ED MADIS 631 - Université de Lille - INRIA

CLASSEMENT DYNAMIQUE ET TRANSLATION-SYNCHRONISATION SUR DES GRAPHES DYNAMIQUES.

DYNAMIC RANKING AND TRANSLATION-SYNCHRONIZATION ON DYNAMIC GRAPHS.

Thèse de Doctorat soutenue le 16 mai 2024 par Eglantine Karlé, dans le but d'obtenir le grade de Docteure en **Mathématiques et leurs interactions** devant un jury composé de :

Valentin Patilea *Professeur, ENSAI Rennes*	Président du jury
Christine Keribin *Maîtresse de conférences, Université Paris-Saclay*	Rapporteure
Clément Marteau *Professeur, Université de Lyon 1*	Rapporteur
Mylène Maïda *Professeur, Université de Lille*	Examinatrice
Cristian Preda *Professeur, Université de Lille*	Directeur de Thèse
Hemant Tyagi *Chargé de recherche, INRIA Lille*	Co-Directeur de Thèse

École doctorale MADIS 631
Spécialité : Mathématiques Appliquées
Unité de recherche : INRIA, Université de Lille, Laboratoire de mathématiques Painlevé, 59650 Villeneuve d'Ascq, France

SUPERVISORS:
Hemant Tyagi
Cristian Preda

LOCATION:
Lille

TIME FRAME:
Mai 2024

ABSTRACT

Ranking and comparing arise in many real life applications, such as sports tournaments or recommendation systems. In these domains, datasets are composed of pairwise comparisons between a collection of items that can also be summarized into a comparison graph. Many parametric statistical models for ranking were introduced, such as the Bradley-Terry-Luce model, where the items are supposed to have a latent strength. The ranking are then derived from this strengths. Numerous estimation algorithms have been analyzed over the past decades, for example the maximum-likelihood or the spectral method. However, most of them do not account for the temporal aspect of the data. Indeed, ranking and personal preferences can evolve over time. In order to include temporality, we will consider a sequence of comparison graphs, or dynamic graphs, that gathers data at different time instances.

We will first study an extension of the BTL model to this dynamic setting, introduced by Bong et al. under a local Lipschitz assumption on the strengths. Our algorithm is based on a nearest-neighbor approach and on the Rank Centrality algorithm, a classic estimation method in the static case. We will show ℓ_2 and ℓ_∞ bounds for our estimator and show our algorithm performance on both synthetic and real data.

In a second part, we will introduce a dynamic version of the Translation-Synchronization model under a global smoothness assumption. We will propose two estimators, one based on a smoothness-penalized least squares approach and the other based on projection onto the low frequency eigenspace of a suitable smoothness operator. We will show that both method give consistent estimators. We also display the performance of our algorithms on synthetic and real datasets.

RÉSUMÉ

Le classement et les comparaisons apparaissent dans de nombreuses applications de la vie de tous les jours, telles que les tournois sportifs ou les systèmes de recommandation. Dans ces domaines, les données sont souvent composées de comparaisons appairées entre un ensemble d'objets, qui peuvent être résumées dans un graphe de comparaison. De nombreux modèles paramétriques pour les problèmes

de classement ont été introduits, tels que le modèle de Bradley-Terry-Luce (BTL), dans lequel on suppose que les objets possèdent une qualité sous-jacente. Le classement est ensuite déduit des scores de qualité. De nombreux algorithmes d'estimation ont été analysés ces dernières années, comme l'algorithme de maximum de vraisemblance ou une méthode spectrale. Cependant, les classements et préférences personnelles peuvent évoluer avec le temps. Afin de prendre cela en compte, nous allons considérer une suite de graphes de comparaison, ou un graphe dynamique, qui rassemble les données à divers points de temps.

Nous allons d'abord étudier une extension du modèle BTL au cas dynamique, introduit par Bong et al. en supposant que les scores de qualités sont Lipschitz. Notre algorithme est basé sur la méthode des plus proches voisins et sur l'algorithme de Rank Centrality, méthode d'estimation bien connue dans le cas statique. Nous fournissons des bornes ℓ_2 et ℓ_∞ pour notre estimateur et montrons les performances de notre algorithme sur des données réelles et synthétiques.

Dans un second temps, nous introduirons une version dynamique du modèle de Translation-Synchronisation sous une une hypothèse globale de régularité. Nous proposerons deux estimateurs, le premier basé sur une approche des moindres carrés avec une pénalité traduisant la régularité, et le deuxième basé sur la projection sur l'espace des vecteurs propres de basse fréquence d'un opérateur de régularité approprié. Nous montrerons que ces deux méthodes donnent des estimateurs consistants. Nous montrerons à nouveau les performances de nos algorithmes sur des données réelles et synthétiques.

REMERCIEMENTS

Avant toute chose, il me faut remercier Hemant et Cristian, à l'origine de ce projet et m'ayant accueillie au sein de l'équipe Modal. Je remercie Hemant pour avoir guidé nos réflexions et corrigé les preuves, ébauches d'article et ce manuscrit. Merci aussi à Cristian pour son accompagnement tout au long de la thèse, plus humain que théorique mais tout aussi important.

Je tiens à remercier les rapporteurs, Christine Kéribin et Clément Marteau, d'avoir accepté de se pencher sur mon travail. Je remercie aussi Mylène Maïda et Valentin Patilea d'avoir gentiment accepté de faire partie de mon jury.

J'ai une pensée pour l'équipe Modal au complet, et en particulier pour mes collègues doctorants, post-doctorants et stagiaires. Il aurait été bien difficile d'arriver au bout de ce projet faramineux sans votre soutien et sans nos célèbres pauses cafés. Je remercie particulièrement Ernesto, pour son implication dans notre travail commun et pour m'avoir supporté dans notre bureau. Merci à Issam, Etienne, Axel, et tous les autres, nous voilà finalement au bout!

Je conclus ces remerciements en dédiant ce travail à ma famille : Papa, Maman, Louis-Marie et Eléonore, et tous les autres qu'il serait trop long de nommer ici. Cette thèse est aussi un peu la vôtre. Malgré les hauts et les bas ainsi que la distance qui nous séparaient, vous êtes restés à mes côtés. Sans même comprendre ce que représentait mon travail, vous avez su me donner la motivation nécessaire pour avancer. Merci aussi de m'avoir fait prendre des pauses, loin du travail rigoureux que représente une thèse. C'est avec vous autour d'un bon repas ou d'une partie de tarot chez Grand-Pa Grand-Mi que les moments difficiles et les efforts s'adoucissent. Je pourrais vous remercier plus longuement, mais votre présence dans ma vie ne s'arrêtant pas à la fin de cette thèse, je continuerai de le faire à l'avenir de vive voix.

CONTENTS

LIST OF TABLES

NOTATIONS

Symbol	*Definition*
$\|A\|_2$	Spectral norm, i.e., the largest singular value, of a matrix A
$\|A\|_F$	Frobenius norm of a matrix A
$\lambda_n(A) \leqslant \cdots \leqslant \lambda_1(A)$	Real eigenvalues of a $n \times n$ matrix A
$A^\dagger$	Moose-Penrose pseudo-inverse of a matrix A
$A \preccurlyeq B$	$B - A$ is positive semi-definite, for A, B symmetric matrices
$\mathbf{1}_n$	All one's vector of size n
I_n	$n \times n$ identity matrix
$\otimes$	Kronecker product of two matrice
$a \wedge b$	Minimum of a, b for a, b real numbers
$a \vee b$	Minimum of a, b for a, b real numbers
$c, C, \tilde{C}$ etc.	Absolute constants
$a \lesssim b$	There exists a constant $C > 0$ such that $a \leqslant Cb$
$a \asymp b$	$a \lesssim b$ and $b \lesssim a$
ι	Unit imaginary number

INTRODUCTION

Ranking and comparing are a part of everyday life, whether it is to choose which movie to watch, which football team to bet on or which candidate to elect. Although easy to do on a small amount of items, it is much more difficult for the brain to perform such a task for tens or hundreds of candidates. That is why algorithms and estimation methods on ranking were introduced.

STATISTICAL APPROACH OF RANKING

A ranking on a collection of n items simply consists of a permuation of the integers $1, \ldots, n$, indicating the order of preference or qualities of the items of interest. Hence, the simplest model for ranking is the uniform model, where all the $n!$ permutations are equally probable. However, this model does not fit most of the real life applications as sports tournaments or recommendation systems, because some of the items to rank are significantly better than others. Different non-uniform models have then been introduced, based either on order statistics [44], pairwise comparisons [5], distances between permutations [11], or stagewise decompositions of the ranking process [15]. We will focus in this thesis on models based on pairwise comparisons. We observe data in the form of pairwise comparisons between the items to rank, indicating preferences between the compared items. These comparisons can then be presented as a graph, as shown in Figure 2, where we present the preferences over a set of movies.

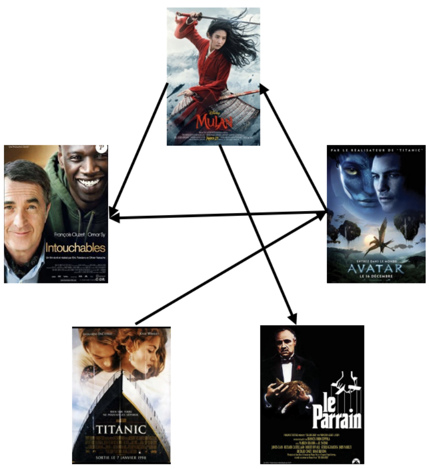


Figure 2: Comparison graph over a set of 5 movies. Arrow points to the preferred movie between the pair.

In particular, we will focus on the Bradley-Terry-Luce model [5] (or BTL for short). This model has already been studied extensively in the case of a single comparison graph, through regularized maximum-likelihood estimation (abbreviated as MLE)[9] but also through a spectral method, namely Rank Centrality [36], and the Least Squares method [18]. Note that optimal bounds on the ℓ_2 and ℓ_∞ error rates were recently achieved by Chen et al. [9] in this static setting.

DYNAMIC RANKING

Ranking is by nature a dynamic problem, as personal preferences evolve with time. It can be a result of the evolution of personal taste but it is also dependent on the apparition (or disparition) of new items to rank. Indeed, the movie industry is perpetually evolving, as new movies come out each week and filming technologies become more advanced. Hence, personal preferences change as one ages. We can also think about football competitions, that possess a promotion system between different leagues. Then, a team which does not perform will be relegated to a different league and can no longer play against the top teams. These different dynamic aspects of the data have been the motivation for the introduction of dynamic ranking models, especially based on sports related data. However, most of the papers do not conduct a theoretical analysis of the models. Our motivation in this thesis was to fill this gap by studying extensions of static ranking models to the dynamic case.

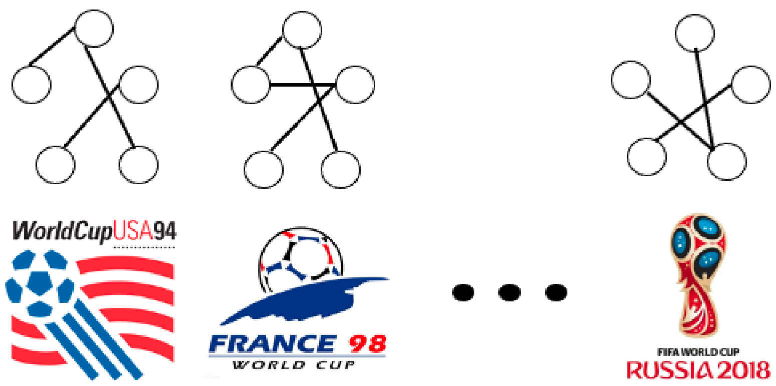


Figure 3: Sequence of comparison graphs representing every FIFA World Cup.

In the dynamic setting, data consist of a sequence of comparison graph as in Figure 3 instead of a single comparison graph. Hence to study dynamic ranking models, a classical idea is to consider observations at different timepoints and to aggregate them in a single graph. Indeed, assuming that the intrinsinc qualities of an item evolve smoothly with time, the comparison results at close timepoints will be similar. This idea has been used in the theoretical work of Bong et al [4]. They proposed a dynamic version of the BTL model and a regularized MLE algorithm. We propose something similar in Chapter 2.

TRANSLATION SYNCHRONIZATION

Synchronization problems have many applications, as clocks synchronization in wireless systems, computer vision or clustering. In such problems, the goal is to recover group elements $g_i \in G$ given noisy pairwise informations $g_i g_j^{-1}$, where g_j^{-1} denotes the group inverse. Different applications correspond to different groups G, examples being $\mathbb{Z}_2$ for clustering or $\mathbb{R}$ for clocks synchronization. Interestingly, these problems can also be linked to ranking. By mapping the ranks $1, \dots, n$ to the angles of the upper semi-circle $[0, 2\pi]$, one can then use the results of Singer [43] on Angular Synchronization, for $G = SO(2)$. This observation is the foundation of the Sync-Rank algorithm [10]. Although not studied theoretically, this algorithm performs particularly well on real datasets. Another interesting model is the Translation Synchronization model, where $G = \mathbb{R}$. Considering that each item possess an underlying strength, one can then derive a ranking of the items by ranking their strengths. Then, the Translation Synchronization model posits that observations are noisy version of the strength differences. One can then recover a ranking by estimating the strenght if each item. This model was studied by Huang et al. [21], there exist no extension of this model to the dynamic setting. We propose in Chapter 3 a dynamic Translation Synchronization model and algorithms for estimating the latent strengths.

OUTLINE OF THE THESIS

In this thesis, we consider different models for dynamic ranking and provide a thorough theoretical analysis for each of them.

In Chapter 1, we will introduce the main models and methods this manuscript is based on. We will first present static models for ranking, with a particuliar focus on the Bradley-Terry-Luce model, as we will study its extension to the dynamic setting later on. We then shift our focus to dynamic extensions of classical models. Finally, we present Synchronization problems and some estimation methods. We focus specifically on Angular Synchronization, as ranks can be mapped to angles on the upper half circle. It is the basis of the Sync-Rank algorithm, developped by Cucuringu [10]. We also detail the Translation Synchronization model, that we extend to the dynamic case in Chapter 3.

Chapter 2 describes the work around the Dynamic BTL model. We study an analogous model to Bong et al. [4] under a smoothness assumption on the items strengths. Our idea is to aggregate the temporal information using a nearest neighbour approach, only using data collected at suitably close timepoint from our time of estimation t. We then average this information and create a single comparison graph, to which we apply the static Rank Centrality method [36]. We detail the model and the estimation method in Section 2.1. We then provide a theoretical analysis of this model and derive ℓ_2 and ℓ_∞ error bounds. We conclude this chapter with experiments on synthetic and real datasets, disclosing the performance of our algorithm.

We present in Chapter 3 a Dynamic Translation Synchronization model, derived from the static Translation Synchronization model [21]. We analyse this model under a global smoothness assumption, assuming that the quadratic variation of the strengths is upper bounded. We provide two estimators, inspired by the Laplacian smoothing and Laplacian eigenmaps estimators [40]. The first one can be studied by solving a smoothness penalized least squares problem (see Section 3.3). The second estimator is obtained via a two-step process, that enforces the smoothness by projecting the estimator onto a "smooth space" (see Section 3.4). We provide for both estimators ℓ_2 error bounds and also conduct experiments on synthetic and real datasets to display the performance of our estimators.

Finally, we present perspectives for future works in Chapter 4.

LIST OF PUBLICATIONS

This thesis is based on the following publications.

- Eglantine Karlé and Hemant Tyagi. “Dynamic ranking with the btl model: a nearest neighbor based rank centrality method.” In: *Journal of Machine Learning Research* 24.269 (2023), pp. 1–57
- Ernesto Araya, Eglantine Karlé, and Hemant Tyagi. “Dynamic ranking and translation synchronization.” In: *Information and Inference: A Journal of the IMA* 12.3 (2023), pp. 2224–2266

1 BACKGROUND ON RANKING AND SYNCHRONIZATION

In this Chapter, we provide an overview of the literature on ranking and synchornization problems. We present models of interest for ranking, in particular the Bradley-Terry-Luce model and associated estimation methods. We then describe the problem of dynamic ranking and the corresponding state-of-the art models and methods. Finally, we present the synchronization problem and how it can be realted to ranking.

1.1 INTRODUCTION TO RANKING IN THE STATIC CASE

1.1.1 *Pairwise comparisons*

Models on ranking based on pairwise comparisons are popular as they reflect the majority of the datasets existing in real life. Indeed, in order to rank objects, a natural way to do so is to compare them directly by pairs, as done in sports tournaments for example. In addition, pairwise comparison data naturally reduce the bias contained in ranking information, as shown in [1]. As a matter of fact, asking reviewers to rank n items directly is similar to asking them to give scores to each item (e.g. between 1 and 5 stars); such scores being relative. This level of relativity does not exist in pairwise data.

Although not always the case, the majority of models based on pairwise comparisons are parametric models. It is assumed that each item has some latent strength, representing their intrinsinc qualities, and on which the probability of pairwise preferences depends. To study such models, the tool of choice is usually graph theory as the data can be represented as a comparison graph $G([n], \mathcal{E})$. The vertices of this graph denote the items and the edges $\mathcal{E}$ indicate whether two items were compared or not. Some estimation methods used for these models rely heavily on this graph structure, such as the Rank Centrality method [36] or the weighted least-squares algorithm of Hendrickx et al [18].

1.1.2 *BTL model and estimation methods*

Ranking models were first introduced in the statistical and psychological literature. Amongst the more popular models, most of them assume that the pairwise observations y_{ij} are independent random vectors such that

$$y_{ij} = \alpha_{ij} + \varepsilon_{ij},$$

with α_{ij} the expected result for a comparison of the pair $\{i, j\}$ and ε_{ij} denoting the noise. The noise variables ε_{ij} are supposed to be independent random variables. In 1929, a binomial version of this model was introduced by Zermelo [48], which has then been popularized by Bradley and Terry [5] and Luce [30]. Let us then detail this model, called Bradley-Terry-Luce model. We suppose that item i has a latent strength $w_i^* \in \mathbb{R}^+$. The probabilities of preference are given as follows.

$$\mathbb{P}(\text{item } i \text{ preferred to item } j) := \frac{w_i^*}{w_i^* + w_j^*}. \tag{1.1.1}$$

Then, for each pair of compared items (i, j), we perform L independent comparisons denoted $y_{ij}^{(l)}$ for $l \in [L]$. The BTL model posits that $y_{ij}^{(l)}$ are independent Bernoulli variables with probability $\frac{w_i^*}{w_i^*+w_j^*}$. In the following, we will consider their empirical mean $y_{ij} = \frac{1}{L}\sum_{l=1}^{L} y_{ij}^{(l)}$.

In order to estimate the ranks, one then only need to estimate the latent strengths w_i^*. This problem has already been thoroughly studied via MLE estimation [9] but also through other methods, such as the spectral method [36] or least-squares methods [18]. Let us describe briefly each method and the theoretical guarantees obtained for each of them.

MAXIMUM LIKELIHOOD ESTIMATION. As a state of the art method of inference, the MLE was the first method used to analyse the BTL model in the original article [5]. The negative log-likelihood associated with this problem is written as follows.

$$\mathcal{L}(w, y) = -\sum_{(i,j)\in\vec{\mathcal{E}}} y_{ij} \log \frac{w_i}{w_i + w_j} + (1 - y_{ij}) \log \frac{w_j}{w_i + w_j}. \tag{1.1.2}$$

This problem was studied for a general connected comparison graph as well as for Erdös-Renyi graphs [1] $G \sim \mathcal{G}(n, p)$. Let us denote $\hat{w}$ the maximum likelihood estimator of the true BTL weights w^* and d_{max}, d_{min} the maximal and minimal degree of the comparison graph.

The analysis of the Erdös-Renyi case has been refined over the years until 2019, when Chen et al. [9] provided optimal error bounds under

1 An Erdös-Renyi graph $G \sim \mathcal{G}(n, p)$ is a random graph with n vertices in which each pair of vertices $\{i, j\}$ are connected by an edge with probability p.

the assumption that the comparison graph is connected. In particular, they showed that with high probability,

$$\|\log \hat{w} - \log w^*\|_2^2 \leqslant O\left(\frac{1}{pL}\right) \quad \text{and} \quad \|\log \hat{w} - \log w^*\|_\infty^2 \leqslant O\left(\frac{\log n}{npL}\right). \tag{1.1.3}$$

In the general case, Li et al. [29] have recently shown that with high probability,

$$\|\log \hat{w} - \log w^*\|_2^2 \lesssim \frac{d_{max} n}{L} \quad \text{and} \quad \|\log \hat{w} - \log w^*\|_\infty \lesssim \sqrt{\frac{n d_{max}}{L}} + \sqrt{\frac{n d_{max}^2}{L d_{min}^2}} \tag{1.1.4}$$

They also provide a minimax ℓ_∞ bound, that matches the upper bound (1.1.3) in the specific case of Erdös-Renyi graphs. Note that for Erdös-Renyi graphs, (1.1.4) do not achieve the same optimal bounds as (1.1.3).

RANK CENTRALITY. Negahban et al. [36] introduced in 2015 a spectral method, named Rank Centrality, to study the BTL model. The main idea is to build a transition matrix using the observations y_{ij} and then to consider its leading left eigenvector as an estimator of the normalized weights. Then, denoting $\pi^* = \frac{w^*}{\|w^*\|_1}$ the normalized vector of true scores, and $\hat{\pi}$ its estimate and assuming that each comparison between two items has been performed L times (for L large enough), it holds with high probability (w.h.p) that [8, 9]

$$\frac{\|\hat{\pi} - \pi^*\|_2}{\|\pi^*\|_2} \lesssim \frac{1}{\sqrt{npL}} \quad \text{and} \quad \frac{\|\hat{\pi} - \pi^*\|_\infty}{\|\pi^*\|_\infty} \lesssim \sqrt{\frac{\log n}{npL}}.$$

As for the MLE, Rank Centrality estimation achieves optimal guarantees in this setting.

Neghaban et al. [36] also provided error bounds for a general connected graph G. Denoting d_{max} the maximum degree of a vertex in G, it holds w.h.p. that for L large enough,

$$\frac{\|\hat{\pi} - \pi^*\|_2}{\|\pi^*\|_2} \lesssim \sqrt{\frac{\log n}{L d_{max}}}.$$

LEAST-SQUARE METHOD. The BTL model was also recently studied by Hendrickx et al. [18] using a weigthed least-squares algorithm. Denoting $\hat{V}_{ij} = \frac{y_{ij}}{y_{ji}} + \frac{y_{ji}}{y_{ij}} + 2$ to be the estimated variance of the observations $\log \frac{y_{ij}}{y_{ji}}$, their method consists in solving the following weighted least-squares problem.

$$\underset{w \in \mathbb{R}^n}{\operatorname{argmin}} \sum_{(i,j) \in \mathcal{E}} \frac{\left(\log \frac{y_{ij}}{y_{ji}} - \log \frac{w_i}{w_j}\right)^2}{\sqrt{\hat{V}_{ij}}}. \tag{1.1.5}$$

Denoting L_γ the Laplacian of the observation graph G with weights $\gamma_{ij} = \frac{1}{(w_i^* + w_j^*)^2}$ on edge $\{i, j\}$, and $L_\gamma^\dagger$ its Moore-Penrose inverse, it holds with high probability that

$$\mathbb{E}[\sin^2(\hat{w}, w^*)] \lesssim \frac{\mathrm{Tr}(L_\gamma^\dagger)}{L \|w^*\|_2^2}.$$

This result also comes with a matching lower bound on the estimation error, up to absolute constants.

1.1.3 *Other models*

Although BTL is one of the most popular models for ranking, other parametric models have also been studied in the literature. Two popular models are the Plackett-Luce model and the Thurstone model.

PLACKETT-LUCE MODEL. Recall that w_i^* denotes the strength of an item i, the discrete Plackett-Luce model [31] posits that for any item i, the probability of prefering item i over any alternatives in a set A is given by

$$\mathbb{P}(i|A) = \frac{w_i^*}{\sum_{j \in A} w_j^*}.$$

As for the BTL model, it has been thoroughly studied via Maximum Likelihood estimation [22] and spectral estimation [26], showing that both estimators are consistent.

THURSTONE MODEL. Another state of the art model for ranking is the Thurstone model, introduced in 1927 [44]. This model posits that the probabilities of preference are Gaussian functions of the weights w_i^*. A throrough analysis of this model was provided by Ennis [13] for various behavioral tasks, including ranking. It has been shown that maximum likelihood estimation and spectral method provide consistent estimators in this setting [13].

NON PARAMETRIC MODELS. All the aformentionned models are parametric models, which present some limitations. Indeed, parametric models often fail to fit accurately to real life data. Moreover, the MLE algorithm of Chen et al. [9] estimates accurately the weights under some regularity conditions that are not satisfy in real life dataset. Thus, non parametric models were introduced more recently to analyse ranking problems. These models often rely on the notion of strong stochastic transitivity (SST), as in the works of Shah et al. [42],[41]. This assumption implies the existence of an underlying ranking of the items such that if the items i, j, l are ranked in this order, then it holds that

$$\mathbb{P}(\text{item } i \text{ preferred to } j) \geqslant \mathbb{P}(\text{item } i \text{ preferred to } l).$$

Under this assumption, they provide guarantees on the recovery of the probability of preferences, achieving the same rate as in standard parametric models. However in this setting, it is non-trivial to compute the minimax-optimal estimator.

1.2 DYNAMIC RANKING

Most of the domains involving ranking can be related to dynamic datasets. Indeed, think about sports results, or individual preferences and opinons, they all vary along with time. However, the majority of the work revolving around ranking ignores this temporal component and focuses on the case of a unique time of observation, hence using a single comparison graph. To include time in a more dynamic setting, different models have been considered, mostly motivated by sports datasets (NFL, NBA, chess, etc.). These models are for the most part simple extensions of classical models for ranking, applied to a sequence of observation graphs. More precisely, Figure 4 shows how to extend a classical parametric model $\mathcal{M}$ to the dynamic setting, where G_t denotes the graph of observations at time t, and $w(t)$ are the parameters influencing the result of the comparisons at time t.

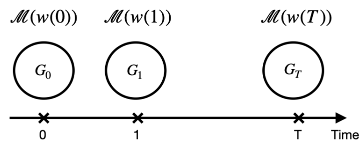


Figure 4: Extension of the model $\mathcal{M}$ to a dynamic setting.

Some works adapt classical ranking model, such as the BTL model [7],[17], the win-lose score [34] or the Poisson log-linear model [12]. Another considered approach was to use Bayesian inference on several datasets including the NBA and male tennis [32]. Glickman [16] proposed a predictive state-space model for NFL results. Note that none of these works provide statistical guarantees for the proposed methods but show how they perform on real datasets. From a theoretical point of view, very few articles have analyzed ranking models for different dynamic settings. The main works proposed extensions of existing static models following the principle of Figure 4. Moreover, they also defined precisely how to incorporate the temporal aspect of the data as assumptions on the time evolution of the model parameters. Let us describe two of the principal works done in this direction.

DYNAMIC BTL MODEL. Bong et al. [4] proposed a smoothly evolving dynamic BTL model, which is an extended version of the static BTL model presented in (1.1.1). They consider the logit[2] version of the BTL model

$$\text{logit}\left(\mathbb{P}(i \text{ beats } j \text{ at time } t)\right) = \beta_i^*(t) - \beta_j^*(t) \tag{1.2.1}$$

where $\beta^*(t)$ represents the vector of scores at time t with $w^*(t) = \exp(\beta^*(t))$. In this setting, the grid $\mathcal{T}$ of observations can be non-uniform, and the number of comparisons $L_{ij}(t)$ made for each pair $\{i, j\}$ at each time $t \in [0, 1]$ can vary. The pairwise comparison data at each time $t' \in \mathcal{T}$ are gathered into a matrix $X(t')$ where $X_{ij}(t')$ is the number of times i beat j at time t'. In order to use the temporal aspect of the data, they smooth the data using a kernel function. More precisely, to estimate the ranks at time t, they first compute the smoothed data as

$$\tilde{X}(t) = \sum_{t'=0}^{T} W_h(t', t) X(t')$$

where W_h is a kernel function with bandwidth h. Then, β^* is estimated by minimizing the negative log-likelihood, i.e.,

$$\underset{\beta : \sum_{i=1}^n \beta_i = 0}{\text{argmin}} \ \hat{R}(\beta; t) \equiv \underset{\beta : \sum_{i=1}^n \beta_i = 0}{\text{argmin}} \left(\sum_{i,j : i \neq j} \tilde{X}_{ij}(t) \log\left(1 + \exp(\beta_j(t) - \beta_i(t))\right) \right) \tag{1.2.2}$$

using a proximal gradient descent algorithm. If the matrix $|\tilde{X}(t)|$ of entrywise absolute values of $\tilde{X}(t)$ is considered to be the adjacency matrix of a weighted directed graph, then the strong connectivity of this graph is sufficient for the unique existence of the solution of (1.2.2). The main assumptions needed for their theoretical analysis are the following.

- The probabilities $\mathbb{P}(i \text{ beats } j \text{ at time } t)$ are Lispchitz functions of time $t \in [0, 1]$ for all $i \neq j \in [n]$ (same as Assumption 1).
- Each pair of teams $\{i, j\}$ has been compared at least at one time point $t' \in \mathcal{T}$. Translated to our notation, this means that the union graph $\cup_{t' \in \mathcal{T}} G_{t'}$ is complete.

2 $\text{logit}(x) = \ln\left(\frac{x}{1-x}\right)$ for $x \in]0, 1[$

Bong et al. derive bounds in the ℓ_∞ norm for estimating β^*. For a bandwidth $h \gtrsim \left(\frac{\log n}{T}\right)^{\frac{1}{3}}$, denoting $\hat{\beta}(t)$ to be the "MLE" estimator[3] (i.e., the solution of (1.2.2)), it is shown [4, Theorem 5.2] that with high probability,

$$\left\|\hat{\beta}(t) - \beta^*(t)\right\|_\infty \lesssim \delta_h(t) + \left(\frac{\log n}{nT}\right)^{\frac{1}{3}}$$

where $\delta_h(t)$ is denoted to be a discrepancy parameter in [4], and is small when all the teams play the same number of games at all times. Additionally, they also derive a rate for the uniform error over all time instants $t \in [0, 1]$. For $h \gtrsim \left(\frac{\log(nT^3)}{T}\right)^{\frac{1}{3}}$, it is shown [4, Theorem 5.3] that with high probability,

$$\sup_{t\in[0,1]} \left\|\hat{\beta}(t) - \beta(t)\right\|_\infty \lesssim \sup_{t\in[0,1]} \delta_h(t) + \left(\frac{\log(nT)}{T}\right)^{\frac{1}{3}}.$$

We will study in Chapter 2 a similar smoothly evolving dynamic BTL model using a spectral algorithm. Note that in our work, we will require analoguous assumptions as in [4]. More precisely, we assume that the probabilities are Lipschitz functions of time, as done by Bong et al. [4]. We also suppose that the union of the comparison graphs needs to be sufficiently connected(see more details in Section 2.1.)

NON-TRANSITIVE MATRIX RECOVERY [28]. Another important work in the dynamic setting is that of Li and Wakin [28]. In this work, they recover the pairwise comparison matrix $X(T)$ from noisy linear measurements of the matrices $X(t)$ for $i \in [T]$. They assume that the matrix $X(t)$ depends on two matrices $S(t)$, gathering the factors influencing the outcomes at time t.

$$X(t) = S(t)Q^\top - QS(t)^\top \quad \forall t \in [T].$$

The matrix Q contains information on factors that do not depend on time. The time-dependent data are contained in the matrix $S(t)$ and evolve under the following generative model.

$$S_t = S_{t-1} + E_t \quad \forall t \in [T]$$

3 Note that (1.2.2) is not the MLE for the dynamic BTL model with Lipschitz evolving win/loss probabilities (Assumption 1). It is in fact the MLE for the static BTL model, applied to the kernel-smoothed observations.

where E_t is an innovation matrix with i.i.d. centered Gaussian entries. In the case of a single factor ($r = 1$) with $Q \in \mathbb{R}^n$ being the all ones vector, the observations are for each pair of items $\{i, j\}$,

$$X_{ij}(t) = s_{t,i} - s_{t,j} \quad \forall t \in [T].$$

The outcomes of the comparisons, as in the BTL, only depend on the strength of each item at time t. Then by recovering the matrix $X(T)$, one can subsequently also derive a ranking of the items (see for eg., [49]). Denoting M to be the number of measurements available at each time t and $\hat{X}(T)$ the estimation of X computed as the solution of an optimization problem, it holds with high probability that

$$\|X(T) - \hat{X}(T)\|_F^2 \lesssim \max\left(n^2\sqrt{\frac{\log n}{MT}}, \frac{n^3 \log n}{MT}\right).$$

1.3 SYNCHRONIZATION PROBLEM

As discussed in the Introduction, synchronization has many applications such as computer vision or clocks synchronization but also ranking. Let us discuss here in more details the problem of Angular Synchronization [43] and of Translation Synchronization [21] as both of them are linked to ranking.

1.3.1 *Angular Synchronization*

Angular synchronization was first introduced and analyzed by Singer [43] as the problem of synchronization over the group $G = SO(2)$. In this problem, the goal is to recover angles from noisy measurements of angles differences. More precisely, denoting $\theta_1^*, \dots, \theta_n^* \in [0, 2\pi]$ to be the true angles, the angle differences θ_{ij} are obtained as follows.

$$\theta_{ij} = (\theta_i^* - \theta_j^* + \varepsilon_{ij}) \mod 2\pi,$$

where ε_{ij} denotes the noise. Then, to estimate the values θ_i^*, one needs to solve the maximization problem,

$$\max_{x \in C^n} x^H A x \tag{1.3.1}$$

where x^H denotes the conjugate transpose of x, $A_{ij} = e^{\iota\theta_{ij}} \mathbf{1}_{\{i,j\} \in \mathcal{E}}$ and $C^n := \{z \in \mathbf{C}^n \mid |z_i| = 1 \forall i\}$. Several relaxation of this NP-hard problem were proven to be efficient in the recovery of θ^*, such as the semi-definite programming relaxation (SDP) or the spectral relaxation [3]. Note that $X = xx^H$ is Hermitian positive semidefinite, has unit diagonal entries and is of rank one. Hence, dropping the rank constraint, the SDP relaxation of (1.3.1) can be written as

$$\max_{X \in \mathbf{C}^{n\times n}} \operatorname{Tr}(AX). \tag{1.3.2}$$

It has been shown [3] that if (1.3.2) admits a solution of rank 1, then this solution is a consistant estimate of a solution of (1.3.1) under a Wiegner noise model.

Another method used to solve (1.3.1) is to solve the following relaxation.

$$\max_{\|x\|^2=n} x^H A x. \tag{1.3.3}$$

The solution of (1.3.3) is the top eigenvector of A. Singer [43] has shown that this eigenvector was an accurate estimate of the vector $e^{\iota\theta^*}$.

SYNC-RANK. Synchronization over $SO(2)$ has been related to ranking as ranks $\{1, \dots, n\}$ can be mapped to the half-circle $[0, \pi]$. Hence, the rank differences can also be interpreted as angles differences, which is the exact setting of angular synchronization. This method, called SyncRank, has been computationally studied in [10]. The main idea is to create a matrix from the pairwise observations of angle differences θ_{ij}. We denote

$$A_{ij} = \begin{cases} e^{\iota\theta_{ij}} & \text{if} \{i, j\} \in \mathcal{E} \\ 0 & \text{otherwise.} \end{cases}$$

Then one obtains the estimates $\hat{v}_i = e^{\iota\hat{\theta}_i}$ as the normalized top eigenvector of A (as suggested by the works of Singer [43]). One can then extract the estimated angles $\hat{\theta}_i$, and then the corresponding ranks $\hat{r}$. The final estimation of the ranks will be the circular permutation of $\hat{r}$ that minimizes the number of upsets with respect to the observations.

This method, although performant on data, has not yet been studied from a theoretical point of view. It is also important to note that there exist no dynamic version of the Sync-Rank algorithm.

1.3.2 *Translation Synchronization*

In the setting of Translation Synchronization, the ranks can be mapped to the real line $\mathbb{R}$, corresponding to strength of each item. The least-squares approach of Hendrickx et al. [19],[18] is related to Translation Synchronization as the BTL model can we linked to this problem. Indeed, equation (1.1.5) can be seen as translation synchronization problem (see Remark 6 in Chapter 3 for more details) where the goal is to estimate the probabilities

$$\operatorname{logit}\left(\mathbb{P}\left(i \text{ prefered to } j\right)\right) = \log w_i^* - \log w_j^*. \tag{1.3.4}$$

As recalled in Section 2.1.2, they provide minimax rate and computational guarantees for their algorithm.

The problem of Translation Synchronization has been analyzed by Huang et al. [21]. The model posits that pairwise observations y_{ij} are noisy measurements of the strength differences $w_i^* - w_j^*$.

$$y_{ij} = w_i^* - w_j^* + \varepsilon_{ij}. \tag{1.3.5}$$

They propose an iterative algorithm that solves a truncated least-squares problem at each step, with a geometrically decaying truncation parameter. They provide ℓ_∞ bounds for the case of deterministic noise as well as for random biased noise. Note that for deterministic noise, it is shown that the estimation error is only bounded by the level of noise (up to constants), hence leading to exact recovery of the weights in the noiseless case. It is also shown that their method performs well on the Netflix Prize dataset.

We will study a dynamic version of this Translation Synchronization problem in the spirit of Figure 4, by applying the static model (1.3.5) at each time t in Chapter 3. In order to use the temporal aspect of the data, we will also assume that the weights $w_i^*(t)$ satisfy a global smoothness assumption (see Assumption 2), so that the quadratic variation of the strengths is upper bounded.

DENOISING SMOOTH SIGNALS OVER A GRAPH. We will propose in Chapter 3 two algorithms that jointly recover the weights. Our techniques are inspired from the literature on signal denoising in graphs. In particular, we draw inspiration from [40], where the authors consider a single graph $G = ([n], \mathcal{E})$, a ground truth signal $x^* \in \mathbb{R}^n$ and study the problem of estimating x^* under the observation model

$$y = x^* + \epsilon, \tag{1.3.6}$$

where ϵ is centered random noise. Denoting L to be the unnormalized Laplacian matrix of G, it is assumed that the quadratic variation of x^* (i.e., ${x^*}^\top L x^*$) is not large which means that the signal does not change quickly between neighboring vertices. This is directly linked to our smoothness assumption above. It is also equivalent to saying that x^* lies close to the subspace spanned by the eigenvectors corresponding to the small eigenvalues of L (i.e., the 'low frequency' part of L). In

[40], the authors show for sufficiently smooth x^* and with G assumed to be the grid graph that linear estimators such as *Laplacian smoothing* and *Laplacian eigenmaps* attain the minimax rate for estimating x^* in the ℓ_2 norm. We will introduce in Chapter 3 two estimators for the Dynamic TranSync model that are related to the Laplacian smoothing and Laplacian eigenmaps estimators.

2 DYNAMIC RANKING WITH THE BTL MODEL

In this chapter, we will study the dynamic Bradley-Terry-Luce model, introduced in Bong et al [4] (see (1.2.1), under a local smoothness assumption. We will adapt the classic Rank Centrality algorithm to the dynamic setting and provide ℓ_2 and ℓ_∞ error bounds on our estimator. Section 2.1 presents the dynamic BTL setup and our algorithm. We gather our main results in Section 2.2 and describe the ℓ_2 and ℓ_∞ analysis in Sections 2.3 and 2.4. Our theoretical results are then illustrated through experiments presented in Section 2.5.

2.1 PROBLEM SETUP AND ALGORITHM

NOTATION For any probability vector $\pi \in \mathbb{R}^n$ with strictly positive entries, we define the vector norm $\|x\|_\pi = \sqrt{\sum_{i=1}^n \pi_i x_i^2}$. For a matrix A, the corresponding induced matrix norm is then defined as $\|A\|_\pi = \sup_{\|x\|_\pi = 1} \|x^\top A\|_\pi$. Note that some simple inequalities follow from these definitions.

$$\sqrt{\pi_{\min}}\,\|x\|_2 \leqslant \|x\|_\pi \leqslant \sqrt{\pi_{\max}}\,\|x\|_2 \quad \text{and} \quad \sqrt{\frac{\pi_{\min}}{\pi_{\max}}}\,\|A\|_2 \leqslant \|A\|_\pi \leqslant \sqrt{\frac{\pi_{\max}}{\pi_{\min}}}\,\|A\|_2 \,. \tag{2.1.1}$$

2.1.1 *BTL model in a dynamic setting*

Let us formally introduce our model for dynamic pairwise comparisons, inspired by the Bradley-Terry-Luce (BTL) model. We consider a set of items $[n] = \{1, 2, \ldots, n\}$, with a certain quality at each time $t \in [0, 1]$, represented by the weight vector $w_t^* = (w_{t,1}^*, \ldots, w_{t,n}^*)^\top \in \mathbb{R}^n$ with $w_{t,i}^* > 0$ for each $i \in [n]$.

Our data consists of pairwise comparisons on this set of items at times t' on a regular grid $\mathcal{T} := \left\{\frac{i}{T} \middle| i = 0, \ldots, T\right\}$. The outcomes at each $t' \in \mathcal{T}$ are gathered into an undirected comparison graph $G_{t'} = ([n], \mathcal{E}_{t'})$ where $\mathcal{E}_{t'}$ is the set of edges. While the set of items $[n]$ is supposed to be the same throughout, the compared items, i.e. the set of edges $\mathcal{E}_{t'}$ can change with time.

To model such data, we use the BTL model at each time $t' \in \mathcal{T}$. This model posits that the probability that an item j wins over an item i is proportional to its strength. At each $t' \in \mathcal{T}$, for each pair of compared items $\{i, j\} \in \mathcal{E}_{t'}$, we perform L independent comparisons.

Their outcomes are independent Bernoulli variables, defined for $l \in \{1, \dots, L\}$ by $y_{ij}^{(l)}(t')$ where

$$\mathbb{P}(y_{ij}^{(l)}(t') = 1) = \frac{w^*_{t',j}}{w^*_{t',i} + w^*_{t',j}}.$$

The proportion of times j won over i at time t' is given by $y_{ij}(t') = \frac{1}{L}\sum_{l=1}^{L} y_{ij}^{(l)}(t')$ and the corresponding true proportion is denoted by (for any $t \in [0,1]$)

$$y^*_{ij}(t) := \mathbb{E}[y_{ij}(t)] = \frac{w^*_{t,j}}{w^*_{t,i} + w^*_{t,j}} \quad \forall i \neq j.$$

SMOOTH EVOLUTION OF PAIRWISE OUTCOMES. Our goal is to recover w^*_t at any time $t \in [0,1]$. Suppose for the moment that t is on the grid, then if G_t is connected, w^*_t is identifiable up to a positive scaling. However in our dynamic setting G_t can be very sparse, and is not necessarily connected. Therefore for meaningful recovery of w^*_t, we need to make additional assumptions on the evolution of the pairwise outcomes over time. To this end, we make the following smoothness assumption.

Assumption 1 (Lipschitz smoothness). *There exists* $M \geqslant 0$ *such that*

$$|y^*_{ij}(t) - y^*_{ij}(t')| \leqslant M|t - t'| \quad \forall t, t' \in [0,1], \quad i \neq j \in [n]. \tag{2.1.2}$$

This assumption suggests that the pairwise outcomes at nearby time instants are similar, hence it is plausible that w^*_t (at any $t \in [0,1]$) could be estimated by utilizing the data lying in a neighborhood of t. To formalize this intuition, let us define a neighborhood at any time t by

$$\mathcal{N}_\delta(t) := \left\{ t' \in \mathcal{T} \mid |t - t'| \leqslant \frac{\delta}{T} \right\}. \tag{2.1.3}$$

where $\delta \in [0, T]$. Note that if $\delta < \frac{1}{2}$, then there exists some values of t for which $\mathcal{N}_\delta(t)$ is empty; hence we will consider $\delta \in [1/2, T]$. It is easy to verify that $\delta \leqslant |\mathcal{N}_\delta(t)| \leqslant 4\delta$ for all $t \in [0,1]$.

