Εθνικο Μετσοβιο Πολυτεχνειο
Σχολη Ηλεκτρολογων Μηχανικων και Μηχανικων Υπολογιστων
ΔΠΜΣ Επιστημη Δεδομενων και Μηχανικη Μαθηση

Αποδοτικοί Αλγόριθμοι Εκμάθησης Χρονικά Μεταβαλλόμενων Κατανομών Κατάταξης

Μεταπτυχιακη Διπλωματικη Εργασια

Δημητριος Οικονομου

Επιβλεπων: Δημήτριος Φωτάκης
Αν. Καθηγητής Ε.Μ.Π.

Αθήνα, Μάρτιος 2022

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
ΔΠΜΣ Επιστήμη Δεδομένων και Μηχανική Μάθηση

Αποδοτικοί Αλγόριθμοι Εκμάθησης Χρονικά Μεταβαλλόμενων Κατανομών Κατάταξης

Μεταπτυχιακη Διπλωματικη Εργασια

Δημητριος Οικονομου

Επιβλέπων: Δημήτριος Φωτάκης
Αν. Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή στις 23 Μαρτίου 2022.

(Υπογραφή)	*(Υπογραφή)*	*(Υπογραφή)*
....	....	....
Δημήτριος Φωτάκης	Αριστείδης Παγουρτζής	Γεώργιος Στάμου
Αν. Καθηγητής Ε.Μ.Π.	Καθηγητής Ε.Μ.Π.	Καθηγητής Ε.Μ.Π.

Αθήνα, Μάρτιος 2022

(Υπογραφή)

....
Δημήτριος Οικονόμου
Πτυχιούχος Μαθηματικός Ε.Κ.Π.Α.

Περιληψη

Το ερώτημα της κατάταξης αντικειμένων από δυαδικές συγκρίσεις αποτελεί αντικείμενο ενδιαφέροντος εδώ και πολύ καιρό. Αυτό το πρόβλημα έχει πολλές εφαρμογές στην πραγματική ζωή, όπως στα συστήματα συστάσεων ή σε κατατάξεις ομάδων σε αθλητικές δραστηριότητες. Επιπλέον, σε πολλές περιπτώσεις είναι θεμιτό να έχουμε ένα “σκορ” για κάθε αντικείμενο έτσι ώστε να μπορούμε να κατανοήσουμε καλύτερα την “βαρύτητα” της κατάταξης.

Σε αυτή την διπλωματική εργασία, δουλέυουμε κυρίως με το δημοφιλές μοντέλο των Bradley-Terry-Luce (BTL) όπου κάθε αντικείμενο έχει ένα σχετικό σκορ που καθορίζει το αποτέλεσμα της σύγκρισης σύμφωνα με μια δοκιμή Bernoulli. Αυτό είναι το Static BTL μοντέλο, όπως έχει περιγραφεί στο [NOS17]. Στη συνέχεια, παρουσιάζουμε διάφορες επεκτάσεις αυτού του βασικού μοντέλου. Ξεκινάμε με το Dynamic BTL μοντέλο, όπου υποθέτουμε ότι τα σκορ εξελίσσονται με το πέρασμα του χρόνο. Αυτό εμφανίστηκε πρώτη φορά στο [KT21]. Στην εργασία μας, επεκτείνουμε το μοντέλο τους, υποθέτοντας ότι τα γραφήματα σύγκρισης είναι θετικά συσχετισμένα γραφήματα σύγκρισης. Αποδεικνύουμε ότι ο αλγόριθμός τους λύνει επίσης και αυτό το επεκτεταμένο μοντέλο. Έπειτα ερευνούμε το Adversarial BTL μοντέλο. Σε αυτό το μοντέλο υποθέτουμε ότι υπάρχει ένας αντίπαλος (adversary) που λέει ψέμματα για κάποια από τα δυαδικά αποτελέσματα. Η πρώτη αναφορά αυτού του μοντέλο ήταν στο [Aga+20], ωστόσο παρουσιάζουμε ένα ισχυρότερο θεώρημα το οποιο εμπεριέχει το Static BTL μοντέλο. Τέλος εισάγουμε την έννοια του Dynamic Adversarial BTL μοντέλου το οποιο γενικεύει και ενωποιεί κάθε ένα από τα προηγούμενα μοντέλα. Επίσης συνδυάζουμε τον αλγόριθμο του dynamic μοντέλου με τον αλγόριθμο του adversarial μοντέλου ώστε να λάβουμε έναν αποδοτικό αλγόριθμο για το γενικευμένο μας μοντέλο.

Λέξεις Κλειδιά

Θεωρία Μάθησης, Στατιστική Μάθηση, Κατανομές Κατάταξης, Μοντέλο BTL, Στοχαστικές Διεργασίες, Τυχαία Γραφήματα, Θεωρία Πιθανοτήτων

Abstract

The question of ranking items from pairwise comparisons has been a subject of interest for a very long time. This problem has many real world applications, such as recommendation systems or ranking teams in a sports event. Moreover, in many cases it is desirable to have a “score” for each item in order to understand the “intensity” of the ranking.

In this thesis, we are working with the popular Bradley-Terry-Luce (BTL) model in which each item has an associated score which determines the outcome of a pairwise comparison according to a Bernoulli trial. This is the Static BTL model as descibed in [NOS17]. We describe the Spectral Ranking algorithm that gives efficient estimates of the BTL scores. Next we present possible extensions of this base model. We start with the Dynamic BTL model where we assume that the scores are evolving over time. This was first introduced in [KT21]. In our work we extend their model by assuming positively correlated comparison graphs. We prove that their algorithm also solves the extended setup. Next we are exploring the Adversarial BTL model. In this model we assume that there is an adversary that lies for some of the pairwise outcomes. The first mention of this model was in [Aga+20], however we present a stronger theorem which also encapsulates the Static BTL model. Finally we introduce the Dynamic Adversarial BTL model which generilizes and unifies each one of the previous models. We also combine the algorithm for the dynamic model and the algorithm for the adversarial model in order to give an efficient algorithm for our most general model.

Keywords

Learning Theory, Statistical Learning, Ranking Distributions, BTL Model, Stochastic Processes, Random Graphs, Probability Theory

ΕΥΧΑΡΙΣΤΙΕΣ

Αρχικά, θα ήθελα να ευχαριστήσω θερμά τον επιβλέποντα καθηγητή αυτής της διπλωματικής κύριο Δημήτρη Φωτάκη, για την ευκαιρία που μου έδωσε να ασχοληθώ με το συγκεκριμένο θέμα αλλά και για την έμπνευση και το ενδιαφέρον που μου καλλιέργησε κατά τη διάρκεια της συνεργασίας μας.

Ιδιαίτερες ευχαριστίες θα ήθελα να αποδώσω στον συνεργάτη και φίλο Άλκη Καλαβάση, υποψήφιο διδάκτορα με τον κύριο Φωτάκη. Η συμβολή του στην διεκπεραίωση της εργασίας ήταν καθοριστική.

Ακόμα, θα ήθελα να ευχαριστήσω τον καθηγητή κύριο Παγουρτζή καθώς και τον καθηγητή κύριο Στάμου που αποτελούν, μαζί με τον καθηγητή κύριο Φωτάκη, την τριμελή εξεταστική επιτροπή της παρούσας εργασίας.

Τέλος, θα ήθελα να ευχαριστήσω την αδελφή μου, Ακριβή, καθώς επίσης και τους γονείς μου, Παναγιώτη και Ιωάννα, για τη συνεχή στήριξη και υπομονή τους, όχι μόνο κατά την εκπόνηση της συγκεκριμένης διπλωματικής εργασίας, αλλά και καθόλη τη διάρκεια των προπτυχιακών και μεταπτυχιακών σπουδών μου.

Δημήτρης Οικονόμου
Αθήνα, Μάρτιος 2022

…στην αδελφή μου και στους γονείς μου

Contents

List of Figures

List of Algorithms

Chapter 1

Introduction

1.1 Problem Formulation and Related Work

The problem that we are dealing with in this thesis can be stated as follows: Given n items, we want to order the items based on partial orderings provided through many samples. Very often the available data that is presented to us is in the form of pairwise comparisons. From such partial preferences in the form of comparisons, we frequently wish to deduce not only the order of the underlying objects, but also the scores associated with the objects themselves so as to deduce the intensity of the resulting preference order.

There are many real world applications where one would like to have global rankings from such partial data:

- In rating responses of an online search engine to search queries ([Kaz11]).
- In marketing, we want to know the preferences of consumers about (many) products ([GCD81]).
- In machine learning we want to labeling data for the training of algorithms ([Hin+12], [Ray+10], [Den+09]).
- In crowdsourcing platforms such as Amazon Mechanical Turk ([Kha+11], [LR11], [Von+08]).
- In peer-grading in massive open online courses ([Pie+13]).
- In competitive sports such as chess or online gaming ([HMG06], [Ros07]).
- Counting the number of malaria parasites in an image of a blood smear ([LAF12]).

Most rating based systems rely on users to provide explicit numeric scores for their interests. While these assumptions have led to a lot of theoretical research for item recommendations based on matrix completion ([CR09], [KMO09], [NW12]) arguably numeric scores provided by individual users are generally inconsistent. Furthermore, in a number of learning contexts as illustrated above, explicit scores are not available.

Over the years, there have been proposed many distributions over ranking permutations:

- Plackett-Luce model ([Luc59], [Pla75]).
- Mallows model ([Mal57]).
- Bradley-Terry-Luce (BTL) model ([BT52], [Luc59]).

In this thesis we will focus in the BTL model. Suppose that we have n items of interest. We assume that there is a latent weight (or score) $\mathbf{w}^* = \left(w_1^*, \dots, w_n^*\right)^\top \in \mathbb{R}_+^n$ associated with each item $i \in [n]$. We also assume that each pair of items is being compared L times. Let Y_{ij}^l denote the outcome of the l-th comparison of the pair i and j, such that $Y_{ij}^l = 1$ if j is preferred over i and 0 otherwise. Then the BTL model assumes that

$$Y_{ij}^l \sim \text{Bernoulli}\left(\frac{w_j^*}{w_i^* + w_j^*}\right).$$

Finally we create the comparison graph $G = ([n], E)$, where $[n]$ represents the n items and i and j are compared if and only if $(i, j) \in E$. Note that the graph G has to be assumed connected, otherwise there would be no way to compare items that belong in different connected components. Our aim is to estimate the BTL weights $\mathbf{w}$ and rank the items accordingly.

There have been many papers that explore and try to solve the above problem. To name a few of them:

- In [For57] they provide an analysis based on Maximum Likelihood Estimators (MLE).
- In [AS11] they give a generalized Borda Count algorithm based on [Bor84].
- In [NOS17] they present an efficient algorithm that is based on random walks on Markov Chains. In [APA18] they modify the previous algorithm for faster convergence.

While the model we just described has many good theoretical guarantees as well as applications ([TVV04]) it can be somewhat restrictive. Hence there have been numerous tries to extend this frame work. We will focus in the following two.

Dynamic BTL. In the dynamic setting we assume that we have a time grid $\mathcal{T}$ and the BTL weights vary over time. Hence now we have a sequence of comparison grahps $\{G_t\}$ instead of just one. Our aim is that given a time instance t in the time grid $\mathcal{T}$, estimate the weights $\mathbf{w}(t)$. This model is essentially described in [KT21]. Another model that has a dynamic element is the one proposed by [Bon+20]. In particular, they consider the logit version of BTL model. One more related model appears in [LW21]. This model aims at recovering a pairwise comparison matrix $X(T)$ at a time T from noisy linear measurements.

Adversarial BTL. In the adversarial setting we assume that there is an adversary with complete knowledge of the BTL weights and the comparison graph and then he gives us a corrupted version of both of them. This problem has been studied in [Aga+20]. Another adversarial corruption model similar to the one that we have been discussing, has been studied in the computer vision literature ([Gol+16], [HLV18]). All of these frameworks are very closely related to robust estimation theory in classical statistics, in particular, the $\varepsilon-$contamination model of Huber ([Hub65], [Hub92]) and its generalizations ([Dia+17], [Dia+18], [Dia+19]).

Finally this field of research on ranking distributions has been quite popular in the National Technical University of Athens. Many student have explored and expanded this area in recent years. For reference we cite some those works: [Kal19], [Mou19], [Sta20] and [Mam22].

1.2 Our Contributions

The main contributions of this thesis are the following:

1. We propose a more general dynamic BTL model. We start by introducing the notion of *positively correlated* random graphs and we assume that these represent the comparisons graphs. We prove that the algorithm given in [KT21] works also in our setting and we prove the corresponding theorem.

2. We modify the algorithm presented in [Aga+20] and we prove an appropriate theorem in this setting. With our result it is clear that when we have no corruption we get the base model and the base result.

3. Finally we propose a novel BTL setting, where we combine the dynamic setting and the adversarial setting. This model has as special cases all the other models we will discuss. Finally, we give an algorithm that solves the problem and we prove a generalization of the theorem presented in the dynamic setting and the theorem presented in the adversarial setting.

1.3 Organization of the thesis

Chapter 2: Mathematical Tools In this chapter we lay all the necessary mathematical foundations for the following chapters. We start with a brief overview of the basic facts of Markov chains. Next we review the basic concentration inequalities as well as others that are based on sub Gaussian random variables. We also introduce the notion of Erdős-Rényi random graphs. Finally we construct a new family of such random graph, called positively correlated random graphs, that we are using for some of the models of the following chapters.

Chapter 3: The Static BTL Model. In this chapter we introduce the classical (static) BTL model for global rankings from pairwise comparisons, as described in [NOS17] and [Che+19]. We begin by formulating the problem setting and we explain the Spectral Ranking Algorithm. Next we prove the main theorem of this framework. In the end we verify the correctness of the algorithm with numerical experiments.[1]

Chapter 4: The Dynamic BTL Model In this chapter we discuss a possible generalization of the Static BTL model. In the dynamic setting we add the element of time in our problem. In particular, the BTL weights vary over a time grid and the comparison graphs are all independent. Building upon the work of [KT21] we make the weaker assumption that the graphs are positively correlated, a construction that we saw in Chapter 2. Next we modify the Spectral Ranking algorithm into our setup and we prove the main theorem of this framework. Finally we present novel numerical experiments.

[1] We follow the generation of synthetic data as in [NOS17] but we have written our own code (in Python).

Chapter 5: The Static Adversarial BTL Model In this chapter we explore another possible generalization of the Static BTL model. In this setting, we assume that there is an adversary that lies about some of the outcomes of the pairwise comparisons. Based on the work of [Aga+20], we provide an algorithm that solves the problem. Moreover, we prove a new theorem that has as a special case the theorem from Chapter 3.

Chapter 6: The Dynamic Adversarial BTL Model In this chapter we combine the models that we described in Chapter 4 and Chapter 5. We explain the whole framework in detail and we combine some of the previous algorithms in order to give a novel algorithm that solves the problem in the most general case. Finally we provide a new theorem that encapsulates all the main theorems of the previous models as special cases.

Appendix A: Technical Tools In this chapter we present detailed proofs for the results from the main chapters. In particular, we provide two results about the ℓ_2 norms of a special kind of random matrices and vectors. In addition, we introduce the notion of the spectral graph and we prove a useful proposition. Finally, we define the $\|\cdot\|_\pi$ norm and using this norm we formulate and prove the Eigenvector Perturbation theorem.

Notation

For the mathematical analysis of this thesis we are going to use the following notation.

- For any natural number n, let $[n] = \{1, \dots, n\}$.
- With lower case, bold-faced letters we denote vectors, e.g. $\mathbf{v}$.
- With upper case, bold-faced letters we denote matrices, e.g. $\mathbf{A}$.
- With A_{ij} we denote the (i, j) element of matrix $\mathbf{A}$.
- We use $c_1, c_2, k_1, k_2, C_1, C_2, \dots$ to denote absolute constants.
- With $\|\mathbf{v}\|_2$ we denote the ℓ_2 norm of the vector $\mathbf{v}$, and with $\|\mathbf{v}\|_1$ we denote the ℓ_1 norm.
- We use $\|\mathbf{A}\|_2$ to denote the spectral norm of matrix $\mathbf{A}$ and $\|\mathbf{A}\|_F$ to denote the Frobenius norm.

Chapter 2

Mathematical Tools

In this chapter we are going to review some basic mathematical results from probability theory and stochastic processes as well as from random graphs theory. We are going to use them later in many proofs.

2.1 Markov Chains

In this section we are going to give a brief overview of the basics of discrete time Markov chains. We are going to use these results all the time in the subsequent chapters. The contents of this section are based on [LP17] and [Lou15].

Definition 2.1.1. Let Ω be a finite state space. A sequence of random variables $X_0, X_1, \dots$ is a *Markov chain* with state space Ω if for all $v_0, \dots, v_{n-1}, x, y \in \Omega$ and all $n \in \mathbb{N}$ we have that

$$\mathbb{P}\left[X_{n+1} = y \mid X_0 = v_0, \dots, X_{n-1} = v_{n-1}, X_n = x\right] = \mathbb{P}\left[X_{n+1} = y \mid X_n = x\right]. \tag{2.1}$$

Remark 2.1.2. A Markov chain is called *time-homogeneous* if the right hand side of Equation (2.1) does not depend from n. This is the case for most applications. Then we can define the *probability transition matrix* $\mathbf{P}$ as

$$P_{x,y} = \mathbb{P}\left[X_{n+1} = y \mid X_n = x\right].$$

From now on when we use the term "Markov chain" we imply that it is a time homogeneous Markov chain. Note that $\mathbf{P}$ is a *stochastic matrix*, i.e.

$$\sum_{y \in \Omega} P_{x,y} = 1.$$

Definition 2.1.3. Let $\Omega = \{x_1, \dots, x_n\}$. Then the distribution at time n is given by

$$\boldsymbol{\pi}_n = \left(\pi_n(x_1), \dots, \pi_n(x_n)\right),$$

where

$$\pi_n(y) = \mathbb{P}\left[X_n = y\right].$$

It is easy to prove that $\boldsymbol{\pi}_{n+1} = \boldsymbol{\pi}_n \mathbf{P}$ for all $n \in \mathbb{N}$. These are known as the *Chapman-Kolmogorov* equations.

Definition 2.1.4. A Markov chain in a finite state space Ω is:

- *irreducible* if for any two states $x, y \in \Omega$, there exists a time step n such that $P^{(n)}(x, y) > 0$.
- *aperiodic* if, for any state x, it holds that $\gcd\{n : P^{(n)}(x, x) > 0\} = 1$.
- *ergodic* if it is both irreducible and aperiodic.

Definition 2.1.5. A *stationary distribution* $\boldsymbol{\pi} \in \mathbb{R}^n$ for a (finite) Markov chain with transition matrix $\mathbf{P}$ is defined as the leading left eigenvector of $\mathbf{P} \in \mathbb{R}^{n \times n}$

$$\boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi}.$$

Theorem 2.1.6. *An ergodic Markov chain has a unique stationary distribution.*

Theorem 2.1.7. *For a finite ergodic Markov we have the (entrywise) convergence*

$$\lim_{n \to \infty} \boldsymbol{\pi}_n = \boldsymbol{\pi}.$$

Definition 2.1.8. A Markov chain is *time reversible* if there exists a distribution $\boldsymbol{\pi}$ that satisfies the detailed balanced equations:

$$\pi(x)P(x, y) = \pi(y)P(y, x) \text{ for all } x, y \in \Omega.$$

Then $\boldsymbol{\pi}$ is the stationary distribution.

2.2 Concentration Inequalities

All of the algorithms that we are going to work with, are randomized algorithms which means that they produce the correct result most of the time, i.e. *with high probablity*. In order to analyze these algorithms we are going to need the following measure concentration inequalities.

First we begin with the elementary Markov's inequality.

Theorem 2.2.1 (Markov)**.** *Let $X \geq 0$ be a non negative random variable and let $a > 0$ be a real number. Then we have*

$$\mathbb{P}[X \geq a] \leq \frac{\mathbb{E}[X]}{a}.$$

An application of Markov's inequality is the Chernoff bound.

Theorem 2.2.2 (Chernoff)**.** *Let $S_n = \sum_{i=1}^n X_i$ where $X_1, \ldots, X_n$ are independent random variables such that $X_i \sim \text{Be}(p_i)$. Note that $\mu = \mathbb{E}[S_n] = \sum_{i=1}^n p_i$. Then*

$$\mathbb{P}[|S_n - \mu| > t\mu] \leq 2\exp\left\{\frac{-\mu t^2}{3}\right\},$$

for $t \in (0, 1)$.

A powerful and useful generalization of the Chernoff bound is Hoeffding's inequality.

Theorem 2.2.3 (Hoeffding, [Hoe63])**.** *Let $X_1, \dots, X_n$ be independent random variables such that $X_i \in [a_i, b_i]$. Also let $S_n = \sum_{i=1}^n X_i$. Then*

$$\mathbb{P}\left[S_n - \mathbb{E}\left[S_n\right] > t\right] \leq \exp\left\{\frac{-2t^2}{\sum_{i=1}^n \left(b_i - a_i\right)^2}\right\}$$

and

$$\mathbb{P}\left[\left|S_n - \mathbb{E}\left[S_n\right]\right| > t\right] \leq 2\exp\left\{\frac{-2t^2}{\sum_{i=1}^n \left(b_i - a_i\right)^2}\right\}$$

for all $t > 0$.

Another generalization of the Chernoff bound as well as Hoeffding's inequality is Bernstein's inequalities. It is actually a family of inequalities but here we present the following.

Theorem 2.2.4 (Bernstein, [Ber37])**.** *Let $X_1, \dots, X_n \in \mathbb{R}$ be independent random variables, each satisfying $\mathbb{E}\left[X_i\right] = 0$ and $|X_i| \leq B$ almost surely. Then for any $t > 0$*

$$\mathbb{P}\left[\left|\sum_{i=1}^n X_i\right| > t\right] \leq 2\exp\left\{\frac{-3t^2}{6\sum_{i=1}^n \mathbb{E}\left[X_i^2\right] + 2Bt}\right\}.$$

Furthermore, there is a generalization of the previous inequality for random matrices.

Theorem 2.2.5 (Matrix Bernstein Inequality, [Tro12]: Theorem 1.5)**.** *Let $Z_1, \dots, Z_n \in \mathbb{R}^{d_1 \times d_2}$ be independent random matrices, each satisfying $\mathbb{E}\left[Z_i\right] = 0$ and $\|Z_i\|_2 \leq B$ almost surely. Then for any $t > 0$*

$$\mathbb{P}\left[\left\|\sum_{i=1}^n Z_i\right\|_2 > t\right] \leq (d_1 + d_2)\exp\left\{\frac{-3t^2}{6\nu + 2Bt}\right\},$$

where $\nu = \max\left\{\left\|\mathbb{E}\left[\sum_{i=1}^n Z_i^\top Z_i\right]\right\|_2, \left\|\mathbb{E}\left[\sum_{i=1}^n Z_i Z_i^\top\right]\right\|_2\right\}$.

2.2.1 Sub-Gaussian Random Variables

Sometimes we need more refined results about measure concentration. Thus we need to strengthen our assumptions. One very common assumption that occurs quite naturally is to let the random variables be sub-Gaussian.

Definition 2.2.6. A random variable $X \in \mathbb{R}$ is said to be *sub-Gaussian with variance proxy* $\sigma^2 > 0$ if

$$\mathbb{P}\left[|X| > t\right] \leq 2\exp\left(\frac{-t^2}{2\sigma^2}\right), \quad \forall t > 0.$$

In this case we write $X \sim \operatorname{subG}\left(\sigma^2\right)$.

Sub-Gaussian random variables have "strong tail decay". The next lemma makes this claim precise.

Lemma 2.2.7. *Let $X \sim \operatorname{subG}\left(\sigma^2\right)$. Then for any positive integer $p \geq 1$ we have*

$$\mathbb{E}\,|X|^p \leq \left(2\sigma^2\right)^{p/2} p\Gamma\left(p/2\right).$$

In particular,

$$\left(\mathbb{E}\,|X|^p\right)^{1/p} \leq \sqrt{p}\sigma e^{1/e}, \quad p \geq 2,$$

and $\mathbb{E}\,|X| \leq \sigma\sqrt{2\pi}$ and $\mathbb{E}\,X^2 \leq 4\sigma^2$.

