LEARNING LATENT COMMUNITY STRUCTURES IN NETWORK-BASED DATA

Ruituo Fan

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Statistics and Operations Research.

Chapel Hill
2020

Approved by:

Shankar Bhamidi

Nicolas Fraiman

Andrew Nobel

Jan Hannig

Vladas Pipiras

ABSTRACT

RUITUO FAN: Learning Latent Community Structures in Network-based Data
(Under the direction of Shankar Bhamidi, Nicolas Fraiman and Andrew Nobel)

In this thesis we study two models that incorporate latent group structure related to networks. In particular, for the first part we introduce a new multitype recursive tree model called Community Modulated Recursive Tree (CMRT) that assigns group labels to vertices in a way similar to the popular stochastic block model for random graphs. Then we introduce a closely related population dependent branching process, and proceed to derive some of CMRT's asymptotic properties based on that, including limiting degree distribution, a tightness result for maximal degree and almost sure convergence of height. For the second part, we study a collection of random processes driven by certain latent community structure in a network and show that global optimum of K-means criterion can recover the groups exactly with high probability given enough observations across time. We shall also discuss other algorithms, and their performance is assessed in a simulation study. For the third part, we focus on a vector autoregressive model driven by stochastic block model, as a special case under the framework considered in the second part, but with change points. We show that this model can be studied under the structural break framework, given that the community structure is fixed and known (or can be recovered from algorithms). We also propose an algorithm for the general case where both communities and edge probabilities change across time, and its performance is compared with other methods in numerical experiments.

To my parents

ACKNOWLEDGEMENTS

Looking back on my past five years in Chapel Hill, I cannot but feel indebted to everyone who helped me one way or another that led me to this dissertation. It is hard to sum up in a few pages my gratitude, but I shall make my humble attempt.

First and foremost, I would like to thank my parents for their unconditional love and authoritative style. I was always, and still is in a sense, the curious and sometimes rebellious kid. Without their support, I would have become a lost wanderer in the forest of knowledge.

Next I wish to express my deepest gratitude to my advisors Shankar, Nico and Andrew, not only for their professional guidance throughout my doctoral study, but also for other invaluable lessons. I am grateful to Shankar for being a talented yet modest mentor, always willing to provide me with advice, from his own experience and also other giants of the field. I should mention his delightful laughter as well. I am grateful to Nico for introducing me to my first research topic, and being a patient listener always offering to help. I am grateful to Andrew for our many enjoyable conversations filled with his wisdom of life.

I also want to thank other members from my committee, Dr. Jan Hannig and Dr. Vladas Pipiras, for sparing time out of their busy schedule to read my thesis in this turbulent time. Special thanks are due to Dr. Pipiras, for his tremendous help on the latter two parts of dissertation, as an expert in time series related field. I am thankful to other faculty member in the department as well, especially Dr. Chuanshu Ji, for his support during my first year here.

Besides that, I would like to pay regards to all my teachers. Special appreciation goes to Jingfen Su, my math instructor in middle school, and Dr. Dayue Chen from Peking University, my instructor for the stochastic process course in college. Although I am afraid I ended up not taking the latter's advice, who is renowned for his opinion against studying abroad (he received his Ph.D. degree from UCLA under the direction of Dr. Thomas M. Liggett).

Last but not least, thanks are due to other individuals who contributed to my achievements:

- To my friends in the department. Special thanks alphabetically go to Haipeng Gao, Meilei Jiang, Weiwei Li, Hongsheng Liu, Yiyun Luo, Zhengling Qi, Huijun Qian and Ruoyu Wu, for making my life livable in this countryside.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND SYMBOLS

URT   Uniform Recursive Tree

CMRT   Community Modulated Recursive Trees

w.h.p.   with high probability

$N_k(n)$ number of vertices of out-degree k in a recursive tree of size n

M_n maximal degree in a recursive tree of size n

H_n height of a recursive tree of size n

$S_k(n)$ size of the subtree rooted at k in a recursive tree of size n

$\mathcal{U}_n$ URT of size n

$\mathcal{T}_n$ CMRT of size n

CHAPTER 1

Introduction

Graphs representing real systems are not regular like. They are objects where order coexists with disorder. —— *Santo Fortunato [41]*

Given a population, oftentimes we can divide it into groups such that objects within the same group are more similar to each other than outsiders. In different contexts, these groups may be referred to as classes, types, clusters, blocks, modules, communities, etc. And in some cases groups are known explicitly, and the focus is on predicting group assignment based on a set of covariates. This problem is known as classification under the framework of supervised learning. In other cases however, the groups are not given and have to be learned. And a central issue in this unsupervised learning problem is that of clustering, which aims to partition the population into groups consisting of similar objects. To these ends, various algorithms have been developed: the distribution-based Gaussian mixture model [103], the centroid-based K-means algorithm [65], the dissimilarity-based hierarchical clustering [55], just to name a few. Naturally, there is no universal clustering scheme that works for all problems, and successful algorithms have to adapt to the population of interest. Throughout this thesis, we shall focus on problems motivated by a special class of population, namely that of networks and graphs.

Over the past decades, the world has seen an emergence of networks in various fields such as computer science, biology, sociology and physics. Driven by the need to understand these networks, many classic statistical problems have been reformulated in the context of networks and became active research fields on their own. As a canonical example, the problem of clustering for networks now takes on the name "community detection", borrowing the terminology from social science. These "communities" can be found in many networks representing real systems. In sociology from which the term is coined, people are organized by hierarchies of communities, from families to nations, and these communities have been studied for decades [26][42]. In computer science, websites form communities based on relevance of topics and these communities may be detected by hyperlinks for example [40]. The advent of internet era has also given rise to online communities such as those in social networking service (SNS). In biology, communities

may correspond to compartments in food webs [58], or modules in protein-protein interaction networks [100]. In all of the examples given above, there typically exist more "links" within each community, and fewer between different communities. However, the communities per se are not defined through these links, but rather that we hope to detect them using information encoded in the links.

Community detection is important for multiple reasons [41], and we will just list a few here. First, networks have grown so large, to the size of millions or even billions of nodes, that we need better tools just to represent them, and community structure can provide a parsimonious way to achieve that purpose. This may, for example, lead to better visualization of large networks. Moreover, many real world network systems display a hierarchical structure characterized by nested communities, *e.g.* the pyramidal organization in business. In these cases, communities are essential building blocks of the whole system. Second, communities can reveal structural properties of the network. Nodes sharing a large number of edges with other community members may play a central role in maintaining the community structure, such as hubs in traffic networks and opinion leaders in the network of media users. Nodes lying on the boundaries between communities, on the other hand, may serve crucial mediation or communication role. Finally, communities also has important practical implications, *e.g.* in building efficient recommendation system for online retailers, or designing better data structures for navigation. After all, these applications are intrinsically related to communities being a useful tool to better understand networks.

The study of community structure can be traced back to Stuart Rice [91], who in 1927 did a research on the identification of blocs in small political parties based on voting behavior. Now, almost a century later, sociologists, computer scientists, biologists, physicists, statisticians, etc. all enter the game and much work has been done to develop various community detection algorithms. While clustering algorithms for non-network data like hierarchical clustering can be carried over to community detection, these generic methods are known to work not so well for real-world network data [81] and this has led to active development of clustering algorithms tailored for networks. In modern days, any practitioner who wishes to conduct some kind of community detection has to choose from a mass of algorithms. To list just a few popular ones, there are the divisive algorithm based on edge betweenness proposed by Girvan and Newman [46][82], a large variety of modularity-based methods starting from the seminal work of Girvan and Newman [82], and spectral clustering [92].

While there is some truth to the saying "clustering is in the eye of the beholder"[38], it is still important to have models under which algorithms can be assessed through their performance. As an example, popular clustering algorithms such as expectation-maximization (EM) and K-means are known to perform well under Gaussian mixture model [66][103]. In the context of community detection, stochastic block model arguably plays a similar role.

In random graph theory, stochastic block model (SBM) generates a graph as follows. First, vertices are partitioned into two or more groups (usually referred to as blocks). And in many cases the group labels are chosen independently for each vertex according to a vector specifying the probabilities of belonging to each group. After the labels are determined, each pair of vertices are connected independently at random with probability that depends only on their group labels.

To assess the performance of community detection algorithms, most results focus on either exact recovery (also called perfect recovery), which seeks to recover the true partition with high probability, or partial recovery, which only requires a constant proportion of objects to be clustered correctly. For stochastic block model with n nodes, both recovery problems depend on how edge probabilities scale as n increases. The cases where edge probabilities are of order $\Theta(1)$, $\Theta(\log n/n)$ and $\Theta(1/n)$ are called dense, semi-sparse and sparse, respectively. As graph gets sparser, it becomes harder to recover the communities, and two types of barriers show up: statistical and computational. Statistically, one is interested in when recovery becomes impossible. For partial recovery, the barrier is shown to lie in the sparse domain under the special case of two equal-sized communities [71][76], and for exact recovery of (any number of) equal-sized communities it is in the semi-sparse domain [2]. For both problems the thresholds are known explicitly, under which recovery becomes impossible. On the computational side of things, one is interested in the existence of feasible algorithms (*e.g.* polynomial-time) that can get the job done. Typical algorithms studied in the literature include spectral methods [59][71][19][61], belief propagation [29][77] and semi-definite programming (SDP) [1][48][49][4].

In the two models we consider however, we shall deal with two classes of population that are related to yet different from the random graphs stochastic block model is used to model. The first population is random recursive trees. Contrary to usual graphs, here we have built-in hierarchical structure and the problem is connected to that of information flow on trees [78]. The second model is for collections of Gaussian processes with network-based correlations. In this case we do not have access to the network itself and aim to extract information about latent

3

communities from noisy signals observed across time. And as we shall see, the problem here can be reformulated to relate to Gaussian mixture model.

1.1 Community Modulated Recursive Tree

A recursive tree of size n is a rooted tree labeled $\{1, 2, ..., n\}$ such that paths from the root to any vertex have increasing labels. The word "recursive" comes from the fact that these trees can be generated recursively and are used to model a growing population. Therefore, instead of thinking about a single tree, it is better to have in mind a collection of growing trees, which are sometimes referred to as a recursive tree without specifying the size. In these cases a recursive tree is viewed as a stochastic process of nested labeled trees. As an example, consider a private club where members meet weekly. At each meeting, one member is asked to invite a new member next week. Then as long as everyone accepts the invitation and no one quits, the club will keep growing and we can model the invitation process by a recursive tree. More precisely, we treat members as vertices, and label them by the order in which they join the club. Each time a newcomer arrives, we connect that person to the member who invited her. It is clear that the random graphs generated this way form a recursive tree. Moreover, if the member who sends out invitation each week is selected uniformly at random, we get the so-called uniform recursive tree (URT).

From a statistical point of view, URT serves as a natural null model for recursive trees. Given that, the next thing one may look for is possible alternatives. For our work, we introduce an alternative to URT that allows for group structure in the spirit of the celebrated stochastic block model. Just as stochastic block model can be viewed as an alternative to the Erdős-Rényi model, where each pair of vertices are connected independently with the same probability, our model, which we shall refer to as Community Modulated Recursive Tree (CMRT), generalizes URT in a similar way to allow for latent group structure. Specifically, whenever a new vertex gets added in CMRT, it is assigned type i with probability p_i. Then it chooses type j with probability q_{ij} and connects to one vertex of that type uniformly at random. When there is only one type, CMRT reduces to the usual URT.

Since URT has been studied extensively in the literature and many of its properties are well-known, it makes sense to compare CMRT with URT in hope of understanding their difference. Therefore we look at three important properties of random trees, namely limiting degree distribution, maximal degree and height. For CMRT of two types, we derive the limiting degree

distribution explicitly and note that it is different from that of URT except for certain special cases. For maximal degree and height however, we show that they both scale like $\log n$ asymptotically in non-trivial cases, just as in URT. In fact, we are able to establish almost convergence of height to the same limit as that of URT, except for some special cases. We shall also mention the study of subtree sizes and the related root-finding algorithm.

1.2 Recovering clusters from covariance structure

Under the Gaussian setting, Gaussian mixture model is one of most prevalent models around. The model consists of n i.i.d. observations, each distributed as a mixture of K Gaussian distributions called components. There are also n latent variables specifying the component for which each observation belongs to. Typically these latent variables are sampled from a multinomial distribution.

While there are many things one can do with Gaussian mixture model, we are particularly interested in using it to assess performance of clustering algorithms through error rate. As could be expected, separation between Gaussian components in the mixture is an essential factor here. Intuitively, the larger the gaps are between means of components compared to variances, the easier it is to recover the true partition.

From the point of modeling group structure, we see that groups in a mixture model are characterized by difference in parameters of their distributions, while the observations themselves are independent. In our work we take a different perspective and consider a collection of correlated Gaussian vectors with the same mean. We index their coordinates by time, making them Gaussian processes. In fact, our model is motivated by a vector autoregressive model where the coefficient matrix is specified by an adjacency matrix with certain community structure. The correlation we consider here are two-fold. First, each process is correlated across time, so the vector we observe will have dependent coordinates. Second, the processes themselves are correlated based on some group structure. Roughly speaking, we assume that processes within the same group are more correlated than those from different groups. Then under certain conditions, we are able to show that global optimum of K-means criterion can recover the true group partition exactly given enough observations across time.

We shall also consider algorithms applied to the sample covariance matrix, including K-means, spectral clustering and an iterative algorithm, similar in spirit to Lloyd's algorithm for

K-means, that fits a blockwise constant approximation to a given matrix. Performance of these algorithms are then assessed in a simulation study.

1.3 SBM-driven VAR model with change points

As a special case and motivating example for what we consider in the previous part, we focus on a vector autoregressive model where the coefficient matrix is a scaled adjacency matrix generated from certain stochastic block model. In addition, we introduce change points into the model, and allow both communities and edge probabilities to change across time. Our goal is to detect the true change points, given that their number is known explicitly.

Guðmundsson [47] studied a similar and more general model (with normalization instead of scaling) without change points, and showed that a spectral clustering based method can partially recover the communities. While one may also choose to use other algorithms, *e.g.* K-means based methods discussed in previous part, existing theoretical guarantees typically break down. On the other hand, in cases where the community structure is fixed and known (or can be recovered from algorithms), we show that this SBM-driven VAR model can be studied under the structural break framework.

We will also propose a change point detection algorithm for the general case where both communities and edge probabilities vary across time, and compare its performance with other methods using simulated data.

CHAPTER 2

Community Modulated Recursive Tree

Trees are everywhere. From real botanical trees seen in daily life, to abstract data structures used in computer science, these hierarchical forms first created by nature are also adored by us humans. Quite naturally, trees are often grown recursively. In probability models for rooted trees labeled $\{1, 2, ..., n\}$, this means that paths from root to any vertex have increasing labels. These recursive trees have been studied for decades, with applications to epidemics [74], pyramid schemes [43][44], convex hull algorithms [70] and modeling family trees of preserved copies of ancient or medieval texts [80], where vertices represents people or texts that arrive chronologically and are labeled by time.

In many applications, uniform recursive tree (URT) is the model actually used. As any recursive tree, it can be defined recursively. First, URT of size 1 is just a single vertex 1. Assuming that URT of size n is already defined, then URT of size $n + 1$ is a tree formed by attaching vertex $n + 1$ uniformly at random to a vertex in a URT of size n. It is not hard to verify that this actually produces a recursive tree. Alternatively, a URT of size n can be formed by choosing uniformly at random a tree from all recursive trees of size n.

Of course there are other alternatives for modeling recursive trees. For example, preferential attachment [13], because of its power law degree distribution, is favored in the complex network community where real data exhibits this scale-free property. There are also variants of URT that introduce choices to the attachment rule. Instead of choosing one existing vertex, these models choose k previous vertices (with or without replacement) as candidates, and connect the new vertex to one of them based on certain optimization criterion. D'Souza et al. [37] investigated criteria such as selecting the candidate with maximal or minimal depth (*i.e.* distance from the root) or maximal degree, and Mahmoud [68] discussed models where the candidate with largest or smallest label is chosen. These models typically lead to improvement in statistics used as the optimization criterion (*e.g.* asymptotic maximal degree becomes larger if candidate with largest degree is chosen). Related to these models there is the so called scaled attachment random recursive tree (SARRT) [34], where at the n-th step the new vertex n is connected

to vertex $\lfloor nX_n \rfloor$, with $X_1, X_2, \ldots$ being a sequence of i.i.d. random variables taking value in $[0, 1)$. When X_n follows uniform distributions, SARRT reduces to the usual URT. And when the random variable tends to take smaller values, new vertices are more likely to connect to existing vertices with small labels.

However, all of these alternatives are still homogeneous in the sense that attachments are made according to the same rule for each vertex, no matter how complicated. In hope of modeling heterogeneity, we introduce an alternative to URT that incorporates group structure so that vertices in different groups behave differently. Just as stochastic block model can be seen as a heterogeneous alternative for the Erdős-Rényi model, our model, which we shall refer to as Community Modulated Recursive Tree (CMRT), generalizes URT in a similar way to allow for latent group labels.

The rest of this chapter is organized as follows: Section 2.1 reviews related work on recursive trees and highlights some of the main findings. Section 2.2 gives a brief introduction to branching processes and their applications to the study of random trees. Section 2.3 introduces CMRT and derives results on its asymptotic properties, including limiting degree distribution, maximal degree and height. Section 2.4 studies structure of subtrees in CMRT, with potential application to root-finding algorithms. Finally, Section 2.5 extends the limiting degree distribution results in Section 2.3 to two variants of the model.

2.1 Literature review

In this section we shall review known results on various global and local properties of recursive trees. The summary here is by no means comprehensive and for a more detailed survey on classic results interested readers are referred to Smythe and Mahmoud [98]. Unless otherwise specified, results stated here are for URT.

2.1.1 Terminology

We gather here terminology on rooted trees we shall use extensively later on.

First, a rooted tree is just a tree with one vertex identified as the root. And since trees are connected (undirected) graphs with no cycles, there exists a unique path from any given vertex v to the root. Any vertex $u \neq v$ in this path is called an antecedent of v, and v its descendant. In addition, if u and v are adjacent, we call u the parent of v and v its child. The length of this path (*i.e.* the number of vertices in the path) is called the depth of v. And the maximal depth

among all vertices in a rooted tree is called its height. Note that depth essentially measures the graph distance between v and the root, and height is the number of levels if one visualizes the whole tree in a hierarchical way. Finally, the subgraph induced by v and all of its descendants is called the subtree rooted at v.

For degree, people typically consider out-degrees, *i.e.* the number of children associated with each vertex. This is consistent with the definition of out-degree in directed graph, if one treats each edge in a rooted tree as directed from the parent to its child. In fact, there are authors who view recursive trees as directed graphs, but we shall stay with the undirected definition. It is easy to see that out-degree of the root is the same as its degree, and for other vertices out-degree is equal to degree minus 1. In addition, maximal degree is defined as the largest out-degree among all vertices.

2.1.2 Degree distribution

For graphs, degree is one of the most important quantities to look at and trees are no exception. Throughout we shall use $N_k(n)$ $(k \geq 0)$ to denote the number of vertices with out-degree k in a recursive tree of size n. Na and Rapoport [79] first derived recursive formula for the expected values $\mathbb{E}\, N_k(n)$ and from that showed $\lim_{n \to \infty} \mathbb{E}\, N_k(n)/n = 1/2^{k+1}$. Moon [74] later proved this convergence in probability as well.

More recently, Janson [54] established a central limit result:

$$n^{-\frac{1}{2}}(N_k(n) - 2^{-k-1}n) \xrightarrow{d} V_k$$

jointly for all $k \geq 0$, where any finite subset of $\{V_k\}_{k \geq 0}$ is jointly Gaussian and covariances can be computed from a generating function. This rather strong result is proved using generalized Pólya urns, where vertices of different degrees are treated as balls of different colors. And the main tool applied there is a functional limit theorem derived by the author earlier [53].

2.1.3 Internodal distance

For trees, distance is measured by the usual graph distance. Given any two vertices u and v, the distance between them is length of the unique path from u to v. These internodal distances are some of the first properties studied in recursive trees. Because of its prevalence, some of the key results are summarized here though we will not study internodal distance in CMRT.

Let d_{ij} denote the distance between vertex i and j, and $D_n := d_{1n}$ be the depth of vertex n. Moon [74] was able to find recursive formulas for expectation and variance of d_{ij} and solved them exactly, yielding the following result:

$$\mathbb{E}\, d_{ij} = \sum_{k=1}^{i} \frac{1}{k} + \sum_{l=1}^{j-1} \frac{1}{l} + \frac{1}{i} - 2$$

$$\mathrm{Var}(d_{ij}) = (1 - \frac{4}{i}) \sum_{k=1}^{i} \frac{1}{k} + \sum_{l=1}^{j-1} \frac{1}{l} - 3 \sum_{k=1}^{i} \frac{1}{k^2} - \sum_{l=1}^{j-1} \frac{1}{l^2} + 4 + \frac{3}{i} - \frac{1}{i^2}.$$

As a corollary, we have $\mathbb{E}\, D_n \sim \log n$ and $\mathrm{Var}(D_n) \sim \log n$.

Devroye [32] proved the asymptotic normality of D_n using theory of records. Mahmoud [67] later gave a more elementary proof by computing moment generating function of $D_n^* = (D_n - \log n)/\sqrt{\log n}$ from the exact distribution of D_n and showed that it converges to that of a standard normal variable.

2.1.4 Extremal statistics

Let M_n and H_n denote respectively the maximal degree and height in a recursive tree with size n. Devroye and Lu [35] showed that maximal degree, when scaled properly, converges both almost surely and in $\mathcal{L}_1$:

$$\frac{M_n}{\log_2 n} \xrightarrow{\text{a.s.}} 1 \quad \text{and} \quad \lim_{n\to\infty} \frac{\mathbb{E}\, M_n}{\log_2 n} = 1.$$

The proof involves writing the degree of each vertex as a sum of indicator functions, the derivation of tail bounds from that and some sort of union bound device.

As for the height, Pittel [88] proved that

$$\frac{H_n}{\log n} \xrightarrow{\text{a.s.}} e$$

using results on first birth problems of Crump-Mode-Jagers branching processes [57]. The proof involves embedding the recursive tree into a continuous time process, which is also crucial in our work. In fact, our proof for convergence of height essentially follows the same procedure. We shall give a more detailed discussion about this technique later in Section 2.2.

10

2.1.5 Structure of subtrees

Asymptotic properties of subtrees in URT can be studied through two observations [69]. First, distribution of the subtree rooted at vertex k, conditioning on its size, has the same distribution as a URT of that size. Second, if we let $S_k(n)$ be size of the subtree rooted at k in a recursive tree of size n, then

$$\frac{S_k(n)}{n} \xrightarrow{\text{a.s.}} \text{Beta}(1, k-1).$$

To see that, note the problem can be formulated into a Pólya urn. Specifically, treat vertex k as a red ball and $1, 2, ..., k-1$ as blue balls. Following the rule of Pólya urns, at each time step a ball is chosen uniformly at random and put back, together with another new ball of the same color. It can be shown that $S_k(n)$ has the same distribution as the number of red balls when there are a total of n balls. The rest follows immediately from classic results on asymptotics of Pólya urns.

In light of these two observations, asymptotic distribution of a statistic of the subtree rooted at some vertex k is a mixture of asymptotic distributions corresponding to URTs of various sizes with Beta$(1, k-1)$ as the mixing density. For concrete examples, see Smythe and Mahmoud [98].

2.1.6 Preferential attachment tree

The preferential attachment tree and its variants have also been studied extensively in the literature. In linear preferential attachment trees, instead of uniform attachment, a new vertex connects to an existing vertex with probability proportional to α plus that vertex's out-degree, where $\alpha > 0$ is a parameter. The linear preferential attachment tree is known to have a limiting power law distribution with exponent depending on the model parameter α [18]. And maximal degree, when scaled by $n^{\frac{1}{1+\alpha}}$, converges a.s. to a positive random variable as n goes to infinity [75]. As for height, in the same paper as that of URT Pittel [88] proved that $H_n / \log n$ converges a.s. to a positive constant dependent on α.

In terms of models with multiple types, which shall be our focus, an extension of preferential attachment based on genealogy of a multitype branching process has been considered in [93]. And more related to the Community Modulated Preferential Attachment (CMPA) tree that we

shall study, Deijfen and Fitzner [30] computed the limiting degree distribution for a special case of CMPA with two types heuristically and conducted a simulation study.

2.2 Branching processes applied to trees

> *It is becoming folklore, however, that global properties such as the maximum of some quantity in the tree, are not easily amenable to combinatorial analysis, but follow from probabilistic reasoning.* —— *Smythe and Mahmoud [98]*

Since a key step in our proof is to embed the recursive tree into a continuous-time process related to branching processes, we will include a short introduction here, with a focus on their applications to the study of random trees.

In discrete time, the customary formulation is that of Galton-Watson process, often used to model reproduction of a population. Mathematically, it consists of a sequence of random variables $Z_0, Z_1, Z_2, ...$ representing population sizes of the zeroth, first, second, ... generation. By default one assumes that $Z_0 = 1$. Then the process is defined recursively via:

$$Z_{n+1} = \sum_{i=1}^{Z_n} \xi_{n,i}.$$

Here $\xi_{n,i}$ denotes the number of children born to the i-th individual in the n-th generation, and are i.i.d. over all $n \in \mathbb{N}$ and $i \in \{1, 2, ..., Z_n\}$. The equation essentially states that individuals reproduce independently following the same probability distribution, and the $(n+1)$-th generation consists of all children born to the n-th generation.

The most basic result on these processes is about extinction probability, formally defined as $\mathbb{P}(\exists n : Z_n = 0)$. Note that if $Z_N = 0$ for some N, then $Z_n = 0$ holds for any $n > N$ as well. Thus we have $\mathbb{P}(\exists n : Z_n = 0) = \mathbb{P}(\lim_{n \to \infty} Z_n = 0)$ and the latter can be used as an alternative definition for extinction probability. Let $p_k = \mathbb{P}(Z_n = k)$ for $k = 0, 1, 2, ...$ be the probability that an individual in the population gives birth to exactly k children, and $\mu = \sum_{k=1}^{\infty} k p_k$ be the mean of this offspring distribution. Then the theorem states:

Theorem 2.1. *When $\mu \leq 1$, we have $\mathbb{P}(\exists n : Z_n = 0) = 1$ except for the degenerate case where $p_1 = 1$. Otherwise if $\mu > 1$, then $\mathbb{P}(\exists n : Z_n = 0) < 1$. Moreover, in the latter case we have $\mathbb{P}(\exists n : Z_n = 0) + \mathbb{P}(\lim_{n \to \infty} Z_n = \infty) = 1$.*

In other words, a Galton-Watson process becomes extinct when each individual gives birth to at most one child, except for the case of course, when all individuals have exactly one child. What might seem more interesting is the fact that the process either goes extinct or explode in terms of population size. And accordingly, a process is called subcritical, critical and supercritical respectively when $\mu < 1$, $\mu = 1$ and $\mu > 1$.

In connection to random trees, Galton-Watson processes are largely used to study global properties such as height [31]. Interested readers are directed to Devroye [33] for a nice summary of these results. Now, we shall not dwell too much in discrete time and will proceed to consider branching processes in continuous time, which are more relevant to us.

As a starting point, consider a Crump-Mode-Jagers (CMJ) branching process [27]. It starts with a single ancestor at time $t = 0$ and the number of its children by time t follows some counting process $Z(t)$. All individuals, from the time of their births, reproduce independently of each other according to random processes with the same (joint) distribution as $Z(\cdot)$. In general $Z(t)$ can be arbitrary, and when between-birth intervals are exponentially distributed with parameters $\lambda_1, \lambda_2, ...$ we call it a Poisson CMJ branching process.

To produce a rooted random tree from a CMJ branching process, one simply treat the ancestor as the root and connect each individual to its parent. Moreover, the random tree generated is recursive as long as births do not occur at the same time, which is satisfied by the Poisson CMJ process. We give two special cases here. When $\lambda_i \equiv 1$, each individual has the same probability giving birth to the next child in the population, or equivalently, we choose a parent uniformly at random when growing a tree. This leads to a uniform recursive tree. When $\lambda_i = i + 1$, parents are picked with probabilities proportional to the number of children (*i.e.* out-degree in the rooted tree) plus one. This leads to the so-called plane-oriented tree, sometimes discussed under the broader context of preferential attachment.

Now, given the embedding of a recursive tree into some CMJ branching process, in theory we can transform any problem about the original tree into its counterpart with respect to the branching process: out-degrees become the number of children and height becomes the number of generations. Similar to Galton-Watson process, CMJ branching process has been used to derive global properties of recursive trees, notably their height [88].