NEIGHBORHOOD GRAPH. For any time $t \in [0,1]$, the data contained in the neighborhood $\mathcal{N}_\delta(t)$ can be gathered into a union graph, defined as $G_\delta(t) = ([n], \mathcal{E}_\delta(t))$ where $\mathcal{E}_\delta(t) = \cup_{t' \in \mathcal{N}_\delta(t)} \mathcal{E}_{t'}$. The maximum and minimum degree of a vertex in $G_\delta(t)$ will be denoted by $d_{\max,\delta}(t)$ and $d_{\min,\delta}(t)$ respectively. For each $i \neq j$, it will also be useful to denote the time instants in $\mathcal{N}_\delta(t)$ where i and j are compared as

$$\mathcal{N}_{ij,\delta}(t) = \left\{t' \in \mathcal{N}_\delta(t) | \, \{i,j\} \in \mathcal{E}_{t'}\right\},$$

along with the quantities

$$N_{\max,\delta}(t) = \max_{\{i,j\}\in\mathcal{E}_\delta(t)} |\mathcal{N}_{ij,\delta}(t)| \quad \text{and} \quad N_{\min,\delta}(t) = \min_{\{i,j\}\in\mathcal{E}_\delta(t)} |\mathcal{N}_{ij,\delta}(t)|.$$

Note that alternatively,

$$\mathcal{E}_\delta(t) = \left\{\{i,j\} : i \neq j, \, |\mathcal{N}_{ij,\delta}(t)| \geqslant 1\right\}.$$

GENERAL RECOVERY IDEA. Given the above setup, a general idea for recovering w^*_t is to first form the union graph $G_\delta(t)$, and to then compute the statistics

$$\bar{y}_{ij}(t) := \frac{1}{|\mathcal{N}_{ij,\delta}(t)|} \sum_{t'\in\mathcal{N}_{ij,\delta}(t)} y_{ij}(t') \quad \text{for} \quad \{i,j\} \in \mathcal{E}_\delta(t). \tag{2.1.4}$$

Suppose for convenience that $L \to \infty$, we have for each $\{i,j\} \in \mathcal{E}_\delta(t)$ that

$$\bar{y}_{ij}(t) \xrightarrow{L\to\infty} \bar{y}^*_{ij}(t) = \frac{1}{|\mathcal{N}_{ij,\delta}(t)|} \sum_{t'\in\mathcal{N}_{ij,\delta}(t)} y^*_{ij}(t').$$

Due to Assumption 1 we know that

$$\left|\bar{y}^*_{ij}(t) - y^*_{ij}(t)\right| \leqslant M\delta/T, \qquad \text{for } \{i,j\} \in \mathcal{E}_\delta(t),$$

hence if $\delta = o(T)$, then for each $\{i,j\} \in \mathcal{E}_\delta(t)$, we have that $\bar{y}^*_{ij}(t)$ converges to $y^*_{ij}(t)$ as $T \to \infty$. Moreover, if the corresponding sequence of graphs $G_\delta(t)$ is connected, the identifiability of w^*_t (up to a positive scaling) is ensured. The connectivity requirement on $G_\delta(t)$ is of course much weaker than requiring the individual graph(s) to be connected, which is also a key difference between the static and dynamic settings. Also note that the computed statistcs $\bar{y}_{ij}(t)$ are nearest neighbor estimators as they average the data over a suitable neighborhood of t. More generally, one could also compute $\bar{y}_{ij}(t)$ via kernel smoothing as considered in [4]. As will be shown later, even if $L = 1$, consistent recovery of w^*_t is possible provided that $\delta = o(T)$, $\delta \overset{T\to\infty}{\longrightarrow} \infty$ and that every $\{i,j\} \in \mathcal{E}_\delta(t)$ satisfies $|\mathcal{N}_{ij,\delta}(t)| \overset{T\to\infty}{\longrightarrow} \infty$.

With the above discussion in mind, a general scheme for recovering w^*_t for any given $t \in [0,1]$ would be to first form $G_\delta(t)$ for a suitable choice of the parameter δ, then compute the statistics $\bar{y}_{ij}(t)$ as in (2.1.4), and finally apply any existing method for the static case using the comparison graph $G_\delta(t)$ and the data $(\bar{y}_{ij}(t))_{\{i,j\}\in\mathcal{E}_\delta(t)}$. We will focus on the Rank Centrality algorithm of Negahban et al. [36] – a popular spectral algorithm known to achieve state of the art performance – and adapt it to our dynamic setting.

2.1.2 *Spectral dynamic ranking*

The Rank Centrality [36] method is based on the connection between pairwise comparisons and a random walk on a directed graph. In the static case, a faithful estimation of the weight vector is given by the stationary distribution of the Markov chain induced by a suitably constructed transition matrix. In the dynamic setting, this method can be adapted for estimating w_t^* by now constructing the transition matrix using $G_\delta(t)$ and the data $(\bar{y}_{ij}(t))_{\{i,j\}\in\mathcal{E}_\delta(t)}$, as in (2.1.4).

More precisely, we define a transition matrix $\hat{P}(t)$ on this graph using $(\bar{y}_{ij}(t))_{\{i,j\}\in\mathcal{E}_\delta(t)}$, with

$$\hat{P}_{ij}(t) = \begin{cases} \frac{\bar{y}_{ij}(t)}{d_\delta(t)} = \frac{1}{d_\delta(t)}\left(\frac{1}{|\mathcal{N}_{ij,\delta}(t)|}\sum_{t'\in\mathcal{N}_{ij,\delta}(t)} y_{ij}(t')\right) & \text{if} \quad \{i,j\}\in\mathcal{E}_\delta(t) \\ 1-\frac{1}{d_\delta(t)}\left(\sum_{k\neq i}\frac{1}{|\mathcal{N}_{ik,\delta}(t)|}\sum_{t'\in\mathcal{N}_{ik,\delta}(t)} y_{ik}(t')\right) & \text{if} \quad i=j \\ 0 & \text{otherwise} \end{cases} \tag{2.1.5}$$

where $d_\delta(t) \geqslant d_{\max,\delta}(t)$ is a suitably chosen normalization term. Then, one can easily verify that at each time t, $\hat{P}(t)$ is a transition matrix ($\hat{P}(t)$ is stochastic) corresponding to a Markov chain on a finite state space. Thus there always exists at least one stationary distribution associated to $\hat{P}(t)$. Moreover, stochastic matrices admit 1 as leading eigenvalue and so a candidate as the stationary distribution is its leading left eigenvector, i.e.

$$\hat{\pi}(t)^\top = \hat{\pi}(t)^\top \hat{P}(t).$$

Besides, the vector of true weights w_t^* we want to recover can be seen as the stationary distribution of a transition matrix on the union graph. Specifically, denoting $\pi^*(t) = \frac{w_t^*}{\sum_{i=1}^n w_{t,i}^*}$, one can easily show that $\pi^*(t)$ is the stationary distribution of the transition matrix

$$\bar{P}_{ij}(t) = \begin{cases} \frac{1}{d_\delta(t)}\frac{w_{t,j}^*}{w_{t,i}^*+w_{t,j}^*} & \text{if} \quad \{i,j\}\in\mathcal{E}_\delta(t) \\ 1-\frac{1}{d_\delta(t)}\sum_{k\neq i}\frac{w_{t,k}^*}{w_{t,i}^*+w_{t,k}^*} & \text{if} \quad i=j \\ 0 & \text{otherwise} \end{cases} \tag{2.1.6}$$

since $\bar{P}(t)$ and $\pi^*(t)$ verify the detailed balance equation of reversibility [27]

$$\bar{P}_{ij}(t)\pi_i^*(t) = \bar{P}_{ji}(t)\pi_j^*(t) \quad \forall i,j \in [n].$$

One can reasonably expect $\hat{\pi}(t)$ to be close to $\pi^*(t)$ as they are stationary distributions of $\hat{P}(t)$ and $\bar{P}(t)$ respectively, the latter of which are expected to be close. Indeed, one has the following bias-variance trade-off

$$\hat{P}(t) - \bar{P}(t) = \underbrace{\hat{P}(t) - \mathbb{E}[\hat{P}(t)]}_{\text{variance}} + \underbrace{\mathbb{E}[\hat{P}(t)] - \bar{P}(t)}_{\text{bias}}.$$

where the variance term is typically expected to decrease with δ (due to averaging over $\mathcal{N}_\delta(t)$) while the bias term will scale as $O(\delta/T)$ (due to the smoothness assumption 1). Hence for a suitably chosen $\delta = o(T)$ we will then have (for n, T large enough) $\hat{P}(t) \approx \bar{P}(t)$, which implies $\hat{\pi}(t) \approx \pi^*(t)$.

Remark 1. *For meaningful recovery, the vector $\pi^*(t)$ clearly has to be unique. This is the case if the associated Markov chain is irreducible which in turn is ensured by the connectivity of the underlying graph (here, the union graph $G_\delta(t)$) and the strict positivity of the weights on its edges [27]. The condition on the weights is guaranteed in our setup since $\bar{P}_{ij}(t) > 0$ for each $\{i,j\} \in \mathcal{E}_\delta(t)$ (indeed, $w_{t,i}^* > 0$ for each $t \in [0,1]$ and $i \in [n]$).*

Based on the above discussion we can outline the steps of our method for ranking in the dynamic setting in the form of Algorithm 1. Our goal now is to establish conditions under which $\hat{\pi}(t)$ is close

---

Algorithm 1 Spectral algorithm for dynamic ranking (Dynamic Rank Centrality)

1: **Input:** Grid $\mathcal{T} \subset [0,1]$, and a given time $t \in [0,1]$. For each $t' \in \mathcal{T}$: comparison graph $G_{t'}$, results of comparisons as statistics $(y_{ij}(t'))_{\{i,j\}\in\mathcal{E}_{t'}}$.
2: Form the neighbourhood graph $G_\delta(t) = ([n], \mathcal{E}_\delta(t))$ where $\mathcal{E}_\delta(t) = \cup_{t'\in\mathcal{N}_\delta(t)}\mathcal{E}_{t'}$, and $\mathcal{N}_\delta(t)$ is as in (2.1.3).
3: Compute the transition matrix $\hat{P}(t)$ as in (2.1.5) with $d_\delta(t) \geqslant d_{max,\delta}(t)$.
4: Compute the leading left eigenvector $\hat{\pi}(t)$ of $\hat{P}(t)$.
5: **Output:** $\hat{\pi}(t) \in \mathbb{R}^n$.

---

to $\pi^*(t)$ under the ℓ_2 and ℓ_∞ norms. These results are summarized in the next section. In particular, we will strive to establish consistency results (i.e., the error approaching zero) when the grid size $T \to \infty$.

ADDITIONAL DEFINITIONS. Before proceeding, we need to define some additional quantities, some being related to the union graph $G_\delta(t)$, which will appear in the following sections. Let $\mathcal{L}_\delta(t) =$

$D_\delta^{-1}(t)A_\delta(t)$ denote the random walk Laplacian of $G_\delta(t)$ [6] , where $D_\delta(t)$ is the diagonal matrix of vertex degrees, and $A_\delta(t)$ is its adjacency matrix. We denote $\xi_\delta(t) = 1 - \lambda_{max}(\mathcal{L}_\delta(t))$ where

$$\lambda_{max}(\mathcal{L}_\delta(t)) := \max\{\lambda_2(\mathcal{L}_\delta(t)), -\lambda_n(\mathcal{L}_\delta(t))\},$$

is the second largest eigenvalue (in absolute value) of $\mathcal{L}_\delta(t)$. Note that $\mathcal{L}_\delta(t)$ has real eigenvalues since it is similar to the symmetric Laplacian $D_\delta^{-1/2}(t)A_\delta(t)D_\delta^{-1/2}(t)$. Let us also denote

$$b(t) := \max_{i,j\in[n]} \frac{w^*_{t,i}}{w^*_{t,j}} \text{ for all } t \in [0,1]$$

where we will require that $b(t)$ is finite for each $t \in [0,1]$.

2.2 MAIN RESULTS

In Section 2.2.1, we present bounds on the ℓ_2 error $\|\hat{\pi}(t) - \pi^*(t)\|_2$, while Section 2.2.2 contains our bounds on the ℓ_∞ error $\|\hat{\pi}(t) - \pi^*(t)\|_\infty$. A summary of the notation used in the paper is outlined in tabular form in Section 2.7.

2.2.1 ℓ_2 *error bound*

For a given sequence of graphs $(G(t'))_{t'\in\mathcal{T}}$, the following theorem provides an explicit ℓ_2 error bound (holding w.h.p) which in particular highlights the dependence on parameters related to the union graph $G_\delta(t)$, the grid size T and the neighborhood size δ. The proofs of results in this section are outlined in Section 2.3.

Theorem 1. *For any given* $t \in [0,1]$*, suppose that* $\delta \in [\frac{1}{2}, T]$ *is such that* $n \geqslant c_1 \log n$ *and* $\xi_\delta(t) > 0$ *for some constant* $c_1 > 0$*. Choosing* $d_\delta(t) \geqslant d_{max,\delta}(t)$*, there exist constants* $\tilde{C_1} \geqslant 15, \tilde{C_2} \geqslant 1$ *such that if*

$$\tilde{C_1}\sqrt{\frac{N_{max,\delta}(t)d_{max,\delta}(t)\log n}{Ld_\delta^2(t)N^2_{min,\delta}(t)}} + 4\frac{M\delta|\mathcal{E}_\delta(t)|}{Td_{max,\delta}(t)} \leqslant \frac{\xi_\delta(t)d_{min,\delta}(t)}{8d_\delta(t)b^{7/2}(t)}, \quad (2.2.1)$$

then it holds with probability at least $1 - O(n^{-10})$ *that*

$$\frac{\|\hat{\pi}(t) - \pi^*(t)\|_2}{\|\pi^*(t)\|_2} \leqslant 32\frac{M\delta|\mathcal{E}_\delta(t)|b^{7/2}(t)d_\delta(t)}{T\xi_\delta(t)d_{min,\delta}(t)d_{max,\delta}(t)} + 8\tilde{C_2}\frac{b^{9/2}(t)}{\xi_\delta(t)d_{min,\delta}(t)}\sqrt{\frac{N_{max,\delta}(t)d_{max,\delta}(t)}{LN^2_{min,\delta}(t)}}. \quad (2.2.2)$$

Let us make the following observations.

1. The first term in the RHS of (2.2.2) corresponds to the bias and arises from the regularity assumption 1, while the second term

therein is the variance term. Moreover, note that the error depends on δ – either explicitly, or through certain quantities such as $d_{max,\delta}(t), N_{min,\delta}(t)$ etc. In order to obtain a more explicit dependence in terms of δ, we will need to make specific assumptions on the graphs $G_{t'}$, $t' \in \mathcal{T}$. Below, we will consider the setting where the graphs are Erdös-Renyi graphs and derive explicit conditions on δ that lead to consistency with respect to T.

2. The condition (2.2.1) arises from the eigenvector perturbation result in [9, Theorem 5.1] which requires the noise term (i.e., $\hat{P}(t) - \bar{P}(t)$) to be small compared to the spectral gap of $\bar{P}(t)$ (i.e., $1 - \lambda_{max}(\bar{P}(t))$), see (2.3.5).

3. In the static case, we have $t = t'$ for some $t' \in \mathcal{T}$ and only the graph $G_{t'}$ is observed. Then $M = 0$ and $\delta = 1/2$, so $N_{min,\delta}(t)$, $N_{max,\delta}(t) \equiv 1$. Denoting $d_{min}(t), d_{max}(t), \xi(t)$ to be the corresponding quantities with the δ suffix suppressed, condition (2.2.1) is satisfied for L large enough. Moreover, the error bound is then $O(\frac{b^{9/2}(t)}{\xi(t)d_{min}(t)}\sqrt{\frac{d_{max}(t)}{L}})$ which matches the ℓ_2 bound of Negahban et al. [36, Theorem 1] with the $\sqrt{\log n}$ factor therein removed, but with an extra $b^2(t)$ factor.

Remark 2. *The term $N_{max,\delta}(t)$ is admittedly counterintuitive and is an artifact of the proof technique. In particular, this occurs due to certain concentration inequalities used within the proof (e.g., Lemma 3). To our knowledge, similar issues would arise in the static case for analyzing the Rank Centrality method if each comparison $\{i, j\}$ was made L_{ij} times. This would occur, for instance, in the proof of [36, Lemma 3] due to Hoeffding's inequality, in which case both the minimum and maximum of the L_{ij}'s would appear in the bounds.*

Now we consider the important case where the comparison graphs are Erdös-Renyi graphs, i.e., $G_{t'} = \mathcal{G}(n, p(t'))$ for each $t' \in \mathcal{T}$. It is easily seen that the union graph $G_\delta(t)$ is also then Erdös-Renyi denoted by $\mathcal{G}(n, p_\delta(t))$ where $p_\delta(t)$ is given by

$$p_\delta(t) = 1 - \prod_{t' \in \mathcal{N}_\delta(t)} (1 - p(t')). \tag{2.2.3}$$

In this setting, the bound in Theorem 1 can be simplfied using concentration results for parameters related to $G_\delta(t)$ (see Lemma 21). Specifically, we have that if $p_\delta(t) \gtrsim \log n/n$, then w.h.p

$$\frac{np_\delta(t)}{2} \leqslant d_{max,\delta}(t), d_{min,\delta}(t) \leqslant \frac{3np_\delta(t)}{2}; \quad \xi_\delta(t) \geqslant \frac{1}{2}; \quad |\mathcal{E}_\delta(t)| \leqslant 2n^2 p_\delta(t).$$

In particular, we will choose the normalization factor $d_\delta(t) = 3np_\delta(t)$ which is a valid choice (w.h.p). Lemma 21 also states that if

$p_{\delta,sum}(t) := \sum_{t' \in \mathcal{N}_\delta(t)} p(t') \gtrsim \log n$, then $N_{max,\delta}(t), N_{min,\delta}(t) \asymp p_{\delta,sum}(t)$ w.h.p. These considerations lead to the following simplification of Theorem 1.

Theorem 2. *Suppose that $G_{t'} \sim \mathcal{G}(n, p(t'))$ for all $t' \in \mathcal{T}$ so that $G_\delta(t) \sim \mathcal{G}(n, p_\delta(t))$ (for any given $t \in [0,1]$) with $p_\delta(t)$ as in (2.2.3), and denote $p_{\delta,sum}(t) := \sum_{t' \in \mathcal{N}_\delta(t)} p(t')$. Choosing $d_\delta(t) = 3np_\delta(t)$, let $\delta \in [\frac{1}{2}, T]$ be such that $n \geqslant c_1 \log n$, $np_\delta(t) \geqslant c_0 \log n$, and $p_{\delta,sum}(t) \geqslant c_2 \log n$ with constant $c_1 > 0$ as in Theorem 1, and constants $c_0, c_2 \geqslant 1$. Then for constants $\tilde{C_1}, \tilde{C_2}$ as in Theorem 1, if*

$$2\tilde{C_1}\sqrt{\frac{\log n}{Lnp_\delta(t)p_{\delta,sum}(t)}} + 16\frac{M\delta n}{T} \leqslant \frac{1}{96 b^{7/2}(t)} \tag{2.2.4}$$

holds, we have with probability at least $1 - O(n^{-10})$ that

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

The following remarks are in order.

1. As can be seen, the bias term is $O(\frac{n\delta}{T})$ while the variance term scales as $O(\frac{1}{\sqrt{Lnp_\delta(t)p_{\delta,sum}(t)}})$. Hence if $p_{\delta,sum}(t)$ grows with δ, then the variance error will reduce as δ increases. Furthermore, if $\delta = o(T)$ and δ increases with T then it would imply that the LHS of (2.2.4) decreases with T, and hence (2.2.4) will be satisfied for T sufficiently large.

2. In the static case we observe a single comparison graph $G_{t'}$ (for $t' \in \mathcal{T}$) with $t = t'$. Thus $p_{\delta,sum}(t) \equiv 1$ and the condition $p_{\delta,sum}(t) \gtrsim \log n$ is not needed, while $p_\delta(t) = p(t)$. Hence, if $p(t) \gtrsim \frac{\log n}{n}$ and L is suitably large, the ℓ_2 error is bounded by $O(\frac{1}{\sqrt{Lnp(t)}})$, which corresponds to the bound obtained by Chen et al. [9, Theorem 5.2]. So our result is coherent with existing results for the static case for Erdös-Renyi graphs.

3. The choice $d_\delta(t) = 3np_\delta(t)$ ensures that $d_{max,\delta}(t) \leqslant d_\delta(t)$ w.h.p. In fact, we could have chosen $d_\delta(t)$ to be a constant ($\geqslant 1$) multiple of $d_{max,\delta}(t)$ as well. However for the ℓ_∞ analysis later on, it will be crucial to choose $d_\delta(t)$ as a constant times $np_\delta(t)$ for technical reasons arising in the analysis. Similar considerations for the choice of the normalization factor exist in the static setting as well (see [8, 9]). Note that this choice of $d_\delta(t)$ requires us to know $p_\delta(t)$, but in case we don't know $p_\delta(t)$ in practice, we can instead use its empirical estimate which can be easily computed.

We now derive an appropriate choice for δ that leads to an ℓ_2 error rate of $O(T^{-1/3})$. To this end, we first need to explicitly show the dependence on δ for $p_\delta(t), p_{\delta,sum}(t)$. Let us assume for simplicity that

$$p_{\min} := \min_{t' \in \mathcal{T}} p(t') > 0. \tag{2.2.5}$$

Since $\delta \leqslant |\mathcal{N}_\delta(t)| \leqslant 4\delta$, we have for all $t \in [0,1]$ that $p_{\delta,sum}(t) \geqslant \delta p_{\min}$. Besides, as shown in Proposition 4, $p_\delta(t) \gtrsim \min\{1, \delta p_{\min}\}$. Hence if $\delta p_{\min} \gtrsim \log n$ then it implies

$$p_{\delta,sum}(t) \gtrsim \log n \quad \text{and} \quad p_\delta(t) \gtrsim 1 \; (\geqslant \log n/n)$$

meaning that the conditions on $p_\delta(t)$ and $p_{\delta,sum}(t)$ in Theorem 2 are satisfied.

Remark 3. *The condition $p_{\delta,sum}(t) \gtrsim \log n$ is needed to ensure that $N_{\min,\delta}(t), N_{\max,\delta}(t)$ concentrate around $p_{\delta,sum}(t)$, as shown in Lemma 21. This is in fact a strong condition as requiring $\delta p_{\min} \gtrsim \log n$ imposes that the union graph $G_\delta(t)$ is complete (see proof of Lemma 21). We will show in Section 2.2.3 how to weaken this assumption.*

Corollary 1. *Under the same notations as in Theorem 2, for all $t \in [0,1]$ suppose that $n \gtrsim \log n$ and $p_{\min}$ is as in (2.2.5).*
Choosing $\delta = \min\left\{\frac{(b(t))^{\frac{2}{3}}}{(2M)^{\frac{2}{3}} n (Lp_{\min})^{\frac{1}{3}}} T^{2/3}, T\right\}$ and $d_\delta(t) = 3np_\delta(t)$, if T is such that $\delta \gtrsim \frac{\log n}{p_{\min}}$ and

$$\frac{1}{b^{7/2}(t)} \gtrsim \sqrt{\frac{\log n}{Lnp_{\min}}} \max\left\{\frac{(2M)^{1/3} n^{1/2} (Lp_{\min})^{1/6}}{b^{1/3}(t) T^{1/3}}, \frac{1}{\sqrt{T}}\right\}$$
$$+ Mn \min\left\{\frac{b^{2/3}(t)}{(2M)^{2/3} n (Lp_{\min})^{1/3} T^{1/3}}, 1\right\},$$

then with probability at least $1 - O(n^{-10})$,

$$\frac{\|\hat{\pi}(t) - \pi^*(t)\|_2}{\|\pi^*(t)\|_2} \lesssim Mnb^{7/2}(t) \min\left\{\frac{b^{2/3}(t)}{(2M)^{2/3} n (Lp_{\min})^{1/3} T^{1/3}}, 1\right\}$$
$$+ \frac{b^{9/2}(t)}{\sqrt{Lnp_{\min}}} \max\left\{\frac{(2M)^{1/3} n^{1/2} (Lp_{\min})^{1/6}}{b^{1/3}(t) T^{1/3}}, \frac{1}{\sqrt{T}}\right\}.$$

The following observations are useful to note.

1. When $M > 0$, Corollary 1 states that for $\delta \asymp T^{2/3}$, if n, T are large enough (thus ensuring that all the stated conditions are satisfied)), then w.h.p $\|\hat{\pi}(t) - \pi^*(t)\|_2 = O(T^{-1/3})$. This matches the rate for the pointwise risk for estimating univariate Lipschitz functions (see for e.g. [37, Theorem 1.3.1]).

2. If $M = 0$ then $\delta = T$ which makes sense since $y^*_{ij}(t)$ is a constant function for each $i \neq j$. Indeed, the problem is then the same as the setting where the comparison graph is $\cup_{t' \in \mathcal{T}} G_{t'}$, and we observe (a potentially different number of) i.i.d pairwise outcomes for each given edge in this graph. In this case, the corollary states that provided T is large enough, the ℓ_2 error is $\lesssim \frac{b^{9/2}(t)}{\sqrt{Lnp_{\min} T}}$. This is logically faster than the $T^{-1/3}$ nonparametric rate, and is analogous to the optimal ℓ_2 bound for Erdös Renyi graphs in the static setting (see [9, Theorem 5.2]).

2.2.2 ℓ_∞ *error bound*

We now discuss our results for bounding the ℓ_∞ error $\|\hat{\pi}(t) - \pi^*(t)\|_\infty$ at any given time t. Such bounds are particularly desirable in the context of ranking as they lead to guarantees for recovering the ranks of the items. We will assume that all the comparison graphs (at each $t' \in \mathcal{T}$) are Erdös-Renyi graphs. The following theorem is the ℓ_∞ counterpart of Theorem 2, the proofs of results in this section are outlined in Section 2.4.

Theorem 3. *Under the notation and assumptions of Theorem 2, there exists a constant* $\tilde{C_3} \geqslant 1$ *such that if additionally*

$$96 b^{\frac{5}{2}}(t) \left(\frac{4M\delta}{T} + \tilde{C_3} \sqrt{\frac{\log n}{n p_\delta(t)}} \right) \leqslant \frac{1}{2}, \tag{2.2.6}$$

then there exist constants $\tilde{C_4}, \tilde{C_5}, \tilde{C_6} \geqslant 1$ *such that with probability at least* $1 - O(n^{-9})$,

$$\frac{\|\hat{\pi}(t) - \pi^*(t)\|_\infty}{\|\pi^*(t)\|_\infty} \leqslant \left(\tilde{C_5} \gamma_{n,\delta}(t) \sqrt{\frac{\log n}{Lnp_\delta(t) p_{\delta,sum}(t)}} + \tilde{C_6} \frac{Mn\delta b^{\frac{7}{2}}(t)}{T} \right) \times \frac{12 b_{\max,\delta}(t)}{1 - \tilde{C_4} b_{\max,\delta}(t) \sqrt{\frac{\log n}{n p_\delta(t)}}},$$

where $\gamma_{n,\delta}(t) := (1 + \frac{b^{\frac{5}{2}}(t)}{\sqrt{\log n}} \max\{b^2(t), \frac{\log n}{\sqrt{n p_\delta(t)}}\})$, $b_{\max,\delta}(t) := \max_{t' \in \mathcal{N}_\delta(t)} b(t')$.

As before for Theorem 2, let us interpret Theorem 3 for the static setting where $t = t'$ for some $t' \in \mathcal{T}$, and only $G_{t'}$ is observed. Then Theorem 3 states that if $np(t) \gtrsim b^5(t) \log n$, and n, L are large enough, then w.h.p, the ℓ_∞ error is

$$\frac{\|\hat{\pi}(t) - \pi^*(t)\|_\infty}{\|\pi^*(t)\|_\infty} \lesssim b(t) \left(1 + \frac{b^{\frac{5}{2}}(t)}{\sqrt{\log n}} \max\{b^2(t), \frac{\log n}{\sqrt{np(t)}}\} \right) \sqrt{\frac{\log n}{Lnp(t)}}.$$

Hence if $b(t) = O(1)$ then the bound is $O(\sqrt{\frac{\log n}{Lnp(t)}})$ which matches the corresponding bound of Chen et al. [9, Theorem 3].

Let us now denote

$$b_{\max} := \max_{t' \in [0,1]} b(t'), \quad \gamma_n(t) := \left(1 + \frac{b^{\frac{5}{2}}(t)}{\sqrt{\log n}} \max\{b^2(t), \frac{\log n}{\sqrt{n}}\}\right) \tag{2.2.7}$$

so that $b_{\max,\delta}(t) \leqslant b_{\max}$. Since $p_\delta(t) \gtrsim 1$ provided that $\delta p_{\min} \gtrsim \log n$, hence $\gamma_{n,\delta}(t) \lesssim \gamma_n(t)$.

Remark 4. *Suppose that for every* $t \in [0,1]$ *and* $i \neq j$, $y^*_{ij}(t) > y^*_{\min}$, *with* $y^*_{\min} \in (0, \frac{1}{2})$. *Then one can show that* $b(t)$ *is a Lipschitz function with constant* $\frac{M}{y^{*2}_{\min}}$. *Note that*

$$\frac{1}{y^*_{ij}(t)} = \frac{w^*_{t,j}}{w^*_{t,i} + w^*_{t,j}} \implies b_{\max} = \frac{1}{y^*_{\min}} - 1.$$

Hence, if $y^*_{\min} \gtrsim 1$, *then* $b_{\max} = O(1)$ *and consequently for every* $t' \in [0,1]$, $b(t') = O(1)$.

Then as for the ℓ_2 case, one can derive a value for δ that leads to a ℓ_∞ error rate of $T^{-\frac{1}{3}}$.

Corollary 2. *Under the same notations as in Theorem 3, for all* $t \in [0,1]$ *suppose that* $n \gtrsim \log n$, $p_{\min}$ *is as in* (2.2.5) *and* $b_{\max}, \gamma_n(t)$ *are as as in* (2.2.7). *Choosing* $\delta = \min\left\{ \frac{(\gamma_n(t))^{\frac{2}{3}} (\log n)^{\frac{1}{3}}}{(2M)^{\frac{2}{3}} n b^{\frac{7}{3}}(t) (Lp_{\min})^{\frac{1}{3}}} T^{2/3}, T \right\}$ *and* $d_\delta(t) = 3np_\delta(t)$, *if* T *is such that* $\delta \gtrsim \frac{\log n}{p_{\min}}$ *and*

$$\min\left\{ \left(\frac{(\gamma_n(t))^{\frac{2}{3}} (M \log n)^{\frac{1}{3}}}{2^{\frac{2}{3}} n b^{\frac{7}{3}}(t) (Lp_{\min})^{\frac{1}{3}}} \right) \frac{1}{T^{1/3}}, M \right\} + \sqrt{\frac{\log n}{n}} \lesssim \frac{1}{b^{5/2}(t)},$$

then with probability at least $1 - O(n^{-9})$,

$$\begin{aligned} \frac{\|\hat{\pi}(t) - \pi^*(t)\|_\infty}{\|\pi^*(t)\|_\infty} \lesssim & \left(\frac{b_{\max}}{1 - b_{\max}\sqrt{\frac{\log n}{n}}} \right) \\ & \times \Bigg[M n b^{7/2}(t) \min\left\{ \frac{\gamma_n^{2/3}(t) (\log n)^{1/3}}{(2M)^{2/3} n b^{7/3}(t) (Lp_{\min})^{1/3} T^{1/3}}, 1 \right\} \\ & + \gamma_n(t) \sqrt{\frac{\log n}{L n p_{\min}}} \max\left\{ \frac{(2M)^{1/3} n^{1/2} b^{7/6}(t) (Lp_{\min})^{1/6}}{\gamma_n^{1/3}(t) (\log n)^{1/6} T^{1/3}}, \frac{1}{\sqrt{T}} \right\} \Bigg]. \end{aligned}$$

Note that when $M > 0$ and $b_{\max} = O(1)$, which implies that $b(t) = O(1)$ and $\gamma_n(t) = O(1)$ (see Remark 4), then Corollary 2 asserts that for $\delta = \Theta(T^{2/3})$, if n, T are large enough, then w.h.p $\|\hat{\pi}(t) - \pi^*(t)\|_\infty = O(T^{-1/3})$. This matches the rate for the pointwise risk for estimating univariate Lipschitz functions (see for e.g. [37, Theorem 1.3.1]).

2.2.3 *Different construction of* $G_\delta(t)$

As noted in Remark 3, the condition $p_{\delta,sum}(t) \gtrsim \log n$ appearing in the results of Sections 2.2.1, 2.2.2 is quite strict as it implicitly imposes that the union graph $G_\delta(t)$ is complete. Indeed, this condition comes from our need to bound the quantities $|\mathcal{N}_{ij,\delta}(t)|$ with high probability (see Lemma 21). It appears to be difficult to get meaningful concentration bounds on $N_{\min,\delta}(t), N_{\max,\delta}(t)$ in the sparser regime where $\delta p_{\min} = o(\log n)$.

One way to avoid these difficulties is to construct a graph such that $|\mathcal{N}_{ij,\delta}(t)|$ is already controlled for any $i \neq j$. To this end, we will consider the graph $\tilde{G}_\delta(t) = G([n], \tilde{\mathcal{E}}_\delta(t))$ in Algorithm 1 where

$$\tilde{\mathcal{E}}_\delta(t) := \left\{\{i,j\} : |\mathcal{N}_{ij,\delta}(t)| \in \left[\max\left(1, \frac{p_{\delta,sum}(t)}{2}\right), \max\left(2p_{\delta,sum}(t), 6\log\frac{4}{\delta p_{\min}}\right)\right]\right\}. \tag{2.2.8}$$

Lemma 1. *Suppose that there exist* $p_{\min} > 0$ *and* $p_{\max} < \frac{1}{2}$ *such that for any* $t' \in \mathcal{T}$, $p_{\min} \leqslant p(t') \leqslant p_{\max}$. *Let* $\tilde{G}_\delta(t) = G([n], \tilde{\mathcal{E}}_\delta(t))$ *be constructed as in* (2.2.8). *Then* $\tilde{G}_\delta(t) \sim G(n, \tilde{p}_\delta(t))$, *and the following is true.*

1. *If* $\delta p_{\min} \geqslant 3$, *then* $\tilde{p}_\delta(t) \geqslant 1 - 2e^{-\frac{1}{4}}$ *(hence* $\tilde{G}_\delta(t)$ *is dense) and*
$$\tilde{\mathcal{E}}_\delta(t) := \left\{\{i,j\} : |\mathcal{N}_{ij,\delta}(t)| \in \left[\frac{p_{\delta,sum}(t)}{2}, 2p_{\delta,sum}(t)\right]\right\}.$$
2. *If* $\delta p_{\max} \leqslant \frac{1}{8}$, *then* $\tilde{p}_\delta(t) \in \left[\frac{\delta p_{\min}}{4}, 8\delta p_{\max}\right]$ *and*
$$\tilde{\mathcal{E}}_\delta(t) := \left\{\{i,j\} : |\mathcal{N}_{ij,\delta}(t)| \in \left[1, 6\log\frac{4}{\delta p_{\min}}\right]\right\}.$$

The proof of Lemma 1 is outlined in Section 2.10. We now show how this can be used to obtain error bounds in the regimes where $\delta p_{\min} = o(\log n)$. For simplicity, we will only show this for the ℓ_2 error, but the idea can be used in an analogous manner to obtain ℓ_∞ bounds as well[1].

As for the simple union graph $G_\delta(t)$, one can introduce analoguous notations for some specific quantities related to the graph $\tilde{G}_\delta(t)$. Let us denote $\tilde{d}_{\min,\delta}(t)$ (resp. $\tilde{d}_{\max,\delta}(t)$) to be the minimal (resp. maximal) vertex degree in $\tilde{G}_\delta(t)$, and $\tilde{\xi}_\delta(t)$ to be the spectral gap of the random walk Laplacian of $\tilde{G}_\delta(t)$. Also, let

$$\tilde{N}_{\min,\delta}(t) := \min_{\{i,j\}\in\tilde{\mathcal{E}}_\delta(t)} |\mathcal{N}_{ij,\delta}(t)| \quad \text{and} \quad \tilde{N}_{\max,\delta}(t) := \max_{\{i,j\}\in\tilde{\mathcal{E}}_\delta(t)} |\mathcal{N}_{ij,\delta}(t)|.$$

1 This is omitted due to space considerations.

Then choosing $\tilde{d}_\delta(t) = 3n\tilde{p}_\delta(t)$, one can employ Algorithm 1 using the data gathered in $\tilde{G}_\delta(t)$, and consequently derive error bounds for the cases $\delta p_{\min} \gtrsim 1$ (dense regime) and $\delta p_{\max} \lesssim 1$ (sparse regime). Note that if $\tilde{\mathcal{E}}_\delta(t)$ is non-empty – which will happen w.h.p – it implies lower (resp. upper) bounds on $\tilde{N}_{\min,\delta}(t)$ (resp. $\tilde{N}_{\max,\delta}(t)$) from (2.2.8).

Remark 5. *To construct $\tilde{G}_\delta(t)$ in practice, we would need to estimate $p(t')$ for each $t' \in \mathcal{T}$ in order to have estimates for $p_{\delta,sum}(t)$ and $p_{\min}$. This was not the case when working with the union graph $G_\delta(t)$.*

Theorem 4 (Dense regime)**.** *Suppose that $n \gtrsim \log n$ and $G_{t'} \sim G(n, p(t'))$ for all $t' \in \mathcal{T}$ so that $\tilde{G}_\delta(t) \sim G(n, \tilde{p}_\delta(t))$. Choosing $\tilde{d}_\delta(t) = 3n\tilde{p}_\delta(t)$ in Algorithm 1 with the graph $\tilde{G}_\delta(t)$ and $\delta = \min\left\{T, \frac{T^{2/3} b^{2/3}(t)}{n(2M)^{2/3}(Lp_{\min})^{1/3}}\right\}$, if T is such that $\delta p_{\min} \geqslant 3$ and*

$$\frac{1}{b^{7/2}(t)} \gtrsim \sqrt{\frac{\log n}{Lnp_{\min}}} \max\left\{\frac{(2M)^{1/3} n^{1/2}(Lp_{\min})^{1/6}}{b^{1/3}(t)T^{1/3}}, \frac{1}{\sqrt{T}}\right\} + Mn \min\left\{\frac{b^{2/3}(t)}{(2M)^{2/3} n (Lp_{\min})^{1/3} T^{1/3}}, 1\right\},$$

then with probability at least $1 - O(n^{-10})$,

$$\frac{\|\hat{\pi}(t) - \pi^*(t)\|_2}{\|\pi^*(t)\|_2} \lesssim Mnb^{7/2}(t) \min\left\{\frac{b^{2/3}(t)}{(2M)^{2/3} n (Lp_{\min})^{1/3} T^{1/3}}, 1\right\} + \frac{b^{9/2}(t)}{\sqrt{Lnp_{\min}}} \max\left\{\frac{(2M)^{1/3} n^{1/2}(Lp_{\min})^{1/6}}{b^{1/3}(t)T^{1/3}}, \frac{1}{\sqrt{T}}\right\}.$$

Proof. The proof of this theorem is analoguous to the proof of Corollary 1. We simply use Lemma 1 (first part) and Lemma 21 for the new union graph $\tilde{G}_\delta(t)$ in order to bound $\tilde{d}_{\min,\delta}(t)$, $\tilde{d}_{\max,\delta}(t)$, $\tilde{\xi}_\delta(t)$, $\tilde{N}_{\min,\delta}(t)$, $\tilde{N}_{\max,\delta}(t)$ and $|\tilde{\mathcal{E}}_\delta(t)|$, w.h.p. □

Observe that Theorem 4 gives the same error bound as in Corollary 1, but with the relatively milder condition $\delta p_{\min} \gtrsim 1$. From Lemma 1, note that this means that $\tilde{G}_\delta(t)$ is still dense. The following theorem provides an error bound in a sparser regime where $\frac{\log n}{np_{\min}} \lesssim \delta \lesssim \frac{1}{p_{\max}}$.

Theorem 5 (Sparse regime)**.** *Suppose that $n \gtrsim \log n$ and $G_{t'} \sim G(n, p(t'))$ for all $t' \in \mathcal{T}$ so that $\tilde{G}_\delta(t) \sim G(n, \tilde{p}_\delta(t))$. Choosing $\tilde{d}_\delta(t) = 3n\tilde{p}_\delta(t)$ in*

Algorithm 1 with the graph $\tilde{G}_\delta(t)$ *and* $\delta = \min\left\{T, \frac{T^{2/3} b^{2/3}(t)(\log n)^{1/3}}{n(2M)^{2/3}(Lp_{\min})^{1/3}}\right\}$, *if* T *is such that* $\frac{\log n}{np_{\min}} \lesssim \delta \lesssim \frac{1}{p_{\max}}$ *and*

$$\frac{1}{b^{7/2}(t)} \gtrsim \frac{\log n}{\sqrt{Lnp_{\min}}} \max\left\{\frac{(2M)^{1/3} n^{1/2} (Lp_{\min})^{1/6}}{b^{1/3}(t) T^{1/3} (\log n)^{1/6}}, \frac{1}{\sqrt{T}}\right\} \tag{2.2.9}$$

$$+ Mn \min\left\{\frac{b^{2/3}(t)(\log n)^{1/3}}{(2M)^{2/3} n (Lp_{\min})^{1/3} T^{1/3}}, 1\right\},$$

then with probability at least $1 - O(n^{-10})$,

$$\frac{\|\hat{\pi}(t) - \pi^*(t)\|_2}{\|\pi^*(t)\|_2} \lesssim Mnb^{7/2}(t) \min\left\{\frac{b^{2/3}(t)(\log n)^{1/3}}{(2M)^{2/3} n (Lp_{\min})^{1/3} T^{1/3}}, 1\right\}$$
$$+ b^{9/2}(t) \sqrt{\frac{\log n}{Lnp_{\min}}} \max\left\{\frac{(2M)^{1/3} n^{1/2} (Lp_{\min})^{1/6}}{b^{1/3}(t) T^{1/3} (\log n)^{1/6}}, \frac{1}{\sqrt{T}}\right\}.$$

Proof. Similar proof technique as Corollary 1 using Lemmas 1 and 21 for $\tilde{G}_\delta(t)$. □

The above error bound has an extra $(\log n)^{1/3}$ factor as compared to that of Corollary 1, however the dependence on T is unchanged. The condition $\frac{\log n}{np_{\min}} \lesssim \delta \lesssim \frac{1}{p_{\max}}$ implies both a lower and an upper bound on T, and is of course feasible provided $p_{\min}$ and $p_{\max}$ are of the same order. Let us instantiate Theorems 4 and 5 on an example to see the conditions therein more clearly.

Example 1. *Let us choose* $L = 1$, $b(t) \asymp 1$, $M \asymp 1$ *and* $p_{\min}, p_{\max} \asymp \frac{1}{n}$.

1. *(Theorem 5) For* $\delta = \frac{T^{2/3}(\log n)^{1/3}}{n^{2/3}}$, *if* $n(\log n)^{5/2} \lesssim T \lesssim \sqrt{\frac{n^5}{\log n}}$, *then it holds with high probability that*

$$\frac{\|\hat{\pi}(t) - \pi^*(t)\|_2}{\|\pi^*(t)\|_2} \lesssim \left(\frac{n \log n}{T}\right)^{1/3}.$$

2. *(Theorem 4) For* $\delta = (\frac{T}{n})^{2/3}$, *if* $T \gtrsim n^{5/2}$, *then it holds with high probability that*

$$\frac{\|\hat{\pi}(t) - \pi^*(t)\|_2}{\|\pi^*(t)\|_2} \lesssim \left(\frac{n}{T}\right)^{1/3}.$$

2.3 ℓ_2-ANALYSIS OF THE SPECTRAL ESTIMATOR

We now describe the main ideas that lead to the ℓ_2 bound in Theorem 1. We will essentially proceed in three steps following the ideas in [9].

$$\|\hat{\pi}(t) - \pi^*(t)\|_2 \overset{(i)}{\leqslant} \frac{1}{\sqrt{\pi^*_{\min}(t)}} \|\hat{\pi}(t) - \pi^*(t)\|_{\pi^*(t)}$$
$$\overset{(ii)}{\leqslant} \frac{8d_\delta(t)b^{7/2}(t)}{\xi_\delta(t)d_{\min,\delta}(t)} \left\|\pi^*(t)^\top(\hat{P}(t) - \bar{P}(t))\right\|_2 \tag{2.3.1}$$
$$\overset{(iii)}{\leqslant} \frac{8d_\delta(t)b^{7/2}(t)}{\xi_\delta(t)d_{\min,\delta}(t)} \tag{2.3.2}$$
$$\times \left(4\frac{M\delta|\mathcal{E}_\delta(t)|}{Td_{\max,\delta}(t)} + \tilde{C_2}\sqrt{\frac{N_{\max,\delta}(t)d_{\max,\delta}(t)b^2(t)}{Ld^2_\delta(t)N^2_{\min,\delta}(t)}}\right) \|\pi^*(t)\|_2 \,. \tag{2.3.3}$$

(i) The first step is easy to verify, due to the definition of the norm $\|\cdot\|_{\pi^*(t)}$

$$\|\pi^*(t) - \hat{\pi}(t)\|^2_{\pi^*(t)} = \sum_{i=1}^{n} \pi^*_i(t)(\pi^*_i(t) - \hat{\pi}_i(t))^2 \geqslant \pi^*_{\min}(t)\,\|\pi^*(t) - \hat{\pi}(t)\|^2_2 \,.$$

(ii) This is given by the combination of Lemmas 2, 3 and 4 which in turn are derived using [9, Theorem 8] and [36, Lemma 6].

(iii) For this step, we can decompose $\hat{P}(t) - \bar{P}(t) = \Delta(t) + \Delta_1(t)$ as in (2.3.6). Then one has to bound $\left\|\pi^*(t)^\top\Delta(t)\right\|_2$ and $\left\|\pi^*(t)^\top\Delta_1(t)\right\|_2$. The second term is completely deterministic and can be bounded using Assumption 1. A bound on $\left\|\pi^*(t)^\top\Delta(t)\right\|_2$ is found following the same steps as in the proof of [8, Theorem 9].

2.3.1 *Proof of Theorem 1*

A bound on $\|\pi^*(t) - \hat{\pi}(t)\|_{\pi^*(t)}$ is given by [9, Theorem 8], which is recalled in Section 4.1.1. Denoting $\lambda_{\max}(\bar{P}(t))$ to be the second largest eigenvalue of $\bar{P}(t)$ in absolute value, i.e.,

$$\lambda_{\max}(\bar{P}(t)) = \max\left\{\lambda_2(\bar{P}(t)), -\lambda_n(\bar{P}(t))\right\},$$

this theorem gives the bound

$$\|\hat{\pi}(t) - \pi^*(t)\|_{\pi^*(t)} \leqslant \frac{\left\|\pi^*(t)^\top(\hat{P}(t) - \bar{P}(t))\right\|_2}{1 - \lambda_{\max}(\bar{P}(t)) - \left\|\hat{P}(t) - \bar{P}(t)\right\|_{\pi^*(t)}} \tag{2.3.4}$$

provided that the following condition holds.

$$\left\|\hat{P}(t) - \bar{P}(t)\right\|_{\pi^*(t)} < 1 - \lambda_{\max}(\bar{P}(t)). \tag{2.3.5}$$

First let us note that these eigenvalues are real, and so (2.3.5) is well defined. Indeed, denoting $\Pi^*(t) = \operatorname{diag}(\pi^*(t))$ and $S = \Pi^*(t)^{1/2}\bar{P}(t)\Pi^*(t)^{-1/2}$, S is similar to $\bar{P}(t)$, and S is symmetric due to the reversibility of $\bar{P}(t)$.