Proof. We have

$$\begin{aligned}\mathbb{E}|X|^p &= \int_0^{+\infty} \mathbb{P}\left[|X|^p > t\right] dt \\ &= \int_0^{+\infty} \mathbb{P}\left[|X| > t^{1/p}\right] dt \\ &\le 2\int_0^{+\infty} e^{\frac{-t^{2/p}}{2\sigma^2}} dt \\ &= \left(2\sigma^2\right)^{p/2} p \int_0^{+\infty} e^{-u} u^{p/2-1} du, \quad u = \frac{t^{2/p}}{2\sigma^2} \\ &= \left(2\sigma^2\right)^{p/2} p\Gamma\left(p/2\right).\end{aligned}$$

The second statement follows from $\Gamma(p/2) \le (p/2)^{p/2}$ and $p^{1/p} \le e^{1/e}$ for $p \ge 2$. It yields

$$\left(\left(2\sigma^2\right)^{p/2} p\Gamma\left(p/2\right)\right)^{1/p} \le p^{1/p}\sqrt{\frac{2\sigma^2 p}{2}} \le e^{1/e}\sigma\sqrt{p}.$$

Moreover, for $p = 1$, we have $\sqrt{2}\Gamma(1/2) = \sqrt{2\pi}$ and for $p = 2$ it is $\Gamma(1) = 1$. □

The previous lemma motivates the following definition.

Definition 2.2.8. Let X be a random variable. The sub-Gaussian norm $\|\cdot\|_{\psi_2}$ is defined as

$$\|X\|_{\psi_2} = \sup_{p\ge 1} p^{-1/2}\left(\mathbb{E}|X|^p\right)^{1/p}.$$

Note that if $X \sim \mathrm{subG}\left(\sigma^2\right)$, then $\|X\|_{\psi_2} \le 3\sigma < +\infty$.

One of the most powerful inequalities regarding sub-Gaussian random variables is the Hanson-Wright inequality.

Theorem 2.2.9 (Hanson-Wright, [RV13]: Theorem 1.1)**.** *Let $X = (X_1, \dots, X_n) \in \mathbb{R}^n$ be a random vector with independent components X_i which satisfy $\mathbb{E}[X_i] = 0$ and $\|X_i\|_{\psi_2} \le K$, i.e. each component X_i is a sub-Gaussian random variable. Let A be a $n \times n$ matrix. Then, for every $t > 0$,*

$$\mathbb{P}\left[\left|X^\top AX - \mathbb{E}\left[X^\top AX\right]\right| > t\right] \le 2\exp\left\{-c\min\left(\frac{t^2}{K^4\|A\|_F^2}, \frac{t}{K^2\|A\|_2}\right)\right\},$$

for some constant $c > 0$.

Corollary 2.2.10. *Let $X = (X_1, \dots, X_n) \in \mathbb{R}^n$ be a random vector with independent components X_i which satisfy $\mathbb{E}[X_i] = 0$ and $\|X_i\|_{\psi_2} \le K$. Then, for every $t > 0$,*

$$\mathbb{P}\left[\left|\sum_{i=1}^n X_i^2 - \mathbb{E}\left[\sum_{i=1}^n X_i^2\right]\right| > t\right] \le 2\exp\left\{-\frac{ct}{K^2}\min\left(\frac{t}{nK^2}, 1\right)\right\},$$

for some constant $c > 0$.

Proof. Apply Theorem 2.2.9 with $A = I_n$. □

2.3 Erdős-Rényi Random Graphs

In all of the models that we are going to see, it will always exist a *random graph*. In particular, we are going to encounter the Erdős-Rényi random graph. We start with a definition.

Definition 2.3.1 (Erdős-Rényi random graph, [FK16]: Sec 1.1)**.** Fix $0 \leq p \leq 1$. Start with an empty graph with vertex set $[n]$ and perform $\binom{n}{2}$ Bernoulli experiments inserting edges independently with probability p.

Notation 2.3.2. We use $G \sim \mathcal{G}(n,p)$ to denote that G is an Erdős-Rényi random graph on n vertices with probability p.

One of the most important properties of this class of random graphs is given by the following theorem.

Theorem 2.3.3 (Erdős-Rényi, [ER60])**.** *Let $G \sim \mathcal{G}(n,p)$ be an Erdős-Rényi graph. If $p \geq c\frac{\log n}{n}$ for some sufficiently large constant $c > 1$, then G is almost surely connected.*

Another useful result that we are going to use, is the following lemma.

Lemma 2.3.4 ([KT21]: Lemma 11)**.** *Let $G \sim \mathcal{G}(n,p)$ be an Erdős-Rényi graph. Let*

1. $\mathcal{A}_1 = \left\{\frac{np}{2} \leq d_{\min} \leq d_{\max} \leq \frac{3np}{2}\right\}$,
2. $\mathcal{A}_3 = \left\{|E| \leq 2n^2 p\right\}$,
3. $\mathcal{A}_2 = \left\{\xi > \frac{1}{2}\right\}$, *where $\xi = \xi(G)$ is the spectral gap*[1] *of the graph G.*

Then there is a constant $c > 1$ such that if $p \geq c\frac{\log n}{n}$, then $\mathbb{P}\left[\mathcal{A}_i\right] \geq 1 - O\left(n^{-10}\right)$ for $i = 1,2,3$.

2.3.1 Time dependent Erdős-Rényi graphs

In this section we are going to construct an *evolution* of random graphs through time.

Construction 2.3.5. Let $G_1 \sim \mathcal{G}(n,p_1)$ be an Erdős-Rényi random graph. We will inductively construct a finite sequence $\{G_t\}_{t=1}^{T}$ of "correlated" Erdős-Rényi graphs.

Suppose that we already have $G_t = ([n], E_t)$. We construct G_{t+1} as follows: The set of vertices of G_{t+1} is the set $[n]$ and let $\alpha_t, p_{t+1} \in [0,1]$. We call a_t the *similarity coefficient* of the graph G_{t+1} with respect to the graph G_t. Let (i,j) be a pair of vertices. We have to decide whether this pair will be an edge in G_{t+1} or not. Firstly, we look at the *state* of (i,j) in G_t, i.e. we check whether $(i,j) \in E_t$ or $(i,j) \notin E_t$. Then with probability $\alpha_t \in [0,1]$ we keep the same state in G_{t+1} and with probability $1-\alpha_t$ we change the pair's state according to the rule: with probability p_{t+1} we add the edge $(i,j) \in E_{t+1}$ and with probability $1-p_{t+1}$ we don't add the edge. We do this procedure for all possible pairs of vertices. Hence we construct a graph $G_{t+1} = ([n], E_{t+1})$.

Remark 2.3.6. The Construction 2.3.5 is essentially the same with the construction described in [ODo14]: Chapter 2.4.

Remark 2.3.7. The graph G_{t+1} satisfies the following properties:

[1] See: Definition A.2.1

1. By construction we have the Markov property:

$$\mathbb{P}\left[(i,j)\in E_{t+1}|(i,j)\in E_t,\ldots,(i,j)\in E_1\right]=\mathbb{P}\left[(i,j)\in E_{t+1}|(i,j)\in E_t\right]$$
$$\mathbb{P}\left[(i,j)\notin E_{t+1}|(i,j)\notin E_t,\ldots,(i,j)\notin E_1\right]=\mathbb{P}\left[(i,j)\notin E_{t+1}|(i,j)\notin E_t\right]$$

2. It is easy to see that:

$$\begin{aligned}\mathbb{P}\left[(i,j)\in E_{t+1}|(i,j)\in E_t\right]&=\alpha_t+(1-\alpha_t)\,p_{t+1}\\ \mathbb{P}\left[(i,j)\notin E_{t+1}|(i,j)\in E_t\right]&=(1-\alpha_t)\,(1-p_{t+1})\\ \mathbb{P}\left[(i,j)\in E_{t+1}|(i,j)\notin E_t\right]&=(1-\alpha_t)\,p_{t+1}\\ \mathbb{P}\left[(i,j)\notin E_{t+1}|(i,j)\notin E_t\right]&=\alpha_t+(1-\alpha_t)\,(1-p_{t+1})\end{aligned}$$

3. Note that $G_{t+1}\sim\mathcal{G}\left(n,\alpha_t p_t+(1-\alpha_t)\,p_{t+1}\right)$. Indeed

$$\begin{aligned}\mathbb{P}\left[(i,j)\in E_{t+1}\right]&=\mathbb{P}\left[(i,j)\in E_t\right]\cdot\mathbb{P}\left[(i,j)\in E_{t+1}|(i,j)\in E_t\right]\\ &\quad+\mathbb{P}\left[(i,j)\in E_t\right]\cdot\mathbb{P}\left[(i,j)\in E_{t+1}|(i,j)\notin E_t\right]\\ &=p_t\left(\alpha_t+(1-\alpha_t)\,p_{t+1}\right)+(1-p_t)\,(1-\alpha_t)\,p_{t+1}\\ &=\alpha_t p_t+(1-\alpha_t)\,p_{t+1}\end{aligned}$$

4. We have the following special cases:

   - If $\alpha_t=0$ then $G_{t+1}\sim\mathcal{G}(n,p_{t+1})$ and the random graphs G_t and G_{t+1} are independent, i.e. $\mathbb{P}\left[(i,j)\in E_{t+1}\cap E_t\right]=\mathbb{P}\left[(i,j)\in E_{t+1}\right]\mathbb{P}\left[(i,j)\in E_t\right]$.
   - If $\alpha_t=1$ then $G_{t+1}=G_t$ and obviously $G_{t+1}\sim\mathcal{G}(n,p_t)$.

The union of all the graphs G_t constructed above has some interesting properties.

Proposition 2.3.8. *Let $\{G_t\}_{t=1}^T$ be a finite sequence of random graphs, constructed as above. Then the union graph $G=\cup_{t=1}^T G_t$ is an Erdős-Rényi graph with probability*

$$p=1-(1-p_1)\prod_{t=1}^{T-1}\left(1-(1-\alpha_t)\,p_{t+1}\right).$$

Proof. We have

$$\begin{aligned}\mathbb{P}\left[(i,j)\in\bigcup_{t=1}^T E_t\right]&=1-\mathbb{P}\left[(i,j)\notin\bigcap_{t=1}^T E_t\right]\\ &=1-\prod_{t=1}^T\mathbb{P}\left[(i,j)\notin E_t|(i,j)\notin E_{t-1},\ldots,(i,j)\notin E_1\right]\\ &=1-\mathbb{P}\left[(i,j)\notin E_1\right]\prod_{t=2}^T\mathbb{P}\left[(i,j)\notin E_t|(i,j)\notin E_{t-1}\right]\\ &=1-(1-p_1)\prod_{t=2}^T\left(\alpha_{t-1}+(1-\alpha_{t-1})\,(1-p_t)\right)\\ &=1-(1-p_1)\prod_{t=1}^{T-1}\left(1-(1-\alpha_t)\,p_{t+1}\right).\end{aligned}$$

□

Proposition 2.3.9. *Let $\{G_t\}_{t=1}^T$ be a finite sequence of random graphs, constructed as above. Suppose that*

$$p_1 + \sum_{t=1}^{T-1} (1 - a_t)\, p_{t+1} \geq \log n - \log (n - c \log n)\,,$$

for some sufficiently large constant $c > 1$. Then the union graph $G = \cup_{t=1}^T$ is almost surely connected.

Proof. By Proposition 2.3.8 we have that $G \sim \mathcal{G}(n,p)$ with

$$p = 1 - (1 - p_1) \prod_{t=1}^{T-1} (1 - (1 - \alpha_t)\, p_{t+1})\,.$$

Hence by Theorem 2.3.3 it is enough to prove that $p \geq \frac{c \log n}{n}$. We have

$$\begin{aligned} p &= 1 - (1 - p_1) \prod_{t=1}^{T-1} (1 - (1 - \alpha_t)\, p_{t+1}) \\ &\geq 1 - e^{-p_1} \prod_{t=1}^{T-1} e^{-(1-\alpha_t)p_{t+1}} \\ &= 1 - e^{-p_1 - \sum_{t=1}^{T-1} (1-\alpha_t) p_{t+1}} \\ &> 1 - e^{-(\log n - \log(n - c \log n))} \\ &= 1 - \frac{n - c \log n}{n} \\ &= \frac{c \log n}{n}, \end{aligned}$$

as desired. □

2.4 Positively Correlated Erdős-Rényi Graphs

In this section we present an interesting special case of Construction 2.3.5. We are going to use this new concept in our models in Chapter 4 and in Chapter 6.

In this special case we assume that all the similarity coefficients α_t are equal. In particular, let $\{G_1\}_{t=1}^T$ be a finite sequence constructed as previously, with $\alpha_t = \alpha$ for all $t = 1, \dots, T-1$. Then we have the following two edge cases:

- If $\alpha = 0$, then $G_t \sim \mathcal{G}(n, p_t)$ for all $t = 1, \dots, T$ and moreover all the pairs G_t, G_{t+1} are independent random graphs.
- If $\alpha = 1$, then $G_t = G_1 \sim \mathcal{G}(n, p_1)$ for all $t = 1, \dots, T$.

Then we have the following definition.

Definition 2.4.1. Let $\{G_1\}_{t=1}^T$ be a finite sequence constructed as previously, with $\alpha_t = \alpha$ for all $t = 1, \dots, T-1$. If $\alpha \in [0,1]$, then we call G_t's *positively correlated* random graphs.

Using the results in previous section we get the following corollary.

Corollary 2.4.2. *Let $\{G_1\}_{t=1}^T$ be a finite sequence of positively correlated random graphs. If $\sum_{t=1}^T p_t \geq \frac{\log n - \log(n - c \log n)}{1-\alpha}$ then the union graph $G = \cup_{t=1}^T$ is almost surely connected. In particular if $p_t \geq \frac{\log n - \log(n - c \log n)}{(1-\alpha)T}$ for all $t = 1, \dots, T$, then G is almost surely connected.*

Proof. It follows from Proposition 2.3.9, since

$$
\begin{aligned}
p_1 + \sum_{t=1}^{T-1} (1 - a_t) p_{t+1} &\geq (1 - \alpha) \sum_{t=1}^{T} p_t \\
&\geq \log n - \log(n - c \log n).
\end{aligned}
$$

□

Chapter 3

The Static BTL Model

In this chapter we aim to formally present the basic problem setup of ranking from binary comparisons. Furthermore, we introduce the Spectral Ranking Algorithm, an attempt at solving the problem efficiently. Subsequently, we provide theoretical guarantees that the algorithm works with high probability. In the end we show the effectiveness of the algorithm through numerical experimentation.

3.1 Problem Setup

Preference Scores. When comparing pairs of items from n items of interest, represented as $[n] = \{1, \dots, n\}$, the Bradley-Terry-Luce (BTL) model assumes that there is a latent weight (or score) $\mathbf{w}^* = \left(w_1^*, \dots, w_n^*\right)^\top \in \mathbb{R}_+^n$ associated with each item $i \in [n]$. The outcome of a comparison for pair of items i and j is determined only by the corresponding weights w_i^* and w_j^*. We also introduce the *condition number* as

$$b := \frac{w_{\max}^*}{w_{\min}^*}.$$

We assume that b is a fixed constant independent of n.

Comparison Graph. We assume that the comparisons between items are governed by a *comparison graph* $G = ([n], E)$, where $[n]$ represents the n items of interest. The items i and j are compared if and only if $(i, j) \in E$. The set of edges E is taken to be a subset of $\{(i, j) \in [n] \times [n] \mid i < j\}$. Throughout this chapter we assume that G is drawn from the Erdős-Rényi random graph $\mathcal{G}(n, p)$. Of course, the graph G has to be connected, otherwise there would be no way to compare items that belong in different connected components. Hence by Theorem 2.3.3, from now on we assume that $p \geq c\frac{\log n}{n}$ for a sufficiently large constant $c > 1$.

Pairwise Comparisons. For each $(i, j) \in E$, we assume that L independent comparisons take place between items i and j. Let Y_{ij}^l denote the outcome of the l-th comparison of the pair i and j, such that $Y_{ij}^l = 1$ if j is preferred over i and 0 otherwise. Then the BTL model assumes that

$$Y_{ij}^l \sim \text{Bernoulli}\left(\frac{w_j^*}{w_i^* + w_j^*}\right)$$

Furthermore, it is assumed that the random variables Y_{ij}^l are independent of one another for all i, j and l. Now let $\mathbf{y} = \left\{y_{ij} | (i,j) \in E\right\}$, where

$$y_{ij} = \frac{1}{L}\sum_{l=1}^{L} Y_{ij}^l$$

is the fraction of wins of j over i. By convention, we set $Y_{ji}^l = 1 - Y_{ij}^l$ for all $(i,j) \in E$. Then obviously $y_{ji} = 1 - y_{ij}$. In the same fashion, we denote

$$y_{ij}^* = \frac{w_j^*}{w_i^* + w_j^*}$$

and $y_{ji}^* = 1 - y_{ij}^*$ for all $(i,j) \in E$. Now we can turn the comparison graph G into a weighted graph by assigning the weight y_{ij} for all $(i,j) \in E$. Note that the weighted graph G contains all the information of our data.

Goal. The BTL model as we have described it so far is invariant under the scaling of the weights $\mathbf{w}^*$, so an n-dimensional representation of the scores is not unique. To get a unique representation we let

$$\boldsymbol{\pi}^* = \frac{\mathbf{w}^*}{\|\mathbf{w}^*\|_1}.$$

The goal is to *learn* (or at least *estimate*) the normalized weight vector $\boldsymbol{\pi}^*$ and then rank all the items according to $\boldsymbol{\pi}^*$.

3.2 Spectral Ranking Algorithm

As we have already mentioned there are many approaches that attempt to solve the above problem. Here we present one of the most recent and most powerful attempts, the *Spectral Ranking Algorithm*. The idea, which shares many similarities with the PageRank Algorithm [Pag+99], is to create a random walk on the comparison graph G and then calculate the stationary distribution of this random process.

In particular, consider the following stochastic matrices: Let $\mathbf{P} = \left[P_{ij}\right] \in \mathbb{R}_+^{n \times n}$ be the *comparison transition matrix* with

$$P_{ij} = \begin{cases} \frac{1}{d_{\max}} y_{ij} & \text{if } (i,j) \in E \text{ or } (j,i) \in E \\ 1 - \frac{1}{d_{\max}} \sum_{k \in N_G(i)} y_{ik} & \text{if } i = j \\ 0 & \text{otherwise,} \end{cases}$$

where d_{max} is the maximum degree of the comparison graph G and $N_G(i)$ is the set of neighbors of the node i in G.

Also let $\mathbf{P}^* = \left[P^*_{ij}\right] \in \mathbb{R}^{n\times n}_+$ be the *preference transition matrix* with

$$P^*_{ij} = \begin{cases} \dfrac{1}{d_{\max}} y^*_{ij} & \text{if } (i,j) \in E \text{ or } (j,i) \in E \\ 1 - \dfrac{1}{d_{\max}} \displaystyle\sum_{k\in N_G(i)} y^*_{ik} & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$$

Note that $\mathbb{E}\left[y_{ij}\right] = y^*_{ij}$, so by the Strong Law of Large Numbers, $y_{ij} \to y^*_{ij}$ almost surely when $L \to \infty$. Similarly for the transition matrices (the convergence is entrywise) $\mathbf{P} \to \mathbf{P}^*$. Moreover it is easy to see that the normalized weight vector $\boldsymbol{\pi}^* = \left(\pi^*_1, \dots, \pi^*_n\right)^\top \in \mathbb{R}^n_+$ is the stationary distribution of the Markov chain induced by the matrix $\mathbf{P}^*$, since it satisfies the detailed balance equation $\pi^*_i P^*_{ij} = \pi^*_j P^*_{ji}$ for all $i, j \in [n]$. As a result, it is reasonable to expect that the stationary distribution of the empirical version $\mathbf{P}$ to form a good estimate of $\boldsymbol{\pi}^*$, provided the sample size L is sufficiently large.

This motivates the following algorithm:

Algorithm 1: Spectral Ranking Algorithm for the Static BTL Model.

Input: The comparison graph G and the statistics $\mathbf{y}$.
Output: An estimate $\boldsymbol{\pi} \in \mathbb{R}^n_+$ of the true normalized weight vector $\boldsymbol{\pi}^*$.

1 Compute the comparison transition matrix $\mathbf{P}$ as shown above.
2 Compute the leading left eigenvector $\boldsymbol{\pi}$ of $\mathbf{P}$.

3.3 Main Result

The purpose of this section is to prove that the Spectral Ranking Algorithm actually works, i.e. it provides a good estimate of $\boldsymbol{\pi}^*$.

Theorem 3.3.1 ([NOS17]: Theorem 2, [Che+19]: Theorem 9)**.** *Suppose that $p \geq k_1 \frac{\log n}{n}$ for a suitable constant $k_1 > 0$. Then there is a constant $C_1 > 0$ such that if*

$$L \geq c_1 \frac{b^7 \log n}{np}, \tag{3.1}$$

for some constant $c_1 > 0$, then with probability at least $1 - \frac{1}{\text{poly}(n)}$, one has

$$\frac{\|\boldsymbol{\pi} - \boldsymbol{\pi}^*\|_2}{\|\boldsymbol{\pi}^*\|_2} \leq C_1 \frac{b^{9/2}}{\sqrt{npL}}. \tag{3.2}$$

Notation 3.3.2. Throughout this section let $\boldsymbol{\Delta} = \mathbf{P} - \mathbf{P}^*$.

We need the following lemmas.

Lemma 3.3.3 ([NOS17]: Lemma 3)**.** *There exist a constant C_1 such that with probability at least $1 - \frac{1}{\text{poly}(n)}$, one has*

$$\|\boldsymbol{\Delta}\|_2 \leq C_1 \sqrt{\frac{\log n}{L d_{\max}}}$$

Proof. This is a direct application of Proposition A.1.1: Let $L_{ij} = L$ and

$$Z_{ij}^l = \begin{cases} \frac{1}{Ld_{\max}} \left(Y_{ij}^l - y_{ij}^* \right) & \text{if } (i,j) \in E \text{ or } (j,i) \in E \\ 0 & \text{otherwise,} \end{cases}$$

where $Y_{ij}^l \sim \text{Bernoulli}\left(y_{ij}^*\right)$. Then $B = \frac{1}{Ld_{\max}}$ and $N_{\max} = d_{\max}$ and we get exactly what we want. □

Lemma 3.3.4 ([Che+19]: Section C.2)**.** *Since $n \geq \frac{c_1}{p} \log n$ we have that with probability at least $1 - \frac{1}{\text{poly}(n)}$ it holds*

$$\left\| \boldsymbol{\pi}^{*\top} \left(\mathbf{P} - \mathbf{P}^* \right) \right\|_2 \leq C_2 \frac{b}{\sqrt{Ld_{\max}}} \left\| \boldsymbol{\pi}^* \right\|_2 .$$

Proof. This is a direct application of Proposition A.1.2: Construct $\boldsymbol{\Delta}$ as in the previous proof and let $\mathbf{a} = \boldsymbol{\pi}^*$. Then $\varphi = b$ and we get exactly what we want. □

Lemma 3.3.5 ([NOS17]: Lemma 4)**.** *If*

$$L \geq C_3 \frac{b^7 d_{\max}}{\xi^2 d_{\min}^2} \log n,$$

for some sufficiently large constant $C_3 > 0$, then with probability at least $1 - \frac{1}{\text{poly}(n)}$ we have

$$\gamma := 1 - \lambda_{\max}\left(\mathbf{P}^*\right) - \left\| \mathbf{P} - \mathbf{P}^* \right\|_{\boldsymbol{\pi}^*} \geq \frac{\xi d_{\min}}{4b^3 d_{\max}},$$

where ξ is the spectral gap of the graph G.

Proof. By Proposition A.2.2 with $(\mathbf{P}, \boldsymbol{\pi}) = (\mathbf{P}^*, \boldsymbol{\pi}^*)$, we have that

$$\gamma := 1 - \lambda_{\max}\left(\mathbf{P}^*\right) - \left\| \mathbf{P} - \mathbf{P}^* \right\|_{\boldsymbol{\pi}^*} \geq \frac{\xi d_{\min}}{2b^3 d_{\max}} - \left\| \mathbf{P} - \mathbf{P}^* \right\|_{\boldsymbol{\pi}^*} .$$

But by the innitial assumption on L and Lemma 3.3.3, with probability at least $1 - \frac{1}{\text{poly}(n)}$ we have,

$$\begin{aligned} \left\| \mathbf{P} - \mathbf{P}^* \right\|_2 &\leq \sqrt{b} \left\| \boldsymbol{\Delta} \right\|_{\boldsymbol{\pi}^*} \\ &\leq C_1 \sqrt{\frac{b \log n}{Ld_{\max}}} \\ &\leq \frac{\xi d_{\min}}{4b^3 d_{\max}}. \end{aligned}$$

As a result,

$$\begin{aligned} \gamma &\geq \frac{\xi d_{\min}}{2b^3 d_{\max}} - \frac{\xi d_{\min}}{4b^3 d_{\max}} \\ &= \frac{\xi d_{\min}}{4b^3 d_{\max}}. \end{aligned}$$

□

Now we have all the tools to prove the main theorem.