This continuous-time embedding technique is not limited to the study of global properties and also sees success in local properties such as degree distribution. Historically, most results on degree distribution were derived via combinatorics. However, the instant one asks for more than

the distribution itself and wants, for example, to establish some kind of central limit theorem, combinatorics quickly becomes insufficient. In these cases the problems of interest are often formed as an urn scheme. The embedding hereto can be traced back to Athreya and Karlin [7], who first introduced the embedding of urns into continuous-time Markov branching processes. For URT, Janson [54] was able to prove central limit results for its joint degree distribution using a functional limit theorem of generalized Pólya urns that builds on continuous time embedding. Moreover, as we shall see in our application, continuous-time embedding is particularly useful when the attachment rule is too complicated for combinatorics. For an example other than our own, see Bhamidi et al. [15] where a change point is present in preferential attachment model.

2.3 Community Modulated Recursive Tree and its asymptotic properties

We shall first focus on a special case of CMRT with two types (say A and B) for simplicity. Extension to the general case is discussed in Section 2.3.10.

2.3.1 Model formulation

Recall that a uniform recursive tree (URT) can be constructed as follows: starting with a root vertex, at each step choose uniformly at random an existing vertex, and add a new vertex connected to the chosen vertex. This procedure yields a growing tree-valued process, which we shall denote by $\{\mathcal{U}_n\}_{n\geq 1}$, where $\mathcal{U}_n$ is the random recursive tree given by the process as it reaches size n. Then $\{\mathcal{U}_n\}_{n\geq 1}$ is a URT and $\mathcal{U}_n$ a URT of size n.

Now we are ready to introduce Community Modulated Recursive Tree (CMRT) with two types:

- Start with a type A vertex 1 and a type B vertex 2. We shall refer to them as roots of each type. A CMRT of size 2 is the tree containing vertex 1 and 2, with an edge between them.

- Given a CMRT of size $n-1$ ($n \geq 3$), a new vertex n is added to the tree at the next time step and assigned type A with probability $p \in (0,1]$ and type B with probability $1-p$. Note here that we do not lose any generality by excluding the case where $p = 0$ because one can just switch the types.

- Vertex n then chooses its own type with probability $q \in [0,1]$ and the other type with probability $1-q$. Similar to stochastic block model, we call it assortative when $q \geq 1/2$

14
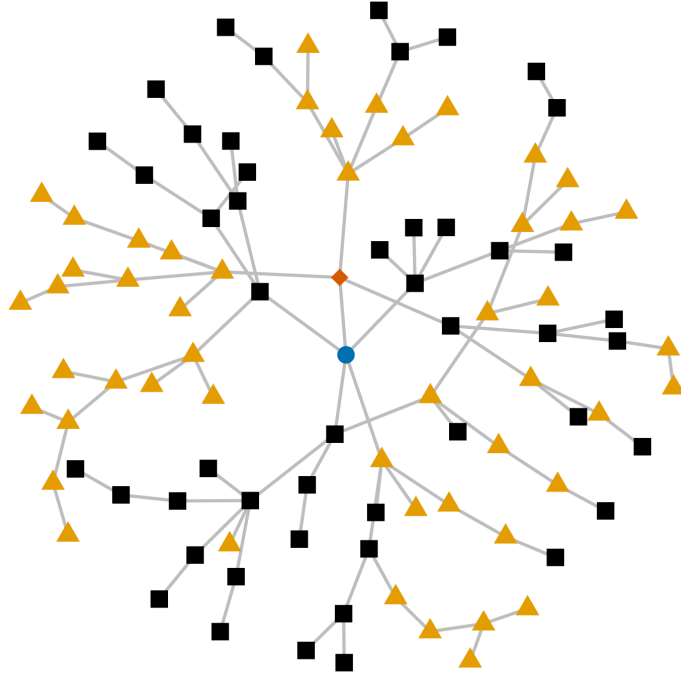
Figure 2.1: A (two-type) Community Modulated Recursive Tree of size 100, with parameters $p = 0.5$ and $q = 0.8$. The type A and B roots are plotted as blue circle and orange diamond respectively, and other type A and B vertices as black squares and yellow triangles.

and disassortative when $q < 1/2$. In general the probability for a vertex to choose its own type may vary across types. Here we assume that they are the same for simplicity and the general case is discussed in Section 2.3.10.

- Finally vertex n chooses uniformly at random an existing vertex of that chosen type, and connect to it, forming a CMRT of size n.

The process then uses the above dynamics recursively to yield a growing tree-valued process, which we shall denote by $\{\mathcal{T}_n\}_{n \geq 2}$, where $\mathcal{T}_n$ is the random recursive tree given by the process as it reaches size n. Then we call $\{\mathcal{T}_n\}_{n \geq 2}$ a Community Modulated Recursive Tree (CMRT) and $\mathcal{T}_n$ a CMRT of size n.

Since both vertices 1 and 2 are called roots, notions of out-degree and depth change slightly for CMRT. Out-degree is still defined as the number of children, and in CMRT that is equal to degree minus 1 for all vertices including the roots. Depth of a vertex in this case is the distance between that vertex and the nearest root. Although one can define things differently, asymptotically we will get the same results.

2.3.2 Special cases

1. When $p = 1$ and $q = 1$, CMRT becomes a URT consisting of type A vertices plus an additional type B vertex (*i.e.* the type B root) connecting to the type A root. Thus one may consider this case to be degenerate.

2. When $p = 1$ and $q \neq 1$, all new vertices are of type A and have a fix probability $1 - q$ to connect to the unique type B vertex. This case is degenerate too in the sense that one can construct such a CMRT from a URT. To do so, take a URT consisting of type A vertices, and add a single type B vertex connected to the root of that URT. Treat this type B vertex as the type B root. Then remove each edge between type A vertices independently with probability $1 - q$. For each edge removed this way, add an edge between the corresponding child and the type B root. The recursive tree formed this way has the same distribution as a CMRT with $p = 1$ and $q \neq 1$.

3. When $q = 1$, CMRT can be partitioned into two disjoint subtrees, one for each type, if one removes the edge between the two roots. Furthermore, each of these subtrees, conditioning on its size, has the same distribution as a URT of that size. However, existence of this edge makes it non-trivial to separate the two subtrees even in this simple case. Although not fully comparable, we note that in stochastic block model the recovery problem (*i.e.* recovering the block assignment for each vertex) becomes trivial if each block is connected and the probability to connect to vertices in other blocks is zero.

4. When $q = 0$, each new vertex will connect to a vertex of the other type. Therefore we would observe an alternating behavior if we know the types: for any path, the types of vertices along this path will be alternating.

2.3.3 Continuous time embedding

In this section we will introduce a continuous time embedding which plays a crucial role in later proofs. To fix ideas, we shall refer to "vertices" in our continuous time process as "individuals" to differentiate them from vertices in the corresponding discrete time recursive trees. To state the embedding result, we need to introduce a (two-type) continuous time process which we shall refer to as a **population dependent** branching process (**pdBP**). As we shall see, this process is not a branching process in the common sense, not even a time inhomogeneous
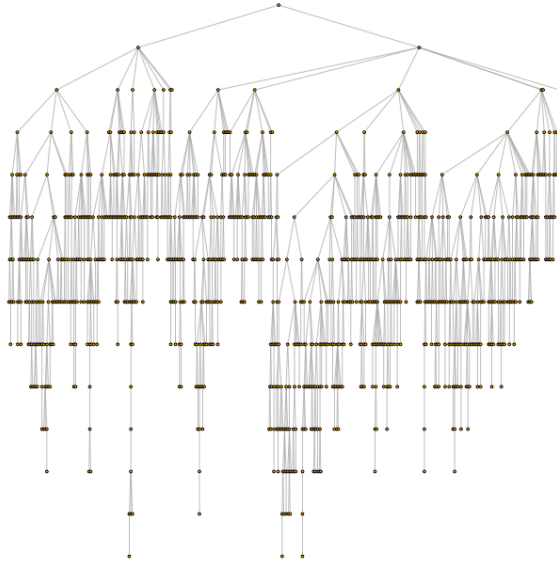
Figure 2.2: A URT of size 1000.

one, since reproduction processes of different individuals are not independent. We now describe this in more details:

Initialization: start with two individuals at time $t = 0$, with one of each type, which we shall refer to as "ancestors". For any time $t \geq 0$ let $n_A(t), n_B(t)$ denote the number of type A and type B individuals respectively. We have $n_A(0) = n_B(0) = 1$. For future reference denote by $\mathcal{F}(t)$ the σ-field generated by the process until time t and let $\{\mathcal{F}(t) : t \geq 0\}$ be the natural filtration of the process.

Types: Each individual in the system has a type $\in \{A, B\}$ and lives forever, while giving birth to other type A and B individuals (which we shall refer to as its "offsprings").

Reproduction: At any time t, the rates at which a living individual gives birth are as follows:

- For a type A individual, it gives birth to type A individuals at rate $r_{AA}(t) = q$ and type B individuals at rate $r_{AB}(t) = (1-p)(1-q)/p$.

- For a type B individual, it gives birth to type A individuals at rate

$$r_{BA}(t) = \frac{n_A(t)}{n_B(t)} \cdot (1-q)$$

17
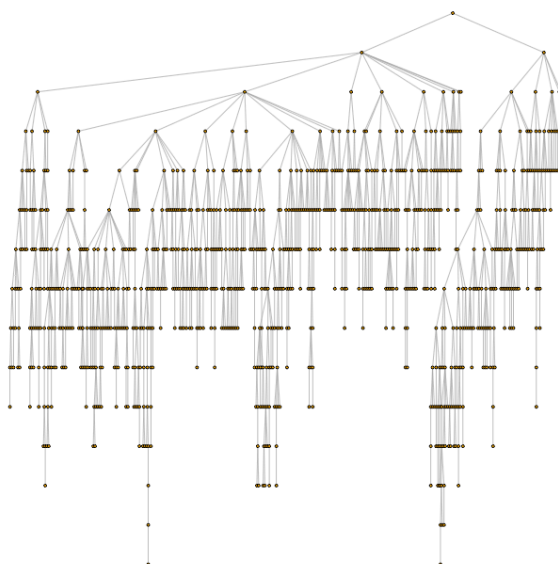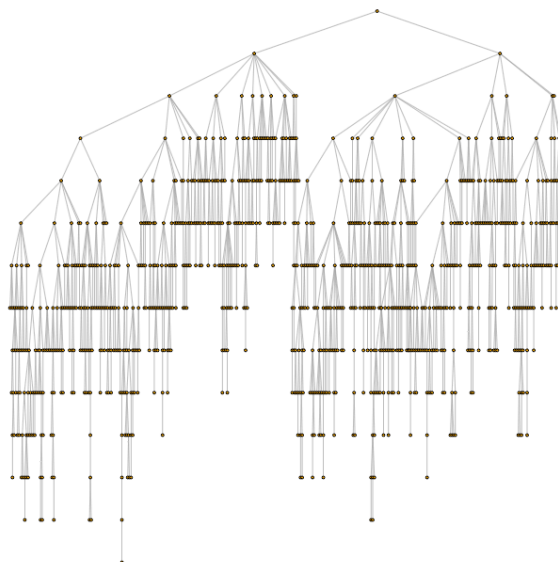
Figure 2.3: CMRTs of size 1000 with p=0.5, q=1 (top) and 0.5 (bottom). The type A root 1 is used as the root to make these plots.

and type B individuals at rate

$$r_{BB}(t) = \frac{n_A(t)}{n_B(t)} \cdot \frac{q(1-p)}{p}.$$

Write $n(t) := n_A(t) + n_B(t)$ for the total number of individuals alive at time t. Let $\mathrm{pdBP}(t)$ denote the recursive tree corresponding to the genealogical structure of the process until time t: treat each individual as a vertex and ancestors as roots, label them according to birth order, and include edges between each vertex and all of its offsprings. For roots, label the type A root as 1, the type B one as 2, and add an edge between them.

Now we are ready to state the following embedding result:

Lemma 2.2. *Consider the process* $\{\mathrm{pdBP}(t) : t \geq 0\}$ *above and for fixed* $n \geq 2$ *define the stopping time* $T_n = \inf\{t \geq 0 : n(t) = n\}$*. Then* $\mathrm{pdBP}(T_n) \overset{d}{=} \mathcal{T}_n$ *where* $\{\mathcal{T}_n\}_{n \geq 2}$ *is a CMRT. In fact,* $\{\mathrm{pdBP}(T_n)\}_{n \geq 2} \overset{d}{=} \{\mathcal{T}_n\}_{n \geq 2}$ *as processes.*

We postpone the proof to Section 2.3.5.

2.3.4 Asymptotics for CMRT

The first result we have deals with limiting degree distribution.

Theorem 2.3. *For each fixed k, let $N_k(n)$ denote the number of vertices with out-degree k in* $\mathcal{T}_n$*. Then*

$$\frac{N_k(n)}{n} \xrightarrow{P} p_k$$

where

$$p_k := \frac{p}{1+r_A^*}\left(\frac{r_A^*}{1+r_A^*}\right)^k + \frac{1-p}{1+r_B^*}\left(\frac{r_B^*}{1+r_B^*}\right)^k. \tag{2.1}$$

Here

$$r_A^* := q + \frac{(1-p)(1-q)}{p} \quad \text{and} \quad r_B^* := \frac{p(1-q)}{1-p} + q.$$

Note that when $p = 1$ the second term in p_k should be interpreted as 0.

Remark 2.1. This limiting degree distribution is a mixture of two geometric distribution shifted by 1 (to the left), with parameters $1/(1+r_A^*)$ and $1/(1+r_B^*)$. It is not hard to see that this is identical to that of URT if and only if $p = 1/2$ or $q = 1$.

Also, note that when $p \neq 1/2$ and $q \neq 1$, the limiting proportion of leaves p_0 satisfies:

$$p_0 - \frac{1}{2} = \frac{p}{1 + r_A^*} + \frac{1 - p}{1 + r_B^*} - \frac{1}{2} = \frac{(2p - 1)^2 (1 - q)^2}{2(1 - (1 - 2p)^2 q^2)} > 0.$$

Thus one characteristic of CMRT is a larger proportion of leaves. In fact, $p_0 \to 1$ as $p \to 1$ and $q \to 0$. For this extreme case, $\mathcal{T}_n$ looks like a "star" where $n - 1$ type A vertices are connect to the type B root.

To better understand Theorem 2.3, we shall make the following heuristic calculations which are made possible by continuous time embedding. First we consider cases where $p < 1$ and note that in discrete time (*i.e.* for $\{\mathcal{T}_n\}_{n \geq 2}$) the proportions of type A and B vertices converge almost surely to p and $1 - p$ respectively as $n \to \infty$ by strong law of large numbers. Turning back to continuous time (*i.e.* for $\{\mathrm{pdBP}(t)\}_{t \geq 0}$), one may imagine similar things to hold (and even prove it rigorously using stopping times). So we can reasonably expect

$$\frac{n_A(t)}{n_B(t)} \approx \frac{p}{1 - p}$$

to hold for large t. This suggests the following approximations to the rates defined in Section 2.3.3:

$$r_{BA}(t) \approx \frac{p}{1 - p} \cdot (1 - q), \qquad r_{BB}(t) \approx \frac{p}{1 - p} \cdot \frac{q(1 - p)}{p} = q.$$

Combine the rates we get $r_{AA}(t) + r_{AB}(t) = r_A^*$ and $r_{BA}(t) + r_{BB}(t) \approx r_B^*$. This suggests that one might be able to approximate our population-dependent branching process with a time homogeneous multitype branching process with rates given by the approximations above. By results from Jagers and Nerman [52], for such an approximating process, the type of a randomly chosen individual v in $\mathrm{pdBP}(t)$ and its age have limiting distribution $\pi \times \mathrm{Exponential}(\alpha)$ as $t \to \infty$, where $\times$ denotes product measure, π is a discrete measure on $\{A, B\}$ with $\pi(A) = p$ and $\pi(B) = 1 - p$, and $\alpha = 1$ is the Malthusian parameter.

Thus we have by simple calculus that

$$\mathbb{P}(v \text{ is of type A and has } k \text{ offsprings}) \approx p \int_0^\infty \mathbb{P}(\mathrm{Poisson}(r_A^* s) = k) e^{-s} \mathrm{d}s = \frac{p}{1 + r_A^*} \left(\frac{r_A^*}{1 + r_A^*}\right)^k.$$

Here s in the integral denotes the age of v, and the number of its offsprings born in a time period of length s follows a Poisson distribution with parameter $r_A^* s$ (given that v is of type A). Similar calculation can be done for type B individuals.

One may check a similar approximation for URT, for which the corresponding process is a (unitype) continuous time branching process with Poisson offspring process (say of rate 1). The limiting age distribution for a randomly chosen individual v, again by results from Jagers and Nerman [52], is Exponential(1) and simple calculus readily gives

$$\mathbb{P}(v \text{ has } k \text{ offsprings}) \approx \int_0^\infty \mathbb{P}(\text{Poisson}(s) = k)e^{-s}\mathrm{d}s = 2^{-k-1}.$$

This is exactly the limiting degree distribution of URT.

Now suppose $p = 1$, using the construction in special case 2 of Section 2.3.2 we have that for a randomly chosen individual v of type A in pdBP(t)

$$\mathbb{P}(v \text{ has } k \text{ offsprings}) \approx \sum_{i=0}^\infty 2^{-k-i-1}\binom{k+i}{k}q^k(1-q)^i = \frac{q^k}{2^{k+1}}\sum_{i=0}^\infty \binom{k+i}{k}(\frac{1-q}{2})^i = \frac{1}{q+1}(\frac{q}{q+1})^k.$$

Here the infinite sum is calculated using binomial series, which is equivalent to summing up the probability mass function of a negative binomial distribution in this case.

As oftentimes is the case in mathematics, these intuitive calculations agree with the actual theorem. In fact, the last step of our formal proof ends up with exactly the same integral. This, to some extent, shows the power of continuous time embedding in transforming intuition into fact.

Based on this limiting degree distribution, we can derive consistent estimators of model parameters p and q. One will need at least two statistics to do this and what we choose here are $N_0(n)$, the number of leaves, and $N_1(n)$, the number of vertices with out-degree 1:

Corollary 2.4. *Given that $p \neq 1/2$ and $q \neq 1$, there exist consistent estimators $\hat{p}$ and $\hat{q}$ for p and q that can be computed by solving a quadratic equation that depends only on $N_0(n)$ and $N_1(n)$.*

Remark 2.2. In practice, with a large number of vertices, one might want to use a subsample to estimate $N_0(n)$ and $N_1(n)$ instead of counting the exact number. As long as one replaces $N_0(n)$ and $N_1(n)$ with their consistent estimators, $\hat{p}$ and $\hat{q}$ will remain consistent.

The second and third results we have deal with two global statistics of the tree: maximal degree and height.

Theorem 2.5. *Let M_n denote the maximal degree in $\mathcal{T}_n$. Then when $p, q \neq 1$, there exist constant $C_1, C_2 > 0$ that depend only on p and q such that*

$$\liminf_{n \to \infty} \frac{M_n}{\log n} \geq C_1 \quad and \quad \limsup_{n \to \infty} \frac{M_n}{\log n} \leq C_2 \quad a.s.$$

When $q = 1$,
$$\frac{M_n}{\log n} \xrightarrow{a.s.} \frac{1}{\log 2}$$

and when $p = 1$ but $q \neq 1$,
$$\frac{M_n}{n} \xrightarrow{a.s.} 1 - q.$$

Theorem 2.6. *Let H_n denote the height of $\mathcal{T}_n$. Then when $p \neq 1$ or $q = 1$,*

$$\frac{H_n}{\log n} \xrightarrow{a.s.} e.$$

When $p = 1$ but $q \neq 1$,

$$\liminf_{n \to \infty} \frac{H_n}{\log n} \geq qe \quad and \quad \limsup_{n \to \infty} \frac{H_n}{\log n} \leq e \quad a.s.$$

Remark 2.3. When $p = 1$ and $q = 0$, $\mathcal{T}_n$ looks like a "star" where $n - 1$ type A vertices are connect to the type B root. And height of this tree is $H_n = 1$. In this case $\lim_{n \to \infty} H_n / \log n = 0$.

2.3.5 Proof for continuous-time embedding and some basic properties

In this section we shall first give a proof of the continuous-time embedding (Lemma 2.2).

Proof: Assume that $\{\text{pdBP}(T_n)\}_{2 \leq n \leq k} \overset{d}{=} \{\mathcal{T}_n\}_{2 \leq n \leq k}$ for a fixed integer $k \geq 2$ (which holds for $k = 2$ by definition). Conditioning on $\{\text{pdBP}(T_n)\}_{2 \leq n \leq k}$, it can be checked using properties of exponential distribution that the probability for the next individual born to be of type A is

$$\frac{n_A(T_n) r_{AA}(T_n) + n_B(T_n) r_{BA}(T_n)}{n_A(T_n)(r_{AA}(T_n) + r_{AB}(T_n)) + n_B(T_n)(r_{BA}(T_n) + r_{BB}(T_n))} = p$$

and the probability for the next type A individual born to have a parent of type A is

$$\frac{n_A(T_n)r_{AA}(T_n)}{n_A(T_n)r_{AA}(T_n) + n_B(T_n)r_{BA}(T_n)} = q.$$

Also, given a type $\in \{A, B\}$, the probability for each individual of this type to give birth to the next type A individual is equal. Similarly one can check the corresponding probability for type B individuals. Thus the dynamics of $\text{pdBP}(T_{k+1})$ conditioning on $\{\text{pdBP}(T_n)\}_{2 \leq n \leq k}$ is the same as that of $\mathcal{T}_{k+1}$ conditioning on $\{\mathcal{T}_n\}_{2 \leq n \leq k}$. Therefore by induction we have the desired result.

$\blacksquare$

Now that we have the continuous-time embedding, we shall proceed to derive some of its basic properties that will come in handy. In what follows, we shall assume that natural filtration $\{\mathcal{F}(t) : t \geq 0\}$ is used throughout.

Lemma 2.7. *The process $\left\{e^{-t}n_A(t)\right\}_{t \geq 0}$ is an $\mathbb{L}^2$-bounded positive martingale. In particular there exists a strictly positive finite random variable W such that*

$$e^{-t}n_A(t) \xrightarrow{\text{a.s.}} W, \qquad as\ t \to \infty.$$

Remark 2.4. As will be evident from the calculation below, the marginal distribution of $n_A(\cdot)$ is identical to that of a rate one Yule process starting with a single individual. In particular the limit random variable $W \overset{d}{=} \exp(1)$. Recall that a Yule process with rate λ is a time-inhomogeneous Poisson process with birth rate λi, where i is the current population size.

Proof: First we introduce some preliminary notations that will be used extensively throughout the proofs. Recall that for a jump diffusion $\{X(t) \in \mathbb{R}^n\}_{t \geq 0}$, its infinitesimal generator $\mathcal{A}$ is defined for functions $f : \mathbb{R}^n \to \mathbb{R}$ by

$$\mathcal{A}f(x) = \lim_{t \to 0^+} \frac{1}{t}\left(\mathbb{E}(f(X(t))|X(0) = x) - f(x)\right)$$

if the limit exists. Then by Dynkin's formula (see Øksendal and Sulem [83, Chapter 1.3] for a formulation) and Markov property of jump diffusion we have that $\left\{X(t) - \int_0^t \mathcal{A}X(s)ds\right\}_{t \geq 0}$ is a martingale. Note that if the diffusion term is not present and the jump part is an inhomogeneous Poisson process, we have that $\mathcal{A}X(t) = \delta(X(t))\lambda(X(t))$ where $\delta(x)$ and $\lambda(x)$ are size and intensity of jump when the process is at $x \in \mathbb{R}^n$.

Now we are ready to introduce some martingales. Denote $\exp(-t)n_A(t)$ by $\tilde{n}_A(t)$. Using the rates in Section 2.3.3 we have

$$\mathcal{A}\tilde{n}_A(t) = e^{-t}n_A(t) - e^{-t}n_A(t) = 0$$

and

$$\mathcal{A}\tilde{n}_A^2(t) = e^{-2t}(2n_A(t) + 1)n_A(t) - 2e^{-2t}n_A^2(t) = e^{-2t}n_A(t).$$

Thus by Dynkin's formula we have that both $\{\tilde{n}_A(t) : t \geq 0\}$ and

$$\tilde{n}_A^2(t) - \int_0^t e^{-2s}n_A(s)ds, \qquad t \geq 0$$

are martingales. Taking expectation of both martingales we get $\mathbb{E}(n_A(t)) = \exp(t)$ and $\mathbb{E}(\tilde{n}_A^2(t)) = 2 - \exp(-t)$. Therefore $\{\tilde{n}_A(t) : t \geq 0\}$ is $\mathbb{L}^2$ bounded and the second statement follows from standard martingale convergence theorem.

∎

Lemma 2.8. *Define $Z(t) := pn_B(t) - (1-p)n_A(t)$. Then $\{Z(t) : t \geq 0\}$ is a martingale and further $e^{-t}Z(t) \xrightarrow{a.s.} 0$. This implies $e^{-t}n_B(t) \xrightarrow{a.s.} (1-p)W/p$ where W is as in Lemma 2.7.*

Proof: Again using the rates in Section 2.3.3 we have

$$\mathcal{A}Z(t) = p\frac{1-p}{p}n_A(t) - (1-p)n_A(t) = 0$$

and

$$\mathcal{A}Z^2(t) = p^2\mathcal{A}n_B^2(t) + (1-p)^2\mathcal{A}n_A^2(t) - 2p(1-p)\mathcal{A}n_An_B(t),$$

with

$$\mathcal{A}n_B^2(t) = (2n_B(t) + 1)\frac{1-p}{p}n_A(t), \quad \mathcal{A}n_A^2(t) = (2n_A(t) + 1)n_A(t)$$

and

$$\mathcal{A}n_An_B(t) = n_A(t)\frac{1-p}{p}n_A(t) + n_B(t)n_A(t).$$

Thus by some elementary algebra and Dynkin's formula we have that both $\{Z(t) : t \geq 0\}$ and

$$M(t) = Z^2(t) - \int_0^t (1-p)n_A(s)ds, \qquad t \geq 0$$

are martingales. From Lemma 2.7 we get that $\mathbb{E}(n_A(s)) = \exp(s)$. Taking expectation of $\mathbb{E}(M(t))$ shows

$$\mathbb{E}(Z^2(t)) = (1-p)(e^t - 1) + (2p-1)^2.$$

Now apply Markov's inequality to $\exp(-4 \log n)Z^2(2 \log n)$ we have for any $\epsilon > 0$:

$$\mathbb{P}\left(e^{-4 \log n}Z^2(2 \log n) > \epsilon\right) \leq \frac{e^{-4 \log n}((1-p)(e^{2 \log n} - 1) + (2p-1)^2)}{\epsilon} = \frac{1-p}{n^2 \epsilon} + \frac{4p^2 - 3p}{n^4 \epsilon}.$$

Thus by the Borel-Cantelli lemma $\exp(-2 \log n)Z(2 \log n) \xrightarrow{\text{a.s.}} 0$. Since we know by Lemma 2.7 that $\exp(-2 \log n)n_A(2 \log n) \xrightarrow{\text{a.s.}} W$, we get $\exp(-2 \log n)n_B(2 \log n) \xrightarrow{\text{a.s.}} (1-p)W/p$.

Finally for any $t > 0$ we can find a positive integer n such that $2 \log n \leq t < 2 \log(n+1)$. By monotonicity of $n_B(\cdot)$ we have $n_B(2 \log n) \leq n_B(t) < n_B(2 \log(n+1))$ and further

$$e^{-2 \log(n+1)}n_B(2 \log n) \leq e^{-t}n_B(t) < e^{-2 \log n}n_B(2 \log(n+1)).$$

Since the left hand side

$$e^{-2 \log(n+1)}n_B(2 \log n) = e^{-2 \log n}n_B(2 \log n) \cdot \frac{n^2}{(n+1)^2} \xrightarrow{\text{a.s.}} \frac{(1-p)W}{p}$$

and similarly the right hand side converges a.s. to the same limit, we have $\exp(-t)n_B(t) \xrightarrow{\text{a.s.}} (1-p)W/p$. This immediately implies $\exp(-t)Z(t) \xrightarrow{\text{a.s.}} 0$.

