

**To be determined**

by

Runze Tang

A dissertation submitted to The Johns Hopkins University in conformity with the  
requirements for the degree of Doctor of Philosophy.

Baltimore, Maryland

July, 2017

© Runze Tang 2017

All rights reserved

# **Abstract**

Abstract goes here.

Primary Reader: Carey Priebe

Secondary Reader: Minh Tang

# Acknowledgments

Thanks!

# Dedication

This thesis is dedicated to ...

# Contents

<b>Abstract</b>	ii
<b>Acknowledgments</b>	iii
<b>List of Tables</b>	vii
<b>List of Figures</b>	viii
<b>1 Introduction</b>	1
1.1 Section . . . . .	1
1.1.1 Subsection . . . . .	1
1.2 Optional table of contents heading . . . . .	2
1.2.1 Another subsection . . . . .	2
1.2.1.1 Subsubsection . . . . .	2
<b>2 Random Graph Models</b>	4
2.1 Graphs and Random Graphs . . . . .	5

## CONTENTS

2.1.1	Basic Concepts of Graphs . . . . .	5
2.1.2	Random Graphs . . . . .	6
2.2	Unweighted Random Graph Models . . . . .	7
2.2.1	Independent Edge Model . . . . .	7
2.2.2	Random Dot Product Graph . . . . .	8
2.2.3	Stochastic Blockmodel . . . . .	12
2.3	Weighted Random Graph Models . . . . .	16
2.3.1	Weighted Independent Edge Model . . . . .	16
2.3.2	Weighted Random Dot Product Graph . . . . .	17
2.3.3	Weighted Stochastic Blockmodel . . . . .	18
<b>3</b>	<b>A Law of Large Graphs</b>	<b>20</b>
3.1	Another Section . . . . .	20
<b>4</b>	<b>Robust Generalizations of the Law of Large Graphs</b>	<b>21</b>
<b>5</b>	<b>Discussion</b>	<b>22</b>
<b>Vita</b>		<b>25</b>

# List of Tables

# List of Figures

- 2.1 Example illustrating the stochastic blockmodel. The parameters are given in Example 2.2.6. The left figure shows the probability matrix  $P$  with  $K = 5$  blocks and  $n = 200$  vertices and the right figure shows an adjacency matrix  $A$  sampled under  $\text{SBM}(P)$ . While  $A$  is a noisy version of  $P$ , much of the structure of  $P$  is preserved in  $A$ , a property we will exploit in our estimation procedure. . . . .

14

# Chapter 1

## Introduction

Introduction.

A citation ?. A citation without brackets ?. Multiple citations ???.

### 1.1 Section

This is a section. Here's a reference to a different section: 1.1.1.

#### 1.1.1 Subsection

This is a subsection.

**Table 1.1:** This is a caption.

A	B
a1	b1
a2	b2
a3	b3

Table 1.1 ... continued	
A	B
a4	b4

## 1.2 Section with linebreaks in the name

This is another section.

### 1.2.1 Another subsection

#### 1.2.1.1 Subsubsection

##### 1.2.1.1.1 Heading level below subsubsection

And I quote:

La la la.

No indent after end of quote.

Another paragraph with a list:

- Item 1

- Item 2

## CHAPTER 1. OPTIONAL RUNNING CHAPTER HEADING

Again, we don't indent here.

# Chapter 2

## Random Graph Models

Our main focus of this work is a collection of random graphs. As a first step, we will introduce some models for a single random graph in this chapter. These important components will not only lead to our model for a collection of graphs but also motivate our estimators later in Chapter 3 and Chapter 4.

We start the chapter with some basic concepts of graphs and random graphs in Section 2.1. Then in Section 2.2, random graph models for unweighted graphs are introduced. In particular, it is concerned with independent edge model, random dot product graph, and stochastic blockmodel. Their relationship will also be discussed. Section 2.3 generalizes all three models introduced in Section 2.2 so that they can be adapted to weighted graphs.

## 2.1 Graphs and Random Graphs

In this section we will introduce some basic concepts of graphs and random graphs.

