Boston University Theses & Dissertations

2013

Network data analysis

BOSTON UNIVERSITY

GRADUATE SCHOOL OF ARTS AND SCIENCES

Dissertation

NETWORK DATA ANALYSIS

by

WESTON VILES

B.S., University of Maine, 2006
M.A., University of Maine, 2008

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2013

Approved by

First Reader

Eric D. Kolaczyk, PhD
Professor of Mathematics

Second Reader

Mark A. Kramer, PhD
Assistant Professor of Mathematics

Third Reader

Luis E. Carvalho, PhD
Assistant Professor of Mathematics

# Acknowledgments

Thank you, everyone.

Rocío González: mi amor.

Luis Carvalho and Ian Johnston: my brothers.

Yaonan Zhang and Lijun Peng: my sisters.

Ander Steele, Ross Sweet, and Ikemefuna Agbanusi: my original homies.

Lauren Viles: my strength.

Cheryl and David Viles: my courage.

Lingzi Lu:

My thesis is written in your honor.

I miss you very much, my friend ♡

...just trying to get by.

NETWORK DATA ANALYSIS

(Order No.)

WESTON VILES

Boston University, Graduate School of Arts and Sciences, 2013

Major Professor: Eric D. Kolaczyk, Professor of Mathematics

ABSTRACT

Frequently the graphs used to represent networks are inferred from data and, surprisingly, the uncertainty in their inferred topology is typically ignored. This dissertation makes significant contributions toward remedying this issue in three main ways.

Let $\eta(G)$ be a function of a network, i.e. a network characteristic, of the graph G and let $\hat{G}$ be an inferred network. I provide an analytical method for approximating the probability $\mathbb{P}\left(\eta\left(\hat{G}\right) - \eta(G) \leq x\right)$ via Stein's method for a broad class of functions, the so-called *motif* count statistics. The first contribution of my research is to quantify the propagation of error from network uncertainty to network characterization, a problem for which no results currently exist. Particular attention is given to the case of network density.

When modeling the functional connectivity of a brain, dynamic networks may be inferred from ECoG data. Of interest is an understanding of the process of coalescence and fragmentation of these functional brain networks during epileptic seizure. The second contribution of my research is to develop a means for tracing the community structures of an inferred dynamic network as they are born, merged, fractured, and die over time. In particular, I developed an algorithm which is robust to edge noise, yields well-defined spatio-temporal communities, and facilitates the connection of communities with a range of network motifs of various orders.

Models exist in which the rate of the emergence of a giant connected component, i.e. percolation, is manipulated through means of edge connection rules. However, such mod-

els are unable to account for (i) edge removal and (ii) the error associated with inferred networks, i.e. type-I and II errors. The third contribution of my research is to develop a model for graph percolation which allows for birth, death, and edge noise. I developed a method which accommodates both edge removal and a noisy edge set and propose a statistical hypothesis test to identify the rate of percolation. The novel approach of monitoring the relative size of the second component is found to be more robust than the first when distinguishing between percolation rates in the presence of noise.

Contents

List of Tables

List of Figures

List of Symbols and Abbreviations

G	Graph
V	Vertex Set
E	Edge Set
DPP ...	Dynamic Plex Percolation
CPM ...	Clique Percolation Method
QD	Quantile Difference
QR	Quantile Range
G_t	Time Series of Graphs
$G_t^{\dagger}$	Enhanced Time Series of Graphs
$MSGCC$	Maximum Size of the Giant Connected Component
H_0	Null Hypothesis
H_1	Alternative Hypothesis
$\hat{G}$	Estimator for Graph G
δ	Network Density
$\hat{\delta}$	Network Density Estimator
α	Probability of Type-I Error
β	Probability of Type-II Error
$\sim$	Follows in Probability Distribution
$O(\cdot)$	Big-O Notation

Chapter 1

Introduction

The analysis of networks and network data is widespread across scientific disciplines. Biologists, sociologists, computer scientists, economists, physicists, statisticians, and mathematicians require both theoretical and practical results in order to recognize and better understand the patterns in their relational data. For example, a biologist is interested in gene-gene interactions, a sociologist, in friendship relationships, a computer scientist, in World Wide Web hyperlink pointers, and a neuroscientist, in brain interactions. It has been the work of physicists, statisticians, and mathematicians dating back to at least Euler and his solution to the Bridges of Konigsberg Problem to analyze networks. In the following I focus on the *analysis of network data* which may be broadly considered to be "a core set of methods and models for the analysis of measurements that are either of or from a system conceptualized as a network." [Kol09].

Much work remains to more completely develop the theory and methods of the statistical analysis of network data. The network analysis literature frequently fails to fully quantify and understand the error associated with the inference of networks from data or "noisy" observations. My thesis focuses on the development of methods which may allow us to better understand the effect of noise (or error associated) in the construction of networks from data on the the inference of higher-order network characteristics.

In Section 2, I address the canonical statistical problem of estimation and inference. Few, if any, 'classical' results exist about the probability distribution of a network characteristic estimator. I develop a means by which to prove weak convergence of the probability distribution of a broad class of network characteristic estimators to an approximating probability distribution and establish the rate of said convergence. The class of motif counts, to be defined later, are fundamental to the characterization of a network. Examples of motifs include edges and triangles. The most basic, yet non-trivial, characteristic of a graph is perhaps its density defined simply to be the fraction of edges and vertex pairs. Naturally, during the construction of a network based upon inference from network data, some edges are missed and some are incorrectly identified causing a corruption in the motif count of number of edges which, as a result, corrupts the network density characteristic. Another example follows from the notion of transitivity of a network which is roughly a notion of the density of fully-connected vertex triples, i.e. triangles, in a graph. The work of [Wat99] and their so-called "small-world" network models incorporate the idea that we live in a small world in which many social relationships exhibit locally a triangle-like quality.

The methods I rely upon to approximate the probability distribution of a motif count estimator, and in particular the network density estimator, are non-classical. Using the Stein methodology I am able to determine a rate of convergence, and thus prove weak convergence in the Kolomogorov metric - even in the case of arbitrary dependence among the inferred edges, of a network motif count to an appropriately parameterized Skellam probability distribution. I illustrate the results in the small-world network setting.

In Section 3.1, I have developed a dynamic community detection algorithm, the Dynamic Plex Percolation method or (DPP), designed to detect and trace communities through a time series of network graphs. The novel and fundamental contributions of the DPP to the dynamic community detection or, for that matter, the static community detection literature is that noise in the edge set is directly addressed and that the evolution of network communities, to be defined later, is made explicit. Ultimately, the algorithm operates on a time series of network graphs and is able to recover dynamic mesoscopic structures of the network. The algorithm identifies structures in the series of networks which are more strongly connected than simple connectivity but not so tightly connected as cliques.

It is believed that changes in synchrony of voltage activity between brain regions is the manifestion of an epileptic seizure. The authors of [MAK08] and [MAKC10] analyze electrocorticogram (ECoG) recordings to produce functional networks of the brain pre-, present-, and post-seizure with the intention of better understanding the genesis of the seizure. Generally speaking, the strength of association between voltage recordings on the cortical surface of the brain are measured and if said strength is statistically significant then the vertices representing these location of the brain are connected by an edge. As such, the data can be represented as a time series of graphs and, by tracing the community structure across time with DPP, I propose a new tool to characterize the formation and fragmentation of communities during seizure. This ability may lead to a broader understanding of the origination of an epileptic seizure both in time and location and, hopefully, may provide new insights toward preventative measures for neuroscientists and neurosurgeons. I illustrate the method by tracing dynamic communities in a series of simulated random graphs.

In Section 4, we here consider different criteria for distinguishing between Erdos-Renyi (ER) and Achlioptas percolation under noise. A statistical model of explosive percolation is constructed allowing for the birth and death of edges as well as the presence of noise in the observations. This graph-valued stochastic process exhibits the Markov property, and is composed of a non-stationary stochastic process, which is treated as a latent process; and of a corresponding observed stochastic graph process, which is corrupted by Type I and Type II errors. This produces a hidden Markov graph model. We show that for certain choices of parameters controlling the noise model, the standard ER percolation is indistinguishable from the Achlioptas model. In this setting, we compare two different criteria for discriminating between these two percolation models, based on a quantile difference (QD) of the first component's size and on the maximal size of the second largest component. We show through data simulations that this second criterion outperforms the QD of the first component's size, in terms of discriminatory power. The maximal size of the second component therefore provides a useful statistic for distinguishing between the ER and Achlioptas models of percolation, under realistic conditions of birth/death and noise.

The potential application of the proposed models for percolation detection is discussed.

The results of this thesis are both theoretical and numerical, with practical applications to real-world data. The fundamental results of this thesis are contained in Sections 2, 3.1, and 4 with the proofs of results and technical details placed in Section 6.

Chapter 2

Uncertainty Propagation from Network Inference to Network Characterization

2.1 Introduction

The analysis of network data is widespread across scientific disciplines. Technological and infrastructure, social, biological, and information networks are a few of the network classes which have employed such analyses. Much work remains, however, to more fully develop the theory and methods of statistical analysis of network data. We refer the reader to the book [Kol09] for a comprehensive survey of the state of the science of statistical analysis of network data.

The focus of the this chapter is upon the canonical statistical problem of estimation and inference, but in the context of graphs. Frequently graphs used to represent networks are inferred from data and, surprisingly, the uncertainty incurred is ignored when characterizing the higher-level properties of these networks (e.g. density, clustering, centrality, etc.), see [Kol09]. Furthermore, said inference is commonly performed in an ad-hoc fashion; formal inference is less common but far more desirable for its ability to yield probabilistic statements on estimation accuracy.

A basic and fundamental problem in statistics is to identify or approximate the probability distribution of the discrepancy between a population parameter of interest and an estimator thereof to quantify estimation accuracy. We consider a variant of this problem in a network context wherein an underlying network-graph G possesses a network characteristic $\eta(G)$ of interest. If $\hat{G}$ is a network-graph inferred from data then $\eta\left(\hat{G}\right)$ is a natural and common estimator of $\eta(G)$, i.e. a plug in estimator. Let the discrepancy between these two quantities be defined as $D = \eta\left(\hat{G}\right) - \eta(G)$. Presently, and to our knowledge, there exist no general results making precise probabilistic statements about D.

In this work, we are the first to consider the propagation of error from network inference to network characterization, for a broad class of network motif counts. In the class of network motif counts we include motif counts scaled by a constant factor. Even for the most basic, non-trivial motif count of all, the network density, no machinery has yet been delevoped to measure in probabilistic terms the quantity $\hat{\delta} - \delta$ where, in this case, $\hat{\delta} = \eta\left(\hat{G}\right)$ and $\delta = \eta(G)$. We focus on the quantity $\eta\left(\hat{G}\right) - \eta(G)$ for the class of network motif counts and provide both finite sample and asymptotic results on an approximating probability distribution. Specifically, we present a Stein operator for the Skellam probability distribution and, in a manner consistent with the Stein methodology, we derive the rate of weak convergence of an arbitrary difference of sums of binary random variables to an appropriately parameterized Skellam distribution. We then apply our result to the class of network motif counts and characterize the rate of weak convergence to the Skellam distribution of the quantity $\eta\left(\hat{G}\right) - \eta(G)$.

2.2 Background

Suppose that $G = (V, E)$ is a network graph with $|V| = p$. We assume the vertex set V is known but that the edge set E is unknown. However, we assume there is information by which to infer E, as a set $\hat{E}$, yielding an inferred network graph $\hat{G}$. A principled approach to formally inferring an edge set on a known set of vertices is to perform $\binom{p}{2}$ hypothesis tests, perhaps based on vertex attributes, one for each of the vertex pairs $\{i, j\} \in V^{(2)}$ in the network graph G, of the form

$$H_0 : \{i, j\} \notin E \text{ versus } H_1 : \{i, j\} \in E$$

where

$$V^{(2)} = \{\{i,j\} : i,j \in V; i < j\}.$$

These tests amount to a collection of $\binom{p}{2}$ random variables $\mathbf{Y} = \left\{Y_{ij} : \{i,j\} \in V^{(2)}\right\}$ defined by

$$Y_{ij} = \begin{cases} 1 & \text{if } H_0 \text{ is rejected} \\ 0 & \text{if } H_0 \text{ is not rejected.} \end{cases}$$

In this way, the inferred edge set $\hat{E}$ is defined to be

$$\hat{E} = \left\{\{i,j\} \in V^{(2)} : Y_{ij} = 1\right\}.$$

Naturally, the set $\hat{E}$ is stochastic, as $\mathbf{Y} = \left\{Y_{ij} : \{i,j\} \in V^{(2)}\right\}$ is a collection of random variables, since associated with each hypothesis test there is a probability of test error. In particular, if $E^c = V^{(2)} \setminus E$ and $\{i,j\} \in E^c$ then $Y_{ij} \sim \text{Bernoulli}\,(\alpha_{ij})$, where α_{ij} is the probability of Type-I error for the test on vertex pair $\{i,j\} \in E^c$. On the other hand, if $\{i,j\} \in E$ then $Y_{ij} \sim \text{Bernoulli}\,(1 - \beta_{ij})$, where β_{ij} is the probability of Type-II error for the test on vertex pair $\{i,j\} \in E$.

The way in which we aim to probabilistically quantify the accuracy of $\hat{\delta} = \left|\hat{E}\right| / V^{(2)}$, as an estimator of the network characteristic $\delta = |E| / V^{(2)}$, is via the distribution of the discrepancy $D = \hat{\delta} - \delta$, where the latter may be written as

$$D = \hat{\delta} - \delta = \frac{\left|\hat{E}\right| - |E|}{\left|V^{(2)}\right|} = \frac{\sum_{\{i,j\} \in E^c} Y_{ij} - \sum_{\{i,j\} \in E} (1 - Y_{ij})}{\binom{p}{2}}.$$

Note that if λ_1 and λ_2 are defined to be

$$\lambda_1 = \mathbb{E}\left[\sum_{\{i,j\} \in E^c} Y_{ij}\right] \quad \text{and} \quad \lambda_2 = \mathbb{E}\left[\sum_{\{i,j\} \in E} (1 - Y_{ij})\right]$$

then λ_1 is the expected number of false-positives inferred and λ_2 is the expected number of false-negatives inferred in $\hat{G}$. It follows immediately that $\mathbb{E}D = (\lambda_1 - \lambda_2) \cdot \binom{p}{2}^{-1}$, so that $\hat{\delta}$

is an unbiased estimator of δ precisely when $\lambda_1 = \lambda_2$. This case will be of primary interest to us, although many of the intermediate results we develop will be for general λ_1 and λ_2.

The numerator

$$ND = \sum_{\{i,j\} \in E^c} Y_{ij} - \sum_{\{i,j\} \in E} (1 - Y_{ij})$$

of D is a difference of two sums of binary random variables. Therefore, one might hypothesize that ND has roughly the probability distribution of a difference of two binomial random variables and, furthermore, for tractability, attempt a Gaussian approximation. Recall, however, the inferred network is derived by the performance of $\binom{p}{2}$ hypothesis tests, clearly a multiple testing scenario. As we know, we must therefore adjust each individual type-I error rate to accommodate for multiple testing and, as such, the expected values of the binary random variables of ND decrease as p, the number of vertices in G, grows. Such an observation invalidates the assumptions of the binomial and, furthermore, the Gaussian approximation. Beyond this, a normal approximation may be entirely inappropriate since the probability mass function of ND may be skewed. In our case it is true and, in fact, necessary. In order for the asymptotics, as demonstrated below, to hold, one must require a decrease in the probability of type-I error which invalidates the usual normal approximation assumptions.

Alternatively, under our intended asymptotic regime, it is more plausible that individually the two sums of ND may each be well approximated in distribution by appropriately parameterized Poissons. If it is true that individually a Poisson approximation is appropriate then it is reasonable to pursue a Skellam approximation for ND. The Skellam probability distribution is the probability distribution of $Z_1 - Z_2$ where Z_1 and Z_2 are independent Poisson random variables. With this is mind, an may attempt a Skellam approximation to ND but must be mindful of the dependence structure of the collection of random variables $\{Y_{ij}\}$.

Finally, recall the Stein method for Poisson approximation, see [BHJ92], and how it may be applied to upper bound the error incurred by approximating the probability distribution of an arbitrary sum of binary random by the Poisson distribution. For example, if

$$V = \sum_{k=1}^{n} X_k$$

where $X_k \sim \text{Bernoulli}(p_k)$ and $Y \sim \text{Poisson}\left(\sum_{k=1}^{n} p_k\right)$ then the crucial quantity in identifying the quality of a Poisson approximation is

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P}(V \leq x) - \mathbb{P}(Y \leq x) \right|.$$

The Stein method provides means for bounding the above quantity no matter the dependencies among the collection of the binary random variables thus being a very powerful tool compared to usual Fourier methods. We, however, are interested in the Skellam distribution. In the following section we discuss in detail our derivation of the Stein method for the Skellam distribution and apply our new results to approximating the probability distribution of D.

2.3 Main Results

The probability mass function of $W \sim \text{Skellam}(\lambda_1, \lambda_2)$ is

$$\mathbb{P}(W = k) = e^{-(\lambda_1+\lambda_2)} \left(\sqrt{\frac{\lambda_1}{\lambda_2}} \right)^{k} I_k \left(2\sqrt{\lambda_1 \lambda_2} \right) \text{ for } k \in \mathbb{Z},$$

where $I_k\left(2\sqrt{\lambda_1\lambda_2}\right)$ is the modified Bessel function of the first kind with index k and argument $2\sqrt{\lambda_1\lambda_2}$. Let $\left\{ \{I_k\}_{k=1}^{n}, \{J_k\}_{k=1}^{m} \right\}$ be a collection of binary random variables with $\mathbb{E}I_k = p_k$ for $k = 1, \ldots, n$ and $\mathbb{E}J_k = q_k$ for $k = 1, \ldots, m$. If V is defined to be

$$V = \sum_{k=1}^{n} I_k - \sum_{k=1}^{m} J_k$$

then the crucial quantity to determine the accuracy of a Skellam approximation to the probability distribution of V is

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P}\left(V \leq x\right) - \mathbb{P}\left(W \leq x\right) \right|. \tag{2.1}$$

Note that Equation (2.1) is the Kolmogorov-Smirnov distance, which we will henceforth write as $KS\left(V, W\right)$. In order to upper bound $KS\left(V, W\right)$ we employ the Stein method.

The first step required by Stein's method is to determine an operator $\mathcal{A}\left[f\left(k\right)\right]$ such that

$$\mathbb{E}\mathcal{A}\left[f\left(W\right)\right] = 0 \quad \text{if and only if} \quad W \sim \text{Skellam}\left(\lambda_1, \lambda_2\right)$$

for any bounded function $f : \mathbb{Z} \mapsto \mathbb{R}$. This is accomplished through the following result.

Theorem 2.1. *The Stein operator $\mathcal{A}$ for the Skellam$\left(\lambda_1, \lambda_2\right)$ distribution is*

$$\mathcal{A}\left[f\left(k\right)\right] = \lambda_1 f\left(k+1\right) - k f\left(k\right) - \lambda_2 f\left(k-1\right). \tag{2.2}$$

The second step required by Stein's method is to set

$$\mathcal{A}\left[f\left(k\right)\right] = g_x\left(k\right) \tag{2.3}$$

for some test function $g_x\left(k\right)$ and solve for the function f. Note that, as with any difference equation, the function f is actually a family of functions. For the test function g_x, we choose

$$g_x\left(k\right) = 1\left\{k \leq x\right\} - \mathbb{P}\left(W \leq x\right) \tag{2.4}$$

for reasons that will become apparent in Theorem 2.2 below. The Stein equation (2.3), and a solution to it such as f_x in Theorem 2.5, for the Skellam distribution will be used, as with a Poisson approximation, to upper bound $KS\left(V, W\right)$.

To begin bounding the error incurred by a Skellam approximation to the probability distribution of a network density estimator we start with the following statement.

Theorem 2.2. *With g_x as in Equation* (2.4)*, let f_x be a solution to* (2.3)*. Then*

$$\begin{aligned} KS(V,W) &= \sup_{x\in\mathbb{R}} \left|\mathbb{P}(V \le x) - \mathbb{P}(W \le x)\right| \\ &= \sup_{x\in\mathbb{R}} \left|\mathbb{E}\left\{\lambda_1 f_x(V+1) - V f_x(V) - \lambda_2 f_x(V-1)\right\}\right|. \end{aligned}$$

The following corollary of Theorem 2.2 is of more direct use in bounding $KS(V,W)$:

Corollary 2.3. *Suppose that for each k, we construct the random variables S_k and S'_k for $k \in \{1,2,\ldots,n\}$ on a common probability space and then we construct the random variables T_k and T'_k for $k \in \{1,2,\ldots,m\}$ on a common probability space such that all together*

$$S_k \stackrel{d}{=} V \quad \text{and} \quad S'_k + 1 \stackrel{d}{=} V|I_k = 1.$$

and

$$T_k \stackrel{d}{=} V \quad \text{and} \quad T'_k - 1 \stackrel{d}{=} V|J_k = 1.$$

It follows that

$$KS(V,W) \le \Delta f \left\{ \sum_{k=1}^{n} p_k \mathbb{E}\left|S_k - S'_k\right| + \sum_{k=1}^{m} q_k \mathbb{E}\left|T_k - T'_k\right| \right\}. \tag{2.5}$$

where

$$\Delta f = \sup_{j\in\mathbb{Z}} \left|f_x(j+1) - f_x(j)\right|.$$

That Corollary 2.3 follows as a consequence of Theorem 2.2 is analogous to Theorem 1.B in [BHJ92], to which we refer the reader for details.

Another corollary of Theorem 2.2 follows.

Corollary 2.4. *Suppose that for each k the collection of random variables $\{\{I_k\}_{k=1}^n, \{J_k\}_{k=1}^m\}$ are independent then*

$$KS(V,W) \le \Delta f \left\{ \sum_{k=1}^{n} p_k^2 + \sum_{k=1}^{m} q_k^2 \right\}.$$

What is important to note here is that, by assuming independence, we are not gaining anything over the usual rate of convergence for the Poisson approximation found in [BHJ92]. However, and as is demonstrated later by appealing to Corollary 2.3, we must and can do better than simple Poisson approximation since the collection of random variables $\{\{I_k\}_{k=1}^{n}, \{J_k\}_{k=1}^{m}\}$ need not necessarily be independent among themselves. That is, for example, I_1 and J_1 may not be independent and thus the Poisson approximation as founded in [BHJ92] is invalidated.

We return to Equation (2.3) and state the solution f_x which satisfies the difference equation.

Theorem 2.5. *A solution to the difference equation*

$$\lambda_1 f(k+1) - kf(k) - \lambda_2 f(k-1) = g_x(k)$$

is

$$f_x(m) = \begin{cases} (-1)^m \left(\sqrt{\frac{\lambda_2}{\lambda_1}}\right)^m I_m \left[(-1)^c \left(\sqrt{\frac{\lambda_1}{\lambda_2}}\right)^c \frac{1}{I_c} f_x(c) \right. & \\ \left. + \frac{1}{\sqrt{\lambda_1\lambda_2}} \sum_{n=c}^{m-1} (-1)^{n+1} \frac{1}{I_n I_{n+1}} \sum_{k=-\infty}^{n} \left(\sqrt{\frac{\lambda_1}{\lambda_2}}\right)^k I_k g_x(k)\right] & \text{if } m > c \\ (-1)^m \left(\sqrt{\frac{\lambda_2}{\lambda_1}}\right)^m I_m \left[(-1)^c \left(\sqrt{\frac{\lambda_1}{\lambda_2}}\right)^c \frac{1}{I_c} f_x(c) \right. & \\ \left. - \frac{1}{\sqrt{\lambda_1\lambda_2}} \sum_{n=m}^{c-1} (-1)^{n+1} \frac{1}{I_n I_{n+1}} \sum_{k=-\infty}^{n} \left(\sqrt{\frac{\lambda_1}{\lambda_2}}\right)^k I_k g_x(k)\right] & \text{if } m < c \end{cases}$$

for any initial condition $(c, f_x(c))$ with $c \in \mathbb{Z}$ and $f_x(c) \in \mathbb{R}$.

The quantity Δf plays a key role in Corollaries 2.3 and 2.4. The following theorem provides an upper bound for it in the special case that $\lambda_1 = \lambda_2 = \lambda$.

Theorem 2.6. *The quantity Δf is upper bounded as*

$$\Delta f = \sup_{x\in\mathbb{R}} \sup_{j\in\mathbb{Z}} |f_x(j+1) - f_x(j)| \leq \frac{83}{\lambda}.$$

2.3.1 Returning to Networks

We return to a network context and recall that $\hat{\delta}$ is an unbiased network density estimator if $\lambda_1 = \lambda_2$, where recall that

$$\lambda_1 = \mathbb{E}\left[\sum_{\{i,j\}\in E^c} Y_{ij}\right] \quad \text{and} \quad \lambda_2 = \mathbb{E}\left[\sum_{\{i,j\}\in E} (1 - Y_{ij})\right].$$

If we suppose that $\mathbb{E}Y_{ij} = \alpha$ for all $\{i,j\} \in E^c$ and $\mathbb{E}(1 - Y_{ij}) = \beta$ for all $\{i,j\} \in E$ then we have that if $\lambda_1 = \lambda_2$ then

$$\alpha\left[\binom{p}{2} - |E|\right] = |E|\,\beta$$

or, equivalently,

$$\beta = \frac{1-\delta}{\delta}\alpha \tag{2.6}$$

where, again, δ is the network density. Recall that the definition of a "sparse" network is one in which $|E| \in O(p)$ in which case, $\delta \to 0$ as $p \to \infty$. It is believed in the scientific literature that many real-world networks are indeed sparse however we choose to consider the case that δ is bounded away from zero. The problem of rescaling in the case of sparse networks requires a more subtle handling of the asymptotic results and, therefore, is not treated here. None-the-less, as our result is the first of its kind, it represents a significant improvement in the understanding of complex networks.

On the other hand, if δ is bounded away from 0 and 1 as $p \to \infty$ then the problem is non-trivial.

On the topic of asymptotics when $\lambda_1 = \lambda_2$, let us suppose that the density of a network on p vertices δ_p is a function of p and that

$$\lim_{p\to\infty} \delta_p = \delta \in (0,1).$$

Obviously, $\alpha, \beta \in [0,1]$ so the feasible values for (α, β) is the line in Equation 2.6

$$(1-\delta)\,\alpha = \delta\beta$$

intersected with the region $[0,1] \times [0,1]$.

Before we take the limit $p \to \infty$, however, let us again consider the case that $\lambda_1 = \lambda_2$, call the common value λ. Before the limit, we have

$$\beta_p = \frac{1-\delta_p}{\delta_p}\alpha_p \tag{2.7}$$

and, furthermore, that

$$\lambda_p = \binom{p}{2}\left[1-\delta_p\right]\alpha_p. \tag{2.8}$$

The implication for the bound is, in the case that the indicator random variables are independent and $\alpha_{ij} = \alpha_p$ for all $\{i,j\} \notin E$ and $\beta_{ij} = \beta_p$ for $\{i,j\} \in E$, that we have

$$\begin{aligned}
KS\left(V,W\right) &\leq \frac{83}{\lambda}\left[\sum_{\{i,j\}\notin E}\alpha_{ij}^2 + \sum_{\{i,j\}}\beta_{ij}^2\right] \\
&= \frac{83}{\binom{p}{2}\left[1-\delta_p\right]\alpha_p}\left\{\left[\binom{p}{2} - |E_p|\right]\alpha_p^2 + |E_p|\,\beta_p^2\right\} \\
&= 83\left[\alpha_p + \beta_p\right] \\
&= 83\left[\alpha_p + \frac{1-\delta_p}{\delta_p}\alpha_p\right] \\
&= 83\alpha_p\left[1 + \frac{1-\delta_p}{\delta_p}\right] \\
&= 83\left(\frac{\delta_p}{\alpha_p}\right)^{-1}.
\end{aligned}$$

The final quantity is the inverse of the signal-to-noise ratio. That is, for example, suppose that we have a clique for all p so that $\delta_p = 1$. In this case, the quality of a Skellam approximation to the probability distribution of a network density estimator is entirely governed by the magnitude of the density relative to the type-I error rate. That is, the signal, i.e. δ_p, must not dominate the noise, i.e. α_p, in order for $KS\left(V,W\right) \to 0$ as $p \to \infty$.

Our result is important even in the independent case in the sense that we achieve a $O\left(\lambda^{-1}\right)$ bound in the KS distance between our estimator. Naturally, now we can handle dependence but, if in the case of independence among the collection of indicators, we can state the following.

Theorem 2.7. *Let X, Y, Z, and W be real-valued random variables taking at most countably infinite many values on the same support S. Suppose that*

$$d(X,Z) := \sup_{t\in\mathbb{R}} \left|\mathbb{P}\left[X \leq t\right] - \mathbb{P}\left[Z \leq t\right]\right| < \frac{\epsilon}{2},$$

and

$$d(Y,W) = \sup_{t\in\mathbb{R}} \left|\mathbb{P}\left[Y \leq t\right] - \mathbb{P}\left[W \leq t\right]\right| < \frac{\epsilon}{2},$$

and that (X,Y) and (Z,W) are independent. Then,

$$d(X-Y, Z-W) < \epsilon.$$

In particular, in statistics, when performing an independent 2-sample test for the population proportion, when $\hat{p}_1 \sim Bin\left(p_1, \sqrt{\frac{p_1(1-p_1)}{n_1}}\right) \approx Poisson(n_1 p_1)$ with KS-distance with $\frac{\epsilon}{2}$, and similarly for $\hat{p}_2$, then the sampling distribution of $\hat{p}_1 - \hat{p}_2$ under the null hypothesis is approximately a $Skellam(np, np)$ where $p = \frac{n_1p_1+n_2p_2}{n_1+n_2}$ and has total error ϵ to the true sampling distribution. We present the proof below.

Proof of Theorem 2.7. Let $\{s_n\}_{n=-\infty}^{\infty} = S$ be an ordering of S. Since (X,Y) are independent:

$$\mathbb{P}\left[X - Y \leq t\right] = \sum_{n=-\infty}^{\infty} \mathbb{P}\left[X \leq t + s_n\right] \mathbb{P}\left[Y = s_n\right].$$

and similarly

$$\mathbb{P}\left[Z - W \leq t\right] = \sum_{n=-\infty}^{\infty} \mathbb{P}\left[Z \leq t + s_n\right] \mathbb{P}\left[W = s_n\right].$$

So,

$$\begin{aligned}
&\left|\mathbb{P}\left[X - Y \leq t\right] - \mathbb{P}\left[Z - W \leq t\right]\right| \\
&= \left|\sum_{n=-\infty}^{\infty} \mathbb{P}\left[X \leq t + s_n\right] \mathbb{P}\left[Y = s_n\right] - \sum_{n=-\infty}^{\infty} \mathbb{P}\left[Z \leq t + s_n\right] \mathbb{P}\left[W = s_n\right]\right| \\
&\leq \left|\sum_{n=-\infty}^{\infty} \left(\mathbb{P}\left[X \leq t + s_n\right] - \mathbb{P}\left[Z \leq t + s_n\right]\right) \mathbb{P}\left[Y = s_n\right]\right| \\
&\qquad\qquad + \left|\sum_{n=-\infty}^{\infty} \mathbb{P}\left[Z \leq t + s_n\right] \left(\mathbb{P}\left[Y = s_n\right] - \mathbb{P}\left[W = s_n\right]\right)\right| \\
&\leq \sup_n \left|\mathbb{P}\left[X \leq t + s_n\right] - \mathbb{P}\left[Z \leq t + s_n\right]\right| + \left|\sum_{n=-\infty}^{\infty} \mathbb{P}\left[Z \leq t + s_n\right] \left(\mathbb{P}\left[Y = s_n\right] - \mathbb{P}\left[W = s_n\right]\right)\right|
\end{aligned}$$

Next, notice that $\mathbb{P}\left[Y = s_n\right] = \mathbb{P}\left[Y \leq s_n\right] - \mathbb{P}\left[Y \leq s_{n-1}\right]$, and similarly for W, so that defining $A(n) = \mathbb{P}\left[Y \leq s_n\right] - \mathbb{P}\left[W \leq s_n\right]$,

$$\begin{aligned}
&\left|\sum_{n=-\infty}^{\infty} \mathbb{P}\left[Z \leq t + s_n\right] \left(\mathbb{P}\left[Y = s_n\right] - \mathbb{P}\left[W = s_n\right]\right)\right| \\
&= \left|\sum_{n=-\infty}^{\infty} \mathbb{P}\left[Z \leq t + s_n\right] \left(\mathbb{P}\left[Y \leq s_n\right] - \mathbb{P}\left[Y \leq s_{n-1}\right] - \left(\mathbb{P}\left[W \leq s_n\right] - \mathbb{P}\left[W \leq s_{n-1}\right]\right)\right)\right| \\
&= \left|\sum_{n=-\infty}^{\infty} \mathbb{P}\left[Z \leq t + s_n\right] \left(\mathbb{P}\left[Y \leq s_n\right] - \mathbb{P}\left[W \leq s_n\right] - \left(\mathbb{P}\left[Y \leq s_{n-1}\right] - \mathbb{P}\left[W \leq s_{n-1}\right]\right)\right)\right| \\
&= \left|\sum_{n=-\infty}^{\infty} \mathbb{P}\left[Z \leq t + s_n\right] \left(A(n) - A(n-1)\right)\right| \\
&= \left|\sum_{n=-\infty}^{\infty} A(n) \left(\mathbb{P}\left[Z \leq t + s_n\right] - \mathbb{P}\left[Z \leq t + s_{n+1}\right]\right)\right| \\
&\leq \sum_{n=-\infty}^{\infty} \left|A(n)\right| \left|\mathbb{P}\left[Z \leq t + s_n\right] - \mathbb{P}\left[Z \leq t + s_{n+1}\right]\right| \\
&\leq \sup_n \left|A(n)\right|
\end{aligned}$$

Thus,

$$
\begin{aligned}
&|\mathbb{P}[X - Y \le t] - \mathbb{P}[Z - W \le t]| \\
&\le \sup_n |\mathbb{P}[X \le t + s_n] - \mathbb{P}[Z \le t + s_n]| + \sup_n |\mathbb{P}[Y \le s_n] - \mathbb{P}[W \le s_n]|
\end{aligned}
$$

Taking supremums over t now yields the result. □

This result means that the usual Poisson approximation to the Binomial is sufficient in the case of independence but we must, in a network context, treat the dependence issue delicately. We maintain the same asymptotics as in the independence case but can now handle dependence.