To prove (2.3.5), we will use results similar to [36, Lemma's 3, 4]. The main idea is to write the following decomposition

$$\hat{P}(t) = \bar{P}(t) + \underbrace{\hat{P}(t) - \hat{P}^*(t)}_{\Delta(t)} + \underbrace{\hat{P}^*(t) - \bar{P}(t)}_{\Delta_1(t)} \tag{2.3.6}$$

where $\hat{P}^*(t) = \mathbb{E}\hat{P}(t)$ whose entries are given by

$$\hat{P}^*_{ij}(t) = \begin{cases} \dfrac{1}{d_\delta(t)}\left(\dfrac{1}{|\mathcal{N}_{ij,\delta}(t)|}\displaystyle\sum_{t'\in\mathcal{N}_{ij,\delta}(t)}\dfrac{w^*_{t',j}}{w^*_{t',i}+w^*_{t',j}}\right) & \text{if} \quad \{i,j\}\in\mathcal{E}_\delta(t) \\ 1 - \dfrac{1}{d_\delta(t)}\displaystyle\sum_{k\neq i}\dfrac{1}{|\mathcal{N}_{ik,\delta}(t)|}\sum_{t'\in\mathcal{N}_{ik,\delta}(t)}\dfrac{w^*_{t',k}}{w^*_{t',i}+w^*_{t',k}} & \text{if} \quad i=j \\ 0 & \text{otherwise.} \end{cases}$$

We now provide bounds on $\|\Delta_1(t)\|_2$ and $\|\Delta(t)\|_2$ in Lemma's 2 and 3 respectively. The proofs of all results in this section are outlined in Section 2.8.

Lemma 2 (Bound on $\|\Delta_1(t)\|_2$)**.** *It holds that* $\|\Delta_1(t)\|_2 \leqslant 4\frac{M\delta|\mathcal{E}_\delta(t)|}{T d_{\max,\delta}(t)}$.

Lemma 3 (Bound on $\|\Delta(t)\|_2$)**.** *There exists a constant* $\tilde{C}_1 \geqslant 15$ *such that*

$$\|\Delta(t)\|_2 \leqslant \tilde{C}_1\sqrt{\frac{N_{\max,\delta}(t)d_{\max,\delta}(t)\log n}{L d^2_\delta(t) N^2_{\min,\delta}(t)}}$$

with probability at least $1 - O(n^{-10})$.

The proof of Lemma 2 follows from the smoothness condition in Assumption 1, while the proof of Lemma 3 follows the proof steps of [36, Lemma 3]. Next, we show that if $\xi_\delta(t) > 0$ (which implies that $G_\delta(t)$ is connected) and if the perturbation $\left\|\hat{P}(t) - \bar{P}(t)\right\|_2$ is sufficiently small, then we can ensure (2.3.5).

Lemma 4. *Denoting* $\rho(t) = \lambda_{\max}(\bar{P}(t)) + \left\|\hat{P}(t) - \bar{P}(t)\right\|_2\sqrt{\frac{\pi^*_{\max}(t)}{\pi^*_{\min}(t)}}$, *recall that* $\xi_\delta(t) = \lambda_{max}(\mathcal{L}_\delta(t))$ *(where* $\mathcal{L}_\delta(t)$ *is the Laplacian of* $G_\delta(t)$*) and* $b(t) := \max_{i,j}\frac{w^*_{t,i}}{w^*_{t,j}} = \frac{\pi^*_{\max}(t)}{\pi^*_{\min}(t)}$. *If* $\xi_\delta(t) > 0$ *and* $d_\delta(t) \geqslant d_{\max,\delta}(t)$, *then we have that* $1 - \lambda_{\max}(\bar{P}(t)) \geqslant \frac{\xi_\delta(t)d_{\min,\delta}(t)}{4d_\delta(t)b^3(t)}$. *Moreover, if*

$$\left\|\hat{P}(t) - \bar{P}(t)\right\|_2 \leqslant \frac{\xi_\delta(t)d_{\min,\delta}(t)}{8d_\delta(t)b^{7/2}(t)} \tag{2.3.7}$$

then it holds that

$$1 - \rho(t) \geqslant \frac{\xi_\delta(t) d_{\min,\delta}(t)}{8 d_\delta(t) b^3(t)} > 0.$$

The statement is analogous to that of [36, Lemma 4]. The bound on $1 - \lambda_{\max}(\bar{P}(t))$ is clearly the crucial statement, and requires using [36, Lemma 6]. We remark in passing that the dependence on $b(t)$ is $b^3(t)$ in Lemma 4, we could not verify the dependence stated in [36, Lemma 6] (which is $b^2(t)$). For completeness, we outline the proof of Lemma 4 in Section 2.8.

Condition (2.3.7) is ensured via Lemma's 2 and 3 (with high probability) whenever (2.2.1) holds. Then, (2.1.1) readily implies that

$$1 - \lambda_{\max}(\bar{P}(t)) - \left\|\hat{P}(t) - \bar{P}(t)\right\|_{\pi^*(t)} \geqslant 1 - \rho(t) \geqslant \frac{\xi_\delta(t) d_{\min,\delta}(t)}{8 d_\delta(t) b^3(t)} > 0, \tag{2.3.8}$$

thus ensuring (2.3.5). Using (2.3.4) and (2.3.8) we finally obtain (2.3.1) as follows.

$$\begin{aligned} \|\hat{\pi}(t) - \pi^*(t)\|_{\pi^*(t)} &\leqslant \frac{8 d_\delta(t) b^3(t)}{\xi_\delta(t) d_{\min,\delta}(t)} \left\|\pi^*(t)^\top (\hat{P}(t) - \bar{P}(t))\right\|_{\pi^*(t)} \\ &\leqslant \frac{8 d_\delta(t) b^{7/2}(t)}{\xi_\delta(t) d_{\min,\delta}(t)} \left\|\pi^*(t)^\top (\hat{P}(t) - \bar{P}(t))\right\|_2 \end{aligned} \tag{2.3.9}$$

where the last inequality uses (2.1.1).

Finally, we can bound $\left\|\pi^*(t)^\top (\hat{P}(t) - \bar{P}(t))\right\|_2$ using the decomposition in (2.3.6) along with the triangular inequality, leading to

$$\left\|\pi^*(t)^\top (\hat{P}(t) - \bar{P}(t))\right\|_2 \leqslant \left\|\pi^*(t)^\top \Delta(t)\right\|_2 + \left\|\pi^*(t)^\top \Delta_1(t)\right\|_2.$$

Bounds on $\left\|\pi^*(t)^\top \Delta(t)\right\|_2$ and $\left\|\pi^*(t)^\top \Delta_1(t)\right\|_2$ are provided in the following lemma's.

Lemma 5.

$$\left\|\pi^*(t)^\top \Delta_1(t)\right\|_2 \leqslant \|\pi^*(t)\|_2 \|\Delta_1(t)\|_2 \leqslant 4 \frac{M\delta|\mathcal{E}_\delta(t)|}{T d_{\max,\delta}(t)} \|\pi^*(t)\|_2. \tag{2.3.10}$$

The statement follows directly from Lemma 2.

Lemma 6. *There exist constants $c_1 > 0, \tilde{C_2} \geqslant 1$ such that if $n \geqslant c_1 \log n$ then with probability at least $1 - O(n^{-10})$, we have that*

$$\left\|\pi^*(t)^\top \Delta(t)\right\|_2 \leqslant \tilde{C_2} \sqrt{\frac{N_{\max,\delta}(t) d_{\max,\delta}(t) b^2(t)}{L d_\delta^2(t) N_{\min,\delta}^2(t)}} \|\pi^*(t)\|_2.$$

The proof of Lemma 6 follows the ideas in the proof of [9, Theorem 9]. Applying these bounds in (2.3.9) finally leads to the stated bound in Theorem 1.

2.3.2 *Proof of Theorem 2*

This theorem follows directly from Theorem 1 and from the propreties of Erdös-Renyi graphs gathered in Lemma 21. Using (2.2.4) and Lemma 21 along with the choice $d_\delta(t) = 3np_\delta(t)$, it holds with probability at least $1 - O(n^{-10})$ that

$$\begin{aligned}\tilde{C_1}\sqrt{\frac{N_{\max,\delta}(t)d_{\max,\delta}(t)\log n}{Ld_\delta^2(t)N^2_{\min,\delta}(t)}} + 4\frac{M\delta|\mathcal{E}_\delta(t)|}{Td_{\max,\delta}(t)} &\leqslant 2\tilde{C_1}\sqrt{\frac{\log n}{Lnp_\delta(t)p_{\delta,sum}(t)}} + 16\frac{M\delta n}{T} \\ &\leqslant \frac{1}{96b^{\frac{7}{2}}(t)} \\ &\leqslant \frac{\xi_\delta(t)d_{\min,\delta}(t)}{8d_\delta(t)b^{\frac{7}{2}}(t)}.\end{aligned}$$

Hence, condition (2.2.1) is satisfied with high probability and Theorem 1 implies that

$$\frac{\|\hat{\pi}(t) - \pi^*(t)\|_2}{\|\pi^*(t)\|_2} \leqslant 32\frac{M\delta|\mathcal{E}_\delta(t)|d_\delta(t)b^{7/2}(t)}{T\xi_\delta(t)d_{\min,\delta}(t)d_{\max,\delta}(t)} + 8\tilde{C_2}\frac{b^{9/2}(t)}{\xi_\delta(t)d_{\min,\delta}(t)}\sqrt{\frac{N_{\max,\delta}(t)d_{\max,\delta}}{LN^2_{\min,\delta}(t)}}$$

Again, using Lemma 21, we can simplify the above bound so that with probability at least $1 - O(n^{-10})$,

$$\begin{aligned}\frac{\|\hat{\pi}(t) - \pi^*(t)\|_2}{\|\pi^*(t)\|_2} &\leqslant 1536\frac{M\delta n^2p_\delta(t)b^{7/2}(t)}{Tnp_\delta(t)} + 64\tilde{C_2}\frac{b^{9/2}(t)}{np_\delta(t)}\sqrt{\frac{3p_{\delta,sum}(t)np_\delta(t)}{Lp^2_{\delta,sum}(t)}} \\ &\leqslant 1536\frac{M\delta n b^{7/2}(t)}{T} + 64\tilde{C_2}b^{9/2}(t)\sqrt{\frac{3}{Lnp_\delta(t)p_{\delta,sum}(t)}}.\end{aligned}$$

2.3.3 *Proof of Corollary 1*

Since $p_{\delta,sum}(t) \geqslant p_{\min}\delta$, therefore the condition $\delta \gtrsim \frac{\log n}{p_{\min}}$ implies $p_{\delta,sum}(t) \gtrsim \log n$, as well as $p_\delta(t) \gtrsim 1$ (due to Proposition 4), thus satisfying the requirements of Theorem 2. Additionally, δ is required to satisfy $\delta \leqslant T$, and also condition (2.2.4), i.e.,

$$\sqrt{\frac{\log n}{Ln\delta p_{\min}}} + \frac{M\delta n}{T} \lesssim \frac{1}{b^{\frac{7}{2}}(t)}. \tag{2.3.11}$$

If δ satisfies the three aforementioned conditions, and if $n \gtrsim \log n$, we have with probability at least $1 - O(n^{-10})$ the ℓ_2 bound

$$\frac{\|\hat{\pi}(t) - \pi^*(t)\|_2}{\|\pi^*(t)\|_2} \lesssim \frac{M\delta n b^{\frac{7}{2}}(t)}{T} + \frac{b^{\frac{9}{2}}(t)}{\sqrt{Ln\delta p_{\min}}}. \tag{2.3.12}$$

The optimal choice of $\delta \in (0, T]$ that minimizes the RHS of (2.3.12) is easily verified to be

$$\delta^* = \min\left\{\frac{(b(t)T)^{\frac{2}{3}}}{(2M)^{\frac{2}{3}} n (Lp_{\min})^{\frac{1}{3}}}, T\right\}.$$

Now it remains to ensure that δ^* satisfies the previously stated conditions on δ. Clearly $\delta^* \leqslant T$, and condition (2.3.11) are equivalent to the stated conditions on T in the corollary. Hence provided $\delta \asymp \delta^*$, T satisfies the stated conditions, and $n \gtrsim \log n$, we arrive at the stated ℓ_2 bound in the corollary.

2.4 ℓ_∞-ANALYSIS OF THE SPECTRAL ESTIMATOR

The main goal of this section is to present the steps of the proof of Theorem 3. We will follow the steps of the analysis carried out by Chen et al. [9], and adapt it to our setting. The proofs of all results from this section are provided in Section 2.9.

As $\hat{\pi}(t)$ and $\pi^*(t)$ are stationary distributions of the transition matrices $\hat{P}(t)$ and $\bar{P}(t)$, then denoting $P_{.m}$ to be the m^{th} column of a matrix P, it holds for all $m \in [n]$ that,

$$\begin{aligned}
\hat{\pi}_m(t) - \pi^*_m(t) &= \left(\hat{\pi}(t)^\top \hat{P}(t)\right)_m - \left(\pi^*(t)^\top \bar{P}(t)\right)_m \\
&= \hat{\pi}(t)^\top \hat{P}_{.m}(t) - \pi^*(t)^\top \bar{P}_{.m}(t) \\
&= \pi^*(t)^\top (\hat{P}_{.m}(t) - \bar{P}_{.m}(t)) + (\hat{\pi}(t) - \pi^*(t))^\top \hat{P}_{.m}(t) \\
&= \underbrace{\pi^*(t)^\top (\hat{P}_{.m}(t) - \hat{P}^*_{.m}(t))}_{:=I_0^m} + \underbrace{\pi^*(t)^\top (\hat{P}^*_{.m}(t) - \bar{P}_{.m}(t))}_{:=I_1^m} + \underbrace{(\hat{\pi}_m(t) - \pi^*_m(t))\hat{P}_{mm}(t)}_{:=I_2^m} \\
&+ \sum_{j:j\neq m} (\hat{\pi}_j(t) - \pi^*_j(t))\hat{P}_{jm}(t).
\end{aligned}$$

Let us first focus on bounding $|I_0^m|, |I_1^m|$ and $|I_2^m|$. The last term will be treated carefully due to the occurrence of some statistical dependencies therein. Since $\hat{P}^*(t) = \mathbb{E}[\hat{P}(t)]$, hence $|I_0^m|$ can be bounded using Hoeffding's inequality.

Lemma 7. *Suppose that* $np_\delta(t) \geqslant c_0 \log n$ *and* $p_{\delta,sum}(t) \geqslant c_2 \log n$ *for constants* $c_0, c_2 \geqslant 1$ *from Lemma 21. Then there exists a constant* $C_1 \geqslant 1$ *such that with probability at least* $1 - O(n^{-9})$*,*

$$|I_0^m| \leqslant C_1 \sqrt{\frac{\log n}{Lnp_\delta(t)p_{\delta,sum}(t)}} \|\pi^*(t)\|_\infty \quad \forall m \in [n]. \tag{2.4.1}$$

The second term I_1^m can be bounded easily using Assumption 1.

Lemma 8. *With probability at least* $1 - O(n^{-10})$*, we have*

$$|I_1^m| \leqslant \frac{2M\delta}{T} \|\pi^*(t)\|_\infty \quad \forall m \in [n]. \tag{2.4.2}$$

The next lemma shows that with high probability, $|I_2^m| \lesssim \|\hat{\pi}(t) - \pi^*(t)\|_\infty$ for all $m \in [n]$.

Lemma 9. *Recall that* $b_{\max,\delta}(t) = \max_{t \in \mathcal{N}_\delta(t)} b(t)$ *and suppose that* $np_\delta(t) \geqslant c_0 \log n$ *and* $p_{\delta,sum}(t) \geqslant c_2 \log n$ *for constants* $c_0, c_2 \geqslant 1$ *from Lemma 21. Then there exists a constant* $C_2 \geqslant 1$ *such that it holds with probability at least* $1 - O(n^{-9})$ *that*

$$|I_2^m| \leqslant \left(1 - \frac{1}{12b_{\max,\delta}(t)} + C_2\sqrt{\frac{\log n}{Lnp_\delta(t)p_{\delta,sum}(t)}}\right) \|\hat{\pi}(t) - \pi^*(t)\|_\infty \quad \forall m \in [n]. \tag{2.4.3}$$

The last term to bound is more difficult to handle due to the statistical dependency between $\hat{\pi}(t)$ and $\hat{P}(t)$. The idea is then to use the same "leave-one-out" trick as in [9] and introduce a new matrix $\hat{P}^{(m)}(t)$ with entries given by (for all $i \neq j$)

$$\hat{P}_{ij}^{(m)}(t) = \begin{cases} \hat{P}_{ij}(t) & \text{if } i \neq m, j \neq m \\ \frac{p_\delta(t)}{d_\delta(t)} \frac{w_{t,j}^*}{w_{t,j}^* + w_{t,i}^*} & \text{if } i = m \text{ or } j = m. \end{cases}$$

To ensure that $\hat{P}^{(m)}(t)$ is a transition matrix, its diagonal entries are defined as

$$\hat{P}_{ii}^{(m)}(t) = 1 - \sum_{k \neq i} \hat{P}_{ik}^{(m)}(t).$$

Here, the m^{th} line and column of $\hat{P}(t)$ have been replaced by their expectation, unconditionally of the union graph $G_\delta(t)$. Let us denote $\hat{\pi}^{(m)}(t)$ to be the leading left eigenvector of $\hat{P}^{(m)}(t)$. This vector is now statistically independent of the connectivity and the pairwise comparison outputs involving the m^{th} item, and one can reasonably expect that it is close to $\pi^*(t)$. Hence, we can decompose the last term as

$$\sum_{j:j\neq m} \left(\hat{\pi}_j(t) - \pi_j^*(t)\right) \hat{P}_{jm}(t) = \underbrace{\sum_{j:j\neq m} \left(\hat{\pi}_j(t) - \hat{\pi}_j^{(m)}(t)\right) \hat{P}_{jm}(t)}_{:=I_3^m} + \underbrace{\sum_{j:j\neq m} \left(\hat{\pi}_j^{(m)}(t) - \pi_j^*(t)\right) \hat{P}_{jm}(t)}_{:=I_4^m}.$$

To bound $\left|I_3^m\right|$, note that the Cauchy-Schwarz inequality implies

$$|I_3^m| \leqslant \sqrt{\sum_{j:j\neq m} \hat{P}_{jm}(t)^2}\,\left\|\hat{\pi}(t) - \hat{\pi}^{(m)}(t)\right\|_2.$$

Since for all $j \neq m$, $\hat{P}_{jm}(t) \leqslant \frac{1}{d_\delta(t)}$, hence $\left|I_3^m\right| \leqslant \frac{\sqrt{d_{\max,\delta}(t)}}{d_\delta(t)} \left\|\hat{\pi}(t) - \hat{\pi}^{(m)}(t)\right\|_2$. The important step now is to bound $\left\|\hat{\pi}(t) - \hat{\pi}^{(m)}(t)\right\|_2$ since $d_{\max,\delta}(t) \leqslant d_\delta(t)$ w.h.p (when $G_{t'} \sim \mathcal{G}(n, p(t'))$ for all $t' \in \mathcal{N}_\delta(t)$) due to Lemma 21. This is shown in the following lemma.

Lemma 10. *Suppose that* $np_\delta(t) \geqslant c_0 \log n$ *and* $p_{\delta,sum}(t) \geqslant c_2 \log n$ *for constants* $c_0, c_2 \geqslant 1$ *from Lemma 21. Then there exist constants* $C_3, C_6 \geqslant 1$ *such that if*

$$96 b^{\frac{5}{2}}(t)\left(\frac{4M\delta}{T} + C_3\sqrt{\frac{\log n}{np_\delta(t)}}\right) \leqslant \frac{1}{2}, \tag{2.4.4}$$

then it holds with probability at least $1 - O(n^{-9})$ *that for all* $m \in [n]$,

$$\left\|\hat{\pi}^{(m)}(t) - \hat{\pi}(t)\right\|_2 \leqslant 192 b^{\frac{5}{2}}(t)\left(\frac{4M\delta}{T} + C_3\sqrt{\frac{\log n}{Lnp_\delta(t)p_{\delta,sum}(t)}}\right)\|\pi^*(t)\|_\infty + \|\hat{\pi}(t) - \pi^*(t)\|_\infty.$$

Consequently, we have with probability at least $1 - O(n^{-9})$ *that for all* $m \in [n]$,

$$|I_3^m| \leqslant 192\frac{b^{\frac{5}{2}}(t)}{\sqrt{3np_\delta(t)}}\left(\frac{4M\delta}{T} + C_3\sqrt{\frac{\log n}{Lnp_\delta(t)p_{\delta,sum}(t)}}\right)\|\pi^*(t)\|_\infty + \frac{1}{\sqrt{3np_\delta(t)}}\|\hat{\pi}(t) - \pi^*(t)\|_\infty. \tag{2.4.5}$$

Finally, we can bound $\left|I_4^m\right|$ using the statistical independence between $\hat{\pi}^{(m)}(t)$ and $\hat{P}_{\cdot m}(t), \hat{P}_{m\cdot}(t)$.

Lemma 11. *Suppose that* $np_\delta(t) \geqslant c_0 \log n$ *and* $p_{\delta,sum}(t) \geqslant c_2 \log n$ *for constants* $c_0, c_2 \geqslant 1$ *from Lemma 21;* $n \geqslant c_1 \log n$ *for the constant* $c_1 > 0$ *in Theorem 1, and that condition (2.2.4) of Theorem 2 holds. Then there exist* *constants* $C_7, C_8, C_9 \geqslant 1$ *such that with probability at least* $1 - O(n^{-9})$, *we have for all* $m \in [n]$,

$$|I_4^m| \leqslant \left(C_7 \frac{Mn\delta b^{\frac{7}{2}}(t)}{T} + C_8 \frac{b^{\frac{5}{2}}(t) \max\left\{ b^2(t), \frac{\log n}{\sqrt{np_\delta(t)}} \right\}}{\sqrt{Lnp_\delta(t) p_{\delta,sum}(t)}} \right) \|\pi^*(t)\|_\infty + C_9 \sqrt{\frac{\log n}{np_\delta(t)}} \|\hat{\pi}(t) - \pi^*(t)\|_\infty .$$

2.4.1 *Proof of Theorem 3*

As seen in Section 2.4, the bound in Theorem 3 follows from the combination of the bounds on $I_0^m, I_1^m, I_2^m, I_3^m$, and I_4^m. Note that we can identify two types of terms in these bounds – those depending on $\|\pi^*(t)\|_\infty$ and the ones which depend on $\|\hat{\pi}(t) - \pi^*(t)\|_\infty$. Then, our bound can be written as

$$\|\hat{\pi}(t) - \pi^*(t)\|_\infty \leqslant \alpha \|\pi^*(t)\|_\infty + \beta \|\hat{\pi}(t) - \pi^*(t)\|_\infty$$

where $\beta < 1$, which in turn implies $\frac{\|\hat{\pi}(t) - \pi^*(t)\|_\infty}{\|\pi^*(t)\|_\infty} \leqslant \frac{\alpha}{1-\beta}$.

We can lower bound the term $1 - \beta$ as

$$\begin{aligned}
1 - \beta =& 1 - \left(1 - \frac{1}{12 b_{\max,\delta}(t)} + C_2 \sqrt{\frac{\log n}{Lnp_\delta(t) p_{\delta,sum}(t)}} + \frac{1}{\sqrt{np_\delta(t)}} + C_9 \sqrt{\frac{\log n}{np_\delta(t)}} \right) \\
=& \frac{1}{12 b_{\max,\delta}(t)} - C_2 \sqrt{\frac{\log n}{Lnp_\delta(t) p_{\delta,sum}(t)}} - \frac{1}{\sqrt{np_\delta(t)}} - C_9 \sqrt{\frac{\log n}{np_\delta(t)}} \\
\geqslant& \frac{1}{12 b_{\max,\delta}(t)} - \tilde{C} \sqrt{\frac{\log n}{np_\delta(t)}} \quad (\text{ since } p_{\delta,sum}(t) \geqslant \log n)
\end{aligned} \tag{2.4.6}$$

for some constant $\tilde{C} \geqslant 1$.

Concerning the terms in α, one can divide them in two groups depending on whether they depend on T or not. The sum of the terms depending on T is

$$\frac{2M\delta}{T} + 768 \frac{M\delta b^{\frac{5}{2}}(t)}{T\sqrt{3np_\delta(t)}} + C_7 \frac{Mn\delta b^{\frac{7}{2}}(t)}{T} \leqslant \tilde{C_6} \frac{Mn\delta b^{\frac{7}{2}}(t)}{T}$$

for some constant $\tilde{C_6} \geqslant 1$. Furthermore, the sum of the terms which are independent of T is

$$\begin{aligned}&\left(C_1 + 192C_3\frac{b^{\frac{5}{2}}(t)}{\sqrt{3np_\delta(t)}}\right)\sqrt{\frac{\log n}{Lnp_\delta(t)p_{\delta,sum}(t)}} + C_8\frac{b^{\frac{5}{2}}(t)\max\left\{b^2(t), \frac{\log n}{\sqrt{np_\delta(t)}}\right\}}{\sqrt{Lnp_\delta(t)p_{\delta,sum}(t)}}\\ &= \left(C_1 + 192C_3\frac{b^{\frac{5}{2}}(t)}{\sqrt{3np_\delta(t)}} + C_8\frac{b^{\frac{5}{2}}(t)}{\sqrt{\log n}}\max\left\{b^2(t), \frac{\log n}{\sqrt{np_\delta(t)}}\right\}\right)\sqrt{\frac{\log n}{Lnp_\delta(t)p_{\delta,sum}(t)}}\\ &\leqslant \tilde{C_5}\underbrace{\left(1 + \frac{b^{\frac{5}{2}}(t)}{\sqrt{\log n}}\max\left\{b^2(t), \frac{\log n}{\sqrt{np_\delta(t)}}\right\}\right)}_{=:\gamma_{\delta,n}(t)}\sqrt{\frac{\log n}{Lnp_\delta(t)p_{\delta,sum}(t)}}\end{aligned}$$

for some constant $\tilde{C_5} \geqslant 1$, since $np_\delta(t) \geqslant \log n$. Thus we arrive at the bound

$$\alpha \leqslant \tilde{C_5}\gamma_{n,\delta}(t)\sqrt{\frac{\log n}{Lnp_\delta(t)p_{\delta,sum}(t)}} + \tilde{C_6}\frac{Mn\delta b^{\frac{7}{2}}(t)}{T}. \tag{2.4.7}$$

Combining (2.4.6) and (2.4.7), we readily arrive at the stated bound in Theorem 3.

2.4.2 *Proof of Corollary 2*

As in the proof of Corollary 1, since $p_{\delta,sum}(t) \geqslant p_{\min}\delta$, the condition $\delta \gtrsim \frac{\log n}{p_{\min}}$ implies $p_{\delta,sum}(t) \gtrsim \log n$ and $p_\delta(t) \gtrsim 1$. Moreover, δ has to statisfy $\delta \leqslant T$ and condition (2.2.6), i.e.

$$\frac{M\delta}{T} + \sqrt{\frac{\log n}{n}} \lesssim \frac{1}{b^{\frac{5}{2}}(t)}.$$

If δ satisfies these assumptions and if $n \gtrsim \log n$, we have with probability at least $1 - O(n^{-9})$,

$$\frac{\|\hat{\pi}(t) - \pi^*(t)\|_\infty}{\|\pi^*(t)\|_\infty} \lesssim \frac{b_{\max}}{1 - b_{\max}\sqrt{\frac{\log n}{n}}}\left(\gamma_n(t)\sqrt{\frac{\log n}{Ln\delta p_{\min}}} + \frac{Mn\delta b^{\frac{7}{2}}(t)}{T}\right). \tag{2.4.8}$$

The optimal choice of $\delta \in (0, T]$ minimizing the RHS of (2.4.8) is given by

$$\delta^* = \min\left\{\frac{(\gamma_n(t)T)^{\frac{2}{3}}(\log n)^{\frac{1}{3}}}{(2M)^{\frac{2}{3}}nb^{\frac{7}{3}}(t)(Lp_{\min})^{\frac{1}{3}}}, T\right\}.$$

It now remains to check that δ^* satisfies the aforementioned conditions on δ. Clearly $\delta^* \leqslant T$ while $\delta^* \gtrsim \frac{\log n}{p_{\min}}$ and condition (2.2.6) are ensured for the stated conditions on T. Hence, for $\delta \asymp \delta^*$, if T satisfies the stated conditions and $n \gtrsim \log n$, we arrive at the stated expression for the ℓ_∞ error bound in the corollary.

2.5 EXPERIMENTS

We now empirically evaluate the performance of our method via numerical experiments[2] on synthetic data, and on a real dataset. We will in particular compare our method with the MLE approach of Bong et al. [4], as it is for now the only other method we are aware of that theoretically analyzes a 'smoothly evolving' dynamic BTL model as us.

2.5.1 *Synthetic data*

The synthetic data is generated as follows.

1. For $i \in [n]$, we simulate the strength of item i across the grid $w_i^* \in \mathbb{R}^{|\mathcal{T}|}$ using the same gaussian process $GP(\mu_i, \Sigma_i)$ as in [4].
   - $\mu_i = (\mu_i(0), \dots, \mu_i(T))$ with $\mu_i(t) \sim \mathcal{N}(0, 0.1)$ for all $t \in \mathcal{T}$, and
   - Σ_i is a Toeplitz symmetric matrix, with its coefficients in the first row defined as $\Sigma_{i,1t} = 1 - \frac{t}{T+1}$.

   We then define $w_i^* = \exp\left(GP(\mu_i, \Sigma_i)\right) \in \mathbb{R}^{|\mathcal{T}|}$ for each $i \in [n]$.
2. For all $t \in \mathcal{T}$, we simulate an Erdös-Renyi comparison graph $G(n, p(t))$ with $p(t)$ chosen randomly from the interval $[\frac{1}{n}, \frac{\log n}{n}]$. We check that the union graph of all the data on the grid $\mathcal{T}$ is connected. Indeed, it is a sufficient condition for the existence at all times $t \in [0, 1]$ of a δ such that the union graph $G_\delta(t)$ is connected (which is required for the ranking recovery).
3. For all $t \in \mathcal{T}$, for all $1 \leqslant i < j \leqslant n$ and for all $l \in [L]$, we draw the outcomes of the comparisons as $y_{ij}^{(l)}(t) \sim \mathcal{B}\left(\frac{w_{t,j}^*}{w_{t,i}^* + w_{t,j}^*}\right)$ and define $y_{ji}^{(l)}(t) = 1 - y_{ij}^{(l)}(t)$.

Starting with an initial value of δ as in Corollary 1 ($\delta \approx T^{\frac{2}{3}}$), we increase its value till the union graph $G_\delta(t)$ is connected. We then recover the weight vectors w_t^* for all $t \in \mathcal{T}$ using Algorithm 1. This process is repeated over 60 Monte Carlo runs.

Apart from the MLE approach [4], we also evaluate against an adaptation of the simple Borda Count method from the static setting to the dynamic setup. Let us describe briefly those two methods.

MLE METHOD. The analysis of Bong et al [4] relies on a kernel smoothing of the data followed by a maximum likelihood estimation. We choose the bandwidth parameter $h = T^{-3/4}$ for the kernel smoothing, as suggested in their synthetic experiments[3]. An alternative is to use a cross-validation procedure to tune h, but we avoid this method for computational reasons (see Tables 1, 2).

BORDA COUNT. This method, analysed by Ammar and Shah [1] in the static case, gives a score to each item based on its win rate. To estimate the

2 Code available at : https://github.com/karle-eglantine/Dynamic_Rank_Centrality

3 see their GitHub repository https://github.com/shamindras/bttv-aistats2020

scores at time t in our dynamic setup, we compute the win rate of each item i using the neighborhood $\mathcal{N}_\delta(t)$ as

$$s(t,i) = \frac{\sum_{t' \in \mathcal{N}_\delta(t)} \text{Number of times } i \text{ has won at time } t'}{\sum_{t' \in \mathcal{N}_\delta(t)} \text{Number of times } i \text{ has been compared at time } t'}.$$

The scores $(s(t,i))_{i \in [n]}$ then yield a ranking of the items at time t (the higher the winrate for an item, the better its rank). Note that the neighborhood used to compute these scores is the same neighborhood used in the DRC estimation. Hence, the parameter δ is chosen as its theoretical optimal value $\delta^* \simeq T^{2/3}$.

RANKING ERROR. In order to compare the rankings produced by these three methods, we compute for all them an estimation error with the error metric defined by Negahban [36]. Let $\pi^* = \frac{w^*}{\|w^*\|_1}$ denote the normalized true weight vector and σ denote an estimated ranking, with $\sigma_i < \sigma_j \Leftrightarrow i$ is better than j. The error metric is then defined[4] as follows.

$$D_{\pi^*}(\sigma) = \sqrt{\frac{1}{2n\,\|\pi^*\|_2^2} \sum_{i<j} (\pi_i^* - \pi_j^*)^2 \mathbf{1}_{(\pi_i^* - \pi_j^*)(\sigma_i - \sigma_j) > 0}}.$$

It has been shown in [36, Lemma 1] that this error criterion is related to the ℓ_2 error that we have bounded in Theorem 1. If the ranking σ comes from a weight vector $\hat{\pi}$, then

$$D_{\pi^*}(\sigma) \leqslant \frac{\|\pi^* - \hat{\pi}\|_2}{\|\pi^*\|_2}. \tag{2.5.1}$$

Although (2.5.1) doesn't necessarily require $\hat{\pi}$ to satisfy $\|\hat{\pi}\|_1 = 1$, we will impose this in what follows since $\hat{\pi}$ will be the strength estimates returned by the methods being compared.

RESULTS. The results are summarized below.

1. In Figure 5, we consider T ranging from 10 to 150, fix $L = 5$ and take $n = 100$ or $n = 400$. We plot the mean ℓ_2 error $\|\pi^* - \hat{\pi}\|_2$ versus T for our Algorithm 1 (dubbed DRC for Dynamic Rank Centrality), as well as the MLE method (since Borda Count is not designed for recovering the latent weight vector w^*). One can observe that the ℓ_2 error decreases with T for both DRC and MLE and they achieve similar performance, which is consistent with the theoretical results developed in the present paper, and in [4]. The error bars illustrate that the variance of the errors typically decreases with T. We note that the error curves for $n = 400$ are lower than those for $n = 100$, and the variance of the errors are also slightly smaller for $n = 400$.

2. We show in Figure 6 the evolution of the mean ranking error $D_{\pi^*}(\sigma)$ as a function of T, where for each T the mean is taken across all time instants in $\mathcal{T}$, and all Monte Carlo runs. We compute this error for the DRC, MLE and Borda Count methods, with σ, π denoting the es-

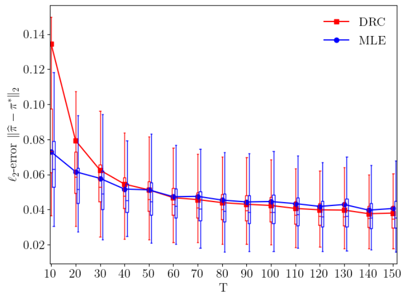
4 The quantities π^*, σ and $\hat{\pi}$ are obviously defined at the time instant t where the estimation is being carried out, however we suppress the dependence on t for ease of exposition.

timated ranks and weights by these algorithms. Using (2.5.1), this implies that $D_{\pi^*}(\sigma)$ should decrease with T which is what we observe in Figure 6a. As for the ℓ_2 error, MLE and DRC have similar performance. The Borda Count method performs well for rank recovery, as its error is only slightly worse than the other methods.

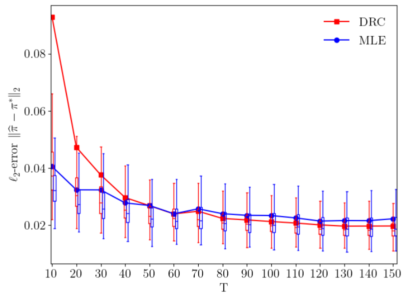
3. We plot the evolution of the ℓ_∞ error $\|\pi^* - \hat{\pi}\|_\infty$ versus T for $n = 100$ and $n = 400$ in Figure 7. We observe that the errors decrease with T, as shown theoretically, and that both MLE and DRC methods perform similarly.

4. In Figure 8, we show that the optimal value derived theoretically for δ is close to the numerically optimal δ. The minimal ℓ_2 error is indeed obtained for δ close to $\delta^* \simeq T^{2/3}$.

5. As the DRC and MLE methods performs similarly in term of the estimation error, it is of interest to look at their computational cost. We track the running time of both methods, for different values of n and T, as shown in Tables 1 and 2. We can see that the DRC method is far quicker than the MLE. Indeed, the MLE method solves an optimisation problem via a gradient descent algorithm whereas the DRC method only solves an eigenvalue problem.



(a) Evolution of the ℓ_2 error for $n = 100$

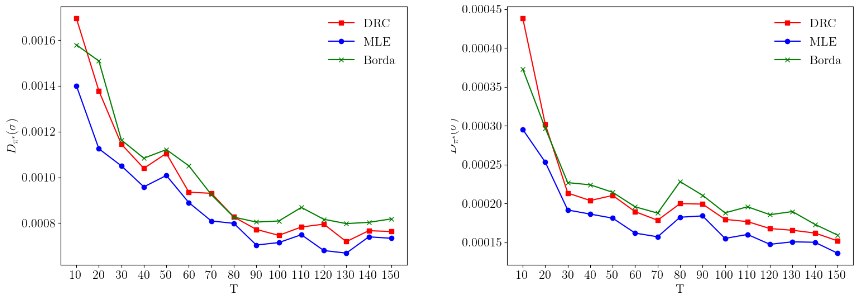


(b) Evolution of the ℓ_2 error for $n = 400$

Figure 5: Evolution of the ℓ_2 error with T for Dynamic Rank Centrality, the MLE and Borda Count method for $n = 100$ and $n = 400$. The results are averaged over the grid $\mathcal{T}$ as well as 60 Monte Carlo runs.

2.5.2 *Real dataset: NFL data*

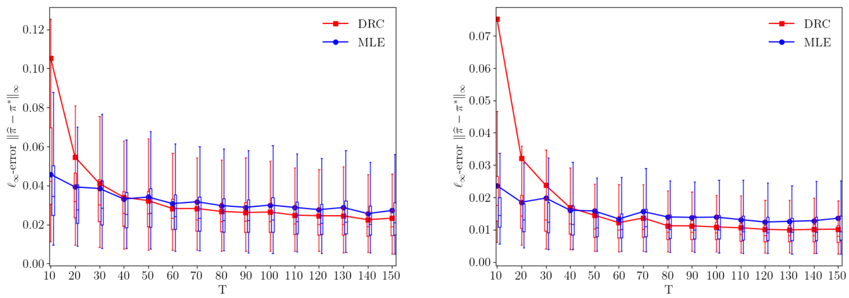
We now evaluate our method on a real dataset which consists of the results of National Football League (NFL) games for each season between 2009 and 2015, that are available in the `nflWAR` package [47]. The aim is to recover the ranking of the $n = 32$ teams at the end of each season, which contains $T = 16$ rounds. The dataset is hence composed of 16 comparison graphs with 32 nodes each, and comparison outcomes $(y_{ij}(t))_{i,j,t}$, where $y_{ij}(t) = 1$ if team j beat team i during in round t. We fit our model to this data by estimating the underlying strengths (and therefore the ranks) of the teams, at the end of a season. To do so, we tune the parameter δ using a Leave-One-Out Cross-Validation (LOOCV) procedure, described below.



(a) Evolution of the error metric $D_w(\sigma)$ for $n = 100$

(b) Evolution of the error metric $D_w(\sigma)$ for $n = 400$

Figure 6: Evolution of the errors $D_w(\sigma)$ with T for Dynamic Rank Centrality, the MLE and Borda Count method for $n = 100$ and $n = 400$. The results are averaged over the grid $\mathcal{T}$ as well as 60 Monte Carlo runs.



(a) $n = 100$

(b) $n = 400$

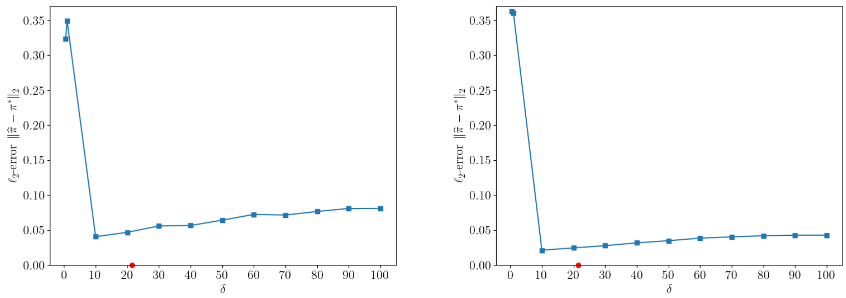
Figure 7: Evolution of the ℓ_∞ error with T for Dynamic Rank Centrality and MLE method for $n = 100$ and $n = 400$. The results are averaged over the grid $\mathcal{T}$ as well as 60 Monte Carlo runs.

T	10	20	30	40	50
DRC	0.48±0.01	0.97±0.02	1.53±0.02	2.13±0.03	2.83±0.04
MLE	2.17±0.84	3.69±0.52	6.06±0.72	9.14±1.23	12.2±2.42

T	60	70	80	90	100
DRC	3.61±0.03	4.44±0.05	5.36±0.07	6.44±0.07	7.38±0.66
MLE	15.67±3.2	21.61±4.16	34.62±20.68	27.35±3.08	39.5±15.93

T	110	120	130	140	150
DRC	8.25±0.14	9.52±0.14	10.38±0.18	11.85±0.13	13.2±0.19
MLE	56.86±38.95	48.69±9.54	50.27±4.35	55.93±2.62	65.99±5.97

Table 1: Average running time and associated standard deviations (in seconds) of DRC and MLE methods for $n = 100$. Results are averaged over 20 Monte Carlo runs.



(a) $n = 100$ (b) $n = 400$

Figure 8: Performance of the Dynamic Rank Centrality method for different value of parameter δ, with $T = 100$, $n = 100$ and $n = 400$. We highlight in red on the x-axis the theoretical optimal value of parameter $\delta = T^{2/3}$. The results are averaged over the grid $\mathcal{T}$ as well as 20 Monte Carlo runs.

T	10	20	30	40	50
DRC	5.41± 0.41	10.21±0.06	16.07±1.86	20.93±0.18	26.56± 0.18
MLE	37.04±35.7	68.92±16.68	130.84±39.06	148.84±27.88	251.99±175.51

T	60	70	80	90	100
DRC	32.58±0.40	39.55±1.52	46.24±1.33	53.25± 0.65	59.58±2.28
MLE	292.74±93.97	337.47±50.92	553.51±306.46	583.90±297.78	551.54± 56.71

Table 2: Average running time and associated standard deviations (in seconds) of DRC and MLE methods for $n = 400$. Results are averaged over 15 Monte Carlo runs.

1. Fix a list of potential values of δ and compute for each of them the associated estimates of the strength $\hat{\pi}$ at the end of the season.
2. For every possible values of δ, repeat the following steps several times.
   - Select randomly a game during the season, identified by a time t and the pair of compared teams $\{i, j\}$.
   - Consider the dataset obtained by removing the outcome $y_{ij}(t)$ of this game. Use this dataset to compute the estimated strength $\tilde{\pi}(t)$ at this time t.
   - Compute the prediction error $\left\| y_{ij}(t) - \frac{\tilde{\pi}_j(t)}{\tilde{\pi}_j(t)+\tilde{\pi}_i(t)} \right\|_2$.

   We then compute the mean of these prediction errors for each value of δ.
3. Finally, we select δ^* which has the smallest mean prediction error, and take as our estimate of the strengths at the end of the season the associated vector $\hat{\pi}$.

An analogous LOOCV procedure is performed in the MLE approach [4] to tune the bandwidth parameter h.

We then compare our estimated ranking with the ones obtained by the MLE method, the Borda Count method and also with the ELO ratings. The latter are reputed to be relevant estimations of the teams qualities and are also openly available [35]. Using the ELO ratings as ground truth, we assess the performance of the methods by comparing the estimated top 10 teams and by computing the correlation between the ELO rankings and the estimated ranks as well as correlations between the ELO ratings and the estimated strengths.

TOP 10 TEAMS. Figure 9 contains the estimation of the top 10 teams by all the methods, for the seasons 2011 to 2015. Considering the ELO ranks as the true ranks, one can observe that the DRC and MLE methods perform similarly, as a majority of the top 10 teams are recovered for each season. The Borda Count does not perform as well as it did on synthetic data; we can observe that it recovers the same ranks for several teams. This is explained by the fact that in this dataset, each team plays a small number of games, and thus the win rates take a finite (and small) number of values. However, it still recovers a large fraction of the top 10 teams.