### 2.1.1 Basic Concepts of Graphs

A *graph* is an ordered pair  $G = (V, E)$  comprising a set  $V$  of vertices together with a set  $E$  of edges. Denote the number of vertices  $|V|$  to be  $n$ . Without loss of generality, we assume  $V = [n] = \{1, 2, \dots, n\}$ . Each edge is associated with two vertices. We say a graph has no self-loops if each edge is associated with two distinct vertices. And a graph is undirected if all its edges have no orientation. Thus for a undirected graph without self-loops, we have the edge set  $E \subset \{\{u, v\} : u, v \in V, u \neq v\}$ . This will be our basic setting throughout the entire paper.

Every graph can be represented in the form of adjacency matrix  $a \in \mathcal{A}$ . For unweighted graphs,  $\mathcal{A} = \{0, 1\}^{n \times n}$  so that the adjacency matrices are binary. For  $i, j \in [n]$ ,  $a_{ij} = 1$  indicates that there is an edge from vertex  $i$  to vertex  $j$  and  $a_{ij} = 0$  otherwise. Note that  $a$  is always symmetric since we assume the graphs to be undirected in this work. Thus  $a_{ij} = 1$  means there is an edge between vertex  $i$  and vertex  $j$  and  $a_{ji} = 1$  as well. For weighted graphs, each edge is assigned with a positive real-valued weight, i.e.  $\mathcal{A} = \mathbb{R}_{\geq 0}^{n \times n}$ . Similarly,  $a_{ij} > 0$  denotes the weight assigned to the edge between vertex  $i$  and vertex  $j$ , and  $a_{ij} = 0$  means there is no edge between these two vertices.

## CHAPTER 2. RANDOM GRAPH MODELS

As mentioned above, we assume the graph has no self-loops and are undirected. So every adjacency matrix  $a$  is hollow and symmetric, that is  $a_{ii} = 0$  for  $i \in [n]$  and  $a_{ij} = a_{ji}$  for  $i, j \in [n]$ . This is always assumed without specific clarifications in later sections.

### 2.1.2 Random Graphs

A random graph is a graph with fixed vertex set whose edges are randomly distributed with respect to some distributions. Mathematically, a random graph  $A : \Omega \mapsto \mathcal{A}$  is a map from the probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  to the space of all adjacency matrices on  $n$  vertices.

**Example 2.1.1 (Erdős-Rényi Graphs)** *The first random graph model is the Erdős-Rényi graphs (ER) introduced by [Gilbert, 1959]. They are unweighted graphs where each edge is present with the same probability  $p \in [0, 1]$  independently. In our setting (undirected graphs without self-loops), that is  $A_{ij} \stackrel{iid}{\sim} \text{Bernoulli}(p)$  for  $i, j \in [n]$  with  $i < j$ . Thus for  $a \in \mathcal{A}$  we have*

$$\mathbb{P}(A = a) = \prod_{i < j} p^{a_{ij}} (1 - p)^{1 - a_{ij}}.$$

In later sections, we will introduce other random graph models which capture different properties of the graphs in practice respectively.

## 2.2 Unweighted Random Graph Models

In this section, we will focus on unweighted graphs and introduce three important models, i.e. independent edge model, random dot product graph, and stochastic blockmodel.

### 2.2.1 Independent Edge Model

As introduced in Example 2.1.1, we see ER model is quite restrictive since all edges follow the same Bernoulli distribution with parameter  $p$ . Here we consider a much more general model.