Proof of Theorem 3.3.1. First we assume that the comparison graph G is a general graph (i.e. there is no randomness) and that

$$L \geq C_3 \frac{b^7 d_{\max}}{\xi^2 d_{\min}^2} \log n. \tag{3.3}$$

In this case we have:

$$\begin{aligned}
\|\pi - \pi^*\|_2 &\leq \frac{1}{\sqrt{\pi^*_{\min}}} \|\pi - \pi^*\|_{\pi^*}, \text{ by Proposition A.3.3} \\
&\leq \frac{1}{\sqrt{\pi^*_{\min}}} \frac{\left\|\pi^{*\top}(\mathbf{P} - \mathbf{P}^*)\right\|_{\pi^*}}{1 - \max\{\lambda_2(\mathbf{P}^*), |\lambda_n(\mathbf{P}^*)|\} - \|\mathbf{P} - \mathbf{P}^*\|_{\pi^*}}, \text{ by Theorem A.3.4} \\
&\leq \frac{1}{\sqrt{\pi^*_{\min}}} \frac{4b^3 d_{\max}}{\xi d_{\min}} \left\|\pi^{*\top}(\mathbf{P} - \mathbf{P}^*)\right\|_{\pi^*}, \text{ by Lemma 3.3.5} \\
&\leq \frac{4b^{7/2} d_{\max}}{\xi d_{\min}} \left\|\pi^{*\top}(\mathbf{P} - \mathbf{P}^*)\right\|_2, \text{ by Proposition A.3.3} \\
&\leq \frac{4b^{7/2} d_{\max}}{\xi d_{\min}} \frac{Cb}{\sqrt{L d_{\max}}} \|\pi^*\|_2, \text{ by Lemma 3.3.3.}
\end{aligned}$$

Hence

$$\frac{\|\pi - \pi^*\|_2}{\|\pi^*\|_2} \leq \frac{4b^{7/2} d_{\max}}{\xi d_{\min}} \frac{Cb}{\sqrt{L d_{\max}}}. \tag{3.4}$$

Now using Lemma 2.3.4, the Equation (3.3) becomes the Equation (3.1), and the Equation (3.4) becomes the Equation (3.2). This finishes the proof. □

3.4 Numerical Experiments

In this section we are going to test how well the Spectral Ranking Algorithm works in practise. In order to do this we are going to create synthetic data under the Static BTL Model which we will then feed to the algorithm. In the end we are going to use error metrics to quantify the “accuracy” of the algorithm.

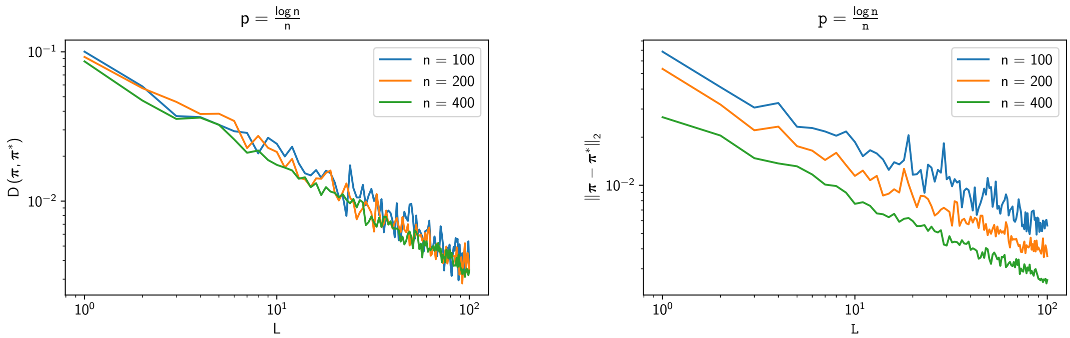
Error Metrics. For the majority of this chapter we have worked with the ℓ_2 norm. Here we introduce another metric, which is better suited for comparing rankings. We define D as the normalized weighted sum of pairs (i, j) whose ordering is incorrect:

$$D(\pi^*, \pi) = \left\{ \frac{1}{2n\|\pi^*\|_2^2} \sum_{i<j} \left(\pi_i^* - \pi_j^*\right)^2 \mathbb{1}_{(\pi_i^* - \pi_j^*)(\pi_i - \pi_j) < 0} \right\}^{1/2}, \tag{3.5}$$

where $\mathbb{1}_{\cdot}$ is an indicator function. Note that this metric is less sensitive to errors between pairs with similar weights. Moreover, we have the following lemma which connects the metric $D(\pi^*, \pi)$ to the bound provided in Theorem 3.3.1. As a result, the same upper bound holds for $D(\pi^*, \pi)$ error.

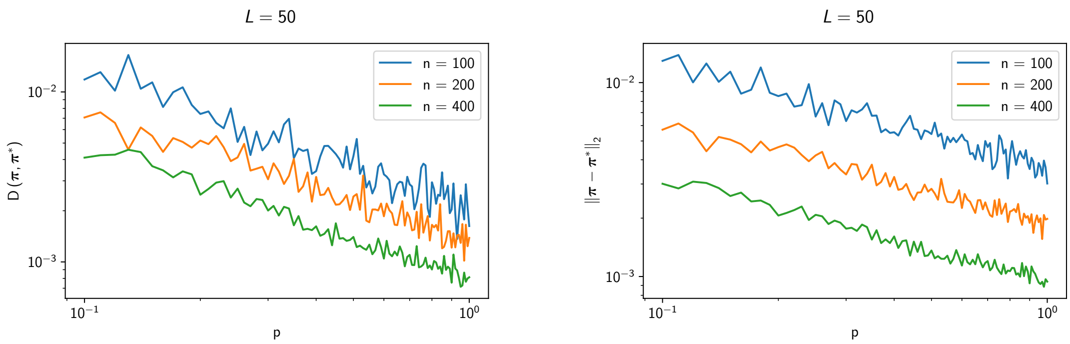
Lemma 3.4.1 ([NOS17]: Lemma 1). *Let π^*, π be probability vectors. Then,*

$$D(\pi^*, \pi) \leq \frac{\|\pi^* - \pi\|_2}{\|\pi^*\|_2}.$$



(a) Error Metric: D (b) Error Metric: ℓ_2

Figure 3.1: Log-Log plots of the evolution of errors as L grows with fixed $p = 5\frac{\log n}{n}$



(a) Error Metric: D (b) Error Metric: ℓ_2

Figure 3.2: Log-Log plots of the evolution of errors as p goes to 1, with fixed $L = 32$

Synthetic data. We generate data according to the BTL model. We follow the experimentation scheme of [NOS17] but we have written our own code[1]. In particular, for a given n and $b > 1$, the weights are constructed as follows:

$$\pi_i^* = b^{(2i-1-n)/2n}, \quad i \in [n].$$

Obviously $\frac{\pi^*_{\max}}{\pi^*_{\min}} < b$.

We present two kinds of plots. In the first one, we let L grow while keeping constant the probability ($p = 5\frac{\log n}{n}$) in the $\mathcal{G}(n, p)$, for $n = 100, 200, 400$. A representative result is depicted in Figure 3.1.

In the second plot, we let p go to 1 while keeping constant the number of comparisons ($L = 50$), for $n = 100, 200, 400$. A representative result is depicted in Figure 3.2.

Acknowledgements

The contents of this chapter are based on [NOS17] and [Che+19]. In particular, the first appearance of a result like Theorem 3.3.1 was in [NOS17]: Theorem 2, but with an extra $\sqrt{\log n}$ factor. Later in [Che+19], the result was further improved by eliminating the $\sqrt{\log n}$ factor.

[1] The code is availiable at: https://github.com/dimoik96/ntua-thesis-code

Chapter 4

The Dynamic BTL Model

In this chapter we present one possible generalization of the Static BTL setup. In this dynamic setting we add the element of time evolution in the BTL weights. We start by carefully introducing the model. Next we explain the Dynamic Spectral Algorithm which attempts at solving the problem efficiently. Subsequently, we provide theoretical guarantees that the algorithm works with high probability. In the end we show the effectiveness of the algorithm through numerical experimentation.

4.1 Problem Setup

Time. The basis of this model is that the weights vary over time. We assume that we have a time grid $\mathcal{T}$ which corresponds to the time evolution. In our case we assume $\mathcal{T}$ to be finite. Let $\mathcal{T} = \{1, \ldots, T\}$.

Preference Scores. As in the static case suppose we are comparing pairs of items from n items of interest, represented as $[n] = \{1, \ldots, n\}$. The Dynamic BTL model assumes that there is a latent weight (or score) $\mathbf{w}^*(t) = (w_1^*(t), \ldots, w_n^*(t))^\top \in \mathbb{R}_+^n$ associated with each item $i \in [n]$ for each timestamp $t \in \mathcal{T}$. The outcome of a comparison for a pair of items i and j at the moment $t \in \mathcal{T}$ is determined only by the corresponding weights $w_i(t)$ and $w_j(t)$. Again we introduce the *condition number* as

$$b(t) := \frac{w_{\max}^*(t)}{w_{\min}^*(t)}.$$

It is easy to see that if we let the weights vary unconditionally between each time moment there would be no way to get good estimates since the variance could be huge. Therefore for a meaningful recovery of $\mathbf{w}^*(t)$, we need to make the additional assumption:

Assumption 4.1.1. There exists $M \geq 0$ such that

$$\left| \frac{w_j^*(t)}{w_i^*(t) + w_j^*(t)} - \frac{w_j^*(t')}{w_i^*(t') + w_j^*(t')} \right| \leq M|t - t'|, \tag{4.1}$$

for all $t, t' \in \mathcal{T}$ and $i \neq j \in [n]$.

In words, the above assumption tell us that the weights of each item do not vary a lot throughout the evolution of time.

Comparison Graphs. We assume that at each time moment $t \in \mathcal{T}$ the comparisons between items are governed by a *comparison graph* $G_t = ([n], E_t)$, where $[n]$ represents the n items of interest. The items i and j are compared at the moment $t \in \mathcal{T}$ if and only if $(i,j) \in E_t$. The sets of edges E_t are assumed subsets of $\{(i,j) \in [n] \times [n] \mid i < j\}$. Throughout this chapter we assume that the graphs $\{G_t\}_{t=1}^T$ are positively correlated Erdős-Rényi graphs, as descibed in Section 2.4. Note that now we are not going to assume that each comparison graph G_t is connected, as we did in the static case. Indeed, in most real world application they are disconnected and very sparse. But we do assume that the union of all G_t is connected. The reason of this assumption will become clear later.

Pairwise Comparisons. For each $(i,j) \in E_t$, we assume that L independent comparisons take place between items i and j at the moment $t \in \mathcal{T}$. Let $Y_{ij}^l(t)$ denote the outcome of the l-th comparison of the pair i and j at $t \in \mathcal{T}$, such that $Y_{ij}^l(t) = 1$ if j is preferred over i and 0 otherwise. Then the Dynamic BTL model assumes that

$$Y_{ij}^l(t) \sim \text{Bernoulli}\left(\frac{w_j^*(t)}{w_i^*(t) + w_j^*(t)}\right)$$

Furthermore, it is assumed that the random variables $Y_{ij}^l(t)$ are independent of one another for all i, j, l and t. Now let $\mathbf{y}(t) = \left\{y_{ij}(t) | (i,j) \in E_t\right\}$, where

$$y_{ij}(t) = \frac{1}{L}\sum_{l=1}^{L} Y_{ij}^l(t)$$

is the fraction of wins of j over i at $t \in \mathcal{T}$. By convention, we set $Y_{ji}^l(t) = 1 - Y_{ij}^l(t)$ for all $(i,j) \in E_t$. Then obviously $y_{ji}(t) = 1 - y_{ij}(t)$. In the same fashion, we denote

$$y_{ij}^*(t) = \frac{w_j^*(t)}{w_i^*(t) + w_j^*(t)}$$

and $y_{ji}^*(t) = 1 - y_{ij}^*(t)$ for all $(i,j) \in E_t$. Now we can turn each comparison graph G_t into a weighted graph by assigning the weight $y_{ij}(t)$ for all $(i,j) \in E_t$. Note that the weighted graphs G_t contain all the information of our data.

Goal. The Dynamic BTL model as we have described it so far is invariant under the scaling of the weights $\mathbf{w}^*(t)$ at each time $t \in \mathcal{T}$, so an n-dimensional representation of the scores is not unique. To get a unique representation we let

$$\boldsymbol{\pi}^*(t) = \frac{\mathbf{w}^*(t)}{\|\mathbf{w}^*(t)\|_1}.$$

The goal is that given a $t \in \mathcal{T}$, we want to *learn* (or at least *estimate*) the normalized weight vector $\boldsymbol{\pi}^*(t)$ and then rank all the items according to $\boldsymbol{\pi}^*(t)$.

Remark 4.1.2. Note that this model is a generalization of the model presented in Chapter 3. Indeed, if we take $G_t = G$ and $M = 0$ for all $t \in \mathcal{T}$ then we recover exactly the Static BTL model.

4.2 Dynamic Spectral Ranking Algorithm

We are going to generalize the Spectral Ranking Algorithm (Section 3.2) in this setting. The first approach that one might try is that given $t \in \mathcal{T}$, apply the Spectral Ranking Algorithm on G_t. This could work if all of the graphs G_t are connected. But as we said, we make no such assumption. Instead, the Assumption 4.1.1 suggests that the pairwise outcomes at close time instants are similar, so it is possible to estimate $\boldsymbol{\pi}(t)$ utilizing the data lying in a neighborhood of t. Since we have assumed that the union of all graphs is connected, there exists a time neighborhood such that the union of all the graphs is connected and thus we can apply the Static Spectral Ranking Algorithm.

We are going to make everything precise. Let

$$N_\delta(t) = \{t-\delta, \dots, t, \dots, t+\delta\} \cap \mathcal{T}, \quad \delta \in \mathbb{N},$$

denote a δ time neighborhood around $t \in \mathcal{T}$. Note that $1 \le |N_\delta(t)| \le 2\delta + 1$. For $\delta \in \mathbb{N}$, let

$$G_t^\delta = \left([n], E_t^\delta\right) \text{ with } E_t^\delta = \bigcup_{t' \in N_\delta(t)} E_{t'},$$

be the union graph that corresponds to the time neighborhood $N_\delta(t)$. Sometimes we will abuse the notation and we will denote the union graph just by $G = ([n], E)$.

Let

$$N_{ij,\delta}(t) = \{t' \in N_\delta(t) | (i,j) \in E_{t'}\},$$

denote the time instances in $N_\delta(t)$ where i and j are being compared. Then for the graph $G = G_t^\delta$, consider

$$\hat{y}_{ij,\delta}(t) = \frac{1}{|N_{ij,\delta}(t)|} \sum_{t' \in N_{ij,\delta}(t)} y_{ij}(t'),$$
$$\hat{y}^*_{ij,\delta}(t) = \frac{1}{|N_{ij,\delta}(t)|} \sum_{t' \in N_{ij,\delta}(t)} y^*_{ij}(t').$$

Let $\mathbf{P}_\delta(t) = \left[P_{ij,\delta}(t)\right] \in \mathbb{R}_+^{n \times n}$ with

$$P_{ij,\delta}(t) = \begin{cases} \dfrac{1}{d_{\delta,\max}} \cdot \hat{y}_{ij,\delta}(t) & \text{if } (i,j) \in E \text{ or } (j,i) \in E \\ 1 - \dfrac{1}{d_{\delta,\max}} \cdot \displaystyle\sum_{k \in N_G(i)} \hat{y}_{ik,d}(t) & \text{if } i = j \\ 0 & \text{otherwise,} \end{cases}$$

where $d_{\delta,\max} = d_{\delta,\max}(t)$ is the maximum degree of the graph $G = G_t^\delta$.

Also let $\mathbf{P}^*(t) = \left[P^*_{ij}(t)\right] \in \mathbb{R}_+^{n \times n}$ with

$$P^*_{ij}(t) = \begin{cases} \dfrac{1}{d_{\delta,\max}} y^*_{ij}(t) & \text{if } (i,j) \in E \text{ or } (j,i) \in E \\ 1 - \dfrac{1}{d_{\delta,\max}} \displaystyle\sum_{k \in N_G(i)} y^*_{ik}(t) & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$$

It is easy to see that the normalized weight vector $\boldsymbol{\pi}^*(t) = \left(\pi_1^*(t), \dots, \pi_n^*(t)\right)^\top \in \mathbb{R}_+^n$ is the stationary distribution of the Markov chain induced by the matrix $\mathbf{P}^*(t)$, since it satisfies the detailed balance equation $\pi_i^*(t)P_{ij}^*(t) = \pi_j^*(t)P_{ji}^*(t)$ for all $i, j \in [n]$.

Hence building upon the previous algorithm we get the following:

Algorithm 2: Spectral Ranking Algorithm for the Dynamic BTL.

Input: The time grid $\mathcal{T}$, a time instance $t \in \mathcal{T}$, the comparison graphs $G_{t'}$ and the statistics $\mathbf{y}(t')$ for all $t' \in \mathcal{T}$

Output: An estimate $\boldsymbol{\pi}(t) \in \mathbb{R}_+^n$ of the true normalized weight vector $\boldsymbol{\pi}^*(t)$.

1 Choose $\delta \in \mathbb{N}$ such that the union graph $G = G_t^\delta$ is connected.
2 Compute the matrix $\mathbf{P}_\delta(t)$ as shown above.
3 Compute the leading left eigenvector $\boldsymbol{\pi}(t)$ of $\mathbf{P}_\delta(t)$.

4.3 Previous Work

The Dynamic BTL model was first introduced in [KT21]. Firstly in this paper it is assumed that

$$T = \left\{\frac{i}{T} : i = 0, \dots, T\right\} \subseteq [0, 1].$$

Secondly and more crucially it is also assumed that the comparison graphs G_t are all pairwise independent in the following sense:

$$\mathbb{P}\left[(i,j) \in E_t \cap E_s\right] = \mathbb{P}\left[(i,j) \in E_t\right]\mathbb{P}\left[(i,j) \in E_s\right] = p_t p_s.$$

With these extra assumptions and using the same algorithm described above, they get the following theorem.

Theorem 4.3.1 ([KT21]: Theorem 2). *Suppose that $G_{t'} \sim \mathcal{G}(n, p_{t'})$ for all $t \in \mathcal{T}$ so that $G_\delta \sim \mathcal{G}(n, p_\delta(t))$ with*

$$p_\delta(t) = 1 - \prod_{t' \in N_\delta(t)} (1 - p_{t'}),$$

and denote $p_{\delta,\mathrm{sum}} = \sum_{t' \in N_\delta(t)} p_{t'}$. Assume that $n \geq c_1 \log n$, $np_\delta(t) \geq c_0 \log n$ and $p_{\delta,\mathrm{sum}}(t) \geq c_2 \log n$ for some constants c_0, c_1, c_2. Then for constants C_1, C_2, if

$$2C_1\sqrt{\frac{\log n}{Lnp_\delta(t)p_{\delta,\mathrm{sum}}(t)}} + 16\frac{M\delta n}{T} \leq \frac{1}{96b^{7/2}(t)}$$

holds, we have with probability at least $1 - \frac{1}{\mathrm{poly}(n)}$ that

$$\frac{\|\boldsymbol{\pi}(t) - \boldsymbol{\pi}^*(t)\|_2}{\|\boldsymbol{\pi}^*(t)\|_2} \leq 1536\frac{M\delta n b^{7/2}(t)}{T} + 64C_2 b^{9/2}(t)\sqrt{\frac{3}{Lnp_\delta(t)p_{\delta,\mathrm{sum}}(t)}}.$$

Remark 4.3.2. Apart from the non-essential difference in the definition of the time grid $\mathcal{T}$, our assumption that the graphs $\{G_t\}$ are positively correlated is more general than the assumption of independence of the graphs $\{G_t\}$. Indeed for $\alpha = 0$ in our model we immediately get the independence model.

Moreover, we argue that our setup is a more natural setting, since usually in real life events, such as football tournaments, the next day's events are not entirely independent from the previous day.

4.4 Main Result

The aim of this section is to prove that the Dynamic Spectral Ranking Algorithm actually works in our more general setting, i.e. it provides a good estimate of $\boldsymbol{\pi}^*(t)$.

Recall that we have assumed that the graphs G_t are positively correlated (Definition 2.4.1). So, by Proposition 2.3.8, the graph $G = G_t^\delta$ is an Erdős-Rényi graph with probability

$$p = 1 - (1 - p_{t-\delta}) \prod_{t'=t-\delta+1}^{t+\delta} (1 - (1-\alpha) p_{t'}) .$$

Theorem 4.4.1. *Suppose that $p \geq k_1 \frac{\log n}{n}$ for a suitable constant $k_1 > 0$. Then there are constants $C_1, C_2 > 0$ such that if*

$$c_1 \sqrt{\frac{\delta \log n}{Lnp}} + c_2 M \delta n \leq b^{-7/2}(t), \tag{4.2}$$

for some constants $c_1, c_2 > 0$, then with probability at least $1 - \frac{1}{\mathrm{poly}(n)}$, one has

$$\frac{\|\boldsymbol{\pi}(t) - \boldsymbol{\pi}^*(t)\|_2}{\|\boldsymbol{\pi}^*(t)\|_2} \leq C_1 \sqrt{\frac{b^9(t)\delta}{Lnp}} + C_2 M \delta n b^{7/2}(t). \tag{4.3}$$

Remark 4.4.2. Note that if $M = 0$ and $\delta = 1$, Theorem 4.4.1 reduces to Theorem 3.3.1.

Remark 4.4.3. By Corollary 2.4.2, the condition $p \geq c_1 \frac{\log n}{n}$ is satisfied if

$$p_{\delta,\mathrm{sum}} = \sum_{t' \in N_\delta(t)} p_{t'} \geq \frac{\log n - \log(n - c\log n)}{1 - a} \geq c \frac{\log n}{n(1-\alpha)}. \tag{4.4}$$

Moreover, the Equation (4.4) is satisfied if $p_{t'} \geq c \frac{\log n}{n(1-\alpha)T}$.

Let $\hat{\mathbf{P}}_\delta(t) = \left[\hat{P}_{ij,\delta}(t)\right] \in \mathbb{R}_+^{n \times n}$ with

$$\hat{P}_{ij,\delta}(t) = \begin{cases} \frac{1}{d_{\delta,\max}} \hat{y}^*_{ij,\delta}(t) & \text{if } (i,j) \in E \text{ or } (j,i) \in E \\ 1 - \frac{1}{d_{\delta,\max}} \sum_{k \in N_G(i)} \hat{y}^*_{ik,\delta}(t) & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$$

Note that $\mathbb{E}\left[y_{ij,\delta}\right] = y^*_{ij,\delta}$, so by the Strong Law of Large Numbers we have that $\mathbf{P}_\delta(t) \to \hat{\mathbf{P}}_\delta(t)$ entrywise, almost surely. Now let $\boldsymbol{\Delta} = \mathbf{P}_\delta(t) - \mathbf{P}^*(t)$. Then

$$\begin{aligned} \boldsymbol{\Delta} &= \left(\mathbf{P}_\delta(t) - \hat{\mathbf{P}}_\delta(t)\right) + \left(\hat{\mathbf{P}}_\delta(t) - \mathbf{P}^*(t)\right) \\ &= \boldsymbol{\Delta}_1 + \boldsymbol{\Delta}_2. \end{aligned}$$

We need the following lemmas.