We may now state the following theorem in the case that the indicator random variables, i.e. edge indicator random variables, are independent.

Theorem 2.8. *If $\{I_k\}_{k=1}^n$ is a collection of independent binary random variables each with expectation $\{p_k\}_{k=1}^n$ and $\{J_k\}_{k=1}^m$ is a collection of independent binary random variables each with expectation $\{q_k\}_{k=1}^m$ then*

$$KS\left(\sum_{k=1}^n I_k - \sum_{k=1}^m J_k, Y - Z\right) \le \frac{\sum_{k=1}^n p_k^2}{2\sum_{k=1}^n p_k} + \frac{\sum_{k=1}^m q_k^2}{2\sum_{k=1}^m q_k} \tag{2.9}$$

where $Y \sim Poisson\left(\sum_{k=1}^n p_k\right)$ and $Z \sim Poisson\left(\sum_{k=1}^m q_k\right)$.

Proof. According to [BC05] if $\{I_k\}_{k=1}^n$ is a collection of independent binary random variables each with expectation $\{p_k\}_{k=1}^n$ then

$$KS\left(\sum_{k=1}^n I_k, Y\right) \le \frac{\sum_{k=1}^n p_k^2}{\sum_{k=1}^n p_k} \tag{2.10}$$

where $Y \sim Poisson\left(\sum_{k=1}^n p_k\right)$. Furthermore, if $\{J_k\}_{k=1}^m$ is a collection of independent binary random variables each with expectation $\{q_k\}_{k=1}^m$ then

$$KS\left(\sum_{k=1}^m J_k, Z\right) \le \frac{\sum_{k=1}^m q_k^2}{\sum_{k=1}^m q_k} \tag{2.11}$$

where $Z \sim Poisson\left(\sum_{k=1}^{m} q_k\right)$. In total, and by appealing to Theorem 2.7

$$KS\left(\sum_{k=1}^{n} I_k - \sum_{k=1}^{m} J_k, Y - Z\right) \leq \frac{\sum_{k=1}^{n} p_k^2}{2\sum_{k=1}^{n} p_k} + \frac{\sum_{k=1}^{m} q_k^2}{2\sum_{k=1}^{m} q_k} \tag{2.12}$$

□

In total, we have the following corollary.

Corollary 2.9. *If* $X = \sum_{\{i,j\}\notin E} Y_{ij}$ *and* $Z = \sum_{\{i,j\}\in E} (1 - Y_{ij})$ *then*

$$KS\left(X - Z, X' - Z'\right) \leq \frac{\sum_{i,j\notin E} \left[\mathbb{E}Y_{ij}\right]^2}{2\sum_{i,j\notin E} \mathbb{E}Y_{ij}} + \frac{\sum_{i,j\notin E} \left[(1 - \mathbb{E}Y_{ij})\right]^2}{2\sum_{i,j\notin E} (1 - \mathbb{E}Y_{ij})} \tag{2.13}$$

where $X' \sim Poisson\left(\sum_{i,j\notin E} \mathbb{E}Y_{ij}\right)$ *and* $Z' \sim Poisson\left(\sum_{i,j\in E} (1 - \mathbb{E}Y_{ij})\right)$.

Finally, as we mentioned in Section 2.1, the above results extend immediately to any network characteristic which is a member of the class of network motif counts. For example, let us consider triangles instead of edges. In accordance with Section 2.2, we would perform $\binom{p}{3}$ hypothesis tests, one for each of the $\{i, j, k\} \in V^{(3)}$ vertex triples in the network graph G of the form

$$H_0 : \{i, j, k\} \notin E^{(3)} \quad \text{versus} \quad H_1 : \{i, j, k\} \in E^{(3)}$$

where $V^{(3)} = \{\{i, j, k\} : i, j, k \in V; i < j < k\}$ and $E^{(3)}$ is the set of vertex triples $\{i, j, k\} \in V^{(3)}$ for which the respective vertex induced subgraph is a complete graph (in this case a triangle). As an analogy, we have been writing $E = E^{(2)}$ which means that the respective vertex induced subgraph on two vertices is a complete graph (in this case an edge).

Finally, the $\binom{p}{3}$ random variables Y_{ijk} are defined accordingly and the results of this section follow immediately. Of course we would have to consider many more covariance terms but, in the end, so long as one was able to control said covariance terms, then

2.3.2 The Watts-Strogatz, Gaussian Measurement Model, and Couplings

In this section we create the foundation upon which a particular example of the Skellam approximation for the numerator of $\hat{\delta} - \delta$, i.e. ND, may be applied. We start with general results, some of which may appear familiar, but are necessary for laying the general framework for the example. In the end, we produce a result which, under the assumption of a Watts-Strogatz random graph model and Gaussian measurement model, demonstrates a bound on the Skellam approximation to the probability distribution of ND. This result appears at the end of the section but the prelimary results are necessary for the final conclusion. Also, we derive a bound for the Kolmogorov distance between the probability distribution of a difference of two sums of possibly dependent binary random variables and the $Skellam\,(\lambda_1, \lambda_2)$ distribution. Specifically, let $\mathbf{X} = \{\mathbf{X}_1, \mathbf{X}_2\}$ be a collection of binary random variables where

$$\mathbf{X}_1 = \{X_{1k} : k = 1, 2, \ldots, n\} \quad \text{and} \quad \mathbf{X}_2 = \{X_{2k} : k = 1, 2, \ldots, m\} \tag{2.14}$$

partition $\mathbf{X}$ and define the random variables X_1 and X_2 as

$$X_1 = \textstyle\sum_{k=1}^{n} X_{1k} \quad \text{and} \quad X_2 = \sum_{k=1}^{m} X_{2k}. \tag{2.15}$$

Finally, let $X = X_1 - X_2$. The following theorem establishes a bound for the Kolmogorov distance K between the probability distributions of X and $W \sim Skellam\,(\lambda_1, \lambda_2)$.

Theorem 2.10. *Let $p_{1k} = \mathbb{E}X_{1k}$ for $k = 1, 2, \ldots, n$ and $p_{2k} = \mathbb{E}X_{2k}$ for $k = 1, 2, \ldots, m$ so that $\lambda_1 = \sum_{k=1}^{n} p_{1k}$ and $\lambda_2 = \sum_{k=1}^{m} p_{2k}$. Then*

$$\begin{aligned} K \quad \leq \quad & M \left\{ \sum_{k=1}^{n} p_{1k}^2 + 2 \sum_{i<j}^{n} \left| Cov\,(X_{1i}, X_{1j}) \right| \right. \\ & \left. + \sum_{k=1}^{m} p_{2k}^2 + 2 \sum_{i<j}^{m} \left| Cov\,(X_{2i}, X_{2j}) \right| \right. \end{aligned} \tag{2.16}$$

$$+ 2\sum_{i=1}^{n}\sum_{j=1}^{m}\left|Cov\left(X_{1i}, X_{2j}\right)\right|\Bigg\}$$

where $M = \sup_{x\in\mathbb{R}}\sup_{k\in\mathbb{Z}}\left|f_x\left(k+1\right) - f_x\left(k\right)\right|$.

It was shown that if $\lambda_1 = \lambda_2$ in the previous section that $M = 83/\lambda$ where λ is the common $\lambda_1 = \lambda_2$ common value.

Recall that the discrepancy D between the estimator $\hat{\delta}$ and the network density δ is defined as $D = \hat{\delta} - \delta$ and may be written as

$$D = \frac{\sum_{\{i,j\}\in E^c} Y_{ij} - \sum_{\{i,j\}\in E}\left(1 - Y_{ij}\right)}{\binom{p}{2}} \tag{2.17}$$

where $|V| = p$ and the collection of $\binom{p}{2}$ random variables $\mathbf{Y} = \left\{Y_{ij} : \{i,j\} \in V^{(2)}\right\}$ are defined by

$$Y_{ij} = \begin{cases} 1 & \text{if } H_0 \text{ is rejected} \\ 0 & \text{if } H_0 \text{ is not rejected.} \end{cases} \tag{2.18}$$

In accordance with Section 2.2, the collection of binary random variables $\mathbf{Y} = \{Y_{ij} : \{i,j\} \in V^{(2)}\}$ may be partitioned as $\mathbf{Y} = \{\mathbf{Y}_1, \mathbf{Y}_2\}$ where

$$\mathbf{Y}_1 = \{Y_{ij} : \{i,j\} \in E^c\} \quad \text{and} \quad \mathbf{Y}_2 = \{1 - Y_{ij} : \{i,j\} \in E\}\,. \tag{2.19}$$

Now, define the random variables Y_1 and Y_2 as

$$Y_1 = \textstyle\sum_{\{i,j\}\in E^c} Y_{ij} \quad \text{and} \quad Y_2 = \sum_{\{i,j\}\in E}\left(1 - Y_{ij}\right) \tag{2.20}$$

so that $\binom{p}{2}D = Y_1 - Y_2$ and $\lambda_1 = \mathbb{E}Y_1$ and $\lambda_2 = \mathbb{E}Y_2$. Recall from Section 2.2 that

$$Y_{ij} \sim \begin{cases} Bernoulli\left(\alpha_{ij}\right) & \text{if } \{i,j\} \in E^c \\ Bernoulli\left(1 - \beta_{ij}\right) & \text{if } \{i,j\} \in E \end{cases} \tag{2.21}$$

where α_{ij} is the probability of Type-I error for the test on vertex pair $\{i,j\} \in E^c$ and β_{ij}

is the probability of Type-II error for the test on vertex pair $\{i, j\} \in E$.

This section contains a demonstration of an application of the results of this in the context of an association network. In particular, we choose an example association network in which vertex pairs are connected by an edge if the attributes corresponding to those vertices have non-zero correlation. Such an object is a so-called correlation network. Furthermore, we assume that the collection of attributes has a multivariate Gaussian distribution. Furthermore, we shall suppose that the underlying network graph is a realization of a Wattz-Strogatz small-world random graph model.

Suppose that $\mathbf{Z}$ is a zero-mean $p-$dimensional random vector with a multivariate Gaussian distribution and population correlation matrix $\Sigma = (\rho_{ij})$, and that $G = (V, E)$ is a network graph with $|V| = p$. We identify with each vertex v_k for $k = 1, 2, \ldots, p$ the k^{th} random element Z_k of $\mathbf{Z}$. In the context of a correlation network graph (see [Kol09]), the measure of similarity between vertex pairs v_i and v_j in this brand of association networks is ρ_{ij}, the Pearson product-moment correlation between Z_i and Z_j. Accordingly, the edge set E of G is defined as

$$E = \left\{ \{i, j\} \in V^{(2)} : \rho_{ij} > 0 \right\}. \tag{2.22}$$

Note that more generally one could define the edge set in a correlation association network as $\left\{\{i, j\} \in V^{(2)} : \rho_{ij} \neq 0\right\}$ but the former definition is chosen here for the sake of illustration and demonstration.

The general construction of the Watts-Strogatz random graph model for a network graph is to begin with p vertices and, arranged in a periodic fashion, join with an edge each vertex to d of its neighbors to each side for some $d << p$. The result is a network graph which is a one-dimensional lattice consisting of pd edges. Note that in the case of correlation networks, an edge between vertices v_i and v_j indicates positive correlation of the attributes respective attributes and, therefore, the one-dimensional lattice represents a cyclic auto-regressive model. Finally, it is then typical to introduce a small number of edges to randomly selected pairs of vertices.

To complete the scenario of our demonstration, we specifically describe the formal inference procedure of Section 2.2. For each of the $\{i,j\} \in V^{(2)}$ vertex pairs in the network graph G we perform the hypothesis test

$$H_0 : \rho_{ij} = 0 \quad \text{versus} \quad H_1 : \rho_{ij} > 0. \tag{2.23}$$

To execute the above $\binom{p}{2}$ hypothesis tests, suppose that $\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_N$ is a random sample from the distribution $\mathbf{Z} \sim N(\mathbf{0}, \Sigma)$. From this sample one may compute the sampling distribution of the unbiased estimator $\hat{\Sigma} = (\hat{\sigma}_{ij})$ of Σ, see [And71], and thus estimate the corresponding correlation matrix with $R = (r_{ij})$ to perform the hypothesis tests in Equation (2.23). The upper-triangular entries r_{ij} of R can serve as a test statistics for the hypothesis tests in Equation (2.23) and, as such, each element of the collection of random variables $\mathbf{Y} = \big\{\{i,j\} \in V^{(2)} : Y_{ij}\big\}$ is defined as

$$Y_{ij} = \begin{cases} 1 & \text{if } r_{ij} > \tau \\ 0 & \text{if } r_{ij} < \tau \end{cases} \tag{2.24}$$

for some rejection threshold $\tau \in \mathbb{R}$. It follows that the inferred edge set is defined as

$$\begin{aligned} \hat{E} &= \Big\{\{i,j\} \in V^{(2)} : Y_{ij} = 1\Big\} & (2.25) \\ &= \Big\{\{i,j\} \in V^{(2)} : r_{ij} > \tau\Big\}. & (2.26) \end{aligned}$$

Again, proceeding along the Stein methodology, the following proof upper-bounds KS in Equation (2.5) by upper-bounding each term in the sum on the right-hand-side of (2.5). First, a preliminary lemma.

Lemma 2.11. *Suppose that $\mathbf{I} = \{I_k : k = 1, 2, \ldots, s\}$ is a collection of binary random variables and let $S = \sum_{k=1}^{s} I_k$. There exists random variables S^j and $S^{j\prime}$ defined on a* *common probability space such that*

$$\mathcal{L}\left(S^j\right) = \mathcal{L}\left(S\right) \quad \textit{and} \quad \mathcal{L}\left(S^{j\prime} + 1\right) = \mathcal{L}\left(S | I_j = 1\right) \ . \tag{2.27}$$

Furthermore, there exists a coupling of S^j and $S^{j\prime}$ with the property

$$\left(\mathbb{E}I_j\right) \mathbb{E}\left|S^j - S^{j\prime}\right| \leq \left(\mathbb{E}I_j\right)^2 + \sum_{k \neq j}^{s} \left|Cov\left(I_j, I_k\right)\right| \tag{2.28}$$

for any $j = 1, 2, \ldots, s$.

Proof. We begin by describing the so-called generalized standard construction, see [AL78]. Suppose that the collection $\mathbf{I}$ has the cumulative distribution function F and define recursively the functions $F_1\left(x\right) = \mathbb{P}\left(I_1 \leq x\right)$ and

$$F_j\left(x | x_1, \ldots, x_{j-1}\right) = \mathbb{P}\left(I_j \leq x | I_k = x_k; 1 \leq k < j\right) \tag{2.29}$$

for each x measurable in $\left(x_1, \ldots, x_{j-1}\right)$ and $j = 2, \ldots, s$. Now, define recursively the functions $G_1\left(x\right) = \inf_{t \in \mathbb{R}} \left\{F_1\left(t\right) \geq x\right\}$ and

$$G_j\left(x | x_1, \ldots, x_{j-1}\right) = \inf_{t \in \mathbb{R}} \left\{F_j\left(t | x_1, \ldots, x_{j-1}\right) \geq x\right\} \tag{2.30}$$

for $j = 2, \ldots, s$. Finally, let $\mathbf{U} = \left\{U_k : k = 1, 2, \ldots, s\right\}$ be a collection of independent $Uniform\left(0, 1\right)$ random variables and construct the collection of random variables $\mathbf{I}^* = \left\{I_k^* : k = 1, 2, \ldots, s\right\}$ according to the following program. First, define recursively the random variables $I_1^* = G_1\left(U_1\right)$ and, given that $I_1^* = x_1$, …, $I_{j-1}^* = x_{j-1}$,

$$I_j^* = G_j\left(U_j | x_1, \ldots, x_{j-1}\right). \tag{2.31}$$

As noted in [AL78], it is immediate that $\mathcal{L}\left(I_1^*\right) = \mathcal{L}\left(I_1\right)$ and since

$$\mathbb{P}\left(I_j^* \leq x | I_i^* = x_i; 1 \leq i < j\right) = \mathbb{P}\left(I_j \leq x | I_i = x_i; 1 \leq i < j\right) \tag{2.32}$$

all finite-dimensional distributions of $\mathbf{I}^*$ and $\mathbf{I}$ coincide so that $\mathcal{L}(I^*) = \mathcal{L}(I)$.

To construct the random variables S^j and $S^{j\prime}$ with the desired property of Equation 2.28, first assume without loss of generality that $j = 1$. Now, let $\mathbf{U}^1 = \{U_k : k = 1, 2, \ldots, s\}$ be a collection of independent $Uniform\,(0,1)$ random variables and define $\mathbf{U}^{1\prime} = \{U_k = 1;\ U_k : k = 2, 3, \ldots, s\}$ to be an almost sure copy of $\mathbf{U}^1$ for all indices $k \neq 1$. Next, execute the generalized standard construction with $\mathbf{U}^1$ to construct the collection $\mathbf{I}^1 = \left\{I_k^1 : k = 1, 2, \ldots, s\right\}$ and then execute the generalized standard construction with $\mathbf{U}^{1\prime}$ to construct the collection $\mathbf{I}^{1\prime} = \left\{I_k^{1\prime} : k = 2, 3, \ldots, s\right\}$. Now define S^1 and $S^{1\prime}$ as

$$S^1 = \textstyle\sum_{k=1}^s I_k^1 \quad \text{and} \quad S^{1\prime} = \sum_{k=2}^s I_k^{1\prime}. \tag{2.33}$$

It is immediate that

$$\mathcal{L}\left(S^1\right) = \mathcal{L}(S) \quad \text{and} \quad \mathcal{L}\left(S^{1\prime} + 1\right) = \mathcal{L}(S | I_1 = 1) \tag{2.34}$$

and that S^1 and $S^{1\prime}$ exist on a common probability space since their randomness is derived from the randomness of U^1 and $U^{1\prime}$ which, in turn, are defined on a common probability space.

To show that S^1 and $S^{1\prime}$ have the desired property of Equation 2.28, note that

$$\begin{aligned} (\mathbb{E}I_1)\,\mathbb{E}\left|S^1 - S^{1\prime}\right| &= (\mathbb{E}I_1)\,\mathbb{E}\left|I_1^1 + \sum_{k=2}^{n}\left(I_k^1 - I_k^{1\prime}\right)\right| && (2.35)\\ &\leq (\mathbb{E}I_1)^2 + \sum_{k=2}^{s}\left(\mathbb{E}I_1^1\right)\mathbb{E}\left|I_k^1 - I_k^{1\prime}\right|. && (2.36) \end{aligned}$$

Note, that by construction of I_k^1 and $I_k^{1\prime}$ with the collections $\mathbf{U}^1$ and $\mathbf{U}^{1\prime}$

$$\mathbb{E}\left|I_k^1 - I_k^{1\prime}\right| = \left|\mathbb{P}\left(I_k^1 = 1\right) - \mathbb{P}\left(I_k^1 = 1 | I_1^{1\prime} = 1\right)\right| \tag{2.37}$$

which implies that

$$(\mathbb{E}I_1)\,\mathbb{E}\left|I_k^1 - I_k^{1\prime}\right| = \left|Cov\left(I_1^1, I_k^{1\prime}\right)\right| = \left|Cov\left(I_1, I_k\right)\right|. \tag{2.38}$$

Thus the claim of Equation 2.28 is verified. □

Proof of Theorem 2.10. Let $\mathbf{X} = \{\mathbf{X}_1, \mathbf{X}_2\}$ be a collection of binary random variables where

$$\mathbf{X}_1 = \{X_{1k} : k = 1, 2, \ldots, n\} \quad \text{and} \quad \mathbf{X}_2 = \{X_{2k} : k = 1, 2, \ldots, m\} \tag{2.39}$$

partition $\mathbf{X}$ and define the random variables

$$X_1 = \textstyle\sum_{k=1}^{n} X_{1k}, \quad \text{and} \quad X_2 = \textstyle\sum_{k=1}^{m} X_{2k} \tag{2.40}$$

and let $X = X_1 - X_2$. Suppose that $p_{1k} = \mathbb{E}X_{1k}$ for $k = 1, 2, \ldots, n$ and $p_{2k} = \mathbb{E}X_{2k}$ for $k = 1, 2, \ldots, m$ and define

$$\mu_1 = \textstyle\sum_{k=1}^{n} p_{1k} \quad \text{and} \quad \mu_2 = \textstyle\sum_{k=1}^{m} p_{2k}. \tag{2.41}$$

Finally let $W \sim Skellam\left(\mu_1, \mu_2\right)$ and note that by Theorems 2.1 and 2.3

$$K = \sup_{x \in \mathbb{R}} \left|\mathbb{E}\left[\mu_1 f_x\left(X+1\right) - X f_x\left(X\right) - \mu_2 f_x\left(X-1\right)\right]\right| \tag{2.42}$$

where, again

$$K = \sup_{x \in \mathbb{R}} \left|\mathbb{P}\left(X \leq x\right) - \mathbb{P}\left(W \leq x\right)\right|. \tag{2.43}$$

Note that Equation (2.42) can be written as

$$K \;=\; \sup_{x \in \mathbb{R}} \left|\sum_{k=1}^{n} \mathbb{E}\left[p_{1k} f_x\left(X+1\right) - X_{1k} f_x\left(X\right)\right]\right. \tag{2.44}$$

$$
\begin{aligned}
&\quad + \sum_{k=1}^{m} \mathbb{E}\left[p_{2k} f_x(X) - X_{2k} f_x(X-1)\right] \Bigg| \\
&= \sup_{x\in\mathbb{R}} \left| \sum_{k=1}^{n} p_{1k} \mathbb{E}\left[f_x\left(X^{1k}+1\right) - \mathbb{E} f_x\left(X^{1k\prime}+1\right)\right] \right. \qquad (2.45) \\
&\quad \left. + \sum_{k=1}^{m} p_{2k} \mathbb{E}\left[f_x\left(X^{2k\prime}-1\right) - f_x\left(X^{2k}-1\right)\right] \right|
\end{aligned}
$$

where, for each $k = 1, 2, \ldots, n$ the random variables X^{1k} and $X^{1k\prime}$ are such that $\mathcal{L}\left(X^{1k}\right) = \mathcal{L}(X)$ and $\mathcal{L}\left(X^{1k\prime}+1\right) = \mathcal{L}(X|X_{1k}=1)$ for each $k = 1, 2, \ldots, m$ the random variables X^{2k} and $X^{2k\prime}$ are such that $\mathcal{L}\left(X^{2k}\right) = \mathcal{L}(X)$ and $\mathcal{L}\left(X^{2k\prime}+1\right) = \mathcal{L}(X|X_{2k}=1)$. By Lemma 2.11 such random variables can be constructed so that X^{1k} and $X^{1k\prime}$ exist on a common probability space for each $k = 1, 2, \ldots, n$ and so that X^{2k} and $X^{2k\prime}$ exist on a common probability space for each $k = 1, 2, \ldots, m$. It follows that

$$
\begin{aligned}
K &\leq \sup_{x\in\mathbb{R}} \left\{ \sum_{k=1}^{n} p_{1k} \mathbb{E}\left| f_x\left(X^{1k}+1\right) - f_x\left(X^{1k\prime}+1\right)\right| \right. \qquad (2.46) \\
&\quad \left. + \sum_{k=1}^{m} p_{2k} \mathbb{E}\left| f_x\left(X^{2k\prime}-1\right) - f_x\left(X^{2k}-1\right)\right| \right\} \\
&\leq M \left\{ \sum_{k=1}^{n} p_{1k} \mathbb{E}\left|X^{1k} - X^{1k\prime}\right| + \sum_{k=1}^{m} p_{2k} \mathbb{E}\left|X^{2k} - X^{2k\prime}\right| \right\} \qquad (2.47)
\end{aligned}
$$

where $M = \sup_{x\in\mathbb{R}} \sup_{k\in\mathbb{Z}} |f_x(k+1) - f_x(k)|$.

Finally, by an application of Lemma 2.11,

$$
\begin{aligned}
K &\leq M \left\{ \sum_{k=1}^{n} p_{1k}^2 + 2\sum_{i<j}^{n} |Cov(X_{1i}, X_{1j})| \right. \qquad (2.48) \\
&\quad + \sum_{k=1}^{m} p_{2k}^2 + 2\sum_{i<j}^{m} |Cov(X_{2i}, X_{2j})| \\
&\quad \left. + 2\sum_{i=1}^{n}\sum_{j=1}^{m} |Cov(X_{1i}, X_{2j})| \right\}
\end{aligned}
$$

is immediate and the claim is proved. □

We are now set to describe our demonstration of these results to the Watts-Strogatz network model discussed earlier. For the sake of illustration, we choose to explore the above framework in the case that $\mathbf{Z}$ has the multivariate Gaussian distribution with population correlation matrix Σ, e.g. $\mathbf{Z} \sim N(\mathbf{0}, \Sigma)$. In this context, the probability distribution of $\hat{\Sigma}$, the unbiased estimator of Σ, and thus the collection of test statistics $\left\{r_{ij} : \{i,j\} \in V^{(2)}\right\}$, is known, although in general, not in closed form, see [And71]. Thus we may, in principle, directly compute the probability and covariance terms on the right-hand side of the bound in Equation 2.16.

The unbiased estimator $\hat{\Sigma}$ of Σ is of the form

$$\hat{\Sigma} = \frac{1}{N-1} \sum_{k=1}^{N} (\mathbf{z}_k - \bar{\mathbf{z}})(\mathbf{z}_k - \bar{\mathbf{z}})' \tag{2.49}$$

where N is the size of the random sample from the probability distribution of $\mathbf{Z}$ and $\mathbf{z}_k$ is the k^{th} observation of the $p-$dimensional random vector and

$$\bar{\mathbf{z}} = \begin{bmatrix} \bar{\mathbf{z}}_1 \\ \bar{\mathbf{z}}_2 \\ \vdots \\ \bar{\mathbf{z}}_p \end{bmatrix} = \frac{1}{N} \sum_{k=1}^{N} \mathbf{z}_k. \tag{2.50}$$

If $\mathbf{Z}$ has the multivariate Gaussian distribution $N(\mu, \Sigma)$ then $\hat{\Sigma}$ has the Wishart $\left(\frac{1}{n}\Sigma, n\right)$ distribution where $n = N-1$, see [And71].

The following lemma is useful for determining the probability distribution of a submatrix of $\hat{\Sigma}$ and, thus, the bivariate probability distributions of r_{ij} and r_{hk}. For proof, see [And71].

Lemma 2.12. *If* $\hat{\Sigma} \sim$ *Wishart* $\left(\frac{1}{n}\Sigma, n\right)$ *and* $\hat{\Sigma}$ *and* Σ *are partitioned as*

$$\hat{\Sigma} = \begin{pmatrix} \hat{\Sigma}_{11} & \hat{\Sigma}_{12} \\ \hat{\Sigma}_{21} & \hat{\Sigma}_{22} \end{pmatrix} \quad \textit{and} \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \tag{2.51}$$

then $\hat{\Sigma}_{11} \sim$ *Wishart* $\left(\frac{1}{n}\Sigma_{11}, n\right)$.

Our interest is, of course, the sample correlation matrix R and the elements therein. We are interested specifically in the the bivariate probability distribution of two sample correlation coefficients r_{ij} and r_{hk} since these determine the values of $\operatorname{Cov}(Y_{ij}, Y_{hk})$ in the bound of Theorem 2.10. Within the association network $G = (V, E)$ these sample correlation coeffiencents are one of two types. Either r_{ij} and r_{hk} are the sample correlations of adjacent vertex pairs or non-adjacent vertex pairs. In other words, either there is a common vertex to the vertex pairs of $\{i, j\}$ and $\{h, k\}$ or there is not. We begin by examining the first case, i.e. there is a vertex in common. The following lemma provides an expression for the bivariate distribution of r_{ij} and r_{hk} when the vertex pairs $\{i, j\}$ and $\{h, k\}$ are adjacent. We may, without loss of generality, suppose that $i, h = 1$, $j = 2$, and $k = 3$ so that we are considering only the upper-left 3×3 submatrix of Σ.

Lemma 2.13. *If $N \geq 4$ then the bivariate probability density of r_{12} and r_{13} is*

$$
\begin{aligned}
p(x, y; \Sigma, n) \quad \propto \quad & \int_{\mathbb{R}}\int_{\mathbb{R}}\int_{\mathbb{R}} \exp\left\{-\sigma^{23}\sqrt{h_{22}}\sqrt{h_{33}}\left(xy - \sqrt{(1-x^2)(1-y^2)}\right)\right\} \\
& \cdot \left[\left(1-x^2\right)\left(1-y^2\right)\right]^{(n-3)/2}\left[1 + \sum_{k=1}^{\infty}\left(\prod_{r=0}^{k-1}\frac{n/2-1+r}{n-2+r}\right)\right. \\
& \left. \cdot \frac{\left(-2\sigma^{23}\sqrt{h_{22}}\sqrt{h_{33}}\sqrt{(1-x^2)(1-y^2)}\right)^k}{k!}\right] \\
& \cdot \exp\left\{-\frac{1}{2}\sum_{k=1}^{3}\sigma^{kk}h_{kk} - \sigma^{12}\sqrt{h_{11}}\sqrt{h_{22}}x - \sigma^{13}\sqrt{h_{11}}\sqrt{h_{33}}y\right\} \\
& \cdot \left(\prod_{k=1}^{3} h_{kk}^{n/2-1}\right) dh_{11}\, dh_{22}\, dh_{33}.
\end{aligned}
\tag{2.52}
$$

where $H = (h_{ij})$ is the upper-left 3×3 submatrix of $\hat{\Sigma}$. Thus, if the upper-left 3×3

submatrix of Σ is of the form

$$\begin{bmatrix} \sigma_{11} & 0 & 0 \\ 0 & \sigma_{22} & 0 \\ 0 & 0 & \sigma_{33} \end{bmatrix} \quad \begin{bmatrix} \sigma_{11} & \sigma_{12} & 0 \\ \sigma_{12} & \sigma_{22} & 0 \\ 0 & 0 & \sigma_{33} \end{bmatrix} \quad \begin{bmatrix} \sigma_{11} & 0 & \sigma_{13} \\ 0 & \sigma_{22} & 0 \\ \sigma_{13} & 0 & \sigma_{33} \end{bmatrix} \tag{2.53}$$

where we denote with σ_{12} and σ_{13} potentially non-zero values in the upper-left 3×3 submatrix of Σ then r_{12} and r_{13} are pairwise independent.

