $1-\frac{\gamma}{4}$.

Next we show that given $X_L < u$ the probability that the sequence continues to obey the

half-progress relation (as defined in Lemma 3.4.4) and drops below $\frac{\gamma}{2}$ is at least $1 - \frac{\gamma}{4}$. Let $\alpha = \frac{(1-u)^2(t-u)^2}{8}$. For $X_i < u$,

$$\mathbb{P}\left(X_{i+1} > \frac{X_i + f(X_i)}{2}\right) < \exp\left(-\alpha m X_i^2\right).$$

We compute the probability that $N + i$ is the first value for which the half-progress relation is not satisfied given $X_L < u$. If $X_i < u$ and the half-progress relation is satisfied at i then $X_{i+1} \leq X_i(1 - \beta)$ where $\beta = \frac{1}{2}(1 - u)(t - u)$. It follows that if the half-progress relation is satisfied for all $j < i$, then $X_{N+i} \leq (1 - \beta)^i u$. Let $L' = \frac{4}{(1-u)(t-u)} \log_2\left(\frac{2u}{\gamma}\right)$. If for all $L \leq i \leq L + L'$, the half-progress relation is satisfied then $X_{L+L'} < u(1-\beta)^{L'} < \frac{\gamma}{2}$. We bound the probability that this does not happen. Let $m \geq \frac{16 \ln\left(\frac{16}{(1-u)(t-u)\gamma}\right)}{(1-u)^2(t-u)^2\gamma}$. For ease of notation, let $c = \ln\left(\frac{8}{\beta\gamma}\right)$. Observe

$$\begin{aligned}
\mathbb{P}\left(X_{L+L'} > \frac{\gamma}{2}\right) &\leq \sum_{i=0}^{L'} \exp\left(-m X_i \alpha\right) \\
&= \sum_{i=1}^{L'} \exp\left(-\frac{2cX_i}{\gamma}\right) \\
&\leq \sum_{i=0}^{L'} \exp\left(-c(1-\beta)^{-(L'-i)}\right) \\
&= \sum_{i=0}^{L'} \exp\left(-c(1-\beta)^i\right) \\
&\leq \sum_{i=0}^{\beta L'} \frac{1}{\beta} \exp\left(-ce^i\right) \\
&\leq \frac{2\exp\left(-c\right)}{\beta} \\
&= \frac{\gamma}{4}.
\end{aligned}$$

Therefore, with probability at least $1 - \frac{\gamma}{2}$, items at level $\Omega(\log \frac{1}{\gamma} + \log \frac{1}{\varepsilon}))$ of an iterative

construction with width m fire with probability at most $\frac{\gamma}{2}$ for $m = \Omega\left(\ln(\frac{1}{\gamma})(\frac{1}{\gamma} + \frac{1}{\varepsilon^2})\right)$. Thus, the iterative construction accurately computes the threshold function with probability at least $(1 - \frac{\gamma}{2})^2 > 1 - \gamma$. □

We give a tighter bound for the accuracy of the finite width construction for functions with quadratic convergence. We now prove Theorem 3.1.8, which can be restated as follows: in order to accurately compute, with probability at least $1 - \gamma$, a t-threshold function for inputs in which the fraction of inputs firing is within ε of t , the width of the levels must be $\Omega\left(\frac{\ln(1/\gamma)}{\varepsilon^2}\right)$.

of Theorem 3.1.8. Let X_i be the fraction of items firing at level X_i. Then $E(X_i) = f(X_{i-1})$. By Corollory 3.2.4, it suffices to consider the case when the fraction of inputs firing is less that $t - \varepsilon$. As proved in Lemma 3.3.2, in expectation, the sequence $X_1, X_2, X_3, \ldots$ convergences to 0. We will show that with probability at least $1 - \gamma$, the sequence reaches 0.

First we apply Lemma 3.4.4. Recall from the proof of Theorem 3.1.4, that the minimum value of $g(p) = \frac{f(p)-p}{p(1-p)(p-t)}$ is 1 on the interval $[0, 1]$. Therefore for such constructions d in the statement of Lemma 3.4.4 is 1. For constructions described in Theorem 3.1.7, the minimum value of $g(p) = \frac{f(p)-p}{p(1-p)(p-t)}$ is $\frac{1}{t}$ on the interval $[0, 1]$, as proved in Lemma 3.2.8. Therefore for such constructions d in the statement of Lemma 3.4.4 is $\frac{1}{t}$. Let u be the constant $0 < u < t$ in quadratic convergence as in Lemma 3.3.2, $m_\ell \geq \frac{8\ln(\frac{4}{u(1-t)\gamma})}{d^2u(1-t)^2(1+(c_1/2)^{\ell-1}\varepsilon^2}$ and $L = \Omega(\frac{1}{\varepsilon})$. Thus, $X_L < u$ with probability at least $1 - \frac{\gamma}{2}$.

Next, we bound the probability given $X_L < u$, that $X_{L+r} = 0$. We say that X_{k+1} *regresses* if $X_{k+1} \geq u$. For $m \geq \frac{l}{u}$ and c_3 as in Lemma 3.3.2, we apply Lemma 3.4.3 and obtain

$$\begin{aligned}\mathbb{P}(X_{k+1} \geq X_k) &= \sum_{i=\lceil \frac{u}{e}n \rceil}^{m} \binom{m}{i} f(p)^i (1 - f(p))^{m-i} \\ &\leq \exp\left(-m\left(p\log\left(\frac{p}{f(p)}\right) + (1-p)\log\left(\frac{1-p}{1-f(p)}\right)\right)\right)\end{aligned}$$

$$
\begin{aligned}
&\leq \exp\left(-m\left(p\log\left(\frac{1}{c_3 p}\right)+(1-p)\log\left(1-\frac{p-f(p)}{1-f(p)}\right)\right)\right)\\
&\leq \exp\left(-m\left(p\log\left(\frac{1}{c_3 p}\right)+(1-p)\left(-\frac{p-f(p)}{1-f(p)}\right)\right)\right)\\
&\leq \exp\left(-m\left(p\log\left(\frac{1}{c_3 p}\right)-p(1-p)\right)\right)\\
&\leq \exp\left(-mp\log\left(\frac{1}{c_3 pe}\right)\right)\\
&\leq (c_3 ue)^l
\end{aligned}
$$

It follows that

$$\mathbb{P}(X_L, X_{L+1}, \ldots X_{L+r} \text{ do not regress}) \geq 1 - r(c_3 ue)^l.$$

Next we bound the probability that given $X_k < u$, $X_{k+1} = 0$.

$$\mathbb{P}(X_{k+1} = 0 | X_k \leq u) = (1 - f(X_k))^m \geq (1-u^2)^m \geq 1 - mu^2 = 1 - lu.$$

Therefore

$$\mathbb{P}(X_{L+r} = 0 | X_L, X_{L+1}, \ldots X_{L+r} \text{ do not regress}) \geq 1 - (lu)^r.$$

Let $l = r = \max\left\{c_3 e, \min\left\{\frac{1}{2u}, \log_2\left(\frac{4}{\gamma}\right)\right\}\right\}$. It follows $\log_2\left(\frac{4}{\gamma}\right) < l\log_2\left(\frac{1}{lu}\right)$ and therefore $\frac{4}{\gamma} < \left(\frac{1}{lu}\right)^l$. We now compute

$$\mathbb{P}(X_{L+r} = 0) \geq 1 - r(c_3 ue)^l - (lu)^r \geq 1 - 2(lu)^l \geq 1 - \frac{\gamma}{2}.$$

We have shown that given $X_L < u$, $\mathbb{P}(X_{L+r} > 0) \leq \frac{\gamma}{2}$. Therefore, with probability at least $1-\gamma$, items at level $\Omega(\log\frac{1}{\gamma} + \log\frac{1}{\varepsilon}))$ do not fire for $m = \Omega\left(\frac{\ln(\frac{1}{\gamma})}{\varepsilon^2}\right)$. □

3.4.1 Exponential and wild constructions

In this section, we analyze exponential constructions, where items are chosen with probabilities proportional to their weights, and the latter decay exponentially with time.

We begin with a precise analysis of the wild construction corresponding to no decay of weights. Each new item is the root of a tree chosen according to the probability distribution with leaves chosen uniformly at random from all existing items and inputs. We now prove Theorem 3.1.9, which can be restated as follows: After $O\left(\frac{n^2}{\varepsilon}\right)$ items are created in a wild iterative construction on n inputs, the next item will accurately compute the threshold function probability less than ε.

of Theorem 3.1.9. Let x_0 be the fraction of inputs that fire and let x_k be the probability an item chosen at random from the n inputs and k constructed items fires. We have

$$x_{k+1} = \frac{(n+k)x_k + f(x_k)}{n+k+1}.$$

Equivalently,

$$x_k - x_{k+1} = \frac{x_k - f(x_k)}{n+k+1}.$$

For $\frac{x_k+f(x_k)}{2} \leq x \leq x_k$,

$$x - f(x) \geq \frac{x_k + f(x_k)}{2} - f(x_k) = \frac{x_k - f(x_k)}{2}.$$

By Corollary 3.2.4, it suffices to consider the case when $x_0 < t$. We will show that x_k converges to zero by analyzing the progress towards zero in phases. Items $k+1, k+2, \ldots k+g$ form a phase if $f(x_{k+g}) \leq \frac{3x_k+f(x_k)}{4}$. We show that the number of items in each phase is at most one more than the number of existing items at the start of the phase (which includes

both the inputs and the items constructed in previous phases). For $g = n + k + 1$ observe

$$x_k - x_{k+g} = \sum_{i=1}^{g} \frac{x_{k+i} - f(x_{k+i})}{j+k+i+1} \geq \frac{g(x_k - f(x_k))}{2(n+k+g+1)} = \frac{x_k - f(x_k)}{4}$$

Therefore $f(x_{k+g}) \leq \frac{3x_k + f(x_k)}{4}$.

Let $\overline{f}(x) = \frac{3x_k+f(x_k)}{4}$ and L be the smallest integer such that $\overline{f}^{(L+1)}(x) \leq \varepsilon$ for all $x \leq t - \frac{1}{n}$ and $\overline{f}^{(L+1)}(x) \geq 1 - \varepsilon$ for all $x \geq t + \frac{1}{n}$. Since $\overline{f}$ exhibits linear divergence, $L = \Omega\left(\log(n) + \log(\frac{1}{\varepsilon})\right)$. Then after L phases, the expected value of an item chosen at random from all items is less than $\overline{f}^{(L)}(x_0)$. The next item created will fire with probability at most $\overline{f}^{(L+1)}(x_0) \leq \varepsilon$. Let N_i be the maximum number of total items after phase i and $N_0 = n$. Using the recurrence relation $N_{i+1} = 2N_i + 1$, we conclude that total number of items in L phases is bounded above by $2^L(n+1)$. Therefore after $\Omega(2^L n) = \Omega(\frac{n^2}{\varepsilon})$ items are created, the next item will fire probability less than ε, as desired. □

Now we discuss exponential constructions with $\alpha > 0$. Suppose we start with n items of which a $y_0 = p$ fraction are set to 1. Let y_j be the probability that the j^{th} item is a 1. At time k, when k items have been created, the weight of the j^{th} item is $e^{-\alpha(k-j)}$ since at step the weight decreases by a multiplicative $e^{-\alpha}$. Since the $(k+1)^{th}$ item is picked according to these weights and the function applied,

$$y_{k+1} = f\left(\frac{ne^{-\alpha k}y_0 + \sum_{j=1}^{k} e^{-\alpha(k-j)}y_j}{ne^{-\alpha k} + \sum_{j=1}^{k} e^{-\alpha(k-j)}}\right).$$

To solve this, as before, let x_k be the probability that a random item chosen from the above distribution is 1, i.e.,

$$x_k = \frac{ne^{-\alpha k}y_0 + \sum_{j=1}^{k} e^{-\alpha(k-j)}y_j}{ne^{-\alpha k} + \sum_{j=1}^{k} e^{-\alpha(k-j)}} \text{ and } y_{k+1} = f(x_k).$$

Therefore,

$$x_{k+1} = \frac{e^{-\alpha}(ne^{-\alpha k} + \sum_{j=1}^{k} e^{-\alpha(k-j)})x_k + f(x_k)}{ne^{-\alpha(k+1)} + \sum_{j=1}^{k+1} e^{-\alpha(k+1-j)}}.$$

To see this, just note that the probability of picking one of the first k items is proportional to the multiplier in the numerator above, while the probability of picking the new item is proportional to 1 (its current weight). The denominator is the sum of all weights. This gives

$$x_k - x_{k+1} = \frac{(x_k - f(x_k))}{ne^{-\alpha(k+1)} + \sum_{j=1}^{k+1} e^{-\alpha(k+1-j)}} \geq (1 - e^{-\alpha})(x_k - f(x_k)).$$

For these x_k's to represent the true probabilities, and for this analysis to be valid, the x_k's should be close to their expectation. If we start at distance ε from the threshold, then the deviation from expectation in the beginning should be smaller than ε. For this, it will suffice to set $\alpha = O(1/\varepsilon^2)$ so that the weights decay only by a constant factor after $O(1/\varepsilon^2)$ steps.

3.5 Learning

So far we have studied the realizability of thresholds via neurally plausible simple iterative constructions. These constructions were based on prior knowledge of the target threshold. Here we study the learnability of thresholds from examples. It is important that the learning algorithm should be neurally plausible and not overly specialized to the learning task. We believe the simple results presented here are suggestive of considerably richer possibilities.

We begin with a one-shot learning algorithm. We show that given a single example of a string $X \in \{0,1\}^n$ with $\|X\|_1 = tn$, we can build an iterative tree that computes a t-threshold function with high probability. Let T_1 and T_2 be the building block trees in the construction given in Theorem 3.1.2.

This simple algorithm has the guarantee stated in Theorem 3.1.10, which follows from Theorem 3.1.2.

Now we consider a more complex scenario, where multiple labeled examples are presented.

LearnThreshold(L, m, X)**:**
Input: Levels parameter L, a string $X \in \{0,1\}^n$ such that $\|X\|_1 = tn$, width parameter m.
Output: A finite realization of iterative tree with width m.

For each level j from 1 to L, apply the following iteration m times:
(level 0 consists of the input items X)

1. Pick a random input item i.
2. If $X_i = 1$ then let $T = T_1$, else let $T = T_2$.
3. Pick 3 items from the previous level.
4. Build T with these items as leaves.

We assume that the labels $+, -$, correspond to a uniform threshold function, but are noisy close to the threshold, i.e., the probability of the label being positive goes from 0 to 1 monotonically near the threshold. This is a natural model of noise in the literature. In this setting, we propose to modify the learning algorithm as follows. On the first example it learns an iterative threshold as above. The prediction of the learned structure is given by uniformly sampling a random top-level item. A future example is first evaluated using the existing structure. If the label produced is correct, then no change is made to the structure. Otherwise, if the true label is positive, a new iterative tree is built (and the prediction of the structure would sample from all existing top-level items). If the true label is negative, then top-level items that predicted positive are each destroyed independently with a fixed probability (say 1/2). This effectively builds a structure with a monotone probability of predicting that the label is positive.

This construction highlights two aspects of interest to cognition: (1) the first example presented is crucial and (2) updates to the existing structure become less and less frequent (they are made only on errors).

3.6 Proofs

3.6.1 Properties of achievable polynomials

Proof. (of Lemma 3.2.1.) We proceed by induction on the degree of $A(x)$. For $d = 1$, $A(x) = x$ is the only polynomial in $\mathcal{A}$ and all the above properties hold. Next assume all the properties hold for polynomials of degree less than d. Let $A(x)$ be an achievable polynomial of degree d. Then the root of the tree for A, which we call T_A, is either an AND or an OR operation. In the former case, $A = B \cdot C$ and in the latter case $A = B + C - B \cdot C$ where $B, C \in \mathcal{A}$ and B has degree k and C has degree $d - k$ for $0 < k < d$. In either case, the first three properties follow trivially from the inductive hypothesis. For item (5), let T_B and T_C be trees that correspond to B and C respectively. Then T_B and T_C have k and $d - k$ leaves respectively. Since T_A is T_B adjoined with T_C with an AND or OR operation, T_A has d leaves. □

Proof. (of Lemma 3.2.2.) Proceed by induction. The only achievable polynomial of degree 1 is $f(x) = x$, so the statement clearly holds. Next, assume $|a_{l'}| \leq d^{l'}$ holds for all $l' < l$. Let f be a degree d achievable polynomial. We may assume $f = g + h - gh$ or $f = gh$ where g and h are achievable polynomials with degree k and $d - k$ respectively where $k \leq \frac{l}{2}$. First consider the case when $f = g + h - gh$, meaning the root of the tree corresponding to f is an OR operation. Observe

$$\begin{aligned}
|a_\ell(f)| &= \left|a_\ell(g) + a_\ell(h) - \sum_{i=1}^{l-1} a_i(g)a_{l-i}(h)\right| \\
&\leq k^l + (d-k)^l + \sum_{i=1}^{l-1} k^i(d-k)^{l-i} \\
&\leq k^l + (d-k)^l + (l-1)\max_i\{k^i(d-k)^{l-i}\} \\
&\leq k^l + (d-k)^l + l(k)(d-k)^{l-i}
\end{aligned}$$

$$\leq ((d-k)+k)^l$$
$$= d^l.$$

Next consider the case when $f = gh$, meaning the root of the tree corresponding to f is an AND operation. Observe that

$$|a_\ell(f)| = \Big|\sum_{i=1}^{l-1} a_i(g)a_{l-i}(h)\Big| \leq \sum_{i=1}^{l-1} k^i(d-k)^{l-i} < d^l.$$

□

of Lemma 3.2.3. Proceed by induction on the number of leaves of the tree. For a tree on one leaf, the statement holds trivially. Without loss of generality, assume that the root of tree A is an AND operation. Then $f_A(x) = a_1(x)a_2(x)$ and $f_B(x) = b_1(x) + b_2(x) - b_1(x)b_2(x)$ where the trees corresponding to a_1 and b_1 are complements and the trees corresponding to a_2 and b_2 are also complements. By the inductive hypothesis, $a_1(p) = 1 - b_1(1-p)$ and $a_2(p) = 1 - b_2(1-p)$. Observe

$$\begin{aligned} 1 - f_A(p) &= 1 - a_1(p)a_2(p) \\ &= 1 - (1 - b_1(1-p))(1 - b_2(1-p)) \\ &= b_1(1-p) + b_2(1-p) - b_1(1-p)b_2(1-p) \\ &= f_B(1-p). \end{aligned}$$

□

The proof of Lemma 3.2.6 will use the following elementary inequality.

Lemma 3.6.1. *For $x \in (0,1)$, and any integer $k \geq 0$, $1 - kx < (1-x)^k < 1 - kx + \binom{k}{2}x^2$.*

of Lemma 3.2.6. It suffices to show that $g(x) = f_{A_k}(x) - x$ has a zero on the interval

$\left(\frac{1}{k^2}, \frac{1}{k(k-1)}\right)$. We will show that $g(1/k^2) < 0$ and $g(1/(k(k-1))) > 0$ and apply the intermediate value theorem. Using Lemma 3.6.1, for $x = 1/k^2$,

$$(1-(1-x)^k)^2 - x < (1-(1-kx))^2 - x = k^2x^2 - x = 0.$$

Similarly, for $x = 1/(k(k-1))$,

$$\begin{aligned}(1-(1-x)^k)^2 - x &> \left(1 - \left(1 - kx + \frac{k(k-1)x^2}{2}\right)\right)^2 - x \\ &= \left(\frac{1}{k-1} - \frac{1}{2k(k-1)}\right)^2 - \frac{1}{k(k-1)} \\ &= \frac{1}{(k-1)^2}\left(1 - \frac{1}{2k}\right)^2 - \frac{1}{k(k-1)} \\ &= \frac{1}{(k-1)^2}\left(1 - \frac{1}{k} + \frac{1}{4k^2} - (1 - \frac{1}{k})\right) = \frac{1}{4k^2(k-1)^2} > 0.\end{aligned}$$

It follows from Corollary 3.2.4 that f_{B_k} has a fixed point in the interval $\left(1 - \frac{1}{k(k-1)}, 1 - \frac{1}{k^2}\right)$. Uniqueness follows from Lemma 3.2.7. □

Proof. (of Lemma 3.2.7.) By definition

$$g(p) = \frac{f(p) - p}{p(1-p)(p-t)} = \frac{\alpha(1 - 2p^{k-1} + p^{2k-1}) + (1-\alpha)(1 - 2p^k + p^{2k+1})}{(1-p)(t-p)}.$$

Since 1 and t are fixed points of $f(p)$, $(1-p)$ and $(t-p)$ divide $f(p) - p$. Therefore, we may write $g = a_0 + a_1 p \cdots + a_{2k-1}p^{2k-1}$ polynomial. We claim that all coefficients of g are positive. Note that

$$(t-p)\sum_{i=0}^{2k-1} a_i p^i =$$

$$\alpha(1 + p \cdots + p^{k-2} - p^{k-1} - p^k \cdots - p^{2k-2}) + (1-\alpha)(1 + p \cdots + p^{k-1} - p^k - p^{k+1} \cdots - p^{2k}).$$

Observe

$$\begin{aligned}
a_i &= \frac{1}{t} && \text{for } i = 0 \\
a_i &= \frac{a_{i-1} + 1}{t} && \text{for } 1 \leq i \leq k-2 \\
a_i &= \frac{a_{i-1} + 1 - 2\alpha}{t} && \text{for } i = k-1 \\
a_i &= \frac{a_{i-1} - 1}{t} && \text{for } k \leq i \leq 2k-2 \\
a_i &= \frac{a_{i-1} - (1-\alpha)}{t} && \text{for } i = 2k-1.
\end{aligned}$$

Note that $\frac{1}{t} = a_0 < a_1 < \cdots < a_{k-2}$, so $a_i > 0$ for all $0 \leq i \leq k-2$. Next observe that $a_{2k-1} = 1 - \alpha$ since comparing the coefficients of the p^{2k} terms on both sides gives $-a_{2k-1} = -(1-\alpha)$. It follows that $a_{2k-2} = t(1-\alpha) + 1 - \alpha$. For all $k \leq i \leq 2k-2$, $a_{i-1} = ta_i + 1$. Therefore $a_i > 0$ for all $k-1 \leq i \leq 2k-3$.

Since all coefficients of g are positive, all derivatives are increasing. In particular the first derivative of $g(p)$ is increasing on the interval $(0,1)$. Therefore $g(p) \geq g(0) = \frac{1}{t}$ for all $p \in [0,1]$. □

Proof. (of Lemma 3.2.8.) By Corollary 3.2.4, it suffices to prove the theorem for $\phi - 1 \leq t < 1$. By Lemma 3.2.6, f_{B_k} has a single fixed point in the range $\left(1 - \frac{1}{k(k-1)}, 1 - \frac{1}{k^2}\right)$. Let b_k be the fixed point f_{B_k}. Note that for $0 < p < 1$,

$$f_{B_{k+1}}(p) < f_{B_k}(p).$$

It follows that $b_k < b_{k+1}$. We obtain an increasing sequence $b_2, b_3, b_4 \ldots$ that converges to 1. Let k be the value for which $b_k \leq t < b_{k+1}$. Let $f = \alpha f_{B_k} + (1-\alpha) f_{B_{k+1}}$ where α is chosen so that f has fixed point t.

Next we show that $t - f(p) \geq (t-p)\left(1 + \frac{p(1-p)}{t}\right)$. By Lemma 3.2.7, $g(p) = \frac{f(p)-p}{p(1-p)(p-t)} \geq \frac{1}{t}$ for $0 \leq p \leq 1$. Therefore

$$t - f(p) = (t-p)(1 + p(1-p)g(p)) \geq (t-p)\left(1 + \frac{p(1-p)}{t}\right).$$

□

3.6.2 Quadratic convergence for trees with five leaves

Proof. (of Theorem 3.1.4B.) By construction $f_{V_1}(p) = p^2 + p^3 - p^5$ and $f_{V_2}(p) = 6p^2 - 9p^3 + 5p^4 - p^5$, so

$$f_R(p) = \frac{1 + t - 2t^2 - t^3}{t(1-t)}p^2 + \frac{-2 + t + t^2 + 2t^3}{t(1-t)}p^3 + \frac{1 - t^2 - t^3}{t(1-t)}p^4 - p^5.$$

We apply Lemma 3.3.2. First note that $0, t$, and 1 are fixed points. Let p be the fraction of input items firing. It show convergence to 0 when $p \leq t - \frac{1}{n}$. By Corollary 3.2.4, convergence to 1 for $p \geq t + \frac{1}{n}$ follows from the complementary construction.

First we show linear divergence from t. Let $g(p) = \frac{f(p)-p}{p(1-p)(p-t)} = \frac{1}{t} - \frac{t^2+t-1}{t(t-1)}p + p^2 = \frac{1}{t} + p(-\frac{t^2+t-1}{t(t-1)} + p)$. We claim that $g(p) \geq 1$ for $0 \leq p \leq 1$. If $-\frac{t^2+t-1}{t(t-1)} + p \geq 0$, then $g(p) \geq \frac{1}{t} \geq 1$. If $-\frac{t^2+t-1}{t(t-1)} + p < 0$, then $g(p) \geq \frac{1}{t} - \frac{t^2+t-1}{t(t-1)} + 1 = \frac{1}{1-t} \geq 1$. Thus as in the proof of part A, $c_1 = 1 + u(1-t)$ satisfies the first linear divergence condition. Next we show quadratic convergence. Let $u = 1/7$. Note that

$$f(p) \leq p^2(6 - 9p + 5p^2 - p^3) < 6p^2.$$

Since $6(1/7) < 1$, taking $c_3 = 6$ satisfies the first condition of quadratic convergence. Thus, we may apply Lemma 3.3.2 to conclude that in expectation items at level $\Omega(\log n + \log k)$ of the iterative construction compute a t-threshold function with probability at least $1 - 2^{-k}$.

It remains to show that no construction using trees with five leaves will yield quadratic convergence to a t-threshold function for t outside the range $0.26 \lesssim t \lesssim 0.74$. A t-threshold function with quadratic convergence must satisfy the following five constraints: (i) $f(0) = 0$, (ii) $f(1) = 1$, (iii) $f(t) = t$, (iv) $f'(0) = 0$, (v) $f'(1) = 0$. Such a function will have the form:

$$z_{d,t}(p) = \frac{1 + t - (3 + d)t^2 + dt^3}{(1 - t)t} p^2 + \frac{-2 + (2 + d)t + (2 + d)t^2 - 2dt^3}{(1 - t)t} p^3 + \frac{1 - (2 + 2d)t + dt^2 + dt^3}{(1 - t)t} p^4 + dp^5.$$

Since each achievable polynomial has leading coefficient -1 or 1, if $z_{d,t}(p)$ can written as a convex combination of achievable polynomials of degree five, then

$$z_{d,t}(p) = \beta z_{-1,t}(p) + (d + \beta) z_{1,t}(p),$$

where $z_{-1,t}(p)$ and $z_{1,t}(p)$ are convex combinations of achievable polynomials of degree five with leading coefficient -1 and 1 respectively and $0 \leq \beta \leq 1$. Thus, it suffices to determine the values of t for which $z_{-1,t}(p)$ is achievable through convex combinations and the values of t for which $z_{1,t}(p)$ is achievable through convex combinations.

Claim: Let $\alpha(t) = \frac{-1+5t-4t^2+t^3}{5t(t-1)}$. If the function $z_{-1,t}(p)$ is achievable through convex combinations then $0 \leq \alpha(t) \leq 1$, meaning $0.26 \lesssim t \lesssim 0.74$.

Notice that achievable polynomials of degree five with leading coefficient -1 have coefficient $a_3 \geq -9$ (see Table 3.1). It follows that

$$\frac{-2 + t + t^2 + 2t^3}{t(1 - t)} \geq -9 \quad \text{and} \quad 2(-1 + 5t - 4t^2 + t^3) \geq 0,$$

so $\alpha(t) \geq 0$. Next, note that the coefficient a_4 of $z_{-1,t}(p)$ must be non-negative (see Table

3.1). It follows that

$$\frac{1-t^2-t^3}{t(1-t)} \geq 0 \quad \text{and} \quad -1+5t-4t^2+t^3 \leq 5t-5t^2,$$

so $\alpha(t) \leq 1$.

Claim: Let $\gamma(t) = 1-2t-t^2+t^3$ and $\beta(t) = 1-3t^2+t^3$. If the function $z_{1,t}(p)$ is achievable through convex combinations, then $\gamma(t) \leq 0$ and $\beta(t) \geq 0$, meaning $.445 \lesssim t \lesssim .653$

Assume $z_{-1,t}(p)$ is achievable through convex combinations. Notice that for degree five achievable polynomials with $a_1 = 0$ and $a_5 = 1$, $-4 \leq a_4 \leq -2$. It follows that

$$-4 \leq \frac{1-4t+t^2+t^3}{(1-t)t} \leq -2,$$

so $\gamma(t) \leq 0$ and $\beta(t) \geq 0$.

Now consider $t \lesssim 0.26$ or $t \gtrsim 0.74$. By the above claims, $z_{-1,t}(p)$ and $z_{1,t}(p)$ are not achievable through convex combinations. It follows that $z_{d,t}(p)$ is not achievable through convex combinations, meaning no construction on trees with five leaves that converges quadratically to a t-threshold function for $t \lesssim 0.26$ or $t \gtrsim 0.74$. □

3.7 Discussion

We have seen that very simple, distributed algorithms requiring minimal global coordination and control can lead to stable and efficient constructions of important classes of functions. Our work raises several interesting questions.

1. What are the ways in which threshold functions are applied in cognition? Object recognition is one application of threshold functions in cognition. For instance, suppose we have items representing features such as "trunk," "grey," "wrinkled skin," and "big ears," and an item representing our concept of an "elephant." If a certain threshold of

items representing the features we associate with an elephant fire, then the "elephant" item will fire. This structure lends itself to a hierarchical organization of concepts that is consistent with the fact that as we learn, we build on our existing set of knowledge. For example, when a toddler learns to identify an elephant, he does not need to re-learn how to identify an ear. The item representing "ear" already exists and will fire as a result of some threshold function created when the toddler learned to identify ears. Now the item representing "ear" may be used as an input as the toddler learns to identify elephants and other animals.

2. What is an interesting model and neurally plausible algorithm for learning threshold functions of k relevant input items? In this scenario, the input is a set of sparse binary strings of length n representing examples in which at least tk of k relevant items are firing. The output is an iterative tree that computes a t-threshold function on the k relevant items. We can formulate the previously described example of learning to identify an elephant as an instance of this problem. Each time the toddler sees an example of an elephant, many features associated with elephant will fire in addition to some features that are not associated with elephants. There may also be features associated with an elephant that are not present in this example and therefore not firing. A learning algorithm must rely on information about the items that are currently firing to learn both the set of relevant items and a threshold function on this set of items. It might also be beneficial to utilize prediction, as e.g., done by [PV15b].

3. To what extent can general linear threshold functions with general weights be constructed/learned by cortical algorithms?

4. A concrete question is whether the construction of Theorem 3.1.7 is optimal, similar to the optimality of the constructions in Theorem 3.1.4.

5. Our construction of a probabilistic threshold function raises the question of what

monotone functions $g : \{0,1\}^n \to [0,1]$ can be realized by iterative constructions.

6. A simple way to include non monotone Boolean functions with the same constructions as we study here, would be to have input items together with their negations (as in e.g., [Sav90]). What functions can be realized this way, using a distribution on a small set of fixed-size trees?

CHAPTER 4
APPROXIMATING SPARSE GRAPHS: THE RANDOM OVERLAPPING COMMUNITIES MODEL

This Section describes joint work with Santosh Vempala and appears as [PV18].

4.1 Introduction and overview of results

The goal of this direction of research is to approximate sparse graphs and sequences of sparse graphs (with average degree unbounded and $o(n)$). What is a good summary of a very large graph? Besides simple statistics like its size and edge density, one would like to know the chance of finding small subgraphs (e.g., triangles), to estimate global properties (e.g., the size of the minimum or maximum cut), and to be able to produce a smaller graph of desired size with similar properties as the original. One approach is to define a random graph model, a simple description of a probability distribution over all graphs, such that a graph drawn from the model will likely have similar properties as the graph of interest.

A more powerful approach to graph approximation is via Szemerédi's regularity lemma [Sze76], as discussed in Sections 1.1.2 and 2.1. The regularity lemma's consequences are striking— one can approximate the homomorphism density of any fixed size graph (from the left or right), the size of any cut to within additive error; the partition itself can be constructed algorithmically and is easy to sample.

Such a theory is missing for sparse graphs (with $o(n^2)$ edges). All the approximations for the dense case produce the trivial approximation of the empty graph. While there is an intricately developed theory for bounded-degree graphs that allows one to describe the limits of sequences of such graphs[1], it is not algorithmically tractable, and it does not extend

[1]One can approximate a bounded-degree graph as a distribution over local neighborhood structures, i.e.,

to graph sequences when the degree can grow with the size of the graph. Moreover, as we will presently see, the known objects for approximating dense graphs (block models, regularity decompositions, graphons) are inherently unable to approximate sparse graphs. The main motivation for this paper is to understand what properties of sparse graphs (resp. graph sequences) can be succinctly approximated and to provide a model (limit object) for them. Existing theories are limited in what they can achieve for families of graphs which are neither dense nor bounded-degree. In particular, they seem unable to answer the following representative question [LV14]:

What is the limit of the sequence of hypercube graphs?

We focus on approximating the simple cycle and closed walk counts of sparse graphs and graph sequences appropriately normalized. These counts encode information about the local structure of the graph and are related to its spectral properties; the number of closed k-walks in a graph is equal to the k^{th} moment of the graph's eigenspectrum. For dense graphs, stochastic block models and graphons approximate both homomorphism densities and cut norm. However the standard cut norm is not useful for sparse graphs as the norm tends to zero. Moreover, natural normalizations do not seem to work either, i.e., they either go to zero or distinguish hypercubes of different sizes.

Another reason we focus on cycle and walk counts rather than cuts is that approximating local structure is of particular interest in practice. A widely-used technique for inferring the structure and function of a real-world graph is to observe overrepresented motifs, i.e., small subgraphs that appear frequently. Recent work describes the overrepresented motifs of a variety of graphs including transcription regulation graphs, protein-protein interaction

the probability that the r-neighborhood of a vertex is a particular graph. For any r, this is a finite description and it captures homomorphism densities and appears as a limit of a bounded-degree graph sequence.

graphs, the rat visual cortex, ecological food webs, and the internet (WWW), [YL+04, Alo07, Son+05, Mil+02]. The type of overrepresented motifs has been shown to be correlated with the graph's function [Mil+02]. A model that produces graphs with high motif counts is necessary for approximating graphs whose function depends on the abundance of a particular motif.

Limitations of previous approaches for capturing the cycle and walk counts. Previous approaches do not provide a meaningful way to approximate the small cycle counts of sparse graphs. A regularity style partition or stochastic block model inherently cannot approximate the number of triangles unless the rank of the block model grows nearly linearly with the size of the graph as shown in the following simple observation.

Proposition 4.1.1. *Let M be a symmetric matrix with entries in $[0, 1]$ such that each row sum is at most d. Then the expected number of simple k-cycles in a graph obtained by sampling M is at most $d^k rank(M)$.*

The proposition follows by observing that the expected number of simple k-cycles is at most the trace of M^k. In particular, any rank r approximation of the d-dimensional hypercube where each vertex has degree $O(d)$ has fewer than $O(rd^4)$ simple four-cycles, whereas the hypercube has $2^d d^2$ of them.

The local neighborhood distribution approach is hopeless for this setting since the degree is not bounded and therefore there are infinitely many r-neighborhoods [BS01]. Other methods designed for the sparse but not bounded degree setting do not produce a satisfactory limit object for the sequence of hypercubes. The theory of L^p graphons generalizes the graphon to a range of sparse settings [Bor+14]. While the L^p graphon gives approximations for a generalized notion of cut metric for sparse graphs, graphs sampled from the L^p graphon limit of the sequence may have very different normalized subgraph counts than the sequence (i.e. no "counting lemma" is possible). Frenkel redefined homomorphism density with a different

normalization based on the size of the subgraph, but this notion does not help distinguish the limiting number of non-tree structures for sequences of graphs with degree tending to infinity [Fre16]. The recently developed notion of *graphex* [VR16b, Bor+17b] is the limit object for sequences of sampling convergent graphs. However, any sequence of nearly d-regular graphs with $d = o(n)$ is sampling convergent since the sampled object according to this notion is a set of isolated edges with high probability. Therefore the graphex cannot distinguish between different graph sequences that are nearly regular.

Another natural approach to constructing a graph with high simple cycle density is to repeatedly add simple cycles on a randomly chosen subset of vertices. However, this process yields low cycle to edge ratios for sparse graphs. For example, a graph on n vertices with average degree less than $\sqrt{n}$ built by randomly adding triangles will have a triangle-to-edge-ratio at most 2/3. (See Theorem 4.7.1.) In [New09] Newman considers a similar approach which produces graphs with varied degree sequences and triangle-to-edge ratio strictly less than 1/3. However, it is not hard to construct graphs with arbitrarily high triangle ratio (growing with the size of the graph).

Normalizing closed walk and cycle counts. In order to meaningfully compare the closed walk counts and cycle counts between graphs of different sizes, it is necessary to normalize the counts. For dense graphs, homomorphism density of a subgraph H in a graph on n vertices is the number of copies of a subgraph H divided by $n^{|v(H)|}$. This normalization is natural because it gives the probability H is present on a random subset of vertices. For the sparse case, this normalization causes the homomorphism density of all subgraphs tend to zero, and so we must define a different normalization.

When considering a sequence of graphs, we can find a proper normalization of the closed walk counts by looking at the rate of growth of the counts. A graph that locally looks like a d-ary tree has approximately $d^{k/2}$ closed k-walks at each vertex for k even. Therefore

the appropriate normalization of the closed k-walk counts for a sequence of such graphs is $nd^{k/2}$. We will see in Section 4.3.1 that this normalization is also natural for the sequence of hypercubes. A sequence of sparse graphs in which each vertex's local neighborhood is dense (e.g., a collection of d cliques of size d), the appropriate normalization for the walk counts is nd^{k-1}. We define the *sparsity exponent* of a sequence to measure the rate of growth of the number of closed k-walks in the sequence. Let $W_k(G)$ be the number of simple cycles of length k in a graph G. [2]

Definition 4.1.2. For $0 < \alpha \leq 1$, we define the α normalized closed k-walk count as

$$W_k(G, \alpha) = \frac{W_k(G)}{nd^{1+\alpha(k-2)}}$$

where $n = |V(G)|$ and d is the average degree of G.

Definition 4.1.3 (sparsity exponents)**.** Let (G_i) be a sequence of graphs. Let

$$\alpha = \inf_{b \in [1/2,1]} \left\{ b \mid \lim_{i \to \infty} W_j(G_i, b) \text{ exists for all } j \right\}$$

be the sparsity exponent of the sequence. For $k \geq 3$, let

$$\alpha_k = \inf_{b \in [1/2,1]} \left\{ b \mid \lim_{i \to \infty} W_j(G_i, b) \text{ exists for all } j \leq k \right\}$$

be the k-sparsity exponent of the sequence.

We define the minimum of the sparsity exponent to be $1/2$ because all d-regular graphs have at least $Cat_{k/2} nd^{k/2}$ closed k-walks obtained from tracing trees. For two sequences of graphs with matching degrees, a higher sparsity exponent indicates denser local neighborhoods and therefore more closed walks.

[2] Denoting the number of closed k-walks by $W_k(G)$, and the eigenvalues of the adjacency matrix of G as $\lambda_1(G) \geq \lambda_2(G), \cdots \geq \lambda_n(G)$, we have $W_k(G) = \sum_{i=1}^{n} \lambda_i^k$.

While the sparsity exponent gives a natural way to normalize the closed walk counts for a sequence of sparse graphs, it does not help determine the appropriate normalization factor for approximating an individual graph (not contextualized in a sequence). When approximating an individual graph, we instead choose to focus on the number of simple k-cycles denoted $C_k(G)$ and normalize by the number of edges in the graph. Throughout this paper, for convenience we refer to a simple k-cycle as a k-cycle. For example, under this convention each triangle is counted 6 times because there are 6 closed walks that traverse a triangle.

As described in the previous section, approximating $W_k(G)$ and $C_k(G)$ for sparse graphs is already out-of-reach for known methods that work well in the dense and bounded-degree settings. The main contribution of this paper is the following model which can approximate the normalized closed walk and cycle counts for a large class of graphs.

The Random Overlapping Communities Model. We introduce a simple generalization of the Erdős-Rényi model that can approximate the normalized cycle and walk counts of a wide range of sparse graphs. The *Random Overlapping Communities (ROC)* model generates graphs that are the union of many relatively dense random communities. A *community* is an instance of an Erdős-Rényi graph $G_{s,q}$ (or a bipartite Erdős-Rényi graph $G_{s/2,s/2,q}$) on a set of s randomly chosen vertices. A ROC graph is the union of many randomly selected communities that overlap, so every vertex is a member of multiple communities. The number, size and density of communities are drawn from a distribution. Figure 4.1 illustrates this construction.

An instantiation of the ROC model is given by a distribution $\mathcal{D}$ on triples (s, q, b) where s is an integer, $0 \leq q \leq 1$ and $b < s$ is an integer indicating bipartiteness. A graph of a desired size n and expected degree d is generated by repeatedly selecting a triple (s, q, b) from the distribution $\mathcal{D}$ and picking each vertex with probability s/n and adding a random graph of edge density q in the subgraph. We refer to each such structure as a community. If

$b > 0$, then the subset of s vertices is partitioned into two subsets of expected size $b, s - b$ and edges are added only between these subsets. For this paper we will set $b = 0$ or $b = s/2$. See Section 4.4.1 for a formal definition of the model. We note that the ROC model can be viewed as a generalization of the Erdős-Rényi model (with $s = n$ and $q = d/n$) and maintains the property that it is easy to sample given its defining parameters.

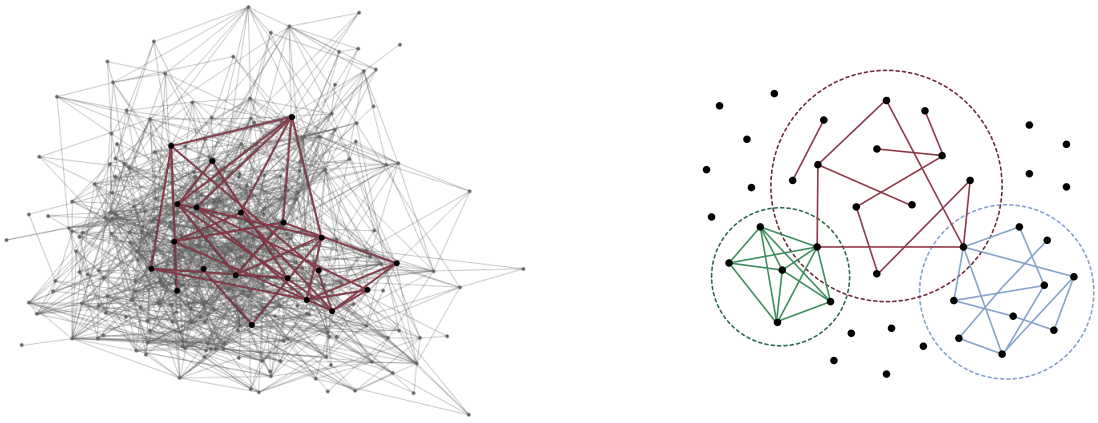

Figure 4.1: Left: in each step of the construction of a ROC(n, d, s, q) graph, an instance of $G_{s,q}$ is added on a set of s randomly selected vertices. Right: three communities of a ROC graph.

Organization. The paper has two main objectives: to show that ROC is an effective approximation for individual sparse graphs (Section 4.2) and to develop a theory of sparse graph limits in which the ROC model is a natural limit object (Sections 4.3 to 4.5). These two parts are self-contained and may be read independently. We end with a discussion of limitations of the model, possible extensions, and open questions (Section 4.6). In the remainder of this section we summarize the results.

4.1.1 ROC for approximating a single graph

In Section 4.2, we show that the ROC model can approximate the triangle-to-edge and four-cycle-to-edge ratios of a graph, and can be tuned to exhibit high clustering coefficient (the probability two randomly selected neighbor of a random vertex are adjacent). Since these properties are of interest in practice, the model may be of use in real-world contexts. In addition, we introduce a variant of the model that produces graphs with varied degree distributions. For a comparison of the ROC model to existing models used in practice see Section 4.2.4.

First, we show that for almost all triangle-to-edge and four-cycle-to-edge ratios arising from some graph, there exists a single community size s density q such that the ROC model produces graphs with these ratios, *simultaneously*. Moreover, the vanishing set of triangle and four-cycle ratio pairs not achievable exactly can be approximated to within a small error.

Theorem 4.1.4.

1. *Let H be a graph and let $c_i = C_k(H)/|E(H)|$ for $i = 3, 4$. Then $c_3(c_3/2 - 1) \leq c_4$.*

2. *For any c_3 and c_4 such that $c_3^2 \leq 2c_4$, and $d = o(n^{1/3})$, the random graph $G \sim ROC\,(n, d, \mathcal{D})$ where $\mathcal{D}$ is the distribution with support one on $s = \frac{2c_4^2}{c_3^3}$ and $q = \frac{c_3^2}{2c_4}$ has*

$$\lim_{n\to\infty} \frac{2\,\mathbb{E}(C_3(G))}{nd} = c_3 \quad \text{and} \quad \lim_{n\to\infty} \frac{2\,\mathbb{E}(C_4(G))}{nd} = c_4.$$

Theorem 4.2.2 gives conditions for determining when it is possible to construct a ROC family that matches a vector of k-cycle-to-edge ratios. These conditions are related to the conditions for determining when the ROC model is the limit object for a sequence of graphs.

Modeling the clustering coefficient of real-world graphs. In Theorem 4.2.3, we prove the average clustering coefficient of a ROC graph (with one community type) is

approximately sq^2/d, meaning that tuning the parameters s and q with d fixed yields wide range of clustering coefficients for a fixed density. Furthermore, Theorem 4.2.4 describes the inverse relationship between degree and clustering coefficient in ROC graphs, a phenomena observed in protein-protein interaction graphs, the internet, and various social networks [Ste+05, Mah+06, Mis+07, Ahn+07].

Diverse degree distributions and the DROC model. We also introduce an extension of our model which produces graphs that match a target degree distribution in expectation. The extension uses the Chung-Lu configuration model: given a degree sequence $d_1, \dots d_n$, an edge is added between each pair of vertices v_i and v_j with probability $\frac{d_i d_j}{\sum_{i=1}^{n} d_i}$, yielding a graph where the expected degree of vertex v_i is d_i [CL02]. In the DROC model, a modified Chung-Lu random graph is placed instead of an E-R random graph in each iteration. Instead of normalizing the probability an edge is selected in a community by the sum of the degrees in the community, the normalization constant is the expected sum of the degrees in the community.

4.1.2 ROC as a limit object for sparse graph sequences

We show that the ROC model is a limit for several interesting sequences of graphs and a give a characterization of sequences which are the limits of ROC. To state our results, we first define the convergence of sparse graph sequences and their limits. We consider convergence first for each k and then for all positive integers k, referring to the latter as full convergence. In Section 4.3, we compute the limits for the hypercube sequence, the rook's graph sequence (a family of strongly regular graphs) as well as for Erdős-Rényi random graphs.

Definition 4.1.5 (k-convergent)**.** Let (G_i) be a sequence of graphs with k-sparsity exponent α_k. The sequence (G_i) is k-convergent if $\lim_{i\to\infty} W_j(G_i, \alpha_k)$ exists for all $j \leq k$. We let $w_j = \lim_{i\to\infty} W_j(G_i, \alpha_k)$ and say $(w_3, w_4, \dots, w_k)$ is the k-limit of the graph sequence (G_i).

Definition 4.1.6 (fully convergent)**.** Let (G_i) be a sequence of graphs with sparsity exponent α. We say the sequence is fully convergent if $\lim_{i\to\infty} W_j(G_i, \alpha)$ exists for all j. We let $w_j = \lim_{i\to\infty} W_j(G_i, \alpha)$ and say $(w_3, w_4, \dots)$ is the limit of the graph sequence (G_i).

Informally, we say that a ROC family (a distribution on triples) achieves the limit of a convergent sequence of graphs (G_i) if the normalized expected number of walks in a graph drawn from the ROC family matches the limit of (G_i). We consider the ROC family that achieves the limit to be a limit object of (G_i). Since a particular limit may be achieved by many ROC families, the limit object for a sequence is not unique.

We now formalize the notion of a ROC family achieving a vector of normalized counts as its limit. We use *achievable* to describe when a ROC family realizes a k-limit, *fully achievable* to describe when a ROC family realizes a limit, and *totally k-achievable* to describe the weaker notion that any subsequence of a limit is achievable by a ROC family.

Definition 4.1.7 (k-achievable, totally k-achievable, fully achievable)**.**

1. The k-limit $(w_3, w_4, \dots w_k)$ of a sequence of graphs with sparsity exponent α is *k-achievable by ROC* if there exists a ROC family $\mathcal{D}$ such that for all $3 \le j \le k$, when $d \to \infty$ and $d = o\left(n^{1/((1-a)k+2a-1)}\right)$

$$\lim_{n\to\infty} \frac{\mathbb{E}(W_j\left(ROC(n, d, \mathcal{D})\right)}{nd^{1+\alpha(j-2)}} = w_j.$$

2. The limit of a sequence of graphs *totally k-achievable by ROC* if every k-limit of the sequence is achievable (possibly with a different choice for each k).

3. The limit $(w_3, w_4, \dots)$ of a sequence of graphs with sparsity exponent α is *fully achievable by ROC* if there exists a ROC family $\mathcal{D}$ such that for all $j \ge 3$, when $d \to \infty$ and $o\left(n_i^{\varepsilon}\right)$

for all ε if $a < 1$ and $d_i = o(n_i)$ for $a = 1$

$$\lim_{n\to\infty} \frac{\mathbb{E}(W_j\left(ROC(n, d, \mathcal{D})\right))}{nd^{1+\alpha(j-2)}} = w_j.$$

Roughly speaking, the degree upper bounds ensure that the overwhelming majority of simple cycles are contained entirely in single communities. In Theorem 4.4.13 we show that the probability that the normalized closed walk counts of $G \sim ROC(n, d, \mathcal{D})$ deviate from the family's limit vanishes as $d \to \infty$. Moreover Corollary 4.4.14 gives conditions on n_i and d_i which guarantee that a sequence (G_i) with $G_i \sim ROC(n_i, d_i, \mathcal{D})$ almost surely converges to limit vector achieved by the family.

Results for convergent sequences. We begin with the limit of the sequence of hypercube graphs, answering the question raised by [LV14].

Theorem 4.1.8. *The limit of the sequence of hypercube graphs is totally k-achievable by ROC.*

This theorem generalizes to sequences of Hamming cubes and Cayley graphs of $(\mathbb{Z} \bmod k\mathbb{Z})^d$ (Corollary 4.5.7). These sequences have the same limit as the hypercube sequence and therefore are achieved by the same ROC family.

The next theorem is about a sequence of strongly regular graphs called rook's graphs (the Cartesian product of two complete graphs, see Lemma 4.3.2).

Theorem 4.1.9. *The limit of the sequence of rook's graphs is fully achievable by ROC.*

We also discuss the convergence of sequences of Erdős-Rényi random graphs (Lemma 4.3.6), demonstrating the limits of some sequences cannot be achieved exactly by ROC familes, but they can be approximated to arbitrarily small error.

Achievability and the Stieltjes condition. A limit vector L is achieved by a ROC family when the normalized expected walk counts of a graph sampled from the ROC family match L up to terms that vanish as the size of the sampled graph grows. The number of closed walks in a ROC graph is related to its simple cycles counts, and the expected simple cycle counts are the moments of a distribution determined by the ROC parameters. Therefore, determining which vectors can be achieved by a ROC family is closely related to the Stieltjes' moment problem: given a sequence whether there exists a discrete distribution with positive support with that moment sequence? This is the classical Stieltjes moment problem, whose solution is characterized by the definition below (see Lemma 4.4.22).

Definition 4.1.10 (Stietljes conditions)**.** The Hankel matrices of a sequence μ are

$$H_{2s}^{(0)} = \begin{pmatrix} \mu_0 & \mu_1 & \dots & \mu_s \\ \mu_1 & & & \\ \vdots & & \ddots & \vdots \\ & & & \\ \mu_s & & \dots & \mu_{2s} \end{pmatrix} \quad \text{and} \quad H_{2s+1}^{(1)} = \begin{pmatrix} \mu_1 & \mu_2 & \dots & \mu_{s+1} \\ \mu_2 & & & \\ \vdots & & \ddots & \vdots \\ & & & \\ \mu_{s+1} & & \dots & \mu_{2s+1} \end{pmatrix}.$$

1. The vector $\mu = (\mu_0, \mu_1, \dots \mu_n)$ satisfies the Stieltjes condition if

$$det\left(H_{2s}^{(0)}\right) \geq 0 \text{ for all } 0 \leq 2s \leq n \quad \text{and} \quad det\left(H_{2s+1}^{(1)}\right) \geq 0 \text{ for all } 1 \leq 2s+1 \leq n,$$

and for k the smallest integer such that $det\left(H_{2k}^{(0)}\right) = 0$ or $det\left(H_{2k+1}^{(1)}\right) = 0$,

$$det\left(H_{2i}^{(1)}\right) = 0 \quad \text{and} \quad det\left(H_{2i+1}^{(1)}\right) = 0 \quad \text{for all} \quad k \leq i \leq n.$$

2. The infinite vector $\mu = (\mu_0, \mu_1, \dots)$ satisfies the full Stieltjes condition if the above statements hold for all n.

In ROC families that produce sequences of graphs with sparsity exponent greater than 1/2, the counts of simple cycles dominate the total closed walk counts. Therefore, a limit is achievable when it is possible to construct a ROC family with normalized simple cycle counts that match the desired normalized closed walk counts. The cycle counts are dominated by the cycles contained entirely in one community. Every community contributes even cycles, but only the non-bipartite communities contribute to the odd cycle counts. In the following theorems, the parameters s_i and t_i count the number of simple i-cycles in non-bipartite and bipartite communities respectively, and the parameter γ indicates the expected fraction of communities that are non-bipartite.

Theorem 4.1.11 (achievability with sparsity exponent $> 1/2$)**.** *A limit vector* $(w_3, w_4, \dots w_k)$ *is achievable by ROC with sparsity exponent greater than* $1/2$ *if and only if there exists* $\gamma \in [0,1]$, $s_0, s_1, \dots s_k, t_0, t_2, \dots t_{2\lfloor \frac{k}{2} \rfloor} \in \mathbb{R}^+$, $s_2, t_2 \leq 1$ *such that* $(s_0, s_1, s_2, \dots s_k)$ *and* $(t_0, t_2, \dots t_{2\lfloor \frac{k}{2} \rfloor})$ *satisfy the Stieltjes condition and for* $3 \leq j \leq k$

$$w_j = \begin{cases} \gamma s_j & j \text{ odd} \\ \gamma s_j + (1-\gamma) t_j & j \text{ even.} \end{cases}$$

Approximating sequences with sparsity exponent 1/2 is more complicated because the number of simple cycles can be of the same order as the number of closed walks that are not simple cycles in ROC families that produce graph sequences with sparsity exponent 1/2. In Theorem 4.4.3 we prove that the polynomial T given in Definition 4.4.1 describes the relationship between simple cycle counts and closed walks counts in ROC graphs. Moreover we show that T describes the relationship between simple cycle counts and closed walk counts in *locally regular* graphs in which each vertex is in the same number of cycles. A limit with sparsity exponent 1/2 is achievable when it is possible to construct a ROC family with normalized simple cycle counts that match the inverse of this polynomial T applied to the

desired normalized closed walk counts.

Theorem 4.1.12 (achievability with sparsity exponent 1/2)**.** *Let* $T((c_3, c_4, \dots c_k)) = (w_3, w_4, \dots w_k)$ *be the transformation of a vector given in Definition 4.4.1. The limit vector* $(w_3, w_4, \dots w_k)$ *is achievable by ROC with sparsity exponent* $1/2$ *if and only if there exists* $\gamma \in [0, 1]$, $s_0, s_1, s_2, \dots s_k, t_0, t_2, \dots t_{2\lfloor \frac{k}{2} \rfloor} \in \mathbb{R}^+$, $s_2, t_2 \leq 1$ *such that* $(s_0, s_1, s_2, \dots s_k)$ *and* $(t_0, t_2, \dots t_{2\lfloor \frac{k}{2} \rfloor})$ *satisfy the Stieltjes condition and for* $3 \leq j \leq k$

$$c_j = \begin{cases} \gamma s_j & j \text{ odd} \\ \gamma s_j + (1-\gamma) t_j & j \text{ even.} \end{cases}$$

The analogous theorems for the full achievability of limits by ROC require the full Stieltjes condition. See Theorems 4.4.19 and 4.4.20 in Section 4.4.3. The Stieltjes condition also determines when a vector of k-cycle-to-edge ratios can be matched by a ROC family, demonstrating the relevance of our method for different normalizations (Theorem 4.2.2). All 4-limits can be achieved by a ROC model; however not all k-limits can be achieved. In Section 4.6.1 we give an example of a graph sequences with a 6-limit that cannot be achieved by a ROC family.

Theorem 4.1.13. *The limit* (w_3, w_4) *of any convergent sequence of graphs with increasing degree is achieved by a ROC family.*

We have already seen that any realizable triangle and four-cycle count normalized by the number of edges can be approximated by a ROC model with one community type.

4.2 Approximating a graph with a ROC

In this section we consider the utility of the ROC model for approximating individual graphs. First we show that almost all pairs of triangle-to-edge and four-cycle-to-edge ratio can be

approximated with a ROC graph with one community type and give conditions for when a vector of k-cycle-to-edge ratios can be achieved by ROC generally. (In Appendix B, we analyze the connectivity of ROC graphs.) Then, we shift our focus to modeling real-world graphs with ROC. In Section 4.2.2 we show that ROC graphs exhibit high clustering coefficient and an inverse relationship between clustering coefficient and degree, a phenomena observed in real world networks. In Section 4.2.3 we introduce an extension of the ROC model that produces graphs with varied degree distributions. Finally we end the section by comparing the ROC model to existing models used in practice (Section 4.2.4).

Often in this section, we focus on the special case of the ROC model when the distribution of communities $\mathcal{D}$ is taken to be a single community s and density q. When this is clear from context, we denote the model $ROC(n, d, s, q)$. In terms of the formal parameterization of the ROC model given in Section 4.4.1, $a = 0$ and μ is the the distribution with support one on $m_i = s$, $q_i = q$ and $\beta_i = 0$.

4.2.1 The k-cycle-to-edge ratios of ROC graphs

In this section we prove Theorem 4.1.4, which states that most triangle-to-edge and four-cycle-to-edge ratios can be approximated simultaneously by the ROC model on one community. Then we prove Theorem 4.2.2, which describes more generally when it is possible to match all j-cycle-to-edge ratios up to some k with the ROC model.

To begin we consider the ROC model with all communities of size s and density q. The following lemma describes the k-cycle-to-edge ratios of ROC graphs in this setting. The lemma is a special case of Corollary 4.4.12.

Lemma 4.2.1. *Let $G \sim ROC(n, d, s, q)$. Then*

$$R_k = \lim_{n\to\infty} \frac{2\,\mathbb{E}(C_k(G))}{nd} = 2s^{k-2}q^{k-1} \text{ for } d = o(n^{1/(k-1)}).$$

By varying s and q, we can construct a ROC graph that achieves any ratio of triangles to edges or any ratio of four-cycles to edges. By setting $s = \sqrt{\log(n)}/4$ and $q = 1$, we obtain a family of graphs with the hypercube four-cycle-to-edge ratio $\log(n)/4$, something not possible with any existing random graph model.