Lemma 4.4.4. *There exists a constant $C_1 \geq 1$ such that with probability at least $1 - \frac{1}{\text{poly}(n)}$, one has*

$$\|\mathbf{\Delta}_1\|_2 \leq C_1 \sqrt{\frac{N_{\delta,\max} \log n}{N_{\delta,\min}^2 d_{\delta,\max} L}},$$

where

$$N_{\delta,\max} = \max_{(i,j)\in E_t^\delta} \left|N_{ij,\delta}(t)\right| \text{ and } N_{\delta,\min} = \min_{(i,j)\in E_t^\delta} \left|N_{ij,\delta}(t)\right|$$

Proof. This is a direct application of Proposition A.1.1: Let $L_{ij} = LN_{ij,\delta}$ and

$$Z_{ij}^l = \begin{cases} \frac{1}{Ld_{\delta,\max}\left|N_{ij,\delta}(t)\right|}\left(Y_{ij}^l - y_{ij}^*\right) & \text{if } (i,j) \in E \text{ or } (j,i) \in E \\ 0 & \text{otherwise,} \end{cases}$$

where $Y_{ij}^l \sim \text{Bernoulli}\left(y_{ij}^*\right)$. Then $L_{\max} = LN_{\delta,\max}$, $B = \frac{1}{Ld_{\delta,\max}N_{\delta,\max}}$ and $N_{\max} \leq d_{\delta,\max}$ and we get exactly what we want. □

Lemma 4.4.5 ([KT21]: Lemma 1)**.** *We have that*

$$\|\mathbf{\Delta}_2\|_2 \leq 4\frac{M\delta\,|E|}{d_{\delta,\max}}.$$

Proof. The entries of $\mathbf{\Delta}_2 = \left[\Delta_{2,ij}\right]$ are given by

$$\Delta_{2,ij} = \begin{cases} \frac{1}{d_{\delta,\max}\left|N_{ij,\delta}(t)\right|} \cdot \left(\hat{y}_{ij,\delta}^*(t) - y_{ij}^*(t)\right) & \text{if } (i,j) \in E \text{ or } (j,i) \in E \\ -\sum_{k\in N_G(i)} \Delta_{2,ik} & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$$

By Assumption 4.1.1 for all $(i,j) \in E$ or $(j,i) \in E$ we have that

$$\left|\Delta_{2,ij}\right| \leq \frac{1}{d_{\delta,\max}\left|N_{ij,\delta}(t)\right|} \sum_{t'\in N_{ij,\delta}(t)} \left|y_{ij}^*(t') - y_{ij}^*(t)\right| \tag{4.5}$$

$$\leq \frac{M}{d_{\delta,\max}\left|N_{ij,\delta}(t)\right|} \sum_{t'\in N_{ij,\delta}(t)} |t' - t| \tag{4.6}$$

$$\leq \frac{M\delta}{d_{\delta,\max}}. \tag{4.7}$$

Now let $\mathbf{D}_1$ be the diagonal matrix containing the elements $\Delta_{2,ii}$ and $\mathbf{D}_2 = \mathbf{\Delta}_2 - \mathbf{D}_1$. Since $\mathbf{D}_1$ is diagonal we have

$$\begin{aligned} \|\mathbf{\Delta}_2\|_2 &\leq \|\mathbf{D}_1\|_2 + \|\mathbf{D}_1\|_2 \\ &\leq \max_i \left|\Delta_{2,ii}\right| + \|\mathbf{D}_2\|_F\,. \end{aligned}$$

By Equation (4.7) we have that

$$\|\mathbf{D}_2\|_F \leq 2\frac{M\delta\,|E|}{d_{\delta,\max}}.$$

Now, in order to bound $\|\mathbf{D}_1\|_2$ we note that

$$\begin{aligned}
|\Delta_{2,ii}| &= \left|-\sum_{k\neq i}\Delta_{2,ik}\right| \\
&\leq d_{\delta,\max}\max_{k\neq i}|\Delta_{2,ik}| \\
&\leq d_{\delta,\max}\max_{k\neq i}\frac{M\delta}{d_{\delta,\max}} \\
&\leq M\delta.
\end{aligned}$$

Hence $\|\mathbf{D}_1\|_2 \leq M\delta$. As a result

$$\begin{aligned}
\|\mathbf{D}_2\|_2 &\leq M\delta\left(1+\frac{2\,|E|}{d_{\delta,\max}}\right) \\
&\leq 4\frac{M\delta\,|E|}{d_{\delta,\max}}.
\end{aligned}$$

□

Lemma 4.4.6. *There exist constants c_1, C_1 such that if $n \geq c_1 \log n$, then with probability at least $1-\frac{1}{\mathrm{poly}(n)}$ we have that*

$$\left\|\boldsymbol{\pi}^*(t)^\top\boldsymbol{\Delta}_1\right\|_2 \leq C_1\sqrt{\frac{N_{\delta,\max}b^2(t)}{Ld_{\delta,\max}N_{\delta,\min}^2}}\,\|\boldsymbol{\pi}^*(t)\|_2\,.$$

Proof. This is a direct application of Proposition A.1.2: Construct $\boldsymbol{\Delta}$ as in Lemma 4.4.4 and let $\mathbf{a} = \boldsymbol{\pi}^*(t)$. □

Lemma 4.4.7. *If there exist constants C_1, C_2 such that*

$$C_1\sqrt{\frac{N_{\delta,\max}\log n}{N_{\delta,\min}^2 d_{\delta,\max}L}} + C_2\frac{M\delta|E|}{d_{\delta,\max}} \leq \frac{\xi d_{\delta,\min}}{4b^{7/2}(t)d_{\delta,\max}}, \tag{4.8}$$

then with probability at least $1-\frac{1}{\mathrm{poly}(n)}$ we have

$$1-\lambda_{\max}\left(\mathbf{P}^*(t)\right) - \|\boldsymbol{\Delta}\|_{\boldsymbol{\pi}^*(t)} \geq \frac{\xi d_{\delta,\min}}{4b(t)^3 d_{\delta,\max}},$$

where ξ is the spectral gap of the graph G.

Proof. By Proposition A.2.2 with $(\mathbf{P},\boldsymbol{\pi}) = (\mathbf{P}^*(t),\boldsymbol{\pi}^*(t))$, we have that

$$1-\lambda_{\max}\left(\mathbf{P}^*(t)\right) - \left\|\mathbf{P}_\delta(t)-\mathbf{P}^*(t)\right\|_{\boldsymbol{\pi}^*(t)} \geq \frac{\xi d_{\delta,\min}}{2b(t)^3 d_{\delta,\max}} - \left\|\mathbf{P}_\delta(t)-\mathbf{P}^*(t)\right\|_{\boldsymbol{\pi}^*(t)}.$$

But by Lemma 4.4.4, Lemma 4.4.5 and Equation (4.8),with probability at least $1-\frac{1}{\mathrm{poly}(n)}$

we have,

$$
\begin{aligned}
\|\mathbf{P}_\delta(t) - \mathbf{P}^*(t)\|_2 &\le \|\boldsymbol{\Delta}_1\|_2 + \|\boldsymbol{\Delta}_2\|_2 \\
&\le \sqrt{b(t)}\, \|\boldsymbol{\Delta}_1\|_{\boldsymbol{\pi}^*(t)} + 4\frac{M\delta\,|E|}{d_{\delta,\max}} \\
&\le C_1\sqrt{\frac{b(t)N_{\delta,\max}\log n}{N_{\delta,\min}^2 d_{\delta,\max} L}} + C_2\frac{M\delta\,|E|}{d_{\delta,\max}} \\
&\le \sqrt{b(t)}\left(C_1\sqrt{\frac{N_{\delta,\max}\log n}{N_{\delta,\min}^2 d_{\delta,\max} L}} + C_2\frac{M\delta\,|E|}{d_{\delta,\max}}\right) \\
&\le \frac{\xi d_{\delta,\min}}{4b^3(t)d_{\delta,\max}}.
\end{aligned}
$$

As a result,

$$
\begin{aligned}
1-\lambda_{\max}\left(\mathbf{P}^*(t)\right) - \|\boldsymbol{\Delta}\|_{\boldsymbol{\pi}^*(t)} &\ge \frac{\xi d_{\delta,\min}}{2b(t)^3 d_{\delta,\max}} - \frac{\xi d_{\delta,\min}}{4b(t)^3 d_{\delta,\max}} \\
&= \frac{\xi d_{\delta,\min}}{4b(t)^3 d_{\delta,\max}}.
\end{aligned}
$$

□

Now we have all the tools to prove the main theorem.

Proof of Theorem 4.4.1. First we assume that G is a general graph (i.e. there is no randomness) and that

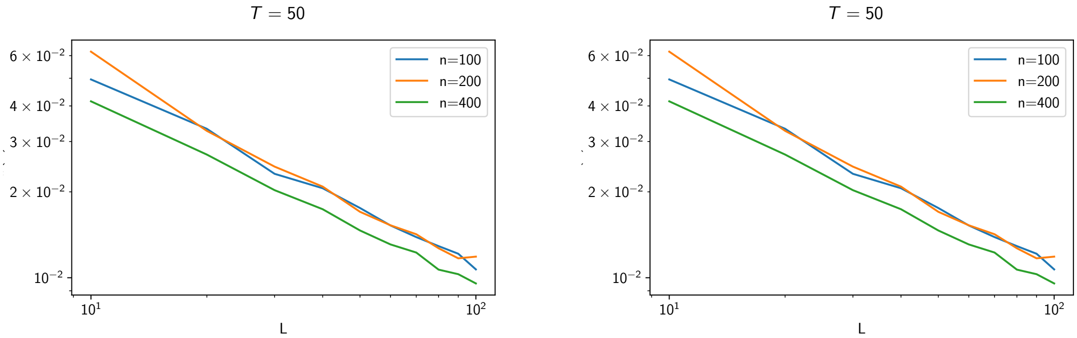
$$
C_1\sqrt{\frac{N_{\delta,\max}\log n}{N_{\delta,\min}^2 d_{\delta,\max} L}} + 4\frac{M\delta|E|}{d_{\delta,\max}} \le \frac{\xi d_{\delta,\min}}{4b^{7/2}(t)d_{\delta,\max}}. \tag{4.9}
$$

In this case we have:

$$
\begin{aligned}
\|\boldsymbol{\pi}(t)-\boldsymbol{\pi}^*(t)\|_2 &\le \frac{1}{\sqrt{\pi^*_{\min}(t)}}\|\boldsymbol{\pi}(t)-\boldsymbol{\pi}^*(t)\|_{\boldsymbol{\pi}^*(t)}, \text{ by Proposition A.3.3} \\
&\le \frac{1}{\sqrt{\pi^*_{\min}(t)}}\frac{\left\|\boldsymbol{\pi}^*(t)^\top\boldsymbol{\Delta}\right\|_{\boldsymbol{\pi}^*}}{1-\max\left\{\lambda_2(\mathbf{P}^*), |\lambda_n\left(\mathbf{P}^*\right)|\right\} - \|\boldsymbol{\Delta}\|_{\boldsymbol{\pi}^*}}, \text{ by Theorem A.3.4} \\
&\le \frac{1}{\sqrt{\pi^*_{\min}(t)}}\frac{4b^3(t)d_{\delta,\max}}{\xi d_{\delta,\min}}\left\|\boldsymbol{\pi}^*(t)^\top\boldsymbol{\Delta}\right\|_{\boldsymbol{\pi}^*}, \text{ by Lemma 4.4.7} \\
&\le \frac{4b^{7/2}(t)d_{\delta,\max}}{\xi d_{\delta,\min}}\left\|\boldsymbol{\pi}^*(t)^\top\boldsymbol{\Delta}\right\|_2, \text{ by Proposition A.3.3} \\
&\le \frac{4b^{7/2}(t)d_{\delta,\max}}{\xi d_{\delta,\min}}\left(\left\|\boldsymbol{\pi}^*(t)^\top\boldsymbol{\Delta}_1\right\|_2 + \left\|\boldsymbol{\pi}^*(t)^\top\boldsymbol{\Delta}_2\right\|_2\right) \\
&\le \frac{4b^{7/2}(t)d_{\delta,\max}}{\xi d_{\delta,\min}}\left(C_1\sqrt{\frac{N_{\delta,\max}b^2(t)}{Ld_{\delta,\max}N_{\delta,\min}^2}} + \|\boldsymbol{\Delta}_2\|_2\right)\|\boldsymbol{\pi}^*(t)\|_2, \text{ by Lemma 4.4.6}
\end{aligned}
$$

Hence by Lemma 4.4.5 we get

$$
\frac{\|\boldsymbol{\pi}(t)-\boldsymbol{\pi}^*(t)\|_2}{\|\boldsymbol{\pi}^*(t)\|_2} \le \frac{4b^{7/2}(t)d_{\delta,\max}}{\xi d_{\delta,\min}}\left(C_1\sqrt{\frac{N_{\delta,\max}b^2(t)}{Ld_{\delta,\max}N_{\delta,\min}^2}} + 4\frac{M\delta\,|E|}{d_{\delta,\max}}\right). \tag{4.10}
$$



(a) Error Metric: D (b) Error Metric: ℓ_2

Figure 4.1: Log-Log plots of the evolution of errors as L grows with fixed $T = 50$

Now using the results in Lemma 2.3.4 and the fact that $1 \leq N_{\delta,\min} \leq N_{\delta,\max} \leq |N_\delta(t)| \leq 2\delta + 1 \leq 3\delta$, the Equation (4.9) becomes Equation (4.2), and the Equation (4.10) becomes the Equation (4.3). This finishes the proof. □

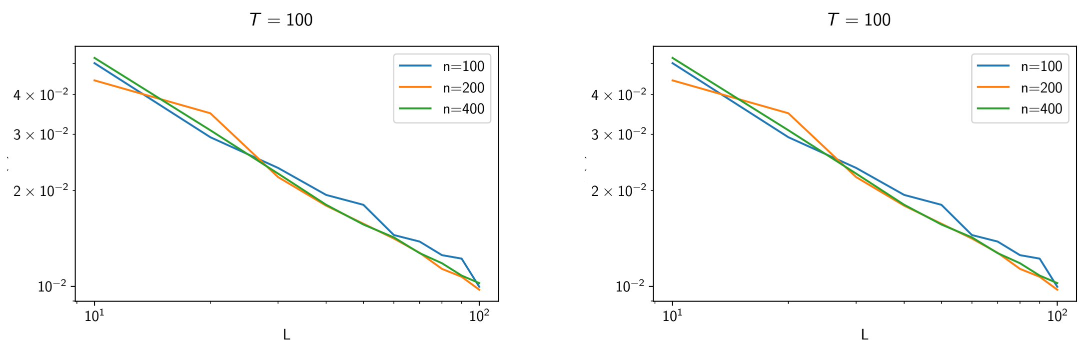
4.5 Numerical Experiments

In this section we are going to test[1] how well the Dynamic Spectral Ranking Algorithm works in practise. In order to do this we are going to create synthetic data under the Dynamic BTL Model which we will then feed to the algorithm. For the evaluation we are going to use the average ℓ_2 norm over all time instances as well as the average metric D, defined in Equation (3.5), over all time instances.

Synthetic Data. Let $\mathcal{T} = [T]$ and $n \in \mathbb{N}$. We define the true BTL weights as follows: We start with a *base weight vector* $\boldsymbol{\pi}^*$ as in the static case, ie $\pi_i^* =^{(2i-1-n)/2n}$ for some $b > 1$. Now all the true BTL weights are generated according to the normal $\mathcal{N}(\boldsymbol{\pi}^*, I_n)$. For the generation of the graphs G_t they are created as independent Erdős-Rényi graphs with probabilities $p_t \in \left[\frac{1}{n}, \frac{\log n}{n}\right]$. Then we create the statistics $\mathbf{y}(t)$ as defined by the Dynamic BTL Model.

We present two sets of plots. In both of them, we let L grow while keeping the time window constant (for $T = 50$ and $T = 100$), for $n = 100, 200, 400$. The results are depicted in Figure 4.1 and in Figure 4.2.

[1] The code is availiable at: https://github.com/dimoik96/ntua-thesis-code



(a) Error Metric: D (b) Error Metric: ℓ_2

Figure 4.2: Log-Log plots of the evolution of errors as L grows with fixed $T = 100$

Chapter 5

The Static Adversarial BTL Model

In this chapter we present another possible generalization of the Static BTL setup. In this static adversarial setting we add an adversary that is untruthful about some of the pairwise outcomes. We formally present the model and then we introduce the Static Adversarial Spectral Ranking Algorithm. Furthermore, we provide theoretical guarantees that the algorithm works with high probability.

5.1 Problem Setup

Consider the Static BTL Model, as presented in Chapter 3. To reiterate, suppose that we have n items that we want to compare and that there is a latent weight $\mathbf{w}^* \in \mathbb{R}_+^n$ associated with each item. The outcome of each pairwise comparison depends only on the weights w_i^* and w_j^*. We are also given a comparison graph $G^* = ([n], E^*)$ where $(i, j) \in E^*$ if and only if the items i and j have been compared. This is an Erdős-Rényi random graph. Moreover, for each $(i, j) \in E^*$, we assume that L independent comparisons take place between items i and j. We turn the comparison graph G^* into a weighted graph by assigning as weights y_{ij} the fraction of wins of j over i, i.e.

$$y_{ij} = \frac{1}{L} \sum_{l=1}^{L} Y_{ij}^l,$$

where

$$Y_{ij}^l \sim \text{Bernoulli}\left(\frac{w_j^*}{w_i^* + w_j^*}\right)$$

Also, as before, let

$$y_{ij}^* = \frac{w_j^*}{w_i^* + w_j^*}.$$

Now we consider a contamination model where an adversary has a complete knowledge of the truthful comparison graph G^*, as well as the true weights $\mathbf{w}^*$. This adversary can subsequently contaminate some fraction of E^* by adding new edges with arbitrary weights, deleting and corrupting existing edges and weigths. As a result, we receive as input a contaminated comparison graph $G = ([n], E)$.

Let $E_{uc} = E^* \cap E$ be the set of uncorrupted edges and $E_c = E \setminus E_{uc}$ be the set of corrupted or newly added edges. Note that $E_d = E^* \setminus E$ is the set of edges deleted by the adversary. Then the problem can be formulated as follows: Let G^* be a truthful comparison graph. Given a corrupted comparison graph G can we estimate the true BTL weights $\mathbf{w}$?

5.1.1 The contamination model for Erdős-Rényi graphs

Notation 5.1.1. Let $G = (V, E)$ be a graph and let $(S, V \setminus S)$ be a *cut*, with $S \subseteq V$. Denote the set of edges in E that "cross" the cut $(S, V \setminus S)$ by $E(S)$. Note that if $S = \{u\}$ then $E(u) := E(\{u\})$ is the set of edges in E that are incident on u, i.e. $E(u) = N_G(u)$.

In this general setting, i.e. when there is no assumption about the nature of the corruption of the graph G^*, we have the following theorem which characterizes which corrupted graphs can be recovered:

Theorem 5.1.2 ([Aga+20]: Theorem 1). *Given any arbitrary comparison graph $G = ([n], E)$ as input, it is possible to uniquely identify the true weights $\boldsymbol{\pi}^*$ in the limit $L \to \infty$, if and only if for every cut $(S, V \setminus S)$ it holds*

$$|E_{uc}(S)| > |E_c(S)|.$$

Now we present a more specific contamination model that is more suitable for the structure of an Erdős-Rényi graph. Let $G^* = ([n], E^*) \sim \mathcal{G}(n, p)$ be an Erdős-Rényi graph that corresponds to a truthful comparison graph. Of course, we assume that $p \geq c\frac{\log n}{n}$ because G^* has to be connected. We want to construct a *contaminated version* of G^* given a *contamination rate* γ in a canonical way.

Definition 5.1.3. Let $\gamma \in [0, 1)$ be the contamination rate. Consider the set $\Gamma(G^*, \gamma)$ defined as the set of all graphs $G = ([n], E)$ such that

$$\forall u \in [n] : |E_d(u) \cup E_c(u)| \leq \gamma |E^*(u)|,$$

where $E_c(u)$ is some subset of $E(u)$ (this set represents the edges that have corrupted weights) and $E_d(u) = E^*(u) \setminus E(u)$ (this set represents the deleted edges). We refer to $\Gamma(G^*, \gamma)$ as the set of all γ- *contaminated versions of* G^*.

Remark 5.1.4. Note that $\Gamma(G^*, \gamma) \neq \emptyset$ for all $\gamma \in [0, 1)$, since $\Gamma(G^*, \gamma) \subseteq \Gamma(G^*, \gamma')$ for $0 \leq \gamma \leq \gamma' < 1$ and $\Gamma(G^*, 0) = \{G^*\}$.

Now let's see what is the connection between the degrees of the truthful graph and the degrees of the contaminated graph.

Lemma 5.1.5. *Let $G \in \Gamma(G^*, \gamma)$ and let d_i^* and d_i denote the degree of vertex i of graph G^* and G, respectively. Then*

$$(1-\gamma)d_i^* \leq d_i \leq (1+\gamma)d_i^*.$$

Hence

$$(1-\gamma)|E^*| \leq |E| \leq (1+\gamma)|E^*| \text{ and } (1-\gamma)d^*_{\min} \leq d_{\min} \leq d_{\max} \leq (1+\gamma)d^*_{\max}.$$

Proof. Fix $i \in [n]$. Let

- $a_1 = |E(u) \setminus E^*(u)|$ be the number of new edges,

- $a_2 = |E^*(u) \setminus E(u)|$ be the number of deleted edges,
- $a_3 = |E_c(u) \cap E^*(u)|$ be the number of existing edges that are contaminated,
- $a_4 = |E^*(u) \setminus ((E_c(u) \cap E^*(u)) \cup (E^*(u) \setminus E(u)))|$ be the number of existing edges that are not contaminated.

Then it is easy to see that $d_i = a_1 + a_3 + a_4$, $d_i^* = a_2 + a_3 + a_4$ and $a_1 + a_2 + a_3 \le \gamma(a_2 + a_3 + a_4)$, since $G \in \Gamma(G^*, \gamma)$. Hence

$$\begin{aligned} d_i &= a_1 + a_3 + a_4 \\ &\le a_1 + a_2 + a_3 + a_4 \\ &\le \gamma(a_2 + a_3 + a_4) + a_4 \\ &\le (1+\gamma)(a_2 + a_3 + a_4) \\ &= (1+\gamma)d_i^*. \end{aligned}$$

Similarly $d_i \ge (1-\gamma)d_i^*$. □

Remark 5.1.6. As in the previous proof we can show that if $G \in \Gamma(G^*, \gamma)$ then $|E_c(u)| \le \gamma|E(u)|$ for all u.

Goal. To sum up, the problem setup is that we assume that there is a truthful comparison Erdős-Rényi graph G^* (which we have no access) and given a contamination rate γ we have access to a connected γ-contaminated version G. We want to estimate the normalized true BTL weights

$$\pi^* = \frac{\mathbf{w}^*}{\|\mathbf{w}^*\|_1}.$$

Remark 5.1.7. Note that this model is a generalization of the model presented in Chapter 3. Indeed, if we take $\gamma = 0$ then $\Gamma(G^*, 0) = \{G^*\}$, so we recover exactly the Static BTL model.

In this contamination setting for Erdős-Rényi random graphs, Theorem 5.1.2 takes the following form.

Theorem 5.1.8 ([Aga+20]: Theorem 2). *Let $\varepsilon > 0$ and $0 < \gamma < 1/4 - \varepsilon$. Then there exists a sufficiently large constant $c > 1$, such that if $G^* \sim \mathcal{G}(n,p)$ with $p \ge c\frac{\log n}{n}$ and $G \in \Gamma(G^*, \gamma)$ then with probability at least $1 - \frac{1}{\text{poly}(n)}$, the cut-majority condition described in Theorem 5.1.2 is satisfied for every cut in G, and as a consequence, the true weights π^* are uniquely identifiable as $L \to \infty$. Conversely, if the corruption rate $\gamma \ge 1/4 + \varepsilon$, then with probability at least $1 - \frac{1}{\text{poly}(n)}$, there exists a choice of adversarial corruption such that the cut-majority condition described in Theorem 5.1.2 is violated for at least one cut in G, rendering the true weights unidentifiable, even as $L \to \infty$.*

5.2 Previous Work

In this section we go over some previous work on this problem. In particular, we briefly explain the Adversarially Robust Recovery algorithm as presented in [Aga+20], that attempts to solve the problem in hand. The basic idea is to somehow find which edges are most probably corrupted and remove them. Then we "hope" that the resulting graph is connected and apply the Accelerated Spectral Ranking Algorithm ([APA18]).

Now let's make everything precise. Let $p \geq c_1 \frac{\log n}{n}$, for some sufficiently large constant $c_1 > 0$ and let $G^* = ([n], E^*) \sim \mathcal{G}(n,p)$ be an Erdős-Rényi graph that corresponds to a truthful comparison graph. Let $0 \leq \gamma \leq \gamma_{LP} = c_2 \frac{\log(np)}{\log n}$ be a contamination rate, where c_2 is a constant. Now let $G \in \Gamma(G^*, \gamma)$ be a connected contaminated version of G. In order to identify which edges have been corrupted we introduce the variables $x(e) \in [0,1]$ which indicate whether an edge is corrupted. Intuitively they can be interpreted as follows: the higher the value of $x(e)$ the higher the probability that e is corrupted.

Definition 5.2.1 ([Aga+20]: Definition 3)**.** Given a (simple) cycle $C = (v_1, \ldots, v_l, v_1)$ of length l in G, we call C approximately consistent if

$$\frac{1-(2l-1)\varepsilon_L}{1+\varepsilon_L} \leq \prod_{i=1}^{l} \frac{y_{v_i,v_{i+1}}}{y_{v_{i+1},v_i}} \leq \frac{1+\varepsilon_L}{1-(2l-1)\varepsilon_L},$$

where $\varepsilon_L = (1+b)\sqrt{\log n / L}$, and inconsistent otherwise. Let $\mathbb{C}$ denote the set of all inconsistent cycles in G.

Now we consider the following Linear Program (LP) which identifies corrupted edges:

$$\min \sum_{e \in E} x(e)$$

$$\begin{aligned}
\text{subject to:} \quad & \sum_{e \in C} x(e) \geq 1, \quad \forall C \in \mathbb{C} \\
& \sum_{e \in E(u)} x(e) \leq \gamma\,|E(u)| \leq \gamma_{LP}\,|E(u)|\,, \quad \forall u \in [n] \\
& 0 \leq x(e) \leq 1, \quad \forall e \in E.
\end{aligned}$$

Lemma 5.2.2 ([Aga+20]: Lemma 1)**.** *The above LP is solvable in $O\left(n^{2+o(1)} d_{\text{avg}}^6\right)$ time where d_{avg} is the average degree of G.*

Now that we "know" which edges are corrupted, we prune the given graph G according to a solution of the above LP, as follows: Given any feasible solution $\mathbf{x}$ to the above LP, let $E_{lpr} = \{e \in E : x(e) \geq \log(np)/(5 \log n)\}$ be the set of edges with "large" $x(e)$ values. Then we delete all edges in E_{lpr} from G, resulting in a "cleaned" comparison graph $\widetilde{G} = \left([n], \widetilde{E}\right)$. The pruned graph is connected with high probability, as shown by the following lemma.