Proof. Recall that $\hat{\Sigma}$ is defined as

$$\hat{\Sigma} = \frac{1}{N-1} \sum_{k=1}^{N} (\mathbf{x}_k - \bar{\mathbf{x}}) (\mathbf{x}_k - \bar{\mathbf{x}})' \tag{2.54}$$

where N is the size of the random sample from the probability distribution of $\mathbf{Z}$ and $\mathbf{z}_k$ is the k^{th} observation of the $p-$dimensional random vector and

$$\bar{\mathbf{z}} = \begin{bmatrix} \bar{\mathbf{z}}_1 \\ \bar{\mathbf{z}}_2 \\ \vdots \\ \bar{\mathbf{z}}_p \end{bmatrix} = \frac{1}{N} \sum_{k=1}^{N} \mathbf{z}_k. \tag{2.55}$$

According to Lemma 2.12, if $N \geq 4$ then the probability density function (pdf) of the upper-left 3×3 submatrix H of $\hat{\Sigma}$ is

$$p(H; \Sigma, n) \propto |H|^{\frac{1}{2}(n-p'-1)} \exp\left\{-\frac{1}{2}\text{tr}\left(\Sigma^{-1} H\right)\right\} \text{ for } H \text{ positive definite} \tag{2.56}$$

where $n = N - 1$ and $p' = 3$ and H is the upper-left 3×3 submatrix of R. Since we are specifically interested in the sample correlation matrix R and the elements therein we perform the transformation

$$H \mapsto \sqrt{\text{diag}(H)} R \sqrt{\text{diag}(H)}. \tag{2.57}$$

The Jacobian matrix J of the transformation is a $\binom{p'}{2} \times \binom{p'}{2}$ matrix. In fact, J is diagonal so that the determinant of the transformation is simply a product. The Jacobian determinant is

$$|J| = \prod_{k=1}^{p'} h_{kk}^{(p'-1)/2}. \tag{2.58}$$

Thus, the joint probability distribution of H and R is

$$p\left(R, H; \Sigma, n\right) \propto |R|^{(n-p'-1)/2} \exp\left\{-\frac{1}{2}\mathrm{tr}\left(\Sigma^{-1}H\right)\right\}\left(\prod_{k=1}^{p'} h_{kk}^{n/2-1}\right). \tag{2.59}$$

Denote $\Sigma^{-1} = \left(\sigma^{ij}\right)$ and $H = (h_{ij})$ so that

$$\begin{aligned} p\left(R, H; \Sigma, n\right) \quad \propto \quad & |R|^{(n-p'-1)/2} \exp\left\{-\frac{1}{2}\sum_{k=1}^{p'} \sigma^{kk} h_{kk}\right. \\ & \left. - \sum_{i=1}^{p'-1}\sum_{j=i+1}^{p} \sigma^{ij}\sqrt{h_{ii}}\sqrt{h_{jj}}r_{ij}\right\}\left(\prod_{k=1}^{p'} h_{kk}^{n/2-1}\right). \end{aligned} \tag{2.60}$$

Naturally, this means that

$$\begin{aligned} p\left(R; \Sigma, n\right) \quad \propto \quad & \int_{\mathbb{R}}\int_{\mathbb{R}}\int_{\mathbb{R}} |R|^{(n-p'-1)/2} \exp\left\{-\frac{1}{2}\sum_{k=1}^{p'} \sigma^{kk} h_{kk}\right. \\ & \left. - \sum_{i=1}^{p'-1}\sum_{j=i+1}^{p'} \sigma^{ij}\sqrt{h_{ii}}\sqrt{h_{jj}}r_{ij}\right\}\left(\prod_{k=1}^{p'} h_{kk}^{n/2-1}\right) dh_{11}\, dh_{22}\, dh_{33}. \end{aligned} \tag{2.61}$$

The goal is to determine the bivariate density of (r_{12}, r_{13}). Let $x = r_{12}$, $y = r_{13}$, and $z = r_{23}$. This will require marginalizing $z = r_{23}$. Of course, we must then compute the determinant

$$|R| \quad = \quad \begin{vmatrix} 1 & x & y \\ x & 1 & z \\ y & z & 1 \end{vmatrix} \tag{2.62}$$

$$= \quad -\left(z - xy\right)^2 + \left(1 - x^2\right)\left(1 - y^2\right). \tag{2.63}$$

The roots of the above quadratic are

$$(z_0, z_1) = xy \pm \sqrt{\left(1-x^2\right)\left(1-y^2\right)}. \tag{2.64}$$

This means, to marginalize $z = r_{23}$ we must compute the integral

$$\begin{aligned} I_z &= \int_{xy-\sqrt{(1-x^2)(1-y^2)}}^{xy+\sqrt{(1-x^2)(1-y^2)}} \left[\left(1-x^2\right)\left(1-y^2\right) - (z-xy)^2\right]^{(n-4)/2} \\ &\quad \cdot \exp\left\{-\sigma^{23}\sqrt{h_{22}}\sqrt{h_{33}}z\right\} dz. \end{aligned} \tag{2.65}$$

Using substitution we find that

$$\begin{aligned} p\left(x, y; \Sigma, n\right) &\propto \int_{\mathbb{R}^+}\int_{\mathbb{R}^+}\int_{\mathbb{R}^+} \exp\left\{-\sigma^{23}\sqrt{h_{22}}\sqrt{h_{33}}\left(xy - \sqrt{\left(1-x^2\right)\left(1-y^2\right)}\right)\right\} \\ &\quad \cdot \left[\left(1-x^2\right)\left(1-y^2\right)\right]^{(n-3)/2}\left[1 + \sum_{k=1}^{\infty}\left(\prod_{r=0}^{k-1}\frac{n/2-1+r}{n-2+r}\right)\right. \\ &\quad \left. \cdot \frac{\left(-2\sigma^{23}\sqrt{h_{22}}\sqrt{h_{33}}\sqrt{\left(1-x^2\right)\left(1-y^2\right)}\right)^k}{k!}\right] \\ &\quad \cdot \exp\left\{-\frac{1}{2}\sum_{k=1}^{3}\sigma^{kk}h_{kk} - \sigma^{12}\sqrt{h_{11}}\sqrt{h_{22}}x - \sigma^{13}\sqrt{h_{11}}\sqrt{h_{33}}y\right\} \\ &\quad \cdot \left(\prod_{k=1}^{3} h_{kk}^{n/2-1}\right) dh_{11}\, dh_{22}\, dh_{33}. \end{aligned} \tag{2.66}$$

The above is difficult to analyze aside from the fact that if $\sigma^{23} = 0$ and at least $\sigma^{13} = 0$ or $\sigma^{12} = 0$ then r_{12} and r_{13} are pairwise independent. If $\sigma^{23} = 0$ then $\sigma_{12}\sigma_{13} - \sigma_{11}\sigma_{23} = 0$. Now, suppose that $\sigma_{23} = 0$ so that it must be that $\sigma_{12} = 0$ or $\sigma_{13} = 0$. For example, if $\sigma_{23} = 0$ and $\sigma_{13} = 0$ then $\sigma^{23} = 0$ and $\sigma^{13} = 0$. In this case, Equation (2.66) reduces to

$$\begin{aligned} p\left(x, y; \Sigma, n\right) &\propto \left(1-y^2\right)^{(n-3)/2}\int_{\mathbb{R}^+}\int_{\mathbb{R}^+}\left(1-x^2\right)^{(n-3)/2}\left(\prod_{k=1}^{2} h_{kk}^{n/2-1}\right) \\ &\quad \cdot \exp\left\{-\frac{1}{2}\sum_{k=1}^{2}\sigma^{kk}h_{kk} - \sigma^{12}\sqrt{h_{11}}\sqrt{h_{22}}x\right\} dh_{11}\, dh_{22} \end{aligned} \tag{2.67}$$

$$\propto \quad p\left(x;\Sigma,n\right)p\left(y;n\right). \tag{2.68}$$

It follows that the upper-left 3×3 submatrix of Σ can be one of the three forms in Equations (2.69) in order for r_{12} and r_{13} to be pairwise independent.

$$\begin{bmatrix} \sigma_{11} & 0 & 0 \\ 0 & \sigma_{22} & 0 \\ 0 & 0 & \sigma_{33} \end{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{12} & 0 \\ \sigma_{12} & \sigma_{22} & 0 \\ 0 & 0 & \sigma_{33} \end{bmatrix} \begin{bmatrix} \sigma_{11} & 0 & \sigma_{13} \\ 0 & \sigma_{22} & 0 \\ \sigma_{13} & 0 & \sigma_{33} \end{bmatrix} \tag{2.69}$$

In Equation (2.69) we denote with σ_{12} and σ_{13} potentially non-zero values in the upper-left 3×3 submatrix of Σ. □

Next, we examine the second case, i.e. there is not a vertex in common. The following lemma provides the topolgies of the association network graph $G=(V,E)$ in terms of the covariance matrix Σ in which r_{ij} and r_{hk} are pairwise independent where the vertex pairs $\{V_i,V_j\}$ and $\{V_h,V_k\}$ are not adjacent.

Lemma 2.14. *If $N\geq5$ the upper-left 4×4 submatrix of Σ is of one of the forms*

$$\begin{bmatrix} \sigma_{11} & 0 & 0 & 0 \\ 0 & \sigma_{22} & 0 & 0 \\ 0 & 0 & \sigma_{33} & 0 \\ 0 & 0 & 0 & \sigma_{44} \end{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{12} & 0 & 0 \\ \sigma_{12} & \sigma_{22} & 0 & 0 \\ 0 & 0 & \sigma_{33} & \sigma_{34} \\ 0 & 0 & \sigma_{34} & \sigma_{44} \end{bmatrix} \begin{bmatrix} \sigma_{11} & 0 & 0 & 0 \\ 0 & \sigma_{22} & \sigma_{23} & \sigma_{24} \\ 0 & \sigma_{23} & \sigma_{33} & \sigma_{34} \\ 0 & \sigma_{24} & \sigma_{34} & \sigma_{44} \end{bmatrix} \tag{2.70}$$

where we denote with σ_{ij} for $1\leq i<j\leq4$ potentially non-zero values, then r_{12} and r_{34} are pairwise independent.

Proof. Suppose that $N\geq5$ and that $p'=4$. By following along the same lines as the proof of Lemma 2.13, we find that

$$p\left(R;\Sigma,n\right) \quad \propto \quad \int_{\mathbb{R}^+}\int_{\mathbb{R}^+}\int_{\mathbb{R}^+}\int_{\mathbb{R}^+} |R|^{(n-p'-1)/2}\exp\left\{-\frac{1}{2}\sum_{k=1}^{p'}\sigma^{kk}h_{kk}\right.$$

$$-\sum_{i=1}^{p'-1}\sum_{j=i+1}^{p'}\sigma^{ij}\sqrt{h_{ii}}\sqrt{h_{jj}}r_{ij}\Bigg\}\left(\prod_{k=1}^{p'}h_{kk}^{n/2-1}\right)dh_{11}\,dh_{22}\,dh_{33}\,dh_{44}. \quad (2.71)$$

Suppose that the upper-left 4×4 submatrix of Σ is of the form

$$\begin{bmatrix} \sigma_{11} & 0 & 0 & 0 \\ 0 & \sigma_{22} & \sigma_{23} & \sigma_{24} \\ 0 & \sigma_{23} & \sigma_{33} & \sigma_{34} \\ 0 & \sigma_{24} & \sigma_{34} & \sigma_{44} \end{bmatrix} \quad (2.72)$$

or, in terms of the association network $G=(V,E)$, an isomorphic subgraph to that induced by the matrix above. In Equation (2.72) we denote with σ_{23}, σ_{24} and σ_{44} potentially non-zero values in the upper-left 4×4 submatrix of Σ. In this way, $\sigma^{12}=\sigma^{13}=\sigma^{14}=0$. This means that Equation (2.71) reduces to

$$\begin{aligned} p\left(R;\Sigma,n\right) \quad \propto \quad & \int_{\mathbb{R}^+}\int_{\mathbb{R}^+}\int_{\mathbb{R}^+}\int_{\mathbb{R}^+}|R|^{(n-5)/2}\exp\Bigg\{-\frac{1}{2}\sum_{k=1}^{4}\sigma^{kk}h_{kk} \\ & -\sum_{i=2}^{3}\sum_{j=i+1}^{4}\sigma^{ij}\sqrt{h_{ii}}\sqrt{h_{jj}}r_{ij}\Bigg\}\left(\prod_{k=1}^{4}h_{kk}^{n/2-1}\right)dh_{11}\,dh_{22}\,dh_{33}\,dh_{44} \quad (2.73) \end{aligned}$$

The goal is to determine the bivariate probability distribution of r_{12} and r_{34}. Let $u=r_{12}$, $v=r_{13}$, $w=r_{14}$, $x=r_{23}$, $y=r_{24}$, and $z=r_{34}$. This will require marginalizing $v=r_{13}$, $w=r_{14}$, $x=r_{23}$, and $y=r_{24}$. Written with the new labels, Equation (2.73) becomes

$$\begin{aligned} p\left(u,v,w,x,y,z;\Sigma,n\right) \quad \propto \quad & \int_{\mathbb{R}^+}\int_{\mathbb{R}^+}\int_{\mathbb{R}^+}\int_{\mathbb{R}^+}\begin{vmatrix} 1 & u & v & w \\ u & 1 & x & y \\ v & x & 1 & z \\ w & y & z & 1 \end{vmatrix}^{(n-5)/2} \\ & \cdot\exp\Bigg\{-\frac{1}{2}\sum_{k=1}^{4}\sigma^{kk}h_{kk}-\sigma^{23}\sqrt{h_{22}h_{33}}\cdot x \quad (2.74) \\ & -\sigma^{24}\sqrt{h_{22}h_{44}}\cdot y-\sigma^{34}\sqrt{h_{33}h_{44}}\cdot z\Bigg\} \quad (2.75) \end{aligned}$$

$$\cdot \left(\prod_{k=1}^{4} h_{kk}^{n/2-1} \right) dh_{11}\, dh_{22}\, dh_{33}\, dh_{44}. \tag{2.76}$$

We begin by marginalizing $w = r_{14}$. The determinant $|R|$ is a quadratic function of w given by

$$|R| = aw^2 + bw + c \tag{2.77}$$

where

$$a = -\left(1 - x^2\right) \tag{2.78}$$

$$b = 2\left(uy + vz - vxy - uxz\right) \tag{2.79}$$

$$\begin{aligned} c = {} & 1 - u^2 - v^2 - x^2 - y^2 - z^2 + u^2 z^2 + v^2 y^2 \\ & + 2\left(uvx + xyz - uvyz\right). \end{aligned} \tag{2.80}$$

Thus, to marginalize $w = r_{14}$ we must compute the integral

$$\begin{aligned} I_w = {} & \int_{\frac{b}{2a} + \frac{\sqrt{b^2-4ac}}{2a}}^{\frac{b}{2a} - \frac{\sqrt{b^2-4ac}}{2a}} \left[a \left(w + \frac{b}{2a} - \frac{\sqrt{b^2 - 4ac}}{2a} \right) \right. \\ & \left. \left(w + \frac{b}{2a} + \frac{\sqrt{b^2 - 4ac}}{2a} \right) \right]^{(n-5)/2} dw. \end{aligned} \tag{2.81}$$

Using substitution, we find that

$$I_w \propto \frac{\left(b^2 - 4ac\right)^{(n-4)/2}}{(-a)^{(n-3)/2}}. \tag{2.82}$$

Now, $b^2 - 4ac$ is a quadratic function of v given by

$$b^2 - 4ac = dv^2 + ev + f \tag{2.83}$$

where

$$d = 4\left[(z-xy)^2 - \left(1-y^2\right)\left(1-x^2\right)\right] \tag{2.84}$$

$$e = 8u\left[(y-xz)(z-xy) + \left(1-x^2\right)(x-yz)\right] \tag{2.85}$$

$$f = 4u^2\left[(y-xz)^2 - \left(1-x^2\right)\left(1-z^2\right)\right] \tag{2.86}$$

$$-4\left(1-x^2\right)\left[(x-yz)^2 - \left(1-y^2\right)\left(1-z^2\right)\right]. \tag{2.87}$$

Thus, to marginalize $v = r_{13}$ we must compute the integral

$$I_v = \int_{\frac{e}{2d}+\frac{\sqrt{e^2-4df}}{2d}}^{\frac{e}{2d}-\frac{\sqrt{e^2-4df}}{2d}} \left[d\left(v+\frac{e}{2d}-\frac{\sqrt{e^2-4df}}{2d}\right)\left(v+\frac{e}{2d}+\frac{\sqrt{e^2-4df}}{2d}\right)\right]^{(n-4)/2} dv. \tag{2.88}$$

Using substitution, we find that

$$I_v \propto \frac{\left(e^2-4df\right)^{(n-3)/2}}{(-d)^{(n-2)/2}}. \tag{2.89}$$

Finally, note that

$$e^2 - 4df \propto \left(1-u^2\right)\left(1-x^2\right)\left[1-x^2-y^2-z^2+2xyz\right]^2 \tag{2.90}$$

so that the joint pdf of $u = r_{12}$, $x = r_{23}$, $y = r_{24}$, and $z = r_{34}$ can be factored as

$$p(u,x,y,z;\Sigma,n) \propto p(u;n)\,p(x,y,z;\Sigma,n) \tag{2.91}$$

As a result, $u = r_{12}$ and $z = r_{34}$ are pairwise independent.

Similar computations may be carried out to show that r_{12} and r_{34} are pairwise independent when the upper-left 4×4 submatrix of Σ is of the form

$$\begin{bmatrix} \sigma_{11} & 0 & 0 & 0 \\ 0 & \sigma_{22} & 0 & 0 \\ 0 & 0 & \sigma_{33} & 0 \\ 0 & 0 & 0 & \sigma_{44} \end{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{12} & 0 & 0 \\ \sigma_{12} & \sigma_{22} & 0 & 0 \\ 0 & 0 & \sigma_{33} & \sigma_{34} \\ 0 & 0 & \sigma_{34} & \sigma_{44} \end{bmatrix} \tag{2.92}$$

or, in terms of the association network $G = (V, E)$, an isomorphic subgraph to that induced by the matrices above. Again, in the above we have indicated the potentially non-zero values in the upper-left 4×4 submatrix of Σ. $\square$

Evidently, the value of $\mathrm{Cov}\,(Y_{ij}, Y_{hk})$ depends in large part upon the topology of the network graph G. For this reason, we choose to continue our demonstration on the specific topology, i.e. a Wattz-Strogatz network model as outlined in Section 2.3.2, for a network graph G in the following. Table 2.1 displays the total number of 3×3 and 4×4 subgraphs isomorphicly unique when keeping the anchor vertex pairs $\{i, j\} \in V^{(2)}$ and $\{h, k\} \in V^{(2)}$ fixed when $G = (V, E)$ is a one-dimensional lattice in which $G = (V, E)$ has $p = |V|$ vertices and each vertex is connected by an edge to d of its neighbors on each side for $d << p$. In addition to the counts of each isomorphic subgraph in G, Table 2.1 indicates whether the value of $\mathrm{Cov}\,(Y_{ij}, Y_{hk})$ is zero or non-zero based upon the conclusions of Lemmas 2.13 and 2.14. Figures 2.1, 2.2, and 2.3 display the topologies of the upper-left 3×3 and 4×4 topologies corresponding to the ID's of Table 2.1.

It is important to observe that the highest order of the number of subgraphs with non-zero $\mathrm{Cov}\,(Y_{ij}, Y_{hk})$ is $O\left(p^2\right)$. Now the order of the covariance terms needs to be addressed after random edges have been added per the Watts-Strogatz network model.

Lemma 2.15. *If $G = (V, E)$ is an Erdos-Renyi random graph, see [Bol85], on $|V| = p$ vertices and probability of an edge joining a vertex pair is $q \in O\left(p^{-1}\right)$ then the number of*

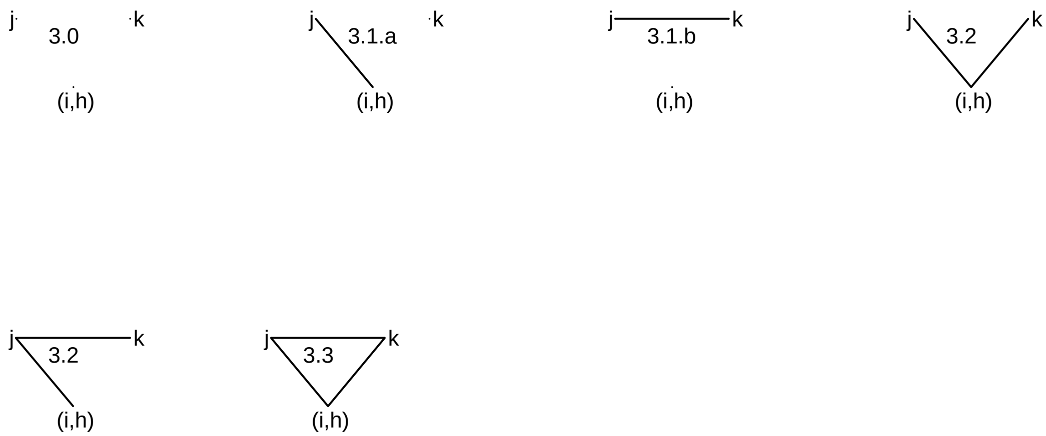

Figure 2.1: Configurations When $\{i, j\}$ and $\{h, k\}$ Are Adjacent

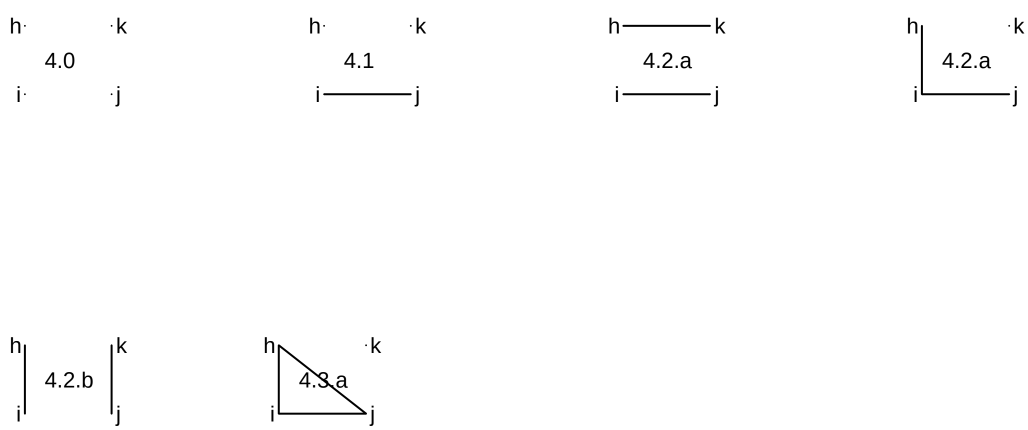

Figure 2.2: Configurations When $\{i, j\}$ and $\{h, k\}$ Are Non-adjacent

ID	#Vertices	#Edges	#Subgraphs	$\mathrm{Cov}(Y_{ij}, Y_{hk})$
3.0	3	0	$\frac{p(p-3d-2)(p-3d-1)}{2}$	0
3.1.a	3	1	$2dp^2 - (5d+3)\,dp$	0
3.1.b	3	1	$dp^2 - \frac{(5d+3)}{2}p$	$*$
3.2	3	2	$\frac{3}{2}(d+1)\,dp$	$*$
3.3	3	3	$\frac{3}{2}d\,(d-1)\,p$	$*$
4.0	4	0	$\frac{p(p-4d-1)(p-4d-2)(p-4d-3)}{8}$	0
4.1	4	1	$\frac{3d}{2}p^3 - \frac{3d(7d+4)}{2}p^2 + \frac{d\left(37d^2+42d+11\right)}{2}p$	0
4.2.a	4	2	$\frac{d(4d+3)}{2}p^2 - \frac{d\left(13d^2+17d+5\right)}{2}p$	0
4.2.b	4	2	$d^2p^2 - d^2\,(3d+2)\,p$	$*$
4.3.a	4	3	$\frac{3}{2}d\,(d-1)\,p^2 - \frac{1}{2}d\,(8d+5)\,(d-1)\,p$	0
4.3.b	4	3	$\frac{1}{2}d\,(2d+1)\,(d+1)\,p$	$*$
4.4	4	4	$(d+1)\,d\,(d-1)\,p$	$*$
4.5	4	5	$\frac{1}{2}(d+1)\,d\,(d-1)\,p$	$*$
4.6	4	6	$\frac{1}{2}d\,(d-1)\,(d-2)\,p$	$*$

Table 2.1: Number of Uniquely Isomorphic 3×3 and 4×4 Subgraphs; $*$ Indicates a non-zero value of $\mathrm{Cov}(Y_{ij}, Y_{hk})$ when the attributes of the vertices in the association network graph $G = (V, E)$ has the multivariate Gaussian distribution

non-zero $\mathrm{Cov}(Y_{ij}, Y_{hk})$ *is* $O_P\left(p^2\right)$ *which immediately implies that if*

$$\left|\mathbb{P}\left(Y_{ij} = 1 | Y_{hk} = 1\right) - \mathbb{P}\left(Y_{ij} = 1\right)\right| \overset{p \to \infty}{\to} 0 \tag{2.93}$$

then $K \overset{p \to \infty}{\to} 0$.

Proof. We wish to show that for any $\epsilon > 0$ there exist $M_\epsilon > 0$ and $p_0 > 0$ such that for all $p > p_0$

$$\mathbb{P}\left(\frac{CP}{p^2} \leq M_\epsilon\right) \geq 1 - \epsilon \tag{2.94}$$

or, equivalently,

$$\mathbb{P}\left(CP > M_\epsilon p^2\right) \leq \epsilon \tag{2.95}$$

where CP represents the number of subgraphs which are contributing coveriance (C) in the probabilistic model (P). The P_{ij} refers to the configurations in which the subgraphs of Figures 2.1, 2.2, 2.3.

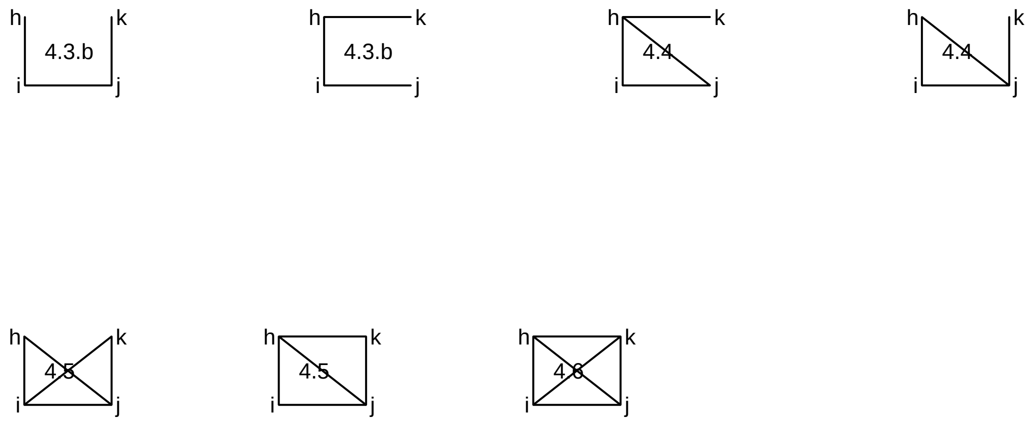

Figure 2.3: Configurations When $\{i, j\}$ and $\{h, k\}$ Are Non-adjacent

Define the quantity CP' as

$$CP' = 3\left(P_{31} + P_{32} + P_{33} + P_{42} + P_{43} + P_{44} + P_{45} + P_{46}\right). \tag{2.96}$$

It is clear that $CP' \geq CP$ which implies that

$$\mathbb{P}\left(CP > M_\epsilon p^2\right) \leq \mathbb{P}\left(CP' > M_\epsilon p^2\right) \tag{2.97}$$

which, by using Markov's inequality, we further have that

$$\mathbb{P}\left(CP > M_\epsilon p^2\right) \leq \frac{\mathbb{E}CP'}{M_\epsilon p^2}. \tag{2.98}$$

Now, we compute the parts of the expectation in the upper-bound.

$$\begin{aligned}
\mathbb{E}P_{31} &= 3L_{30}p\left(1-p\right)^2 + 2L_{31}\left(1-p\right)^2 \\
\mathbb{E}P_{32} &= 3L_{30}p^2\left(1-p\right) + 2L_{31}p\left(1-p\right) + L_{32}\left(1-p\right)
\end{aligned}$$

$$
\begin{aligned}
\mathbb{E}P_{33} &= L_{30}p^3 + L_{31}p^2 + L_{32}p + L_{33} \\
\mathbb{E}P_{42} &= 15L_{40}p^2(1-p)^4 + 5L_{41}p(1-p)^4 + L_{42}(1-p)^4 \\
\mathbb{E}P_{43} &= 20L_{40}p^3(1-p)^3 + 10L_{41}p^2(1-p)^3 + 4L_{42}p(1-p)^3 + L_{43}(1-p)^3 \\
\mathbb{E}P_{44} &= 15L_{40}p^4(1-p)^2 + 10L_{41}p^3(1-p)^2 + 6L_{42}p^2(1-p)^2 \\
&\quad +3L_{43}p(1-p)^2 + L_{44}(1-p)^2 \\
\mathbb{E}P_{45} &= 6L_{40}p^5(1-p) + 5L_{41}p^4(1-p) + 4L_{42}p^3(1-p) \\
&\quad 3L_{43}p^2(1-p) + 2L_{44}p(1-p) + L_{45}(1-p)^2 \\
\mathbb{E}P_{46} &= L_{40}p^6 + L_{41}p^5 + L_{42}p^4 + L_{43}p^3 + L_{44}p^2 + L_{45}p + L_{46}
\end{aligned}
$$

Suppose that p is defined as $p = c/p$ for some $c \geq 0$. From this choice it follows that there exists $M_1 > 0$ and $p_1 > 0$ such that for all $p > p_1$

$$\frac{\mathbb{E}CP'}{p^2} \leq M_1 \tag{2.99}$$

or, in other words,

$$\mathbb{E}CP' \in O\left(p^2\right). \tag{2.100}$$

As a consequence, we further the upper-bound to

$$\mathbb{P}\left(\frac{CP}{p^2} > M_\epsilon\right) \leq \frac{M_1}{M_\epsilon} \tag{2.101}$$

for $p > p_1$. Now, for a choice of

$$M_\epsilon = \frac{M}{\epsilon} \tag{2.102}$$

and $p_0 = p_1$, we have that

$$CP \in O_P\left(p^2\right) \tag{2.103}$$

or, similarly, there exists $M > 0$ such that

$$\lim_{p\to\infty} \mathbb{P}\left(\frac{CP}{p^2} \leq M\right) = 1. \tag{2.104}$$

The Watts-Strogatz small-world network model has $O_P(p)$ vertex pairs connected with an edge and $O_P\left(p^2\right)$ vertex pairs disconnected. Suppose that the probability of type-I error $\alpha = \alpha(p)$ is a function of p and that $\alpha \in O\left(p^{-1/2}\right)$. This implies that $\lambda_1 \in O\left(p^{3/2}\right)$ so that

$$\frac{\sum_{\{i,j\}\in E^c} \alpha_{ij}^2 + \sum_{\{i,j\}\in E} \beta_{ij}^2}{\lambda} \in O\left(p^{-1/2}\right). \tag{2.105}$$

Now, according to Table 2.1 and to the results previously stated,

$$\begin{aligned}
\frac{\sum\sum\limits_{\{i,j\}\neq\{h,k\}\in V^{(2)}} |Cov\left(Y_{ij}, Y_{hk}\right)|}{\lambda} &\in O\left(\sqrt{p}\,|Cov\left(Y_{ij}, Y_{hk}\right)|\right) &(2.106)\\
&\in O\left(|\mathbb{P}\left(Y_{ij}=1|Y_{hk}=1\right) - \mathbb{P}\left(Y_{ij}=1\right)|\right) &(2.107)
\end{aligned}$$

since $\alpha \in O\left(p^{-1/2}\right)$ and where $\{i,j\}, \{h,k\} \in E^c$. It follows that if

$$|\mathbb{P}\left(Y_{ij}=1|Y_{hk}=1\right) - \mathbb{P}\left(Y_{ij}=1\right)| \overset{p\to\infty}{\to} 0 \tag{2.108}$$

where, in the above, we have chosen $\lambda_1 = \lambda_2 = \lambda$, then $K \to 0$ and the claim is proved upon appealing to Theorem 2.10. □

2.4 Discussion

The statistical analysis of networks has seen tremendous theoretical and methodological gains in recent years.

The problem is (a) that researchers effectively infer networks without formal confirmation and (b) there exists little theory for the impact of noisy observations on network inference. That is, many infer a network and call it “the network”. Arguably a statistician should always first consider in any estimation or inference problem the probability distribution of any estimator.

We have provided a means to build the foundation of inference for a class of network statistics. Our methods address the propagation of error from network inference to network characterization, a topic which deserves more attention.

Proving Theorem 2.6 in Section 2.3, requiring the analysis of alternating sums of differences of the ratios of modified Bessel functions of the first kind, was of no small expenditure. The literature on quantities of this sort is lacking and, hence, we were required to develop methods of our own. These methods, of independent interest, are published in [PB13].

Chapter 3

Robust Dynamic Community Detection

3.1 Introduction

Approximately 50 million people worldwide (3 million in the United States) have epilepsy. More than 50 percent of those suffer from localization-related epilepsy. Unfortunately, 25 to 35 percent of these patients continue to have seizures despite maximal anticonvulsant therapy, see [JDW06], and furthermore those receiving medications may suffer from its considerable side effects. In many cases, regional surgical resection may reduce or cure seizures, but this procedure remains an option of last resort as it carries significant risk and is only curative in approximately 40 to 70 percent of neocortical epilepsy cases, see [neo90]. To better treat this devastating illness requires a deeper understanding of the seizure, including improved brain imaging and quantitative characterization.

Despite widely variable etiologies, epilepsy is often characterized as a disease of brain rhythms — a paroxysmal cerebral dysrhythmia, see [GF] — as brain voltage activity typically reveals stereotyped dynamics, see [KC12a]. Localization related seizures are an emergent event that typically begins focally, are believed to be due to an imbalance of excitation and inhibition, and spreads throughout the brain network. From this simplistic perspective a secondary generalized seizure (that spreads from a focus to include a large portion of the brain network) can be understood as beginning in a circumscribed region, then recruiting other brain areas through a cascade of spreading activity that includes both pathological and normal brain tissue. Presumably, this spread manifests as synchronized activity throughout the brain network, and classically seizures have been thought to represent a hypersynchronous state, see [dCG], [BB99], and [PJ54]. However, recent observations challenge the assertion that seizures are hypersynchronous events at larger spatial scales, see [Jer05] and [SK], in in vitro recordings of experimental seizures, see [NT], in human temporal lobe seizures, see [BF].

Functional networks have provided new insights into the dynamic evolution of the synchronized neuronal activity of seizure. Representing the statistical associations between dynamic activity recorded from separate brain areas, see [K.94], functional networks have been applied to many different types of multivariate neuroimaging data at different spatial and temporal scales, see [MAK10]. From high-density, multielectrode voltage recordings, complex functional network patterns emerge during siezure whose quantitative understanding requires sophisticated techniques. Many tools exist to characterize network topology, and analysis of human brain networks has resulted in many recent observations, as summarized in detailed reviews, see [PJ11] and [BB09]. In general, identification of the network characteristics that support epileptic seizures remains an active research area, yet some prevailing results have emerged, including the importance of small-world topologies [NTJ04], [PSC07], and [WHX06] (although see [BSK10]), and [ALA10].

Recent observations suggest a specific pattern of organization in functional networks during seizure [KC12b]. At seizure onset, brain voltage activity tends to become more coupled [KC12b], resulting in the coalescence of network components to form a dominant community. Then, during the seizure, these dominant communities tend to fragment; brain regions tend to decouple their activities, and global community structures break down. Finally, preceding seizure termination, brain regions recouple, and a dominant community then forms [KC12b]. However, this understanding of coalescence and fragmentation of is based on static network analysis tools applied to the dynamic functional networks inferred during a seizure. A more appropriate characterization would directly analyze the rich, evolving structure of the dynamic functional networks of seizure.

Following a decade of work in the context of static networks, in recent years there have emerged a number of proposals to perform community detection in dynamic networks (e.g., [Bas13]). These methods typically apply an algorithm for static graphs to define candidate communities at a fixed moment in time, and then define time-varying communities by "stitching together" consecutive static communities. For example, the clique percolation method (CPM) [1] defines communities in static time slices by the extent to which a clique (e.g., a triangle) can be walked over the graph, and then communities at successive time points are sewn together using a rule based on overlap of vertex subsets. More recent proposals have begun to operate directly on the dynamic network itself.

However, within the context of functional connectivity networks and epilepsy, existing dynamic community detection tools are inadequate for two reasons. First, while most existing methods typically use implicit notions of communities (e.g., defined via optimization of some cost criterion), we instead would like to extract communities based on more explicit notions as time-evolving aggregations of smaller motifs, consistent with the hypothesis that functional network motifs are important evolutionary outcomes in the neuronal architectures of highly evolved species, see [RM02]. Second, existing methods are unable to explicitly account for edge noise (i.e., false positive/negative inferences of edge status) a factor in any set of networks inferred from biological data like the invasive brain voltage measurements utilized here.