∎

Remark 2.5. This lemma essentially proves that $n_B(t)/n_A(t) \xrightarrow{\text{a.s.}} (1-p)/p$, which is used in Section 2.3.4 for heuristic calculations. As noted there, the corresponding statement in discrete time (*i.e.* the original tree-valued process) follows easily from strong law of large numbers. Judging from this point alone, it might seem that continuous-time embedding is making things more complicated. However, as we shall see later, all this hard work is well worth the effort and the embedding really facilitates our analysis of degree asymptotics.

Lemma 2.9. *The population size process $n(t)$ satisfies $e^{-t}n(t) \xrightarrow{\text{a.s.}} W + (1-p)W/p := W_\infty$. In particular, the sequence of stopping times T_n satisfy*

$$T_n - \log n \xrightarrow{\text{a.s.}} -\log(W_\infty),$$

where W_∞ is a strictly positive finite random variable.

The first statement here is a direct corollary of Lemma 2.7 and 2.8, and the second statement follows by replacing t with T_n. Note that the first statement essentially says that the Malthusian parameter for process $\{n(t) : t \geq 0\}$ is 1.

2.3.6 Proof for limiting degree distribution

In this section we shall prove Theorem 2.3.

To work in continuous time, we need to reformulate Theorem 2.3 via the embedding. Let $N_{k,A}(n)$ and $N_{k,B}(n)$ denote the number of type A and B vertices with out-degree k in $\mathcal{T}_n$. The plan is as follows: first we focus on type A vertices and prove the result below, and as similar result holds for type B vertices, using $N_k(n) = N_{k,A}(n) + N_{k,B}(n)$ we completes the proof. Note that when $p = 1$, there is no need to consider type B vertices.

Theorem 2.10. *For a fixed integer $k > 0$, let $n_{k,A}(t)$ denote the number of type A individuals with k offsprings in* $\mathrm{pdBP}(t)$. *Then*

$$\frac{n_{k,A}(t)}{n(t)} \xrightarrow{P} p_{k,A}, \qquad as \ t \to \infty.$$

Here

$$p_{k,A} := p \int_0^\infty \mathbb{P}(\mathrm{Poisson}(r_A^* s) = k) e^{-s} ds = \frac{p}{1 + r_A^*} \left(\frac{r_A^*}{1 + r_A^*} \right)^k$$

where r_A^* is the total reproduction rate of type A individuals:

$$r_A^* = q + \frac{(1-p)(1-q)}{p}. \tag{2.2}$$

Thus by the embedding in Lemma 2.2 for $\{\mathcal{T}_n\}_{n \geq 2}$ we have

$$\frac{N_{k,A}(n)}{n} \xrightarrow{P} p_{k,A}, \qquad as \ n \to \infty.$$

Proof: Throughout the proof we work with the continuous time embedding. For any fixed constant $0 < a < t$, let $n_{k,A}[t - a, t]$ be the number of type A individuals born in the interval $[t - a, t]$ that have exactly k offsprings by time t. Given Lemma 2.9, it is enough to show the following two propositions.

Proposition 2.11.

$$\limsup_{a\to\infty} \limsup_{t\to\infty} e^{-t}\left(n_{k,A}(t) - n_{k,A}[t-a,t]\right) = 0, \qquad a.s.$$

Proof: Since the population size $n_A(t)$ grows exponentially, most type A individuals are born after time $t-a$. Indeed, $n_{k,A}(t) - n_{k,A}[t-a,t] = n_{k,A}(t-a) \leq n_A(t-a)$, and by Lemma 2.7 we have

$$\limsup_{t\to\infty} e^{-t}\left(n_{k,A}(t) - n_{k,A}[t-a,t]\right) \leq e^{-a} \lim_{t\to\infty} e^{-(t-a)} n_A(t-a) = e^{-a}W, \qquad a.s.$$

Letting $a \to \infty$ proves the proposition.

∎

Proposition 2.12. *Recall the random variable W in Lemma 2.8. Then for each fixed $a > 0$, we have*

$$e^{-t} n_{k,A}[t-a,t] \xrightarrow{P} W \int_0^a \mathbb{P}(\mathrm{Poisson}(r_A^* s) = k)e^{-s}ds$$

as $t \to \infty$.

This assertion needs some work and the proof follows a similar procedure as that in Bhamidi et al. [15, Section 4.2.2]. First, recall from Lemma 2.7 that $n_A(t) \approx We^t$ for large t. For our proof we will need a finer concentration result that goes as follows:

Lemma 2.13.

$$\mathbb{P}\left(\sup_{t-a\leq s\leq t} |n_A(s) - We^s| < \sqrt{te^t}\right) \to 1$$

as $t \to \infty$ where W is as in Lemma 2.7. Equivalently, we shall say that w.h.p. as $t \to \infty$, $\sup_{t-a\leq s\leq t} |n_A(s) - W\exp(s)| < \sqrt{t\exp(t)}$.

Proof: First note that

$$e^{-s} n_A(s) - e^{-(t-a)} n_A(t-a), \qquad s \geq t-a$$

is a martingale by Lemma 2.7 and recall that $\mathbb{E}(\exp(-2s)n_A^2(s)) = 2 - \exp(-s)$ from the proof.

Now fix any $T > t-a$, Doob's L^2-maximal inequality (here we use the version stated in Øksendal [83, Theorem 3.2.4], see also Karatzas and Shreve [56, Theorem 3.8] for a proof) applied to the above martingale gives, for any $C > 0$:

$$\mathbb{P}\left(\sup_{t-a\leq s\leq T}|e^{-s}n_A(s)-e^{-(t-a)}n_A(t-a)|>C\right)\leq\frac{\mathbb{E}(e^{-T}n_A(T)-e^{-(t-a)}n_A(t-a))^2}{C^2}\quad(2.3)$$

Recall that $\{\tilde{n}_A(t):=\exp(-t)n_A(t)\}_{t\geq0}$ is a martingale so we have

$$\mathbb{E}(\tilde{n}_A(T)\tilde{n}_A(t-a))=\mathbb{E}\left(\tilde{n}_A(t-a)\,\mathbb{E}\left(\tilde{n}_A(T)|\tilde{n}_A(t-a)\right)\right)=\mathbb{E}(\tilde{n}_A^2(t-a))=\mathbb{E}(e^{-2(t-a)}n_A^2(t-a)).$$

It follows that

$$\mathbb{E}(e^{-T}n_A(T)-e^{-(t-a)}n_A(t-a))^2=\mathbb{E}(e^{-2(t-a)}n_A^2(t-a))-\mathbb{E}(e^{-2T}n_A^2(T))=e^{-(t-a)}-e^{-T}.$$

Plug this in (2.3) we get

$$\mathbb{P}\left(\sup_{t-a\leq s\leq T}|e^{-s}n_A(s)-e^{-(t-a)}n_A(t-a)|>C\right)\leq\frac{e^{-(t-a)}-e^{-T}}{C^2}\quad(2.4)$$

Now, let $T\to\infty$ and use the a.s. convergence result from Lemma 2.7 to yield

$$\mathbb{P}\left(|W-e^{-(t-a)}n_A(t-a)|>C\right)\leq\frac{e^{-(t-a)}}{C^2}.$$

On the other hand, let $T=t$ in (2.4) we get

$$\mathbb{P}\left(\sup_{t-a\leq s\leq t}|e^{-s}n_A(s)-e^{-(t-a)}n_A(t-a)|>C\right)\leq\frac{e^{-(t-a)}-e^{-t}}{C^2}\leq\frac{e^{-(t-a)}}{C^2}.$$

Finally, combine the above two inequalities we have

$$\mathbb{P}\left(\sup_{t-a\leq s\leq t}|e^{-t}n_A(t)-W|>2C\right)\leq\frac{2e^{-(t-a)}}{C^2}.$$

Let $C=\sqrt{t\exp(-t)}/2$ and after some simple algebraic manipulation we get the wanted result.

∎

Now that we have Lemma 2.13, we will proceed to approximate the integral in Proposition 2.12 by $e^{-t}n_{k,A}[t-a,t]$. To do so, we divide the interval $[t-a,t]$ into intervals of length $\delta:=e^{-t/3}$ and denote by $\mathcal{I}_i(1\leq i\leq a/\delta)$ the i-th interval $[t-a+(i-1)\delta,t-a+i\delta]$ (for simplicity we

treat a/δ as if it is an integer but this should not matter). Let $t_i = t - a + (i-1)\delta$ and write $n_A(\mathcal{I}_i)$ for the number of type A individuals born in $\mathcal{I}_i$.

On the event given in Lemma 2.13, we have

$$|n_A(t_i + \delta) - n_A(t_i) - We^{t_i}(e^\delta - 1)| < 2\sqrt{te^t}.$$

Since $W \exp(t_i)(\exp(\delta) - 1) = \delta W \exp(t_i) + o_p(\sqrt{\exp(t)})$, we can further get

$$\mathbb{P}\left(\bigcap_{i=1}^{a/\delta} \left\{ |n_A(\mathcal{I}_i) - \delta W e^{t_i}| < 3\sqrt{te^t} \right\}\right) \to 1 \qquad (2.5)$$

as $t \to \infty$. This shows that the number of type A individuals born in $\mathcal{I}_i$ is approximately $\delta W e^{t_i}$.

Unfortunately, even individuals born in the same interval could start reproducing at different times (*i.e.* have distinct birth times), and we need to at least align these individuals. To facilitate the analysis, for each type A individual born in $\mathcal{I}_i$, we call it "good" if it gives birth to no offspring (type A or B) in $\mathcal{I}_i$ and "bad" otherwise. Since the interval is small, we expect most individuals born to be good:

Lemma 2.14. *There exists a constant $M > 0$ such that*

$$\mathbb{P}\left(\bigcap_{i=1}^{a/\delta} \left\{ n_A^{bad}(\mathcal{I}_i) < MWte^{t/3} \right\}\right) \to 1$$

as $t \to \infty$, where $n_A^{bad}(\mathcal{I}_i)$ is the number of bad individuals born in $\mathcal{I}_i$.

Proof: Call a bad individual in $\mathcal{I}_i$ a "direct" bad individual if it is an offspring of individuals born before $\mathcal{I}_i$ and write $n_A^{dir}(\mathcal{I}_i)$ for the number of direct bad individuals born in $\mathcal{I}_i$. As non-direct bad individuals in $\mathcal{I}_i$ must have an antecedent that is direct bad in $\mathcal{I}_i$, we have $n_A^{bad}(\mathcal{I}_i) \le n_A^{desc}(\mathcal{I}_i)$, where $n_A^{desc}(\mathcal{I}_i)$ is the number of descendants of direct bad individuals in $\mathcal{I}_i$.

Note that a direct bad individual in $\mathcal{I}_i$ has to satisfy the following two conditions:

- First it has to be an offspring of individuals born before $\mathcal{I}_i$ and the number of such offsprings is at most $n_A(\mathcal{I}_i)$.

29

- Secondly it has to give birth to at least one offspring in $\mathcal{I}_i$ and the corresponding (conditional) probability is at most

$$p_{bad} := \mathbb{P}(\mathrm{Exp}(r_A^*) \leq \delta) \sim r_A^* \delta.$$

Combining the above conditions we have

$$n_A^{dir}(\mathcal{I}_i) \preceq \mathrm{Bin}(n_A(\mathcal{I}_i), p_{bad}).$$

where $\preceq$ denotes stochastic dominance. By (2.5) we have that w.h.p. as $t \to \infty$, $W \exp(2t/3)/2 \leq n_A(\mathcal{I}_i) \leq 2W \exp(2t/3)$ for all i. Condition on this event and use Hoeffding inequality for binomial distribution we get

$$\mathbb{P}\left(n_A^{dir}(\mathcal{I}_i) > n_\delta p_{bad} + \sqrt{n_\delta \log(n_\delta)} \right) \leq \frac{1}{n_A^2(\mathcal{I}_i)} \leq \frac{4}{W^2 \exp(4t/3)}$$

where $n_\delta := 2W \exp(2t/3)$. Since there exists $C_1 > 0$ such that $n_\delta p_{bad} + \sqrt{n_\delta \log(n_\delta)} < C_1 W t e^{t/3}$ for large enough t, and

$$\frac{a}{\delta} \cdot \frac{4}{W^2 \exp(4t/3)} \to 0$$

as $t \to \infty$, using a union bound we have

$$\mathbb{P}\left(\bigcap_{i=1}^{a/\delta} \left\{ n_A^{dir}(\mathcal{I}_i) \leq C_1 W t e^{t/3} \right\} \right) \to 1 \tag{2.6}$$

as $t \to \infty$.

Conditioning on the event in (2.6) we have

$$n_A^{desc}(\mathcal{I}_i) \preceq \sum_{j=1}^{n_A^{dir}(\mathcal{I}_i)} Y_j(\delta) \preceq \sum_{j=1}^{C_1 W t \exp(t/3)} Y_j(\delta)$$

where $\{Y_j(\cdot) : j \geq 1\}$ is an sequence of *i.i.d.* Yule process with rate r_A^*. Since $Y_j(t)$ follows a geometric distribution with parameter $\exp(-r_A^* t)$, we have that for any positive integer C_2:

$$\mathbb{P}(Y_j(\delta) \geq C_2) = (1 - \exp(-r_A^* \delta))^{C_2 - 1} \leq (r_A^* \delta)^{C_2 - 1}$$

Using a union bound it follows (for simplicity the conditional is suppressed) that

$$\mathbb{P}(n_A^{desc}(\mathcal{I}_i) \geq C_1 C_2 W t e^{t/3}) \leq C_1 W t e^{t/3} \, \mathbb{P}(Y_j(\delta) \geq C_2) \leq C_1 W (r_A^*)^{C_2-1} t (e^{t/3})^{2-C_2}.$$

Apply union bound once again we get

$$\mathbb{P}\left(\bigcup_{i=1}^{a/\delta} \left\{ n_A^{desc}(\mathcal{I}_i) \geq C_1 C_2 W t e^{t/3} \right\} \right) \leq \frac{a}{\delta} \cdot C_1 W (r_A^*)^{C_2-1} t (e^{t/3})^{2-C_2} \to 0$$

as $t \to \infty$ for $C_2 \geq 4$. Recall that $n_A^{bad}(\mathcal{I}_i) \leq n_A^{desc}(\mathcal{I}_i)$ and we can get rid of the conditional that was suppressed with the aid of (2.6). Then letting $M = 4C_1$ completes the proof.

∎

Now that we get the bad individuals under control, we can turn our attention to those good individuals. To start, combine Lemma 2.14 with (2.5) we have

$$\mathbb{P}\left(\bigcap_{i=1}^{a/\delta} \left\{ |n_A^{good}(\mathcal{I}_i) - \delta W e^{t_i}| < 4\sqrt{t e^t} \right\} \right) \to 1 \tag{2.7}$$

as $t \to \infty$, where $n_A^{good}(\mathcal{I}_i)$ is the number of good individuals in $\mathcal{I}_i$.

Next, note that good individuals in $\mathcal{I}_i$ reproduce independently at rate r_A^* starting from time $t_i + \delta = t_{i+1}$. In particular the probability that a good individual has k offsprings by time t is

$$g_k(t - t_{i+1}) := \mathbb{P}(\text{Poisson}(r_A^*(t - t_{i+1})) = k).$$

Here $t - t_{i+1}$ refers to the time left until time t. Since reproductions are all independent, we have (conditioning on pdBP(t_{i+1}))

$$n_{k,A}^{good}(\mathcal{I}_i) \stackrel{\text{d}}{=} \text{Bin}(n_A^{good}(\mathcal{I}_i), g_k(t - t_{i+1})) \tag{2.8}$$

where $n_{k,A}^{good}(\mathcal{I}_i)$ is the number of good individuals in $\mathcal{I}_i$ that have k offsprings by time t. Similar to the proof of (2.6), by (2.7), (2.8) and Hoeffding inequality for binomial distribution together with a union bound, there exists $C > 0$ such that

$$\mathbb{P}\left(\bigcap_{i=1}^{a/\delta} \left\{ |n_{k,A}^{good}(\mathcal{I}_i) - \delta W e^{t_i} g_k(t - t_{i+1})| < CW \log(W) t e^{t/3} \right\} \right) \to 1 \tag{2.9}$$

as $t \to \infty$.

With both good and bad individuals under control, we are ready to prove what we started out for. Note that

$$\sum_{i=1}^{a/\delta} n_{k,A}^{good}(\mathcal{I}_i) \leq n_{k,A}[t-a,t] \leq \sum_{i=1}^{a/\delta} [n_{k,A}^{good}(\mathcal{I}_i) + n_A^{bad}(\mathcal{I}_i)].$$

For the bad individuals we know from Lemma 2.14 that

$$e^{-t} \sum_{i=1}^{a/\delta} n_A^{bad}(\mathcal{I}_i) \xrightarrow{P} 0 \qquad \text{as} \qquad t \to \infty.$$

On the other hand, for the good individuals we have from (2.9) that

$$e^{-t} \sum_{i=1}^{a/\delta} n_{k,A}^{good}(\mathcal{I}_i) - \sum_{i=1}^{a/\delta} \delta W e^{-(t-t_i)} g_k(t-t_{i+1}) \xrightarrow{P} 0 \qquad \text{as} \qquad t \to \infty.$$

Finally, note that by definition of Riemann integral we know

$$\sum_{i=1}^{a/\delta} \delta W e^{-(t-t_i)} g_k(t-t_{i+1}) \to W \int_0^a \mathbb{P}(\text{Poisson}(r_A^* s) = k) e^{-s} ds \qquad \text{as} \qquad t \to \infty.$$

This completes the proof for Proposition 2.12.

∎

2.3.7   Estimation of model parameters

In this section we derive consistent estimators for model parameters p and q based on the limiting degree distribution (i.e. Corollary 2.4).

Specifically, from Theorem 2.3 we know

$$\frac{N_0(n)}{n} \xrightarrow{P} \frac{p}{1+r_A^*} + \frac{1-p}{1+r_B^*}$$

and

$$\frac{N_1(n)}{n} \xrightarrow{P} \frac{p r_A^*}{(1+r_A^*)^2} + \frac{(1-p) r_B^*}{(1+r_B^*)^2}.$$

Let $\theta_A := 1/(1 + r_A^*)$ and $\theta_B := 1/(1 + r_B^*)$. If we can solve for unique $\hat{p}$, $\hat{\theta}_A$ and $\hat{\theta}_B$ (up to a switch between two types) that satisfy the following equations

$$m_1 := \frac{N_0(n)}{n} = \hat{p}\hat{\theta}_A + (1 - \hat{p})\hat{\theta}_B \qquad (2.10)$$

$$m_2 := \frac{N_0(n) - N_1(n)}{n} = \hat{p}\hat{\theta}_A^2 + (1 - \hat{p})\hat{\theta}_B^2, \qquad (2.11)$$

then $\hat{p}$, $\hat{\theta}_A$ and $\hat{\theta}_B$ are consistent estimators of p, θ_A and θ_B. We can further solve for a consistent estimator $\hat{q}$ for q from either $\hat{\theta}_A$ or $\hat{\theta}_B$ together with $\hat{p}$. Note that all estimators are consistent because (as we shall see) they are continuous functions of m_1 and m_2.

To solve three unknowns from two equations (2.10) and (2.11), we need a third equation

$$\frac{\hat{p}}{\hat{\theta}_A} + \frac{1 - \hat{p}}{\hat{\theta}_B} = 2 = p(1 + r_A^*) + (1 - p)(1 + r_B^*) \qquad (2.12)$$

where the second equality can be verified from the definitions of r_A^* and r_B^*.

From (2.10) we have
$$\hat{p} = \frac{m_1 - \hat{\theta}_B}{\hat{\theta}_A - \hat{\theta}_B}. \qquad (2.13)$$

Plug it back to (2.11) and (2.12) we get $m_1(\hat{\theta}_A + \hat{\theta}_B) - \hat{\theta}_A\hat{\theta}_B = m_2$ and $\hat{\theta}_A + \hat{\theta}_B - m_1 = 2\hat{\theta}_A\hat{\theta}_B$. Then we can solve for

$$\hat{\theta}_A\hat{\theta}_B = \frac{m_2 - m_1^2}{2m_1 - 1} \quad \text{and} \quad \hat{\theta}_A + \hat{\theta}_B = \frac{2m_2 - m_1}{2m_1 - 1},$$

so $\hat{\theta}_A$ and $\hat{\theta}_B$ are roots of the following quadratic equation:

$$x^2 - \frac{2m_2 - m_1}{2m_1 - 1}x + \frac{m_2 - m_1^2}{2m_1 - 1} = 0.$$

After we solve for $\hat{\theta}_A$ and $\hat{\theta}_B$ we can compute $\hat{p}$ from (2.13) and

$$\hat{q} = \frac{\hat{p} - \hat{\theta}_A}{(2\hat{p} - 1)\hat{\theta}_A} = \frac{1 - \hat{p} - \hat{\theta}_B}{(1 - 2\hat{p})\hat{\theta}_B}$$

from definitions of θ_A and θ_B.

Note that these estimators fail when $m_1 - 1/2$, $m_2 - m_1^2$ or $2m_2 - m_1$ is negative. However, these cases are unlikely to occur for CMRT as all three quantities have positive limits as $n \to \infty$

(recall that m_1 and m_2 both depend on n). When $m_1 = 1/2$, these estimators also fail: from the remark under Theorem 2.3 we know m_1 has limit $1/2$ if and only if $p = 1/2$ or $q = 1$, and parameters p, q are not identifiable from the limiting degree distribution in these special cases.

2.3.8 Proof for maximal degree

In this section we shall prove Theorem 2.5. Throughout the proof we work with the continuous time embedding unless otherwise noted.

First we consider cases where $p, q \neq 1$. By Lemma 2.9 and Egoroff's Theorem, given any $\epsilon > 0$, we can choose $K > 0$ such that (the dependence of K on ϵ is suppressed throughout):

$$\mathbb{P}\left(\sup_n |T_n - \log n| < K\right) > 1 - \epsilon. \tag{2.14}$$

To ease notations, write $T_n^+ = \log n + K$ and $T_n^- = \log n - K$.

Lower bound: Here it is enough to consider just the type A root. Recall from Section 2.3.3 that this vertex reproduces at constant rate

$$r_A^* = q + \frac{(1-p)(1-q)}{p}. \tag{2.15}$$

Denote by $D(t)$ the out-degree of type A root at time t. Then $D(t)$ follows a Poisson($r_A^* t$) distribution. So for any $0 < \gamma < 1$, by standard tail bound for Poisson distribution we have

$$\mathbb{P}\left(D(T_n) \leq \gamma r_A^* \log n\right) \leq \mathbb{P}\left(D(T_n^-) \leq \gamma r_A^* \log n\right) \leq \exp\{-M \log n\} \tag{2.16}$$

for large enough n, conditioning on the event in (2.14). Here $M > 0$ is a constant that depends on both γ and r_A^*.

Denote by $M(t)$ the maximal number of offsprings an individual has by time t. Since $M(t) \geq D(t)$, it follows from (2.16) that

$$\mathbb{P}\left(M(T_n) \leq \gamma r_A^* \log n\right) \leq n^{-M} \tag{2.17}$$

conditioning on the event in (2.14).

Let $\{n_k\}_{k \geq 1}$ be an increasing sequence of positive integers such that $n_k \geq k^{2/M}$. Then from (2.17) we know

$$\mathbb{P}\left(M(T_{n_k}) \leq \gamma r_A^* \log n_k\right) \leq \frac{1}{k^2} \tag{2.18}$$

for large k, conditioning on the event in (2.14). By Borel-Cantelli lemma and (2.14) it follows that

$$\mathbb{P}\left(\liminf_{k \to \infty} \frac{M(T_{n_k})}{\log n_k} \leq \gamma r_A^*\right) \leq \varepsilon.$$

From that we can also get

$$\mathbb{P}\left(\liminf_{n \to \infty} \frac{M(T_n)}{\log n} \leq \gamma r_A^*\right) \leq \varepsilon.$$

Recall that $M_n \overset{d}{=} M(T_n)$ from the embedding, let $\gamma \to 1$ and $\varepsilon \to 0$ we have

$$\mathbb{P}\left(\liminf_{n \to \infty} \frac{M_n}{\log n} \geq r_A^*\right) = 1.$$

This completes the proof for the lower bound part, with $C_1 = r_A^*$.

Alternatively, we provide here another proof for the lower bound part that works when $q \neq 0$. Ideas used in the proof shall become useful later on.

Consider type A individuals alone. We call a type A individual "pure-blooded" if all of its antecedents are type A individuals. Define $\text{pdBP}_A(t)$ to be the branching process consisting of all pure-blooded type A individuals in $\text{pdBP}(t)$. Let $T_{n,A} := \inf\{t \geq 0 : |\text{pdBP}_A(t)| = n\}$. Note that pure-blooded type A individuals give birth to new pure-blooded type A individuals at constant rate q, mimicking the results for CMRT (*i.e.* Lemma 2.2 and Lemma 2.9) it is not hard to see that:

- $\{\text{pdBP}_A(T_{n,A})\}_{n \geq 1} \overset{d}{=} \{\mathcal{U}_n\}_{n \geq 1}$ as processes, where $\{\mathcal{U}_n\}_{n \geq 1}$ is a URT.

- The sequence of stopping times $T_{n,A}$ satisfy

$$T_{n,A} - \frac{1}{q} \log n \overset{\text{a.s.}}{\longrightarrow} -\log(W_A) \tag{2.19}$$

for a finite positive random variable $W_A \overset{d}{=} \exp(1)$.

Since we can transform results for URT into that of the branching process $\mathrm{pdBP}_A(t)$ through embedding, using Devroye and Lu [35] we get

$$\frac{M_A^{pure}(T_{n,A})}{\log n} \xrightarrow{\text{a.s.}} \frac{1}{\log 2}$$

where $M_A^{pure}(t)$ is the maximal number of offsprings a pure-blooded type A individual has by time t. Similar to (2.14) we have that by (2.19) and Egoroff's Theorem, given any $\epsilon > 0$, we can choose $K_A > 0$ such that (again for simplicity the dependence of K_A on ϵ is suppressed throughout):

$$\mathbb{P}\left(\sup_n |T_{n,A} - \frac{1}{q}\log n| < K_A\right) > 1 - \epsilon. \qquad (2.20)$$

For each $t > 0$, let $n^*(t)$ be the positive integer satisfying $\frac{1}{q}\log(n^*(t)) + K_A \leq t < \frac{1}{q}\log(n^*(t) + 1) + K_A$. Conditioning on the event in (2.20) we have

$$\frac{M_A^{pure}(t)}{t} \geq \frac{M_A^{pure}(\frac{1}{q}\log(n^*(t)) + K_A)}{\frac{1}{q}\log(n^*(t)+1) + K_A} \geq \frac{M_A^{pure}(T_{n^*(t),A})}{\frac{1}{q}\log(n^*(t)+1) + K_A} \xrightarrow{\text{a.s.}} \frac{q}{\log 2}$$

as $t \to \infty$, and it follows that

$$\mathbb{P}\left(\liminf_{t\to\infty} \frac{M_A^{pure}(t)}{t} \geq \frac{q}{\log 2}\right) > 1 - \epsilon.$$

Denote by $M(t)$ the maximal number of offsprings an individual has by time t. Let $\epsilon \to 0$ in the above inequality and note $M(t) \geq M_A^{pure}(t)$, we have

$$\mathbb{P}\left(\liminf_{t\to\infty} \frac{M(t)}{t} \geq \frac{q}{\log 2}\right) = 1.$$

Finally, since

$$\frac{T_n}{\log n} \xrightarrow{\text{a.s.}} 1, \qquad (2.21)$$

by Lemma 2.9, and $M_n \overset{d}{=} M(T_n)$ from the embedding, we get

$$\mathbb{P}\left(\liminf_{n\to\infty} \frac{M_n}{\log n} \geq \frac{q}{\log 2}\right) = 1.$$

This completes the proof for the lower bound part when $q \neq 0$, with $C_1 = \frac{q}{\log 2}$.