Lemma 5.2.3 ([Aga+20]: Lemma 2)**.** *With probability at least $1 - \frac{1}{\text{poly}(n)}$, we have that the pruned graph $\widetilde{G}$ is connected and furthermore contains **no** edges from*

$$E_A = \left\{ (i,j) \in E : \left| y_{ij} - y_{ij}^* \right| > \ell_{n,p} \sqrt{\frac{\log n}{L}} \right\},$$

where $\ell_{n,p} = 4\left(4 + \frac{\log n}{\log(np)}\right)(1+b)$.

At this point, since we have a connected graph $\widetilde{G}$ we can apply the Accelerated Spectral Ranking Algorithm ([APA18]). Hence we have Algorithm 3.

This algorithm satisfies the following.

Algorithm 3: Adversarially Robust Recover Algorithm for the Static Adversarial BTL Model.

Input: Items $[n]$, graph $G = (V, E)$, parameters p, b and L.
Output: An estimate $\widetilde{\pi} \in \mathbb{R}_+^n$ of the true normalized weight vector $\widetilde{\pi}^*$.

1 Construct a LP as above and solve it: $\mathbf{x} \leftarrow$ solution of LP.
2 For all $(i,j) \in E$, let $\hat{x}(i,j) = \mathbb{1}_{\{x(i,j) \geq \log(np)/(5\log n)\}}$.
3 If $\hat{x}(i,j) = 1$, then remove the edge (i,j) and create the graph $\widetilde{G}$.
4 Return the output of Accelerated Spectral Ranking ([APA18]) algorithm on this pruned dataset.

Theorem 5.2.4 ([Aga+20]: Theorem 3)**.** *Given an input comparison graph $G = (V, E)$ conforming to the contamination model described in Section 5.1.1 with Erdős-Rényi graph parameter $p \geq \frac{k\log n}{n}$ for any k larger than some sufficiently large constant, true BTL weights π^*, and number of samples per pair L; if the corruption rate per vertex $\gamma \leq \frac{\log(np)}{125\log n}$, then there is an efficient algorithm that, with probability at least $1 - \frac{1}{\text{poly}(n)}$, recovers an estimate π such that*

$$\|\pi - \pi^*\|_1 \leq Cb^2 \log b \sqrt{\frac{\log n}{L}},$$

for an absolute constant C.

5.3 Static Adversarial Spectral Ranking Algorithm

The main drawback in the previous algorithm is that in Theorem 5.2.4 there is no immediate reduction of the result into the Static BTL Theorem (Theorem 3.3.1) when $\gamma = 0$, as one might would anticipate. Below we modify the previous algorithm by applying the Spectral Ranking Algorithm (Algorithm 1) in the last step, instead of the Accelerated Spectral Ranking Algorithm.

Recall that we have a connected pruned graph $\widetilde{G} = ([n], \widetilde{E})$. Let $\widetilde{\mathbf{P}} = \left[\widetilde{P}_{ij}\right] \in \mathbb{R}_+^{n\times n}$ with

$$\widetilde{P}_{ij} = \begin{cases} \frac{1}{d_{\max}} \cdot y_{ij} & \text{if } (i,j) \in \widetilde{E} \text{ or } (j,i) \in \widetilde{E} \\ 1 - \frac{1}{d_{\max}} \sum_{k \in N_{\widetilde{G}}(i)} y_{ik} & \text{if } i = j \\ 0 & \text{otherwise,} \end{cases}$$

where y_{ij} are the corrupted statistics, i.e. the weights of the corrupted graph G, and $d_{\max} = d_{\max}(G)$ is the maximum degree of the graph G. Note that $d_{\max}(G) \geq d_{\max}(\widetilde{G})$.

Also let $\widetilde{\mathbf{P}}^* = [\widetilde{P}^*_{ij}] \in \mathbb{R}_+^{n\times n}$ with

$$\widetilde{P}^*_{ij} = \begin{cases} \frac{1}{d_{\max}} \cdot y^*_{ij} & \text{if } (i,j) \in \widetilde{E} \text{ or } (j,i) \in \widetilde{E} \\ 1 - \frac{1}{d_{\max}} \sum_{k \in N_{\widetilde{G}}(i)} y^*_{ik} & \text{if } i = j \\ 0 & \text{otherwise,} \end{cases}$$

It is easy to see that the normalized weight vector $\widetilde{\boldsymbol{\pi}}^* = \boldsymbol{\pi}^*$ is the stationary distribution of $\widetilde{\mathbf{P}}^*$, since it satisfies the detailed balance equation $\widetilde{\pi}_i^* \widetilde{P}_{ij}^* = \widetilde{\pi}_j^* \widetilde{P}_{ji}^*$ for all $i, j \in [n]$.

In summary, we have the following algorithm:

Algorithm 4: Spectral Ranking Algorithm for the Static Adversarial BTL Model.

Input: The corrupted comparison graph G and the corrupted statistics $\mathbf{y}$.
Output: An estimate $\widetilde{\boldsymbol{\pi}} \in \mathbb{R}_+^n$ of the true normalized weight vector $\widetilde{\boldsymbol{\pi}}^*$.

Construct a LP as above and solve it: $\mathbf{x} \leftarrow$ solution of LP.
For all $(i,j) \in E$, let $\hat{x}(i,j) = \mathbb{1}_{\{x(i,j) \geq \log(np)/(5\log n)\}}$.
If $\hat{x}(i,j) = 1$, then remove the edge (i,j) and create the graph $\widetilde{G}$.
Compute the matrix $\widetilde{\mathbf{P}}$ as shown above.
Compute its leading left eigenvector $\widetilde{\boldsymbol{\pi}}$.

5.4 Main Result

Now we show that the Static Adversarial Spectral Ranking Algorithm actually works, i.e. it provides a good estimate of $\widetilde{\boldsymbol{\pi}}^* = \boldsymbol{\pi}^*$.

Theorem 5.4.1. *Suppose that $p \geq k_1 \frac{\log n}{n}$ and $\gamma \leq k_2 \frac{\log(np)}{\log n}$ for suitable constants $k_1, k_2 > 0$. Then there are constants $C_1, C_2 > 0$ such that if*

$$L \geq b^7 \left(c_1 \sqrt{\frac{\log n}{(1-\gamma)np}} + c_2 \ell_{n,p} \frac{\gamma}{1-\gamma} \sqrt{n^2 \log n} \right)^2, \tag{5.1}$$

for some constants $c_1, c_2 > 0$, then with probability at least $1 - \frac{1}{\mathrm{poly}(n)}$, one has

$$\frac{\|\widetilde{\boldsymbol{\pi}} - \widetilde{\boldsymbol{\pi}}^*\|_2}{\|\widetilde{\boldsymbol{\pi}}^*\|_2} \leq C_1 \frac{b^{9/2}}{\sqrt{(1-\gamma)npL}} + C_2 \ell_{n,p} \frac{\gamma}{1-\gamma} \sqrt{\frac{n^2 \log n}{L}}, \tag{5.2}$$

where $\ell_{n,p} = 4\left(4 + \frac{\log n}{\log(np)}\right)(1+b)$.

Remark 5.4.2. For $\gamma = 0$, i.e. when we have no corruption, Theorem 5.4.1 reduces to Theorem 3.3.1. Note that in this case $\widetilde{G}$ is a just a connected subgraph of G^* so the basic algorithm will still give us an estimate of the truthful BTL weights $\mathbf{w}$.

The rest of the chapter is dedicated to a proof of the above theorem. Let $G = G_{uc} \cup G_c$, where $G_{uc} = ([n], E_{uc})$ is the subgraph of G that contains all the uncorrupted edges of G and $G_c = ([n], E_c)$ are the rest of them, the corrupted ones. Similarly, let $\widetilde{G} = \widetilde{G}_{uc} \cup \widetilde{G}_c$, with $\widetilde{G}_{uc} = ([n], \widetilde{E}_{uc})$ and $\widetilde{G}_c = ([n], \widetilde{E}_c)$ be the uncorrupted and corrupted decomposition of the pruned graph. Note that even if some corrupted edges were pruned, there is still the possibility that there are some left. By Remark 5.1.6 we get $d_i(G_c) \leq \gamma d_i$ and obviously $d_i(\widetilde{G}_c) \leq d_i(G_c)$, since we are not adding any new edges. Hence $d_i(\widetilde{G}_c) \leq \gamma d_i$, so $d_{\max}(\widetilde{G}_c) \leq \gamma d_{\max}$.

Let $\widetilde{\mathbf{P}}_{uc} = \left[\widetilde{P}_{uc,ij}\right] \in \mathbb{R}_{+}^{n\times n}$ with

$$\widetilde{P}_{uc,ij} = \begin{cases} \dfrac{1}{d_{\max}} \cdot y_{ij} & \text{if } (i,j) \in \widetilde{E}_{uc} \text{ or } (j,i) \in \widetilde{E}_{uc} \\ 1 - \dfrac{1}{d_{\max}} \displaystyle\sum_{k \in N_{\widetilde{G}_{uc}}(i)} y_{ik} & \text{if } i = j \\ 0 & \text{otherwise,} \end{cases}$$

and $\widetilde{\mathbf{P}}_{c} = \left[\widetilde{P}_{c,ij}\right] \in \mathbb{R}_{+}^{n\times n}$ with

$$\widetilde{P}_{c,ij} = \begin{cases} \dfrac{1}{d_{\max}} \cdot y_{ij} & \text{if } (i,j) \in \widetilde{E}_{c} \text{ or } (j,i) \in \widetilde{E}_{c} \\ -\dfrac{1}{d_{\max}} \displaystyle\sum_{k \in N_{\widetilde{G}_{c}}(i)} y_{ik} & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$$

Note that $\widetilde{\mathbf{P}} = \widetilde{\mathbf{P}}_{uc} + \widetilde{\mathbf{P}}_{c}$.

Similarly let $\widetilde{\mathbf{P}}^{*}_{uc} = [\widetilde{P}^{*}_{uc,ij}] \in \mathbb{R}_{+}^{n\times n}$ with

$$\widetilde{P}^{*}_{uc,ij} = \begin{cases} \dfrac{1}{d_{\max}} \cdot y^{*}_{ij} & \text{if } (i,j) \in \widetilde{E}_{uc} \text{ or } (j,i) \in \widetilde{E}_{uc} \\ 1 - \dfrac{1}{d_{\max}} \displaystyle\sum_{k \in N_{\widetilde{G}_{uc}}(i)} y^{*}_{ik} & \text{if } i = j \\ 0 & \text{otherwise,} \end{cases}$$

and $\widetilde{\mathbf{P}}^{*}_{c} = [\widetilde{P}^{*}_{c,ij}] \in \mathbb{R}_{+}^{n\times n}$ with

$$\widetilde{P}^{*}_{c,ij} = \begin{cases} \dfrac{1}{d_{\max}} \cdot y^{*}_{ij} & \text{if } (i,j) \in \widetilde{E}_{c} \text{ or } (j,i) \in \widetilde{E}_{c} \\ -\dfrac{1}{d_{\max}} \displaystyle\sum_{k \in N_{\widetilde{G}_{c}}(i)} y^{*}_{ik} & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$$

Note that $\widetilde{\mathbf{P}}^{*} = \widetilde{\mathbf{P}}^{*}_{uc} + \widetilde{\mathbf{P}}^{*}_{c}$.

Now let

$$\mathbf{\Delta} = \widetilde{\mathbf{P}} - \widetilde{\mathbf{P}}^{*} = \left(\widetilde{\mathbf{P}}_{uc} - \widetilde{\mathbf{P}}^{*}_{uc}\right) + \left(\widetilde{\mathbf{P}}_{c} - \widetilde{\mathbf{P}}^{*}_{c}\right) = \mathbf{\Delta}_{uc} + \mathbf{\Delta}_{c}.$$

Then

$$\|\mathbf{\Delta}\|_2 \leq \|\mathbf{\Delta}_{uc}\|_2 + \|\mathbf{\Delta}_{c}\|_2 .$$

For the proof we need the following lemmas.

Lemma 5.4.3. *There exists a constant $C_1 \geq 1$ such that with probability at least $1 - \frac{1}{\mathrm{poly}(n)}$, it holds*

$$\|\mathbf{\Delta}_{uc}\|_2 \leq C_1 \sqrt{\frac{\log n}{d_{\max} L}}.$$

Proof. Apply Proposition A.1.1 with $L_{ij} = L$ and

$$Z_{ij}^l = \begin{cases} \frac{1}{Ld_{\max}} \left(Y_{ij}^l - y_{ij}^*\right) & \text{if } (i,j) \in E \text{ or } (j,i) \in E \\ 0 & \text{otherwise,} \end{cases}$$

where $Y_{ij}^l \sim \text{Bernoulli}\left(y_{ij}^*\right)$. Then $L_{\max} = L$, $B = \frac{1}{d_{\max}L}$ and $N_{\max} = d_{\max}(\widetilde{G}_{uc}) \leq d_{\max}(\widetilde{G}) \leq d_{\max}(G)$. □

Lemma 5.4.4. *It holds that*

$$\|\boldsymbol{\Delta}_c\|_2 \leq 4\gamma\ell_{n,p}\frac{|E|}{d_{\max}}\sqrt{\frac{\log n}{L}}.$$

Proof. The entries of $\boldsymbol{\Delta}_c$ are given by

$$(\Delta_c)_{ij} = \begin{cases} \frac{1}{d_{\max}} \cdot \left(y_{ij} - y_{ij}^*\right) & \text{if } (i,j) \in \widetilde{E}_c \text{ or } (j,i) \in \widetilde{E}_c \\ -\sum\limits_{k \in N_{\widetilde{G}_c}(i)} (\Delta_c)_{ik} & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$$

By Lemma 5.2.3 we have that

$$\left|y_{ij} - y_{ij}^*\right| \leq \ell_{n,p}\sqrt{\frac{\log n}{L}}, \quad \forall (i,j),(j,i) \in \widetilde{E}_c.$$

Let $\mathbf{D}$ be the diagonal matrix containing the elements $(\boldsymbol{\Delta}_c)_{ii}$ and $\mathbf{D}' = \boldsymbol{\Delta}_c - \mathbf{D}$. As $\mathbf{D}$ is diagonal we have

$$\|\boldsymbol{\Delta}_c\|_2 \leq \|\mathbf{D}\|_2 + \|\mathbf{D}'\|_2 \leq \max_i |(\Delta)_{ii}| + \|\mathbf{D}'\|_F.$$

Let us bound $\|\mathbf{D}'\|_F$. We have that

$$\left|(\Delta_c)_{ij}\right| \leq \frac{1}{d_{\max}}\ell_{n,p}\sqrt{\frac{\log n}{L}},$$

so

$$\begin{aligned} \|\mathbf{D}'\|_F &\leq \frac{2\left|\widetilde{E}_c\right|}{d_{\max}}\ell_{n,p}\sqrt{\frac{\log n}{L}} \\ &\leq \frac{2\gamma|E|}{d_{\max}}\ell_{n,p}\sqrt{\frac{\log n}{L}}, \end{aligned}$$

since

$$2\left|\widetilde{E}_c\right| = \sum d_i\left(\widetilde{G}_c\right) \leq \sum \gamma d_i = 2\gamma|E|.$$

In order to bound $\|\mathbf{D}\|_2$, we simply note that

$$\begin{aligned} |(\Delta_c)_{ii}| &= \left|-\sum_{k \in N_{\widetilde{G}_c}(i)} (\Delta_c)_{ik}\right| \\ &\leq d_{\max}(\widetilde{G}_c)\max_{j \neq i}\left|(\Delta_c)_{ij}\right| \\ &\leq \frac{d_{\max}(\widetilde{G}_c)}{d_{\max}}\ell_{n,p}\sqrt{\frac{\log n}{L}} \\ &\leq \gamma\ell_{n,p}\sqrt{\frac{\log n}{L}}. \end{aligned}$$

Hence

$$\|\boldsymbol{\Delta}_c\|_2 \le \left(1 + \frac{2|E|}{d_{\max}}\right)\gamma\ell_{n,p}\sqrt{\frac{\log n}{L}} \le \frac{4|E|}{d_{\max}}\gamma\ell_{n,p}\sqrt{\frac{\log n}{L}}. \qquad \square$$

Lemma 5.4.5. *Since $n \ge \frac{k_1}{p}\log n$ we have that with probability at least $1 - \frac{1}{\text{poly}(n)}$ it holds*

$$\left\|\widetilde{\boldsymbol{\pi}}^{*\top}\boldsymbol{\Delta}_{uc}\right\|_2 \le \frac{Cb}{\sqrt{d_{\max}L}}\|\widetilde{\boldsymbol{\pi}}^*\|_2 .$$

Proof. This is a direct application of Proposition A.1.2: Construct $\boldsymbol{\Delta}$ as in the proof of Lemma 5.4.3 and let $\mathbf{a} = \widetilde{\boldsymbol{\pi}}^*$ to get exactly what we want. $\square$

Lemma 5.4.6. *If there exist constants C_1, C_2 such that*

$$C_1\sqrt{\frac{\log n}{d_{\max}L}} + C_2\gamma\ell_{n,p}\frac{|E|}{d_{\max}}\sqrt{\frac{\log n}{L}} \le \frac{\tilde{\xi}\tilde{d}_{\min}}{4b^{7/2}\tilde{d}_{max}}, \tag{5.3}$$

then with probability at least $1 - \frac{1}{\text{poly}(n)}$ we have

$$1 - \lambda_{\max}(\widetilde{\mathbf{P}}^*) - \|\boldsymbol{\Delta}\|_{\widetilde{\boldsymbol{\pi}}^*} \ge \frac{\tilde{\xi}\tilde{d}_{\min}}{4b^{7/2}\tilde{d}_{max}},$$

where $\tilde{\xi}$ is the spectral gap of the pruned graph $\widetilde{G}$.

Proof. Use Proposition A.2.2 in combination with Lemma 5.4.3 and finish the proof as in Lemma 3.3.5. $\square$

Lemma 5.4.7 ([Aga+20]: Lemma 7)**.** *Let $\widetilde{G}$ be the pruned graph. Then there is a constant C_1 such that with probability at least $1 - \frac{1}{\text{poly}(n)}$ one has*

$$\tilde{\xi} \ge C_1,$$

where $\tilde{\xi}$ is the spectral gap of $\widetilde{G}$. Moreover, there is a constant C_2 such that

$$\max_{i,j\in[n]}\frac{\tilde{d}_i}{\tilde{d}_j} \le C_2.$$

Now we have all the tools to prove the main theorem.

Proof of Theorem 5.4.1. We have

$$\begin{aligned}
\|\widetilde{\boldsymbol{\pi}} - \widetilde{\boldsymbol{\pi}}^*\|_2 &\le \frac{1}{\sqrt{\widetilde{\pi}^*_{\min}}}\|\widetilde{\boldsymbol{\pi}} - \widetilde{\boldsymbol{\pi}}^*\|_{\widetilde{\boldsymbol{\pi}}^*}, \text{ by Proposition A.3.3} \\
&\le \frac{1}{\sqrt{\widetilde{\pi}^*_{\min}}}\frac{\left\|\widetilde{\boldsymbol{\pi}}^{*\top}\boldsymbol{\Delta}\right\|_{\widetilde{\boldsymbol{\pi}}^*}}{1 - \lambda_{\max}\left(\widetilde{\mathbf{P}}^*\right) - \|\boldsymbol{\Delta}\|_{\widetilde{\boldsymbol{\pi}}^*}}, \text{ by Theorem A.3.4} \\
&\le \frac{1}{\sqrt{\widetilde{\pi}^*_{\min}}}\frac{4b^3\tilde{d}_{\max}}{\tilde{\xi}\tilde{d}_{\min}}\left\|\widetilde{\boldsymbol{\pi}}^{*\top}\boldsymbol{\Delta}\right\|_{\widetilde{\boldsymbol{\pi}}^*}, \text{ by Lemma 5.4.6} \\
&\le \frac{4b^{7/2}\tilde{d}_{\max}}{\tilde{\xi}\tilde{d}_{\min}}\left\|\widetilde{\boldsymbol{\pi}}^{*\top}\boldsymbol{\Delta}\right\|_2, \text{ by Proposition A.3.3} \\
&\le \frac{4b^{7/2}\tilde{d}_{\max}}{\tilde{\xi}\tilde{d}_{\min}}\left(\left\|\widetilde{\boldsymbol{\pi}}^{*\top}\boldsymbol{\Delta}_{uc}\right\|_2 + \left\|\widetilde{\boldsymbol{\pi}}^{*\top}\boldsymbol{\Delta}_c\right\|_2\right)
\end{aligned}$$

Hence by Cauchy-Schwarz, Lemma 5.4.5 and Lemma 5.4.4 we have

$$\|\widetilde{\boldsymbol{\pi}} - \widetilde{\boldsymbol{\pi}}^*\|_2 \leq \frac{4b^{7/2}\tilde{d}_{\max}}{\tilde{\xi}\tilde{d}_{\min}}\left(\frac{Cb}{\sqrt{d_{\max}L}} + 4\gamma\ell_{n,p}\frac{|E|}{d_{\max}}\sqrt{\frac{\log n}{L}}\right)\|\widetilde{\boldsymbol{\pi}}^*\|_2 \,. \tag{5.4}$$

Now by Lemma 5.4.7, Lemma 2.3.4 and Lemma 5.1.5, Equation (5.3) turns into Equation (5.1) and Equation (5.4) turns into Equation (5.2), as wanted. □

Chapter 6

The Dynamic Adversarial BTL Model

In this chapter we are going to unify all the BTL models that we have discussed so far. The new model called *Dynamic Adversarial BTL Model* is essentially the Dynamic BTL Model with the presence of an adversary. In particular, in this setup all of the previous models are just a special case of this one. Then we will combine Algorithm 2 with Algorithm 4 into one algorithm, named *Dynamic Adversarial Spectral Ranking Algorithm*. Moreover, we are going to prove that our algorithm works with high probability.

6.1 Problem Setup

Consider the Dynamic BTL Model, as discussed in Chapter 4. To reiterate, suppose that we have a time grid $\mathcal{T}$ and n items that we want to compare. For each time instance $t \in \mathcal{T}$ we assume that here is a latent weight $\mathbf{w}^*(t) \in \mathbb{R}^n_+$ associated with each item. The outcome of each pairwise comparison at the time t depends only on the weights $w_i^*(t)$ and $w_j^*(t)$. Moreover, in order to get a meaningful recovery, we make the additional assumption:

Assumption 6.1.1. There exists $M \geq 0$ such that

$$\left| \frac{w_j^*(t)}{w_i^*(t) + w_j^*(t)} - \frac{w_j^*(t')}{w_i^*(t') + w_j^*(t')} \right| \leq M |t - t'|, \tag{6.1}$$

for all $t, t' \in \mathcal{T}$ and $i \neq j \in [n]$.

We are also given a sequence of comparison graphs $G_t^* = ([n], E_t^*)$ where $(i,j) \in E_t^*$ if and only if the items i and j have been compared at the time instance $t \in \mathcal{T}$. Furthermore, we assume that the graphs $\{G_t^*\}_{t=1}^T$ are positively correlated Erdős-Rényi graphs, as descibed in Section 2.4. We make no assumption about the connectivity of the graphs G_t^*. Moreover, for each $(i,j) \in E_t^*$, we assume that L independent comparisons take place between items i and j at $t \in \mathcal{T}$. Then we turn each comparison graph G_t^* into a weighted graph by assigning as weights $y_{ij}(t)$ the fraction of wins of j over i at $t \in \mathcal{T}$, i.e.

$$y_{ij}(t) = \frac{1}{L} \sum_{l=1}^{L} Y_{ij}^l(t),$$

where

$$Y^l_{ij}(t) \sim \text{Bernoulli}\left(\frac{w^*_j(t)}{w^*_i(t)+w^*_j(t)}\right).$$

Also, as before, let

$$y^*_{ij}(t) = \frac{w^*_j(t)}{w^*_i(t)+w^*_j(t)}.$$

Adversary. Now we assume that there is an adversary having complete knowledge of all truthful graphs G^*_t, as well as the true weights $\mathbf{w}^*(t)$. This adversary can subsequently contaminate some fraction of each E^*_t by adding new edges with arbitrary weights, deleting and corrupting existing edges and weigths. More formally, after fixing the truthful graphs G^*_t the adversary chooses $G_t \in \Gamma(G^*_t, \gamma_t)$, γ_t-contaminated versions of G^*_t. Finally, we assume that the union of all the contaminated graphs $\cup_{t=1}^T G_t$ is connected.