Moreover, it is possible to achieve a given ratio by larger, sparser communities or by smaller, denser communities. For example communities of size 50 with internal density 1 produce the same triangle ratio as communities of size 5000 with internal density 1/10. Figure 4.2 illustrates the range of s and q that achieve various triangle and four-cycle ratios. Note that it is possible to achieve $R_3 = 3$ and $R_4 \in \{100, 50, 25\}$ but not $R_3 = 3$ and $R_4 \in \{3, 10\}$.

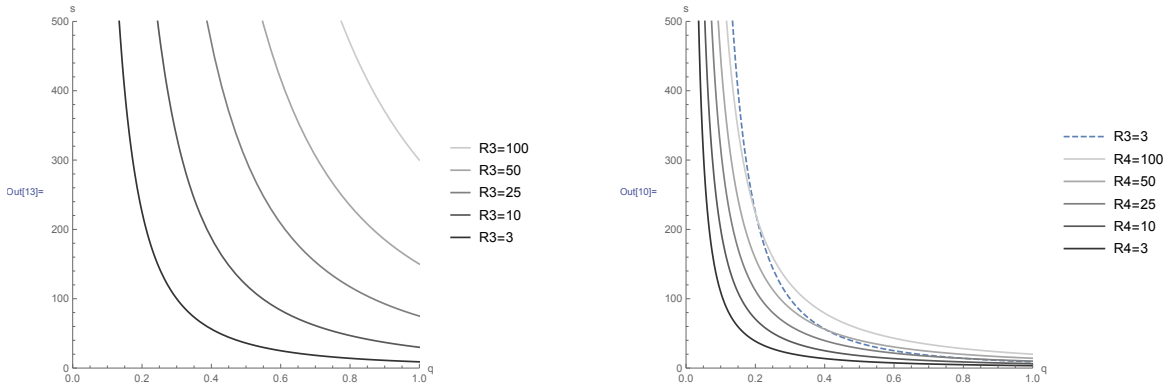

Figure 4.2: Left: A wide range of s and q yield the same R_3 and R_4 ratio (left and right respectively).

We apply Lemma 4.2.1 to prove Theorem 4.1.4, which states that any non-zero triangle and four-cycle ratios satisfying $c_3^2 \leq 2c_4$ can be approximated with the ROC model. For every graph with triangle and four-cycle ratios in the narrow range $c_3(c_3/2 - 1) \leq c_4 \leq c_3^2/2$, there exists a ROC construction that matches c_3 and can approximate c_4 by $c_3^2/4$, i.e., up to an additive error $c_3/8$ (or multiplicative error of at most $1/(c_3/2 - 1)$ which goes to zero as c_3 increases).

Proof. (of Theorem 4.1.4.) (1) For clarity of this proof we refer to the number triangle and four-cycle structures (not counted as walks). Under this convention, the number of triangles

is $T_3 = C_3(H)/6$ and the number of four-cycles is $T_4 = C_4(H)/8$. Note $T_3 = |E(H)|c_3/6$ and $T_4 = c_4|E(H)|/8$. For each edge in H, let t_e be the number of triangles containing e, so $\sum_{e \in E(H)} t_e = 3T_3 = c_3|E(H)|/2$. If triangles abc and abd are present, then so is the four-cycle $acbd$. This four-cycle may also be counted via triangles cad and cdb. Therefore $T_4 \geq \frac{1}{2}\sum_{e \in E(H)} \binom{t_e}{2}$. This expression is minimized when all t_e are equal. We therefore obtain

$$\frac{c_4|E(H)|}{8} = T_4 \geq \frac{|E(H)|}{2}\binom{c_3/2}{2} = \frac{c_3(c_3/2-1)|E(H)|}{8}.$$

It follows that $\frac{c_3(c_3/2-1)}{c_4} \leq 1$.

(2) Since the hypothesis guarantees $q \leq 1$, applying Lemma 4.2.1 to $G \sim ROC\left(n, d, \frac{2c_4^2}{c_3^3}, \frac{c_3^2}{2c_4}\right)$ implies the desired statements. □

By increasing the support size of the distribution over communities, it is possible to achieve a wider range of k-cycle-to-edge-ratios. The following theorem shows that the condition for determining whether a vector of k-cycle-to-edge ratios of a graph can be matched by a ROC family is closely related to determining if a limit of a sequence of graphs is achievable by a ROC family.

Theorem 4.2.2. *There exists a ROC family such that for $G \sim ROC(n, d, \mathcal{D})$ with $d = o\left(n^{1/(k-1)}\right)$*

$$\lim_{n\to\infty} \frac{2\,\mathbb{E}(C_j(G))}{nd} = c_j \quad \text{for } 3 \leq j \leq k$$

if and only if there exists $\gamma \in [0,1]$, $s_0, s_1, s_2, \ldots s_k, t_0, t_2, \ldots t_{2\lfloor \frac{k}{2} \rfloor} \in \mathbb{R}^+$, $s_2, t_2 \leq 1$ such that $(s_0, s_1, s_2, \ldots s_k)$ and $(t_0, t_2, \ldots t_{2\lfloor \frac{k}{2} \rfloor})$ satisfy the Stieltjes condition and for $3 \leq j \leq k$

$$c_j/2 = \begin{cases} \gamma s_j & j \text{ odd} \\ \gamma s_j + (1-\gamma)t_j & j \text{ even.} \end{cases}$$

The proof of the above theorem is a slight modification of the proofs of the main limit

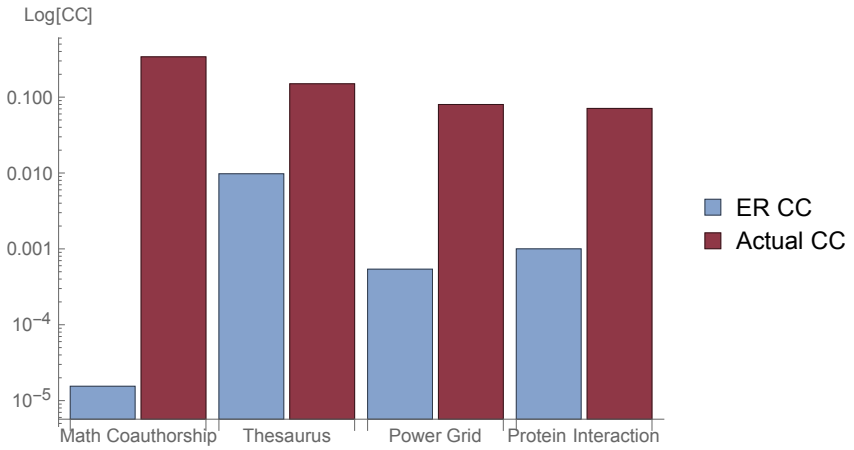

Figure 4.3: The clustering coefficient in real world graphs is much greater than that of an E-R random graph of the same density. Data from Table 3.1 of [New03].

achievability theorems (Theorem 4.1.12, Theorem 4.1.11), and so we give the proof in Section 4.4.3. Later we show that a ROC family that achieves a normalized closed walk count limit is parameterized so that the community sizes grow with d^a for some constant $a \in [1/2, 1]$. In the proof of the above theorem we show that ROC family that approximates a vector of k-cycle-to-edge ratios will have constant community sizes.

4.2.2 Approximating clustering coefficient

Closely related to the density of triangles is the clustering coefficient at a vertex v, the probability two randomly selected neighbors are adjacent:

$$C(v) = \frac{|\{\{a, b\} : a, b \in N(v), a \sim b\}|}{deg(v)(deg(v) - 1)/2}.$$

Equivalently the clustering coefficient is twice the ratio of the number of triangles containing v to the degree of v squared. Figure 4.3 illustrates the markedly high clustering coefficients of real-world graphs as compared with Erdős-Rényi (E-R) graphs of the same density. We show that the ROC model can be tuned to produce graphs with a variety of clustering coefficients at any density. The proofs in this section are quite technical and left to Section 4.2.2.

Theorem 4.2.3 gives an approximation of the expected clustering coefficient when the degree and average number of communities per vertex grow with n. The exact statement is given in Lemma 4.2.7 of Section 4.2.2, and bounds in a more general setting are given by Equation (4.4).

Theorem 4.2.3. *Let $C(v)$ denote the clustering coefficient of a vertex v with degree at least 2 in a graph drawn from $ROC(n, d, s, q)$ with $d = o(\sqrt{n})$, $d < (s-1)qe^{sq}$, $d = \omega(sq \log \frac{nd}{s})$, $s^2 q = \omega(1)$, and $sq = o(d)$. Then*

$$\mathbb{E}(C(v)) = (1 + o(1)) \frac{sq^2}{d}.$$

Unlike in E-R graphs in which local clustering coefficient is independent of degree, higher degree vertices in ROC graphs have lower clustering coefficient. High degree vertices tend to be in more communities, and thus the probability two randomly selected neighbors are in the same community is lower. Figure 4.4 illustrates the relationship between degree and clustering coefficient, the degree distribution, and the clustering coefficient for two ROC graphs with different parameters and the E-R random graph of the same density.

Theorem 4.2.4. *Let $C(v)$ denote the clustering coefficient of a vertex v in a graph drawn from $ROC(n, d, s, q)$ with $d = o(\sqrt{n})$, $s = \omega(1)$ and $deg(v) \geq 2sq$. Then*

$$\mathbb{E}(C(v) \mid deg(v) = r) = \frac{sq^2}{r} (1 + o_r(1))$$

Clustering coefficient proofs

Remark 4.2.5. Theorem 4.2.3 gives bounds on the expected clustering coefficient up to factors of $(1 + o(1))$. The clustering coefficient at a vertex is only well-defined if the vertex has degree at least two. Given the assumption in Theorem 4.2.3 that $d = \omega(sq \log \frac{nd}{s})$,

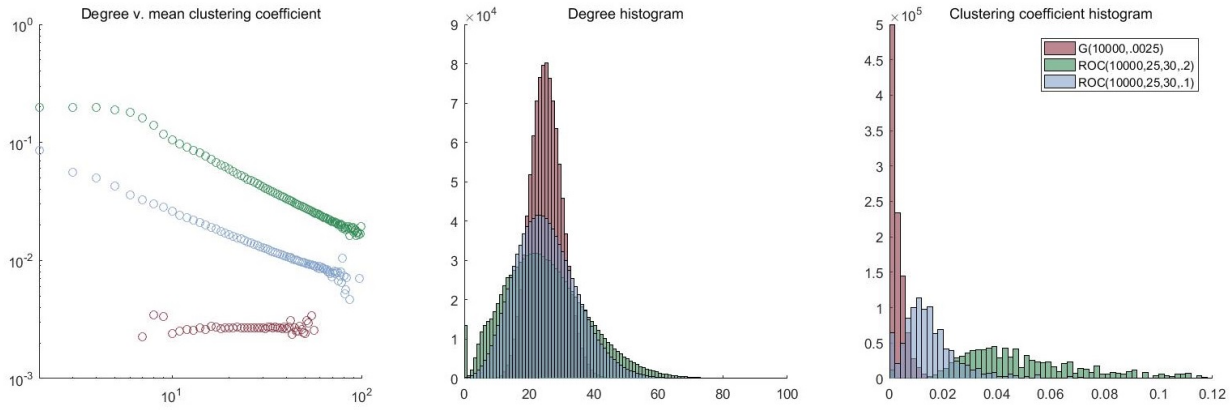

Figure 4.4: A comparison of the degree distributions and clustering coefficients of 100 graphs with average degree 25 drawn from each $G_{10000,0.0025}$, $ROC(10000, 25, 30, 0.2)$, and $ROC(10000, 25, 30, 0.1)$. The mean clustering coefficients are 0.00270, 0.06266, and 0.01595 respectively.

$d < (s-1)qe^{sq}$, and $s = \omega(1)$, Lemma 4.2.6 implies that the fraction of vertices of degree strictly less than two is $o(1)$. Therefore we ignore the contribution of these terms throughout the computations for Theorem 4.2.3 and supporting Lemma 4.2.7. In addition we divide by $deg(v)^2$ rather than by $deg(v)(deg(v)-1)$ in the computation of the clustering coefficient since this modification only affects the computations up to a factor of $(1+o(1))$.

Lemma 4.2.6. *If $d = \omega(sq \log \frac{nd}{s})$, $s = \omega(1)$, $s = o(n)$, and $d < (s-1)qe^{sq}$, then a graph from $ROC(n, d, s, q)$ a.a.s. has no vertices of degree less than 2.*

Proof. Theorem 4.7.2 implies there are no isolated vertices a.a.s. We begin by computing the probability a vertex has degree one.

$$\begin{aligned}
\mathbb{P}(deg(v) = 1) &= \sum_{i=1}^{\frac{nd}{s^2q}} \mathbb{P}(v \text{ is in } i \text{ communities}) q(1-q)^{si-1} \\
&= \sum_{i=1}^{\frac{nd}{s^2q}} \binom{\frac{nd}{s(s-1)q}}{i} \left(\frac{s}{n}\right)^i \left(1 - \frac{s}{n}\right)^{\frac{nd}{s(s-1)q} - i} q(1-q)^{si-1} \\
&\leq (1+o(1)) \sum_{i=1}^{\frac{nd}{s^2q}} \left(\frac{nd}{s(s-1)q}\right)^i \left(\frac{s}{n}\right)^i e^{-\frac{d}{sq} + \frac{si}{n}} qe^{-qsi+q}
\end{aligned}$$

$$= (1 + o(1))\, qe^{-\frac{d}{sq}} \sum_{i=1}^{\frac{nd}{s^2q}} \left(\frac{de^{-sq}}{(s-1)q} \right)^i$$
$$= O\left(\frac{de^{-sq-\frac{d}{sq}}}{s} \right)$$

Let X be a random variable that represents the number of degree one vertices of a graph drawn from $ROC(n, d, s, q)$. When $d = \omega(sq \log \frac{nd}{s})$, we obtain

$$\mathbb{P}(X > 0) \leq \mathbb{E}(X) = O\left(\frac{nde^{-sq-\frac{d}{sq}}}{s} \right) = o(1).$$

□

Lemma 4.2.7. *Let $C(v)$ denote the clustering coefficient of a vertex v of degree at least 2 in a graph drawn from $ROC(n, d, s, q)$ with $d = o(\sqrt{n})$ and $d = \omega(sq \log \frac{nd}{s})$. Then*

$$\mathbb{E}(C(v)) = (1 + o(1)) \left(\sum_{i=1}^{\frac{nd}{s^2q}} \binom{\frac{nd}{s^2q}}{i} \left(\frac{s}{n}\right)^i \left(1 - \frac{s}{n}\right)^{\frac{nd}{s^2q}-i} \frac{s(s-1)q^3k}{(sqk + 2 - 2q)^2} \right).$$

Proof. For ease of notation, we ignore factors of $(1 + o(1))$ throughout as described in Remark 4.2.5. First we compute the expected clustering coefficient of a vertex from an $ROC(n, d, s, q)$ graph given v is contained in precisely k communities. Let $X_1, \ldots X_k$ be random variables representing the degree of v in each of the communities, $X_i \sim Bin(s, q)$. We have

$$\mathbb{E}(C(v)|\ v \text{ in } k \text{ communities }) = \mathbb{E}\left(\frac{\sum_{i=1}^k X_i(X_i - 1)q}{\left(\sum_{i=1}^k X_i\right)^2} \right) \tag{4.1}$$
$$= qk\, \mathbb{E}\left(\frac{X_1(X_1 - 1)}{(sq(k-1) + X_1)^2} \right)$$

$$= qk\,\mathbb{E}\left(\frac{X_1^2}{\left(sq(k-1)+X_1\right)^2}\right) - qk\,\mathbb{E}\left(\frac{X_1}{\left(sq(k-1)+X_1\right)^2}\right).$$

Write $X_1 = \sum_{i=1}^{s} y_i$ where $y_i \sim Bernoulli(q)$. Using linearity of expectation and the independence of the $y_i's$ we have

$$\mathbb{E}\left(\frac{X_1}{\left(sq(k-1)+X_1\right)^2}\right) = s\,\mathbb{E}\left(\frac{y_1}{\left(sq(k-1)+(s-1)q+y_1\right)^2}\right) \;= \frac{sq}{\left(sq(k-1)+(s-1)q+1\right)^2},$$

and

$$\begin{aligned}
\mathbb{E}\left(\frac{X_1^2}{\left(sq(k-1)+X_1\right)^2}\right) &= \mathbb{E}\left(\frac{\left(\sum_{i=1}^{s} y_i\right)^2}{\left(sq(k-1)+\sum_{i=1}^{s} y_i\right)^2}\right) \\
&= s\,\mathbb{E}\left(\frac{y_1^2}{\left(sq(k-1)+q(s-1)+y_1\right)^2}\right) \\
&\quad + s(s-1)\,\mathbb{E}\left(\frac{\left(y_1y_2\right)^2}{\left(sq(k-1)+(s-2)q+y_1+y_2\right)^2}\right) \\
&= \frac{sq}{\left(sq(k-1)+q(s-1)+1\right)^2} + \frac{s(s-1)q^2}{\left(sq(k-1)+(s-2)q+2\right)^2}.
\end{aligned}$$

Substituting in these values into Equation (4.1), we obtain

$$\mathbb{E}(C(v)|v \in k \text{ communities }) = qk\left(\frac{s(s-1)q^2}{\left(sq(k-1)+(s-2)q+2\right)^2}\right) = \frac{s(s-1)q^3k}{\left(sqk+2-2q\right)^2}. \quad (4.2)$$

Let M be the number of communities a vertex is in, so $M \sim Bin\left(\frac{nd}{s^2q}, \frac{s}{n}\right)$. It follows

$$\begin{aligned}
\mathbb{E}(C(v)) &= \sum_{i=1}^{\frac{nd}{s^2q}} \mathbb{P}(\ v \text{ in } k \text{ communities }\,)\,\mathbb{E}(C(v)|\ v \text{ in } k \text{ communities }\,) \\
&= \sum_{i=1}^{\frac{nd}{s^2q}} \binom{\frac{nd}{s^2q}}{i} \left(\frac{s}{n}\right)^i \left(1-\frac{s}{n}\right)^{\frac{nd}{s^2q}-i} \frac{s(s-1)q^3k}{\left(sqk+2-2q\right)^2}.
\end{aligned}$$

□

The proof of Theorem 4.2.3, relies on the follow two lemmas regarding expectation of binomial random variables.

Lemma 4.2.8. *Let $X \sim Bin(n,p)$. Then*

1. $\mathbb{E}\left(\frac{1}{X+1} \mid X \geq 1\right) = \frac{1-(1-p)^{n+1}-(n+1)p(1-p)^n}{p(n+1)}$ *and*
2. $\mathbb{E}\left(\frac{1}{X+1}\right) = \frac{1-(1-p)^{n+1}}{p(n+1)}$.

Proof. Observe

$$\begin{aligned}\mathbb{E}\left(\frac{1}{X+1} \mid X \geq 1\right) &= \sum_{i=1}^{n} \binom{n}{i} \frac{p^i(1-p)^{n-i}}{i+1} \\ &= \frac{1}{p(n+1)} \sum_{i=1}^{n} \binom{n+1}{i+1} p^{i+1}(1-p)^{n-i} \\ &= \frac{1-(1-p)^{n+1}-(n+1)p(1-p)^n}{p(n+1)}.\end{aligned}$$

Similarly

$$\mathbb{E}\left(\frac{1}{X+1}\right) = \sum_{i=0}^{n} \binom{n}{i} \frac{p^i(1-p)^{n-i}}{i+1} = \frac{1}{p(n+1)} \sum_{i=0}^{n} \binom{n+1}{i+1} p^{i+1}(1-p)^{n-i} = \frac{1-(1-p)^{n+1}}{p(n+1)}.$$

□

Lemma 4.2.9. *Let $X \sim Bin(n,p)$. Then*

$$\mathbb{E}\left(\frac{1}{X} \mid X \geq 1\right) \leq \frac{1}{p(n+1)}\left(1 + \frac{3}{p(n+2)}\right).$$

Proof. Note that when $X \geq 1$,

$$\frac{1}{X} \leq \frac{1}{X+1} + \frac{3}{(X+1)(X+2)}.$$

By Lemma 4.2.8,

$$\mathbb{E}\left(\frac{1}{X+1} \mid X \geq 1\right) \leq \frac{1}{p(n+1)}. \tag{4.3}$$

We compute

$$\begin{aligned}
\mathbb{E}\left(\frac{1}{(X+1)(X+2)} \mid X \geq 1\right) &= \sum_{i=1}^{n} \frac{\binom{n}{i} p^i (1-p)^{n-i}}{(i+1)(i+2)} \\
&= \frac{1}{p^2(n+2)(n+1)} \sum_{i=1}^{n} \binom{n+2}{i+2} p^{i+2} (1-p)^{n-i} \\
&\leq \frac{1}{p^2(n+2)(n+1)}.
\end{aligned}$$

Taking expectation of Equation (4.3) gives

$$\mathbb{E}\left(\frac{1}{X} \mid X \geq 1\right) \leq \frac{1}{p(n+1)} \left(1 + \frac{3}{p(n+2)}\right).$$

□

Proof. (of Theorem 4.2.3.) For ease of notation, we ignore factors of $(1 + o(1))$, as described in Remark 4.2.5. It follows from Equation (5.24) in the proof of Lemma 4.2.7 that

$$\frac{q}{k+1} \leq \mathbb{E}(C(v) | v \in k \text{ communities }) \leq \frac{q}{k},$$

where the left inequality holds when $q(s-1) \geq 5$.

We now compute upper and lower bounds on $\mathbb{E}(C(v))$, assuming v is in some community. Let M be the random variable indicating the number of communities containing v, $M \sim$

$Bin\left(\frac{nd}{s(s-1)q}, \frac{s}{n}\right)$. It follows

$$\mathbb{E}(C(v)) = \sum_{k=1}^{\frac{nd}{s^2q}} \mathbb{P}(M=k)\,\mathbb{E}(C(v)|M=k)$$

$$q\,\mathbb{E}\left(\frac{1}{M+1} \mid M \geq 1\right) \leq \mathbb{E}(C(v)) \leq q\,\mathbb{E}\left(\frac{1}{M} \mid M \geq 1\right).$$

Applying Lemmas 4.2.8 and 4.2.9 to the lower and upper bounds respectively, we obtain

$$\frac{q\left(1-\left(1-\frac{s}{n}\right)^{\frac{nd}{s(s-1)q}+1} - \left(\frac{nd}{s(s-1)q}+1\right)\left(1-\frac{s}{n}\right)^{\frac{nd}{s(s-1)q}}\right)}{\frac{d}{(s-1)q}+\frac{s}{n}} \leq \mathbb{E}(C(v)) \leq \frac{q}{\frac{d}{(s-1)q}+\frac{s}{n}}\left(1+\frac{3}{\frac{d}{(s-1)q}+\frac{2s}{n}}\right)$$

which for $s = o(n)$ simplifies to

$$(1+o(1))\,\frac{(s-1)q^2}{d}\left(1-\frac{nd}{s(s-1)q}e^{-d/((s-1)q)}\right) \leq \mathbb{E}(C(v)) \leq \frac{(s-1)q^2}{d}\left(1+\frac{(s-1)q}{d}\right)(1+o(1))\,. \tag{4.4}$$

Under the assumptions $s^2q = \omega(1)$ and $sq = o(d)$, we obtain our desired result

$$\mathbb{E}(C(v)) = (1+o(1))\left(\frac{sq^2}{d}\right).$$

□

The following lemma will be used in the proof of Theorem 4.2.4.

Lemma 4.2.10. *The X be a nonnegative integer drawn from the discrete distribution with density proportional to $f(x) = x^{r-x}e^{-ax}$. Let $z = \operatorname{argmax} f(x)$. Then*

$$\mathbb{P}\big(|x-z| \geq 2t\sqrt{z}\big) \leq e^{-t+1}.$$

Proof. First we observe that f is logconcave:

$$\frac{d^2}{dx^2} \ln f(x) = \frac{d}{dx}(-a + \frac{r}{x} - 1 - \ln x) = -\frac{r}{x^2} - \frac{1}{x}$$

which is nonpositive for all $x \geq 0$. We will next bound the standard deviation of this density, so that we can use an exponential tail bound for logconcave densities. To this end, we estimate $\max f$. Setting its derivative to zero, we see that at the maximum, we have

$$a + 1 = \frac{r}{x} - \ln x. \tag{4.5}$$

The maximizer z is very close to

$$\frac{r}{(a+1) + \ln \frac{r}{(a+1)+\ln(r/(a+1))}}, \tag{4.6}$$

and the maximum value z satisfies $z^{r-z}e^{-az} = z^r e^{-r+z}$. Now we consider the point $z + \delta$ where $f(z + \delta) = f(z)/e$, i.e.,

$$\frac{(z+\delta)^{r-z-\delta}e^{-az-a\delta}}{z^{r-z}e^{-az}} \leq e^{-1}.$$

The LHS is

$$\begin{aligned}\left(1 + \frac{\delta}{z}\right)^{r-z} z^{-\delta} \left(1 + \frac{\delta}{z}\right)^{-\delta} e^{-a\delta} &\leq e^{\delta(\frac{r}{z} - 1 - a - \ln z)} e^{-\frac{\delta^2}{z}} \\ &\leq e^{-\frac{\delta^2}{z}}\end{aligned}$$

where in the second step we used the optimality condition (4.5). Thus for $\delta = (1 + o(1))\sqrt{z}$, $f(x + \delta) \leq f(x)/e$. By logconcavity (which says that for any x, y and any $\lambda \in [0, 1]$, we have

$f(\lambda x + (1-\lambda)y) \geq f(x)^{\lambda} f(y)^{1-\lambda})$ we have

$$f(x+\delta) = f\left(\left(1-\frac{1}{t}\right)x + \frac{1}{t}(x+t\delta)\right) \geq f(x)^{1-1/t} f(x+t\delta)^{1/t}$$

for any $t \geq 1$. It follows

$$f(x+t\delta) \leq f(x)/e^t \tag{4.7}$$

for all t (since we can apply the same argument for $z-\delta$). Taking $x = z$ in Equation (4.7) and using the observation $\sum_{x\in\mathbb{Z}^+} f(x) \geq f(z)$, it follows that

$$\mathbb{P}\big(x = z + t\sqrt{z}\big) \leq e^{-t} \quad \text{and} \quad \mathbb{P}\big(x = z - t\sqrt{z}\big) \leq e^{-t}$$

and so

$$\mathbb{P}\big(|x-z| \geq t\sqrt{z}\big) \leq 2e^{-t} \leq e^{-t+1}.$$

□

Proof. (of Theorem 4.2.4). Let M denote the number of communities a vertex v is selected to participate in. We can write

$$\begin{aligned}\mathbb{E}(C(v)|deg(v)=r) &= \sum_{k=\frac{r}{s}}^{r} \mathbb{E}(C(v)|deg(v)=r, M=k)\,\mathbb{P}(M=k|deg(v=r) \\ &= \sum_{k=\frac{r}{s}}^{r} \mathbb{E}(C(v)|deg(v)=r, M=k)\,\mathbb{P}(deg(v)=r|M=k)\frac{\mathbb{P}(M=k)}{\mathbb{P}(deg(v)=r)}.\end{aligned}$$

First we compute the expected clustering coefficient of a degree r vertex given that it is k communities:

$$\mathbb{E}(C(v)|deg(v)=r \text{ and } M=k) = \frac{\sum_{i\neq j, i,j\in N(v)} q\left(\mathbb{P}(i,j \text{ part of same community})\right)}{deg(v)\left(deg(v)-1\right)} = \frac{q}{k}.$$

Next we note that M is a drawn from a binomial distribution, and the degree of v is drawn from a sum of k binomials, each being $Bin(s,q)$. Therefore,

$$\mathbb{P}(M=k)\,\mathbb{P}(deg(v)=r|M=k) = \binom{\frac{nd}{s(s-1)q}}{k}\left(\frac{s}{n}\right)^k\left(1-\frac{s}{n}\right)^{\frac{nd}{s(s-1)q}-k}\binom{sk}{r}q^r(1-q)^{sk-r}.$$

Using this we obtain

$$\begin{aligned}
\mathbb{E}(C(v)|deg(v)=r) &= \frac{\sum_{k=\frac{r}{s}}^{r}\frac{q}{k}\,\mathbb{P}(M=k)\,\mathbb{P}(deg(v)=r|M=k)}{\sum_{k=\frac{r}{s}}^{r}\mathbb{P}(M=k)\,\mathbb{P}(deg(v)=r|M=k)}\\
&= (1+o(1))\,q\frac{\sum_{k=\frac{r}{s}}^{r}\frac{1}{k}\cdot\left(\frac{d}{(s-1)qk}\right)^k e^{-\frac{d}{(s-1)q}+\frac{sk}{n}}\left(\frac{skq}{r}\right)^r e^{-qsk+qr}}{\sum_{k=\frac{r}{s}}^{r}\left(\frac{d}{(s-1)qk}\right)^k e^{-\frac{d}{(s-1)q}+\frac{sk}{n}}\left(\frac{skq}{r}\right)^r e^{-qsk+qr}}\\
&= (1+o(1))\,q\frac{\sum_{k=\frac{r}{s}}^{r}\frac{1}{k}\cdot\left(\frac{d}{(s-1)q}\right)^k k^{r-k}e^{-qsk}}{\sum_{k=\frac{r}{s}}^{r}\left(\frac{d}{(s-1)q}\right)^k k^{r-k}e^{-qsk}}. \qquad (4.8)
\end{aligned}$$

Writing $a = qs - \ln(d/(s-1)q)$, this is

$$q\frac{\sum_{k=\frac{r}{s}}^{r}\frac{1}{k}\cdot k^{r-k}e^{-ak}}{\sum_{k=\frac{r}{s}}^{r}k^{r-k}e^{-ak}}.$$

Therefore Equation (4.8) is the same as $q\,\mathbb{E}(1/x)$ when x is a nonnegative integer drawn from the discrete distribution with density proportional to $f(x) = x^{r-x}e^{-ax}$. We let z be as in Equation (4.6) of Lemma 4.2.10, so $z \approx \frac{r}{sq}$. We use Lemma 4.2.10 to bound

$$\begin{aligned}
\mathbb{E}\left(\left|\frac{1}{x}-\frac{1}{z}\right|\right) &\le \sum_{t=1}^{\infty}\left(\frac{1}{z}-\frac{1}{z+t\sqrt{z}}\right)e^{-t} + \sum_{t=1}^{\sqrt{z}-1}\left(\frac{1}{z-t\sqrt{z}}-\frac{1}{z}\right)e^{-t}\\
&= \sum_{t=1}^{\infty}\frac{t\sqrt{z}e^{-t}}{z(z+t\sqrt{z})} + \sum_{t=1}^{\sqrt{z}-1}\frac{t\sqrt{z}e^{-t}}{z(z-t\sqrt{z})}\\
&\le \frac{1}{z}\sum_{t=1}^{\infty}\frac{te^{-t}}{\sqrt{z}+1} + \frac{\sqrt{z}}{z}\left(\sum_{t=1}^{\sqrt{z}/3}\frac{3te^{-t}}{2z} + \sum_{t=\sqrt{z}/3}^{\sqrt{z}-1}te^{-t}\right)
\end{aligned}$$

$$= \frac{O(1)}{z\sqrt{z}} + \frac{O(1)}{z\sqrt{z}} + O\left(\frac{\sqrt{z}}{3}e^{-\frac{\sqrt{z}}{3}}\right) = \frac{O(1)}{z\sqrt{z}}.$$

Using this and approximating z by $\frac{r}{sq}$, the expectation of x with respect to the density proportional to f can be estimated:

$$q\,\mathbb{E}\left(\frac{1}{x}\right) = \frac{q}{z}\left(1 + O\left(\frac{1}{\sqrt{z}}\right)\right) = (1+o(1))\,\frac{sq^2}{r}\left(1 + O\left(\sqrt{\frac{sq}{r}}\right)\right) = (1+o_r(1))\frac{sq^2}{r}$$

as claimed.

□

4.2.3 Varied degree distributions: the DROC extension

In this section we introduce an extension of our model which produces graphs that match a target degree distribution in expectation. In each iteration a modified Chung-Lu random graph is placed instead of an E-R random graph.

DROC(n, D, s, q).

Input: number of vertices n, target degree sequence $D = t(v_1), \ldots t(v_n)$ with mean d.

Output: a graph on n vertices where vertex v_i has expected degree $t(v_i)$.

Repeat $n/((s-1)q)$ times:

1. Pick a random subset S of vertices (from $\{1, 2, \ldots, n\}$) by selecting each vertex with probability s/n.
2. Add a modified C-L random graph on S, i.e., for each pair in S, add the edge between them independently with probability $\frac{qt(v_i)t(v_j)}{sd}$; if the edge already exists, do nothing.

Theorem 4.2.11. *Given a degree distribution D with mean d and $\max_i t(v_i)^2 \leq \frac{sd}{q}$, DROC$(n, D, s, q)$ yields a graph where vertex v_i has expected degree $t(v_i)$.*

We require $\max_i t(v_i)^2 \leq \frac{sd}{q}$ to ensure that the probability each edge is chosen is at most 1. In the DROC model the number of communities a vertex belongs to is independent of target degree $t(v)$. When $t(v) > \frac{sd}{q}$, if v participates in the average number of communities and is connected to all vertices in each of its communities, it likely will not reach degree $t(v)$. Therefore when s is low and q is high, the DROC model is less able to capture degree distributions with long upper tails. Moreover, when s is low and q is high, there will be more isolated vertices in a DROC graph since the expected fraction of isolated vertices is at least $(1 - s/n)^{n/(q(s-1))}$. In Theorem 4.2.13 we show that when s is low and q is high the clustering coefficient is largest. In this regard the DROC model is somewhat limited; it may not be possible to achieve some very high clustering coefficients while simultaneously capturing the upper tail of the degree distribution and avoiding isolated vertices.

The following corollary shows that it is possible to achieve a power law degree distribution with the DROC model for power law parameter $\gamma > 2$. We use $\zeta(\gamma) = \sum_{n=1}^{\infty} n^{-\gamma}$ to denote the Riemann zeta function.

Corollary 4.2.12. *Let $D \sim \mathcal{D}_\gamma$ be the power law degree distribution defined as follows:*

$$\mathbb{P}(t(v_i) = k) = \frac{k^{-\gamma}}{\zeta(\gamma)},$$

for all $1 \leq i \leq n$. If $\gamma > 2$ and

$$\frac{s}{q} = \omega(1)\frac{\zeta(\gamma)}{\zeta(\gamma - 1)} n^{\frac{1}{\gamma - 1}},$$

then with high probability D satisfies the conditions of Theorem 4.2.11, and therefore can be used to produce a DROC graph.

Taking the distribution D_d with $t(v) = d$ for all v in the DROC model does not yield $ROC(n, d, s, q)$. The model $DROC(n, D_d, s, q)$ is equivalent to $ROC(n, d, s, \frac{qd}{s})$.

By varying s and q we can control the clustering coefficient of a $DROC$ graph.

Theorem 4.2.13. *Let $C(v)$ denote the clustering coefficient of a vertex v in graph drawn from $DROC(n, D, s, q)$ with $\max t(v_i)^2 \leq \frac{sd}{q}$, $s = \omega(1)$, $s/n = o(q)$, and $t = t(v)$. Then*

$$\mathbb{E}(C(v)) = (1 + o(1)) \frac{\left(\sum_{u \in V} t(u)^2\right)^2}{d^3 n^2 s} \left((1 - e^{-t})^2 q^2 + c_t q^3\right),$$

where $c_t \in [0, 6.2)$ is a constant depending on t.

Equation Eq. (4.10) in the proof of the theorem gives a precise statement of the expected clustering coefficient conditioned on community membership.

DROC proofs

Proof. (of Theorem 4.2.13.) Let v be a vertex with target degree $t = t(v)$, and let k denote the number communities containing v. First we claim $deg(v) \sim Bin\left((s-1)k, \frac{tq}{s}\right)$. Let s be an arbitrary vertex of a community S containing v.

$$\mathbb{P}(s \sim v \text{ in } S) = \sum_{u \in V} \mathbb{P}(s = u)\, \mathbb{P}(v \sim u \text{ in } S) = \sum_{u \in V} \frac{1}{n} \frac{t(u)tq}{ds} = \frac{tq}{s}.$$

A vertex in k communities has the potential to be adjacent to $(s-1)k$ other vertices, and each adjacency occurs with probability tq/s.

Next, let N_u be the event that a randomly selected neighbor of vertex v is vertex u. We compute

$$\mathbb{P}(N_u) = \sum_r \frac{\mathbb{P}(u \sim v \mid deg(v) = r)\, \mathbb{P}(deg(v) = r)}{r}$$

$$
\begin{aligned}
&= \sum_r \frac{\mathbb{P}(u \sim v)\,\mathbb{P}(deg(v) = r \mid u \sim v)}{r} \\
&= \mathbb{P}(u \sim v)\,\mathbb{E}\left(\frac{1}{deg(v)} \mid u \sim v\right) \\
&= (1 + o(1))\left(\frac{s}{n}\right)^2 \frac{n}{(s-1)q} \frac{t(u)tq}{sd}\left(\frac{1 - e^{-tqk}}{tkq}\right) \qquad (4.9) \\
&= (1 + o(1))\,\frac{t(u)\left(1 - e^{-tqk}\right)}{qkdn}.
\end{aligned}
$$

To see Equation (4.9), note that by the first claim $\mathbb{E}\left(\frac{1}{deg(v)} \mid u \sim v\right) = \mathbb{E}\left(\frac{1}{X+1}\right)$ where $X \sim Bin\left((s-1)k - 1, \frac{tq}{s}\right)$. Applying Lemma 4.2.8 and assuming $s = \omega(1)$, we obtain

$$
\mathbb{E}\left(\frac{1}{deg(v)} \mid u \sim v\right) = \frac{1 - (1 - \frac{tq}{s})^{(s-1)k}}{((s-1)k)\frac{tq}{s}} = (1 + o(1))\,\frac{1 - e^{-tqk}}{tkq}.
$$

Now we compute the expected clustering coefficient conditioned on the number of communities the vertex is part of under the assumption that $s/n = o(q)$. Observe

$$
\begin{aligned}
\mathbb{E}(C(v) \mid v \text{ in } k \text{ communities}) &= \sum_{u,w} N_u N_w\,\mathbb{P}(u \sim w \mid u \sim v \text{ and } w \sim v) \\
&= \sum_{u,w} \frac{t(u)t(w)\left(1 - e^{-tqk}\right)^2}{(qkdn)^2}\left(\frac{1}{k} + \left(\frac{s}{n}\right)^2 \frac{n}{(s-1)q}\right)\frac{t(u)t(w)q}{sd} \\
&= (1 + o(1))\,\frac{\left(1 - e^{-tqk}\right)^2\left(\sum_{u \in V} t(u)^2\right)^2}{qd^3k^3n^2s}. \qquad (4.10)
\end{aligned}
$$

Next compute the expected clustering coefficient without conditioning on the number of communities. To do so we need to compute the expected value of the function $f(k) = \frac{(1-e^{-kqt})^2}{k^3}$. We first use Taylor's theorem to give bounds on $f(k)$. For all k, there exists some $z \in [1/q, k]$ such that

$$
f(k) = f\left(\frac{1}{q}\right) + f'\left(\frac{1}{q}\right)\left(k - \frac{1}{q}\right) + \frac{f''(z)}{2}\left(k - \frac{1}{q}\right)^2.
$$

Note that for $z \in [1/q, k]$

$$\begin{aligned} f''(z) &= \frac{12(1-e^{-kqt})^2}{k^5} - \frac{12e^{-kqt}(1-e^{-kqt})qt}{k^4} + \frac{2e^{-2kqt}q^2t^2}{k^3} - \frac{2e^{-kqt}(1-e^{-kqt})q^2t^2}{k^3} \\ &\leq \frac{12(1-e^{-kqt})^2}{k^5} + \frac{2e^{-2kqt}q^2t^2}{k^3} \\ &\leq q^5\left(12 + 2t^2e^{-2t}\right), \end{aligned}$$

and

$$f''(z) \geq 0.$$

It follows that

$$f\left(\frac{1}{q}\right) + f'\left(\frac{1}{q}\right)\left(k - \frac{1}{q}\right) \leq f(k) \leq f\left(\frac{1}{q}\right) + f'\left(\frac{1}{q}\right)\left(k - \frac{1}{q}\right) + q^5\left(6 + t^2e^{-2t}\right)\left(k - \frac{1}{q}\right)^2. \tag{4.11}$$

Let $M \sim Bin(n/(sq), s/n)$ be the random variable for the number of communities a vertex v is part of. (Since $s = \omega(1)$ replacing the number of communities by $n/(sq)$ changes the result by a factor of $(1 + o(1))$.) We use Equation (4.11) to give bounds on the expectation of $f(M)$,

$$\begin{aligned} \mathbb{E}(f(M)) &\leq \mathbb{E}\left(f\left(\frac{1}{q}\right) + f'\left(\frac{1}{q}\right)\left(M - \frac{1}{q}\right) + q^5\left(12 + 2t^2e^{-2t}\right)\left(M - \frac{1}{q}\right)^2\right) \\ &= (1-e^{-t})^2q^3 + \frac{1}{q}\left(1 - \frac{s}{n}\right)q^5\left(6 + t^2e^{-2t}\right) \\ &\leq (1-e^{-t})^2q^3 + q^4\left(6 + t^2e^{-2t}\right) \end{aligned}$$

and

$$\mathbb{E}(f(M)) \geq \mathbb{E}\left(f\left(\frac{1}{q}\right) + f'\left(\frac{1}{q}\right)\left(M - \frac{1}{q}\right)\right) = (1-e^{-t})^2q^3.$$

Therefore $\mathbb{E}(f(M)) = (1-e^{-t})^2 q^3 + c_t q^4$ for some constant $c_t \in [0, 6.2)$.

Finally, we compute

$$\begin{aligned}\mathbb{E}(C(v)) &= \sum_k \mathbb{P}(M=k) \frac{\left(1-e^{-tqk}\right)^2 \left(\sum_{u\in V} t(u)^2\right)^2}{qd^3k^3n^2s} \\ &= \frac{\left(\sum_{u\in V} t(u)^2\right)^2}{qd^3n^2s} \mathbb{E}(f(M)) \\ &= (1+o(1)) \frac{\left(\sum_{u\in V} t(u)^2\right)^2}{d^3n^2s} \left((1-e^{-t})^2 q^2 + c_t q^3\right).\end{aligned}$$

□

Proof. (of Corollary 4.2.12.) Let $d = mean(D)$. We compute

$$\mathbb{E}(d) = \sum_{k=1}^{\infty} \frac{k^{-\gamma+1}}{\zeta(\gamma)} = \frac{\zeta(\gamma-1)}{\zeta(\gamma)}.$$

Next we claim that with high probability the maximum target degree of a vertex is at most $t_0 = n^{2/(\gamma-1)}$. Let X be the random variable for the number of indices i with $t(v_i) > k_0$.

$$\begin{aligned}\mathbb{P}\Big(\max_i t(v_i) > t_0\Big) \le \mathbb{E}(X) &= n\,\mathbb{P}(t(v_1) > t_0) \le n \sum_{i=t_0+1}^{\infty} \frac{i^{-\gamma}}{\zeta(\gamma)} \\ &\le n \int_{i=t_0}^{\infty} \frac{i^{-\gamma}}{\zeta(\gamma)} = \left(\frac{1}{\zeta(\gamma)(\gamma-1)}\right) n t_0^{1-\gamma} = o(1).\end{aligned}$$

It follows that $\max_i t(v_i)^2 \le n^{\frac{1}{\gamma-1}}$, and so $\max_i t(v_i)^2 \le \frac{sd}{q}$. □

4.2.4 Comparison to other random graph models.

The ROC model captures any pair of triangle-to-edge and four-cycle-to edge ratios simultaneously, and the DROC model can exhibit a wide range of degree distributions with high clustering coefficient. Previous work [HK02], [ORS13], and [Rav+02] provides models that produce

power law graphs with high clustering coefficients. Their results are limited in that the resulting graphs are restricted to a limited range of power-law parameters, and are either deterministic or only analyzable empirically. In contrast, the DROC model is a fully random model designed for a variety of degree distributions (including power law with parameter $\gamma > 2$) and can provably produce graphs with a range of clustering coefficient. The algorithm presented in [Vol04] produces graphs with tunable degree distribution and clustering, but unlike ROC graphs, there is no underlying community structure and the resultant graphs do not exhibit the commonly observed inverse relationship between degree and clustering coefficient.

The Block Two-Level Erdős and Rényi (BTER) model produces graphs with scale-free degree distributions and random dense communities [SKP12]. However, the communities in the BTER model do not overlap; all vertices are in precisely one E-R community and all other edges are added during a subsequent configuration model phase of construction. Moreover, in the BTER model community membership is determined by degree, which ensures that all vertices in a BTER community have similar degree. In contrast, the degree distribution within a DROC community is a random sample of the entire degree distribution.

Mixed membership stochastic block models have traditionally been applied in settings with overlapping communities [Air+08], [KN11], [Air+06]. The ROC model differs in two key ways. First, unlike low-rank mixed membership stochastic block models, the ROC model can produce sparse graphs with high triangle and four-cycle ratios. As discussed in the introduction, the over-representation of particular motifs in a graph is thought to be fundamental for its function, and therefore modeling this aspect of local structure is important. Second, in a stochastic block model the size and density of each community and the density between communities are all specified by the model. As a result, the size of the stochastic block model must grow with the number of communities, but the ROC model maintains a succinct description. This observation suggests the ROC model may be better

suited for graphs in which there are many communities that are similar in structure, whereas the stochastic block model is better suited for graphs with a small number of communities with fundamentally different structures.

4.3 Convergent sequences of sparse graphs

In this section, we discuss a few example sequences to illustrate the notions of convergence. Section 4.3.2 focuses on the convergence of sequences of random graphs. Not all sequences of graphs converge or contain a convergent subsequence according to our defintion; see Section 4.6.1 for an example.

4.3.1 Hypercube and rook graphs

We begin with the hypercube sequence, which directly motivates this paper.

Lemma 4.3.1 (hypercube limit)**.** *The d-dimensional hypercube is a graph on 2^d vertices, each labeled with a string in $\{0,1\}^d$. Two vertices are adjacent if the Hamming distance of their labels is 1. Let (G_d) be the sequence of d-dimensional hypercubes. The sparsity exponent of the sequence (G_d) is $1/2$ and the sequence is fully convergent with limit $(w_3, w_4, \dots)$ where*

$$w_k = \begin{cases} (k-1)!! & \text{for } k \text{ even} \\ 0 & \text{for } k \text{ odd.} \end{cases}$$

The k-limit of the sequence is $(w_3, w_4, \dots w_k)$.

Proof. We claim that for k even $W_k(G_d) = (k-1)!! n d^{k/2} + o\left(n d^{k/2}\right)$ where $n = 2^d$. Each hypercube edge (u, v) corresponds to a one coordinate difference between the labels of u and v. We think of k-walks on the hypercube as length k strings where the i^{th} character indicates which of the d coordinates is changed on the i^{th} edge of the walk. In closed

walks each coordinate that is changed is changed back, so every coordinate appearing in the corresponding string appears an even number of times. Therefore at most $k/2$ coordinates appear in the string. Let Y_i be the number of length k strings with i distinct characters that correspond to a closed k-walk. Since there are d possible coordinates, there are $\binom{d}{i}$ ways to select the i characters and so $Y_i = \Theta(d^i)$. Therefore

$$W_k(G_d) = nY_{k/2} + o\left(nd^{k/2}\right).$$

There are $\binom{d}{k/2}$ ways to select the coordinates to change and $k!/2^{k/2}$ length k strings where $k/2$ characters appear twice. Thus

$$Y_{k/2} = d^{k/2}\frac{k!}{2^{k/2}\left(\frac{k}{2}\right)!} + o\left(d^{k/2}\right) = (k-1)!!d^{k/2} + o\left(d^{k/2}\right),$$

and the claim follows.

Note that there are no odd closed walks in the hypercube because it is bipartite. Therefore $W_k(G) = 0$ for k odd. It follows that the sparsity exponent is $1/2$ and the limit vector is as stated. □

Our second example is a strongly regular family with a different sparsity exponent.

Lemma 4.3.2 (rook's graph limit)**.** *The rook graph G_k on k^2 vertices is the Cartesian product of two cliques of size k. (Viewing the vertices as the squares of a $k \times k$ chessboard, the edges represent all legal moves of the rook.) Let (G_k) be the sequence on rook graphs. The sparsity exponent of (G_k) is 1 and the sequence is fully convergent with limit $(w_3, w_4, \dots)$ where $w_j = 2^{2-j}$.*

Proof. The rook's graph is the strongly regular graph on $n = k^2$ vertices with degree $d = 2k-2$ such that each pair of adjacent vertices have $\lambda = k-2$ common neighbors and each pair of non-adjacent vertices have $\mu = 2$ common neighbors. The classical result [BH12] states that

the eigenspectrum of a strongly regular graph is

d with multiplicity 1,

$\frac{1}{2}\left((\lambda-\mu)+\sqrt{(\lambda-\mu)^2+4(d-\mu)}\right)$ with multiplicity $\frac{1}{2}\left((n-1)-\frac{2d+(v-1)(\lambda-\mu)}{\sqrt{(\lambda-\mu)^2+4(d-\mu)}}\right)$, and

$\frac{1}{2}\left((\lambda-\mu)-\sqrt{(\lambda-\mu)^2+4(d-\mu)}\right)$ with multiplicity $\frac{1}{2}\left((n-1)+\frac{2d+(v-1)(\lambda-\mu)}{\sqrt{(\lambda-\mu)^2+4(d-\mu)}}\right)$.

Therefore the eigenspectrum of the rook graph G_k is

$2k-2$ with multiplicity 1, -2 with multiplicity $(k-1)^2$, and $k-2$ with multiplicity $2k-2$.

We compute

$$W_j(G_k) = (2k-2)^j + (k-1)^2(-2)^j + (2k-2)(k-2)^j = 2k^{j+1} + O\left(k^j\right).$$

Therefore

$$\lim_{k\to\infty} W_j(G_k, 1) = \lim_{k\to\infty} \frac{2k^{j+1} + O\left(k^j\right)}{k^2(2k-2)^{j-1}} = 2^{2-j}.$$

□

4.3.2 Almost sure convergence for sequences of random graphs

By an abuse of notation, we say that a sequence of random graphs converges to a limit vector L if a sequence of graphs drawn from the sequence of random graph models almost surely converges L. Lemma 4.3.4 gives a method for showing that a sequence of random graphs converges, which we apply to describe the limits of sequence of E-R graphs (Lemma 4.3.6). We will again apply Lemma 4.3.4 when we discuss the convergence of sequences of ROC graphs (Theorem 4.4.13 and Corollary 4.4.14).

Definition 4.3.3 (convergence of random graph sequences)**.** Let $M = (M_i)$ be a sequence of random graph models. Let S be a sequence of graphs (S_i) where $S_i \sim M_i$. We say the sequence of random graphs M converges to L if a sequence S drawn from M almost surely converges to L.

Lemma 4.3.4. *Let $M = (M_i)$ be a sequence of random graph models, and let (S_i) be a sequence of graphs where $S_i \sim M_i$. Let $\varepsilon > 0$ and $A_{i,\varepsilon,\alpha}(w_j)$ be the event that $|W_j(S_i, \alpha) - w_j| \geq \varepsilon$.*

1. *If for all j and $\varepsilon > 0$*
$$\sum_{i=1}^{\infty} \mathbb{P}(A_{i,\varepsilon,\alpha}(w_j)) < \infty,$$
then M converges to $L = (w_3, w_4, \dots)$ with sparsity exponent α.

2. *If the above hypothesis holds for all $j \leq k$, then M k-converges to $L = (w_3, w_4, \dots w_k)$ with k-sparsity exponent α.*

3. *Let $D(S_i)$ be the random variable for the average degree of a vertex in S_i, let $d_i = \mathbb{E}(D(S_i))$, and let n_i be the number of vertices of S_i. If $\lim_{i\to\infty} \frac{\mathbb{E}(W_j(S_i))}{n_i d_i^{1+\alpha(j-2)}} = w_j$ then there exists an index i_0 and a constant C such that*
$$\sum_{i=1}^{\infty} \mathbb{P}(A_{i,\varepsilon,\alpha}(w_j)) \leq C + \sum_{i=i_0}^{\infty} \frac{\mathsf{Var}(D(S_i))}{d_i^2} + \frac{\mathsf{Var}(W_j(S_i))}{\left(n_i d_i^{1+\alpha(j-2)}\right)^2}.$$

Proof. We begin with (1) and (2). Fix j. To show that $W_j(S_n, \alpha) \to w_j$ almost surely, it suffices to show that for all $\varepsilon > 0$, $\mathbb{P}(A_{i,\varepsilon,\alpha}(w_j)$ occurs infinitely often$) = 0$. By the Borel Cantelli Lemma $\sum_{n=1}^{\infty} \mathbb{P}(A_{i,\varepsilon,\alpha}(w_j)) < \infty$ implies $\mathbb{P}(A_{i,\varepsilon,\alpha}(w_j)$ occurs infinitely often$) = 0$. Statements (1) and (2) follow from the fact that a countable intersection of almost sure events occurs almost surely.

For (3), we apply Lemma 4.3.5 which bounds the probability $W_j(S_i, \alpha)$ deviates from expectation by separately bounding the probabilities that the number of edges and the number of closed j-walks in S_i deviate from expectation. Let $g_i = w_j - \frac{\mathbb{E}(W_j(G_i))}{n_i d_i^{1+\alpha(j-2)}}$, and so $g_i = o(1)$. Let i_0 be such that for all $i \geq i_0$, $|g_i| < \varepsilon/4$, and let $c = \min\{\delta, \varepsilon/4\}$. By Lemma 4.3.4(3) for all $i \geq i_0$

$$\mathbb{P}(A_{i,\varepsilon,\alpha}(w_j)) \leq \frac{1}{c^2}\left(\frac{\mathsf{Var}(D(G_i))}{d_i^2} + \frac{\mathsf{Var}(W_j(G_i))}{\left(n_i d_i^{1+\alpha(j-2)}\right)^2}\right).$$

The claim follows from the observation that

$$\sum_{i=1}^{\infty} \mathbb{P}(A_{i,\varepsilon,\alpha}(w_j)) \leq i_0 + \frac{1}{c^2}\sum_{i=i_0}^{\infty} \frac{\mathsf{Var}(D(S_i))}{d_i^2} + \frac{\mathsf{Var}(W_j(S_i))}{\left(n_i d_i^{1+\alpha(j-2)}\right)^2}.$$

□

Lemma 4.3.5. *Let S be a random graph on n vertices. Let $\varepsilon > 0$ and $A_{\varepsilon,\alpha}(w_j)$ be the event that $|W_j(S, \alpha) - w_j| \geq \varepsilon$. Let $D(S)$ be the random variable for the average degree of a vertex in S, and let $d = \mathbb{E}(D(S))$. Let $g = w_j - \frac{\mathbb{E}(W_j(S))}{nd^{1+\alpha(j-2)}}$. For $|g| < \varepsilon/2$, $\delta = \min\left\{\frac{\varepsilon}{2j(w_j+\varepsilon)}, \frac{1}{2(j-1)^2}\right\}$ and $\lambda = \varepsilon/2 - |g|$,*

$$\mathbb{P}(A_{\varepsilon,\alpha}(w_j)) \leq \frac{\mathsf{Var}(D(S))}{\delta^2 d^2} + \frac{\mathsf{Var}(W_j(S))}{\lambda^2\left(nd^{1+\alpha(j-2)}\right)^2}.$$

Proof. Observe that if $A_{i,\varepsilon,\alpha}(w_j)$ holds, then for any $\delta > 0$ at least one of the following events hold:

(a) $|D(S) - d| > \delta d$

(b) $W_j(S) \geq (w_j + \varepsilon)\left(n\left(d(1-\delta)\right)^{1+\alpha(j-2)}\right)$

(c) $W_j(S) \leq (w_j - \varepsilon)\left(n\left(d(1+\delta)\right)^{1+\alpha(j-2)}\right)$.

When (a) does not hold

$$\frac{W_j(S)}{n\left(d(1+\delta)\right)^{1+\alpha(j-2)}} \leq W_j(S,\alpha) \leq \frac{W_j(S)}{n\left(d(1-\delta)\right)^{1+\alpha(j-2)}}.$$

Assume (a) does not hold and $A_{\varepsilon,\alpha}(w_j)$. If $W_j(S,\alpha) \geq w_j + \varepsilon$ then (b) holds. If $W_j(S,\alpha) \leq w_j - \varepsilon$ then (c) holds. The observation follows.

We now give a bound on the probability of (b) or (c). Let $\gamma^- = 1-(1-\delta)^{1+\alpha(j-2)}$ and $\gamma^+ = (1+\delta)^{1+\alpha(j-2)} - 1$. We write $w_j = \frac{\mathbb{E}(W_j(S))}{nd^{1+\alpha(j-2)}} + g$. Statement (b) becomes

$$W_j(S) - \mathbb{E}(W_j(S)) \geq \left(\varepsilon + g - \gamma^-\left(w_j+\varepsilon\right)\right) nd^{1+\alpha(j-2)},$$

and statement (c) becomes

$$W_j(S) - \mathbb{E}(W_j(S)) \leq \left(\gamma^+\left(w_j-\varepsilon\right) - \varepsilon + g\right) nd^{1+\alpha(j-2)}.$$

Under the assumptions that $\delta = \min\left\{\frac{\varepsilon}{2j(w_j+\varepsilon)}, \frac{1}{2(j-1)^2}\right\}$ and $\alpha \leq 1$,

$$\gamma^- = 1-(1-\delta)^{1+\alpha(j-2)} \leq \delta(1+\alpha(j-2)) < \delta j \leq \frac{\varepsilon}{2(w_j+\varepsilon)}$$

$$\gamma^+ = (1+\delta)^{1+\alpha(j-2)} - 1 \leq \delta(j-1) + \sum_{i=2}^{j-1} \delta^i \binom{j-1}{i} \leq \delta(j-1) + 2\delta^2(j-1)^2 < \delta j \leq \frac{\varepsilon}{2(w_j-\varepsilon)}.$$

Let $\lambda = \varepsilon/2 - |g|$ and note

$$\varepsilon + g - \gamma^-\left(w_j+\varepsilon\right) \geq \lambda \quad \text{ and } \quad \varepsilon - g - \gamma^+\left(w_j-\varepsilon\right) \geq \lambda.$$

It follows from Chebyshev's inequality that

$$\mathbb{P}((b) \text{ or } (c)) \leq \mathbb{P}\left(|W_j(S) - \mathbb{E}(W_j(S))| \geq c(\delta) nd^{1+\alpha(j-2)}\right) \leq \frac{\mathsf{Var}(W_j(S))}{\left(\lambda nd^{1+\alpha(j-2)}\right)^2}.$$

Finally, we apply Chebyshev's inequality to bound the probability of (a), apply a union bound for the event $A_{i,\varepsilon,\alpha}(w_j)$, and obtain

$$\mathbb{P}(A_{\varepsilon,\alpha}(w_j)) \leq \mathbb{P}((a)) + \mathbb{P}((b) \text{ or } (c)) \leq \frac{\mathsf{Var}(D(S))}{\delta^2 d^2} + \frac{\mathsf{Var}(W_j(S))}{\lambda^2 \left(nd^{1+\alpha(j-2)}\right)^2}.$$

□

Our final example is sequences of Erdős-Rényi random graphs. These demonstrate some of the subtler issues with defining limits.