Upper bound: First we consider the simpler type A individuals. Given $\{\mathrm{pdBP}(t)\}_{t\geq 0}$, we couple it with another process where: whenever a type A individual is born to a type B

individual in $\{\text{pdBP}(t)\}_{t\geq 0}$, chose uniformly at random a living type A individual, and treat the newborn individual as an offspring of that chosen individual. In this new process, if we look at type A individuals alone, it is not hard to see that they give birth to new type A individuals at constant rate 1. Denote by $M_A(t)$ and $\widetilde{M}_A(t)$ respectively the maximal number of offsprings a type A individual has by time t in $\{\text{pdBP}(t)\}_{t\geq 0}$ and new process. Using the same argument as in the proof for lower bound we see that

$$\mathbb{P}\left(\limsup_{t\to\infty} \frac{\widetilde{M}_A(t)}{t} \leq \frac{1}{\log 2}\right) = 1.$$

Since $\widetilde{M}_A(t) \geq M_A(t)$, this immediately implies

$$\mathbb{P}\left(\limsup_{t\to\infty} \frac{M_A(t)}{t} \leq \frac{1}{\log 2}\right) = 1.$$

Denote by $M_{n,A}$ the maximal degree of type A vertices in $\mathcal{T}_n$. Then it follows from (2.21) and $M_{n,A} \overset{d}{=} M_A(T_n)$ that

$$\mathbb{P}\left(\limsup_{n\to\infty} \frac{M_{n,A}}{\log n} \leq \frac{1}{\log 2}\right) = 1.$$

Next we consider the more complicated type B individuals. For a given type B individual v and time $T \in [0, t)$, write $n_v[T, t]$ for the number of offsprings this individual produced in time interval $[T, t]$. This is a pure birth process with rate

$$c\frac{n_A(t)}{n_B(t)} := \left(1 - q + \frac{(1-p)q}{p}\right)\frac{n_A(t)}{n_B(t)}.$$

Therefore for a fixed T, the following process

$$X(t) = n_v[T, t] - \int_T^t c\frac{n_A(s)}{n_B(s)}ds, \qquad t \geq T, \tag{2.22}$$

is a martingale (here the infinitesimal generator is exactly the rate).

To handle the variability of $X(t)$, we will need its predictable quadratic variation process $\langle X \rangle(t)$. Note that

$$\mathcal{A}X^2(t) = \mathcal{A}n_v^2[T, t] + \mathcal{A}\left(\int_T^t c\frac{n_A(s)}{n_B(s)}ds\right)^2 - 2\mathcal{A}\left(n_v[T, t]\int_T^t c\frac{n_A(s)}{n_B(s)}ds\right)$$

with

$$\mathcal{A}n_v^2[T,t] = (2n_v[T,t]+1)c\frac{n_A(t)}{n_B(t)},$$

$$\mathcal{A}\left(\int_T^t c\frac{n_A(s)}{n_B(s)}ds\right)^2 = 2\left(\int_T^t c\frac{n_A(s)}{n_B(s)}ds\right)c\frac{n_A(t)}{n_B(t)}$$

and

$$\mathcal{A}\left(n_v[T,t]\int_T^t c\frac{n_A(s)}{n_B(s)}ds\right) = n_v[T,t]c\frac{n_A(t)}{n_B(t)} + \left(\int_T^t c\frac{n_A(s)}{n_B(s)}ds\right)c\frac{n_A(t)}{n_B(t)}.$$

It then follows from some elementary algebra that $\mathcal{A}X^2(t) = cn_A(t)/n_B(t)$ and

$$\langle X\rangle(t) = \int_T^t c\frac{n_A(s)}{n_B(s)}ds, \qquad t \geq T. \tag{2.23}$$

Now use Lemma 2.8 to choose T such that

$$\mathbb{P}\left(\sup_{t\geq T}\left|\frac{n_A(t)}{n_B(t)} - \frac{p}{1-p}\right| > \varepsilon\right) \leq \varepsilon. \tag{2.24}$$

Also, define the stopping time

$$S = \inf\left\{t \geq T : \left|\frac{n_A(t)}{n_B(t)} - \frac{p}{1-p}\right| > \varepsilon\right\}.$$

Observe that by our choice of T we have $\mathbb{P}(S < \infty) \leq \varepsilon$. The idea here is to bound $n_A(t)/n_B(t)$ around $p/(1-p)$ after some finite time T, and show that what happened before time T does not have a noticeable effect in the long run.

Recall that we write $T_n^+ = \log n + K$ and $T_n^- = \log n - K$. Consider the process $\{X(t \wedge S) : t \geq T\}$ and note that for n large enough we have $T_n^- \geq T$. By the exponential martingale inequality from Liptser and Shiryayev [64, Section 4.13, Theorem 5] with choices

$$K = 2 \quad \text{and} \quad \varphi(t) = c\left(\frac{p}{1-p} + \varepsilon\right)t,$$

we have for any $\delta > 0$,

$$\mathbb{P}\left(\sup_{t\leq T_n^+\wedge S} X(t) \geq \delta c\left(\frac{p}{1-p} + \varepsilon\right)T_n^+\right) \leq \exp\left(-\kappa c\left(\frac{p}{1-p} + \varepsilon\right)T_n^+\right)$$

$$+ \mathbb{P}\left(\langle X\rangle(T_n^+ \wedge S) \geq 2\varphi(T_n^+)\right), \tag{2.25}$$

where $\kappa = (\delta + 2)\log\frac{\delta+2}{2} - \delta$. By definition of S and the expression of $\langle X \rangle(\cdot)$ in (2.23), we have that with probability one

$$\langle X \rangle(T_n^+ \wedge S) \leq \varphi(T_n^+),$$

so the second term on the right hand side of (2.25) vanishes. Further, using the expression of $X(\cdot)$ in (2.22) we get

$$\mathbb{P}\left(n_v[T, T_n^+ \wedge S] \geq (\delta + 1)c\left(\frac{p}{1-p} + \varepsilon\right)T_n^+\right) \leq \exp\left(-\kappa c\left(\frac{p}{1-p} + \varepsilon\right)T_n^+\right). \qquad (2.26)$$

For $a < b$, denote by $M_B[a, b]$ the maximal number of offsprings produced by a type B individual in time interval $[a, b]$ and let $M_B(t) = M_B[0, t]$. Then using a union bound and note that $M_B[T, T_n \wedge S] \leq M_B[T, T_n^+ \wedge S]$ with probability at least $1 - \varepsilon$, we have

$$\mathbb{P}\left(M_B[T, T_n \wedge S] \geq (\delta + 1)c\left(\frac{p}{1-p} + \varepsilon\right)T_n^+\right) \leq \varepsilon + n\exp\left(-\kappa c\left(\frac{p}{1-p} + \varepsilon\right)T_n^+\right).$$

Further, from our choice of T in (2.24) it follows that

$$\mathbb{P}\left(M_B[T, T_n] \geq (\delta + 1)c\left(\frac{p}{1-p} + \varepsilon\right)T_n^+\right) \leq 2\varepsilon + n\exp\left(-\kappa c\left(\frac{p}{1-p} + \varepsilon\right)T_n^+\right).$$

Next, note that there exists $L > 0$ such that

$$\mathbb{P}\left(M_B(T) \geq L\right) < \varepsilon.$$

As $M_B(T_n) \leq M_B(T) + M_B[T, T_n]$, combining the above results readily yields

$$\mathbb{P}\left(M_B(T_n) \geq L + (\delta + 1)c\left(\frac{p}{1-p} + \varepsilon\right)T_n^+\right) \leq 3\varepsilon + n\exp\left(-\kappa c\left(\frac{p}{1-p} + \varepsilon\right)T_n^+\right). \qquad (2.27)$$

Denote by $M_{n,B}$ the maximal degree of type B vertices in $\mathcal{T}_n$. Letting $\varepsilon \to 0$ in (2.27) and note $M_{n,B} \overset{d}{=} M_B(T_n)$, we have

$$\mathbb{P}\left(M_{n,B} \geq L + (\delta + 1)\frac{cp}{1-p}T_n^+\right) \leq n\exp\left(-\frac{\kappa cp}{1-p}T_n^+\right). \qquad (2.28)$$

Recall that $\kappa = (\delta + 2)\log\frac{\delta+2}{2} - \delta$. So given any $\widetilde{\varepsilon} > 0$, for $\delta > \max\{2e^2 - 2, \frac{(2+\widetilde{\varepsilon})(1-p)}{cp} - 4\}$ we have $\frac{\kappa cp}{1-p}T_n^+ > (2 + \widetilde{\varepsilon})\log n$ for large enough n, which makes the right hand side of (2.28)

39

summable. By Borel-Cantelli lemma this implies

$$\mathbb{P}\left(\limsup_{n\to\infty}\frac{M_{n,B}}{\log n}\leq(\delta+1)\frac{cp}{1-p}+\widetilde{\varepsilon}\right)=1.$$

Letting $\widetilde{\varepsilon}\to 0$ we get

$$\mathbb{P}\left(\limsup_{n\to\infty}\frac{M_{n,B}}{\log n}\leq C_2\right)=1$$

where $C_2=\max\{\frac{(2e^2-1)cp}{1-p},2-\frac{3cp}{1-p}\}$. This completes the proof for the upper bound part.

This left us with the cases where $p=1$ or $q=1$. When $q=1$, as noted in Section 2.3.2, $\{\mathcal{T}_n\}_{n\geq 2}$ looks like two disjoint URTs connected by a single edge between roots. By strong law of large numbers, with probability one these two subtrees have sizes proportional to p and $1-p$ asymptotically. Therefore, appealing to existing results on maximal degree of URT [35] we have

$$\frac{M_n}{\log n}\xrightarrow{\text{a.s.}}\frac{1}{\log 2}.$$

Last we consider the special case where $p=1$ but $q\neq 1$. We shall use the same notations as defined before. First, note that what we proved for type A individuals still holds. Specifically, recall that

$$\mathbb{P}\left(\limsup_{n\to\infty}\frac{M_{n,A}}{\log n}\leq\frac{1}{\log 2}\right)=1.$$

However, we can no longer define T by (2.24). In fact, consider the unique type B vertex (*i.e.* the type B root) in $\mathcal{T}_n$ and note that all type A vertices other than the root have a fixed probability $1-q$ to connect to the type B root independently. By strong law of large numbers applied to binomial distribution we have

$$\frac{M_{n,B}}{n}\xrightarrow{\text{a.s.}}1-q,\quad\text{and therefore}\quad\frac{M_n}{n}\xrightarrow{\text{a.s.}}1-q.$$

This completes the proof for Theorem 2.5.

$\blacksquare$

2.3.9 Proof for height

In this section we shall prove Theorem 2.6. Instead of proving the result from scratch using continuous time martingales as what we did for maximal degree, we present here a proof that appeals to existing results on first birth problem of branching processes.

Once again, we consider first cases where $p, q \neq 1$. The basic idea here is still the same, *i.e.* to bound $n_A(t)/n_B(t)$ around $p/(1-p)$ after some finite time and prove that what happened in the beginning does not really matter in the long term.

To obtain strong convergence, it is enough to prove that for any $\delta, \varepsilon > 0$,

$$\limsup_{n \to \infty} \mathbb{P}\left(\sup_{k \geq n} |\frac{H_k}{\log k} - e| > \delta\right) < \varepsilon \tag{2.29}$$

Define the event

$$E_1 = \left\{ \sup_{t \geq T} \left| \frac{n_A(t)}{n_B(t)} - \frac{p}{1-p} \right| < \eta \frac{p}{1-p} \right\}. \tag{2.30}$$

where $0 < \eta < 1$ is any given constant. By Lemma 2.8 there exists $T > 0$ such that $\mathbb{P}(E_1) > 1 - \varepsilon/3$. Also, choose $N \in \mathbb{Z}^+$ such that $\mathbb{P}(n(T) > N) < \varepsilon/3$, and define another event $E_2 = \{n(T) \leq N\}$. Moreover, choose T_n^+ and T_n^- in the same way as in Section 2.3.8, with ε replaced by $\varepsilon/3$, and define our last "good" event $E_3 = \{T_n^- < T_n < T_n^+ \text{ for all } n \in \mathbb{Z}^+\}$. In what follows, we condition on the event $E_1 \cap E_2 \cap E_3$ and note that $\mathbb{P}(E_1 \cap E_2 \cap E_3) > 1 - \varepsilon$ (for simplicity the conditional is suppressed throughout).

On this event, we have at most N individuals alive at time T and the ratio of $n_A(t)$ to $n_B(t)$ is bounded around $p/(1-p)$ after that time. For a fixed individual v alive at time T and $t > 0$, denote by $H_v(t)$ the height of the subtree root at v in pdBP$(T + t)$. To bound $H_v(t)$, we now construct two processes. Recall our continuous time process as defined in Section 2.3.3, and consider a process where each type B individual gives birth to type A individuals at rate $(1 - \eta)p(1 - q)/(1 - p)$ and type B individuals at rate $(1 - \eta)q$, while everything else stays the same. Denote by $H_{min}(t)$ the height of this tree at time t, $B_{min}(n)$ the time when the first individual in the n-th generation is born, and define $H_{max}(t)$ similarly using $1 + \eta$ instead of $1 - \eta$ in the rates. Since the reproduction rates of type A individuals are constants, and those of type B individuals only depend on $n_A(t)/n_B(t)$, we have $H_{min}(t) \preceq H_v(t) \preceq H_{max}(t)$ where $\preceq$ denotes stochastic dominance as processes (*i.e.* for any monotone increasing functional f we have $f(H_{min}(t)) \preceq f(H_v(t)) \preceq f(H_{max}(t))$ where $\preceq$ denotes the usual stochastic dominance).

From Biggins [17, Theorem 2] we know that

$$\lim_{n \to \infty} \frac{B_{min}(n)}{n} = \gamma_{min},$$

where γ_{min} can be calculated following the procedure given in the paper. First, compute the matrix $\Phi(\theta)$ with entries

$$\Phi_{ij}(\theta) = \theta \int_0^\infty e^{-\theta t} \mathbb{E}(Z_{ij}(t)) dt = \frac{r_{ij}}{\theta}.$$

Here $Z_{ij}(t)$ denotes the number of type j individuals born to a type i individual by time t, and r_{ij} denotes the rate at which a type i individual gives birth to type j individuals. Then take the largest eigenvalue $\phi(\theta) = \lambda_{min}/\theta$ of $\Phi(\theta)$. In our case we have

$$\lambda_{min} = \frac{(2 - \eta)q + \sqrt{\eta^2 q^2 + 4(1 - \eta)(1 - q)^2}}{2}$$

by calculation. Finally, define

$$\mu(a) = \inf \left\{ e^{\theta a} \phi(\theta) : \theta > 0 \right\},$$

and compute $\gamma_{min} = \inf \left\{ a : \mu(a) \geq 1 \right\} = 1/(\lambda_{min} e)$.

Since

$$B_{min}(H_{min}(t)) \leq t \leq B_{min}(H_{min}(t) + 1),$$

dividing by $H_{min}(t)$ and letting $t \to \infty$ we get

$$\frac{H_{min}(t)}{t} \xrightarrow{\text{a.s.}} \lambda_{min} e. \tag{2.31}$$

Similarly we have

$$\frac{H_{max}(t)}{t} \xrightarrow{\text{a.s.}} \lambda_{max} e, \tag{2.32}$$

where

$$\lambda_{max} = \frac{(2 + \eta)q + \sqrt{\eta^2 q^2 + 4(1 + \eta)(1 - q)^2}}{2}.$$

As the eigenvalues are continuous with respect to η, and $\lambda_{max} = \lambda_{min} = 1$ when $\eta = 0$, we can choose η in (2.30) small enough such that both $\lambda_{max} - 1$ and $1 - \lambda_{min}$ are smaller than $\delta/3$.

Remark 2.6. Since we will need to generalize the result to CMRT with more types, we include here an alternative argument using Perron-Frobenius theory of positive matrices [73, Chapter 8]. Note that when $\eta = 0$, the rate matrix consisting of r_{ij}'s is a positive matrix (*i.e.* a matrix

where all entries are strictly positive)

$$\begin{pmatrix} q & \frac{(1-p)(1-q)}{p} \\ \frac{p(1-q)}{1-p} & q \end{pmatrix}$$

with left eigenvector $(p, 1 - p)$ corresponding to eigenvalue $\lambda = 1$. Since this eigenvector has strictly positive coordinates, we have that $\lambda = 1$ is the unique largest eigenvalue of that rate matrix. Therefore the largest eigenvalue of rate matrix is continuous with respect to η around $\eta = 0$, and the previous result we established on λ_{max} and λ_{min} follows.

By (2.31), (2.32) and stochastic dominance (as processes) we see that

$$\lim_{s \to \infty} \mathbb{P} \left(\inf_{t \geq s} \frac{H_v(t)}{t} < \lambda_{min} e - \frac{\delta}{3} \right) = 0 \quad \text{and} \quad \lim_{s \to \infty} \mathbb{P} \left(\sup_{t \geq s} \frac{H_v(t)}{t} > \lambda_{max} e + \frac{\delta}{3} \right) = 0$$

hold for all $v \in V$, where V denotes the set of individuals alive at time T.

Since $H_n \geq H_v(T_n^- - T)$, we have

$$\limsup_{n \to \infty} \mathbb{P} \left(\inf_{k \geq n} \frac{H_k}{T_k^- - T} < \lambda_{min} e - \frac{\delta}{3} \right) \leq \lim_{s \to \infty} \mathbb{P} \left(\inf_{t \geq s} \frac{H_v(t)}{t} < \lambda_{min} e - \frac{\delta}{3} \right) = 0.$$

On the other hand, note that $H_n \leq \max_{u \in V} H_u(T_n^+ - T) + N$. Using union bound with $|V| \leq N$ we get

$$\limsup_{n \to \infty} \mathbb{P} \left(\sup_{k \geq n} \frac{H_k - N}{T_k^+ - T} > \lambda_{max} e + \frac{\delta}{3} \right) \leq \sum_{u \in V} \lim_{s \to \infty} \mathbb{P} \left(\sup_{t \geq s} \frac{H_u(t)}{t} > \lambda_{max} e + \frac{\delta}{3} \right) = 0.$$

With

$$\left| \frac{H_k}{T_k^- - T} - \frac{H_k}{\log k} \right| < \frac{\delta}{3} \quad \text{and} \quad \left| \frac{H_k - N}{T_k^+ - T} - \frac{H_k}{\log k} \right| < \frac{\delta}{3}$$

for large enough k, by our choice of η and triangle inequality it follows that

$$\limsup_{n \to \infty} \mathbb{P} \left(\sup_{k \geq n} \left| \frac{H_k}{\log k} - e \right| > \delta \right) = 0. \tag{2.33}$$

Finally, do not forget that we are conditioning on the event $E_1 \cap E_2 \cap E_3$, which occurs with probability at least $1 - \varepsilon$. Therefore we have (2.29) as desired.

Once again we are left with cases where $p = 1$ or $q = 1$. When $q = 1$, recall that $\{\mathcal{T}_n\}_{n \geq 2}$ looks like two disjoint URTs in this case. Using results for URT [88] we have

$$\frac{H_n}{\log n} \xrightarrow{\text{a.s.}} e.$$

Last we turn to cases where $p = 1$ but $q \neq 1$. Consider only pure-blooded type A individuals as in Section 2.3.8 and use the same argument there we have

$$\mathbb{P}\left(\liminf_{n \to \infty} \frac{H_n}{\log n} \geq qe\right) = 1.$$

For upper bound we construct $\{\mathcal{T}_n\}_{n \geq 2}$ from a URT $\{\mathcal{U}_n\}_{n \geq 1}$ as described in special cases of Section 2.3.2 and note that by construction the height of $\mathcal{T}_n$ is at most equal to that of $\mathcal{U}_n$. From results for URT [88] we know that

$$\mathbb{P}\left(\limsup_{n \to \infty} \frac{H_n}{\log n} \leq e\right) = 1.$$

This completes the proof for Theorem 2.6.

■

2.3.10 Extension to general case

To extend our results to more types, first we need to define Community Modulated Recursive Tree (CMRT) with $K > 2$ types precisely. As in CMRT with two types, we start with one vertex of each type to make sure new vertices always have a vertex to connect to. When $K = 2$, these two vertices must be adjacent. However, when $K > 2$ one has to specify the starting tree.

The choice we make is as follows. To construct a CMRT with $K > 1$ types:

- Take a URT of size K and randomly assign each vertex to one of the K types based on uniform random permutation. We shall refer to these vertices as roots of each type.

- Given a CMRT of size $n - 1$ $(n \geq K + 1)$, a new vertex n is added to the tree at the next time step and assigned type i $(1 \leq i \leq K)$ with probability $p_i \in [0, 1]$. W.L.O.G. we assume that $p_1 = \max_{1 \leq i \leq K} p_i > 0$.

- Vertex n then chooses type j $(1 \leq i \leq K)$ with probability $q_{ij} \in [0, 1]$.

- Finally vertex n chooses uniformly at random an existing vertex of that chosen type, and connect to it, forming a CMRT of size n.

The process then uses the above dynamics recursively to yield a growing tree-valued process, which we denote by $\{\mathcal{T}_n\}_{n \geq K}$, where $\mathcal{T}_n$ is the random recursive tree given by the process as it reaches size n. Then we call $\{\mathcal{T}_n\}_{n \geq K}$ a Community Modulated Recursive Tree (CMRT) with K types and $\mathcal{T}_n$ a CMRT of size n.

Accordingly, we will need a population-dependent branching process with K types as follows:

Initialization: start with K individuals at $t = 0$, with one of each type. For any time $t \geq 0$ and $1 \leq i \leq K$, let $n_i(t)$ be the number of type i individuals. We have $n_i(0) = 1$. Denote by $\mathcal{F}(t)$ the σ-field generated by the process until time t.

Types: Each individual in the system has a type $\in \{1, 2, ..., K\}$ and lives forever, while giving birth to individuals of all types.

Reproduction: At any time t, a living type i individual gives birth to type j $(1 \leq j \leq K)$ individuals at rate:

$$r_{ij}(t) = \frac{n_1(t)}{n_i(t)} \cdot \frac{p_j q_{ji}}{p_1}.$$

Then using the same arguments as in the proofs for CMRT with two types we can prove similar results in the general case:

Theorem 2.15. *For each fixed k, let $N_k(n)$ denote the number of vertices with out-degree k in $\mathcal{T}_n$. Then*

$$\frac{N_k(n)}{n} \xrightarrow{P} c_k$$

where

$$c_k := \sum_{i=1}^{K} \frac{p_i}{1 + r_i} \left(\frac{r_i}{1 + r_i}\right)^k.$$

Here

$$r_i = \frac{1}{p_i} \sum_{j=1}^{K} p_j q_{ji}.$$

When $p_i = 0$ the i-th term in c_k should be interpreted as 0.

Remark 2.7. Similar to CMRT with two types, the limiting degree distribution is a mixture of K shifted geometric distributions, and coincides with that of URT if and only if the balance equation $\sum_{j=1}^{K} p_j q_{ji} = p_i$ holds for all $1 \leq i \leq K$. Note that $\sum_{j=1}^{K} p_j q_{ji}$ is the probability for the

parent of a new vertex to be of type i. So the balance equation essentially states that the type distribution for parents is identical to that of children. For the continuous time embedding, the condition implies that all individuals will reproduce at approximately the same rate once the population stabilizes.

Theorem 2.16. *Let M_n denote the maximal degree in $\mathcal{T}_n$. Then when $p_1 < 1$ and $q_{ii} < 1$ for some $1 \leq i \leq K$, there exists constants C_1 and C_2 that depend only on p_i and q_{ij} such that*

$$\liminf_{n \to \infty} \frac{M_n}{\log n} \geq C_1 \quad and \quad \limsup_{n \to \infty} \frac{M_n}{\log n} \leq C_2 \quad a.s.$$

When $q_{ii} = 1$ for all $1 \leq i \leq K$,

$$\frac{M_n}{\log n} \xrightarrow{a.s.} \frac{1}{\log 2}$$

and when $p_1 = 1$ but $q_{11} < 1$,

$$\frac{M_n}{n} \xrightarrow{a.s.} \max_{2 \leq i \leq K} q_{1i}.$$

Theorem 2.17. *Let H_n denote the height of $\mathcal{T}_n$. Then when $p_1 < 1$ or $q_{11} = 1$,*

$$\frac{H_n}{\log n} \xrightarrow{a.s.} e.$$

When $p_1 = 1$ and $q_{11} < 1$,

$$\liminf_{n \to \infty} \frac{H_n}{\log n} \geq q_{11} e \quad and \quad \limsup_{n \to \infty} \frac{H_n}{\log n} \leq e \quad a.s.$$

2.4 Structure of subtrees

As mentioned in Section 2.1.5, subtree rooted at a fixed vertex in URT has conditional distribution identical to URT of the same size. This is obvious from the continuous embedding since rates are constant. For CMRT however, similar result does not hold. Still, one can ask if subtrees in CMRT retain some properties of CMRT. In this section, we shall consider CMRT with two types and prove that the ratio between type A and B vertices is retained asymptotically in subtrees. We will also relate the convergence of branch sizes to root finding algorithms.

2.4.1 Ratios between population size of different types

For a fixed positive integer k, let $S_k(n)$ be size of the subtree rooted at k in $\mathcal{T}_n$, and $S_k^A(n)$ and $S_k^B(n)$ be the number of type A and B vertices in this subtree. As the type sequence is i.i.d. Bernoulli random variables, we have that ratio between population size of different types in $\mathcal{T}_n$ converges a.s. to $(1-p)/p$. We shall prove that this holds for all subtrees as well:

Theorem 2.18. *For any fixed positive integer k,*

$$\frac{S_k^B(n)}{S_k^A(n)} \xrightarrow{\text{a.s.}} \frac{1-p}{p}.$$

Proof: Once again we shall work with the continuous embedding in Section 2.3.3.

Denote by $s_k^A(t)$ and $s_k^B(t)$ the number of type A and B descendants of individual k in pdBP(t), and define $Y(t) := (1-p)s_k^A(t) - ps_k^B(t)$ and $Z(t) := (1-p)n_A(t) - pn_B(t)$.

Denote $f(t) := \mathbb{E}\, Y^2(t)$, by calculations similar to those in Lemma 3.2 we can compute $f'(t)$. Then take its absolute value and we get:

$$
\begin{aligned}
f'(t) \leq & 2|2q-1|f(t) + 2|2q-1|\,\mathbb{E}\,|Y(t)Z(t)| + |2pq-p-q|\,\mathbb{E}\,|Z(t)| \\
& + |1-p-q+2pq|\,\mathbb{E}\,|Z(t)s_k^B(t)| + p\,\mathbb{E}\,s_k^B(t) \\
\leq & 2|2q-1|f(t) + 2|2q-1|\sqrt{\mathbb{E}\,|Y(t)|^2\,\mathbb{E}\,|Z(t)|^2} + |2pq-p-q|\,\mathbb{E}\,|Z(t)| \\
& + |1-p-q+2pq|\sqrt{\mathbb{E}\,|Z(t)|^2\,\mathbb{E}\,|n_B(t)|^2} + p\,\mathbb{E}\,n_B(t) \\
\leq & 2|2q-1|f(t) + C_1 e^{t/2}\sqrt{f(t)} + C_2 e^t
\end{aligned}
$$

for some positive constant C_1 and C_2. Here we use results obtained in Section 2.3.5 on moments of $n_B(t)$ and $Z(t)$.

Since $f(t)$ is strictly positive, we can define $g(t) = \sqrt{f(t)}$ and derive:

$$g'(t) \leq |2q-1|g(t) + C_3 e^{t/2} + \frac{C_4 e^t}{g(t)} \tag{2.34}$$

for $C_3 = C_2/2$ and $C_4 = C_3/2$. W.L.O.G suppose $C_3, C_4 > 1$. We want to show that $g(t) \leq C(\exp(t/2) + t\exp(\lambda t)), \ \forall t \geq 0$ for some $C > 0$, where $\lambda = |2q-1|$.

For given $t \geq 0$, note that $g(t)$ is deterministic and let

$$s = \sup\{0 \leq t^* \leq t : g(t^*) \leq e^{t^*/2}\}.$$

If $s = t$ then letting $C = 1$ we have $g(t) \le C \exp(t/2)$ trivially. Otherwise we have $s < t$ and by continuity $g(s) = \exp(s/2)$. From (2.34) we get

$$g(t) - g(s) = g(t) - e^{s/2} \le \int_s^t \lambda g(u) du + (C_3 + C_4) \int_s^t e^{u/2} du,$$

and

$$g(t) \le \int_s^t \lambda g(u) du + C_5 e^{t/2}$$

follows, where $C_5 = \max\{1, (C_3 + C_4)/2\}$.

By Gronwall's inequality,

$$g(t) \le C_5 e^{t/2} + \lambda C_5 \int_s^t e^{u/2} e^{\lambda(t-u)} du.$$

If $\lambda \ge 1/2$ we have

$$g(t) \le C_5 e^{t/2} + \lambda C_5 t e^{\lambda t},$$

and if $\lambda < 1/2$ we have

$$g(t) \le \frac{C_5}{1 - 2\lambda} e^{t/2}.$$

Letting $C = \max\{C_5, C_5/(1 - 2\lambda)\}$ we have $g(t) \le C(\exp(t/2) + t \exp(\lambda t))$.