In this chapter we introduce a new method of dynamic community detection — dynamic plex percolation (DPP)— that connects static communities to aggregations of a variety of common motifs in a natural and flexible manner, that defines dynamic communities across time through an explict notion of temporal progression of these motifs, and that is demonstrably robust to edge noise. We then illustrate the utility of this method through application to invasive brain voltage recordings made from a human subject during seizure.

3.2 Background

Existing (both static and dynamic) community detection methods provide powerful tools for understanding and characterizing networks. However, these methods typically are (i) unable to account for the edge noise (e.g. false positive/negative edge status) that is inherently present in “functional networks” inferred from noisy data, and (iia) either discover dynamic communities by sewing together discovered static communities in an ad-hoc manner or (iib) by using an implicit notion of community (e.g. by optimizing some cost criterion), see [PJMO10], [AWP], [TYBW10], and [GP07]. These shortcommings potentially limit our understanding of inferred networks and, in particular, the functional networks inferred from brain voltage data of interest here.

What is primarily lacking in the static community dectection is robustness to the noise that is inherent in an inferred edge set. Specifically, one is typically less able to directly control type-II error with a fixed dataset. In the methods section of this chapter I discuss a solution to this problem and demonstrate its capacity for robustness to error, specifically type-II error.

Dynamically, I seek a method that is locally based instead of globally based. In this sense, we mean that, for example in the context of a social network, if two people cease to be friends then we should not necessarily fracture the community structure. The Clique Percolation Method (CPM) as defined by the authors of [GP07] achieves the goal of a local definition of community structure statically but additional advancements are required for dynamic community detection; this point is discussed in the following. I propose a method which is both mathematically elegant and intuitive, thus providing both rigor and interpratibility. Methods which optimize a criterion, such as minimizing graph conductance, are computationally tractable and mathematically elegant via spectral methods but lack an intuitive explanation, see [Bol98]. Furthermore, optimizing a crition can cause a local definition of a community to fail and is, therefore, undesireable since communities are local by nature. The last criterion I seek is a method which explicity traces dynamic communities in time rather than defining an ad-hoc “stitching” mechanism to link static communities. Based upon the above I define the *dynamic plex percolation* (DPP) method of dynamic community detection.

The DPP acheives each of the goals (i-iib) but additionaly, and of equal importance, the proposed method detects network motifs. A network motif of size n is a connected subgraph on m vertices. A vision of networks, in particular that of a brain functional network, as described by the authors of [LVZ05], is of interconnected microscopic patterns (motifs) coalescing to form mesoscopic patterns called “themes” which then themselves interconnect to encompass the macroscopic network. In the following, a demonstration of the DPP’s ability to detect signifcant motifs of neurological networks, those as cited by [LVZ05] and [RM02], is demonstrated.

3.3 Methods

Let $G = (V, E)$ be a graph with vertex set V and edge set E, and let $\{G_t\}_{t=1}^{n}$ denote a sequence of n graphs indexed by t. Our goal is to first identify time-indexed sequences of subgraphs $\left\{H_t^{(k_t)}; k_t \in \mathbb{Z}^+\right\}_{t=1}^{n}$ such that for each $H_t^{(k_t)}$ the vertices within are more cohesive, than to vertices outside. We use the notation k_t to mean that, for ultimate generality, one could choose to tune the requisite level of cohesiveness over time for declaring community structure. Following this, our second goal is to identify a sequence $H_{t-1}^{k_{t-1}}, H_t,^{k_t}, H_{t+1}^{k_{t+1}}$ such that a similar cohesiveness has evolved in time.

Definition 3.1. *A subgraph H on n vertices is a $k-plex$ if*

$$deg\,(v) \geq n - k \quad \textit{for all vertices } v \in H$$

for $n > k$.

Remark 3.2. *It is straight-forward to recognize that a $1-plex$ is a clique. This shows us that a plex is a generalization of a clique and, therefore, provides a more flexible notion of cohesiveness.*

The authors of [RM02] found that the diamond and multi-layer perceptron motifs, as shown in Figure 3.1, occured significantly more frequently in biological and technological networks than in random networks conditional on the smaller subgraphs. In con-

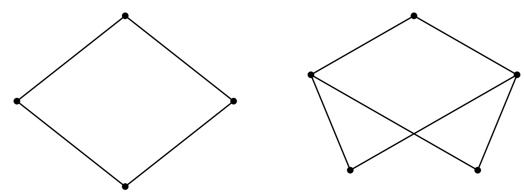

Figure 3.1: The Diamond (left) and Multi-Layer Perceptron (right) Motifs

trast with the existing CPM, the proposed DPP method is able to identify or, in other words, aggregate many of the motifs the authors of [RM02] established as significant into themes/communities. Table 3.2 is a display of the wide-ranging detection abilities of the DPP with respect to these statistically significant motifs.

The authors of [LVZ05] discussed the notion that themes, i.e. the mesoscopic structures of a network or, in other words communities, are an aggregation of the microscopic motifs. Here, we use the notion of a k-plex to represent a motif. Recall that a plex of order (n, k) is a subgraph H on $n > k$ vertices in which $deg\,(v) \geq n - k$ for all $v \in H$. To aggregate the motifs into dynamic themes or communities (herein the terms "themes" and "communities" are used interchangably) I define a community by the extent to which a plex of a given order can "walk" both within each graph G_t and across temporally adjacent graphs G_t and G_{t+1}, where movement necessarily must be from one plex sub-graph to another in such a way that the two differ by at most one vertex, see [GP07]. Movement across time is facilitated by fully connecting each pair of subsets vertices in G_t and G_{t+1}, for every t. That is, each vertex $v \in G_t$ is connected by an edge to its matured self in G_{t+1} the results of which can be visualized as a railroad track. Note that this technique for multi-slice networks has been utilized by the authors of [PJMO10]. Next, if an edge $\{v_1, v_2\} \in G_t$ exists again in G_{t+1} then the vertices $v_1 \in G_t$ and $v_2 \in G_{t+1}$ are connected by an edge and likewise for $v_2 \in G_t$ and $v_1 \in G_{t+1}$ the result of which is converting the "railroad track" into a proper bridge for plex "walking". The enhanced sequence $\{G_t\}_{t=1}^{n}$ is the foundation upon which the plexes walk in space and time, and thus dynamic communities are defined. It is important to note that the enhanced graph sequence is independent of the choice of both plex size and order.

The notion of "walking" in **space** is similar to the CPM of [GP07], however, the notion of walking in **time** is novel. Moreover, the fundamental units being plexes and motifs, rather than cliques, drastically alters the detectable dynamic communities, as illustrated below. This novel approach for walking in time is more explicit han existing methods, in the sense that it is deterministic and interpretable as demonstrated in Table 3.3. The result of explicitly defining a means for plex walking through space and time may be conceptualized as an evolving series of "tubes" which represent the dynamic networks evolving in time.

3.4 Communities as Aggregations of Network Motifs

As mentioned in the previous subsection, a network is a coalescence of mesoscopic themes or communities which themselves are an aggregation of microscopic motifs, see Table 3.1. The authors of [RM02] hypothesize motifs in neuronal functional networks to be important

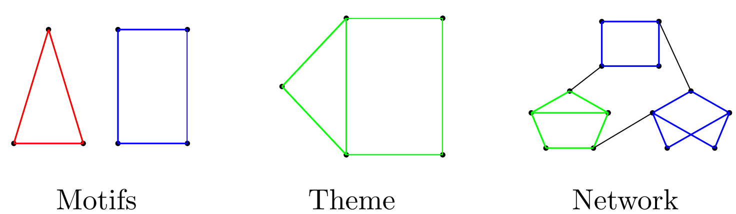

Table 3.1: The Microscopic, Mesoscopic, Macroscopic Perspective. The two motifs (left, red and blue) unit to form a theme (middle, green), which forms one component of a network (right)

to the brain achritectures of highly evolved species. The authors of [LVZ05] concluded that higher order connection patterns, which they termed themes, i.e. an aggregation of motifs, were significantly discovered in neurological functional networks. We show in Table 3.2 that all of the non-trivial $n = 5$ motifs are at least one of, or an interleaving of, a plex of order $(3, 1)$, $(4, 2)$, $(5, 3)$, or $(5, 2)$.

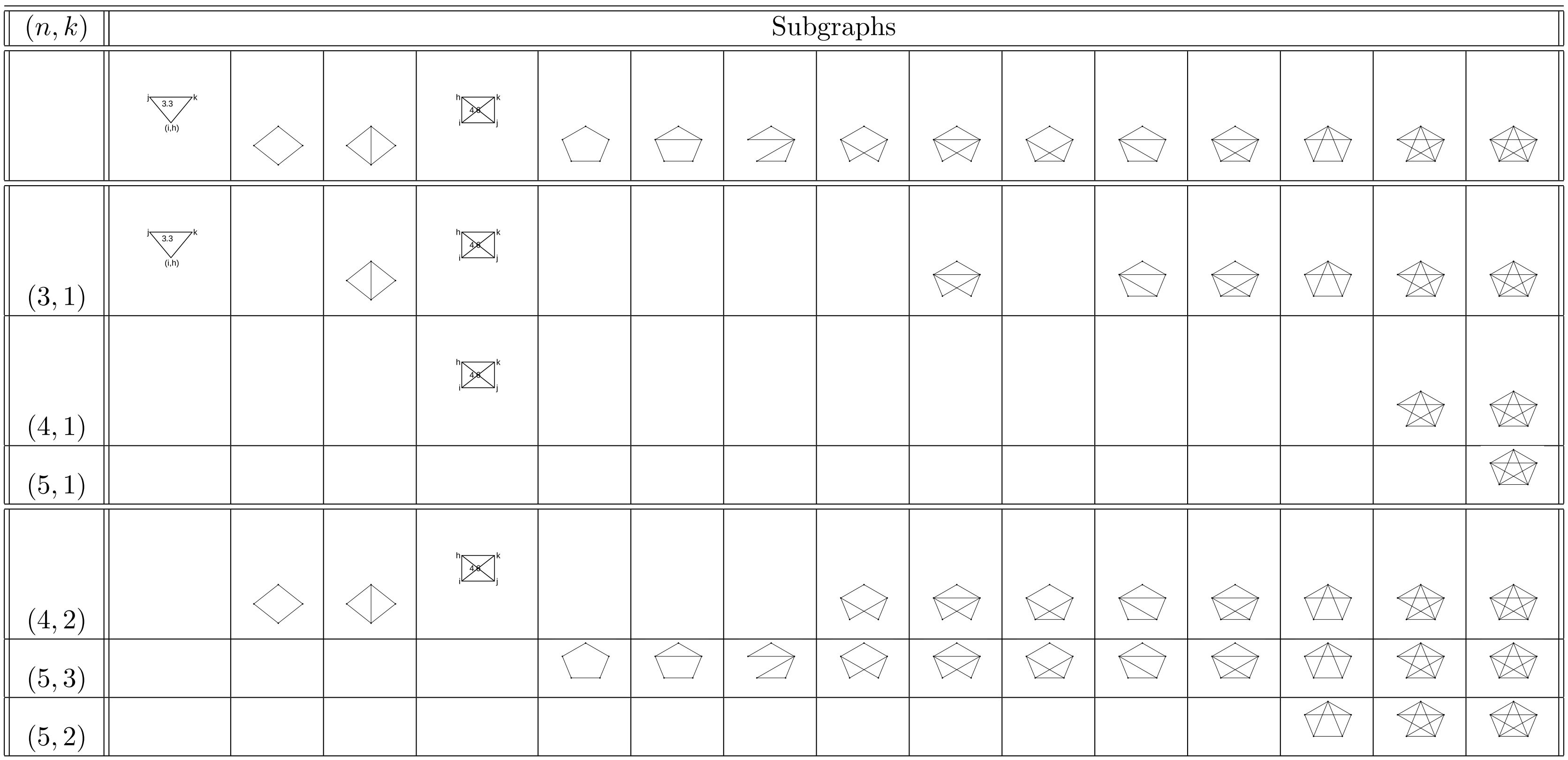

Table 3.2: Network Motifs Detectable via Plexes. By (n, k) we mean a plex of order k on n vertices. For instance, in the above we display all subgraphs on $n = 3, 4, 5$ vertices which can be discovered by order $k = 1, 2, 3$ plexes. The top row represents the motifs which we prefer to discover and which will form the themes as an aggregation thereof. The next six rows represent the order (n, k) plex which can discover such a motif in the respective column.

3.5 Dynamic Communities Through Progression of Motifs

The simple example in Table 3.3 illustrates how plexes walking across time lead to more sensible dynamic communities than existing approaches, such as "stitching" [GP07]. Briefly, in

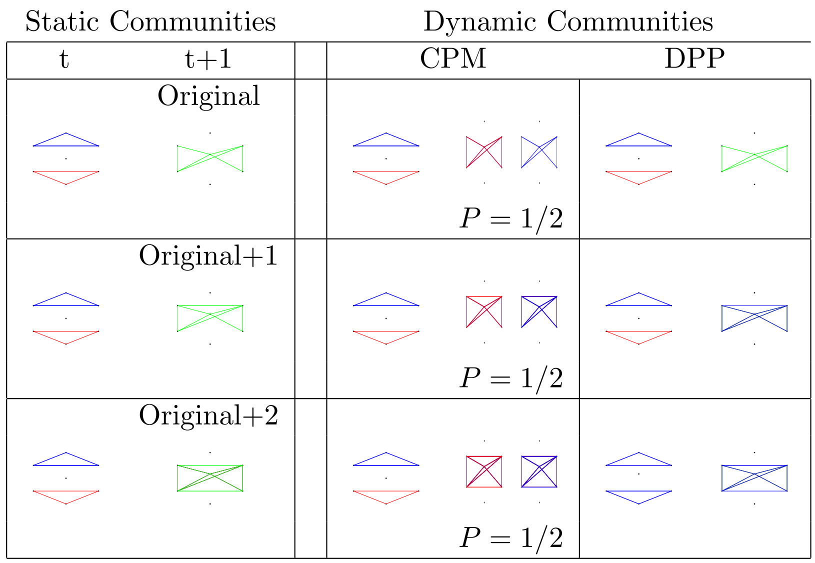

Table 3.3: An Order $(3, 1)$ Plex Walking in Time. Displayed above in the first column are static communities discovered both by DPP and CPM at time t and $t+1$. At time $t+1$ we consider three different choices of networks. In the second column are the dynamic communities discovered by CPM. Notice that (i) the structure of the graph at $t+1$ is different in all three cases but (ii) the labeling of dynamic community is still non-deterministic and arbitrary. In the third column are the dynamic communities discovered by DPP. The key point is that by allowing for the plex template to walk in time we acheive an explicit and deterministic dynamic community.

stiching, the set of static communities of G_t are paired with the static communities of G_{t+1} based upon relative overlap. However, the usage of this method gives rise to two important issues: (i) In some cases, the stiching algorithm cannot decide from which community in G_t a community in G_{t+1} arose as shown in Table 3.3 . And, (ii) if two communities "merge" then the authors of [GP07] declare that the smaller community "comes to the end of its life at t, and is swallowed by the [larger community]". Similarly, if a community fractures, the authors of [GP07] would declare a new community to be born and only allow the "larger" of the decendants to maintain the former label.

Both (i) and (ii) are important distinctions. For case (i) consider Table 3.3 and specifically note that the three rows each have unique static community structure in the "Original" column . However, in the CPM, the label that continues from time t to $t+1$ is chosen arbitrarily (i.e., by a fair coin flip) from the existing community labels at time t. This choice was made from the authors of [GP07]. In constrast, the DPP method uniquely and consistently the label which proceeds from times t to $t+1$. In the first row, there exist no bridges so no labels proceed. In the second row, the upper triangle has an edge persist from time t to $t+1$ so the decendant is labelled as the upper triangle (in this case, the color blue). In the third row, both the upper and lower triangles at time t have an equal share in the community at $t+1$ so we propagate one of the labels back to time t. This "backwards labeling" may seem counterintuitive. However, the merger of two communities is simply a split in reverse time, and therefore the matching of labels becomes quite intuititive. The key distinction between CPM and DPP in the third row is that CPM declares the lower triangle to come to the end of its life while for the DPP its label has always been blue. In conclusion, labeling of dynamic communities using the DPP is explicit and deterministic.

3.6 Robustness to Noise

In neuroscience, "functional networks" represent statistical associations between brain areas. In these inferred networks, the edges represent the assignment of some significant level of assocation between a pair of vertices; therefore, the assignment of edges is inherently noisy. Take for example the networks constructed by the authors of [MAK08] in which an edge between two vertices, wherein a vertex represents an electrode on the cortical surface of a patient's brain, indicates that the voltage recordings of the two electrodes are significantly correlated. Noise in the original signal, in this case the voltage readings, propagates to a noisy edge set in which false positives and false negatives occur with some probability. Now, understanding that the edge set is noisy and that any community detection algorithm uses as input the edge set, it becomes clear that one must take care in designing a community detection algorithm that is robust to noise, specifically false negatives. False negatives, or type-II errors, can have a substanial shattering effect on the detection of communities within a network graph. It is important to note that sparseness or density does not play a role here. No matter the topology of the original graph we need to specifically control for Type-II errors. What is most important is that the motifs are discovered.

Consider for example Table 3.4 in which the true network graph is depicted in the top row. Note the discrepancy between the left and right columns. Prior to any false negatives,

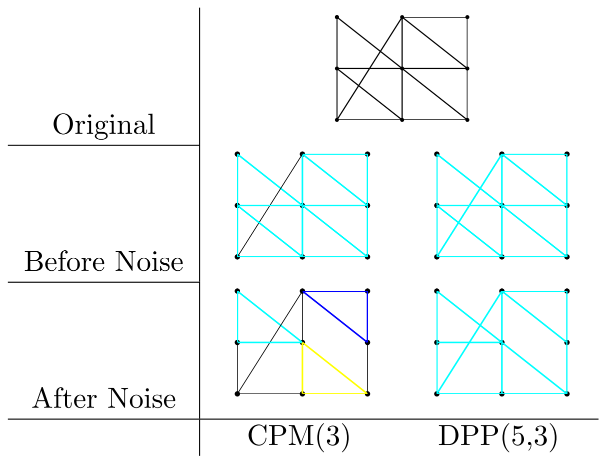

Table 3.4: Static Community Detection: A Vastly Different Result with $\mathbb{P}\,(\text{type-II error})\,0.125$. The first row displays the original network and the second displays the communities detected by the CPM order $(3, 1)$ and DPP order $(5, 3)$. The third row displays the detected communities by the same methods but after two edges have been removed. It is important to notice that the CPM now detects three distinct communities from what is truly one community while, on the other hand, the DPP the original true community remains intact.

the CPM(3) and the DPP(5,3) place every vertex into a single community. However, following two out of sixteen $(2/16 = 0.125)$ edge removals, i.e. probability of type-II error being 0.125, even the most flexible of the CPM algorithms shatters one community into three where as the DPP has managed to correctly identify, despite the noise in the edge set, the community structure, demonstrating a robustness to noise. We have specifically demonstrated type-II errors in the figure since type-I errors may be more directly controlled and, therefore, less of an issue when considering noisy networks. For example, Type I errors can be controlled through a Bonferroni correction.

To further illustrate the robustness of the DPP method to noise, I begin with a 5-clique and remove edges one at a time to directly affect a false negative in the edge set.The results are shown in Table 3.6. The column headed by (n, k) represents the orders which discover the motif in the second column as a community. Examinination of the "Discovery Rate", defined in Table 3.6, that the DPP is able to sustain a much greater type-II error rate in the discovery of network communities.

In Table 3.5 in both figures the curves, represent the percent of vertices which remain in the 5-clique chain giant component using CPM(4), DPP(4,2), and DPP(5,3). The 5-clique chain, i.e. the true graph, is a graph on 35 vertices and is a chain of 5-cliques each of which overlap by four vertices or, in other words, overlap by a 4-clique. In the true graph an order $(5, 1)$ plex would walk from end-to-end with no difficulty. However, as we can see, immediately upon some edge removals, i.e. type-II errors, the allowance for plexes of order (n, k) with $k > 1$ provides robustness and the giant community remains intact. Both DPP(4,2) and DPP(5,3) are able to sustain much greater type-II error rates and detect the giant community (originally the 5-clique chain).

Further illustrations of the DPP methods robustness to noise are shown in Table 3.5. To see the point further demonstrated, consider Table 3.6.

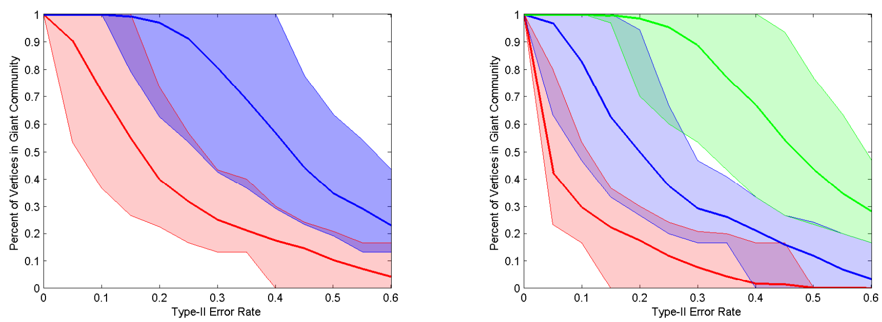

Table 3.5: Percent of Vertices in Giant Community. Left Pane: Red= $CPM\,(4)$ Blue= $DPP\,(4,2)$. Right Pane: Red= $CPM\,(4)$ Blue= $DPP\,(4,2)$ Green= $DPP\,(5,3)$. Each curve represents the mean number of vertices which remain in the giant connected component after edge deletion. The color bands represent empirical confidince intervals over the simulations. That is, the color bands are the 0.975 and 0.025 upper and lower quantiles of the simulation study.

			Discovery Rate		False Negative	False Positive	
# Edges	Motif	(n,k)	CPM Rate	DPP Rate	$\mathbb{P}$ (Type-II Error)	$\mathbb{P}$ (Type-I Error)	FDR
10		$(3,1)$, $(4;2,1)$, $(5;3,2,1)$	1	1	0	1	0.5
9		$(3,1)$, $(4;2,1)$, $(5;3,2)$	2/3	1	0.1	0.8	0.444
8		$(3,1)$, $(4,2)$, $(5;2,3)$	1/3	1	0.2	0.6	0.375
		$(3,1)$, $(4,2)$, $(5,3)$	1/3	2/3			
7		$(3,1)$, $(4,2)$, $(5,3)$	1/3	2/3	0.3	0.4	0.286
		$(4,2)$, $(5,3)$	0	2/3			
		$(3,1)$, $(4,2)$, $(5,3)$	1/3	2/3			
6		$(5,3)$	0	1/3	0.4	0.2	0.167
		$(5,3)$	0	1/3			
		$(4,2)$, $(5,3)$	0	2/3			
5		$(5,3)$	0	1/3	0.5	0	0

Table 3.6: Robustness to Noise. Note that $(n;k_1,k_2)$ is shorthand for order (n,k_1) and order (n,k_2) plex. It is clear from the center column that the DPP rate of discovery of motifs retains a much higher level in the duration of edge removal. For instance, at 0.4 false negative rate the DPP drastically out performs the CPM

3.7 Application to ECoG Data

Having developed a dynamic community detection method robust to noise, I now illustrate the application of this method to real-world data. To do so, I apply the DPP method to determine the dynamic community structure of a time series of networks inferred from invasive brain voltage data collected from a patient suffering from epileptic seizures. The duration of the time series is approximately 241 seconds with a seizure duration of approximately 74 seconds. Our aim is to better understand the functional connectivity and organization of a brain at seizure onset and offset; ideally, this deeper understanding will provide new insights into the mechanisms of seizure, and novel treatment strategies by Sydney Cash and colleagues at the Massachusetts General Hospital.

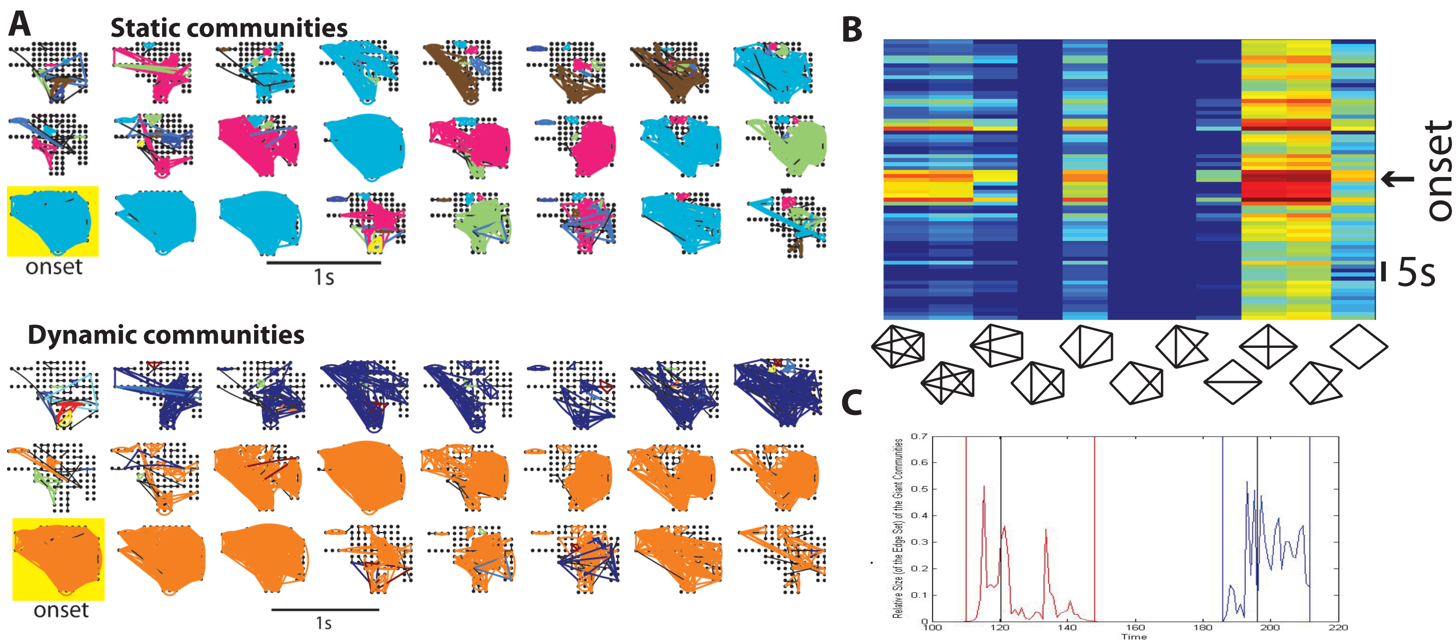

Figure 3.2: Tracking of dynamic communities at seizure onset and termination. (A) Standard clique percolation method (top) and novel time-varying plex percolation method (bottom) applied to ECoG data near seizure onset (yellow box). Time progresses from top-left, to bottom-right; edges with the same color belong to the same community. In the standard method, communities lack consistency across time, while in the proposed method an evolution of dynamic communities (orange edges) precedes seizure onset and continues into seizure. (B) Relative frequency (color) of various sub-graphs (i.e., motifs) representable by the underlying plex defining communities, as a function of time (i.e., y-axis) near seizure onset. The variation in representation of plexes of similar size suggests the limitation of using only a single clique in the standard method. (C) Relative size of the largest dynamic community at seizure onset (orange) and termination (red). Note that each community appears preceding the seizure event.

The static community analysis of the seizure is depicted in Panel A of Figure 3.2 ; note that, in each subfigure, the black circles indicate nodes, and the lines edges. Colored lines represent edges that belong to the same community, and black lines edges that do not belong to any community. Analysis of the dynamic community structure reveals the emergence of a giant dynamic community (orange community) that begins to emerge approximately ten seconds before the clinically declared onset of seizure (yellow shaded box). We depict in Panel C the emergence of the orange dynamic community. The red curve in Panel C represents the relative size of the orange community, that is, the percentage of vertices which belong to the orange community. It is evident that this giant community representing the largest hypersynchronous regions of the brain emerged ten seconds before seizure onset. It is also interesting to note that this giant dynamic connected community occupied, at its peaks, approximately 0.5, 0.4, and 0.3 of the vertices in the network before the community came to the end of its life approximately 25 seconds after seizure onset. We also find that a giant community (colored blue) precedes the onset community only to, in a matter of seconds, dissipate allowing for the orange community to dominate through seizure onet. A simple observation to be made is that large communities emerge and dissipate in time. That the dominant community at seizure onset formed ten seconds prior to the clinical manifestation of the seize suggests intriguing possibilities for seizure detection and treatment. For example, could the seizure have been prevented by the disruption of the formation of this community? Our algorithm affords clinical researchers a powerful to begin exploring these types of questions.

Lastly, consider Panel B in which the horizontal axis represents the various 5 vertex motifs of interest, two of which (numbers 9 and 10, the diamond and the multi-layer perceptron), were those described in Figure 3.1. According to this heat map, and in agreement with [RM02], these motifs arise with significant regularity. In particular, it is important to note that at seizure onset these motifs are prevelant.

Our method has provided a intuitive and interpretable method for tracing dynamic communities in time. My future work and research includes using the discovered communities as meta-data to be analyzed. For example, we ask, do the “tubes” possess a particular topological property such as genus $1, 2, \ldots$ during seizure? The DPP method provides a powerful to investigate the dynamic functional networks of seizure and address these types of questions.

Chapter 4

Percolation Under Noise

4.1 Introduction

The emergence of a giant connected component (GCC) in a random graph was studied at least as early 1960 by the authors of [ER60] who considered a random graph model which is perhaps over-simplified but mathematically tractable. The authors quantified the rate at which the GCC emerges in a random graph or, in other words, the type of phase transition. Phase transitions are frequently observed in nature yet the rate of said transitions are typically not easy to understand. Phase transitions may occur in either a so-called first-order, or discontinuous manner, or in a so-called second-order, or continuous fashion. For a hallmark example, consider the triple-point of the states of water. The transitions from solid to liquid and then from liquid to gas away from the triple-point are second order while, at the triple-point, the transition from solid to gas is first order.

The authors of [DA09] had believed to have shown a random graph mechanism which could produce a first-order phase transition in the so-called "explosive percolation" random graph generation but, in later work, the authors of [RW11] refuted this result and instead were able to show that explosive percolation is indeed continuous and, therefore, second-order.

Existing work in the literature has sought to define a graph-generating model in which a first-order phase transition can be observed and has met with moderate success. In these graph generating models, one begins with a graph G_0 of n isolated vertices and at each time step t adds a single edge between two disconnected vertex pairs according to some rule. At time $t = 0$ the graph is fully disconnected but, of course, after $t = \binom{n}{2}$ time steps, the graph is fully connected. The authors of [ER60], for example, simply connect a randomly chosen disconnected vertex pair. On the other hand, the authors of [DA09] define a rule in which two pairs of vertex pairs are selected and an edge is added between the vertex pair which minimizes the product of the size of the components to which its endpoints would connect. Naturally, there exist a plethora of choices for generation, see [RW11] for examples, but the method for creating percolation has remained generative in the sense that at each time step a single edge is added; never is one deleted and the fact that *exactly* one edge is added is deterministic. Of course, the location of the added edge is random.

I generalize the standard rule of deterministically adding one edge at each time step in two ways. First, it is a natural extension to, instead of adding an edge at time step t with probability 1, add an edge with probability p. I then specify removing an edge with probablity q. Second, in practice we rarely see the "true" network and instead observe a network corrupted by noise. Therefore, we specify that in addition to the underlying birth and death process governed by the parameter p, what is observed is corrupted by the probabilities of types-I and -II error, denoted by α and β, respectively. My enhanced model is more complicated than the "pure birth" model outlined earlier but clearly better represents not only nature but the reality of observation and inference.

I will show that the distinction between first- and second-order percolation, that is the discontinuity or continuity, respectively, of the percolation curves, in the case of birth and death, i.e. $p < 1$, and observation error, i.e. $\alpha, \beta > 0$, is corrupted to the point that one cannot differentiate between the ER [ER60] random graph generative process and the Achlioptas [DA09] product rule random graph generative process.

Recall the inferred networks from ECog data from previous sections. Our birth and death with observation error is an appropriate (although still simplified) model for network graph percolation at seizure onset, see Figure 4.1. The relative size of the giant connected component in Figure 4.1 is represented by the light-gray plot whereas the black trace is a

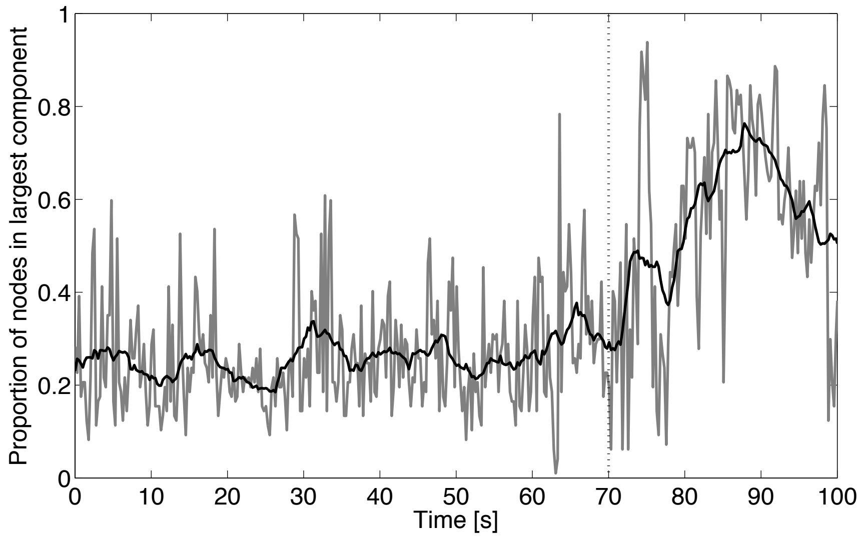

Figure 4.1: Proportion of Nodes in the Giant Connected Component At Seizure Onset

smoothed version. At time point 70 (the vertical dashed line) seizure has been clinically determined to begin for the patient. It is evident that in the time interval of roughly seizure onset plus-or-minus twenty seconds there exists a nearly four-fold increase in the relative size of the giant component.