Lemma 4.3.6 (Erdős-Rényi sequence)**.** *Let $(G_n) \sim G\left(n^{2\ell}, n^{2-2\ell}\right)$ for $\ell > 1$. We denote the j^{th} Catalan number $Cat_j = \frac{1}{j+1}\binom{2j}{j}$.*

1. *For $k < 2\ell$, the k-sparsity exponent of (G_n) is $1/2$ and the k-limit is $(w_3, w_4, \dots w_k)$ where $w_j = 0$ for i odd and $w_j = Cat_{i/2}$ for i even.*

2. *For $k = 2\ell$, the k-sparsity exponent of (G_n) is $1/2$ and the k-limit is $(w_3, w_4, \dots, w_{k-1}, \overline{w}_k)$ where $w_j = 0$ for odd j, $w_j = Cat_{j/2}$ for even j, and $\overline{w}_k = w_k + 1$.*

3. *For $k > 2\ell$, the sparsity exponent of (G_n) is $\frac{k-\ell-1}{k-2}$ and the k-limit is $(w_3, w_4, \dots, w_k)$ where $w_j = 0$ for $j < k$, $w_k = 1$.*

4. *The sparsity exponent of (G_n) is 1 and the limit is $(0, 0, \dots)$.*

First we compute the expectation and variance of the number of closed i-walks in a E-R random graph.

Lemma 4.3.7. *Let $G \sim G(n, d/n)$. Let $W_j(G)$ be the random variable for the number of closed j walks in G. Then*

$$\mathbb{E}(W_j(G)) = d^j + Cat_{j/2} n d^{\lfloor j/2 \rfloor} + \Theta\left(d^{j-1} + n d^{\lfloor j/2 \rfloor - 1}\right)$$

$$\mathsf{Var}(W_j(G)) = \Theta\left(d^{2j-1} + n^2 d^{2\lfloor j/2 \rfloor - 1} + n d^{\lfloor j/2 \rfloor + j - 1}\right).$$

Proof. Let $W_j^{a,b}(G_n)$ be the number of closed j walks involving a vertices and b edges, so $b \leq j$ and either $a \leq b$ (the walk contains a cycle) or $a = b+1$ and $b \leq j/2$ (the walk traces a tree). Let $f(a,b,j)$ be the number of closed j-walks with b total edges on a labeled vertices $1, 2, \ldots a$ such that the order in which the vertices are first visited is $1, 2, \ldots a$. Note $f(j,j,j) = 1$ and $f(j/2+1, j/2, j) = Cat_{j/2}$ for j even (because there are Cat_b ordered trees on b edges, see [Sta97]). Let $\zeta(j)$ be one if j is even and zero otherwise. We split the sum based on whether the walk contains a cycle or traces a tree and compute

$$\begin{aligned}
\mathbb{E}(W_j(G)) &= \sum_{b=1}^{j}\sum_{a=1}^{b+1}\mathbb{E}\left(W_j^{a,b}(G)\right) = \sum_{b=1}^{j}\sum_{a=1}^{b+1} f(a,b,j)\frac{n!}{(n-a)!}\left(\frac{d}{n}\right)^b \\
&= \sum_{b=3}^{j}\sum_{a=1}^{b} f(a,b,j)\frac{n!}{(n-a)!}\left(\frac{d}{n}\right)^b + \sum_{b=1}^{\lfloor j/2 \rfloor} f(b+1,b,j)\frac{n!}{(n-(b+1))!}\left(\frac{d}{n}\right)^b \\
&= d^j + \zeta(j) Cat_{j/2} n d^{j/2} + \Theta\left(d^{j-1} + n d^{\lfloor j/2 \rfloor - 1}\right).
\end{aligned}$$

To find the variance of $W_j(G)$ we compute the expectation squared. Let $P_j^{a,b}(G)$ be the number of pairs of closed j walks involving a total of a vertices and b edges. Let $g(a,b,j)$ be the number of pairs of closed j-walks with b total edges on a labeled vertices $1, 2, \ldots a$ such that the order in which the vertices are first visited is $1, 2, \ldots a$ when the first walk is traversed then the second walk. Note $g(2j, 2j, j) = 1$ and $g(j+2, j, j) = \left(Cat_{j/2}\right)^2$ (since there are $\left(Cat_b\right)^2$ ways to pick two disjoint ordered trees on b edges).

We split the sum based on whether both walks contain a cycle, or both trace trees, or one traces a tree and one traces a cycle and compute

$$\mathbb{E}\left(W_j(G)^2\right) = \sum_{b=1}^{2j}\sum_{a=1}^{b+2}\mathbb{E}\left(P_j^{a,b}(G)\right) = \sum_{b=1}^{2j}\sum_{a=1}^{b+2} g(a,b,j)\frac{n!}{(n-a)!}\left(\frac{d}{n}\right)^b$$

$$
\begin{aligned}
&= \sum_{b=1}^{2j} \sum_{a=1}^{b} g(a,b,j) \frac{n!}{(n-a)!} \left(\frac{d}{n}\right)^b + \sum_{b=1}^{2\lfloor j/2 \rfloor} g(b+2,b,j) \frac{n!}{(n-(b+2))!} \left(\frac{d}{n}\right)^b \\
&\quad + \sum_{b=1}^{2j} \sum_{a=1}^{b+1} g(a,b,j) \frac{n!}{(n-a)!} \left(\frac{d}{n}\right)^b \\
&= d^{2j} + \zeta(j) \left(Cat_{j/2}\right)^2 n^2 d^j + \Theta\left(d^{2j-1} + n^2 d^{2\lfloor j/2 \rfloor - 1} + n d^{\lfloor j/2 \rfloor + j - 1}\right).
\end{aligned}
$$

It follows

$$
\mathsf{Var}(W_j(G)) = \mathbb{E}\big(W_j(G)^2\big) - \mathbb{E}(W_j(G))^2 = \Theta\left(d^{2j-1} + n^2 d^{2\lfloor j/2 \rfloor - 1} + n d^{\lfloor j/2 \rfloor + j - 1}\right).
$$

□

Now we use these computations and apply Lemma 4.3.4 to prove Lemma 4.3.6.

Proof. (of Lemma 4.3.6). By Lemma 4.3.7

$$
\mathbb{E}(W_j(G_n)) = \begin{cases} n^{2j} + Cat_{j/2} n^{2\ell + j} + o\left(n^{2j} + n^{2\ell + 2\lfloor j/2 \rfloor}\right) & j \text{ is even} \\ n^{2j} + o\left(n^{2j} + n^{2\ell + 2\lfloor j/2 \rfloor}\right) & j \text{ is odd.} \end{cases}
$$

We compute the k-sparsity exponent

$$
\alpha_k = \inf_{a \in [1/2,1]} \left\{ a \mid \mathbb{E}(W_j(G_n)) = O\left(n^{2\ell + 2 + 2\alpha(j-2)}\right) \text{ for all } j \leq k \right\} = \max\left\{ \frac{1}{2}, \frac{k - \ell - 1}{k-2} \right\},
$$

and the sparsity exponent

$$
\alpha = \inf_{a \in [1/2,1]} \left\{ a \mid \mathbb{E}(W_j(G_n)) = O\left(n^{2\ell + 2 + 2\alpha(j-2)}\right) \text{ for all } j \right\} = 1.
$$

Note for each of the cases outlined in the statement, $w_j = \lim_{i \to \infty} \frac{\mathbb{E}(W_j(G_n))}{n^{2\ell + 2 + 2\alpha(j-2)}}$ where α is the corresponding sparsity exponent. To prove convergence for cases 1-3, we apply

Lemma 4.3.4(2) and for case 4 we apply Lemma 4.3.4(1). By Lemma 4.3.4(3), it remains to show that

$$\sum_{n=n_0}^{\infty}\left(\frac{\mathsf{Var}(D(G_n))}{d_n^2}+\frac{\mathsf{Var}(W_j(G_n))}{\left(n^{2\ell+2+2\alpha(j-2)}\right)^2}\right)<\infty. \tag{4.12}$$

Note $D(G_n)\sim Bin\left(\binom{n^{2\ell}}{2},n^{2-2\ell}\right)$, and so $\mathsf{Var}(D(G_n))=\binom{n^{2\ell}}{2}n^{2-2\ell}\left(1-n^{2-2\ell}\right)$ and $d_n=\binom{n^{2\ell}}{2}n^{2-2\ell}$. It follows

$$\sum_{n=2}^{\infty}\frac{\mathsf{Var}(D(G_n))}{d_n^2}=\sum_{n=2}^{\infty}\frac{\binom{n^{2\ell}}{2}n^{2-2\ell}\left(1-n^{2-2\ell}\right)}{\left(\binom{n^{2\ell}}{2}n^{2-2\ell}\right)^2}\leq\sum_{n=2}^{\infty}2n^{-2\ell-2}<\infty.$$

We show the sum of the variance term is finite by considering the cases separately. By Lemma 4.3.7,

$$\mathsf{Var}(W_j(G_n))=\Theta\left(n^{4j-2}+n^{4\ell+4\lfloor j/2\rfloor-2}+n^{2\ell+2\lfloor j/2\rfloor+2j-2}\right),$$

and so

$$X:=\frac{\mathsf{Var}(W_j(G_n))}{\left(n^{2\ell+2+2\alpha(j-2)}\right)^2}=\Theta\left(n^{4j-6-4\ell-4\alpha(j-2)}+n^{-2+(j-2)(2-4\alpha)}+n^{-2\ell+(j-2)(3-4\alpha)}\right).$$

For (1) and (2), $\alpha=1/2$, $j\leq k\leq 2\ell$, and so $X=\Theta\left(n^{2j-4\ell-2}+n^{-2}+n^{-2\ell+j-2}\right)=O\left(n^{-2}\right)$. For (3), $\alpha=\frac{k-\ell-1}{k-2}$, $k>2\ell$, and $j\leq k$. Since $\alpha\geq\frac{j-\ell-1}{j-2}$, $\Theta\left(n^{4j-6-4\ell-4\alpha(j-2)}\right)=O\left(n^{-2}\right)$. Since $\alpha>1/2$, $\Theta\left(n^{-2+(j-2)(2-4\alpha)}\right)=O\left(n^{-2}\right)$. Since $\alpha>1/2$ and $k>2\ell$, $\Theta\left(n^{-2\ell+(j-2)(3-4\alpha)}\right)=O\left(n^{-2\ell+k-2}\right)=O\left(n^{-2}\right)$. It follows that $X=O\left(n^{-2}\right)$. For (4) $\alpha=1$ and $j\geq 3$, and so $X=\Theta\left(n^{2-4\ell}+n^{2-2j}+n^{-2\ell-j+2}\right)=O\left(n^{-2}\right)$. Therefore in all cases

$$\sum_{n=2}^{\infty}\frac{\mathsf{Var}(W_j(G_n))}{\left(n^{2\ell+2+2\alpha(j-2)}\right)^2}=\sum_{n=2}^{\infty}O\left(n^{-2}\right)<\infty,$$

and the statement follows from Equation (4.12). □

4.4 Approximating a convergent sequence by a ROC

4.4.1 A parameterization of the ROC model

In this section, we introduce a parametrization of the ROC model which will be particularly convenient in proofs. The distribution $\mathcal{D}$ of a ROC family is specified by a number $a \in [0, 1]$ and a distribution μ on triples (m_i, q_i, β_i) with probability μ_i for the i^{th} triple. Communities are generated by repeatedly picking a triple from the distribution μ. When $\beta_i = 0$, the community has expected size $s = m_i d^a$ and density q_i. If $\beta_i = 1$, indicating balanced bipartiteness, the community is defined on a bipartite graph with $m_i d^a$ vertices expected in each class.

ROC(n, d, μ, a).

Input: number of vertices n, degree d, $a \in [0,1]$, and μ a distribution on a finite set of triples (m_i, q_i, β_i) where (m_i, q_i, β_i) is selected with probability μ_i, $m_i > 0$, $\sum_i \mu_i = 1$, $\beta_i \in \{0,1\}$, $0 \le q_i \le 1$, and $\max m_i d^a \le n$. Let B be the set of indices i such that $\beta_i = 1$ and B^c be the set of indices i such that $\beta_i = 0$. Let $x = 1/\left(\sum_{i \in B^c} \mu_i m_i^2 q_i + 2\sum_{i \in B} \mu_i m_i^2 q_i\right)$.

Output: a graph on n vertices with expected degree d.

Repeat xnd^{1-2a} times:

1. Randomly select a pair (m_i, q_i, β_i) from μ with probability μ_i.
2. If $\beta_i = 0$
   (a) Pick a random subset S of vertices (from $\{1, 2, \ldots, n\}$) by selecting each vertex independently with probability $m_i d^a/n$.
   (b) Add the random graph $G_{|S|,q_i}$ on S, i.e., for each pair in S, add the edge between them independently with probability q_i; if the edge already exists, do nothing.

   If $\beta_i = 1$
   (a) Pick a random subset S of vertices (from $\{1, 2, \ldots, n\}$) by selecting each vertex independently with probability $2m_i d^a/n$. For each vertex that is in S randomly assign it to either S_1 or S_2.
   (b) Add the bipartite random graph $G_{|S_1|,|S_2|,q_i}$ on S, i.e., for each pair $u \in S_1$ and $v \in S_2$, add the edge between them independently with probability q_i; if the edge already exists, do nothing.

The parameters (n, d, μ, a) are valid ROC parameters if the conditions described under *input* in the box hold. A ROC family $\mathcal{D} = (\mu, a)$ refers to the set of ROC models with parameters μ and a and any valid n and d.

The sparsity exponent of a sequence determines the parameter a of the ROC family that achieves the limit vector. If a vector is achievable with sparsity exponent α then the ROC family that achieves the vector will have parameter $a = \alpha$ unless $\alpha = 1/2$ and vector is the Catalan vector ($w_j = 0$ for j odd and $w_j = Cat_{j/2}$ for j even). In this case any ROC family with $a < 1/2$ achieves the vector.

4.4.2 Cycles, walks, and limits of ROC's

In this section we describe a combinatorial relationship between closed walk counts and simple cycle counts that appears in graphs in which each vertex is in approximately the same number of simple cycles (Definition 4.4.1). Throughout this paper, for convenience we refer to a simple k-cycle as a k-cycle. For example, under this convention a k-cycle graph has $2k$ k-cycles because there are $2k$ distinct closed walks that traverse a k-cycle.

Definition 4.4.1 (cycle-walk transform)**.** Let

$$\mathcal{S}_k = \left\{ \{(a_1, t_1), (a_2, t_2), \ldots, (a_j, t_j)\} \mid \sum_{i=1}^{j} a_i t_i = k, a_i \neq a_j \text{ for } i \neq j, a_i, t_i \in \mathbb{Z}^+, a_i > 1 \right\}.$$

Define $T((c_3, c_4, \ldots c_n)) = (w_3, w_4, \ldots w_n)$ as the invertible transform

$$w_k = \sum_{S \in \mathcal{S}_k} \frac{k!}{(\prod t_i!)(k + 1 - \sum t_i)!} \prod_{i=1}^{j} (c_{a_i})^{t_i}.$$

The transform T is analogously defined for infinite count vectors.

Remark 4.4.2. The first few terms of T are illustrated below:

$$w_3 = c_3$$

$$w_4 = 2 + c_4$$

$$w_5 = c_5 + 5c_3$$

$$w_6 = c_6 + 6c_4 + 3c_3^2 + 5$$

$$w_7 = 21c_3 + 7c_3c_4 + c_7 + 7c_5$$

$$w_8 = 8c_3c_5 + 28c_3^2 + c_8 + 8c_6 + 28c_4 + 4c_4^2 + 14$$

$$w_9 = 9c_3c_6 + 84c_3 + 12c_3^3 + 9c_7 + 36c_5 + 9c_4c_5 + c_9 + 72c_3c_4$$

$$w_{10} = 42 + 5c_5^2 + 180c_3^2 + 45c_3^2c_4 + 10c_4c_6 + 90c_3c_5 + 10c_8 + c_{10} + 45c_6 + 10c_3c_7 + 120c_4 + 45c_4^2.$$

We see that T is invertible by using induction to show that each c_j is completely determined by the vector $(w_3, \dots w_j)$. Note $c_3 = w_3$ is completely determined. Assume $c_3, \dots c_j$ have been completely determined. Note that $w_{j+1} = c_{j+1} + f((c_3, c_4, \dots c_j))$ for some function f. Since w_{j+1} is given and f is a function of values that are already determined, there is only one choice for c_{j+1}.

In Section 4.4.2 we derive the coefficient of $\prod_{i=1}^{j}(c_{a_i})^{t_i}$ in T by counting the number of walk structures that can be decomposed into $t_1, t_2, \dots t_j$ cycles of lengths $a_1, a_2, \dots a_j$ respectively. In Section 4.4.2, we define class of *locally regular* graphs in which each vertex is in the same number of cycles, and then show this class of graphs exhibits the relationship between cycles and closed walks given in Definition 4.4.1.

In Section 4.4.2, we prove Theorem 4.4.3, which describes the limit achieved by a ROC family (μ, a). The parameter a plays an important role. When $a < 1/2$, the closed walks that trace trees dominate the closed walk count, so the limit is the Catalan sequence. When $a > 1/2$ the closed walks that trace simple cycles dominate the closed walk count, so the

limit is the normalized number of expected simple cycles. However, when $a = 1/2$, cycles, trees, and other walk structures are all of the same order, and so the relationship given in Definition 4.4.1 appears in the limit.

Theorem 4.4.3. *Let (μ, a) be a ROC family. Let B be the set of all i such that $\beta_i = 1$, let B^c be the set of all i such that $\beta_i = 0$, and let $x = 1/\left(\sum_{i\in B^c} \mu_i m_i^2 q_i + 2\sum_{i\in B} \mu_i m_i^2 q_i\right)$. Define*

$$c(k) = \begin{cases} 1 & k = 2 \\ x\sum_{i\in B^c} \mu_i (m_i q_i)^k & k \text{ odd and } k \geq 3 \\ x\sum_{i\in B^c} \mu_i (m_i q_i)^k + 2x\sum_{i\in B} \mu_i (m_i q_i)^k & k \text{ even and } k \geq 4. \end{cases}$$

Let $Cat_n = \frac{1}{n+1}\binom{2n}{n}$ denote the n^{th} Catalan number, and let T be as given in Definition 4.4.1.

1. *If $a < 1/2$, the ROC family fully achieves the limit $(0, Cat_2, 0, Cat_3, \dots)$ with sparsity exponent $1/2$.*

2. *If $a = 1/2$, the ROC family fully achieves the limit $T((c(3), c(4), c(5), \dots))$ with sparsity exponent $1/2$.*

3. *If $a > 1/2$, the ROC family fully achieves the limit $(c(3), c(4), c(5), \dots)$ with sparsity exponent a.*

The ROC family achieves the corresponding length $k-2$ prefix as its k-limit with the same k-sparsity exponent for $k \geq 4$.

In Section 4.4.2, we show that the probability the normalized walk count $W_j(G, \alpha)$ of a ROC graph $G \sim ROC(n, d, \mu, a)$ deviates from w_j in the limit achieved by the family vanishes as d grows (Theorem 4.4.13). Corollary 4.4.14 gives conditions that guarantee that a sequence of graphs drawn from a common ROC family converges almost surely to the limit achieved by the family.

The cycle structure of closed walks

In order to count the number of closed walks in a graph, we divide the closed walks into classes based on the structure of the cycles appearing in the closed walk and then count the number of closed walks in each class. Each class is defined by a "cycle permutation" in which each non-zero character represents the first step of a cycle within the walk and each zero represents a step in a cycle that has already begun (Definition 4.4.4). In Lemma 4.4.5, we show that the number of cycle permutations corresponding to a walk made of $t_1, t_2, \ldots t_j$ cycles of lengths $a_1, a_2, \ldots a_j$ respectively is the coefficient of $\prod_{i=1}^{j} (c_{a_i})^{t_i}$ in the cycle-walk transform.

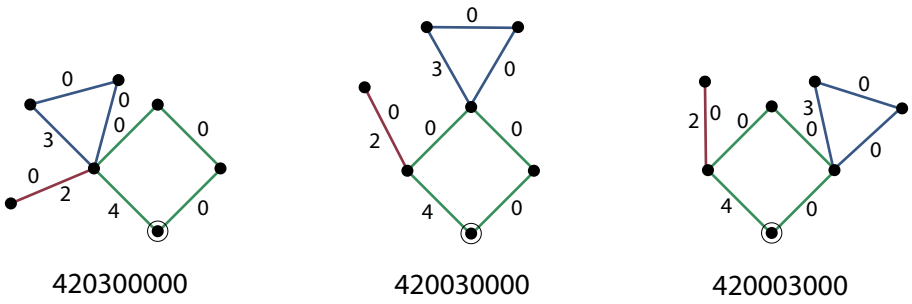

Figure 4.5: The above walks begin and end at the circled vertex and proceed left to right. Each is labeled with its cycle permutation.

Definition 4.4.4 (cycle permutation)**.** Follow the procedure below to label each step of a closed k-walk $\mathcal{W} = (r_1, r_2, \ldots r_k)$ with a label and define the "cycle permutation" P of $\mathcal{W}$ as the labels of the steps in order of traversal.

1. Repeat until all steps are labeled:

   Traverse $\mathcal{W}$ skipping a step r_i if it has already been labeled. Let u be the first repeated vertex on this traversal. The modified walk must have traversed a cycle $r_i, r_{i+1}, \ldots r_{i+j-1}$ starting at u. Label the first step r_i with the length of the cycle. Label all other steps with zero.

2. Traverse $\mathcal{W}$ and let P be the string of labels of the steps as they are traversed.

The following lemma enumerates the cycle permutations using bijection between cycle permutations and generalized Dyck paths (Definition 4.4.6).

Lemma 4.4.5. *Let $\mathcal{S}_k$ be as given in Definition 4.4.1. For each $S \in \mathcal{S}_k$, let M_s be the multiset where a_i appears t_i times and there are $k - \sum_i t_i$ zeros. Let P_S be the set of all permutations of M_s such that the following property holds for all $2 \leq i \leq k+1$:*

$$\sum_{\ell \in N(s)} (\ell - 1) \geq z_i$$

where $N(s)$ is the multiset of non-zero labels that appear before the i^{th} label of the permutation and z_i is the number of times zero occurs before the i^{th} label of the permutation. The set of cycle permutations is $\bigcup_{s \in S} P_S$ and

$$|P_S| = \frac{k!}{(\prod t_i!)(k+1-\sum t_i)!}.$$

To compute the size of P_S in the above lemma we use a bijection between permutations in P_S and generalized sub-diagonal Dyck paths, whose cardinality is given in Lemma 4.4.7.

Definition 4.4.6 (generalized Dyck path, see [Ruk11])**.** A generalized Dyck path p is a sequence of n vertical steps of height one and $k \leq n$ horizontal steps with positive integer lengths $\ell_1, \ell_2, \ldots \ell_k$ satisfying $\sum_{i=1}^{k} \ell_i = n$ on a $n \times n$ grid such that no vertical step is above the diagonal.

Lemma 4.4.7 (from [Ruk11])**.** *Let D the set of generalized Dyck paths on a $(k - \sum t_i) \times (k - \sum t_i)$ grid that are made up of t_i horizontal steps of length $a_i - 1$ and $k - \sum t_i$ vertical steps of length 1. Then*

$$|D| = \frac{k!}{(\prod t_i!)(k+1-\sum t_i)!}.$$

Proof. (of Lemma 4.4.5) First we show the set of cycle permutations is $\bigcup_{s\in S} P_S$. Let P be a cycle permutation of some closed walk $\mathcal{W}$ of length k. Let $s = \{(a_1, t_1), (a_2, t_2), \dots (a_j, t_j)\} \in S$ where $\{a_1, a_2, \dots a_j\}$ are the non-zero labels of P and each a_i appears t_i times in P (so $\sum a_i t_i = k$). To see that P is in P_S we show that for all $2 \leq i \leq k+1$:

$$\sum_{\ell \in N(s)} (\ell - 1) \geq z_i$$

where $N(s)$ is the multiset of non-zero labels that appear before the i^{th} label of P and z_i is the number of times zero occurs before the i^{th} label of P. Before the i^{th} step of the walk suppose non-zero labels $\ell_1, \dots \ell_k$ have been traversed. The only steps labeled with a zero that have been traversed must be part of a cycle corresponding to one of the labels $\ell_1, \dots \ell_k$. Since ℓ_i labels a cycle of length ℓ_i at most $\sum(\ell_i - 1)$ zero steps have been traversed. Thus $P \in P_S$.

Next we claim that any $P \in \bigcup P_S$ corresponds to a closed walk $\mathcal{W}$. Let $T(s) = \sum t_i$ be the number of non-zero values in each permutation in P_S. We show that for any $k = \sum a_i t_i$ all permutations in P_S correspond to closed walks by induction on $T(s)$. Note for any k and $T(s) = t_1 = 1$, there is one permutation in P_S, $k = a_1$ followed by $k-1$ zeros. This permutation corresponds to a k-cycle. Assume that if $T(s') < T(s)$ then each string in $P_{s'}$ corresponds to a closed walk. We show each $P \in P_S$ corresponds to a closed walk. Consider the last non-zero value of the permutation P. Without loss of generality, suppose this value is a_j and that it occurs at the i^{th} coordinate of P. Since $P \in P_S$, there must be at least $a_j - 1$ zeros to the right of a_j. Removing a_j and $a_j - 1$ zeros to its right in P yields a valid sequence $P' \in P_{s'}$ where $s' = (a_1, \dots a_j, t_1, \dots t_{j-1}, t_j - 1)$ and $k' = k - a_j$. By the inductive hypothesis, P' corresponds to a closed walk $\mathcal{W}'$ of length $\sum a_i t_i - a_j$. Add a cycle of length a_j in between the $(i-1)^{st}$ and i^{th} steps of $\mathcal{W}'$ to obtain a closed walk $\mathcal{W}$ of length $k = \sum a_i t_i$. We have shown that $\bigcup P_S$ is the set of all cycle permutations for closed walks.

We compute the size of $|P_S|$ by constructing a bijection between permutations in P_S

and a set of subdiagonal generalized Dyck paths. Let $s = \{(a_1, t_1), (a_2, t_2), \dots (a_j, t_j)\} \in S$. Let D the set of subdiagonal generalized Dyck paths on a $(k - \sum t_i) \times (k - \sum t_i)$ grid that are made up of t_i horizontal steps of length $a_i - 1$ and $k - \sum t_i$ vertical steps of length 1. Each non-zero label a_i of P_S corresponds to a horizontal step of length $a_i - 1$ and each zero label of P_S corresponds to a vertical step of length one. Consider the map between permutations and generalized Dyck paths based on this correspondence. The condition that for all $2 \leq i \leq k+1$ $\sum_{\ell \in N(s)} (\ell - 1) \geq z_i$ translates to the generalized Dyck path not crossing the diagonal. Thus, the correspondence is a bijection between P_S and D. Lemma 4.4.7 implies that $|P_S| = |D| = \frac{k!}{(\prod t_i!)(k+1-\sum t_i)!}$. □

Walk and cycle counts in locally regular graphs

We show that the polynomial relating cycles and closed walks given in Definition 4.4.1 governs the relationship between cycles and closed walks in graphs where each vertex is in approximately the same number of cycles.

Definition 4.4.8 (k-locally regular, essentially k-locally regular)**.** Let $C_k(G, v)$ denote the the number of k-cycles at vertex v in G.

1. A graph G is k-locally regular if it is regular and $C_j(G, v) = C_j(G, u)$ for all $u, v \in V(G)$ and $j \leq k$.

2. A sequence of graphs (G_i) with $d_i \to \infty$ and k-sparsity exponent a is essentially k-locally regular if $C_j(G_i, v) - C_j(G_i, u) = o\left(d_i^{j/2}\right)$ for all $u, v \in V(G)$ and $j \leq k$.

Theorem 4.4.9. *Let G be a k-locally regular graph on n vertices with degree d. Let $c_k = C_k(G)/(nd^{k/2})$ and $w_k = W_k(G)/(nd^{k/2}) = W_k(G, 1/2)$ where $W_k(G)$ and $C_k(G)$ denote the number of closed k-walks and simple k-cycles in G respectively. Then*

$$w_k = T((c_3, c_4, \dots c_k)).$$

Proof. We count the number closed walks in G at a vertex v by partitioning the closed walks into sets based on their cycle permutation and computing the size of each partition class. Let $S = \{(a_1, t_1), (a_2, t_2), \ldots, (a_j, t_j)\} \in \mathcal{S}_k$ and $P \in P_S$ as defined in Lemma 4.4.5. Define X_P as the number of walks with cycle permutation P at v in G. Let $t = \sum t_i$ be the number of non-zero values in P and let $N(P, i)$ denote the i^{th} non-zero value of the string P. Let $C_k(G, u)$ denote the number of k-cycles at u. Since G is locally regular $C_k(G, u) = c_k d^{k/2}$ for all k. It follows

$$X_P = \prod_{\ell=1}^{t} C_{N(P,\ell)}(G, u) = \prod_{i=1}^{j} \left(c_{a_i} d^{a_i/2}\right)^{t_i} = d^{k/2} \prod_{i=1}^{j} \left(c_{a_i}\right)^{t_i}. \tag{4.13}$$

Summing over all $P \in P_S$ and all vertices we obtain

$$W_k(G) = nd^{k/2} \sum_{S \in \mathcal{S}} |P_S| \prod_{i=1}^{j} \left(c_{a_i}\right)^{t_i}, \quad \text{equivalently} \quad w_k = \sum_{S \in \mathcal{S}} |P_S| \prod_{i=1}^{j} \left(c_{a_i}\right)^{t_i}.$$

The statement follows directly from Lemma 4.4.5. □

Theorem 4.4.10. *Let (G_r) be a sequence of essentially k-locally regular graphs with n_r vertices and degree $d_r \to \infty$, sparsity exponent $1/2$, and k-limit $(w_3, w_4, \ldots w_k)$. Let $C_j(G_r)$ be the number of j-cycles in G_r. Then $(w_3, w_4, \ldots w_k) = T((c_3, c_4, \ldots c_k))$ where*

$$c_j = \lim_{s \to \infty} \frac{C_j(G_r)}{n_r d_r^{j/2}}.$$

Proof. We follow the proof of Theorem 4.4.9 until line Equation (4.13). Since the G_s is approximately locally regular rather than locally regular, we have the weaker guarantee that $C_k(G_i, u) = \frac{C_k(G_i)}{n_i} + o\left(d_i^{j/2}\right)$. It follows that for $G = G_r$, $n = n_r$, and $d = d_r$,

$$X_P = \prod_{\ell=1}^{t} C_{N(P,\ell)}(G, u) = \prod_{i=1}^{j} \left(\frac{C_{a_i}(G)}{n} + o\left(d^{a_i/2}\right)\right)^{t_i} = d^{k/2} \prod_{i=1}^{j} \left(\frac{C_{a_i}(G)}{nd^{a_i/2}}\right)^{t_i} + o\left(d^{k/2}\right).$$

Summing over all $P \in P_S$ and all vertices we obtain

$$W_k(G_r) = n_r d_r^{k/2} \sum_{S \in \mathcal{S}} |P_S| \prod_{i=1}^{j} \left(\frac{C_{a_i}(G_r)}{n_r d_r^{a_i/2}} \right)^{t_i} + o\left(n_r d_r^{k/2}\right).$$

Therefore

$$w_k = \lim_{r\to\infty} \frac{W_k(G_r)}{n_r d_r^{k/2}} = \lim_{r\to\infty} \sum_{S \in \mathcal{S}} |P_S| \prod_{i=1}^{j} \left(\frac{C_{a_i}(G_r)}{n_r d_r^{a_i/2}} \right)^{t_i} + o_r(1) = \sum_{S \in \mathcal{S}} |P_S| \prod_{i=1}^{j} (c_{a_i})^{t_i},$$

and the statement follows directly from Lemma 4.4.5. □

Limits achieved by ROC families

We prove Theorem 4.4.3, which describes the limits of ROC families. The following lemma gives the expected number of closed walks by permutation type.

Lemma 4.4.11. *Let $S = \{(a_1, t_1), (a_2, t_2), \ldots, (a_j, t_j)\} \in \mathcal{S}_k$ as defined in Lemma 4.4.5, and let $t = \sum_{i=1}^{j} t_i$. Let $X_S(G)$ be the random variable for the number of walks with a permutation type in P_S in $G \sim ROC(n, d, \mu, a)$ where $d = o\left(n^{1/((1-a)k+2a-1)}\right)$. Then for the function c as given in Theorem 4.4.3*

1. *For $a < 1$, $\mathbb{E}(X_S(G)) = |P_S| \left(\prod_{i=1}^{j} c(a_i)^{t_i}\right) nd^{(1-2a)t+ak} + o\left(nd^{(1-2a)t+ak}\right)$.*
2. *For $a = 1$ and $t = 1$, $\mathbb{E}(X_S(G)) = |P_S| \left(\prod_{i=1}^{j} c(a_i)^{t_i}\right) nd^{(1-2a)t+ak} + o\left(nd^{(1-2a)t+ak}\right)$.*
3. *For $a = 1$ and $t > 1$, $\mathbb{E}(X_S(G)) = \Theta\left(nd^{(1-2a)t+ak}\right)$.*

Taking $S = \{(k, 1)\}$ in the above lemma gives the number of simple k-cycles in a ROC graph. The following corollary describes the cycle counts when the community size is a constant independent of d.

Corollary 4.4.12. *Let $G \sim ROC(n, d, \mu, 0)$ Then for $d = o\left(n^{\frac{1}{k-1}}\right)$,*

$$\mathbb{E}(C_k(G)) = c(k)nd + o(nd).$$

Proof. Let $P \in P_S$, and let $X_P(G)$ be the random variable for the number of walks in G with permutation type P. We show that $\mathbb{E}(X_P(G))$ is the same for each $P \in P_S$ and so

$$\mathbb{E}(X_S(G)) = \sum_{P \in P_S} \mathbb{E}(X_P(G)) = |P_S| \, \mathbb{E}(X_P(G)). \tag{4.14}$$

To compute the expectation of $X_P(G)$, we apply linearity of expectation to indicator random variables representing each possible walk. We define a possible walk as (i) an ordered set of vertices $(v_1, \dots v_k)$ such that the walk $v_1, v_2, \dots v_k$ is closed and has cycle permutation P and (ii) an ordered set of communities $(u_1, \dots u_k)$, $u_i \in [xnd^{1-2a}]$. The walk exists if for each $1 \leq i \leq k-1$, the vertices v_i and v_{i+1} are adjacent by an edge that was added in the $(u_i)^{th}$ community in the construction of G. The probability a possible walk exists in G depends on how often the community labels $(u_1, \dots u_k)$ change between adjacent vertices.

Let A be the set of possible walks in which each cycle is assigned a distinct community, each edge is labeled with the community assigned to its cycle, and there are $k - t + 1$ distinct vertices. We write $X_P(G) = A_P(G) + B_P(G)$ where $A_P(G)$ is the random variable for the number of walks in A that appear in G and $B_P(G)$ is the random variable for the number of walks that appear in G and are not in A. We compute $\mathbb{E}(A_P(G))$ and show that $\mathbb{E}(B_P(G)) = o\left(\mathbb{E}(A_P(G))\right)$ in cases (1) and (2) and $\mathbb{E}(B_P(G)) = \Theta\left(\mathbb{E}(A_P(G))\right)$ in case (3).

Claim 1: $\mathbb{E}(A_P(G)) = \left(\prod_{i=1}^{j} c(a_i)^{t_i}\right) nd^{(1-2a)t+ak} + o\left(nd^{(1-2a)t+ak}\right)$.

We write $A_P(G)$ as the sum of random variables $A_W(G)$ that indicate if a walk $W \in A$ is in G. We show that $\mathbb{E}(A_W(G))$ is the same for all W in A and so

$$\mathbb{E}(A_P(G)) = |A| \, \mathbb{E}(A_W(G)) = |A| \, \mathbb{P}(A_W(G)).$$

We now compute $\mathbb{P}(A_W(G))$. Let $z_1, \dots z_t$ be the non-zero characters of P ordered by first appearance. Let A_ℓ be the event that all edges in the cycle corresponding z_ℓ were added in the community assigned to z_ℓ, which we denote y_ℓ. The probability of A_ℓ depends on the community type (m_i, q_i, β_i) of y_ℓ. We compute

$$\mathbb{P}(A_\ell) = \sum_i \mathbb{P}(\text{ specified cycle appears in community } y_\ell \mid \ y_\ell \text{ is type } i)\, \mathbb{P}(\ y_\ell \text{ is type } i).$$

It follows that

$$\mathbb{P}(A_\ell) = \begin{cases} \sum_{i \in B} 2 \left(\frac{m_i d^a}{n}\right)^{z_\ell} q_i^{z_\ell} \mu_i + \sum_{i \in B^c} \left(\frac{m_i d^a}{n}\right)^{z_\ell} q_i^{z_\ell} \mu_i & z_\ell \geq 3 \\ \\ \sum_{i \in B} 2 \left(\frac{m_i d^a}{n}\right)^{2} q_i \mu_i + \sum_{i \in B^c} \left(\frac{m_i d^a}{n}\right)^{2} q_i \mu_i & z_\ell = 2. \end{cases}$$

Equivalently, $\mathbb{P}(A_\ell) = d^{az_\ell} c(z_\ell)/(xn^{z_\ell})$. The event that the walk W appears in G is the intersection of the events $A_1, \dots, A_t$. Note that these events are independent because the communities $y_1, y_2, \dots y_t$ are distinct. It follows

$$\mathbb{P}(A_W(G)) = \prod_{\ell=1}^{t} \mathbb{P}(A_\ell) = \frac{d^{ak} \prod_{i=1}^{j} c(a_i)^{t_i}}{n^k x^t}.$$

Next we compute the size of A. There are $\frac{(xnd^{1-2a})!}{(xnd^{1-2a}-t)!} = (xnd^{1-2a})^t + o\left((xnd^{1-2a})^t\right)$ ways to select t distinct communities and $\frac{n!}{n-(k-t+1)!} = n^{k-t+1} + o\left(n^{k-t+1}\right)$ ways to select the the vertices for $W \in A$. The claim follows,

$$\begin{aligned} \mathbb{E}(A_P(G)) &= \left((xnd^{1-2a})^t + o\left((xnd^{1-2a})^t\right)\right)\left(n^{k-t+1} + o\left(n^{k-t+1}\right)\right)\left(\frac{d^{ak} \prod_{i=1}^{j} c(a_i)^{t_i}}{n^k x^t}\right) \\ &= nd^{(1-2a)t+ak}\left(\prod_{i=1}^{j} c(a_i)^{t_i}\right) + o\left(nd^{(1-2a)t+ak}\right). \end{aligned} \tag{4.15}$$

Claim 2: In cases (1) and (2), $\mathbb{E}(B_P(G)) = o\left(nd^{(1-2a)t+ak}\right)$, and in case (3) $\mathbb{E}(B_P(G)) = \Theta\left(nd^{(1-2a)t+ak}\right)$.

Before computing $\mathbb{E}(B_P(G))$, we introduce notation to describe the features of possible walks that are not in A. Let $z_1, \ldots, z_t$ be the non-zero characters of P, so the walk is composed of cycles of lengths $z_1, \ldots z_t$. Let m_ℓ be the number of vertices in the cycle corresponding z_ℓ that do not appear in the cycles corresponding to $z_1, \ldots z_{\ell-1}$. Let λ_i be the number of community edge labels in the cycle corresponding to z_ℓ that do not appear as community edge labels in any cycle corresponding to $z_1, \ldots z_{\ell-1}$. For each community assignment u_i, v_i and v_{i+1} must both be in community u_i if the possible walk exists in G. We say the i^{th} edge "assigns" the community u_i to the vertices v_i and v_{i+1}; if two consecutive edges have the same community label, then the common end is assigned to the same community twice. Let Γ_ℓ be the number of vertex-community assignments from the cycle corresponding to z_ℓ that are not assigned in cycles corresponding to $z_1, \ldots z_{\ell-1}$. Let $m = \sum_i m_i \leq k - t + 1$, $\lambda = \sum_i \lambda \leq k$, $\Gamma = \sum \Gamma_i$, and j be the number of indices i such that $\lambda_i \geq 2$. Let $\mathcal{P}$ denote the parameters $\{\lambda_i, m_i, \Gamma_i\}$, and let $B_{\mathcal{P}}(G)$ be the number of possible walks with the parameters $\mathcal{P}$. There are $\Theta\left(\left(nd^{1-2a}\right)^\lambda\right)$ ways to select the communities, $\Theta\left(n^m\right)$ ways to select the vertices. The probability a vertex is an assigned community is $\Theta\left(\frac{d^a}{n}\right)$. It follows that

$$\mathbb{E}(W_{\mathcal{P}}(G)) = \Theta\left(\left(nd^{1-2a}\right)^\lambda n^m \left(\frac{d^a}{n}\right)^\Gamma\right). \tag{4.16}$$

Next we show that for any set of parameters $\mathcal{P}$, $\Gamma \geq m + \lambda + j - 1$. First we describe relationships between λ_ℓ, Γ_ℓ, z_ℓ, and m_ℓ in different settings.

1. If there are precisely λ_ℓ community labels in the cycle corresponding to z_ℓ then

$$\Gamma_\ell \geq z_\ell + \lambda_\ell \qquad \lambda_\ell \geq 2$$

$$\Gamma_\ell = z_\ell \qquad \lambda_\ell = 1.$$

If there are $\lambda_\ell \geq 2$ communities assigned to edges in the cycle, then there are at least λ_ℓ vertices where the adjacent edges are assigned different communities. These vertices contribute $2\lambda_\ell$ vertex-community assignments and the remaining $z_\ell - \lambda_\ell$ vertices are also assigned a community. If $\lambda_\ell = 1$, then each vertex is assigned to the one community.

2. If there are more than λ_ℓ community labels in the cycle corresponding to z_ℓ then

$$\Gamma_\ell \geq m_\ell + \lambda_\ell + 1 \qquad \lambda_\ell \geq 1$$
$$\Gamma_\ell \geq m_\ell \qquad \lambda_\ell = 0.$$

The m_i new vertices must be assigned at least one community. When $\lambda_\ell \neq 0$, there must be at least $\lambda_\ell + 1$ vertices in which (i) both adjacent edges are labeled with two different first appearing communities or (ii) one adjacent edge is labeled with a first appearing community and one adjacent edge is labeled with a community that has already appeared. If such a vertex is a new vertex then this vertex has a total of two community assignments. If such a vertex has appeared before, it has not been previously assigned to a first appearing community, so this contributes one community assignment.

Since the first vertex of a cycle corresponding to z_ℓ for $\ell \geq 2$ has already been visited, $z_\ell \geq m_\ell + 1$ for $\ell \geq 2$. Therefore when $\ell \geq 2$

$$\Gamma_\ell \geq m_\ell + \lambda_\ell + 1 \qquad \lambda_\ell \geq 2$$
$$\Gamma_\ell = m_\ell + \lambda_\ell \qquad \lambda_\ell \leq 1,$$

and for $\ell = 1$,

$$\Gamma_1 \geq m_1 + \lambda_1 \qquad \lambda_1 \geq 2$$

$$\Gamma_1 = m_1 \qquad\qquad \lambda_1 = 1.$$

Summing the above inequalities over ℓ yields the observation that $\Gamma \geq m + \lambda + j - 1$. Note also that $m \leq k - t + 1$. Equation Equation (4.16) becomes

$$\mathbb{E}(W_{\mathcal{P}}(G)) = \Theta\left(\left(nd^{1-2a}\right)^{\lambda} n^m \left(\frac{d^a}{n}\right)^{m+\lambda+j-1}\right) = \Theta\left(n^{1-j} d^{(1-a)\lambda + a(k-t+j)}\right). \tag{4.17}$$

Next consider a walk that is not in A. There must either be (i) a cycle that has at least two new community labels and so $j \geq 1$ or (ii) fewer than t total community labels, so $\lambda \leq t - 1$. If $t = 1$ (the walk is a simple cycle), case (ii) does not occur because there must be at least one community label. In case (i), $j \geq 1$, $\lambda \leq k - t + j$ and $t \geq 1$, and so Equation Equation (4.17) becomes

$$\mathbb{E}(W_{\mathcal{P}}(G)) = \Theta\left(n^{1-j} d^{k-t+j}\right) = \begin{cases} \Theta\left(d^{k-t}\right) & j = 1 \\ o\left(d^{k-t}\right) & j \geq 2. \end{cases} \tag{4.18}$$

Equivalently if $\mathcal{P}$ is type (i), then

$$\mathbb{E}(W_{\mathcal{P}}(G)) = nd^{(1-2a)t+ak} O\left(n^{-1} d^{(1-a)k+2a-1}\right) = o\left(nd^{(1-2a)t+ak}\right).$$

In case (ii), then $\lambda \leq t - 1$ and $j \geq 0$, and so Equation Equation (4.17) becomes

$$\mathbb{E}(W_{\mathcal{P}}(G)) = \Theta\left(n^{1-j} d^{(1-a)\lambda + a(k-t+j)}\right) = \begin{cases} \Theta\left(nd^{(1-2a)t+ak+a-1}\right) & \lambda = t-1 \text{ and } j = 0 \\ o\left(nd^{(1-2a)t+ak+a-1}\right) & \lambda \leq t-2 \text{ or } j \geq 1. \end{cases} \tag{4.19}$$

Equivalently if $\mathcal{P}$ is type (ii), then

$$\mathbb{E}(W_{\mathcal{P}}(G)) = O\left(n^{1-j}d^{(1-a)(t-1)+a(k-t+j)}\right) = nd^{(1-2a)t+ak}O\left(d^{a-1}\right).$$

Therefore for any $\mathcal{P}$ that is not in A,

$$\mathbb{E}(W_{\mathcal{P}}(G)) = \begin{cases} o\left(nd^{(1-2a)t+ak}\right) & a < 1 \text{ or } a = 1 \text{ and } t = 1 \\ \Theta\left(nd^{(1-2a)t+ak}\right) & a = 1 \text{ and } t > 1. \end{cases}$$

Note $\mathbb{E}(B_P(G)) = \sum_{\mathcal{P}} \mathbb{E}(W_{\mathcal{P}}(G))$. Since the number of sets of valid parameters $\mathcal{P}$ is constant, claim 2 follows.

The computation of $\mathbb{E}(X_P(G)) = \mathbb{E}(A_P(G)) + \mathbb{E}(B_P(G))$ did not rely on any information about P besides that $P \in P_S$. Therefore, equation *Equation* (4.14) holds and the statement of the lemma follows directly from claims 1 and 2. □

Proof. (of Theorem 4.4.3) For $a < 1$, Lemma 4.4.11 implies

$$\mathbb{E}(W_k(G)) = \sum_{S \in \mathcal{S}} |P_s| \left(\prod_{i=1}^{j} c(a_i)^{t_i}\right) nd^{(1-2a)t+ak} + o\left(nd^{(1-2a)t+ak}\right) \tag{4.20}$$

We now collect the highest order terms of Equation (4.20) for different values of a. Recall $\sum a_i t_i = k$ and $a_i \geq 2$, so $1 \leq \sum t_i \leq \frac{k}{2}$.

Case 1: $a \in (0, 1/2)$**.** The highest order term of Equation (4.20) is from $S \in \mathcal{S}$ with $a_1 = 2$ and $t_1 = \frac{k}{2}$ for k even and $S \in \mathcal{S}$ with $a_1 = 3$, $a_2 = 2$, $t_1 = 1$, $t_2 = \frac{k-3}{2}$ for k odd. For even k, Equation Equation (4.20) becomes

$$\mathbb{E}(W_k(G)) = nd^{k/2}\frac{k!}{\left(\frac{k}{2}\right)!\left(\frac{k}{2}+1\right)!}c(2)^{k/2} + o\left(nd^{k/2}\right) = (Cat_{k/2})nd^{k/2} + o(nd^{k/2}).$$

For odd k, Equation Equation (4.20) becomes

$$\mathbb{E}(W_k(G)) = O\left(nd^{ak+(1-2a)\frac{k-1}{2}}\right) = O\left(nd^{\frac{k-1}{2}+a}\right) = o(nd^{k/2}).$$

It follows that the ROC family (μ, a) achieves the Catalan vector with sparsity exponent $1/2$, and achieves the $k-2$ length prefix with k-sparsity exponent $1/2$ for all $k \geq 4$.

Case 2: $a = 1/2$**.** Each $S \in \mathcal{S}$ contributes a term of order $d^{k/2}$ to Equation Equation (4.20). Therefore

$$\mathbb{E}(W_k(G)) = \left(\sum_{S \in \mathcal{S}_k} \frac{k!}{(\prod t_i!)(k+1-\sum t_i)!} \prod_i c(a_i)^{t_i}\right) nd^{k/2} + o\left(nd^{k/2}\right) = w(k)nd^{k/2} + o\left(nd^{k/2}\right).$$

It follows that the ROC family (μ, a) achieves the limit $(w_3, w_4, \dots)$ with sparsity exponent $1/2$, and achieves the $k-2$ length prefix with k-sparsity exponent $1/2$ for all $k \geq 3$.

Case 3: $a \in (1/2, 1)$**.** The highest order term of Equation (4.20) is from $S \in \mathcal{S}$ with $a_1 = k$ and $t_1 = 1$. Therefore Equation Equation (4.20) becomes

$$\mathbb{E}(W_k(G)) = c(k)nd^{1+a(k-2)} + o\left(nd^{1+a(k-2)}\right).$$

It follows that the ROC family (μ, a) achieves the limit $(c_3, c_4, \dots)$ with sparsity exponent a, and achieves the $k-2$ length prefix with k-sparsity exponent a for all $k \geq 3$.

Case 4: $a = 1$**.** For $S \in \mathcal{S}$ with $t = \sum t_i$, the number of walks with permutation type in the set P_S is $\Theta\left(nd^{(1-2a)t+ak}\right)$. Therefore the walks contributing the highest order terms correspond to $S \in \mathcal{S}$ with $a_1 = k$ and $t_1 = 1$. By parts 2 and 3 of Lemma 4.4.11, we have

$$\mathbb{E}(W_k(G)) = c(k)nd^{k-1} + o\left(nd^{k-1}\right).$$

It follows that the ROC family (μ, a) achieves the limit $(c_3, c_4, \dots)$ with sparsity exponent 1,

and achieves the $k-2$ length prefix with k-sparsity exponent 1 for all $k \geq 3$. $\square$

The convergence of sequences of ROC graphs

Definition 4.1.7 states that the vector achieved by a ROC family is the expected walk count of a ROC graph from that family normalized with respect to *expected* degree. We now justify this definition by showing that for $G \sim ROC(n, d, \mu, a)$, the probability that the normalized closed walk count $W_j(G, \alpha)$ deviates from the limit w_j achieved by the family tends to zero as d grows (Theorem 4.4.13). Moreover, we show that the sequence $G_i \sim ROC(n_i, d_i, \mu, a)$ almost surely converges to the limit achieved by the family when n_i and d_i grow sufficiently fast (Corollary 4.4.14).

Theorem 4.4.13. *Let $(w_3, w_4, \dots)$ be the limit achieved by the ROC family (μ, a). Let $G \sim ROC(n, d, \mu, a)$ where $d = o\left(n^{1/((1-a)k+2a-1)}\right)$ and $|w_j - \frac{\mathbb{E}(W_j(G))}{nd^{1+\alpha(j-2)}}| < \varepsilon/2$. Then for $\alpha = \max\{a, 1/2\}$,*

$$\mathbb{P}(|W_j(G, \alpha) - w_j| > \varepsilon) = f(d, a)$$

where

$$f(d,a) = \begin{cases} O\left(d^{-1+2a} + \frac{d^{k/2-1}}{n} + \frac{d^{(k-1)(2a-1)}}{n}\right) & a < 1/2 \\ O\left(\frac{d^{k/2-1}}{n} + d^{-1/2}\right) & a = 1/2 \\ O\left(d^{1-2a} + \frac{d^{(1-a)(k-2)}}{n}\right) & 1/2 < a < 1 \\ O\left(d^{-1} + \frac{d}{n}\right) & a = 1. \end{cases}$$

Corollary 4.4.14. *Let $(w_3, w_4, \dots)$ be the limit achieved by the ROC family (μ, a). Let $G_i \sim ROC(n_i, d_i, \mu, a)$ where $d_i = o\left(n_i^{1/((1-a)k+1-2a)}\right)$ and $f(d_i, a)$ is defined for G_i as in Theorem 4.4.13. If $\sum_{i=1}^{\infty} f(d_i, a) < \infty$ the sequence of graphs (G_i) converges to the limit $(w_3, w_4, \dots)$ with sparsity exponent $\alpha = \max\{a, 1/2\}$.*

Achieving normalized and unnormalized closed walk counts. A ROC family (μ, a) that achieves the limit of a sequence of graphs (G_i) with the appropriate sparsity exponent is a sampleable model that produces graphs in which the normalized closed walk counts match the limit up to an error term that tends to zero as the size of the sampled graph grows. The following remark describes when a sequence of graphs drawn from the ROC model also matches the unnormalized closed walk counts of the sequence (G_i) term by term. The remark is stated for k-convergence and k-limits, but an analogous statement holds for full convergence and limits.

Remark 4.4.15. Let (G_i) be a sequence of graphs each with n_i vertices and average degree d_i such that (G_i) is k-convergent with k-limit L and k-sparsity exponent α. Suppose the ROC family (μ, a) achieves the limit L with sparsity exponent α.

1. If $d_i = o\left(n_i^{1/((1-a)k+2a-1)}\right)$ then the sequence (H_i) with $H_i \sim ROC(n_i, d_i, \mu, a)$ has the property that for sufficiently large i and $j \leq k$, in expectation G_i and H_i have the same average degree and number of closed j-walks up to lower order terms with respect to d_i.

2. It is possible to construct other sequences (H_i) with $H_i \sim ROC(n_i, f(n_i), \mu, a)$ such that for sufficiently large i, in expectation H_i and G_i have different edge densities, but have the same normalized number of closed walks up to lower order terms.

To prove Theorem 4.4.13, we will apply Lemma 4.3.5 , which bounds the probability that the normalized walk count deviates from expectation in terms of the probability that the number of edges deviates and the probability that the walk count deviates. Lemmas 4.4.16 and 4.4.17 compute these quantities.

Lemma 4.4.16. *Let $G \sim ROC(n, d, \mu, a)$. Let $D(G)$ be the random variable for the average degree of G. Then*

$$\mathbb{E}(D(G)) = d \quad \text{and} \quad \mathsf{Var}(D(G)) = \Theta\left(\frac{d^{1+a}}{n}\right).$$

Proof. Let $D(G) = \frac{1}{n}\sum_{v,w,u} X_{u,v,w}$ where $X_{u,v,w}$ is an indicator random variable for the event that the edge w, v is added in the u^{th} community. Note

$$\mathbb{E}(X_{u,v,w}) = \mathbb{P}(X_{u,v,w}) = \sum_{i\in B^c} \mu_i \left(\frac{m_i d^a}{n}\right)^2 q_i + \sum_{i\in B} \mu_i 2 \left(\frac{m_i d^a}{n}\right)^2 q_i = \frac{d^{2a}}{xn^2}.$$

There are $n(n-1)$ pairs w, v and xnd^{1-2a} communities u. Therefore

$$\mathbb{E}(D(G)) = \frac{1}{n}\sum_{v,w,u} \mathbb{E}(X_{u,v,w}) = \frac{1}{n} n(n-1)xnd^{1-2a}\frac{d^{2a}}{xn^2} = d - \frac{d}{n}.$$

Next we compute the expected pairs of edges, $\mathbb{E}(D(G)^2)$. We fix the potential edge defined by vertices a and b and community x and sum over all other potential edges.

$$\begin{aligned}
\mathbb{E}\big(D(G)^2\big) &= \left(\frac{1}{n^2}\right) n(n-1)(xnd^{1-2a}) \sum_{u,v,w} \mathbb{P}(X_{u,v,w} \text{ and } X_{a,b,x}) \\
&= x(n-1)d^{1-2a}\Bigg(\sum_{u,v,w\neq x} \mathbb{P}(X_{u,v,w})\,\mathbb{P}(X_{a,b,x}) + \sum_{u,v\notin\{a,b\},x} \mathbb{P}(X_{u,v,w})\,\mathbb{P}(X_{a,b,x}) \\
&\quad + 2\sum_{u=a,v\neq b,x} \mathbb{P}(X_{u,v,w} \text{ and } X_{a,b,x}) + \mathbb{P}(X_{a,b,x})\Bigg) \\
&= x(n-1)d^{1-2a}\left(\frac{d^{2a}}{xn^2}\right)\left(n(n-1)\left(xnd^{1-2a}-1\right)\left(\frac{d^{2a}}{xn^2}\right) + (n-2)(n-3)\left(\frac{d^{2a}}{xn^2}\right)\right. \\
&\quad \left. + (n-3)\Theta\left(\frac{d^a}{n}\right) + 1\right) \\
&= \left(d - \frac{d}{n}\right)^2 + \Theta\left(\frac{d^{1+a}}{n}\right).
\end{aligned}$$

It follows that

$$\mathsf{Var}(D(G)) = \mathbb{E}\big(D(G)^2\big) - \mathbb{E}(D(G))^2 = \Theta\left(\frac{d^{1+a}}{n}\right).$$

□

Lemma 4.4.17. *Let $G \sim ROC(n, d, \mu, a)$ with $d = o\left(n^{1/((1-a)k+2a-1)}\right)$. Let $W_k(G)$ be the*

random variable for the number of closed k-walks in G. Then

$$\mathsf{Var}(W_k(G)) = \begin{cases} O\left(\left(nd^{k/2}\right)^2\left(d^{-1+2a} + \frac{d^{k/2-1}}{n} + \frac{d^{(k-1)(2a-1)}}{n}\right)\right) & a < 1/2 \\ O\left(\left(nd^{k/2}\right)^2\left(\frac{d^{k/2-1}}{n} + d^{-1/2}\right)\right) & a = 1/2 \\ O\left(\left(nd^{1+a(k-2)}\right)^2\left(d^{1-2a} + \frac{d^{(1-a)(k-2)}}{n}\right)\right) & 1/2 < a < 1 \\ O\left(\left(nd^{k-1}\right)^2\left(d^{-1} + \frac{d}{n}\right)\right) & a = 1. \end{cases}$$

Proof. We give an upper bound on $\mathbb{E}(W_k(G)^2)$ by counting the expected number of pairs of walks. Let $P'_k(G)$ be the random variable for the number of pairs of k-walks in G that do not intersect, and let $P''_k(G)$ be the random variable for the number of pairs of k-walks in G that do intersect. Note that two k-walks that intersect can be thought of as a $2k$ walk. The expected number of $2k$ walks in G is $\Theta\left(nd^{1+a(2k-2)}\right)$ (see Theorem 4.4.3), and so $\mathbb{E}(P''_k(G)) = \Theta\left(nd^{1+a(2k-2)}\right)$.