Thus we have $\mathbb{E} Y^2(t) \le 2C^2(\exp(t) + t^2 \exp(2\lambda t))$. Note that C_i $(i = 1, 2, ..., 5)$ do not depend on p and q, so we can take C as a constant independent of p and q as well. For large enough $K > 0$ using Markov inequality and Borel-Cantelli lemma just as in Lemma 2.8 we get $\exp(-K \log n) Y(K \log n) \xrightarrow{\text{a.s.}} 0$. Note that $K \log n \le t < K \log(n+1)$,

$$
|e^{-t} Y(t)| \le \max\Big\{ n^{-K} |(1-p) s_k^A(K \log(n+1)) - p s_k^B(K \log n)|,
$$
$$
n^{-K} |(1-p) s_k^A(K \log n) - p s_k^B(K \log(n+1))| \Big\}. \tag{2.35}
$$

For the first term on the right hand side we have

$$
|(1-p) s_k^A(K \log(n+1)) - p s_k^B(K \log n)| \le |Y(K \log(n+1))| + p|s_k^B(K \log(n+1)) - s_k^B(K \log n)|
$$
$$
\le |Y(K \log(n+1))| + p|n_B(K \log(n+1)) - n_B(K \log n)|.
$$

and second term can be bounded similarly to yield

$$|e^{-t}Y(t)| \leq n^{-K}(p|n_B(K\log(n+1)) - n_B(K\log n)| + \max\{|Y(K\log n)|, Y(K\log(n+1))|\}) \xrightarrow{\text{a.s.}} 0.$$

Here we use the facts that $(\frac{n}{n+1})^K \to 1$ and $\{n^{-K}n_B(K\log n)\}$ is a.s. Cauchy by the previously established a.s. convergence of $\exp(-t)n_B(t)$. It follows that $e^{-t}Y(t) \xrightarrow{\text{a.s.}} 0$ as $t \to \infty$.

Since $|s_k^A(t_1) - s_k^A(t_2)| \leq |n_A(t_1) - n_A(t_2)|$ and $e^{-t}n_A(t) \xrightarrow{\text{a.s.}} W$, $\{e^{-t}s_k^A(t)\}$ must be a Cauchy sequence a.s. and $e^{-t}s_k^A(t) \xrightarrow{\text{a.s.}} V$ for some nonnegative r.v. V. Note that $W \sim Exp(1)$ is positive a.s. so we also have $s_k^A(t)/n_A(t) \xrightarrow{\text{a.s.}} U$ for some nonnegative r.v. U. Then $e^{-t}Y(t) = e^{-t}\big((1-p)s_k^A(t) - ps_k^B(t)\big) \xrightarrow{\text{a.s.}} 0$ as $t \to \infty$ readily gives $s_k^B(t)/s_k^A(t) \xrightarrow{\text{a.s.}} (1-p)/p$.

■

2.4.2   Convergence of subtree sizes

We observe that types of vertice $1, 2, ..., k$ are not specified in Section 2.4.1. Therefore, the results there actually hold more generally. To state this precisely, we need to define some terms. For $K \geq 2$ and $i \leq K$, denote by $\mathcal{T}_n^{(i,K)}$ the tree containing vertex i after removing all edges between vertices $\{1, 2, ..., K\}$. Then let $N_A(n)$ (or $N_A^{(i,K)}(n)$) and $N_B(n)$ (or $N_B^{(i,K)}(n)$) be the number of type A and B vertices in $\mathcal{T}_n$ (or $\mathcal{T}_n^{(i,K)}$). For fixed K and a partition $X \cup Y$ of $\{1, 2, ..., K\}$, let a_X (or a_Y) and b_X (or b_Y) be the number of type A and type B vertices in X (or Y). Define

$$N_A^X(n) := \sum_{i \in X} N_A^{(i,K)}(n), \quad N_B^X(n) := \sum_{i \in X} N_B^{(i,K)}(n).$$

Then conditioning on (a_X, b_X, a_Y, b_Y), it can be proved using the same argument as in Section 2.4.1 that $N_A^X(n)/N_A(n)$ and $N_B^X(n)/N_B(n)$ converge a.s. to the same nonnegative r.v. U. It follows that $(N_A^X(n) + N_B^X(n))/n$ also converges a.s. to U. Here the law of U depends only on (a_X, b_X, a_Y, b_Y), so we shall use $U(a_X, b_X, a_Y, b_Y)$ to denote the law of the limiting r.v..

It is not hard to see by monotonicity that the following lemma holds:

Lemma 2.19. *If $x \leq x'$, $y \leq y'$, $z \geq z'$, $w \geq w'$, then $U(x, y, z, w) \preceq U(x', y', z', w')$, where $\preceq$ denotes stochastic dominance.*

Let $b_X = b_Y = 0$ and $p = q = 1$, which cannot happen in CMRT because the type B root is always present. Still we can see that $U(a_X, 0, a_Y, 0)$ is well defined in this case and has a

Beta(a_X, a_Y) distribution. Therefore, it seems natural to compare $U(x, x, y, y)$ with Beta(x, y). Indeed, we hypothesized the following moment bounds:

Hypothesis 2.20. *For any positive integer k, we have*

$$\mathbb{E}\, U^k(x, x, y, y) \leq \mathbb{E}\, V^k$$

where $V \sim Beta(x, y)$.

Much work has been done in an attempt to prove this hypothesis, and we include a discussion in Appendix A. For the rest of this section, we will show that this hypothesis, if it were indeed true, has important implications for understanding branching structure of CMRT.

First we examine subtrees rooted at the two roots, *i.e.* $\mathcal{T}_n^{(1,2)}$ and $\mathcal{T}_n^{(2,2)}$. Since we know that both $|\mathcal{T}_n^{(1,2)}|/n$ and $|\mathcal{T}_n^{(2,2)}|/n$ converges a.s. to some nonnegative r.v., we can expect to obtain tail bounds on the limiting distribution.

Let τ be the stopping time n when the first type A vertex is added to $\mathcal{T}_n^{(2,2)}$. Then conditioning on $\mathcal{T}_\tau$ we have that $|\mathcal{T}_n^{(1,2)}|/n$ converges a.s. to r.v. $U(N_A^{(1,2)}(\tau), N_B^{(1,2)}(\tau), N_A^{(2,2)}(\tau), N_B^{(2,2)}(\tau))$. Note that $N_A^{(1,2)}(\tau), N_B^{(1,2)}(\tau) \leq \tau - 2$, $N_A^{(2,2)}(\tau) = 1$ and $N_B^{(2,2)}(\tau) \geq 1$, so the r.v. above is stochastically dominated by $U(\tau - 2, \tau - 2, 1, 1)$ using Corollary 2.19. Therefore, conditioning on τ we have by Markov inequality and Hypothesis 2.20 that

$$\limsup_{n \to \infty}\ \mathbb{P}(|\mathcal{T}_n^{(1,2)}| \geq (1 - \varepsilon)n) \leq \frac{\mathbb{E}\, U^k(\tau - 2, \tau - 2, 1, 1)}{(1 - \varepsilon)^k} \leq \frac{\tau - 2}{(\tau - 2 + k)(1 - \varepsilon)^k}$$

for all positive integer k.

Let $k = \lceil 1/\varepsilon \rceil - \tau + 2$ we get

$$\limsup_{n \to \infty}\ \mathbb{P}(|\mathcal{T}_n^{(1,2)}| \geq (1 - \varepsilon)n) \leq (\tau - 2)\varepsilon(1 - \varepsilon)^{\tau - 1 - 1/\varepsilon} \leq 4(\tau - 2)\varepsilon$$

for $\varepsilon < 1/2$.

On the other hand, let $C_1 = p(1 - q)$ and note the following fact. At time $t + 1$, a type A vertex is added and connect to a type B vertex with probability C_1, and there are at most $t - 1$ type B vertices to choose from. Since the subtree rooted at vertex 2 has at least one type B

vertex (*i.e.* the root itself), we have

$$\mathbb{P}(\tau > t + 1) \le (1 - \frac{C_1}{t - 1}) \, \mathbb{P}(\tau > t)$$

It follows that

$$\mathbb{P}(\tau > t) \le \prod_{i=1}^{t-2}(1 - \frac{C_1}{i}) = \frac{\Gamma(t - 1 - C_1)}{\Gamma(t - 1)\Gamma(1 - C_1)}.$$

Then using Stirling's approximation we know there exists a universal constant $C_2 > 0$ such that

$$\mathbb{P}(\tau > t) \le \frac{C_2}{(t - 1 - C_1)^{C_1}} \le \frac{C_2}{(t - 2)^{C_1}}.$$

Combine the above results we have (unconditionally)

$$\limsup_{n \to \infty} \; \mathbb{P}(|\mathcal{T}_n^{(1,2)}| \ge (1 - \varepsilon)n) \le 4(t - 2)\varepsilon + \frac{C_2}{(t - 2)^{C_1}}$$

for all $t \le 2 + \lceil 1/\varepsilon \rceil$. Let $t = \lceil 1/\varepsilon^{1/(1+C_1)} \rceil$ we get

$$\limsup_{n \to \infty} \; \mathbb{P}(|\mathcal{T}_n^{(1,2)}| \ge (1 - \varepsilon)n) \le C_3 \varepsilon^{\frac{C_1}{1 + C_1}}$$

for some universal constant $C_3 > 0$.

Similarly we can establish bounds for $|\mathcal{T}_n^{(2,2)}|$. Then from

$$\limsup_{n \to \infty} \; \mathbb{P}\left\{|\mathcal{T}_n^{(1,2)}| \ge (1 - \varepsilon)n \text{ or } |\mathcal{T}_n^{(1,2)}| \le \varepsilon n\right\} = \limsup_{n \to \infty} \; \mathbb{P}(|\mathcal{T}_n^{(1,2)}| \ge (1 - \varepsilon)n)$$
$$+ \limsup_{n \to \infty} \; \mathbb{P}(|\mathcal{T}_n^{(2,2)}| \ge (1 - \varepsilon)n),$$

we get

$$\limsup_{n \to \infty} \; \mathbb{P}\left\{|\mathcal{T}_n^{(1,2)}|/n \notin (\varepsilon, (1 - \varepsilon))\right\} \le C^* \varepsilon^{\frac{(1-q)\omega}{1 + (1-q)\omega}} \tag{2.36}$$

where $\omega = \min(p, 1 - p)$ and $C^* > 0$ is some universal constant.

Next we consider general subtrees $\mathcal{T}_n^{(k,K)}$ for $1 \le k \le K$. Intuitively we would expect that these subtrees should be small when K is large. We shall see that this is indeed true assuming the moment bounds given in Hypothesis 2.20.

Using Hoeffding's inequality for binomial r.v., we know that with probability at least $1 - 2/k^2$, $\min(N_A(k), N_B(k)) \ge K\omega - \sqrt{K \log K}$. Denote $K\omega - \sqrt{K \log K}$ by M. Then for any

$1 \leq k \leq K$, we have that $|\mathcal{T}_n^{(k,K)}|/n$ converges a.s. to a r.v. stochastically dominated by $U(1,1,M,M)$. By Markov inequality and Hypothesis 2.20 we see that

$$\limsup_{n\to\infty} \mathbb{P}\left\{|\mathcal{T}_n^{(k,K)}| \geq \varepsilon n\right\} \leq \frac{\mathbb{E}\, U^2(1,1,M,M)}{\varepsilon^2} \leq \frac{2}{M(M+1)\varepsilon^2}.$$

Using union bound we get

$$\limsup_{n\to\infty} \mathbb{P}\left\{\exists 1 \leq k \leq K : |\mathcal{T}_n^{(k,K)}| \geq \varepsilon n\right\} \leq \frac{2K}{M(M+1)\varepsilon^2} \sim \frac{2}{K\omega^2\varepsilon^2} \qquad (2.37)$$

where the approximation holds for large K. This shows that all subtrees $\mathcal{T}_n^{(k,K)}$ are relatively small when K is large. Note that we can always take higher moments in Markov inequality and get better bounds asymptotically (with respect to ε), but the coefficient will increase accordingly.

2.4.3 Root-finding algorithm

As shown by Bubeck et.al. [21], convergence of subtree sizes in URT has important implication for root-finding algorithms. Following their notations, given a (unrooted) tree T and one of its vertex v, let $\psi(v)$ be size of the largest subtree rooted at a child of v when regarding v as the root of T. Then given an integer K, let $H(K)$ be the set of K vertices with the smallest values of $\psi(v)$ (with ties broken arbitrarily). Bubeck et.al. [21] showed that $H(K)$ gives a root-finding algorithm for URT. In this section, we shall apply results in Section 2.4.2 to show that $H(K)$ also gives a root-finding algorithm for CMRT, assuming Hypothesis 2.20.

First observe directly from (2.36) that

$$\limsup_{n\to\infty} \mathbb{P}\left\{\psi(1) \geq (1-\varepsilon)n \text{ or } \psi(2) \geq (1-\varepsilon)n\right\} \leq \limsup_{n\to\infty} \mathbb{P}\left\{|\mathcal{T}_n^{(1,2)}|/n \notin (\varepsilon, (1-\varepsilon))\right\}$$
$$\leq C^* \varepsilon^{\frac{(1-q)\omega}{1+(1-q)\omega}}$$

where $\omega = \min(p, 1-p)$ and $C^* > 0$ is some universal constant.

Next we proceed to bound $\psi(i)$ for large i. Fix a large K and assume $i \geq K$. We shall use the following inequality from Bubeck et.al. [21]:

$$\psi(i) \geq \min_{1\leq k\leq K} \sum_{j=1, j\neq k}^{K} |\mathcal{T}_n^{(j,K)}| = n - \max_{1\leq k\leq K} |\mathcal{T}_n^{(k,K)}|.$$

Then using (2.37) we have

$$\limsup_{n\to\infty} \mathbb{P}\left\{\exists i > K : \psi(i) < (1-\varepsilon)n\right\} \leq \limsup_{n\to\infty} \mathbb{P}\left\{\exists 1 \leq k \leq K : |\mathcal{T}_n^{(k,K)}| \geq \varepsilon n\right\}$$

$$\leq \frac{2K}{M(M+1)\varepsilon^2} \sim \frac{2}{K\omega^2\varepsilon^2}$$

Finally, combine these results we get

$$1 - \liminf_{n\to\infty} \mathbb{P}(\{1,2\} \subseteq H(K)) \leq C^* \varepsilon^{\frac{(1-q)\omega}{1+(1-q)\omega}} + \frac{2K}{M(M+1)\varepsilon^2}.$$

Therefore if $K \geq 1/\varepsilon^3$, we have

$$\mathbb{P}(\{1,2\} \subseteq H(K)) = 1 - O(\varepsilon^{\frac{(1-q)\omega}{1+(1-q)\omega}})$$

as $\varepsilon \to 0$, where $\omega = \min(p, 1-p)$. This shows that $H(K)$ gives a confidence set for roots in CMRT.

2.5 Extension to other variants of the model

There are other variants of the model that could be of potential interest, and we shall focus on two of them in this section.

Construct a random recursive tree with $K > 1$ types as follows:

- Take a URT of size K and randomly assign each vertex to one of the K types based on uniform random permutation. We shall refer to these vertices as roots of each type.

- Given a tree of size $n-1$ ($n \geq K+1$), a new vertex n is added to the tree at the next time step and assigned type i ($1 \leq i \leq K$) with probability $p_i \in [0,1]$. W.L.O.G. we assume that $p_1 = \max_{1\leq i\leq K} p_i > 0$.

- Vertex n then connects to an existing type j ($1 \leq j \leq K$) vertex with probability proportional to $\omega_{ij} > 0$. Specifically, for an existing type j vertex v, the probability for vertex n to connect to v is given by

$$\frac{\omega_{ij}}{\sum_{l=1}^K n_l \omega_{il}},$$

where n_l is the number of existing type l ($1 \leq l \leq K$) vertices. We shall refer to these ω_{ij} as weights.

The process then uses this dynamics recursively to yield a growing tree-valued process, which we denote by $\left\{\tilde{\mathcal{T}}_n\right\}_{n \geq K}$, where $\tilde{\mathcal{T}}_n$ is the random recursive tree given by the process as it reaches size n. Then we call $\left\{\tilde{\mathcal{T}}_n\right\}_{n \geq K}$ a Community Weighted Recursive Tree (CWRT) with K types and $\tilde{\mathcal{T}}_n$ a CWRT of size n.

Similar to CMRT, we can define a population-dependent branching process with K types for CWRT as follows:

Initialization: start with K individuals at $t = 0$, with one of each type. For any time $t \geq 0$ and $1 \leq i \leq K$, let $n_i(t)$ be the number of type i individuals. We have $n_i(0) = 1$. Denote by $\mathcal{F}(t)$ the σ-field generated by the process until time t.

Types: Each individual in the system has a type $\in \{1, 2, ..., K\}$ and lives forever, while giving birth to individuals of all types.

Reproduction: At any time t, a living type i individual gives birth to type j $(1 \leq j \leq K)$ individuals at rate:
$$r_{ij}(t) = \frac{n_1(t)}{p_1} \cdot \frac{p_j \omega_{ji}}{\sum_{l=1}^{K} \omega_{jl} n_l(t)}.$$

Then following the same procedure as in the proof for CMRT we can derive the limiting degree distribution of CWRT:

Theorem 2.21. *For each fixed k, let $N_k(n)$ denote the number of vertices with out-degree k in $\tilde{\mathcal{T}}_n$. Then*
$$\frac{N_k(n)}{n} \xrightarrow{P} \tilde{c}_k$$

where
$$\tilde{c}_k := \sum_{i=1}^{K} \frac{p_i}{1 + \tilde{r}_i} \left(\frac{\tilde{r}_i}{1 + \tilde{r}_i}\right)^k.$$

Here
$$\tilde{r}_i = \sum_{j=1}^{K} \frac{p_j \omega_{ji}}{\sum_{l=1}^{K} p_l \omega_{jl}}.$$

Remark 2.8. Similar to CMRT, the limiting degree distribution of CWRT is a mixture of K geometric distributions, with a different set of parameters. As a special case, when $\omega_{ij} \equiv 1$, the limiting distribution becomes identical to that of URT.

Next we consider a "preferential attachment" version of CMRT which we shall refer to as Community Modulated Preferential Attachment (CMPA) model. To construct a CMPA tree

with K types, we follow the same rule as that of CMRT, except for the attachment step. Given that vertex n is of type i and chooses to connect to vertices of type j, instead of choosing uniformly among vertices of type j, vertex n now connects to a type j vertex v with probability proportional to $D_v + \alpha_{ij}$, where D_v is the out-degree of v, and α_{ij} are positive parameters. We denote by $\{\mathcal{T}_n^*\}_{n \geq K}$ a Community Modulated Preferential Attachment (CMPA) tree with K types and $\mathcal{T}_n^*$ a CMPA tree of size n.

Accordingly, we need to adjust the reproduction rates for our population-dependent branching process as follows:

Reproduction: For any time $t \geq 0$, $1 \leq i, j \leq K$ and individual v, let $n_{ij}(t)$ be the number of type j individuals born to type i individuals, and $d_v(t)$ be the number of offsprings born to individual v. Then at any time t, a living type i individual v gives birth to type j $(1 \leq j \leq K)$ individuals at rate:

$$r_v(t, j) = \frac{n_1(t)}{p_1} \cdot \frac{p_j q_{ji}(d_v(t) + \alpha_{ji})}{\alpha_{ji} n_i(t) + \sum_{l=1}^{K} n_{il}(t)}. \tag{2.38}$$

Note here that $\sum_{l=1}^{K} n_{il}(t)$ is exactly the total number of offsprings born to type i individuals.

Let $p_i^* = \sum_{j=1}^{K} p_j q_{ji}$ be the probability for the parent of a new vertex to be of type i. Following the same procedure as in the proof for CMRT we can derive the corresponding limiting degree distribution:

Theorem 2.22. *For each fixed k, let $N_k(n)$ denote the number of vertices with out-degree k in $\mathcal{T}_n^*$. Then*

$$\frac{N_k(n)}{n} \xrightarrow{P} c_k^*$$

where

$$c_k^* := \sum_{i=1}^{K} \frac{p_i}{\nu_i} \frac{\Gamma(k + \alpha_i)\Gamma(1/\nu_i + \alpha_i)}{\Gamma(\alpha_i)\Gamma(k + 1 + 1/\nu_i + \alpha_i)}. \tag{2.39}$$

Here

$$\nu_i = \sum_{j=1}^{K} \nu_{ji}$$

with

$$\nu_{ji} = \frac{p_j q_{ji}}{\alpha_{ji} p_i + p_i^*},$$

and

$$\alpha_i = \frac{\sum_{j=1}^{K} \nu_{ji}\alpha_{ji}}{\sum_{j=1}^{K} \nu_{ji}} = \frac{\sum_{j=1}^{K} \nu_{ji}\alpha_{ji}}{\nu_i}.$$

And $\Gamma(\cdot)$ *is the gamma function. When* $p_i^* = 0$, *the* i-*th term in* c_k^* *should be interpreted as* p_i *when* $k = 0$, *and* 0 *otherwise.*

Since the proof mimics that of CMRT, we will only include here the computation for the final step. Recall that for CMRT we have the integral

$$\int_0^\infty \mathbb{P}(\text{Poisson}(r_A^* s) = k)e^{-s} ds = \frac{1}{1 + r_A^*} \left(\frac{r_A^*}{1 + r_A^*} \right)^k.$$

For CMPA, the (time-homogeneous) Poisson process is replaced by a time-inhomogeneous Poisson process. To get the birth rates, observe that similar to the proof of CMRT we have $n_i(t)/n(t) \xrightarrow{\text{a.s.}} p_i$ and $n_{ij}(t)/n(t) \xrightarrow{\text{a.s.}} p_j q_{ji}$ as $t \to \infty$. Therefore from (2.38) the reproduction of a type i individual v is approximately

$$\sum_{j=1}^{K} \frac{p_j q_{ji}(d_v(t) + \alpha_{ji})}{\alpha_{ji} p_i + \sum_{l=1}^{K} p_l q_{li}} = \nu_i(d_v(t) + \alpha_i).$$

This leads to a pure birth process with rate $\nu_i(m + \alpha_i - 1)$, where m is the current population size. We shall denote this process by $\{\tilde{Y}(\nu_i, t) : t \geq 0\}$ (with $\tilde{Y}(\nu_i, 0) = 1$). Then similar to CMRT we need to compute the integral $\int_0^\infty \mathbb{P}(\tilde{Y}(\nu_i, s) = k + 1)e^{-s} ds$.

Since transition probability function of $\{\tilde{Y}(\nu_i, t) : t \geq 0\}$ can be computed explicitly (see *e.g.* [94, Proposition 6.1]), we have

$$\mathbb{P}(\tilde{Y}(\nu_i, s) = k + 1) = \frac{\Gamma(k + \alpha_i)}{\Gamma(\alpha_i)\Gamma(k+1)} e^{-\alpha_i \nu_i s}(1 - e^{-\nu_i s})^k.$$

Plugging it into the integral we get

$$\int_0^\infty \mathbb{P}(Y(\nu_i, s) = k + 1)e^{-s} ds = \frac{\Gamma(k + \alpha_i)}{\Gamma(\alpha_i)\Gamma(k+1)} \int_0^\infty e^{-\alpha_i \nu_i s}(1 - e^{-\nu_i s})^k e^{-s} ds.$$

Let $x = e^{-\nu_i s}$ and by change of variable we have

$$\int_0^\infty e^{-\alpha_i \nu_i s}(1 - e^{-\nu_i s})^k e^{-s} ds = \frac{1}{\nu_i} \int_0^1 x^{1/\nu_i + \alpha_i - 1}(1 - x)^k dx = \frac{1}{\nu_i} \frac{\Gamma(k+1)\Gamma(1/\nu_i + \alpha_i)}{\Gamma(k + 1 + 1/\nu_i + \alpha_i)}.$$

This leads to the constant in (2.39).

Remark 2.9. Similar to CMRT, the limiting degree distribution of CMPA tree is a mixture of K distributions, and coincides with that of preferential attachment model [18] when $\alpha_{ji} \equiv \alpha$ and the balance equation $p_i^* = \sum_{j=1}^{K} p_j q_{ji} = p_i$ holds for all $1 \leq i \leq K$. Moreover, for large k, using Stirling's approximation we get $c_k^* = O(k^{-1-1/\nu})$ where $\nu = \max_i \nu_i$. Consider the special case where $\alpha_{ji} \equiv \alpha$. we have $\sum_{i=1}^{K} p_i^* = \sum_{j=1}^{K} p_j \sum_{i=1}^{K} q_{ji} = \sum_{j=1}^{K} p_j = 1$, and $\sum_{i=1}^{K} p_i^*(1/\nu_i - 1) = \alpha \sum_{i=1}^{K} p_i = \alpha$. It follows that $-1 - 1/\nu \geq -2 - \alpha$ and equality holds if and only if the balance equation holds. Therefore, the limiting degree distribution of CMPA tree typically has a heavier tail than that of preferential attachment tree with the same parameter α. The tail is heavier when p_i^* is larger compared to p_i, which agrees with intuition.

CHAPTER 3
Recovering clusters from covariance structure

Clustering, a central problem in unsupervised learning, aims to partition a set of objects into similar groups. Among the various algorithms developed to this end, many are based on the K-means criterion, *e.g.* the standard iterative algorithm first proposed in 1957 by Stuart Lloyd of Bell Labs [65]. To analyze these algorithms theoretically and computationally, one popular choice is to consider their performance under Gaussian mixture model. Under this framework, we have independent observations from a mixture of Gaussian components, and the goal is to examine the clustering error rate using certain algorithm.

However, given that the goal of K-means is to minimize within-cluster sum of squares, it might be more natural to think about covariances instead of means. Intuitively, K-means should work when covariances within the same group are larger than that between different groups. In this work we take this prospective and consider a collection of correlated Gaussian vectors that have certain latent group structure. For better illustration we index the dimensions by time and treat these vectors as processes. Specifically, we have n Gaussian processes and each process is observed at the same T time points. Moreover, each process belongs to one of the K groups, and information about group labels are encoded in the covariance structure. Our goal is to recover the groups using as few observations as possible.

In this chapter we shall provide conditions under which exact recovery from covariance structure is possible via K-means. Compared with previous results, which mostly focus on the mean structure and require observations to be independent, we assume that no information is contained in the group means and impose covariance structure based on group partition instead. For a motivating example, consider the following vector autoregression (VAR) model with n nodes as dimensions. Write $\mathbf{X}(t) = (X_1(t), X_2(t), ..., X_n(t))'$ and $\boldsymbol{\varepsilon}(t) = (\varepsilon_1(t), \varepsilon_2(t), ..., \varepsilon_n(t))'$ for the observations and white noise at time t. Suppose

$$\mathbf{X}(t+1) = A\mathbf{X}(t) + \boldsymbol{\varepsilon}(t) \tag{3.1}$$

where A is the adjacency matrix of some graph with certain community structure. For example, consider a simple case where each node belongs to one of the K groups, and the $(i,j)^{th}$ entry of A is equal to $1/n$ if i and j are in the same group, and 0 otherwise. Here the normalizing factor n is needed to ensure stationarity of the process.

Then as a special case of the model we shall consider in the main theorem, we have the following exact recovery result:

Theorem 3.1. *Suppose each group has size at least rn for $r > 0$. Then for all $n \geq 1$ and $T \geq Cn \log n$, with probability at least $1 - n^{-2}$, global optimum of K-means is unique and recovers the K groups exactly. Here $C = C(r, K) > 0$ is a constant that depends on r and K.*

As an application, (3.1) can be used to model brain activity, where the n processes represent signals observed from voxels and the K groups represent brain regions. Indeed, PECOK algorithm [24], based on convex relaxation of K-means, is used to analyze fMRI data [23], and theoretical guarantees are given under a variable clustering scheme called G-block covariance model. However, G-block covariance model assumes that observations are independent across time, and in this sense (3.1) could be a more reasonable choice for brain data.

In general, we do not require a generative model like (3.1) and instead impose conditions on covariances. And as in Theorem 3.1, instead of any specific algorithm, we work directly with the global optimum of K-means.

We shall also conduct a simulation study to compare performance of various algorithms, including K-means, spectral clustering and an iterative algorithm, similar in spirit to Lloyds algorithm for K-means, that fits a blockwise constant approximation to the sample covariance matrix.

The rest of this chapter is organized as follows: Section 3.1 gives an overview on related work. Section 3.2 formulates the model and introduces assumptions we make on the covariance structure. Section 3.3 states the main theorem, with applications to several special cases. Section 3.4 contains the proofs. Section 3.5 discusses algorithms based on sample covariance matrix. And Section 3.6 is devoted to simulation study.