	2011				2012			
Ranks	ELO	DRC	MLE	Borda	ELO	DRC	MLE	Borda
1	GB	GB	GB	GB	NE	ATL	HOU	ATL,DEN
2	NE	SF	SF	SF,NE,NO	DEN	SEA	ATL	
3	NO	NO	NO		GB	SF	SF	HOU,SEA,NE,IND,GB
4	PIT	NE	NE		SF	IND	CHI	
5	BAL	PIT	DET	BAL,PIT	ATL	NE	GB	
6	SF	BAL	BAL		SEA	GB	NE	
7	ATL	KC	PIT	DET,ATL,HOU	NYG	HOU	DEN	
8	PHI	DET	HOU		CIN	MIN	SEA	SF,CIN
9	SD	HOU	CHI		BAL	DEN	BAL	
10	HOU	NYG	ATL	TEN,NYG,CIN	HOU	NYG	IND	NYG,MIN,WAS,CHI,BAL

	2013				2014				2015			
Ranks	ELO	DRC	MLE	Borda	ELO	DRC	MLE	Borda	ELO	DRC	MLE	Borda
1	SEA	CAR	SEA	CAR	SEA	NE	DEN	NE,DAL,GB	SEA	CAR	CAR	CAR
2	SF	SEA	DEN	SF,SEA	NE	DAL	ARI		CAR	ARI	DEN	ARI
3	NE	SF	NO		DEN	GB	NE		ARI	DEN	NE	NE,DEN,CIN
4	DEN	NO	KC	CIN,DEN,NE	GB	ARI	SEA	SEA,IND,DEN	KC	PIT	CIN	
5	CAR	NE	SF		DAL	DEN	DAL		DEN	NE	ARI	
6	CIN	ARI	NE		PIT	SEA	GB		NE	CIN	GB	MIN,KC
7	NO	NYJ	IND	PHI	BAL	KC	PHI	ARI,BAL,PIT,DET	PIT	ATL	MIN	
8	ARI	CIN	CAR	PIT	IND	SD	SD		CIN	SEA	KC	SEA,GB,PIT,NYJ
9	IND	MIA	ARI	ARI,SD,NYG,IND	ARI	CIN	DET		GB	KC	PIT	
10	SD	PIT	CIN		CIN	PIT	KC		MIN	MIN	SEA	

Figure 9: The top 10 teams for seasons 2011 to 2015, using ELO ranks, DRC, the MLE, and Borda Count. Teams highlighted in yellow for a particular recovery method are teams appearing in the top 10 list for ELO rankings. Teams highlighted in green are recovered at the same rank as in the ELO rankings.

CORRELATION WITH ELO RANKINGS. As all of the methods estimate the ranking of the teams, one can also compute the Kendall rank correlation[5]

5 The Kendall rank correlation [25] between two rankings $x, y \in \{1, \dots, n\}$ is defined as

$$\tau := \frac{2}{n(n-1)} \sum_{i<j} \mathrm{sign}(x_i - x_j)\mathrm{sign}(y_i - y_j).$$

- If $\tau = 1$, x and y are the same.
- If $\tau = -1$, x and y are reversed.
- If $\tau = 0$, the correlation between x and y is no better than the correlation between 2 random rankings.

between the estimated rankings of each method with the ELO ranking (considered as the true ranks). The results, gathered in Table 3, show that the methods have similar performance.

	2011	2012	2013	2014	2015
DRC	-0.092	0.104	0.116	0.245	-0.201
MLE	-0.052	0.161	0.125	0.125	-0.084
Borda	0.1	0.104	0.08	0.209	-0.036

Table 3: Kendall rank correlations between the ELO ranks and the estimated ranks by DRC, MLE and Borda Count methods at the end of each seasons. All of the methods perform similarly.

CORRELATION WITH ELO RATINGS. The ELO ratings provide a vector of underlying ground truth strengths for each team. Then, as the DRC and MLE methods estimate the underlying strengths, one can also compute the correlation between the estimated strengths and the ELO ratings (all normalized such that their ℓ_1 norm is 1). These results are gathered in Table 4 and show that the DRC performs much better than the MLE for all seasons. In particular, it shows that even if the ranks are not perfectly recovered, the estimated strengths by DRC are positively correlated with the ELO ratings.

	2011	2012	2013	2014	2015
DRC	0.425	0.518	0.284	0.478	0.414
MLE	0.092	-0.237	-0.337	-0.026	0.002

Table 4: Correlations between the normalized ELO ratings and the DRC and MLE estimated strengths at the end of each seasons. Note here that DRC performs much better than the MLE.

2.6 COMPARISON WITH RELATED WORK

We now provide a detailed discussion with closely related work for dynamic ranking. As mentioned in Section 1.2, existing theoretical results for the dynamic ranking setup are limited, and the only works we are aware of are the recent results of Bong et al. [4] and of Li and Wakin [28]. We now discuss the results of Bong et al in comparison to our work, as both works focus on a dynamic version of the BTL model. The dynamic BTL model studied by Bong et al. [4] is closely related to the one we presented in Section 2.1. Let us recall their model (see Section 1.2 for more details).

$$\operatorname{logit}(\mathbb{P}(i \text{ beats } j \text{ at time } t)) = \beta_i^*(t) - \beta_j^*(t)$$

where $\beta^*(t)$ represents the vector of scores at time t with $w^*(t) = \exp(\beta^*(t))$. The main differences with our model are that (a) the grid $\mathcal{T}$ can be non-uniform, and (b) the number of comparisons $L_{ij}(t)$ made for each pair $\{i, j\}$ at each time $t \in [0, 1]$ can vary. We assumed the grid $\mathcal{T}$ to be uniform only for simplicity, our analysis can be easily extended to handle the non-uniform setting as long as $\mathcal{T}$ is "sufficiently regular". Similarly, one could potentially extend our analysis to handle varying number of comparisons $L_{ij}(t)$, although the current proof technique will lead to the presence of the maximum of the $L_{ij}(t)$'s (recall Remark 2). The assumption $L_{ij}(t) = L$ was only made for simplicity in the proofs as is typically done for the static setting (see for eg., [8, Remark 3]). The main assumptions needed for their theoretical analysis are the following.

- The probabilities $\mathbb{P}(i \text{ beats } j \text{ at time } t)$ are Lispchitz functions of time $t \in [0, 1]$ for all $i \neq j \in [n]$ (same as Assumption 1 in this chapter).
- Each pair of teams $\{i, j\}$ has been compared at least at one time point $t' \in \mathcal{T}$. Translated to our notation, this means that the union graph $\cup_{t' \in \mathcal{T}} G_{t'}$ is complete.

We remark that this last assumption is a stronger assumption than the connectivity assumption on the union graph $G_\delta(t)$ that we required in our analysis.

Let us recall the ℓ_∞ bounds they obtained in order to compare them with our results. Denoting $\hat{\beta}(t)$ to be the "MLE" estimator (i.e., the solution of (1.2.2)), then

- for a bandwidth $h \gtrsim \left(\frac{\log n}{T}\right)^{\frac{1}{3}}$, it holds w.h.p. [4, Theorem 5.2]
$$\left\|\hat{\beta}(t) - \beta^*(t)\right\|_\infty \lesssim \delta_h(t) + \left(\frac{\log n}{nT}\right)^{\frac{1}{3}};$$
- for $h \gtrsim \left(\frac{\log(nT^3)}{T}\right)^{\frac{1}{3}}$, it holds w.h.p. [4, Theorem 5.3]
$$\sup_{t\in[0,1]} \left\|\hat{\beta}(t) - \beta(t)\right\|_\infty \lesssim \sup_{t\in[0,1]} \delta_h(t) + \left(\frac{\log(nT)}{T}\right)^{\frac{1}{3}}.$$

Note that they also recover the $T^{-1/3}$ rate for pointwise estimation of Lipschitz functions.

While we provide pointwise estimation error bounds for any given $t \in [0, 1]$, it is also possible to extend our results to obtain error bounds holding uniformly over all $t \in [0, 1]$. The first main idea here would be to observe that there are $O(T)$ different number of neighborhoods in the construction of $\hat{P}(t)$ in (2.1.5), which implies that there are $O(T)$ different possible values of $\hat{\pi}(t)$ over all $t \in [0, 1]$. Secondly, one can verify that $\pi^*(t)$ is a Lipschitz function of t. These two facts together with a union bound argument can be used to establish ℓ_2 and ℓ_∞ bounds holding uniformly over $t \in [0, 1]$, with probability at least $1 - O(Tn^{-c})$ where c is a suitably large constant. In our analysis, we had taken c to be 10 (resp. 9) for the ℓ_2 (resp. ℓ_∞) analysis, but it can be any suitably large value, at the expense of worsening the other constants in the accompanying theorems in Section 2.2.

2.7 SUMMARY OF NOTATION

Symbol	*Definition*
$\mathcal{T}$	Uniform grid of $[0,1]$, with size $T+1$.
$G_{t'} = ([n], \mathcal{E}_{t'})$	Undirected comparison graph at time $t' \in \mathcal{T}$
$w_t^* \in \mathbb{R}^n$	Ground-truth strengths at time $t \in [0,1]$
$y_{ij}^*(t)$ and $y_{ij}(t)$	(resp. population and empirical) Fraction of times j beats i at time t
M	Lipschitz constant of $y_{ij}^*(t)$
L	Number of independent comparisons made for each $\{i,j\} \in \mathcal{E}_{t'}$
$\mathcal{N}_\delta(t)$	Neighborhood of size δ, see (2.1.3)
$\mathcal{N}_{ij,\delta}(t)$	$\{t' \in \mathcal{N}_\delta(t) \vert \ \{i,j\} \in \mathcal{E}_{t'}\}$
$G_\delta(t) = ([n], \mathcal{E}_\delta(t))$	"Union graph" where $\mathcal{E}_\delta(t) = \big\{\{i,j\} : i \neq j, \vert\mathcal{N}_{ij,\delta}(t)\vert \geqslant 1\big\}$
$N_{max,\delta}(t)$ and $N_{min,\delta}(t)$	$N_{max,\delta}(t) = \max_{\{i,j\}\in\mathcal{E}_\delta(t)} \vert\mathcal{N}_{ij,\delta}(t)\vert$ and $N_{min,\delta}(t) = \min_{\{i,j\}\in\mathcal{E}_\delta(t)} \vert\mathcal{N}_{ij,\delta}(t)\vert$
$\hat{P}(t), \bar{P}(t) \in \mathbb{R}^{n\times n}$	Empirical and population transition matrices resp., see (2.1.5) and (2.1.6)
$\hat{\pi}(t), \pi^*(t) \in \mathbb{R}^n$	Stationary distributions of $\hat{P}(t)$ and $\bar{P}(t)$ resp.
$d_{max,\delta}(t), d_{min,\delta}(t)$	Maximum (resp. minimum) degree of $G_\delta(t)$
$d_\delta(t)$ $(\geqslant d_{max,\delta}(t))$	Normalization term in (2.1.5)
$b(t)$	$\max_{i,j\in[n]} \frac{w_{t,i}^*}{w_{t,j}^*}$ for all $t \in [0,1]$
$b_{max,\delta}(t)$	$\max_{t'\in\mathcal{N}_\delta(t)} b(t')$
$\mathcal{L}_\delta(t)$	Random walk Laplacian of $G_\delta(t)$
$\xi_\delta(t)$	$\xi_\delta(t) = 1 - \lambda_{max}(\mathcal{L}_\delta(t))$ where $\lambda_{max}(\mathcal{L}_\delta(t))$ is the second largest eigenvalue (in absolute value) of $\mathcal{L}_\delta(t)$
$G_{t'} = \mathcal{G}(n, p(t'))$	Erdös-Renyi graph at $t' \in \mathcal{T}$ with parameter $p(t') \in [0,1]$
$p_\delta(t)$	$1 - \prod_{t'\in\mathcal{N}_\delta(t)}(1-p(t'))$, see (2.2.3)
$p_{\delta,sum}(t)$	$\sum_{t'\in\mathcal{N}_\delta(t)} p(t')$
p_{min}, p_{max}	$p_{min} := \min_{t'\in\mathcal{T}} p(t')$ and $p_{max} := \max_{t'\in\mathcal{T}} p(t')$

Table 5: Summary of symbols used throughout the chapter along with their definitions.

2.8 PROOFS OF RESULTS IN SECTION 2.3

2.8.1 *Proof of Lemma 2*

Recall that the entries of $\Delta_1(t)$ are given by

$$\Delta_{1,ij}(t) = \begin{cases} \dfrac{1}{d_\delta(t)\vert\mathcal{N}_{ij,\delta}(t)\vert} \displaystyle\sum_{t'\in\mathcal{N}_{ij,\delta}(t)} \left(y_{ij}^*(t') - y_{ij}^*(t)\right) & \text{if} \quad \{i,j\} \in \mathcal{E}_\delta(t), \\ -\displaystyle\sum_{k\neq i} \Delta_{1,ik}(t) & \text{if} \quad i = j, \\ 0 & \text{otherwise.} \end{cases}$$

Let us denote $D_1(t)$ to be the diagonal matrix containing the elements $\Delta_{1,ii}(t)$ and $\overline{\Delta_1}(t) = \Delta_1(t) - D_1(t)$. As $D_1(t)$ is diagonal, we then have

$$\|\Delta_1(t)\|_2 \leqslant \|D_1(t)\|_2 + \left\|\overline{\Delta_1}(t)\right\|_2 \leqslant \max_i \left|\Delta_{1,ii}(t)\right| + \left\|\overline{\Delta_1}(t)\right\|_F.$$

Let us first bound $\left\|\overline{\Delta_1}(t)\right\|_F$. Assumption 1 implies that for any $\{i,j\} \in \mathcal{E}_\delta(t)$,

$$\begin{aligned}\left|\overline{\Delta_{1,ij}}(t)\right| \leqslant & \frac{1}{d_\delta(t)|\mathcal{N}_{ij,\delta}(t)|} \sum_{t' \in \mathcal{N}_{ij,\delta}(t)} \left|y^*_{ij}(t') - y^*_{ij}(t)\right| \\ \leqslant & \frac{1}{d_\delta(t)|\mathcal{N}_{ij,\delta}(t)|} \sum_{t' \in \mathcal{N}_{ij,\delta}(t)} M\left|t' - t\right| \\ \leqslant & \frac{M\delta}{Td_\delta(t)},\end{aligned}$$

which in turn implies $\left\|\overline{\Delta_1}(t)\right\|_F \leqslant 2\frac{M|\mathcal{E}_\delta(t)|\delta}{Td_\delta(t)}$. In order to bound $\|D_1(t)\|_2$, we simply note that

$$\left|D_{1,ii}(t)\right| = \left|-\sum_{j\neq i} \Delta_{1,ij}(t)\right| \leqslant d_{\max,\delta}(t) \max_{j\neq i} \left|\Delta_{1,ij}(t)\right| \leqslant d_{\max,\delta}(t)\frac{M\delta}{Td_\delta(t)} \leqslant \frac{M\delta}{T}$$

since $d_\delta(t) \geqslant d_{\max,\delta}(t)$. Hence $\|D_1(t)\|_2 \leqslant \frac{M\delta}{T}$, and so

$$\|\Delta_1(t)\|_2 \leqslant \frac{M\delta}{T}\left(1 + 2\frac{|\mathcal{E}_\delta(t)|}{d_\delta(t)}\right) \leqslant \frac{M\delta}{T}\left(1 + 2\frac{|\mathcal{E}_\delta(t)|}{d_{\max,\delta}(t)}\right) \leqslant 4\frac{M\delta|\mathcal{E}_\delta(t)|}{Td_{\max,\delta}(t)},$$

since $1 \leqslant 2\frac{|\mathcal{E}_\delta(t)|}{d_{\max,\delta}(t)}$.

□

2.8.2 *Proof of Lemma 3*

Our goal is to bound $\|\Delta(t)\|_2$ for the random matrix $\Delta(t) = \hat{P}(t) - \bar{P}(t)$, whose entries are

$$\Delta_{ij}(t) = \begin{cases} \dfrac{1}{d_\delta(t)|\mathcal{N}_{ij,\delta}(t)|} \displaystyle\sum_{t' \in \mathcal{N}_{ij,\delta}(t)} \left(y_{ij}(t') - y^*_{ij}(t')\right) & \text{if} \quad \{i,j\} \in \mathcal{E}_\delta(t) \\ -\displaystyle\sum_{k\neq i} \Delta_{ik}(t) & \text{if} \quad i = j \\ 0 & \text{otherwise.} \end{cases}$$

Once again $\Delta(t)$ can be decomposed as $\Delta(t) = D(t) + \overline{\Delta}(t)$ where $D(t)$ is a diagonal matrix and $\overline{\Delta}(t)$ contains its off-diagonal coefficients. Since

$$\|\Delta(t)\|_2 \leqslant \|D(t)\|_2 + \left\|\overline{\Delta}(t)\right\|_2,$$

we will bound $\|D(t)\|_2$ and $\left\|\overline{\Delta}(t)\right\|_2$ using ideas from the proof of [36, Lemma 3].

2.8.2.1 *Bound on* $\|D(t)\|_2$

The matrix $D(t)$ is diagonal, hence $\|D(t)\|_2 = \max_{i\in[n]} |D_{ii}(t)|$ where the diagonal entries are given by

$$D_{ii}(t) = -\sum_{k\neq i}\sum_{t'\in\mathcal{N}_{ik,\delta}(t)}\sum_{l=1}^{L}\frac{1}{L|\mathcal{N}_{ik,\delta}(t)|d_\delta(t)}(y_{ik}^{(l)}(t') - y_{ik}^*(t')).$$

This is a sum of at most $Ld_{max,\delta}(t)N_{max,\delta}(t)$ independent and centered variables, which almost surely lie in the interval $\left[-\frac{1}{Ld_\delta(t)N_{min,\delta}(t)}, \frac{1}{Ld_\delta(t)N_{min,\delta}(t)}\right]$. Thus, by applying Hoeffding's inequality (see Theorem 13 in Section 4.1.1) we obtain for any $s > 0$

$$\mathbb{P}\left(|D_{ii}(t)| > s\right) \leqslant 2\exp\left(-\frac{2s^2L^2d_\delta^2(t)N_{min,\delta}^2(t)}{4Ld_{max,\delta}(t)N_{max,\delta}(t)}\right) = 2\exp\left(-\frac{Ld_\delta^2(t)N_{min,\delta}^2(t)s^2}{2N_{max,\delta}(t)d_{max,\delta}(t)}\right)$$

which implies via a union bound that

$$\mathbb{P}(\|D(t)\|_2 > s) \leqslant 2n\exp\left(-\frac{Ld_\delta^2(t)N_{min,\delta}^2(t)s^2}{2N_{max,\delta}(t)d_{max,\delta}(t)}\right).$$

Choosing $s = c_1\sqrt{\frac{N_{max,\delta}(t)d_{max,\delta}(t)\log n}{Ld_\delta^2(t)N_{min,\delta}^2(t)}}$ where $c_1 \geqslant 2$ is a constant, we have that

$$\|D(t)\|_2 \leqslant c_1\sqrt{\frac{N_{max,\delta}(t)d_{max,\delta}(t)\log n}{Ld_\delta^2(t)N_{min,\delta}^2(t)}}$$

with probability at least $1 - 2n^{1-\frac{c_1^2}{2}}$.

2.8.2.2 *Bound on* $\|\overline{\Delta}(t)\|_2$ *when* $d_{max,\delta}(t)N_{max,\delta}(t) \leqslant \log n$

In order to bound $\|\overline{\Delta}(t)\|_2$, we first recall the standard inequality $\|M\|_2 \leqslant \sqrt{\|M\|_1\|M\|_\infty}$, and also $\|M\|_1 = \|M^\top\|_\infty$. Since $\overline{\Delta}(t)$ is skew-symmetric, it follows that

$$\|\overline{\Delta}(t)\|_2 \leqslant \sqrt{\|\overline{\Delta}(t)\|_1\|\overline{\Delta}(t)^\top\|_1} \leqslant \|\overline{\Delta}(t)\|_1.$$

Our aim now is to bound $\|\overline{\Delta}(t)\|_1$. Recall that for every $i \neq j$,

$$\overline{\Delta}_{ij}(t) = \begin{cases} \frac{1}{d_\delta(t)|\mathcal{N}_{ij,\delta}(t)|}\sum_{t'\in\mathcal{N}_{ij,\delta}(t)}\left(y_{ij}(t') - y_{ij}^*(t')\right) & \text{if} \quad \{i,j\}\in\mathcal{E}_\delta(t) \\ 0 & \text{otherwise.}\end{cases}$$

Now denoting for all i,

$$R_i(t) = \sum_{j\neq i}\sum_{t'\in\mathcal{N}_{ij,\delta}(t)}\left|\sum_{l=1}^{L}\frac{1}{Ld_\delta(t)|\mathcal{N}_{ij,\delta}(t)|}y_{ij}^{(l)}(t') - y_{ij}^*(t')\right|,$$

clearly $\|\overline{\Delta}\|_2 \leqslant \max_{i\in[n]} \sum_{j=1}^{n} |\overline{\Delta}_{ij}(t)| \leqslant \max_{i\in[n]} R_i(t)$ (since $\overline{\Delta}_{ii}(t) = 0$). In order to bound $R_i(t)$ for any given $i \in [n]$, let us first denote

$$\mathcal{S}_i = \left\{ (\xi_{jt'}^{(i)})_{\{i,j\}\in\mathcal{E}_{t'}, t'\in\mathcal{N}_{ij,\delta}(t)} \mid \xi_{jt'}^{(i)} \in \{-1,1\} \right\}.$$

Since the random variables $\frac{\xi_{jt'}^{(i)}}{|\mathcal{N}_{ij,\delta}(t)|}\left(y_{ij}^{(l)}(t') - y_{ij}^{*}(t')\right)$ lie almost surely in the interval $[-\frac{1}{N_{\min,\delta}(t)}, \frac{1}{N_{\min,\delta}(t)}]$, we obtain via a simple union bound argument, along with Hoeffding's inequality, that

$$\begin{aligned}
\mathbb{P}(R_i(t) > s) &\leqslant \sum_{\xi^{(i)}\in\mathcal{S}_i} \mathbb{P}\left(\sum_{j\neq i} \sum_{t'\in\mathcal{N}_{ij,\delta}(t)} \xi_{jt'}^{(i)} \sum_{l=1}^{L} \frac{1}{Ld_\delta(t)|\mathcal{N}_{ij,\delta}(t)|}(y_{ij}^{(l)}(t') - y_{ij}^{*}(t')) > s \right) \\
&\leqslant \sum_{\xi^{(i)}\in\mathcal{S}_i} \mathbb{P}\left(\sum_{j\neq i} \sum_{t'\in\mathcal{N}_{ij,\delta}(t)} \xi_{jt'}^{(i)} \sum_{l=1}^{L} \frac{1}{|\mathcal{N}_{ij,\delta}(t)|}(y_{ij}^{(l)}(t') - y_{ij}^{*}(t')) > Ld_\delta(t)s \right) \\
&\leqslant \sum_{\xi^{(i)}\in\mathcal{S}_i} \exp\left(-\frac{L^2 d_\delta^2(t) N_{\min,\delta}^2(t) s^2}{2Ld_{\max,\delta}(t)N_{\max,\delta}(t)} \right) \\
&\leqslant \sum_{\xi^{(i)}\in\mathcal{S}_i} \exp\left(-\frac{L d_\delta^2(t) N_{\min,\delta}^2(t) s^2}{2d_{\max,\delta}(t)N_{\max,\delta}(t)} \right).
\end{aligned}$$

Since $|\mathcal{S}_i| \leqslant 2^{d_{\max,\delta}(t)N_{\max,\delta}(t)}$ for each $i \in [n]$, we obtain

$$\begin{aligned}
\mathbb{P}(R_i(t) > s) &\leqslant 2^{d_{\max,\delta}(t)N_{\max,\delta}(t)} \exp\left(-\frac{L d_\delta^2(t) N_{\min,\delta}^2(t) s^2}{2d_{\max,\delta}(t)N_{\max,\delta}(t)} \right) \\
&= \exp\left(d_\delta(t)N_{\max,\delta}(t)\ln 2 - \frac{L d_\delta^2(t) N_{\min,\delta}^2(t) s^2}{2d_{\max,\delta}(t)N_{\max,\delta}(t)} \right).
\end{aligned}$$

Applying a union bound over $[n]$ now leads to

$$\mathbb{P}(\|\overline{\Delta}(t)\|_2 > s) \leqslant n \exp\left(d_{\max,\delta}(t)N_{\max,\delta}(t)\ln 2 - \frac{L d_\delta^2(t) N_{\min,\delta}^2(t) s^2}{2d_{\max,\delta}(t)N_{\max,\delta}(t)} \right).$$

Finally, since $d_{\max,\delta}(t)N_{\max,\delta}(t) \leqslant \log n$ by assumption we obtain for any constant $c_2 > 2$ that

$$\mathbb{P}\left(\|\overline{\Delta}(t)\|_2 > c_2\sqrt{\frac{N_{\max,\delta}(t)d_{\max,\delta}(t)\log n}{Ld_\delta^2(t)N_{\min,\delta}^2(t)}} \right) \leqslant n^{2-\frac{c_2^2}{2}}.$$

2.8.2.3 *Bound on* $\|\overline{\Delta}(t)\|_2$ *when* $d_{\max,\delta}(t)N_{\max,\delta}(t) \geqslant \log n$

In this case, we will use Bernstein's inequality for random matrices (see Theorem 14 in Section 4.1.1). For each $i < j \in [n]$, let $Z^{ij,l}(t')$ be a $n \times n$ matrix with all entries equal to zero except for $Z_{ij}^{ij,l}(t')$ and $Z_{ji}^{ij,l}(t')$ defined as follows.

$$Z_{ij}^{ij,l}(t') = \frac{C_{ij}^{(l)}(t')}{Ld_\delta(t)|\mathcal{N}_{ij,\delta}(t)|} \quad \text{and} \quad Z_{ji}^{ij,l}(t') = -\frac{C_{ij}^{(l)}(t')}{Ld_\delta(t)|\mathcal{N}_{ij,\delta}(t)|}$$

where $C_{ij}^{(l)}(t') = y_{ij}^{(l)}(t') - y_{ij}^{*}(t')$ if $\{i,j\} \in \mathcal{E}_\delta(t)$, which is a centered Bernoulli variable, and $C_{ij}^{(l)}(t') = 0$ otherwise. Hence, $\mathbb{E}[Z^{ij,l}(t')] = 0$. Now we can write

$$\overline{\Delta}(t) = \sum_{\substack{i<j \\ \{i,j\}\in\mathcal{E}_\delta(t)}} \sum_{t'\in\mathcal{N}_{ij,\delta}(t)} \sum_{l=1}^{L} Z^{ij,l}(t').$$

To apply Bernstein's inequality, we first note that $\left\|Z^{ij,l}(t')\right\|_2$ is uniformly bounded for all t', i, j.

$$\left\|Z^{ij,l}(t')\right\|_2 \leqslant \left\|Z^{ij,l}(t')\right\|_F = \frac{\sqrt{2}\left|C_{ij}^{(l)}(t')\right|}{Ld_\delta(t)|\mathcal{N}_{ij,\delta}(t)|} \leqslant \frac{\sqrt{2}}{Ld_\delta(t)N_{\min,\delta}(t)} =: B.$$

Since $Z^{ij,l}(t')$ are skew-symmmetric and independent matrices, we also have that

$$\nu = \left\|\mathbb{E}[\overline{\Delta}(t)\overline{\Delta}(t)^\top]\right\|_2 = \left\| \sum_{\substack{i<j \\ \{i,j\}\in\mathcal{E}_\delta(t)}} \sum_{t'\in\mathcal{N}_{ij,\delta}(t)} \sum_{l=1}^{L} -\mathbb{E}[Z^{ij,l}(t')^2] \right\|_2 .$$

The matrix $(Z^{ij,l}(t'))^2$ only has two non-zero entries, which are at the locations (i,i) and (j,j), and they are equal to $-\frac{C_{ij}^{(l)}(t')^2}{L^2 d_\delta^2(t)|\mathcal{N}_{ij,\delta}(t)|^2}$. Then its expectation (which also has only two non zero entries) satisfies

$$-\mathbb{E}[Z^{ij,l}(t')^2]_{i,i} = -\mathbb{E}[Z^{ij,l}(t')^2]_{j,j} = \frac{\mathbb{E}[C_{ij}^{(l)}(t')^2]}{L^2 d_\delta^2(t)|\mathcal{N}_{ij,\delta}(t)|^2} \leqslant \frac{Var(C_{ij}^{(l)}(t'))}{L^2 d_\delta^2(t) N_{\min,\delta}^2(t)} \leqslant \frac{1}{4L^2 d_\delta^2(t) N^2_{m}}$$

since $Var(C_{ij}^{(l)}(t')) = y_{ij}^{*}(t')(1 - y_{ij}^{*}(t')) \leqslant 1/4$. Furthermore, $\mathbb{E}[\overline{\Delta}(t)\overline{\Delta}(t)^\top]$ is a diagonal matrix with positive entries, hence

$$\begin{aligned}\nu &= \max_{k\in[n]} \left| \left(\sum_{\substack{i<j \\ \{i,j\}\in\mathcal{E}_\delta(t)}} \sum_{t'\in\mathcal{N}_{ij,\delta}(t)} \sum_{l=1}^{L} -\mathbb{E}[Z^{ij}(t')^2] \right)_{kk} \right| \\ &\leqslant \frac{Ld_{\max,\delta}(t)N_{\max,\delta}(t)}{4L^2 d_\delta^2(t) N_{\min,\delta}^2(t)} = \frac{N_{\max,\delta}(t)d_{\max,\delta}(t)}{4Ld_\delta^2(t)N_{\min,\delta}^2(t)}.\end{aligned}$$

Finally applying Bernstein's inequality, it holds for any $s > 0$ that

$$\begin{aligned}\mathbb{P}(\left\|\overline{\Delta}(t)\right\|_2 > s) &\leqslant 2n \exp\left(-\frac{3s^2}{6\nu + 2Bs}\right) \\ &\leqslant 2n \exp\left(-\frac{3s^2}{\frac{6N_{\max,\delta}(t)d_{\max,\delta}(t)}{4Ld_\delta^2(t)N_{\min,\delta}^2(t)} + \frac{2s\sqrt{2}}{Ld_\delta(t)N_{\min,\delta}(t)}}\right) \\ &\leqslant 2n \exp\left(-\frac{6Ld_\delta^2(t)N_{\min,\delta}^2(t)s^2}{3N_{\max,\delta}(t)d_{\max,\delta}(t) + 4\sqrt{2}N_{\min,\delta}(t)d_\delta(t)s}\right).\end{aligned}$$

Choosing $s = c_3\sqrt{\frac{N_{\max,\delta}(t)d_{\max,\delta}(t)\log n}{Ld_\delta^2(t)N_{\min,\delta}^2(t)}}$ for a constant $c_3 \geqslant 2$ and since $d_{\max,\delta}(t)N_{\max,\delta}(t) \geqslant \log n$ by assumption, it holds that

$$\mathbb{P}\left(\left\|\overline{\Delta}(t)\right\|_2 > c_3\sqrt{\frac{N_{\max,\delta}(t)d_{\max,\delta}(t)\log n}{Ld_\delta^2(t)N_{\min,\delta}^2(t)}}\right) \leqslant 2n^{1-\frac{6c_3^2}{3+4\sqrt{2}c_3}}. \quad (2.8.1)$$

Finally, we observe that up to a multiplicative constant, the bounds for $\|D(t)\|_2$ and $\left\|\overline{\Delta}(t)\right\|_2$ are the same. Thus, for constants $c_1, c_2, c_3 \geqslant 2$, we have

$$\|\Delta(t)\|_2 \leqslant (c_1 + \max\{c_2, c_3\})\sqrt{\frac{N_{\max,\delta}(t)d_{\max,\delta}(t)\log n}{Ld_\delta^2(t)N_{\min,\delta}^2(t)}}.$$

with probability at least $1 - 2n^{1-\frac{c_1^2}{2}} - \max\left\{n^{2-\frac{c_2^2}{2}}, 2n^{1-\frac{6c_3^2}{3+4\sqrt{2}c_3}}\right\}$. In simpler words, there exists a constant $\tilde{C_1} \geqslant 15$ such that with probability at least $1 - O(n^{-10})$,

$$\|\Delta(t)\|_2 \leqslant \tilde{C_1}\sqrt{\frac{N_{\max,\delta}(t)d_{\max,\delta}(t)\log n}{Ld_\delta^2(t)N_{\min,\delta}^2(t)}}. \quad (2.8.2)$$

□

2.8.3 *Proof of Lemma 4*

The steps below follow the proof of [36, Lemma 4] and are provided for completeness. We will first show $1 - \lambda_{\max}(\bar{P}(t)) \geqslant \frac{\xi_\delta(t)d_{\min,\delta}(t)}{4d_\delta(t)b^3(t)}$ using [36, Lemma 6]. To this end, let us denote $Q_{ij}(t) = \frac{1}{d_i(t)}\mathbf{1}_{\{i,j\}\in\mathcal{E}_\delta(t)}$ where $d_i(t)$ denotes the degree of the ith vertex in $G_\delta(t)$ and $\mathbf{1}$ is an indicator variable. Clearly, $Q(t)$ has a unique stationary distribution as it is irreducible (since $G_\delta(t)$ is connected), defined as $\mu_i(t) = \frac{d_i(t)}{\sum_{j=1}^n d_j(t)}$. It satisfies the detailed balanced equations of reversibility

$$\mu_i(t)Q_{ij}(t) = \mu_j(t)Q_{ji}(t).$$

Using [36, Lemma 6], we obtain

$$\frac{1-\lambda_{\max}(\bar{P}(t))}{1-\lambda_{\max}(Q(t))} \geqslant \frac{\alpha}{\beta} \quad \text{with } \alpha = \min_{\{i,j\}\in\mathcal{E}_\delta(t)} \frac{\pi_i^*(t)\bar{P}_{ij}(t)}{\mu_i(t)Q_{ij}(t)}, \quad \beta = \max_i \frac{\pi_i^*(t)}{\mu_i(t)}. \quad (2.8.3)$$

Note that the transition matrix Q is equal to the Laplacian $\mathcal{L}_\delta(t)$ of the graph $G_\delta(t)$ and so $1 - \lambda_{\max}(Q(t)) = \xi_\delta(t)(t)$. We now need to suitably bound α and β.

Denoting $d_i(t)$ to be the degree of node i in $G_\delta(t)$, clearly $\sum_{j=1}^n d_j(t) = 2|\mathcal{E}_\delta(t)|$. Hence we have

$$\mu_i(t)Q_{ij}(t) = \frac{1}{\sum_{k=1}^n d_k(t)} \leqslant \frac{1}{|\mathcal{E}_\delta(t)|} \text{ for } \{i,j\}\in\mathcal{E}_\delta(t) \quad \text{and} \quad \mu_i(t) = \frac{d_i(t)}{2|\mathcal{E}_\delta(t)|} \geqslant \frac{d_{\min,\delta}(t)}{2|\mathcal{E}_\delta(t)|} \text{ for } i\in[n]$$

It is also easy to see that

$$\pi_i^*(t)\bar{P}_{ij}(t) \geqslant \frac{1}{2d_\delta(t)nb^2(t)} \text{ for } \{i,j\} \in \mathcal{E}_\delta(t), \quad \text{and} \quad \pi_i^*(t) \leqslant \frac{b(t)}{n} \text{ for } i \in [n]$$

thus leading to the bounds

$$\alpha \geqslant \frac{|\mathcal{E}_\delta(t)|}{2d_\delta(t)nb^2(t)} \quad \text{and} \quad \beta \leqslant \frac{2b(t)|\mathcal{E}_\delta(t)|}{nd_{\min,\delta}(t)}.$$

Plugging this in (2.8.3) leads to

$$1-\lambda_{\max}(\bar{P}(t)) \geqslant \frac{\xi_\delta(t)|\mathcal{E}_\delta(t)|nd_{\min,\delta}(t)}{4d_\delta(t)nb^3(t)|\mathcal{E}_\delta(t)|} \geqslant \frac{\xi_\delta(t)d_{\min,\delta}(t)}{4d_\delta(t)b^3(t)} > 0$$

which together with (2.3.7) readily leads to the stated lower bound on $1-\rho(t)$.

□

2.8.4 *Proof of Lemma 6*

We will bound $\left\|\pi^*(t)^\top\Delta(t)\right\|_2$, using the same ideas as in the proof of [9, Theorem 9]. Recall that the entries of $\Delta(t)$ are given by

$$\Delta_{ij}(t) = \begin{cases} \frac{1}{d_\delta(t)|\mathcal{N}_{ij,\delta}(t)|}\sum_{t'\in\mathcal{N}_{ij,\delta}(t)} (y_{ij}(t') - y_{ij}^*(t')) & \text{if} \quad \{i,j\} \in \mathcal{E}_\delta(t), \\ -\sum_{k\neq i}\Delta_{ik}(t) & \text{if} \quad i=j, \\ 0 & \text{otherwise.} \end{cases}$$

Let $\Delta^l(t)$ denote the lower triangular part of $\Delta(t)$ and $\Delta^u(t)$ its upper triangular part, both having zeros on the diagonal. Let us define the diagonal matrices

$$\Delta^{\text{diag},l}(t) = -\operatorname{diag}\left(\sum_{j<i}\Delta_{ij}(t)\right)_{i=1\ldots n} \quad \text{and} \quad \Delta^{\text{diag},u}(t) = -\operatorname{diag}\left(\sum_{j>i}\Delta_{ij}(t)\right)_{i=1\ldots n}. \tag{2.8.4}$$

Note that both these matrices have independent entries. The matrix $\Delta(t)$ is now decomposed as

$$\Delta(t) = \Delta^l(t) + \Delta^u(t) + \Delta^{\text{diag},l}(t) + \Delta^{\text{diag},u}(t)$$

and the triangle inequality implies

$$\begin{aligned}\left\|\pi^*(t)^\top\Delta(t)\right\|_2 \leqslant &\underbrace{\left\|\pi^*(t)^\top\Delta^l(t)\right\|_2}_{=I_l} + \underbrace{\left\|\pi^*(t)^\top\Delta^u(t)\right\|_2}_{=I_u} + \underbrace{\left\|\pi^*(t)^\top\Delta^{\text{diag},l}(t)\right\|_2}_{=I_{\text{diag},l}} \\ &+ \underbrace{\left\|\pi^*(t)^\top\Delta^{\text{diag},u}(t)\right\|_2}_{=I_{\text{diag},u}}.\end{aligned} \tag{2.8.5}$$

The goal is to bound each of these terms separately, using the same method for all of them. Let us describe the steps for the first term $I_l = \left\|\pi^*(t)^\top \Delta^l(t)\right\|_2$. For all $j \in [n]$,

$$\left[\pi^*(t)^\top \Delta^l(t)\right]_j = \sum_{i=1}^{n} \pi_i^*(t)\Delta_{ij}^l(t) = \sum_{i>j} \pi_i^*(t)\Delta_{ij}(t).$$

For any pair $\{i,j\} \in \mathcal{E}_\delta(t)$, $\Delta_{ij}(t)$ is a sum of at most $LN_{\max,\delta}(t)$ independent centered random variables, thus $\left[\pi^*(t)^\top \Delta^l(t)\right]_j$ is a sum of at most $LN_{\max,\delta}(t)d_j^l(t)$ independent centered random variables, where $d_j^l(t)$ is defined as

$$d_j^l(t) := |\{i \in [n] \mid \{i,j\} \in \mathcal{E}_\delta(t) \text{ and } i > j\}|.$$

By applying Hoeffding's inequality we obtain for any $s > 0$,

$$\begin{aligned}\mathbb{P}(\left|\left[\pi^*(t)^\top \Delta^l(t)\right]_j\right| > s) &\leqslant 2\exp\left(-\frac{2L^2N_{\min,\delta}^2(t)d_\delta^2(t)s^2}{4LN_{\max,\delta}(t)d_j^l(t)\left\|\pi^*(t)\right\|_\infty^2}\right)\\ &\leqslant 2\exp\left(-\frac{LN_{\min,\delta}^2(t)d_\delta^2(t)s^2}{2N_{\max,\delta}(t)d_j^l(t)\left\|\pi^*(t)\right\|_\infty^2}\right).\end{aligned}$$

Hence, $\left[\pi^*(t)^\top \Delta^l(t)\right]_j$ can be seen as a sub-Gaussian random variable with sub-Gaussian norm at most $c_1\sigma_j$ with $c_1 > 0$ a constant, and $\sigma_j^2 = \frac{2d_j^l(t)N_{\max,\delta}(t)\|\pi^*(t)\|_\infty^2}{Ld_\delta^2(t)N_{\min,\delta}^2(t)}$. As $d_j^l(t) < d_{\max,\delta}(t)$, it holds that

$$\sigma_j^2 \leqslant 2\frac{N_{\max,\delta}(t)d_{\max,\delta}(t)}{Ld_\delta^2(t)N_{\min,\delta}^2(t)}\left\|\pi^*(t)\right\|_\infty^2$$

Denoting $\sigma^2 = 2c_1\frac{N_{\max,\delta}(t)d_{\max,\delta}(t)}{Ld_\delta^2(t)N_{\min,\delta}^2(t)}\left\|\pi^*(t)\right\|_\infty^2$, then we see that $\pi^*(t)^\top\Delta^l(t)$ is a random vector with independent, centered σ- sub-gaussian entries. Then,

$$I_l^2 = \left\|\pi^*(t)^\top \Delta^l(t)\right\|_2^2 = \sum_{j=1}^{n}[\pi^*(t)^\top \Delta^l(t)]_j^2$$

is a quadratic form of a sub-gaussian vector, so its expectation satisfies $\mathbb{E}[I_l^2] \leqslant c_2 n\sigma^2$ for some constant $c_2 > 0$.

Using the Hanson-Wright inequality [39, Theorem 1.1], we have for any $s > 0$ that

$$\mathbb{P}(|I_l^2 - \mathbb{E}[I_l^2]| > s) \leqslant 2\exp\left(-c_3\min\left(\frac{s^2}{\sigma^4 n}, \frac{s}{\sigma^2}\right)\right) = 2\exp\left(-c_3\frac{s}{\sigma^2}\min\left(\frac{s}{\sigma^2 n}, 1\right)\right)$$

for $c_3 > 0$ a constant. Choosing $s = \sqrt{\frac{C_1}{c_3}}\sigma^2\sqrt{n\log n}$ with $C_1 \geqslant 1$ a constant, note that $\frac{s}{n\sigma^2} = \sqrt{\frac{C_1}{c_3}}\sqrt{\frac{\log n}{n}} \leqslant 1$ if $n \geqslant \frac{C_1}{c_3}\log n$, and so,

$$\mathbb{P}\left(|I_l^2 - \mathbb{E}[I_l^2]| > \sqrt{\frac{C_1}{c_3}}\sigma^2\sqrt{n\log n}\right) \leqslant 2\exp^{-C_1\log n} = 2n^{-C_1}.$$

Then, it holds with probability atleast $1 - O(n^{-C_1})$ that

$$\begin{aligned}
I_l^2 &\leqslant \mathbb{E}[I_l^2] + \sigma^2\sqrt{n\log n} \\
&\leqslant c_2 n\sigma^2 + \sqrt{\frac{C_1}{c_3}}\sigma^2\sqrt{n\log n} \\
&\leqslant n\sigma^2(c_2 + \sqrt{\frac{C_1}{c_3}}\sqrt{\frac{\log n}{n}}) \\
&\leqslant C_2\frac{nN_{\max,\delta}(t)d_{\max,\delta}(t)}{Ld_\delta^2(t)N^2_{\min,\delta}(t)}\|\pi^*(t)\|_\infty^2 \\
&\leqslant C_2\frac{N_{\max,\delta}(t)d_{\max,\delta}(t)b^2(t)}{Ld_\delta^2(t)N^2_{\min,\delta}(t)}\|\pi^*(t)\|_2^2
\end{aligned}$$

as $\|\pi^*(t)\|_2^2 \geqslant n\pi^*_{\min}(t)^2 \geqslant \frac{n}{b^2(t)}\pi^*_{\max}(t)^2 = \frac{n}{b^2(t)}\|\pi^*(t)\|_\infty^2$.

The same analysis leads to the same bound for $I_u^2, I^2_{diag,l}, I^2_{diag,u}$. Thus for any constant $C_1 \geqslant 1$, there exists a constant $C_2 \geqslant 1$ such that it holds with probability greater than $1 - O(n^{-C_1})$ that

$$\left\|\pi^*(t)^\top\Delta(t)\right\|_2 \leqslant C_2 b(t)\sqrt{\frac{N_{\max,\delta}(t)d_{\max,\delta}(t)}{Ld_\delta^2(t)N^2_{\min,\delta}(t)}}\|\pi^*(t)\|_2\,. \tag{2.8.6}$$

2.9 PROOFS OF RESULTS IN SECTION 2.4

The proofs are largely inspired from the proofs in [9, Section C], which are adapted to our problem setup.

2.9.1 *Proof of Lemma 7*

Consider any given collection of graphs $(G_{t'})_{t'\in\mathcal{N}_\delta(t)}$. Then following the same ideas as in [9, Lemma 2], we have for any given $m \in [n]$ that

$$\begin{aligned}
I_0^m &= \sum_{j=1}^n \pi_j^*(t)(\hat{P}_{jm}(t) - \hat{P}^*_{jm}(t)) \\
&= \sum_{j:j\neq m}\pi_j^*(t)(\hat{P}_{jm}(t) - \hat{P}^*_{jm}(t)) + \pi^*_m(t)(\hat{P}_{mm}(t) - \hat{P}^*_{mm}(t)) \\
&= \sum_{j:j\neq m}(\pi_j^*(t) + \pi_m^*(t))(\hat{P}_{jm}(t) - \hat{P}^*_{jm}(t)) \\
&= \sum_{j:j\neq m}(\pi_j^*(t) + \pi_m^*(t))\sum_{t'\in\mathcal{N}_{jm,\delta}(t)}\sum_{l=1}^L\left(\frac{y_{jm}^{(l)}(t') - y^*_{jm}(t')}{Ld_\delta(t)|\mathcal{N}_{jm,\delta}(t)|}\right).
\end{aligned}$$

Applying Hoeffding's inequality, it holds for any constant $C_1 \geqslant 2$,

$$\mathbb{P}(|I_0^m| > C_1\sqrt{\frac{N_{\max,\delta}(t)d_{\max,\delta}(t)\log n}{Ld_\delta^2(t)N^2_{\min,\delta}(t)}}\|\pi^*(t)\|_\infty) \leqslant 2n^{-\frac{c_1^2}{2}}.$$

If $G_{t'} \sim \mathcal{G}(n, p(t'))$ for all $t' \in \mathcal{N}_\delta(t)$, then using Lemma 21, it holds with probability at least $1 - O(n^{-10})$ that $\frac{N_{\max,\delta}(t)d_{\max,\delta}(t)}{d_\delta^2(t)N^2_{\min,\delta}(t)} \leqslant \frac{4}{3np_\delta(t)p_{\delta,sum}(t)}$.