**Definition 2.2.1 (Independent Edge Model)** *Under an independent edge model (IEM) proposed by Bollobás et al. [2007], for  $i, j \in [n]$  and  $i < j$ , the edge between vertex  $i$  and vertex  $j$  is present with probability  $p_{ij} \in [0, 1]$  independently, i.e.  $A_{ij} \stackrel{\text{ind}}{\sim} \text{Bernoulli}(p_{ij})$ . Let  $P = (p_{ij})_{i,j=1}^n \in [0, 1]^{n \times n}$  be the parameter matrix consists of all probabilities for Bernoulli distributions, then the model is denoted by IEM( $P$ ). Thus for  $a \in \mathcal{A}$  we have*

$$\mathbb{P}(A = a) = \prod_{i < j} p_{ij}^{a_{ij}} (1 - p_{ij})^{1 - a_{ij}}.$$

Note that the graphs considered in this paper are always assumed to be undirected without self-loops, thus the parameter matrix  $P$  needs to be symmetric and hollow. However, for convenience, we still define the parameters to be an  $n$ -by- $n$  matrix while

## CHAPTER 2. RANDOM GRAPH MODELS

only  $\binom{n}{2}$  of them are effective.

Compared to the ER model, IEM allows the probabilities of the existence of the edges to vary while keeping the independence. Different probabilities give IEM the freedom to include a wide range of random graph distributions, but definitely not all of them because of independence assumption. The most general model is a multinomial distribution with  $2^{\binom{n}{2}}$  choices from all possible unweighted and undirected graphs with no self-loops on  $n$  vertices, which imposes no assumptions on the graph structure. While such model is instructive sometimes, IEM will be the most general model we consider in this paper.

### 2.2.2 Random Dot Product Graph

For a graph, the adjacencies between vertices generally depend on unobserved properties of the corresponding vertices. For instance, in a social network setting, people are more likely to be friends if they have shared interests; in a connectomics setting, the two brain regions with similar properties will have similar connectivity patterns compared to other regions of the brain. The latent positions model (LPM) proposed by [Hoff et al., 2002] captures such structure, where each vertex is associated with a latent position that influences the adjacencies for that vertex. In particular, we are interested in the case that the latent positions are random in this work.

**Definition 2.2.2 (Latent Position Model)** *Let the latent position for vertex  $i$  be*

## CHAPTER 2. RANDOM GRAPH MODELS

a random variable  $X_i : \Omega \mapsto \mathcal{X}$ , where  $\mathcal{X}$  is the latent space and  $i \in [n]$ . Let the link function be a symmetric map  $\kappa : \mathcal{X}^2 \mapsto [0, 1]$ . Then under a latent position model (LPM), for  $i, j \in [n]$  and  $i < j$ , conditioned on their respective latent positions  $X_i$  and  $X_j$ ,

$$A_{ij}|(X_i, X_j) \stackrel{\text{ind}}{\sim} \text{Bernoulli}(\kappa(X_i, X_j)).$$

Thus for  $a \in \mathcal{A}$  we have

$$\mathbb{P}(A = a | X_1, \dots, X_n) = \prod_{i < j} \kappa(X_i, X_j)^{a_{ij}} (1 - \kappa(X_i, X_j))^{1-a_{ij}}.$$

Let  $X_1, \dots, X_n \stackrel{iid}{\sim} F$  for some distribution  $F$  on  $\mathcal{X}$ , then the model is denoted by  $\text{LPM}(\mathcal{X}, F, \kappa)$ .

Among all the LPMs, we are particularly interested in the random dot product graph (RDPG) [Nickel, 2007, Young and Scheinerman, 2007].

**Definition 2.2.3 (Random Dot Product Graph)** Let the latent position for vertex  $i$  be a random variable  $X_i : \Omega \mapsto \mathcal{X} \subset \mathbb{R}^d$ , where  $i \in [n]$  and  $\mathcal{X}$  is the latent space such that  $x^\top y = \sum_{i=1}^d x_i y_i \in [0, 1]$  for  $x, y \in \mathcal{X}$ . Then under a random dot product graph (RDPG), for  $i, j \in [n]$  and  $i < j$ , conditioned on their respective latent positions  $X_i$  and  $X_j$ ,

$$A_{ij}|(X_i, X_j) \stackrel{\text{ind}}{\sim} \text{Bernoulli}(X_i^\top X_j).$$

## CHAPTER 2. RANDOM GRAPH MODELS

Thus for  $a \in \mathcal{A}$  we have

$$\mathbb{P}(A = a | X_1, \dots, X_n) = \prod_{i < j} (X_i^\top X_j)^{a_{ij}} (1 - X_i^\top X_j)^{1-a_{ij}}.$$

Let  $X_1, \dots, X_n \stackrel{iid}{\sim} F$  for some distribution  $F$  on  $\mathcal{X}$ , then the model is denoted by RDPG( $\mathcal{X}, F$ ). Typically, we write all latent positions together as  $X = [X_1, \dots, X_n]^\top \in \mathbb{R}^{n \times d}$ , where the  $i$ -th row  $X_i^\top$  corresponds to the latent position for vertex  $i$ .