To compute $P'_k(G)$ we recall the partition of possible walks with permutation type P_S into sets $A, B(i)$ and $B(ii)$ as described in the proof of Lemma 4.4.11. Let $S(t)$ be the set of $S \in \mathcal{S}_k$ such that $\sum t_i = t$. For $S \in S(t)$, the expected number of walks with type A is at most $nd^{(1-2a)t+ak}\left(\prod_{i=1}^{j} c(a_i)^{t_i}\right)$ (see Equation (4.15)). The expected number of walks with type $B(i)$ is $\Theta\left(d^{k-t}\right)$ (see Equation (4.18)), and the expected number of walks with type $B(ii)$ is $\Theta\left(nd^{(1-2a)t+ak+a-1}\right)$ when $t \neq 1$ and 0 when $t = 1$ (see Equation (4.19)). Therefore

$$\mathbb{E}(P'_k(G)) \leq \left(\sum_{t=1}^{\lfloor k/2 \rfloor} \sum_{S \in S(t)} |P_S| nd^{(1-2a)t+ak}\left(\prod_{i=1}^{j} c(a_i)^{t_i}\right) + \Theta\left(d^{k-t} + \zeta_t nd^{(1-2a)t+ak+a-1}\right)\right)^2$$

where $\zeta_t = 0$ if $t = 1$ and $\zeta_t = 1$ otherwise. We simplify and obtain

$$\mathbb{E}(P_k'(G)) = \begin{cases} \mathbb{E}(W_k(G))^2 + O\left(nd^{k/2}\left(nd^{k/2-1+2a} + d^{k-1} + nd^{k/2+a-1}\right)\right) & a < 1/2 \\ \mathbb{E}(W_k(G))^2 + O\left(nd^{k/2}\left(d^{k-1} + nd^{ak+a-1}\right)\right) & a = 1/2 \\ \mathbb{E}(W_k(G))^2 + O\left(nd^{1+a(k-2)}\left(nd^{2+a(k-4)} + d^{k-1} + nd^{1+a(k-3)}\right)\right) & 1/2 < a < 1 \\ \mathbb{E}(W_k(G))^2 + O\left(nd^{k-1}\left(nd^{k-2} + d^{k-1} + nd^{k-2}\right)\right) & a = 1. \end{cases}$$

Finally we compute

$$\mathsf{Var}(W_k(G)) = \mathbb{E}\left(W_k(G)^2\right) + \mathbb{E}(W_k(G))^2 = \mathbb{E}(P_k'(G)) + \mathbb{E}(P_k''(G)) - \mathbb{E}(W_k(G))^2$$

$$= \begin{cases} O\left(nd^{k/2}\left(nd^{k/2-1+2a} + d^{k-1} + nd^{k/2+a-1}\right)\right) + O\left(nd^{1+a(2k-2)}\right) & a < 1/2 \\ O\left(nd^{k/2}\left(d^{k-1} + nd^{ak+a-1}\right)\right) + O\left(nd^{k}\right) & a = 1/2 \\ O\left(nd^{1+a(k-2)}\left(nd^{2+a(k-4)} + d^{k-1} + nd^{1+a(k-3)}\right)\right) + O\left(nd^{1+a(2k-2)}\right) & 1/2 < a < 1 \\ O\left(nd^{k-1}\left(nd^{k-2} + d^{k-1} + nd^{k-2}\right)\right) + O\left(nd^{2k-1}\right) & a = 1, \end{cases}$$

and the statement follows by simplifying the above expressions. □

We now prove Theorem 4.4.13 by applying Lemma 4.3.5.

Proof. (of Theorem 4.4.13) Let $g = w_j - \frac{\mathbb{E}(W_j(G))}{nd^{1+\alpha(j-2)}}$, $\delta = \min\left\{\frac{\varepsilon}{2j(w_j+\varepsilon)}, \frac{1}{2(j-1)^2}\right\}$ and $\lambda = \varepsilon/2 - |g|$. By Lemmas 4.3.5, 4.4.16 and 4.4.17,

$$\mathbb{P}(|W_j(G,\alpha) - w_j| > \varepsilon) \leq \frac{\mathsf{Var}(D(G))}{\delta^2 d^2} + \frac{\mathsf{Var}(W_j(G))}{\lambda^2\left(nd^{1+\alpha(j-2)}\right)^2} \leq O\left(\frac{d^{a-1}}{n}\right) + f(d,a) = f(d,a).$$

□

Corollary 4.4.14 follows directly from Theorem 4.4.13 and part 3 of Lemma 4.3.4.

4.4.3 Conditions for ROC achievable limits

In this section we address the questions: for which vectors L does there exist a ROC family that achieves limit or k-limit L with sparsity exponent α? We first show that all 4-limits are achievable in Section 4.4.3, then describe necessary and sufficient conditions for a limit vector (of any length) to be achievable in Section 4.4.3. Finally Lemma 4.4.25 in Section 4.4.3 gives a convenient criterion for determining when the Stietljes condition is satsified.

Achievability of (w_3, w_4)

In this section we prove Theorem 4.1.13, which states that any (w_3, w_4) that is a limit of a sequence of graphs with increasing degree can be achieved by a ROC family. In fact, the requirement that degree increases is only needed for the case in which the 4-sparsity exponent is $1/2$.

Lemma 4.4.18. *Let $C_j(G)$ and $W_j(G)$ denote the number of simple j-cycles and closed j-walks of a graph G respectively. For any graph G on n vertices with average d*

$$W_4(G) \geq \frac{W_3(G)^2}{nd} \quad \text{and} \quad C_4(G) \geq C_3(G)\left(\frac{C_3(G)}{nd} - 1\right).$$

Proof. For each directed edge $e = (u, v)$ let t_e be the number of walks that traverse a triangle with first edge (u, v). For each edge (u, v) we can construct t_e^2 four walks including $\binom{t_e}{2}$ four cycles as follows. Select two triangles (u, v, a) and (u, v, b). The closed walk (u, b, v, a) is a closed four walk. When $a \neq b$ the walk is a four cycle. Note $C_3(G) = W_3(G)$. It follows that

$$W_4(G) \geq \sum_{e \in E(G)} t_e^2 \geq nd\left(\frac{W_3(G)}{nd}\right)^2 \quad \text{and} \quad C_4(G) \geq \sum_{e \in E(G)} \binom{t_e}{2} \geq nd\left(\frac{C_3(G)}{nd}\right)\left(\frac{C_3(G)}{nd} - 1\right).$$

□

Note the first part of the lemma also follows from the observation that for any set of λ_i,

$$\left(\sum_i \lambda_i^2\right)\left(\sum_i \lambda_i^4\right) \geq \left(\sum_i \lambda_i^3\right)^2.$$

Using these properties, we now prove Theorem 4.1.13.

Proof. (of Theorem 4.1.13) Let α be the 4-sparsity exponent of the sequence.

Case 1: $\alpha > 1/2$. By Lemma 4.4.18 for each graph G_i in the sequence satisfies

$$W_4(G_i, \alpha) \geq W_3(G_i, \alpha)^2.$$

It follows that $w_4 \geq w_3$. If $w_3 \neq 0$, the ROC family (μ, a) where μ is the distribution with support one on $m = w_4^2/w_3^3$ and $q = w_3^2/w_4$ achieves the limit (w_3, w_4). If $w_3 = 0$, the bROC family (μ, a) where μ is the distribution with support one on $m = w_4$ and $q = 1$ achieves the limit (w_3, w_4).

Case 2: $\alpha = 1/2$. It suffices to show that the cycle counts $T((w_3, w_4)) = (w_3, w_4 - 2)$ are the moments of some distribution. By Lemma 4.4.18, $C_4(G) \geq C_3(G)\left(\frac{C_3(G)}{nd} - 1\right) = W_3(G)\left(\frac{W_3(G)}{nd} - 1\right)$. Let $T_4(G)$ be the number of closed four walks that trace a path of length two. The number of two paths is $\sum_v \binom{deg(v)}{2} \geq n\binom{d}{2}$, and each two path contributes four closed four walks. Each edge contributes two closed four walks. It follows that

$$W_4(G) = C_4(G) + 4T_4(G) + nd \geq W_3(G)\left(\frac{W_3(G)}{nd} - 1\right) + 2nd(d-1) + nd,$$

and so

$$W_4(G_i, 1/2) \geq W_3(G_i, 1/2)^2 + 2 + O\left(\frac{1}{\sqrt{d}}\right).$$

Therefore $w_4 \geq w_3^2 + 2$. If $w_3 \neq 0$, the ROC family (μ, a) where μ is the distribution with support one on $m = (w_4 - 2)^2/w_3^3$ and $q = w_3^2/(w_4 - 2)$ achieves the limit (w_3, w_4). If $w_3 = 0$,

the bROC family (μ, a) where μ is the distribution with support one on $m = w_4 - 2$ and $q = 1$ achieves the limit (w_3, w_4). □

Achievability of limits of general sequences

In this section we prove Theorems 4.1.11 and 4.1.12, which characterize achievable k-limits for sparsity exponent greater than half and half respectively. Additionally, we prove the analogous characterization for full achievability, as stated in the following theorems.

Theorem 4.4.19 (full achievability with sparsity exponent $> 1/2$)**.** *A limit vector* $(w_3, w_4, \dots)$ *is achievable by ROC with sparsity exponent greater than* $1/2$ *if and only if there exists* $\gamma \in [0, 1]$, $s_0, s_1, \dots, t_0, t_2, \dots \in \mathbb{R}^+$, $s_2, t_2 \leq 1$ *such that* $(s_0, s_1, s_2, \dots)$ *and* $(t_0, t_2, \dots)$ *satisfy the full Stieltjes condition and for all* $j \geq 3$

$$w_j = \begin{cases} \gamma s_j & j \text{ odd} \\ \gamma s_j + (1-\gamma) t_j & j \text{ even.} \end{cases}$$

Theorem 4.4.20 (full achievability with sparsity exponent $1/2$)**.** *Let* $T((c_3, c_4, \dots c_k)) = (w_3, w_4, \dots w_k)$ *be the transformation of a vector given in Definition 4.4.1. The limit vector* $(w_3, w_4, \dots w_k)$ *is achievable by ROC with sparsity exponent* $1/2$ *if and only if there exists* $\gamma \in [0, 1]$, $s_0, s_1, \dots, t_0, t_2, \dots \in \mathbb{R}^+$, $s_2, t_2 \leq 1$ *such that* $(s_0, s_1, s_2, \dots)$ *and* $(t_0, t_2, \dots)$ *satisfy the full Stieltjes condition and for all* $j \geq 3$

$$c_j = \begin{cases} \gamma s_j & j \text{ odd} \\ \gamma s_j + (1-\gamma) t_j & j \text{ even.} \end{cases}$$

The question underlying achievability is how to determine when a vector is the vector of normalized cycle counts of some ROC family. Note that the normalized cycle counts $(c(3), c(4), \dots c(k))$ of the family $ROC(n, d, \mu, a)$ are the moments of a discrete probability

distribution over values determined by m_i, q_i and β_i scaled by x. The question of whether a vector can be realized as the vector of normalized cycle counts for some ROC family is a slight variant of the Stieltjes moment problem, which gives necessary and sufficient conditions for a sequence to be the moment sequence of some distribution with positive support.

Our question differs in two key ways. First, the second moment is not directly specified; instead we obtain an upper bound on the second moment from the restriction that

$$x\left(\sum_{i\in B^c}\mu_i m_i^2 p_i + 2\sum_{i\in B}\mu_i m_i^2 p_i\right) = 1.$$

Second, for achievability of k-limits we are interested in when a vector is the prefix of some moment sequence.

The proofs of Theorems 4.4.19 and 4.4.20 rely on the classical solution to the Stieltjes moment problem (Lemma 4.4.21), and the proofs of Theorems 4.1.11 and 4.1.12 use a variant for truncated moment vectors (Lemma 4.4.22). We use these lemmas to show Lemma 4.4.23 and Lemma 4.4.24 which together with Theorem 4.4.3 directly imply the necessary and sufficient conditions given in Theorems 4.1.11, 4.1.12, 4.4.19 and 4.4.20. Finally we prove Lemma 4.4.25 which gives a sufficient local condition to guarantee that a sequence can be extended to satisfy the Stieltjes condition. The proof of this lemma establishes the semi-definiteness of Hankel matrices of sequences satisfying a logconcavity condition.

Lemma 4.4.21 (Stieltjes moment problem, see [ST43]). *A sequence $\mu = (\mu_0, \mu_1, \mu_2, \dots)$ is the moment sequence of a distribution with finite positive support if there exists $\{(x_i, t_i)\}$ with $x_i, t_i > 0$ such that $\sum_{i=1}^{k} x_i t_i^{\ell} = \mu_{\ell}$. A sequence μ is a moment sequence with positive support of size k if and only if*

$$\Delta_i^{(0)} > 0 \text{ and } \Delta_i^{(1)} > 0 \text{ for all } i < k \quad \text{and} \quad \Delta_i^{(0)} = \Delta_i^{(1)} = 0 \text{ for all } i \geq k$$

where Δ is given in Definition 4.1.10.

The next lemma follows, we give a proof for convenience.

Lemma 4.4.22 (truncated Stieltjes moment problem)**.** *A vector $\mu = (\mu_0, \mu_1, \mu_2, \ldots, \mu_n)$ is the truncated moment sequence of a distribution with finite positive support if there exists $\{(x_i, t_i)\}$ with $x_i, t_i > 0$ such that $\sum_{i=1}^{k} x_i t_i^{\ell} = \mu_\ell$ for all $\ell \leq n$. A vector μ is a truncated moment sequence with finite positive support if and only if the Stietljes condition given in Definition 4.1.10 is satisfied.*

Proof. Lemma 4.4.21 directly implies that a truncated moment vector μ satisfies the Stieltjes condition. To show that a vector μ satisfying the Stieltjes condition is a truncated moment vector with finite positive support, it suffices to construct $\mu_{n+1}, \mu_{n+2}, \ldots$ such that the complete vector satisfies the hypotheses of Lemma 4.4.21. Given μ_{2s}, take μ_{2s+1} to be a value that makes $\Delta_s^{(1)} = 0$. Given μ_{2s+1}, take μ_{2s+2} to be the value that makes $\Delta_{s+1}^{(0)} = 0$. □

Lemma 4.4.23. *There exists s_0, s_1, s_2 with $s_2 \leq 1$ such that $(s_0, s_1, s_2, a_3, \ldots a_n)$ satisfies the Stieltjes condition if and only if there exists x_i, m_i, q_i with $x_i, m_i > 0$ and $0 \leq q_i \leq 1$ satisfying*

1. $\sum x_i m_i^2 q_i = 1$
2. $\sum x_i (m_i q_i)^j = a_j \quad$ *for all* $3 \leq j \leq n$.

Similarly, there exists s_0, s_1, s_2 with $s_2 \leq 1$ such that $(s_0, s_1, s_2, a_3, \ldots)$ satisfies the full Stieltjes condition if and only if there exists x_i, m_i, q_i with $x_i, m_i > 0$ and $0 \leq q_i \leq 1$ satisfying (1) and (2) for all $j \geq 3$.

Proof. First assume $(s_0, s_1, s_2, a_3, \ldots a_n)$ satisfies the Stieltjes condition (or $(s_0, s_1, s_2, a_3, \ldots)$ satisfies the full Stieltjes condition) and $s_2 \leq 1$. By Lemma 4.4.22 (or Lemma 4.4.21) there

exists a discrete distribution on $(t_1, t_2, \dots t_k)$ where t_i has mass x_i, $t_i > 0$, and

$$\sum x_i t_i^j = \begin{cases} a_j & 3 \le j \le n \text{ (or } j \ge 3) \\ s_i & 0 \le j \le 2. \end{cases}$$

Let $q_i = s_2$ for all i, and $m_i = t_i/s_2$ for all i. Observe

$$\sum x_i (m_i q_i)^j = \sum x_i \left(\frac{t_i s_2}{s_2}\right)^j = a_j \text{ for all } 3 \le j \le n \text{ (or for all } j \ge 3)$$

$$\sum x_i m_i^2 q_i = \sum \frac{x_i t_i^2}{s_2} = 1.$$

Next assume there exists x_i, m_i, q_i satisfying the given conditions. Let $s_j = \sum x_i t_i^j$ for $j \in \{1, 2, 3\}$ and $t_i = m_i q_i$. Note

$$\sum \alpha_i t_i^j = \sum x_i (m_i q_i)^j = a_j \text{ for all } 3 \le j \le n \text{ (or for all } j \ge 3),$$

and so $(s_0, s_1, s_2, a_3, \dots a_n)$ (or $(s_0, s_1, s_2, a_3, \dots)$) is a moment vector of a finite distribution with positive support. It follows by Lemma 4.4.22 (or Lemma 4.4.21) that the moment vector satisfies the (full) Stieltjes condition. To see that $s_2 \le 1$, let $q = \max_i q_i$ and observe

$$s_2 = \sum x_i m_i^2 q_i^2 \le q \sum x_i m_i^2 q_i = q \le 1.$$

□

Lemma 4.4.24. *There exists s_0, s_2 with $s_2 \le 1$ such that $(s_0, s_2, a_4, \dots a_n)$ satisfies the Stieltjes condition if and only if there exists x_i, m_i, q_i with $x_i, m_i > 0$ and $0 \le q_i \le 1$ satisfying*

1. $2 \sum x_i m_i^2 q_i = 1$
2. $2 \sum x_i (m_i q_i)^{2j} = a_{2j}$ *for all* $2 \le j \le n$.

Similarly, there exists s_0, s_2 with $s_2 \leq 1$ such that $(s_0, s_2, a_4, \dots a_n)$ satisfies the full Stieltjes condition if and only if there exists x_i, m_i, q_i with $x_i, m_i > 0$ and $0 \leq q_i \leq 1$ satisfying (1) and (2) for all $j \geq 4$.

Proof. First assume $(s_0, s_2, a_4, a_6, \dots a_n)$ satisfies the Stieltjes condition (or $(s_0, s_2, a_4, a_6, \dots)$ satisfies the full Stieltjes condition) and $s_2 \leq 1$. It follows that $\left(\frac{s_0}{2}, \frac{s_2}{2}, \frac{a_4}{2}, \frac{a_6}{2}, \dots \frac{a_n}{2}\right)$ satisfies the Stieltjes condition (or $\left(\frac{s_0}{2}, \frac{s_2}{2}, \frac{a_4}{2}, \frac{a_6}{2}, \dots\right)$ satisfies the full Stieltjes condition) because multiplying all entries of a matrix by a positive number does not change the sign of the determinant. By Lemma 4.4.22 (or Lemma 4.4.21) there exists a discrete distribution on $(t_1, t_2, \dots t_k)$ where t_i has mass x_i, $t_i > 0$, and

$$2\sum x_i t_i^j = \begin{cases} a_{2j} & 2 \leq j \leq n \text{ (or } j \geq 4t) \\ s_{2j} & 0 \leq j \leq 1. \end{cases}$$

Let $q_i = s_2$, and $m_i = \sqrt{t_i}/s_2$ for all i. Observe

$$2\sum x_i (m_i q_i)^{2j} = \sum x_i t_i^j = a_{2j} \text{ for all } 2 \leq j \leq n \text{ (or } j \geq 4)$$

$$2\sum x_i m_i^2 q_i = \frac{2}{s_2}\sum x_i t_i = 1.$$

Next assume there exists x_i, m_i, q_i satisfying the given conditions. Let $s_0 = 2\sum x_i$ and $t_i = (m_i q_i)^2$ for all i. Note

$$\sum x_i t_i^j = \sum \frac{x_i}{s_0}(m_i q_i)^{2j} = \frac{a_j}{2} \text{ for all } 2 \leq j \leq n \text{ (or } j \geq 4).$$

Let $q = \max_i q_i$, $s_1 = 2\sum x_i t_i$, and observe

$$s_1 = 2\sum x_i t_i = 2\sum x_i m_i^2 q_i^2 \leq 2q \sum x_i m_i^2 q_i = q \leq 1.$$

It follows that $\left(\frac{s_0}{2}, \frac{s_2}{2}, \frac{a_4}{2}, \ldots \frac{a_n}{2}\right)$ is a moment vector (or $\left(\frac{s_0}{2}, \frac{s_2}{2}, \frac{a_4}{2}, \ldots\right)$ is a moment vector), and therefore by Lemma 4.4.22 (or Lemma 4.4.22) satisfies the (full) Stieltjes condition. It follows that $(s_0, s_1, a_2, \ldots a_n)$ also satisfies the Stieltjes condition (or $(s_0, s_1, a_2, \ldots)$ also satisfies the full Stieltjes condition) because multiplying all entries of a matrix by a positive number does not change the sign of the determinant. □

Proof. (of Theorems 4.1.11, 4.1.12, 4.4.19 and 4.4.20) First assume the vectors of s_i and t_i satisfy the hypotheses. Then by Lemma 4.4.23, there exists (x_i, m_i, q_i) satisfying $\sum x_i(m_i q_i)^j = s_j$ and $\sum x_i m_i^2 q_i = 1$. For each triple add the triple $(m_i, q_i, \beta_i = 0)$ to the distribution. Soon we will specify the corresponding probability μ_i. By Lemma 4.4.24, there exists (x_i, m_i, q_i) satisfying $2\sum x_i(m_i q_i)^{2j} = t_{2j}$ and $2\sum x_i m_i^2 q_i = 1$. For each triple add the triple $(m_i, q_i, \beta_i = 1)$ to the distribution. Let B be the set of indices i such that $\beta_i = 1$ and B^c be the set of indices i such that $\beta_i = 0$. Let $z = \sum_{i\in B} x_i\gamma + \sum_{i\in B^c} x_i(1-\gamma)$. We now define μ by assigning probabilities to triples (m_i, q_i, β_i). If $i \in B^c$, let $\mu_i = x_i\gamma/z$. If $i \in B$, let $\mu_i = x_i(1-\gamma)/z$. Note $\sum \mu_i = 1$, and therefore (μ, a) is a well-defined ROC family with

$$x = 1/\left(\sum_{i\in B^c} \mu_i m_i^2 q_i + 2\sum_{i\in B} \mu_i m_i^2 q_i\right) = z.$$

Theorem 4.4.3 implies that the family achieves the desired limit with sparsity exponent a.

Suppose the limit is achievable by some ROC family (μ, a). Let $\gamma = x\sum_{i\in B^c} \mu_i m_i^2 q_i$, and so $1-\gamma = 2x\sum_{i\in B} \mu_i m_i^2 q_i$. For each $i \in B^c$, let $x_i = x\mu_i/\gamma$. Note $\sum_{i\in B^c} x_i m_i^2 q_i = 1$, and so by Lemma 4.4.23, the vector with $s_j = \sum_{i\in B^c} \mu_i(m_i q_i)^j$ satisfies the Stietljes condition. For each $i \in B$, let $x_i = x\mu_i/(1-\gamma)$. Note $2\sum_{i\in B} x_i m_i^2 q_i = 1$, and so by Lemma 4.4.24, the vector with $t_{2j} = \sum_{i\in B} \mu_i(m_i q_i)^{2j}$ satisfies the Stietljes condition. Theorem 4.4.3 implies that c_j or w_j is the appropriate combination of s_j and t_j. □

Finally we show that a similar argument proves the condition for when it is possible to match a k-cycle-to-edge vector with a ROC family.

Proof. (of Theorem 4.2.2) First assume the vectors of s_i and t_i satisfy the hypotheses. Then by Lemma 4.4.23, there exists (x_i, m_i, q_i) satisfying $\sum x_i(m_i q_i)^j = s_j$ and $\sum x_i m_i^2 q_i = 1$. For each triple add the triple $(m_i, q_i, \beta_i = 0)$ to the distribution. Soon we will specify the corresponding probability μ_i. By Lemma 4.4.24, there exists (x_i, m_i, q_i) satisfying $2\sum x_i(m_i q_i)^{2j} = t_{2j}$ and $2\sum x_i m_i^2 q_i = 1$. For each triple add the triple $(m_i, q_i, \beta_i = 1)$ to the distribution. Let B be the set of indices i such that $\beta_i = 1$ and B^c be the set of indices i such that $\beta_i = 0$. Let $z = \sum_{i\in B} x_i\gamma + \sum_{i\in B^c} x_i(1-\gamma)$. We now define μ by assigning probabilities to triples (m_i, q_i, β_i). If $i \in B^c$, let $\mu_i = x_i\gamma/z$. If $i \in B$, let $\mu_i = x_i(1-\gamma)/z$. Note $\sum \mu_i = 1$, and therefore $(\mu, 0)$ is a well-defined ROC family with

$$x = 1/\left(\sum_{i\in B^c} \mu_i m_i^2 q_i + 2\sum_{i\in B} \mu_i m_i^2 q_i\right) = z.$$

Note $c(j) = c_j/2$ by construction. For $G \sim ROC(n, d, \mu, 0)$ and $d = o\left(n^{\frac{1}{k-1}}\right)$, Corollary 4.4.12 implies that $\mathbb{E}(C_j(G)) = \frac{c_j}{2}nd + o(nd)$. The statement follows.

For the other direction, there is some ROC family that achieves the limit. Since the number of j-cycles grows with nd for all j, we must have $a = 0$. Let $(\mu, 0)$ be the ROC family. Let $\gamma = x\sum_{i\in B^c} \mu_i m_i^2 q_i$, and so $1 - \gamma = 2x\sum_{i\in B} \mu_i m_i^2 q_i$. For each $i \in B^c$, let $x_i = x\mu_i/\gamma$. Note $\sum_{i\in B^c} x_i m_i^2 q_i = 1$, and so by Lemma 4.4.23, the vector with $s_j = \sum_{i\in B^c} \mu_i(m_i q_i)^j$ satisfies the Stietljes condition. For each $i \in B$, let $x_i = x\mu_i/(1-\gamma)$. Note $2\sum_{i\in B} x_i m_i^2 q_i = 1$, and so by Lemma 4.4.24, the vector with $t_{2j} = \sum_{i\in B} \mu_i(m_i q_i)^{2j}$ satisfies the Stietljes condition. Theorem 4.4.3 implies that $c_j/2$ is the appropriate combination of s_j and t_j. □

Simple criterion for the the Stieltjes condition.

The following lemma provides a convenient criterion that implies the Stieltjes condition. In particular, we use this show that the limit of hypercube sequence is totally k-achievable (Theorem 4.1.8).

Lemma 4.4.25. *Let $s_1, s_2, \ldots s_k$ be a vector with $s_1 > 0$ satisfying $s_x s_y < s_a s_b$ for all $1 \leq a < x \leq y < b$. Then there exists $s_0 > 0$ such that $(s_0, s_1, s_2, \ldots s_k)$ satisfies the Stieltjes condition.*

The following is the key lemma for proving Lemma 4.4.25.

Lemma 4.4.26. *Let $s_1, s_2, \ldots s_k$ be a vector with $s_1 > 0$ and $s_x s_y < s_a s_b$ for all $1 \leq a < x \leq y < b$. Let H be the $\lfloor \frac{k+1}{2} \rfloor \times \lfloor \frac{k+1}{2} \rfloor$ with $H_{ij} = s_{i+j-1}$. Then all leading principal minors of H have positive determinant.*

Proof. Let H_k denote the k^{th} leading principal minor of H. We show that $det(H_k) > 0$ by induction on k. Note $det(H_1) = s_1 > 0$. Next assume $det(H_{k-1}) > 0$. Write $H_k = AB$ where A and B are in the form displayed here.

$$H_k = \begin{pmatrix} s_1 & s_2 & s_3 & \ldots & s_k \\ s_2 & s_3 & & & \vdots \\ s_3 & & & & \vdots \\ \vdots & & & \ddots & \vdots \\ s_k & \ldots & \ldots & \ldots & s_{2k-1} \end{pmatrix} = \begin{pmatrix} s_1 & s_2 & \ldots & s_{k-1} & 0 \\ s_2 & & & \vdots & 0 \\ \vdots & & \ddots & \vdots & \vdots \\ s_{k-1} & \ldots & \ldots & s_{2k-3} & 0 \\ 0 & 0 & \ldots & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & \ldots & 0 & x_1 \\ 0 & 1 & & \vdots & x_2 \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & \ldots & \ldots & 1 & x_{k-1} \\ s_k & s_{k+1} & \ldots & s_{2k-2} & s_{2k-1} \end{pmatrix}.$$

Note $det(A) = det(H_{k-1})$, which is positive by the inductive hypothesis. It follows there exists a unique solution of real values $x_1, x_2, \ldots x_{k-1}$ so that $H_k = AB$. Since $det(H_k) = det(A)det(B)$ and $det(A) > 0$, it suffices to show that $det(B) > 0$ to prove the inductive hypothesis.

Note $det(B) = s_{2k-1} - L$ where

$$L = \begin{pmatrix} s_k & s_{k+1} & \ldots & s_{2k-2} \end{pmatrix} \begin{pmatrix} x_1 & x_2 & \ldots & x_{k-1} \end{pmatrix}^T.$$

By construction of A and B,

$$\begin{pmatrix} s_1 & s_2 & \dots & s_{k-1} \\ s_2 & & & \vdots \\ \vdots & & \ddots & \vdots \\ s_{k-1} & \dots & \dots & s_{2k-3} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{k-1} \end{pmatrix} = \begin{pmatrix} s_k \\ s_{k+1} \\ \vdots \\ s_{2k-2} \end{pmatrix}. \tag{4.21}$$

For $i \in [0, k-2]$, define

$$\alpha_i = \frac{s_{k+i}}{s_{i+1} + \cdots + s_{k-1+i}},$$

and let $\alpha = \max_{i \in [0,k-2]} \alpha_i$. Therefore for all $i \in [0, k-2]$

$$x_{i+1} s_{k+i} = \alpha_i x_{i+1}(s_{i+1} + \cdots + s_{k-1+i}) \leq \alpha x_{i+1}(s_{i+1} + \cdots + s_{k-1+i}).$$

Summing the above equation over all $i \in [0, k-2]$ and applying equality Equation (4.21) yields

$$L \leq \alpha(s_k + s_{k+1} + \cdots + s_{2k-2}).$$

To prove $det(B) = s_{2k-1} - L > 0$ we show that for all $i \in [0, k-2]$, $\alpha_i(s_k + s_{k+1} + \cdots + s_{2k-2}) < s_{2k-1}$, or equivalently

$$s_{k+i}(s_k + s_{k+1} + \cdots + s_{2k-2}) < s_{2k-1}(s_{i+1} + \cdots + s_{k-1+i}). \tag{4.22}$$

Note by assumption $s_{k+i} s_{k+j} < s_{2k-1} s_{i+j+1}$ for all $i, j \in [0, k-2]$. Therefore the j^{th} term on the left side of Equation (4.22) is less than the j^{th} term on the right side of Equation (4.22), and so Equation (4.22) holds. □

Proof. (of Lemma 4.4.25.) Let $H^{(0)}$ and $H^{(1)}$ be defined for the vector $(s_0, s_1, \dots s_k)$ as in Definition 4.1.10. Lemma 4.4.26 implies that all leading principal minors of $H^{(1)}$ have positive

determinant. It remains to show that there exists $s_0 > 0$ such that all leading principal minors of $H^{(0)}$ have positive determinant.

Let H' be $H^{(0)}$ with the first row and column deleted. The i^{th} leading principal determinant of $H^{(1)}$ has the form $s_0 h_i + b_i$ where h_i is the $(i-1)^{st}$ principal determinant of H'. (Taking the determinant via expansion of the first row makes this clear.) Note that Lemma 4.4.26 applied to the vector $(s_2, s_3, \dots s_k)$ guarentees that each $h_i > 0$. Therefore, it is possible to pick s_0 sufficiently large such that all principal determinants $s_0 h_i + b_i$ of $H^{(1)}$ are positive. □

4.5 ROC achievable limits

4.5.1 Hypercube

In this section, we prove Theorem 4.1.8, which states that the limit of the hypercube sequence is totally k-achievable, and discuss the limits of several related graph sequences. First we provide the ROC parameters which achieve the small k-limits of the hypercube.

Remark 4.5.1. The ROC family $(\mu, 1/2)$ where μ is the distribution with support size one on $(8, 1/4, 1)$ achieves the 6-limit of the hypercube sequence. To achieve longer limits, the distribution will have larger support.

We now prove Theorem 4.1.8. Recall from Lemma 4.3.1 that the sparsity exponent of the hypercube sequence is 1/2. Therefore, to prove the theorem we apply Theorem 4.1.12, which states the vector $(w_3, w_4, \dots w_k)$ can be achieved by ROC if the normalized cycle count vector $(c_3, c_4, \dots c_k)$ corresponding to the transform T can be extended to satisfy the Stieltjes condition. The following lemma gives the cycle vector for the hypercube.

Lemma 4.5.2. *Recall from Lemma 4.3.1 that the limit of the hypercube sequence* (G_d) *is* $(w_3, w_4, \dots)$ *where*

$$w_j = \begin{cases} (j-1)!! & \text{for } j \text{ even} \\ 0 & \text{for } j \text{ odd.} \end{cases}$$

For T the cycle transform given in Definition 4.4.1*, $T((0, s_2, 0, s_3, 0, \dots)) = (0, w_4, 0, w_6, 0, \dots)$ where $s_1 = 1$ and $s_n = (n-1)\sum_{j=1}^{n-1} s_j s_{n-j}$.*

Remark 4.5.3. This sequence is A000699 in OEIS: $1, 1, 4, 27, 248, 2830, \dots$

Proof. (of Lemma 4.5.2) Note that the hypercube is vertex transitive so the sequence of hypercubes is essentially k-locally regular. Therefore, Theorem 4.4.10 implies $C((0, w_4, 0, w_6, 0, \dots)) = (c_3, c_4, c_5, \dots)$ where

$$c_i = \lim_{d\to\infty} \frac{C_i(G_d)}{2^d d^{i/2}}$$

where $C_i(G_d)$ is the normalized number of i cycles at a vertex. Instead of applying polynomial operations to the vector $(0, w_4, 0, w_6, 0, \dots)$ to obtain $(c_3, c_4, c_5, \dots)$ we directly compute $C_k(G_d)$, the number of cycles in the d-dimensional hypercube graph. As in Lemma 4.3.1, we think of k-walks on the hypercube as length k strings where the i^{th} character indicates which of the d coordinates is changed on the i^{th} edge of the walk. For closed walks each coordinate that is changed must be changed back, so each coordinate that appears in the string must appear an even number of times. For $1 \le i \le k/2$, let Z_i be the number of such strings of length k that involve i coordinates and correspond to a k-cycle on the hypercube graph. Since there are d coordinates $Z_i = \Theta(d^i) = o(d^{k/2})$ for $i < k/2$ and so

$$C_k(G_d) = nZ_{k/2} + o\left(nd^{k/2}\right)$$

where $n = 2^d$ is the number of vertices.

We compute $Z_{k/2}$ by constructing a correspondence between length k strings with $k/2$ characters each appearing twice that represent cycles and irreducible link diagrams. A link diagram is defined as $2n$ points in a line with n arcs such that each arc connects precisely two distinct points and each point is in precisely one arc. The arcs define a complete pairing of the interval $[1, 2n]$. A link diagram is reducible if there is a subset of $j < n$ arcs that form

a complete pairing of a subinterval of $[1, 2n]$ and irreducible otherwise. Let S be the set of k length strings in which the characters $1, 2, \ldots k/2$ each appear twice and the first appearance of character i occurs before the first appearance of j for all $i < j$. Let L be the set link diagrams L on k points. Let $\overline{S} \subseteq S$ be the subset of strings that correspond to cycles on the hypercube and let $\overline{L} \subseteq L$ be the set of irreducible link diagrams.

We construct a bijection $f : S \to L$ and show f restricted to $\overline{S}$ gives a bijection between $\overline{S}$ and $\overline{L}$. Given $s \in S$, we construct a corresponding link diagram $f(s) \in L$ by labeling k points so that the i^{th} point is labeled with the i^{th} character of s and then drawing an arc between each pair of points with the same label. Note f is a bijection. (To produce $f^{-1}(\ell)$ label the arcs $1, 2, \ldots k/2$ by order of their left endpoints. Label each point with the label of its arc and read off the string of the labels.) It remains to show that $f(s) \in \overline{L}$ if and only if $s \in \overline{S}$. We prove the contrapositive. Suppose $s \notin \overline{S}$. Then the hypercube closed walk corresponding to s is not a cycle. Therefore there exists j and steps $i, i+1, \ldots i+j$ of the walk that make a $j/2$ cycle. (Here by convention traversing an edge twice is a 2-cycle.) Since $i, i+1, \ldots i+j$ form a cycle, each coordinate that was changed between step i and step $i+j$ must have been changed back. Therefore each character that appears in the interval $[i, i+j]$ appears appears twice. It follows $f(s)$ has a complete pairing of the subinterval $[i, i+j]$ and therefore is reducible. Next suppose $\ell \notin \overline{L}$. Then there exists a subinterval $[i, i+j] \neq [1, 2n]$ with a complete pairing. Therefore the walk corresponding to $f^{-1}(\ell)$ is not a cycle because the walk visits the same vertex before step i and after step $i+j$. It follows that $f^{-1}(\ell) \notin \overline{S}$.

Thus $|\overline{L}| = |\overline{S}|$. See [Ste78] for a proof that $|L| = s_{k/2}$. There are $d(d-1)\ldots(d-k/2+1) = d^{k/2} + o\left(d^{k/2}\right)$ ways to select the $k/2$ coordinates in the order they will be changed. Thus $Z_{k/2} = d^{k/2}|S| + o\left(d^{k/2}\right) = s_{k/2}d^{k/2} + o\left(d^{k/2}\right)$, and therefore

$$C_k(G_d) = s_{k/2}nd^{k/2} + o\left(nd^{k/2}\right).$$

□

Lemma 4.5.4 (from [Ste78]). *Let* $s_1 = 1$ *and* $s_n = (n-1)\sum_{j=1}^{n-1} s_j s_{n-j}$. *Let* $s_1 = 1$ *and* $s_n = (n-1)\sum_{j=1}^{n-1} s_j s_{n-j}$. *Then*

$$s_{n+1} > (2n+1)s_n \quad \textit{for} \quad n \geq 4$$

$$s_{n+1} < (2n+2)s_n \quad \textit{for} \quad n \geq 1.$$

Lemma 4.5.5. *Let* $s_1 = 1$ *and* $s_n = (n-1)\sum_{j=1}^{n-1} s_j s_{n-j}$. *Let* $s_1 = 1$ *and* $s_n = (n-1)\sum_{j=1}^{n-1} s_j s_{n-j}$. *Then there exists* $s_0 > 0$ *such that* $(s_0, s_1, s_2, \ldots s_k)$ *satisfies the Stieltjes condition.*

Proof. We apply Lemma 4.4.25 which says a vector with $s_x s_y < s_a s_b$ for all $1 \leq a < x \leq y < b$ can be can be extended to satisfy the Stieltjes condition. We show this conditions holds for the infinite vector.

First we consider the case when $y \geq 4$. By Lemma 4.5.4

$$s_b > \frac{(2b-1)!!}{(2y-1)!!}s_y \quad \text{and} \quad s_x < \frac{(2x)!!}{(2a)!!}s_a.$$

Note $\frac{(2x)!!}{(2a)!!} < \frac{(2b-1)!!}{(2y-1)!!}$, and therefore

$$s_x s_y < \frac{(2x)!!}{(2a)!!}s_a s_y < \frac{(2b-1)!!}{(2y-1)!!}s_a s_y < s_a s_b.$$

We consider the remaining three cases separately. For $a = 1, x = 2, y = 3$, $b \geq 4$ so $s_b \geq 27$. Therefore $s_x s_y = 4 < 27 \leq s_a s_b$. For $a = 1, x = 3, y = 3$, $b \geq 4$ so $s_b \geq 27$. Therefore $s_x s_y = 16 < 27 \leq s_a s_b$. For $a = 1, x = 2, y = 2$, $b \geq 3$ so $s_b \geq 4$. Therefore $s_x s_y = 1 < 4 \leq s_a s_b$.

□

Proof. (of Theorem 4.1.8) Follows directly from Lemma 4.5.2, Lemma 4.5.5, and Theorem 4.1.12. □

Remark 4.5.6. The limit of the sequence of hypercubes is not fully achievable.

Proof. For a ROC family $(\mu, 1/2)$ with $m = \max_i m_i$, the w_k coordinate in the limit vector is at most $x(2m)^k$. However the w_k coordinate in the hypercube sequence is $(k-1)!! = \Theta\left(\left(\frac{k}{e}\right)^{k/2}\right)$. Therefore there is no μ which achieves the full hypercube limit vector. □

Generalized hypercubes

Two generalizations of the hypercube have the same limit and therefore are also totally k-achievable.

Corollary 4.5.7 (Hypercube generalizations)**.** *The following sequences of graphs (G_d) converge with sparsity exponent 1/2 to the same limit as the hypercube sequence.*

1. *(Hamming generalization) Let G_d be the graph on vertex set $\{0, 1, \ldots, k-1\}^d$ where two vertices are adjacent if the Hamming distance between their labels is one.*

2. *(Cayley generalization) Let G_d be the graph on vertex set $\{0, 1, \ldots, k-1\}^d$ where two vertices are adjacent if their labels differ by a standard basis vector.*

Proof. Since the Hamming and Cayley sequences are locally regular, it suffices to show that the sequences have sparsity exponent $1/2$ and the same vector of normalized cycle counts as the hypercube. Let D denote the degree of the graph G_d, so for the Hamming graph $D = d(k-1)$ and for the Cayely graph $D = 2d$.

First we show that both sequences have sparsity exponent $1/2$ by showing that each vertex is in $O\left(D^{i/2}\right)$ i-cycles (locally regularity guarantees the walk counts are of the same order). We count i-cycles by grouping them according to the number of coordinate positions changed during the cycle, as in the proof of Lemma 4.5.2. The highest order term comes

from i cycles in which $i/2$ coordinates are changed. Therefore the number of i-cycles at each vertex is $O\left(d^{i/2}\right) = O\left(D^{i/2}\right)$ and it follows that the sparsity exponent is $1/2$.

Next we compute the cycle vector $c_i = \lim_{d\to\infty} \frac{C(G_d)}{nD^{i/2}}$. The number of i-cycles at a vertex that involve changing fewer than $i/2$ coordinates is $o(D^{i/2})$, so such cycles do not contribute to c_i. Therefore, while there are odd cycles in the Hamming and Cayley graphs, $c_i = 0$ for i odd. We now count the number of i-cycles at a vertex that involve changing $i/2$ coordinates. As described in Lemma 4.5.2 there are $s_{i/2}d^{i/2}$ ways to select $i/2$ coordinates and change them in a manner that corresponds to a cycle. In the hypercube, there is only one way to change a single coordinate, so the total number of cycles at a vertex is $s_{i/2}d^{i/2}$.

In the Hamming graph there are $k-1$ ways to change a coordinate since there are k possibilities for each coordinate. Therefore, for the Hamming sequence and i even

$$c_i = \lim_{d\to\infty} \frac{s_{i/2}(k-1)^{i/2}}{nD^{i/2}} = \lim_{d\to\infty} = s_{i/2}.$$

In the Cayley graph there are two ways to change a single coordinate (either add one or subtract one). Therefore, for the Cayley sequence and i even

$$c_i = \lim_{d\to\infty} \frac{s_{i/2}2^{i/2}}{nD^{i/2}} = \lim_{d\to\infty} = s_{i/2}.$$

□

Remark 4.5.8. The above corollary shows that same ROC family (μ, a) achieves the k-limit of the sequence of hypercubes, and the closely related Hamming and Cayley generalizations. While these sequences all have the same ROC family limit object, the ROC family can produce sequences of ROC graphs (G_d), $G_d \sim ROC(n_d, d_d, \mu, a)$, unique to each of these settings by varying relationship between n_d and d_d. A sequence with $n_d = 2^d$ and $d_d = d$ will match the edge density and unnormalized walk counts of the hypercube, whereas a sequence

with $n_d = k^d$ and $d_d = d(k-1)$ or $d_d = 2d$ will match the edge density and unnormalized walk counts of the Hamming or Cayley generalization respectively.

4.5.2 Rook graph

We now prove Theorem 4.1.9 which states that the limit of the sequence of rook's graphs is fully achievable.

Proof. (of Theorem 4.1.9.) Recall from Lemma 4.3.2 that the sequence of (G_k) has sparsity exponent 1 and converges to the vector with $w_i = 2^{2-i}$. By Theorem 4.4.3, the ROC family with $a = 1$ and μ the distribution that selects $m = 1/2$ and $q = 1$ with probability 1 achieves this limit. □

4.5.3 Erdős-Rényi sequences

We consider ROC approximations of the sequences of Erdős-Rényi graphs given in Lemma 4.3.6.

Theorem 4.5.9. *Let $\ell > 1$. Let $(G_n) \sim G(n^{2\ell}, n^{2\ell-2})$.*

1. *For $k < 2\ell$, the k-limit of (G_n) is achieved by any ROC family with $a < 1/2$.*
2. *For $k \geq 2\ell$, the k-limit of (G_n) is not k-achievable by any ROC family. However, for any $\varepsilon > 0$, there exists a ROC k-achievable vector that is L_∞ distance at most ε from the k-limit.*
3. *The sparsity exponent of (G_n) is 1 and the limit is $(0, 0, \dots)$. This limit is not ROC fully achievable. However, for any $\varepsilon > 0$, there exists a ROC fully achievable vector that is L^∞ distance at most ε from $(0, 0, \dots)$.*

Proof. For $k < 2\ell$ the sparsity exponent of (G_n) is $1/2$ and the k-limit is $(w_3, w_4, \dots w_k)$ where $w_i = 0$ for i odd and $w_i = Cat_{i/2}$ for i even. By Theorem 4.4.3, this is the limit for any ROC family with $a < 1/2$.

For $k = 2\ell$, the k-sparsity exponent of (G_n) is $1/2$ and the k-limit is $(w_3, w_4, \ldots, w_{k-1}, \overline{w}_k)$ where $w_i = 0$ for odd i, $w_i = Cat_{i/2}$ for even i, and $\overline{w}_k = w_k + 1$. By Theorem 4.4.3, in order to approximate the vector it is necessary to have μ be such that $x \sum \mu_i (m_i q_i)^j = c_j$ where $(c_3, c_4, \ldots c_k) = T(w_3, w_4, \ldots, w_{k-1}, \overline{w}_k)$ is the cycle transform, so $c_j = 0$ for $j < k$ and $c_k = 1$. Since $\mu_i, q_i, m_i > 0$, $c_k = 1$ implies $c_j \neq 0$ for all $j < k$. Therefore, the vector cannot be achieved exactly by ROC.

We now show that it is possible to achieve a vector that is arbitrarily close to the desired vector with respect to the L_∞ metric. Note that for μ the distribution on one point $m = \delta^{\frac{1-k}{k-2}}$ and $q = \delta$, the resulting ROC family $(\mu, 1/2)$ has

$$c_j = m^{j-2} q^{j-1} = \delta^{\frac{(j-2)(1-k)}{k-2} + j - 1}.$$

Therefore $c_k = 1$ and c_j for $j < k$ can be made arbitrarily small by decreasing δ. To achieve L_∞ distance ε, choose δ small enough so that $\max_{j<k} w_j = max_j T(c_3, c_4, \ldots c_k)_j < \varepsilon$.

For $k > 2\ell$, the sparsity exponent of (G_n) is $\frac{k-\ell-1}{k-2}$ and the k-limit is $(w_3, w_4, \ldots, w_k)$ where $w_i = 0$ for $i < k$, $w_k = 1$. Therefore, by Theorem 4.4.3, to approximate the vector we likewise need μ be such that $x \sum \mu_i (m_i q_i)^j = w_j$ where $c_j = w_j = 0$ for $j < k$ and $w_k = c_k = 1$, and the result is as in the previous case.

Similarly, we can approximate the vector $(0, 0, \ldots)$ with sparsity exponent 1 up to arbitrarily small error with respect to the L_∞ distance. Note that for μ the distribution on one point m and q, the ROC family $(\mu, 1)$ has $w_k = m^{k-2} q^{k-1}$. Therefore it is possible to achieve error ε by selecting m and q such that $\max_k m^{k-2} q^{k-1} < \varepsilon$. □

4.6 Discussion

4.6.1 Limitations of the ROC model

We have shown that the ROC model is an insightful limit object for many sequences of graphs; the model is succinct and can be easily sampled to produce graphs with the same normalized walk counts as the sequence up to terms that disappear as the size of the sampled graph grows. Theorems 4.1.11, 4.1.12, 4.4.19 and 4.4.20 give necessary and sufficient conditions for when a limit sequence can be achieved. The natural next question is whether all graphs sequences converge to a limit that can be achieved by a ROC family. The answer to this question is no. There are both sequences of graphs that are not convergent in any of the senses we have defined, and convergent sequences of graphs with limits that are not achievable by a ROC family. We discuss such examples in this section.

Non-convergent graph sequences. Not all sequences of graphs converge or have a convergent subsequence. Consider the following sequence of ROC graphs drawn from different ROC families.

Example 4.6.1. Let $a \in [1/2, 1]$ and let μ_i be the distribution on one point $m_i = i$ and $q_i = 1$. Let (G_i) be a sequence of graphs with $G_i \sim ROC(n_i, d_i, \mu_i, a)$ such that d_i satisfies the degree conditions given in Definition 4.1.7. The sequence (G_i) is not k-convergent for any k and is not fully convergent.

Proof. By Theorem 4.4.3, $\mathbb{E}(W_3(G_i)) = m_i q_i^2 n_i d_i^{1+a(k-2)} + o\left(n_i d_i^{1+a(k-2)}\right)$. Therefore $W_3(G_i, a) = i + \varepsilon(i)$ and for all $\alpha > a$ $W_3(G_i, \alpha) = 0 + \varepsilon(i)$ where $\varepsilon(i)$ is an error term that vanishes with high probability as i tends to infinity. It follows that almost surely the sparsity exponent and k-sparsity exponent are a and the sequence $W_3(G, a)$ does not converge. Thus, the sequence is almost surely not k-convergent or fully convergent. □

Limit sequences that are not achievable by any ROC model We give a sequence of graphs with increasing degree that converges to a limit that is not achievable by any ROC family and provide a method for producing such sequences. First we give a necessary condition on (w_3, w_4, w_5, w_6) for it to appear as the prefix of a limit vector achievable by a ROC family.

Lemma 4.6.2. *If (w_3, w_4, w_5, w_6) is a prefix of a k-limit that can be achieved by ROC family with sparsity exponent $> 1/2$, then*

$$w_3^2 w_6 \geq w_5^3.$$

Proof. By Theorem 4.1.11, there exists $\gamma \in [0, 1]$, s_0, s_1, and $s_2 \leq 1$ such that $(s_0, s_1, \ldots s_k)$ satisfies the Stieltjes condition and $w_j = s_j \gamma$ for j odd. This condition implies that $H_{2s}^{(0)}$ and $H_{2s+1}^{(1)}$ as defined in Definition 4.1.10 are positive semidefinite. It follows that the principal minors

$$\begin{pmatrix} s_4 & s_5 \\ s_5 & s_6 \end{pmatrix} = \begin{pmatrix} s_4 & \frac{w_5}{\gamma} \\ \frac{w_5}{\gamma} & s_6 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} s_3 & s_4 \\ s_4 & s_5 \end{pmatrix} = \begin{pmatrix} \frac{w_3}{\gamma} & s_4 \\ s_4 & \frac{w_5}{\gamma} \end{pmatrix}$$

of $H_{2s}^{(0)}$ and $H_{2s+1}^{(1)}$ respectively have non-negative determinant. Therefore

$$\gamma^2 s_4 s_6 \geq w_5^2 \quad \text{and} \quad \gamma^2 s_4^2 \leq w_3 w_5,$$

and so $\gamma s_6 \geq \frac{w_5^{3/2}}{\sqrt{w_3}}$. Since $w_6 \geq \gamma s_6$, the statement follows. □

Next we show how to construct a sequence of graphs with increasing degree that fails this condition. This construction is due to Shyamal Patel.

Lemma 4.6.3. *Let G_0 be a graph. We construct a sequence G_i as follows. Let G_i be the graph with adjacency matrix $\begin{pmatrix} A_{i-1} & A_{i-1} \\ A_{i-1} & A_{i-1} \end{pmatrix}$ where A_{i-1} is the adjacency matrix of G_{i-1}. Let*

$w_j = W_j(G_0, 1)$. *Then for each* i,

$$W_j(G_i, 1) = w_j,$$

and so (G_i) *converges to* $(w_3, w_4, w_5, \dots)$ *with sparsity exponent 1.*

Proof. Let G_0 be a graph on n vertices with average degree d. Let A_0 be the adjacency matrix of G_0 and let $\lambda_1 \geq \lambda_2 \geq \dots \lambda_\ell$ be the non-zero eigenvalues of A_0. Note that if λ is an eigenvalue of A_{i-1} with eigenvector v, then 2λ is an eigenvalue of A_i with eigenvector $\begin{bmatrix} v \\ v \end{bmatrix}$. Since the adjacency matrix A_i has the same rank as the adjacency matrix of A_{i-1}, the set of non-zero eigenvalues of A_i is precisely the set of non-zero eigenvalues of A_{i-1} in which each is doubled. Therefore A_i has non-zero eigenvalues $2^i\lambda_1, 2^i\lambda_2, \dots 2^i\lambda_\ell$. Note G_i has $G_0 2^i$ vertices and average degree $2^i d$. Therefore

$$W_j(G_i, 1) = \frac{\sum_{b=1}^{\ell}(2^i\lambda_b)^j}{2^i n(2^i d)^{j-1}} = \frac{\sum_{b=1}^{\ell}(\lambda_b)^j}{nd^{j-1}} = w_j.$$

□

The lemma implies that if there is a graph with $W_j(G, 1) = w_j$, then there is a sequence of graphs (G_i) with increasing degree that converges to this limit with sparsity exponent 1. Taking G_0 to be a girth four graph yields a sequence (G_i) with a limit vector that violates the condition of Lemma 4.6.2. This sequence is dense since $d_i = \Theta(n_i)$. However, we can construct a sparser sequence (G_i') from (G_i) with the same limit by taking each G_i' to be the union of disjoint copies of G_i.

Lemma 4.6.4. *Let* (G_i) *be a convergent sequence of graphs with sparsity exponent* α. *Let* (t_i) *be a sequence of positive integers, and let* (G_i') *be a graph sequence in which* G_i' *consists of* t_i *disjoint copies of* G_i. *Then* (G_i') *achieves the same limit as* G_i *with the same sparsity*

exponent.

Proof. Note that G_i and G_i' both have average degree d_i and the number of vertices in G_i' is $n_i' = t_i n_i$. Note also that $W_k(G_i') = t_i W_k(G_i)$. It follows that

$$W_k(G_i', \alpha) = \frac{W_j(G_i')}{n_i' d_i'^{1+a(j-2)}} = \frac{W_j(G_i)}{n_i d_i^{1+a(j-2)}}.$$

□

We use Lemmas 4.6.2 and 4.6.3 to construct a family of sequences with arbitrary sparsity that are not achievable by the ROC model. This example implies that there is no class of densities for which the ROC model can capture all 6-limits of sequences with the specified density.

Example 4.6.5. Let G_0 be the five cycle. Then the sequence (G_i) defined as in Lemma 4.6.3 converges to a limit that cannot be achieved by any ROC family. This limit $(0, 3/4, 1/8, 5/8)$ is at constant distance from any achievable limit. Moreover, there exists a sequence (G_i') with the same limit and $d_i' = f(n_i')$ for any function $f(n) = o(n)$. To see this, apply Lemma 4.6.4 to the sequence (G_i) with $t_i = f^{-1}(n_i)/n_i$.

4.6.2 ROC as a model for real-world graphs.

Modeling a graph as the union of relatively dense communities has explanatory value for many real-world settings, in particular for social and biological networks. Social networks can naturally be thought of as the union of communities where each community represents a shared interest or experience (e.g. school, work, or a particular hobby); the conceptualization of social networks as overlapping communities has been studied in [PBV07], [XSL11]. Protein-protein interaction networks can also be modeled by overlapping communities, each representing a group of proteins that interact with each other in order to perform a specific cellular process.

Analyses of such networks show proteins are involved in multiple cellular processes, and therefore overlapping communities define the structure of the underlying graph [ABL10], [Kro+06], [BH03].

Our model therefore may be a useful tool for approximating large graphs. It is often not possible to test algorithms on graphs with billions of vertices (such as the brain, social graphs, and the internet). Instead, one could use the DROC model to generate a smaller graph with same clustering coefficient and degree distribution as the large graph, and then optimize the algorithm in this testable setting. Further study of such a small graph approximation could provide insight into the structure of the large graph of interest.

Moreover, the ROC model could be used as a null hypothesis for testing properties of a real-world networks known to have community structure. It is established practice to compare real-world graphs to various random graph models to understand the non-random aspects of its structure ([CG08, Son+05, New05, NSW01]). The ROC model is particularly well-suited to be the null hypothesis graph for graphs with known community structure. Comparing such a network to a ROC network would differentiate between properties of the network that are artifacts of community structure and those that are unique to the graph.

4.6.3 ROC as a limit object

We have seen that the ROC model provides a sampleable approximation of the limits of many sparse graph sequences, in particular the hypercube sequence. Our metric was defined in terms of a vector of closed walk counts of each length appropriately normalized. This vector is a natural choice because closed walk counts are equivalent to the moments of the eigenspectrum, and the normalization factor encodes average density of local neighborhoods. Our findings suggest that measuring closed walk counts and approximating with the ROC model is a promising beginning to a complete theory for describing the limits of sparse graph sequences (in particular those with roughly uniform degree and are not captured by

graphexes). We end with a discussion of future directions that illustrate the potential of the ROC model and address current limitations of the theory.

Distance and convergence. Our notion of convergence based on normalized closed walk count vectors differs from other notions of graph convergence in two key ways. First, our theory does not provide an inherent metric for describing the distance between two graphs. The normalization factor used to determine the convergence of a sequence of graphs depends on the rate of growth of the closed walk counts in the sequence. Therefore, it not clear which normalization factor α to use when comparing the closed walk vectors of just two graphs. Second, due to this flexibility in normalization parameter α, the space of all vectors of normalized closed walk counts is unbounded, and so it is possible to construct sequences of graphs with no convergent subsequence (as in Section 4.6.1). In contrast the set of local profiles and the set of graphons are compact, so every sequence of graphs in these settings has a convergent subsequence. Further investigation is necessary to meaningfully extend the ROC theory to the context of approximating a small set of graphs rather than a sequence, and to the context of non-convergent sequences.

Capturing cuts. While a graph H drawn from the ROC model may capture the closed walk counts of a graph G, there is no guarantee that H and G will have similar cuts. (The in the local profile approach for bounded degree graphs also succeeds at encoding local properties and fails to capture the global property of cuts.) For example, consider a convergent sequence of connected graphs (G_i) and a sequence of graphs (G_i') where G_i' is a collection of disjoint copies of G_i. Lemma 4.6.4 implies (G_i) and (G_i') have the same limit; however the cuts in these sequences greatly differ. Moreover the cuts of a ROC graph drawn from the family that achieves the limit need not have cuts that match either G_i or G_i'.

In general, even if the moments of the eigenspectra of two graphs match, their spectral gaps and precise set of eigenvalues may greatly differ. In Appendix C we discuss a different

approximation of the spectrum of the hypercube graph. It is not of constant size (the size of the approximation grows with d for a hypercube of size 2^d), but it captures the d distinct eigenvalues of the hypercube precisely (and therefore the minimum cut). On the other hand, the approximation does not preserve information about the multiplicities of the eigenvalues, and hence does not capture the moments.

An extension of the ROC model. We imagine the following extension to the ROC model that has the potential to encode information about the cuts of a graphs and give a finer grained approximation of local structure while also maintaining the approximation of closed walk counts. Begin with a partition of the vertex set, and for each community type specify a distribution over partition classes. Then, when adding a community to a ROC graph, instead of selecting community membership from the entire vertex set with equal probability, select vertices for the community based on the corresponding distribution over partition classes. This modification has the potential to better approximate cuts because it is possible to control the number of edges between partition classes.

Moreover, the above modification will likely be a better approximation for graphs that are not necessarily close to locally regular. Currently a ROC approximation produces a graph in which each vertex is in approximately the number of closed walks as an average vertex in the target graph. However, the target graph could be made up of several types of vertices where all vertices of a given type have the same local closed walk count vector. A ROC approximation captures the weighted average of these vectors, but does not retain information about the distribution over such vectors. An expanded theory, perhaps including the above modification, could create graph approximations that capture the distribution of local closed walk counts vectors at each vertex.

Achievability of all limits. As demonstrated in Section 4.6.1, not all limit sequences can be achieved by a ROC model. In particular, the model may not be able to capture the limits

of sequences of girth five graphs because the density of the communities need to produce many five cycles will also produce many three and four cycles. This problem could be resolved by generalizing the model so that communities may have structure other than E-R random graphs. Alternately, the aforementioned approach of adding ROCs between partition classes might provide sufficient flexibility to achieve a wider range of limits.

4.6.4 Additional open questions.

1. A further generalization involves adding particular subgraphs from a specified set according to some distribution instead of E-R graphs in each step (e.g., perfect matchings or Hamiltonian paths). Does doing so allow for greater flexibility in tuning the number of various types of motifs present (not just triangles and four-cycles)?

2. Can the DROC model be extended to produce graphs with arbitrary clustering coefficients and degree distributions (that have long upper tails)? A modification of the DROC model could be that vertices with higher target degrees are more likely to join each community.

3. A fundamental question in the study of graphs is how to identify relatively dense clusters. For example, clustering protein-protein interaction networks is a useful technique for identifying possible cellular functions of proteins whose functions were otherwise unknown [Ste+05, Kro+06]. An algorithm designed specifically to identify the communities in a graph drawn from the ROC model has potential to become a state-of-the-art algorithm for clustering real-world networks with overlapping community structure.