3.1 Literature review

3.1.1 Notation

For an $N \times N$ square matrix $A = (a_{ij})_{i,j=1}^N$, we write $\text{Tr}(A) = \sum_{i=1}^N a_{ii}$ for its trace, $||A||_F = (\sum_{i,j=1}^N a_{ij}^2)^{1/2} = \sqrt{\text{Tr}(A'A)}$ for its Frobenus norm, $||A||_{op} = \inf\{c \geq 0 : ||A\nu|| \leq c||\nu||$ for all $\nu \in \mathbb{R}^N\}$ for its operator norm (here $||\cdot||$ denotes the usual Euclidean norm in $\mathbb{R}^N$), $||A||_\infty = \max_i \sum_{j=1}^N |a_{ij}|$ for its L_∞ norm (*i.e.* maximal absolute row sum), and $||A||_{max} = \max_{i,j} |a_{ij}|$ for its max norm. Denote by I_N an identity matrix of order N and by

$$\delta(i,j) = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{otherwise.} \end{cases}$$

the delta function.

3.1.2 Recovery rate of clustering algorithms

Under the framework of Gaussian mixture model, much work has been done to estimate the model parameters, but less was devoted to the study of clustering error rate. In the context of K-means, theoretical guarantees for the standard Lloyd's algorithm and other variants were given under different signal to noise conditions. For Lloyd's algorithm (with appropriate initialization), Lu and Zhou [66] give the minimax optimal rate in partial recovery. Another line of work is based on the convex relaxation of K-means proposed by Peng and Wei [85], where various exact recovery results [95] [62] have been derived and generalized to partial recovery [45]. There are partial recovery results based on other convex relaxations as well [39].

Since estimation of model parameters is not an issue in recovery problem, most of these results apply to the more general sub-Gaussian mixture model. Also, these results are often stated in terms of certain signal to noise ratio (SNR). For simplicity let us consider a Gaussian mixture with K groups of equal size n/K and covariance Σ (in general the sizes and covariance matrices need not be the same). Use θ to denote the minimum Euclidean distance between means of different components and define the following two notions of SNR: $s_1^2 = \min\{\frac{\theta^2}{||\Sigma||_{op}}, \frac{n\theta^4}{K||\Sigma||_F^2}\}$ and $s_2^2 = \theta^2/||\Sigma||_{op}$. Then for perfect recovery, Royer [95] provides a sufficient condition $s_1^2 \geq C \max\{K, \log n\}$ for some constant $C > 0$. For partial recovery, Giraud and Verzelen [45] prove that the proportion of mis-clustered observations decay exponentially

with respect to s_1^2 as long as $s_1^2 \geq CK$ for some constant $C > 0$. Lu and Zhou [66] and Fei and Chen [39] prove similar exponential decay rate with respect to s_2^2 under stronger minimal SNR conditions. For our work we do not have a notion of SNR, but as we shall see in Section 3.13, our results are comparable with that of Royer [95] under special cases.

Performance of these algorithms have been analyzed under other probabilistic model as well, *e.g.* the stochastic ball model [9] [51]. Closer to our work, Bunea et al. [24] derive exact recovery results based on the aforementioned PECOK algorithm under G-block covariance model. In the deterministic setting, Kumar and Kannan [60] and Awasthi and Sheffet [10] prove that under certain proximity or separation conditions, Lloyd's algorithm with an initialization step will recover all groups exactly with high probability. These results have the advantage of not requiring any distributional assumption. However, they do not apply to the high dimensional setting we shall consider.

3.1.3 K-means based methods

Let $\mathbf{X} \in \mathbb{R}^{n \times p}$ be a $n \times p$ matrix, whose rows $\mathbf{X}_i$ ($i = 1, ..., n$) are observations to be clustered. The K-means objective function with respect to a partition $\{1, 2, ..., n\} = \cup_{a=1}^K C_a$ is given by the within-cluster sum of squares (WCSS):

$$WCSS(\{C_a\}_{a=1}^K) := \sum_{a=1}^K \sum_{i \in C_a} ||\mathbf{X}_i - \bar{\mathbf{X}}^{(a)}||_2^2 \tag{3.2}$$

where $\bar{\mathbf{X}}^{(a)} = \sum_{i \in C_a} \mathbf{X}_i / |C_a|$ are the cluster means, and $||\cdot||_2$ denotes the usual Euclidean norm. K-means algorithm aims to find the partition that minimizes the above objective function, given the number of clusters K.

The K-means optimization problem is NP-hard in worst cases, and most practical algorithms are only guaranteed to find local optimums, *e.g.* the standard Lloyd's algorithm [65], sometimes also referred to as "naive K-means". This iterative algorithm takes K initial cluster centers (usually chosen from the n observations), and alternative between two steps. In the assignment step, each observation is assigned to the nearest cluster center in terms of Euclidean distance and this leads to K clusters. In the update step, cluster centers are updated as means of the K clusters formed in the assignment step. There are many variants of K-means, and some of them has better theoretical guarantee than just local optimums. For example, k-means++

[6] includes a randomized seeding step to choose the initial cluster centers, and a competitive provable upper bound on the objective function (3.2) can be established accordingly.

As another alternative, semi-definite programming (SDP) relaxations of K-means are also well-developed. While there are many variants in the literature, we are only going to introduce the one used in [45] to derive the state-of-the-art recovery results.

Note first that any partition $\{1, 2, ..., n\} = \cup_{a=1}^{K} C_a$ can be represented uniquely by a $n \times n$ matrix Z such that $Z_{ij} = 1/|C_a|$ if i and j are in the same group C_a, and $Z_{ij} = 0$ otherwise. Peng and Wei [85] show that minimizing (3.2) is equivalent to the following optimization problem:

$$\begin{aligned} \min \quad & < \mathbf{XX'}, Z > \\ & Z \text{ is symmetric}, Z \geq 0, \\ & Z^2 = Z, Z1 = 1, \mathrm{Tr}(Z) = K. \end{aligned} \qquad (3.3)$$

Here $Z \geq 0$ means that the inequality holds entry-wise, *i.e.* all entries of Z are nonnegative. Given a maximizer $\hat{Z}$, clusters can be constructed by grouping i and j together for $\hat{Z}_{ij} \neq 0$.

To relax (3.3), Peng and Wei replace the non-convex constraint $Z^2 = Z$ by positive semi-definiteness and consider the following SDP problem:

$$\begin{aligned} \min \quad & < \mathbf{XX'}, Z > \\ & Z \text{ is positive semi-definite}, \\ & Z \geq 0, Z1 = 1, \mathrm{Tr}(Z) = K. \end{aligned} \qquad (3.4)$$

To construct the clusters from maximizer $\hat{Z}$ of (3.4), one has to apply a rounding procedure. In the original work [85] Peng and Wei use a procedure based on principal component analysis, and in [45] this is replaced by clustering rows of $\hat{Z}$ using an approximate K-medoids algorithm.

3.1.4 G-block covariance model

In this section we give a brief introduction to the G-block covariance model considered in [22] as a framework for variable clustering, and relate that to the covariance-based clustering problem we shall study.

Let $X = (X_1, ..., X_p)$ be a p-dimensional zero-mean random vector, and suppose its covariance matrix Σ has the following block structure:

$$\Sigma = Z\Omega Z' + \Gamma, \tag{3.5}$$

where the $p \times K$ matrix Z is the membership matrix corresponding to partition $\{1, 2, ..., p\} = \cup_{a=1}^{K} G_a$ with entries $Z_{ja} := \mathbb{1}\{j \in G_a\}$, Ω is a $K \times K$ symmetric matrix, and Γ is a $p \times p$ diagonal matrix with $\Gamma_{jj} = \gamma_a \geq 0$ for $j \in G_a$.

An important sub-class of G-block covariance model is studied in [24], when

$$X_j = M_a + \varepsilon_j, \quad \text{for all } j \in G_a, \tag{3.6}$$

where $M = (M_1, ..., M_K)$ is a K-dimensional zero-mean vector independent of the p-dimensional zero-mean error vector $\varepsilon = (\varepsilon_1, ..., \varepsilon_p)$. In addition, $\text{Cov}(M) = \Omega$, and ε is assumed to have independent coordinates with variances given by $\text{Cov}(\varepsilon) = \Gamma$. Not all G-block covariance model is of this form because Ω is not necessarily positive semi-definite in general.

For the general G-block covariance model (3.5), separation between clusters can be measured by

$$\text{MCord}(\Sigma) := \min_{i \nsim j} \max_{k \neq i, j} |\Sigma_{ik} - \Sigma_{jk}|,$$

where $i \sim j$ is the equivalence relation induced by partition $\{G_a\}$, i.e. $i \sim j$ when i and j belong to the same block. $\text{MCord}(\Sigma) = 0$ if and only if all off-diagonal entries of Σ are identical. For the special case (3.6), separation can be the within-between group covariance gap, which is equivalent to distance between the latent means M_a:

$$\Delta(\Omega) := \min_{a \neq b}(\Omega_{aa} + \Omega_{bb} - 2\Omega_{ab}) = \min_{a \neq b} \mathbb{E}(M_a - M_b)^2.$$

$\Delta(\Omega) = 0$ if and only if all coordinates of M are identical.

Suppose we have n i.i.d. observations $X^{(1)}, ..., X^{(n)}$ of X from a G-block covariance model, the goal of variable clustering is to recover the true partition $\{G_a\}$. Note that here the block structure is encoded in the covariance matrix, which fits into our framework. Indeed, if we index the observations by time and consider p processes $X_j^{(t)}$ ($j = 1, ..., p$), variable clustering is equivalent to clustering the processes. However, these processes are independent across time.

In terms of perfect recovery, the CORD algorithm introduced in [22] is minimax-optimal with respect to the separation measure $\mathrm{MCord}(\Sigma)$ in the general case (3.5), and a penalized convex relaxation of K-means called PECOK is near minimax-optimal with respect to $\Delta(\Omega)$ in the special case (3.6).

3.1.5 Stochastic block network VAR model

In this section we give an introduction to a class of vector autoregressive (VAR) model considered in [47], in which coefficient matrices are specified by certain generalized stochastic block model. For simplicity we shall only describe the undirected case.

Let $\mathbf{X}(t) = (X_1(t), X_2(t), ..., X_n(t))'$ be a n-dimensional vector, with dynamics given by the following VAR(p) model:

$$\mathbf{X}(t) = \sum_{l=1}^{p} \Phi_l \mathbf{X}(t-l) + \varepsilon(t)$$

where Φ_l $(l = 1, ..., p)$ are $n \times n$ coefficient matrices, and $\{\varepsilon(t)\}$ is a zero-mean covariance-stationary stochastic process.

Assume that Φ_l is specified by a latent undirected weighted graph $\mathcal{G}_l$. Let A_l be the $n \times n$ adjacency matrix of $\mathcal{G}_l$, with $(A_l)_{ij}$ equal to the edge weight w_{ij} if there exists an edge between vertex i and j, and 0 otherwise. Define also the $n \times n$ diagonal degree matrix D_l, with $(D_l)_{ii} = \sum_{j=1}^{n} (A_l)_{ij}$, and the symmetric normalized Laplacian matrix $L_l = D_l^{-1/2} A_l D_l^{-1/2}$. Then Φ_l is defined as follows:

$$\Phi_l = \phi_l D_l^{-1/2} A_l D_l^{-1/2} = \phi_l L_l,$$

where $\phi_l > 0$ is called the stationarity parameter. Assume further that $\mathcal{G}_l$ are i.i.d. realizations of a generalized stochastic block model defined below.

Definition 3.2 ((Generalized stochastic block model [89])). *Let Z be the $n \times K$ membership matrix, with entries $Z_{ia} = 1$ if vertex i belongs to community a, and 0 otherwise. Let P be the $K \times K$ symmetric edge probability matrix, and Θ be the $n \times n$ diagonal matrix of nonnegative vertex-specific weights. Let F be a probability distribution supported on a bounded interval.*

In a generalized stochastic block model with n vertices, the probability that an edge exists between vertex i and j which belong to community a and b, respectively, is $\Theta_{ii}\Theta_{jj}P_{ab} \leq 1$, and all edges are formed independently. Moreover, each edge (i, j) is assigned a weight w_{ij} that follows the distribution F, independent of all edges and other weights. Write $\mathcal{G} \sim GSBM(Z, P, \Theta, F)$ when a random graph $\mathcal{G}$ is generated this way.

Note that P and Θ are only unique up to a multiplicative constant, so we can normalize Θ such that Θ_{ii} sums to 1 within each community.

Remark 3.1. If one treats the observations at each vertex, namely $\{X_i(t)\}$, as individual processes, we get n correlated stationary processes, and the covariance matrix of $\mathbf{X}(t)$ has an underlying block structure. This fits into our framework perfectly. However, as we shall see in simulation study, algorithms tailored to this VAR model perform much better than more generic methods like K-means.

As in [47], we call the VAR model generated above the stochastic block network VAR(p) model. And community detection aims to recover the block structure given a sample of observations $\mathbf{X}(t)$ ($t = 1, ..., T$). To this end Guðmundsson proposes an algorithm called VAR Blockbuster. First, derive the ordinary least squares (OLS) estimates of coefficient matrices. Write $\mathbf{R}(t) = (\mathbf{X}(t-1)', ..., \mathbf{X}(t-p)')'$ for the lagged values. Then the OLS estimate $\hat{\Phi} = (\hat{\Phi}_1, ..., \hat{\Phi}_p)$ is given by

$$\hat{\Phi} = \Big(\sum_{t=p+1}^{T} \mathbf{X}(t)\mathbf{R}(t)' \Big) \Big(\sum_{t=p+1}^{T} \mathbf{R}(t)\mathbf{R}(t)' \Big)^{-1}$$

when $\sum_{t=p+1}^{T} \mathbf{R}(t)\mathbf{R}(t)'$ is invertible. Compute the sum of symmetrized coefficient matrices $\sum_{l=1}^{p}(\hat{\Phi}_l + \hat{\Phi}_l')$ and gather eigenvectors corresponding to its K largest eigenvalues into a $n \times K$ matrix $\hat{U}$. Finally, apply k-means++ [6] to the row-normalized version of $\hat{U}$.

Guðmundsson establishes the following clustering error rate bound for the VAR Blockbuster algorithm:

Theorem 3.3. *Consider a stochastic block network VAR(p) model with fixed Θ and F. Define $p_n = \min_a P_{aa}$. Let r be the proportion of mis-clustered vertices. Then under Assumption 3.4 and 3.5 we have:*

$$r = O(\frac{n}{T} + \frac{\log n}{np_n})$$

with high probability as $n \to \infty$, if $T = \Omega(n^{2/\gamma - 1})$, where γ is as in (g) of Assumption 3.5.

The first assumption here concerns edge probabilities:

Assumption 3.4. *P is positive definite and $p_n = \min_a P_{aa} = \Omega(\log n/n)$.*

Recall here that $p_n = \Theta(\log n/n)$ is the "semi-sparse" domain, the sparsest regime where exact recovery is possible.

Assumption 3.5. *Let* $\{\mathbf{X}(t)\}$ *be a stochastic block network VAR(p) model. Write* $G = \{\mathcal{G}_1, ..., \mathcal{G}_p\}$. *Define covariance matrices* $\Sigma_\varepsilon = \mathrm{Cov}(\varepsilon(t)) = \mathbb{E}(\varepsilon(t)\varepsilon(t)')$ *and* $\Sigma_{\mathbf{X}} = \mathbb{E}(\mathbf{X}(t)\mathbf{X}(t)'|G)$. *Assume that:*

(a) $\{\varepsilon(t)\}$ *is covariance-stationary and serially uncorrelated.*

(b) $\mathbb{E}(\mathbf{X}(t)\varepsilon(t)'|G) = \mathbf{0}$ *for all* t.

(c) $||\Sigma_\varepsilon||_{op} < C_1$ *and* $||\Sigma_\varepsilon^{-1}||_{op} < C_2$, *where* $C_1, C_2 > 0$ *are constants.*

(d) $\sum_{l=1}^p \phi_l < 1$.

(e) $\{\Sigma_{\mathbf{X}}^{-1/2}\mathbf{X}(t)\}$ *is strongly mixing with mixing coefficient*

$$\alpha(m) \leq e^{-c_1 m^{\gamma_1}}$$

for all positive integer m, *where* $c_1, \gamma_1 > 0$ *are constants.*

(f) For arbitrary n-dimensional vector c with unit Euclidean norm,

$$\mathbb{P}(|c'\Sigma_{\mathbf{X}}^{-1/2}\mathbf{X}(t)| > s|G) \leq e^{1-(s/c_2)^{\gamma_2}}$$

for all $s > 0$ *and* t, *where* $c_2, \gamma_2 > 0$ *are constants.*

(g) $\gamma < 1$ *where* $1/\gamma = 1/\gamma_1 + 1/\gamma_2$.

Assumptions (a),(b) and (d) guarantee that $\{\mathbf{X}(t)\}$ is covariance-stationary. (c) bounds eigenvalues of Σ_ε from both sides. The remaining assumptions are needed for the concentration inequality of Merlevàde [72] to apply. Note that the mixing coefficient in (e) is defined conditioning on G. Specifically, let $\mathcal{F}_{-\infty}^t = \sigma\{\Sigma_{\mathbf{X}}^{-1/2}\mathbf{X}(s) : -\infty \leq s \leq t\}$ and $\mathcal{F}_t^\infty = \sigma\{\Sigma_{\mathbf{X}}^{-1/2}\mathbf{X}(s) : t \leq s \leq \infty\}$. Then the α-mixing coefficient is defined as

$$\alpha(m) := \sup_t \sup_{A \in \mathcal{F}_{-\infty}^t, B \in \mathcal{F}_{t+m}^\infty} |\mathbb{P}(A \cap B|G) - \mathbb{P}(A|G)\mathbb{P}(B|G)|.$$

Remark 3.2. In the special case where $\{\varepsilon(t)\}$ is Gaussian, we have $\gamma_1 = 1$ and $\gamma_2 = 2$ from [72, Corollary 3]. It follows that $\gamma = 2/3$ and the condition in Theorem 3.3 becomes $T = \Omega(n^{2/\gamma-1}) = \Omega(n^2)$. Suppose further that $\inf_n p_n > 0$, then Theorem 3.3 states that the number of mis-clustered vertices is of order $O(\log n)$, which is the best bound one could hope for from the theorem.

3.2 Model formulation and assumptions

We use $X_i(t)$ to denote value of process i observed at time t, and write $\mathbf{X}_i = (X_i(1), X_i(2), ..., X_i(T))'$ for process i. The first assumption we need is joint Gaussianity of these processes.

Assumption 3.6. $\{\mathbf{X}_i\}_{i=1}^n$ *are jointly Gaussian.*

Let $\{1, 2, ..., n\} = \cup_{a=1}^K G_a$ corresponds to the partition of n processes into K groups such that processes in group a have indices in G_a. Let $n_a^G = |G_a|$ be the size of group a. We assume that each group has size of order n:

Assumption 3.7. $\min_a \ n_a^G \geq rn/K$ *for some* $r \in (0, 1]$.

The normalizing factor K here ensures that r does not necessarily depend on K, and $r = 1$ if and only if we have K groups of equal size n/K.

For simplicity we center and scale the process and introduce some notations to better describe the group structure:

Assumption 3.8. *For all* $1 \leq i \leq n$ *and* $1 \leq t \leq T$,

$$\mathbb{E}(X_i(t)) = 0.$$

For all $1 \leq i, j \leq n$ *with* $i \in G_a$, $j \in G_b$ *for some* $1 \leq a, b \leq K$, *and* $1 \leq t_1, t_2 \leq T$,

$$\mathrm{Cov}(X_i(t_1), X_j(t_2)) = \delta(i, j)\delta(t_1, t_2) + \rho_{ab}(t_1, t_2) + \xi_{ij}(t_1, t_2). \tag{3.7}$$

Here $\rho_{ab}(t_1, t_2)$ and $\xi_{ij}(t_1, t_2)$ are fixed parameters. $\rho_{ab}(t_1, t_2)$ should be interpreted as the part of autocovariance that can be explained by the group labels (similar to a "groupwise average"), while $\xi_{ij}(t_1, t_2)$ is some kind of perturbation.

For convenience we define $T \times T$ matrices $P(i, j) = \mathrm{Cov}(\mathbf{X}_i, \mathbf{X}_j) - \delta(i, j)I_T$, $\Xi_{ij} = (\xi_{ij}(t_1, t_2))_{t_1, t_2=1}^T$ and $P_{ab} = (\rho_{ab}(t_1, t_2))_{t_1, t_2=1}^T$ for $1 \leq i, j \leq n$ and $1 \leq a, b \leq K$. Then (3.7) can be rewritten as

$$\mathrm{Cov}(\mathbf{X}_i, \mathbf{X}_j) = \delta(i, j)I_T + P_{ab} + \Xi_{ij} \tag{3.8}$$

for $i \in G_a$ and $j \in G_b$.

To recover the groups, we need the following assumptions on the covariance structure:

Assumption 3.9. *For all $1 \leq i, j \leq n$ with $i \in G_a$, $j \in G_b$ for some $1 \leq a, b \leq K$, and $1 \leq t, t_1, t_2 \leq T$,*

$$\rho_{ab}(t_1, t_2) \geq |\xi_{ij}(t_1, t_2)| \geq 0, \tag{3.9}$$

$$\sum_{b=1}^{K} ||P_{ab}||_\infty \leq \frac{M}{n}, \ \text{ for some constant } M > 0, \tag{3.10}$$

$$\rho_{aa}(t, t) + \rho_{bb}(t, t) - 2\rho_{ab}(t, t) \geq \frac{2\gamma}{n}, \ \text{ for some constant } \gamma > 0 \text{ when } a \neq b, \tag{3.11}$$

$$||\Xi_{ij}||_\infty \leq \frac{r\gamma}{4K^3 n^2}. \tag{3.12}$$

Note that from (3.10) and (3.11) we have $\gamma \leq M$.

We briefly explain the motivation behind these assumptions. First, (3.9) ensures nonnegative correlation between observations at different points and time. This condition can be relaxed and it is obvious from the proof that we only need sign of $\rho_{ab}(t_1, t_2)$ to be the same for all $1 \leq a, b \leq K$, given any pair of time points t_1 and t_2. Violation of this assumption can cause larger variability in the objective function of K-means and makes recovery harder. (3.11) tells us that correlation within the same group is stronger than that between different groups. Since K-means aims to minimize within-cluster sum of squares, larger within group covariance will encourage the optimum to include points from the same group in the same cluster. (3.10) controls long-range dependence of processes across time. If observations at different time are highly correlated, we gain little information by adding more time points. Note that instead of the sum over b, we can assume $||P_{ab}||_\infty \leq \frac{M}{nK}$ instead. However, as (3.11) and examples in Section 3.3 suggest, $||P_{aa}||_\infty$ is usually much larger compared to $||P_{ab}||_\infty$ with $b \neq a$. In cases where they are of the same order, we can choose M as a multiple of K. Finally, (3.12) controls magnitude of the perturbation terms.

Although M and γ are not necessarily independent of n, K and T, the results are more natural when that is indeed the case. To see that, note that (in this case) (3.10) and (3.11) readily give $\rho_{aa}(t, t) = O(1/n)$ for large n. Moreover, from (3.12) we know that $\xi_{ij}(t_1, t_2) = O(1/n^2)$ for fixed K. It follows that $\text{Var}(X_i(t)) = 1 + \rho_{aa}(t, t) + \xi_{ii}(t, t) = 1 + O(1/n)$ for $i \in G_a$, and $\rho_{ab}(t_1, t_2) + \xi_{ij}(t_1, t_2)$ is approximately equal to the auto-correlation function. Order $1/n$ here comes from the normalizing factor n in (3.10), and we shall see in Section 3.13 that this is a reasonable choice.

3.3 Main theorem and special cases

Theorem 3.10. *Under Assumptions 3.6, 3.7, 3.8 and 3.9, for all $n \geq 1$ and $T \geq CK^5 n \log n$, with probability at least $1 - n^{-2K^2}$, global optimum of K-means is unique and recovers the K groups exactly. Here $C = C(r, M, \gamma) > 0$ is a constant that depends on r, M and γ.*

3.3.1 Time-independent case

For a special case where the process is independent across time, consider:

$$X_i(t) = \mu_a(t) + \varepsilon_i(t) \text{ if } i \in G_a \tag{3.13}$$

Here $\mu_a(t) \sim N(0, \frac{1}{n})$ and $\varepsilon_i(t) \sim N(0,1)$ are mutually independent.

In this case Assumption 3.6, 3.8 and 3.9 hold without further assumptions if we let $\rho_{ab}(t_1, t_2) = \frac{1}{n}\delta(a,b)\delta(t_1, t_2)$, $\xi_{ij}(t_1, t_2) = 0$, $M = 1$ and $\gamma = 1$.

Note that if we condition on $\mu_k(t)$, we actually have a Gaussian mixture model here, and the minimum Euclidean distance between means of different components θ satisfies $\theta^2 = O(T/n)$. Recall from Section 3.1 that signal to noise ratio $s_1^2 = \min\{\frac{\theta^2}{||\Sigma||_{op}}, \frac{n\theta^4}{K||\Sigma||_F^2}\}$. In this case $\Sigma = I_T$ and $s_1^2 = O(T/n)$ if one treats K as a constant. From Royer [95] we know that perfect recovery can be achieved by the Peng-Wei convex relaxation of K-means when $s_1^2 \geq C \log n$ for some constant $C > 0$. This is exactly the same condition as given by Theorem 3.10 with respect to n.

In general, under time-independence our model is similar to the G-block covariance model described in Section 3.1.4, for which (3.13) is also a special case. So the case above also shows a relationship between G-block covariance model and Gaussian mixture model.

3.3.2 Block VAR model

Write $\mathbf{X}(t) = (X_1(t), X_2(t), ..., X_n(t))'$ and $\boldsymbol{\varepsilon}(t) = (\varepsilon_1(t), \varepsilon_2(t), ..., \varepsilon_n(t))'$. Consider the following vector autoregression (VAR) model:

$$\mathbf{X}(t+1) = A\mathbf{X}(t) + \boldsymbol{\varepsilon}(t) \tag{3.14}$$

where $\boldsymbol{\varepsilon}(t) \sim N(\mathbf{0}, \mathbf{I}_n)$ are mutually independent. Suppose entries of the coefficient matrix A are specified by $A_{i,j} = p_{ab}$ when $i \in G_a$ and $j \in G_b$. Consider the following assumptions:

Assumption 3.11.

$$p_{ab} \geq 0. \tag{3.15}$$

$$n \sum_{b=1}^{K} p_{ab} \leq 1 - \tau. \tag{3.16}$$

$$p_{aa} - p_{ab} \geq \frac{\eta}{n}, \quad when \ a \neq b. \tag{3.17}$$

Here $a, b \in \{1, 2, ..., K\}$ and $0 \leq \tau \leq 1$, $\eta > 0$ are constants.

Then we have the following proposition:

Proposition 3.12. *Under assumption 3.7, Assumption 3.11 implies Assumptions 3.8 and 3.9 with $M = 2K/\tau^2$ and $\gamma = r\eta^2/K$.*

In the simple case of (3.1), we have $p_{ab} = \delta(a, b)/n$ and Assumption 3.11 obviously holds.

3.3.3 Stationary case

The previous two examples are both special cases of stationary processes. In general, (with a slight abuse of notation) if $\rho_{ab}(t_1, t_2) = \rho_{ab}(|t_1 - t_2|)$ is a function of $|t_1 - t_2|$, we have

$$P_{ab} = \begin{pmatrix} \rho_{ab}(0) & \rho_{ab}(1) & \cdots & \rho_{ab}(T-1) \\ \rho_{ab}(1) & \rho_{ab}(0) & \cdots & \rho_{ab}(T-2) \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{ab}(T) & \rho_{ab}(T-2) & \cdots & \rho_{ab}(0) \end{pmatrix}. \tag{3.18}$$

Also, (3.10) in Assumption 3.9 can be replaced by

$$\sum_{b=1}^{K} \left(\rho_{ab}(0) + 2 \sum_{t=1}^{\infty} \rho_{ab}(t) \right) \leq \frac{M}{n}, \tag{3.19}$$

which is a more natural condition on autocovariances in the stationary case. In fact, from (3.18) we know that (3.19) is equivalent to the stronger assumption that (3.10) jointly holds for all $T > 0$.

Note that we only require $\rho_{ab}(t_1, t_2)$ to be stationary here, so the process is not necessarily stationary. For that to be the case, $\xi_{ij}(t_1, t_2)$ also has to be a function of $|t_1 - t_2|$.