From the definition above, we see RDPG is actually a special case of LPM with latent space  $\mathcal{X} \subset \mathbb{R}^d$  such that  $x^\top y \in [0, 1]$  for  $x, y \in \mathcal{X}$  and link function  $\kappa(x, y) = x^\top y$ . If  $d$  is much smaller than the number of vertices  $n$ , which is likely to be the case in practice, RDPG is then a more parsimonious model compared to IEM, requiring only  $n \cdot d$  parameters rather than  $\binom{n}{2}$ .

The direction and magnitude of the latent position, which are determined by properties of the corresponding vertex, are the most important factors in RDPG due to the nature of dot product. Vertices with latent positions pointing in similar directions are more likely to have an edge between them compared to those with different directions. Similarly, the magnitude of the latent positions encodes the vertices' overall tendency to form edges. A larger magnitude potentially leads to more edges incident with the vertex.

Conditioned on the latent positions  $X$ , the RDPG now can be considered to be an IEM( $P$ ) with  $P = XX^\top$ , i.e. an edge between vertex  $i$  and vertex  $j$  is present with

## CHAPTER 2. RANDOM GRAPH MODELS

probability  $P_{ij} = X_i^\top X_j$ . Note that the probability matrix  $p$  is the outer product of the latent position matrix  $X$  with itself. This imposes two important properties on  $P$  under RDPG, namely that  $P$  is positive semidefinite (PSD) and  $\text{rank}(P) = \text{rank}(X) \leq d$ . On the other hand, this also suggests that for certain circumstances when the probability matrix  $P$  might not be positive semi-definite, one may want to use some other LPM which preserves the low-rank property instead.

**Example 2.2.4** Consider a LPM with link function  $\kappa(x, y) = x^\top Ky$  where  $K \in \{-1, 0, 1\}^{d \times d}$  is a diagonal matrix with  $K_{ii} = 1$  for  $i \leq d'$  and  $K_{ii} = -1$  otherwise for  $1 < d' < d$ . Unlike RDPG, this LPM can be applied to the situation where the low-rank probability matrix  $P$  is indefinite. For example  $P \in [0, 1]^{n \times n}$  with  $d$  non-zero eigenvalues  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{d'} > 0 > \lambda_{d'+1} \geq \dots \geq \lambda_d$ .

The LPM in Example 2.2.4 above is a natural extension of RDPG to the non-PSD case. We will see later that it motivates another estimator analogous to the one we are going to propose. Details are discussed in Remark ??.

As one may notice, RDPG is non-identifiable. For any orthonormal matrix  $W \in \mathbb{R}^{d \times d}$ , the rotated latent positions  $XW$  is equivalent to the original  $X$  since  $P = XX^\top = (XW)(XW)^\top$ . In later sections, we will sometimes consider the equivalent class of latent positions up to rotations instead.

Importantly, since  $P$  is the outer product of the latent positions  $X$  in RDPG, it also motivates the low-rank estimator based on spectral decomposition which will be discussed in details later.

### 2.2.3 Stochastic Blockmodel

One of the most important structures for graphs is the community structure in which vertices are clustered into different communities such that vertices of the same community behave similarly. This structural property is captured by the stochastic blockmodel (SBM) [Holland et al., 1983], where each vertex is assigned to a block and the probability that an edge exists between two vertices depends only on their respective block memberships.