Two key notions to keep in mind while viewing Figure 4.1 are, first, that the network graph characteristic (see Section 2) of relative size of the giant connected component is computed from noisy voltage readings in the ECoG and, thus, the error is propagated to the graph characteristic. Second, behind the noisy model is the truth, which we believe to be regions of the brain that are functionally coupled and decoupled. This observation motivates our use of a hidden Markov model (HMM) for our percolation experiments on random graphs.

4.2 Methods

I begin by defining a graph-valued Markov chain

$$G_t = \{(V, E_t) : t = 0, 1, \ldots\} \tag{4.1}$$

with an initial state G_0 of $|V| = n$ isolated vertices, i.e. $|E_0| = 0$. Now, define the sequence $\{Y_t : t = 0, 1, \ldots\}$ of binary random variables according to the probability transition matrix in Table 4.1 for some $p, q \in (0, 1)$. Note that, in particular, if $p + q = 1$ then the sequence

	$Y_{t+1} = 0$	$Y_t = 1$
$Y_t = 0$	$1 - p$	p
$Y_t = 1$	q	$1 - q$

Table 4.1: Probabilty Transition Matrix for $\{Y_t\}$

$\{Y_t\}$ is not only Markovian but independent. Alternatively, we need not be restricted to a constant p and q in the sense that $p = p(t)$ and $q = q(t)$ may, in general, be time inhomogeneous.

Now, independent of the sequence $\{Y_t\}$, let $X_t(v_i, v_j)$ denote the edge status, a binary value. That is,

$$X_t(v_i, v_j) = \begin{cases} 1 & \text{if an edge exists between } v_i \text{ and } v_j \\ 0 & \text{if no edge exists between } v_i \text{ and } v_j. \end{cases} \tag{4.2}$$

It is important to note that the sequence of collections of random variables

$$\left\{X_t(v_i, v_j) : t = 0, 1, \ldots; (v_i, v_j) \in V^{(2)}\right\}$$

with $V^{(2)}$ being the set of all vertex pairs has dependence across both space and time. In space, for example, suppose that $Y_t = 1$ then precisely one edge must be added which means that if both $X_t(v_i, v_j) = 0$ and $X_t(v_h, v_k) = 0$ then at most one will transition to 1. In time, for example, if $Y_t = 1$ and $X_t(v_i, v_j) = 1$ then, since we are only considering simple graphs, $X_{t+1}(v_i, v_j) = 1$. The edge status $X_t(v_i, v_j)$ is our latent edge status of the HMM.

4.3 Network Evolution Processes

In this section I will discuss two mechanisms, describing in both cases what exists in the literature and my contributions.

4.3.1 Erdos-Renyi (ER) Percolation

The mechanism which the authors of [ER60] utilized sets $p = 1$ and, for each time t, choose uniformly at random among the vertex pairs currently disconnected, i.e. the vertex pair (v_i, v_j) for which $X_t(v_i, v_j) = 0$, to connect with an edge. Conditional on the sequence $\{Y_t\}$, the probability transition matrix of the sequence of the collection of edge statuses $\{X_t(v_i, v_j)\}$ is as in Table 4.2. In contrast, I allow, as stated earlier, for $0 < p < 1$ which

	$X_{t+1}(v_i, v_j) = 0$	$X_{t+1}(v_i, v_j) = 1$
$X_t(v_i, v_j) = 0$	$\frac{\binom{n}{2} - \lvert E_t \rvert - 1}{\binom{n}{2} - \lvert E_t \rvert}$	$\frac{1}{\binom{n}{2} - \lvert E_t \rvert}$
$X_t(v_i, v_j) = 1$	$\frac{1}{\lvert E_t \rvert}$	$\frac{\lvert E_t \rvert - 1}{\lvert E_t \rvert}$

Table 4.2: Probability Transition Matrix for $\{X_t(v_i, v_j)\}$ Conditional on $\{Y_t\}$.

instead gives rise to the probability transition matrix in Table 4.3. Since it is allowed for

	$X_{t+1}(v_i, v_j) = 0$	$X_{t+1}(v_i, v_j) = 1$
$X_t(v_i, v_j) = 0$	$\frac{\binom{n}{2} - \lvert E_t \rvert - 1\{Y_t = 1\}}{\binom{n}{2} - \lvert E_t \rvert}$	$\frac{1\{Y_t = 1\}}{\binom{n}{2} - \lvert E_t \rvert}$
$X_t(v_i, v_j) = 1$	$\frac{1\{Y_t = 0\}}{\lvert E_t \rvert}$	$\frac{\lvert E_t \rvert - 1\{Y_t = 0\}}{\lvert E_t \rvert}$

Table 4.3: Probability Transition Matrix for $\{X_t(v_i, v_j)\}$ Conditional on $\{Y_t\}$.

$p < 1$, this means that an edge will be removed from G_t with probability $1 - p$. In order to be congruent with the process of adding an edge, if $Y_t = 0$ and an edge must be deleted, the edge chosen to be removed is selected uniformly at random. In this way, it is seen that it is analagous to adding an edge to the complement graph of G_t.

4.3.2 Achlioptas Product Rule (PR) Percolation

The probability transition matrix of $X_t(v_i, v_j)$ is not as straight forward in the case of PR percolation as it was for the ER for reasons that will imminently become clear. The authors of [DA09], as was in the case of ER, set $p = 1$ and choose a disconnected vertex pair to be connected at each time t but in a clever way so as to slow the rate of percolation. The authors choose two vertex pairs (v_i, v_j) and (v_k, v_h), both of which at time t are disconnected, uniformly at random. Denote with C_s for $s = i, j, k, h$ the connected components to which v_s belong. The authors choose the vertex pair to be connected by the following rule. Connect (v_i, v_j) with an edge if $|C_i|\,|C_j| < |C_k|\,|C_h|$ and otherwise connect (v_k, v_h). The term "product rule" should now be clear.

Remark 4.1. *The intuition behind choosing such a rule is to delay the onset of percolation. Consider again the standard ER model in which a disconnected vertex pair is chosen uniformly at random to be connected and suppose that, in the following simple example,* $|V| = 6$ *and* G_t *is composed of three disconnected cliques* K_1*,* K_2*, and* K_3*. The probability of merging* K_1 *with* K_2 *is* $\frac{2}{11}$*,* K_1 *with* K_3 *is* $\frac{3}{11}$*, and* K_2 *with* K_3 *is* $\frac{6}{11}$*. It is intuitive that the larger "bodies" or components are attracted by a multiplying, i.e. a quasi-squared or gravity like force. Intuitively, by choosing to use the PR mechanism, the authors of [DA09] gave a repulsive force to the larger components and a stronger gravitational force to the smaller components. This intuition gave rise to my choice of a statistic, to be discussed later, for discriminating between the ER and PR models in the presence of noise.*

Of course, as was the case with my modified ER model, I choose $0 < p < 1$ for the probability of an edge to be added into the network but, in the same way as in the PR, I select the placement of the edge according to the minimization of the product rule. Conversely, if an edge is to be deleted, as will occur with probability $1 - p$, then I choose to delete an edge in a manner which is analogous to the process of adding and edge. That is, I randomly select two vertex pairs (v_i, v_j) and (v_k, v_h), both of which at time t are connected, uniformly at random and automatically disconnect them to produce $G_t^{\dagger}$. At this point two edges have been removed but only one removal is permitted so I replace the one which minimizes the product rule for $G_t^{\dagger}$. In this way, the deletion of an edge has the effect of fracturing the larger connected components.

4.3.3 Percolation Under Network Evolution

Here we introduce the notion of a percolation curve. Consider Figure 4.2 and, in particular, the main panes of both plots. Moving from left to right in the main panes one will notice,

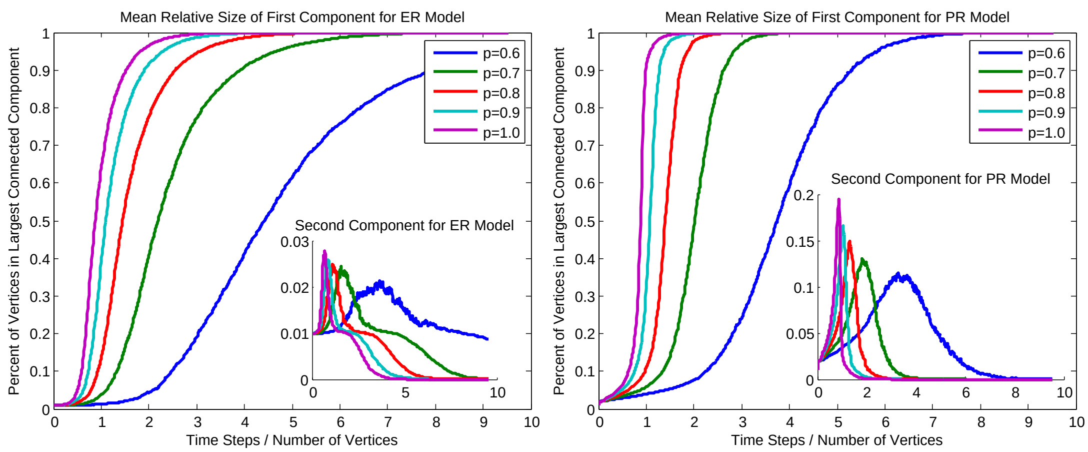

Figure 4.2: Mean Percolation Curve with Birth and Death for ER (left) and PR (right) Models. In the plots above $|V| = 100$.

via the legend, that the birth rate is decreased across the range $\{1, 0.9, 0.8, 0.7, 0.6\}$. The plotted curves represent a mean over 5000 simulations of the percentage of vertices which belong to a giant connected component versus t/n. The effect of choosing $p < 1$ becomes immediately clear in the sense that percolation is slowed and, evidently, the $p = 1$ curve is scaled rightward as p decreases.

Returning to Remark 4.1, recall that the ER mechanism has the effect of gravitationally attracting larger components to become connected while the PR mechanism has the effect of “gravitationally” repelling larger components from becoming connected. Note the insets of Figure 4.2 and note the plotted lines which represent the relative size of the second largest connected component (SGCC). We see that the relative size of the SGCC is, independently of p, an order of magnitude larger for the PR model compared with the ER model. The intuition is that moderately large connected components emerge but they repel each other while they attract the smaller components.

4.4 Noisy Observations

In this section I describe a time-independent error process which corrupts the signal, i.e. the percolation curves. Consider Figure 4.3 and the curves in the main pane. The bold

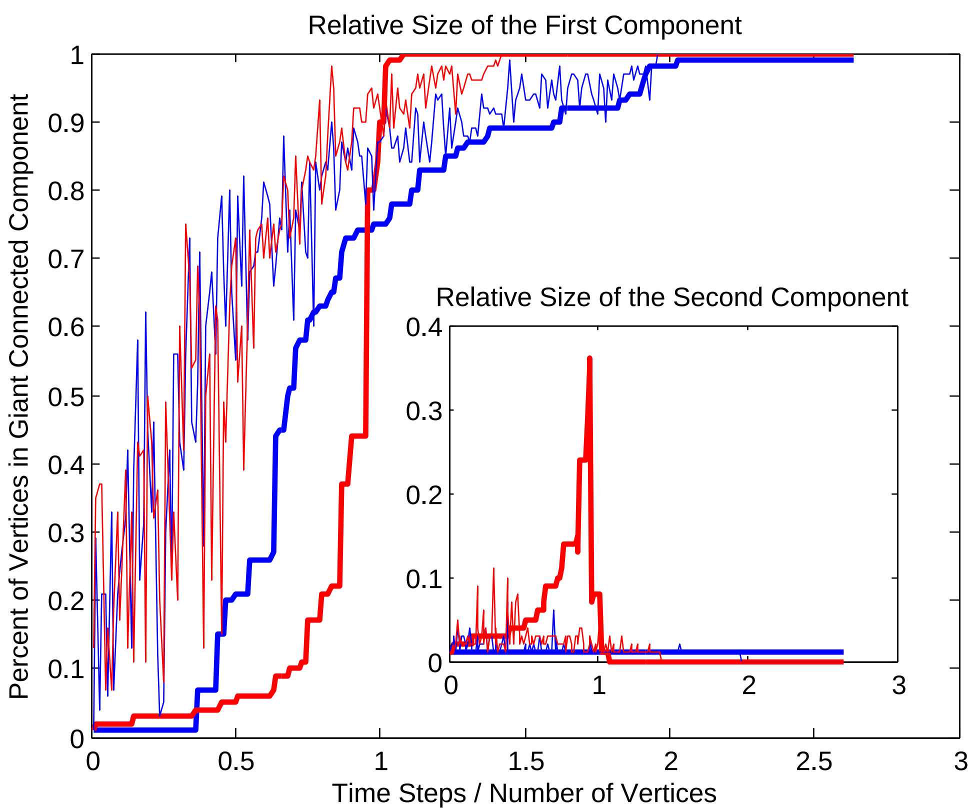

Figure 4.3: Sample Paths of Percolation Curves

curves in the main pane are a **single** sample path of the ER and PR percolation curves, the steepest of which represents the PR mechanism with $p = 1$ and no noise. What was displayed in Figure 4.2 was an average of 5000 such sample paths. Purely based upon visual inspection, the steepness of the curve is enough to decipher a difference between the ER and PR mechanisms. I sought to determine the ability one may possess in deciphering the two curves when either $\alpha > 0$ or $\beta > 0$ or both. According to the thin lines in Figure 4.3 for a single sample path, there is potentially no hope to decipher between the ER and PR percolation curves. The thin lines in the main pane were generated with $p = 1$, $\alpha = 0.0125$, and $\beta = 0.01$ on a graph with $|V| = 100$.

In practice, one observes not many sample paths, but one sample path. The effect of noise via simulations of the ER and PR mechanisms in Figure 4.3 further illustrate the importance of the single sample issue and real data example in Figure 4.1. Simply put, we began by trying to answer the question: is this sample path, which has been corrupted by noisy observations, generated by an ER mechanism or some process with accelerated percolation such as the PR model, or explosive percolation?

4.5 Detecting Percolation

Explosive percolation is expected to produce a sharper phase transition than a typical second-order phase transition model such as the ER. As stated earlier, it is quite clear that when $(p, \alpha, \beta) = (1, 0, 0)$ visual inspection is sufficient to distinguish the two mechanisms based solely on the percolation curve of the GCC. Translating this observation into a statistic, I define a statistic called the "quantile distance" QD which, for values $0 \leq z_1 < z_2 \leq 1$, is defined to be

$$QD\left(z_1, z_2\right) = Q\left(z_2\right) - Q\left(z_1\right)$$

where

$$Q\left(z\right) = \min_{t \in \{0,1,\ldots\}} \left\{F\left(t\right) \geq z\right\}$$

and where F is a percolation curve of a GCC. Note that in the case that $(p, \alpha, \beta) = (1, 0, 0)$ then the percolation curves F are monotonically increasing so F is a proper "cumulative density function" and, therefore, the function Q can be thought of as a proper "quantile function". Of course as p varies away from one and/or α and β vary away from zero then the function F is no longer guaranteed to be monotonic. However, I choose the measure QD to be as such since the particular case of pure birth and no noise will thus provide a baseline. For the choice of $z_1 = 0.05$ and $z_2 = 0.75$ I have simulated 10000 percolation curves on a network with $|V| = 100$ vertices for both the ER and PR models with $(p, \alpha, \beta) = (1, 0, 0)$ and generated the empirical probability densities, i.e. histograms, for our QD statistic and plotted them in the main pane of Figure 4.4. It is no surprise

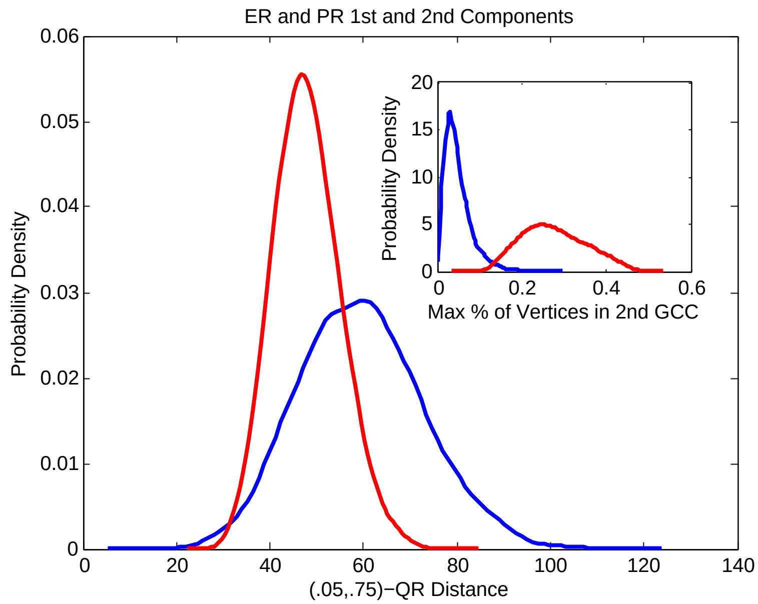

Figure 4.4: Empirical Probability Density Functions. In the main figure we see the probability densities of the 1st component QR distance with the ER in blue and the PR in red. Similarly, in the inset we see the ER in blue and the PR in red but instead for the size of the second component statistic.

that the QD statistic supports a large amount of probability mass on lower integer values due to the sharp increase in the percolation curves of the GCC.

Formally, our goal is to test the hypothesis

$$H_0 : \text{Percolation is ER} \quad \text{versus} \quad H_1 : \text{Percolation is not ER}. \tag{4.3}$$

According to standard wisdom, a proper test statistic is potentially $QD\,(z_1, z_2)$ for $(z_1, z_2) = (0.05, 0.75)$. However, consider the heavy curves in the ROC plot in Figure 4.5. Clearly, although there is power for the test in Equation (4.3), one can do much better by using the

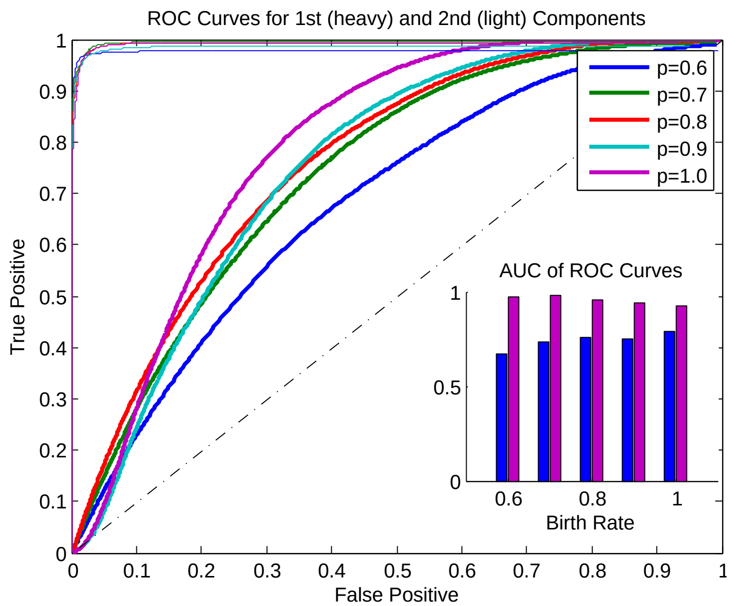

Figure 4.5: ROC and AUC for Birth and Death $p = \{0.6, \ldots, 1.0\}$

maximum achieved value of the SGCC. It is clear from the inset plot of Figure 4.4 that the densities for the SGCC have support that are nearly disjoint which, of course, should produce much greater power for a test which incorporates the test statistic $MSGCC$, defined to be the maximum size of the SGC.

We vary α and β and see results which further illustrate our understanding that noisy generative models require information beyond the size of the GCC. Recall that in the previous plots p, i.e. the birth rate, was varied. However, consider Figure 4.6 and, in particular, the right hand pane. It becomes evident that there is no possibility to achieve much statistical power using the QD statistic alone. However, as demonstrated in Figure 4.7, the relative size of the second component statistic provides an improvement in this scenario. Over a range of α and β, i.e. Type-I and -II, errors we can clearly see that a hypothesis test based upon the second component can acheive higher power, specifically for lower levels of Type-I error.

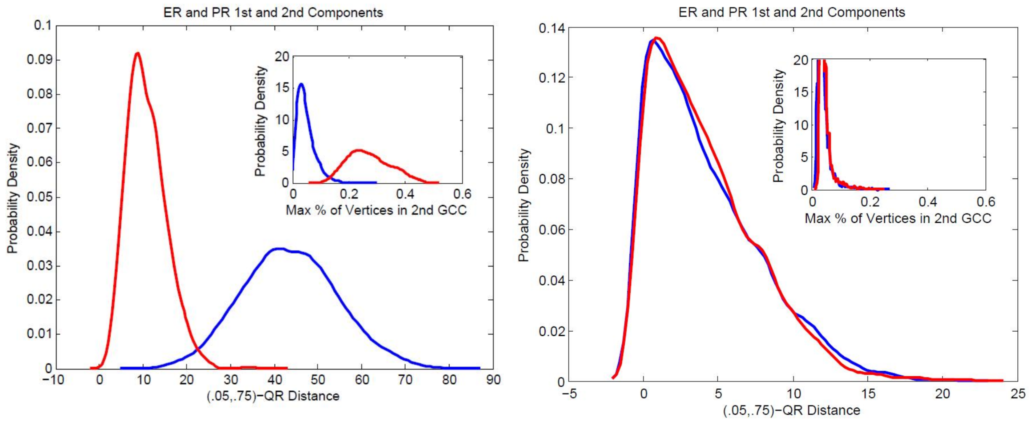

Figure 4.6: Left: The probability density of the QD statistic (main) and $MSGCC$ statistic (inset) for $(p, \alpha, \beta) = (1, 0, 0)$. Right: The probability density of the QD statistic (main) and $MSGCC$ statistic (inset) for $(p, \alpha, \beta) = (1, .0125, .075)$.

4.6 Discussion

In the above we discussed extension to existing models of percolation by allowing for both edge deletion and noisy observations. These modeling extensions have been articulated within a hidden Markov model graph process. Moreover, we have compared different summary statistics for distinguishing between the ER and PR models, i.e. standard and explosive percolation, respectively. It is important to note that when considering real data we rarely observe fully connected networks and, in particular, at seizure onset we observe the relative size of a giant connected component to be near our choice of quantiles. We demonstrated that our second-component statistic is more powerful in this real-world setting. Ultimately, we will apply our test to real data in an effort to determine the rate of seizure onset.

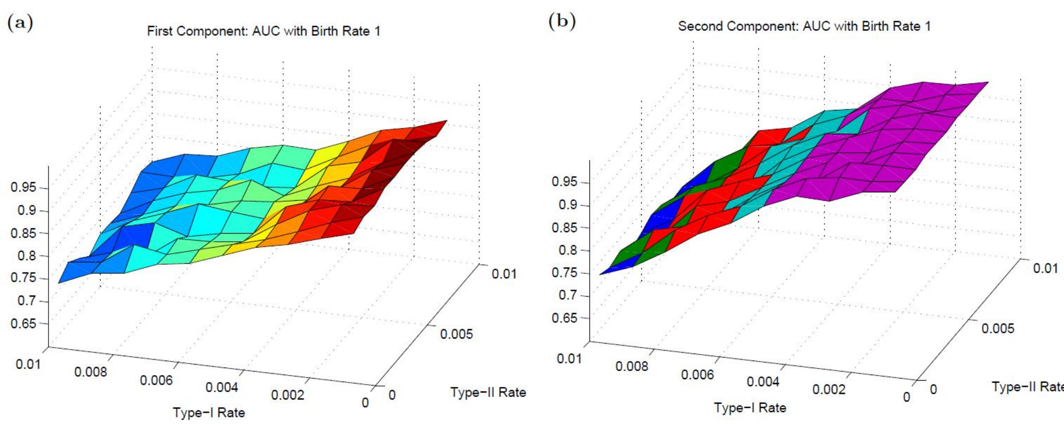

Figure 4.7: AUC over a range of error (both Type-I and -II) and birth rate 1

Chapter 5

Conclusion

We have seen that the error incurred from the propagation of noisy observations to network network inference and, furthermore, to network characterization must not be simply ignored. I made three major contributions in the direction of understanding and handling the error incurred from network inference. My first contribution was probabilistic in nature and provided a central limit theorem type result for network characteristics. My second project was more algorithmic and provided a means for the detection of dynamic communities in a time series of networks. Finally, my third project was a numerical study of the effect of noise on network percolation which completes the story that we have written here: that networks with edge error require special treatment. The results we have provided are vital for researchers and, furthermore, open the door for further results in this direction.

A statistics 101 student will (hopefully) tell you that $\bar{x}$ is not the population mean μ and will instead report a confidence interval for μ. I have developed a "central-limit" type theorem for the network density which, in the future, may give rise to the ability for one to construct a confidence interval for the network density. Remaining future work in this area is the development of methods to construct said confidence intervals and to further generalize the class of network characteristics for which we may approximate the probability distribution of an estimator thereof.

To better understand the origin of an epileptic seizure could profoundly impact treatment strategies. Our dynamic community method provides a possible route toward tracing backward in time the origin of a seizure episode to its source. Future work in this direction is to obtain meta-data, in the sense of community structure, for analysis. The topology of the communities pre-, during, and post-seizure may exhibit distinct behaviors of splitting and merging. Such analysis has the potential for a break-through in the understanding of seizure organization.

The impact of death and noise on the emergence of a giant connected component in standard models has not been studied. We made advances in the direction of understanding how noise can distort the assessment of the first connected component. We propose a statistic that has the potential to provide significantly more power and goes beyond observing the first component. Future work includes making the non-parametric test that we propose to include more theoretical results; at the moment we have primarily numerical and non-parametric results but we have laid the framework for the theoretical solution to a significant problem. Ultimately, to apply our techniques to data is our primary goal so that the rate of seizure onset can be more fully understood.

Chapter 6

Appendix

In this section I provide the proofs and technical details omitted in the main text of my thesis.

6.1 Proofs for Section 2

In this section we will present proofs that coorespond with previously stated results. The goal of this section is to verify all claims.

6.1.1 The Stein Operator

We begin with the operator which was first introduced in Theorem 2.3

$$\mathcal{A}\left[f\left(k\right)\right] = \lambda_1 f\left(k+1\right) - kf\left(k\right) - \lambda_2 f\left(k-1\right)$$

for any bounded $f : \mathbb{Z} \mapsto \mathbb{R}$ with the intent of showing that the random variable $W \sim \text{Skellam}\left(\lambda_1, \lambda_2\right)$ if and only if $\mathbb{E}\mathcal{A}\left[f\left(W\right)\right] = 0$ for any bounded function $f : \mathbb{Z} \mapsto \mathbb{R}$.

We begin with the necessity direction and the computation of

$$\begin{aligned}
\mathbb{E}\mathcal{A}\left[f\left(W\right)\right] &= \mathbb{E}\left[\lambda_1 f\left(W+1\right) - Wf\left(W\right) - \lambda_2 f\left(W-1\right)\right] \\
&\propto \sum_{k=-\infty}^{\infty} \left[\lambda_1 f\left(k+1\right) - kf\left(k\right) - \lambda_2 f\left(k-1\right)\right] \left(\sqrt{\frac{\lambda_1}{\lambda_2}}\right)^k I_k
\end{aligned}$$

where $\propto$ is to be read as "proportional to". By a result of [AS72] we have that

$$I_{k-1} - I_{k+1} = \frac{k}{\sqrt{\lambda_1 \lambda_2}} I_k$$

or, in other words,

$$\sqrt{\lambda_1 \lambda_2} \frac{I_{k-1}}{I_k} - \sqrt{\lambda_1 \lambda_2} \frac{I_{k+1}}{I_k} = k. \tag{6.1}$$

This means that

$$\begin{aligned}
\mathbb{E}\mathcal{A}\left[f(W) 1\{W \leq n\}\right] &\propto \sum_{k=-\infty}^{n} \left[\sqrt{\frac{\lambda_1}{\lambda_2}} f(k+1) - \frac{I_{k-1}}{I_k} f(k) + \right. \\
&\quad \left. \frac{I_{k+1}}{I_k} f(k) - \sqrt{\frac{\lambda_2}{\lambda_1}} f(k-1)\right] \left(\sqrt{\frac{\lambda_1}{\lambda_2}}\right)^k I_k \\
&= \sum_{k=-\infty}^{n} \left[\left(\sqrt{\frac{\lambda_1}{\lambda_2}}\right)^{k+1} I_k f(k+1) - \left(\sqrt{\frac{\lambda_1}{\lambda_2}}\right)^k I_{k-1} f(k)\right] \\
&\quad + \sum_{k=-\infty}^{n} \left[\left(\sqrt{\frac{\lambda_1}{\lambda_2}}\right)^k I_{k+1} f(k) - \left(\sqrt{\frac{\lambda_1}{\lambda_2}}\right)^{k-1} I_k f(k-1)\right] \\
&= \left(\sqrt{\frac{\lambda_1}{\lambda_2}}\right)^{n+1} I_n f(n+1) + \left(\sqrt{\frac{\lambda_1}{\lambda_2}}\right)^n I_{n+1} f(n).
\end{aligned}$$

Now, since f is bounded then

$$\lim_{n\to\infty} \left\{\left(\sqrt{\frac{\lambda_1}{\lambda_2}}\right)^{n+1} I_n f(n+1) + \left(\sqrt{\frac{\lambda_1}{\lambda_2}}\right)^n I_{n+1} f(n)\right\} = 0$$

which proves the claim.

To prove sufficiency, we begin with $\mathbb{E}\mathcal{A}[f(W)] = 0$ and suppose that $f_k(j) = 1\{j = k\}$ for some $j \in \mathbb{Z}$ in which case

$$\lambda_1 p(k-1) - kp(k) - \lambda_2 p(k+1) = 0$$

where $p(k) = \mathbb{P}(W = k)$. An ansatz of

$$p(k) = \left(\sqrt{\tfrac{\lambda_1}{\lambda_2}}\right)^k I_k\left(2\sqrt{\lambda_1\lambda_2}\right) \quad \text{and} \quad p(k) = \left(\sqrt{\tfrac{\lambda_1}{\lambda_2}}\right)^k K_k\left(2\sqrt{\lambda_1\lambda_2}\right)$$

solves this non-linear difference equation. From the theory of differential (difference) equations [PBH06], we know that the solution to the above is of the form

$$p(k) = C_1 S(k) + C_2 T(k)$$

for some constants $C_1, C_2 \in \mathbb{R}$ and linearly independent functions $S, T : \mathbb{Z} \mapsto \mathbb{R}$. It is easy to show that

$$S(k) = \left(\sqrt{\frac{\lambda_1}{\lambda_2}}\right)^k I_k$$

and

$$T(k) = \left(\sqrt{\frac{\lambda_1}{\lambda_2}}\right)^k K_k$$

where $I_k = I_k\left(2\sqrt{\lambda_1\lambda_2}\right)$ and $K_k = K_k\left(2\sqrt{\lambda_1\lambda_2}\right)$ are the modified Bessel functions of the first and second kinds. Now, to determine the constants C_1 and C_2 we appeal to the fact that $\sum_{k=-\infty}^{\infty} p(k) = 1$. Since $I_k, K_k > 0$ for all $k \in \mathbb{Z}$ and $\sum_{k=-\infty}^{\infty} K_k = \infty$ it must be that $C_2 = 0$. Now, consider the generating function

$$e^{\frac{z}{2}(t+1/t)} = \sum_{k=-\infty}^{\infty} t^k I_k(z)$$

which means that

$$\begin{aligned} C_1 &= \frac{1}{\sum_{k=-\infty}^{\infty} \left(\sqrt{\frac{\lambda_1}{\lambda_2}}\right)^k I_k\left(2\sqrt{\lambda_1\lambda_2}\right)} \\ &= \frac{1}{e^{\sqrt{\lambda_1\lambda_2}\left(\sqrt{\frac{\lambda_1}{\lambda_2}}+\sqrt{\frac{\lambda_2}{\lambda_1}}\right)}} \\ &= e^{-(\lambda_1+\lambda_2)} \end{aligned}$$

so that

$$p\left(k\right)=e^{-\left(\lambda_1+\lambda_2\right)}\left(\sqrt{\frac{\lambda_1}{\lambda_2}}\right)^k I_k\left(2\sqrt{\lambda_1\lambda_2}\right)$$

so that $W\sim\text{Skellam}\left(\lambda_1,\lambda_2\right)$.

6.1.2 The Stein Solution

Since the left-hand-side may be easier to bound than the right-hand-side of the following equation, consider the solution to

$$\lambda_1 f\left(k+1\right)-kf\left(k\right)-\lambda_2 f\left(k-1\right)=g\left(k\right),$$

for some function $g:\mathbb{Z}\mapsto\mathbb{R}$, with boundary condition

$$\lim_{k\to-\infty}\left(\sqrt{\frac{\lambda_1}{\lambda_2}}\right)^k I_k f\left(k\right)=0. \tag{6.2}$$

The goal here is to bound the right-hand-side but we will instead bound the left-hand-side. The method we will use the similar to that of the Poisson approximation to the binomial distribution. That is, we will compute a Lipschitz constant for the function f on the left-hand-side, define a coupling of random variables, and proceed to bound the left-hand-side.