Further if we assume no perturbation altogether, *i.e.* $\xi_{ij}(t_1, t_2) = 0$, we may replace Assumptions 3.8 and 3.9 with the following assumptions:

Assumption 3.13.

$$\mathbb{E}(X_i(t)) = 0.$$

$$\mathrm{Var}(X_i(t)) = 1 + \rho_{aa}(0) \ \ when \ i \in G_a.$$

$$\mathrm{Cov}(X_i(t), X_j(t)) = \rho_{ab}(0) \ \ when \ i \neq j, i \in G_a \ and \ j \in G_b.$$

$$\mathrm{Cov}(X_i(t_1), X_j(t_2)) = \rho_{ab}(|t_1 - t_2|) \ \ when \ t_1 \neq t_2, i \in G_a \ and \ j \in G_b.$$

(3.20)

Assumption 3.14.

$$\rho_{ab}(t) \geq 0. \tag{3.21}$$

$$n \sum_{b=1}^{K} (|\rho_{ab}(0)| + 2 \sum_{t=1}^{\infty} |\rho_{ab}(t)|) \leq M, \quad for \ some \ constant \ M > 0. \tag{3.22}$$

$$\rho_{aa}(0) + \rho_{bb}(0) - 2\rho_{ab}(0) \geq \frac{2\gamma}{n}, \quad for \ some \ constant \ \gamma > 0 \ when \ a \neq b. \tag{3.23}$$

Here $a, b \in \{1, 2, ..., K\}$.

In this case we have the following theorem:

Theorem 3.15. *Under Assumptions 3.6, 3.7, 3.13 and 3.14, for all $n \geq 1$ and $T \geq CK^5 n \log n$, with probability at least $1 - n^{-2K^2}$, global optimum of K-means is unique and recovers the K groups exactly. Here $C = C(r, M, \gamma) > 0$ is a constant that depends on r, M and γ.*

3.4   Proofs

In this section we give a proof for Theorem 3.15. Proof for the more general Theorem 3.10 essentially follows the same argument, with lengthier computations. For completeness we still include a proof for Theorem 3.10 in Appendix B.

3.4.1   Proof of Theorem 3.15

Recall that K-means aims to partition the n nodes into sets $\{C_a\}_{a=1}^K$ that minimize the within-cluster sum of squares:

$$WCSS(\{C_a\}_{a=1}^K) := \sum_{a=1}^K \sum_{i \in C_a} ||\mathbf{X}_i - \bar{\mathbf{X}}^{(j)}||_2^2$$

where $\bar{\mathbf{X}}^{(j)} = \sum_{i \in C_a} \mathbf{X}_i / |C_a|$ are the cluster means.

Our goal here is to show that under the assumptions, with probability close to 1, global optimum is uniquely achieved when the clusters coincide with the latent groups, i.e. $\{C_a\}_{a=1}^{K} = \{G_b\}_{b=1}^{K}$. For clarity we shall refer to $\{C_a\}_{a=1}^{K}$ as clusters and $\{G_b\}_{b=1}^{K}$ as groups throughout.

We proceed in two steps. First we show that $WCSS(\{C_a\}_{a=1}^{K}) > WCSS(\{G_b\}_{b=1}^{K})$ with high probability for a given partition $\{C_a\}_{a=1}^{K} \neq \{G_b\}_{b=1}^{K}$. Then we use a union bound on all possible partitions to establish the desired result.

Step 1: For a given partition $\{C_a\}_{a=1}^{K}$, write $n_a^C = |C_a|$ and $n_{ab} = |C_a \cap G_b|$ $(a, b \in \{1, 2, ..., K\})$. First note that

$$
\begin{aligned}
WCSS(\{C_a\}_{a=1}^{K}) &:= \sum_{a=1}^{K} \sum_{i \in C_a} ||\mathbf{X}_i - \bar{\mathbf{X}}^{(j)}||_2^2 = \sum_{a=1}^{K} \left(\sum_{i \in C_a} ||\mathbf{X}_i||_2^2 - |C_a| \cdot ||\bar{\mathbf{X}}^{(j)}||_2^2 \right) \\
&= \sum_{i=1}^{n} ||\mathbf{X}_i||_2^2 - \sum_{a=1}^{K} |C_a| \cdot ||\bar{\mathbf{X}}^{(j)}||_2^2 \\
&= \sum_{i=1}^{n} ||\mathbf{X}_i||_2^2 - \sum_{a=1}^{K} \frac{||\sum_{i \in C_a} \mathbf{X}_i||_2^2}{|C_a|}.
\end{aligned}
$$

Therefore we have

$$
\Delta := WCSS(\{C_a\}_{a=1}^{K}) - WCSS(\{G_b\}_{b=1}^{K}) = \sum_{b=1}^{K} \frac{||\sum_{i \in G_b} \mathbf{X}_i||_2^2}{n_b^G} - \sum_{a=1}^{K} \frac{||\sum_{i \in C_a} \mathbf{X}_i||_2^2}{n_a^C}.
$$

Instead of expressing Δ as a quadratic form in $\{\mathbf{X}_i\}_{i=1}^{n}$ with nT dimensions, we shall view it as a quadratic form in group and cluster means with $2KT$ dimensions. Formally, let $\mathbf{Y}_a = \sum_{i \in G_a} \mathbf{X}_i / \sqrt{n_a^G}$ for $1 \leq a \leq K$ be the scaled group means and $\mathbf{Z}_b = \sum_{i \in C_b} \mathbf{X}_i / \sqrt{n_b^C}$ for $1 \leq b \leq K$ be the scaled cluster means. Then $\Delta = \sum_{a=1}^{K} ||\mathbf{Y}_a||_2^2 - \sum_{b=1}^{K} ||\mathbf{Z}_b||_2^2 = \mathbf{U}'\Omega\mathbf{U}$, where

$$
\Omega = \begin{pmatrix} I_{KT} & \\ & -I_{KT} \end{pmatrix}
$$

and $\mathbf{U} = (\mathbf{Y}_1', ..., \mathbf{Y}_K', \mathbf{Z}_1', ..., \mathbf{Z}_K')'$.

We proceed to derive the distribution of $\mathbf{U}$ from (3.20). Specifically, we know that $\mathbf{U}$ is multivariate Gaussian with mean $\mathbf{0}$ and covariance matrix

$$\Sigma = \begin{pmatrix} \Sigma_{11}^G & \cdots & \Sigma_{1K}^G & \tilde{\Sigma}_{11} & \cdots & \tilde{\Sigma}_{K1} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \Sigma_{K1}^G & \cdots & \Sigma_{KK}^G & \tilde{\Sigma}_{1K} & \cdots & \tilde{\Sigma}_{KK} \\ \tilde{\Sigma}_{11}' & \cdots & \tilde{\Sigma}_{1K}' & \Sigma_{11}^C & \cdots & \Sigma_{1K}^C \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \tilde{\Sigma}_{K1}' & \cdots & \tilde{\Sigma}_{KK}' & \Sigma_{K1}^C & \cdots & \Sigma_{KK}^C \end{pmatrix} \tag{3.24}$$

where $\Sigma_{bd}^G = \mathrm{Cov}(\mathbf{Y}_b, \mathbf{Y}_d)$, $\Sigma_{ac}^C = \mathrm{Cov}(\mathbf{Z}_a, \mathbf{Z}_c)$ and $\tilde{\Sigma}_{ab} = \mathrm{Cov}(\mathbf{Z}_a, \mathbf{Y}_b)$. For clarity in what follows we shall use a, c to index the clusters and b, d to index the groups. These matrices can be computed explicitly using (3.20). Covariances between scaled group means are

$$\Sigma_{bb}^G = \mathrm{Cov}(\mathbf{Y}_b) = \frac{1}{n_b^G} \mathrm{Cov}(\sum_{i \in G_b} \mathbf{X}_i) = \frac{1}{n_b^G}(n_b^G I_T + (n_b^G)^2 P_{bb}) = I_T + n_b^G P_{bb}, \tag{3.25}$$

and similarly

$$\Sigma_{bd}^G = \sqrt{n_b^G n_d^G} P_{bd} \tag{3.26}$$

when $b \neq d$. By symmetry we have $\Sigma_{bd}^G = \Sigma_{db}^G$.

Covariances between scaled cluster means are

$$\begin{aligned} \Sigma_{aa}^C = \mathrm{Cov}(\mathbf{Z}_a) = \frac{1}{n_a^C} \mathrm{Cov}(\sum_{i \in C_a} \mathbf{X}_i) &= \frac{1}{n_a^C}(n_a^C I_T + \sum_{1 \leq b,d \leq K} n_{ab} n_{ad} P_{bd}) \\ &= I_T + \frac{1}{n_a^C} \sum_{1 \leq b,d \leq K} n_{ab} n_{ad} P_{bd}, \end{aligned} \tag{3.27}$$

and similarly

$$\Sigma_{ac}^C = \frac{1}{\sqrt{n_a^C n_c^C}} \sum_{1 \leq b,d \leq K} n_{ab} n_{cd} P_{bd}, \tag{3.28}$$

when $a \neq c$. Again by symmetry we have $\Sigma_{ac}^C = \Sigma_{ca}^C$.

Finally covariances between scaled group and cluster means are

$$\tilde{\Sigma}_{ab} = \text{Cov}(\mathbf{Z}_a, \mathbf{Y}_b) = \frac{1}{\sqrt{n_a^C n_b^G}} \text{Cov}\left(\sum_{i \in C_a} \mathbf{X}_i, \sum_{j \in G_b} \mathbf{X}_j\right) = \frac{1}{\sqrt{n_a^C n_b^G}}\left(n_{ab} I_T + \sum_{d=1}^{K} n_b^G n_{ad} P_{bd}\right)$$

$$= \frac{n_{ab}}{\sqrt{n_a^C n_b^G}} I_T + \sqrt{\frac{n_b^G}{n_a^C}} \sum_{d=1}^{K} n_{ad} P_{bd}. \tag{3.29}$$

Going back to the quadratic form $\Delta = \mathbf{U}'\Omega\mathbf{U}$, we decorrelate $\mathbf{U}$ so that Δ can be expresses in terms of a white noise. Let $\boldsymbol{\varepsilon}$ be multivariate Gaussian with mean $\mathbf{0}$ and covariance matrix I_{2KT} such that $\mathbf{U} = \Sigma^{1/2}\boldsymbol{\varepsilon}$. Then we have $\mathbb{E}\,\Delta = \mathbb{E}\,\mathbf{U}'\Omega\mathbf{U} = \mathbb{E}\,\boldsymbol{\varepsilon}'\Sigma^{1/2}\Omega\Sigma^{1/2}\boldsymbol{\varepsilon} = \text{Tr}(\Sigma^{1/2}\Omega\Sigma^{1/2}) = \text{Tr}(\Omega\Sigma)$. Block matrix algebra readily gives

$$\Omega\Sigma = \begin{pmatrix} \Sigma_{11}^G & \cdots & \Sigma_{1K}^G & \tilde{\Sigma}_{11} & \cdots & \tilde{\Sigma}_{K1} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \Sigma_{K1}^G & \cdots & \Sigma_{KK}^G & \tilde{\Sigma}_{1K} & \cdots & \tilde{\Sigma}_{KK} \\ -\tilde{\Sigma}_{11}' & \cdots & -\tilde{\Sigma}_{1K}' & -\Sigma_{11}^C & \cdots & -\Sigma_{1K}^C \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ -\tilde{\Sigma}_{K1}' & \cdots & -\tilde{\Sigma}_{KK}' & -\Sigma_{K1}^C & \cdots & -\Sigma_{KK}^C \end{pmatrix}. \tag{3.30}$$

And from that we know

$$
\mathbb{E}\,\Delta = \mathrm{Tr}(\Omega\Sigma) = \sum_{b=1}^{K} \mathrm{Tr}(\Sigma_{bb}^{G}) - \sum_{a=1}^{K} \mathrm{Tr}(\Sigma_{aa}^{C})
$$

$$
= KT + T\sum_{b=1}^{K} n_{b}^{G}\rho_{bb}(0) - KT - T\sum_{a=1}^{K} \frac{1}{n_{a}^{C}} \sum_{1\le b,d\le K} n_{ab}n_{ad}\rho_{bd}(0)
$$

(by (3.25) and (3.27))

$$
= T\left(\sum_{a=1}^{K}\sum_{b=1}^{K} n_{ab}\rho_{bb}(0) - \sum_{a=1}^{K}\frac{1}{n_{a}^{C}}\sum_{1\le b,d\le K} n_{ab}n_{ad}\rho_{bd}(0)\right)
$$

(using $n_{b}^{G} = \sum_{a=1}^{K} n_{ab}$)

$$
= T\left(\sum_{a=1}^{K}\frac{1}{n_{a}^{C}}\sum_{1\le b,d\le K} n_{ab}n_{ad}\rho_{bb}(0) - \sum_{a=1}^{K}\frac{1}{n_{a}^{C}}\sum_{1\le b,d\le K} n_{ab}n_{ad}\rho_{bd}(0)\right)
$$

(using $n_{a}^{C} = \sum_{d=1}^{K} n_{ad}$)

$$
= \frac{1}{2}\sum_{a=1}^{K}\sum_{\substack{b,d=1\\b\ne d}}^{K} \frac{n_{ab}n_{ad}}{n_{a}^{C}}(\rho_{bb}(0) + \rho_{dd}(0) - 2\rho_{bd}(0))T
$$

$$
\ge \left(\sum_{a=1}^{K}\sum_{\substack{b,d=1\\b\ne d}}^{K}\frac{n_{ab}n_{ad}}{n_{a}^{C}}\right)\frac{\gamma T}{n}. \quad \text{(by (3.23) in Assumption 3.14)}
$$

(3.31)

To make this lower bound more amenable to analyze, we introduce a measure s to quantify the difference between $\{C_a\}_{a=1}^{K}$ and $\{G_b\}_{b=1}^{K}$. W.L.O.G. we can reorder $\{C_a\}_{a=1}^{K}$ and $\{G_b\}_{b=1}^{K}$ such that $n_{aa} = \max_{a',b'\ge a}\{n_{a'b'}\}$. Let

$$
s = \max_{a}\{n_a^{C} - n_{aa}\}. \tag{3.32}
$$

Then we have the following lower bound on δ:

Lemma 3.16.
$$
\delta \ge \frac{rs\gamma}{2}\frac{T}{K^2 n}.
$$

Since $\{C_a\}_{a=1}^{K} \ne \{G_b\}_{b=1}^{K}$, we have $s \ge 1$ and $\delta > 0$. This should come as no surprise: since covariances within groups are larger than those between groups, we should expect larger within-group sum of squares when clusters coincide with the groups.

Define $\delta := \mathbb{E}\Delta > 0$. We should expect $\mathbb{P}(\Delta \le 0)$ to be small. To bound this probability, we use Hanson-Wright inequality in the form introduced by Rudelson and Vershynin[96] :

Lemma 3.17 ((Hanson-Wright inequality [96])). *Let $\varepsilon = (\varepsilon_1, ..., \varepsilon_N)' \in \mathbb{R}^N$ be a random vector with independent, mean zero, sub-Gaussian coordinates. Let A be an $N \times N$ matrix. Then for all $\lambda \ge 0$,*

$$\mathbb{P}(\varepsilon' A \varepsilon \le \mathbb{E}(\varepsilon' A \varepsilon) - \lambda) \le \exp\{-c_1 \min\{\frac{\lambda^2}{||\Sigma^{1/2}\Omega\Sigma^{1/2}||_F^2}, \frac{\lambda}{||\Sigma^{1/2}\Omega\Sigma^{1/2}||_{op}}\}\}$$

where c_1 depends on the largest sub-Gaussian norm among coordinates $\varepsilon_1, ..., \varepsilon_N$.

Using Hanson-Wright inequality with $\lambda = \delta$, we get

$$\mathbb{P}(\Delta = \varepsilon' \Sigma^{1/2}\Omega\Sigma^{1/2}\varepsilon \le 0) \le \exp\{-c_1 \min\{\frac{\delta^2}{||\Sigma^{1/2}\Omega\Sigma^{1/2}||_F^2}, \frac{\delta}{||\Sigma^{1/2}\Omega\Sigma^{1/2}||_{op}}\}\}, \qquad (3.33)$$

where $c_1 > 0$ is some universal constant because coordinates of ε are i.i.d. $N(0,1)$ random variables. Note that if we replace Assumption 3.6 by (joint) sub-Gaussianity of $\{\mathbf{X}_i\}_{i=1}^n$, we have that $\mathbf{U}$ is sub-Gaussian. However, it is not clear if there exists a random vector ε with independent coordinates such that $\mathbf{U} = \Sigma^{1/2}\varepsilon$.

The following two lemmas bound the two norms in (3.33) from above.

Lemma 3.18. $||\Sigma^{1/2}\Omega\Sigma^{1/2}||_{op} \le 2K(M+1)$, *where M is the constant in* (3.22) *of Assumption 3.14.*

Lemma 3.19. $||\Sigma^{1/2}\Omega\Sigma^{1/2}||_F^2 \le K(1 + \frac{2}{r\gamma} + \frac{2M^2}{\gamma})\delta$.

Using Lemma 3.18 and 3.19 we get

$$\mathbb{P}(\Delta \le 0) \le \exp\{-c_2\frac{\delta}{K}\}, \qquad (3.34)$$

where $c_2 = c_2(r, M, \gamma) > 0$ is a constant that depends on r, M and γ. This completes step 1 of the proof.

Step 2: Now, for fixed $\{n_{ab}\}_{a,b=1}^K$, the number of possible partitions $\{C_a\}_{a=1}^K$ is at most

$$\prod_{b=1}^K \binom{n_b^G}{n_{1k}, ..., n_{Kk}} \le \prod_{b=1}^K n^{n_b^G - n_{bb}} = n^{\sum_{a=1}^K (n_a^C - n_{aa})} \le n^{Ks}.$$

Here s as defined in (3.32) is a function of $\{n_{ab}\}_{a,b=1}^K$. Moreover, we have that for fixed $s > 0$, the number of possible partitions $\{C_a\}_{a=1}^K$ is at most

$$\sum_{\{n_{ab}\}_{j,k=1}^K} n^{Ks} \leq (n+1)^{K^2} n^{Ks} \leq n^{2K^2} n^{Ks},$$

as $n_{ab} \in \{0, 1, ..., n\}$ and $1 \leq a, b \leq K$.

Using (3.34), Lemma 3.16 and union bound we get

$$
\begin{aligned}
\mathbb{P}(\Delta \leq 0 \text{ for any } \{C_a\}_{a=1}^K \neq \{G_b\}_{b=1}^K) &\leq \sum_{s=1}^n n^{2K^2} n^{Ks} \exp\{-c_2 \frac{r\gamma s T}{2K^3 n}\} \\
&\leq \sum_{s=1}^n \exp\{2K^2 s(\log n - \frac{c_2 r \gamma}{4K^5} \frac{T}{n})\} \\
&\leq \sum_{s=1}^n \exp\{2K^2 s(1 - Cc_2 r\gamma/4)\log n\} \\
&\leq n \exp\{2K^2(1 - Cc_2 r\gamma/4)\log n\} \\
&\leq \exp\{2K^2(2 - Cc_2 r\gamma/4)\log n\}.
\end{aligned}
\tag{3.35}
$$

Here we used the fact that $T \geq CK^5 n \log n$.

Choose $C = \frac{12}{c_2 r\gamma}$ and the proof is complete. ∎

3.4.2 Proof of Lemma 3.16

First note that $n_b^G \geq rn/K$ by Assumption 3.7. If there exists (a', b', d') such that $n_{a'b'} \geq n_{b'}^G/K$ and $n_{a'd'} \geq n_{d'}^G/K$, then we have

$$\sum_{\substack{b,d=1 \\ b \neq d}}^K \sum_{j=1}^K \frac{n_{ab} n_{ad}}{n_a^C} \geq \frac{n_{a'b'}(n_{a'}^C - n_{a'b'})}{n_{a'}^C} \geq \frac{n_{a'b'} n_{a'd'}}{n_{a'b'} + n_{a'd'}} \geq \frac{rn}{2K^2} \geq \frac{rs}{2K^2}.$$

Here we used the fact that $n_{a'b'}(n_{a'}^C - n_{a'b'})/n_{a'}^C$ is increasing in $n_{a'}^C$ and $n_{a'b'} n_{a'd'}/(n_{a'b'} + n_{a'd'})$ is increasing in both $n_{a'b'}$ and $n_{a'd'}$.

If such (a', b', d') does not exist, we have $n_{ab} < n_b^G/K$ whenever $b > a$. Otherwise let a' be the smallest a such that there exists $b' > a$ with $n_{ab'} \geq n_{b'}^G/K$. Choose that b' so that $n_{a'b'} \geq n_{b'}^G/K$. Moreover, for all $a < a'$, $n_{ab} < n_b^G/K$ whenever $b > a$. Set $a = a'$ to yield $n_{aa'} < n_{a'}^G/K$ for all $a < a'$. Since $\sum_{a=1}^K n_{aa'} = n_{a'}^G$, there must exist $a \geq a'$ such that $n_{aa'} \geq n_{a'}^G/K$. Thus we

have $n_{a'a'} = \max_{a,b \geq a'}\{n_{ab}\} \geq \max_{a \geq a'}\{n_{aa'}\} \geq n'_a/K$, and (a', b', a') satisfies the condition of (a', b', d'), a contradiction.

Furthermore, we have $n_{aa} \geq n_a^G/K$ for all $1 \leq a \leq K$. Otherwise there exists a' such that $n_{a'}^G/K > n_{a'a'} = \max_{a,b \geq a'}\{n_{ab}\} \geq \max_{a \geq a'}\{n_{aa'}\}$. But we have $n_{aa'} < n_{a'}/K$ whenever $a < a'$ and together we get $n_{aa'} < n_{a'}^G/K$ for all a, contradicting the fact that $\sum_{a=1}^K n_{aa'} = n_{a'}^G$.

Let $a' = \text{argmax}_a\{n_a^C - n_{aa}\}$. We have $n_{a'a'} \geq n_{a'}^G/K$ and it follows that

$$\sum_{\substack{b,d=1 \\ b \neq d}}^K \sum_{a=1}^K \frac{n_{ab}n_{ad}}{n_a^C} \geq \frac{n_{a'a'}(n_{a'}^C - n_{a'a'})}{n_{a'}^C} = \frac{sn_{a'a'}}{n_{a'}^C} \geq \frac{sn_{a'a'}}{n} \geq \frac{sn_{a'}^G}{nK} \geq \frac{rs}{K^2}.$$

Lemma 3.16 then follows.

∎

3.4.3 Proof of Lemma 3.18

Since operator norm is sub-multiplicative and $||\Omega||_{op} = 1$, we have

$$||\Sigma^{1/2}\Omega\Sigma^{1/2}||_{op} \leq ||\Sigma^{1/2}||_{op}||\Omega||_{op}||\Sigma^{1/2}||_{op} = ||\Sigma||_{op}.$$

Next, note that we can upper-bound the operator norm by the L_∞ norm. And from (3.22) in Assumption 3.14 we have

$$n\sum_{d=1}^K ||P_{bd}||_\infty \leq M \tag{3.36}$$

for $1 \leq b, d \leq K$. We use this inequality to bound L_∞ norms of components in Σ part by part. For Σ_{bd}^G in (3.25) and (3.26) we have

$$\sum_{d=1}^K ||\Sigma_{bd}^G||_\infty \leq 1 + \sum_{d=1}^K \sqrt{n_b^G n_d^G}||P_{bd}||_\infty \leq 1 + n\sum_{d=1}^K ||P_{bd}||_\infty \leq 1 + M.$$

using $n_b^G, n_d^G \leq n$ and (3.36).

Similarly, for Σ_{ac}^C in (3.27) and (3.28) we have

$$
\begin{aligned}
\sum_{c=1}^{K} ||\Sigma_{ac}^C||_\infty &\leq 1 + \sum_{c=1}^{K} \frac{1}{\sqrt{n_a^C n_c^C}} \sum_{1 \leq b,d \leq K} n_{ab} n_{cd} ||P_{bd}||_\infty \\
&\leq 1 + n \sum_{c=1}^{K} \frac{1}{n_a^C n_c^C} \sum_{1 \leq b,d \leq K} n_{ab} n_c^C ||P_{bd}||_\infty \\
&\quad (\text{using } n_a^C, n_c^C \leq n \text{ and } n_{cd} \leq n_c^C) \\
&= 1 + nK \sum_{b=1}^{K} \frac{n_{ab}}{n_a^C} \sum_{d=1}^{K} ||P_{bd}||_\infty \\
&\leq 1 + KM \sum_{b=1}^{K} \frac{n_{ab}}{n_a^C} = 1 + KM. \quad (\text{by (3.36)}).
\end{aligned}
$$

and for $\tilde{\Sigma}_{ab}$ in (3.29) note that $n_{ab} \leq \min\{n_a^C, n_b^G\} \leq \sqrt{n_a^C n_b^G}$ and $n_b^G \leq n$, we get

$$
\begin{aligned}
||\tilde{\Sigma}_{ab}||_\infty &\leq \frac{n_{ab}}{\sqrt{n_a^C n_b^G}} + \sqrt{\frac{n_b^G}{n_a^C}} \sum_{d=1}^{K} n_{ad} ||P_{bd}||_\infty \\
&\leq 1 + n \sum_{d=1}^{K} ||P_{bd}||_\infty \leq 1 + M.
\end{aligned}
$$

It follows that $\sum_{a=1}^{K} ||\tilde{\Sigma}_{ab}||_\infty \leq K(1+M)$ and

$$
||\Sigma||_{op} \leq K(1+M) + 1 + KM \leq 2K(M+1)
$$

and this proves the lemma.

∎

3.4.4   Proof of Lemma 3.19

First note that

$$
||\Sigma^{1/2} \Omega \Sigma^{1/2}||_F^2 = \text{Tr}(\Sigma^{1/2} \Omega \Sigma \Omega \Sigma^{1/2}) = \text{Tr}(\Omega \Sigma \Omega \Sigma) \tag{3.37}
$$

by properties of traces. And from (3.30) we can compute the trace as

$$
\text{Tr}(\Omega \Sigma \Omega \Sigma) = \sum_{1 \leq b,d \leq K} \text{Tr}((\Sigma_{bd}^G)^2) + \sum_{1 \leq a,c \leq K} \text{Tr}((\Sigma_{ac}^C)^2) - 2 \sum_{1 \leq a,b \leq K} \text{Tr}(\tilde{\Sigma}_{ab}^2).
$$

Divide this trace into two parts and write $\text{Tr}(\Omega\Sigma\Omega\Sigma) = I_1 + I_2$, where

$$I_1 := \sum_{1 \le b,d \le K} \text{Tr}((\Sigma_{bd}^G)^2) - \sum_{1 \le a,b \le K} \text{Tr}(\tilde{\Sigma}_{ab}^2)$$

and

$$I_2 := \sum_{1 \le a,c \le K} \text{Tr}((\Sigma_{ac}^C)^2) - \sum_{1 \le a,b \le K} \text{Tr}(\tilde{\Sigma}_{ab}^2).$$

Plugging in (3.25), (3.26) and (3.29) we get

$$I_1 = \sum_{b=1}^{K} \text{Tr}(I_T + 2n_b^G P_{bb}) + \sum_{1 \le b,d \le K} n_b^G n_d^G \text{Tr}(P_{bd}^2)$$
$$- \sum_{1 \le a,b \le K} \text{Tr}\left(\frac{n_{ab}^2}{n_a^C n_b^G} I_T + \frac{2n_{ab}}{n_a^C} \sum_{d=1}^{K} n_{ad} P_{bd}\right) - \sum_{1 \le a,b \le K} \frac{n_b^G}{n_a^C} \text{Tr}\left(\left(\sum_{d=1}^{K} n_{ad} P_{bd}\right)^2\right).$$

Divide it further into $I_1 = J_{11} + J_{12}$, where the linear term

$$J_{11} := \sum_{b=1}^{K} \text{Tr}(I_T + 2n_b^G P_{bb}) - \sum_{1 \le a,b \le K} \text{Tr}\left(\frac{n_{ab}^2}{n_a^C n_b^G} I_T + \frac{2n_{ab}}{n_a^C} \sum_{d=1}^{K} n_{ad} P_{bd}\right)$$

and the quadratic term

$$J_{12} := \sum_{1 \le b,d \le K} n_b^G n_d^G \text{Tr}(P_{bd}^2) - \sum_{1 \le a,b \le K} \frac{n_b^G}{n_a^C} \text{Tr}\left(\left(\sum_{d=1}^{K} n_{ad} P_{bd}\right)^2\right).$$

For the linear term we have

$$J_{11} := \sum_{b=1}^{K} \text{Tr}(I_T + 2n_b^G P_{bb}) - \sum_{1 \le a,b \le K} \text{Tr}\left(\frac{n_{ab}^2}{n_a^C n_b^G} I_T + \frac{2n_{ab}}{n_a^C} \sum_{d=1}^{K} n_{ad} P_{bd}\right)$$

$$= T \sum_{1 \le a,b \le K} \left(\frac{n_{ab}}{n_b^G} - \frac{n_{ab}^2}{n_a^C n_b^G}\right) + \sum_{1 \le a,b,d \le K} \frac{n_{ab} n_{ad} \text{Tr}(P_{bb} + P_{dd} - 2P_{bd})}{n_a^C}$$

$$\left(\text{using } \text{Tr}(I_T) = T \text{ and } n_b^G = \sum_{a=1}^{K} n_{ab}\right)$$

$$= T\left(\sum_{a=1}^{K} \sum_{\substack{b,d=1 \\ b \ne d}}^{K} \frac{n_{ab} n_{ad}}{n_a^C n_b^G} + \sum_{a=1}^{K} \sum_{\substack{b,d=1 \\ b \ne d}}^{K} \frac{n_{ab} n_{ad}(\rho_{bb}(0) + \rho_{dd}(0) - 2\rho_{bd}(0))}{n_a^C}\right)$$

$$\left(\text{using } n_a^C - n_{ab} = \sum_{d \ne b} n_{ad}, \ \text{Tr}(P_{bd}) = T\rho_{bd}\right)$$

From (3.31) and using Assumption 3.7 we know that

$$T\sum_{a=1}^{K}\sum_{\substack{b,d=1\\b\neq d}}^{K}\frac{n_{ab}n_{ad}}{n_a^C n_b^G}\leq\frac{KT}{rn}\sum_{a=1}^{K}\sum_{\substack{b,d=1\\b\neq d}}^{K}\frac{n_{ab}n_{ad}}{n_a^C}\leq\frac{K}{r\gamma}\delta$$

and

$$T\sum_{a=1}^{K}\sum_{\substack{b,d=1\\b\neq d}}^{K}\frac{n_{ab}n_{ad}(\rho_{bb}(0)+\rho_{dd}(0)-2\rho_{bd}(0))}{n_a^C}=2\delta.$$

Combine these terms we get $J_{11}\leq(2+\frac{K}{r\gamma})\delta$.