Goal. The goal is that given a $t \in \mathcal{T}$, we want to *estimate* the normalized weight vector

$$\pi^*(t) = \frac{\mathbf{w}^*(t)}{\|\mathbf{w}^*(t)\|_1},$$

using the corrupted comparison graphs $\{G_t\}_{t=1}^T$.

Remark 6.1.2. Note that this model is indeed a generalization of all the previous models.

1. If $G^*_t = G^*$, $\gamma_t = \gamma$ and $G_t = G \in \Gamma(G^*, \gamma)$ for all $t \in \mathcal{T}$ and $M = 0$, then we get the Static Adversarial BTL Model, presented in Chapter 5.
2. If $\gamma_t = 0$, then $G_t = G^*_t$ for all $t \in \mathcal{T}$. Hence we get the Dynamic BTL Model, presented in Chapter 4.
3. If $G^*_t = G^*$ and $\gamma_t = 0$ then we have that $G_t = G^*$. If moreover $M = 0$, then we get the Static BTL model, presented in Chapter 3.

6.2 Dynamic Adversarial Spectral Ranking Algorithm

We are going to generalize the Spectral Ranking Algorithm (Section 3.2) in this setting. The idea is to consider the union of the graphs G_t in a time neighborhood such that the union is connected and then apply the Static Adversarial Spectral Ranking Algorithm (Algorithm 4).

Recall that $N_\delta(t) = \{t-\delta, \dots, t, \dots, t+\delta\} \cap \mathcal{T}$ for some $\delta \in \mathbb{N}$. Now let $G^\delta_t = ([n], E^\delta_t)$, where $E^\delta_t = \cup_{t' \in N_\delta(t)} E_{t'}$, be the union graph that corresponds to the time neighborhood $N_\delta(t)$. We will abuse the notation and we will denote the union graph just by G. Note that the union graph G is actually a contaminated version of G^*, the union of the truthful comparison graphs, according to the following proposition.

Proposition 6.2.1. *Let $G^*_t = ([n], E^*_t)$ for $t = 1, \dots, T$ be a sequence of (truthful) graphs and let $G_t = ([n], E_t) \in \Gamma(G^*_t, \gamma_t)$ be a γ_t-contaminated version of G^*_t for all $t = 1, \dots, T$. Let $G^* = ([n], E^*) = \cup_{t=1}^T G^*_t = ([n], \cup_{t=1}^T E^*_t)$ and $G = ([n], E) = \cup_{t=1}^T G_t = ([n], \cup_{t=1}^T E_t)$ be the truthful union graph and the contaminated union graph respectively. Then $G \in \Gamma(G^*, \gamma)$ for $\gamma = \sum_{t=1}^T \gamma_t$, or in words the union of contaminated version of truthful graphs is a contaminated version of the union of truthful graphs.*

Proof. Since $G_t \in \Gamma(G_t^*, \gamma_t)$ we have $|E_{t,d}(u) \cup E_{t,c}(u)| \leq \gamma_t |E_t^*(u)|$ for all $u \in [n]$. Now let $E_c(u) = \cup_{t=1}^T E_{t,c}(u)$ and note that $E^*(u) = \cup_{t=1}^T E_t^*(u)$ and $E(u) = \cup_{t=1}^T E_t(u)$. We have

$$\begin{aligned} E_d(u) &= E^*(u) \setminus E(u) \\ &= \left(\bigcup_{t=1}^T E_t^*(u) \right) \setminus \left(\bigcup_{t=1}^T E_t(u) \right) \\ &\subseteq \bigcup_{t=1}^T E_t^*(u) \setminus E_t(u) \\ &= \bigcup_{t=1}^T E_{t,d}(u) \end{aligned}$$

Hence

$$\begin{aligned} |E_d(u) \cup E_c(u)| &\leq |\bigcup_{t=1}^T E_{t,d}(u) \cup \bigcup_{t=1}^T E_{t,c}(u)| \\ &= |\bigcup_{t=1}^T \big(E_{t,d}(u) \cup E_{t,c}(u)\big)| \\ &\leq \sum_{t=1}^T |E_{t,d}(u) \cup E_{t,c}(u)| \\ &\leq \sum_{t=1}^T \gamma_t |E_t^*(u)| \\ &\leq \left(\sum_{t=1}^T \gamma_t \right) |E^*(u)|, \end{aligned}$$

as wanted. □

Recall that $N_{ij,\delta}(t) = \{t' \in N_\delta(t) | (i,j) \in E_{t'}\}$ is the set of the time instances in $N_\delta(t)$ where i and j are being compared. Then for the union graph $G = G_t^\delta$, consider

$$\begin{aligned} y_{ij,\delta}(t) &= \frac{1}{|N_{ij,\delta}(t)|} \sum_{t' \in N_{ij,\delta}(t)} y_{ij}(t'), \\ y_{ij,\delta}^*(t) &= \frac{1}{|N_{ij,\delta}(t)|} \sum_{t' \in N_{ij,\delta}(t)} y_{ij}^*(t'), \end{aligned}$$

where $y_{ij}(t')$ are the corrupted statistics, i.e. the weights of the corrupted graphs $G_{t'}$.

Notation 6.2.2. At this point we fix a time instance $t \in \mathcal{T}$. Assume that we want to estimate the true BTL weights $\boldsymbol{\pi}^*(t)$. From now on we will drop t (and sometimes δ) from the notation in order to not obfuscate the presentation.

Now that we know that $G \in \Gamma(G^*, \gamma)$ we can follow the procedure described in Section 5.3, i.e. create a LP and prune from the graph G all the edges with "high" $x(e)$. Let $\widetilde{G} = \big([n], \widetilde{E}\big)$ be the pruned subgraph of the union graph G. $\widetilde{G}$ is connected with high probability, see Lemma 5.2.3.

Let $\widetilde{\mathbf{P}}_\delta = \left[\widetilde{P}_{ij,\delta}\right] \in \mathbb{R}_+^{n\times n}$ with

$$\widetilde{P}_{ij,\delta} = \begin{cases} \frac{1}{d_{\delta,\max}} \cdot y_{ij,\delta} & \text{if } (i,j) \in \widetilde{E} \text{ or } (j,i) \in \widetilde{E} \\ 1 - \frac{1}{d_{\delta,\max}} \cdot \sum_{k \in N_{\widetilde{G}}(i)} y_{ik,\delta} & \text{if } i = j \\ 0 & \text{otherwise,} \end{cases}$$

where $d_{\delta,max}$ is the maximum degree of the *unpruned* union graph G.

Also let $\widetilde{\mathbf{P}}^* = \left[\widetilde{P}^*_{ij}\right] \in \mathbb{R}_+^{n\times n}$ with

$$\widetilde{P}^*_{ij} = \begin{cases} \frac{1}{d_{\delta,\max}} \cdot y^*_{ij} & \text{if } (i,j) \in \widetilde{E} \text{ or } (j,i) \in \widetilde{E} \\ 1 - \frac{1}{d_{\delta,\max}} \sum_{k \in N_{\widetilde{G}}(i)} y^*_{ik} & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$$

Observe that $\widetilde{\mathbf{P}}^*$ is time reversible with stationary distribution $\widetilde{\boldsymbol{\pi}}^* = \boldsymbol{\pi}^*$, since it satisfies the detailed balance equation $\widetilde{\pi}^*_i \widetilde{P}^*_{ij} = \widetilde{\pi}^*_j \widetilde{P}^*_{ji}$ for all $i, j \in [n]$.

In summary, we have the following algorithm:

Algorithm 5: Spectral Ranking Algorithm for the Dynamic Adversarial BTL Model.

Input: The time grid $\mathcal{T}$, a time instance $t \in \mathcal{T}$, the corrupted graphs $G_{t'}$ and the corrupted statistics $\mathbf{y}(t')$ for all $t' \in \mathcal{T}$

Output: An estimate $\widetilde{\boldsymbol{\pi}} \in \mathbb{R}_+^n$ of the true normalized weight vector $\widetilde{\boldsymbol{\pi}}^*$.

1. Choose $\delta \in \mathbb{N}$ such that the union graph $G = G_t^\delta$ is connected.
2. Construct and solve a LP with respect to G as in Section 5.3. Let $\mathbf{x}$ be a solution.
3. For all $(i,j) \in E$, let $\hat{x}(i,j) = \mathbb{1}_{\{x(i,j) \geq \log(np)/(5\log n)\}}$.
4. If $\hat{x}(i,j) = 1$, then remove the edge (i,j) from G and create the graph $\widetilde{G}$.
5. Compute the matrix $\widetilde{\mathbf{P}}_\delta$ as shown above.
6. Compute the leading left eigenvector $\widetilde{\boldsymbol{\pi}}$ of $\widetilde{\mathbf{P}}_\delta$.

6.3 Main Results

First we present a modification of Theorem 5.1.2 for our setup. This theorem gives the optimal information theoretic bound for γ in our setup.

Theorem 6.3.1. *Let $G_t^* \sim \mathcal{G}(n, p_t)$ for $t = 1, \dots, T$ be a sequence of Erdős-Rényi that correspond to truthful comparison graphs. Let $G_t \in \Gamma(G_t^*, \gamma_t)$, for all $t = 1, \dots, T$. Let $\varepsilon > 0$ and suppose that $\sum_t \gamma_t < \frac{1}{4} - \varepsilon$. Then there exists a sufficiently large constant $c > 1$ such that if $\sum_t p_t \geq \log n - \log(n - c\log n)$, then with probability at least $1 - 1/\text{poly}(n)$, the cut-majority condition described in Theorem 5.1.2 is satisfied for every cut in $G = \cup_t G_t$, and as a consequence, the true weights $\mathbf{w}$ are uniquely identifiable as $L \to \infty$.*

Proof. This follows immediately from Proposition 2.3.9, Proposition 6.2.1 and Theorem 5.1.8. □

Now we show that the Dynamic Adversarial Spectral Ranking Algorithm actually works, i.e. it provides a good estimate of $\widetilde{\pi}^* = \pi^*$. Recall that G^* is the truthful union graph, which is an Erdős-Rényi graph with probability p. Then G is the contaminated union graph, which is a γ-contaminated version of G. Finally, $\widetilde{G}$ is the pruned graph. The parameter p is given by

$$p = 1 - (1 - p_{t-\delta}) \prod_{t'=t-\delta+1}^{t+\delta} (1 - (1-\alpha)\, p_{t'})$$

and the parameter γ is given by

$$\gamma = \sum_{t'=t-\delta}^{t+\delta} \gamma_{t'}.$$

Theorem 6.3.2. *Suppose that $p \geq k_1 \frac{\log n}{n}$ and $\gamma \leq k_2 \frac{\log(np)}{\log n}$ for suitable constants $k_1, k_2 > 0$. Then there are constants $C_1, C_2, C_3 > 0$ such that if*

$$c_1 \sqrt{\frac{\delta \log n}{(1-\gamma)Lnp}} + c_2 \ell_{n,p} \frac{\gamma}{1-\gamma} \sqrt{\frac{n^2 \log n}{L}} + c_3 M n \delta \leq b^{-7/2}, \tag{6.2}$$

for some constants $c_1, c_2, c_3 > 0$, then with probability at least $1 - \frac{1}{\text{poly}(n)}$, one has

$$\frac{\|\widetilde{\pi} - \widetilde{\pi}^*\|_2}{\|\widetilde{\pi}^*\|_2} \leq C_1 \sqrt{\frac{b^9 \delta}{(1-\gamma)Lnp}} + C_2 \ell_{n,p} \frac{\gamma}{1-\gamma} \sqrt{\frac{n^2 \log n}{L}} + C_3 M n b^{7/2} \delta, \tag{6.3}$$

where $\ell_{n,p} = 4\left(4 + \frac{\log n}{\log(np)}\right)(1+b)$.

Remark 6.3.3.

1. For $G_t^* = G^*$, $\gamma_t = \gamma$, $G_t = G \in \Gamma(G^*, \gamma)$ and $M = 0$, the Theorem 6.3.2 reduces to Theorem 5.4.1.
2. For $\gamma_t = 0$ for all $t \in \mathcal{T}$ we get $\gamma = 0$ so the Theorem 6.3.2 reduces to Theorem 4.4.1.
3. For $G_t^* = G^*$ and $\gamma_t = 0$, the Theorem 6.3.2 reduces to Theorem 3.3.1.

The rest of the chapter is dedicated to a proof of the above theorem. Let $\widehat{\widetilde{\mathbf{P}}}_\delta = \left[\widehat{\widetilde{P}}_{ij,\delta}\right] \in \mathbb{R}_+^{n\times n}$ with

$$\widehat{\widetilde{P}}_{ij,\delta} = \begin{cases} \frac{1}{d_{\delta,\max}} \cdot y_{ij,\delta}^* & \text{if } (i,j) \in \widetilde{E} \text{ or } (j,i) \in \widetilde{E} \\ 1 - \frac{1}{d_{\delta,\max}} \cdot \sum_{k \in N_{\widetilde{G}}(i)} y_{ik,d}^* & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$$

As in Chapter 5 let $G = G_{uc} \cup G_c$ with $G_{uc} = ([n], E_{uc})$ and $G_c = ([n], E_c)$ and also let $\widetilde{G} = \widetilde{G}_{uc} \cup \widetilde{G}_c$ with $\widetilde{G}_{uc} = ([n], \widetilde{E}_{uc})$ and $\widetilde{G}_c = ([n], \widetilde{E}_c)$ be the uncorrupted and corrupted decompositions of the union graph and the pruned union graph. By Remark 5.1.6 we get $d_i(G_c) \leq \gamma d_i$ and obviously $d_i(\widetilde{G}_c) \leq d_i(G_c)$, since we are not adding any new edges. Hence $d_i(\widetilde{G}_c) \leq \gamma d_i$ and $d_{\max}(\widetilde{G}_c) \leq \gamma d_{\max}$.

Let $\widetilde{\mathbf{P}}_{uc,\delta} = \left[\widetilde{P}_{ij,uc,\delta}\right] \in \mathbb{R}_+^{n\times n}$ with

$$\widetilde{P}_{ij,uc,\delta} = \begin{cases} \frac{1}{d_{\delta,\max}} \cdot y_{ij,\delta} & \text{if } (i,j) \in \widetilde{E}_{uc} \text{ or } (j,i) \in \widetilde{E}_{uc} \\ 1 - \frac{1}{d_{\delta,\max}} \sum_{k \in N_{\widetilde{G}_{uc}}(i)} y_{ik,\delta} & \text{if } i = j \\ 0 & \text{otherwise,} \end{cases}$$

and $\widetilde{\mathbf{P}}_{c,\delta} = \left[\widetilde{P}_{ij,c,\delta}\right] \in \mathbb{R}_+^{n\times n}$ with

$$\widetilde{P}_{ij,c,\delta} = \begin{cases} \frac{1}{d_{\delta,\max}} \cdot y_{ij,\delta} & \text{if } (i,j) \in \widetilde{E}_{c} \text{ or } (j,i) \in \widetilde{E}_{c} \\ -\frac{1}{d_{\delta,\max}} \sum_{k \in N_{\widetilde{G}_{c}}(i)} y_{ik,\delta} & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$$

Note that $\widetilde{\mathbf{P}}_{\delta} = \widetilde{\mathbf{P}}_{uc,\delta} + \widetilde{\mathbf{P}}_{c,\delta}$.

Similarly, let $\widehat{\widetilde{\mathbf{P}}}_{uc,\delta}(t) = \left[\widehat{\widetilde{P}}_{ij,uc,\delta}\right] \in \mathbb{R}_+^{n\times n}$ with

$$\widehat{\widetilde{P}}_{ij,uc,\delta} = \begin{cases} \frac{1}{d_{\delta,\max}} \cdot y^*_{ij,\delta} & \text{if } (i,j) \in \widetilde{E}_{uc} \text{ or } (j,i) \in \widetilde{E}_{uc} \\ 1 - \frac{1}{d_{\delta,\max}} \sum_{k \in N_{\widetilde{G}_{uc}}(i)} y^*_{ik,\delta} & \text{if } i = j \\ 0 & \text{otherwise,} \end{cases}$$

and $\widehat{\widetilde{\mathbf{P}}}_{c,\delta} = \left[\widehat{\widetilde{P}}_{ij,c,\delta}\right] \in \mathbb{R}_+^{n\times n}$ with

$$\widehat{\widetilde{P}}_{ij,c,\delta} = \begin{cases} \frac{1}{d_{\delta,\max}} \cdot y^*_{ij,\delta} & \text{if } (i,j) \in \widetilde{E}_{c} \text{ or } (j,i) \in \widetilde{E}_{c} \\ -\frac{1}{d_{\delta,\max}} \sum_{k \in N_{\widetilde{G}_{c}}(i)} y^*_{ik,\delta} & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$$

Note that $\widehat{\widetilde{\mathbf{P}}}_{\delta}(t) = \widehat{\widetilde{\mathbf{P}}}_{uc,\delta} + \widehat{\widetilde{\mathbf{P}}}_{c,\delta}$.

Let

$$\begin{aligned} \mathbf{\Delta} &= \widetilde{\mathbf{P}}_{\delta} - \widetilde{\mathbf{P}}^*_{\delta} \\ &= \left(\widetilde{\mathbf{P}}_{\delta} - \widehat{\widetilde{\mathbf{P}}}_{\delta}\right) + \left(\widehat{\widetilde{\mathbf{P}}}_{\delta} - \widetilde{\mathbf{P}}^*_{\delta}\right) \\ &= \left(\widetilde{\mathbf{P}}_{uc,\delta} - \widehat{\widetilde{\mathbf{P}}}_{uc,\delta}\right) + \left(\widetilde{\mathbf{P}}_{c,\delta} - \widehat{\widetilde{\mathbf{P}}}_{c,\delta}\right) + \left(\widehat{\widetilde{\mathbf{P}}}_{\delta} - \widetilde{\mathbf{P}}^*_{\delta}\right) \\ &= \mathbf{\Delta}_1 + \mathbf{\Delta}_2 + \mathbf{\Delta}_3. \end{aligned}$$

For the proof we need the following lemmas.

Lemma 6.3.4. *There exists a constant $C_1 \geq 1$ such that with probability at least $1 - \frac{1}{\text{poly}(n)}$, it holds*

$$\|\mathbf{\Delta}_1\|_2 \leq C_1 \sqrt{\frac{N_{\delta,\max} \log n}{N^2_{\delta,\min} d_{\delta,\max} L}}.$$

Proof. Apply Proposition A.1.1 with $L_{ij} = LN_{ij,\delta}$ and

$$Z_{ij}^{l} = \begin{cases} \frac{1}{Ld_{\delta,\max}\left|N_{ij,\delta}\right|}\left(Y_{ij}^{l} - y_{ij}^{*}\right) & \text{if } (i,j) \in \widetilde{E}_{uc} \text{ or } (j,i) \in \widetilde{E}_{uc} \\ 0 & \text{otherwise,} \end{cases}$$

where $Y_{ij}^{l} \sim \text{Bernoulli}\left(y_{ij}^{*}\right)$. Then $L_{\max} = LN_{\delta,\max}$, $B = \frac{1}{Ld_{\delta,\max}N_{\delta,\max}}$ and

$$N_{\max} = d_{\max}(\widetilde{G}_{uc}) \leq d_{\max}(\widetilde{G}) \leq d_{\max}(G). \qquad \square$$

Lemma 6.3.5. *It holds that*

$$\|\mathbf{\Delta}_2\|_2 \leq 4\gamma\ell_{n,p}\frac{|E|}{d_{\delta,\max}}\sqrt{\frac{\log n}{L}}.$$

Proof. The entries of $\mathbf{\Delta}_2$ are given by

$$(\Delta_2)_{ij} = \begin{cases} \frac{1}{d_{\delta,\max}} \cdot \left(y_{ij,\delta}(t) - y_{ij,\delta}^{*}(t)\right) & \text{if } (i,j) \in \widetilde{E}_c \text{ or } (j,i) \in \widetilde{E}_c \\ -\sum_{k \in N_{\widetilde{G}_c}(i)} (\Delta_2)_{ik} & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$$

By Lemma 5.2.3 we have that

$$\begin{aligned} \left|y_{ij,\delta} - y_{ij,\delta}^{*}\right| &\leq \frac{1}{\left|N_{ij,\delta}\right|}\sum_{t' \in N_{ij,\delta}}\left|y_{ij} - y_{ij}^{*}\right| \\ &\leq \ell_{n,p}\sqrt{\frac{\log n}{L}}, \end{aligned}$$

for all $(i,j) \in \widetilde{E}_c$. Let $\mathbf{D}$ be the diagonal matrix containing the elements $(\Delta_2)_{ii}$ and $\mathbf{D}' = \mathbf{\Delta}_2 - \mathbf{D}$. As $\mathbf{D}$ is diagonal we have

$$\|\mathbf{\Delta}_2\|_2 \leq \|\mathbf{D}\|_2 + \|\mathbf{D}'\|_2 \leq \max_i\left|(\Delta_2)_{ii}\right| + \|\mathbf{D}'\|_F.$$

Let us bound $\|\mathbf{D}'\|_F$. We have that

$$\left|(\Delta_2)_{ij}\right| \leq \frac{1}{d_{\delta,\max}}\ell_{n,p}\sqrt{\frac{\log n}{L}},$$

so

$$\|\mathbf{D}'\|_F \leq \frac{2\left|\widetilde{E}_c\right|}{d_{\delta,\max}}\ell_{n,p}\sqrt{\frac{\log n}{L}} \leq \frac{2\gamma|E|}{d_{\delta,\max}}\ell_{n,p}\sqrt{\frac{\log n}{L}},$$

since $2\left|\widetilde{E}_c\right| = \sum d_i\left(\widetilde{G}_c\right) \leq \sum \gamma d_i = 2\gamma|E|$. In order to bound $\|\mathbf{D}\|_2$, we simply note that

$$\begin{aligned} \left|(\Delta_2)_{ii}\right| &= \left|-\sum_{k \in N_{\widetilde{G}_c}(i)} (\Delta_2)_{ik}\right| \\ &\leq d_{\max}\left(\widetilde{G}_c\right)\max_{j \neq i}\left|(\Delta_2)_{ij}\right| \\ &\leq \frac{d_{\max}\left(\widetilde{G}_c\right)}{d_{\delta,\max}}\ell_{n,p}\sqrt{\frac{\log n}{L}} \\ &\leq \gamma\ell_{n,p}\sqrt{\frac{\log n}{L}}. \end{aligned}$$

Hence

$$\|\mathbf{\Delta}_2\|_2 \leq \left(1 + \frac{2\,|E|}{d_{\delta,\max}}\right)\gamma\ell_{n,p}\sqrt{\frac{\log n}{L}} \leq \frac{4\,|E|}{d_{\delta,\max}}\gamma\ell_{n,p}\sqrt{\frac{\log n}{L}}. \qquad \square$$

Lemma 6.3.6. *It holds that*

$$\|\mathbf{\Delta}_3\|_2 \leq 4\frac{M\delta\,|E|}{d_{\delta,\max}}.$$

Proof. It is the same proof as in Lemma 4.4.5. □

Lemma 6.3.7. *Since* $n \geq \frac{k_1}{p}\log n$ *we have that with probability at least* $1 - \frac{1}{\mathrm{poly}(n)}$ *it holds*

$$\left\|\widetilde{\boldsymbol{\pi}}^{*\top}\mathbf{\Delta}_1\right\|_2 \leq Cb\sqrt{\frac{N_{\delta,\max}}{Ld_{\delta,\max}N^2_{\delta,\min}}}\,\|\widetilde{\boldsymbol{\pi}}^*\|_2\,.$$

Proof. This is a direct application of Proposition A.1.2: Construct $\mathbf{\Delta}$ as in the proof of Lemma 6.3.4 and let $\mathbf{a} = \widetilde{\boldsymbol{\pi}}^*$ to get exactly what we want. □

Lemma 6.3.8. *If there exist constants* C_1, C_2, C_3 *such that*

$$C_1\sqrt{\frac{N_{\delta,\max}\log n}{N^2_{\delta,\min}d_{\delta,\max}L}} + C_2\gamma\ell_{n,p}\frac{|E|}{d_{\delta,\max}}\sqrt{\frac{\log n}{L}} + C_3\frac{M\delta\,|E|}{d_{\delta,\max}} \leq \frac{\tilde{\xi}\tilde{d}_{\min}}{4b^{7/2}\tilde{d}_{max}}, \tag{6.4}$$

then with probability at least $1 - \frac{1}{\mathrm{poly}(n)}$ *we have*

$$1 - \lambda_{\max}(\widetilde{\mathbf{P}}^*) - \|\mathbf{\Delta}\|_{\widetilde{\boldsymbol{\pi}}^*} \geq \frac{\tilde{\xi}\tilde{d}_{\min}}{4b^{7/2}\tilde{d}_{max}},$$

where $\tilde{\xi}$ *is the spectral gap of the pruned graph* $\widetilde{G}$.