First, multiply both sides by $\left(\sqrt{\frac{\lambda_1}{\lambda_2}}\right)^k I_k$ and sum over $k\in\{-\infty,\ldots,n\}$ to get

$$\left(\sqrt{\frac{\lambda_1}{\lambda_2}}\right)^{n+1} I_n f\left(n+1\right)+\left(\sqrt{\frac{\lambda_1}{\lambda_2}}\right)^n I_{n+1} f\left(n\right)=\frac{1}{\sqrt{\lambda_1\lambda_2}}\sum_{k=-\infty}^{n}\left(\sqrt{\frac{\lambda_1}{\lambda_2}}\right)^k I_k g\left(k\right).$$

Note that in order for the integrating factor of $\left(\sqrt{\frac{\lambda_1}{\lambda_2}}\right)^k I_k$ to create a telescoping series one must appeal to the recurrence in Equation (6.1) and the boundary condition of Equation (6.2). Now multiply both sides by $(-1)^{n+1}\frac{1}{I_n I_{n+1}}$ and sum over $n\in\{c,c+1,\ldots,m\}$ for $m>c$ and over $n\in\{m,m+1,\ldots,c-1\}$ for some initial condition $c\in\mathbb{Z}$ and $f\left(c\right)\in\mathbb{R}$ to get

$$f(m)=\begin{cases}(-1)^m\left(\sqrt{\frac{\lambda_2}{\lambda_1}}\right)^m I_m\left[(-1)^c\left(\sqrt{\frac{\lambda_1}{\lambda_2}}\right)^c\frac{1}{I_c}f(c)\right. & \\ \left.+\frac{1}{\sqrt{\lambda_1\lambda_2}}\sum_{n=c}^{m-1}\frac{(-1)^{n+1}}{I_nI_{n+1}}\sum_{k=-\infty}^{n}\left(\sqrt{\frac{\lambda_1}{\lambda_2}}\right)^k I_k g(k)\right] & \text{if } m>c\\ (-1)^m\left(\sqrt{\frac{\lambda_2}{\lambda_1}}\right)^m I_m\left[(-1)^c\left(\sqrt{\frac{\lambda_1}{\lambda_2}}\right)^c\frac{1}{I_c}f(c)\right. & \\ \left.-\frac{1}{\sqrt{\lambda_1\lambda_2}}\sum_{n=m}^{c-1}\frac{(-1)^{n+1}}{I_nI_{n+1}}\sum_{k=-\infty}^{n}\left(\sqrt{\frac{\lambda_1}{\lambda_2}}\right)^k I_k g(k)\right] & \text{if } m<c.\end{cases}$$

Note that if

$$g(k)=1\{k\le x\}-\mathbb{P}(W\le x)$$

then

$$\sum_{k=-\infty}^{n}\left(\sqrt{\frac{\lambda_1}{\lambda_2}}\right)^k I_k g(k)=\begin{cases}e^{\lambda_1+\lambda_2}\mathbb{P}(W\le n)\,\mathbb{P}(W>x) & \text{if } n\le x\\ e^{\lambda_1+\lambda_2}\mathbb{P}(W\le x)\,\mathbb{P}(W>n) & \text{if } n\ge x\end{cases}$$

since, for example if $n\le x$

$$\begin{aligned}\sum_{k=-\infty}^{n}\left(\sqrt{\frac{\lambda_1}{\lambda_2}}\right)^k I_k g(k) &= e^{\lambda_1+\lambda_2}\sum_{k=-\infty}^{n}\mathbb{P}(W=k)\,g(k)\\ &= e^{\lambda_1+\lambda_2}\sum_{k=-\infty}^{n}\mathbb{P}(W=k)\left[1\{k\le x\}-\mathbb{P}(W\le x)\right]\\ &= e^{\lambda_1+\lambda_2}\left[\mathbb{P}(W\le\min\{n,x\})-\mathbb{P}(W\le x)\,\mathbb{P}(W\le n)\right]\\ &= e^{\lambda_1+\lambda_2}\mathbb{P}(W\le n)\left[1-\mathbb{P}(W\le x)\right]\\ &= e^{\lambda_1+\lambda_2}\mathbb{P}(W\le n)\,\mathbb{P}(W>x).\end{aligned}$$

and the case that $n\ge x$ is similar. This means that

$$f(m)=\begin{cases}(-1)^m\left(\sqrt{\frac{\lambda_2}{\lambda_1}}\right)^m I_m\left[(-1)^c\left(\sqrt{\frac{\lambda_1}{\lambda_2}}\right)^c\frac{1}{I_c}f(c)\right. & \\ \left.+\frac{e^{\lambda_1+\lambda_2}}{\sqrt{\lambda_1\lambda_2}}\sum_{n=c}^{m-1}\frac{(-1)^{n+1}}{I_nI_{n+1}}\mathbb{P}(W\le\min\{n,x\})\,\mathbb{P}(W>\max\{n,x\})\right] & \text{if } m>c\\ (-1)^m\left(\sqrt{\frac{\lambda_2}{\lambda_1}}\right)^m I_m\left[(-1)^c\left(\sqrt{\frac{\lambda_1}{\lambda_2}}\right)^c\frac{1}{I_c}f(c)\right. & \\ \left.-\frac{e^{\lambda_1+\lambda_2}}{\sqrt{\lambda_1\lambda_2}}\sum_{n=m}^{c-1}\frac{(-1)^{n+1}}{I_nI_{n+1}}\mathbb{P}(W\le\min\{n,x\})\,\mathbb{P}(W>\max\{n,x\})\right] & \text{if } m<c.\end{cases}$$

The choice for initial condition $(c, f(c))$ is naturally open but we choose $c = \lambda_2 - \lambda_1$ and

$$
\begin{aligned}
f(c) = & \left[\frac{1}{\lambda_1}\left(\sqrt{\frac{\lambda_2}{\lambda_1}}\right)^c \frac{1}{I_c}\sum_{k=-\infty}^{c}\left(\sqrt{\frac{\lambda_1}{\lambda_2}}\right)^k I_k g(k)\right. \\
& \left. + \frac{1}{\lambda_2}\left(\sqrt{\frac{\lambda_2}{\lambda_1}}\right)^c \frac{1}{I_c}\sum_{k=-\infty}^{c-1}\left(\sqrt{\frac{\lambda_1}{\lambda_2}}\right)^k I_k g(k)\right] \\
& \times \left[\left(\sqrt{\frac{\lambda_2}{\lambda_1}}\frac{I_{c+1}}{I_c} + 1\right) + \left(1 + \sqrt{\frac{\lambda 1}{\lambda_2}}\frac{I_{m-1}}{I_m}\right)\right]^{-1}.
\end{aligned}
$$

Now, in the special case that $\lambda_1 = \lambda_2 = \lambda$ then

$$
\begin{aligned}
f(m) = (-1)^m I_m & \left[\frac{1}{I_0} f(0)\right. \\
& + I(m > 0)\frac{e^{2\lambda}}{\lambda}\sum_{n=0}^{m-1}\frac{(-1)^{n+1}}{I_n I_{n+1}}\mathbb{P}(W \leq \min\{n, x\})\,\mathbb{P}(W > \max\{n, x\}) \\
& \left. - I(m < 0)\frac{e^{2\lambda}}{\lambda}\sum_{n=m}^{c-1}\frac{(-1)^{n+1}}{I_n I_{n+1}}\mathbb{P}(W \leq \min\{n, x\})\,\mathbb{P}(W > \max\{n, x\})\right]
\end{aligned}
\tag{6.3}
$$

and $c = 0$, and

$$
\begin{aligned}
f(0) &= \frac{e^{2\lambda}}{\lambda}\frac{1}{I_0}\left[\mathbb{P}(W \leq \min\{0, x\})\,\mathbb{P}(W > \max\{0, x\})\right. \\
& \quad \left. + \mathbb{P}(W \leq \min\{-1, x\})\,\mathbb{P}(W > \max\{-1, x\})\right] \\
& \quad \times \left[\left(1 + \frac{I_1}{I_0}\right) + \left(\frac{I_{-1} I_0}{+} 1\right)\right] \\
&= \frac{e^{2\lambda}}{2\lambda}\frac{1}{I_0 + I_1}\left[\mathbb{P}(W \leq \min\{0, x\})\,\mathbb{P}(W > \max\{0, x\})\right. \\
& \quad \left. + \mathbb{P}(W \leq \min\{-1, x\})\,\mathbb{P}(W > \max\{-1, x\})\right].
\end{aligned}
$$

If $x \leq -1$ then

$$
\begin{aligned}
f(0) &= \frac{e^{2\lambda}}{2\lambda}\frac{\mathbb{P}(W \leq x)}{I_0 + I_1}\left[\mathbb{P}(W > 0) + \mathbb{P}(W > -1)\right] \\
&= \frac{e^{2\lambda}}{2\lambda}\frac{\mathbb{P}(W \leq x)}{I_0 + I_1}
\end{aligned}
$$

and if $x \geq 0$ then

$$\begin{aligned} f(0) &= \frac{e^{2\lambda}}{2\lambda}\frac{\mathbb{P}(W > x)}{I_0 + I_1}\left[\mathbb{P}(W \leq 0) + \mathbb{P}(W \leq -1)\right] \\ &= \frac{e^{2\lambda}}{2\lambda}\frac{\mathbb{P}(W > x)}{I_0 + I_1}. \end{aligned}$$

Note that if $m > 0$ then

$$\begin{aligned} f(m+1) &= (-1)^{m+1} I_{m+1}\Bigg[\frac{1}{I_0} f(0) \\ &\quad + \frac{e^{2\lambda}}{\lambda}\sum_{n=0}^{m}\frac{(-1)^{n+1}}{I_n I_{n+1}}\mathbb{P}(W \leq \min\{n, x\})\,\mathbb{P}(W > \max\{n, x\})\Bigg] \\ &= (-1)^{m+1} I_{m+1}\Bigg[\frac{1}{I_0} f(0) \\ &\quad - \frac{e^{2\lambda}}{\lambda}\frac{1}{I_0 I_1}\mathbb{P}(W \leq \min\{0, x\})\,\mathbb{P}(W > \max\{0, x\}) \\ &\quad + \frac{e^{2\lambda}}{\lambda}\sum_{n=1}^{m}\frac{(-1)^{n+1}}{I_n I_{n+1}}\mathbb{P}(W \leq \min\{n, x\})\,\mathbb{P}(W > \max\{n, x\})\Bigg] \\ &= (-1)^{m+1} I_{m+1}\Bigg[\frac{1}{I_0} f(0) \\ &\quad - \frac{e^{2\lambda}}{\lambda}\frac{1}{I_0 I_1}\mathbb{P}(W \leq \min\{0, x\})\,\mathbb{P}(W > \max\{0, x\}) \\ &\quad - \frac{e^{2\lambda}}{\lambda}\sum_{n=0}^{m-1}\frac{(-1)^{n+1}}{I_{n+1} I_{n+2}}\mathbb{P}(W \leq \min\{n+1, x\})\,\mathbb{P}(W > \max\{n+1, x\})\Bigg] \end{aligned}$$

and if $m < 0$ then

$$\begin{aligned}
f(m-1) = (-1)^{m-1}\, I_{m-1} \Bigg[& \frac{1}{I_0} f(0) \\
& - \frac{e^{2\lambda}}{\lambda} \sum_{n=m-1}^{-1} \frac{(-1)^{n+1}}{I_n I_{n+1}} \mathbb{P}\left(W \leq \min\{n,x\}\right) \mathbb{P}\left(W > \max\{n,x\}\right) \Bigg] \\
= (-1)^{m-1}\, I_{m-1} \Bigg[& \frac{1}{I_0} f(0) \\
& - \frac{e^{2\lambda}}{\lambda} \frac{1}{I_0 I_1} \mathbb{P}\left(W \leq \min\{-1,x\}\right) \mathbb{P}\left(W > \max\{-1,x\}\right) \\
& - \frac{e^{2\lambda}}{\lambda} \sum_{n=m-1}^{-2} \frac{(-1)^{n+1}}{I_n I_{n+1}} \mathbb{P}\left(W \leq \min\{n,x\}\right) \mathbb{P}\left(W > \max\{n,x\}\right) \Bigg] \\
= (-1)^{m-1}\, I_{m-1} \Bigg[& \frac{1}{I_0} f(0) \\
& - \frac{e^{2\lambda}}{\lambda} \frac{1}{I_0 I_1} \mathbb{P}\left(W \leq \min\{-1,x\}\right) \mathbb{P}\left(W > \max\{-1,x\}\right) \\
& + \frac{e^{2\lambda}}{\lambda} \sum_{n=m}^{-1} \frac{(-1)^{n+1}}{I_{n-1} I_n} \mathbb{P}\left(W \leq \min\{n-1,x\}\right) \mathbb{P}\left(W > \max\{n-1,x\}\right) \Bigg].
\end{aligned}$$

6.1.3 Bounding The First Difference : $m > 0$

Our goal is to compute a Lipschitz constant for the function f. Specifically, that is, we wish to determine a bound on $f(k+1) - f(k)$. The method that we use is highly analytical and requires much effort. We will write the expression of $f(k+1) - f(k)$ and then partition the expression into pieces which can be analyzed.

We start with the case that $m > 0$. There is no reason, in particular, to choose $m > 0$, but one case forms the base case to which we will refer for each of the other cases. That is, $m > 0$ is the case for which we find a bound and then we show transformations of the other cases to the $m > 0$ case to respectively bound. So,

$$f(m+1) - f(m) = (-1)^{m+1} \Big\{ I_{m+1} \Big[\frac{1}{I_0} f(0)$$

$$- \frac{e^{2\lambda}}{\lambda} \frac{1}{I_0 I_1} \mathbb{P}\left(W \leq \min\{0,x\}\right) \mathbb{P}\left(W > \max\{0,x\}\right) \Big] + I_m \frac{1}{I_0} f(0) \Bigg\}$$

$$- (-1)^{m+1} \frac{e^{2\lambda}}{\lambda} \sum_{n=0}^{m-1} (-1)^{n+1} \Bigg\{ \frac{I_{m+1}}{I_{n+1} I_{n+2}} \mathbb{P}\left(W \leq \min\{n+1,x\}\right) \mathbb{P}\left(W > \max\{n+1,x\}\right)$$

$$- \frac{I_m}{I_n I_{n+1}} \mathbb{P}\left(W \leq \min\{n,x\}\right) \mathbb{P}\left(W > \max\{n,x\}\right) \Bigg\}$$

Suppose that $x \leq 0$ so that we have if m is even define

$$\begin{aligned} S(m) &= \frac{\mathbb{P}(W \leq x)}{\lambda} \sum_{n=0}^{m-1} (-1)^{n+1} \left\{ \frac{I_{m+1}}{I_{n+2}} H(n+1) - \frac{I_m}{I_{n+1}} H(n) \right\} \\ &= \frac{\mathbb{P}(W \leq x)}{\lambda} \sum_{n=1,3,\ldots}^{m-1} \Bigg\{ \frac{I_{m+1}}{I_{n+2}} H(n+1) - \frac{I_m}{I_{n+1}} H(n) \\ &\quad - \frac{I_{m+1}}{I_{n+1}} H(n) + \frac{I_m}{I_n} H(n-1) \Bigg\} \end{aligned}$$

where $H(n) = \frac{\mathbb{P}(W>n)}{\mathbb{P}(W=n)}$. If m is odd then we have

$$\begin{aligned} S(m) &= \frac{\mathbb{P}(W \leq x)}{\lambda} \Bigg[\sum_{n=1,3,\ldots}^{m-2} \Bigg\{ \frac{I_{m+1}}{I_{n+2}} H(n+1) - \frac{I_m}{I_{n+1}} H(n) \\ &\quad - \frac{I_{m+1}}{I_{n+1}} H(n) + \frac{I_m}{I_n} H(n-1) \Bigg\} + (-1)^m \left(H(m+1) - H(m) \right) \Bigg] \end{aligned}$$

Note that

$$H(n) = \frac{I_{n+1}}{I_n} + \frac{I_{n+1}}{I_n} H(n+1)$$

so that

$$\begin{aligned} H(n+1) &= \frac{I_n}{I_{n+1}} H(n) - 1 \\ H(n-1) &= \frac{I_n}{I_{n-1}} + \frac{I_n}{I_{n-1}} H(n) \end{aligned}$$

which means that

$$
\begin{aligned}
S(m) &= \frac{\mathbb{P}(W \le x)}{\lambda} \left\{ \sum_{n=1,3,\ldots}^{m-1} \left\{ \frac{I_{m+1}}{I_{n+2}} \left[\frac{I_n}{I_{n+1}} H(n) - 1 \right] - \frac{I_m}{I_{n+1}} H(n) \right.\right. \\
&\quad \left.\left. - \frac{I_{m+1}}{I_{n+1}} H(n) - \frac{I_m}{I_n} \left[\frac{I_n}{I_{n-1}} + \frac{I_n}{I_{n-1}} H(n) \right] \right\} \right\} \\
&= \frac{\mathbb{P}(W \le x)}{\lambda} \left\{ \sum_{n=1,3,\ldots}^{m-1} H(n) \left\{ \frac{I_{m+1} I_n}{I_{n+1} I_{n+2}} - \frac{I_m}{I_{n+1}} - \frac{I_{m+1}}{I_{n+1}} + \frac{I_m}{I_{n-1}} \right\} \right. \\
&\quad \left. + \sum_{n=1,3,\ldots}^{m-1} \left\{ \frac{I_{m+1}}{I_{n+2}} - \frac{I_m}{I_{n-1}} \right\} \right\}.
\end{aligned}
$$

It turns out that one may now write

$$
\begin{aligned}
|f(m+1) - f(m)| &= \left| S(m) + I_m \frac{1}{I_0} f(0) + I_{m+1} \left\{ \frac{1}{I_0} f(0) \right.\right. \\
&\quad \left.\left. - \frac{e^{2\lambda}}{\lambda} \frac{1}{I_0 I_1} \mathbb{P}(W \le \min\{0,x\}) \mathbb{P}(W > \max\{0,x\}) \right\} \right|
\end{aligned}
$$

At this point we start to bound by first using a triangle inequality to find

$$
\begin{aligned}
|f(m+1) - f(m)| &= \frac{\mathbb{P}(W \le x)}{\lambda} \left\{ \left| \sum_{n=1,3,\ldots}^{m-1} \frac{I_{m+1}}{I_{n+2}} - \frac{I_m}{I_{n-1}} \right| \right. \\
&\quad + \sum_{n=1,3,\ldots}^{m-1} H(n) \left| \frac{I_{m+1} I_n}{I_{n+1} I_{n+2}} - \frac{I_m}{I_{n+1}} - \frac{I_{m+1}}{I_{n+1}} + \frac{I_m}{I_{n-1}} \right| \\
&\quad + \left| I_m \frac{1}{I_0} f(0) + I_{m+1} \left\{ \frac{1}{I_0} f(0) \right.\right. \\
&\quad \left.\left. - \frac{e^{2\lambda}}{\lambda} \frac{1}{I_0 I_1} \mathbb{P}(W \le \min\{0,x\}) \mathbb{P}(W > \max\{0,x\}) \right\} \right|
\end{aligned} \tag{6.4}
$$

if m is odd and

$$
\begin{aligned}
|f(m+1) - f(m)| &= \frac{\mathbb{P}(W \le x)}{\lambda} \left\{ \left| \sum_{n=1,3,\ldots}^{m-1} \frac{I_{m+1}}{I_{n+2}} - \frac{I_m}{I_{n-1}} \right| \right. \\
&\quad + \sum_{n=1,3,\ldots}^{m-1} H(n) \left| \frac{I_{m+1} I_n}{I_{n+1} I_{n+2}} - \frac{I_m}{I_{n+1}} - \frac{I_{m+1}}{I_{n+1}} + \frac{I_m}{I_{n-1}} \right|
\end{aligned} \tag{6.5}
$$

$$
\begin{aligned}
&+\left|I_m \frac{1}{I_0} f(0) + I_{m+1}\left\{\frac{1}{I_0} f(0)\right.\right.\\
&\left.\left.-\frac{e^{2\lambda}}{\lambda}\frac{1}{I_0 I_1}\mathbb{P}(W \le \min\{0,x\})\,\mathbb{P}(W > \max\{0,x\})\right\}\right|\\
&+|H(n+1) - H(n)|\}
\end{aligned}
$$

if m is even. According to each of the three lines of Equation (6.4), we name the terms T_1, T_2, T_3, and T_4 and bound each in succession in the following theorems.

Theorem 6.1.

$$
T_1 = \sup_{m\in\mathbb{N}^+}\left|\sum_{n=1,3,\ldots}^{m-1} \frac{I_{m+1}}{I_{n+2}} - \frac{I_m}{I_{n-1}}\right| \le 8 \tag{6.6}
$$

for $\lambda \ge 1$.

Proof. Note that

$$
\begin{aligned}
\sum_{n=1,3,\ldots}^{m-1} \frac{I_{m+1}}{I_{n+2}} - \frac{I_m}{I_{n-1}} &= 1 - \frac{I_m}{I_0} + \sum_{n=1,\ldots}^{m-3} \frac{I_{m+1}}{I_{n+2}} - \sum_{n=3,\ldots}^{m-1} \frac{I_m}{I_{n-1}}\\
&= 1 - \frac{I_m}{I_0} + \sum_{n=1,3,\ldots}^{m-3}\left(\frac{I_{m+1}}{I_{n+2}} - \frac{I_m}{I_{n+1}}\right)\\
&= 1 - \frac{I_m}{I_0} + \sum_{n=1,3,\ldots}^{m-3}\frac{I_m}{I_{n+2}}\left(\frac{I_{m+1}}{I_m} - \frac{I_{n+2}}{I_{n+1}}\right).
\end{aligned}
$$

Now, since $\frac{I_{k+1}}{I_k}$ is a decreasing function for $k \ge 0$, see [Bar13], we have that

$$
\begin{aligned}
T_1 &\le \left|1 - \frac{I_m}{I_0}\right| + \sum_{n=1,3,\ldots}^{m-3}\frac{I_m}{I_{n+2}}\left|\frac{I_{m+1}}{I_m} - \frac{I_{n+2}}{I_{n+1}}\right|\\
&= 1 - \frac{I_m}{I_0} + \sum_{n=1,3,\ldots}^{m-3}\frac{I_m}{I_{n+2}}\left(\frac{I_{n+2}}{I_{n+1}} - \frac{I_{m+1}}{I_m}\right).
\end{aligned}
$$

According to [Bar13]

$$
\frac{2\lambda}{k+1+\sqrt{4\lambda^2+(k+1)^2}} \le \frac{I_{k+1}}{I_k} \le \frac{2\lambda}{k+\sqrt{4\lambda^2+k^2}}
$$

for any $k \geq 0$ so that

$$
\begin{aligned}
\frac{I_{n+2}}{I_{n+1}} - \frac{I_{m+1}}{I_m} &\leq \frac{2\lambda}{n+1+\sqrt{4\lambda^2+(n+1)^2}} \frac{2\lambda}{m+1+\sqrt{4\lambda^2+(m+1)^2}} \\
&= -\left(\sqrt{1+\left(\frac{n+1}{2\lambda}\right)^2} - \frac{n+1}{2\lambda}\right) \\
&\quad -\left(\sqrt{1+\left(\frac{m+1}{2\lambda}\right)^2} - \frac{m+1}{2\lambda}\right) \\
&= \frac{m-n}{2\lambda} + \left(\sqrt{1+\left(\frac{n+1}{2\lambda}\right)^2} - \sqrt{1+\left(\frac{m+1}{2\lambda}\right)^2}\right) \\
&= \frac{m-n}{2\lambda} + \frac{\left(\frac{n+1}{2\lambda}\right)^2 - \left(\frac{m+1}{2\lambda}\right)^2}{\left(\sqrt{1+\left(\frac{n+1}{2\lambda}\right)^2} + \sqrt{1+\left(\frac{m+1}{2\lambda}\right)^2}\right)} \\
&= \frac{m-n}{2\lambda} + \frac{\left(\frac{n+m+2}{2\lambda}\right)\left(\frac{n-m}{2\lambda}\right)}{\left(\sqrt{1+\left(\frac{n+1}{2\lambda}\right)^2} - \sqrt{1+\left(\frac{m+1}{2\lambda}\right)^2}\right)} \\
&= \frac{m-n}{2\lambda} - \frac{\left(\frac{n+m+2}{2\lambda}\right)\left(\frac{m-n}{2\lambda}\right)}{\left(\sqrt{1+\left(\frac{n+1}{2\lambda}\right)^2} - \sqrt{1+\left(\frac{m+1}{2\lambda}\right)^2}\right)} \\
&= \frac{m-n}{2\lambda}\left[1 - \frac{\left(\frac{m+n+2}{2\lambda}\right)}{\left(\sqrt{1+\left(\frac{n+1}{2\lambda}\right)^2} - \sqrt{1+\left(\frac{m+1}{2\lambda}\right)^2}\right)}\right].
\end{aligned}
$$

Now, $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ so that

$$
\begin{aligned}
\frac{I_{n+2}}{I_{n+1}} - \frac{I_{m+1}}{I_m} &\leq \frac{m-n}{2\lambda}\left[1 - \frac{\left(\frac{m+n+2}{2\lambda}\right)}{\left(\sqrt{1+\left(\frac{n+1}{2\lambda}\right)^2} - \sqrt{1+\left(\frac{m+1}{2\lambda}\right)^2}\right)}\right] \\
&\leq \frac{m-n}{2\lambda}\left[1 - \frac{\left(\frac{m+n+2}{2\lambda}\right)}{2+\frac{m+n+2}{2\lambda}}\right] \\
&= \frac{m-n}{2\lambda} \frac{2}{2+\frac{m+n+2}{2\lambda}} \\
&= 2\frac{m-n}{4\lambda+m+n+2}
\end{aligned}
$$

$$\leq \frac{m-n}{2\lambda}.$$

Now recall that

$$\begin{aligned}
\frac{I_{k+1}}{I_k} &\leq \frac{2\lambda}{k+\sqrt{4\lambda^2+k^2}} \\
&= \sqrt{1+\left(\frac{k}{2\lambda}\right)^2} - \frac{k}{2\lambda}.
\end{aligned}$$

If $k \leq 2\lambda$ then we can Taylor expand the square root to second order and ignore the remainder since the square root function is concave to give

$$\begin{aligned}
\sqrt{1+\left(\frac{k}{2\lambda}\right)^2} - \frac{k}{2\lambda} &\leq 1+\frac{1}{2}\left(\frac{k}{2\lambda}\right)^2 - \frac{k}{2\lambda} \\
&\leq 1+\frac{1}{2}\frac{k}{2\lambda} - \frac{k}{2\lambda} \\
&= 1-\frac{k}{4\lambda} \\
&\leq e^{-\frac{k}{4\lambda}}.
\end{aligned}$$

On the other hand, if $k > 2\lambda$ then

$$\sqrt{1+\left(\frac{k}{2\lambda}\right)^2} - \frac{k}{2\lambda} = \frac{k}{2\lambda}\left[\sqrt{1+\left(\frac{2\lambda}{k}\right)^2} - 1\right]$$

and so the same procedure may be applied to obtain

$$\begin{aligned}
\sqrt{1+\left(\frac{k}{2\lambda}\right)^2} - \frac{k}{2\lambda} &\leq \frac{k}{2\lambda}\left[1+\frac{1}{2}\left(\frac{2\lambda}{k}\right)^2 - 1\right] \\
&= \frac{\lambda}{k}.
\end{aligned}$$

All of this implies that

$$T_1 \leq \left|1-\frac{I_m}{I_0}\right| + \sum_{n=1,3,\ldots}^{m-3} \frac{I_m}{I_{n+2}}\left|\frac{I_{m+1}}{I_m} - \frac{I_{n+2}}{I_{n+1}}\right|$$

$$
\begin{aligned}
&\leq 1+\sum_{n=1,3,\ldots}^{m-3} \frac{I_m}{I_{n+2}}\left(\frac{I_{n+2}}{I_{n+1}}-\frac{I_{m+1}}{I_m}\right) \\
&= 1+\sum_{n=1,3,\ldots}^{m-3} \prod_{k=n+2}^{m-1} \frac{I_{k+1}}{I_k}\left(\frac{I_{n+2}}{I_{n+1}}-\frac{I_{m+1}}{I_m}\right) \\
&\leq \begin{cases} 1+\sum_{n=1,3,\ldots}^{m-3} \frac{m-n}{2\lambda} \prod_{k=n+2}^{m-1} e^{-\frac{k}{2\lambda}} & \text{if } m-1 \leq 2\lambda \\ 1+\sum_{n=1,3,\ldots}^{2\lambda-2} \frac{m-n}{2\lambda} \prod_{k=n+2}^{2\lambda} e^{-\frac{k}{4\lambda}} \prod_{k=2\lambda+1}^{m-1} \frac{\lambda}{k} & \\ +2\sum_{n=2\lambda-1,odd}^{m-3} \frac{m-n}{4\lambda+m+n+2} \prod_{k=n+2}^{m-1} \frac{\lambda}{k} & \text{if } m-1 > 2\lambda. \end{cases}
\end{aligned}
$$

where in the last line when the sum is split from 1 to $2\lambda-2$ and $2\lambda-1$ to $m-3$ over the odds we used for notational simplicity that $2\lambda-2$ is the smallest odd integer less than or equal to $m-3$.

Now, we require further bounds. Note that when $m-1 \leq 2\lambda$

$$
\begin{aligned}
\prod_{k=n+2}^{m-1} e^{-\frac{k}{2\lambda}} &= \exp\left(-\frac{1}{4\lambda}\sum_{k=n+2}^{m-1} k\right) \\
&= \exp\left(-\frac{1}{8\lambda}\left[(m-1)\,m-(n+1)\,(n+2)\right]\right) \\
&= \exp\left(-\frac{1}{8\lambda}(m+n+1)\,(m-n-2)\right) \\
&\leq \exp\left(-\frac{1}{8\lambda}(m-n-2)^2\right).
\end{aligned}
$$

In addition we have when $n \geq 2\lambda-1$ and $m-1 > 2\lambda$ that both

$$
\begin{aligned}
\prod_{k=2\lambda+1}^{m-1} \frac{\lambda}{k} &\leq \left(\frac{\lambda}{2\lambda+1}\right)^{m-1-(2\lambda+1)+1} \\
&\leq \left(\frac{1}{2}\right)^{m-1-2\lambda}
\end{aligned}
$$

and

$$
\prod_{k=n+2}^{m-1} \frac{\lambda}{k} \leq \left(\frac{\lambda}{n+2}\right)^{m-1-(n+2)+1}
$$

$$
\begin{aligned}
&= \left(\frac{\lambda}{n+2}\right)^{m-n-2} \\
&\leq \left(\frac{\lambda}{2\lambda+1}\right)^{m-n-2} \\
&\leq \left(\frac{1}{2}\right)^{m-n-2}.
\end{aligned}
$$

Implementing these bounds, we obtain,

$$
\begin{aligned}
&\left|\sum_{n=1,3,\ldots}^{m-1} \frac{I_{m+1}}{I_{n+2}} - \frac{I_m}{I_{n-1}}\right| \\
&\leq \begin{cases}
1 + \sum_{n=1,3,\ldots}^{m-3} \exp\left(-\frac{1}{8\lambda}(m-n-2)^2\right) \frac{m-n}{2\lambda} & m-1 \leq 2\lambda \\
1 + \displaystyle\sum_{n=1,3,\ldots}^{2\lambda-2} \left[\exp\left(-\frac{1}{8\lambda}(2\lambda-n-2)^2\right)\left(\frac{1}{2}\right)^{m-1-2\lambda}\right] \frac{m-n}{2\lambda} & \\
\quad + \displaystyle\sum_{n=2\lambda-1,\text{n odd}}^{m-3} \left(\frac{1}{2}\right)^{m-n-2} \cdot \left[2\frac{m-n}{4\lambda+m+n+2}\right] & m-1 > 2\lambda
\end{cases} \\
&\leq \begin{cases}
1 + \frac{1}{2\lambda}\sum_{n=1,3,\ldots}^{m-3} (m-n)\exp\left(-\frac{1}{8\lambda}(m-n-2)^2\right) & m-1 \leq 2\lambda \\
1 + \dfrac{1}{2\lambda} \cdot \left(\dfrac{1}{2}\right)^{m-1-2\lambda} \displaystyle\sum_{n=1,3,\ldots}^{2\lambda-2} (m-n)\exp\left(-\frac{1}{8\lambda}(2\lambda-n-2)^2\right) & \\
\quad + \dfrac{2m}{4\lambda+m+2} \displaystyle\sum_{n=2\lambda-1,\text{n odd}}^{m-3} \left(\frac{1}{2}\right)^{m-n-2} & m-1 > 2\lambda
\end{cases} \\
&\leq \begin{cases}
1 + \frac{1}{2\lambda}\sum_{n=1,3,\ldots}^{m-3} (m-n)\exp\left(-\frac{1}{8\lambda}(m-n-2)^2\right) & m-1 \leq 2\lambda \\
1 + \dfrac{1}{2\lambda} \cdot \left(\dfrac{1}{2}\right)^{m-1-2\lambda} \displaystyle\sum_{n=1,3,\ldots}^{2\lambda-2} (m-n)\exp\left(-\frac{1}{8\lambda}(2\lambda-n-2)^2\right) & \\
\quad + \dfrac{m}{4\lambda+m+2}\left(1 - \left(\dfrac{1}{2}\right)^{m-1-2\lambda}\right) & m-1 > 2\lambda
\end{cases}
\end{aligned}
$$

where we note that in the case of $m-1 > 2\lambda$ in the line above, it is an average of two bounds with probabilities $\left(1 - \left(\frac{1}{2}\right)^{m-1-2\lambda}\right)$ and $\left(\frac{1}{2}\right)^{m-1-2\lambda}$. Now, we show that if $m-1 \leq 2\lambda$

that

$$\frac{1}{2\lambda}\sum_{n=1,3,\ldots}^{m-3}(m-n)\exp\left(-\frac{(m-n-2)^2}{8\lambda}\right)\leq\frac{4\lambda+2\sqrt{\frac{\lambda}{e}}+2\sqrt{2\pi\lambda}}{2\lambda}.$$