4. A ROC graph H that approximates a target graph G has similar closed walk counts as G. To what extent does this similarly imply that algorithms will behave similarly on G and H? For instance, can we analyze the behavior of random walks or percolation of

ROC graphs? How does this behavior compare to the behavior of the same process on other graphs with the same closed walk counts?

5. Moreover, the asymptotic thresholds for properties of ROC graphs have yet to be studied. (See [FK15] for a survey on E-R random graphs.) Which phase transitions appearing in E-R random graphs also appear in ROC graphs? Does every nontrivial monotone property have a threshold?

4.7 ROC model supplementary section

4.7.1 Limitations of previous approaches

Theorem 4.7.1. *Let G be a graph on n vertices obtained by repeatedly adding triangles on sets of three randomly chosen vertices. If the average degree is less than $\sqrt{n}$, the expected ratio of triangles to edges is at most 2/3.*

Proof. Let t be the number of triangles added and d the average degree, so $d = 6t/n$. To ensure that $d < \sqrt{n}$, $t < n^{3/2}/6$. The total number of triangles in the graph is $t + (d/n)^3\binom{n}{3} = t + d^3/6 = t + 36t^3/n^3$. It follows that the expected ratio of triangles to edges is at most

$$\frac{t + 36\left(\frac{t}{n}\right)^3}{3t} \leq \frac{2}{3}.$$

□

Proof. (of Proposition 4.1.1) Let $\sigma_1 \dots \sigma_{rank(M)}$ denote the eigenvalues of M.

$$\begin{aligned}
\mathbb{E}(\#k\text{-cycles}) &= \sum_{i_1 \neq i_2 \cdots \neq i_k} M_{i_1 i_2} M_{i_2 i_3} \dots M_{i_k i_1} \\
&\leq Tr(M^k) \\
&= \sum_{i=1}^{\text{rank}(M)} \sigma_i^k
\end{aligned}$$

$$\leq \operatorname{rank}(M) d^k.$$

□

4.7.2 Connectivity of the ROC model

We describe the thresholds for connectivity for ROC graphs with one community type, $ROC(n, d, s, q)$. A vertex is isolated if it is has no adjacent edges. A community is isolated if it does not intersect any other communities. Here we use the abbreviation a.a.s. for asympotically almost surely. An event A_n happens a.a.s. if $\mathbb{P}(A_n) \to 1$ as $n \to \infty$.

Theorem 4.7.2. *For $(s-1)q(\ln n + c) \leq d \leq (s-1)qe^{sq}(1-\varepsilon)$, a graph from $ROC(n, d, s, q)$ a.a.s. has at most $(1+o(1))\frac{e^{-c}}{1-\varepsilon}$ isolated vertices.*

Proof. We begin by computing the probability a vertex is isolated,

$$\begin{aligned}
\mathbb{P}(v \text{ is isolated}) &= \sum_{i=0}^{\frac{nd}{s^2q}} \mathbb{P}(v \text{ is in } i \text{ communities})(1-q)^{si} \\
&= (1+o(1)) \sum_{i=1}^{\frac{nd}{s^2q}} \binom{\frac{nd}{s(s-1)q}}{i} \left(\frac{s}{n}\right)^i \left(1 - \frac{s}{n}\right)^{\frac{nd}{s(s-1)q} - i} e^{-sqi} \\
&\leq (1+o(1))\, e^{-\frac{d}{(s-1)q}} \sum_{i=0}^{\frac{nd}{s^2q}} \left(\frac{de^{-sq+\frac{s}{n}}}{(s-1)q}\right)^i \\
&= (1+o(1))\, e^{-\frac{d}{(s-1)q}} \sum_{i=1}^{\frac{nd}{s^2q}} \left(\frac{de^{-sq}}{(s-1)q}\right)^i \\
&= (1+o(1)) \left(e^{-\frac{d}{(s-1)q}}\right) \left(\frac{1}{1-\varepsilon}\right).
\end{aligned}$$

Let X be a random variable that represents the number of isolated vertices of a graph

drawn from $ROC(n, d, s, q)$. We compute

$$\mathbb{P}(X > 0) \leq \mathbb{E}(X) = (1 + o(1))\, n \left(e^{-\frac{d}{(s-1)q}}\right) \left(\frac{1}{1-\varepsilon}\right) = (1 + o(1)) \left(\frac{e^{-c}}{1-\varepsilon}\right).$$

□

Theorem 4.7.3. *A graph from $ROC(n, d, s, q)$ with $s = o(\sqrt{n})$ has no isolated communities a.a.s. if*

$$\frac{d}{q} > \log \frac{nd}{s^2 q}.$$

Proof. We construct a "community graph " and apply the classic result that $G(n, p)$ will a.a.s. have no isolated vertices when $p > (1 + \epsilon) \log n/n$ for any $\epsilon > 0$ [ER59]. In the "community graph " each vertex is a community and there is an edge between two communities if they share at least one vertex; a ROC graph has no isolated communities if and only if the corresponding "community graph " is connected. The probability two communities don't share a vertex is $\left(1 - \left(\frac{s}{n}\right)^2\right)^n$. Since communities are selected independently, the "community graph " is an instance of $G\left(\frac{nd}{s(s-1)q}, \left(1 - \left(\frac{s}{n}\right)^2\right)^n\right)$. By the classic result, approximating the parameters by $\frac{nd}{s^2 q}, 1 - e^{s^2/n}$, this graph is connected when

$$1 - e^{-s^2/n} > \frac{\log \frac{nd}{s^2 q}}{\frac{nd}{s^2 q}}.$$

Since $s = o(\sqrt{n})$ is small, the left side of the inequality is approximately s^2/n, yielding the equivalent statement

$$\frac{d}{q} > \log \frac{nd}{s^2 q}.$$

□

Note that the threshold for isolated vertices is higher, meaning that if a ROC graph a.a.s has no isolated vertices, then it a.a.s has no isolated communities. These two properties

together imply the graph is connected.

CHAPTER 5
SAMPLING FROM SPARSE GRAPHS WITH OVERLAPPING COMMUNITIES AND HETEROGENEOUS DEGREES

This section describes joint work with Christian Borgs, Jennifer Chayes, Souvik Dhara, and Subhabrata Sen.

5.1 Introduction

Large networks are ubiquitous in modern scientific applications. Empirical studies on their structural properties reveal that real-world networks exhibit characteristic traits such as sparsity, heavy-tailed degree distribution, and community structure. The study of simple models for these real networks, as well as formal inference regarding latent characteristics based on network data, has witnessed rapid growth in Statistics and Machine Learning in recent years. In this context, it is often assumed that the practitioner has access to the entire network of interest. Unfortunately, this is often invalid in numerous practical applications since computational limitations or privacy considerations may restrict the data available for scientific inquiry. In this context, it is often reasonable to assume that one observes only a small sub-sample of the underlying network. Statistical inference regarding the properties of the underlying graph, based on these sub-samples, poses an outstanding challenge. Some preliminary investigations have been done in this regard [KW18b, KW18a], yet, a lot remains to be understood.

Motivated by this challenge, in this paper, we study a specific notion of vertex sampling, referred to as p-sampling (see Definition 5.1.1 below). This notion has its roots in the theory of sparse graph limits [VR16a, Bor+17a], and it is now understood that this sampling procedure

often furnishes non-trivial information regarding the structure of the underlying graph. In this paper, we exhibit the power of p-samples through two complimentary results. First, we use p-samples to establish direct conceptual connections among various network models with an overlapping community structure, proposed in the recent literature (see Corollary 5.1.2 and Theorem 5.2.7). Second, we demonstrate the power of p-samples in detecting the presence of a latent community structure in the underlying graph (see Theorem 5.1.3).

Networks with an underlying community structure have attracted significant attention recently. In this context, the question of recovering the latent characteristics of canonical models like the Stochastic Block Model (SBM) has inspired an explosion of activity in recent years. We refer the interested reader to [Abb17b] for a survey of recent breakthroughs in this domain. In parallel, it has also been recognized that simple latent communities might be too simplistic for many applications. For instance, individuals are usually members of multiple overlapping communities in social networks and proteins may be involved with multiple cellular processes. This has prompted the introduction of simple random graph models with an overlapping community structure [Air+06, BKN11a], and diverse methods for fitting these models [Air+14] have been developed.

In this paper, we study properties of p-samples drawn from graphs with an overlapping community structure. To present our results in its most general form, we find it particularly convenient to introduce a new random graph, which we call the *Community Configuration Model* (CCM). In addition to latent overlapping communities, this model is particularly flexible, and can fit graphs with given degrees. The CCM is a generalization of [Kry19] because it allows for both bipartite and non-bipartite community structure. In Section 5.3, we see how this model has natural connections to other popular models proposed in the prior Statistics and Machine Learning literature.

The CCM generalizes the configuration model (CM), classically studied in the combinatorics literature [BC78, Bol80]. The model is determined by two parameters: D a sequence of

vectors containing colored half-edge counts, and a matching M that describes how to pair the half-edges by color. Let $D = (d_{v_1}, \dots d_{v_n})$ be a sequence of degree vectors $d_v = (d_v^1, d_v^2, \dots d_v^k)$ where d_v^i denotes the number of color i half-edges at vertex v . We define a matching on the set of colors in which each color in $i \in [k]$ is either matched to itself or precisely one other color $j \in [k] \setminus \{i\}$. Let M be the $k \times k$ permutation matrix that describes the matching (i.e. $M_{ij} = M_{ji} = 1$ if and only if color i is matched to color j). We require that $\sum_v d_v^i = \sum_v d_v^j$ for all $i \neq j$ such that $M_{ij} = 1$, and $\sum_v d_v^i$ is even for all i such that $M_{ii} = 1$. We construct $G \sim CCM(D, M)$ as follows. For each $i < j$ with $M_{ij} = 1$, we sequentially select a half-edge of color i, and pair it uniformly at random with an unpaired half-edge of color j, and continue until all the half-edges of color i and j are exhausted. The paired half-edges constitute the edges of G.

The traditional SBM, the mixed membership models of [Air+06, BKN11a], as well as the CCM introduced above, construct graphs on a fixed number of vertices. In stark contrast, motivated by the work of Caron and Fox [CF17], Herlau et al. [HSM16] have recently introduced graph models with a latent community structure based on exchangeable point process. We refer to this model as the HSM model for random graphs henceforth in the paper. The HSM model generates random graphs with community structure and varied degree distributions. Given $T > 0$, $k \geq 1$, and a measure ρ on $(0, \infty)$, we generate a Poisson point process on $\{1, \cdots, k\} \times (0, \infty)$ with intensity $T \cdot \lambda \times \rho$, where λ denotes the counting measure on $\{1, \cdots, k\}$. Given a realization of the point process $\{(c_i, v_i) : i \geq 1\}$, we generate a graph with the Poisson points as the vertices. A symmetric function $f(c_k, c_\ell)$ describes the affinity between communities. We add edges independently for each pair of vertices, such that the distribution of the number of edges between vertex i and j is $\text{Pois}(f(c_i, c_j)v_i v_j)$. Finally, we throw away the isolated vertices. We denote this random graph $HSM(\rho, f, T)$.

The HSM model, at first glance, looks very different compared to the finite network models such as CCM introduced above. Our results, stated below, establish direct conceptual

links between these classes of models, using the notion of p-sampling for graphs.

5.1.1 The convergence of p-samples from CCM graphs

We start by delineating the notion of p-sampling for graphs.

Definition 5.1.1 (p-sampling)**.** The p-sampled subgraph of a multigraph G, denoted $\mathrm{Smpl}(G, p)$, is an unlabeled random graph obtained by sampling each vertex independently with probability p, retaining the induced edges on the chosen vertices, and deleting the isolated vertices in the induced subgraph.

At a high level, we establish that appropriate p-samples of CCM graphs converge weakly to HSM-like models. The formal statement requires the full power of *sampling convergence*, a notion of graph convergence introduced recently in [Bor+17a]. We introduce this notion and state our general result in Section 5.2.2. For the curious reader, we present here a corollary of our general result, relating p-samples of certain CCMs to the HSM model introduced above.

To this end, we first introduce a sequence of multidimensional degree measures, which play a central role in our subsequent analysis. Let $D = (D_n)_{n\geq 1}$ denote a sequence of degree vector sequences, in which the n^{th} degree vector sequence contains $N(n)$ degree vectors of length k,

$$D_n = \left(\left(d_v^1, d_v^2, \dots d_v^k\right)\right)_{v=1}^{N(n)}.$$

We let

$$\ell_n^i = \sum_{v=1}^{N(n)} d_v^i \quad \text{and} \quad \ell_n = \sum_{i=1}^{k} \ell_n^i.$$

We say D_n is a valid sequence with respect to the matching M, if

$$\begin{cases} \ell_n^i \text{ is even} & \text{for all } i \in [k] \text{ s.t. } M_{ii} = 1 \\ \ell_n^i = \ell_n^j & \text{for all } i, j \in [k] \text{ s.t. } M_{ij} = 1. \end{cases} \tag{5.1}$$

Throughout this paper, when we refer to $CCM(D_n, M)$, we implicitly assume that D_n is a valid sequence with respect to the matching M and that $d_i \in \mathbb{Z}^+$ for all degrees. We define the k-dimensional measure

$$\rho_n = \frac{1}{\sqrt{\ell_n}} \sum_{v=1}^{N(n)} \delta_{\tilde{d}_v} \quad \text{where} \quad \tilde{d}_v = \left(\frac{d_v^1}{\sqrt{\ell_n^1}}, \ldots \frac{d_v^k}{\sqrt{\ell_n^k}} \right).$$

The following corollary of Theorem 5.2.7 describes a family of configuration models whose p-samples converge in distribution to an HSM graph process with k communities.

Corollary 5.1.2. *Let $f : [k]^2 \to \mathbb{R}^+$ be a symmetric function and ρ^* be a measure on $(0, \infty)$. For $s \in [k]$, let C^s be the length $k(k+1)/2$ affinity vector for community s, which is indexed by pairs $i, j \in [k]$ with $i \leq j$. Let*

$$C_{ij}^s = \begin{cases} \sqrt{f(s,j)} & i = s \\ 0 & i \neq s. \end{cases}$$

Let ρ be the $k(k+1)/2$ dimensional measure such that for any bounded continuous function $\phi : \mathbb{R}^{k(k+1)/2} \to \mathbb{R}$ with compact support,

$$\int \phi d\rho = \frac{1}{k} \sum_{s=1}^{k} \int \phi(xC^s) d\rho^*(x).$$

Let $(D_n)_{n \geq 1}$ be a sequence of degree vector sequences of the form $D_n = \left(d_v^1, d_v^2, \ldots, d_v^{k(k+1)/2} \right)_{v=1}^{N(n)}$, and let ρ_n be the corresponding degree measure. Let $G_n \sim CCM(D_n, I)$, where I denotes the $k(k+1)/2 \times k(k+1)/2$ identity matrix. Assume that $\rho_n \to \rho$ vaguely on $(\mathbb{R}^+)^{k(k+1)/2} \setminus \{0\}$ and for all $i \in [k(k+1)/2]$

$$\lim_{n \to \infty} \int_{(\mathbb{R}^+)^{k(k+1)/2}} x_i \wedge 1 d\rho_n = \int_{(\mathbb{R}^+)^{k(k+1)/2}} x_i \wedge 1 d\rho.$$

Then for any $T \in \mathbb{R}^+$, almost surely,

$$\mathrm{Smpl}(G_n, T/\sqrt{2e(G_n)}) \xrightarrow{d} HSM(\rho^*, f, T)$$

where $e(G)$ denotes the number of edges G.

An analogous connection was developed in the prior work [Bor+], where the Caron-Fox model was realized as the p-sample from a configuration model random graph. Our result allows a re-interpretation of the HSM model— it arises naturally under vertex sub-sampling from specific CCM models, and the measure ρ arises naturally from the degree distribution of the underlying graph.

The family of CCMs whose p-samples converge to the HSM model are restricted in two ways: (i) there is no bipartite community structure and (ii) there is no correlation between x_i, the predictor of total degree, and community membership. General degree measures for CCMs can encode overlapping community structure and exhibit correlations between community membership and degree, for instance by requiring that all half-edges corresponding to a particular community are on high degree vertices. In Theorem 5.2.7, we show that the p-samples for any convergent sequence of CCMs converge to a graphex, a more general HSM-like object first introduced in [VR15].

5.1.2 Community detection under p-sampling

Given the interpretation of the HSM as p-samples of CCM, it is natural to turn our attention to non-asymptotic properties of p-sampling. Here, we initiate a study into the following question:

Do p-samples capture the presence of an underlying community structure?

To formalize this question, we fix $t > 0$ and the degree sequence $D^* = \left(\left(d_i^r, d_i^b\right)\right)_{i=1}^n$. We define $d_i = d_i^r + d_i^b$, and set $D = (d_i)_{i=1}^n$. Setting $\ell = \sum_i d_i$, $\ell^r = \sum_i d_i^r$, $\ell^b = \sum_i d_i^b$, we define

the parameter

$$\beta(D^*) = \frac{1}{2\ell} \sum_{i<j} \left(\frac{d_i^r d_j^r}{\ell^r - 1} + \frac{d_i^b d_j^b}{\ell^b - 1} - \frac{d_i d_j}{\ell - 1} \right)^2 .$$

The non-centrality parameter β measures the prominence of the community structure— the larger it is, the easier it is to distinguish the two hypotheses.

We observe a graph $G^s \sim \mathrm{Smpl}_t(G)$, and for any $\varepsilon > 0$, consider the following hypothesis testing problem:

$$H_0 : G \sim CM(D) \quad \text{vs.} \quad H_1 : G \sim CCM(D^*), \beta(D^*) \geq \varepsilon.$$

For any multigraph G, let d_i denote the degree of vertex i, $\ell/2$ the number of edges, and X_{ij} the number of multiedges between i and j. Define the statistic

$$S(G) = \ell^2 \sum_{i<j} \left(X_{ij} - \frac{d_i d_j}{\ell} \right)^2 . \tag{5.2}$$

This statistic is intimately related to the modularity statistic of Newman and Girvan [NG04]. For the hypothesis testing problem under consideration, we define $\tilde{\ell} = 2|E(G^s)|$ and reject the null whenever $S(G^s) \leq \tilde{\ell}^3/2 + \varepsilon\tilde{\ell}^3$ where $\tilde{\ell}$ is twice the number of non-loop edges in G^s. The following theorem provides non-asymptotic bounds on the error probabilities of this statistical procedure.

Theorem 5.1.3. *Let $D = (d_i)$ be a degree sequence, and let $D^* = \left(\left(d_i^r, d_i^b\right)\right)$ be a corresponding colored sequence, so $d_i = d_i^r + d_i^b$. Let $\ell = \sum d_i$, $\ell^r = \sum d_i^r$, and $\ell^b = \sum d_i^b$. Assume $\beta(D^*) \geq \varepsilon$. Let λ be a parameter satisfying $\lambda \geq 64/\varepsilon + 64$, $\lambda = o\left(\ell^{1/4}\sqrt{\varepsilon}\right)$, $\lambda = \Omega\left(\sqrt{\log \ell}\right)$, and $d_i \leq \lambda\sqrt{\ell}$, $d_i^r \leq \lambda\sqrt{\ell_r}$, $d_i^b \leq \lambda\sqrt{\ell_b}$ for all vertices. Let $G^s \sim \mathrm{Smpl}_t(G)$ for $t = \Omega(\lambda^3)$. There exists*

constants $c, \bar{c}$ such that the hypothesis test on G^s is correct with probability at least

$$1 - O\left(\exp\left(-\frac{1}{32}\sqrt{\frac{t}{\lambda^3}}\right) + \exp\left(-\bar{c}\varepsilon\sqrt{\frac{t}{\lambda^3}}\right) + \exp\left(\frac{-c\ell\varepsilon^2}{\lambda^4}\right) + \frac{\ell^2}{\lambda}\exp\left(-2e^2\lambda^2\right)\right).$$

At a high-level, the theorem implies that in case $t \gg \lambda^3$, t-samples can consistently distinguish these hypotheses. As a point of comparison, Theorem 5.7.1 describes the hypothesis test given access to the whole graph rather than just one sample. The proof of Theorem 5.1.3 involves the following steps.

- Show that $S(G)$ concentrates around $\ell^3/2 + 2\beta\ell^3$ if $G \sim CM$, while it concentrates around $\ell^3/2$ if $G \sim CCM$.
- Show $\mathbb{E}(S(G^s)) \approx t^6/\ell^3 S(G)$. To this end, we express $S(G)$ as a polynomial in multigraph counts, and quantify the effect of p-sampling on S.
- Show that $S(G^s)$ concentrates around its mean. Since the terms of multigraph counts in the polynomial $S(G)$ do not correspond to multigraphs on the same number of vertices, these terms scale differently under p-sampling. To show concentration, we must use the Kim-Vu [KV00] concentration inequalities for arbitrary polynomials of indicator variables.
- Show $\mathbb{E}\left(\tilde{\ell}\right) \approx t^2$ and use the Kim-Vu inequality to show concentration.

These steps imply that $S(G^s)$ concentrates around $\tilde{\ell}^3/2$ when $G \sim CM$ and around $\tilde{\ell}^3/2 + 2\beta\tilde{\ell}^3$ when $G \sim CCM$.

The second step of the analysis outlined above, relating $\mathbb{E}(S(G^s))$ to $S(G)$, is extremely general, and holds for deterministic graphs G, and for any S which is a polynomial in the multigraph counts. We collect here a very general statement, in anticipation of its future usefulness in other analogous contexts.

Definition 5.1.4. We define $H(F, G)$ as the number edge-labeled homomorphisms of F in G. Let Ψ the set of injective maps $\psi : V(F) \to V(G)$, and X_{ij} be the indicator function for $\{i, j\} \in E(G)$. Then

$$H(F, G) = \sum_{\psi \in \Psi} \prod_{\{i,j\} \in E(F)} X_{\psi(i)\psi(j)}.$$

Lemma 5.1.5. *Let G be a graph with maximum degree $\lambda\sqrt{\ell}$ such that there are at most λ^2 edges between any pair of vertices, where $\ell = 2E(G)$. Let $S : G \to \mathbb{R}$ be the linear combination of injective homomorphism counts,*

$$S(G) = \sum_{i=1}^{k} \alpha_i H(F_i, G),$$

where for each i, $\alpha_i \in \mathbb{R}^+$ and F_i is a fixed graph on v_i vertices and e edges with no isolated vertices. For $G^s \sim \mathrm{Smpl}_t(G)$ with $t \geq \lambda$ and $m \leq \lambda^2$, there exists a constant c such that

$$\left(\frac{t}{\sqrt{\ell}}\right)^{v_1} S(G) \leq \mathbb{E}(S(G^s)) \leq \left(\frac{t}{\sqrt{\ell}}\right)^{v_1} S(G) + c\lambda^{2e-v_2} t^{v_2},$$

where $v_1 = \max_{i \in [k]} v_i$ and $v_2 = \max_{i \in [k] \,|\, v_i \neq v_1} v_i$.

In Lemma 5.7.11 we give concentration results for $H(F, G^s)$ using the Kim-Vu concentration inequality (Lemma 5.8.12).

5.1.3 Organization

In Section 5.2 we describe the limit of general sequences of CCMs. In Section 5.2.1 we give the requisite formal definitions of sampling convergence and graphexes, and then in Section 5.2.2 we state our main theorem describing the graphex limit of CCMs (Theorem 5.2.7). In Section 5.3, we compare the CCM to popular models in the Statistical and Machine Learning literature and, in doing so, establish that the CCM encompasses a broad class of these models. In Section 5.4 we describe the accuracy of hypothesis testing p-samples for a class of power

law graphs. The proofs of the sampling convergence and hypothesis testing results are given in Section 5.6 and Section 5.7 respectively. Moreover, in Section 5.7.1 we describe the accuracy of the hypothesis for community structure given access to the whole graph rather than a p-sample.

5.2 Sampling convergence, graphexes, and the limit of CCMs

5.2.1 An overview of sampling convergence and graphexes

In this section, we give a brief introduction to multigraphexes and describe how they are the limiting objects for sampling convergent graph sequences. This overview is based on the detailed account given in [Bor+].

A multigraphex $\mathcal{W}$ is a triple of functions $\mathcal{W} = (W, S, I)$ that give instructions for how to construct a family of random graphs. The multigraphon function W is defined over an arbitrary feature space $\Omega \times \Omega$ equipped with a measure. To produce the associated random graph, vertices are placed on Ω according to a Poisson point process. Then for each pair of vertices, edges are added according to W. We define the sequence space $\ell_1 = \{(x_i)_{i\geq 1} \mid x_i \in \mathbb{R}_+ \forall i, \sum_{i=0}^{\infty} x_i < \infty\}$. The multi-star function $S : \Omega \to \ell_1$ governs how we add star edges between vertices established by the multigraphon and new vertices. The dust function $I \in \ell_1$ produces isolated edges on new vertices. In order to guarantee that this process almost surely produces a finite graph, we impose integrability conditions on W, S, and I and delete isolated vertices.

Definition 5.2.1 (Multigraphex)**.** A multigraphex is a triple $\mathcal{W} = (W, S, I)$ such that $I \in \ell_1$, $S : \Omega \mapsto \ell_1$ is a measurable function, and $W : \Omega^2 \times \mathbb{N}_0 \mapsto \mathbb{R}_+$ is a measurable function satisfying $W(x, y, k) = W(y, x, k)$, $\sum_{k=0}^{\infty} W(x, y, k) = 1$, for any $x, y \in \Omega$ and $k \in \mathbb{N}_0$. We will assume throughout that, $\min\{\sum_{k\geq 1} S(\cdot, k), 1\}$ is integrable. Further, setting $\mu_W(\cdot) = \int (1 - W(\cdot, y, 0))\mathrm{d}y$, we assume that

- $\Lambda(\{x : \mu_W(x) = \infty\}) = 0$ and $\Lambda(\{x : \mu_W(x) > 1\}) < \infty$,
- $\int (1 - W(x, y, 0))\mathbb{1}\{\mu_W(x) \leq 1\}\mathbb{1}\{\mu_W(y) \leq 1\}\mathrm{d}y\mathrm{d}x < \infty$,
- $\int (1 - W(x, x, 0))\mathrm{d}x < \infty$.

Each of the three multigraphex functions may produce multiedges; we now describe the simple graph analog of each. A graphon W is the special case of a multigraphon in which $W(x, y, k) = 0$ for $k \geq 2$. In this case, the graphon can be described as a function $W : \Omega^2 \to [0, 1]$ where $W(x, y) = W(x, y, 1)$. Similarly, when the star function is simple $S(x, k) = 0$ for $k \geq 2$ and all $x \in \mathbb{R}_+$, and so we write $S : \Omega \mapsto \mathbb{R}_+$, where $S(x) = S(x, 1)$. Last, in the case where $I(k) = 0$ for $k \geq 2$, we represent the simple dust function $I = I(1)$ as a constant. When W, S, and I are all simple, the multigraphex is a graphex[Jan17, Bor+18] and the random graphs produced are simple.

In this paper, we exclusively consider cases where W is a multigraphon, but I and S are simple. Whenever we specify a star function $S : \mathbb{R}_+ \mapsto \mathbb{R}_+$ or an isolated edge constant I, we assume that these describe a star function or an edge sequence with $S(x, k)$ or $I(k) = 0$ for $k \geq 2$.

Informally, the *multigraphex process* $\mathrm{GP}_t(\mathcal{W})$ describes how to sample from a multigraphex $\mathcal{W}$ to obtain a graph of a desired size (approximately). The parameter t tunes the intensity of the point process on Ω (used to create the vertices associated with the multigraphex) and therefore determines the expected number of edges in the random graph.

Definition 5.2.2. Let $\mathfrak{G}_f$ denote the space of finite graphs. The multigraphex process $\mathrm{GP}_t(\mathcal{W})$ is $\mathfrak{G}_f$-valued stochastic process obtained as follows. Consider a single Poisson process $\{v_j\}_{j\geq 1}$ of rate t on $\mathbb{R}_+$, and add edges according to the following procedure:

▷ for $i \neq j$, connect v_i and v_j with n_{ij} edges, where $\mathsf{P}(n_{ij} = r) = W(v_i, v_j, r)$;

▷ for each j, add n_j self-loops to v_j, where $\mathsf{P}(n_j = r) = W(v_j, v_j, r)$;

▷ for each j add a multi-star to v_j by adding edges of multiplicity r at a rate $tS(v_j, r)$;

▷ add isolated edges of multiplicity r with rate $t^2 I(r)$.

Discard all isolated vertices (as well as all labels), and output the resulting unlabeled graph.

Finally, we define sampling convergence of a sequence of multigraphs to a multigraphex. For techincal reasons, throughout this paper we let $e(G)$ denote the number of edges in G that are not self-loops.

Definition 5.2.3 (Convergence to multigraphex)**.** A sequence of (multi)graphs $(G_n)_{n\geq 1}$ is said to converge to a (multi)graphex $\mathcal{W}$ if $(\mathrm{Smpl}(G_n, t/\sqrt{2e(G_n)}))_{n\geq 1}$ converges in distribution to $\mathrm{GP}_t(\mathcal{W})$ for all $t > 0$.

Weak convergence in the space of graphs finite graphs equipped with the discrete topology $\mathfrak{G}_f$ is equivalent to convergence in total variation distance.

5.2.2 Main result: the graphex limit of CCMs

In this section we define convergence for a sequence of community configuration models and write down the corresponding graphex limit. Recall the $CCM(D_n, M)$ random graph defined in the introduction.

Definition 5.2.4. Let $D = (D_n)_{n\geq 1}$ be a sequence of degree vector sequences. Define the k-dimensional CCM measure

$$\rho_n = \frac{1}{\sqrt{\ell_n}} \sum_{v=1}^{N(n)} \delta_{\tilde{d}_v} \quad \text{where} \quad \tilde{d}_v = \left(\frac{d_v^1}{\sqrt{\ell_n^1}}, \dots \frac{d_v^k}{\sqrt{\ell_n^k}} \right).$$

Definition 5.2.5. The sequence of k-dimensional measures (μ_n) converges as a degree measure to (μ, a) if

- $\mu_n \to \mu$ vaguely on $(\mathbb{R}^+)^k \setminus \{0\}$

- For $i \in [k]$, $\lim_{n\to\infty} \int_{(\mathbb{R}^+)^k \setminus \{0\}} x_i \wedge 1 d\mu_n = b_i$
- $a = b - \int h(x) d\mu$ where $b = (b_1, \dots b_k)$ and $h(x)_i = x_i \wedge 1$.

Definition 5.2.6. For a k-dimensional measure μ, $k \times k$ matrix X, and k-dimensional vector a, we define the bilinear graphex $\mathscr{W}(\mu, a, X) = (W, S, I)$. Let the feature space be $\mathbb{R}_+^k$ equipped with the measure μ. Let $p(k; \lambda)$ denote the probability that a Poisson λ random variable takes value k. For $x, y \in \mathbb{R}_+^k$, define

$$W(x, y, k) = \begin{cases} p(k; \mu(x)^T X \mu(y)) & x \neq y \\ p(k; \mu(x)^T X \mu(x)/2) & x = y \end{cases}$$

$$S(x) = a^T X \mu(x), \quad I = a^T X a/2.$$

Note that different triples (μ, a, X) may yield the same bilinear graphex , i.e. $\mathscr{W}(\mu', a', X') = \mathscr{W}(\mu, a, X)$ for $(\mu, a, X) \neq (\mu', a', X')$. See Section 5.5 for further discussion.

Theorem 5.2.7. *Let $D = (D_n)_{n\geq 1}$ be a sequence of degree sequences with $\ell_n = \Omega(\log(n))$. Suppose the corresponding CCM measure (ρ_n) converges as a degree measure to (ρ, a). Let $(G_n)_{n\geq 1}$ be a sequence of graphs $G_n \sim CCM(D_n, M)$. Then almost surely $(G_n)_{n\geq 1}$ converges to the graphex $\mathscr{W}(\rho, a, M)$.*

Proof strategy

We show that the multigraphex is the limit object of sampling convergent graph sequences by showing convergence of the corresponding random adjacency measures.

Random adjacency measures. The background on random adjacency measures presented here is a summary of the more detailed exposition of [Bor+].

Definition 5.2.8 (Random adjacency measure)**.** An adjacency measure ξ is a locally finite measure such that $\xi(A \times B) = \xi(B \times A)$ for all $A, B \in \mathscr{B}(\mathbb{R}_+)$ and $\xi \in \mathcal{N}(\mathbb{R}_+^2)$. A random

adjacency measure is ξ is a $\mathcal{N}(\mathbb{R}_+^2)$ valued random variable such that ξ is almost surely an adjacency measure.

Observe that without loss of generality, any random adjacency measure ξ may be expressed as

$$\xi = \sum_{ij} \beta_{ij} \delta_{(\alpha_i, \alpha_j)}$$

for $\beta_{ij} \in \mathbb{N}_0$. Given an adjacency measure ξ, one can obtain an unlabeled graph $\mathcal{G}(\xi)$ as follows. Define a countable vertex set where vertex i is labeled by α_i. We obtain $\mathcal{G}(\xi)$ by constructing β_{ij} many edges between the vertices labeled α_i and α_j, deleting the isolated vertices, and erasing the vertex labels.

The function $\mathcal{G}(\cdot)$ will allow us to interpolate between the random adjacency measure of the graphex (defined below) and the multigraphex process $GP_t(\mathcal{W})$. (the $\mathfrak{G}_f$ valued stochastic process given in Definition 5.2.2).

Definition 5.2.9 (Random adjacency measure of a multigraphex)**.** Given any multigraphex $\mathcal{W} = (W, S, I)$, define $\xi_{\mathcal{W}}$, the *random adjacency measure generated by* $\mathcal{W}$ as follows:

$$\begin{aligned}
\xi_{\mathcal{W}} = &\sum_{i \neq j} \zeta_{ij} \delta_{(\theta_i, \theta_j)} + \sum_i \zeta_{ii} \delta_{(\theta_i, \theta_i)} + \sum_{j,k} g(\theta_j, \chi_{jk}) \big(\delta_{(\theta_j, \sigma_{jk})} + \delta_{(\sigma_{jk}, \theta_j)} \big) \\
&+ \sum_k h(\eta_k'') \big(\delta_{(\eta_k, \eta_k')} + \delta_{(\eta_k', \eta_k)} \big), \\
\zeta_{ij} = r, \quad &\text{if } \sum_{l=0}^{r-1} W(v_i, v_j, l) \leq U_{\{i,j\}} \leq \sum_{l=0}^{r} W(v_i, v_j, l), \\
g(\theta_j, \chi_{jk}) = r, \quad &\text{if } \sum_{l=0}^{r-1} S(v_j, l) \leq \chi_{jk} \leq \sum_{l=0}^{r} S(v_j, l), \\
h(\eta_k'') = r, \quad &\text{if } \sum_{l=0}^{r-1} I(l) \leq \eta_k'' \leq \sum_{l=0}^{r} I(l).
\end{aligned}$$

where $(U_{\{i,j\}})_{i,j \geq 1}$ is a collection of independent uniform[0,1] random variables, $\{(\theta_j, v_j)\}_{j \geq 1}$, $\{(\chi_{jk}, \sigma_{jk})\}_{k \geq 1}$ for all $j \geq 1$ are unit rate Poisson point processes on $\mathbb{R}_+^2$, and $(\eta_k, \eta_k', \eta_k'')_{k \geq 1}$

is a unit rate Poisson point processes on $\mathbb{R}_+^3$, where all the above Poisson point processes are independent of each other and $(U_{\{i,j\}})_{i,j\geq 1}$.

The adjacency measure $\xi_{\mathcal{W}}$ associated with the multigraphex $\mathcal{W} = (W, S, I)$ naturally defines a $\mathfrak{G}_f$ valued stochastic process by considering the corresponding unlabeled graphs; this is precisely the multigraphex process $\mathrm{GP}_t(\mathcal{W})$ defined in Definition 5.2.2. For a point process ξ, let us denote by $\xi|_A$ the measure ξ restricted to A.

Definition 5.2.10 (Multigraphex process, equivalent definition)**.** For any given multigraphex $\mathcal{W}$ we define the *multigraphex process generated by* $\mathcal{W}$ as the $\mathfrak{G}_f$-valued stochastic process $(\mathrm{GP}_t(\mathcal{W}))_{t\geq 0}$ where $\mathrm{GP}_t(\mathcal{W}) = \mathcal{G}(\xi_{\mathcal{W}}|_{[0,t]^2})$.

Next we define the random adjacency measure of a multigraph (Definition 5.2.11), and then we show that sampling convergence of a graph sequence is equivalent to convergence of these random adjacency measures.

Definition 5.2.11 (Random labeling adjacency measure for multigraphs)**.** Let M be a distribution over finite graphs. The random s-labeling adjacency measure $\mathrm{Lbl}_s(M)$, is generated as follows. Draw a graph G from M. For each vertex v, independently choose a value U_v uniformly from $[0, s]$ and associate the label U_v to the vertex. Then $\mathrm{Lbl}_s(M) := \sum_{v,w\in \mathrm{V}(G)} n_{vw}\delta_{(U_v,U_w)}$, where n_{vw} denotes the number of edges between vertices v and w in G. The canonical labeling of M, denoted by $\mathrm{Lbl}(M)$, refers to the case $s = \sqrt{2E(G)}$.

We also define the random labeling adjacency measure for random multigraphs, which we will need for the proof of Theorem 5.2.7. Note that $\mathrm{Lbl}(M)$ is not equal to the random adjacency measure formed by selecting $G \sim M$ and then choosing an adjacency measure from $\mathrm{Lbl}(G)$ unless all graphs drawn from M have the same number of edges.

Definition 5.2.12 (Random labeling adjacency measure for random multigraphs)**.** Let M be a distribution over finite multigraphs. The random s-labeling adjacency measure $\mathrm{Lbl}_s(M)$,

is generated as follows. Draw a graph G from M. For each vertex v, independently choose a value U_v uniformly from $[0, s]$ and associate the label U_v to the vertex. Then $\mathrm{Lbl}_s(M) := \sum_{v,w\in \mathrm{V}(G)} n_{vw}\delta_{(U_v,U_w)}$, where n_{vw} denotes the number of edges between vertices v and w in G. The canonical labeling of M, denoted by $\mathrm{Lbl}(M)$, refers to the case $s = \sqrt{2\,\mathbb{E}(E(G))}$.

Finally Proposition 5.2.13 establishes that the distributional limit of the random adjacency measures of a sequence of graphs is the random adjacency measure of the multigraphex limit.

Proposition 5.2.13. *Consider a sequence of multigraphs $(G_n)_{n\geq 1}$ with $e(G_n) < \infty$ for all $n \geq 1$ and $\lim_{n\to\infty} e(G_n) = \infty$. Then the following are equivalent:*

- *$(G_n)_{n\geq 1}$ is sampling convergent.*
- *$(\mathrm{Lbl}(G_n))_{n\geq 1}$ converges in distribution as random variables in $\mathcal{N}(\mathbb{R}^2_+)$.*

Moreover, if the distributional limits of $(\mathrm{Smpl}(G_n, r/\sqrt{2e(G_n)}))_{n\geq 1}$ and $(\mathrm{Lbl}(G_n))_{n\geq 1}$ are given by H_r and ξ, then $\mathrm{Lbl}_r(H_r) \overset{d}{=} \xi|_{[0,r)^2}$. Further, ξ is extremal. Therefore, there exists a multigraphex $\mathcal{W}$ (non-random) such that $\xi \overset{d}{=} \xi_{\mathcal{W}}$, and $(G_n)_{n\geq 1}$ is sampling convergent to $\mathcal{W}$.

Proof outline for Theorem 5.2.7. By Proposition 5.2.13, it suffices to show that with high probability $\mathrm{Lbl}\,(G_n) \to \xi_{\mathcal{W}}$ where $G_n \sim CCM(D_n, M)$ and $\xi_{\mathcal{W}}$ is random adjacency measure given by the graphex $\mathcal{W} = \mathscr{W}(\rho, a, M)$, as in Definition 5.2.9. We show this via the following:

$$\mathrm{Lbl}\,(G_n) \overset{\text{Close whp by } \textit{Lemma } 5.6.7}{\longleftrightarrow} \mathrm{Lbl}\,(CCM_n) \overset{\text{Close by } \textit{Lemma } 5.6.8}{\longleftrightarrow} \xi^P_{\rho_n,M} \overset{\text{Converges by } \textit{Lemma } 5.6.4 \text{ } \textit{Fact } 5.8.7}{\longrightarrow} \xi^P_{\rho,M} \overset{\text{Equal by } \textit{Lemma } 5.6.5}{\longleftrightarrow} \xi_{\mathcal{W}},$$

where we denote $CCM_n = CCM(D_n, M)$.

The random adjacency measure $\xi^P_{\rho_n,M}$ is a variant of the random labeling adjacency measure $\mathrm{Lbl}\,(CCM(D_n, M))$ in which the number of edges between each pair of vertices is determined independently by a Poisson rather than by a matching. To formalize this, we

define the k-dimensional random measure S_n that describes the degree distribution of the vertices of $CCM(D_n, M)$ with respect to the following random vertex labeling. Label each vertex with an independent uniform value from $[0, \sqrt{\ell_n}]$. For $A \subset \mathbb{R}^+$, let

$$S_n(A) = \sum_v \tilde{d}_v \mathbb{1}\{v \in A\}. \tag{5.3}$$

We define

$$\xi^P_{\rho_n, M}(A \times B) \sim \begin{cases} \text{Pois}(S_n(A)^T M S_n(B)) & A \neq B \\ \text{Pois}(S_n(A)^T M S_n(A)/2) & A = B \end{cases}. \tag{5.4}$$

Next we define the completely random measure ω that will describe the degree measure arising from the limiting graphex. Let $\{(\sigma_i, x_i)\}$ be drawn from a Poisson point process with mean intensity $d\lambda \times d\rho$. Let

$$\omega(A) = a\lambda(A) + \sum x_i \mathbb{1}\{\sigma_i \in A\}. \tag{5.5}$$

Finally, we define

$$\xi^P_{\rho, M}(A \times B) \sim \begin{cases} \text{Pois}(\omega(A)^T M \omega(B)) & A \neq B \\ \text{Pois}(\omega(A)^T M \omega(A)/2) & A = B \end{cases}, \tag{5.6}$$

analogous to $\xi^P_{\rho_n, M}$ introduced above.

5.3 Connections to established models with overlapping community structure

Random graph models with an overlapping community structure have been studied extensively in Statistics and Machine Learning in recent years. In this Section, we look at some natural alternative graph models with heterogeneous degrees and overlapping community structure,

proposed in prior literature. Upon viewing them through the lens of sampling convergence, we discover some connections among these models.

5.3.1 The mixed membership configuration model MMCM

In a seminal paper, [Air+06] introduced the Mixed Membership Stochastic Block Model (MMSBM), now a canonical model for networks with overlapping communities. In this model, each vertex is labeled with probability distribution over features (or communities), and each pair of features is assigned an affinity (a value in $(0, 1)$) that describes how likely it is for vertices with those two features to be connected. A MMSBM graph is created by starting with the complete graph and deleting edges as follows. For each edge (u, v), u and v draw a feature according to their respective probability distributions. Then the edge (u, v) remains with probability equal to the affinity between these two selected features and is discarded otherwise. Traditionally, the probability distributions and affinities are constants independent of the number of vertices. Therefore, since the algorithm begins with a complete graph and the probability of deleting edges is constant, the resultant MMSBM graph is dense ($\Theta(n^2)$ edges). The model may be extended to produce sparse graphs by choosing affinities that decay to zero as the number of vertices n diverges. We note that while this model naturally produces graphs with overlapping community structure, the practitioner does not have a precise control on the degree sequence of the graph produced.

To address this issue, we introduce the mixed membership configuration model (MMCM), which combines the properties of MMSBM with those of the Configuration Model (CM). In particular, the MMSBM allows for more freedom to model a variety of sparse degree distributions. The MMCM pairs colored half-edges according to a CM (without regard to the colors of the half edges at this step), and then retains each edge independently according to the the affinity of the colored half-edges forming this edge. Formally, fix $k \geq 1$, and let $D = (d_{v_1}, \dots d_{v_n})$ be a sequence of degree vectors $d_v = (d_v^1, d_v^2, \dots d_v^k)$ where d_v^i denotes the

number of color i half-edges at vertex v. Let B be a symmetric $k \times k$ matrix with entries in $[0, 1]$ describing the affinities between communities. Construct a graph by pairing up the half-edges as in a usual configuration model (ignoring the colors). Then for each edge, keep the edge with probability B_{ij} where i and j are the colors of the half-edges and delete the edge otherwise. Let $MMCM(D, B)$ denote the probability distribution over all graphs constructed by this process.

We note that both the CCM and the MMCM produce sparse graphs with given degrees and overlapping communities, albeit using slightly different procedures. It is thus natural to expect that if the parameters in these models are matched appropriately, sub-samples of these random graphs should approximately "look the same," and so the limits of convergent sequences of MMCM graphs are also bilinear graphexes. Our next result formalizes this heuristic; we describe the bilinear graphex limit of convergent sequences of MMCM graphs. Then we show how to appropriately match MMCM and CCM parameters so that the corresponding graphex limits are the same.

In order to describe the bilinear graphex limit of MMCM sequences, we must define a MMCM degree measure, analogous to Definition 5.2.4.

Definition 5.3.1. Let $D = (D_n)_{n\geq 1}$ be a sequence of length k degree vector sequences for the MMCM with $k \times k$ matrix B. Let ℓ_n be the number of half-edges (before deletion). Let $e_n = \mathbb{E}(|E(G_n)|)$ for $G_n \sim MMCM(D_n, B)$. Define the k-dimensional MMCM measure

$$\gamma_n = \frac{1}{e_n} \sum_{v=1}^{N(n)} \delta_{\hat{d}_v} \quad \text{where} \quad \hat{d}_v = \left(\frac{d_v^1}{\sqrt{\ell_n}}, \dots \frac{d_v^k}{\sqrt{\ell_n}} \right).$$

Theorem 5.3.2. *Let (D_n, B) be MMCM parameters with $\ell_n = \Omega(\log(n))$. Suppose the corresponding MMCM measure γ_n converges as a degree measure to (γ, a). Let $(H_n)_{n\geq 1}$ be a sequence of graphs $H_n \sim MMCM(D_n, B)$. Then almost surely $(H_n)_{n\geq 1}$ converges to the bilinear graphex $\mathscr{W}(\gamma, a, B)$.*

Next we show that given a CCM or MMCM, it is easy to construct a corresponding MMCM or CCM respectively that yields a graphs with similar p-samples.

Definition 5.3.3. Let D be a k-dimensional degree sequence on n vertices for an CCM with matching matrix M. Let $\Gamma_{C\to M}(D, M) = (\tilde{D}, B)$ be the corresponding MMCM parameters obtained as following. With respect to D, define $\ell^i = \sum_{v=1}^n d_v^i$ and $\ell = \sum_{i=1}^k \ell^i$. Let the affinity matrix $B = M$ and $\tilde{D}$ be the k-dimensional degree sequence on n vertices with

$$\bar{d}_v^i = \frac{d_v^i \left(\sum_{i=1}^k \sqrt{\ell^i}\right)}{\sqrt{\ell^i}}.$$

Definition 5.3.4. Let D be a k-dimensional degree sequence on n vertices for an MMCM with affinity matrix B. Let $\Gamma_{M\to C}(D, M) = (\tilde{D}, M)$ be the corresponding CCM parameters with k^2 colors obtained as following. Index the colors by pairs ij, for $i, j \in [k]$. Let M be the matching such that color ij matches to color ji. With respect to D, let $\ell^i = \sum_{v=1}^n d_v^i$ and $\ell = \sum_{i=1}^k d_v^i$. Let $\tilde{D}$ be the k-dimensional degree sequence on n vertices with

$$\bar{d}_v^{ij} = \frac{B_{ij}\ell^i d_v^i}{\ell}.$$

Note that it is possible that these definitions yield a corresponding model in which the degree vectors contain fractional half-edge counts. While it does not make sense to construct a configuration model in this context, it is possible to define CCM and MMCM measures ρ and γ for degree distributions with fractional half-edge counts. We show that under the correspondences $\Gamma_{C\to M}$ and $\Gamma_{M\to C}$, the random adjacency measures $\xi_{\rho,M}^P$ and $\xi_{\gamma,B}^P$ are the same (Lemmas 5.8.16 and 5.8.17). These are the random adjacency measures arising from the variants of the models in which edges are constructed via independent Poissons rather through a configuration model pairing procedure, see Equations (5.4) and (5.15). Lemmas 5.8.16 and 5.8.17 and Theorems 5.2.7 and 5.3.2, directly imply the following corollaries.

Corollary 5.3.5. *Let (D_n, M) be CCM parameters, and let $(\tilde{D}, B) = \Gamma_{C\to M}(D_n, M)$ be corresponding MMCM parameters. Let $G_n \sim CCM(D_n, M)$ and $H_n \sim MMCM(\tilde{D}, B)$. Then (G_n) converges almost surely if and only if and (H_n) converges almost surely. Moreover, when the sequences converge they have the same bilinear graphex limit.*

Corollary 5.3.6. *Let (D_n, B) be MMCM parameters and let $(\tilde{D}, M) = \Gamma_{M\to C}(D_n, B)$ be corresponding CCM parameters. Let $G_n \sim CCM(D_n, M)$ and $H_n \sim MMCM(\tilde{D}, B)$. Then (G_n) converges almost surely if and only if and (H_n) converges almost surely. Moreover, when the sequences converge they have the same bilinear graphex limit.*

We imagine that there exists a bijection between families of CCM and MMCM degree measures such that each pair has a unique bilinear graphex limit. However, even understanding the class of CCM measures that yield the same bilinear graphex is beyond the scope of this paper, see Section 5.5 for further discussion.

5.3.2 The Ball-Karrer-Newman (BKN) model

The BKN model, introduced in [BKN11b], is specified by a set of vertices, each with a vector of features $(\theta_{u1}, \theta_{u2}, \dots \theta_{uk})$ where θ_{uj} measures the affiliation of vertex u with community j. The number of edges of color ℓ between vertex u and w is given by $\text{Pois}(\theta_{u\ell}\theta_{w\ell})$ if $u \neq w$ and $\text{Pois}(\theta_{u\ell}\theta_{u\ell}/2)$ if $u = w$. Several variants of the BKN model have been studied in the literature; [KN11] discusses a similar model without mixed membership, and [COL09] is the simple graph version of this model.

Intuitively, the BKN model is a variant of the CCM in which the number of edges between each pair of vertices is given by an independent Poisson, instead of being paired according to a configuration model. Given a BKN model on k colors, we can describe the corresponding

CCM on k colors with the identity matching and

$$d_u^j = \theta_{uj} \sum_v \theta_{vj}.$$

Lemma 5.6.1 implies that in this CCM, for any two fixed vertices u, w, the number of color j edges joining them is approximately

$$\text{Pois}\left(\frac{d_u^j d_w^j}{\sum_v d_v^j}\right) = \text{Pois}\left(\theta_{uj}\theta_{wj}\right).$$

This clarifies the correspondence between these models. The proof ideas of Theorem 5.2.7 can be adapted easily to be applicable in this context. Indeed, if we have $\{(\theta_{u1}, \cdots, \theta_{uk}) : 1 \le u \le n\}$ such that the corresponding CCM degrees $\{d_u^j : 1 \le j \le k, 1 \le u \le n\}$ satisfy the hypotheses of Theorem 5.2.7, this sequence of BKN random graphs will also converge to the same graphex limit almost surely.

5.3.3 Todeschini-Miscouridou-Caron (TMC) model

The TMC model [TMC16b] uses the framework of exchangeable point processes, analogous to the HSM model [HSM16], and produces simple graphs with an overlapping community structure. The model is specified by two parameters: a constant $T > 0$ and a measure ρ on $\mathbb{R}_+^k$ where k denotes the number of communities. Given T and ρ, we sample a Poisson process $\{(w_{i1}, \cdots, w_{ik}) : i \ge 1\}$ on $\mathbb{R}^k$ with intensity $T\mathrm{d}\rho$. Each $i \ge 1$ now corresponds to a node in a graph, where w_{ij} measures the affiliation of vertex i with community j. For each pair of distinct vertices i, j, an edge is added independently with probability $1-\mathsf{exp}\left(-2\sum_{k=1}^p w_{ik}w_{jk}\right)$. Finally, the isolated vertices are discarded. The TMC model extends the HSM model [HSM16] by allowing the vertices to have overlapping community assignments.

An equivalent description of the TMC model is as follows. Sample $\{(w_{i1}, \cdots, w_{ik}) : i \ge 1\}$ as a Poisson process on $\mathbb{R}^k$ with intensity $T\mathrm{d}\rho$. Next, for each pair of distinct vertices i, j, add

Pois($\sum_k w_{ik}w_{jk}$) multi-edges independently. We call this graph the Multi-TMC model. The usual TMC model is obtained by collapsing the multigraph— one replaces every multi-edge by a simple edge. Note that the Multi-TMC model is very similar to the BKN model, albeit constructed using point processes.

Analogous to Corollary 5.1.2, Theorem 5.2.7 allows us to relate the Multi-TMC model to p-samples from an appropriate CCM. In particular, consider a CCM (D_n, M) with $M = I$, and D_n converging to $(\rho, 0)$. Theorem 5.2.7 immediately implies that

$$\text{Smpl}(CCM(D_n, I), T/\sqrt{2E(CCM(D_n, I))})$$

converges weakly a.s. to the Multi-TMC model. Finally, this establishes that the TMC model can arise as a result of two sequential operations— draw a p-sample from an appropriate CCM, and then replace every multi-edge in the graph with a simple graph.

5.4 Hypothesis testing on power law CCMs

We now illustrate the behavior of testing for the presence of communities based on samples when the degree distribution follows the quantiles of a truncated power-law with parameter between one and two. To choose the value of the parameter ε in Theorem 5.1.3, we need to compute the value of the parameter β.

Let F be the cumulative distribution function on $\mathbb{N}$ such that $(1-F)(x) = c_F x^{-(\tau-1)}$, where $\tau \in (1, 2)$. Let $F^{-1}(y) = \inf\{x : F(x) \leq y\}$. We consider the case when $d_i = (1 - F)^{-1}(i/n)$, i.e., choose according to the quantiles of the distribution. Next we truncate the distribution by dropping vertices with degrees more than M, and let $V = \{i : d_i > M\}$. If $\sum_{i \in V} d_i$ is odd, we add a dummy vertex of degree to make the sum of degrees even. We define D^* the degree distribution with community structure as follows. If i is odd, α fraction of the d_i half-edges are red and $(1 - \alpha)$ fraction of the d_i half-edges are blue. If i is even, α fraction of the d_i

half-edges are blue and $(1-\alpha)$ fraction of the d_i half-edges are red. In the case where αd_i is not integer, we randomly select the color of the final half-edge with probability $\alpha d_i - \lfloor \alpha d_i \rfloor$. Throughout, we write $a_n \sim b_n$ to denote that $\lim_{n\to\infty} a_n/b_n = 1$, and write c as a generic notation for a constant whose value can be different between expressions. Assuming $\alpha \neq 1/2$, we compute the bias parameter

$$\begin{aligned}
\beta(D^*) &= \frac{1}{2\ell} \sum_{i<j} \left(\frac{d_i^r d_j^r}{\ell^r - 1} + \frac{d_i^b d_i^b}{\ell^b - 1} - \frac{d_i d_j}{\ell} \right)^2 \\
&\sim \frac{c}{2\ell} \left(\sum_{\substack{i<j \\ i=j \bmod 2}} \frac{\left(2\left(\alpha^2 + (1-\alpha)^2\right) - 1\right)^2 d_i^2 d_j^2}{\ell^2} + \sum_{\substack{i<j \\ i\neq j \bmod 2}} \frac{\left(4\alpha(1-\alpha) - 1\right)^2 d_i^2 d_j^2}{\ell^2} \right) \\
&\sim \frac{c\left(4\alpha^2 - 4\alpha + 1\right)^2}{\ell^3} \sum_{i<j} d_i^2 d_j^2. \qquad (5.7)
\end{aligned}$$

Next we approximate $\sum_{i<j} d_i^2 d_j^2/\ell^3$. Note that smaller i corresponds to larger degrees; for any $\varepsilon_n \to 0$, $d_i \sim (c_F n/i)^{1/(\tau-1)}$ uniformly over $i \leq \varepsilon_n n$. Note also that $d_i \leq M$ implies that $i \geq cnM^{-(\tau-1)}$. Therefore

$$\ell = \sum_{i\in V} d_i = \sum_{\varepsilon_n n \leq i \leq n} d_i + \sum_{cnM^{-(\tau-1)} \leq i < \varepsilon n} d_i \sim cn + cn^{1/(\tau-1)} \sum_{i \geq cnM^{-(\tau-1)}} i^{-1/(\tau-1)} = cnM^{2-\tau}$$

$$\sum_{i\in V} d_i^2 \sim cn + cn^{2/(\tau-1)} \sum_{i \geq cnM^{-(\tau-1)}} i^{-2/(\tau-1)} = nM^{3-\tau}.$$

We take $M = n^{1/\tau}$ and apply Theorem 5.1.3 with $\lambda = n^{1/\tau - 1/2}$. Since $\tau \in (1,2)$, we have $\lambda \to \infty$. We compute,

$$\frac{\left(\sum_{i\in V} d_i^2\right)^2}{\ell^3} = \frac{M^\tau}{n} = \Theta(1).$$

Therefore by 5.7,

$$\beta(D^*) = \left(4\alpha^2 - 4\alpha + 1\right)^2 \Theta(1). \tag{5.8}$$

Thus, the bound in Theorem 5.1.3 reduces to

$$1 - O\left(\exp\left(-\frac{1}{32}\sqrt{\frac{t}{\lambda^3}} \right) + \exp\left(-cn^{2(\tau-1)/\tau}\right) + \frac{\ell^2}{\lambda} \exp\left(-2e^2 n^{-1+2/\tau}\right) \right).$$

5.5 Discussion

To summarize, our work investigates p-sampling in the context of graphs with arbitrary degree distributions and overlapping communities. To this end, we introduced the community configuration model, a canonical random graph exhibiting the desired characteristics, and derived our results in this setting. Our results are two-fold:

(1) We derive sufficient conditions for p-samples of a sequence of CCM graphs to converge in distribution almost surely and describe the corresponding graphex limit object. This result establishes a new connection between finite network models (such as the CCM, SBM, BKN models) and random graphs generated by exchangeable point processes (graphex, Caron-Fox, HSM); the latter arise as the p-samples from finite network models.

(2) We formulate a hypothesis test for CCM graphs and establish that one p-sample from a graph is sufficient to detect whether the graph exhibits overlapping community structure. This results indicates that (sufficiently large) p-samples of a graph indeed retain non-trivial structural information about the graph.

We conclude by collecting here some natural follow up questions arising from our investigations, and a few broad interesting directions for future research enquiry.

Necessary conditions for convergence and identifiability. Theorem 5.2.7 gives sufficient conditions on a sequence of degree measures that guarantee the almost sure convergence of the corresponding CCM graph sequences. It is natural to wonder whether these conditions are also necessary. Indeed, in the special case of the configuration model ($k = 1$), [Bor+, Theorem 1.1] establishes that the sufficient conditions of Theorem 5.2.7 are, in fact, necessary for sampling convergence. However, a close scrutiny of the proof of Theorem 5.2.7 suggests that our conditions are potentially sub-optimal for $k > 1$. To see this, recall the proof strategy for Theorem 5.2.7 described earlier. The most critical step in our proof establishes that $\xi^P_{\rho_n,M} \to \xi^P_{\rho,M}$ in distribution. Our assumptions facilitate this step of the argument; however, it is possible to establish this under weaker assumptions.