Proof. Use Proposition A.2.2 in combination with Lemma 6.3.4 and finish the proof as in Lemma 4.4.7. □

Now we have all the tools to prove the main theorem.

Proof of Theorem 6.3.2. We have

$$\begin{aligned}
\|\widetilde{\boldsymbol{\pi}} - \widetilde{\boldsymbol{\pi}}^*\|_2 &\leq \frac{1}{\sqrt{\widetilde{\pi}^*_{\min}}}\,\|\widetilde{\boldsymbol{\pi}} - \widetilde{\boldsymbol{\pi}}^*\|_{\widetilde{\boldsymbol{\pi}}^*}\,, \text{ by Proposition A.3.3}\\
&\leq \frac{1}{\sqrt{\widetilde{\pi}^*_{\min}}}\frac{\left\|\widetilde{\boldsymbol{\pi}}^{*\top}\mathbf{\Delta}\right\|_{\widetilde{\boldsymbol{\pi}}^*}}{1 - \lambda_{\max}\left(\widetilde{\mathbf{P}}^*\right) - \|\mathbf{\Delta}\|_{\widetilde{\boldsymbol{\pi}}^*}}, \text{ by Theorem A.3.4}\\
&\leq \frac{1}{\sqrt{\widetilde{\pi}^*_{\min}}}\frac{4b^3\tilde{d}_{\max}}{\tilde{\xi}\tilde{d}_{\min}}\left\|\widetilde{\boldsymbol{\pi}}^{*\top}\mathbf{\Delta}\right\|_{\widetilde{\boldsymbol{\pi}}^*}, \text{ by Lemma 5.4.6}\\
&\leq \frac{4b^{7/2}\tilde{d}_{\max}}{\tilde{\xi}\tilde{d}_{\min}}\left\|\widetilde{\boldsymbol{\pi}}^{*\top}\mathbf{\Delta}\right\|_2, \text{ by Proposition A.3.3}\\
&\leq \frac{4b^{7/2}\tilde{d}_{\max}}{\tilde{\xi}\tilde{d}_{\min}}\left(\left\|\widetilde{\boldsymbol{\pi}}^{*\top}\mathbf{\Delta}_1\right\|_2 + \left\|\widetilde{\boldsymbol{\pi}}^{*\top}\mathbf{\Delta}_2\right\|_2 + \left\|\widetilde{\boldsymbol{\pi}}^{*\top}\mathbf{\Delta}_3\right\|_2\right)
\end{aligned}$$

Hence by Cauchy-Schwarz, Lemma 6.3.5, Lemma 6.3.6 and Lemma 6.3.7 we have

$$\|\widetilde{\boldsymbol{\pi}} - \widetilde{\boldsymbol{\pi}}^*\|_2 \leq \frac{4b^{7/2}\tilde{d}_{\max}}{\tilde{\xi}\tilde{d}_{\min}} \left(Cb\sqrt{\frac{N_{\delta,\max}}{Ld_{\delta,\max}N_{\delta,\min}^2}} + 4\gamma\ell_{n,p}\frac{|E|}{d_{\delta,\max}}\sqrt{\frac{\log n}{L}} + 4\frac{M\delta\,|E|}{Td_{\delta,\max}} \right) \|\widetilde{\boldsymbol{\pi}}^*\|_2 \,. \tag{6.5}$$

Now by Lemma 5.4.7, Lemma 2.3.4, Lemma 5.1.5 and the fact that $1 \leq N_{\delta,\min} \leq N_{\delta,\max} \leq |N_\delta(t)| \leq 2\delta + 1 \leq 3\delta$, Equation (6.4) turns into Equation (6.2) and Equation (6.5) turns into Equation (6.3), as wanted. □

Chapter 7

Conclusion

7.1 Summary

In this thesis we gave an overview of the most recent results in the theory of ranking distribution. In particular, we presented the Static, Dynamic and Adversarial BTL models and we gave algorithms that solve each problem efficiently. Moreover, we proposed a more general, unified model, the *Dynamic Adversarial BTL model*, where each of the previous models is just a special case of our setup. Finally, we provided an algorithm that solves the most general problem and we proved that it works with high probability.

7.2 Future Work

Here are a few questions that have arisen during the writing of this thesis.

- The Theorem 5.4.1 makes the "unnatural" assumption that $\gamma \leq \gamma_{LP} = O\left(\frac{\log(np)}{\log n}\right)$. But from Theorem 5.1.8 we know that the maximum γ is $1/4$. Why do we have this gap? Can we improve it?
- Can we extend our results to other models other than the BTL Model? Another popular ranking model is the Mallows Model ([Mal57]).
- Another possible question to examine is whether we can use a different kind of random graphs. In this thesis we worked only with Erdős-Rényi graphs. However, there are many more kinds of random graphs, such as the Barabási–Albert Model, the Bianconi–Barabási Model, the Random Geometric Graph (RGG) and the Random Exponential Graph (REG).
- Throught our work in the dynamic setting we have only worked with discrete time grids. Can we generalize our results using a *continuous* time grid such as $T = [0, 1]$?
- Another interesting direction is to examine other kinds of adversarial corruption.

Appendix A

Technical Tools

In this chapter we give detailed proofs of some technical results that we repeatedly use in the main chapters.

A.1 Results about a special kind of random matrices

Proposition A.1.1. *Let $n \geq 1$ and $L_{i,j} \in \mathbb{N}$ for all $i < j \in [n]$. Let $\mathbf{\Delta} = \left[\Delta_{ij}\right] \in \mathbb{R}^{n \times n}$ be a matrix defined by*

$$\Delta_{ij} = \begin{cases} \sum_{l=1}^{L_{ij}} Z_{ij}^{l} & \text{if } i < j \\ -\Delta_{ji} & \text{if } i > j \\ -\sum_{k \neq i} \Delta_{ik} & \text{if } i = j \end{cases},$$

where Z_{ij}^{l} are random variables such that:

- *Z_{ij}^{l} are independent for all i, j, l.*
- *$\mathbb{E}\left[Z_{ij}^{l}\right] = 0$.*
- *$\left|Z_{ij}^{l}\right| \leq B$.*

Then there exists a constant $C \geq 12$ such that with probability at least $1 - \frac{1}{\text{poly}(n)}$, it holds

$$\|\mathbf{\Delta}\|_2 \leq C\sqrt{B^2 N_{\max} L_{\max} \log n}, \tag{A.1}$$

where

$$N_i = \left\{j \in [n] \setminus \{i\} \mid \Delta_{ij} \neq 0\right\}, N_{\max} = \max_i |N_i| \text{ and } L_{\max} = \max_{i,j} L_{ij}.$$

Proof. Let $\mathbf{D} = \text{diag}\{\Delta_{11}, \dots, \Delta_{nn}\}$ be the diagonal matrix with entries the main diagonal of $\mathbf{\Delta}$ and let $\mathbf{\Delta}' = \mathbf{\Delta} - \mathbf{D}$. Then $\mathbf{\Delta} = \mathbf{D} + \mathbf{\Delta}'$, so $\|\mathbf{\Delta}\|_2 \leq \|\mathbf{D}\|_2 + \|\mathbf{\Delta}'\|_2$. Note that $\mathbf{\Delta}'$ is skew-symmetric. We will bound both $\|\mathbf{D}\|_2$ and $\|\mathbf{\Delta}'\|_2$ by $C\sqrt{B^2 N_{\max} L_{\max} \log n}$.

Bounding $\|\mathbf{D}\|_2$: Since $\mathbf{D}$ is diagonal we have $\|\mathbf{D}\|_2 = \max_i |\Delta_{ii}|$, and moreover by definition it is $\Delta_{ii} = -\sum_{k \in N_i} \sum_{l=1}^{L_{ik}} Z_{ik}^l$. Hence by Hoeffding's inequality (Theorem 2.2.3) we get

$$\begin{aligned}
\mathbb{P}\left[|\Delta_{ii}| > t\right] &\leq 2\exp\left\{-\frac{2t^2}{\sum_{k \in N_i}\sum_{l=1}^{L_{ik}} (2B)^2}\right\} \\
&= 2\exp\left\{-\frac{t^2}{2B^2 \sum_{k \in N_i} L_{ik}}\right\} \\
&\leq 2\exp\left\{-\frac{t^2}{2B^2 N_{\max} L_{\max}}\right\}.
\end{aligned}$$

Then for $t = C\sqrt{B^2 N_{\max} L_{\max} \log n}$ we have:

$$\mathbb{P}\left[|\Delta_{ii}| > C\sqrt{B^2 N L_{\max} \log n}\right] \leq 2n^{-C^2/2}.$$

Now using the union bound and the above inequality we get:

$$\begin{aligned}
\mathbb{P}\left[\|\mathbf{D}\|_2 > C\sqrt{B^2 N L_{\max} \log n}\right] &\leq \sum_{i=1}^{n} \mathbb{P}\left[|\Delta_{ii}| > C\sqrt{B^2 N L_{\max} \log n}\right] \\
&\leq n \cdot 2n^{-C^2/2} \\
&= 2n^{-(C^2/2-1)}.
\end{aligned}$$

Since $C^2/2 - 1 \geq 1$ we have that with probability at least $1 - \frac{1}{\text{poly}(n)}$ it holds

$$\|\mathbf{D}\|_2 \leq C\sqrt{B^2 N L_{\max} \log n}. \tag{A.2}$$

Thus we have the wanted bound for $\|\mathbf{D}\|_2$.

Bounding $\|\mathbf{\Delta}'\|_2$: We will discriminate two cases. Firstly, assume that $N_{\max} \leq \log n$.

Recall the following standard inequality

$$\|\mathbf{\Delta}'\|_2 \leq \sqrt{\|\mathbf{\Delta}'\|_1 \|\mathbf{\Delta}'\|_\infty} = \|\mathbf{\Delta}'\|_\infty,$$

since $\mathbf{\Delta}'$ is skew-symmetric. Let

$$R_i = \sum_{j \neq i} \left|\Delta_{ij}\right| = \sum_{j \in N_i} \left|\sum_{l=1}^{L_{ij}} Z_{ij}^l\right|$$

be the absolute row sum of the $\mathbf{\Delta}'$. Then by definition $\|\mathbf{\Delta}'\|_\infty = \max_i R_i$. Moreover let $\mathcal{S}_i = \left\{(\rho_1, \dots, \rho_{N_i}) | \rho_j \in \{-1, 1\}\right\}$. Obviously $|\mathcal{S}_i| = 2^{|N_i|}$. Now using the union bound and

Hoeffding's inequality (Theorem 2.2.3) we have

$$\begin{aligned}
\mathbb{P}\left[R_i > t\right] &\leq \sum_{\rho\in\mathcal{S}_i} \mathbb{P}\left[\sum_{j\in N_i} \rho_j \sum_{l=1}^{L_{ij}} Z_{ij}^l > t\right] \\
&\leq \sum_{\rho\in\mathcal{S}_i} \exp\left\{-\frac{2t^2}{\sum_{j\in N_i}\sum_{l=1}^{L_{ij}} (2B)^2}\right\} \\
&\leq \sum_{\rho\in\mathcal{S}_i} \exp\left\{-\frac{t^2}{2B^2 N_{\max} L_{\max}}\right\} \\
&= 2^{|N_i|} \exp\left\{-\frac{t^2}{2B^2 N_{\max} L_{\max}}\right\} \\
&\leq \exp\left\{N_{\max}\log 2 - \frac{t^2}{2B^2 N_{\max} L_{\max}}\right\}.
\end{aligned}$$

Setting $t = \frac{C}{2}\sqrt{B^2 N_{\max} L_{\max}\left(\log n + N_{\max}\log 2\right)}$ in the above we get

$$\begin{aligned}
&\mathbb{P}\left[R_i > \frac{C}{2}\sqrt{B^2 N_{\max} L_{\max} (\log n + N_{\max}\log 2)}\right] \\
&\leq \exp\left\{N_{\max}\log 2 - \frac{C^2/4B^2 N_{\max} L_{\max}\left(\log n + N_{\max}\log 2\right)}{2B^2 N_{\max} L_{\max}}\right\} \\
&= \exp\left\{N_{\max}\log 2 - \frac{C^2\left(\log n + N_{\max}\log 2\right)}{8}\right\} \\
&= n^{-C^2/8} 2^{N_{\max}(1-C^2/8)} \\
&\leq n^{-C^2/8},
\end{aligned}$$

since $C \geq 2\sqrt{2}$. Finally, by the union bound and the previous inequality we have

$$\mathbb{P}\left[\|\mathbf{\Delta}'\|_\infty > \frac{C}{2}\sqrt{B^2 N L_{\max}\left(\log n + N\log 2\right)}\right] \leq n \cdot n^{-C^2/8} = n^{-(C^2/8-1)}.$$

Hence with probability at least $1 - \frac{1}{\text{poly}(n)}$ we have

$$\begin{aligned}
\|\mathbf{\Delta}'\|_2 &\leq \|\mathbf{\Delta}'\|_\infty \\
&\leq \frac{C}{2}\sqrt{B^2 N_{\max} L_{\max}\left(\log n + N\log 2\right)} \\
&\leq \frac{C}{2}\sqrt{B^2 N_{\max} L_{\max}\left(\log n + \log n\log 2\right)} \\
&\leq C\sqrt{B^2 N_{\max} L_{\max}\log n},
\end{aligned}$$

as wanted.

Now assume that $N_{\max} \geq \log n$. For each $i < j$ with $j \in N_i$, let $\mathbf{U}_{ij}^l \in \mathbb{R}^{n\times n}$ with all entries equal to 0 except for

$$\left(\mathbf{U}_{ij}^l\right)_{ij} = Z_{ij}^l \text{ and } \left(\mathbf{U}_{ij}^l\right)_{ji} = -Z_{ij}^l.$$

Then

$$\boldsymbol{\Delta}' = \sum_{\substack{i<j \\ j\in N_i}} \sum_{l=1}^{L_{ij}} \mathbf{U}_{ij}^l$$

and also $\mathbb{E}\left[\mathbf{U}_{ij}^l\right] = 0$ and $\left\|\mathbf{U}_{ij}^l\right\|_2 \leq \left\|\mathbf{U}_{ij}^l\right\|_F \leq \sqrt{2}B$. Since $\mathbf{U}_{ij}^l$ are skew-symmetric and independent matrices we have

$$\begin{aligned}
\nu &= \max\left\{\left\|\mathbb{E}\left[\sum_{\substack{i<j \\ j\in N_i}} \sum_{l=1}^{L_{ij}} \left(\mathbf{U}_{ij}^l\right)^\top \mathbf{U}_{ij}^l\right]\right\|_2, \left\|\mathbb{E}\left[\sum_{\substack{i<j \\ j\in N_i}} \sum_{l=1}^{L_{ij}} \mathbf{U}_{ij}^l \left(\mathbf{U}_{ij}^l\right)^\top\right]\right\|_2\right\} \\
&= \left\|\sum_{\substack{i<j \\ j\in N_i}} \sum_{l=1}^{L_{ij}} \mathbb{E}\left[\left(\mathbf{U}_{ij}^l\right)^2\right]\right\|_2.
\end{aligned}$$

But the matrices $\left(\mathbf{U}_{ij}^l\right)^2$ are diagonals with only two non zero entries, which are at the positions (i,i) and (j,j), and they are equal to $\left(Z_{ij}^l\right)^2$. Then

$$\mathbb{E}\left[\left(\mathbf{U}_{ij}^l\right)^2\right]_{i,i} = \mathbb{E}\left[\left(\mathbf{U}_{ij}^l\right)^2\right]_{j,j} = \mathbb{E}\left[\left(Z_{ij}^l\right)^2\right] \leq B^2.$$

Thus $\sum_{\substack{i<j \\ j\in N_i}} \sum_{l=1}^{L_{ij}} \mathbb{E}\left[\left(\mathbf{U}_{ij}^l\right)^2\right]$ is a diagonal matrix, so

$$\nu = \max_{k\in[n]} \left|\left(\sum_{\substack{i<j \\ j\in N_i}} \sum_{l=1}^{L_{ij}} \mathbb{E}\left[\left(\mathbf{U}_{ij}^l\right)^2\right]\right)_{kk}\right| \leq N_{\max} L_{\max} B^2.$$

Finally applying Matrix Bernstein inequality (Theorem 2.2.5) for $t = C\sqrt{B^2 N_{\max} L_{\max} \log n}$:

$$\begin{aligned}
\mathbb{P}\left[\|\boldsymbol{\Delta}'\|_2 > C\sqrt{B^2 N L_{\max} \log n}\right] &\leq 2n \exp\left\{-\frac{3\left(C\sqrt{B^2 N_{\max} L_{\max} \log n}\right)^2}{6\nu + 2\sqrt{2}B\left(C\sqrt{B^2 N_{\max} L_{\max} \log n}\right)}\right\} \\
&\leq 2n \exp\left\{-\frac{3C^2 B^2 N_{\max} L_{\max} \log n}{6 N_{\max} L_{\max} B^2 + 2\sqrt{2}BC\sqrt{B^2 N_{\max} L_{\max} \log n}}\right\} \\
&\leq 2n \exp\left\{-\frac{3CB^2 N_{\max} L_{\max} \log n}{6 N_{\max} L_{\max} B^2 + 2\sqrt{2}B^2\sqrt{N_{\max} L_{\max} \log n}}\right\} \\
&\leq 2n \exp\left\{-\frac{3C N_{\max} L_{\max} \log n}{6 N_{\max} L_{\max} + 2\sqrt{2}\sqrt{N_{\max}^2 L_{\max}}}\right\} \\
&= 2n \exp\left\{-\frac{3CB^2}{6 + 2\sqrt{2}B^2 L_{\max}^{-1/2}} \log n\right\} \\
&\leq 2n \exp\left\{-\frac{3C}{6 + 2\sqrt{2}} \log n\right\}
\end{aligned}$$

$$= 2n \cdot n^{-\frac{3C}{6+2\sqrt{2}}}$$
$$= 2n^{-\left(\frac{3C}{6+2\sqrt{2}}-1\right)}.$$

But $\frac{3C}{6+2\sqrt{2}} - 1 \geq 1$, since $C \geq 6$, so with probability at least $1 - \frac{1}{\text{poly}(n)}$ we have

$$\|\mathbf{\Delta}'\|_2 \leq C\sqrt{B^2 N_{\max} L_{\max} \log n}.$$

Hence for both cases ($N_{\max} \leq \log n$ and $N_{\max} \geq \log n$) we have that

$$\|\mathbf{\Delta}'\|_2 \leq C\sqrt{B^2 N_{\max} L_{\max} \log n}. \tag{A.3}$$

Now by Equation (A.2) and Equation (A.3) we get

$$\|\mathbf{\Delta}\|_2 \leq 2C\sqrt{B^2 N_{\max} L_{\max} \log n},$$

with probability at least $1 - \frac{1}{\text{poly}(n)}$. This finishes the proof. □

Proposition A.1.2. *Assume matrix $\mathbf{\Delta}$ as in the previous proposition and let $\mathbf{a} \in \mathbb{R}_+^n$. There exist positive constants C, c_0 such that if $n \geq c_0 \log n$ then with probability at least $1 - \frac{1}{\text{poly}(n)}$ it holds*

$$\left\|\mathbf{a}^\top \mathbf{\Delta}\right\|_2 \leq C\varphi B\sqrt{N_{\max} L_{\max}}\, \|\mathbf{a}\|_2\,, \tag{A.4}$$

where $\varphi = \max_{ij} \frac{a_i}{a_j} = \frac{a_{\max}}{a_{\min}}$.

Proof. Recall that

$$\Delta_{ii} = -\sum_{j \neq i} \Delta_{ij} = -\Delta_{ii}^{l} - \Delta_{ii}^{u},$$

where $\Delta_{ii}^{\text{lower}} = \sum_{j:j<i} \Delta_{ij}$ and $\Delta_{ii}^{\text{upper}} = \sum_{j:j<i} \Delta_{ij}$. So

$$\mathbf{\Delta} = \mathbf{\Delta}_{\text{lower}} + \mathbf{\Delta}_{\text{upper}} + \mathbf{\Delta}_{\text{diag}}^{\text{lower}} + \mathbf{\Delta}_{\text{diag}}^{\text{upper}},$$

where $\mathbf{\Delta}_{\text{lower}}$ ($\mathbf{\Delta}_{\text{upper}}$ respectively) is the lower (upper respectively) triangular part of $\mathbf{\Delta}$ excluding the diagonal and

$$\mathbf{\Delta}_{\text{diag}}^{\text{lower}} = -\text{diag}\left(\Delta_{11}^{\text{lower}}, \dots, \Delta_{nn}^{\text{lower}}\right) \text{ and } \mathbf{\Delta}_{\text{diag}}^{\text{upper}} = -\text{diag}\left(\Delta_{11}^{\text{upper}}, \dots, \Delta_{nn}^{\text{upper}}\right).$$

Hence we get

$$\left\|\mathbf{a}^\top \mathbf{\Delta}\right\|_2 \leq \left\|\mathbf{a}^\top \mathbf{\Delta}_{\text{lower}}\right\|_2 + \left\|\mathbf{a}^\top \mathbf{\Delta}_{\text{upper}}\right\|_2 + \left\|\mathbf{a}^\top \mathbf{\Delta}_{\text{diag}}^{\text{lower}}\right\|_2 + \left\|\mathbf{a}^\top \mathbf{\Delta}_{\text{diag}}^{\text{upper}}\right\|_2.$$

Let $I_{\text{lower}} = \left\|\mathbf{a}^\top \mathbf{\Delta}_{\text{lower}}\right\|_2$. Note that the j-th component of $\mathbf{a}^\top \mathbf{\Delta}_{\text{lower}}$ can be expressed as

$$\left[\mathbf{a}^\top \mathbf{\Delta}_{\text{lower}}\right]_j = \sum_{i:i>j} a_i \Delta_{ij}.$$

Let $N_j^{\text{lower}} = \left|\left\{(i,j) \,|i > j \text{ and } i \in N_j\right\}\right|$. Then by Hoeffding's Inequality (Theorem 2.2.3) we get

$$\begin{aligned}
\mathbb{P}\left[\left|\left[\mathbf{a}^\top \mathbf{\Delta}_{\text{lower}}\right]_j\right| > t\right] &\leq 2\exp\left\{-\frac{2t^2}{\sum_{\substack{i>j \\ i\in N_j}} \sum_{l=1}^{L_{ij}} \left(2B\,\|\mathbf{a}\|_\infty\right)^2}\right\} \\
&\leq 2\exp\left\{-\frac{2t^2}{4N_j^{\text{lower}} L_{\max} B^2 \|\mathbf{a}\|_\infty^2}\right\} \\
&= 2\exp\left\{-\frac{t^2}{2N_j^{\text{lower}} L_{\max} B^2 \|\mathbf{a}\|_\infty^2}\right\}.
\end{aligned}$$

Hence $\left[\mathbf{a}^\top \mathbf{\Delta}_{\text{lower}}\right]_j$ is a sub Gaussian random variable with variance proxy

$$\sigma_j^2 = dN_j^{\text{lower}} L_{\max} B^2 \|\mathbf{a}\|_\infty^2 \leq N_{\max} L_{\max} B^2 \|\mathbf{a}\|_\infty^2 = \sigma^2.$$

Note that

$$\begin{aligned}
\mathbb{E}\left[I_{\text{lower}}^2\right] &= \mathbb{E}\left[\sum_{j=1}^{n}\left[\mathbf{a}^\top \mathbf{\Delta}_{\text{lower}}\right]_j^2\right] \\
&\leq 4\sum_{j=1}^{n}\sigma_j^2 \\
&\leq 4n\sigma^2
\end{aligned}$$

Since the entries of $\mathbf{a}^\top \mathbf{\Delta}_{\text{lower}}$ are independent, by the Hanson-Wright inequality (Corollary 2.2.10) we have

$$\mathbb{P}\left[\left|I_{\text{lower}}^2 - \mathbb{E}\left[I_{\text{lower}}^2\right]\right| > t\right] \leq 2\exp\left\{-\frac{ct}{9\sigma^2}\min\left(\frac{t}{9n\sigma^2}, 1\right)\right\},$$

for some constant $c > 0$. For $t = \sqrt{\frac{81C}{c}}\sigma^2\sqrt{n\log n}$, for some $C \geq 1$, we have

$$\begin{aligned}
&\mathbb{P}\left[\left|I_{\text{lower}}^2 - \mathbb{E}\left[I_{\text{lower}}^2\right]\right| > \sqrt{\frac{81C}{c}}\sigma^2\sqrt{n\log n}\right] \\
&\leq 2\exp\left\{-\frac{c\sqrt{\frac{81C}{c}}\sigma^2\sqrt{n\log n}}{9\sigma^2}\min\left(\frac{\sqrt{\frac{81C}{c}}\sigma^2\sqrt{n\log n}}{9n\sigma^2}, 1\right)\right\}
\end{aligned}$$

If $\log n \leq \left(\frac{c}{C}\right) n$ we get

$$\mathbb{P}\left[\left|I_{\text{lower}}^2 - \mathbb{E}\left[I_{\text{lower}}^2\right]\right| > \sqrt{\frac{81C}{c}}\sigma^2\sqrt{n\log n}\right] \leq 2n^{-C}, C \geq 1$$

So with probability at least $1 - \frac{1}{\text{poly}(n)}$ we have

$$
\begin{aligned}
I_{\text{lower}}^2 &\leq \mathbb{E}\left[I_{\text{lower}}^2\right] + \sqrt{\frac{81C}{c}}\sigma^2\sqrt{n\log n} \\
&\leq 4n\sigma^2 + \sqrt{\frac{81C}{c}}\sigma^2\sqrt{n\log n} \\
&\leq n\sigma^2\left(4 + \sqrt{\frac{81C}{c}}\sqrt{\frac{\log n}{n}}\right) \\
&\leq n\sigma^2(4+9) \\
&= 13n\sigma^2 \\
&= 13nN_{\max}L_{\max}B^2\|\mathbf{a}\|_\infty^2 \\
&\leq 13\varphi^2B^2N_{\max}L_{\max}\|\mathbf{a}\|_2^2,
\end{aligned}
$$

since $\|\mathbf{a}\|_2^2 \geq na_{\min}^2 = n\frac{a_{\max}^2}{\varphi^2} = \frac{n}{\varphi^2}\|\mathbf{a}\|_\infty^2$, where $\varphi = \frac{a_{\max}}{a_{\min}}$.
Working similarly for the other terms we get the desired result. □

A.2 Spectral gaps

In this section we introduce the notion of *spectral gap* and we prove that random walks on connected graphs have strictly positive spectral gaps.

Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ be an irreducible stochastic matrix. Let

$$1 = \lambda_1(\mathbf{A}) \geq \lambda_2(\mathbf{A}) \geq \ldots \lambda_n(\mathbf{A})$$

be its eigenvalues in a decreasing order. By the Perron-Frobenius Theorem, the spectral radius $\rho(\mathbf{A}) = \max_i |\lambda_i(\mathbf{A})|$ is equal to 1 and it corresponds to the unique eigenvalue $\lambda_1(\mathbf{A}) = 1$. We denote with $\lambda_{\max}(\mathbf{A})$ the second largest absolute value of eigenvalues, i.e.

$$\lambda_{\max}(\mathbf{A}) = \max_{i=2,\ldots,n}|\lambda_i(\mathbf{A})| = \max\{\lambda_2(\mathbf{A}), -\lambda_n(\mathbf{A})\} > 0.$$

Definition A.2.1. We denote the *spectral gap* of **A** as

$$\xi = \xi(\mathbf{A}) = 1 - \lambda_{\max}(\mathbf{A}).$$

The following proposition associate the spectral gap of random walks on graphs to the spectral gap of the Laplacian[1] of the graph.

Proposition A.2.2. *Let $\mathbf{P} = \left[P_{ij}\right] \in \mathbb{R}^{n\times n}$ be a reversible Markov chain with stationary distribution $\pi \in \mathbb{R}^n$, defined on a finite set $[n]$ representing random walks on a graph $G = ([n], E)$, i.e. $P_{ij} = 0$ if $(i,j) \notin E$. Then*

$$1 - \lambda_{\max}(\mathbf{P}) \geq \frac{\xi d_{\min}}{2b^3 d_{\max}} > 0,$$

where ξ is the spectral gap of the Laplacian of the graph G and $b = \max_{i,j}\frac{\pi_i}{\pi_j}$. Moreover $d_{\max}$ ($d_{\min}$ respectively) is the maximum (minimum respectively) degree of the graph G.

[1]Note that the Laplacian $\mathbf{L} = D^{-1}A$ is a stochastic matrix.

In order to prove the above proposition we will use the following lemma which is a special case of a more general result from [DS93].

Lemma A.2.3 ([NOS17]: Lemma 6). *Let* (Q_1, μ_1) *and* (Q_2, μ_2) *be reversible Markov chains on a finite set* $[n]$ *representing random walks on a graph* $G = ([n], E)$, *i.e.* $Q_1(i,j) = Q_2(i,j) = 0$ *if* $(i,j) \notin E$. *Let*

$$\alpha = \max_{(i,j)\in E} \frac{\mu_2(i)Q_2(i,j)}{\mu_1(i)Q_1(i,j)} \text{ and } \beta = \max_{i\in[n]} \frac{\mu_2(i)}{\mu_1(i)}.$$

Then

$$\frac{1-\lambda_{\max}(Q_2)}{1-\lambda_{\max}(Q_1)} \geq \frac{\alpha}{\beta}.$$

Proof of Proposition A.2.2. Apply the previous lemma with $(Q_2, \mu_2) = (\mathbf{P}^*, \boldsymbol{\pi}^*)$ and

$$Q_1(i,j) = \begin{cases} \frac{1}{d_i} & \text{if } (i,j) \in E \\ 0 & \text{otherwise} \end{cases}.$$

Note that Q_2 is a reversible Markov chain with $\mu_1(i) = \frac{d_i}{2|E|}$, since $\mu_1(i)Q_1(i,j) = \frac{1}{2|E|}$ if $(i,j) \in E$ and 0 otherwise. Observe that Q_1 is actually the Laplacian of G, so $1-\lambda_{\max}(Q_1) = \xi$. Now we have

$$\begin{aligned} \beta &= 2|E| \max_i \frac{w_i}{d_i} \\ &\leq \frac{2|E|w_{\max}}{d_{\min}} \\ &\leq \frac{2b|E|}{nd_{\min}} \end{aligned}$$

and

$$\begin{aligned} 2nw_{\max}^2 w_i w_j &\geq 2nw_{\max}^2 w_{\min}^2 \\ &\geq 2w_{\max} w_{\min}^2 \\ &\geq (w_i + w_j)w_{\min}^2, \end{aligned}$$

so $\pi^*(i)P^*_{ij} = \frac{w_i w_j}{d_{\max}(w_i+w_j)} \geq \frac{1}{2nb^2 d_{\max}}$. Hence

$$1-\lambda_{\max}(\mathbf{P}^*) \geq \frac{\xi d_{\min}}{2b^3 d_{\max}},$$

as wanted. □

A.3 Eigenvector perturbation

In this section we introduce a new matrix norm associated to a probability vector $\boldsymbol{\pi}$. Using this norm we are going to see an important result for the eigenvector perturbation for probability transition matrices.

Definition A.3.1. Let $\boldsymbol{\pi} \in \mathbb{R}^n$ be a strictly positive probability vector. Then we define the inner product space indexed by $\boldsymbol{\pi}$ as a vector space in $\mathbb{R}^n$ endowed with the inner product

$$\langle \mathbf{x}, \mathbf{y} \rangle_{\pi} = \sum_{i=1}^{n} \pi_i x_i y_i.$$

The corresponding vector norm and the induced matrix norm are defined respectively as

$$\|\mathbf{x}\|_{\pi} = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle_{\pi}} = \sqrt{\sum_{i=1}^{n} \pi_i x_i^2} \text{ and } \|\mathbf{A}\|_{\pi} = \sup_{\|\mathbf{x}\|=1} \|\mathbf{x}^\top \mathbf{A}\|_{\pi}.$$

Remark A.3.2. The $\|\cdot\|_{\pi}$-norm can be viewed as a generalization of the $\|\cdot\|_2$-norm. In particular, if $\boldsymbol{\pi} = \frac{1}{n}(1, \ldots, 1)^\top$, then

$$\|x\|_{\pi} = \frac{\|x\|_2}{\sqrt{n}}.$$

The next proposition associates the $\|\cdot\|_{\pi}$-norm to the $\|\cdot\|_2$-norm.

Proposition A.3.3. *The following inequalities hold:*

- $\sqrt{\pi_{\min}}\|\mathbf{x}\|_2 \le \|\mathbf{x}\|_{\pi} \le \sqrt{\pi_{\max}}\|\mathbf{x}\|_2$
- $\sqrt{\frac{\pi_{\min}}{\pi_{\max}}}\|\mathbf{A}\|_2 \le \|\mathbf{A}\|_{\pi} \le \sqrt{\frac{\pi_{\max}}{\pi_{\min}}}\|\mathbf{A}\|_2$

Now we can state the main theorem. This theorem can be a treated as the analogue of the famous Davis-Kahan $\sin\Theta$ theorem ([DK70]).

Theorem A.3.4 ([Che+19]: Theorem 8)**.** *Suppose that* $\mathbf{P}$*,* $\hat{\mathbf{P}}$*, and* $\mathbf{P}^*$ *are probability transition matrices with stationary distributions* $\boldsymbol{\pi}$*,* $\hat{\boldsymbol{\pi}}$*,* $\boldsymbol{\pi}^*$*, respectively. Also, assume that* $\mathbf{P}^*$ *represents a reversible Markov chain. When*

$$\|\hat{\mathbf{P}} - \mathbf{P}^*\|_{\pi^*} < 1 - \max\{\lambda_2(\mathbf{P}^*), |\lambda_n(\mathbf{P}^*)|\},$$

it holds that

$$\|\boldsymbol{\pi} - \hat{\boldsymbol{\pi}}\|_{\pi^*} \le \frac{\|\boldsymbol{\pi}^\top(\mathbf{P} - \hat{\mathbf{P}})\|_{\pi^*}}{1 - \max\{\lambda_2(\mathbf{P}^*), |\lambda_n(\mathbf{P}^*)|\} - \|\hat{\mathbf{P}} - \mathbf{P}^*\|_{\pi^*}}.$$

We include the proof for completeness.

Proof. We write

$$\begin{aligned}
\boldsymbol{\pi}^\top - \hat{\boldsymbol{\pi}}^\top &= \boldsymbol{\pi}^\top \mathbf{P} - \hat{\boldsymbol{\pi}}^\top \mathbf{P} \\
&= \boldsymbol{\pi}^\top (\mathbf{P} - \hat{\mathbf{P}}) + (\boldsymbol{\pi} - \hat{\boldsymbol{\pi}})^\top \hat{\mathbf{P}} \\
&= \boldsymbol{\pi}^\top (\mathbf{P} - \hat{\mathbf{P}}) + (\boldsymbol{\pi} - \hat{\boldsymbol{\pi}})^\top \mathbf{P}^* + (\boldsymbol{\pi} - \hat{\boldsymbol{\pi}})^\top (\hat{\mathbf{P}} - \mathbf{P}^*) \\
&= \boldsymbol{\pi}^\top (\mathbf{P} - \hat{\mathbf{P}}) + (\boldsymbol{\pi} - \hat{\boldsymbol{\pi}})^\top (\mathbf{P}^* - \mathbf{1}\boldsymbol{\pi}^{*\top}) + (\boldsymbol{\pi} - \hat{\boldsymbol{\pi}})^\top (\hat{\mathbf{P}} - \mathbf{P}^*),
\end{aligned}$$

where $\mathbf{1} \in \mathbb{R}^{n\times 1}$ is the column vector whose all entries are ones and $\boldsymbol{\pi}^\top \mathbf{1} = 1$ for all probality vectors $\boldsymbol{\pi}$. Hence we get

$$\|\boldsymbol{\pi} - \hat{\boldsymbol{\pi}}\|_{\pi^*} \le \|\boldsymbol{\pi}^\top (\mathbf{P} - \hat{\mathbf{P}})\|_{\pi^*} + \|\boldsymbol{\pi} - \hat{\boldsymbol{\pi}}\|_{\pi^*} \|\mathbf{P}^* - \mathbf{1}\boldsymbol{\pi}^{*\top}\|_{\pi^*} + \|\boldsymbol{\pi} - \hat{\boldsymbol{\pi}}\|_{\pi^*} \|\hat{\mathbf{P}} - \mathbf{P}^*\|_{\pi^*}$$

Now observe that $\|\mathbf{P}^* - \mathbf{1}\boldsymbol{\pi}^{*\top}\|_{\pi^*} = \max\{\lambda_2(\mathbf{P}^*), |\lambda_n(\mathbf{P}^*)|\}$ and use the given condition to finish the proof. □

For $\mathbf{P} = \mathbf{P}^*$ in the previous theorem we get (after renaming): If $\|\mathbf{P} - \mathbf{P}^*\|_{\pi^*} < 1 - \max\{\lambda_2(\mathbf{P}^*), |\lambda_n(\mathbf{P}^*)|\}$ then

$$\|\pi - \pi^*\|_{\pi^*} \leq \frac{\left\|\pi^{*\top}(\mathbf{P} - \mathbf{P}^*)\right\|_{\pi^*}}{1 - \max\{\lambda_2(\mathbf{P}^*), |\lambda_n(\mathbf{P}^*)|\} - \|\mathbf{P} - \mathbf{P}^*\|_{\pi^*}}.$$

Acknowledgements

The Proposition A.1.1 is based on techniques used in [NOS17]: Lemma 3. The Proposition A.1.2 is based on techniques used in [Che+19]: Section C2.

Bibliography

[Aga+20] Arpit Agarwal, Shivani Agarwal, Sanjeev Khanna, and Prathamesh Patil. "Rank aggregation from pairwise comparisons in the presence of adversarial corruptions". In: *International Conference on Machine Learning*. PMLR. 2020, pp. 85–95.

[APA18] Arpit Agarwal, Prathamesh Patil, and Shivani Agarwal. "Accelerated spectral ranking". In: *International Conference on Machine Learning*. PMLR. 2018, pp. 70–79.

[AS11] Ammar Ammar and Devavrat Shah. "Ranking: Compare, don't score". In: *2011 49th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE. 2011, pp. 776–783.

[Arr51] Kenneth Joseph Arrow. "Social choices and individual values". In: (1951).

[Ber37] Sergei N Bernstein. "On certain modifications of Chebyshev's inequality". In: *Doklady Akademii Nauk SSSR* 17.6 (1937), pp. 275–277.

[Bon+20] Heejong Bong, Wanshan Li, Shamindra Shrotriya, and Alessandro Rinaldo. "Nonparametric Estimation in the Dynamic Bradley-Terry Model". In: *International Conference on Artificial Intelligence and Statistics*. PMLR. 2020, pp. 3317–3326.

[Bor84] JC de Borda. "Mémoire sur les élections au scrutin". In: *Histoire de l'Academie Royale des Sciences pour 1781 (Paris, 1784)* (1784).

[BT52] Ralph Allan Bradley and Milton E Terry. "Rank analysis of incomplete block designs: I. The method of paired comparisons". In: *Biometrika* 39.3/4 (1952), pp. 324–345.

[CR09] Emmanuel J Candès and Benjamin Recht. "Exact matrix completion via convex optimization". In: *Foundations of Computational mathematics* 9.6 (2009), pp. 717–772.

[Che+13] Xi Chen, Paul N Bennett, Kevyn Collins-Thompson, and Eric Horvitz. "Pairwise ranking aggregation in a crowdsourced setting". In: *Proceedings of the sixth ACM international conference on Web search and data mining*. 2013, pp. 193–202.

[Che+19] Yuxin Chen, Jianqing Fan, Cong Ma, and Kaizheng Wang. "Spectral method and regularized MLE are both optimal for top-K ranking". In: *Annals of statistics* 47.4 (2019), p. 2204.

[DK70] Chandler Davis and William Morton Kahan. "The rotation of eigenvectors by a perturbation. III". In: *SIAM Journal on Numerical Analysis* 7.1 (1970), pp. 1–46.

[Den+09] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. "Imagenet: A large-scale hierarchical image database". In: *2009 IEEE conference on computer vision and pattern recognition*. Ieee. 2009, pp. 248–255.

[DS93] Persi Diaconis and Laurent Saloff-Coste. "Comparison theorems for reversible Markov chains". In: *The Annals of Applied Probability* 3.3 (1993), pp. 696–730.

[Dia+19] Ilias Diakonikolas, Gautam Kamath, Daniel Kane, Jerry Li, Ankur Moitra, and Alistair Stewart. "Robust estimators in high-dimensions without the computational intractability". In: *SIAM Journal on Computing* 48.2 (2019), pp. 742–864.

[Dia+17] Ilias Diakonikolas, Gautam Kamath, Daniel M Kane, Jerry Li, Ankur Moitra, and Alistair Stewart. "Being robust (in high dimensions) can be practical". In: *International Conference on Machine Learning*. PMLR. 2017, pp. 999–1008.

[Dia+18] Ilias Diakonikolas, Gautam Kamath, Daniel M Kane, Jerry Li, Ankur Moitra, and Alistair Stewart. "Robustly learning a gaussian: Getting optimal error, efficiently". In: *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*. SIAM. 2018, pp. 2683–2702.

[ER60] Paul Erdős and Alfréd Rényi. "On the evolution of random graphs". In: *Publ. Math. Inst. Hung. Acad. Sci* 5.1 (1960), pp. 17–60.

[For57] Lester R Ford Jr. "Solution of a ranking problem from binary comparisons". In: *The American Mathematical Monthly* 64.8P2 (1957), pp. 28–33.

[FK16] Alan Frieze and Michał Karoński. *Introduction to random graphs*. Cambridge University Press, 2016.

[Gol+16] Thomas Goldstein, Paul Hand, Choongbum Lee, Vladislav Voroninski, and Stefano Soatto. "Shapefit and shapekick for robust, scalable structure from motion". In: *European Conference on Computer Vision*. Springer. 2016, pp. 289–304.

[GCD81] Paul E Green, J Douglas Carroll, and Wayne S DeSarbo. "Estimating choice probabilities in multiattribute decision making". In: *Journal of Consumer Research* 8.1 (1981), pp. 76–84.

[HOX14] Bruce Hajek, Sewoong Oh, and Jiaming Xu. "Minimax-optimal inference from partial rankings". In: *Advances in Neural Information Processing Systems* 27 (2014).

[HLV18] Paul Hand, Choongbum Lee, and Vladislav Voroninski. "Shapefit: Exact location recovery from corrupted pairwise directions". In: *Communications on Pure and Applied Mathematics* 71.1 (2018), pp. 3–50.

[HMG06] Ralf Herbrich, Tom Minka, and Thore Graepel. "TrueSkill: a Bayesian skill rating system". In: *Advances in neural information processing systems* 19 (2006).

[Hin+12] Geoffrey Hinton, Li Deng, Dong Yu, George E Dahl, Abdel-rahman Mohamed, Navdeep Jaitly, Andrew Senior, Vincent Vanhoucke, Patrick Nguyen, Tara N Sainath, et al. "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups". In: *IEEE Signal processing magazine* 29.6 (2012), pp. 82–97.

[Hoe63] Wassily Hoeffding. "Probability Inequalities for Sums of Bounded Random Variables". In: *Journal of the American Statistical Association* 58.301 (1963), pp. 13–30.

[Hub65] Peter J Huber. "A robust version of the probability ratio test". In: *The Annals of Mathematical Statistics* (1965), pp. 1753–1758.

[Hub92] Peter J Huber. "Robust estimation of a location parameter". In: *Breakthroughs in statistics*. Springer, 1992, pp. 492–518.

[Hun04] David R Hunter. "MM algorithms for generalized Bradley-Terry models". In: *The annals of statistics* 32.1 (2004), pp. 384–406.

[JO20] Ayush Jain and Alon Orlitsky. "Optimal robust learning of discrete distributions from batches". In: *International Conference on Machine Learning*. PMLR. 2020, pp. 4651–4660.

[Kal19] Alkis Kalavasis. "On Learning Rankings from Noisy Samples". MA thesis. National Technical University of Athens, 2019.

[KT21] Eglantine Karlé and Hemant Tyagi. "Dynamic Ranking with the BTL Model: A Nearest Neighbor based Rank Centrality Method". In: *arXiv preprint, arXiv: 2109.13743* (2021).

[Kaz11] Gabriella Kazai. "In search of quality in crowdsourcing for search engine evaluation". In: *European conference on information retrieval*. Springer. 2011, pp. 165–176.

[KMO09] Raghunandan Keshavan, Andrea Montanari, and Sewoong Oh. "Matrix completion from noisy entries". In: *Advances in neural information processing systems* 22 (2009).

[Kha+11] Firas Khatib, Frank DiMaio, Seth Cooper, Maciej Kazmierczyk, Miroslaw Gilski, Szymon Krzywda, Helena Zabranska, Iva Pichova, James Thompson, Zoran Popović, et al. "Crystal structure of a monomeric retroviral protease solved by protein folding game players". In: *Nature structural & molecular biology* 18.10 (2011), pp. 1175–1177.

[LR11] Andrew SID Lang and Joshua Rio-Ross. "Using Amazon Mechanical Turk to transcribe historical handwritten documents". In: *Code4Lib Journal* 15 (2011).

[LP17] David A Levin and Yuval Peres. *Markov chains and mixing times*. Vol. 107. American Mathematical Society, 2017.

[LW21] Shuang Li and Michael B Wakin. "Recovery guarantees for time-varying pairwise comparison matrices with non-transitivity". In: *arXiv preprint arXiv:2106.09151* (2021).

[Lou15] Michail Loulakis. *Stochastic Processes*. Kallipos, 2015.

[Luc59] R. Duncan Luce. *Individual Choice Behavior: A Theoretical Analysis*. New York: Wiley, 1959.

[LAF12] Miguel Angel Luengo-Oroz, Asier Arranz, and John Frean. "Crowdsourcing malaria parasite quantification: an online game for analyzing images of infected thick blood smears". In: *Journal of medical Internet research* 14.6 (2012), e2338.

[Mal57] Colin L Mallows. "Non-null ranking models. I". In: *Biometrika* 44.1/2 (1957), pp. 114–130.

[Mam22] Aikaterini Mamali. "Distribution Learning from Truncated Samples". MA thesis. National Technical University of Athens, 2022.

[Mou19] Argyris Mouzakis. "Learning Techniques for Ranking Distributions". MA thesis. National Technical University of Athens, 2019.

[NOS17] Sahand Negahban, Sewoong Oh, and Devavrat Shah. “Rank centrality: Ranking from pairwise comparisons”. In: *Operations Research* 65.1 (2017), pp. 266–287.

[NW12] Sahand Negahban and Martin J Wainwright. “Restricted strong convexity and weighted matrix completion: Optimal bounds with noise”. In: *The Journal of Machine Learning Research* 13.1 (2012), pp. 1665–1697.

[Nic+85] Jean Antoine Nicolas et al. *Essai sur l’application de l’analyse à la probabilité des décisions rendues à la pluralité des voix*. de l’Imprimerie Royale, 1785.

[ODo14] Ryan O’Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2014.

[Pag+99] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. *The PageRank citation ranking: Bringing order to the web.* Tech. rep. Stanford InfoLab, 1999.

[Pie+13] Chris Piech, Jonathan Huang, Zhenghao Chen, Chuong Do, Andrew Ng, and Daphne Koller. “Tuned models of peer assessment in MOOCs”. In: *arXiv preprint arXiv:1307.2579* (2013).

[Pla75] Robin L Plackett. “The analysis of permutations”. In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 24.2 (1975), pp. 193–202.

[Ray+10] Vikas C Raykar, Shipeng Yu, Linda H Zhao, Gerardo Hermosillo Valadez, Charles Florin, Luca Bogoni, and Linda Moy. “Learning from crowds.” In: *Journal of machine learning research* 11.4 (2010).

[Ros07] Daniel Ross. *Arpad Elo and the Elo rating system*. http://en.chessbase.com/post/arpad-elo-and-the-elo-rating-system. 2007.

[RV13] Mark Rudelson and Roman Vershynin. “Hanson-wright inequality and subgaussian concentration”. In: *Electronic Communications in Probability* 18 (2013), pp. 1–9.

[Sha+15] Nihar Shah, Sivaraman Balakrishnan, Joseph Bradley, Abhay Parekh, Kannan Ramchandran, and Martin Wainwright. “Estimation from pairwise comparisons: Sharp minimax bounds with topology dependence”. In: *Artificial Intelligence and Statistics*. PMLR. 2015, pp. 856–865.

[Sha+14] Nihar B Shah, Sivaraman Balakrishnan, Joseph Bradley, Abhay Parekh, Kannan Ramchandran, and Martin Wainwright. “When is it Better to Compare than to Score?” In: *arXiv preprint arXiv:1406.6618* (2014).

[Sta20] Konstantinos Stavropoulos. “Learning rankings from incomplete samples”. MA thesis. National Technical University of Athens, 2020.

[TVV04] Kalyan T Talluri, Garrett Van Ryzin, and Garrett Van Ryzin. *The theory and practice of revenue management*. Vol. 1. Springer, 2004.

[Tro12] Joel A Tropp. “User-friendly tail bounds for sums of random matrices”. In: *Foundations of computational mathematics* 12.4 (2012), pp. 389–434.

[Von+08] Luis Von Ahn, Benjamin Maurer, Colin McMillen, David Abraham, and Manuel Blum. “recaptcha: Human-based character recognition via web security measures”. In: *Science* 321.5895 (2008), pp. 1465–1468.

Index