If we reindex then we have

$$\begin{aligned}\frac{1}{2\lambda}\sum_{n=1,3,\ldots}^{m-3}(m-n)\exp\left(-\frac{(m-n-2)^2}{8\lambda}\right) &= \frac{1}{2\lambda}\sum_{u=1,3,\ldots}^{m-3}(u+2)\,e^{-\frac{1}{2\sigma^2}u^2}\\ &= \frac{1}{2\lambda}\left[\sum_{u=1,3,\ldots}^{m-3}ue^{-\frac{1}{2\sigma^2}u^2}+2\sum_{u=1,3,\ldots}^{m-3}e^{-\frac{1}{2\sigma^2}u^2}\right]\end{aligned}$$

where $\sigma = 2\sqrt{\lambda}$. Note that both of these terms are Gaussian-like, where the first one is like the expectation of a Gaussian and the second is the probability. Proceeding in this fashion by bounding by an integral using right-handed Riemann sums, the "probability" term becomes

$$2\sum_{u=1,3,\ldots}^{m-3}e^{-\frac{u^2}{2\sigma^2}}\leq 2\int_0^{m-3}e^{-\frac{u^2}{2\sigma^2}}du\leq 2\frac{\sigma\sqrt{2\pi}}{2}=2\sqrt{2\pi\lambda}.$$

For the "expected value" term, some care is needed since the index of the sum can go past σ which is the critical point for the function $f(x) = xe^{-\frac{x^2}{2\sigma^2}}$, and so we cannot just use a left or right Riemann sum. We split the sum up into two pieces: one up to $\sigma_* := \lfloor\sigma\rfloor$ and the other starting from $\sigma^* := \lceil\sigma\rceil$, using a left-sum for the former and a right sum for the latter. In either case, there will be a single term leftover corresponding to either $f(\sigma^*)$ or $f(\sigma_*)$ since we are summing over the odds. In either case, we can always bound this by $f(\sigma) = \sigma e^{-\frac{1}{2}}$.

$$\begin{aligned}\sum_{u=1,3,\ldots}^{m-3}ue^{-\frac{u^2}{2\sigma^2}} &= \sum_{u=1,3,\ldots}^{\sigma_*}ue^{-\frac{u^2}{2\sigma^2}}+\sum_{u=\sigma^*,\text{n odd}}^{m-3}ue^{-\frac{u^2}{2\sigma^2}}+\sigma e^{-\frac{1}{2}}\\ &\leq \int_1^{\sigma_*}ue^{-\frac{u^2}{2\sigma^2}}du+\int_{\sigma^*}^{m-3}ue^{-\frac{u^2}{2\sigma^2}}du+\frac{\sigma}{\sqrt{e}}\\ &\leq \int_0^{\infty}ue^{-\frac{u^2}{2\sigma^2}}du+\frac{\sigma}{\sqrt{e}}\end{aligned}$$

$$
\begin{aligned}
&= \sigma^2 + \frac{\sigma}{\sqrt{e}} \\
&= 4\lambda + 2\sqrt{\frac{\lambda}{e}}.
\end{aligned}
$$

Lastly, for the second sum, reindexing with $u = 2\lambda - n - 2$ and summing over all integers,

$$
\begin{aligned}
\sum_{n=1,3,\dots}^{2\lambda-2} (m-2\lambda)\exp\left(-\frac{1}{8\lambda}(2\lambda-n-2)^2\right) &= (m-2\lambda)\sum_{n=1,3,\dots}^{2\lambda-2} \exp\left(-\frac{1}{8\lambda}(2\lambda-n-2)^2\right) \\
&\leq (m-2\lambda)\sum_{u=0}^{2\lambda-1} e^{-\frac{u^2}{2\sigma^2}} \\
&= (m-2\lambda)\left(1+\int_0^{2\lambda-1} e^{-\frac{u^2}{2\sigma^2}}\,du\right) \\
&\leq (m-2\lambda)\left(1+\frac{\sigma\sqrt{2\pi}}{2}\right) \\
&= (m-2\lambda)\left(1+\sqrt{2\pi\lambda}\right)
\end{aligned}
$$

Substituting in these bounds,

$$
\begin{aligned}
&\left|\sum_{n=1,3,\dots}^{m-1} \frac{I_{m+1}}{I_{n+2}} - \frac{I_m}{I_{n-1}}\right| \\
&\leq \begin{cases} 1 + \frac{4\lambda+2\sqrt{\frac{\lambda}{e}}+2\sqrt{2\pi\lambda}}{2\lambda} & m-1 \leq 2\lambda \\ 1 + \left(\frac{1}{2}\right)^{m-1-2\lambda} \dfrac{4\lambda+2\sqrt{\frac{\lambda}{e}}+2\sqrt{2\pi\lambda}+(m-2\lambda)(1+\sqrt{2\pi\lambda})}{2\lambda} \\ \quad + \dfrac{m}{4\lambda+m+2}\left(1-\left(\dfrac{1}{2}\right)^{m-1-2\lambda}\right) & m-1 > 2\lambda \end{cases} \\
&\leq \begin{cases} 1 + \frac{4\lambda+2\sqrt{\frac{\lambda}{e}}+2\sqrt{2\pi\lambda}}{2\lambda} & m-1 \leq 2\lambda \\ 1 + \max\left\{\frac{m}{4\lambda+m+2}, \frac{2\lambda+\sqrt{\frac{\lambda}{e}}+\sqrt{2\pi\lambda}}{\lambda}\right\} + \left(\frac{1}{2}\right)^{m-1-2\lambda} \frac{(m-2\lambda)(1+\sqrt{2\pi\lambda})}{2\lambda} & m-1 > 2\lambda \end{cases} \\
&\leq \begin{cases} 3 + \frac{C'}{\sqrt{\lambda}} & m-1 \leq 2\lambda \\ 3 + \frac{C'}{\sqrt{\lambda}} + \left(\frac{1}{2}\right)^{m-1-2\lambda} \frac{(m-2\lambda)(1+\sqrt{2\pi\lambda})}{2\lambda} & m-1 > 2\lambda \end{cases}
\end{aligned}
$$

where $C' = \frac{1}{\sqrt{e}} + \sqrt{2\pi} \approx 3.11$. To bound the final term, note that the function $f(x) = \left(\frac{1}{2}\right)^{x-2\lambda}(x-2\lambda)$ is monotone decreasing for $x > 2\lambda + \frac{1}{\ln(2)}$. Thus, we bound by taking $m = 2\lambda + 2$ so that,

$$\left(\frac{1}{2}\right)^{m-1-2\lambda} \frac{(m-2\lambda)(1+\sqrt{2\pi\lambda})}{2\lambda} \leq \frac{1+\sqrt{2\pi\lambda}}{2\lambda}.$$

Finally then,

$$\left| \sum_{n=1,3,\ldots}^{m-1} \frac{I_{m+1}}{I_{n+2}} - \frac{I_m}{I_{n-1}} \right| \leq 3 + \frac{C}{\sqrt{\lambda}} + \frac{1}{2\lambda}$$

where $C = C' + \sqrt{\frac{\pi}{2}} \approx 4.366$. This finally shows that

$$\frac{\mathbb{P}(W \leq x)}{\lambda} \left| \sum_{n=1,3,\ldots}^{m-1} \frac{I_{m+1}}{I_{n+2}} - \frac{I_m}{I_{n-1}} \right| \leq \frac{8}{\lambda} \tag{6.7}$$

for $\lambda \geq 1$ $\square$

Now, we bound T_2.

Theorem 6.2.

$$T_2 = \sum_{n=1,3,\ldots}^{m-1} H(n) \left| \frac{I_{m+1}I_n}{I_{n+1}I_{n+2}} - \frac{I_m}{I_{n+1}} - \frac{I_{m+1}}{I_{n+1}} + \frac{I_m}{I_{n-1}} \right| \leq 73 \tag{6.8}$$

for $\lambda \geq 1$.

Proof. Start with the quantity Q_1 defined as

$$Q_1 = \frac{I_{m+1}I_n}{I_{n+1}I_{n+2}} - \frac{I_m}{I_{n+1}} - \frac{I_{m+1}}{I_{n+1}} + \frac{I_m}{I_{n-1}}$$

and note that

$$Q_1 = \frac{I_{m+1}I_n}{I_{n+1}I_{n+2}} - \frac{I_m}{I_{n+1}} - \frac{I_{m+1}}{I_{n+1}} + \frac{I_m}{I_{n-1}}$$

$$
\begin{aligned}
&= \frac{I_{m+1}}{I_{n+1}I_{n+2}}\left(I_n - I_{n+2}\right) - \frac{I_m}{I_{n+1}I_{n-1}}\left(I_{n-1} - I_{n+1}\right) \\
&= \frac{1}{\lambda}\left\{\left(n+1\right)\frac{I_{m+1}}{I_{n+2}} - n\frac{I_m}{I_{n+1}}\frac{I_n}{I_{n-1}}\right\}.
\end{aligned}
$$

This means that

$$
\begin{aligned}
T_2 &= \sum_{n=1,3,\ldots}^{m-1} H\left(n\right)\left|\frac{I_{m+1}I_n}{I_{n+1}I_{n+2}} - \frac{I_m}{I_{n+1}} - \frac{I_{m+1}}{I_{n+1}} + \frac{I_m}{I_{n-1}}\right| \\
&\leq \frac{1}{\lambda}H\left(m-1\right)\left|m - \left(m-1\right)\frac{I_m}{I_{m-2}}\right| + \sum_{n=1,3,\ldots}^{m-3} H\left(n\right)\left|\frac{I_{m+1}I_n}{I_{n+1}I_{n+2}} - \frac{I_m}{I_{n+1}} - \frac{I_{m+1}}{I_{n+1}} + \frac{I_m}{I_{n-1}}\right|.
\end{aligned}
$$

Now, note that

$$
m - \left(m-1\right)\frac{I_{m-1}}{I_{m-2}} = m\left(1 - \frac{I_{m-1}}{I_{m-2}}\right) + \frac{I_{m-1}}{I_{m-2}} \geq 0
$$

and since $m > 1$ then we have

$$
H\left(m-1\right)\left(m - \left(m-1\right)\frac{I_{m-1}}{I_{m-2}}\right) \leq H\left(m-1\right)m.
$$

Now, using the bounds derived in [PB13] and, again, since $m > 1$ we have

$$
H\left(m-1\right)\left(m - \left(m-1\right)\frac{I_{m-1}}{I_{m-2}}\right) \leq H\left(m-1\right)m \leq 16\lambda.
$$

Now we must determine a bound for Q_1 for $1 \leq n \leq m-3$. For this, note that

$$
\begin{aligned}
\left|\left(n+1\right)\frac{I_{m+1}}{I_{n+2}} - n\frac{I_m}{I_{n+1}}\frac{I_n}{I_{n-1}}\right| &= \frac{I_m}{I_{n+2}}\left|\left(n+1\right)\frac{I_{m+1}}{I_m} - n\frac{I_{n+2}}{I_{n+1}}\frac{I_n}{I_{n-1}}\right| \\
&= \frac{I_m}{I_{n+2}}\left|\left(n+1\right)\left(\frac{I_{m+1}}{I_m} - \frac{I_{n+2}}{I_{n+1}}\right) + \frac{I_{n+2}}{I_{n+1}}\left(\left[n+1\right] - n\frac{I_n}{I_{n-1}}\right)\right| \\
&\leq \frac{I_m}{I_{n+2}}\left(n+1\right)\left(\frac{I_{n+2}}{I_{n+1}} - \frac{I_{m+1}}{I_m}\right) + \frac{I_m}{I_{n+1}} + \frac{I_m}{I_{n+1}}n\left(1 - \frac{I_n}{I_{n-1}}\right).
\end{aligned}
$$

According to [Bar13] we have the Turanian

$$
0 < I_k^{-} I_{k-1}I_{k+1} < \frac{I_k^2}{k+1}
$$

for all $k > -1$. With some algebra we have that

$$1 < \frac{I_k^2}{I_{k-1}I_{k+1}}$$

which implies that

$$\begin{aligned}
\frac{I_m}{I_{n+1}}\left(1-\frac{I_n}{I_{n-1}}\right) &= \frac{I_m}{I_{n+1}} - I_m\frac{I_n}{I_{n-1}I_{n+1}} \\
&= \frac{I_m}{I_{n+1}} - \frac{I_m}{I_n}\frac{I_n^2}{I_{n-1}I_{n+1}} \\
&\leq \frac{I_m}{I_{n+1}} - \frac{I_m}{I_n}
\end{aligned}$$

and

$$Q_1 \leq \frac{I_m}{I_{n+2}}(n+1)\left(\frac{I_{n+2}}{I_{n+1}} - \frac{I_{m+1}}{I_m}\right) + \frac{I_m}{I_{n+1}} + n\left(\frac{I_m}{I_{n+1}} - \frac{I_m}{I_n}\right).$$

Now, note that

$$\begin{aligned}
T_2 - 16 &\leq \frac{1}{\lambda}\sum_{n=1,3,\ldots}^{m-3} H(n)\left[\frac{I_m}{I_{n+2}}(n+1)\left(\frac{I_{n+2}}{I_{n+1}} - \frac{I_{m+1}}{I_m}\right) + \frac{I_m}{I_{n+1}}\right. \\
&\quad \left. + n\left(\frac{I_m}{I_{n+1}} - \frac{I_m}{I_n}\right)\right] \\
&= \frac{1}{\lambda}\left\{\sum_{n=1,3,\ldots}^{m-3} H(n)\left[\frac{I_m}{I_{n+2}}(n+1)\left(\frac{I_{n+2}}{I_{n+1}} - \frac{I_{m+1}}{I_m}\right) + \frac{I_m}{I_{n+1}}\right]\right. \\
&\quad \left. + \sum_{n=1,3,\ldots}^{m-3} H(n)\, n\left(\frac{I_m}{I_{n+1}} - \frac{I_m}{I_n}\right)\right\} \\
&\leq \frac{1}{\lambda}\left\{\sum_{n=1,3,\ldots}^{m-3} H(n)\left[\frac{I_m}{I_{n+2}}(n+1)\left(\frac{I_{n+2}}{I_{n+1}} - \frac{I_{m+1}}{I_m}\right) + \frac{I_m}{I_{n+1}}\right]\right\} \\
&\quad + 8\sum_{n=1,3,\ldots}^{m-3}\left(\frac{I_m}{I_{n+1}} - \frac{I_m}{I_n}\right) \\
&= \frac{1}{\lambda}\left\{\sum_{n=1,3,\ldots}^{m-3} H(n)\left[\frac{I_m}{I_{n+2}}(n+1)\left(\frac{I_{n+2}}{I_{n+1}} - \frac{I_{m+1}}{I_m}\right) + \frac{I_m}{I_{n+1}}\right]\right\} \\
&\quad + 8\left(\frac{I_m}{I_{m-2}} - \frac{I_m}{I_1}\right)
\end{aligned}$$

$$\leq \quad \frac{1}{\lambda}\left\{\sum_{n=1,3,\ldots}^{m-3} H\left(n\right)\left[\frac{I_m}{I_{n+2}}\left(n+1\right)\left(\frac{I_{n+2}}{I_{n+1}}-\frac{I_{m+1}}{I_m}\right)+\frac{I_m}{I_{n+1}}\right]\right\}$$
$$+8$$

where we have again used the result of [PB13]. This means that now we have

$$T_2 \leq 24 + \frac{1}{\lambda}\sum_{n=1,3,\ldots}^{m-3} H\left(n\right)\left|\frac{I_m}{I_{n+2}}\left(n+1\right)\left(\frac{I_{n+2}}{I_{n+1}}-\frac{I_{m+1}}{I_m}\right)+\frac{I_m}{I_{n+1}}\right|.$$

and $\lambda > \frac{1}{2}$, there exists a positive integer λ' such that

$$\lambda = \lambda' + \epsilon, \quad \epsilon \in \Big(-\frac{1}{2}, \frac{1}{2}\Big].$$

Thus,

$$2\lambda = 2\lambda' + 2\epsilon$$

so that

$$2\lambda' - 1 < 2\lambda \leq 2\lambda' + 1.$$

Now,

$$\frac{1}{\lambda}\left|\sum_{n=1,3,\ldots}^{m-1} H(n)\left[(n+1)\frac{I_{m+1}}{I_{n+2}} - n\frac{I_m}{I_{n+1}}\frac{I_n}{I_{n-1}}\right]\right|$$

$$\leq 24 + \frac{1}{\lambda}\sum_{n=1,3,\ldots}^{m-3} H(n)\left[\frac{I_m}{I_{n+2}}(n+1)\left(\frac{I_{n+2}}{I_{n+1}} - \frac{I_{m+1}}{I_m}\right) + \frac{I_m}{I_{n+1}}\right]$$

$$= \begin{cases} 24 + \frac{1}{\lambda}\sum_{n=1,3,\ldots}^{m-3} H(n)\left[\frac{I_m}{I_{n+2}}(n+1)\left(\frac{I_{n+2}}{I_{n+1}} - \frac{I_{m+1}}{I_m}\right) + \frac{I_m}{I_{n+1}}\right] & m-1 \leq 2\lambda \\ 24 + \frac{1}{\lambda}\left[\sum_{n=1,\mathrm{odd}}^{2\lambda'-3} H(n)\left[\frac{I_m}{I_{n+2}}(n+1)\left(\frac{I_{n+2}}{I_{n+1}} - \frac{I_{m+1}}{I_m}\right) + \frac{I_m}{I_{n+1}}\right]\right. & \\ \left. + \sum_{n=2\lambda'-1,\mathrm{odd}}^{m-3} H(n)\left[\frac{I_m}{I_{n+2}}(n+1)\left(\frac{I_{n+2}}{I_{n+1}} - \frac{I_{m+1}}{I_m}\right) + \frac{I_m}{I_{n+1}}\right]\right] & m-1 > 2\lambda \end{cases}$$

$$\leq \begin{cases} 24 + \frac{1}{\lambda}\sum_{n=1,3,\ldots}^{m-3} H(n)\left[\prod_{j=n+2}^{m-1}\frac{I_{j+1}}{I_j}(n+1)\frac{m-n}{2\lambda} + \prod_{j=n+1}^{m-1}\frac{I_{j+1}}{I_j}\right] & m-1 \leq 2\lambda \\ 24 + \dfrac{1}{\lambda}\left[\displaystyle\sum_{n=1,\mathrm{odd}}^{2\lambda'-3} H(n)\left[\prod_{j=n+2}^{2\lambda'}\frac{I_{j+1}}{I_j}\prod_{j=2\lambda'+1}^{m-1}\frac{I_{j+1}}{I_j}(n+1)\frac{m-n}{2\lambda}\right.\right. & \\ \left. \quad + \displaystyle\prod_{j=n+1}^{2\lambda'}\frac{I_{j+1}}{I_j}\prod_{j=2\lambda'+1}^{m-1}\frac{I_{j+1}}{I_j}\right] & m-1 > 2\lambda \\ \left. \quad + \displaystyle\sum_{n=2\lambda'-1,\mathrm{odd}}^{m-3} H(n)\left[\prod_{j=n+2}^{m-1}\frac{I_{j+1}}{I_j}2\frac{m-n}{4\lambda+m+n+2} + \prod_{j=n+1}^{m-1}\frac{I_{j+1}}{I_j}\right]\right] & \end{cases}$$

$$\leq \begin{cases} 24 + \frac{1}{\lambda}\sum_{n=1,3,\ldots}^{m-3} H(n)\left[e^{-\frac{(m-n-2)^2}{8\lambda}}(n+1)\frac{m-n}{2\lambda} + e^{-\frac{(m-n-1)^2}{8\lambda}}\right] & m-1 \leq 2\lambda \\ 24 + \dfrac{1}{\lambda}\left[\displaystyle\sum_{n=1,\mathrm{odd}}^{2\lambda'-3} H(n)\left[e^{-\frac{(2\lambda'-n-2)^2}{8\lambda}}\left(\frac{1}{2}\right)^{m-1-2\lambda'}(n+1)\frac{m-n}{2\lambda}\right.\right. & \\ \left. \quad + e^{-\frac{(2\lambda'-n-1)^2}{8\lambda}}\left(\frac{1}{2}\right)^{m-1-2\lambda'}\right] & m-1 > 2\lambda \\ \left. \quad + \displaystyle\sum_{n=2\lambda'-1,\mathrm{odd}}^{m-3} H(n)\left[\left(\frac{1}{2}\right)^{m-n-2}\frac{2m}{4\lambda+m+2} + \left(\frac{1}{2}\right)^{m-n-1}\right]\right] & \end{cases}$$

where we have used that

1. $$\prod_{k=n+2}^{m-1}\frac{I_{k+1}}{I_k} \leq \prod_{k=n+2}^{m-1} e^{-\frac{k}{4\lambda}} \leq \exp\left(-\frac{1}{8\lambda}(m-n-2)^2\right)$$

   when $m-1 \leq 2\lambda$.

2.
$$\prod_{k=2\lambda+1}^{m-1} \frac{I_{k+1}}{I_k} \leq \prod_{k=2\lambda+1}^{m-1} \frac{\lambda}{k} \leq \left(\frac{1}{2}\right)^{m-1-2\lambda}.$$

$$\prod_{k=n+2}^{m-1} \frac{I_{k+1}}{I_k} \leq \prod_{k=n+2}^{m-1} \frac{\lambda}{k} \leq \left(\frac{1}{2}\right)^{m-n-2}$$

when $n \geq 2\lambda - 1$ and $m - 1 > 2\lambda$.

and that

$$\frac{I_{n+2}}{I_{n+1}} - \frac{I_{m+1}}{I_m} \leq 2\frac{m-n}{4\lambda + m + n + 2} \leq \frac{m-n}{2\lambda}.$$

Again, by [PB13] $H(n) \leq \frac{8\lambda}{n}$ so that

$$\begin{aligned}
&\frac{1}{\lambda} \sum_{n=2\lambda'-1,\text{odd}}^{m-3} H(n) \left[\left(\frac{1}{2}\right)^{m-n-2} \frac{2m}{4\lambda + m + 2} + \left(\frac{1}{2}\right)^{m-n-1} \right] \\
&\leq \frac{8}{2\lambda' - 1} \left[\frac{m}{4\lambda + m + 2} \left(1 - \left(\frac{1}{2}\right)^{m-2\lambda'-1}\right) + \frac{1 - \left(\frac{1}{2}\right)^{m-2\lambda'-1}}{4} \right] \\
&\leq \frac{10}{2\lambda' - 1} \left(1 - \left(\frac{1}{2}\right)^{m-2\lambda'-1}\right)
\end{aligned}$$

Next, for any $m \leq 2\lambda'$

$$
\begin{aligned}
\frac{1}{\lambda}\sum_{n=1,3,\ldots}^{m-3} H(n)e^{-\frac{(m-n-1)^2}{8\lambda}} &\leq \frac{1}{\lambda}\sum_{n=1,3,\ldots}^{m-3} \frac{8\lambda}{n-\frac{1}{2}+\sqrt{\frac{32\lambda}{\pi}}} e^{-\frac{(m-n-1)^2}{8\lambda}} \\
&\leq \frac{8}{\frac{1}{2}+\sqrt{\frac{32\lambda}{\pi}}}\sum_{n=1}^{m-3} e^{-\frac{(m-n-1)^2}{8\lambda}} \\
&= \frac{8}{\frac{1}{2}+\sqrt{\frac{32\lambda}{\pi}}}\sum_{n=2}^{m-2} e^{-\frac{u^2}{8\lambda}} \\
&\leq \frac{8}{\frac{1}{2}+\sqrt{\frac{32\lambda}{\pi}}}\int_0^{m-2} e^{-\frac{u^2}{8\lambda}} \\
&\leq \frac{8(2\sqrt{\lambda})(\sqrt{2\pi})}{\frac{1}{2}+\sqrt{\frac{32\lambda}{\pi}}}\frac{1}{(2\sqrt{\lambda})\sqrt{2\pi}}\int_0^{\infty} e^{-\frac{u^2}{8\lambda}} \\
&\leq \frac{8(2\sqrt{\lambda})(\sqrt{2\pi})}{\frac{1}{2}+\sqrt{\frac{32\lambda}{\pi}}}\frac{1}{2} \\
&= \frac{8\sqrt{2\pi}\sqrt{\lambda}}{\frac{1}{2}+\sqrt{\frac{32\lambda}{\pi}}} \\
&\leq 8\sqrt{2\pi}\sqrt{\frac{\pi}{32}} \approx 6.28
\end{aligned}
$$

so that we obtain the bound

$$
\begin{cases}
24+\frac{1}{\lambda}\sum_{n=1,3,\ldots}^{m-3} H(n)e^{-\frac{(m-n-2)^2}{8\lambda}}(n+1)\frac{m-n}{2\lambda}+8\sqrt{2\pi}\sqrt{\frac{\pi}{32}} & m-1\leq 2\lambda \\
24+\dfrac{1}{\lambda}\displaystyle\sum_{n=1,\mathrm{odd}}^{2\lambda'-3} H(n)\left[e^{-\frac{(2\lambda'-n-2)^2}{8\lambda}}\left(\frac{1}{2}\right)^{m-1-2\lambda'}(n+1)\frac{m-n}{2\lambda}\right] & \\
\quad +8\sqrt{2\pi}\sqrt{\dfrac{\pi}{32}}\left(\dfrac{1}{2}\right)^{m-2\lambda'-1}+\dfrac{10}{2\lambda'-1}\left(1-\left(\dfrac{1}{2}\right)^{m-2\lambda'-1}\right) & m-1>2\lambda
\end{cases}
$$

This means that, by using previous results, we have

$$\begin{aligned}\frac{1}{\lambda}\sum_{n=1,3\ldots}^{m-3} H(n)e^{\frac{(m-n-2)^2}{8\lambda}}(n+1)\frac{m-n}{2\lambda} &\leq 8\sum_{n=1,3\ldots}^{m-3}\frac{n+1}{n-\frac{1}{2}+\sqrt{\frac{32\lambda}{\pi}}}e^{\frac{(m-n-2)^2}{8\lambda}}\frac{m-n}{2\lambda}\\ &\leq 8\frac{1}{2\lambda}\sum_{n=1,3,\ldots}^{m-3} e^{\frac{(m-n-2)^2}{8\lambda}}(m-n)\\ &\leq 8\left(\frac{4\lambda+2\sqrt{\frac{\lambda}{e}}+2\sqrt{2\pi\lambda}}{2\lambda}\right)\end{aligned}$$

Using the same bound on $H(n)$ and employing a similar argument,

$$\begin{aligned}&\frac{1}{\lambda}\sum_{n=1,\text{odd}}^{2\lambda'-3} H(n)\left[e^{-\frac{(2\lambda'-n-2)^2}{8\lambda}}(n+1)\frac{m-n}{2\lambda}\right]\\ &\qquad\leq 8\left(\frac{4\lambda+2\sqrt{\frac{\lambda}{e}}+2\sqrt{2\pi\lambda}+(m-2\lambda')\sqrt{2\pi\lambda}}{2\lambda}\right).\end{aligned}$$

Putting this all together, we obtain the bound

$$\begin{cases} 24+8\left(\frac{4\lambda+2\sqrt{\frac{\lambda}{e}}+2\sqrt{2\pi\lambda}}{2\lambda}\right)+8\sqrt{2\pi}\sqrt{\frac{\pi}{32}} & m-1\leq 2\lambda \\ 24+\left[8\left(\dfrac{4\lambda+2\sqrt{\frac{\lambda}{e}}+2\sqrt{2\pi\lambda}+(m-2\lambda')\sqrt{2\pi\lambda}}{2\lambda}\right)\right. & \\ \qquad \left.+8\sqrt{2\pi}\sqrt{\dfrac{\pi}{32}}\right]\left(\dfrac{1}{2}\right)^{m-2\lambda'-1}+\dfrac{10}{2\lambda'-1}\left(1-\left(\dfrac{1}{2}\right)^{m-2\lambda'-1}\right) & m-1>2\lambda \end{cases}$$

$$\leq\begin{cases} 24+8\left(2+\frac{1}{\sqrt{e}}+\sqrt{2\pi}\right)+8\sqrt{2\pi}\sqrt{\frac{\pi}{32}} & m-1\leq 2\lambda \\ 24+\left[8\left(2+\dfrac{1}{\sqrt{e}}+\sqrt{2\pi}+\sqrt{2\pi}+(m-2\lambda')\sqrt{\dfrac{\pi}{2}}\right)\right. & \\ \qquad \left.+8\sqrt{2\pi}\sqrt{\dfrac{\pi}{32}}\right]\left(\dfrac{1}{2}\right)^{m-2\lambda'-1}+\dfrac{10}{2\lambda'-1}\left(1-\left(\dfrac{1}{2}\right)^{m-2\lambda'-1}\right) & m-1>2\lambda \end{cases}$$

$$\leq\begin{cases} C & m-1\leq 2\lambda \\ 24+\max\{C',\frac{10}{2\lambda'-1}\}+2\sqrt{\frac{\pi}{2}}(m-2\lambda')\left(\frac{1}{2}\right)^{m-2\lambda'} & m-1>2\lambda \end{cases}$$

where $C'=8\left(2+\frac{1}{\sqrt{e}}+\sqrt{2\pi}\right)+8\sqrt{2\pi}\sqrt{\frac{\pi}{32}}\approx 47.18$

$C=24+8\left(2+\frac{1}{\sqrt{e}}+\sqrt{2\pi}\right)+8\sqrt{2\pi}\sqrt{\frac{\pi}{32}}\approx 71.18.$

Now, note that $f(x)=\left(\frac{1}{2}\right)^{x-2\lambda'}(x-2\lambda')$ is monotone decreasing for $x>2\lambda'+\frac{1}{\ln(2)}$. Thus, we may take $m=2\lambda'+1$, at which $(m-2\lambda')\left(\frac{1}{2}\right)^{m-2\lambda'}\leq\frac{1}{2}$.

Finally, noting that $\lambda'\geq 1$, in which case $C'>10$, the bound becomes $C+\sqrt{\frac{\pi}{2}}\approx$ 72.433. □

This means that

$$\frac{\mathbb{P}\left(W\leq x\right)}{\lambda}T_2\leq\frac{83}{\lambda} \tag{6.9}$$

Now we bound T_3.

Theorem 6.3.

$$\begin{aligned}
T_3 &= \left| I_m \frac{1}{I_0} f(0) + I_{m+1} \left\{ \frac{1}{I_0} f(0) - \frac{e^{2\lambda}}{\lambda} \frac{1}{I_0 I_1} \mathbb{P}(W \leq \min\{0, x\}) \mathbb{P}(W > \max\{0, x\}) \right\} \right| \\
&\in O\left(\frac{1}{\lambda}\right).
\end{aligned}$$

Proof. Define $p_k = \mathbb{P}(W = k)$ and note that

$$\begin{aligned}
\frac{1}{2p_0}\left[\frac{p_m + p_{m+1}}{p_0 + p_1} - p_{m+1}\frac{2\mathbb{P}(W > 0)}{p_1}\right] &= \frac{1}{2p_0}\left[\frac{p_{m+1}\left[p_1 - 2\mathbb{P}(W > 0)(p_0 + p_1)\right] + p_m p_1}{p_1 (p_0 + p_1)}\right] \\
&= \frac{1}{2p_0 p_1 (p_0 + p_1)} \left[p_{m+1}\left[p_1\left[1 - 2\mathbb{P}(W > 0)\right] - 2\mathbb{P}(W > 0) p_0\right] + p_1 p_m\right] \\
&= \frac{1}{2p_0 p_1 (p_0 + p_1)} \left[p_{m+1} p_0 \left[p_1 - 2\mathbb{P}(W > 0)\right] + p_1 p_m\right] \\
&= \frac{1}{2p_0 p_1 (p_0 + p_1)} \left[p_{m+1} p_0 \left[p_0 + p_1 - 1\right] + p_1 p_m\right] \\
&\geq \frac{1}{2p_0 p_1 (p_0 + p_1)} \left[p_1 p_m - p_0 p_{m+1}\right] \\
&\geq 0
\end{aligned}$$

since

$$\frac{p_1}{p_0} \geq \frac{p_{m+1}}{p_m} \tag{6.10}$$

for $m \geq 0$. Now,

$$\begin{aligned}
\frac{1}{2p_0}\left[\frac{p_m + p_{m+1}}{p_0 + p_1} - p_{m+1}\frac{2\mathbb{P}(W > 0)}{p_1}\right] &= \frac{1}{2p_0 p_1}\left[\frac{p_1}{p_0 + p_1}(p_m + p_{m+1}) - p_{m+1} 2\mathbb{P}(W > 0)\right] \\
&\leq \frac{1}{2p_0 p_1}\left[\frac{p_m + p_{m+1}}{2} - p_{m+1} 2\mathbb{P}(W > 0)\right] \\
&= \frac{1}{2p_0 p_1}\left[\frac{p_m + p_{m+1}}{2} - 2p_{m+1}\mathbb{P}[W > 0]\right] \\
&= \frac{1}{4p_0 p_1}\left[p_m + p_{m+1} - 4p_{m+1}\mathbb{P}[W > 0]\right] \\
&= \frac{1}{4p_0 p_1}\left[p_m + p_{m+1} - 2p_{m+1}\mathbb{P}[W > 0] - 2p_{m+1}\mathbb{P}[W > 0]\right] \\
&= \frac{1}{4p_0 p_1}\left[p_m + p_{m+1}(1 - 2\mathbb{P}[W > 0]) - 2p_{m+1}\mathbb{P}[W > 0]\right]
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{4p_0p_1}\left[p_m + p_{m+1}p_0 - 2p_{m+1}\mathbb{P}\left[W > 0\right]\right] \\
&= \frac{1}{4p_0p_1}\left[p_m + p_{m+1}p_0 + p_{m+1}(1 - 2\mathbb{P}\left[W > 0\right]) - p_{m+1}\right] \\
&= \frac{1}{4p_0p_1}\left[p_m + p_{m+1}p_0 + p_{m+1}p_0 - p_{m+1}\right] \\
&= \frac{1}{4p_0p_1}\left[p_m + 2p_{m+1}p_0 - p_{m+1}\right] \\
&\le \frac{1}{4p_0p_1}\left[p_{m-1} + 2p_{m+1}p_0 - p_{m+1}\right] \\
&= \frac{1}{4p_0p_1}\left[2p_{m+1}p_0 + \frac{m}{\lambda}p_m\right] \\
&\le \frac{1}{4p_0p_1}\left[2p_1p_0 + \frac{1}{\lambda}p_1\right] \\
&= \frac{1}{4p_0}\left[2p_0 + \frac{1}{\lambda}\right] \\
&= \frac{1}{2} + \frac{1}{4p_0\lambda} \\
&= \frac{1}{2} + \frac{1}{4\lambda}\left(3 + \beta\frac{\frac{1}{2} + \sqrt{1}4 + \frac{8\lambda}{\pi\beta}}{4\lambda}\right) \\
&\le 1
\end{aligned}
$$

where we have used a result due to [PB13].

for any m, see [AS72]. $\square$

Finally, for T_4, it is clear that

$$
\begin{aligned}
T_4 &= H\left(m\right) - H\left(m+1\right) \\
&= \frac{\mathbb{P}\left(W > m\right)}{\mathbb{P}\left(W = m\right)} - \frac{\mathbb{P}\left(W > m+1\right)}{\mathbb{P}\left(W = m+1\right)} \\
&= \frac{1}{\mathbb{P}\left(W = m\right)}\left[\mathbb{P}\left(W > m\right) - \mathbb{P}\left(W > m+1\right)\right] \\
&= \frac{\mathbb{P}\left(W = m+1\right)}{\mathbb{P}\left(W = m\right)} \\
&\le 1
\end{aligned}
$$

and, at the same time $H\left(m\right) - H\left(m+1\right) \ge 0$ so, we have that $T_4 \le 1$.