Hence after a union bound over $[n]$, we conclude that there exists a suitably large constant $C_2 \geqslant 1$ such that

$$\mathbb{P}\left(\forall m \in [n] : |I_0^m| > C_2 \sqrt{\frac{\log n}{Lnp_\delta(t)p_{\delta,sum}(t)}} \|\pi^*(t)\|_\infty\right) \leqslant O(n^{-9}).$$

□

2.9.2 *Proof of Lemma 8*

For any given collection of graphs $(G_{t'})_{t' \in \mathcal{N}_\delta(t)}$, recall that the entries of $\hat{P}^*(t) - \bar{P}(t)$ have already been bounded in the proof of Lemma 2 as follows

$$\left|\hat{P}^*_{ij}(t) - \bar{P}_{ij}(t)\right| \leqslant \begin{cases} \frac{M\delta}{Td_\delta(t)} & \text{if } i \neq j, \{i,j\} \in \mathcal{E}_\delta(t) \\ \frac{M\delta}{T} & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases} \tag{2.9.1}$$

This implies for all $m \in [n]$ that

$$|I_1^m| = \left|\sum_{j:j\neq m} \pi_j^*(t)(\hat{P}^*_{jm}(t) - \bar{P}_{jm}(t)\right| \leqslant \|\pi^*(t)\|_\infty \left(d_{max,\delta}(t)\frac{M\delta}{Td_\delta(t)} + \frac{M\delta}{T}\right).$$

Due to Lemma 21, we know that $3np_\delta(t) = d_\delta(t) \geqslant d_{max,\delta}(t)$ holds with probability at least $1 - O(n^{-10})$. This leads to the statement in Lemma 8.

□

2.9.3 *Proof of Lemma 9*

We will follow the steps in [9, Lemma 3]. Recalling that $I_2^m = (\hat{\pi}_m(t) - \pi_m^*(t))\,\hat{P}_{mm}(t)$, let us bound $\hat{P}_{mm}(t)$ using the decomposition

$$\hat{P}_{mm}(t) = \left(\hat{P}_{mm}(t) - \hat{P}^*_{mm}(t)\right) + \hat{P}^*_{mm}(t).$$

For any given collection of graphs $(G_{t'})_{t' \in \mathcal{N}_\delta(t)}$, note that

$$\hat{P}_{mm}(t) - \hat{P}^*_{mm}(t) = -\frac{1}{Ld_\delta(t)} \sum_{j:j\neq m} \sum_{t' \in \mathcal{N}_{jm,\delta}(t)} \sum_{l=1}^{L} \frac{1}{|\mathcal{N}_{jm,\delta}(t)|} \left(y_{mj}^{(l)}(t') - y^*_{mj}(t')\right)$$

is a sum of independent, centered random variables. Then Hoeffding's inequality implies for each $m \in [n]$,

$$\mathbb{P}\left(\left|\hat{P}_{mm}(t) - \hat{P}^*_{mm}(t)\right| > C_1 \sqrt{\frac{N_{max,\delta}(t) d_{max,\delta}(t) \log n}{Ld_\delta^2(t) N^2_{min,\delta}(t)}}\right) \leqslant 2n^{-\frac{C_1^2}{2}}$$

for any constant $C_1 \geqslant 2$.

If $G_{t'} \sim \mathcal{G}(n, p(t'))$ for all $t' \in \mathcal{N}_\delta(t)$, then once again combining this result with Lemma 21, we see (after a union bound over $[n]$) that there exists a suitably large constant $C_2 \geqslant 1$ such that with probability at least $1-O(n^{-9})$,

$$\left|\hat{P}_{mm}(t) - \hat{P}^*_{mm}(t)\right| \leqslant C_2\sqrt{\frac{\log n}{Lnp_\delta(t)p_{\delta,sum}(t)}}, \quad \forall m \in [n].$$

Now we bound $\left|\hat{P}^*_{mm}(t)\right|$. Recalling that $b_{\max,\delta}(t) = \max_{t'\in\mathcal{N}_\delta(t)} b(t')$ and $b(t') = \max_{i,j\in[n]} \frac{w^*_{t',i}}{w^*_{t',j}}$, we have for any given collection of graphs $(G_{t'})_{t'\in\mathcal{N}_\delta(t)}$ that all $m \in [n]$,

$$\begin{aligned}
\hat{P}^*_{mm}(t) &= 1 - \sum_{j:j\neq m} \frac{1}{d_\delta(t)|\mathcal{N}_{jm,\delta}(t)|} \sum_{t'\in\mathcal{N}_{jm,\delta}(t)} y^*_{mj}(t') \\
&= 1 - \sum_{j:j\neq m}\sum_{t'\in\mathcal{N}_{jm,\delta}(t)} \frac{1}{d_\delta(t)|\mathcal{N}_{jm,\delta}(t)|}\left(\frac{w^*_{t'j}}{w^*_{t'j}+w^*_{t'm}}\right) \\
&\leqslant 1 - \sum_{j:j\neq m}\sum_{t'\in\mathcal{N}_{jm,\delta}(t)} \frac{1}{d_\delta(t)|\mathcal{N}_{jm,\delta}(t)|}\left(\frac{1}{1+b(t')}\right) \\
&\leqslant 1 - \frac{d_{\min,\delta}(t)}{d_\delta(t)}\frac{1}{2b_{\max,\delta}(t)}.
\end{aligned}$$

If $G_{t'} \sim \mathcal{G}(n, p(t'))$ for all $t' \in \mathcal{N}_\delta(t)$, then we have using Lemma 21 that $d_{\min,\delta}(t) \geqslant np_\delta(t)/2$ with probability at least $1-O(n^{-10})$, this in turn implies that $\left|\hat{P}^*_{mm}(t)\right| \leqslant 1 - \frac{1}{12b_{\max,\delta}(t)}$ for all $m \in [n]$.

Hence, we conclude that there exists a suitably large constant $C_2 \geqslant 1$ such that with probability at least $1-O(n^{-9})$,

$$|I_2^m| \leqslant \left(1 - \frac{1}{12b_{\max,\delta}(t)} + C_2\sqrt{\frac{\log n}{Lnp_\delta(t)p_{\delta,sum}(t)}}\right)\|\hat{\pi}(t)-\pi^*(t)\|_\infty, \quad \forall m \in [n].$$

□

2.9.4 *Proof of Lemma 10*

This proof follows the lines of the proof of [9, Lemma 4]. Consider any given collection of graphs $(G_{t'})_{t'\in\mathcal{N}_\delta(t)}$. Using the properties of $\|.\|_{\pi^*(t)}$ in (2.1.1) and Theorem 11 with the matrices $\hat{P}(t), \hat{P}^{(m)}(t), \bar{P}(t)$, it holds for all $m \in [n]$ that

$$\begin{aligned}
\left\|\hat{\pi}(t) - \hat{\pi}^{(m)}(t)\right\|_2 &\leqslant \frac{\left\|\hat{\pi}^{(m)}(t)^\top(\hat{P}^{(m)}(t) - \hat{P}(t))\right\|_{\pi^*(t)}}{\sqrt{\pi^*_{\min}(t)}\left(1-\lambda_{\max}(\bar{P}(t)) - \left\|\hat{P}(t)-\bar{P}(t)\right\|_{\pi^*(t)}\right)} \\
&\leqslant \sqrt{\frac{\pi^*_{\max}(t)}{\pi^*_{\min}(t)}}\left(\frac{8d_\delta(t)b^3(t)}{\xi_\delta(t)d_{\min,\delta}(t)}\right)\left\|\hat{\pi}^{(m)}(t)^\top(\hat{P}^{(m)}(t)-\hat{P}(t))\right\|_2
\end{aligned} \tag{2.9.2}$$

where we used (2.3.8).

In order to bound $\left\|\hat{\pi}^{(m)}(t)^\top\left(\hat{P}^{(m)}(t)-\hat{P}(t)\right)\right\|_2$, we need to get rid of the statistical dependency between $\hat{P}^{(m)}(t)$ and $\hat{\pi}^{(m)}(t)$. To do so, let us introduce $\hat{P}^{(m),\mathcal{G}}(t)$, defined for $i \neq j$ as

$$\hat{P}_{ij}^{(m),\mathcal{G}}(t) = \begin{cases} \hat{P}_{ij}(t) & \text{if } i \neq m, j \neq m \\ \frac{1}{d_\delta(t)} \frac{w^*_{t,j}}{w^*_{t,j}+w^*_{t,i}} \mathbb{1}_{\{i,j\}\in\mathcal{E}_\delta(t)} & \text{if } i = m \text{ or } j = m. \end{cases}$$

Its diagonal entries are defined as

$$\hat{P}_{ii}^{(m),\mathcal{G}}(t) = 1 - \sum_{k:k\neq i} \hat{P}_{ik}^{(m),\mathcal{G}}(t)$$

to ensure that $\hat{P}^{(m),\mathcal{G}}(t)$ is a transition matrix. Note that the m^{th} line and columns of $\hat{P}^{(m)}(t)$ are the expectations of those in $\hat{P}^{(m),\mathcal{G}}(t)$. Moreover, $\hat{P}_{.m}^{(m),\mathcal{G}}(t) = \bar{P}_{.m}(t)$ and $\hat{P}_{m.}^{(m),\mathcal{G}}(t) = \bar{P}_{m.}(t)$. Now starting with the decomposition

$$\hat{\pi}^{(m)}(t)^\top(\hat{P}^{(m)}(t)-\hat{P}(t)) = \underbrace{\hat{\pi}^{(m)}(t)^\top\left(\hat{P}(t)-\hat{P}^{(m),\mathcal{G}}(t)\right)}_{J_1^m} + \underbrace{\hat{\pi}^{(m)}(t)^\top\left(\hat{P}^{(m),\mathcal{G}}(t)-\hat{P}^{(m)}(t)\right)}_{J_2^m}$$

we will separately bound $\|J_1^m\|_2$, $\|J_2^m\|_2$.

2.9.4.1 *Bound on* $\|J_1^m\|_2$

Consider any given collection of graphs $(G_{t'})_{t'\in\mathcal{N}_\delta(t)}$. Since $\hat{P}_{.m}^{(m),\mathcal{G}}(t) = \bar{P}_{.m}(t)$, we have the decomposition

$$\begin{aligned} J_{1,m}^m &= \sum_{j=1}^n \hat{\pi}_j^{(m)}(t)\left(\hat{P}_{jm}(t)-\bar{P}_{jm}(t)\right) \\ &= \sum_{j:j\neq m} \hat{\pi}_j^{(m)}(t)\left(\hat{P}_{jm}(t)-\bar{P}_{jm}(t)\right) + \hat{\pi}_m^{(m)}(t)\left(\hat{P}_{mm}(t)-\bar{P}_{mm}(t)\right) \\ &= \sum_{j:j\neq m} (\hat{\pi}_j^{(m)}(t)+\hat{\pi}_m^{(m)}(t))\left(\hat{P}_{jm}(t)-\bar{P}_{jm}(t)\right) \\ &= \sum_{j:j\neq m} (\hat{\pi}_j^{(m)}(t)+\hat{\pi}_m^{(m)}(t))\left(\hat{P}_{jm}(t)-\hat{P}^*_{jm}(t)\right) + \sum_{j:j\neq m} (\hat{\pi}_j^{(m)}(t)+\hat{\pi}_m^{(m)}(t))\left(\hat{P}^*_{jm}(t)-\bar{P}_{jm}(t)\right). \end{aligned}$$

Since $\hat{\pi}^{(m)}(t)$ is independent of $(\hat{P}_{jm}(t)-\hat{P}^*_{jm}(t))_{j\neq m}$, the first term can be bounded using Hoeffding's inequality as

$$\mathbb{P}\left(\left|\sum_{j:j\neq m} (\hat{\pi}_j^{(m)}(t)+\hat{\pi}_m^{(m)}(t))\left(\hat{P}_{jm}(t)-\hat{P}^*_{jm}(t)\right)\right| > C_1\left\|\hat{\pi}^{(m)}(t)\right\|_\infty \sqrt{\frac{N_{\max,\delta}(t)d_{\max,\delta}(t)\log n}{Ld_\delta^2(t)N^2_{\min,\delta}(t)}}\right) \leqslant 2n^{-\frac{C_1^2}{2}}$$

for any constant $C_1 \geqslant 2$. Moreover, due to (2.9.1), we have for all m that

$$\sum_{j:j\neq m} (\hat{\pi}_j^{(m)}(t)+\hat{\pi}_m^{(m)}(t))\left(\hat{P}^*_{jm}(t)-\bar{P}_{jm}(t)\right) \leqslant 2d_{\max,\delta}(t)\left(\frac{M\delta}{Td_\delta(t)}\right)\left\|\hat{\pi}^{(m)}(t)\right\|_\infty.$$

Hence for any given $m \in [n]$, it holds with probability at least $1 - O(n^{-\frac{c_1^2}{2}})$ that

$$|J_{1,m}^m| \leqslant \left(\frac{2d_{\max,\delta}(t)M\delta}{Td_\delta(t)} + C_1\sqrt{\frac{N_{\max,\delta}(t)d_{\max,\delta}(t)\log n}{Ld_\delta^2(t)N_{\min,\delta}^2(t)}} \right) \left\| \hat{\pi}^{(m)}(t) \right\|_\infty . \tag{2.9.3}$$

Let us now bound the other coefficients of J_1^m. For any $j \neq m$,

$$\begin{aligned} J_{1,j}^m &= \sum_{i:i\neq j} \hat{\pi}_i^{(m)}(t)(\hat{P}_{ij}(t) - \hat{P}_{ij}^{(m),\mathcal{G}}(t)) + \hat{\pi}_j^{(m)}(t)\left(\hat{P}_{jj}(t) - \hat{P}_{jj}^{(m),\mathcal{G}}(t)\right) \\ &= \hat{\pi}_m^{(m)}(t)\left(\hat{P}_{mj}(t) - \hat{P}_{mj}^{(m),\mathcal{G}}(t)\right) + \hat{\pi}_j^{(m)}(t)\left(\hat{P}_{jj}(t) - \hat{P}_{jj}^{(m),\mathcal{G}}(t)\right) \\ &= (\hat{\pi}_m^{(m)}(t) + \hat{\pi}_j^{(m)}(t))\left(\hat{P}_{mj}(t) - \hat{P}_{mj}^*(t) + \hat{P}_{mj}^*(t) - \bar{P}_{mj}(t)\right) \\ \implies \left|J_{1,j}^m\right| &\leqslant 2\left\|\hat{\pi}^{(m)}(t)\right\|_\infty \left(\left|\hat{P}_{mj}(t) - \hat{P}_{mj}^*(t)\right| + \frac{M\delta}{Td_\delta(t)} \right) \end{aligned}$$

where we used (2.9.1) and the fact that the m^{th} rows of $\hat{P}^{(m),\mathcal{G}}(t)$ and $\bar{P}(t)$ are identical. Using Hoeffding's inequality and a union bound, we have

$$\mathbb{P}\left(\max_{j\neq m} \left|\hat{P}_{mj}(t) - \hat{P}_{mj}^*(t)\right| > C_2\sqrt{\frac{N_{\max,\delta}(t)\log n}{Ld_\delta^2(t)N_{\min,\delta}^2(t)}} \right) \leqslant 2n^{1-\frac{c_2^2}{2}}$$

for any constant $C_2 \geqslant 2$. Hence, for any $j \neq m$, we have with probability at least $1 - O(n^{1-\frac{c_2^2}{2}})$ that

$$\left|J_{1,j}^m\right| \leqslant \begin{cases} 2\left\|\hat{\pi}^{(m)}(t)\right\|_\infty \left(\frac{M\delta}{Td_\delta(t)} + C_2\sqrt{\frac{N_{\max,\delta}(t)\log n}{Ld_\delta^2(t)N_{\min,\delta}^2(t)}} \right) & \text{if } \{j,m\} \in \mathcal{E}_\delta(t) \\ 0 & \text{otherwise.} \end{cases} \tag{2.9.4}$$

Finally combining (2.9.3), (2.9.4) along with Lemma 21, we see after taking a union bound over $[n]$ that when $G_{t'} \sim \mathcal{G}(n, p(t'))$ for all $t' \in \mathcal{N}_\delta(t)$, then there exists a suitably large constant $C_3 \geqslant 1$ such that with probability at least $1 - O(n^{-9})$,

$$\|J_1^m\|_2 \leqslant \left(\frac{4M\delta}{T} + C_3\sqrt{\frac{\log n}{Lnp_\delta(t)p_{\delta,sum}(t)}} \right) \left\|\hat{\pi}^{(m)}(t)\right\|_\infty \quad \forall m \in [n]. \tag{2.9.5}$$

2.9.4.2 *Bound on* $\left\|J_2^m\right\|_2$

Consider again any given collection of graphs $(G_{t'})_{t'\in\mathcal{N}_\delta(t)}$. Note that the off-diagonal entries of $\hat{P}^{(m)}(t) - \hat{P}^{(m),\mathcal{G}}(t)$ are non zero if and only if they belong to the m^{th} row /column. Then for $i \neq j$ with $i = m$ or $j = m$, we have

$$\left(\hat{P}^{(m)}(t) - \hat{P}^{(m),\mathcal{G}}(t)\right)_{ij} = \frac{1}{d_\delta(t)}\left(p_\delta(t) - \mathbf{1}_{\{i,j\}\in\mathcal{E}_\delta(t)}\right) y_{ij}^*(t)$$

Moreover, the diagonal entries are given as

$$\left(\hat{P}^{(m)}(t) - \hat{P}^{(m),\mathcal{G}}(t)\right)_{ii} = \begin{cases} -\frac{1}{d_\delta(t)} \sum_{j:j\neq m} \left(p_\delta(t) - \mathbf{1}_{\{m,j\}\in\mathcal{E}_\delta(t)}\right) y^*_{mj}(t) & \text{if } i = m \\ -\frac{1}{d_\delta(t)} \left(p_\delta(t) - \mathbf{1}_{\{i,m\}\in\mathcal{E}_\delta(t)}\right) y^*_{im}(t) & \text{if } i \neq m. \end{cases}$$

Let us first show that $\pi^*(t)^\top \left(\hat{P}^{(m)}(t) - \hat{P}^{(m),\mathcal{G}}(t)\right) = 0$ by considering the cases below.

1. For any $j \neq m$,

$$\begin{aligned} \left[\pi^*(t)^\top \left(\hat{P}^{(m)}(t) - \hat{P}^{(m),\mathcal{G}}(t)\right)\right]_j =& \frac{\pi^*_m(t)}{d_\delta(t)}(p_\delta(t) - \mathbf{1}_{\{m,j\}\in\mathcal{E}_\delta(t)})y^*_{mj}(t) \\ &- \frac{\pi^*_j(t)}{d_\delta(t)}(p_\delta(t) - \mathbf{1}_{\{j,m\}\in\mathcal{E}_\delta(t)})y^*_{jm}(t) \\ =& \frac{p_\delta(t) - \mathbf{1}_{\{m,j\}\in\mathcal{E}_\delta(t)}}{d_\delta(t)} \left(\pi^*_m(t)y^*_{mj}(t) - \pi^*_j(t)y^*_{jm}(t)\right) \\ =& 0. \end{aligned}$$

   This equality comes from the definition of $\pi^*(t)$ and $y^*_{jm}(t)$.

2. For $j = m$,

$$\begin{aligned} &\left[\pi^*(t)^\top \left(\hat{P}^{(m)}(t) - \hat{P}^{(m),\mathcal{G}}(t)\right)\right]_m \\ =& -\frac{\pi^*_m(t)}{d_\delta(t)} \sum_{k:k\neq m} (p_\delta(t) - \mathbf{1}_{\{m,k\}\in\mathcal{E}_\delta(t)})y^*_{mk}(t) + \frac{1}{d_\delta(t)} \sum_{k:k\neq m} \pi^*_k(t)(p_\delta(t) - \mathbf{1}_{\{k,m\}\in\mathcal{E}_\delta(t)})y^*_{km}(t) \\ =& \frac{1}{d_\delta(t)} \sum_{k:k\neq m} \left(p_\delta(t) - \mathbf{1}_{\{k,m\}\in\mathcal{E}_\delta(t)}\right)(\pi^*_k(t)y^*_{km}(t) - \pi^*_m(t)y^*_{mk}(t)) \\ =& 0. \end{aligned}$$

Hence, $\pi^*(t)^\top \left(\hat{P}^{(m)}(t) - \hat{P}^{(m),\mathcal{G}}(t)\right) = 0$ and it holds that

$$J_2^m = \left(\hat{\pi}^{(m)}(t) - \pi^*(t)\right)^\top \left(\hat{P}^{(m)}(t) - \hat{P}^{(m),\mathcal{G}}(t)\right).$$

Now for $j \neq m$,

$$\begin{aligned} \left|J^m_{2,j}\right| =& \left|\left(\hat{\pi}^{(m)}_m(t) - \pi^*_m(t)\right)\left(\frac{p_\delta(t) - \mathbf{1}_{\{m,j\}\in\mathcal{E}_\delta(t)}}{d_\delta(t)}\right) y^*_{mj}(t) - \left(\hat{\pi}^{(m)}_j(t) - \pi^*_j(t)\right)\left(\frac{p_\delta(t) - \mathbf{1}_{\{j,m\}\in\mathcal{E}_\delta(t)}}{d_\delta(t)}\right) y^*_{jm}(t)\right| \\ \leqslant& \begin{cases} \frac{4}{d_\delta(t)} \left\|\hat{\pi}^{(m)}(t) - \pi^*(t)\right\|_\infty & \text{if } \{j,m\} \in \mathcal{E}_\delta(t) \\ 2\frac{p_\delta(t)}{d_\delta(t)} \left\|\hat{\pi}^{(m)}(t) - \pi^*(t)\right\|_\infty & \text{if } \{j,m\} \notin \mathcal{E}_\delta(t). \end{cases} \end{aligned}$$

Moreover, we can bound $\left|J^m_{2,m}\right|$ as follows.

$$\begin{aligned}\left|J^m_{2,m}\right| &= \Bigg| -\left(\hat{\pi}^{(m)}_m(t) - \pi^*_m(t)\right) \sum_{j:j\neq m} \frac{1}{d_\delta(t)} \left(p_\delta(t) - \mathbf{1}_{\{m,j\}\in\mathcal{E}_\delta(t)}\right) y^*_{mj}(t) \\ &\quad + \sum_{j:j\neq m} \left(\hat{\pi}^{(m)}_j(t) - \pi^*_j(t)\right) \frac{1}{d_\delta(t)} \left(p_\delta(t) - \mathbf{1}_{\{j,m\}\in\mathcal{E}_\delta(t)}\right) y^*_{jm}(t)\Bigg| \\ &\leqslant |J^m_3| + |J^m_4|\end{aligned}$$

where

$$J^m_3 = \sum_{j:j\neq m} \left(\hat{\pi}^{(m)}_m(t) - \pi^*_m(t)\right) \frac{1}{d_\delta(t)} \left(p_\delta(t) - \mathbf{1}_{\{m,j\}\in\mathcal{E}_\delta(t)}\right) y^*_{mj}(t),$$

and

$$J^m_4 = \sum_{j:j\neq m} \left(\hat{\pi}^{(m)}_j(t) - \pi^*_j(t)\right) \frac{1}{d_\delta(t)} \left(p_\delta(t) - \mathbf{1}_{\{j,m\}\in\mathcal{E}_\delta(t)}\right) y^*_{jm}(t).$$

When $G_{t'} \sim \mathcal{G}(n, p(t'))$ for all $t' \in \mathcal{N}_\delta(t)$, then denoting $\xi^{(m)}_j = \left(\hat{\pi}^{(m)}_m(t) - \pi^*_m(t)\right) \frac{y^*_{mj}(t)}{d_\delta(t)}$, a bound on J^m_3 can be given by Bernstein's inequality, introducing the random variables $Z^{(m)}_j = \xi^{(m)}_j (p_\delta(t) - \mathbf{1}_{\{j,m\}\in\mathcal{E}_\delta(t)})$. Since for all $j \neq m$,

$$\left|Z^{(m)}_j\right| \leqslant \frac{1}{d_\delta(t)} \left\|\hat{\pi}^{(m)}(t) - \pi^*(t)\right\|_\infty, \quad \mathbb{E}[Z^{(m)}_j]^2 \leqslant \frac{\left\|\hat{\pi}^{(m)}(t) - \pi^*(t)\right\|^2_\infty}{d^2_\delta(t)} p_\delta(t)$$

hence there exists a constant $c \geqslant 2$ such that it holds with probability at least $1 - 2n^{-3c/2}$ that

$$|J^m_3| \leqslant c \frac{\sqrt{np_\delta(t)\log n} + \log n}{d_\delta(t)} \left\|\hat{\pi}^{(m)}(t) - \pi^*(t)\right\|_\infty. \tag{2.9.6}$$

The same bound holds for $\left|J^m_4\right|$. Using Lemma 21, we then see that there exists a suitably large constant $C_4 \geqslant 1$ such that with probability at least $1 - O(n^{-10})$,

$$\|J^m_2\|_2 \leqslant \left(C_4 \frac{\sqrt{np_\delta(t)\log n} + \log n}{d_\delta(t)} + \frac{4\sqrt{d_\delta(t)}}{d_\delta(t)} + \frac{2p_\delta(t)\sqrt{n}}{d_\delta(t)}\right) \left\|\hat{\pi}^{(m)}(t) - \pi^*(t)\right\|_\infty.$$

Taking the union bound over $[n]$ and performing a minor simplification of the previous bound, we finally observe that with with probability at least $1 - O(n^{-9})$,

$$\|J^m_2\|_2 \leqslant C_5 \sqrt{\frac{\log n}{np_\delta(t)}} \left\|\hat{\pi}^{(m)}(t) - \pi^*(t)\right\|_\infty \quad \forall m \in [n]. \tag{2.9.7}$$

for some suitably large constant $C_5 \geqslant 1$.

2.9.4.3 *Putting it together*

Using Lemma 21, we know that with probability at least $1 - O(n^{-10})$,

$$\sqrt{\frac{\pi^*_{\max}(t)}{\pi^*_{\min}(t)}}\left(\frac{8d_\delta(t)b^3(t)}{\xi_\delta(t)d_{\min,\delta}(t)}\right) \leqslant 96b^{\frac{7}{2}}(t). \tag{2.9.8}$$

Finally, combining (2.9.2), (2.9.5), (2.9.7) and (2.9.8), we have that with probability at least $1 - O(n^{-9})$, it holds for all $m \in [n]$,

$$\begin{aligned}
\left\|\hat{\pi}^{(m)}(t) - \hat{\pi}(t)\right\|_2 \leqslant& 96b^{\frac{5}{2}}(t)(\|J_1^m\|_2 + \|J_2^m\|_2) \\
\leqslant& 96b^{\frac{5}{2}}(t)\left(\frac{4M\delta}{T} + C_3\sqrt{\frac{\log n}{Lnp_\delta(t)p_{\delta,sum}(t)}}\right)\left\|\hat{\pi}^{(m)}(t)\right\|_\infty \\
&+ 96C_5 b^{\frac{5}{2}}(t)\sqrt{\frac{\log n}{np_\delta(t)}}\left\|\hat{\pi}^{(m)}(t) - \pi^*(t)\right\|_\infty \\
\leqslant& 96b^{\frac{5}{2}}(t)\left(\frac{4M\delta}{T} + C_3\sqrt{\frac{\log n}{Lnp_\delta(t)p_{\delta,sum}(t)}}\right)\|\pi^*(t)\|_\infty \\
&+ 96b^{\frac{5}{2}}(t)\left(\frac{4M\delta}{T} + C_3\sqrt{\frac{\log n}{Lnp_\delta(t)p_{\delta,sum}(t)}} + C_5\sqrt{\frac{\log n}{np_\delta(t)}}\right) \\
&\times \left\|\hat{\pi}^{(m)}(t) - \pi^*(t)\right\|_\infty \\
\leqslant& 96b^{\frac{5}{2}}(t)\left(\frac{4M\delta}{T} + C_3\sqrt{\frac{\log n}{Lnp_\delta(t)p_{\delta,sum}(t)}}\right)\|\pi^*(t)\|_\infty \\
&+ 96b^{\frac{5}{2}}(t)\left(\frac{4M\delta}{T} + C_6\sqrt{\frac{\log n}{np_\delta(t)}}\right)\left\|\hat{\pi}^{(m)}(t) - \pi^*(t)\right\|_\infty
\end{aligned}$$

for some suitably large constant $C_6 \geqslant 1$. Condition (2.4.4) further implies that for all $m \in [n]$,

$$\left\|\hat{\pi}^{(m)}(t) - \hat{\pi}(t)\right\|_2 \leqslant 96b^{\frac{5}{2}}(t)\left(\frac{4M\delta}{T} + C_3\sqrt{\frac{\log n}{Lnp_\delta(t)p_{\delta,sum}(t)}}\right)\|\pi^*(t)\|_\infty + \frac{1}{2}\left\|\hat{\pi}^{(m)}(t) - \pi^*(t)\right\|_\infty.$$

Finally, the triangular inequality

$$\left\|\hat{\pi}^{(m)}(t) - \pi^*(t)\right\|_\infty \leqslant \left\|\hat{\pi}^{(m)}(t) - \hat{\pi}(t)\right\|_2 + \|\hat{\pi}(t) - \pi^*(t)\|_\infty$$

implies that

$$\begin{aligned}
&\left\|\hat{\pi}^{(m)}(t) - \hat{\pi}(t)\right\|_2 \\
\leqslant& 96b^{\frac{5}{2}}(t)\left(\frac{4M\delta}{T} + C_3\sqrt{\frac{\log n}{Lnp_\delta(t)p_{\delta,sum}(t)}}\right)\|\pi^*(t)\|_\infty + \frac{1}{2}\left(\left\|\hat{\pi}^{(m)}(t) - \hat{\pi}(t)\right\|_2 + \|\hat{\pi}(t) - \pi^*(t)\|_\infty\right) \\
\leqslant& 192b^{\frac{5}{2}}(t)\left(\frac{4M\delta}{T} + C_3\sqrt{\frac{\log n}{Lnp_\delta(t)p_{\delta,sum}(t)}}\right)\|\pi^*(t)\|_\infty + \|\hat{\pi}(t) - \pi^*(t)\|_\infty.
\end{aligned}$$

2.9.5 *Proof of Lemma 11*

Since we follow the same steps as in [8], let us introduce some new notations in our particular setup. For all $t \in \mathcal{T}$, all $i \neq j$ and all $l \in [L]$, let us denote

$$\tilde{y}_{ij}^{(l)}(t) \sim \mathcal{B}\left(y_{ij}^*(t)\right) \quad \text{and} \quad \tilde{y}_{ij}(t) = \frac{1}{L}\sum_{l=1}^{L} \tilde{y}_{ij}^{(l)}(t). \tag{2.9.9}$$

so for all $t' \in \mathcal{N}_\delta(t)$, we have

$$y_{ij}(t') = \tilde{y}_{ij}(t')\mathbb{1}_{\{i,j\}\in\mathcal{E}_{t'}} \leqslant \tilde{y}_{ij}(t')\mathbb{1}_{\{i,j\}\in\mathcal{E}_\delta(t)}.$$

Then $\left|I_4^m\right|$ can be bounded as

$$\begin{aligned}
|I_4^m| &= \left|\sum_{j:j\neq m}\left(\hat{\pi}_j^{(m)}(t) - \pi_j^*(t)\right)\frac{1}{Ld_\delta(t)|\mathcal{N}_{jm,\delta}(t)|}\sum_{t'\in\mathcal{N}_{jm,\delta}(t)}\sum_{l=1}^{L} y_{jm}^{(l)}(t')\right| \\
&\leqslant \sum_{j:j\neq m}\left|\hat{\pi}_j^{(m)}(t) - \pi_j^*(t)\right|\left[\frac{1}{Ld_\delta(t)|\mathcal{N}_{jm,\delta}(t)|}\sum_{t'\in\mathcal{N}_{jm,\delta}(t)}\sum_{l=1}^{L}\tilde{y}_{jm}^{(l)}(t')\right]\mathbb{1}_{\{j,m\}\in\mathcal{E}_\delta(t)} \\
&\leqslant \frac{N_{\max,\delta}(t)}{N_{\min,\delta}(t)}\sum_{j:j\neq m}\frac{1}{d_\delta(t)}\left|\hat{\pi}_j^{(m)}(t) - \pi_j^*(t)\right|\mathbb{1}_{\{j,m\}\in\mathcal{E}_\delta(t)}.
\end{aligned}$$

When $G_{t'} \sim \mathcal{G}(n, p(t'))$ for all $t' \in \mathcal{N}_\delta(t)$, then we know from Lemma 21 that

$$\frac{N_{\max,\delta}(t)}{N_{\min,\delta}(t)} \leqslant 4 \tag{2.9.10}$$

holds with probability at least $1 - O(n^{-10})$. Hence, to bound $\left|I_4^m\right|$, it suffices to bound

$$I_5^m := \sum_{j:j\neq m}\frac{1}{d_\delta(t)}\left|\hat{\pi}_j^{(m)}(t) - \pi_j^*(t)\right|\mathbb{1}_{\{j,m\}\in\mathcal{E}_\delta(t)}.$$

To this end, let us first denote $G_\delta^{(m)}(t)$ to be the union graph $G_\delta(t)$ where the item m has been removed, and $\tilde{y}(t)$ to be the variables computed in (2.9.9). Then we have by triangle inequality that

$$I_5^m \leqslant \mathbb{E}[I_5^m|G_\delta^{(m)}(t), \tilde{y}] + \left|I_5^m - \mathbb{E}[I_5^m|G_\delta^{(m)}(t), \tilde{y}]\right|. \tag{2.9.11}$$

The expectation can be bound using Cauchy-Schwarz inequality.

$$\begin{aligned}
\mathbb{E}[I_5^m|G_\delta^{(m)}(t), \tilde{y}] &= \sum_{j\neq m}\frac{1}{d_\delta(t)}\left|\hat{\pi}_j^{(m)}(t) - \pi_j^*(t)\right| p_\delta(t) \\
&\leqslant \frac{\sqrt{n}p_\delta(t)}{d_\delta(t)}\left\|\hat{\pi}^{(m)}(t) - \pi^*(t)\right\|_2 \\
&\leqslant \frac{1}{3\sqrt{n}}\left\|\hat{\pi}^{(m)}(t) - \pi^*(t)\right\|_2 \\
&\leqslant \frac{1}{3\sqrt{n}}\left\|\hat{\pi}^{(m)}(t) - \hat{\pi}(t)\right\|_2 + \frac{1}{3\sqrt{n}}\left\|\hat{\pi}(t) - \pi^*(t)\right\|_2.
\end{aligned}$$

A bound on the first term is given by Lemma 10 and by Theorem 2 for the second term. Hence there exist constants $C_3, \tilde{C}_2 \geqslant 1$ so that with probability at least $1 - O(n^{-9})$, we have for all $m \in [n]$ that

$$\begin{aligned}
&\mathbb{E}[I_5^m | G_\delta^{(m)}(t), \tilde{y}] \\
\leqslant & 64 b^{\frac{5}{2}}(t) \left(\frac{4M\delta}{T\sqrt{n}} + C_3 \sqrt{\frac{\log n}{Ln^2 p_\delta(t) p_{\delta,sum}(t)}} \right) \|\pi^*(t)\|_\infty + \frac{1}{3\sqrt{n}} \|\hat{\pi}(t) - \pi^*(t)\|_\infty \\
& + \frac{1}{\sqrt{n}} \left(521 \frac{M\delta n b^{\frac{7}{2}}(t)}{T} + 64 \tilde{C}_2 b^{\frac{9}{2}}(t) \sqrt{\frac{1}{Lnp_\delta(t) p_{\delta,sum}(t)}} \right) \|\pi^*(t)\|_2 \\
\leqslant & 64 b^{\frac{5}{2}}(t) \left(\frac{4M\delta}{T\sqrt{n}} + C_3 \sqrt{\frac{\log n}{Ln^2 p_\delta(t) p_{\delta,sum}(t)}} \right) \|\pi^*(t)\|_\infty + \frac{1}{3\sqrt{n}} \|\hat{\pi}(t) - \pi^*(t)\|_\infty \\
& + \left(512 \frac{M\delta n b^{\frac{7}{2}}(t)}{T} + 64 \tilde{C}_2 b^{\frac{9}{2}}(t) \sqrt{\frac{1}{Lnp_\delta(t) p_{\delta,sum}(t)}} \right) \|\pi^*(t)\|_\infty \\
\leqslant & \left(C_3' \frac{Mn\delta b^{\frac{7}{2}}(t)}{T} + C_4' \frac{b^{\frac{9}{2}}(t)}{\sqrt{Lnp_\delta(t) p_{\delta,sum}(t)}} \right) \|\pi^*(t)\|_\infty + \frac{1}{3\sqrt{n}} \|\hat{\pi}(t) - \pi^*(t)\|_\infty
\end{aligned} \tag{2.9.12}$$

for some constants $C_3', C_4' \geqslant 1$.

To bound the deviation term in (2.9.11), we will use Bernstein's inequality. Then, there exists a constant $c \geqslant 1$ such that for every $m \in [n]$,

$$\mathbb{P}\left(\left| I_5^m - \mathbb{E}[I_5^m | G_\delta^{(m)}(t), \tilde{y}] \right| \geqslant c \frac{\sqrt{np_\delta(t) \log n} + \log n}{d_\delta(t)} \left\| \hat{\pi}^{(m)}(t) - \pi^*(t) \right\|_\infty \right) \leqslant 2n^{-10}.$$

Hence using a union bound over $[n]$, we see that there exists a suitably large constant $c' \geqslant 1$ so that with probability at least $1 - O(n^{-9})$,

$$\left| I_5^m - \mathbb{E}[I_5^m | G_\delta^{(m)}(t), \tilde{y}] \right| \leqslant c' \sqrt{\frac{\log n}{np_\delta(t)}} \left\| \hat{\pi}^{(m)}(t) - \pi^*(t) \right\|_\infty \quad \forall m \in [n]. \tag{2.9.13}$$

It remains to bound $\left\| \hat{\pi}^{(m)}(t) - \pi^*(t) \right\|_\infty$. Using Lemma 10 along with the triangular inequality

$$\left\| \hat{\pi}^{(m)}(t) - \pi^*(t) \right\|_\infty \leqslant \left\| \hat{\pi}^{(m)}(t) - \hat{\pi}(t) \right\|_2 + \|\hat{\pi}(t) - \pi^*(t)\|_\infty ,$$

we have that with probability at least $1 - O(n^{-9})$,

$$\left\| \hat{\pi}^{(m)}(t) - \pi^*(t) \right\|_\infty \leqslant 192 b^{\frac{5}{2}}(t) \left(\frac{4M\delta}{T} + C_3 \sqrt{\frac{\log n}{Lnp_\delta(t) p_{\delta,sum}(t)}} \right) \|\pi^*(t)\|_\infty + 2 \|\hat{\pi}(t) - \pi^*(t)\|_\infty .$$

Upon plugging this in (2.9.13), the latter simplifies to

$$\begin{aligned}
&\left| I_5^m - \mathbb{E}[I_5^m | G_\delta^{(m)}(t), \tilde{y}] \right| \\
\leqslant & C_3'' \sqrt{\frac{\log n}{np_\delta(t)}} \|\hat{\pi}(t) - \pi^*(t)\|_\infty + C_4'' b^{\frac{5}{2}}(t) \sqrt{\frac{\log n}{np_\delta(t)}} \left(\frac{4M\delta}{T} + \sqrt{\frac{\log n}{Lnp_\delta(t) p_{\delta,sum}(t)}} \right) \|\pi^*(t)\|_\infty
\end{aligned} \tag{2.9.14}$$

for some constants $C_3'', C_4'' \geqslant 1$.

Finally, combining (2.9.10), (2.9.11), (2.9.12) and (2.9.14), we conclude that with probability at least $1 - O(n^{-9})$, it holds for all $m \in [n]$ that

$$|I_4^m| \leqslant \left(C_7 \frac{Mn\delta b^{\frac{7}{2}}(t)}{T} + C_8 \frac{b^{\frac{5}{2}}(t) \max\left\{ b^2(t), \frac{\log n}{\sqrt{np_\delta(t)}} \right\}}{\sqrt{Lnp_\delta(t) p_{\delta,sum}(t)}} \right) \|\pi^*(t)\|_\infty + C_9 \sqrt{\frac{\log n}{np_\delta(t)}} \|\hat{\pi}(t) -$$

for some constants $C_7, C_8, C_9 \geqslant 1$.

2.10 PROOF OF LEMMA 1

Since $|\mathcal{N}_{ij,\delta}(t)|$ are i.i.d random variables for each $i < j$, hence $\tilde{G}_\delta(t)$ is distributed as an Erdös-Renyi graph with probability $\tilde{p}_\delta(t)$ defined as

$$\tilde{p}_\delta(t) = \mathbb{P}\left(|\mathcal{N}_{12,\delta}(t)| \in \left[\max\left\{1, \frac{p_{\delta,sum}(t)}{2}\right\}, \max\left\{ 2p_{\delta,sum}(t), 6\log\frac{4}{\delta p_{\min}} \right\} \right] \right).$$

Note that $\delta p_{\min} \leqslant |\mathcal{N}_\delta(t)| p_{\min} \leqslant p_{\delta,sum}(t) \leqslant |\mathcal{N}_\delta(t)| p_{\max} \leqslant 4\delta p_{\max}$, thus it follows that

- if $\delta p_{\min} \geqslant 3$, then $\tilde{\mathcal{E}}_\delta(t) := \left\{ \{i,j\} : |\mathcal{N}_{ij,\delta}(t)| \in \left[\frac{p_{\delta,sum}(t)}{2}, 2p_{\delta,sum}(t) \right] \right\}$;
- if $\delta p_{\max} \leqslant \frac{1}{8}$, then $\tilde{\mathcal{E}}_\delta(t) := \left\{ \{i,j\} : |\mathcal{N}_{ij,\delta}(t)| \in \left[1, 6\log\frac{4}{\delta p_{\min}} \right] \right\}$.

Let us now bound $\tilde{p}_\delta(t)$ in each case using Chernoff bounds (see Theorem 12).

- If $\delta p_{\min} \gtrsim 1$, then Chernoff's bound implies that $\tilde{p}_\delta(t) \geqslant 1 - 2e^{-\frac{p_{\delta,sum}(t)}{12}} \geqslant 1 - 2e^{-1/4}$.
- If $\delta p_{\max} < \frac{1}{8}$, then

$$\tilde{p}_\delta(t) = \mathbb{P}\left(|\mathcal{N}_{12,\delta}(t)| \in \left[1, 6\log\frac{4}{\delta p_{\min}} \right] \right) = p_\delta(t) - \mathbb{P}\left(|\mathcal{N}_{12,\delta}(t)| > 6\log\frac{4}{\delta p_{\min}} \right). \tag{2.10.1}$$

Using Chernoff's bound, it holds that

$$\mathbb{P}\left(|\mathcal{N}_{12,\delta}(t)| > 6\log\frac{4}{\delta p_{\min}} \right) \leqslant \frac{\delta p_{\min}}{4}. \tag{2.10.2}$$

Moreover, $\delta p_{\max} < \frac{1}{8}$ implies that $p_{\delta,sum}(t) \leqslant \frac{1}{2}$, and so, one can bound $p_\delta(t)$ using Proposition 3 as follows.

$$\frac{\delta p_{\min}}{2} \leqslant 1 - e^{-\delta p_{\min}} \leqslant p_\delta(t) \leqslant 1 - e^{-8\delta p_{\max}} \leqslant 8\delta p_{\max}. \tag{2.10.3}$$

Finally, combining (2.10.1),(2.10.2) and (2.10.3), it holds that $\tilde{p}_\delta(t) \geqslant \frac{\delta p_{\min}}{4}$. The upper bound on $\tilde{p}_\delta(t)$ comes from (2.10.3) and the fact that $\tilde{p}_\delta(t) \leqslant p_\delta(t)$.

3 DYNAMIC RANKING AND TRANSLATION SYNCHRONIZATION

In this chapter we consider the problem of estimating the latent strengths of a set of n items from noisy pairwise measurements in a dynamic setting. In particular, we propose a dynamic version of the TranSync model [21] by placing a global smoothness assumption on the evolution of the latent strengths. We propose and analyze two estimators for this problem and obtain ℓ_2 estimation error rates for the same. Experiment results on both synthetic data and real data sets are presented.

3.1 PROBLEM SETUP AND ALGORITHMS

3.1.1 *The Dynamic TranSync model*

Let us introduce formally our model for dynamic pairwise comparisons, inspired by the TranSync model [21]. Our data consists of pairwise comparisons on a set of items $[n] = \{1, 2, \ldots, n\}$ at different times $t \in \mathcal{T} = \left\{\frac{k}{T} | k = 0, \ldots, T\right\}$, where $\mathcal{T}$ is a uniform grid on the interval $[0, 1]$. At each time $t \in \mathcal{T}$, we denote the observed comparison graph $G_t = ([n], \mathcal{E}_t)$ where $\mathcal{E}_t$ is the set of undirected edges. It will be useful to denote $\overrightarrow{\mathcal{E}_t} = \{(i,j) | \{i,j\} \in \mathcal{E}_t,\ i < j\}$ as the corresponding set of directed edges. We assume that the set of items $[n]$ is the same throughout, but the set of compared items $\mathcal{E}_t$ can change with time.