**Definition 2.2.5 (Stochastic Blockmodel)** Consider a  $K$ -block stochastic blockmodel (SBM) with block probability matrix  $B \in [0, 1]^{K \times K}$ , and the vector of block memberships  $\tau \in [K]^n$ , where for each  $i \in [n]$ ,  $\tau_i = k$  means vertex  $i$  is a member of block  $k$ . We have for  $i, j \in [n]$  and  $i < j$ ,

$$A_{ij} \stackrel{ind}{\sim} \text{Bernoulli}(B_{\tau_i, \tau_j}).$$

Thus for  $a \in \mathcal{A}$  we have

$$\mathbb{P}(A = a) = \prod_{i < j} B_{\tau_i, \tau_j}^{a_{ij}} (1 - B_{\tau_i, \tau_j})^{1-a_{ij}}.$$

Such model is denoted by  $\text{SBM}(\tau, B)$ .

In some scenarios, the block probability matrix  $B$  and block membership vector  $\tau$  in SBM are assumed to be random. For example, each vertex can be assigned to

## CHAPTER 2. RANDOM GRAPH MODELS

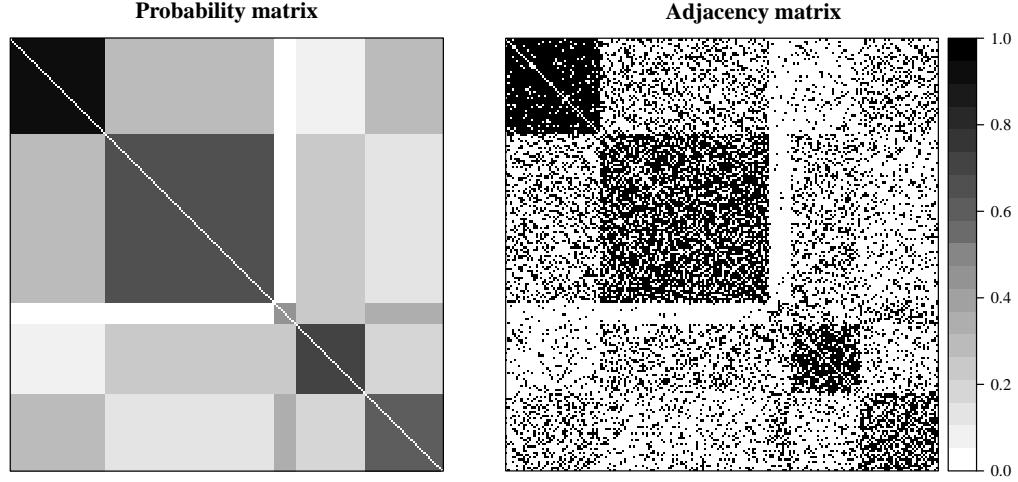
a block independently according to a probability vector  $\rho \in (0, 1)^K$  with  $\sum_{k=1}^K \rho_k = 1$  such that  $\mathbb{P}(\tau_i = k) = \rho_k$ . Nevertheless, the definition above still holds when conditioning on  $B$  and  $\tau$ . And then the SBM now can be considered to be an IEM( $P$ ) with  $P = B_{\tau, \tau}$ , i.e. an edge between vertex  $i$  and vertex  $j$  is present with probability  $P_{ij} = B_{\tau_i, \tau_j}$ .

**Example 2.2.6** Consider a 5-block SBM with

$$B = \begin{bmatrix} 0.90 & 0.27 & 0.05 & 0.10 & 0.30 \\ 0.27 & 0.67 & 0.02 & 0.26 & 0.14 \\ 0.05 & 0.02 & 0.44 & 0.25 & 0.33 \\ 0.10 & 0.26 & 0.25 & 0.70 & 0.18 \\ 0.30 & 0.14 & 0.33 & 0.18 & 0.58 \end{bmatrix}, \quad \rho = \begin{bmatrix} 0.22 & 0.39 & 0.05 & 0.16 & 0.18 \end{bmatrix}.$$

We sample a graph with  $n = 200$  vertices under this SBM and plot the corresponding probability matrix  $P = B_{\tau, \tau}$  and the adjacency matrix  $A$  in Figure 2.1. While  $A$  is a noisy version of  $P$ , the structure of 25 blocks can be seen clearly in both figures as a result of 5 different blocks among vertices.