A major conceptual hurdle in making progress in this direction arises from our lack of understanding of identifiability for CCM graphex processes. Note that for $k > 1$ there might exist $\rho \neq \rho'$ such that $\xi^P_{\rho,M} \stackrel{d}{=} \xi_{\rho',M}$. Even in the special case of identity matching, the question is currently intractable. In fact, for a CCM with two colors and the identity matching, there are many different degree measures that yield the same graphex limit. Trivially, the graphex corresponding to the limit of the CCM in which all edges are red is the same as if the CCM had all edges blue. More generally, given a one-dimensional degree measure (one color), let ρ_α be the degree measure corresponding to coloring α fraction of the half-edges red and $(1 - \alpha)$ fraction of the half-edges blue at each vertex. For any $\alpha, \alpha' \in [0, 1]$, the corresponding graphexes are the same, i.e. $\mathscr{W}(\rho_\alpha, a, I) = \mathscr{W}(\rho_{\alpha'}, a, I)$. In this case, there is measure preserving bijection between ρ_α and $\rho_{\alpha'}$ that is a rotation about the origin in $\mathbb{R}^2$. We conjecture that when the degree measure is two dimensional, $\mathscr{W}(\rho_\alpha, a, I) = \mathscr{W}(\rho_{\alpha'}, a, I)$ if and only if ρ_α and $\rho_{\alpha'}$ can be obtained from one another by a rotation. Resolving the question of identifiability in this setting might provide valuable clues regarding the optimal convergence criteria alluded to above. We believe that characterizing these conditions for sampling convergence in Theorem 5.2.7 would be extremely interesting, and leave this for

future research.

General hypothesis testing and information theoretic limits. The hypothesis test based on $S(\cdot)$ considered in Theorem 5.1.3 requires that the size of the sample grow with ℓ; the sampling parameter $t = \Omega((\log \ell)^{3/2} + \Delta^3)$ where $\Delta\sqrt{\ell^i}$ is an upper bound on the color i degree of a vertex. It would be intriguing to have information theoretic lower bounds for this problem, i.e., how large must t be to achieve accuracy $1 - o_\ell(1)$ for CCMs on two colors with the identity matching? What are the corresponding information theoretic limits if the desired accuracy is at least $1/2 + \varepsilon$ for some fixed $\varepsilon > 0$? We conjecture that there exists an algorithm that determines whether $G \sim CM$ or $G \sim CCM$ given only access to $G^s \sim \mathrm{Smpl}(G, t)$ such that $\lim_{t\to\infty} \lim_{\ell\to\infty} \mathbb{P}(\text{ error }) = 0$.

In a different direction, our test statistic $S(\cdot)$ is designed for CCMs with two colors and the identity matching. A natural extension would be to look at the case of CCMs on k colors and arbitrary matching matrices. More generally, given a p-sample from $\mathscr{W}_0$ or $\mathscr{W}_1$, when is it possible to consistently distinguish between these two alternatives? We leave these questions for future research.

Property testing via p-samples Our hypothesis testing result demonstrates that the community structure of CCM graphs is maintained under p-sampling. What other properties can be inferred by studying one p-sample?

More generally, property testing on graphs has been studied extensively (see [Lov12]). In the context of dense graphs, the goal is to determine with high probability whether a graph G satisfies property $\mathcal{P}$ or is ε-far from satisfying property $\mathcal{P}$, meaning it is not possible to obtain a graph satisfying property $\mathcal{P}$ by adding or deleting at εn^2 edges. A property is said to be testible if there exists an algorithm whose run time depends only on ε that determines with accuracy at least $2/3$ whether G satisfies $\mathcal{P}$ or is ε-far from satisfying $\mathcal{P}$. A graph property P testible if and only if knowing a regularity partition of the graph suffices

to determine whether G is close to satisfying $\mathcal{P}$ (Theorem 2 of [Alo+09]). An outstanding direction for future research concerns the generalization of these ideas to the sparse graph setting. Specifically, can we characterize the testible properties given access to one p-sample? The Szemerédi Regularity lemma forms the cornerstone of property testing on dense graphs—it would indeed be fascinating if extensions are possible in the sparse setting, based on the graphex regularity lemma recently established in [Bor+18].

5.5.1 A CCM "regularity" lemma

We imagine that community configuration models can approximate sparse graphs, analogous to how stochastic block models approximate dense graphs. Instead of using the cut metric or homomorphism densities as a metric, we consider sampling distributions, a generalization of homomorphism densities introduced in [**Bor17**]. Let $\mathrm{Smpl}_t(G)$ be the random graph as defined above. For a random graph model M, let $\mathrm{Smpl}_t(M)$ be the graph obtained by selecting G from M, then returning $\mathrm{Smpl}_t(G)$. In order to model simple graphs, we define the erased community configuration model (ECCM) as a CCM in which all self-loops are deleted and multi-edges are condensed into an edge.

Conjecture 5.5.1. *Let $t > 0$ and $\varepsilon > 0$. For any graph G, there exists an erased community configuration model ECCM on $k = f(t, \varepsilon)$ colors such that*

$$d_{TV}\left(\mathcal{L}\left(\mathrm{Smpl}_t(G)\right), \mathcal{L}\left(\mathrm{Smpl}_t(ECCM)\right)\right) \leq \varepsilon.$$

If true, the above conjecture would be a configuration model "regularity" lemma. Like the traditional regularity lemma, the number of groups needed is a constant depending only on the error parameter (and not the size of the graph). However, the total size of the description for the approximating ECCM would grow linearly with the size of the graph since it is necessary to store the degree distribution. Such a lemma could give fast approximation

algorithms and improve our understanding of property testing in sparse graphs.

5.6 Sampling convergence: proofs

In this section we prove Theorem 5.2.7 and lay the foundation for the proof of Theorem 5.3.2, which describe the bilinear graphex limits of CCM and MMCM sequences respectively. We begin by introducing lemmas that describe how edges pair and functions concentrate in the configuration model and its variants (Section 5.6.1). Next we develop a general framework that allows us to prove that the random adjacency measures based on the CCM and MMCM degree measures converge to the appropriate bilinear graphexes (Section 5.6.2). The proofs of Theorems 5.2.7 and 5.3.2 are similar; the former is given in Section 5.6.3 and latter in Section 5.8.1.

5.6.1 Pairing and switching lemmas

The Poisson Pairing Lemma essentially says that distribution of the number of edges between formed between sets of half-edges under a configuration model is similar in total variation distance to the corresponding distribution when the number of edges between each pair of vertices i, j is determined by an independent Poisson with intensity $d_i d_j/\ell$.

Lemma 5.6.1 (Poisson Pairing)**.** *Let $S_1, \dots S_k$ be disjoint sets of half edges of some configuration model $CM(d)$ where $\ell = \sum_i d_i$. Let $s_i = |S_i|$ and $s = \sum_{i=1}^k s_i$. Let $E, \bar{E}$ be random variables whose values are vectors of length $(k^2 + k)/2$ indexed by pairs ij where $i \leq j$, $i, j \in [k]$. For $G \sim CM(d)$, define $\bar{E}(G)$ as follows*

$$\bar{E}(G)_{ij} = \#\textit{edges between } S_i \textit{ and } S_j \textit{ in } G.$$

Let $\mathcal{L}(\bar{E})$ be the distribution of $\bar{E}(G)$ when $G \sim CM(d)$. Let $\mathcal{L}(E)$ be the distribution over

vectors E where each entry takes values independently as follows

$$E_{ij} \sim \begin{cases} \text{Pois}\left(\frac{s_i s_j}{\ell}\right) & i \neq j \\ \text{Pois}\left(\frac{s_i(s_i-1)}{2\ell}\right) & i = j. \end{cases}$$

Then

$$d_{TV}\left(\mathcal{L}\left(\bar{E}\right), \mathcal{L}\left(E\right)\right) \leq \frac{5s^3}{(\ell-2s)^2} + \frac{2\ell s}{(\ell-s)^2}\left(\frac{1}{2} + log\left(\frac{3s^2}{\ell-s}\right)\right).$$

Proof. Consider a labeling of the half-edges in $S_1, S_2 \ldots S_k$ from 1 to s in which the half-edges in S_i appear before the half-edges in S_j for $i < j$. Let β_j be the number of half-edges that appear before the half-edges in S_j, so

$$\beta_j = \sum_{i=1}^{j-1} s_i.$$

Consider a sequential pairing of the half-edges where at step t if half-edge t is not yet paired, we pair it with another unpaired half-edge chosen uniformly at random. We define a sequence of $((k^2+k)/2)$-dimensional unit vectors (I_t), in which I_t describes how the t^{th} half-edge is paired. For ease of notation we describe the coordinates of the vectors as pairs ij where $i \leq j$, $i, j \in [k]$. Let $I_t = e_{ij}$ represent the event that the half-edge t was not paired previously, is in S_i, and pairs to a half-edge in S_j in step t. Let $I_t = 0$ represent the event the half-edge t was previously paired or paired to a half-edge in $\bigcap_{i=1}^{k} S_i^c$ in step t. Note $\bar{E}(G) = \sum_{t=1}^{s} I_t$, so $\mathcal{L}\left(\sum I_t\right) \stackrel{d}{=} \mathcal{L}\left(\bar{E}\right)$.

To approximate $\sum I_t$, we construct a sequence of independent Poisson random variables $(\hat{I}_t)$, define a coupling with (I_t), and apply Stein's method.

Coupling. To construct the sequence $\hat{I}_t$, we use the following algorithm that sequentially pairs half-edges with some replacement. We maintain sets of half-edges $S_1, S_2, \ldots S_k$ and

$\bigcap_{i=1}^{k} S_i^c$. At step t, we pair the half-edge t (call it e_t) with a uniformly chosen half-edge f_t from these sets, and then replace f_t in the corresponding set $S_1, S_2, \ldots S_k$ and $\bigcap_{i=1}^{k} S_i^c$. We do not replace e_t. Set $\hat{I}_t = e_{ij}$ if the half-edge t is in S_i and is paired with a half-edge in S_j, and zero otherwise. We say the original copy of each half-edge is "non-bad." A copy of half-edge is declared "bad" if any previous copy of the edge was paired to a non-bad edge or if the edge was selected as r_t as described in the third case below. We couple (I_t) and $(\hat{I}_t)$ as follows:

- If both e_t and f_t are non-bad, set $I_t = \hat{I}_t$
- If $e_t = t$ is bad, set $I_t = 0$
- If e_t is non-bad and f_t is bad, choose a non-bad half-edge uniformly at random (leaving out $e_1, \ldots, e_t$) and call it r_t. Set $I_t = e_{ij}$ if $e_t \in S_i$ and $r_t \in S_j$, and set $I_t = 0$ otherwise. This is independent of $\hat{I}_t$. Declare r_t bad.

First we show $d_{TV}\left(\mathcal{L}\left(\sum_t I_t\right), \mathcal{L}\left(\sum_t \hat{I}_t\right)\right)$ is small by computing a bound on the probability that $\sum_t I_t$ and $\sum_t \hat{I}_t$ differ. If $I_t \neq \hat{I}_t$, then f_t is bad or e_t is bad.

If e_t is bad at step t, then $e_t = f_{t'}$ or $e_t = r_{t'}$ for some $t' \leq t$. The probability $e_t = f_{t'}$ is at most $1/(\ell - t)$. If $e_t = r_{t'}$ then $f_{t'}$ is bad, which happens with probability at most $t/(\ell - t)$, and e_t was picked as the replacement, which happens with probability at most $1/(\ell - 2t)$. Using the facts that there are at most s previous steps t' and $t \leq s$, we compute

$$\mathbb{P}(e_t \text{ is bad}) \leq s\left(\frac{1}{\ell - s} + \frac{s}{\ell - s}\frac{1}{\ell - 2s}\right) = \frac{s}{\ell - 2s}.$$

Given that e_t is bad, $I_t \neq \hat{I}_t$ only if $\hat{I}_t \neq 0$. Independent of whether e_t is bad, $\mathbb{P}\left(\hat{I}_t \neq 0\right) \leq s/(\ell - t)$. It follows that

$$\mathbb{P}\left(e_t \text{ bad and } I_t \neq \hat{I}_t\right) \leq \left(\frac{s}{\ell - 2s}\right)^2. \tag{5.9}$$

Since at most one new edge is deemed bad at each step, the probability that f_t is bad is at most $\frac{s}{\ell_n - 2s}$. We use the fact that $\hat{I}_t$ and I_t are independent when f_t is bad to compute

$$\begin{aligned}
&\mathbb{P}\Big(f_t \text{ is bad and } \hat{I}_t \neq I_t\Big) \leq \mathbb{P}\Big(f_t \text{ is bad and } (\hat{I}_t \neq 0 \text{ or } I_t \neq 0)\Big) \\
&= \mathbb{P}(f_t \text{ bad }) \Big(1 - \mathbb{P}\Big(\hat{I}_t = 0 \text{ and } I_t = 0 \mid f_t \text{ is bad}\Big)\Big) \\
&= \mathbb{P}(f_t \text{ bad }) \left(1 - \mathbb{P}\Big(\hat{I}_t = 0 \mid f_t \text{ is bad}\Big)\, \mathbb{P}\left(r_t \in \bigcap_{i=1}^{k} S_i^c\right)\right) \\
&\leq \frac{s}{\ell - 2s}\left(1 - \left(1 - \frac{s}{\ell - s}\right)\left(1 - \frac{s}{\ell - 2s}\right)\right) \\
&\leq 2\left(\frac{s}{\ell - 2s}\right)^2 \qquad (5.10)
\end{aligned}$$

Combining Equation (5.9) and Equation (5.10), we obtain

$$\begin{aligned}
d_{TV}\left(\mathcal{L}\left(\sum_{t=1}^{s} \hat{I}_t\right), \mathcal{L}\left(\sum_{t=1}^{s} I_t\right)\right) &\leq \mathbb{P}\left(\sum_{t=1}^{s} \hat{I}_t \neq \sum_{t=1}^{s} I_t\right) \leq \mathbb{P}\Big(\exists t : \hat{I}_t \neq I_t\Big) \\
&\leq \sum_{t=1}^{s} \mathbb{P}\Big(\hat{I}_t \neq I_t\Big) \\
&\leq \frac{3s^3}{(\ell - 2s)^2}. \qquad (5.11)
\end{aligned}$$

Next we apply Stein's method (Lemma 5.8.6) to approximate $\mathcal{L}\left(\sum_t \hat{I}_t\right)$. Let $p(i,j,t) = \mathbb{P}\Big(\hat{I}_t = e_{ij}\Big)$. We compute

$$p(i,j,t) = \begin{cases} \frac{s_j}{\ell - t} & j > i \\ \frac{s_i - (t - \beta_i)}{\ell - t} & i = j \end{cases}$$

Observe

$$\lambda_{ij} = \sum_{t=1}^{s} p(i,j,t) = \begin{cases} \sum_{t=\beta_i+1}^{\beta_i+s_i} \frac{s_j}{\ell - t} \in \left[\frac{s_i s_j}{\ell}, \frac{s_i s_j}{\ell - s}\right] & i \neq j \\ \sum_{t=\beta_i+1}^{\beta_i+s_i} \frac{s_i - (t-\beta_i)}{\ell - t} \in \left[\frac{s_i(s_i-1)}{2\ell}, \frac{s_i(s_i-1)}{2(\ell-s)}\right] & i = j. \end{cases}$$

We compute

$$\sum_{i\le j}\lambda_{ij} \le \sum_{i<j}\frac{s_i s_j}{\ell - s} + \sum_i \frac{s_i^2}{2(\ell-s)} \le \frac{3s\max_i\{s_i\}}{2(\ell-s)} \le \frac{3s^2}{2(\ell-s)},$$

and so

$$c_\lambda = \frac{1}{2} + \log\left(2\sum_{i\le j}\lambda_{ij}\right) \le \frac{1}{2} + \log\left(\frac{3s^2}{\ell-s}\right).$$

Lemma 5.8.6 implies that

$$\begin{aligned} d_{TV}\left(\mathcal{L}\left(\sum_t \hat{I}_t\right), \bigotimes_{i\le j}\mathrm{Pois}\left(\lambda_{ij}\right)\right) &\le \sum_{t=1}^{s} c_\lambda \sum_{i\ge j}\frac{p(i,j,t)^2}{\lambda_{ij}} \\ &= c_\lambda\left(\sum_{i<j}\sum_{t=\beta_i+1}^{\beta_i+s_i}\frac{p(i,j,t)^2}{\lambda_{ij}} + \sum_i\sum_{t=\beta_i+1}^{\beta_i+s_i}\frac{p(i,i,t)^2}{\lambda_i i}\right) \\ &\le c_\lambda\frac{\ell}{(\ell-s)^2}\left(\sum_{i<j}s_j + \sum_i\frac{2s_i-1}{12}\right) \\ &\le \left(\frac{1}{2}+\log\left(\frac{3s^2}{\ell-s}\right)\right)\frac{2\ell s}{(\ell-s)^2}. \end{aligned} \tag{5.12}$$

Recall that for $\lambda < \lambda'$, $X \sim \mathrm{Pois}(\lambda)$, and $Y \sim \mathrm{Pois}(\lambda')$, $d_{TV}(\mathcal{L}(X), \mathcal{Y}) \le \lambda' - \lambda$. It follows by Claim 5.8.3 that

$$\begin{aligned} d_{TV}\left(\bigotimes_{i\le j}\mathrm{Pois}\left(\lambda_{ij}\right), \mathcal{L}\left(E\right)\right) &\le \left(\sum_{i<j}s_i s_j + \sum_i\frac{s_i(s_i-1)}{2}\right)\left(\frac{1}{\ell-s}-\frac{1}{\ell}\right) \\ &\le \frac{3s^2}{2}\left(\frac{s}{\ell(\ell-s)}\right) \le \frac{2s^3}{(\ell-2s)^2} \end{aligned} \tag{5.13}$$

The statement follows from the triangle inequality applied to Equations (5.11) to (5.13). □

Next we prove the Switching Lemma, which establishes concentration under the configuration model for functions on graphs that do not vary much when two edges are switched. Wormwald

proved this lemma in the special case of a d-regular configuration model (see Theorem 2.19 of [Wor+99]); we modify this proof for the more general setting. Let $G \sim G'$ denote that G and G' are graphs that differ only in one "switch" $\{(i,j),(k,\ell)\} = E(G) \setminus E(G')$ and $\{(i,k),(j,\ell)\} = E(G') \setminus E(G)$.

Lemma 5.6.2 (Switching Lemma)**.** *Let $G \sim CCM(D,M)$ where ℓ is the sum of the degrees. Let f be a function on the support of $CCM(D,M)$ such that $|f(G) - f(G')| < b$ when $G \sim G'$. Then*

$$\mathbb{P}\big(\big|f(G) - \mathbb{E}(f(G))\big| > \delta\big) \leq 2\exp\left(\frac{-\delta^2}{\ell b^2}\right).$$

Proof. Let G be a pairing of the half-edges of $CCM(D,M)$ according to the matching rules. Establish a pairing convention where the half-edges are labeled with natural numbers, and in each step the unpaired half-edge with the lowest label is paired to a uniformly random eligible half-edge (according to the matching rule). Let P_0 denote a pairing obtained by following this convention, and let $P_0(m)$ denote the first m edges paired in P_0. Define the Doob martingale

$$Y_m(P_0) = \mathbb{E}(f(G) \mid P_0(m) \subseteq G),$$

so that

$$Y_0(P_0) = \mathbb{E}(f(G)) \quad \text{and} \quad Y_{\ell/2}(P_0) = f(G)$$

when P_0 are the pairings that form G_n. To apply the Azuma-Hoeffding inequality and conclude the lemma, we show that the martingale has differences bounded by b.

For a given P_0 let i be the next half-edge paired after the first m edges are constructed. Let S_j be the set of all pairings that contain $P_0(M) \cup (i,j)$. For any j,k that do not appear in $P_0(m)$ and are the color that half-edge i matches to, there is a one-to-one correspondence between pairings in S_j and pairings in S_k. For $P \in S_j$ there is a $P' \in S_k$ defined by switching $\{(i,j),(k,\ell)\} \leftrightarrow \{(i,k),(j,\ell)\}$ where ℓ is the partner of k in P. Let G and G' be the graphs

associated with P and P'. Since it is equally likely for the pairing to be in any S_k (with k a half-edge that does not appear in $P_0(m)$ and is the color that the half-edge i matches to) and each P and P' is equally likely,

$$|Y_m(P_0) - Y_{m+1}(P_0)| \leq |f(G) - f(G')| \leq b.$$

□

We now provide the analogous lemma for $MMCM$, but leave the proof to Section 5.8.1. Let $G \overset{s}{\sim} G'$ denote that G and G' are graphs that differ only in a subset of a "switch" $E(G) \setminus E(G') \subseteq \{(i,j),(k,\ell)\}$ and $E(G') \setminus E(G) \subseteq \{(i,k),(j,\ell)\}$.

Lemma 5.6.3 (Subset Switching Lemma)**.** *Let $G \sim MMCM(D,B)$ where ℓ is the total degree of D before deletion. Let f be a function on the support of $CCM(D,M)$ such that $|f(G) - f(G')| < b$ when $G \overset{s}{\sim} G'$.*

$$\mathbb{P}\big(\big|f(G) - \mathbb{E}(f(G))\big| > \delta\big) \leq 2\exp\left(\frac{-\delta^2}{\ell b^2}\right).$$

5.6.2 Random adjacency measures based on degree measures

Recall the definitions of S_n, $\xi^P_{\rho_n,M}$, ω, and $\xi^P_{\rho,M}$ given in Equations (5.3) to (5.6). These definitions were built for the CCM framework. Here we define the more general forms of S_n, $\xi^P_{\mu_n,X}$, ω, and $\xi^P_{\mu,X}$ so that we may apply this notation and the corresponding lemmas to the proofs of both Theorems 5.2.7 and 5.3.2.

Let $\mathcal{M}_n$ be a random graph on $N(n)$ vertices that involves D_n, a set of $N(n)$ degree vectors of length k. Let $e_n = \mathbb{E}(|E(G_n)|)$ for $G_n \sim \mathcal{M}_n$. Let μ_n be a measure of the form

$$\mu_n = \frac{1}{\sqrt{2e_n}} \sum_{v=1}^{N(n)} \delta_{d(v)},$$

where $d(v)$ is some function of the corresponding degree sequence D_n.

Next we define a generalized version of S_n for degree measures μ_n of the above form. Label each vertex with an independent uniform value from $[0, \sqrt{2e_n}]$. For $A \subset \mathbb{R}^+$, let

$$S_n(A) = \sum_v d(v) \mathbb{1}\{v \in A\}. \tag{5.14}$$

We define

$$\xi^P_{\mu_n, X}(A \times B) \sim \begin{cases} \text{Pois}(S_n(A)^T X S_n(B)) & A \neq B \\ \text{Pois}(S_n(A)^T X S_n(A)/2) & A = B \end{cases}. \tag{5.15}$$

Assume μ_n converges as a degree measure to μ. Next we define the completely random measure ω that will describe the degree measure in the limiting graphex. Let $\{(\sigma_i, x_i)\}$ be drawn from a Poisson point process with mean intensity $d\lambda \times d\rho$. Let

$$\omega(A) = a\lambda(A) + \sum x_i \mathbb{1}\{\sigma_i \in A\}. \tag{5.16}$$

Finally, we define

$$\xi^P_{\mu, X}(A \times B) \sim \begin{cases} \text{Pois}(\omega(A)^T X \omega(B)) & A \neq B \\ \text{Pois}(\omega(A)^T X \omega(A)/2) & A = B \end{cases}. \tag{5.17}$$

Note that $e_n = \ell_n/2$ for the CCM. Letting $d(v) = \tilde{d}_v$ implies that $\rho_n = \mu_n$, and so the definitions given in Equations (5.3) to (5.6) exactly match the corresponding definitions given in Equations (5.14) to (5.17).

Lemma 5.6.4. *Let S_n and ω be as given in Equations* (5.14) *and* (5.16). *Let $Y = A_1 \sqcup A_2 \cdots \sqcup A_j \subseteq \mathbb{R}^+$. Let S_n^Y be the probability distribution over vectors of length jk of the form $(S_n(A_1) \dots S_n(A_j))$, and ω^Y be the probability distribution over vector of length jk of the form*

$(\omega(A_1)\dots\omega(A_k))$. *Then*

$$S_n^Y \xrightarrow{d} \omega^Y.$$

Proof. We show convergence of the characteristic functions; for any vector $t \in \mathbb{R}^{jk}$,

$$\lim_{n\to\infty} \mathbb{E}\big(\exp\left(\mathrm{i}\langle t, S_n^Y\rangle\right)\big) = \mathbb{E}\big(\exp\left(\mathrm{i}\langle t, \omega^Y\rangle\right)\big).$$

First we claim

$$\mathbb{E}\big(\exp\left(\mathrm{i}\langle t, \omega^Y\rangle\right)\big) = \exp\left(\sum_{\ell=1}^{j} \mathrm{i}\lambda(A_\ell)\langle t_\ell, a\rangle + \lambda(A_\ell)\int \exp\left(\mathrm{i}\langle t_\ell, x\rangle\right) - 1 d\mu\right),$$

where $t = (t_1 \dots t_j)$ and $t_\ell \in \mathbb{R}^k$. Let $X(s) = \omega([0,s))$. Note X is a Lévy process since $X(0) = 0$ almost surely, and X has independent and stationary increments. We compute $\eta(t)$ the Lévy symbol of $X(1)$.

$$\begin{aligned}
\mathbb{E}(\exp\left(\mathrm{i}\langle t, X(1)\rangle\right)) &= \sum_{k=0}^{\infty} \exp\left(\mathrm{i}\langle t, kx + a\rangle\right) \frac{(d\mu(x))^k}{k!} \exp\left(-d\mu(x)\right) \\
&= \exp\left(\mathrm{i}\langle t, a\rangle\right) \sum_{k=0}^{\infty} \frac{\left(\int \exp\left(\mathrm{i}\langle t, x\rangle\right) d\mu(x)\right)^k}{k!} \exp\left(-d\mu(x)\right) \\
&= \exp\left(\mathrm{i}\langle t, a\rangle + \int \exp\left(\mathrm{i}\langle t, x\rangle\right) - 1 d\mu(x)\right),
\end{aligned}$$

so $\eta(t) = \mathrm{i}\langle t, a\rangle + \int \exp\left(\mathrm{i}\langle t, x\rangle\right) - 1 d\mu(x)$. For any Lévy process X, $\mathbb{E}(\exp\left(\mathrm{i}\langle t, X(s)\rangle\right)) = \exp\left(s\eta(t)\right)$ (see [App09]). The claim follows.

Next we compute the characteristic function of S_n^Y

$$\begin{aligned}
\mathbb{E}\big(\exp\left(\mathrm{i}\langle t, S_n^Y\rangle\right)\big) &= \mathbb{E}\left(\exp\left(\mathrm{i}\sum_{\ell=1}^{j}\langle t_\ell, S_n^Y(A_\ell)\rangle\right)\right) \\
&= \mathbb{E}\left(\exp\left(\mathrm{i}\sum_{v}\sum_{\ell=1}^{j}\langle t_\ell, \tilde{d}_v\rangle \mathbb{1}\{v \in A_\ell\}\right)\right)
\end{aligned}$$

$$= \prod_v \left(1 - \frac{\sum_{\ell=1}^{j} \lambda(A_\ell)}{\sqrt{2e_n}} + \sum_{\ell=1}^{j} \frac{\lambda(A_\ell)}{\sqrt{2e_n}} \exp\left(\mathrm{i}\langle t_\ell, \tilde{d}_v\rangle\right)\right)$$
$$= \exp\left((1+o(1))\sum_{\ell=1}^{j} \lambda(A_\ell) \int \exp\left(\mathrm{i}\langle t_\ell, x\rangle\right) - 1 d\mu_n\right).$$

Note

$$\begin{aligned}\lim_{n\to\infty} \int \exp\left(\mathrm{i}\langle t, x\rangle\right) - 1 d\mu_n(x) &= \lim_{n\to\infty} \int \exp\left(\mathrm{i}\langle t, x\rangle\right) - \mathrm{i}\langle t, h(x)\rangle - 1 d\mu_n + \mathrm{i}\int \langle t, h(x)\rangle d\mu_n \\ &= \int \exp\left(\mathrm{i}\langle t, x\rangle\right) - \mathrm{i}\langle t, h(x)\rangle - 1 d\mu + \mathrm{i}\langle t, b\rangle \\ &= \int \exp\left(\mathrm{i}\langle t, x\rangle\right) - 1 d\mu + \mathrm{i}\langle t, a\rangle\end{aligned}$$

The convergence of the left integral follows from Claim 5.8.1 and the convergence of the right integral follows from assumption (2). It follows that

$$\lim_{n\to\infty} \mathbb{E}\left(\exp\left(\mathrm{i}\langle t, S_n^Y\rangle\right)\right) = \exp\left(\sum_{\ell=1}^{j} \mathrm{i}\lambda(A_\ell)\langle t_\ell, a\rangle + \lambda(A_\ell)\int \exp\left(\mathrm{i}\langle t_\ell, x\rangle\right) - 1 d\mu\right),$$

as desired. □

Lemma 5.6.5. *Let $\xi_{\mathcal{W}}$ denote the random adjacency measure associated to the multigraphex $\mathcal{W} = \mathscr{W}(\mu, a, X)$, and $\xi^*_{\mathcal{W}} = \xi_{\mathcal{W}} \mid_{(x,y):y\le x}$. Let ω be as defined in Equation* (5.16)*. For any $A, B \in \mathscr{B}(\mathbb{R}^+)$ with $A \cap B = \emptyset$, conditional on $\{(\theta_i, v_i)\}_{i\ge 1}$*

$$\xi^*_{\mathcal{W}}(A \times A) \sim \mathrm{Pois}(\omega(A)^T X \omega(A)/2)$$

$$\xi_{\mathcal{W}}(A \times B) \sim \mathrm{Pois}(\omega(A)^T X \omega(B)).$$

Proof. Let $\{(\theta_i, v_i)\}_{i\ge 1}$ be a unit rate Poisson process on $\mathbb{R}^2_+$ and set $w_i = \mu(v_i)$, a random point in $(\mathbb{R}^+)^k$ with respect to ρ. Conditionally on $\{(\theta_i, v_i)\}_{i\ge 1}$, Now (5.3) implies that

conditionally on $\{(\theta_i, v_i)\}_{i\geq 1}$,

$$\begin{aligned}
\xi^*_{\mathcal{W}}(A\times A) &= \sum_{i>j} \mathrm{Pois}(w_i^T X w_j)\mathbb{1}\{\theta_i\in A, \theta_j\in A\} + \sum_i \mathrm{Pois}(w_i^T X w_i/2)\mathbb{1}\{\theta_i\in A\} \\
&\quad + \sum_{j,k} \mathbb{1}\{\chi_{jk}\leq a^T X w_j\}\mathbb{1}\{\theta_j\in A, \sigma_{jk}\in A\} + \sum_k \mathbb{1}\{\eta''_k\leq a^2/2\}\mathbb{1}\{\eta_k\in A, \eta'_k\in A\} \\
&= \sum_{i>j} \mathrm{Pois}(w_i^T X w_j)\mathbb{1}\{\theta_i\in A, \theta_j\in A\} + \sum_i \mathrm{Pois}(w_i^T X w_i/2)\mathbb{1}\{\theta_i\in A\} \\
&\qquad + \sum_j \mathrm{Pois}(\Lambda(A) a^T X w_j)\mathbb{1}\{\theta_j\in A\} + \mathrm{Pois}(\Lambda(A)^2 a^T X a/2),
\end{aligned} \tag{5.18}$$

where, by construction, all the $\mathrm{Pois}(\cdot)$ random variables above are mutually independent. Therefore,

$$\begin{aligned}
\xi^*_{\mathcal{W}}(A\times A) &= \mathrm{Pois}\Big(\frac{\Lambda(A)^2 a^T X a}{2} + \Lambda(A)\sum_{i\geq 1} a^T X w_i \mathbb{1}\{\theta_i\in A\} + \frac{1}{2}\Big(\sum_{i\geq 1} w_i^T X w_i \mathbb{1}\{\theta_i\in A\}\Big)\Big) \\
&= \mathrm{Pois}(\omega(A)^T X \omega(A)/2).
\end{aligned} \tag{5.19}$$

Similarly, conditionally on $(w_i, \theta_i)_{i\geq 1}$,

$$\begin{aligned}
\xi_{\mathcal{W}}(A\times B) &= \sum_{i\neq j}\mathrm{Pois}(w_i^T X w_j)\mathbb{1}\{\theta_i\in A,\theta_j\in B\} \\
&\quad+\sum_{j,k}\mathbb{1}\{\chi_{jk}\leq a^T X w_j\}\mathbb{1}\{\theta_j\in A,\sigma_{jk}\in B\}+\sum_{j,k}\mathbb{1}\{\chi_{jk}\leq a^T X w_j\}\mathbb{1}\{\theta_j\in B,\sigma_{jk}\in A\} \\
&\quad+\sum_{k}\mathbb{1}\{\eta''_k\leq a^T X a/2\}\mathbb{1}\{\eta_k\in A,\eta'_k\in B\}+\sum_{k}\mathbb{1}\{\eta''_k\leq a^T X a/2\}\mathbb{1}\{\eta_k\in B,\eta'_k\in A\} \\
&= \sum_{i\neq j}\mathrm{Pois}(w_i^T X w_j)\mathbb{1}\{\theta_i\in A,\theta_j\in B\}+\sum_{j}\mathrm{Pois}(\Lambda(A)a^T X w_j)\mathbb{1}\{\theta_j\in B\} \\
&\qquad+\sum_{j}\mathrm{Pois}(\Lambda(B)a^T X w_j)\mathbb{1}\{\theta_j\in A\}+\mathrm{Pois}(\Lambda(A)\Lambda(B)a^T X a), \\
&= \mathrm{Pois}(\omega(A)^T X\omega(B)).
\end{aligned} \tag{5.20}$$

The stated conditional independence follows by construction. □

The following lemma characterizes weak convergence of random measures in our context. The lemma follows directly from Theorem 11.1.VIII of [DVJ07] after noting that $\mathcal{A}$ is a covering semiring of continuity sets for random adjacency measures arising from graphexes.

Lemma 5.6.6. *Let ξ_n be a sequence of random adjacency measures, and let $\xi_{\mathcal{W}}$ be the random adjacency measure of some graphex $\mathcal{W}$. Let $\mathcal{A}$ be the set of open rectangles in $\mathbb{R}_+^2$ with rational endpoints, $\mathcal{A}=\{(a_1,a_2)\times(a_3,a_4)\,|\,a_1,a_2,a_3,a_4\in\mathbb{Q}\}$. Then ξ_n converges weakly to $\xi_{\mathcal{W}}$ if and only if for any finite family $A_1,\dots A_k\in\mathcal{A}$, the joint distribution $(L_{G_n}(A_1),\dots,L_{G_n}(A_k))$ converges weakly to $(\xi_{\mathcal{W}}(A_1),\dots,\xi_{\mathcal{W}}(A_k))$.*

5.6.3 Proof of Theorem 5.2.7

Lemma 5.6.7. *Let $G_n\sim CCM(D_n,M)$ where ℓ_n is the total degree of D_n. For ease of notation let $\mathrm{Lbl}(G_n)=L_{G_n}$ and $\mathrm{Lbl}(CCM(D_n,M))=L_{CCM}$. Let $A_1,\dots A_k,B_1,\dots B_k\in$*

$\mathscr{B}(\mathbb{R}_+)$, $A = \bigcup_{i=1}^k A_i$, $B = \bigcup_{i=1}^k B_i$, $j \in \mathbb{N}^+$, *and* $\delta > 0$. *Let*

$$P_n(j_1, j_2, \dots j_k) = \mathbb{P}(L_{G_n}(A_1 \times B_1) = j_1 \cap \cdots \cap L_{G_n}(A_k \times B_k) = j_k)$$

$$P_{CCM}(j_1, j_2, \dots j_k) = \mathbb{P}(L_{CCM}(A_1 \times B_1) = j_1 \cap \cdots \cap L_{CCM}(A_k \times B_k) = j_k).$$

$$\mathbb{P}\big(\big|P_n(j_1, j_2, \dots j_k) - P_{CCM}(j_1, j_2, \dots j_k)\big| > \delta\big) \le 2\,\mathsf{exp}\left(\frac{-\delta^2 \ell_n}{16\Lambda(A)^2\Lambda(B)^2}\right).$$

Proof. We apply Lemma 5.6.2 with $f(G) = P_n(j_1, j_2, \dots j_k)$. It suffices to show that $|f(G) - f(G')| \le 4\left(\frac{\Lambda(A)\Lambda(B)}{\ell_n}\right)$ when G and G' differ by a switch $\{(i,j),(k,\ell)\} \leftrightarrow \{(i,k),(j,\ell)\}$. Note that if $L_G(A_r \times B_r)$ and $L_{G'}(A_r \times B_r)$ differ for any $r \in [k]$, then $(v_i, v_j) \in A \times B, (v_k, v_\ell) \in A \times B, (v_i, v_k) \in A \times B$, or $(v_j, v_\ell) \in A \times B$ (where v_i represents the label of the vertex attached to the half-edge i). This happens with probability at most $4\left(\frac{\Lambda(A)\Lambda(B)}{\ell_n}\right)$. □

Lemma 5.6.8. *Let $D = (D_n)_{n\ge1}$ be a sequence of degree sequences in which the sum of the degrees ℓ_n tends to infinity with n. Let $\mathcal{E}_n(S, S')$ denote the number of edges created between the set of half-edges S and S' in the construction of $CCM(D_n, M)$. Consider m disjoint subsets of half-edges $(S_j)_{j\in[m]}$. Let $s_j = (s_j^1, \dots s_j^k)$ the vector where s_j^i denotes the number of half-edges in S_j with color i, let $\bar{s}_j = |S_j|$, and assume $\bar{s}_j = O(\sqrt{\ell_n})$ for $j \in [m]$. Define $(\mathcal{E}_{ij})_{1\le i\le j\le m}$ an independent collection such that*

$$\mathcal{E}_{ij} \sim \begin{cases} \text{Pois}\left(\frac{s_i^T M s_j}{\ell_n}\right) & \text{for } i \ne j \\ \text{Pois}\left(\frac{s_i^T M s_i}{2\ell_n}\right) \end{cases}$$

Let $R_n = \mathcal{L}\left((\mathcal{E}_n(S_i, S_j))_{1\le i\le j\le m}\right)$ and $R = \mathcal{L}\left((\mathcal{E}_{ij})_{1\le i\le j\le m}\right)$ Then as $n \to \infty$, the total variation distance

$$d_{TV}(R_n, R) \to 0. \tag{5.21}$$

Moreover, if the S_j are random disjoint subsets satisfying $\mathbb{E}(\bar{s}_j) = O\left(\sqrt{\ell_n}\right)$ for all $i \in [k]$,

then the error in 5.29 converges to zero in expectation.

Proof. Let S_i^a denote the half-edges in S_i with color a. By an abuse of notation, let M be the set of all pairs (a, b) such that $M_{a,b} = 1$. Note

$$\mathcal{E}_n(S_i, S_j) = \sum_{(a,b)\in M} \mathcal{E}_n\left(S_i^a, S_j^b\right)$$

and

$$\mathcal{E}_{ij} = \sum_{(a,b)\in M} \mathcal{E}_{ij}^{ab} \quad \text{where} \quad \mathcal{E}_{ij}^{ab} = \begin{cases} \text{Pois}\left(\frac{s_i^a s_i^a}{2\ell_n}\right) & i = j, a = b. \\ \text{Pois}\left(\frac{s_i^a s_j^b}{\ell_n}\right) & \text{otherwise.} \end{cases}$$

By Claim 5.8.3,

$$d_{TV}\left(R_n, R\right) \leq \sum_{(a,b)\in M} d_{TV}\left(\mathcal{L}\left((\mathcal{E}_n(S_i^a, S_j^b))_{1\leq i\leq j\leq m}\right), \mathcal{L}\left((\mathcal{E}_{ij}^{ab})_{1\leq i\leq j\leq m}\right)\right).$$

Since the pairing of edges of each color pair in M happens independently, we apply Lemma 5.6.1 to each term in the summand and obtain

$$d_{TV}\left(R_n, R\right) = O\left(\frac{1}{\sqrt{\ell_n}}\right)$$

when $\bar{s}_j = O\left(\sqrt{\ell_n}\right)$ for all j.

Finally we consider the case when S_j are random disjoint subsets satisfying $\mathbb{E}(\bar{s}_j) = O\left(\sqrt{\ell_n}\right)$. By Markov's inequality, $\mathbb{P}\left(\bar{s}_j \geq \ell_n^{5/8}\right) \leq \ell_n^{-1/8}$. It follows that

$$d_{TV}\left(R_n, R\right) \leq \mathbb{P}\left(\exists \bar{s}_j = \Omega\left(\ell_n^{5/8}\right)\right) + \tilde{O}\left(\ell_n^{-1/8}\right) = \tilde{O}\left(\ell_n^{-1/8}\right).$$

□

Finally we prove Theorem 5.2.7.

Proof. (of Theorem 5.2.7). Let $\mathcal{W} = \mathscr{W}(\rho, a, M)$. We show that almost surely $\mathrm{Lbl}(G_n)$ converges weakly to $\xi_{\mathcal{W}}$ and apply Proposition 5.2.13 to conclude the theorem. For ease of notation we denote $\mathrm{Lbl}(G_n) = L_{G_n}$ and $\mathrm{Lbl}(CCM(D_n, M)) = L_{CCM_n}$. By Lemma 5.6.6, it suffices to show that almost surely the joint distribution $\mathcal{L}((L_{G_n}(A_1), \ldots, L_{G_n}(A_k)))$ converges weakly to $\mathcal{L}((\xi_{\mathcal{W}}(A_1), \ldots, \xi_{\mathcal{W}}(A_k)))$ for every finite family $A_1, \ldots A_k \in \mathcal{A}$. Since the set of all such finite families is countable and the countable union of almost sure events is almost sure, it remains to show that the joint distribution converges weakly almost surely for an arbitrary finite family $A_1, \ldots A_k$. We do so via the following claims.

Claim 1: The joint distribution $\mathcal{L}((L_{CCM_n}(A_1), \ldots, L_{CCM_n}(A_k)))$ converges weakly to the distribution $\mathcal{L}((\xi_{\mathcal{W}}(A_1), \ldots, \xi_{\mathcal{W}}(A_k)))$.

Note Lemmas 5.6.4 and 5.6.5 (with $\mu_n = \rho_n$, $\mu = \rho$, $X = M$), and Fact 5.8.7, imply that $\xi^P_{\rho_n, M}$ converges weakly to $\xi_{\mathcal{W}}$, and so by Lemma 5.6.6, $\mathcal{L}\left(\left(\xi^P_{\rho_n,M}(A_1), \ldots, \xi^P_{\rho_n,M}(A_k)\right)\right)$ converges weakly to $\mathcal{L}((\xi_{\mathcal{W}}(A_1), \ldots, \xi_{\mathcal{W}}(A_k)))$. Lemma 5.6.8 implies that

$$d_{TV}\left(\mathcal{L}(L_{CCM_n}(A_1), \ldots, L_{CCM_n}(A_k)), \mathcal{L}\left(\xi^P_{\rho_n,M}(A_1), \ldots, \xi^P_{\rho_n,M}(A_k)\right)\right) \to 0.$$

the claim follows from Fact 5.8.11.

Claim 2: Almost surely

$$d_{TV}(\mathcal{L}((L_{CCM_n}(A_1), \ldots, L_{CCM_n}(A_k))), \mathcal{L}((L_{G_n}(A_1), \ldots, L_{G_n}(A_k)))) \to 0.$$

For ease of notation, let P_n and P_{CCM_n} be as defined in Lemma 5.6.7 for the rectangles $A_1, \ldots A_k$. Lemma 5.6.7 gives

$$\mathbb{P}(|P_n(j_1, \ldots j_k) - P_{CCM_n}(j_1, \ldots j_k)| > \delta) < \mathsf{exp}\left(-c\delta^2 \ell_n\right)$$

where c is a constant depending only on $A_1, \ldots A_k$. Under the assumption that $\ell_n = \omega(\log(n))$,

$$\sum_{n=1}^{\infty} \exp\left(-c\delta^2 \ell_n\right) < \infty,$$

and so the Borel-Cantelli Lemma implies that almost surely there exists n_0 such that for all $n \geq n_0$, $|P_n(j_1, \ldots j_k) - P_{CCM_n}(j_1, \ldots j_k)| < \delta$. There are finitely many combinations of positive integers $(j_1, j_2, \ldots j_k)$ such that $0 \leq j_i \leq \ell_n$, and so almost surely there exists n' such that for all $n \geq n'$, $|P_n(j_1, \ldots j_k) - P_{CCM_n}(j_1, \ldots j_k)| < \delta$ for all $j_1, \ldots j_n$. Consider a (countable) sequence $\delta_i \to 0$. For each δ_i, the above holds almost surely. Since the countable union of almost sure events is almost sure, $|P_n(j_1, \ldots j_k) - P_{CCM_n}(j_1, \ldots j_k)| \to 0$ almost surely.

Finally, Fact 5.8.11 applied to Claims 1 and 2 implies that $\mathcal{L}\left(\left(L_{G_n}(A_1), \ldots, L_{G_n}(A_k)\right)\right)$ converges weakly to $\mathcal{L}\left(\left(\xi_{\mathcal{W}}(A_1), \ldots, \xi_{\mathcal{W}}(A_k)\right)\right)$ almost surely, as desired.

□

5.7 Hypothesis testing: proofs

Theorem 5.1.3 establishes the usefulness of p-samples in detecting community structure in the underlying graph. To this end, we use a test based on the statistic $S(\cdot)$ defined in (5.2). Our first result Theorem 5.7.1 establishes that the modularity based statistic S is naturally adapted to detect the presence of an underlying community structure, given access to the full graph. This result also enables us to compare the performance of the statistic S for the testing problem based on p-samples, as compared to the whole graph. In Section 5.7.1, we state and prove Theorem 5.7.1. The intermediate results in Theorem 5.7.1, such as the separation of the expected value of S-statistic under CM and CCM, also form the key ingredients in the proof of Theorem 5.1.3. In Section 5.7.2, we study the behavior of the S-statistic for the p-sampled graphs using a general theorem stated in Lemma 5.1.5, and hence complete the

proof of Theorem 5.1.3.

5.7.1 Hypothesis testing given access to the whole graph

Recall the setting of the detection problem in Section 5.1. Now we instead observe the whole graph G, and seek to test

$$H_0 : G \sim CM(D) \quad \text{vs.} \quad H_1 : G \sim CCM(D^*), \beta(D^*) \geq \varepsilon.$$

To do so, we compute $S(G, m)$, a truncated variant of the statistic $S(G)$, for the observed graph G, and then use this value to guess which model the graph came from. Let

$$S(G, m) = \sum_{i<j} \ell^2 \left(\tilde{X}_{ij} - \frac{d_i d_j}{\ell} \right)^2 \quad \text{where} \quad \tilde{X}_{ij} = \min\{X_{ij}, m\}.$$

Detecting communities based on truncated modularity:

Let $m = \max\{4e^2\lambda^2, 11 \log \ell\}$. Compute $S(G, m)$.

$$\text{Reject } H_0 \text{ if } S(G, m) \geq \ell^3/2 + \varepsilon\ell^3, \text{ and accept } H_0 \text{ otherwise.} \tag{5.22}$$

The following theorem provides non-asymptotic bounds on the Type-1 and Type-II errors of this hypothesis test.

Theorem 5.7.1. *Let $D = (d_i)_{i=1}^n$ be a degree sequence, and let $D^* = \left(\left(d_i^r, d_i^b\right)\right)_{i=1}^n$ be a corresponding colored sequence, so that $d_i = d_i^r + d_i^b$. Let $\ell = \sum d_i$, $\ell^r = \sum d_i^r$, and $\ell^b = \sum d_i^b$. Assume $\beta(D^*) \geq \varepsilon$. Let λ be a parameter satisfying $\lambda = o\left(\varepsilon\ell^{1/2}\right)$, $d_i \leq \lambda\sqrt{\ell}$, $d_i^r \leq \lambda\sqrt{\ell_r}$, and $d_i^b \leq \lambda\sqrt{\ell_b}$. Let $m = \max\{4e^2\lambda^2, 11 \log \ell\}$. There exists a constant $c > 0$ such that the test in*

(5.22) *is correct with probability at least*

$$1 - 2\exp\left(\frac{-c\varepsilon^2\ell}{\max\{\lambda^4, (\log\ell)^2\}}\right).$$

To prove Theorem 5.7.1, we first show that the mean of $S(G,m)$ is approximately $\ell^3/2$ under the CM and approximately $\ell^3/2 + 2\beta\ell^3$ under the CCM, and then use the Switching Lemmas (Lemmas 5.6.2 and 5.6.3) to show that $S(G,m)$ concentrates around its mean in both cases.

The statistic $S(G)$ plays an important role in the analysis. We compute the expectation of $S(G)$ for the CM and CCM (Lemma 5.7.3) and show that the expectation of $S(G)$ is close to the expectation of $S(G,m)$ (Lemma 5.7.4). We use $S(G,m)$ rather than $S(G)$ as the test statistic because it is easier to show concentration of $S(G,m)$. We may apply the Switching Lemmas to $S(G,m)$ because swapping two edges can change the statistic by at most $\Theta(m)$. However, the Switching Lemmas are not useful for showing concentration of $S(G)$. In the worst case, when there are $\Theta(\max\{d_i\})$ edges between a pair of vertices, swapping two edges may cause the statistic to differ by $\Theta(\max\{d_i\})$.

Remark 5.7.2. In Lemmas 5.7.5 and 5.7.6 below, we give upper bounds on the probability that $S(G,m)$ and $S(G)$ differ for the CM and the CCM respectively. Thus, if we use $S(\cdot)$ instead of $S(\cdot,m)$ in the test (5.22), it would introduce an additional additive error term $O\big(\frac{\ell^2}{\sqrt{m}}\exp(-m/2)\big)$.

Proof of Theorem 5.7.1

In this section we prove Theorem 5.7.1, hypothesis testing given access to the whole graph.

Lemma 5.7.3. *Let $D = (d_i)_{i=1}^n$ be a degree sequence, and let $D^* = \big((d_i^r, d_i^b)\big)_{i=1}^n$ be a corresponding colored sequence, so $d_i = d_i^r + d_i^b$. Let $\ell = \sum d_i$, $\ell^r = \sum d_i^r$, and $\ell^b = \sum d_i^b$.*

Assume $d_i \le \lambda\sqrt{\ell}$, $d_i^r \le \lambda\sqrt{\ell_r}$, *and* $d_i^b \le \lambda\sqrt{\ell_b}$. *Let* $\beta = \beta(D^*)$. *Then there exist small constants* c_0 *and* c_1 *such that*

(i) Let $G \sim CM(D)$. *Then* $\left|\mathbb{E}_{CM}(S(G)) - \frac{\ell^3}{2}\right| \le c_0\lambda\ell^{5/2}$.

(ii) Let $G \sim CCM(D^*, I)$. *Then* $\left|\mathbb{E}_{CCM}(S(G)) - \left(\frac{\ell^3}{2} + 2\beta\ell^3\right)\right| \le c_1\lambda\ell^{5/2}$.

Proof. (i) We begin by computing the expectation and variance of X_{ij} under the CM. We write $X_{ij} = \sum_{t=1}^{d_i} Y_t$ where Y_t is the indicator random variable for the event that the t^{th} half-edge of vertex i is paired to a half-edge of vertex j. Note $\mathbb{E}(Y_t) = d_j/(\ell-1)$, and $\mathbb{E}(Y_tY_s) = d_j(d_j-1)/((\ell-1)(\ell-3))$ for $s \ne t$. Observe

$$\begin{aligned}
\mathbb{E}(X_{ij}) &= \sum_{t=1}^{d_i}\mathbb{E}(Y_t) = \frac{d_id_j}{\ell-1}\\
\mathbb{E}\big(X_{ij}^2\big) &= \sum_{t=1}^{d_i}\mathbb{E}(Y_t) + \sum_{s\ne t}\mathbb{E}(Y_tY_s) = \frac{d_id_j}{\ell-1} + \frac{d_i(d_i-1)d_j(d_j-1)}{(\ell-1)(\ell-3)}\\
\mathsf{Var}(X_{ij}) &= \frac{d_id_j}{\ell-1} + \frac{d_i(d_i-1)d_j(d_j-1)}{(\ell-1)(\ell-3)} - \left(\frac{d_id_j}{\ell-1}\right)^2.
\end{aligned}$$

Therefore there exists a constant $c > 0$ such that for all sufficiently large n $|\,\mathsf{Var}(X_{ij}) - \frac{d_id_j}{\ell}| \le c\lambda d_id_j\ell^{-3/2}$. Observe that

$$\mathbb{E}_{CM}(S(G)) = \ell^2\sum_{i<j}\left(X_{ij} - \frac{d_id_j}{\ell-1}\right)^2 = \ell^2\sum_{i<j}\mathsf{Var}(X_{ij}).$$

Thus,

$$\left|\mathbb{E}_{CM}(S(G)) - \frac{\ell^3}{2}\right| \le \ell^2 c\lambda\ell^{-3/2}\sum_{i<j}d_id_j \le c\lambda\ell^{5/2}.$$

(ii) We begin by computing the expectation and variance of X_{ij} under the CCM. We write $X_{ij} = R_{ij} + B_{ij}$ where R_{ij} and B_{ij} are the number of red and blue edges between i

and j respectively. Note that R_{ij} and B_{ij} are independent random variables describing edge counts under a CM. By the above computations,

$$\begin{aligned}
\mathbb{E}(X_{ij}) &= \frac{d_i^r d_j^r}{\ell^r - 1} + \frac{d_i^b d_j^b}{\ell^b - 1} \\
\mathsf{Var}(X_{ij}) &= \mathsf{Var}(B_{ij}) + \mathsf{Var}(R_{ij}) = \frac{d_i^r d_j^r}{\ell^r} + \frac{d_i^b d_j^b}{\ell^b} + O\left(\frac{\lambda}{(\ell^r)^{3/2}} + \frac{\lambda}{(\ell^b)^{3/2}}\right), \\
\left|\sum_{i<j} \mathsf{Var}(X_{ij}) - \frac{\ell}{2}\right| &\leq c\left(\frac{\lambda\ell^2}{(\ell^r)^{3/2}} + \frac{\lambda\ell^2}{(\ell^b)^{3/2}}\right) \leq c'\sqrt{\ell}\lambda
\end{aligned}$$

for some constants c, c'. We now use the variance-bias decomposition to compute

$$\begin{aligned}
\mathbb{E}_{CCM}(S(G)) &= \ell^2 \sum_{i<j} \mathbb{E}\left(\left(X_{ij} - \frac{d_i d_j}{\ell - 1}\right)^2\right) \\
&= \ell^2 \sum_{i<j} \left(\mathsf{Var}(X_{ij}) + \left(\mathbb{E}(X_{ij}) - \frac{d_i d_j}{\ell - 1}\right)^2\right) \\
&= \ell^2 \sum_{i<j} \mathsf{Var}(X_{ij}) + 2\beta\ell^3
\end{aligned}$$

Thus for some constant c_1,

$$\left|\mathbb{E}_{CCM}(S(G)) - \left(\frac{\ell^3}{2} + 2\beta\ell^3\right)\right| \leq c_1\lambda\ell^{5/2}.$$

□

Lemma 5.7.4. *Let G be a multigraph drawn from a distribution over graphs with $\ell/2$ edges and maximum degree at most $\lambda\sqrt{\ell}$. We say G is "bad" if there exists a pair of vertices with at least m edges between them. Let α be probability that G is bad. Then*

$$\mathbb{E}(S(G)) - 2\alpha\lambda\ell^{7/2} \leq \mathbb{E}(S(G, m)) \leq \mathbb{E}(S(G)) + \alpha\ell^3 m^2.$$

Proof. First we note that if G has maximum degree at most $\lambda\sqrt{\ell}$, then

$$S(G) \leq \ell^2 \sum_{i<j} X_{ij}^2 + \sum_{i<j} d_i^2 d_j^2 \leq \lambda \ell^{7/2} + \lambda^2 \ell^3 \leq 2\lambda \ell^{7/2},$$

where we have used that $\lambda \leq \sqrt{\ell}$ as the max degree cannot exceed the sum of degrees. Since

$$\mathbb{E}(S(G)) = \mathbb{E}(S(G) \mid G \text{ good})\,(1-\alpha) + \mathbb{E}(S(G) \mid G \text{ bad})\alpha,$$

and S is a non-negative random variable, it follows that

$$\mathbb{E}(S(G)) - 2\alpha\lambda\ell^{7/2} \leq \mathbb{E}(S(G) \mid G \text{ good})\,(1-\alpha) \leq \mathbb{E}(S(G)). \tag{5.23}$$

Note that if G is good, then $S(G) = S(G, m)$. Therefore

$$\begin{aligned}\mathbb{E}(S(G,m)) &= \mathbb{E}(S(G,m) \mid G \text{ good})\,(1-\alpha) + \mathbb{E}(S(G,m) \mid G \text{ bad})\alpha \\ &= \mathbb{E}(S(G) \mid G \text{ good})\,(1-\alpha) + \mathbb{E}(S(G,m) \mid G \text{ bad})\alpha \end{aligned} \tag{5.24}$$

Combining Equations (5.23) and (5.24) and the observation that $S(G, m) \leq \ell^3 m^2$ yields the desired statement. □

Lemmas 5.7.5 and 5.7.6 give the values of α (an upper bound on the probability that $S(G, m)$ and $S(G)$ differ) for the CM and the CCM.

Lemma 5.7.5. *Let $G \sim CM(D)$ where $\sum_i d_i = \ell$ and $\max_i d_i \leq \lambda\sqrt{\ell}$. Then for $m \geq 2e^2\lambda^2$, the probability that there exists a pair of vertices in G with at least m edges between them is at most*

$$\frac{\ell^2}{\sqrt{2\pi m}}\,\mathsf{exp}\left(-m\right).$$

Proof. Let A_{ij}^m be the event that there are at least m edges between vertices i and j. Consider

an ordering of the half edges in which all the half edges of i appear first. Generate a G by pairing the lowest unpaired half edge to a random unpaired half edge in each iteration. We use Stirling's approximation to compute

$$\mathbb{P}\big(A_{ij}^m\big) \leq \binom{d_i}{m} \frac{(d_j)^m}{(\ell - d_j)^m} \leq \left(\frac{d_i d_j}{\ell - d_j}\right)^m \frac{e^m}{m^m \sqrt{2\pi m}} \leq \frac{1}{\sqrt{2\pi m}} \left(\frac{2e\lambda^2}{m}\right)^m \tag{5.25}$$

Noting $m \geq 2e^2\lambda^2$ and taking a union bound over all pairs i, j (of which there are at most $n^2 \leq \ell^2$), we obtain the desired upper bound. □

Lemma 5.7.6. *Let $G \sim CCM(D^*)$ where $\sum_i d_i^r = \ell_r$, $\sum_i d_i^b = \ell_b$ and $d_i^k \leq \lambda\sqrt{\ell_k}$ for all i and $k \in \{r, b\}$. For $m \geq 4e^2\lambda^2$, the probability that there exists a pair of vertices in G with at least m edges between them is at most*

$$\frac{2\ell^2}{\sqrt{\pi m}} \exp\left(-m/2\right).$$

Proof. Note that G is the union of two configuration models, one composed of red edges and one composed of blue edges. If there is pair of vertices with at least m edges between them, then there is a pair of vertices in the red or blue configuration model with at least $m/2$ edges between them. Thus applying a union bound to Lemma 5.7.5 with $m/2$ yields the desired result.

□

Finally we show concentration of $S(G, m)$ under the CM and CCM (Lemmas 5.7.7 and 5.7.8 respectively).