To model our data, we use the TranSync model at each time t which posits that the outcome of a comparison between two items is solely determined by their strengths. The strengths of the items at time t are represented by the vector $z_t^* = (z_{t,1}^*, \ldots, z_{t,n}^*)^\top \in \mathbb{R}^n$. For each $t \in \mathcal{T}$ and for every pair of items $\{i,j\} \in \mathcal{E}_t$, we obtain a noisy measurement of the strength difference $z_{t,i}^* - z_{t,j}^*$

$$y_{ij}(t) = z_{t,i}^* - z_{t,j}^* + \epsilon_{ij}(t), \tag{3.1.1}$$

where $\epsilon_{ij}(t)$ are i.i.d. centered subgaussian random variables with ψ_2 norm $\|\epsilon_{ij}(t)\|_{\psi_2}^2 = \sigma^2$ [46]. Let us denote $x^*(t) \in \mathbb{R}^{|\mathcal{E}_t|}$ where

$$x_{ij}^*(t) = z_{t,i}^* - z_{t,j}^*, \quad \{i,j\} \in \mathcal{E}_t$$

and x^* to be formed by column-wise stacking as

$$x^* = \begin{pmatrix} x^*(0) \\ x^*(1) \\ \vdots \\ x^*(T) \end{pmatrix}.$$

Remark 6 (BTL model). *A well-studied model in ranking problems is the BTL model [5], and it has recently been extended to a dynamic setting [4, 24]. It posits that at each time* t*, the probability that an item is preferred to another item only* *depends on their strengths. If* $w^*_{t,i}$ *is the strength of item* i *at time* t*, the quantity* $\frac{w^*_{t,i}}{w^*_{t,i}+w^*_{t,j}}$ *represents the probability that* i *beats* j *at time* t*. More precisely, considering* $z^*_{t,i} = \ln w^*_{t,i}$*, it states that*

$$\text{logit}\left(\mathbb{P}(i \textit{ preferred at } j \textit{ at time } t)\right) = z^*_{t,i} - z^*_{t,j},$$

where $\text{logit}(x) = \ln \frac{x}{1-x}$ *for* $x \in (0, 1)$*. At each time* $t \in \mathcal{T}$ *and for each pair* $\{i, j\} \in \mathcal{E}_t$*, the observations are* L *Bernoulli random variables with parameter equal to the probability that* i *is preferred to* j *at time* t*. Then, taking* $\tilde{y}_{ij}(t)$ *as the mean of all the observations corresponding to this triplet* (t, i, j)*, the fraction* $R_{ij}(t) := \frac{\tilde{y}_{ij}(t)}{\tilde{y}_{ji}(t)}$ *can be viewed as a noisy measurement of* $\frac{w^*_{t,i}}{w^*_{t,j}}$*. Indeed, it was shown in [18] for the analogous static case (we drop* t *from the notation) with* $R_{ij} := \frac{\tilde{y}_{ij}}{\tilde{y}_{ji}}$ *that*

$$\ln R_{ij} = \ln w^*_i - \ln w^*_j + \tilde{\epsilon}(w^*_i, w^*_j, y_{ij}) \tag{3.1.2}$$

where the RHS of (3.1.2) consists of terms in the Taylor expansion of the $\ln$ *function. Note that the noise in the observations (captured by* $\tilde{\epsilon}$ *in (3.1.2)) is no longer zero-mean, contrary to the Dynamic TranSync model, which can introduce some bias in the estimation.*

Remark 7 (Outliers model)**.** *Another well-known model in group synchronization is the outliers model [43]. It posits that for any pair of compared items* $\{i, j\}$*, we observe either the true strength difference or random noise. In our setting, one can introduce the following analogous version of this model,*

$$y_{ij}(t) = (z^*_{t,i} - z^*_{t,j})X_{ij}(t) + (1 - X_{ij}(t))\epsilon_{ij}(t)$$

where $\epsilon_{ij}(t) \sim \mathcal{N}(0, 1)$ *and* $X_{ij}(t) \sim \mathcal{B}(\eta)$*,* $\eta \in]0, 1[$*, denotes the probability of observing the true strength difference, independent of* $\epsilon_{ij}(t)$*. Note that in this model, observations are not centered around* $x^*(t)$*, contrary to (3.1.1). However, we can rewrite* $y_{ij}(t)$ *as*

$$\begin{aligned} y_{ij}(t) &= y_{ij}(t) - \mathbb{E}[y_{ij}(t)] + \mathbb{E}[y_{ij}(t)] \\ &= \eta x^*_{ij}(t) + \underbrace{(1 - X_{ij}(t))\epsilon_{ij}(t) + (X_{ij}(t) - \eta)x^*_{ij}(t)}_{0-\textit{mean, subgaussian}} \end{aligned}$$

which is similar to (3.1.1), but with an additional bias since $\mathbb{E}[y_{ij}(t)] \neq x^*_{ij}(t)$*. This suggests that the analysis of the outliers model could potentially be done following a similar strategy to ours, although the presence of the bias term would likely make the analysis more cumbersome. Note that* η *is unknown, hence one can not use it directly in the estimation procedure.*

Finally, we mention that since $t = k/T$ for an integer $0 \leqslant k \leqslant T$, we will often interchangeably use t and k for indexing purposes.

SMOOTH EVOLUTION OF WEIGHTS. As discussed in the introduction, we will assume that the weights z^*_t do not change, in an appropriate sense, too quickly with t. Since each z^*_t is only identifiable up to a constant shift, the smoothness assumption that we impose needs to be invariant to such transformations. The assumption we make is as follows.

Assumption 2 (Global ℓ_2 smoothness)**.** *Let* $C \in \mathbb{R}^{n \times \binom{n}{2}}$ *to be the edge incidence matrix of the complete graph* K_n*. We assume that*

$$\sum_{k=0}^{T-1} \left\| C^\top (z_k^* - z_{k+1}^*) \right\|_2^2 \leqslant S_T. \tag{3.1.3}$$

The above assumption states that the vector $C^\top z_t^* \in \mathbb{R}^{\binom{n}{2}}$ does not change too quickly "on average", and is analogous to the usual notion of quadratic variation of a function. The regime $S_T = O(T)$ is uninteresting of course, what we are interested in is the situation where $S_T = o(T)$. Also note that (3.1.3) is invariant to a constant shift of z_t^*, as desired. We will then aim to estimate the 'block-centered' version of the vector $z^* \in \mathbb{R}^{n(T+1)}$, where each block z_t^* is shifted by an additive constant such that

$$\frac{1}{n} \sum_{i=1}^{n} z_{t,i}^* = 0 \quad \forall\, t \in \mathcal{T}.$$

In what follows, we will assume w.l.o.g that z^* is block-centered. Finally, denoting $M \in \mathbb{R}^{(T+1) \times T}$ to be the incidence matrix of the path graph on $T+1$ vertices and $E = M^\top \otimes C^\top$, we can write (3.1.3) as

$$\sum_{k=0}^{T-1} \left\| C^\top (z_k^* - z_{k+1}^*) \right\|_2^2 = \|Ez^*\|_2^2 = {z^*}^\top E^\top E z^* \leqslant S_T.$$

Hence z^* lies close to the null space of $E^\top E$, i.e. $\mathcal{N}(E^\top E)$, where this closeness is captured by S_T.

Remark 8 (Other smoothness assumptions and error guarantees)**.** *Local smoothness assumptions have been studied in the literature for other related models, leading to recovery guarantees at a given time* t *(see [4, 24] for dynamic BTL model and also [28] for a non-transitive model with autoregressive noise). Here, we propose a weaker global smoothness assumption, which fits better some real-life data sets, as it allows for jumps in the smoothness. However, under a weaker global assumption, the error guarantees one can provide will also be of the global type, and relatively weaker compared to the local ones. In this work, we will bound the MSE for estimating* z^**.*

3.1.2 *Smoothness-constrained estimators*

In the Dynamic TranSync model (3.1.1), observations are noisy measurements of strength differences with zero-mean noise. A classical approach for recovering the strength vector z^* is to solve the following linear system

$$y_{ij}(t) = z_{t,j} - z_{t,i}, \quad \forall t \in \mathcal{T}, \{i,j\} \in \mathcal{E}_t \tag{3.1.4}$$

in the least-squares sense. To reflect the temporal smoothness of the data, we need to take into account the smoothness constraints from Assumption 2. We will consider two different approaches to incorporate these constraints in (3.1.4).

SMOOTHNESS-PENALIZED LEAST SQUARES. A typical approach for incorporating constraints into a least squares problem involves adding them as a penalty term. Hence, an estimator $\hat{z}$ of the strength vector is given as a solution of the following problem.

$$\hat{z} = \underset{\substack{z_0,\ldots,z_T \in \mathbb{R}^n \\ z_k^\top \mathbf{1}_n = 0}}{\operatorname{argmin}} \sum_{t\in\mathcal{T}} \sum_{(i,j)\in\overrightarrow{\mathcal{E}_t}} \left(y_{ij}(t) - (z_{t,i} - z_{t,j})\right)^2 + \lambda \sum_{k=0}^{T-1} \left\| C^T(z_k - z_{k+1}) \right\|_2^2. \tag{3.1.5}$$

The penalty term promotes smooth solutions and the estimate $\hat{z} \in \mathbb{R}^{n(T+1)}$ is formed by column stacking $\hat{z}_0, \ldots, \hat{z}_T \in \mathbb{R}^n$. If $\lambda = 0$, then each estimate $\hat{z}_k$ only uses the information available at this time instant, through G_k. Hence, the error $\|z^* - \hat{z}\|_2^2$ will typically grow linearly with T (large variance). On the other hand, if λ is very large, then the estimate $\hat{z}$ will be very 'smooth' meaning that $\hat{z}_k$ will be similar for all k. Hence, the error $\|z^* - \hat{z}\|_2^2$ will typically have a large bias. Therefore an intermediate choice of the parameter λ is important to achieve the right bias-variance trade-off.

Remark 9 (Laplacian smoothing). *Note that the above estimator has similarities with the Laplacian smoothing estimator [40] for the model in (1.3.6), where an estimate $\hat{x}$ of x^* is obtained as*

$$\hat{x} = \underset{x}{\operatorname{argmin}} \|y - x\|_2^2 + x^\top L x.$$

Indeed, the Dynamic TranSync model can be compared with the setting in [40] by considering each of the graphs G_t as the vertices of a path graph, for which information is available as the vector of observations $y(t)$ at time t. This motivates solving the penalized least-square problem in (3.1.5) using the smoothness assumption in (3.1.3) as the penalty term. Unlike Laplacian smoothing, the penalty in our case does not involve a Laplacian matrix.

PROJECTION METHOD. Our second approach consists of the following two-stage estimator.

1. *Step* 1: For each $t \in \mathcal{T}$, we compute $\check{z}_t \in \mathbb{R}^n$ as the (minimum ℓ_2 norm) least-squares solution of (3.1.4), and form $\check{z} \in \mathbb{R}^{n(T+1)}$ by column-stacking $\check{z}_0, \ldots, \check{z}_T$.
2. *Step* 2: Let $\mathcal{V}_\tau$ be the space generated by the eigenvectors of $E^\top E$ corresponding to eigenvalues smaller than a threshold $\tau > 0$, and let $P_{\mathcal{V}_\tau}$ be the projection matrix on $\mathcal{V}_\tau$. Then, the estimator for z^* is defined as

$$\hat{z} = P_{\mathcal{V}_\tau} \check{z}. \tag{3.1.6}$$

As will be seen later due to the form of $\mathcal{N}(E^\top E)$, it will hold that each block $\hat{z}_k \in \mathbb{R}^n$ satisfies $\hat{z}_k^\top \mathbf{1}_n = 0$ for $k = 0, \ldots, T$.

Remark 10 (Laplacian eigenmaps). *The above estimator is constructed in a similar fashion as the Laplacian eigenmaps estimator [40] for the model (1.3.6), which is obtained by projection of the observations y onto the space spanned by the smallest eigenvectors (i.e., corresponding to the smallest eigenvalues) of the Laplacian L. In our setting, we do not have direct information on the vertices so we replace it by* *a least-squares solution of (3.1.4) in Step 1. Moreover, we construct the projection matrix using the smallest eigenvectors of $E^\top E$ since z^* lies close to $\mathcal{N}(E^\top E)$.*

3.2 MAIN RESULTS

Here we present our main results, namely Theorems A and B, which correspond to theoretical guarantees for the estimation error of the proposed estimators $\hat{z}$ and $\tilde{z}$, based on smoothness-penalized least squares and the projection method respectively. We choose to write the results in this section in a stylized form for better readability, highlighting the rates with respect to the time parameter T and hiding the dependency on the rest of the parameters, which will be later made explicit in Sections 3.3 and 3.4. Our theoretical results hold under the assumption that each comparison graph G_t is connected, but this is essentially for technical reasons and we believe this requirement can be relaxed, see Remarks 12 and 13 for a more detailed discussion. Simulation results in Section 3.7.1 show that the MSE goes to zero for both the estimators (as T increases) even in the very sparse regime where individual G_t's may be disconnected.

Theorem A (Smoothness-penalized least squares)**.** *Let $\hat{z} \in \mathbb{R}^{n(T+1)}$ be the estimator defined in (3.1.5) where the data is generated by the model (3.1.1) with subgaussian noise parameter σ^2 and with $z^* \in \mathbb{R}^{n(T+1)}$ as the ground truth vector of strength parameters. Suppose additionally that G_t is connected for each $t \in \mathcal{T}$. Under Assumption 2, if $\lambda = \sigma^{\frac{4}{5}}(\frac{T}{S_T})^{2/5}$, it holds with probability larger than $1-\delta$*

$$\|\hat{z} - z^*\|_2^2 \leqslant T^{4/5} S_T^{1/5} \Psi_{LS}(n, \sigma, \delta) + \Psi'_{LS}(n, \sigma, \delta).$$

Here, $\Psi_{LS}(\cdot)$ and $\Psi'_{LS}(\cdot)$ are functions of the parameters of the problem.

A more formal version of this theorem is given by Theorem 6 in Section 3.3, where the functions $\Psi_{LS}(\cdot)$, $\Psi'_{LS}(\cdot)$ are made explicit, up to universal constants.

Theorem B (Projection method)**.** *Let $\hat{z} \in \mathbb{R}^{n(T+1)}$ be the estimator defined in (3.1.6). Assume that each comparison graph in the sequence $(G_k)_{k=0}^{T}$ is connected. If $\tau = \sigma^{-\frac{4}{3}}(\frac{S_T}{T})^{2/3}$, then it holds with probability larger than $1-\delta$ that*

$$\|\hat{z} - z^*\|_2^2 \leqslant T^{2/3} S_T^{1/3} \Psi_{Proj}(n, \sigma, \delta) + \Psi'_{Proj}(n, \sigma, \delta)$$

where $\Psi_{Proj}(\cdot)$ and $\Psi'_{Proj}(\cdot)$ are functions of the parameters of the problem.

For a more formal statement of the previous theorem (with the explicit $\Psi_{Proj}, \Psi'_{Proj}$), we direct the reader to Theorem 10 in Section 3.4. Observe that in the case $S_T = O(T)$, the error in both theorems is of order $O(T)$, which is to be expected, since in this case the smoothness constraint is always satisfied and the statement becomes vacuous. When $S_T = o(T)$, we see for both estimators that the mean squared error

$$\frac{1}{T+1}\|\hat{z} - z^*\|_2^2 = \frac{1}{T+1}\sum_{k=0}^{T+1}\|\hat{z}_k - z_k^*\|_2^2 = o(1) \quad \text{as} \quad T \to \infty.$$

However the bound in Theorem B is better than that in Theorem A in terms of dependence on T, when $S_T = o(T)$. This is due to technical difficulties arising during the control of the bias term in the proof of Theorem A, since $(G_k)_{k=0}^{T}$ is allowed to be any sequence of connected graphs. In spite of this, we are able to obtain the same error-rate as Theorem B under additional assumptions. More specifically, in Theorem 6 we assume that the comparison graphs are non-evolving (see Assumption 3) and obtain a rate that matches that of Theorem B. In Theorem 8, we prove that if the comparison graphs are connected and possibly evolving, and an additional technical condition holds, then we again recover the rate of the projection estimator.

Remark 11 (Error rates – Lipschitz smoothness)**.** *When* $S_T = O(1/T)$*, then* $\|\hat{z} - z^*\|_2^2 = O(T^{1/3})$ *for Theorem B, which matches the optimal rate for estimating a Lipschitz function on* $[0,1]$ *over a uniform grid, w.r.t the squared (empirical)* L_2*-norm [37, Thm.1.3.1]. Indeed, let* $f_i : [0,1] \to \mathbb{R}$ *be Lipschitz functions for* $i = 1,\dots,n$ *and define* $f(t) = (f_1(t),\dots,f_n(t))^\top$ *for each* $t \in [0,1]$*. Then consider the case where* z^* *is defined as*

$$z^*_{t,i} := f_i(t) - \frac{\mathbf{1}^\top f(t)}{n}; \quad i \in [n], t \in \mathcal{T}.$$

Clearly, z^* *is block-wise centered and satisfies Assumption 2 with* $S_T = O(1/T)$.

3.3 SMOOTHNESS-PENALIZED LEAST SQUARES ESTIMATOR ANALYSIS

In this section we obtain theoretical guarantees for the error of the estimator $\hat{z}$ given in (3.1.5). The results in this section will build towards proving Theorem A, starting from the special case when all the graphs are equal in Section 3.3.1. Before proceeding, we introduce some notation that will appear in the analysis.

NOTATION. Let us denote by Q_k the incidence matrix of the graph G_k and Q will denote the $n(T+1) \times \sum_{k=0}^{T} |\mathcal{E}_k|$ block diagonal matrix where the blocks on the diagonal are the matrices Q_k. Similarly, we define the Laplacian at time $t = k/T$ by $L_k := Q_k Q_k^\top$ and L will be defined as the stacked Laplacian, which is the $n(T+1) \times n(T+1)$ block diagonal matrix with blocks L_k on the main diagonal. We define, for $\lambda > 0$, the regularized Laplacian matrix $L(\lambda) := L + \lambda E^\top E$. Let us also define notation for the eigenpairs of matrices $CC^\top, MM^\top$ and L_k.

- $(\lambda_j, v_j)_{j=1}^{n}$ denotes the eigenpairs of $CC^\top$. Observe that $v_n = \frac{1}{n}\mathbf{1}_n$ and $(v_j)_{j=1}^{n-1}$ can be any orthonormal basis of $\text{span}(\mathbf{1}_n)^\perp$. In addition, we have $\lambda_j = n-1$ for $1 \leqslant j \leqslant n-1$ and $\lambda_n = 0$.
- $(\mu_k, u_k)_{k=0}^{T}$ denotes the eigenpairs of $MM^\top$ (path graph on $T+1$ vertices), with $\mu_0 \geqslant \mu_1 \geqslant \mu_2 \geqslant \cdots \geqslant \mu_{T-1} > \mu_T = 0$. Note that $u_T = \text{span}(\mathbf{1}_{T+1})$.
- $(\alpha_{k,j}, a_{k,j})_{j=1}^{n}$ denotes the eigenpairs of L_k, with $\alpha_{k,1} \geqslant \alpha_{k,2} \geqslant \cdots \geqslant \alpha_{k,n} = 0$ for all $k = 0,\dots,T$.

Now, the estimator $\hat{z}$ defined in (3.1.5) can be equivalently defined as a solution of

$$\min_{\substack{z \in \mathbb{R}^{n(T+1)}, \\ z_k^\top \mathbf{1}_n = 0}} \left\| Q^\top z - y \right\|_2^2 + \lambda \|Ez\|_2^2. \tag{3.3.1}$$

For the unconstrained problem, the solutions of (3.3.1) satisfy

$$L(\lambda)z = Qy. \tag{3.3.2}$$

The following lemma provides conditions under which the null space $\mathcal{N}(L(\lambda)) = \text{span}\{e_k \otimes \mathbf{1}_n\}_{k=0}^{T}$ where $e_0, \dots, e_T \in \mathbb{R}^{T+1}$ is a canonical basis of $\mathbb{R}^{T+1}$.

Lemma 12. *If the union graph $G_U := ([n], \cup_{t \in \mathcal{T}} \mathcal{E}_t)$ is connected, it follows for any $\lambda > 0$ that $\mathcal{N}\big(L(\lambda)\big) = \text{span}\{e_k \otimes \mathbf{1}_n\}_{k=0}^{T}$. Here, $e_0, \dots, e_T \in \mathbb{R}^{T+1}$ is a canonical basis of $\mathbb{R}^{T+1}$.*

The proof is outlined in Section 3.6. It follows that if the union graph G_U is connected, then the estimator $\hat{z}$ is given by the following solution of (3.3.2)

$$\hat{z} = L^{\dagger}(\lambda)Qy, \tag{3.3.3}$$

which is uniquely defined and is orthogonal to $\text{span}\{e_k \otimes \mathbf{1}_n\}_{k=0}^{T}$ (in other words, $\hat{z}$ is block-wise centered). Furthermore, the ground truth z^* satisfies by definition $Lz^* = Qx^*$, which implies that $L(\lambda)z^* = Qx^* + \lambda E^\top E z^*$. Since z^* is block-wise centered, i.e., $z^* \perp \mathcal{N}(L(\lambda))$ due to Lemma 12, hence it satisfies

$$z^* = L^{\dagger}(\lambda)\big(Qx^* + \lambda E^\top E z^*\big). \tag{3.3.4}$$

By (3.3.3), (3.3.4) and the triangle inequality, we arrive at the following bound

$$\|\hat{z} - z^*\|_2^2 \lesssim \|L^{\dagger}(\lambda)Q(y - x^*)\|_2^2 + \lambda^2 \|L^{\dagger}(\lambda)E^\top E z^*\|_2^2. \tag{3.3.5}$$

The first term in the RHS of (3.3.5) is the variance term due to noise that will be controlled by a large enough value of λ. The second term is the bias which depends on the smoothness of z^*, and will be controlled by choosing λ to be suitably small. The optimal choice of λ will then achieve the right bias-variance trade-off.

3.3.1 *Warm-up: the non-evolving case*

We will first analyze the case where the comparison graph is the same across all times, as this case will serve as a foundation for our analysis of the time-evolving case. We will refer to this case as non-evolving, which is formally defined by the following assumption (notice that this is different from what we referred to previously as the static case, where $T = 1$).

Assumption 3 (Fixed and connected comparison graph)**.** *Let G_0 be any connected graph on n vertices, we assume that $G_k = G_0$, for all $k = 0, \dots, T$.*

Under Assumption 3, we have $L_k = L_0$ and $Q_k = Q_0$, which implies that the matrices Q, L and $L(\lambda)$ can be written as

$$\begin{aligned} Q &= I_{T+1} \otimes Q_0, \\ L &= I_{T+1} \otimes L_0, \\ L(\lambda) &= I_{T+1} \otimes L_0 + \lambda E^\top E. \end{aligned}$$

SPECTRAL DECOMPOSITION OF $L(\lambda)$. In order to obtain an explicit spectral decomposition for $L(\lambda)$, we choose $v_j := a_{0,j}$ (given that $a_{k,j} = a_{0,j}$ we will simply write a_j in the sequel) for all $j = 1, \dots, n-1$ as the eigenvector basis for $L_{com} := CC^\top$. With this choice, and recalling that $E^\top E = MM^\top \otimes L_{com}$, the eigenpairs of $E^\top E$ associated with nonzero eigenvalues are

$$\big((n-1)\mu_k, u_k \otimes a_j\big)_{j=1,\cdots,n-1,k=0,\cdots,T-1}.$$

Therefore each eigenvalue of the form $(n-1)\mu_k$ has multiplicity $n-1$. Observe that we have the freedom to choose any orthonormal basis as the eigenbasis for I_{T+1}, we opt for the basis given by $\{u_k\}_{k=0}^T$, i.e., the eigenvectors of $MM^\top$. Hence, the eigenpairs of $I_{T+1} \otimes L_0$ for nonzero eigenvalues are given by

$$(\alpha_j, u_k \otimes a_j)_{j=1,\cdots,n-1,k=0,\cdots,T}.$$

Thus, the following decompositions hold.

$$E^\top E = \sum_{k=0}^{T-1}\sum_{j=1}^{n-1}(n-1)\mu_k(u_k u_k^\top \otimes a_j a_j^\top), \tag{3.3.6}$$

$$L = \sum_{k=0}^{T}\sum_{j=1}^{n-1}\alpha_j(u_k u_k^\top \otimes a_j a_j^\top), \tag{3.3.7}$$

$$L(\lambda) = \sum_{k=0}^{T-1}\sum_{j=1}^{n-1}(\alpha_j + (n-1)\lambda\mu_k)(u_k u_k^\top \otimes a_j a_j^\top) + \sum_{j=1}^{n-1}\alpha_j(u_T u_T^\top \otimes a_j a_j^\top). \tag{3.3.8}$$

From (3.3.8) we see directly that $L(\lambda)$ has rank $(n-1)(T+1)$ and that its nullspace is given by $\text{span}\{u_k \otimes a_n\}_{k=0}^T = \text{span}\{e_k \otimes \mathbf{1}_n\}_{k=0}^T$. Given the above notations and setup, we can now present the following bound on the estimation error.

Proposition 1. *Take $\delta \in (0, e^{-1})$, then under Assumptions 2 and 3 it holds with probability larger than $1-\delta$*

$$\|\hat{z} - z^*\|_2^2 \leqslant \Big(\frac{1}{\alpha_{n-1}^2} \vee 1\Big)\lambda S_T \tag{3.3.9}$$
$$+ \sigma^2\alpha_1(1 + 4\log(1/\delta))\Big(\sum_{k=0}^{T-1}\sum_{j=1}^{n-1}\frac{1}{(\alpha_j + \lambda(n-1)\mu_k)^2} + \sum_{j=1}^{n-1}\frac{1}{\alpha_j^2}\Big).$$

The proof is detailed in Section 3.6.2.

CHOICE OF λ. The right hand side of the estimation error bound (3.3.5) can be regarded as the sum of a bias and a variance term, representing an instance of the bias-variance trade-off phenomenon. Proposition 1 gives an error bound where the dependence on λ (and the other parameters of the problem) is explicit. The following lemma helps us further simplify the dependence on λ for the variance term (second term in the RHS of (3.3.9)).

Lemma 13. *We have*

$$\sum_{k=0}^{T-1}\sum_{j=1}^{n-1}\frac{1}{(\alpha_j+\lambda(n-1)\mu_k)^2}\lesssim\frac{T\sqrt{n-1}}{\alpha_{n-1}^{3/2}\sqrt{\lambda}}$$

The proof is outlined in Section 3.6.1. Combining the results of Proposition 1 and Lemma 13, we obtain

$$\|\hat{z}-z^*\|_2^2\lesssim O\big(\lambda S_T+\sigma^2\frac{T}{\sqrt{\lambda}}\big)+\sigma^2\alpha_1\log(1/\delta)\sum_{j=1}^{n-1}\frac{1}{\alpha_j^2},$$

where we recall that the asymptotic notation $O(\cdot)$ hides the dependence on constants and all parameters except for T and σ. It is easy to see that the optimal choice for λ (in terms of T, S_T and σ) is then given by

$$\lambda=\operatorname*{argmin}_{\lambda'}\lambda' S_T+\sigma^2\frac{T}{\sqrt{\lambda'}},\tag{3.3.10}$$

which corresponds to $\lambda=\sigma^{\frac{4}{3}}(\frac{T}{S_T})^{2/3}$. The choice of λ given by (3.3.10) is well defined (there is a unique minimizer) in the case where at least one of the terms S_T and σ is non-zero. On the other hand, for the case $\sigma=S_T=0$, the estimation error is zero for any choice of λ, as is clear from Proposition 1. We will only consider the case when at least one of σ, S_T is non-zero in the sequel. Plugging this in (3.3.9) and using Lemma 13 we arrive at the following bound on the estimation error.

Theorem 6. *Let $\delta\in(0,e^{-1})$. Under Assumptions 2 and 3, choosing $\lambda=\sigma^{\frac{4}{3}}(\frac{T}{S_T})^{2/3}$, it holds with probability larger than $1-\delta$ that*

$$\|\hat{z}-z^*\|_2^2\lesssim\sigma^{\frac{4}{3}}T^{\frac{2}{3}}S_T^{\frac{1}{3}}\Big(\frac{1}{\alpha_{n-1}^2}\vee 1+\frac{\alpha_1\sqrt{n-1}}{\alpha_{n-1}^{3/2}}\log(1/\delta)\Big)+\sigma^2\alpha_1\log(1/\delta)\sum_{j=1}^{n-1}\frac{1}{\alpha_j^2}.$$

The following corollary states the order of the estimation error under different smoothness regimes.

Corollary 3. *Assume that $S_T=O(T^\gamma)$ for some $\gamma<1$. If $\lambda=\sigma^{\frac{4}{3}}(\frac{T}{S_T})^{2/3}$, then it holds with high probability that $\|\hat{z}-z^*\|_2^2=O(\sigma^{\frac{4}{3}}T^{\frac{2+\gamma}{3}}\vee\sigma^2)$. In particular,*

$$\|\hat{z}-z^*\|_2^2=\begin{cases}O(\sigma^{\frac{4}{3}}T^{\frac{2}{3}})\text{ if }S_T=O(1)\text{ and }\lambda=\sigma^{\frac{4}{3}}T^{2/3},\\ O(\sigma^{\frac{4}{3}}T^{\frac{1}{3}})\text{ if }S_T=O(\frac{1}{T})\text{ and }\lambda=\sigma^{\frac{4}{3}}T^{4/3}.\end{cases}$$

3.3.2 *Any sequence of connected graphs*

We now treat the case where the comparison graphs G_k can differ from one another across time, but we assume that G_k is connected for all $k\in\{0,\dots,T\}$. The main idea is to use the results from Section 3.3.1 and the fact that the Laplacian of every connected graph can be bounded (in the Loewner order) by the Laplacian of the complete graph, up to a scalar factor.

A LOEWNER BOUND. If graph G_k is connected for all $k \in \{0, \dots, T\}$, then there exists a sequence $\beta_0, \cdots, \beta_T$ of strictly positive real numbers such that

$$L_k \succcurlyeq \beta_k L_{com}, \tag{3.3.11}$$

where $L_{com} = CC^T$ is the Laplacian matrix of K_n, the complete graph on n vertices, and L_k is the Laplacian of the graph G_k. Notice that when G_k is connected, we can always choose $\beta_k(n-1)$ to be the Fiedler eigenvalue (the second-smallest eigenvalue[1]) of L_k [6]. Denoting $D_\beta = \text{diag}(\beta_0, \dots, \beta_T)$, we have the following semi-definite bound for the stacked Laplacian L

$$L \succcurlyeq D_\beta \otimes L_{com} \succcurlyeq \min_{0 \leqslant k \leqslant T} \beta_k \cdot (I_{T+1} \otimes L_{com}) := \min_{0 \leqslant k \leqslant T} \beta_k \tilde{L}, \tag{3.3.12}$$

which implies that

$$L(\lambda) \succcurlyeq (D_\beta + \lambda M M^T) \otimes L_{com} \succcurlyeq \big(\min_{0 \leqslant k \leqslant T} \beta_k \cdot I_{T+1} + \lambda M M^T \big) \otimes L_{com}. \tag{3.3.13}$$

Let $\beta_k = \lambda_{min}(L_k)/(n-1)$, where $\lambda_{min}(L_k)$ denotes the smallest nonzero eigenvalue of L_k. Then, $\min_{0 \leqslant k \leqslant T} \beta_k = \frac{\lambda_{min}(L)}{n-1}$, where $\lambda_{min}(L)$ is the smallest nonzero eigenvalue of L, and (3.3.13) implies

$$L(\lambda) \succcurlyeq \left(\frac{1}{n-1} \lambda_{min}(L) \cdot I_{T+1} + \lambda M M^T \right) \otimes L_{com} := \tilde{L}(\lambda). \tag{3.3.14}$$

Using these observations, we arrive at the following error bound which is a generalization of Proposition 1 to the general case of evolving graphs. Its proof can be found in Section 3.6.3.

Proposition 2. *Assume that G_k is connected for all $k = 0, \dots, T$. Then under Assumption 2, the estimator $\hat{z}$ given by (3.1.5) satisfies with probability larger than* $1 - \delta$

$$\|\hat{z} - z^*\|_2^2 \lesssim \frac{1}{\lambda_{min}^2(L)} \lambda^2 n S_T + \sigma^2 \|L\|_2 (1 + 4\log(1/\delta)) \Big(\sum_{k=0}^{T-1} \frac{n-1}{(\lambda_{min}(L) + (n-1)\lambda\mu_k)^2} + \frac{n-1}{\lambda_{min}^2(L)} \Big).$$

The main difference between the previous result and Proposition 1 is that here we obtain a bound of order $O(\lambda^2 S_T)$ for the bias term, while Proposition 1 gives a bound of order $O(\lambda S_T)$ for the same term. This will have an impact on the final error rate given in Theorem 7 below.

We choose λ by proceeding as in Section 3.3.1. Indeed, we obtain the bound

$$\sum_{k=0}^{T-1} \frac{n-1}{(\lambda_{min}(L) + (n-1)\lambda\mu_k)^2} \lesssim \frac{T\sqrt{n-1}}{\lambda_{min}(L)^{3/2}\sqrt{\lambda}} \tag{3.3.15}$$

1 It is known for a fact that between all the connected graphs on n vertices, the one that has the smallest Fiedler value is the path graph, which provides a lower bound of order n^{-2} on $\beta_k(n-1)$, for all $k \in \{0, \cdots, T\}$.

in the same manner as in (3.6.3). Plugging it in Proposition 2, we deduce $\lambda = \sigma^{\frac{4}{5}}\left(\frac{T}{S_T}\right)^{\frac{2}{5}}$ to be the optimal choice. This then leads to the following theorem which corresponds to a formal version of Theorem A.

Theorem 7. *Under the assumptions of Proposition 2 and choosing* $\lambda = \sigma^{\frac{4}{5}}\left(\frac{T}{S_T}\right)^{\frac{2}{5}}$, *it holds with probability larger than* $1-\delta$

$$\|\hat{z}-z^*\|_2^2 \lesssim \sigma^{\frac{8}{5}} T^{\frac{4}{5}} S_T^{\frac{1}{5}} \left(\frac{n}{\lambda_{\min}^2(L)} + \frac{\sqrt{n-1}\|L\|_2 \log(1/\delta)}{\lambda_{\min}^{3/2}(L)} \right) + \frac{(n-1)\|L\|_2 \sigma^2 \log(1/\delta)}{\lambda_{\min}^2(L)}.$$

Consequently, if $S_T = O(T^\gamma)$ *for some* $\gamma < 1$, *then* $\|\hat{z}-z^*\|_2^2 = O(\sigma^{\frac{8}{5}} T^{\frac{4}{+}\gamma} 5 \vee \sigma^2)$. *In particular,*

$$\|\hat{z}-z^*\|_2^2 = \begin{cases} O(\sigma^{\frac{8}{5}} T^{\frac{4}{5}}) \text{ if } S_T = O(1) \text{ and } \lambda = \sigma^{\frac{4}{5}} T^{2/5}, \\ O(\sigma^{\frac{8}{5}} T^{\frac{3}{5}}) \text{ if } S_T = O(\frac{1}{T}) \text{ and } \lambda = \sigma^{\frac{4}{5}} T^{4/5}. \end{cases}$$

This theorem reveals the dependence on the parameters of the problem, up to absolute constants. In terms of the notation introduced in the statement of Theorem A, we have

$$\Psi_{LS}(n,\sigma,\delta) = \sigma^{\frac{8}{5}} \left(\frac{n}{\lambda_{\min}^2(L)} + \frac{\|L\|_2 \sqrt{n-1} \log(1/\delta)}{\lambda_{\min}^{3/2}(L)} \right),$$
$$\Psi'_{LS}(n,\sigma,\delta) = \frac{(n-1)\|L\|_2 \sigma^2 \log(1/\delta)}{\lambda_{\min}^2(L)}.$$

The error rates presented in Theorem 7 are not optimal and they do not match those obtained in Theorem 1 for the case of non-evolving graphs. We believe that the correct bound for the square error in this case should be, as in Theorem 1, of order $O(T^{2/3} S_T^{1/3} \vee 1)$. In particular, in the case $S_T = O(1)$ we obtain a rate $O(T^{4/5})$ which should be $O(T^{2/3})$. As will be seen in the proof of Proposition 2, the main technical difficulty to obtain what we believe to be the correct bound lies in the fact that the Loewner ordering is not preserved after taking matrix squares, i.e., (3.3.14) does not automatically imply $L^2(\lambda) \succcurlyeq \tilde{L}^2(\lambda)$. On the other hand, under the following assumption on the stacked Laplacian of the comparison graphs, we will obtain the same rate as in the non-evolving case.

Assumption 4. *There exists a constant* $\kappa > 1$ *such that for all* $\lambda > 0$ *it holds*

$$\frac{1}{\kappa}\left(L^2 + \lambda^2 (E^\top E)^2\right) + \lambda(E^\top E L + L E^\top E) \succcurlyeq 0. \tag{3.3.16}$$

Altough technical in nature, we argue that Assumption 4 is reasonable and rather mild. Indeed, notice that given any pair of symmetric matrices (not necessarily p.s.d) A, B of the same size we have that $\frac{1}{\kappa}(A^2 + B^2) + AB + BA \succcurlyeq 0$ for any $\kappa \leqslant 1$. Notice also that in the non-evolving case we have that $E^\top E L + L E^\top E \succcurlyeq 0$ (as can be seen from (3.3.6) and (3.3.7)) which implies in that case the stronger condition that (3.3.16) is satisfied for every $\lambda > 0$ and every $\kappa > 1$ (hence in particular Assumption 4 holds). It is possible that Assumption 4 holds for an absolute constant $\kappa > 1$ for any stacked

Laplacian L of connected graphs G_k, but we are unable to prove it. The following theorem shows that under Assumption 4 we obtain error rates that match those of Theorem 6.

Theorem 8. *Under the hypotheses of Proposition 2, suppose that Assumption 4 holds. Then choosing* $\lambda = \sigma^{\frac{4}{3}}\left(\frac{T}{S_T}\right)^{\frac{2}{3}}$, *it holds with probability greater than* $1-\delta$ *that*

$$\|\hat{z}-z^*\|_2^2 \lesssim \frac{\sigma^{\frac{4}{3}}\bar{\kappa}}{\bar{\kappa}-1} T^{\frac{2}{3}} S_T^{\frac{1}{3}} \left(\left(\frac{1}{\lambda^2_{\min}(L)} \vee 1\right) + \frac{\sqrt{n-1}\|L\|_2\sigma^2\log(1/\delta)}{\lambda^{3/2}_{\min}(L)}\right) + \frac{(n-1)\|L\|_2\sigma^2\log(1/\delta)}{\lambda^2_{\min}(L)},$$

$$\text{where} \quad \bar{\kappa} := \max\left\{\kappa > 1 : \quad \frac{1}{\kappa}\left(L^2+\lambda^2(E^TE)^2\right)+\lambda(E^TEL+LE^TE) \succcurlyeq 0\right\}.$$

Consequently, if $S_T = O(T^\gamma)$ *for some* $\gamma < 1$, *then* $\|\hat{z}-z^*\|_2^2 = O(\sigma^{\frac{4}{3}}T^{\frac{2+\gamma}{3}} \vee \sigma^2)$. *In particular,*

$$\|\hat{z}-z^*\|_2^2 = \begin{cases} O(\sigma^{\frac{4}{3}}T^{\frac{2}{3}}) \text{ if } S_T = O(1) \text{ and } \lambda = \sigma^{\frac{4}{3}}T^{2/3}, \\ O(\sigma^{\frac{4}{3}}T^{\frac{1}{3}}) \text{ if } S_T = O(\frac{1}{T}) \text{ and } \lambda = \sigma^{\frac{4}{3}}T^{4/3}. \end{cases}$$

The proof of Theorem 8 mimics the argument used to prove Theorem 7. It follows from the following bound, which is analogous to the bound in Proposition 2

$$\|\hat{z}-z^*\|_2^2 \leqslant \frac{\bar{\kappa}}{\bar{\kappa}-1}\left(\frac{1}{\lambda^2_{\min}(L)} \vee 1\right)\lambda S_T + \sigma^2\|L\|_2(1+4\log(1/\delta))\left(\sum_{k=0}^{T-1}\frac{n-1}{(\lambda_{\min}(L)+(n-1)\lambda\mu_k)^2}+\frac{n-1}{\lambda^2_{\min}(L)}\right). \tag{3.3.17}$$

Then using the bound in (3.3.15), the optimal choice for λ follows by an elementary calculation. The bound (3.3.17) is proven by using the second statement of Lemma 16 (which controls the bias term), and Lemma 17 is used to bound the variance term (which is unchanged with respect to Theorem 7). These lemmas can be found in Section 3.6.3.

Remark 12 (Connectedness assumption)**.** *We believe that the analysis can likely be extended to the setup where some of the graphs are disconnected (with the union graph* G_U *being connected). The main technical difficulty in this case comes from the fact that the Loewner bound (3.3.11) now does not hold for a strictly positive* β_k *for all* k. *If some of the* β_k*'s in (3.3.11) are allowed to be* 0, *then* $\min_{0\leqslant k\leqslant T}\beta_k = 0$ *which renders the bound (3.3.13) to be not useful. In the proof of Lemma 17, we need to have a good lower bound on the eigenvalues of* $L(\lambda)$ *in order to bound* $\operatorname{Tr}\left(L^{\dagger 2}(\lambda)\right)$, *but this appears to be very challenging. Moreover, while the first part of Lemma 16 will remain unchanged, it is difficult to see how the second part of Lemma 16 can be adapted for this setup.*

3.4 PROJECTION METHOD ANALYSIS

Let us now obtain bounds for the estimation error of the projection method. To prove Theorem B, we will use an analogous scheme to Section 3.3, i.e., we build upon the non-evolving case, by considering that all the comparison graphs are the same. Before proceeding, it will be useful to introduce some notation specific to the projection approach that will be used in the analysis.

NOTATION. Let us define the following quantities.

- $\mathcal{L}_\tau = \{(k,j) \in \{0,\cdots,T\} \times [n] \text{ s.t } \lambda_j \mu_k \leqslant \tau\}$, where $\tau \in [0,\infty]$, corresponds to the indices for the low-frequency part of the spectrum of $E^\top E$. Note that when $\tau = 0$, it corresponds to the indices for the nullspace of $E^\top E$. On the other hand, when $\tau = \infty$, we have $\mathcal{L}_\infty = \{0,\cdots,T\} \times [n]$.
- $\mathcal{H}_\tau = \{0,\cdots,T\} \times [n] \setminus \mathcal{L}_\tau$ corresponds to the indices for the high-frequency part of the spectrum.
- $\mathcal{V}_\tau = \text{span}\{u_k \otimes a_j\}_{(k,j)\in\mathcal{L}_\tau}$ is the linear space spanned by the low frequency eigenvectors. $P_{\mathcal{V}_\tau}$ is the projection matrix for $\mathcal{V}_\tau$.
- $P_{\mathcal{V}_\tau^\perp}$ is the projection matrix for the orthogonal complement of $\mathcal{V}_\tau$, i.e., $\mathcal{V}_\tau^\perp$.

The previous definitions help formalize the description of $\hat{z}$ in (3.1.6) and we can write, in matrix notation,

$$\check{z} = L^\dagger Q y \quad \text{and} \quad \hat{z} = P_{\mathcal{V}_\tau}\check{z}.$$

Notice that $\mathcal{N}(E^\top E) \subset \mathcal{V}_\tau$ and $\check{z} \perp e_k \otimes \mathbb{1}_n$ for each k. This then implies that

$$\hat{z} = P_{\mathcal{V}_\tau}\check{z} \perp (e_k \otimes \mathbb{1}_n), \quad k = 0,\dots,T,$$

which means that $\hat{z}$ is block-wise centered. Now the latent strength vector z^* is given as the solution of the linear system $Lz = Qx^*$ with z^* assumed to be block-centered. If each G_k is connected, then z^* is uniquely given by

$$z^* = L^\dagger Q x^*.$$

Therefore, since

$$z^* = P_{\mathcal{V}_\tau} z^* + P_{\mathcal{V}_\tau^\perp} z^* = P_{\mathcal{V}_\tau} L^\dagger Q x^* + P_{\mathcal{V}_\tau^\perp} z^*,$$

we obtain the following expression for the estimation error

$$\|\hat{z} - z^*\|_2^2 = \|P_{\mathcal{V}_\tau} L^\dagger Q(y - x^*)\|_2^2 + \|P_{\mathcal{V}_\tau^\perp} z^*\|_2^2. \tag{3.4.1}$$

The first term in the RHS of (3.4.1) is the variance term due to noise that will be controlled by choosing τ to be suitably large. The second term is the bias which depends on the smoothness of z^* and will be controlled by choosing τ to be sufficiently small. Hence the optimal choice of τ will be the one which achieves the right bias-variance trade-off.

Remark 13 (Connectedness assumption)**.** *The assumption that each G_k is connected enables us to uniquely write the block-centered z^* as $z^* = L^\dagger Q x^*$. In case some of the graphs were disconnected, then it is not necessary that $L^\dagger Q x^*$ satisfies the smoothness condition in Assumption 2, even though it is a solution of*

$Lz = Qx^*$. *It is unclear how to proceed with the analysis in this case, and we leave it for future work.*

3.4.1 *Non-evolving case*

The following theorem provides a tail bound for the estimation error of the projection method under Assumption 3 (non-evolving case).