After all of this we can finally state the following proof of Theorem 2.6

Proof.

$$|f(m+1) - f(m)| \in O\left(\frac{83}{\lambda}\right) \tag{6.11}$$

for $m > 0$. □

6.1.4 Other Cases

Throughout, we will write $S(m)$ to be the generic sum which occurs in the difference $f(m+1) - f(m)$. Now, suppose that $x \geq m$ so that

$$S(m) = \frac{\mathbb{P}(W > x)}{\lambda} \sum_{n=1,3,\ldots}^{m-1} \left\{ \frac{I_{m+1}}{I_{n+2}} G(n+1) - \frac{I_m}{I_{n+1}} G(n) - \frac{I_{m+1}}{I_{n+1}} G(n) + \frac{I_m}{I_n} G(n-1) \right\}. \tag{6.12}$$

Define

$$G(n) = \frac{\mathbb{P}(W \leq n)}{\mathbb{P}(W = n)}$$

and note that

$$G(n) = 1 + \frac{I_{n-1}}{I_n} G(n-1)$$

so that

$$\begin{aligned} G(n+1) &= 1 + \frac{I_n}{I_{n+1}} G(n) \\ G(n-1) &= \frac{I_n}{I_{n-1}} G(n) - \frac{I_n}{I_{n-1}}. \end{aligned}$$

This means that

$$\begin{aligned} S(m) = \frac{\mathbb{P}(W > x)}{\lambda} \sum_{n=1,3,\ldots}^{m-1} & \left\{ \frac{I_{m+1}}{I_{n+2}} \left[1 + \frac{I_n}{I_{n+1}} G(n) \right] - \frac{I_m}{I_{n+1}} G(n) - \frac{I_{m+1}}{I_{n+1}} G(n) \right. \\ & \left. + \frac{I_m}{I_n} \left[\frac{I_n}{I_{n-1}} G(n) - \frac{I_n}{I_{n-1}} \right] \right\} \end{aligned}$$

$$= \frac{\mathbb{P}(W > x)}{\lambda} \left\{ \sum_{n=1,3,\ldots}^{m-1} G(n) \left[\frac{I_{m+1} I_n}{I_{n+1} I_{n+2}} - \frac{I_m}{I_{n+1}} - \frac{I_{m+1}}{I_{n+1}} + \frac{I_m}{I_{n-1}} \right] \right.$$
$$\left. + \sum_{n=1,3,\ldots}^{m-1} \left[\frac{I_{m+1}}{I_{n+2}} - \frac{I_m}{I_{n-1}} \right] \right\}.$$

Now, note that for $x \geq m$

$$\begin{aligned} \frac{\mathbb{P}(W > x)}{\lambda} G(n) &= \frac{\mathbb{P}(W > x)}{\lambda} \frac{\mathbb{P}(W \leq n)}{\mathbb{P}(W = n)} \\ &\leq \frac{\mathbb{P}(W \leq x)}{\lambda} \frac{\mathbb{P}(W > n)}{\mathbb{P}(W = n)} && (6.13) \\ &= \frac{\mathbb{P}(W \leq x)}{\lambda} H(n) && (6.14) \end{aligned}$$

which means that we can use immediately the earlier results. Now, if $0 \leq x \leq m-1$ is even then we have

$$\mathbb{P}(W > x) \sum_{n=1,3,\ldots}^{x-1} \left\{ \frac{I_{m+1}}{I_{n+2}} G(n+1) - \frac{I_m}{I_{n+1}} G(n) - \frac{I_{m+1}}{I_{n+1}} G(n) + \frac{I_m}{I_n} G(n-1) \right\}$$
$$+ \mathbb{P}(W \leq x) \sum_{n=x+1,\ldots}^{m-1} \left\{ \frac{I_{m+1}}{I_{n+2}} H(n+1) - \frac{I_m}{I_{n+1}} H(n) - \frac{I_{m+1}}{I_{n+1}} H(n) + \frac{I_m}{I_n} H(n-1) \right\}$$

and if $0 < x \leq m-1$ is odd then we have

$$\mathbb{P}(W > x) \sum_{n=1,3,\ldots}^{x} \left\{ \frac{I_{m+1}}{I_{n+2}} G(n+1) - \frac{I_m}{I_{n+1}} G(n) - \frac{I_{m+1}}{I_{n+1}} G(n) + \frac{I_m}{I_n} G(n-1) \right\}$$
$$+ \mathbb{P}(W \leq x) \sum_{n=x+2,\ldots}^{m-1} \left\{ \frac{I_{m+1}}{I_{n+2}} H(n+1) - \frac{I_m}{I_{n+1}} H(n) - \frac{I_{m+1}}{I_{n+1}} H(n) + \frac{I_m}{I_n} H(n-1) \right\}.$$

We can use the same method on the above as was used in $x \leq 0 \leq m-1$ and $0 < m-1 \leq x$ to bound each by 2 times the respective bounds.

6.1.5 The First Difference $m < 0$

When $m < 0$ then

$$
\begin{aligned}
f(m-1) - f(m) = (-1)^{m-1} \Bigg\{ I_{m-1} \Bigg[\frac{1}{I_0} f(0) \\
- \frac{e^{2\lambda}}{\lambda} \frac{1}{I_0 I_1} \mathbb{P}\left(W \le \min\{-1, x\}\right) \mathbb{P}\left(W > \max\{-1, x\}\right) \Bigg] + I_m \frac{1}{I_0} f(0) \Bigg\} \\
+ (-1)^{m-1} \frac{e^{2\lambda}}{\lambda} \sum_{n=m}^{-1} (-1)^{n+1} \Bigg\{ \frac{I_{m-1}}{I_{n-1} I_n} \mathbb{P}\left(W \le \min\{n-1, x\}\right) \mathbb{P}\left(W > \max\{n-1, x\}\right) \\
- \frac{I_m}{I_n I_{n+1}} \mathbb{P}\left(W \le \min\{n, x\}\right) \mathbb{P}\left(W > \max\{n, x\}\right) \Bigg\}.
\end{aligned}
$$

Now, if $x \ge -1$ then

$$
\begin{aligned}
S(m) &= \frac{\mathbb{P}(W > x)}{\lambda} \sum_{n=m}^{-1} (-1)^{n+1} \left[\frac{I_{m-1}}{I_{n-1}} \frac{\mathbb{P}(W \le n-1)}{\mathbb{P}(W = n)} - \frac{I_m}{I_n} \frac{\mathbb{P}(W \le n)}{\mathbb{P}(W = n+1)} \right] \\
&= \frac{\mathbb{P}(W > x)}{\lambda} \sum_{n=1}^{|m|} (-1)^{n+1} \left[\frac{I_{|m|+1}}{I_{n+1}} \frac{\mathbb{P}(W \le -n-1)}{\mathbb{P}(W = n)} - \frac{I_{|m|}}{I_n} \frac{\mathbb{P}(W \le -n)}{\mathbb{P}(W = -n+1)} \right] \\
&= \frac{\mathbb{P}(W > x)}{\lambda} \sum_{n=1}^{|m|} (-1)^{n+1} \left[\frac{I_{|m|+1}}{I_{n+1}} \frac{\mathbb{P}(W \le -n-1)}{\mathbb{P}(W = n)} - \frac{I_{|m|}}{I_n} \frac{\mathbb{P}(W \le -n)}{\mathbb{P}(W = n-1)} \right].
\end{aligned}
$$

But, now since

$$\mathbb{P}(W \le -n) = \mathbb{P}(W < -n+1) = \mathbb{P}(W > n-1)$$

and

$$\mathbb{P}(W \le -n-1) = \mathbb{P}(W < -n) = \mathbb{P}(W > n)$$

then

$$
\begin{aligned}
S(m) &= \frac{\mathbb{P}(W > x)}{\lambda} \sum_{n=1}^{|m|} (-1)^{n+1} \left[\frac{I_{|m|+1}}{I_{n+1}} H(n) - \frac{I_{|m|}}{I_n} H(n-1) \right] \\
&= \frac{\mathbb{P}(W > x)}{\lambda} \sum_{n=0}^{|m|-1} (-1)^{n} \left[\frac{I_{|m|+1}}{I_{n+2}} H(n+1) - \frac{I_{|m|}}{I_{n+1}} H(n) \right].
\end{aligned}
$$

On the other hand, if $x \leq m-1$ then

$$\begin{aligned} S(m) &= \frac{\mathbb{P}(W \leq x)}{\lambda} \sum_{n=m}^{-1} (-1)^{n+1} \left\{ \frac{I_{m-1}}{I_{n-1}} \frac{\mathbb{P}(W > n-1)}{\mathbb{P}(W = n)} - \frac{I_m}{I_n} \frac{\mathbb{P}(W > n)}{\mathbb{P}(W = n+1)} \right\} \\ &= \frac{\mathbb{P}(W \leq x)}{\lambda} \sum_{n=1}^{|m|} (-1)^{n+1} \left\{ \frac{I_{|m|+1}}{I_{n+1}} \frac{\mathbb{P}(W > -n-1)}{\mathbb{P}(W = n)} - \frac{I_{|m|}}{I_n} \frac{\mathbb{P}(W > -n)}{\mathbb{P}(W = n-1)} \right\}. \end{aligned}$$

Since

$$\mathbb{P}(W > -n) = \mathbb{P}(W \geq -n+1) = \mathbb{P}(W \leq n-1)$$

and

$$\mathbb{P}(W > -n-1) = \mathbb{P}(W \geq -n) = \mathbb{P}(W \leq n)$$

then

$$\begin{aligned} S(m) &= \frac{\mathbb{P}(W \leq x)}{\lambda} \sum_{n=1}^{|m|} (-1)^{n+1} \left\{ \frac{I_{|m|+1}}{I_{n+1}} G(n) - \frac{I_{|m|}}{I_n} G(n-1) \right\} \\ &= \frac{\mathbb{P}(W \leq x)}{\lambda} \sum_{n=0}^{|m|-1} (-1)^{n} \left\{ \frac{I_{|m|+1}}{I_{n+2}} G(n+1) - \frac{I_{|m|}}{I_{n+1}} G(n) \right\} \\ &= \frac{\mathbb{P}(W \leq x)}{\lambda} \sum_{n=1,3,\ldots}^{|m|-1} \left\{ \frac{I_{|m|+1}}{I_{n+2}} G(n+1) \right. \\ &\qquad \left. - \frac{I_{|m|}}{I_{n+1}} G(n) - \frac{I_{|m|+1}}{I_{n+1}} G(n) + \frac{I_{|m|}}{I_n} G(n-1) \right\}. \end{aligned}$$

Note that for $x \leq m-1 \leq -1$ we have

$$\begin{aligned} \mathbb{P}(W \leq x) &\leq \mathbb{P}(W \leq m-1) \\ &= \mathbb{P}(W < m) \\ &= \mathbb{P}(W > |m|) \\ &\leq \mathbb{P}(W > n) \end{aligned}$$

which means that

$$
\begin{aligned}
|S(m)| &\le \frac{\mathbb{P}(W \le x)}{\lambda} \sum_{n=1,3,\ldots}^{|m|-1} \left| \frac{I_{|m|+1}}{I_{n+2}} G(n+1) - \frac{I_{|m|}}{I_{n+1}} G(n) - \frac{I_{|m|+1}}{I_{n+1}} G(n) + \frac{I_{|m|}}{I_n} G(n-1) \right| \\
&\le \frac{\mathbb{P}(W \le x)}{\lambda} \left\{ \sum_{n=1,3,\ldots}^{|m|-1} G(n) \left| \frac{I_{|m|+1} I_n}{I_{n+1} I_{n+2}} - \frac{I_{|m|}}{I_{n+1}} - \frac{I_{|m|+1}}{I_{n+1}} + \frac{I_{|m|}}{I_n} \right| \right. \\
&\quad \left. + \sum_{n=1,3,\ldots}^{|m|-1} \left| \frac{I_{|m|+1}}{I_{n+2}} - \frac{I_{|m|}}{I_{n-1}} \right| \right\} \\
&\le \sum_{n=1,3,\ldots}^{|m|-1} H(n) \left| \frac{I_{|m|+1} I_n}{I_{n+1} I_{n+2}} - \frac{I_{|m|}}{I_{n+1}} - \frac{I_{|m|+1}}{I_{n+1}} + \frac{I_{|m|}}{I_n} \right| + \frac{1}{\lambda} \sum_{n=1,3,\ldots}^{|m|-1} \left| \frac{I_{|m|+1}}{I_{n+2}} - \frac{I_{|m|}}{I_{n-1}} \right|
\end{aligned}
$$

At this point we can use all of the same results from $m > 0$ identically.

6.2 Technical Details of Section 3.1

6.2.1 Introduction

Here we discuss our novel method for dynamic community detection using the *Dynamic Plex Percolation*, as defined in Chapter 3. We make use of the following definition of a plex.

Definition: A plex of order (n, k) is a vertex-induced subgraph S of n vertices from G with the property that

$$\deg(v) \ge n - k \text{ for all } v \in S. \tag{6.15}$$

The following theorem is a core ingredient for the DPP's functionality.

Theorem: A vertex induced subgraph S of $m < n$ vertices of a (n, k) order plex P is a order (m, k) plex.

Proof: We need only to consider $m = n - 1$ since lesser values of m will follow by induction. Let S be a vertex induced subgraph resulting from the removal of one vertex in P. Now, note that the degree of each of the vertices in S will have decreased by at most one. This means that

$$deg\,(v) \geq (n-1) - k \text{ for all } v \in S. \tag{6.16}$$

Since $m = n - 1$ it follows that S is an order $(n-1, k)$ plex.

6.2.2 Static Communities

The static component of DPP is explained first since, ultimately, it is the foundation upon which the dynamic component of DPP is built. We begin with a static network graph G_t and determine the static communities according to the following algorithm. Firstly, we identify the maximal k-plexes. That is, we enumerate the vertex induced subgraphs $S_1, S_2, \ldots$ such that each S_j is an order (n_j, k) plex and if any other vertex is added to S_j it would cease to be a plex.

Secondly, we create a secondary graph G^* in which each vertex represents one of the S_j plexes. An edge between the vertices v_i and v_j representing the plexes S_i and S_j exists in G^* if and only if the number of vertices common to S_i and S_j is no less than $k-1$, i.e.

$$|S_i \cap S_j| \geq k - 1. \tag{6.17}$$

The meaning of an edge between S_i and S_j is that the intersection $S_i \cap S_j$ is a $k-1$-plex. Consider the graph in Figure 6.2.2 and note that the upper and lower "squares" are order $(4, 2)$ plexes with an intersection of 3 vertices. This means that, if G were to be the

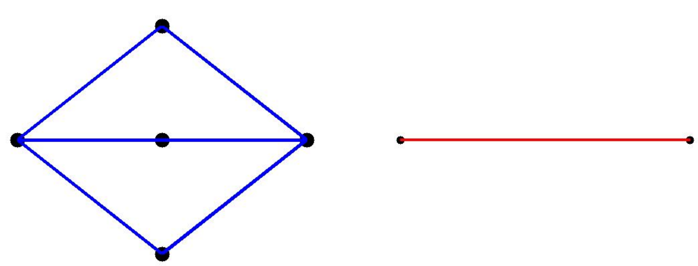

Figure 6.1: (Left) Order $(4, 2)$ Plexes in G and (Right) Vertices in G^* Representing S_1 and S_2 in G

graph in Figure 6.2.2, we would define S_1 and S_2 to be the upper and lower "squares", respectively. We then create an auxillary graph $G^\dagger$ with two vertices, one representing S_1 and one representing S_2. In this auxillary graph we connect the two vertices with an edge since S_1 and S_2 have $4 - 1 = 3$ vertices in common. This implies that with an order $(4, 2)$ plex as our template we are able to find an order $(3, 2)$ plex in the intersection between S_1 and S_2.

Lastly, we enumerate the connected components in G^* as $C_1^*, C_2^*, \ldots$ and apply labels $A, B, \ldots$ correspondingly. We percolate these labels back to the **edges** in the original graph G. To do so, we recall the edges that belong to the plex S_j which belong to the connected component C_h^* and label each of these edges with the corresponding label. It is important to note that an edge may have multiple labels. For example, in Figure 6.2.2, the purple edge belongs to both the blue and red order $(4, 1)$ communities.

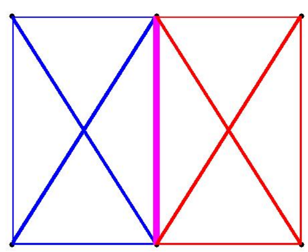

Figure 6.2: Order $(4, 1)$ Communities with Overlap

6.2.3 Dynamic Communities

Now that we have explained the static component of the DPP, we explain the dynamic component of the algorithm. The fundament goal of the dynamic component is to allow for communites to percolate through time in an explicit manner. To allow such percolation three stages exist in the process.

Let us suppose that we have a series of network graphs $\{G_t\}_{t=1}^n$ each of which contains p vertices. We begin by constructing $n - 1$ new graphs which are formed on $2p$ vertices. A new graph $G_t^\dagger$ consists of copies of G_t and G_{t+1} wherein the vertices are labeled $v^{(t)}$ for the vertex copies from G_t and $v^{(t+1)}$ for the vertex copies from G_{t+1}.

Now, we automatically connect with an edge vertices $v_k^{(t)}$ and $v_k^{(t+1)}$ in $G_t^{\dagger}$. We add artificial edges to $G_t^{\dagger}$ according to the rule that if $\left(v_i^{(t)}, v_j^{(t)}\right)$ is an edge in G_t and $\left(v_i^{(t+1)}, v_j^{(t+1)}\right)$ is an edge in G_{t+1} then in $G_t^{\dagger}$ construct the edges $\left(v_i^{(t)}, v_j^{(t+1)}\right)$ and $\left(v_i^{(t+1)}, v_j^{(t)}\right)$. This has the effect of providing a means, or rail-road tracks, for the community, i.e. template, to percolate in time. Figure 6.2.3 illustrates the above. Care must be taken to be sure that a

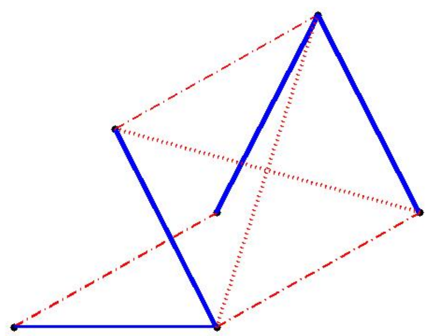

Figure 6.3: Construction of $G_t^{\dagger}$. In this example we see that nodes in G_t are automatically connected immediately to their counterpart in G_{t+1}. Then since the North-Eastern edge persists, the "railroad tracks" are created to the plex to percolate whereas, for the other two vertex pairs, the edge (or non-edge) did not persist so no railroad tracks are constructed.

community is not declared in the enhanced times series of graphs solely as a result of the added edges but this requires only a trivial conditional statement.

In the end, we must match the labels of $G_t^{\dagger}$ and $G_{t+1}^{\dagger}$. Such an endeavor requires little effort. We simply maintain a stack of the natural numbers and assign to each static community in G_1 a number and remove these numbers from the stack and match these numbers to the communities in G_2, that is, we have identified the communities in $G_1^{\dagger}$. We repeat this process and match corresponding labels in G_2 and G_3, i.e. $G_2^{\dagger}$. In this way, we have percolated the labels throughout the entire sequence $\{G_t\}_{t=1}^{n}$.

6.2.4 Pseudocode

This section contains the pseudocode of the algorithm. We begin by formulating the maximal plex enumeration phase of the algorithm. Suppose that A is an $p \times p$ adjacency matrix representing a graph G and we wish to enumerate all maximal order k plexes. We initialize the sets $R = \{\}$, $P = \{1, 2, \ldots, n\}$, $X = \{\}$ and a vector (in the data structure sense) MC. The set R at any stage in the algorithm is a currently growing k plex, the set P are prospective vertices that may be added to those in R to enlarge the growing plex, and the set X are the set of vertices which have already been processed during the recursive algorithm's execution. Finally, the vector MP will contain sets of integer labels for the vertices belonging to a plex. That is, the k^{th} element of MP is a set which contains labels for the vertices which belong to the k^{th} discovered maximal plex. We move now onto the Algorithm 1. We will suppose throughout that MC is a global variable and refer to indices and vertices interchangably. The *Plex* algorithm described in Algorithm 1 is

Algorithm 1 Algorithm for Returning Maximal k plexes

Input: R, P, X, sets of integers and k, an integer
Output: MP, a vector of sets of integers
Plex (R, P, X, k)
if $P = \phi$ and $X = \phi$ and $|R| > k$ **then**
  push back R onto MP
else
  for every $u \in P$ **do**
    define $Rnew$ to be $R \cup u$
    remove u from P
    define Nu to be the vertices in G which are connected to **at least** $|Rnew| - k + 1$ vertices in $Rnew$
    define $crit$ to be the vertices in $Rnew$ which are connected to **exactly** $|Rnew| - k$ vertices in $Rnew$
    if $|crit|$¿0 **then**
      redefine Nu to be the intersection of Nu and the the vertices which are neighbors to all vertices in $crit$
    end if
    Call *Plex* $(Rnew, P \cap Nu, X \cap Nu, k)$
    redefine $X = X \cup u$
  end for
end if

an adaptation of the maximal clique enumeration algorithm first introduced by [BK73]. Now that all maximal plexes of order k have can enumerated for any graph G it needs to be determined which maximal plexes are merged to form the dynamic communities communities. First, however, we describe the static component of the algorithm. Now

---

Algorithm 2 Static Community Detection

Input A, an adjacency matrix and MP, a vector of sets of integers
Output B, a matrix of sets of integers
$StatComm\,(A, MP)$
define n to be $|MC|$
define G^* to be a graph with n vertices enumerated according to the plexes in $|MC|$
compute the connected components of G^*, via breadth-first search, for example
define $C_1^*, C_2,^*, \ldots$ to be the connected components of G^*, of which there are m.
assign labels $1, 2, \ldots$ to $C_1^*, C_2^*, \ldots$ accordingly.
Set B to be a $p \times p$ matrix of sets of integers
for every $k = 1, 2, \ldots, m$ **do**
    for every $MP_j \in C_k^*$ **do**
        for every $i = 1, 2, \ldots, |MP_j| - 1$ **do**
            for every $h = i + 1, \ldots, |MP_j|$ **do**
                if $A\,(i, h) = 1$ **then** set $B\,(i, h)) = B\,(i, h) \cup k$
                end if
            end for
        end for
    end for
end for

---

that the static portion of the code has been described, one must transition to the dynamic phase of the algorithm. At this point, all that remains is label matching. This chore is mundane and does not require any specific details other than if, for example there exist three times points $t = 1, 2, 3$, for example, then we have G_1, G_2, G_3 and $G_1^\dagger, G_2^\dagger$. The overlap in $G_1^\dagger$ and $G_2^\dagger$ is G_2 and, therefore, the dynamic communities community labels need to be matched through the the labelling in timepoint $t = 2$. This is simply a matter of keeping track of which labels coorespond from timestep to timestep.

Algorithm 3 Dynamic Community Detection

Input: A_1, A_2, $p \times p$ adjacency matrices and k, an integer
Output: C, a $p \times p$ matrix of sets of integers
$DynComm\,(A_1, A_2)$
define A as a matrix with the upper $p \times p$ as A_1 and the lower $p \times p$ as A_2
for every $k = 1, 2, \ldots, p$ **do** set $A\,(k, p+k) = 1$ set $A\,(p+k, k) = 1$
end for
for every $i = 1, 2, \ldots, p-1$ **do**
    for every $j = (i+1), \ldots, p$ **do**
        if $A_1\,(i, j) = 1$ and $A_2\,(i, j) = 1$ **then** set $A\,(i, p+j) = 1$ set $A\,(j, p+1) = 1$ set $A\,(p+j, i) = 1$ set $A\,(p+i, j) = 1$
        end if
    end for
end for
set C to be the result of executing $StatComm$ on B.

Bibliography

[AL78] Elja Arjas and Tapani Lehtonen, *Approximating many server queues by means of single server queues*, Mathematics of Operations Research **3** (1978), no. 3, 205–223.

[ALA10] Antiqueira L, Rodrigues FA , van Wijk BC, Costa L da F, and Daffertshofer A, *Estimating complex cortical networks via surface recordings—A critical note*, NeuroImage **53** (2010), no. 2, 439–449.
https://doi.org/10.1016/j.neuroimage.2010.06.018

[And71] T. W. Anderson, *An introduction to multivariate statistical analysis*, second ed., John Wiley and Sons, New York, 1971.

[AS72] Milton Abramowitz and Irene A. Stegun, *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, Dover Publications, New York, 1972.

[AWP] Waugh, AS, Pei, L, Fowler, JH, Mucha, PJ, and Porter, MA. *Party polarization in congress: a network science approach*. ArXiv, (2011).
https://arxiv.org/abs/0907.3509v3

[Bar13] Arpad Baricz, *Bounds for Turánians of modified Bessel functions*, ArXiv (2013). https://arxiv.org/pdf/1311.1450

[Bas13] Bassett, DS et al. *Robust detection of dynamic community structure in networks*, Chaos **23** (2013), no. 1, 013142. https://doi.org/10.1063/1.4790830

[BB99] Wilson Fried Bragin, Engel and Buzsa, *High-frequency oscillations in human brain*, Hippocampus **9** (1999), no. 2, 137–142. https://doi.org/10.1002/(sici)1098-1063(1999)9:2<137::aid-hipo5>3.0.co;2-0

[BB09] Danielle S. Bassett and Edward T. Bullmore, *Human brain networks in health and disease*, Current Opinion in Neurology **22** (2009), no. 4, 340–347.

[BC05] A. D. Barbour and Louis H. Y. Chen, *An introduction to stein's method*, Singapore University Press, New Jersey, 2005.

[BF] Bartolomei F, Wendling F, Régis J, Gavaret M, Guye M, and Chauvel P. *Pre-ictal synchronicity in limbic networks of mesial temporal lobe epilepsy*, Epilepsy Research **61** (2004), no. 1–3, 89–104.
https://doi.org/10.1016/j.eplepsyres.2004.06.006

[BHJ92] A. D. Barbour, Lars Holst, and Svante Janson, *Poisson approximation*, Oxford, New York, 1992.

[BK73] Coen Bron and Joep Kerbosch, *Finding all cliques of an undirected graph*, Proceedings of the ACM **16** (1973), no. 9, 575–579.

[Bol85] Bela Bollobas, *Random graphs*, Academic Press, Boston, 1985.

[Bol98] Bela Bollobas, *Modern graph theory*, Springer Verlag, 1998.

[BSK10] Bialonski S, Horstmann M-T and Lehnertz K, *From brain to earth and climate systems: small-world interaction networks or not*? Chaos **20** (2010), no. 1, 013134. https://doi.org/10.1063/1.3360561

[DA09] Joel Spencer, Dimitris Achlioptas, Raissa M. D'Souza, *Explosive percolation in random networks*, Science **323** (2009), 1453–1455.

[dCG] de Curtis, M and Gnatkovsky, V. *Reevaluating the mechanisms of focal ictogenesis: The role of low-voltage fast activity*, Epilepsia **50** (2009), no. 12, 2514–2525. https://doi.org/10.1111/j.1528-1167.2009.02249.x

[ER60] P. Erdos and A Renyi, *On the evolution of random graphs*, Publication of the Mathematical Institute of the Hungarian Academy of Sciences, 1960, pp. 17–61.

[GF] Gibbs F, Gibbs E, Lennox W. *Epilepsy: a paroxysmal cerebral dysrhythmia*, Epilepsy & Behavior **3** (2002), no. 4, 395–401. https://doi.org/10.1016/s1525-5050(02)00050-1

[GP07] Tama Vicsek Gergely Palla, Alber-Laszlo Barabasi, *Quantifying social group evolution*, Nature **446** (2007), 664–667.

[JDW06] Duncan JS, Sander JW, Sisodiya SM, and Walker MC. *Adult epilepsy,* Lancet **367** (2006), no. 9516, 1087–1100. https://doi.org/10.1016/s0140-6736(06) 68477-8

[Jer05] Jerger K, Weinstein SL, Sauer T, and Schiff SJ., *Multivariate linear discrimination of seizures*, Clinical Neurophysiology **116** (2005), no. 3, 545–551. https://doi.org/10.1016/j.clinph.2004.08.023

[K.94] Friston K, *Functional and effective connectivity in neuroimaging: a synthesis*, Human Brain Mapping **2** (1994), no. 1–2, 56–78. https://doi.org/10.1002/hbm.460020107

[KC12] Kramer MA, and Cash SS. *Epilepsy as a disorder of cortical network organization,* Neuroscientist **18** (2012), no. 4, 360–372. https://doi.org/10.1177/1073858411422754

[Kol09] Eric D. Kolaczyk, *Statistical analysis of network data*, Springer, New York, 2009.

[LVZ05] Zhang, LV, King OD, Wong SL et al., *Motifs, themes and thematic maps of an integrated saccharomyces cerevisiae interaction network*, Journal of Biology **4** (2005), no. 2, 6. https://doi.org/10.1186/jbiol23

[MAK08] Kramer MA, Kolaczyk ED, and Kirsch HE., *Emergent network topology at seizure onset in humans*, Epilepsy Research **79** (2008), 173–186. https://doi.org/10.1016/j.eplepsyres.2008.02.002

[MAK10] Eric D. Kolaczyk et al. Mark A. Kramer, Uri T. Eden, *Coalescence and fragmentation of cortical networks during focal seizures*, The Journal of Neuroscience **30** (2010), no. 30, 10076–10085.

[neo90] *Surgery for epilepsy*, Consensus Statement. National Institutes of Health Consensus Development Conference **8** (1990), no. 2, 1–20. PMID: 2203963

[NT] Netoff TI and Schiff SJ. *Decreased neuronal synchronization during experimental seizures*, Journal of Neuroscience, **22** (2002), no. 16, 7297–7307. https://doi.org/10.1523/jneurosci.22-16-07297.2002

[NTJ04] Arno S Keck T Netoff TI, Clewley R and White JA., *Epilepsy in small-world networks*, Journal of Neuroscience, **24** (2004), no. 37, 8075–8083.

[PB13] Balachandran P, Kolaczyk E, and Viles W. *Exponential-type inequalities involving ratios of the modified Bessel function of the first kind and their applications.* (2013). https://doi.org/10.48550/arXiv.1311.1450

[PBH06] Robert Devaney Paul Blanchard and Glenn Richard Hall, *Differential equations*, Thomson-Brooks, 2006.

[PJ54] Penfield W, and Jasper H. *Epilepsy and the Functional Anatomy of the Human Brain*. Little, Brown, Boston, 1954.

[PJ11] Cohen AL et al. Power JD, *Functional network organization of the human brain*, Neuron **72** (2011), no. 4, 665–678.

[PJMO10] Mucha PJ, Richardson T, Macon K, et al. *Community structure in time-dependent, multiscale, and multiplex networks*, Science **328** (2010), no. 5980, 876–878. https://doi.org/10.1126/science.1184819

[PSC07] Ponten SC, Bartolomei F, and Stam CJ. *Small-world networks and epilepsy: graph theoretical analysis of intracerebrally recorded mesial temporal lobe seizures,* Clinical Neurophysiology, **118** (2007), no. 4, 918–927. https://doi.org/10.1016/j.clinph.2006.12.002

[RM02] Milo R, Shen-Orr S, Itzkovitz S, et al. *Network motifs: Simple building blocks of complex networks,* Science **298** (2002), no. 5594, 824–827. https://doi.org/10.1126/science.298.5594.824

[RW11] Riordan O and Warnke L, *Explosive percolation is continuous*, Science **333** (2011), no. 6040, 322–324. https://doi.org/10.1126/science.1206241

[SK] Elger CE Lehnertz K Schindler K, Leung H. *Assessing seizure dynamics by analysing the correlation structure of multichannel intracranial EEG*, Brain, **130** (2007), pt. 1, 65–77. https://doi.org/10.1093/brain/awl304

[TYBW10] Berger-Wolf, T., Tantipathananandh, C., Kempe, D. (2010). *Dynamic Community Identification*. In Yu, P., Han, J., Faloutsos, C. (eds.) Link Mining: Models, Algorithms, and Applications. Springer, New York. https://doi.org/10.1007/978-1-4419-6515-8_12

[Wat99] Duncan J. Wattz, *Small worlds*, Princeton University Press, Princeton, 1999.

[WHX06] Wu H, Li X and Guan X, *Networking property during epileptic seizure with multi-channel EEG recordings*. In: Wang, J., Yi, Z., Zurada, J.M., Lu, BL., Yin, H. (eds.) Advances in Neural Networks - ISNN 2006. Lecture Notes in Computer Science, volume 3973. Springer, Berlin, Heidelberg. https://doi.org/10.1007/11760191_84

Curriculum Vitae

[redacted]