Lemma 5.7.7. *Let $G \sim CM(D)$ where $\sum_i d_i = \ell$, $\max_i d_i \leq \lambda\sqrt{\ell}$, and $m \geq 2e^2\lambda^2$. Let c_0 be the constant given in Lemma 5.7.3. For $Z \geq c_0\lambda\ell^{5/2} + \ell^5 m^{3/2} \exp\left(-m\right)$, there exists a*

constant $c > 0$ such that

$$\mathbb{P}\left(S(G,m) \geq \frac{\ell^3}{2} + Z\right) \leq 2\exp\left(-\left(\frac{Z}{4\ell^{5/2}m} - \frac{c\lambda}{m} - \frac{\ell^{5/2}\sqrt{m}\exp(-m)}{4}\right)^2\right).$$

Proof. By Lemmas 5.7.3 to 5.7.5,

$$\mathbb{E}(S(G,m)) \leq \frac{\ell^3}{2} + c_0\lambda\ell^{5/2} + \frac{\ell^5 m^{3/2}}{\exp(m)}.$$

Now, when $G \sim G'$, i.e., G and G' differs by at most one switch, then the degrees remain the degrees and ℓ remains the same. Moreover, one switch can change at most two X_{ij}'s, and thus

$$|S(G,m) - S(G',m)| \leq 2\ell^2 \max_{i,j}\left\{2\tilde{X}_{ij} + 1 - \frac{2d_i d_j}{\ell}\right\} \vee 0 \leq 4\ell^2 m,$$

when $G \sim G'$, and so we may apply the Switching Lemma Lemma 5.6.2. We compute

$$\begin{aligned}
\mathbb{P}\left(S(G,m) \geq \frac{\ell^3}{2} + Z\right) &\leq \mathbb{P}\left(|S(G,m) - \mathbb{E}(S(G,m))| \geq Z - c_0\lambda\ell^{5/2} - \frac{\ell^5 m^{3/2}}{\exp(m)}\right) \\
&\leq 2\exp\left(\frac{-\left(Z - c_0\lambda\ell^{5/2} - \ell^5 m^{3/2}\exp(-m)\right)^2}{16\ell^5 m^2}\right) \\
&= 2\exp\left(-\left(\frac{Z}{4\ell^{5/2}m} - \frac{c_0\lambda}{4m} - \frac{\ell^{5/2}\sqrt{m}\exp(-m)}{4}\right)^2\right).
\end{aligned}$$

□

Lemma 5.7.8. *Let $G \sim CCM(D^*, I)$ where $\sum_i d_i = \ell$, $\max_i d_i \leq \lambda\sqrt{\ell}$, and $m \geq 4e^2\lambda^2$. For $Z \geq c_0\lambda\ell^{5/2} + \ell^5 m^{3/2}\exp(-m/2)$, there exist a constant $c > 0$ such that*

$$\mathbb{P}\left(S(G,m) \leq \frac{\ell^3}{2} + 2\beta\ell^3 - Z\right) \leq 2\exp\left(-\left(\frac{Z}{4\ell^{5/2}m} - \frac{c\lambda}{m} - \frac{\ell^3\exp(-m/2)}{4m}\right)^2\right).$$

Proof. By Lemmas 5.7.3, 5.7.4 and 5.7.6,

$$\mathbb{E}(S(G,m)) \geq \frac{\ell^3}{2} + 2\beta\ell^3 - c_1\lambda\ell^{5/2} - 2\ell^{11/2}\exp\left(-m/2\right).$$

Using identical reasoning as in the proof of Lemma 5.7.7, $|S(G,m) - S(G',m)| \leq 4\ell^2 m$ when $G \sim G'$, and so we may apply the Subset Switching Lemma Lemma 5.6.3. We compute

$$\begin{aligned}
&\mathbb{P}\left(S(G,m) \leq \frac{\ell^3}{2} + 2\beta\ell^3 - Z\right) \\
&\leq \mathbb{P}\big(|S(G,m) - \mathbb{E}(S(G,m))| \geq Z - c_1\lambda\ell^{5/2} - 2\ell^{11/2}\exp\left(-m/2\right)\big) \\
&\leq 2\exp\left(\frac{-\left(Z - c_1\lambda\ell^{5/2} - 2\ell^{11/2}\exp\left(-m/2\right)\right)^2}{16\ell^5 m^2}\right) \\
&= 2\exp\left(-\left(\frac{Z}{4\ell^{5/2}m} - \frac{c_1\lambda}{4m} - \frac{2\ell^3\exp\left(-m/2\right)}{4m}\right)^2\right).
\end{aligned}$$

□

We now prove Theorem 5.7.1.

Proof of Theorem 5.7.1. We give an upper bound on the probability the hypothesis test fails. In the case when $G \sim CM(D)$, the test is incorrect only if $S(G,m) \geq \ell^3/2 + \beta\ell^3$. In the case when $G \sim CCM(D)$, the test is incorrect only if $S(G,m) \leq \ell^3/2 + \beta\ell^3$. We now apply Lemmas 5.7.7 and 5.7.8 with $Z = \beta\ell^3$. Note that, since $m \geq 11\log\ell$, $c_0\lambda\ell^{5/2} + \ell^5 m^{3/2}\exp\left(-m/2\right) \leq \ell^{5/2}(c_0\lambda + m^{3/2})$. Under the assumption that $\lambda = o(\ell^{1/2}\varepsilon)$, the conditions of Lemmas 5.7.7 and 5.7.8 are satisfied. The proof now follows. □

5.7.2 Hypothesis testing given access to a sample

In this section, we consider the accuracy of the hypothesis test given access to only one sample $G^s \sim \mathrm{Smpl}(G, t/\sqrt{2e(G)})$, where $e(G)$ is the number of non-loop edges in the multigraph G.

To begin with, we first prove that the number of self-loops is of the order $\sqrt{\ell} = o(\ell)$, and thus, it is enough to consider $G^s \sim \mathrm{Smpl}(G, t/\sqrt{\ell})$.

Lemma 5.7.9. *Let $D = (d_i)$ be a degree sequence, and let $D^* = \left(\left(d_i^r, d_i^b\right)\right)$ be a corresponding colored sequence, so $d_i = d_i^r + d_i^b$. Let $\ell = \sum d_i$, $\ell^r = \sum d_i^r$, and $\ell^b = \sum d_i^b$. Assume $d_i \le \lambda\sqrt{\ell}$, $d_i^r \le \lambda\sqrt{\ell_r}$, and $d_i^b \le \lambda\sqrt{\ell_b}$. Let $L(G)$ be the number of self-loops in G. Then*

$$\mathbb{P}_{CM}\left[L(G) \ge \lambda\sqrt{\ell} + \frac{t\sqrt{\ell}}{4\lambda}\right] \le 2\exp\left(\frac{-t^2}{64\lambda^2}\right), \quad \mathbb{P}_{CCM}\left[L(G) \ge 2\lambda\sqrt{\ell} + \frac{t\sqrt{\ell}}{4\lambda}\right] \le 2\exp\left(\frac{-t^2}{64\lambda^2}\right).$$

Proof. First, let us consider the case for CM. Let X_{ii} be the random variable for the number of self-loops at vertex i, so $L(G) = \sum_i X_{ii}$. It follows that

$$\mathbb{E}_{CM}[L(G)] = \sum_i \mathbb{E}_{CM}[X_{ii}] = \sum_i \frac{d_i(d_i - 1)}{\ell - 1} \le \lambda\sqrt{\ell}.$$

To show concentration of $L(G)$, we apply the Switching Lemma (Lemma 5.6.2). If $G \sim G'$, then $|L(G) - L(G')| \le 2$. It follows that

$$\mathbb{P}_{CM}[|L(G) - \mathbb{E}_{CM}[L(G)]| \ge \delta] \le 2\exp\left(\frac{-\delta^2}{4\ell}\right).$$

It follows that

$$\mathbb{P}_{CM}\left[L(G) \ge \lambda\sqrt{\ell} + \frac{t\sqrt{\ell}}{4\lambda}\right] \le \mathbb{P}_{CM}\left[|L(G) - \mathbb{E}(L(G))| \ge \frac{t\sqrt{\ell}}{4\lambda}\right] \le 2\exp\left(\frac{-t^2}{64\lambda^2}\right).$$

The proof for CCM is identical by observing that

$$\mathbb{E}_{CCM}[L(G)] = \sum_i \mathbb{E}_{CCM}[X_{ii}] = \sum_i \frac{d_i^r(d_i^r - 1)}{\ell^r - 1} + \frac{d_i^b(d_i^b - 1)}{\ell^b - 1} \le 2\lambda\sqrt{\ell}.$$

□

Henceforth, we only consider $G^s \sim \text{Smpl}(G, t/\sqrt{\ell})$. To prove Theorem 5.7.1, we condition on the event A that there is no pair of vertices with more than $m = 4e^2\lambda^2$ edges between them in G (Lemmas 5.7.5 and 5.7.6), and follow these steps:

1. Show that $\mathbb{E}(S(G^s)) \approx t^6/\ell^3 S(G)$. For this, we write $S(G)$ as a polynomial of multigraph counts. The terms corresponding to multigraphs on q vertices are scaled by $(t/\sqrt{\ell})^q$ when they appear in the polynomial $\mathbb{E}(S(G^s))$. The highest order terms of $S(G)$ correspond to multigraphs on six vertices, so $\mathbb{E}(S(G^s)) \approx t^6/\ell^3 S(G)$. See Lemma 5.7.10.

2. Show that $S(G^s)$ concentrates around its mean. Indeed, under the assumption A, we may apply Kim-Vu Lemma 5.8.12 to obtain concentration results, Lemma 5.7.12.

3. Let $\tilde{\ell}$ denote the sum of degrees in G^s. Show that $\tilde{\ell}$ concentrates around t^2, Lemma 5.7.13.

These steps imply that $S(G^s)$ concentrates around $\tilde{\ell}^3/2$ when $G \sim CM$ and around $\tilde{\ell}^3/2 + 2\beta\tilde{\ell}^3$ when $G \sim CCM$.

Proof of Theorem 5.1.3

In this section, we prove Theorem 5.1.3, hypothesis testing given only access to one sample from $\text{Smpl}_t(G)$. First we use Lemma 5.1.5 to show that the modularity statistic $S(G)$ scales approximately as $S(G^s) \approx t^6/\ell^3 S(G)$.

Lemma 5.7.10. *Let G be a multigraph with maximum degree $\lambda\sqrt{\ell}$ such that there are at most $4e^2\lambda^2$ edges between any pair of vertices, where $\ell = 2E(G)$. Let $G^s \sim \text{Smpl}_t(G)$ for $t \geq \lambda$. There exists a constant $c > 0$ such that*

$$\left| \mathbb{E}(S(G^s, m)) - \frac{t^6}{\ell^3} S(G, m) \right| \leq c_0 \lambda^3 t^5.$$

Proof. Let $\tilde{X}_{ij} = \max\{X_{ij}, m\}$ where X_{ij} is the number of edges between vertices i and j in

G. Note that

$$S(G,m) = \sum_{i<j} \left(\tilde{\ell}\tilde{X}_{ij} - \tilde{d}_i\tilde{d}_j\right)^2 = \tag{5.26}$$
$$\sum_{i<j} \tilde{X}_{ij}^2 \left(\sum_{m,n} \tilde{X}_{mn}\right)^2 - 2\tilde{X}_{ij}\left(\sum_{m,n}\tilde{X}_{mn}\right)\left(\sum_k \tilde{X}_{ik}\right)\left(\sum_k \tilde{X}_{jk}\right) + \left(\sum_k \tilde{X}_{ik}\right)^2\left(\sum_k \tilde{X}_{jk}\right)^2.$$

Therefore $S(G,m)$ is the sum of terms of the form $\tilde{X}_{ij}\tilde{X}_{kl}\tilde{X}_{mn}\tilde{X}_{op}$ where $|\{i,j,k,l,m,n,o,p\}| \leq 6$. In other words, $S(G,m)$ is the sum of homomorphism counts of graphs with four edges and at most 6 vertices in $\tilde{G}$, the graph obtained from G by deleting edges so that there are at most m edges between any pair of vertices. Let $S(G,m) = S^+(\tilde{G}) - S^-(\tilde{G})$ where $S^+(\tilde{G})$ contains the terms of Equation (5.26) with positive coefficients, and $S^-(\tilde{G})$ contains the opposite of the terms with negative coefficients.

We apply Lemma 5.1.5 with $e = 4$, $v_1 = 6$ and $v_2 = 5$ to each S^+ and S^- and obtain

$$\left|\mathbb{E}\left[S^+(\tilde{G})\right] - S^+(\tilde{G})\right| \leq c_1\lambda^3 t^5 \quad \text{and} \quad \left|\mathbb{E}\left[S^-(\tilde{G})\right] - S^-(\tilde{G})\right| \leq c_2\lambda^3 t^5.$$

Since $S(G,m) = S^+(\tilde{G}) - S^-(\tilde{G})$, the result follows.

□

Proof of Lemma 5.1.5. First we consider how sampling affects the homomorphism counts. A fixed injective homomorphism from F to G is a homomorphism from F to G^s if and only if the vertices of the image of $V(F)$ survive the sampling process. It follows that

$$\mathbb{E}(H(F_i, G^s)) = H(F_i, G)\left(\frac{t}{\sqrt{\ell}}\right)^{v_i},$$

where $v_i = |V(F_i)|$.

Next we give an upper bound on $H(F_i, G)$ under the maximum degree and maximum number of edges conditions. Let c_i (and v_i) be the number of connected components (and

vertices) of F_i. Consider a fixed spanning forest of F_i in which each component has a labeled "first" edge. There are at most $\left(\frac{\ell}{2}\right)^{c_i}$ ways to pick the first edges of each component. There are at most $(\lambda\sqrt{\ell})^{v_i-2c_i}$ ways to pick the remaining edges in the spanning forest since the maximum degree is $\lambda\sqrt{\ell}$. Since there are at most m edges between any pair of vertices, there are at most $m^{e-v_i+c_i}$ ways to choose the other edges. Under the assumption that $m \leq \lambda^2$,

$$H(F_i, G) \leq a_i \ell^{c_i} (\lambda\sqrt{\ell})^{v_i-2c_i} m^{e-v_i+c_i} = a_i \ell^{v_i/2} \lambda^{2e-v_i},$$

for some constant a_i depending only on F_i. Finally, we compute

$$\mathbb{E}(S(G^s)) = \sum_{i=1}^{k} \alpha_i \, \mathbb{E}(H(F_i, G^s)) = \sum_{i=1}^{k} \alpha_i H(F_i, G) \left(\frac{t}{\sqrt{\ell}}\right)^{v_i}.$$

Since $t \leq \sqrt{\ell}$,

$$\mathbb{E}(S(G^s)) \geq S(G) \left(\frac{t}{\sqrt{\ell}}\right)^{v_1}.$$

Under the assumption that $t \geq \lambda$,

$$\begin{aligned}
\mathbb{E}(S(G^s)) &= \left(\frac{t}{\sqrt{\ell}}\right)^{v_1} S(G) + \sum_{i=1}^{k} \alpha_i H(F_i, G) \left(\left(\frac{t}{\sqrt{\ell}}\right)^{v_i} - \left(\frac{t}{\sqrt{\ell}}\right)^{v_1}\right) \\
&\leq \left(\frac{t}{\sqrt{\ell}}\right)^{v_1} S(G) + \sum_{i=1}^{k} \alpha_i a_i \ell^{v_i/2} \lambda^{2e-v_i} \left(\frac{t}{\sqrt{\ell}}\right)^{v_i} \\
&= \left(\frac{t}{\sqrt{\ell}}\right)^{v_1} S(G) + c_0 \lambda^{2e-v_2} t^{v_2}
\end{aligned}$$

where c_0 is a constant independent of G.

□

The Kim-Vu concentration inequality (Lemma 5.8.12) is the key tool that we use for showing the concentration of $S(G^s, m)$ and $\tilde{\ell}$ once the graph G is fixed (Lemmas 5.7.12 and 5.7.13 respectively). First, we prove Lemma 5.7.11, which uses the Kim-Vu concentration

inequality to establish concentration for homomorphism counts under sampling. Lemma 5.7.12 will follow by applying Lemma 5.7.11.

Lemma 5.7.11. *Let G be a graph with maximum degree $\lambda\sqrt{\ell}$ such that there are at most $4e^2\lambda^2$ edges between any pair of vertices, where $\ell = 2E(G)$. Let $H(F,G)$ be the number of labeled homomorphisms of F in G, and $v = |V(F)|$ and $e = |E(F)|$. Assume $H(F,G) = \Omega\left(t^{v-1}\lambda^{2e-v+1}\right)$. Let $G^s \sim \mathrm{Smpl}_t(G)$ for $t \geq \lambda$. There exists a constant $c > 0$ such that*

$$\mathbb{P}(|H(F,G^s) - \mathbb{E}(H(F,G^s))| \geq \alpha\,\mathbb{E}(H(F,G^s))) \leq O\left(\mathsf{exp}\left(-c\alpha\sqrt{\frac{\mathbb{E}(H(F,G^s))}{t^{v-1}\lambda^{2e-v+1}}}\right)\right).$$

Proof. We can write the random variable $H(F,G^s)$ as a polynomial in the framework of [KV00]

$$Y_H = H(F,G^s) = \sum_{e\in E(H)} w(e)\prod_{s\in e} t_s,$$

where $V(H) = V(G)$, $E(H) = \{S \subset V(G) \mid |S| = v\}$, and

$$w(e) = \sum_{\psi\in\Psi}\prod_{\{i,j\}\in E(F)} X_{\psi(i)\psi(j)}\mathbb{1}\{Im(\psi) = e\},$$

with Ψ given by Definition 5.1.4. We now compute $\mathbb{E}(H)$ and $\mathbb{E}_z(H)$ and apply Kim Vu to the above polynomial Y_H.

Consider a set A. We compute $\mathbb{E}(Y_{H_A})$. The polynomial Y_{H_A} is the sum of terms of the form $w(e)\prod_{s\in e\setminus A} t_s$ where $A \subseteq e$. Let

$$E_r = \{e \in E(H) \mid A \subseteq e, |e \setminus A| = r\} \text{ and } T_r = \sum_{e\in E_r} w(e).$$

It follows that

$$\mathbb{E}(Y_{H_A}) \leq \sum_{r=0}^{|A|} T_r\left(\frac{t}{\sqrt{\ell}}\right)^r.$$

We now bound T_r. Note that T_r is the number of copies of F in G in which all but r vertices are in A. We claim that $T_r \leq c\lambda^{2e-r}t^r$ for some constant c. Consider a partially labeled copy of F in which each vertex of A is the label of a unique vertex of F. There is a constant (independent of G) number of such labelings. We count the number of copies of F in G in which each labeled vertex of F maps to the corresponding vertex in A. To do so, we count the ways to choose the edges and r remaining vertices to obtain a copy of F in G. Let c' be the number of connected components of F that do not contain a labeled vertex. For each such connected component arbitrarily choose a "first edge." There are $\ell^{c'}$ ways to place the first edge in each of these components in G, establishing the placement of $2c'$ vertices in G. Now at least one vertex in each connected component has an established place in G, and we must select $r - 2c'$ additional vertices. Fix a set of $r - 2c'$ edges of F so that each of the $r - 2c'$ vertices that are not yet established (i.e. not in A or adjacent to one of the $2c'$ edges placed) is adjacent to one edge in the set. Since each vertex has maximum degree $\lambda\sqrt{\ell}$, there are at most $(\lambda\sqrt{\ell})^{r-2c'}$ ways to select the placement of this set of edges in G. The remaining $e - r + c'$ edges we need to select to form a copy of F_i in G occur between pairs of vertices that have already been established. Since each pair of vertices has at most λ^2 edges, there are at most $\lambda^{2(e-r+c')}$ ways to select these edges. It follows that for some constant c

$$T_r \left(\frac{t}{\sqrt{\ell}}\right)^r \leq c\ell^{c'} \left(\lambda\sqrt{\ell}\right)^{r-2c'} \lambda^{2(e-r+c')} = ct^r\lambda^{2e-r}.$$

Under the assumption that $\lambda \leq t$, this quantity is maximized when r is largest. Note $r \leq v - |A|$. It follows that

$$\mathbb{E}_z(H) \leq ct^{v-1}\lambda^{2e-v+1}.$$

Under the assumption that $\mathbb{E}(H) = \max\{ct^{v-1}\lambda^{2e-v+1}, \mathbb{E}_0(H)\} = \mathbb{E}(H(F, G'))$, we apply

Lemma 5.8.12 and obtain for some constant c

$$\begin{aligned}
&\mathbb{P}(|H(F,G') - \mathbb{E}(H(F,G'))| \geq \alpha\, \mathbb{E}(H(F,G'))) \\
&\leq \mathbb{P}\left(|H(F,G') - \mathbb{E}(H(F,G'))| \geq 8\left(\alpha c\sqrt{\frac{\mathbb{E}(H(F,G'))}{t^{v-1}\lambda^{2e-v+1}}}\right)\sqrt{\mathbb{E}_z(H)\,\mathbb{E}(H)}\right) \\
&\leq O\left(\exp\left(-c\alpha\sqrt{\frac{\mathbb{E}(H(F,G'))}{t^{v-1}\lambda^{2e-v+1}}}\right)\right)
\end{aligned}$$

□

Lemma 5.7.12. *Let G be a multigraph with maximum degree $\lambda\sqrt{\ell}$ such that there are at most $4e^2\lambda^2$ edges between any pair of vertices, where $\ell = 2E(G)$. Let $G^s \sim \mathrm{Smpl}_t(G)$ for $t \geq \lambda$. Moreover, let $H(F,G) \leq C\ell^3$ for any F with six vertices. Then, there exists a constant $c > 0$ such that*

$$\mathbb{P}\big(|S(G^s,m) - \mathbb{E}(S(G^s,m))| \geq \alpha t^6\big) \leq O\left(\exp\left(-c\alpha\sqrt{\frac{t}{\lambda^3}}\right)\right).$$

Proof. Let $\tilde{G}^s$ denote the graph obtained from G^s by deleting edges so that there are at most m edges between any pair of vertices. As established in the proof of Lemma 5.7.10, $S(G^s,m) = \sum_{i=1}^k \alpha_i^+ H(F_i,\tilde{G}^s) - \sum_{i=1}^k \alpha_i^- H(F_i,\tilde{G}^s)$ where each F_i is a graph with four edges and at most six vertices. It is now enough to show concentration of $H(F_i,\tilde{G}^s)$. We compute

$$\begin{aligned}
&\mathbb{P}\left(|H(F_i,G^s) - \mathbb{E}(H(F_i,G^s))| \geq \mathbb{E}(H(F_i,G^s))\left(\frac{\alpha t^6}{k\,\mathbb{E}(H(F_i,G^s))}\right)\right) \\
&= O\left(\exp\left(\frac{-c\alpha t^6}{\sqrt{t^{v-1}\lambda^{2e-v+1}\,\mathbb{E}(H(F_i,G^s))}}\right)\right) \\
&= O\left(\exp\left(-\bar{c}\alpha\sqrt{\frac{t}{\lambda^3}}\right)\right).
\end{aligned}$$

The first equality follows from Lemma 5.7.11 and the last line follows from the observation

that

$$\mathbb{E}(H(F_i, G^s)) = \frac{t^6}{\ell^3} H(F_i, G) \leq Ct^6.$$

□

Lemma 5.7.13. *Let G be a multigraph with maximum degree at most $\lambda\sqrt{\ell}$ such that there are at most $4e^2\lambda^2$ edges between any pair of vertices, where $\ell = 2E(G)$ and $\lambda \geq 1/2$. Assume G has at most $2\lambda\sqrt{\ell} + \frac{t\sqrt{\ell}}{4\lambda}$ self-loops. Let $G^s \sim \mathrm{Smpl}_t(G)$ for $t \geq 4e^2\lambda^2$. Let $\tilde{\ell} = 2E(G^s)$. Then*

$$\mathbb{P}\left(|\tilde{\ell} - t^2| \geq \frac{t^2}{\lambda}\right) \leq O\left(\exp\left(-\frac{1}{32}\sqrt{\frac{t}{\lambda^3}}\right)\right).$$

Proof. We write $\tilde{\ell}$ as a polynomial and apply Kim-Vu. Note

$$Y_H = \tilde{\ell} = \sum_{i \neq j} 2X_{ij} t_i t_j + \sum_i 2X_{ii} t_i.$$

Therefore

$$\mathbb{E}[\tilde{\ell}] = t^2 + \sum_i X_{ii}\left(\frac{t}{\sqrt{\ell}} - \frac{t^2}{\ell}\right).$$

By the assumption of this lemma, $\sum_i X_{ii} \leq \lambda\sqrt{\ell} + \frac{t\sqrt{\ell}}{4\lambda}$, and thus it follows that

$$t^2 \leq \mathbb{E}[\tilde{\ell}] \leq t^2 + t\lambda + \frac{t^2}{4\lambda}.$$

Note $\mathbb{E}_1(H) \leq 2\lambda\sqrt{\ell}\frac{t}{\sqrt{\ell}} = 2\lambda t$ and $\mathbb{E}_2(H) \leq 8e^2\lambda^2$ since the maximum degree is $\lambda\sqrt{\ell}$ and each $X_{ij} \leq 4e^2\lambda^2$. Therefore $\mathbb{E}_z(H) \leq 2\lambda\sqrt{\ell}$ and $\mathbb{E}_0(H) \leq 2t^2$. By Lemma 5.8.12,

$$\mathbb{P}\left(\left|\tilde{\ell} - \mathbb{E}[\tilde{\ell}]\right| \geq 16\gamma\sqrt{\lambda}t^{3/2}\right) \leq O\left(\exp\left(-\gamma\right)\right).$$

It follows that

$$\begin{aligned}\mathbb{P}\left(|\tilde{\ell}-t^2|\geq\frac{t^2}{\lambda}\right)&\leq\mathbb{P}\left(\left|\tilde{\ell}-\mathbb{E}[\tilde{\ell}]\right|\geq\frac{t^2}{\lambda}-t\lambda-\frac{t^2}{4\lambda}\right)\\&\leq\mathbb{P}\left(\left|\tilde{\ell}-\mathbb{E}[\tilde{\ell}]\right|\geq\frac{t^2}{2\lambda}\right)=\mathbb{P}\left(\left|\tilde{\ell}-\mathbb{E}[\tilde{\ell}]\right|\geq 16\left(\frac{1}{32}\sqrt{\frac{t}{\lambda^3}}\right)\sqrt{\lambda}t^{3/2}\right)\\&\leq O\left(\exp\left(-\frac{1}{32}\sqrt{\frac{t}{\lambda^3}}\right)\right).\end{aligned}$$

□

Finally, we prove Theorem 5.1.3.

Proof of Theorem 5.1.3. Let $G\sim CM(D)$ and $G^s\sim\mathrm{Smpl}(G,t/\sqrt{\ell})$. First we claim that if the following four events hold, then the hypothesis test based on samples is accurate. Let $m=4e^2\lambda^2$, $\tilde{\ell}=2E(G^s)$, and c_0 be the constant given in Lemma 5.7.10.

(1) G has no more m edges between any pair of vertices.

(2) $S(G,m)\leq\frac{\ell^3}{2}+z$.

(3) $S(G^s,m)\leq\mathbb{E}(S(G^s,m))+t^6\left(\beta-\frac{16+32\beta}{\lambda}-\frac{c_0\lambda^3}{t}-\frac{z}{\ell^3}\right)$.

(4) $|\tilde{\ell}-t^2|\leq\frac{t^2}{\lambda}$.

Under assumption (1), we may apply Lemma 5.7.10, which given (2) implies that $\mathbb{E}(S(G^s,m))\leq\frac{t^6}{\ell^3}\left(\frac{\ell^3}{2}+z\right)+c_0\lambda^3t^5=\frac{t^6}{2}+\frac{t^6z}{\ell^3}+c_0\lambda^3t^5$. Therefore under assumption (3),

$$S(G^s,m)\leq\frac{t^6}{2}+t^6\beta-\frac{t^6(16+32\beta)}{\lambda}.$$

Note that under assumption (1), $S(G^s,m)=S(G^s)$. Assumption (4) implies that $t^2\leq\tilde{\ell}\lambda/(1+\lambda)\leq\tilde{\ell}$, and therefore $|t^6-\tilde{\ell}^3|\leq 8\tilde{\ell}^2t^2/\lambda$. Under the hypotheses on t and λ, it now

follows that

$$S(G^s) = S(G^s, m) \leq \left(\tilde{\ell}^3 + \frac{8\tilde{\ell}^2}{\lambda}\right)\left(\frac{1}{2} + \beta\right) - \frac{t^6\,(16 + 32\beta)}{\lambda} \leq \frac{\tilde{\ell}^3}{2} + \beta\tilde{\ell}^3,$$

and so the hypothesis test is accurate.

Next, let $G \sim CCM(D^*, I)$ and $G^s \sim \text{Smpl}_t(G)$. We claim that if the following four events hold, then the hypothesis test is accurate on G^s. Again, let $m = 4e^2\lambda^2$, $\tilde{\ell} = 2E(G^s)$, and c_0 be the constant given in Lemma 5.7.10.

(5) G has no more m edges between any pair of vertices.

(6) $S(G, m) \geq \frac{\ell^3}{2} + 2\beta\ell^3 - z$

(7) $S(G^s, m) \geq \mathbb{E}(S(G^s, m)) - t^6\left(\beta - \frac{16+32\beta}{\lambda} - \frac{c_0\lambda^3}{t} - \frac{z}{\ell^3}\right)$

(8) $|\tilde{\ell} - t^2| \leq \frac{t^2}{\lambda}$.

Under assumption (5), we may apply Lemma 5.7.10, which given (6) implies that $\mathbb{E}(S(G^s, m)) \geq \frac{t^6}{\ell^3}\left(\frac{\ell^3}{2} + 2\beta\ell^3 - z\right) - c_0\lambda^3t^5 = \frac{t^6}{2} + 2\beta t^6 - \frac{t^6 z}{\ell^3} - c_0\lambda^3t^5$. Therefore under assumption (7),

$$S(G^s, m) \geq \frac{t^6}{2} + t^6\beta + \frac{t^6(16 + 32\beta)}{\lambda}.$$

Note that under assumption (1), $S(G^s, m) = S(G^s)$. Assumption (4) implies $|t^6 - \tilde{\ell}^3| \leq 8\tilde{\ell}^2/\lambda$. Under the hypotheses on t and λ, it follows that

$$S(G^s) = S(G^s, m) \geq \left(\tilde{\ell}^3 - \frac{8\tilde{\ell}^2}{\lambda}\right)\left(\frac{1}{2} + \beta\right) + \frac{t^6\,(16 + 32\beta)}{\lambda} \geq \frac{\tilde{\ell}^3}{2} + \beta\tilde{\ell}^3,$$

and so the hypothesis test is accurate.

Under H_0 (or H_1), let A_i be the event that condition (i) does not hold for $i \in [1, 4]$ (or

$i \in [5, 8]$). It follows that

$$\begin{aligned}\mathbb{P}(ERROR) &= \mathbb{P}_{H_0}(\text{rejecting } H_0) + \mathbb{P}_{H_1}(\text{rejecting } H_1) \\ &\leq \mathbb{P}_{H_0}\left[\bigcup_{i=1}^{4} A_i\right] + \mathbb{P}_{H_1}\left[\bigcup_{i=5}^{8} A_i\right] \end{aligned} \tag{5.27}$$

We now use a union bound. By Lemmas 5.7.5 and 5.7.6,

$$\mathbb{P}_{H_0}[A_1] + \mathbb{P}_{H_1}[A_5] \leq \frac{2\ell^2}{\sqrt{m}} \exp\left(-m/2\right) = O\left(\frac{\ell^2}{\lambda} \exp\left(-2e^2\lambda^2\right)\right).$$

Let $z = \ell^3\beta/4$. Applying Lemma 5.7.7 and Lemma 5.7.8 with $Z = z$ (under the assumptions that $\lambda = o\left(\sqrt{\ell}\beta\right)$ and $\lambda = \Omega(\sqrt{\log \ell})$ yields

$$\begin{aligned}\mathbb{P}_{H_0}[A_2] &\leq 2\exp\left(-\left(\frac{Z}{4\ell^{5/2}m} - \frac{c_0\lambda}{m} - \frac{\ell^{5/2}\sqrt{m}\exp\left(-m\right)}{4}\right)^2\right) = O\left(\exp\left(\frac{-c\ell\beta^2}{\lambda^4}\right)\right). \\ \mathbb{P}_{H_1}[A_6] &\leq 2\exp\left(-\left(\frac{Z}{4\ell^{5/2}m} - \frac{c_0\lambda}{m} - \frac{\ell^3\exp\left(-m/2\right)}{4m}\right)^2\right) = O\left(\exp\left(\frac{-c\ell\beta^2}{\lambda^4}\right)\right).\end{aligned}$$

for some constant $c > 0$. Under the assumptions that $\lambda \geq 64/\beta + 64$, $t \geq 4c\lambda^3/\beta$, $m \geq 2\log\frac{16\ell^{3/2}}{\beta}$, Lemma 5.7.12 implies that

$$\begin{aligned}\mathbb{P}[A_3 \mid A_1^c], \mathbb{P}[A_7 \mid A_5^c] &= O\left(\exp\left(-c_1\sqrt{\frac{t}{\lambda^3}}\left(\beta - \frac{16+32\beta}{\lambda} - \frac{c_0\lambda^3}{t} - \frac{z}{\ell^3}\right)\right)\right) \\ &= O\left(\exp\left(-\bar{c}\beta\sqrt{\frac{t}{\lambda^3}}\right)\right)\end{aligned}$$

where $c_1 > 0$ is the constant given in Lemma 5.7.12 and $\bar{c} > 0$ is a constant. Finally we apply **??** and Lemmas 5.7.9 and 5.7.13 (with the assumption that $t^3 \geq 4\lambda$) and obtain

$$\mathbb{P}(A_4 \mid A_1^c), \mathbb{P}(A_8 \mid A_4^c) \leq 2\exp\left(\frac{-t^2}{64\lambda^2}\right) + O\left(\exp\left(-\frac{1}{32}\sqrt{\frac{t}{\lambda^3}}\right)\right)$$

$$= O\left(\exp\left(-\frac{1}{32}\sqrt{\frac{t}{\lambda^3}}\right)\right).$$

The result follows by using the above computations to bound Equation (5.27). □

5.8 Appendix

Claim 5.8.1. *Let ρ and ρ_n be such that $\rho_n \to \rho$ vaguely on $(\mathbb{R}^+)^k$,*

$$\limsup \int_{(\mathbb{R}^+)^k} \langle 1, x\rangle d\rho_n \leq c_1, \quad \textit{and} \int_{(\mathbb{R}^+)^k} \langle 1, \quad x\rangle d\rho \leq c_2.$$

For a bounded function f with $f(x) = o(\langle x, 1\rangle)$ as $x \to 0$, $\int f d\rho_n \to \int f d\rho$.

Proof. Let B_ε denote the L_1 ball of radius ε. We show $\int_{B_\varepsilon} f d\rho_n \to \int_{B_\varepsilon} f d\rho$ as $\varepsilon \to 0$ and $\int_{B_\varepsilon^c} f d\rho_n \to \int_{B_\varepsilon^c} f d\rho$. First note that for any $M \in \mathbb{R}^+$, $M\rho_n(B_M^c) \leq \int_{B_M^c} \langle x, 1\rangle d\rho_n \leq c_1$, and so $\rho_n(B_M^c) \leq c_1/M$. It follows that $\rho(B_M^c) \leq c_1/M$.

We begin by computing the integral away from zero. We approximate f by a sequence of bounded functions (f_M) with bounded support. Let δ be a small constant and $f_{max} = sup_x|f(x)|$. Define $f_M(x)$ to be some sequence of continuous function with the properties

$$f_M(x) = \begin{cases} f(x) & x \in B_M \\ 0 & x \in B_{M+\delta}^c \\ \leq f_{max} & x \in B_M^c \setminus B_{M+\delta}^c. \end{cases}$$

Observe

$$\left|\int_{B_\varepsilon^c} f d\rho_n - \int_{B_\varepsilon^c} f d\rho\right| \leq \left|\int_{B_\varepsilon^c} f d\rho_n - \int_{B_\varepsilon^c} f_M d\rho_n\right| + \left|\int_{B_\varepsilon^c} f_M d\rho_n - \int_{B_\varepsilon^c} f_M d\rho\right| + \left|\int_{B_\varepsilon^c} f_M d\rho - \int_{B_\varepsilon^c} f d\rho\right|$$

$$\leq 2f_{max}\rho_n(B_M^c) + + \left| \int_{B_\varepsilon^c} f_M d\rho_n - \int_{B_\varepsilon^c} f_M d\rho \right| + 2f_{max}\rho(B_M^c)$$
$$\leq \frac{4c_1 f_{max}}{M} + \left| \int_{B_\varepsilon^c} f_M d\rho_n - \int_{B_\varepsilon^c} f_M d\rho \right|.$$

The vague convergence of $\rho_n \to \rho$ and the fact that the above holds for all M together imply $\int_{B_\varepsilon^c} f d\rho_n \to \int_{B_\varepsilon^c} f d\rho$.

Finally observe

$$\int_{B_\varepsilon} f d\rho_n = o\left(\int \langle x, 1\rangle d\rho_n\right) = o(c_1) \quad \int_{B_\varepsilon} f d\rho_n = o\left(\int \langle x, 1\rangle d\rho\right) = o(c_2).$$

Thus, as $\varepsilon \to 0$, $\int_{B_\varepsilon} f d\rho_n \to \int_{B_\varepsilon} f d\rho = 0$. □

Fact 5.8.2. *Let $x_1, \ldots, x_n, y_1, \ldots y_n \in [0,1]$. Then*

$$\prod_{i=1}^n x_i - \prod_{i=1}^n y_i \leq \sum_i |x_i - y_i|.$$

Claim 5.8.3. *Let $X_1, \ldots X_n$ and $Y_1, \ldots Y_n$ be independent random variables taking values in $\mathbb{R}$. Let $\mathcal{X}_j = \left(\sum_{i=1}^j X_j\right) \otimes X_{j+1} \otimes \cdots \otimes X_n$ and $\mathcal{Y}_j = \left(\sum_{i=1}^j Y_i\right) \otimes Y_{j+1} \otimes \cdots \otimes Y_n$. Then for any $j \in [n]$,*

$$d_{TV}\left(\mathcal{L}\left(\mathcal{X}_j\right), \mathcal{L}\left(\mathcal{Y}_j\right)\right) \leq \sum_{i=1}^n d_{TV}\left(\mathcal{L}\left(X_i\right), \mathcal{L}\left(Y_i\right)\right).$$

Proof. First we show the statement for $j = 1$. Using Fact 5.8.2, we compute

$$d_{TV}\left(\mathcal{L}\left(\bigotimes_{i=1}^n X_i\right), \mathcal{L}\left(\bigotimes_{i=1}^n Y_i\right)\right) = \int_{\mathbb{R}^n} \left| \prod_{i=1}^n \mathbb{P}(X_i = z_i) - \prod_{i=1}^n \mathbb{P}(Y_i = z_i) \right| dz$$
$$\leq \int_{\mathbb{R}^n} \sum_{i=1}^n \left| \mathbb{P}(X_i = z_i) - \mathbb{P}(Y_i = z_i) \right| dz$$
$$= \sum_{i=1}^n \int_{\mathbb{R}} \left| \mathbb{P}(X_i = z_i) - \mathbb{P}(Y_i = z_i) \right|$$

$$= \sum_{i=1}^{n} d_{TV}(\mathcal{L}(X_i), \mathcal{L}(Y_i)).$$

Next we show the claim for arbitrary j. Let A be a set in $\mathbb{R}^{n-j}$, and let $B = \{(z_1, \dots z_n) \,|\, (z_1 + \cdots + z_j, z_{j+1}, \dots, z_n) \in A\}$. Observe

$$\begin{aligned} \big|\, \mathbb{P}(\mathcal{X}_j \in A) - \mathbb{P}(\mathcal{Y}_j \in A)| &= \left| \mathbb{P}\left(\bigotimes_{i=1}^{n} X_i \in B \right) - \mathbb{P}\left(\bigotimes_{i=1}^{n} Y_i \in B \right) \right| \\ &\leq \sum_{i=1}^{n} d_{TV}(\mathcal{L}(X_i), \mathcal{L}(Y_i)). \end{aligned}$$

Therefore

$$d_{TV}(\mathcal{X}_j, \mathcal{Y}_j) = \sup_A \big|\, \mathbb{P}(\mathcal{X}_j \in A) - \mathbb{P}(\mathcal{Y}_j \in A)| \leq \sum_{i=1}^{n} d_{TV}(\mathcal{L}(X_i), \mathcal{L}(Y_i)).$$

□

Claim 5.8.4. *Let X and Y be random variables, and $p \in [0,1]$. Then*

$$d_{TV}(\mathcal{L}(Bin(X,p)), \mathcal{L}(Bin(Y,p))) \leq d_{TV}(\mathcal{L}(X), \mathcal{L}(Y)).$$

Proof. Recall

$$d_{TV}(Bin(X,p), Bin(Y,p)) = \sup_A \big|\, \mathbb{P}(Bin(X,p) \in A) - \mathbb{P}(Bin(Y,p) \in A)|.$$

Let A be a set. Observe

$$\begin{aligned} &\big|\, \mathbb{P}(Bin(X,p) \in A) - \mathbb{P}((Y,p) \in A)\big| \\ &= \left| \int \mathbb{P}(X = s)\, \mathbb{P}(Bin(s,p) \in A) - \mathbb{P}(X = s)\, \mathbb{P}(Bin(s,p) \in A) ds \right| \\ &\leq \int \big|\, \mathbb{P}(X = s) - \mathbb{P}(Y = s)\big| ds \leq d_{TV}(\mathcal{L}(X), \mathcal{L}(Y)). \end{aligned}$$

□

Fact 5.8.5. *Let $X \sim \mathrm{Pois}(a)$ and $Z \sim Bin(X, p)$. Then $Z \sim \mathrm{Pois}(ap)$.*

Lemma 5.8.6. *[Bar88] Let $(X_i)_{i=1}^n$ be independent elements of $(\mathbb{Z}^+)^d$ with distributions*

$$\mathbb{P}(X_i = e_j) = p_{ji}, \quad p_i = \sum_{j=1}^{d} p_{ji}, \leq 1 \quad \mathbb{P}(X_i = 0) = 1 - p_i.$$

Let

$$\lambda_j = \sum_{i=1}^{n} p_{ji}, \quad c_\lambda = \frac{1}{2} + \log 2 \sum_{j=1}^{d} \lambda_j, \text{ and } \quad S = \sum_{i=1}^{n} X_i$$

Then

$$d_{TV}\left(\mathcal{L}(S), \bigotimes_{j=1}^{d} \mathrm{Pois}(\lambda_j)\right) \leq \sum_{i=1}^{n} \min\{c_\lambda \sum_{j=1}^{d} p_{ji}^2/\lambda_j, p_i^2\}.$$

Fact 5.8.7. *Let Y_n be a sequence of random variables. If $Y_n \to Y$ in distribution, then $\mathrm{Pois}(Y_n) \to \mathrm{Pois}(Y)$ in distribution.*

Lemma 5.8.8. *[CL06] Let X_i be independent random variables satisfying $X_i \leq M$ for $i \in [n]$. Let $X = \sum_{i=1}^n X_i$ and $Z = \sum_{i=1}^n \mathbb{E}(X_i^2)$. Then*

$$\mathbb{P}(X \geq \mathbb{E}(X) + \lambda) \leq \mathsf{exp}\left(\frac{-\lambda^2}{2(Z + M\lambda/3)}\right).$$

Lemma 5.8.9. *(Chernoff) Let $X \sim Bin(n, p)$. Then for all $\delta > 0$,*

$$\mathbb{P}(X \geq (1+\delta)\,\mathbb{E}(X)) \leq \mathsf{exp}\left(\frac{-\delta^2\,\mathbb{E}(X)}{2+\delta}\right).$$

Lemma 5.8.10 (Rescaling lemma, [Bor+])**.** *Given a sequence of multigraphs $(G_n)_{n\geq 1}$ and real numbers $(\ell_n)_{n\geq 1}$, suppose that $\lim_{n\to\infty} \frac{2e(G_n)}{\ell_n} = c$. Further, let $\mathrm{Lbl}_{\sqrt{\ell_n}}(G_n) \xrightarrow{d} \xi_{\mathcal{W}}$ for some multigraphex $\mathcal{W} = (W, S, I)$. Then $\mathrm{Lbl}(G_n) \xrightarrow{d} \xi'$, where $\xi' = \xi_{\mathcal{W}'}$, $\mathcal{W}' = (W', S', I')$*

with

$$W'(x,y,\cdot) = W(\sqrt{c}x, \sqrt{c}y, \cdot), \quad S'(x,\cdot) = \frac{1}{\sqrt{c}} S(\sqrt{c}x, \cdot), \quad \textit{and} \quad I'(\cdot) = \frac{I(\cdot)}{c}.$$

Fact 5.8.11. *Let X_n be a sequence of random variables that converges weakly to X. Let Y_n be a sequence of random variables such that $d_{TV}\left(\mathcal{L}\left(X_n\right), \mathcal{L}\left(Y_n\right)\right) \to 0$. Then Y_n converges to X weakly.*

Lemma 5.8.12. *[KV00] Let H be a weighted hypergraph with $V(H) = \{1, 2, \ldots n\}$. Each edge $e \in E(H)$ of H has a weight $w(e)$. For $i = 1, \ldots n$, let t_i be an independent $\{0,1\}$ random variable with expected value p_i. Consider a polynomial*

$$Y_H = \sum_{e \in E(H)} w(e) \prod_{s \in e} t_s.$$

For each vertex subset $A \subset V(H)$, define

$$Y_{H_A} = \sum_{e, A \subset e} w(e) \prod_{i \in e \setminus A} t_i.$$

Let $\mathbb{E}_i(H) = \max_{A \subset V(H), |A| = i} \mathbb{E}(Y_{H_A})$, $\mathbb{E}(H) = \max_{i \geq 0} \mathbb{E}_i(H)$, and $\mathbb{E}_z(H) = \max_{i \geq 1} \mathbb{E}_i(H)$. Then

$$\mathbb{P}\left(|Y_H - \mathbb{E}(Y_H)| \geq 8\gamma \sqrt{\mathbb{E}_z(H)\, \mathbb{E}(H)}\right) = O\left(\exp\left(-\gamma\right)\right).$$

5.8.1 Proof of Theorem 5.3.2

We begin by proving the Subset Switching Lemma.

Proof. (of Lemma 5.6.3) Let G be a obtained from $MMCM(D, B)$. Establish a pairing convention where the half-edges are labeled with natural numbers, and in each step the unpaired half-edge with the lowest label is paired to a uniformly random eligible half-edge.

Then deletion happens according to the outcome of the Bernoulli trial determined by the colors of the paired half-edges. Let P_0 denote a pairing obtained by following this convention and G_0 the corresponding graph after the deletion step. Let $P_0(m)$ denote the first m pairs in P_0, and let $G_0(m)$ denote the edges in $P_0(m)$ that survived the deletion step. Define the Doob martingale

$$Y_m(G_0) = \mathbb{E}(f(G) \mid G_0(m) \subseteq G),$$

so that

$$Y_0(G_0) = \mathbb{E}(f(G)) \quad \text{and} \quad Y_{\ell_n/2}(G_0) = f(G)$$

when G_0 are the edges that form G. To apply the Azuma-Hoeffding inequality (**??**) and conclude the lemma, we show that the martingale has differences bounded by b.

For a given P_0 let i be the next half-edge paired after the first m pairings occur. For any j that does not appear in $P_0(m)$, let S_j be the set of all pairings that contain $P_0(M) \cup (i,j)$. For any j,k, we define a one-to-one correspondence between pairings in S_j and pairings in S_k. For $P \in S_j$ there is a $P' \in S_k$ defined by switching $\{(i,j),(k,\ell)\} \leftrightarrow \{(i,k),(j,\ell)\}$ where ℓ is the partner of k in P. Next for each P and corresponding P' we construct a one-to-one correspondence between sets of four graphs $\mathcal{G} = \{G_{00}, G_{10}, G_{01}, G_{11}\}$ and $\mathcal{G}' = \{G'_{00}, G'_{10}, G'_{01}, G'_{11}\}$. The graphs in $\mathcal{G} \cup \mathcal{G}'$ contain the edges of $G_0(m)$, and for each edge in both P and P' the outcome of the deletion step is the same for all the graphs in $\mathcal{G} \cup \mathcal{G}'$. The graph G_{ab} has the edge (i,j) if and only if $a = 1$, and has the edge (k,ℓ) if and only if $b = 1$. Similarly graph G'_{ab} has the edge (i,k) if and only if $a = 1$, and has the edge (j,ℓ) if and only if $b = 1$. Since it is equally likely for the pairing to be in any S_k and it is equally likely that the graph is in $\mathcal{G}$ and $\mathcal{G}'$,

$$|Y_m(G_0) - Y_{m+1}(G_0)| \leq max_{G \in \mathcal{G}, G' \in \mathcal{G}'} \big|f(G) - f(G')\big| \leq b.$$

The final inequality follows from noting that for all $G \in \mathcal{G}$ and $G' \in \mathcal{G}'$, $G \overset{s}{\sim} G'$. $\square$

Lemma 5.8.13. *Let $H_n \sim MMCM(D_n, B)$ where ℓ_n is the sum of half-edges before deletion and ℓ_n^a is sum of half-edges of color a before deletion. Let $E(H_n)$ denote the number of edges in H_n. Assume $\ell_n = \Omega(n)$ and $\ell_n^a = \Theta(\ell_n)$. Almost surely*

$$\lim_{n\to\infty} \frac{2E(H_n)}{\mathbb{E}(E(H_n))} = 1.$$

Proof. We prove the statement by showing that for any $\varepsilon > 0$

$$\sum_{n=1}^{\infty} \mathbb{P}(|E(H_n) - \mathbb{E}(E(H_n))| \geq \varepsilon\, \mathbb{E}(E(H_n))) \leq \infty,$$

and applying the Borel-Cantelli lemma.

To give an upper bound for $\mathbb{P}(|E(H_n) - \mathbb{E}(E(H_n))| \geq \varepsilon\, \mathbb{E}(E(H_n)))$ we (i) use a martingale argument to show concentration of the number of edges with each combination of colors, then (ii) show concentration of the number of edges that remain after the deletion step given the number of edges with each color pair is close to its expectation.

We begin with (i). Let H_n' be a graph drawn from $MMCM(D_n, B)$ before the deletion step. Let $a, b \in [k]$ be a pair of colors. We call an edge before deletion an (a, b) edge if consists of a half-edge of color a and a half-edge of color b. Let $f_{ab}(H_n')$ be the random variable for the number of (a,b) edges in H_n', and so

$$\mathbb{E}(f_{ab}(H_n')) = \begin{cases} \frac{\ell_n^a \ell_n^b}{\ell_n - 1} & a \neq b \\ \frac{(\ell_n^a)^2}{2(\ell_n - 1)} & a = b. \end{cases} \tag{5.28}$$

Note that if $G \sim G'$, then $|f(G) - f(G')| \leq 2$. It follows by Lemma 5.6.2 that

$$\mathbb{P}(|f_{ij}(H_n') - \mathbb{E}(f_{ab}(H_n'))| \geq \delta\, \mathbb{E}(f_{ab}(H_n'))) \leq 2\exp\left(\frac{-\delta^2\, \mathbb{E}(f_{ab}(H_n'))^2}{4\ell_n}\right).$$

We say an initial pairing is "bad" if $|f_{ij}(H_n') - \mathbb{E}(f_{ab}(H_n'))| \geq \delta\, \mathbb{E}(f_{ab}(H_n'))$ for some $a, b \in [k]$ and "good" otherwise. Let $m_n = \min_a \ell_n^a/\ell_n$. We compute

$$\mathbb{P}(\text{ bad pairing }) \leq \sum_{a,b\in[k]} 2\exp\left(\frac{-\delta^2\, \mathbb{E}(f_{ab}(H_n'))^2}{4\ell_n}\right) \leq 2k^2\exp\left(\frac{-\delta m_n^2 \ell_n}{8}\right).$$

Next we show concentration of the number of edges conditioned on a good pairing. First we compute the expected number of edges by summing Equation (5.28),

$$T := \mathbb{E}(E(H_n)) = \sum_{a<b,a,b\in[k]} \frac{\ell_n^a \ell_n^b B_{ab}}{\ell_n - 1} + \sum_{a\in[k]} \frac{\ell_n^a \ell_n^b B_{aa}}{2(\ell_n - 1)}.$$

We write the number of edges in H_n after deletion as a random variable

$$Y = \sum_{a\leq b} \sum_{t=1}^{f_{ab}(H_n')} Y_t^{ab}$$

where Y_t^{ab} is the indicator random variable for the event that the t^{th} (a, b) edge is not deleted. Therefore $Y_t^{ab} \sim Bern(B_{ab})$. If the pairing H_n' is good, then Y is strictly dominated by the random variable

$$Z = \sum_{a\leq b} \sum_{t=1}^{\mathbb{E}(f_{ab}(H_n'))(1+\delta)} Z_t^{ab} \quad \text{where} \quad Z_t^{ab} \sim Bern(B_{ab}).$$

Note $\mathbb{E}(Z) = T\,(1+\delta)$, and so $(1+\varepsilon)T = \frac{1+\varepsilon}{1+\delta}\,\mathbb{E}(Z) \geq (1+\varepsilon-\delta)\,\mathbb{E}(Z)$. We compute via

Lemma 5.8.8

$$\mathbb{P}(Y \geq (1+\varepsilon)T \mid H_n' \text{ good}) \leq \mathbb{P}(Z \geq (1+\varepsilon)T) \leq \mathbb{P}(Z \geq (1+\varepsilon-\delta)\,\mathbb{E}(Z)) \\ \leq \exp\left(\frac{(\varepsilon-\delta)^2\,\mathbb{E}(Z)}{2\left(1+\frac{\varepsilon-\delta}{3}\right)}\right) \leq \exp\left(\frac{-(\varepsilon-\delta)^2\,T}{4}\right).$$

A similar argument shows that

$$\mathbb{P}(Y \leq (1-\varepsilon)T \mid H_n' \text{ good}) \leq \exp\left(-\frac{(\varepsilon-\delta)^2\,T}{4}\right).$$

Finally we compute

$$\mathbb{P}(|E(H_n)-T| \geq \varepsilon T) \leq \mathbb{P}(|E(H_n)-T| \geq \varepsilon T \mid \ H_n' \text{ good }) + \mathbb{P}(H_n' \text{ bad }) \\ \leq 2\exp\left(-\frac{(\varepsilon-\delta)^2\,T}{4}\right) + 2k^2\exp\left(\frac{-\delta m_n^2 \ell_n}{8}\right)$$

Letting $\delta = \varepsilon/2$, it follows that

$$\sum_{n=1}^{\infty} \mathbb{P}(|E(H_n)-T| \geq \varepsilon T) \leq \sum_{n=1}^{\infty} \exp\left(-\Theta\left(\ell_n\right)\right) < \infty.$$

□

Lemma 5.8.14. *Let $H_n \sim MMCM(D_n, B)$ where ℓ_n is the sum of degrees before deletion and $e_n = \sqrt{2\,\mathbb{E}(E(H_n))}$. For ease of notation, denote $\bar{L}_{H_n} = \mathrm{Lbl}\left(H_n, \sqrt{2e_n}\right)$ and $L_{MMCM_n} = \mathrm{Lbl}\left(MMCM(D_n, B)\right)$ Let $A_1, \ldots A_k, B_1, \ldots B_k \in \mathscr{B}(\mathbb{R}_+)$, $A = \bigcup_{i=1}^k A_i$, $B = \bigcup_{i=1}^k B_i$, $j \in \mathbb{N}^+$, and $\delta > 0$. Let*

$$P_n(j_1, j_2, \ldots j_k) = \mathbb{P}\big(\bar{L}_{H_n}\left(A_1 \times B_1\right) = j_1 \cap \cdots \cap \bar{L}_{H_n}\left(A_k \times B_k\right) = j_k\big)$$

$$P_{MMCM}(j_1, j_2, \ldots j_k) = \mathbb{P}(L_{MMCM_n}\left(A_1 \times B_1\right) = j_1 \cap \cdots \cap L_{MMCM_n}\left(A_k \times B_k\right) = j_k).$$

Then

$$\mathbb{P}\big(\big|P_n(j_1, j_2, \dots j_k) - P_{CCM}(j_1, j_2, \dots j_k)\big| > \delta\big) \leq 2\exp\left(\frac{-\delta^2 e_n^2}{16\ell_n \Lambda(A)^2 \Lambda(B)^2}\right).$$

Proof. We apply Lemma 5.6.3 with $f(G) = P_n(j_1, j_2, \dots j_k)$. Let G and G' be such that $E(G) \setminus E(G') \subseteq \{(i,j),(k,\ell)\}$ and $E(G') \setminus E(G) \subseteq \{(i,k),(j,\ell)\}$. Note that if $\bar{L}_G(A_r \times B_r)$ and $\bar{L}_{G'}(A_r \times B_r)$ differ for any $r \in [k]$, then $(v_i, v_j) \in A \times B, (v_k, v_\ell) \in A \times B, (v_i, v_k) \in A \times B$, or $(v_j, v_\ell) \in A \times B$ (where v_i represents the label of the vertex attached to the half-edge i). This happens with probability at most $4\left(\frac{\Lambda(A)\Lambda(B)}{e_n}\right)$. □

Lemma 5.8.15. *Let $D = (D_n)_{n\geq 1}$ be a sequence of degree sequences in which the sum of the degrees (before deletion) ℓ_n tends to infinity with n. Let $\mathcal{E}_n(S, S')$ denote the number of edges created between the set of half-edges S and S' in the construction of $MMCM(D_n, B)$. Consider m disjoint subsets of half-edges $(S_j)_{j\in[m]}$. Let $s_j = (s_j^1, \dots s_j^k)$ the vector where s_j^i denotes the number of half-edges in S_j with color i, let $\bar{s}_j = |S_j|$, and assume $\bar{s}_j = O(\sqrt{\ell_n})$ for $j \in [m]$. (Note that not all the half-edges in S_j will be part of an edge in the graph.) Define $(\mathcal{E}_{ij})_{1\leq i\leq j\leq m}$ an independent collection such that*

$$\mathcal{E}_{ij} \sim \begin{cases} \text{Pois}\left(\frac{s_i^T B s_j}{\ell_n}\right) & \text{for } i \neq j \\ \text{Pois}\left(\frac{s_i^T B s_i}{2\ell_n}\right) \end{cases}$$

Let $R_n = \mathcal{L}\left((\mathcal{E}_n(S_i, S_j))_{1\leq i\leq j\leq m}\right)$ and $R = \mathcal{L}\left((\mathcal{E}_{ij})_{1\leq i\leq j\leq m}\right)$ Then as $n \to \infty$, the total variation distance

$$d_{TV}(R_n, R) \to 0. \tag{5.29}$$

Moreover, if the S_j are random disjoint subsets satisfying $\mathbb{E}(\bar{s}_j) = O\left(\sqrt{\ell_n}\right)$ for all $i \in [k]$, then the error in 5.29 converges to zero in expectation.