As discussed in Section 2.2.2, the probability matrix  $P$  in an RDPG is positive semidefinite. And now we argue that an SBM with a positive semidefinite  $B$  can always be parameterized as an RDPG. Firstly due to the positive semidefiniteness of  $B$ , we can decompose  $B = \nu\nu^\top$  where  $\nu \in \mathbb{R}^{K \times d}$ . Define  $\nu_1, \dots, \nu_K$  such that the rows



**Figure 2.1:** Example illustrating the stochastic blockmodel. The parameters are given in Example 2.2.6. The left figure shows the probability matrix  $P$  with  $K = 5$  blocks and  $n = 200$  vertices and the right figure shows an adjacency matrix  $A$  sampled under  $\text{SBM}(P)$ . While  $A$  is a noisy version of  $P$ , much of the structure of  $P$  is preserved in  $A$ , a property we will exploit in our estimation procedure.

of  $\nu$  are given by  $\nu_1^\top, \dots, \nu_K^\top$ . Then  $\nu_k$  can be regarded as the shared latent position for all vertices assigned to block  $k$ . Define  $X = [X_1, \dots, X_n]^\top = [\nu_{\tau_1}, \dots, \nu_{\tau_n}]^\top \in \mathbb{R}^{n \times d}$ .

Then the SBM now can be parameterized as an RDPG since

$$\mathbb{P}(A_{ij} = 1) = B_{\tau_i, \tau_j} = \nu_{\tau_i}^\top \nu_{\tau_j} = X_i^\top X_j \in [0, 1].$$

**Definition 2.2.7 (SBM as RDPG)** Consider an  $\text{RDPG}(\mathcal{X}, F)$  where  $F$  is a distribution on  $K$  point masses  $\nu_1, \dots, \nu_K \in \mathcal{X}$  such that  $\mathbb{P}(X_i = \nu_k) = \rho_k$  for  $i \in [n]$  and  $k \in [K]$ , where the block proportion vector  $\rho \in (0, 1)^K$  with  $\sum_{k=1}^K \rho_k = 1$ . Then we denote this model by  $\text{SBM}(\rho, B)$  where  $B = \nu^\top \nu$  with  $\nu = [\nu_1, \dots, \nu_K]^\top$ . Similarly, we can define  $\tau \in [K]^n$  such that  $X_i = \nu_{\tau_i}$ .

## CHAPTER 2. RANDOM GRAPH MODELS

For notational convenience we will refer to the sub-model of SBM with positive semidefinite  $B$  as *SBM*. As shown above, the SBM can be regard as an RDPG where all vertices in the same block have identical latent positions. In this work, we will always analyze SBM in an RDPG setting.

**Remark 2.2.8** *As we mentioned, under the SBM, all vertices in the same block have identical latent positions. Rather than allowing vertices differ from each other as RDPG, SBM presumes all nodes within the same block have the same expected degree. To better describe complex networks in some situations, a bunch of generalizations of the SBM have been explored in order to incorporate the local variation of vertices to the block structure. Airoldi et al. [2008] proposed mixed membership stochastic blockmodels, which associates each vertex with multiple blocks with a probability vector rather than a single block as SBM requires. Also, in order to model variation of the expected degrees of different vertices within the same block, Karrer and Newman [2011] proposed degree-corrected SBM, which assigns additional parameters to each vertex to adjust the expected degree relatively. All these generalizations are trying to drag SBM towards RDPG a little bit so that the model can capture variations among vertices while keeping the community structure.*

## 2.3 Weighted Random Graph Models

In this section, we shift our focus to the case where graphs are weighted and generalize the three models introduced in Section 2.2 respectively, i.e. the weighted independent edge model (WIEM) in Section 2.3.1, the weighted random dot product graph model (WRDPG) in Section 2.3.2, and the weighted stochastic blockmodel (WSBM) as a WRDPG in Section 2.3.3.