Theorem 9. *For any $\delta \in (0, e^{-1})$ and $\tau \geqslant 0$, it holds under Assumptions 2 and 3 that with probability at least $1-\delta$*

$$\|\hat{z} - z^*\|_2^2 \lesssim S_T\Big(\frac{1}{\tau} \wedge \frac{1}{(n-1)\mu_{T-1}}\Big) + \alpha_1 \sigma^2 \log(1/\delta) \left(\frac{T+1}{\pi}\sqrt{\frac{\tau}{n-1}} \sum_{j=1}^{n-1} \frac{1}{\alpha_j^2} + \sum_{j=1}^{n-1} \frac{1}{\alpha_j^2} \right).$$

The proof is detailed in Section 3.6.4. We now show how to optimize the choice of the regularization parameter τ.

CHOICE OF τ. Consider the case $S_T = 0$ first. From Theorem 9, it is clear that, in that case, the optimal choice is to set τ to zero. On the other hand, notice that Theorem 9 can be written using asymptotic notation (hiding all the parameters except for T and σ) as

$$\|\hat{z} - z^*\|_2^2 = O\Big(\frac{S_T}{\tau} + \sigma^2 T\sqrt{\tau}\Big),$$

and hence the optimal choice of τ is given by $\tau = \sigma^{-\frac{4}{3}}\left(\frac{S_T}{T}\right)^{\frac{2}{3}}$. Notice that when $\sigma = 0$ and $S_T \neq 0$, the optimal choice is $\tau = \infty$, which gives a zero error as expected. In the case $\sigma = S_T = 0$, the estimation error is also zero as is obvious from the bound above. Below, we focus on the interesting case where at least one of σ, S_T is non-zero. The following result then follows directly from Theorem 9.

Corollary 4. *Choosing $\tau = \sigma^{-\frac{4}{3}}\left(\frac{S_T}{T}\right)^{\frac{2}{3}}$, then for any $\delta \in (0, e^{-1})$, it holds with probability greater than $1-\delta$ that*

$$\|\hat{z} - z^*\|_2^2 \lesssim \sigma^{\frac{4}{3}} T^{\frac{2}{3}} S_T^{\frac{1}{3}} \Big(1 + \frac{\log(1/\delta)}{\pi\sqrt{n-1}} \sum_{j=1}^{n-1} \frac{\alpha_1}{\alpha_j^2}\Big) + \sigma^2 \log(1/\delta) \sum_{j=1}^{n-1} \frac{\alpha_1}{\alpha_j^2}.$$

Consequently, if $S_T = O(T^\gamma)$ for some $\gamma \leqslant 1$, then $\|\hat{z} - z^*\|_2^2 = O(\sigma^{\frac{4}{3}} T^{\frac{2+\gamma}{3}} \vee \sigma^2)$. In particular,

$$\|\tilde{z} - z^*\|_2^2 = \begin{cases} O(\sigma^{\frac{4}{3}} T^{\frac{2}{3}}) \text{ if } S_T = O(1) \text{ and } \tau = \sigma^{-\frac{4}{3}} T^{-2/3}, \\ O(\sigma^{\frac{4}{3}} T^{\frac{1}{3}}) \text{ if } S_T = O(\frac{1}{T}) \text{ and } \tau = \sigma^{-\frac{4}{3}} T^{-4/3}. \end{cases}$$

3.4.2 *Any sequence of connected graphs*

Similar to the analysis for the least-squares approach in Section 3.3.2, we will use the semi-definite bound (3.3.12) to pass from the case of non-evolving graphs to the case where the graphs can differ (but are connected). Given the bound in the Loewner order (3.3.12), we see that the role played by the eigenpairs (α_j, a_j) in Section 3.4.1 is now replaced by the eigenpairs of the Laplacian of the complete graph. Information about the comparison graphs is also encoded in the scalar $\min_{0\leqslant k\leqslant T}\beta_k$ (or in $\lambda_{\min}(L)$), where $\beta_k(n-1)$ is equal to the Fiedler eigenvalue of L_k. The following bound for the estimation error is analogous to Theorem 9 and Corollary 4, and formalizes the statement of Theorem B.

Theorem 10. *Assume that G_k is connected for all $k \in \{0,\dots,T\}$ and let $\lambda_{\min}(L)$ be the smallest nonzero eigenvalue of L. Then for any $\delta \in (0, e^{-1})$ and $\tau \geqslant 0$, it holds under Assumption 2 and with probability larger than $1-\delta$ that*

$$\|\hat{z}-z^*\|_2^2 \lesssim S_T\left(\frac{1}{\tau}\wedge\frac{1}{(n-1)\mu_{T-1}}\right) + \frac{(T+1)\sigma^2\|L\|_2\log(1/\delta)}{\pi}\sqrt{\frac{\tau}{n-1}}\frac{1}{\lambda_{\min}^2(L)} + \sigma^2\|L\|_2\log(1/\delta)\frac{1}{\lambda_{\min}^2(L)}. \tag{3.4.2}$$

Choosing $\tau = \sigma^{-\frac{4}{3}}\left(\frac{S_T}{T}\right)^{2/3}$ leads to the bound

$$\|\hat{z}-z^*\|_2^2 \lesssim \sigma^{\frac{4}{3}}T^{\frac{2}{3}}S_T^{\frac{1}{3}}\left(1+\frac{\|L\|_2\log(1/\delta)}{\pi\sqrt{n-1}}\frac{1}{\lambda_{\min}^2(L)}\right)+\sigma^2\|L\|_2\log(1/\delta)\frac{1}{\lambda_{\min}^2(L)}. \tag{3.4.3}$$

Thus, if $S_T = O(T^\gamma)$ for some $\gamma \leqslant \sigma^2$, then $\|\tilde{z}-z^*\|_2^2 = O(\sigma^{\frac{4}{3}}T^{\frac{2+\gamma}{3}}\vee 1)$. In particular,

$$\|\hat{z}-z^*\|_2^2 = \begin{cases} O(\sigma^{\frac{4}{3}}T^{\frac{2}{3}}) \text{ if } S_T = O(1) \text{ and } \tau = \sigma^{-\frac{4}{3}}T^{-2/3}, \\ O(\sigma^{\frac{4}{3}}T^{\frac{1}{3}}) \text{ if } S_T = O(\frac{1}{T}) \text{ and } \tau = \sigma^{-\frac{4}{3}}T^{-4/3}. \end{cases}$$

In terms of the notation introduced in Theorem B, we have

$$\Psi_{\text{Proj}}(n,\sigma,\delta) = \sigma^{\frac{4}{3}}\left(1+\frac{\|L\|_2\log(1/\delta)}{\pi\sqrt{n-1}}\frac{1}{\lambda_{\min}^2(L)}\right),$$
$$\Psi'_{\text{Proj}}(n,\sigma,\delta) = \sigma^2\|L\|_2\log(1/\delta)\frac{1}{\lambda_{\min}^2(L)}.$$

3.5 PROOFS

3.6 PROOF OF LEMMA 12

Proof. Recall that $(\lambda_i, v_i)_{i=1}^n$ denote the eigenpairs of $CC^\top$ with $\lambda_1 \geqslant \dots \geqslant \lambda_{n-1} > \lambda_n$ its eigenvalues. Since $CC^\top = nI - \mathbf{1}\mathbf{1}^\top$ we know that $v_n = \mathbf{1}_n$ and $\{v_i\}_{i=1}^{n-1}$ is any orthonormal basis for the space orthogonal to span$(\mathbf{1}_n)$. Also recall that $(\mu_k, u_k)_{k=0}^T$ denote the eigenpairs of $MM^\top$ where $u_T = \mathbf{1}_{T+1}$. Now the eigenvectors of $E^\top E$ that have eigenvalue zero are

- $u_T \otimes v_j = \mathbf{1}_{T+1}\otimes v_j$ for $j = 1,\dots,n-1$, and

- $u_k \otimes v_n = u_k \otimes \mathbf{1}_n$ for $k = 0, \dots, T$.

Note that $\{u_k \otimes \mathbf{1}_n\}_{k=0}^T$ lie in $\mathcal{N}(L + \lambda E^\top E)$ and

$$\text{span}\{u_k \otimes \mathbf{1}_n\}_{k=0}^T = \text{span}\{e_k \otimes \mathbf{1}_n\}_{k=0}^T.$$

Since $L + \lambda E^\top E$ is p.s.d, we have that $\mathcal{N}(L + \lambda E^\top E) = \text{span}\{e_k \otimes \mathbf{1}_n\}_{k=0}^T$ iff

$$x^\top (L_{\tilde{V}} + \lambda E^\top E)x > 0, \quad \forall x \ (\neq 0) \in \text{span}^\perp \{e_k \otimes \mathbf{1}_n\}_{k=0}^T. \tag{3.6.1}$$

As the orthogonal complement of $\text{span}\{e_k \otimes \mathbf{1}_n\}_{k=0}^T$ is given by

$$\text{span}^\perp \{e_k \otimes \mathbf{1}_n\}_{k=0}^T = \text{span}\left\{\mathbf{1}_{T+1} \otimes v_j\right\}_{j=1}^{n-1} \oplus \mathcal{N}^\perp(E^\top E)$$

we claim that (3.6.1) translates to establishing that

$$x^\top L x > 0, \quad \forall x \ (\neq 0) \in \text{span}\left\{\mathbf{1}_{T+1} \otimes v_j\right\}_{j=1}^{n-1}. \tag{3.6.2}$$

To prove this claim, we begin by writing any $x \in \text{span}^\perp \{e_k \otimes \mathbf{1}_n\}_{k=0}^T$ as $x = \tilde{x} + x'$ where $\tilde{x} \in \text{span}\left\{\mathbf{1}_{T+1} \otimes v_j\right\}_{j=1}^{n-1}$ and $x' \in \mathcal{N}^\perp(E^\top E)$. Then if $x \neq 0$,

$$\begin{aligned} x^\top (L + \lambda E^\top E)x &= x^\top L x + \lambda x'^\top E^\top E x' \\ &= \begin{cases} \tilde{x}^\top L \tilde{x}\,; & \text{if } x' = 0, \\ > 0 \ (\text{since } \lambda > 0)\,; & \text{if } x' \neq 0, \end{cases} \end{aligned}$$

which establishes the claim since at least one of $\tilde{x}, x' \neq 0$.

To prove (3.6.2), we first observe that $x^\top (L + \lambda E^\top E)x = x^\top L x$, for any $x \in \text{span}\left\{\mathbf{1}_{T+1} \otimes v_j\right\}_{j=1}^{n-1}$. Since $v_1, \dots, v_{n-1}$ is any orthonormal basis for the subspace orthgonal to $\mathbf{1}_n$, we set $x = \mathbf{1}_{T+1} \otimes v$ for any $v (\neq 0)$ lying in that subspace. This gives us

$$x^\top L x = \sum_{t \in \mathcal{T}} v^\top L_t v = v^\top \left(\sum_{t \in \mathcal{T}} L_t\right) v.$$

Now, $\sum_{t \in \mathcal{T}} L_t$ is the Laplacian of the union graph G_U. Since $\sum_{t \in \mathcal{T}} L_t$ has rank $n - 1$ iff G_U is connected (which is true by assumption), we arrive at (3.6.2). □

3.6.1 *Proof of Lemma 13*

Using the fact that $\alpha_j \geqslant \alpha_{n-1}$ for all $j \in \{1, \cdots, n-1\}$, and the explicit form of the non-zero eigenvalues of the path graph on $T + 1$ vertices (see for e.g. [6])

$$\mu_k = 4 \sin^2 \frac{(T - k)\pi}{2(T + 1)}, \quad k = 0, \dots, T - 1;$$

we have that

$$\sum_{k=0}^{T-1} \sum_{j=1}^{n-1} \frac{1}{(\alpha_j + \lambda(n-1)\mu_k)^2} \leqslant (n - 1) \sum_{k=1}^{T} \frac{1}{(\alpha_{n-1} + 4\lambda(n - 1)\sin^2 \frac{k\pi}{2(T+1)})^2} \tag{3.6.3}$$

The last term in the previous inequality can be regarded as a Riemannian sum, and as such it can be bounded by an integral as follows

$$\frac{1}{T+1}\sum_{k=1}^{T}\frac{1}{(\alpha_{n-1}+4\lambda(n-1)\sin^2\frac{k\pi}{2(T+1)})-2}\leqslant\int_0^1\frac{dx}{(\alpha_{n-1}+4\lambda(n-1)\sin^2\frac{\pi x}{2})^2}.$$

Let us focus on the integral term, for which the following bound holds

$$\begin{aligned}\int_0^1\frac{dx}{(\alpha_{n-1}+4\lambda(n-1)\sin^2\frac{\pi x}{2})^2}&\overset{(1)}{=}2\int_0^{\frac{1}{2}}\frac{dx}{(\alpha_{n-1}+4\lambda(n-1)\sin^2\pi x)^2}\\&\overset{(2)}{\leqslant}2\int_0^{\frac{1}{2}}\frac{dx}{(\alpha_{n-1}+\lambda(n-1)\pi^2x^2)^2}\\&\overset{(3)}{=}\frac{2}{\pi\alpha_{n-1}^{3/2}\sqrt{\lambda(n-1)}}I\Big(\frac{\pi}{2}\sqrt{\frac{\lambda(n-1)}{\alpha_{n-1}}}\Big),\end{aligned}$$

where $I(t)=\int_0^t\frac{du}{(1+u^2)^2}$. Equality (1) comes from a change of variables. In (2), we used that $\sin x\geqslant x/2$ for all $x\in[0,\pi/2]$ and (3) results from the change of variable $u=\sqrt{\frac{\lambda(n-1)}{\alpha_{n-1}}}\pi x$. By an elementary calculation, we verify that $I(t)=\frac{\arctan t}{2}+\frac{t}{2(1+t^2)}$. Hence,

$$\begin{aligned}\int_0^1\frac{dx}{(\alpha_{n-1}+4\lambda(n-1)\sin^2\pi x)^2}&\leqslant\frac{1}{\alpha_{n-1}^{3/2}}\frac{\arctan(\frac{\pi}{2}\sqrt{\frac{\lambda(n-1)}{\alpha_{n-1}}})}{\pi\sqrt{\lambda(n-1)}}\\&\quad+\frac{1}{2(\alpha_{n-1}^2+\frac{\pi^2}{2}\alpha_{n-1}(n-1)\lambda)}\\&\lesssim\frac{1}{\alpha_n^{3/2}\sqrt{(n-1)\lambda}}\\&\quad+\frac{1}{\alpha_{n-1}^2+\frac{\pi^2}{4}\alpha_{n-1}(n-1)\lambda},\end{aligned}$$

where the last inequality follows from $|\arctan(x)|\leqslant\frac{\pi}{2}$ for all $x\in\mathbb{R}$. From this, we deduce that

$$\begin{aligned}\sum_{k=0}^{T-1}\sum_{j=1}^{n-1}\frac{1}{(\alpha_j+\lambda(n-1)\mu_k)^2}&\lesssim\frac{T\sqrt{n-1}}{\alpha_{n-1}^{3/2}\sqrt{\lambda}}+\frac{T(n-1)}{\alpha_{n-1}^2+\alpha_{n-1}(n-1)\lambda}\\&=T(n-1)\\&\quad\times\left(\frac{1}{\alpha_{n-1}\sqrt{\alpha_{n-1}(n-1)\lambda}}+\frac{1}{\alpha_{n-1}^2+\alpha_{n-1}(n-1)\lambda}\right)\\&\lesssim\frac{T(n-1)}{\alpha_{n-1}\sqrt{\alpha_{n-1}(n-1)\lambda}}=\frac{T\sqrt{n-1}}{\alpha_{n-1}^{3/2}\sqrt{\lambda}}.\end{aligned}$$

3.6.2 *Proof of Proposition 1*

The proof of Proposition 1 follows from Lemmas 14 and 15 below.

Lemma 14 (Bound on the variance term). *For any* $\delta \in (0, e^{-1})$, *it holds with probability larger than* $1-\delta$ *that*

$$\|L^{\dagger}(\lambda)Q(y-x^*)\|_2^2 \leqslant \sigma^2\alpha_1(1+4\log(1/\delta))\Big(\sum_{k=0}^{T}\frac{1}{(\alpha_j+\lambda(n-1)\mu_k)^2}+\sum_{j=1}^{n-1}\frac{1}{\alpha_j^2}\Big).$$

Proof. Define $\Sigma := (L^{\dagger}(\lambda)Q)^{\top}L^{\dagger}(\lambda)Q = Q^{\top}L^{\dagger 2}(\lambda)Q$. Using [20, Thm.2.1], it holds for any $c > 1$

$$\begin{aligned}\|L^{\dagger}(\lambda)Q(y-x^*)\|_2^2 &\leqslant \sigma^2(\mathrm{Tr}(\Sigma)+2\sqrt{\mathrm{Tr}(\Sigma^2)c}+2\|\Sigma\|_2 c)\\ &\leqslant \sigma^2(1+4c)\,\mathrm{Tr}(\Sigma)\end{aligned} \tag{3.6.4}$$

with probability $1-e^{-c}$ (using the fact that $\mathrm{Tr}(\Sigma^2)\vee\|\Sigma\|_2\,\mathrm{Tr}(\Sigma) \leqslant \mathrm{Tr}(\Sigma)$). On the other hand,

$$\mathrm{Tr}(\Sigma) = \mathrm{Tr}(QQ^{\top}L^{\dagger 2}(\lambda)) \leqslant \|QQ^{\top}\|_2\,\mathrm{Tr}(L^{\dagger 2}(\lambda))$$

where we used the fact that given symmetric p.s.d matrices A, B it holds $\mathrm{Tr}(AB) \leqslant \|A\|_2\,\mathrm{Tr}(B)$. From the spectral decomposition of $L(\lambda)$ (given by (3.3.8)) we deduce that

$$\mathrm{Tr}(L^{\dagger 2}(\lambda)) = \sum_{k=0}^{T-1}\sum_{j=1}^{n-1}\frac{1}{(\alpha_j+\lambda(n-1)\mu_k)^2}+\sum_{j=1}^{n-1}\frac{1}{\alpha_j^2}.$$

Using $\left\|QQ^{\top}\right\|_2 = \alpha_1$ and $c = \log(1/\delta)$ concludes the proof. □

Lemma 15 (Bound on the bias term). *We have*

$$\lambda^2\|L^{\dagger}(\lambda)E^{\top}Ez^*\|_2^2 \leqslant \Big(\frac{1}{\alpha_{n-1}^2}\vee 1\Big)\lambda S_T.$$

Proof. From (3.3.6) and (3.3.8) it is easy to see that

$$\lambda^2\|L^{\dagger}(\lambda)E^{\top}Ez^*\|_2^2 = \sum_{k=0}^{T-1}\sum_{j=1}^{n-1}\Big(\frac{\lambda(n-1)\mu_k}{\alpha_j+\lambda(n-1)\mu_k}\Big)^2\langle z^*, u_k\otimes a_j\rangle^2.$$

On the other hand, the smoothness condition in Assumption 2 can be written as

$$\|Ez^*\|_2^2 = \sum_{k=0}^{T-1}\sum_{j=1}^{n-1}(n-1)\mu_k\langle z^*, u_k\otimes a_j\rangle^2 \leqslant S_T.$$

Defining $b_k := \sum_{j=1}^{n-1}\langle z^*, u_k\otimes a_j\rangle^2$ and using that $\alpha_1 \geqslant \cdots \geqslant \alpha_{n-1} > 0$ we obtain

$$\begin{aligned}\lambda^2\|L^{\dagger}(\lambda)E^{\top}Ez^*\|_2^2 &\leqslant \sum_{k=0}^{T-1}\Big(\frac{\lambda(n-1)\mu_k}{\alpha_{n-1}+\lambda(n-1)\mu_k}\Big)^2 b_k,\\ \lambda\|Ez^*\|_2^2 &= \sum_{k=0}^{T-1}\lambda(n-1)\mu_k b_k \leqslant \lambda S_T.\end{aligned}$$

We now conclude using the following elementary fact with $c = \lambda(n-1)\mu_k$ and $d = \alpha_{n-1}$.

Claim 1. *Let* c, d *be strictly positive real numbers, then*

$$\left(\frac{c}{d+c}\right)^2 \leqslant \left(\frac{1}{d^2} \vee 1\right)c.$$

Proof. We have two cases. For $c \geqslant 1$, we have $\left(\frac{c}{d+c}\right)^2 < 1$ and $\left(\frac{1}{d^2} \vee 1\right)c \geqslant 1$, so the inequality is verified. In the case $c < 1$, we have

$$\left(\frac{c}{d+c}\right)^2 \leqslant \frac{c^2}{d^2} \leqslant \frac{c}{d^2} \leqslant \left(\frac{1}{d^2} \vee 1\right)c.$$

□

□

3.6.3 *Proof of Proposition 2*

Using (3.3.14), the proof of Proposition 2 goes along the same lines as that of Proposition 1. It follows directly from Lemmas 16 (the first statement) and 17 below, which offer a control of the bias and the variance term, respectively.

Lemma 16. *The following is true.*

1. *It holds that*

$$\lambda^2 \|L^\dagger(\lambda)E^T E z^*\|_2^2 \lesssim \frac{\lambda^2 n}{\lambda_{\min}^2(L)} S_T.$$

2. *If* G_k *is connected for all* $k = 0, \cdots, T$ *and* L *satisfies Assumption 4, then for all* $\lambda > 0$ *we have*

$$\lambda^2 \|L^\dagger(\lambda)E^T E z^*\|_2^2 \lesssim \frac{\overline{\kappa}}{\overline{\kappa}-1}\left(\frac{1}{\lambda_{\min}^2(L)} \vee 1\right)\lambda S_T$$

where

$$\overline{\kappa} := \max\left\{\kappa > 1 \text{ } s.t \text{ } \frac{1}{\kappa}(L^2 + \lambda^2(E^T E)^2) + \lambda(E^T E L + L E^T E) \succcurlyeq 0\right\}.$$

Proof. We have

$$\begin{aligned}\lambda^2 \|L^\dagger(\lambda)E^T E z^*\|_2^2 &\leqslant \lambda^2 \|L^\dagger(\lambda)E^T\|_2^2 \|E z^*\|_2^2 \\ &\leqslant \lambda^2 \|L^\dagger(\lambda)\|_2^2 \|E^T\|_2^2 \|E z^*\|_2^2 \\ &\lesssim \frac{\lambda^2 n}{\lambda_{\min}^2(L)} S_T\end{aligned}$$

where in the first inequality we used the submultiplicativity of the operator norm and in the last step we used that $\|E^T\|_2^2 \lesssim n$, $\|Ez^*\|_2^2 \leqslant S_T$ by the smoothness assumption, and $\|L^\dagger(\lambda)\|_2^2 \leqslant \frac{1}{\lambda_{\min}^2(L)}$. This proves the first statement.

To prove the second part, we assume that Assumption 4 holds and $\overline{\kappa}$ is the largest value for which it holds. From this it follows that

$$
\begin{aligned}
(L+\lambda E^T E)^2 &\succcurlyeq \frac{\overline{\kappa}-1}{\overline{\kappa}}(L^2+\lambda^2(E^T E)^2) \\
&\succcurlyeq \frac{\overline{\kappa}-1}{\overline{\kappa}}\left(\frac{\lambda^2_{\min}(L)}{(n-1)^2}\tilde{L}^2+\lambda^2(E^T E)^2\right),
\end{aligned}
$$

where in the last inequality we used (3.3.12) with $\min_{0\leqslant k\leqslant T}\beta_k = \frac{\lambda_{\min}(L)}{n-1}$ and the fact that the eigenvectors of L and $\tilde{L}$ are aligned (because $\tilde{L} = I_{T+1}\otimes L_{com}$ and the eigenvector of L_{com} can be aligned with those of any Laplacian). Since $\lambda > 0$, then by recalling Lemma 12, one can verify that $(L+\lambda E^T E)^2$ and $\tilde{L}^2+\lambda^2(E^T E)^2$ have the same null space. This in turn implies (using Lemma 23) that

$$
\frac{\overline{\kappa}}{\overline{\kappa}-1}\left(\frac{\lambda^2_{\min}(L)}{(n-1)^2}\tilde{L}^2+\lambda^2(E^T E)^2\right)^{\dagger} \succcurlyeq (L+\lambda E^T E)^{\dagger 2} = L^{\dagger 2}(\lambda).
$$

Now arguing similarly to the proof of Lemma 15, we have

$$
\begin{aligned}
\|L^{\dagger}(\lambda)E^T E z^*\|_2^2 &= z^{*T}E^T E L^{\dagger 2}(\lambda)E^T E z^* \\
&\leqslant \frac{\overline{\kappa}}{\overline{\kappa}-1} z^{*T}E^T E\left(\frac{\lambda^2_{\min}(L)}{(n-1)^2}\tilde{L}^2+\lambda^2(E^T E)^2\right)^{\dagger} E^T E z^*, \\
&\overset{(1)}{=} \frac{\overline{\kappa}}{\overline{\kappa}-1}\sum_{k=0}^{T-1}\sum_{j=1}^{n-1}\frac{\lambda^2(n-1)^2\mu_k^2}{\lambda^2_{\min}(L)+\lambda^2(n-1)^2\mu_k^2}\langle z^*, u_k\otimes a_j\rangle^2 \\
&\overset{(2)}{\leqslant} \frac{\overline{\kappa}}{\overline{\kappa}-1}\Big(\frac{1}{\lambda^2_{\min}(L)}\vee 1\Big)\sum_{j=1}^{n-1}\lambda(n-1)\mu_k\langle z^*, u_k\otimes a_j\rangle^2, \\
&\overset{(3)}{\leqslant} \frac{\overline{\kappa}}{\overline{\kappa}-1}\Big(\frac{1}{\lambda^2_{\min}(L)}\vee 1\Big)\lambda S_T
\end{aligned}
$$

where in (1) we used the known spectral expansion of $\tilde{L}$ and $E^T E$, in (2) we used Claim 1 (used in the proof of Lemma 15) with the fact that $\frac{c^2}{c^2+d^2} \leqslant (\frac{c}{c+d})^2$ for $c, d > 0$, and in (3) we used the smoothness assumption. □

When the comparison graphs are not necessarily the same, Lemma 14 is replaced by the following lemma.

Lemma 17. *For any $\delta \in (0, e^{-1})$, it holds with probability greater than $1-\delta$ that*

$$
\begin{aligned}
\|L^{\dagger}(\lambda)Q(y-x^*)\|_2^2 &\leqslant \sigma^2\|L\|_2(1+4\log(1/\delta)) \\
&\times\Big(\sum_{k=0}^{T-1}\frac{n-1}{(\lambda_{\min}(L)+(n-1)\lambda\mu_k)^2}+\frac{n-1}{\lambda^2_{\min}(L)}\Big).
\end{aligned}
$$

Proof. Reasoning as in the proof of Lemma 14 (using [20, Thm. 2.1]) we obtain with probability at least $1-\delta$ that

$$
\|L^{\dagger}Q(y-x^*)\|^2 \leqslant \sigma^2(1+4\log(1/\delta))\,\mathrm{Tr}\left(LL^{\dagger 2}(\lambda)\right). \tag{3.6.5}
$$

Recall from (3.3.14) that $L(\lambda) \succcurlyeq \tilde{L}(\lambda)$, hence we deduce that

$$\mathrm{Tr}\left(LL^{\dagger 2}(\lambda)\right) \leqslant \|L\|_2 \, \mathrm{Tr}\left(L^{\dagger 2}(\lambda)\right) \leqslant \|L\|_2 \, \mathrm{Tr}\left(\tilde{L}^{\dagger 2}(\lambda)\right).$$

Since $\tilde{L}^{\dagger}(\lambda) = \left(\frac{1}{n-1}\lambda_{\min}(L)I_{T+1} + \lambda MM^{\top}\right) \otimes L_{\mathrm{com}}$, it is easy to verify that the nonzero eigenvalues of $\tilde{L}(\lambda)$ correspond to the set $\{\lambda_{\min}(L) + \lambda\mu_k(n-1)\}_{k=0}^{T-1} \cup \{\lambda_{\min}(L)\}$, and each eigenvalue has multiplicity $n-1$. Hence,

$$\mathrm{Tr}\left(\tilde{L}^{\dagger 2}(\lambda)\right) = \sum_{k=0}^{T-1} \frac{n-1}{(\lambda_{\min}(L) + \lambda\mu_k(n-1))^2} + \frac{n-1}{\lambda_{\min}^2(L)}.$$

Plugging this into (3.6.5), the result follows. □

3.6.4 *Proof of Theorem 9*

The proof of Theorem 9 follows directly from the two lemmas below.

Lemma 18. *If z^* is such that $\|Ez^*\|_2^2 \leqslant S_T$, and $\tau \geqslant 0$ then $\|P_{\mathcal{V}_\tau^\perp} z^*\|_2^2 \leqslant S_T(\frac{1}{\tau} \wedge \frac{1}{(n-1)\mu_{T-1}})$.*

Proof. By Parseval's identity,

$$\|P_{\mathcal{V}_\tau^\perp} z^*\|_2^2 = \sum_{(k,j)\in\mathcal{H}_\tau} \langle z^*, u_k \otimes a_j\rangle^2.$$

On the other hand,

$$\begin{aligned}\sum_{(k,j)\in\mathcal{H}_\tau} (n-1)\mu_k\langle z^*, u_k \otimes a_j\rangle^2 &\leqslant \sum_{k=0}^{T-1}\sum_{j=1}^{n-1} (n-1)\mu_k\langle z^*, u_k \otimes a_j\rangle^2 \\ &= \|Ez^*\|_2^2 \leqslant S_T,\end{aligned}$$

where we used the fact $\mu_k \geqslant 0$ in the first inequality, and the assumption $\|Ez^*\|_2^2 \leqslant S_T$ in the second inequality. We first assume that $\tau > 0$. Since $(n-1)\mu_k \geqslant \tau$ for all $(k,j) \in \mathcal{H}_\tau$, we deduce that

$$\|P_{\mathcal{V}_\tau^\perp} z^*\|_2^2 \leqslant \frac{S_T}{\tau}.$$

On the other hand, we have the following inequalities

$$(n-1)\mu_{T-1}\|P_{\mathcal{V}_\tau^\perp} z^*\|^2 \leqslant \|EP_{\mathcal{V}_\tau^\perp} z^*\|^2 \leqslant \|Ez^*\|^2 \leqslant S_T,$$

from which the lemma follows.

□

Lemma 19. *For any $\delta \in (0, e^{-1})$, it holds with probability greater than $1-\delta$*

$$\|P_{\mathcal{V}_\tau} L^{\dagger} Q(y - x^*)\|_2^2 \leqslant 2\sigma^2\alpha_1(1 + 4\log(1/\delta)) \left(\frac{T+1}{\pi}\sqrt{\frac{\tau}{n-1}\sum_{j=1}^{n-1}\frac{1}{\alpha_j^2}} + \sum_{j=1}^{n-1}\frac{1}{\alpha_j^2}\right).$$

Proof. Define

$$\Sigma := (P_{\mathcal{V}_\tau} L^{\dagger} Q)^{\top} P_{\mathcal{V}_\tau} L^{\dagger} Q = Q^{\top} L^{\dagger} P_{\mathcal{V}_\tau} L^{\dagger} Q.$$

Using [20, Thm.2.1], we have for all $c > 1$ and with probability larger than $1 - e^{-c}$ that

$$\|P_{\mathcal{V}_\tau} L^\dagger Q(y - x^*)\|_2^2 \leqslant \sigma^2(1+4c)\operatorname{Tr}(\Sigma). \tag{3.6.6}$$

By the cyclic property of the trace, it holds that

$$\operatorname{Tr}(\Sigma) = \operatorname{Tr}(QQ^\top L^\dagger P_{\mathcal{V}_\tau} L^\dagger) \leqslant \|QQ^\top\|_2 \operatorname{Tr}(L^\dagger P_{\mathcal{V}_\tau} L^\dagger), \tag{3.6.7}$$

using the same property of p.s.d matrices as in the proof of Lemma 14. On the other hand,

$$P_{\mathcal{V}_\tau} = \sum_{(k,j)\in\mathcal{L}_\tau} (u_k \otimes a_j)(u_k \otimes a_j)^\top \text{ and } L^\dagger = \sum_{k=0}^{T}\sum_{j=1}^{n-1} \frac{1}{\alpha_j}(u_k \otimes a_j)(u_k \otimes a_j)^\top.$$

Recall that $\mathcal{L}_\tau = \{(k,j) \in \{0,\cdots,T\} \times [n] \text{ s.t } \lambda_j \mu_k \leqslant \tau\}$ can be written as $\mathcal{L}_1 \cup \mathcal{L}_2$, where

$$\begin{aligned}
\mathcal{L}_1 &= \{k = 0,\cdots,T-1 \text{ s.t } (n-1)\mu_k \leqslant \tau\} \times [n-1],\\
\mathcal{L}_2 &= \{(k,j) \in \{0,\cdots,T\} \times [n] \text{ s.t } \lambda_j\mu_k = 0\} = (\{T\}\times[n]) \cup (\{0,\ldots,T\}\times\{n\}).
\end{aligned}$$

Thus,

$$P_{\mathcal{V}_\tau} L^{\dagger 2} = \sum_{(k,j)\in\mathcal{L}_1} \frac{1}{\alpha_j^2}(u_k \otimes a_j)(u_k \otimes a_j)^\top + \sum_{j=1}^{n-1} \frac{1}{\alpha_j^2}(\mathbf{1}_T \otimes a_j)(\mathbf{1}_T \otimes a_j)^\top.$$

and we obtain the expression

$$\operatorname{Tr}(P_{\mathcal{V}_\tau} L^{\dagger 2}) = \Big(|\{k = 0,\cdots,T-1 \text{ s.t } (n-1)\mu_k \leqslant \tau\}| + 1\Big) \times \sum_{j=1}^{n-1} \frac{1}{\alpha_j^2}. \tag{3.6.8}$$

Recall that $\mu_k = 4\sin^2 \frac{(T-k)\pi}{2(T+1)}$ for $k = 0,\ldots,T-1$ and define the set

$$\mathcal{A} := \{k = 0,\cdots,T-1 \text{ s.t } 4(n-1)\sin^2 \frac{(T-k)\pi}{2(T+1)} \leqslant \tau\}.$$

Now since $\sin x \geqslant x/2$ for $x \in [0,\pi/2]$, we know that

$$\sin^2 \frac{(T-k)\pi}{2(T+1)} \geqslant \frac{(T-k)^2\pi^2}{4(T+1)^2}.$$

This in turn implies that $|\mathcal{A}|$ can be bounded as

$$\begin{aligned}
|\mathcal{A}| &\leqslant \left|\left\{k = 1,\ldots,T : 4(n-1)\frac{\pi^2k^2}{4(T+1)^2} \leqslant \tau\right\}\right|\\
&\leqslant \frac{T+1}{\pi}\sqrt{\frac{\tau}{n-1}}.
\end{aligned}$$

Using this in (3.6.8), we obtain

$$\operatorname{Tr}(P_{\mathcal{V}_\tau} L^{\dagger 2}) \leqslant \left(\frac{T+1}{\pi}\right)\sqrt{\frac{\tau}{n-1}}\sum_{j=1}^{n-1}\frac{1}{\alpha_j^2} + \sum_{j=1}^{n-1}\frac{1}{\alpha_j^2}.$$

Combining this with (3.6.6) and (3.6.7), the result follows by taking $c = \log(1/\delta)$. □

3.6.5 *Proof of Theorem 10*

We first observe that Lemma 18 works under Assumption 2 and hence applies verbatim in this context. We now formulate a result analogous to Lemma 19 dropping the assumption that all the graphs are the same and using the semi-definite bound (3.3.12).

Lemma 20. *Assume that G_k is connected for all $k = 0, \dots, T$ and let $\beta_k(n-1)$ be the Fiedler eigenvalue of L_k. For any $\delta \in (0, e^{-1})$, it holds with probability larger than $1-\delta$*

$$\|P_{\mathcal{V}_\tau} L^\dagger Q(y - x^*)\|_2^2 \leqslant \sigma^2 \|L\|_2 \frac{1}{\min_{0\leqslant k\leqslant T} \beta_k^2}(1 + 4\log(1/\delta)) \times \left(\frac{T+1}{\pi(n-1)}\sqrt{\frac{\tau}{n-1}} + \frac{1}{n-1}\right).$$

Proof. Let $\Sigma := (P_{\mathcal{V}_\tau} L^\dagger Q)^\top P_{\mathcal{V}_\tau} L^\dagger Q = Q^\top L^\dagger P_{\mathcal{V}_\tau} L^\dagger Q$ and recall that $y - x^*$ is subgaussian with parameter σ. Using [20, Thm.2.1] we obtain for any $c > 1$ that with probability larger than $1 - e^{-c}$

$$\begin{aligned}\|P_{\mathcal{V}_\tau} L^\dagger Q(y - x^*)\|_2^2 &\leqslant \sigma^2(\mathrm{Tr}(\Sigma) + 2\sqrt{\mathrm{Tr}(\Sigma^2)c} + 2\|\Sigma\|_2 c) \\ &\leqslant \sigma^2(1+4c)\,\mathrm{Tr}(\Sigma) \leqslant \sigma^2(1+4c)\|QQ^\top\|_2\,\mathrm{Tr}(L^\dagger P_{\mathcal{V}_\tau} L^\dagger),\end{aligned} \tag{3.6.9}$$

where to pass from the first line to the second line we used the fact that $c > 1$ and $\mathrm{Tr}(\Sigma^2) \vee \|\Sigma\|_2\,\mathrm{Tr}(\Sigma) \leqslant \mathrm{Tr}(\Sigma)^2$. Notice that up until this point, the proof does not change with respect to the proof of Lemma 19. Recall from (3.3.12) that $\tilde{L} = I_{T+1} \otimes L_{\mathrm{com}}$ and $L \succcurlyeq (\min_{0\leqslant k\leqslant T} \beta_k)\tilde{L}$. Since the eigenvectors of $L, \tilde{L}$ are aligned (because the eigenvector of L_{com} can be aligned with those of any Laplacian), it follows that

$$\frac{1}{\min_{0\leqslant k\leqslant T} \beta_k^2}\tilde{L}^{\dagger 2} \succcurlyeq L^{\dagger 2}. \tag{3.6.10}$$

Using (3.6.10) and the fact that $P_{\mathcal{V}_\tau}$ is a projection matrix, we obtain

$$\mathrm{Tr}(L^\dagger P_{\mathcal{V}_\tau} L^\dagger) = \mathrm{Tr}(P_{\mathcal{V}_\tau} L^{\dagger 2}) \leqslant \frac{1}{\min_{0\leqslant k\leqslant T} \beta_k^2}\,\mathrm{Tr}(P_{\mathcal{V}_\tau}\tilde{L}^{\dagger 2}). \tag{3.6.11}$$

where the second inequality follows from (3.6.10) and the fact that $\mathrm{Tr}(AB) \geqslant 0$ if $A, B \succcurlyeq 0$. We also have that

$$P_{\mathcal{V}_\tau}\tilde{L}^{\dagger 2} = \sum_{(k,j)\in\mathcal{L}_1} \frac{1}{(n-1)^2}(u_k \otimes a_j)(u_k \otimes a_j)^\top$$

hence proceeding in a manner similar to the proof of Theorem 9 leads to the bound

$$\begin{aligned}\mathrm{Tr}(P_{\mathcal{V}_\tau}\tilde{L}^{\dagger 2}) &= \left(|\{k=0,\cdots,T-1 \text{ s.t } (n-1)\mu_k \leqslant \tau\}|+1\right) \times \frac{1}{n-1} \\ &\leqslant \frac{T+1}{\pi(n-1)}\sqrt{\frac{\tau}{n-1}} + \frac{1}{n-1}.\end{aligned}$$

Combining this with (3.6.9) and (3.6.11), the result follows. □

Proof of Theorem 10. The first inequality of Theorem 10 follows directly from Lemmas 18 and 20. Inequality (3.4.2) follows by choosing τ as in Section 3.4.1. The rest of the proof is a direct consequence of the previous result. □

3.7 EXPERIMENTS

We now empirically evaluate the performance of the proposed methods on both synthetic and real data. The code for all the experiments is available at `https://github.com/karle-eglantine/Dynamic_TranSync`.

3.7.1 *Synthetic data*

GENERATE THE GROUND TRUTH z^*. The first step is to generate the true weight vector z^* such that it satisfies Assumption 2, for different regimes of S_T. Let us recall that this assumption can be written as

$$\|Ez^*\|_2^2 = z^{*\top}E^\top E z^* \leqslant S_T$$

which implies that z^* lies close to the null space of $E^\top E$, as quantified by S_T. Hence, one possibility to simulate z^* is to project any vector onto a space $\mathcal{V}_\varepsilon$ generated by the eigenvectors associated with the smallest eigenvalues of $E^\top E$. Likewise, $\mathcal{V}_\varepsilon$ contains the null space of $E^\top E$ and a few more eigenvectors, depending on S_T. For $\varepsilon > 0$, let us define

$$\mathcal{L}_\varepsilon := \left\{(k,j) \in \{0,\ldots,T\} \times [n] \text{ s.t. } \lambda_j \mu_k \leqslant \varepsilon\right\},$$

then $\mathcal{V}_\varepsilon = \mathrm{span}\{u_k \otimes a_j\}_{(k,j)\in\mathcal{L}_\tau}$. Using similar tools as in the proof of Theorem 9, it can be shown that

$$|\mathcal{L}_\varepsilon| \leqslant T + n + \frac{\sqrt{(n-1)\varepsilon}(T+1)}{\pi}.$$

Hence, we can compute a value of ε such that any vector belonging to $\mathcal{V}_\varepsilon$ satisfies Assumption 2. More precisely, choosing $\varepsilon = \left(\frac{\pi S_T}{(T+1)\sqrt{n-1}}\right)^{2/3}$, it then holds that for any $z \in \mathbb{R}^{n(T+1)}$ such that $\|z\|_2 = 1$,

$$\|EP_{\mathcal{V}_\varepsilon} z\|_2^2 \leqslant S_T,$$

where $P_{\mathcal{V}_\varepsilon}$ denotes the projection matrix onto the space $\mathcal{V}_\varepsilon$. The synthetic data is then generated as follows.

1. Generate $z \sim \mathcal{N}(0, I_{n(T+1)})$ and normalize it such that $\|z\|_2 = 1$.

2. Compute $k_\varepsilon = \lceil T + n + \left(\frac{T+1}{(n-1)\pi}\right)^{2/3} S_T^{1/3} \rceil$ eigenvectors of $E^\top E$, corresponding to its k_ε smallest eigenvalues.
3. Compute $z^* = P_{\mathcal{V}_\varepsilon} z$.
4. Center each block z^*_k, $k = 0, \ldots, T$.

GENERATE THE OBSERVATION DATA. The observations consist of comparison graphs G_t for $t \in \mathcal{T}$ and the associated measurements $y_{ij}(t)$ for each edge $\{i, j\} \in \mathcal{E}_t$.

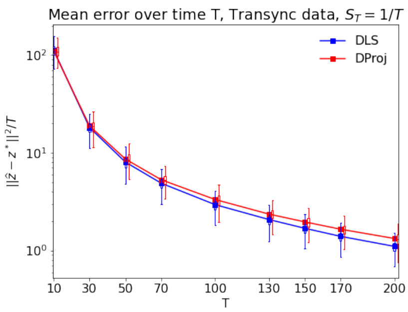
1. For each $t \in \mathcal{T}$, G_t is generated as an Erdös-Renyi graph $G(n, p(t))$, where the probability $p(t)$ is chosen appropriately, as described in Figs. 10, 11 and 12. Meaningful recovery of z^* is only possible if the union of all the graphs is connected, hence we ensure the connectivity of $G_U = (n, \cup_{t \in \mathcal{T}} \mathcal{E}_t)$.
2. For all $t \in \mathcal{T}$ and $\{i, j\} \in \mathcal{E}_t$, we generate a noisy measurement $y_{ij}(t)$ of the strengths difference $z^*_{t,i} - z^*_{t,j}$ as in (3.1.1), using a standard Gaussian noise.

Once the true weights and the observations are generated, one can easily implement our methods using traditional least-squares solvers. In our experiments, we use the least-squares solver `lsqr` from the `scipy` package of Python. Note that in both of our methods, there is a tuning parameter – λ for the Penalized Least-Squares (denoted as DLS later) and τ for the Projection method (denoted DProj). Throughout, we choose $\lambda = (T/S_T)^{2/3}$ and $\tau = (S_T/T)^{2/3}$.

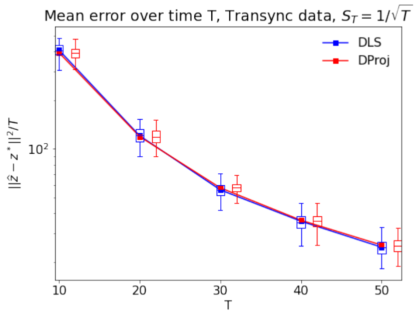
DYNAMIC BTL SET UP. As noted in Remark 6, the Dynamic TranSync model is linked to the Dynamic BTL model. Hence, we will also test numerically the performances of our methods on synthetic data generated according to the BTL model. The ground truth $w^* = \exp(z^*)$ is generated similarly as in the Dynamic TranSync setup, as well as the observation graphs. However the measurements $y_{ij}(t)$ are now generated according to the Dynamic BTL model (see [24] for more details). In this particular setup, we will compare the performance of our methods with two other approaches that focus on dynamic ranking for the BTL model, namely, the Maximum-Likelihood Estimation (MLE) method [4] and the Dynamic Rank Centrality (DRC) method [24]. These methods have shown optimal results when the strength of each item is a Lipschitz function of time, which translates to $S_T = \frac{1}{T}$ in our set up. More generally, the regime of interest in our smoothness assumption is $S_T = o(T)$.