Proof. Let S_i^a denote the half-edges in S_i with color a. Let $\mathcal{E}'(S, S')$ be the number of edges

between S and S' before the deletion step. Note

$$\mathcal{E}_n(S_i, S_j) \sim \sum_{(a,b)\in[k]\times[k]} Bin\left(\mathcal{E}'_n\left(S_i^a, S_j^b\right), B_{ab}\right).$$

Let

$$\mathcal{P}_{ij}^{ab} \sim \begin{cases} \text{Pois}\left(\frac{s_i^a s_i^a}{2\ell_n}\right), & i=j, a=b \\ \text{Pois}\left(\frac{s_i^a s_j^b}{\ell_n}\right) & \text{otherwise.} \end{cases}$$

By Fact 5.8.5

$$\mathcal{E}_{ij} = \sum_{(a,b)\in M} \mathcal{E}_{ij}^{ab} \quad \text{where} \quad \mathcal{E}_{ij}^{ab} \sim Bin\left(\mathcal{P}_{ij}^{ab}, B_{ab}\right).$$

Claim 5.8.3 and Claim 5.8.4 imply

$$d_{TV}\left(R_n, R\right) \leq d_{TV}\left(\mathcal{L}\left(\left(\mathcal{E}'_n\left(S_i^a, S_j^b\right)\right)_{(a,b)\in[k]\times[k],1\leq i\leq j\leq m}\right), \mathcal{L}\left(\left(\mathcal{P}_{ij}^{ab}\right)_{(a,b)\in[k]\times[k],1\leq i\leq j\leq m}\right)\right).$$

Since the initially pairing of the edges (before deletion) occurs according to a configuration model, we can use Lemma 5.6.1 to bound the right hand side. Therefore when $\bar{s}_j = O\left(\sqrt{\ell_n}\right)$ for all j,

$$d_{TV}\left(R_n, R\right) = O\left(\frac{1}{\sqrt{\ell_n}}\right).$$

Finally we consider the case when S_j are random disjoint subsets satisfying $\mathbb{E}(\bar{s}_j) = O\left(\sqrt{\ell_n}\right)$. By Markov's inequality, $\mathbb{P}\left(\bar{s}_j \geq \ell_n^{5/8}\right) \leq \ell_n^{-1/8}$. It follows that

$$d_{TV}\left(R_n, R\right) \leq \mathbb{P}\left(\exists \bar{s_j} = \Omega\left(\ell_n^{5/8}\right)\right) + \tilde{O}\left(\ell_n^{-1/8}\right) = \tilde{O}\left(\ell_n^{-1/8}\right).$$

□

Finally we prove Theorem 5.3.2.

Proof. (of Theorem 5.3.2).

Let $\mathcal{W} = \mathscr{W}(\gamma, a, B)$. We show that almost surely $\mathrm{Lbl}(H_n)$ converges weakly to $\xi_{\mathcal{W}}$ and apply Proposition 5.2.13 to conclude the theorem. By Lemma 5.8.13 and Lemma 5.8.10, to show that $\mathrm{Lbl}(H_n) \to \xi_{\mathcal{W}}$, it suffices to show that $\mathrm{Lbl}(H_n, \sqrt{2e_n}) \to \xi_{\mathcal{W}}$ where $e_n = \mathbb{E}(E(H_n))$.

For ease of notation we denote $\mathrm{Lbl}(H_n, \sqrt{2e_n}) = \bar{L}_{H_n}$ and $\mathrm{Lbl}(MMCM(D_n, B)) = L_{MMCM_n}$. As described in the proof of Theorem 5.2.7, it suffices to show that almost surely the joint distribution $\mathcal{L}\left(\left(\bar{L}_{H_n}(A_1), \ldots, \bar{L}_{H_n}(A_k)\right)\right)$ converges weakly to $\mathcal{L}\left(\left(\xi_{\mathcal{W}}(A_1), \ldots, \xi_{\mathcal{W}}(A_k)\right)\right)$ for an arbitrary finite family $A_1, \ldots A_k$.

Claim 1: The joint distribution $\mathcal{L}\left(\left(L_{MMCM_n}(A_1), \ldots, L_{MMCM_n}(A_k)\right)\right)$ converges weakly to the distribution $\mathcal{L}\left(\left(\xi_{\mathcal{W}}(A_1), \ldots, \xi_{\mathcal{W}}(A_k)\right)\right)$.

Note Lemmas 5.6.4 and 5.6.5 (with $\mu_n = \gamma_n$, $\mu = \gamma$, $X = B$), and Fact 5.8.7, imply that $\xi^P_{\gamma_n, B}$ converges weakly to $\xi_{\mathcal{W}}$, and so by Lemma 5.6.6, $\mathcal{L}\left(\left(\xi^P_{\gamma_n, M}(A_1), \ldots, \xi^P_{\gamma_n, M}(A_k)\right)\right)$ converges weakly to $\mathcal{L}\left(\left(\xi_{\mathcal{W}}(A_1), \ldots, \xi_{\mathcal{W}}(A_k)\right)\right)$. Lemma 5.8.15 implies that

$$d_{TV}\left(\mathcal{L}\left(L_{CCM_n}(A_1), \ldots, L_{CCM_n}(A_k)\right), \mathcal{L}\left(\xi^P_{\rho_n, M}(A_1), \ldots, \xi^P_{\rho_n, M}(A_k)\right)\right) \to 0.$$

The claim follows from Fact 5.8.11.

Claim 2: Almost surely

$$d_{TV}\left(\mathcal{L}\left(\left(\bar{L}_{H_n}(A_1), \ldots, \bar{L}_{H_n}(A_k)\right)\right), \mathcal{L}\left(\left(L_{MMCM_n}(A_1), \ldots, L_{MMCM_n}(A_k)\right)\right)\right) \to 0.$$

Let P_n and P_{MMCM_n} be as defined in Lemma 5.8.14 with rectangles $A_1, \ldots A_k$. The lemma implies

$$\mathbb{P}(|P_n(j_1, \ldots j_k) - P_{MMCM_n}(j_1, \ldots j_k)| > \delta) < \mathsf{exp}\left(-c\delta^2 \ell_n\right)$$

where c is a constant depending only on $A_1, \ldots A_k$. Under the assumption that $\ell_n = \Omega(\log(n))$,

$$\sum_{n=1}^{\infty} \mathsf{exp}\left(-c\delta^2 \ell_n\right) < \infty.$$

The remainder of the proof of this claim follows as Claim 2 in the proof of Theorem 5.2.7.

□

5.8.2 Proofs of Corollaries 5.3.5 and 5.3.6

To prove Corollaries 5.3.5 and 5.3.6, it suffices to show that for corresponding CCM and $MMCM$ parameter pairs the corresponding random adjacency measures $\xi^P_{\rho,M}$ and $\xi^P_{\gamma,B}$ are equal. Lemmas 5.8.16 and 5.8.17 and Theorems 5.2.7 and 5.3.2 directly imply Corollaries 5.3.5 and 5.3.6.

Lemma 5.8.16. *Let (D_n, M) be CCM parameters, and let $(\tilde{D}, B) = \Gamma_{C\to M}(D_n, M)$ be corresponding MMCM parameters. Let ρ and γ be the CCM and MMCM degree measures corresponding to the CCM parameters (D_n, M) and $MMCM$ parameters $(\tilde{D}, B)$ respectively. Then $\gamma = \rho$.*

Proof. Let $\bar{\ell}^i$ and $\bar{\ell}$ be the number of half-edges of color i and total half-edges in the MMCM respectively. We compute

$$\bar{\ell}^j = \sum_{v\in V} \frac{d_v^j \left(\sum_{i=1}^k \sqrt{\ell^i}\right)}{\sqrt{\ell^j}} = \sqrt{\ell^j}\left(\sum_{i=1}^k \sqrt{\ell^i}\right) \quad \text{and} \quad \bar{\ell} = \left(\sum_{i=1}^k \sqrt{\ell^i}\right)^2.$$

Let $\tilde{d}_v$ and $\hat{d}_v$ be as in the definitions of CCM and MMCM measures (Definitions 5.3.1 and 5.2.4). Note that

$$\hat{d}_v^j = \frac{\bar{d}_v^j}{\sqrt{\bar{\ell}}} = \frac{d_v^j}{\sqrt{\ell^j}} = \tilde{d}_v^j.$$

Since $\tilde{d}_v = \hat{d}_v$ and the expected number of edges between v and u is $\tilde{d}_v^T M \tilde{d}_u$ in the CCM and $\hat{d}_v^T M \hat{d}_u$ in the MMCM, it follows that the expected number of edges in the MMCM is $\ell/2$. Thus $\rho = \gamma$. □

Lemma 5.8.17. *Let (D_n, B) be MMCM parameters and let $(\tilde{D}, M) = \Gamma_{M\to C}(D_n, B)$ be corresponding CCM parameters. Let ρ and γ be the CCM and MMCM degree measures*

corresponding to the CCM parameters $(\tilde{D}_n, M)$ *and* $MMCM$ *parameters* (D, B) *respectively. Let* $\xi^P_{\rho,M}$ *and* $\xi^P_{\gamma,B}$ *be as defined in Equation* (5.15)*. Then* $\xi^P_{\rho,M} = \xi^P_{\gamma,B}$.

Proof. Let $\bar{\ell}^{ij}$ denote the number of half-edges of color ij in the CCM. We compute

$$\bar{\ell}^{ij} = \sum_{v \in V} \frac{B_{ij} \ell^i d^i_v}{\ell} = \frac{(\ell^i)^2 B_{ij}}{\ell}.$$

Let $\tilde{d}_v$ and $\hat{d}_v$ be as in the definitions of CCM and MMCM measures (Definitions 5.3.1 and 5.2.4). Note that

$$\hat{d}^{ji}_v \hat{d}^{ij}_u = \frac{\bar{d}^{ji}_u \bar{d}^{ij}_v}{\sqrt{\bar{\ell}^{ij}}\sqrt{\bar{\ell}^{ji}}} = \frac{d^i_v d^j_v B_{ij}}{\ell} = \tilde{d}^i_v \tilde{d}^j_u.$$

It follows that $\hat{d}^T_v B \hat{d}_u = \tilde{d}^T_v M \tilde{d}_u$. Since the expected number of edges between v and u is $\tilde{d}^T_v M \tilde{d}_u$ in the CCM and $\hat{d}^T_v M \hat{d}_u$ in the MMCM, it follows that the expected number of edges is the same across the models.

Finally we show $\xi^P_{\rho,M} = \xi^P_{\gamma,B}$. Note that since the expected number of edges is constant between the two models, we can couple the vertex labeling process in the construction of S (Equation (5.14)) in the definitions of ρ and γ. Since $\hat{d}^T_v B \hat{d}_u = \tilde{d}^T_v M \tilde{d}_u$, it follows that under this coupling $\xi^P_{\rho,M} = \xi^P_{\gamma,B}$. □

CHAPTER 6
A MARKOV CHAIN FOR THE HARD SPHERE MODEL

This Chapter is joint work with Will Perkins and Tyler Helmuth, and appears in [HPP20]. We prove that the single-center dynamics for the hard sphere model at fugacities $\lambda < 2^{1-d}$ mixes rapidly.

6.1 The hard sphere model, single-center dynamics, and critical fugacity

We begin by formally defining the hard sphere model and discussing its importance in mathematics and physics. Then we define single-center dynamics and state our main result that these dynamics are fast mixing for $\lambda < 2^{1-d}$. Finally, we outline how this result implies a lower bound on the critical density and fugacity of the model.

6.1.1 The hard sphere model

The hard sphere model is a simple but fundamental model for monatomic gases. The model has played a starring role in the development of Markov chain Monte Carlo methods since the beginning: the Metropolis algorithm was first applied to the study of the two-dimensional hard sphere model [Met+53]. Its theoretical importance is in part due to the fact that it (conjecturally) possesses a crystalline phase [BL15]. Understanding the phase diagram of the model has presented a significant challenge even at the level of physics [BK11], and mathematical results, including those presented here, are almost exclusively restricted to understanding the low-density phase (see [Ric16] for a notable exception). See [Löw00] for an inspiring introduction and broader overview of the model and its implications.

We now give a more precise definition of the hard sphere model. Let $r = r_d$ be the radius

such that a sphere in d dimensions has volume one. Let $\Lambda \subset \mathbb{R}^d$ be a bounded measurable set. Let

$$\Lambda_{\text{Int}} = \{x \in \Lambda : \text{dist}(x, \Lambda^c) \geq r\}.$$

The hard sphere model on volume Λ at fugacity $\lambda \geq 0$ with *free boundary conditions* is a Poisson point process of intensity λ on Λ_{Int} conditioned on the event that all points are at pairwise distance at least $2r$. We will denote the law of $\mathbf{X}$ by μ_Λ (the dependence on λ will be suppressed). Note that the requirement that spheres lie entirely within Λ instead of just requiring the centers to lie in Λ makes no difference in the infinite volume limit, but it does have a regularizing effect in finite volume.

We will also be interested in the hard sphere model with boundary conditions τ. More precisely, we define $\tau \subseteq \Lambda_{\text{Int}}$ as a set of forbidden locations for centers. The hard sphere model on a volume Λ at fugacity $\lambda \geq 0$ with boundary conditions τ is a Poisson point process of intensity λ on $\Lambda_{\text{Int}} \setminus \tau$ conditioned on the event that all points are at pairwise distance at least $2r$. One possibility is that τ represents the volume blocked by a set of permanently fixed spheres: if Y is a set of centers and $\tau = \Lambda_{\text{Int}} \cap (\cup_{y \in Y} B_{2r}(y))$, then $\Lambda_{\text{Int}} \setminus \tau$ is the set of locations for centers that do not overlap with spheres defined by the centers in Y. Note τ need not have this form. The law of the hard sphere model on Λ with boundary condition τ will be denoted by μ_Λ^τ.

6.1.2 Single-center dynamics

We consider the following Markov chain on Ω_Λ^τ, called the *single-center dynamics*. Given a configuration $X_t \in \Omega_\Lambda^\tau$, form X_{t+1} as follows:

1. Pick $x \in \Lambda$ uniformly at random.

2. With probability $1/(1 + \lambda)$, remove any $y \in X_t$ with $\text{dist}(x, y) < r$; that is, let $X_{t+1} = X_t \setminus B_r(x)$.

3. With probability $\lambda/(1+\lambda)$, attempt to add a center at x. That is, let $X' = X_t \cup \{x\}$. If $X' \in \Omega_\Lambda^\tau$, then set $X_{t+1} = X'$; if not, then set $X_{t+1} = X_t$.

We show in Lemma 6.3.1 below that the stationary distribution of this Markov chain is indeed μ_Λ^τ.

We will use the path coupling theorem (Theorem 6.2.2) to prove that the single-center dynamics are rapidly mixing at fugacities $\lambda < 2^{1-d}$.

Theorem 6.1.1. *Let $\Lambda \subset \mathbb{R}^d$ be compact and measurable, $n = |\Lambda|$, $\gamma \in (0,1)$, and let $\lambda = (1-\gamma)2^{1-d}$. The mixing time of the single-sphere dynamics with fugacity λ on Ω_Λ^τ satisfies*

$$t_{mix}(\varepsilon) \leq \left\lceil \frac{4n(\log(2^{d+2}n) - \log \varepsilon)}{\gamma} \right\rceil .$$

for all boundary conditions τ.

6.1.3 Rapid mixing and critical values

We define the *critical fugacity* $\lambda_c(d)$ as the supremum over λ such that the hard sphere model has a unique infinite volume limit in the sense of van Hove, i.e., such that the set of weak limit points of $\{\mu_{\Lambda,\lambda}\}_\Lambda$ is a singleton set. When $d = 1$, $\lambda_c(d) = \infty$, but it is not known for any $d \geq 2$ whether or not $\lambda_c(d) < \infty$. It is believed that $\lambda_c(d)$ is finite in dimension 3 (and in some or all dimensions $d \geq 4$), while the case $d = 2$ is subtle and remains an active area of investigation in physics [BK11, Tho+17].

We next define the density of the hard sphere model in dimension d at fugacity λ as

$$\rho(\lambda) = \liminf_{n\to\infty} \frac{1}{n} \mathbb{E}_{Q_n,\lambda} |\mathbf{X}| \,, \tag{6.1.1}$$

where Q_n is the d-dimensional cube of volume n centered at the origin and the expectation is with respect to the hard sphere model on Q_n at fugacity λ. The use of liminf in (6.1.1) is

necessary as *a priori* the limit is only known to exist for Lebesgue-a.e. values of λ. We then can define the *critical density* $\rho_c(d)$ of the hard sphere model as $\rho(\lambda_c(d))$ (or as $\lim_{\lambda\to\infty}\rho(\lambda)$ if $\lambda_c = \infty$). That is, $\rho_c(d)$ is the limiting expected packing density of the hard sphere model at the critical fugacity $\lambda_c(d)$.

Theorem 6.1.1 yields an improved lower bounds on the critical fugacity and density when combined with a new continuous analogue of the equivalence of spatial and temporal mixing from lattice spin systems established in [HPP20]. We state these results here to emphasize the consequences of our mixing time result. For proofs, see [HPP20].

Theorem 6.1.2. *For all $d \geq 2$, $\lambda_c(d) \geq 2^{1-d}$.*

Hofer-Temmel [HT19] used disagreement percolation [Ber93] and known bounds on the critical activity of d-dimensional Poisson-Boolean percolation to prove lower bounds on the critical fugacity of the hard sphere model. As $d \to \infty$, this gives a bound of $\lambda_c(d) \geq (1 + o(1))2^{-d}$. Recent work on developing exact sampling algorithms for the hard sphere model using the partial rejection sampling algorithm of Guo and Jerrum [JG a]. Guo and Jerrum showed that this algorithm is efficient in dimension 2 for $\lambda \leq .21027$ and Wellens improved this bound to $\lambda \leq .2344$ [Wel18]. For comparison with the previous results, our bound $\lambda_c(d) \geq 2^{1-d}$ is an improvement of a factor 2 as $d \to \infty$, and of more than 2 compared to the rigorous results in dimension 2.

Applying non-trivial lower bounds on the expected packing density of the hard sphere model from [JJP19] the lower bound on critical fugacity translates the following lower bound on the critical density.

Theorem 6.1.3. *For all $d \geq 2$, $\rho_c(d) \geq \frac{2}{3\cdot 2^d}$. As the dimension d tends to infinity we have $\rho_c(d) \geq (.8526 + o_d(1))2^{-d}$.*

6.2 Markov chain mixing basics

We use the following notation throughout this Chapter. $B_\ell(x)$ denotes the open ball of radius ℓ centered at $x \in \mathbb{R}^d$, and $V_\ell = |B_\ell(x)|$ will denote the volume of this set. In particular, $V_r = 1$. More generally, $|A|$ will denote the Lebesgue measure of $A \subset \mathbb{R}^d$. For $\Lambda \subset \mathbb{R}^d$ the *ℓ-parallel set* $\Lambda^{(\ell)}$ of Λ is $\{x \in \mathbb{R}^d : d(x, \Lambda) \leq \ell\}$. By an abuse of notation, if B is a finite set, we will write $|B|$ for the cardinality of B.

Let Ω denote the state space of a discrete time Markov chain. Let $p^{(0)}$ be the initial probability distribution on Ω, and let $p^{(t)}$ be the distribution after t steps of the Markov chain. Suppose the chain has a unique stationary distribution μ. The mixing time of the chain is a worst-case estimate for the number of steps it takes the Markov chain to approach stationarity. More precisely,

Definition 6.2.1. The mixing time of a Markov chain is

$$t_{\text{mix}}(\varepsilon) = \sup_{p^{(0)} \in \mathcal{P}} \min \left\{ t : \|p^{(t)} - \mu\|_{TV} \leq \varepsilon \right\} \tag{6.2.1}$$

where $\mathcal{P}$ denotes the set of probability distributions on Ω.

A common approach to bounding the mixing time of a Markov chain is to construct a coupling. For our purposes, a coupling of two Markov chains $(X_t)_{t\geq 0}$ and $(Y_t)_{t\geq 0}$ on Ω is a stochastic process $(X_t, Y_t)_{t\geq 0}$ with values in $\Omega \times \Omega$ such that the marginals $(X_t)_{t\geq 0}$ and $(Y_t)_{t\geq 0}$ are faithful copies of the Markov chains, and $X_{t+1} = Y_{t+1}$ whenever $X_t = Y_t$.

The path coupling theorem of Bubley and Dyer says that constructing a coupling for single steps of the Markov chains from certain pairs of configurations in Ω is sufficient for establishing an upper bound on the mixing time. To use this approach, one must represent the state space as the vertex set of a connected finite or infinite graph with a function $\hat{D} \geq 1$ defined on the edges. $\hat{D}$ is called the *pre-metric*. The *path metric* D corresponding to $\hat{D}$ is

the shortest path distance on the graph with edge weights given by $\hat{D}$, i.e.,

$$D(X,Y) = \inf_{\gamma\colon X \to Y} \left\{ \sum_{i=0}^{|\gamma|-1} \hat{D}(\gamma_i, \gamma_{i+1}) \right\}, \tag{6.2.2}$$

where the infimum is over nearest-neighbor paths $\gamma = (\gamma_0, \gamma_1, \ldots, \gamma_{|\gamma|})$ in the graph on Ω with $\gamma_0 = X$ and $\gamma_{|\gamma|} = Y$. To establish a rapid mixing regime for the single-center dynamics we will apply the version of Bubley and Dyer's path coupling theorem stated below. In the theorem, the diameter $\operatorname{diam}(\Omega)$ of Ω is $\sup_{X,Y\in\Omega} D(X,Y)$.

Theorem 6.2.2 ([LP17, Corollary 14.7])**.** *Suppose the state space Ω of a Markov chain is the vertex set of a connected graph, suppose $\hat{D}$ is a pre-metric on this graph, and let D be the corresponding path metric.*

Suppose that for each edge of this graph $\{X_0, Y_0\}$ the following holds: if $p^{(0)}$ and $q^{(0)}$ are the distributions concentrated on the configurations X_0 and Y_0 respectively, then there exists a coupling (X_1, Y_1) of the distributions $p^{(1)}$ and $q^{(1)}$ such that

$$\mathbb{E}\left[D(X_1, Y_1)\right] \leq D(X_0, Y_0)e^{-\alpha} = \hat{D}(X_0, Y_0)e^{-\alpha},$$

where $\mathbb{E}$ is the expectation with respect to the Markov chain. Then

$$t_{\mathrm{mix}}(\varepsilon) \leq \left\lceil \frac{-\log \varepsilon + \log(\operatorname{diam}(\Omega))}{\alpha} \right\rceil.$$

Remark 6.2.3. [LP17, Corollary 14.7] concerns Markov chains on finite state spaces, but the proof applies essentially verbatim to our context.

6.3 Proof that single-center dynamics are fast-mixing

To establish rapid mixing for the single-center dynamics, we follow the approach of Vigoda for the discrete hard-core model on bounded degree graphs [Vig01]. This approach makes use of an extended state space $\Omega^* \supseteq \Omega$. In our setting, let $\Omega_\Lambda^{\tau,*}$ be the collection of all sets of centers $X \subseteq \Lambda_{\text{Int}}$ such that each point in Λ is covered by at most two balls of radius r with a center in X, i.e.

$$X \in \Omega_\Lambda^{\tau,*} \iff \text{for all } x \in \Lambda, \quad |\{y \in X : \text{dist}(x,y) < r\}| \leq 2. \tag{6.3.1}$$

The purpose of this extended state space will become clear below when we introduce a pre-metric. Note that the boundary conditions τ play no role in the definition of $\Omega_\Lambda^{\tau,*}$. Next we extend our definition of the single-center dynamics to $\Omega_\Lambda^{\tau,*}$. At each step of the chain:

1. Pick $x \in \Lambda$ uniformly at random.

2. With probability $1/(1+\lambda)$, remove any $y \in X_t$ with $\text{dist}(x,y) \leq r$. That is, set $X_{t+1} = X_t \setminus B_r(x)$.

3. With probability $\lambda/(1+\lambda)$, attempt to add a center at x. Let $X' = X_t \cup \{x\}$. If $x \in \Lambda_{\text{Int}} \setminus \tau$ and $\text{dist}(x, X_t) \geq 2r$, then set $X_{t+1} = X'$. If not, set $X_{t+1} = X_t$. That is, we add a center at x if it locally satisfies the packing constraints and boundary conditions.

If $X_t \in \Omega_\Lambda^\tau$ then the chain will remain in Ω_Λ^τ and the dynamics agree with the definition given in Section 6.1.2. In addition, a Markov chain that starts in $\Omega_\Lambda^{\tau,*} \setminus \Omega_\Lambda^\tau$ will eventually reach Ω_Λ^τ. Since the chain has a unique invariant measure when considered on the state space Ω_Λ^τ, this shows the chain also has a unique invariant measure on $\Omega_\Lambda^{\tau,*}$, and that the mixing time of the chain on $\Omega_\Lambda^{\tau,*}$ is an upper bound for the mixing time of the chain on Ω_Λ^τ.

Throughout the remainder of this section, we fix the dimension d, the region $\Lambda \subset \mathbb{R}^d$, and the boundary conditions τ. For simplicity we write $\Omega = \Omega_\Lambda^\tau$ and $\Omega^* = \Omega_\Lambda^{\tau,*}$.

Lemma 6.3.1. *The stationary distribution of the single-center dynamics on Ω is the distribution of the hard sphere model on Ω.*

Proof. Consider two distinct configurations $X, Y \in \Omega$. The transition density between X and Y (and vice versa) is non-zero if and only if the symmetric difference $X \Delta Y$ is a singleton. Suppose without loss of generality that $Y = X \cup \{x\}$. Let π denote the density of μ, and let $\pi_U(V)$ denote the transition density from state U to state V. Then $\pi(Y)/\pi(X) = \lambda$, and $\pi_X(Y)/\pi_Y(X) = \lambda$, and so the chain is reversible with respect to the hard sphere measure on Ω. □

Since the single-center dynamics are a Harris recurrent chain, the previous lemma implies that μ is the unique invariant measure for the dynamics on Ω, and that $p^{(t)} \to \mu$ for all initial distributions $p^{(0)}$, see, e.g., [RR04, Section 3.2].

6.3.1 Proof of Theorem 6.1.1

We begin with some preliminary definitions. For $X \in \Omega^*$ let

$$\Gamma(X) = (\Lambda \setminus \Lambda_{\mathrm{Int}}) \cup \tau \cup \left(\bigcup_{x \in X} B_{2r}(x) \right). \tag{6.3.2}$$

This is the 'blocked volume' of a configuration X where an additional center cannot be placed.

For $v \in \Lambda$ we write the ball $B_{2r}(v)$ as the disjoint union of the *occupied* (or *blocked*) set $O_X(v)$ and the *unoccupied* (or *free*) set $U_X(v)$,

$$O_X(v) = B_{2r}(v) \cap \Gamma(X), \qquad U_X(v) = B_{2r}(v) \setminus \Gamma(X). \tag{6.3.3}$$

We now use these notions to define a pre-metric on Ω^*. For $X, Y \in \Omega^*$, we say that X

and Y are adjacent ($X \sim Y$) if X has exactly one more center than Y, and all the centers in X are also in Y (or vice versa). We define a pre-metric $\hat{D}(\cdot,\cdot)$ on adjacent states by

$$\hat{D}(X, X \cup \{v\}) = 2^d - c|O_X(v)|, \qquad c = \frac{\lambda 2^d}{2 + \lambda 2^d}. \tag{6.3.4}$$

For $0 \le \lambda \le 2^{1-d}$, $c \in [0, 1/2]$, and so $\hat{D}(X, X \cup \{v\}) \ge 2^{d-1} \ge 1$. Hence $\hat{D}$ is a pre-metric for such fugacities. Let D be the path metric on Ω^* obtained from $\hat{D}$.

The pre-metric $\hat{D}$ is the continuous analogue of the pre-metric introduced by Vigoda in [Vig01]. Defining the state space to be Ω^* rather than Ω affects the metric D. Consider a simple example with free boundary conditions in which Λ is a ball of radius $3r/2$. Then $\Omega = \emptyset \cup \bigcup_{x \in \Lambda_{\mathrm{Int}}} \{\{x\}\}$. For the state space Ω the graph of adjacent states is a star graph with center $\emptyset$, and so for non-empty distinct $X, Y \in \Omega$, $D(X,Y) = \hat{D}(X,\emptyset) + \hat{D}(Y,\emptyset) = 2^{d+1}$. In contrast, for the state space Ω^*, we have that $D(X,Y) \le \hat{D}(X, X \cup Y) + \hat{D}(Y, X \cup Y) = 2^{d+1}(1-c)$. This is relevant in our proof when we bound the distance between a pair of configurations using the triangle inequality applied with a third configuration that is in $\Omega^* \setminus \Omega$ (see (6.3.12)).

To apply Theorem 6.2.2 we will couple adjacent configurations using the following coupling.

Definition 6.3.2 (The identity coupling for the single-center dynamics)**.** The identity coupling for the single-center dynamics is defined as follows. If X_t and Y_t are separate instances of the single-center dynamics for μ_Λ^τ at time t, we couple them in a Markovian manner via the transition rule

- Choose a point $x \in \Lambda$ uniformly at random.
- With probability $1/(1+\lambda)$, in both X_t and Y_t delete any center in $B_r(x)$ to form X_{t+1} and Y_{t+1} respectively.

- With probability $\lambda/(1+\lambda)$, attempt to add a center at x in both X_t and Y_t.

Consider $X, Y \in \Omega^*$ with $Y = X \cup \{v\}$. Let X' and Y' denote the resultant states after one step of the Markov chains coupled according to the above identity coupling, and let

$$\Delta = D(X', Y') - D(X, Y) \tag{6.3.5}$$

denote the random change in distance between configurations. The next lemma bounds the expectation of Δ.

Lemma 6.3.3. *Let $X, Y \in \Omega^*$ such that $Y = X \cup \{v\}$. Let $\lambda = (1-\gamma)2^{1-d}$, with $\gamma \in (0,1)$. Then*

$$\mathbb{E}[\Delta] \leq \frac{2^d(2c-1)}{n(1+\lambda)} = -\frac{\gamma 2^d}{(2-\gamma)(1+\lambda)n} < 0. \tag{6.3.6}$$

Proof. Let $Y = X \cup \{v\}$. The change in distance Δ is a random variable whose value is function of the current configurations of the chains, the random point w chosen in a single step of the coupling, and whether or not the coupling tries to add or remove spheres. We begin with a case analysis of how Δ changes.

1. Let A_1 be the event the center v is removed from Y, i.e., the chain removes spheres and w lies within distance r of v. The probability of this event is $1/(n(1+\lambda))$. After A_1 occurs, $X' = Y'$, and so $\Delta = -D(X, Y)$. It follows that

$$\mathbb{E}[\Delta \cdot 1_{A_1}] = -\frac{1}{n(1+\lambda)} D(X, Y) = -\frac{2^d - c|O_X(v)|}{n(1+\lambda)} \tag{6.3.7}$$

2. Let A_2 denote the event that a center is added to X but not Y. This occurs when w lies in $U_X(v)$ and the coupling attempts to add a sphere, as $U_X(v)$ is blocked in Y and

not blocked in X. In this case we have $\Delta = D(X \cup \{w\}, Y) - D(X, Y)$. It follows that

$$\mathbb{E}\left[\Delta \cdot 1_{A_2}\right] = \frac{\lambda}{n(1+\lambda)} \int_{U_X(v)} (D(X \cup \{w\}, Y) - D(X, Y))\, dw. \tag{6.3.8}$$

3. Let A_3 be the event that a new center w is added to both X and Y. Note that this event only occurs when $w \in \Lambda \setminus \Gamma(Y)$ and the coupling adds a center. In this case

$$\Delta = -c|\{x \in U_X(v) : \ x \text{ is blocked by the new center } w\}|.$$

For $x \in U_X(v)$, let A_3^x be the event that x becomes blocked by the new center, i.e., that $X' = X \cup \{w\}$, $Y' = Y \cup \{w\}$ and $x \in O_{X \cup \{w\}}(v)$. In order for the event A_3^x to occur, it must be the case that $w \in B_{2r}(x) \setminus \Gamma(Y)$. Hence

$$\begin{aligned}
\mathbb{E}\left[\Delta \cdot 1_{A_3}\right] &= \mathbb{E}\left[\int_{U_x(v)} -c 1_{A_3^x}\, dx\right] \\
&= -\frac{c\lambda}{n(1+\lambda)} \int_{U_x(v)} \int_{\Lambda} 1_{w \in B_{2r}(x) \setminus \Gamma(Y)}\, dw\, dx \\
&= -\frac{c\lambda}{n(1+\lambda)} \int_{U_X(v)} |B_{2r}(x) \setminus \Gamma(Y)|\, dx
\end{aligned} \tag{6.3.9}$$

4. Let A_4 be the event that at least one center is removed in both X and Y, and v is not removed. Let S_w be the set of centers removed; since $w \notin B_r(v)$ we have $S_w = X \cap B_r(w) = Y \cap B_r(w)$. In this case,

$$\Delta = c|\{x \in O_X(v) : \ x \text{ is no longer blocked after } S_w \text{ is removed}\}|.$$

For $x \in O_X(v)$, let A_4^x be the event that $X' = X \setminus S_w$, $Y' = Y \setminus S_w$, and $x \in U_{X \setminus S_w}(v)$. If A_4^x occurs there is a center $b_x \in X$ that is the closest center to x that blocks x. In

particular, $b_x \in S_w$, and hence $w \in B_r(b_x)$. Using this observation we obtain

$$\begin{aligned}\mathbb{E}\left[\Delta \cdot 1_{A_4}\right] &= \mathbb{E}\left[\int_{O_x(v)} c1_{A_4^x}\, dx\right] \\ &\leq \frac{c}{n(1+\lambda)} \int_\Lambda \int_{O_x(v)} 1_{w \in B_r(b_x)}\, dx\, dw \\ &= \frac{c\,|O_X(v)|}{n(1+\lambda)}. \end{aligned} \tag{6.3.10}$$

The events A_1, A_2, A_3, and A_4 are mutually exclusive and exhaustive, so

$$\mathbb{E}[\Delta] = \mathbb{E}\left[\Delta \cdot \sum_{i=1}^{4} 1_{A_i}\right]. \tag{6.3.11}$$

To derive an upper bound on $\mathbb{E}[\Delta]$ we first need to estimate the integrand in (6.3.8). We will use the triangle inequality with the configurations $Y \cup \{w\}$, $X \cup \{w\}$, and Y. Temporarily deferring the justification of the use of the triangle inequality, note that since $c \geq 0$, $D(Y \cup \{w\}, X \cup \{w\}) \leq D(Y, X)$. Further, by definition, $D(Y \cup \{w\}, Y) = 2^d - c|B_{2r}(w) \cap \Gamma(Y)|$. Hence by the triangle inequality

$$\begin{aligned} D(X \cup \{w\}, Y) - D(X, Y) &\leq D(Y \cup \{w\}, X \cup \{w\}) + D(Y \cup \{w\}, Y) - D(X, Y) \\ &\leq 2^d - c|B_{2r}(w) \cap \Gamma(Y)|. \end{aligned} \tag{6.3.12}$$

To justify this use of the triangle inequality we must establish that $X \cup \{v, x\} \in \Omega^*$. Note that no point of Λ is covered by three balls of radius r whose centers are in Y because $Y \in \Omega^*$. No point that is covered by $B_r(x)$ is covered by $B_r(u)$ for some $u \in X$ since x is added to X by the Markov chain. It follows that no point in Λ is covered three times by $Y \cup \{x\}$, i.e., $Y \cup \{x\} \in \Omega^*$.

Inserting the estimates given in (6.3.7)–(6.3.10) into (6.3.11) we obtain

$$\begin{aligned}\mathbb{E}\left[\Delta\right] &\leq \frac{1}{n(1+\lambda)}\Bigg(-(2^d - c|O_X(v)|) + \lambda \int_{U_X(v)} (2^d - c|B_{2r}(w)\cap\Gamma(Y)|)\,dw \\ &\qquad\qquad - c\lambda \int_{U_X(v)} |B_{2r}(x)\setminus\Gamma(Y)|\,dx + c|O_X(v)|\Bigg) \\ &= \frac{1}{n(1+\lambda)}\left(-2^d + 2c|O_X(v)| + \lambda 2^d(1-c)|U_X(v)|\right),\end{aligned}$$

where the last line follows from $|B_{2r}(x)\cap\Gamma(Y)| + |B_{2r}(x)\setminus\Gamma(Y)| = 2^d$. Since $|U_X(v)| + |O_X(v)| = 2^d$ and $2c = \lambda 2^d(1-c)$, it follows that

$$\mathbb{E}\left[\Delta\right] \leq \frac{2^d(2c-1)}{n(1+\lambda)} = -\frac{\gamma 2^d}{(2-\gamma)(1+\lambda)n}\,. \qquad \square$$

Now we deduce Theorem 6.1.1 from Theorem 6.2.2.

Proof of Theorem 6.1.1. First we bound the diameter of Ω^*. Note that if $X \in \Omega^*$ then $|X| \leq 2n$ since each ball covers one unit of volume and each point cannot be covered more than twice. It follows that the combinatorial diameter of the graph representing the states of Ω^* is bounded above by $4n$. For two adjacent states $X \sim Y$, $D(X,Y) \leq \hat{D}(X,Y) \leq 2^d$, and hence $\operatorname{diam}(\Omega^*) \leq n2^{d+2}$.

Next we find a suitable value for α in the statement of Theorem 6.2.2. Let $X_0 = X$ and $Y_0 = X \cup \{v\}$. Applying Lemma 6.3.3 we obtain

$$\begin{aligned}\mathbb{E}\left[D(X_1,Y_1)\right] &= D(X_0,Y_0)\left(1 + \frac{\mathbb{E}\left[\Delta(X_0,Y_0)\right]}{D(X_0,Y_0)}\right) \\ &\leq D(X_0,Y_0)\left(1 - \frac{\gamma}{n(2-\gamma)(1+\lambda)}\right) \\ &\leq D(X_0,Y_0)\,\exp\left[-\frac{\gamma}{n(2-\gamma)(1+\lambda)}\right]\end{aligned}$$

$$\leq D(X_0, Y_0)e^{-\frac{\gamma}{4n}} .$$

The first inequality used that $\mathbb{E}[\Delta] < 0$ and $D(X_0, Y_0) \leq 2^d$, and the last used that $1 + \lambda \leq 2$. Applying Theorem 6.2.2 with $\alpha = \gamma/4n$ gives Theorem 6.1.1. □

REFERENCES

[Abb17a] E. Abbe. "Community detection and stochastic block models: recent developments". In: *arXiv preprint arXiv:1703.10146* (2017).

[Abb17b] E. Abbe. "Community detection and Stochastic Block Models: Recent developments". In: *J. Mach. Learn. Res.* 18.1 (2017), pp. 6446–6531.

[ABL10] Y.-Y. Ahn, J. P. Bagrow, and S. Lehmann. "Link communities reveal multiscale complexity in networks". In: *Nature* 466.7307 (2010), pp. 761–764.

[Ahn+07] Y.-Y. Ahn, S. Han, H. Kwak, S. Moon, and H. Jeong. *Analysis of topological characteristics of huge online social networking services*. 2007.

[Air+14] E. M. Airoldi, D. Blei, E. A. Erosheva, and S. E. Fienberg. *Handbook of Mixed Membership Models and Their Applications*. 1st. Chapman & Hall/CRC, 2014. ISBN: 1466504080, 9781466504080.

[Air+08] E. M. Airoldi, D. M. Blei, S. E. Fienberg, and E. P. Xing. "Mixed membership stochastic blockmodels". In: *Journal of Machine Learning Research* 9.Sep (2008), pp. 1981–2014.

[Air+06] E. M. Airoldi, D. M. Blei, S. E. Fienberg, E. P. Xing, and T. Jaakkola. "Mixed membership stochastic block models for relational data with application to protein-protein interactions". In: *Proceedings of the international biometrics society annual meeting*. 2006, pp. 1–34.

[Alo+09] N. Alon, E. Fischer, I. Newman, and A. Shapira. "A combinatorial characterization of the testable graph properties: it's all about regularity". In: *SIAM Journal on Computing* 39.1 (2009), pp. 143–167.

[AS08] N. Alon and A. Shapira. "Every monotone graph property is testable". In: *SIAM Journal on Computing* 38.2 (2008), pp. 505–522.

[Alo07] U. Alon. "Network motifs: theory and experimental approaches". In: *Nature Reviews Genetics* 8.6 (2007), pp. 450–461.

[App09] D. Applebaum. *Lévy processes and stochastic calculus*. Cambridge university press, 2009.

[Arr+15] R. I. Arriaga, D. Rutter, M. Cakmak, and S. S. Vempala. "Visual Categorization with Random Projection". In: *Neural Computation* 27.10 (2015), pp. 2132–2147.

[AV06] R. I. Arriaga and S. Vempala. "An algorithmic theory of learning: Robust concepts and random projection". In: *Machine Learning* 63.2 (2006), pp. 161–182.

[BH03] G. D. Bader and C. W. Hogue. "An automated method for finding molecular complexes in large protein interaction networks". In: *BMC bioinformatics* 4.1 (2003), p. 2.

[BKN11a] B. Ball, B. Karrer, and M. E. Newman. "Efficient and principled method for detecting communities in networks". In: *Phys. Rev. E* (2011). arXiv: 1104.3590.

[BKN11b] B. Ball, B. Karrer, and M. E. Newman. "Efficient and principled method for detecting communities in networks". In: *Physical Review E* 84.3 (2011), p. 036103.

[Bar88] A. D. Barbour. "Stein's method and Poisson process convergence". In: *Journal of Applied Probability* 25.A (1988), pp. 175–184.

[BC78] E. A. Bender and E. R. Canfield. "The asymptotic number of labeled graphs with given degree sequences". In: *J. Combin. Theory Ser. A* 24.3 (1978), pp. 296–307.

[BS01] I. Benjamini and O. Schramm. "Recurrence of Distributional Limits of Finite Planar Graphs". In: *Electron. J. Probab.* 6 (2001), 13 pp.

[Ber93] J. van den Berg. "A uniqueness condition for Gibbs measures, with applications to the 2-dimensional Ising antiferromagnet". In: *Communications in Mathematical Physics* 152.1 (1993), pp. 161–166.

[BK11] E. P. Bernard and W. Krauth. "Two-step melting in two dimensions: First-order liquid-hexatic transition". In: *Physical Review Letters* 107.15 (2011), p. 155704.

[BL15] X. Blanc and M. Lewin. "The crystallization conjecture: a review". In: *EMS Surveys in Mathematical Sciences* 2.2 (2015), pp. 255–306.

[Bol80] B. Bollobás. "A probabilistic proof of an asymptotic formula for the number of labelled regular graphs". In: *European J. Combin.* 1.4 (1980), pp. 311–316.

[Bop85] R. B. Boppana. "Amplification of probabilistic boolean formulas". In: *Foundations of Computer Science, 1985., 26th Annual Symposium on*. 1985, pp. 20–29.

[Bor+18] C. Borgs, J. T. Chayes, H. Cohn, and L. M. Lovász. "Identifiability for graphexes and the weak kernel metric". In: *arXiv:1804.03277* (2018). arXiv: 1804.03277.

[Bor+17a] C. Borgs, J. T. Chayes, H. Cohn, and V. Veitch. "Sampling perspectives on sparse exchangeable graphs". In: *To appear in Ann. Probab.* (2017). arXiv: 1708.03237.

[Bor+17b] C. Borgs, J. T. Chayes, H. Cohn, and V. Veitch. "Sampling perspectives on sparse exchangeable graphs". In: *arXiv preprint arXiv:1708.03237* (2017).

[Bor+14] C. Borgs, J. T. Chayes, H. Cohn, and Y. Zhao. "An L^p theory of sparse graph convergence I: limits, sparse random graph models, and power law distributions". In: *arXiv preprint arXiv:1401.2906* (2014).

[Bor+] C. Borgs, J. T. Chayes, S. Dhara, and S. Sen. "Limits of sparse configuration models and beyond: Graphexes and Multi-Graphexes". In: *In preparation* ().

[BP05] A. Brodsky and N. Pippenger. "The Boolean functions computed by random Boolean formulas or how to grow the right function". In: *Random Structures & Algorithms* 27.4 (2005), pp. 490–519.

[BH12] A. E. Brouwer and W. H. Haemers. "Strongly regular graphs". In: *Spectra of Graphs*. Springer, 2012, pp. 115–149.

[CF17] F. Caron and E. B. Fox. "Sparse graphs using exchangeable random measures". In: *J. R. Stat .Soc. Series B Stat. Methodol.* 79.5 (2017), pp. 1–44. arXiv: 1401.1137.

[CL02] F. Chung and L. Lu. "Connected components in random graphs with given expected degree sequences". In: *Annals of Combinatorics* 6.2 (2002), pp. 125–145.

[CL06] F. Chung and L. Lu. "Concentration inequalities and martingale inequalities: a survey". In: *Internet Mathematics* 3.1 (2006), pp. 79–127.

[CG08] G. Ciriello and C. Guerra. "A review on models and algorithms for motif discovery in protein–protein interaction networks". In: *Briefings in Functional Genomics and Proteomics* 7.2 (2008), pp. 147–156.

[COL09] A. Coja-Oghlan and A. Lanka. "Finding planted partitions in random graphs with general degree distributions". In: *SIAM Journal on Discrete Mathematics* 23.4 (2009), pp. 1682–1714.

[DVJ07] D. J. Daley and D. Vere-Jones. *An introduction to the theory of point processes: volume II: general theory and structure*. Springer Science & Business Media, 2007.

[DZ92] M. Dubiner and U. Zwick. "Amplification and percolation [probabilistic Boolean functions]". In: *Foundations of Computer Science, 1992. Proceedings., 33rd Annual Symposium on*. 1992, pp. 258–267.

[ER59] P. Erdős and A. Rényi. "On random graphs I". In: *Publ. Math. Debrecen* 6 (1959), pp. 290–297.

[ER60] P. Erdős and A. Rényi. "On the evolution of random graphs". In: *Publ. Math. Inst. Hung. Acad. Sci* 5.1 (1960), pp. 17–60.

[FV09] V. Feldman and L. G. Valiant. "Experience-Induced Neural Circuits That Achieve High Capacity". In: *Neural Computation* 21.10 (2009), pp. 2715–2754.

[FGG09] H. Fournier, D. Gardy, and A. Genitrini. "Balanced and/or Trees and Linear Threshold Functions". In: *Proceedings of the Meeting on Analytic Algorithmics and Combinatorics*. ANALCO '09. New York, New York: Society for Industrial and Applied Mathematics, 2009, pp. 51–57.

[FLZ19] J. Fox, L. M. Lovász, and Y. Zhao. "A fast new algorithm for weak graph regularity". In: *Combinatorics, Probability and Computing* 28.5 (2019), pp. 777–790.

[Fre16] P. E. Frenkel. "Convergence of graphs with intermediate density". In: *arXiv preprint arXiv:1602.05937* (2016).

[Fri86] J. Friedman. "Constructing O($nlogn$) Size Monotone Formulae for the k^{th} Threshold Function of n Boolean Variables". In: *SIAM Journal on Computing* 15.3 (1986), pp. 641–654.

[FK96] A. Frieze and R. Kannan. "The regularity lemma and approximation schemes for dense problems". In: *Foundations of Computer Science, 1996. Proceedings., 37th Annual Symposium on*. IEEE. 1996, pp. 12–20.

[FK15] A. Frieze and M. Karoński. *Introduction to Random Graphs*. Cambridge University Press, 2015.

[Gil59] E. N. Gilbert. "Random graphs". In: *The Annals of Mathematical Statistics* 30.4 (1959), pp. 1141–1144.

[GKS93] S. A. Goldman, M. J. Kearns, and R. E. Schapire. "Exact identification of read-once formulas using fixed points of amplification functions". In: *SIAM Journal on Computing* 22.4 (1993), pp. 705–726.

[GGR98] O. Goldreich, S. Goldwasser, and D. Ron. "Property testing and its connection to learning and approximation". In: *Journal of the ACM (JACM)* 45.4 (1998), pp. 653–750.

[Gow97] W. T. Gowers. "Lower bounds of tower type for Szemerédi's uniformity lemma". In: *Geometric & Functional Analysis GAFA* 7.2 (1997), pp. 322–337.

[HPP20] T. Helmuth, W. Perkins, and S. Petti. *Correlation decay for hard spheres via Markov chains*. 2020. arXiv: 2001.05323 [math.PR].

[HSM16] T. Herlau, M. N. Schmidt, and M. Mørup. "Completely random measures for modelling block-structured sparse networks". In: *Adv Neural Inf Process Syst 29*. Ed. by D. D. Lee, M Sugiyama, U. V. Luxburg, I Guyon, and R Garnett. Curran Associates, Inc., 2016, pp. 4260–4268.

[HT19] C. Hofer-Temmel. "Disagreement percolation for the hard-sphere model". In: *Electronic Journal of Probability* 24 (2019).

[HK02] P. Holme and B. J. Kim. "Growing scale-free networks with tunable clustering". In: *Physical review E* 65.2 (2002).

[HMP06] S. Hoory, A. Magen, and T. Pitassi. "Monotone circuits for the majority function". In: *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*. Springer, 2006, pp. 410–425.

[Jan17] S. Janson. "On convergence for graphexes". In: *arXiv:1702.06389* (2017). arXiv: 1702.06389.

[JJP19] M. Jenssen, F. Joos, and W. Perkins. "On the hard sphere model and sphere packings in high dimensions". In: *Forum of Mathematics, Sigma*. Vol. 7. Cambridge University Press. 2019.

[JG a] M Jerrum and H Guo. "Perfect Simulation of the Hard Disks Model by Partial Rejection Sampling". In: *Annales de l'Institut Henri Poincaré D* (to appear).

[KN11] B. Karrer and M. E. Newman. "Stochastic blockmodels and community structure in networks". In: *Physical Review E* 83.1 (2011), p. 016107.

[KV00] J. H. Kim and V. H. Vu. "Concentration of multivariate polynomials and its applications". In: *Combinatorica* 20.3 (2000), pp. 417–434.

[KW18a] J. M. Klusowski and Y. Wu. "Counting motifs with graph sampling". In: *Proceedings of the 31st Conference On Learning Theory*. Vol. 75. Proceedings of Machine Learning Research. PMLR, 2018, pp. 1966–2011.

[KW18b] J. M. Klusowski and Y. Wu. "Estimating the number of connected components in a graph via subgraph sampling". In: *arXiv:1801.04339* (2018). arXiv: 1801.04339.

[KS96] J. Komlós and M. Simonovits. "Szemerédi's regularity lemma and its applications in graph theory". In: (1996).

[Kro+06] N. J. Krogan, G. Cagney, H. Yu, G. Zhong, X. Guo, A. Ignatchenko, J. Li, S. Pu, N. Datta, A. P. Tikuisis, et al. "Global landscape of protein complexes in the yeast Saccharomyces cerevisiae". In: *Nature* 440.7084 (2006), p. 637.

[Kry19] I. Kryven. "Bond percolation in coloured and multiplex networks". In: *Nature communications* 10.1 (2019), p. 404.

[LP17] D. A. Levin and Y. Peres. *Markov chains and mixing times*. Vol. 107. American Mathematical Soc., 2017.

[Lov12] L. Lovász. *Large Networks and Graph Limits*. Vol. 60. Colloquium Publications. American Mathematical Society, 2012. ISBN: 978-0-8218-9085-1.

[LV14] L. Lovász and S. Vempala. Personal Communication. 2014.

[Löw00] H. Löwen. "Fun with hard spheres". In: *Statistical physics and spatial statistics*. Vol. 554. Springer, 2000, pp. 295–331.

[LMS98] M. Luby, M. Mitzenmacher, and M. A. Shokrollahi. "Analysis of Random Processes via And-Or Tree Evaluation." In: *SODA*. Vol. 98. 1998, pp. 364–373.

[Mah+06] P. Mahadevan, D. Krioukov, M. Fomenkov, X. Dimitropoulos, A. Vahdat, et al. "The Internet AS-level topology: three data sources and one definitive metric". In: *ACM SIGCOMM Computer Communication Review* 36.1 (2006), pp. 17–26.

[Met+53] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. "Equation of state calculations by fast computing machines". In: *The Journal of Chemical Physics* 21.6 (1953), pp. 1087–1092.

[Mil+02] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. "Network motifs: simple building blocks of complex networks". In: *Science* 298.5594 (2002), pp. 824–827.

[Mis+07] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee. *Measurement and analysis of online social networks*. 2007.

[MS56] E. F. Moore and C. E. Shannon. "Reliable circuits using less reliable relays". In: *Journal of the Franklin Institute* 262.3 (1956), pp. 191–208.

[New03] M. E. Newman. "The structure and function of complex networks". In: *SIAM review* 45.2 (2003), pp. 167–256.

[New05] M. E. Newman. "Random graphs as models of networks". In: *Handbook of Graphs and Networks*. Wiley-VCH Verlag GmbH Co. KGaA, 2005, pp. 35–68.

[New09] M. E. Newman. "Random graphs with clustering". In: *Physical review letters* 103.5 (2009), p. 058701.

[NG04] M. E. Newman and M. Girvan. "Finding and evaluating community structure in networks". In: *Physical review E* 69.2 (2004), p. 026113.

[NSW01] M. E. Newman, S. H. Strogatz, and D. J. Watts. "Random graphs with arbitrary degree distributions and their applications". In: *Physical review E* 64.2 (2001), p. 026118.

[ORS13] L. Ostroumova, A. Ryabchenko, and E. Samosvat. "Generalized preferential attachment: tunable power-law degree distribution and clustering coefficient". In: *WAW*. Springer, 2013, pp. 185–202.

[PBV07] G. Palla, A.-L. Barabási, and T. Vicsek. "Quantifying social group evolution". In: *arXiv preprint arXiv:0704.0744* (2007).

[PPV16] C. Papadimitriou, S. Petti, and S. Vempala. "Cortical computation via iterative constructions". In: *Conference on Learning Theory*. 2016, pp. 1357–1375.

[PV15a] C. H. Papadimitriou and S. S. Vempala. "Cortical Computation". In: *Proc. of PODC*. 2015.

[PV15b] C. H. Papadimitriou and S. S. Vempala. "Cortical Learning via Prediction". In: *Proc. of COLT*. 2015.

[PV18] S. Petti and S. S. Vempala. “Approximating Sparse Graphs: The Random Overlapping Communities Model”. In: *arXiv preprint arXiv:1802.03652* (2018).

[Rav+02] E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A.-L. Barabási. “Hierarchical organization of modularity in metabolic networks”. In: *Science* 297.5586 (2002), pp. 1551–1555.

[Ric16] T. Richthammer. “Lower bound on the mean square displacement of particles in the hard disk model”. In: *Communications in Mathematical Physics* 345.3 (2016), pp. 1077–1099.

[RR04] G. O. Roberts and J. S. Rosenthal. “General state space Markov chains and MCMC algorithms”. In: *Probability surveys* 1 (2004), pp. 20–71.

[Ros78] E. Rosch. “Principles of Categorization”. In: *Cognition and Categorization*. Ed. by E. Rosch and B. Lloyd. Lawrence Elbaum Associates, 1978.

[Ros+76] E. Rosch, C. B. Mervis, W. D. Gray, D. M. Johnson, and P. Boyes-braem. “Basic objects in natural categories”. In: *COGNITIVE PSYCHOLOGY* 8 (1976), pp. 382–439.

[RS96] R. Rubinfeld and M. Sudan. “Robust characterizations of polynomials with applications to program testing”. In: *SIAM Journal on Computing* 25.2 (1996), pp. 252–271.

[Ruk11] J. Rukavicka. “On generalized Dyck paths”. In: *The Electronic Journal of Combinatorics* 18.1 (2011), P40.

[Sav87] P. Savicky. “Boolean functions represented by random formulas”. In: *Commentationes Mathematicae Universitatis Carolinae* 28.2 (1987), pp. 397–398.

[Sav90] P. Savicky. “Random boolean formulas representing any boolean function with asymptotically equal probability”. In: *Discrete Mathematics* 83.1 (1990), pp. 95 –103.

[Ser04] R. A. Servedio. “Monotone Boolean formulas can approximate monotone linear threshold functions”. In: *Discrete Applied Mathematics* 142.1–3 (2004). Boolean and Pseudo-Boolean Functions, pp. 181 –187.

[SKP12] C. Seshadhri, T. G. Kolda, and A. Pinar. “Community structure and scale-free collections of Erdős-Rényi graphs”. In: *Physical Review E* 85.5 (2012), p. 056109.

[ST43] J. A. Shohat and J. D. Tamarkin. *The problem of moments*. 1. American Mathematical Soc., 1943.

[Son+05] S. Song, P. J. Sjöström, M. Reigl, S. Nelson, and D. B. Chklovskii. "Highly nonrandom features of synaptic connectivity in local cortical circuits". In: *PLoS biology* 3.3 (2005), e68.

[Sta97] R. P. Stanley. *Enumerative Combinatorics. Vol. 1, vol. 49 of Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1997.

[Ste78] P. R. Stein. "On a class of linked diagrams, I. Enumeration". In: *Journal of Combinatorial Theory, Series A* 24.3 (1978), pp. 357–366.

[Ste+05] U. Stelzl, U. Worm, M. Lalowski, C. Haenig, F. H. Brembeck, H. Goehler, M. Stroedicke, M. Zenkner, A. Schoenherr, S. Koeppen, et al. "A human protein-protein interaction network: a resource for annotating the proteome". In: *Cell* 122.6 (2005), pp. 957–968.

[Sze76] E Szemerédi. "Regular partitions of graphs, Colloques Internationaux CNRS No 260–Problemes Combinatoires et Théorie des Graphes, Orsay." In: *Computer Science Department Worcester Polytechnic Institute Worcester, MA 01609 USA* (1976), pp. 399–401.

[Sze75] E. Szemerédi. "On sets of integers containing k elements in arithmetic progression". In: *Acta Arithmetica* 27.1 (1975), pp. 199–245.

[Tho+17] A. L. Thorneywork, J. L. Abbott, D. G. Aarts, and R. P. Dullens. "Two-dimensional melting of colloidal hard spheres". In: *Physical review letters* 118.15 (2017), p. 158001.

[TMC16a] A. Todeschini, X. Miscouridou, and F. Caron. "Exchangeable Random Measures for Sparse and Modular Graphs with Overlapping Communities". In: *arXiv:1602.02114* (2016). arXiv: 1602.02114.

[TMC16b] A. Todeschini, X. Miscouridou, and F. Caron. "Exchangeable random measures for sparse and modular graphs with overlapping communities". In: *arXiv preprint arXiv:1602.02114* (2016).

[Val84] L. G. Valiant. "Short Monotone Formulae for the Majority Function". In: *Journal of Algorithms* 5.3 (1984).

[Val94] L. G. Valiant. *Circuits of the mind*. Oxford University Press, 1994. ISBN: 978-0-19-508926-4.

[Val00] L. G. Valiant. "A neuroidal architecture for cognitive computation". In: *J. ACM* 47.5 (2000), pp. 854–882.

[Val05] L. G. Valiant. "Memorization and Association on a Realistic Neural Model". In: *Neural Computation* 17.3 (2005), pp. 527–555.

[VR15] V. Veitch and D. M. Roy. "The class of random graphs arising from exchangeable random measures". In: *arXiv:1512.03099* (2015). arXiv: 1512.03099.

[VR16a] V. Veitch and D. M. Roy. "Sampling and Estimation for (sparse) exchangeable graphs". In: *arXiv:1611.00843, To appear in Ann. Stat.* (2016). arXiv: 1611.00843.

[VR16b] V. Veitch and D. M. Roy. "Sampling and Estimation for (Sparse) Exchangeable Graphs". In: *arXiv preprint arXiv:1611.00843* (2016).

[Vig01] E. Vigoda. "A note on the Glauber dynamics for sampling independent sets". In: *The Electronic Journal of Combinatorics* 8.1 (2001), p. 8.

[Vol04] E. Volz. "Random networks with tunable degree distribution and clustering". In: *Physical Review E* 70.5 (2004), p. 056115.

[WS98] D. J. Watts and S. H. Strogatz. "Collective dynamics of "small-world" networks". In: *Nature* 393.6684 (1998), pp. 440–442.

[Wel18] J. Wellens. "A note on partial rejection sampling for the hard disks model in the plane". In: *arXiv preprint arXiv:1808.03367* (2018).

[Wor+99] N. C. Wormald et al. "Models of random regular graphs". In: *London Mathematical Society Lecture Note Series* (1999), pp. 239–298.

[XSL11] J. Xie, B. K. Szymanski, and X. Liu. "Slpa: Uncovering overlapping communities in social networks via a speaker-listener interaction dynamic process". In: *Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on*. IEEE. 2011, pp. 344–349.

[YL+04] E. Yeger-Lotem, S. Sattath, N. Kashtan, S. Itzkovitz, R. Milo, R. Y. Pinter, U. Alon, and H. Margalit. "Network motifs in integrated cellular networks of transcription-regulation and protein-protein interaction". In: *Proceedings of the*

National Academy of Sciences of the United States of America 101.16 (2004), pp. 5934–5939.