As a reminder, for weighted graphs, each edge is assigned with a positive real-valued weight, i.e.  $\mathcal{A} = \mathbb{R}_{\geq 0}^{n \times n}$ . So the adjacency matrices are not binary any more.

### 2.3.1 Weighted Independent Edge Model

As introduced in Section 2.2.1, under IEM( $P$ ), for  $i, j \in [n]$  and  $i < j$ , the edge weight  $A_{ij}$  is distributed from a Bernoulli distribution with parameter  $P_{ij}$  independent of other edges. We first extend IEM to weighted independent edge model (WIEM).

**Definition 2.3.1 (Weighted Independent Edge Model)** *Consider a one-parameter family of distributions  $\mathcal{F} = \{f_\theta : \theta \in \Theta \subset \mathbb{R}\}$ . Let the graph parameters be a matrix  $P \in \Theta^{n \times n} \subset \mathbb{R}^{n \times n}$ . Then under a weighted independent edge model (WIEM) with respect to  $\mathcal{F}$ , for  $i, j \in [n]$  and  $i < j$ , the edge weight between vertex  $i$  and vertex  $j$  is distributed according to  $f_{P_{ij}}$  independently.*

To see that an IEM is a special case of WIEM, let  $\mathcal{F}$  be the collection of Bernoulli distributions and let the graph parameters be a symmetric and hollow matrix  $P \in$

## CHAPTER 2. RANDOM GRAPH MODELS

$[0, 1]^{n \times n}$ . Note that the graphs considered in this paper are undirected without self-loops, thus the parameter matrix  $P$  needs to be symmetric and hollow. However, for convenience, we still define the parameters to be an  $n$ -by- $n$  matrix while only  $\binom{n}{2}$  of them are effective.

### 2.3.2 Weighted Random Dot Product Graph

As discussed in Section 2.2.2, the connectivity between two vertices in a graph generally depends on some hidden properties of the corresponding vertices. Such property is well captured by LPM as well as RDPG, which is a special case of LPM. In this section, we generalize RDPG to a weighted version so that it can model weighted graphs.

**Definition 2.3.2 (Weighted Random Dot Product Graph)** Consider a collection of one-parameter distributions  $\mathcal{F} = \{f_\theta, \theta \in \Theta \subset \mathbb{R}\}$ . The weighted random dot product graph (WRDPG) with respect to  $\mathcal{F}$  is defined as following: Let  $X \in \mathbb{R}^{n \times d}$  be such that  $X = [X_1, X_2, \dots, X_n]^\top$ , where  $X_i \in \mathbb{R}^d$  for all  $i \in [n]$ . The matrix  $X$  is random and satisfies  $\mathbb{P}(X_i^\top X_j \in \Theta) = 1$  for all  $i, j \in [n]$ . Conditioned on  $X$ ,  $A_{ij}$  follows distribution  $f_\theta \in \mathcal{F}$  with parameter  $\theta = X_i^\top X_j$  independent of others for  $i, j \in [n]$  and  $i < j$ .

Under the WRDPG defined above, the parameter matrix  $P = XX^\top \in \mathbb{R}^{n \times n}$  is automatically symmetric because the link function is inner product. Moreover, to

## CHAPTER 2. RANDOM GRAPH MODELS

have symmetric graphs without self-loops, only  $A_{ij}$  ( $i < j$ ) are sampled while leaving the diagonals of  $A$  to be all zeros.

After such extension, WRDPG inherits some properties from RDPG naturally, for example the positive definiteness of  $P$ , non-identifiability, etc. And similarly, since  $P$  is still the outer product of the latent positions  $X$  in WRDPG, it also motivates the low-rank estimator based on spectral decomposition for weighted graphs which will be discussed in later sections.