RESULTS. The results are summarized below.

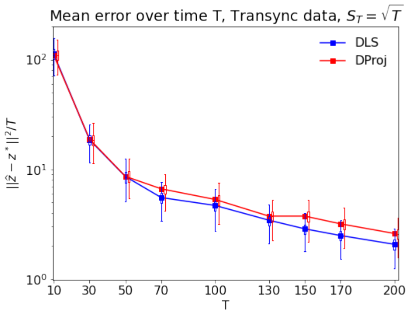
1. In Figure 10, we consider $n = 100$ items and T ranging from 10 to 200, where the data is generated according to the Dynamic TranSync model. The estimation errors are averaged over 40 bootstrap simulations and plotted for different regimes of smoothness S_T. We observe that in every case, DLS and DProj methods give very similar results. As expected, the MSE decreases to zero as T increases in every smoothness regime. Note that the variance of the error, represented by the vertical bars, is also decreasing with T.



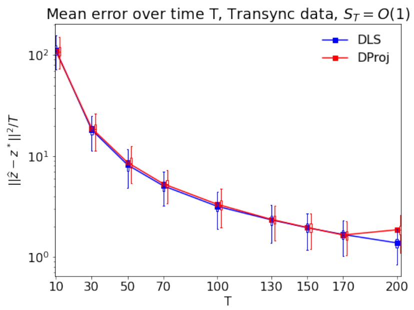
(a) Smoothness $S_T = \frac{1}{T}$



(b) Smoothness $S_T = \frac{1}{\sqrt{T}}$



(c) Smoothness $S_T = \sqrt{T}$



(d) Smoothness $S_T = 1$

Figure 10: MSE versus T for DLS and DProj when the data is generated according to the Dynamic TranSync model for $n = 100$ and graphs are generated as $G(n, p(t))$ with $p(t)$ chosen randomly between $\frac{1}{n}$ and $\frac{\log(n)}{n}$. The results are averaged over the grid $\mathcal{T}$ as well as 40 Monte Carlo runs.

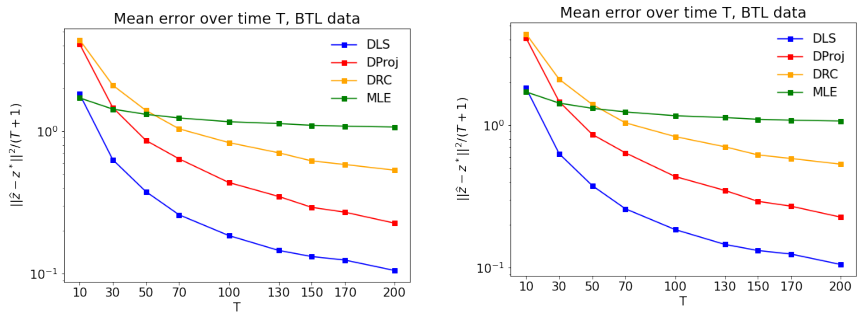
2. In Figure 11, we repeat the same experiments using data simulated according to the Dynamic BTL Model, and compare our results with the DRC and MLE methods for $S_T = \frac{1}{T}$ and $S_T = \frac{1}{\sqrt{T}}$. For both smoothness choices, DLS seems to be the best performing method as its MSE goes to zero at the fastest rate.

3. As a sanity check, we also show in Figure 12 that our methods perform better than the naive least-squares (LS) approach that simply estimates the strength vector individually on each graph G_k. In this case, one needs to impose connectivity on each graph in order to obtain meaningful results using LS. Figure 12 shows that as expected, the MSE is constant with T for the LS method. This illustrates that even when all the graphs are connected, dynamic approaches are better suited to recover the strengths and/or ranking of a set of items.

4. We show in Figure 13 the influence of the sparsity of the input graphs on the estimation. We generate input graphs as $\mathcal{G}(n, p)$ for different values of p and observe that the MSE increases with the sparsity for the DLS method. The Projection method however has a similar performance for all sparsity levels.

5. In Figure 14, we show that the optimal values for the hyperparameters λ and τ derived theoretically are also numerically optimal. Indeed for

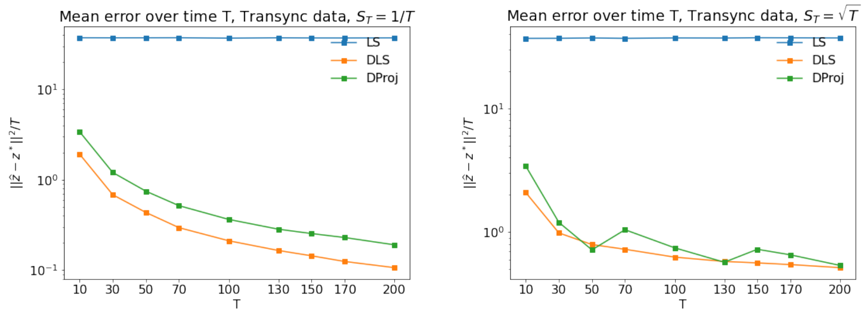
both methods, the MSE is close to its minimum for these choices of parameters.



(a) Smoothness $S_T = \frac{1}{T}$

(b) Smoothness $S_T = \frac{1}{\sqrt{T}}$

Figure 11: MSE versus T for DLS, DProj, DRC and MLE when the data is generated according to the BTL model for $n = 100$, and graphs are $G(n, p(t))$ with $p(t)$ chosen randomly between $\frac{1}{n}$ and $\frac{\log(n)}{n}$. The results are averaged over the grid $\mathcal{T}$ as well as 40 Monte Carlo runs.



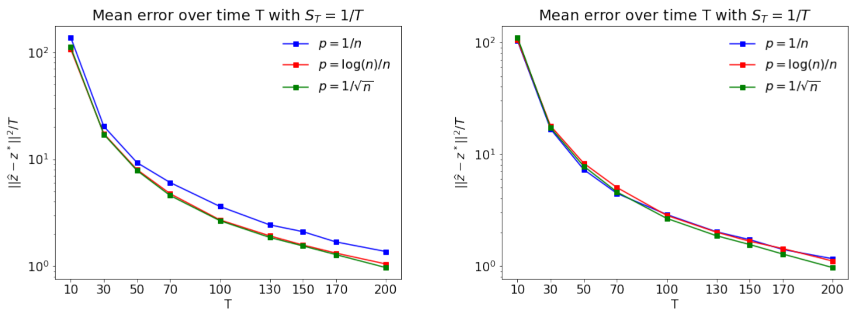
(a) Smoothness $S_T = \frac{1}{T}$

(b) Smoothness $S_T = \sqrt{T}$

Figure 12: Evolution of estimation errors with T for Least-Squares, DLS and DProj method when the synthetic data are generated according to the Dynamic TranSync model for $n = 100$ and the graphs are generated as $G(n, p(t))$ with $p(t) = \frac{\log(n)}{n}$. In particular, we ensure that the individual graphs are all connected. The results are averaged over the grid $\mathcal{T}$ as well as 20 Monte Carlo runs.

3.7.2 *Real data*

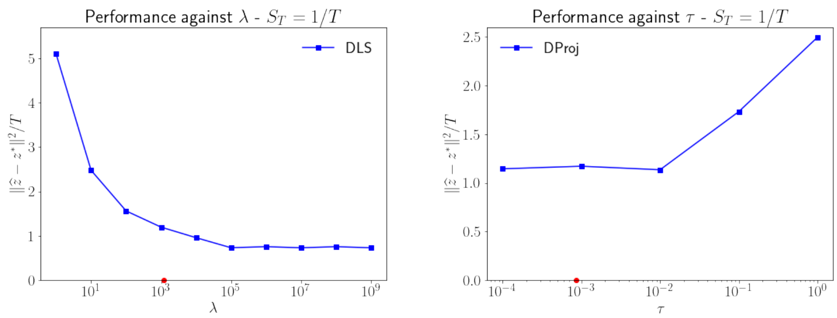
We now provide empirical results on two real data sets – the Netflix Prize data set [23], and Premier League results from season 2000/2001 to season 2017/2018 [49]. In order to assess the performance of our algorithms, we will compute the number of upsets as well as the mean squared error (MSE) for each method. The number of upsets is defined as the number of pairs



(a) DLS Method

(b) DProj Method

Figure 13: Evolution of estimation errors with T for DLS and DProj method when the synthetic data are generated according to the Dynamic TranSync modelfor $n = 100$ and the graphs are generated as $G(n, p)$ for different choices of p. In particular, we ensure that the individual graphs are all connected. The results are averaged over the grid $\mathcal{T}$ as well as 20 Monte Carlo runs.



(a) Performance of DLS

(b) Performance of DProj

Figure 14: Performance of our methods for different values of hyperparameter, with $n = 100$, $T = 200$ and $\sigma = 1$. We highlight in red on the x-axis the optimal values of parameter computed theoretically, $\lambda = \sigma^{4/3}\left(\frac{T}{S_T}\right)^{2/3}$ for the DLS and $\tau = \sigma^{-4/3}\left(\frac{S_T}{T}\right)^{2/3}$ for DProj.

for which the estimated preference is different from the observation. More precisely, denoting $\hat{z}$ to be the estimator of the strengths (for any method),

$$\text{Number of upsets} := \sum_{t \in \mathcal{T}} \sum_{\{i,j\} \in \mathcal{E}_t} \mathbb{1}_{\text{sign}(y_{ij}(t)) \neq \text{sign}(\hat{z}_{t,i} - \hat{z}_{t,j})}.$$

The MSE is defined using the vector of observations y and the estimated strength vector $\hat{z}$ as

$$\text{MSE} = \frac{1}{T+1} \sum_{t \in \mathcal{T}} \sum_{\{i,j\} \in \mathcal{E}_t} \left(y_{ij}(t) - (\hat{z}_{t,i} - \hat{z}_{t,j})\right)^2.$$

The tuning parameters λ, τ will be chosen using the cross validation procedure described below.

Cross validation procedure.

1. Fix a list of possible values of λ (resp. τ).
2. For every possibles value of λ (resp. τ), repeat several times the steps below.
   - For each time t, select randomly a measurement $y_{ij}(t)$ at time t. Let us denote (i_t, j_t) the selected pair of items at time t. We denote $Y_{\text{test}} = \{y_{i_t,j_t}(t) \,|\, t \in \mathcal{T}\}$ to be the set containing this data and $\mathcal{I}_{\text{test}} = \{(t, i_t, j_t) \,|\, t \in \mathcal{T}\}$ to be the set of corresponding indices.
   - Consider the data set where the data in Y_{test} have been removed. Compute the estimator $\hat{z}$ on this smaller data set.
   - Compute the prediction error for both performance criteria
     - MSE: $\frac{1}{T+1} \sum_{(t,i_t,j_t) \in \mathcal{I}_{\text{test}}} (y_{i_t,j_t} - (\hat{z}_{i_t}(t) - \hat{z}_{j_t}(t)))^2$.
     - Mean number of upsets :

$$\frac{1}{T+1} \sum_{(t,i_t,j_t) \in \mathcal{I}_{\text{test}}} \mathbb{1}_{\text{sign}(y_{i_t j_t}(t)) \neq \text{sign}(\hat{z}_{t,i_t} - \hat{z}_{t,j_t})}.$$

   We compute the mean of those errors for each value of the parameter λ (resp. τ).
3. Select $\lambda^\star$ (resp. $\tau^\star$) which minimizes the mean prediction error.
4. Proceed to the estimation with the chosen parameter $\lambda^\star$ (resp. $\tau^\star$).

Netflix prize data set. Netflix has provided a data set containing anonymous customer's ratings of 17770 movies between November 1999 and December 2005. These ratings are on a scale from 1 to 5 stars. From those individual rankings, we need to form pairwise information that satisfies the Dynamic Transync model (3.1.1). Denoting $s_i(t)$ to be the mean score of movie i at time t, computed as the mean rating given to this movie at time t among the customers, we then define for all pair of movies $\{i, j\}$ rated at time t

$$y_{ij}(t) = s_i(t) - s_j(t). \tag{3.7.1}$$

For computational reasons, we choose a subset of 100 movies to rank. We can then use our estimation methods to recover each movie's quality (and then their rank) at any month between November 1999 and December 2005. In order to denoise the observations, we gather the data corresponding to successive months such that all the graphs of merged observations are connected. The merged dataset is then composed of $T = 23$ observation graphs. The associated observations $y_{ij}(t)$ are computed in a similar manner as (3.7.1), using the mean score of the movies across the corresponding merged time points and the customers. For this dataset of merged movies, the cross validation step gives two different values of parameters (λ^*, τ^*) for the two different performance criteria we consider, presented in Table 6a. Using λ^*, τ^*, we then perform the estimation using the naive Least-Squares

	MSE	Upsets
λ^*	131.48	157.89
τ^*	0.42	0.37

(a) Cross-validation results

	MSE	Upsets
LS	3573	0.4954
DLS	3040	0.4984
DProj	3025	0.4976

(b) Performance of the three estimation methods

Table 6: Cross-validation and performance for the chosen parameters for the Netflix dataset. We use the MSE and the mean number of Upsets as our performance criteria.

(LS) approach as well as our two methods, namely DLS and DProj, and compute in both cases the MSE and the mean number of upsets. The results are presented in Table 6b. We note that the mean number of upsets is essentially similar for all the methods, indicating that none of the methods is better than the other for this criteria. However, DLS and DProj improve the performance of the estimation in terms of the MSE criterion.

ENGLISH PREMIER LEAGUE DATASET. This dataset is composed of all the game results from the English Premier League from season 2000-2001 to season 2017-2018. These 18 seasons involve $n = 43$ teams in total, each season seeing 20 teams confront each other across 38 rounds. The observations are the mean scores of the games between a pair of teams within the same season. Similar to the Netflix dataset, we can group the game results from successive seasons, resulting in $T = 9$ observation graphs. Let us denote $\mathcal{T}_k$ to be the set of time-points gathered to form the graph G_k. For each G_k, the corresponding observations are defined for each pair of teams $\{i, j\} \in \mathcal{E}_k$ as

$$y_{ij}(k) = \frac{1}{|\mathcal{T}_k|} \sum_{t \in \mathcal{T}_k} (s_i(t) - s_j(t)),$$

where $s_i(t)$ denotes the mean number of goals scored by team i against team j during the season t. In this case, merging the data does not lead to individual connectivity of the graphs, because of the promotion and relegation of teams at the end of each season. However, the union of all these graphs is connected. As for the Netflix dataset, we perform a cross-validation step to choose the best parameters (λ^*, τ^*), in regards to the performance criterion we consider (MSE or Mean number of upsets). The results of the cross validation are presented in Table 7a. Then, we perform our estimation for these chosen values of the parameters for DLS and DProj. Note that as the individual graphs are not connected, the naive LS method will not give interpretable results,and is only included here for comparison with our methods. As shown in Table 7b, DLS and DProj perform better than LS both in terms of MSE or in Mean number of upsets. Specifically, the number of upsets is improved by 10% with our methods.

SMOOTHNESS OF THE REAL DATA SETS. Our methods rely on the underlying smoothness of the data. However, with real data sets, the ground truth vector z^* is unknown, thus making it difficult to check whether this

	MSE	Upsets
λ^*	24.49	32.65
τ^*	6.43	5.57

(a) Cross-validation results

	MSE	Upsets
LS	0.0024	0.67
DLS	0.0015	0.57
DProj	0.0014	0.58

(b) Performance of the three estimation methods

Table 7: Cross-validation and performances for those chosen parameters for the Premier League dataset. We use the MSE and the mean number of Upsets as our performance criteria. Performance results are presented for the simple LS, DLS and DProj methods.

assumption is satisfied. In order to verify that our data sets are fit for our methods, one can define, as a proxy, an "empirical ground-truth" vector

$$z_{t,i}^{*,emp} := \frac{1}{|\mathcal{N}_{t,i}|} \sum_{j \in \mathcal{N}_{t,i}} y_{ij}(t),$$

where $\mathcal{N}_{t,i}$ denotes the set of neighbors of node i in the graph G_t. For some items that have been compared at all times, we plot the evolution of $z_{t,i}^{*,emp}$ in Figure 15. This figure shows that despite some jumps, the overall evolution of $z_{t,i}^{*,emp}$ is reasonably smooth.

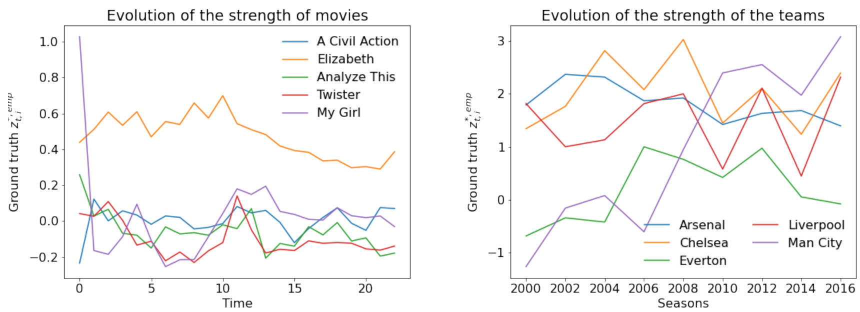


Figure 15: Evolution of the strengths for real data sets

3.8 SUMMARY OF NOTATION

Symbol	*Definition*
$\mathcal{T}$	Uniform grid of $[0,1]$, with size $T+1$.
$G_{t'} = ([n], \mathcal{E}_{t'})$, G_k	Undirected comparison graph at time $t' = \frac{k}{T} \in \mathcal{T}$
$z_t^* \in \mathbb{R}^n$	Ground-truth strengths at time $t \in [0,1]$
$x_{ij}^*(t)$	True strength difference between items i and j at time t
$y_{ij}(t)$	Observed strength difference between items i and j at time t
σ^2	Variance of the subgaussian noise
S_T	Upper bound of the smoothness assumption 2
C	Edge incidence matrix of the complete graph with n vertices
M	Incidence matrix of the path graph on $T+1$ vertices
$E = M^\top \otimes C^\top$	Smoothness operator
$\check{z}^\top = (\check{z}_0^\top, \dots, \check{z}_T^\top)^\top$	Unconstrained least squares estimator, see (3.1.4)
λ	Hyperparameter for the smoothness penalised least squares method
τ	Threshold for the Projection method
$\mathcal{V}_\tau$	Space generated by the eigenvectors of $E^\top E$ corresponding to eigenvalues smaller than τ
$P_{\mathcal{V}_\tau}$	Projection matrix on $\mathcal{V}_\tau$
Q_k	Incidence matrix of the graph G_k
$Q = \mathrm{diag}(Q_0, \dots, Q_T)$	Block diagonal matrix with blocks Q_k
$L_k = Q_k Q_k^\top$	Laplacian matrix of the graph G_k
$L = \mathrm{diag}(L_0, \dots, L_T)$	Block diagonal matrix with blocks L_k
$L(\lambda) = L + \lambda E^\top E$	Regularized Laplacian
$(\lambda_j, \nu_j)_{j=1}^n$	Eigenpairs of $CC^\top$
$(\mu_k, u_k)_{k=0}^T$	Eigenpairs of $MM^\top$
$(\alpha_{k,j}, a_{k,j})_{j=1}^n$	Eigenpairs of L_k
$L_{com} = CC^\top$	Laplacian matrix of the complete graph on n vertices
$\lambda_{min}(L)$	Smallest non-zero eigenvalue of L

Table 8: Summary of symbols used throughout the chapter along with their definitions.

4 CONCLUSION AND PERSPECTIVES

In this thesis, we have provided new analysis of dynamic ranking models. We studied the Dynamic BTL model, introduced by Bong et al. [4], under a local smoothness assumption. We have introduced a spectral algorithm that performs similarly as the MLE estimation presented in [4], but which has a far lower computational cost. We have also introduced a new dynamic model, the Dynamic TranSync model, directly derived from the Translation Synchronization by Huang et al. [21]. We have analysed this model under a global smoothness assumption and introduced two estimation methods. Both estimators are shown to be consistent and performant on both synthetic and real datasets.

As one of the few theoretical works on dynamic ranking, there is still a lot of extensions one can look at in the dynamic setting. Moreover, our models also suffer from some limitations we may want to overcome. We describe here some of those topics of interest.

1. *Independence.* One of the main assumptions of the dynamic BTL model we consider is the independence of the outcomes for all comparisons at a given time t, and the independence across different time points. The latter assumption can in particular be questioned, as the choices of one user across time are typically not going to be independent. It would be interesting to model these dependencies across time, for instance, by modelling them as a Markov process.

2. *Home effect.* Another model, used in the study of sports tournaments by Cattelan et al. [7] adapts the BTL to include a home effect, which is beneficial to the hosting team. Indeed, because of the familiar environment and the supporting public, a team is more likely to win if they play in their stadium. It would be interesting to theoretically analyze such a model.

3. *New estimators for the Dynamic BTL model.* An interesting direction would be to theoretically analyze the performance of the estimators proposed in Chapter 3 for the dynamic BTL model (recall Remark 6). The main difficulty in this regard is that the noise in the measurements is not zero-mean anymore, and it is not easy to see how such a noise term can be handled to give meaningful error bounds.

4. *Disconnected graphs.* As previously discussed in Remarks 12 and 13, we believe that the analysis of the Dynamic TranSync model can be extended to the case where some of the individual graphs are disconnected. It will be interesting to analyze this setup in detail with corresponding error bounds.

5. *Lower bound for the Dynamic TranSync.* The analysis of the Dynamic TranSync model focuses on upper bounds on the estimation error, however it will be interesting to derive lower bounds for the dynamic TranSync model which showcase the optimal dependence on n, T and S_T.

6. *Non-parametric dynamic models.* Considering a parametric model as the BTL or the TranSync model has limitations as it assumes that the preference outcomes depend only on one parameter w^*, which can be seen as a strong transitivity constraint [42]. That is why one can instead use non-parametric models as described in [38, 41, 42], where only a certain transitivity assumption is made on the comparison matrix. It has been shown by Shah et al. [42] that in the static case, the matrix of probabilities can be estimated at the same rate as in a parametric model. As such extensions can be considered in the static setting, it would be interesting to adapt these models to the dynamic case and study their performance.

APPENDIX

4.1 APPENDIX OF CHAPTER 2

4.1.1 *Technical tools*

We collect here some technical results that are used for proving the main results in Chapter 2. We begin by recalling the following result from [9].

Theorem 11 ([9, Theorem 8]). *Suppose that* $P, \hat{P}, P^*$ *are probability transition matrices with stationary distributions* $\pi, \hat{\pi}, \pi^*$ *respectively. We assume that* P^* *represents a reversible Markov chain. When* $\|P - \hat{P}\|_{\pi^\star} < 1 - \max\{\lambda_2(P^\star), -\lambda_n(P^\star)\}$ *holds, then we have*

$$\|\pi - \hat{\pi}\|_{\pi^\star} \leqslant \frac{\|\pi^\top (P - \hat{P})\|_{\pi^\star}}{1 - \max\{\lambda_2(P^\star), -\lambda_n(P^\star)\} - \|P^\star - \hat{P}\|_{\pi^\star}}. \tag{4.1.1}$$

Next, we recall some useful concentration results from probability starting with the classical Chernoff bound.

Theorem 12 (**Chernoff bound**, [33, Theorems 4.4, 4.5]). *Let* $X_1, \ldots, X_n$ *be independent Bernoulli variables with* $\mathbb{P}(X_i = 1) = p_i$. *Let* $X = \sum_{i=1}^n X_i$.

1. *For any* $\mu \geqslant \mathbb{E}[X]$, *the following is true.*
   - *For* $\delta \in (0, 1]$, $\mathbb{P}(X \geqslant (1+\delta)\mu) \leqslant e^{-\delta^2\mu/3}$.
   - *For* $\delta \geqslant 1$, $\mathbb{P}(X \geqslant (1+\delta)\mu) \leqslant e^{-\delta\mu/3}$.
2. *For any* $\mu \leqslant \mathbb{E}[X]$ *and* $\delta \in (0, 1)$, $\mathbb{P}(X \leqslant (1-\delta)\mu) \leqslant e^{-\delta^2\mu/2}$.

Theorem 13 (**Hoeffding's inequality**). *Let* $X_1, \ldots, X_n$ *be a sequence of independent random variables where* $X_i \in [a_i, b_i]$ *for each* $i \in [n]$, *and* $S_n = \sum_{i=1}^n X_i$. *Then*

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

Theorem 14 (**Matrix Bernstein**, [45]). *Let* $Z_1, \ldots, Z_n \in \mathbb{R}^{d_1 \times d_2}$ *be independent, zero-mean random matrices, each satisfying (almost surely)* $\|Z_i\|_2 \leqslant B$. *Then for any* $s \geqslant 0$,

$$\mathbb{P}\left(\left\|\sum_{i=1}^n Z_i\right\|_2 \geqslant s\right) \leqslant (d_1 + d_2) \exp\left(-\frac{3s^2}{6\nu + 2Bs}\right)$$

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

Proposition 3. *The following is true.*

1. $1 - \frac{x}{2} \geqslant e^{-x}$ *for* $x \in [0, 1.59]$
2. $1 - x \leqslant e^{-x}$ *for* $x \in [-1, 1]$.

4.1.2 *Properties of Erdös-Renyi graphs*

Recall that in our setup, we have $T+1$ Erdös-Renyi graphs $G_{t'} \sim \mathcal{G}(n, p(t'))$ for $t' \in \mathcal{T}$. Moreover, for any given $t \in [0,1]$, we have the union graph $G_\delta(t) \sim \mathcal{G}(n, p_\delta(t)))$ with $p_\delta(t)$ as in (2.2.3). Also recall that $p_{\delta,sum}(t) := \sum_{t' \in \mathcal{N}_\delta(t)} p(t')$.

Lemma 21. *Consider the events*

1. $\mathcal{A}_1 = \left\{ \frac{np_\delta(t)}{2} \leqslant d_{\min,\delta}(t) \leqslant d_{\max,\delta}(t) \leqslant \frac{3np_\delta(t)}{2} \right\}$,
2. $\mathcal{A}_2 = \left\{ \xi_\delta(t) > \frac{1}{2} \right\}$,
3. $\mathcal{A}_3 = \left\{ |\mathcal{E}_\delta(t)| \leqslant 2n^2 p_\delta(t) \right\}$,
4. $\mathcal{A}_4 = \left\{ \frac{p_{\delta,sum}(t)}{2} \leqslant N_{\min,\delta}(t) \leqslant N_{\max,\delta}(t) \leqslant 2p_{\delta,sum}(t) \right\}$.

Then the following is true.

1. *There exists a constant* $c_0 \geqslant 1$ *such that if* $p_\delta(t) \geqslant c_0 \frac{\log n}{n}$, *then* $\mathbb{P}(\mathcal{A}_i) \geqslant 1 - O(n^{-10})$ *for* $i = 1,2,3$.
2. *There exists a constant* $c_2 \geqslant 1$ *such that if* $p_{\delta,sum}(t) \geqslant c_2 \log n$, *then* $\mathbb{P}(\mathcal{A}_4) \geqslant 1 - O(n^{-10})$.

Proof.

1. This follows for $\mathcal{A}_1$ from the Chernoff bound (see Section 4.1.1) with $\varepsilon = 1/2$, and the union bound. For $\mathcal{A}_2$, the statement follows directly by applying [36, Lemma 7] to $G_\delta(t)$. For $\mathcal{A}_3$, this is again a standard use of the Chernoff bound with $\mu = \binom{n}{2} p_\delta(t)$ and $\varepsilon = 1/2$.
2. Note that for any $i \neq j \in [n]$, $|\mathcal{N}_{ij,\delta}(t)|$ is a sum of independent Bernoulli random variables and $\mathbb{E}[|\mathcal{N}_{ij,\delta}(t)|] = \sum_{t' \in \mathcal{N}_\delta(t)} p(t') = p_{\delta,sum}(t)$. Then if $p_{\delta,sum}(t) \geqslant c_2 \log n$ for a suitably large constant $c_2 > 0$, it holds with probability at least $1 - O(n^{-12})$ that

$$\frac{p_{\delta,sum}(t)}{2} \leqslant |\mathcal{N}_{ij,\delta}(t)| \leqslant 2p_{\delta,sum}(t).$$

   Now the union bound implies that

$$\mathbb{P}\left(\forall i \neq j : \frac{p_{\delta,sum}(t)}{2} \leqslant |\mathcal{N}_{ij,\delta}(t)| \leqslant 2p_{\delta,sum}(t) \right) \geqslant 1 - O(n^{-10}).$$

   Since $N_{\min,\delta}(t) \geqslant \min_{i \neq j} |\mathcal{N}_{ij,\delta}(t)|$ and $N_{\max,\delta}(t) \leqslant \max_{i \neq j} |\mathcal{N}_{ij,\delta}(t)|$ is always true, hence the statement follows.

□

Proposition 4. *With* $p_{\min}$ *as in* (2.2.5), *we have for all* $t \in [0,1]$ *that* $p_\delta(t) \gtrsim \min\{1, p_{\min}\delta\}$. *Therefore* $p_\delta(t) \gtrsim \frac{\log n}{n}$ *if* $p_{\min}\delta \gtrsim \frac{\log n}{n}$.

Proof. Starting with the definition of $p_\delta(t)$ in (2.2.3), we have using Proposition 3 that

$$p_\delta(t) = 1 - \prod_{t' \in \mathcal{N}_\delta(t)} (1 - p(t')) \geqslant 1 - (1 - p_{\min})^{|\mathcal{N}_\delta(t)|} \geqslant 1 - e^{-p_{\min}|\mathcal{N}_\delta(t)|} \geqslant 1 - e^{-p_{\min}\delta} \tag{4.1.2}$$

where the last inequality follows uses the fact $|\mathcal{N}_\delta(t)| \geqslant \delta$ for all t. Now there are two cases to distinguish.

- If $p_{min}\delta \leqslant 1$, then $1 - e^{-p_{min}\delta} \geqslant \frac{p_{min}\delta}{2}$ using Proposition 3.
- If $p_{min}\delta \geqslant 1$, then $1 - e^{-p_{min}\delta} \geqslant 1 - e^{-1}$.

This concludes the proof. □

4.2 APPENDIX OF CHAPTER 3

4.2.1 *Technical tools*

We recall below two useful and standard results regarding the Löwner ordering. The proof of the next lemma has been adapted from [14].

Lemma 22. *Let* A, B *be two* $N \times N$ *symmetric positive definite matrices, such that* $B \succcurlyeq A$. *Then, it holds* $A^{-1} \succcurlyeq B^{-1}$.

Proof. Given that $B \succcurlyeq A$, it is easy to see that $A^{-1/2}(B - A)A^{-1/2} \succcurlyeq 0$, which is equivalent to $A^{-1/2}BA^{-1/2} \succcurlyeq I$. Assuming A and B are $N \times N$ matrices, let us write the spectral expansion

$$A^{-1/2}BA^{-1/2} = \sum_{i=1}^{N} \mu_i u_i u_i^\top.$$

Noticing that $I = \sum_{i=1}^{N} u_i u_i^\top$, and given the fact $A^{-1/2}BA^{-1/2} \succcurlyeq I$ it is clear that $\mu_i > 1$ for all i, which implies that

$$I = \sum_{i=1}^{N} u_i u_i^\top \succcurlyeq \sum_{i=1}^{N} \mu_i^{-1} u_i u_i^\top = A^{1/2}B^{-1}A^{1/2}.$$

On other hand, using this Loewner inequality, we deduce that

$$B^{-1} = A^{-1/2}(A^{1/2}B^{-1}A^{1/2})A^{-1/2} \preccurlyeq A^{-1/2}IA^{-1/2} = A^{-1}$$

□

The proof of the following lemma is outlined in [50].

Lemma 23. *Let* A, B *be two* $N \times N$ *symmetric positive semidefinite matrices satisfying* $B \succcurlyeq A$ *and* $\mathcal{N}(A) = \mathcal{N}(B)$. *Then it holds* $A^\dagger \succcurlyeq B^\dagger$.

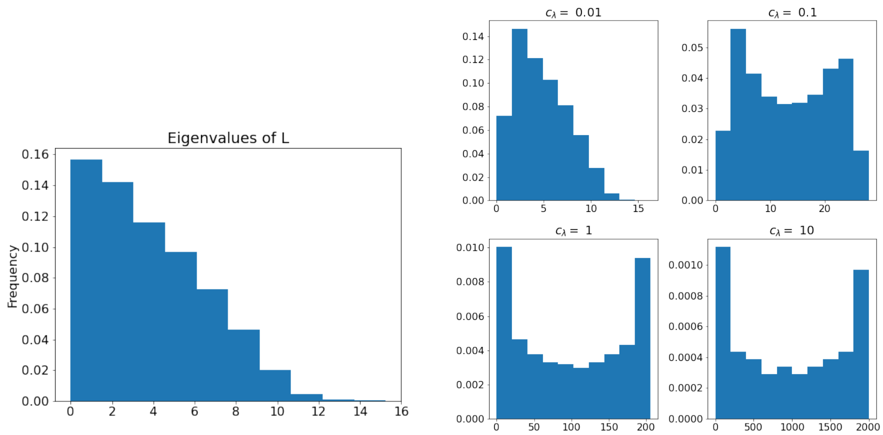
Proof. If A, B are invertible, the result follows direclty from Lemma 22. If they are not, assume that their nullspace $\mathcal{N}(A)$ has dimension $d \geqslant 1$ and their range has dimension $N - d$. Let $\{v_1, \cdots, v_d\}$ be any orthonormal basis of $\mathcal{N}(A)$. It is easy to see that $A + \sum_{i=1}^{d} v_i v_i^\top$ is invertible, and the same is true for $B + \sum_{i=1}^{d} v_i v_i^\top$. Given $B + \sum_{i=1}^{d} v_i v_i^\top \succcurlyeq A + \sum_{i=1}^{d} v_i v_i^\top$, we use Lemma 22, to deduce

$$A^\dagger + \sum_{i=1}^{d} v_i v_i^\top = (A + \sum_{i=1}^{d} v_i v_i^\top)^{-1} \succcurlyeq (B + \sum_{i=1}^{d} v_i v_i^\top)^{-1} = B^\dagger + \sum_{i=1}^{d} v_i v_i^\top,$$

where for the first and last equality we used the orthogonality of the range of A (B has the same range) with respect to the nullspace. From this, the result follows. □

4.2.2 *Additional simulations*

EIGENVALUES OF $L(\lambda)$. The theoretical analysis of the smoothness penalized estimator relies on the knowledge of the eigenvalues of $L(\lambda)$ and how they differ from the eigenvalues of L. We show in Figure 16 the variations of the eigenvalues of these matrices for different choices of λ. We verify experimentally that the addition of a penalty term increases part of the spectrum, which is at the base of our theoretical analysis.

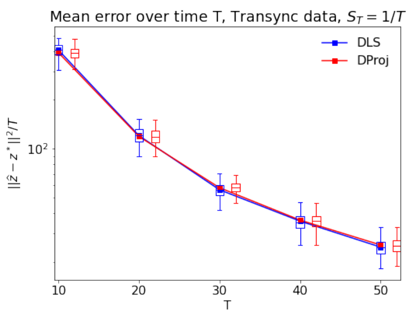


(a) Eigenvalues of L

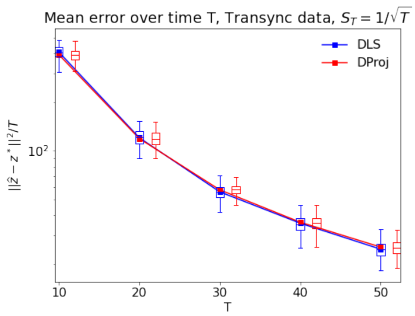
(b) Eigenvalues of $L(\lambda)$

Figure 16: Histogram of eigenvalues of L and $L(\lambda)$ for different values of $\lambda = c_\lambda \left(\frac{T}{S_T}\right)^{2/3}$, for $T = 200$ and $n = 100$. The addition of the penalty term increases the eigenvalues.

SIMULATION FOR $n \gg T$. Figure 17 shows the evolution of the MSE for $n = 200$ and T varying between 10 and 50. We observe that even in the case $n \gg T$, the estimation error goes to 0 as T increases. However, the errors are slightly bigger than in the case $n \ll T$, which is consistent with the theoretical error bounds we presented in this chapter.



(a) Smoothness $S_T = \frac{1}{T}$



(b) Smoothness $S_T = \frac{1}{\sqrt{T}}$

Figure 17: MSE versus T for DLS and DProj for $n \gg T$. Here $n = 200$, T goes from 10 to 50 and data is generated according to the Dynamic TranSync model. Graphs are generated as $G(n, p(t))$ with $p(t)$ chosen randomly between $\frac{1}{n}$ and $\frac{\log(n)}{n}$. The results are averaged over the grid $\mathcal{T}$ as well as 20 Monte Carlo runs.

[1] 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.

[2] Ernesto Araya, Eglantine Karlé, and Hemant Tyagi. “Dynamic ranking and translation synchronization.” In: *Information and Inference: A Journal of the IMA* 12.3 (2023), pp. 2224–2266.

[3] Afonso S. Bandeira, Nicolas Boumal, and Amit Singer. “Tightness of the maximum likelihood semidefinite relaxation for angular synchronization.” In: *Mathematical Programming* 163.1 (2017), pp. 145–167.

[4] 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*. 2020, pp. 3317–3326.

[5] 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.

[6] Andries E Brouwer and Willem H Haemers. *Spectra of graphs*. Springer Science & Business Media, 2011.

[7] Manuela Cattelan, Cristiano Varin, and David Firth. “Dynamic Bradley–Terry modelling of sports tournaments.” In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 62.1 (2013), pp. 135–150.

[8] Pinhan Chen, Chao Gao, and Anderson Y. Zhang. “Partial recovery for top-k ranking: Optimality of MLE and Sub-Optimality of the spectral method.” In: *The Annals of Statistics* 50.3 (2022), pp. 1618 –1652.

[9] Yuxin Chen, Jianqing Fan, Cong Ma, and Kaizheng Wang. “Spectral method and regularized MLE are both optimal for top-K ranking.” In: *The Annals of Statistics* 47.4 (2019), 2204–2235.

[10] Mihai Cucuringu. “Sync-Rank: Robust Ranking, Constrained Ranking and Rank Aggregation via Eigenvector and SDP Synchronization.” In: *IEEE Transactions on Network Science and Engineering* 3.1 (2016), pp. 20–36.

[11] Persi Diaconis. “A Generalization of Spectral Analysis with Application to Ranked Data.” In: *The Annals of Statistics* 17.3 (1989), pp. 949–979. ISSN: 00905364. URL: http://www.jstor.org/stable/2241705 (visited on 08/07/2023).

[12] Regina Dittrich, Brian Francis, and Walter Katzenbeisser. "Temporal dependence in longitudinal paired comparisons." In: (2008).

[13] Daniel M Ennis. *Thurstonian models: Categorical decision making in the presence of noise*. Institute for Perception, 2016.

[14] Epiousios(https://math.stackexchange.com/users/451957/epiousios). *Positive definiteness of difference of inverse matrices*. Mathematics Stack Exchange. URL:https://math.stackexchange.com/q/3018669 (version: 2020-11-20). eprint: https://math.stackexchange.com/q/3018669. URL: https://math.stackexchange.com/q/3018669.

[15] Michael A Fligner and Joseph S Verducci. "Multistage ranking models." In: *Journal of the American Statistical association* 83.403 (1988), pp. 892–901.

[16] Mark E Glickman and Hal S Stern. "A state-space model for National Football League scores." In: *Journal of the American Statistical Association* 93.441 (1998), pp. 25–35.

[17] Alexandra Grand, Regina Dittrich, and Brian Francis. "Markov models of dependence in longitudinal paired comparisons: an application to course design." In: *AStA Advances in Statistical Analysis* 99 (2015), pp. 237–257.

[18] Julien Hendrickx, Alex Olshevsky, and Venkatesh Saligrama. "Minimax rate for learning from pairwise comparisons in the BTL model." In: *International Conference on Machine Learning*. 2020, pp. 4193–4202.

[19] Julien Hendrickx, Alexander Olshevsky, and Venkatesh Saligrama. "Graph Resistance and Learning from Pairwise Comparisons." In: *Proceedings of the 36th International Conference on Machine Learning*. 2019, pp. 2702–2711.

[20] Daniel Hsu, Sham Kakade, and Tong Zhang. "A tail inequality for quadratic forms of subgaussian random vectors." In: *Electronic Communications in Probability* 17 (2012), pp. 1–6.

[21] Xiangru Huang, Zhenxiao Liang, Chandrajit Bajaj, and Qixing Huang. "Translation synchronization via truncated least squares." In: *Advances in neural information processing systems* 30 (2017).

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

[23] Xiaoye Jiang, Lek-Heng Lim, Yuan Yao, and Yinyu Ye. "Statistical ranking and combinatorial Hodge theory." In: *Mathematical Programming* 127.1 (2011), pp. 203–244.

[24] Eglantine Karlé and Hemant Tyagi. "Dynamic ranking with the btl model: a nearest neighbor based rank centrality method." In: *Journal of Machine Learning Research* 24.269 (2023), pp. 1–57.

[25] M. G. Kendall. “A new measure of rank correlation.” In: *Biometrika* 30.1-2 (1938), pp. 81–93.

[26] Ravi Kumar, Andrew Tomkins, Sergei Vassilvitskii, and Erik Vee. “Inverting a steady-state.” In: *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*. 2015, pp. 359–368.

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

[28] Shuang Li and Michael B Wakin. “Recovery Guarantees for Time-varying Pairwise Comparison Matrices with Non-transitivity.” In: *arXiv preprint arXiv:2106.09151* (2021).

[29] Wanshan Li, Shamindra Shrotriya, and Alessandro Rinaldo. ℓ_∞-*Bounds of the MLE in the BTL Model under General Comparison Graphs*. 2022. arXiv: 2110.10825 [math.ST].

[30] Robert D. Luce. *Individual Choice Behavior: A Theoretical Analysis*. Wiley, 1959.

[31] Lucas Maystre and Matthias Grossglauser. “Fast and accurate inference of Plackett–Luce models.” In: *Advances in neural information processing systems*. 2015, pp. 172–180.

[32] Lucas Maystre, Victor Kristof, and Matthias Grossglauser. “Pairwise comparisons with flexible time-dynamics.” In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2019, pp. 1236–1246.

[33] Michael Mitzenmacher and Eli Upfal. *Probability and computing: Randomization and probabilistic techniques in algorithms and data analysis*. Cambridge university press, 2017.

[34] Shun Motegi and Naoki Masuda. “A network-based dynamical ranking system for competitive sports.” In: *Scientific reports* 2.1 (2012), pp. 1–7.

[35] Paine N. *NFL Elo Ratings are Back!* https://fivethirtyeight.com/features/nfl-elo-ratings-are-back/.. 2015.

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

[37] Arkadi Nemirovski. “Topics in non-parametric statistics.” In: *Ecole d’Eté de Probabilités de Saint-Flour* 28 (2000), p. 85.

[38] Ashwin Pananjady, Cheng Mao, Vidya Muthukumar, Martin J. Wainwright, and Thomas A. Courtade. “Worst-case versus average-case design for estimation from partial pairwise comparisons.” In: *The Annals of Statistics* 48.2 (2020), pp. 1072 –1097.

[39] Mark Rudelson, Roman Vershynin, et al. “Hanson-wright inequality and sub-gaussian concentration.” In: *Electronic Communications in Probability* 18 (2013).

[40] Veeranjaneyulu Sadhanala, Yu-Xiang Wang, and Ryan J Tibshirani. "Total variation classes beyond 1d: Minimax rates, and the limitations of linear smoothers." In: *Advances in Neural Information Processing Systems* 29 (2016), 3521–3529.

[41] Nihar B Shah and Martin J Wainwright. "Simple, robust and optimal ranking from pairwise comparisons." In: *The Journal of Machine Learning Research* 18.1 (2017), pp. 7246–7283.

[42] Nihar Shah, Sivaraman Balakrishnan, Aditya Guntuboyina, and Martin Wainwright. "Stochastically transitive models for pairwise comparisons: Statistical and computational issues." In: *International Conference on Machine Learning*. 2016, pp. 11–20.

[43] Amit Singer. "Angular synchronization by eigenvectors and semidefinite programming." In: *Appl. Comput. Harmon. Anal.* 1.30 (2011), pp. 20–36.

[44] Louis L Thurstone. "A law of comparative judgment." In: *Psychological review* 34.4 (1927), p. 273.

[45] Joel A. Tropp. "An Introduction to Matrix Concentration Inequalities." In: *Foundations and Trends® in Machine Learning* 8.1-2 (2015), pp. 1–230.

[46] Roman Vershynin. "Introduction to the non-asymptotic analysis of random matrices." In: *Compressed Sensing: Theory and Applications*. Cambridge University Press, 2012, 210–268.

[47] Ronald Yurko, Samuel Ventura, and Maksim Horowitz. "nflWAR: a reproducible method for offensive player evaluation in football." In: *Journal of Quantitative Analysis in Sports* 15.3 (2019), pp. 163–183.

[48] Ernst Zermelo. "Die berechnung der turnier-ergebnisse als ein maximumproblem der wahrscheinlichkeitsrechnung." In: *Mathematische Zeitschrift* 29.1 (1929), pp. 436–460.

[49] Alexandre d'Aspremont, Mihai Cucuringu, and Hemant Tyagi. "Ranking and synchronization from pairwise measurements via SVD." In: *J. Mach. Learn. Res.* 22 (2021), 19:1–19:63.

[50] jlewk (https://math.stackexchange.com/users/484640/jlewk). *Is it true that* $A \geqslant B$ *implies* $B^\dagger \geqslant A^\dagger$ *for singular positive semi-definite matrices?* Mathematics Stack Exchange. URL:https://math.stackexchange.com/q/36 (version: 2020-05-20). eprint: https://math.stackexchange.com/q/3682798. URL: https://math.stackexchange.com/q/3682798.