### 2.3.3 Weighted Stochastic Blockmodel

**Definition 2.3.3 (Weighted Stochastic Blockmodel)** *Consider a collection of one-parameter distributions  $\mathcal{F} = \{f_\theta, \theta \in \Theta \subset \mathbb{R}\}$ . A  $K$ -block weighted stochastic blockmodel (WSBM) with respect to  $\mathcal{F}$  is defined as following: Let block probability matrix be  $B \in \Theta^{K \times K}$ , and let the vector of block memberships be  $\tau \in [K]^n$ , where for each  $i \in [n]$ ,  $\tau_i = k$  means vertex  $i$  is a member of block  $k$ . Then  $A_{ij}$  follows distribution  $f_\theta \in \mathcal{F}$  with parameter  $\theta = B_{ij}$  independent of others for  $i, j \in [n]$  and  $i < j$ .*

WSBM can be defined in a similar way in the scenario when the block probability matrix  $B$  and block membership vector  $\tau$  are random. As mentioned in Section 2.2.3, all analysis for unweighted graphs is based on RDPG setting. Likewise, in this section we will represent WSBM as WRDPG. Because of the structure of WRDPG, in order to consider WSBM as a WRDPG, the block probability matrix  $B$  needs to be positive

## CHAPTER 2. RANDOM GRAPH MODELS

semidefinite. From now on, we will denote the sub-model of WSBM with positive semi-definite  $B$  as the WSBM.

**Definition 2.3.4 (WSBM as WRDPG)** Consider a WRDPG with respect to  $\mathcal{F}$  where each latent position  $X_i$  can take one of the  $K$  possible values  $\nu_1, \dots, \nu_K$  such that  $\mathbb{P}(X_i = \nu_k) = \rho_k$  for  $i \in [n]$  and  $k \in [K]$ , where the block proportion vector  $\rho \in (0, 1)^K$  with  $\sum_{k=1}^K \rho_k = 1$ . Then this is a WSBM with respect to  $\mathcal{F}$  where  $B = \nu^\top \nu$  with  $\nu = [\nu_1, \dots, \nu_K]^\top$ . Similarly, we can define  $\tau \in [K]^n$  such that  $X_i = \nu_{\tau_i}$ .

As shown above, the WSBM can be regard as a WRDPG where all vertices in the same block have identical latent positions. In later sections when considering weighted graphs, we will always analyze WSBM in a WRDPG setting.

# Chapter 3

## A Law of Large Graphs

### 3.1 Another Section

## **Chapter 4**

# **Robust Generalizations of the Law of Large Graphs**

# **Chapter 5**

## **Discussion**

# Bibliography

Edoardo M Airoldi, David M Blei, Stephen E Fienberg, and Eric P Xing. Mixed membership stochastic blockmodels. *Journal of Machine Learning Research*, 9 (Sep):1981–2014, 2008.

Béla Bollobás, Svante Janson, and Oliver Riordan. The phase transition in inhomogeneous random graphs. *Random Structures & Algorithms*, 31(1):3–122, 2007.

Paul Erdős and Alfréd Rényi. On random graphs, i. *Publicationes Mathematicae (Debrecen)*, 6:290–297, 1959.

Edgar N Gilbert. Random graphs. *The Annals of Mathematical Statistics*, 30(4):1141–1144, 1959.

Peter D Hoff, Adrian E Raftery, and Mark S Handcock. Latent space approaches to social network analysis. *Journal of the american Statistical association*, 97(460):1090–1098, 2002.

## BIBLIOGRAPHY

Paul W Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. Stochastic blockmodels: First steps. *Social networks*, 5(2):109–137, 1983.

Brian Karrer and Mark EJ Newman. Stochastic blockmodels and community structure in networks. *Physical Review E*, 83(1):016107, 2011.

Christine Leigh Myers Nickel. *Random dot product graphs: A model for social networks*, volume 68. 2007.

Stephen J Young and Edward R Scheinerman. Random dot product graph models for social networks. In *Algorithms and models for the web-graph*, pages 138–149. Springer, 2007.

# Vita



Runze Tang received the B.S. degree in mathematics and applied mathematics from the University of Science and Technology of China in 2012, the M.S.E. degree in computer science from the Johns Hopkins University in 2015 and enrolled in the Applied Mathematics and Statistics Ph.D. program at Johns Hopkins University in 2012. Runze has received the Teaching Fellow Recognition Award.