For the quadratic term we have

$$
\begin{aligned}
J_{12}&:=\sum_{1\leq b,d\leq K}n_b^G n_d^G\,\mathrm{Tr}(P_{bd}^2)-\sum_{1\leq a,b\leq K}\frac{n_b^G}{n_a^C}\,\mathrm{Tr}\Big((\sum_{d=1}^{K}n_{ad}P_{bd})^2\Big)\\
&\leq\sum_{1\leq b,d\leq K}n_b^G n_d^G\,\mathrm{Tr}(P_{bd}^2)-\sum_{1\leq a,b\leq K}\frac{n_b^G}{n_a^C}\sum_{d=1}^{K}n_{ad}^2\,\mathrm{Tr}(P_{bd}^2)\quad\text{(by (3.21) in Assumption 3.14)}\\
&=\sum_{1\leq a,b,d\leq K}\frac{n_a^C n_b^G n_{ad}-n_b^G n_{ad}^2}{n_a^C}\,\mathrm{Tr}(P_{bd}^2)\quad\text{(using }n_d^G=\sum_{a=1}^{K}n_{ad}).
\end{aligned}
$$

Since $n^2\,\mathrm{Tr}(P_{bd}^2)\leq Tn^2(\rho_{bd}^2(0)+2\sum_{t=1}^{\infty}\rho_{bd}^2(t))\leq TM^2$ by (3.18) and (3.22) in Assumption 3.14, it follows that $\mathrm{Tr}(P_{bd}^2)\leq TM^2/n^2$ and

$$
\begin{aligned}
J_{12}&\leq\frac{TM^2}{n^2}\sum_{1\leq a,b,d\leq K}\frac{n_a^C n_b^G n_{ad}-n_b^G n_{ad}^2}{n_a^C}\\
&=\frac{TKM^2}{n}\sum_{1\leq a,d\leq K}\frac{n_a^C n_{ad}-n_{ad}^2}{n_a^C}\quad\text{(using }n_b^G\leq n)\\
&\leq\frac{TKM^2}{n}\sum_{a=1}^{K}\sum_{\substack{b,d=1\\b\neq d}}^{K}\frac{n_{ab}n_{ad}}{n_a^C}\quad\text{(using }n_a^C-n_{ad}=\sum_{b\neq d}n_{ab})\\
&\leq\frac{KM^2}{\gamma}\delta.\quad\text{(by (3.31))}
\end{aligned}
$$

Similarly, for I_2 we can plug in (3.27), (3.28) and (3.29) and get

$$
\begin{aligned}
I_2&=\sum_{a=1}^{K}\mathrm{Tr}(I_T+\frac{2}{n_a^C}\sum_{1\leq b,d\leq K}n_{ab}n_{ad}P_{bd})+\sum_{1\leq a,c\leq K}n_a^C n_c^C\,\mathrm{Tr}\Big((\frac{1}{n_a^C n_c^C}\sum_{1\leq b,d\leq K}n_{ab}n_{cd}P_{bd})^2\Big)\\
&\quad-\sum_{1\leq a,b\leq K}\mathrm{Tr}(\frac{n_{ab}^2}{n_a^C n_b^G}I_T+\frac{2n_{ab}}{n_a^C}\sum_{d=1}^{K}n_{ad}P_{bd})-\sum_{1\leq a,b\leq K}\frac{n_b^G}{n_a^C}\,\mathrm{Tr}\Big((\sum_{d=1}^{K}n_{ad}P_{bd})^2\Big)\\
&=J_{21}+J_{22},
\end{aligned}
$$

where the linear term

$$J_{21} := \sum_{a=1}^{K} \text{Tr}(I_T + \frac{2}{n_a^C} \sum_{1 \leq b,d \leq K} n_{ab}n_{ad}P_{bd}) - \sum_{1 \leq a,b \leq K} \text{Tr}(\frac{n_{ab}^2}{n_a^C n_b^G} I_T + \frac{2n_{ab}}{n_a^C} \sum_{d=1}^{K} n_{ad}P_{bd})$$

and the quadratic term

$$J_{22} := \sum_{1 \leq a,c \leq K} n_a^C n_c^C \, \text{Tr}\left((\frac{1}{n_a^C n_c^C} \sum_{1 \leq b,d \leq K} n_{ab}n_{cd}P_{bd})^2\right) - \sum_{1 \leq a,b \leq K} \frac{n_b^G}{n_a^C} \, \text{Tr}\left((\sum_{d=1}^{K} n_{ad}P_{bd})^2\right).$$

The linear term J_{21} can be bounded similarly as J_{11}:

$$\begin{aligned} J_{21} =&\, T \sum_{1 \leq a,b \leq K} (\frac{n_{ab}}{n_b^G} - \frac{n_{ab}^2}{n_a^C n_b^G}) + \sum_{1 \leq a,b,d \leq K} \frac{n_{ab}n_{ad}\,\text{Tr}(2P_{bd} - 2P_{bd})}{n_a^C} \\ =&\, T \sum_{a=1}^{K} \sum_{\substack{b,d=1 \\ b \neq d}}^{K} \frac{n_{ab}n_{ad}}{n_a^C n_b^G} \\ \leq&\, \frac{K}{r\gamma} \delta. \end{aligned}$$

For the quadratic term we reverse the first term back to Frobenius norm and get

$$\begin{aligned} &\sum_{1 \leq a,c \leq K} n_a^C n_c^C \, \text{Tr}\left((\frac{1}{n_a^C n_c^C} \sum_{1 \leq b,d \leq K} n_{ab}n_{cd}P_{bd})^2\right) \\ =& \sum_{1 \leq a,c \leq K} n_a^C n_c^C \|\frac{1}{n_a^C n_c^C} \sum_{1 \leq b,d \leq K} n_{ab}n_{cd}P_{bd}\|_F^2 \\ \leq& \sum_{1 \leq a,c,b,d \leq K} n_{ab}n_{cd}\|P_{bd}\|_F^2 \quad \text{(by convexity of squared Frobenius norm)} \\ =& \sum_{1 \leq b,d \leq K} n_b^G n_d^G \, \text{Tr}(P_{bd}^2). \end{aligned}$$

It readily follows that $J_{22} \leq J_{12} \leq \frac{KM^2}{\gamma}\delta$.

Collect all the above inequalities we get

$$\text{Tr}(\Omega\Sigma\Omega\Sigma) = J_{11} + J_{12} + J_{21} + J_{22} \leq 2(1 + \frac{K}{r\gamma} + \frac{KM^2}{\gamma})\delta \leq K(1 + \frac{2}{r\gamma} + \frac{2M^2}{\gamma})\delta.$$

Lemma 3.19 then follows from (3.37).

∎

3.4.5 Proof of Proposition 3.12

Since (3.16) implies $||A||_\infty \le 1 - \tau$, it follows that $||A||_{op} \le 1 - \tau$, and the process is stationary (in time). For ease of notation we shall follow Section 3.3.3 and use $\rho_{bd}(|t_1 - t_2|)$, $\xi_{ij}(|t_1 - t_2|)$ instead of $\rho_{bd}(t_1, t_2)$, $\xi_{ij}(t_1, t_2)$.

From (3.14) we get

$$\text{Cov}(\mathbf{X}(t_1), \mathbf{X}(t_2)) = A^{|t_1 - t_2|} \sum_{q=0}^{\infty} A^{2q}.$$

This readily implies Assumption 3.8. Specifically, for $i \in G_b$ and $j \in G_d$, $\rho_{bd}(0)$ is equal to the $(i,j)^{th}$ entry of $\sum_{q=1}^{\infty} A^{2q}$ and $\rho_{bd}(t)$ is equal to the $(i,j)^{th}$ entry of $A^t \sum_{q=0}^{\infty} A^{2q}$ for $t > 0$. $\xi_{ij}(t) = 0$ for all $1 \le i, j \le n$.

We now proceed to examine Assumption 3.9. All of these matrices have nonnegative entries by (3.15) so (3.9) holds. To verify (3.10) (or the equivalent (3.19)), we need the following simple lemma:

Lemma 3.20. *For all positive integer q, we have*

$$||A^q||_{max} \le \frac{(1-\tau)^q}{n} \tag{3.38}$$

Proof: From (3.16) we have $p_{bd} \le \frac{1-\tau}{n}$ so (3.38) holds for $q = 1$.

Suppose Lemma 3.20 holds for q. Then we have for the $(i,j)^{th}$ entry of A^{q+1}

$$(A^{q+1})_{i,j} = \sum_{k=1}^{n} A_{i,k}(A^q)_{k,j} \le ||A||_\infty ||A^q||_{max} \le (1-\tau)\frac{(1-\tau)^j}{n} = \frac{(1-\tau)^{q+1}}{n}.$$

It follows that (3.38) holds for all positive integer q by induction.

■

Going back to (3.10), Using Lemma 3.20 we have

$$n \sum_{d=1}^{K} (\rho_{bd}(0) + 2 \sum_{t=1}^{\infty} \rho_{bd}(t)) \le n \sum_{d=1}^{K} (\sum_{q=1}^{\infty} \frac{(1-\tau)^{2q}}{n} + 2 \sum_{t=1}^{\infty} \sum_{q=0}^{\infty} \frac{(1-\tau)^{2q+t}}{n})$$

$$= K \sum_{q=0}^{\infty} (1-\tau)^{2q}((1-\tau)^2 + 2 \sum_{t=1}^{\infty} (1-\tau)^t)$$

$$\le 2K \sum_{q=0}^{\infty} (1-\tau)^q \sum_{t=0}^{\infty} (1-\tau)^t = \frac{2K}{\tau^2}.$$

Therefore (3.10) and (3.19) holds with $M = 2K/\tau^2$.

To verify (3.11), note that for any symmetric $n \times n$ matrix $B = (b_{ij})$, we have

$$(B^2)_{i,i} + (B^2)_{j,j} - 2(B^2)_{i,j} = \sum_{k=1}^{n}(b_{ik}^2 + b_{jk}^2 - b_{ik}b_{jk}) = \sum_{k=1}^{n}(b_{ik} - b_{jk})^2 \geq 0.$$

Recall that for $i \in G_b$ and $j \in G_d$, $\rho_{bd}(0)$ is equal to the $(i,j)^{th}$ entry of $\sum_{q=1}^{\infty} A^{2q}$. Therefore

$$\rho_{bb}(0) + \rho_{dd}(0) - 2\rho_{bd}(0) \geq (A^2)_{i,i} + (A^2)_{j,j} - 2(A^2)_{i,j} = \sum_{k=1}^{K} n_k^G (p_{bk} - p_{dk})^2$$

$$\geq n_b^G(p_{bb} - p_{bd})^2 + n_d^G(p_{dd} - p_{bd})^2$$

$$\geq \frac{2rn}{K}(\frac{\eta}{n})^2 = \frac{2r\eta^2}{Kn}$$

when $b \neq d$. Here the last inequality follows from Assumption 3.7. Thus we have that (3.11) holds with $\gamma = r\eta^2/K$.

Finally (3.12) holds trivially since $\xi_{ij}(t) = 0$ for all $1 \leq i, j \leq n$.

3.5   Other algorithms

Since our goal is to recover the clusters from block structured covariance structure, it makes sense to consider algorithms based on sample covariance matrix $\hat{\Sigma}_{\mathbf{X}}$. And in this section we shall describe three different approaches.

The first approach simply applies K-means clustering to rows of the sample covariance matrix $\hat{\Sigma}^{\mathbf{X}}$. Specifically, given the number of clusters K, it aims to find the partition $\{1, 2, ..., n\} = \cup_{a=1}^{K} C_a$ and K n-dimensional vector $\mathcal{O}_a = (\mathcal{O}_{a,1}, ..., \mathcal{O}_{a,n})$ that minimizes:

$$WCSS(\{C_a\}_{a=1}^{K}) := \sum_{a=1}^{K} \sum_{i \in C_a} \sum_{j=1}^{n} (\hat{\Sigma}_{ij}^{\mathbf{X}} - \mathcal{O}_{a,j})^2. \tag{3.39}$$

The second approach applies spectral clustering to $\hat{\Sigma}^{\mathbf{X}}$, treating it as a weighted graph. Notice however that entries of $\hat{\Sigma}^{\mathbf{X}}$ can be negative in general so a symmetric normalized Laplacian does not always exists. Specifically, we compute the eigenvectors of $\hat{\Sigma}^{\mathbf{X}}$ corresponding to its K eigenvalues that are largest in absolute value, and combine them into a $n \times K$ matrix $\hat{U}$. Then we normalize the rows of $\hat{U}$ so that they have unit Euclidean norms, and apply K-means clustering.

The third approach fits a blockwise constant approximation to $\hat{\Sigma}^{\mathbf{X}}$ (with permutation on the indices). Specifically, it aims to find the partition $\{1, 2, ..., n\} = \cup_{a=1}^{K} C_a$ and $K \times K$ matrix

M that minimizes:

$$l(\{C_a\}_{a=1}^K) := \sum_{a=1}^K \sum_{b=1}^K \sum_{i \in C_a} \sum_{j \in C_b} (\hat{\Sigma}_{ij}^{\mathbf{X}} - M_{ab})^2. \qquad (3.40)$$

Similar to Lloyd's algorithm, we alternative between two steps. In the assignment step, each $i \in \{1, 2, ..., n\}$ is assigned to the cluster that minimizes $l(\{C_a\}_{a=1}^K)$. This is done in a greedy way, $i.e.$ cluster assignment is updated sequentially for $i \in \{1, 2, ..., n\}$, treating membership of other indices and M as fixed. Note that for each i we need to compute the objective function (3.40) K times, each time summing over $2n-1$ terms so each step has time complexity $\Theta(Kn^2)$. In the update step, M is updated as the entry-wise average within and between clusters:

$$\hat{M}_{ab} = \frac{1}{|C_a||C_b|} \sum_{i \in C_a} \sum_{j \in C_b} \hat{\Sigma}_{ij}^{\mathbf{X}}.$$

For initialization we apply naive K-means ($i.e.$ Lloyd's algorithm) to rows of $\hat{\Sigma}^{\mathbf{X}}$. This procedure can be viewed as a biclustering version of Lloyd's algorithm, and finds a local minimum of the objective function (3.40). We shall refer to this algorithm as "block k-means".

3.6   Simulation

For each setting considered below, clustering error rate is measured by the proportion of mis-clusterd indices. 100 independent simulation runs are made and the error rates shown are averages.

3.6.1 G-block covariance model

Consider a G-block covariance model as described in (3.6):

$$X_i(t) = M_a(t) + \varepsilon_i(t), \quad \text{for all } i \in G_a \text{ and } 1 \le t \le T,$$

where $M(t) = (M_1(t), M_2(t))$ is a 2-dimensional Gaussian vector with zero mean and covariance matrix $\begin{pmatrix} 4 & 1 \\ 1 & 2 \end{pmatrix}$, independent of the i.i.d. standard normal white noises $\varepsilon_i(t)$. Here $n = 100$ and $G_1 = \{1, 2, ..., 70\}$ and $G_2 = \{71, 72, ..., 100\}$ are the two blocks. Moreover, $M(t)$ are independent across time therefore so are $\{X_i(t)\}$.

Figure 3.1 shows the error rate as T increases. The four algorithms considered here are kmeans++ and the three sample covariance based methods described in Section 3.5. We can see that kmeans++ on the original data works reasonably well. Performance of kmeans++ and
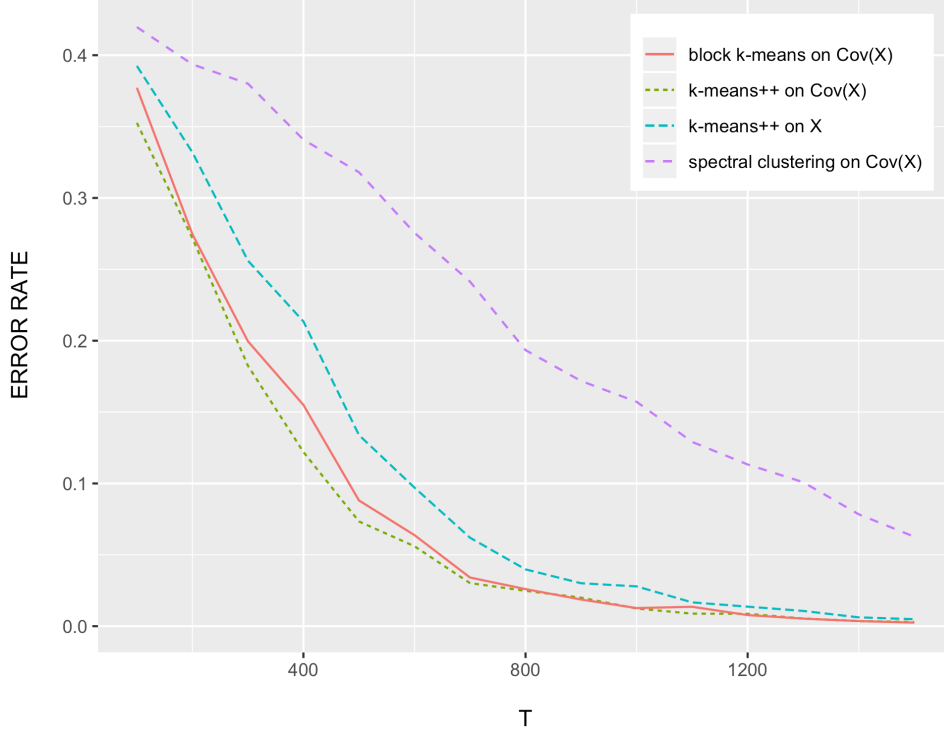
Figure 3.1: Clustering error rate for G-block covariance model

block k-means on sample covariance are comparably the best. However, we remark that perfect recovery is not achieved up to $T = 1500$. Spectral clustering does not seem to work very well in this setting, when $M(t)$ is heteroscedastic. In cases where $M_a(t)$ have similar variances, spectral clustering works reasonably well and is only slightly worse than kmeans++ and block k-means.

3.6.2 G-block covariance model with auto-correlation

Write $\eta_i(t)$ for $X_i(t)$ defined in the previous setting. Consider a G-block covariance model with auto-correlation as follows: $X_1(t) = \eta_1(t)$ and $X_i(t) = 0.5X_i(t-1) + \eta_i(t)$ for $2 \leq t \leq T$. Figure 3.2 shows the error rate as T increases for the same four algorithms. The overall performance degrades in the presence of auto-correlation, especially for spectral clustering.

3.6.3 Stochastic block network VAR model

Consider a stochastic block network VAR model as described in Section 3.1.5: $\mathbf{X}(1) = \varepsilon(1)$ and

$$\mathbf{X}(t) = \phi D^{-1/2} A D^{-1/2} \mathbf{X}(t-1) + \varepsilon(t), \quad \text{for } 2 \leq t \leq T.$$

Here $\{\mathbf{X}(t)\}$ is a n-dimensional stationary process and $\{\varepsilon(t)\}$ have i.i.d. standard normal coordinates. A is an adjacency matrix generated from an undirected unweighted stochastic
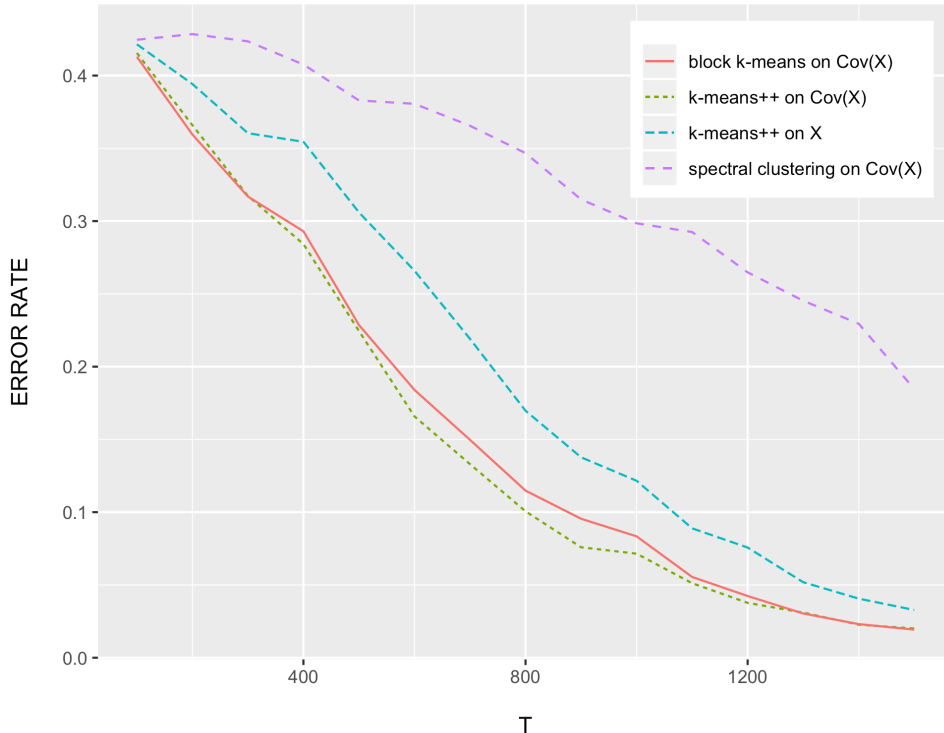
Figure 3.2: Clustering error rate for G-block covariance model with auto-correlation

block model without vertex-specific weights. In our simulation, $n = 100$ and $G_1 = \{1, 2, ..., 60\}$ and $G_2 = \{61, 62, ..., 100\}$ are the two blocks. The edge probability matrix is $\left(\begin{smallmatrix} .8 & .2 \\ .2 & .8 \end{smallmatrix}\right)$, that is the probability of an edge within communities is 0.8 and that between communities is 0.2. D is the diagonal degree matrix with $D_{ii} = \sum_{j=1}^{n} A_{ij}$ so that $D^{-1/2} A D^{-1/2}$ is the symmetric normalized Laplacian. $\phi = 0.9$, and that ensures stationarity of the process.

Figure 3.3 shows the error rate as T increases. We consider three more algorithms here, all based on the ordinary least squares (OLS) estimate of coefficient matrix $\phi D^{-1/2} A D^{-1/2}$, which we shall denote by $\hat{\Phi}_{OLS}$. The three algorithms applies spectral clustering, k-means++ and block k-means to $\hat{\Phi}_{OLS}$, respectively. Note that spectral clustering on $\hat{\Phi}_{OLS}$ is exactly the VAR Blockbuster algorithm in [47]. We can see that all three algorithms are comparable and perform much better than the four generic algorithms. This should not come as a surprise as these algorithms explicitly exploit the VAR model assumption. Among the four generic algorithms considered before, k-means++ on sample covariance matrix performs relatively well compared to the others. However, k-means++ on the original data can achieve a comparable recovery rate for large T. We remark here that for $T = 2000$ the average error rates for algorithms based on $\hat{\Phi}_{OLS}$ are less than 0.002 so that in most cases the blocks are recovered perfectly.
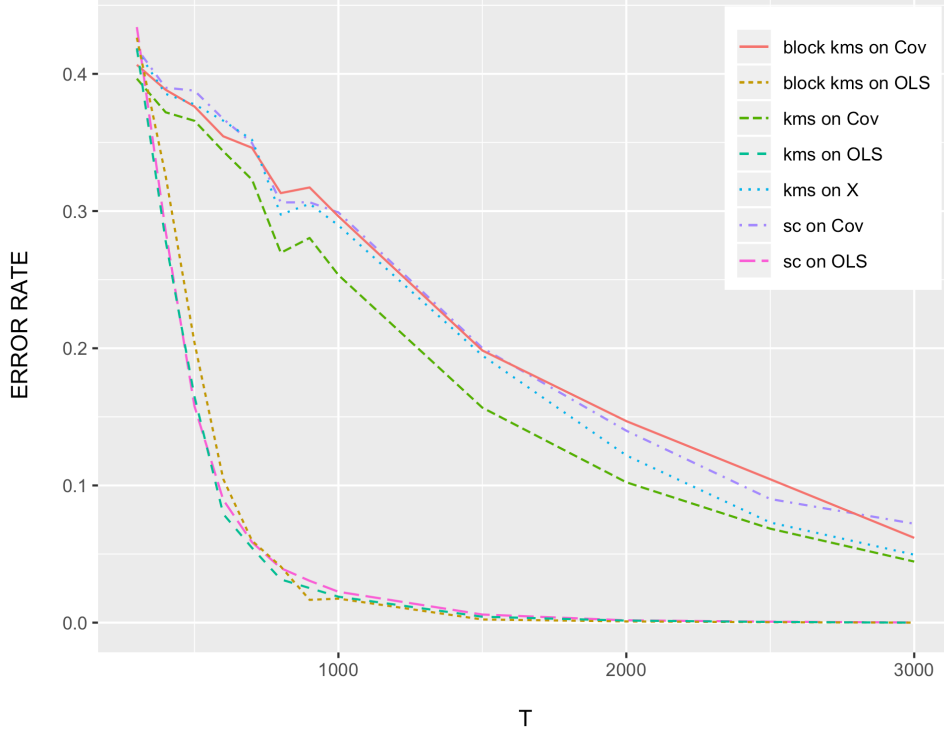
Figure 3.3: Clustering error rate for stochastic block network VAR model. kms stands for k-means and sc for spectral clustering.

3.6.4 SBM-driven VAR model

Consider an unnormalized version of stochastic block network VAR model where the symmetric normalized Laplacian is replaced by the scaled adjacency matrix:

$$\mathbf{X}(t) = \phi \frac{A}{n} \mathbf{X}(t-1) + \varepsilon(t).$$

Here the factor n is required to ensure stationarity of the process, since degrees are of order n in dense graphs. We set $\phi = 1.5$ and the other parameters are same as in the previous stochastic block network VAR model setting. Figure 3.4 shows the error rate as T increases.

We can see that OLS based algorithms still perform significantly better than generic algorithms that do not assume a VAR model. However, spectral clustering suffers in this setting, presumably as an effect of unnormalized adjacency matrix on the eigenstructure. Moreover, the discrepancy between performance of k-means++ on sample covariance and other generic algorithms becomes larger. We remark here that for $T = 3000$ the average error rates for k-means++ and block k-means applied to OLS are less than 0.001 so that in most cases the blocks are recovered perfectly.
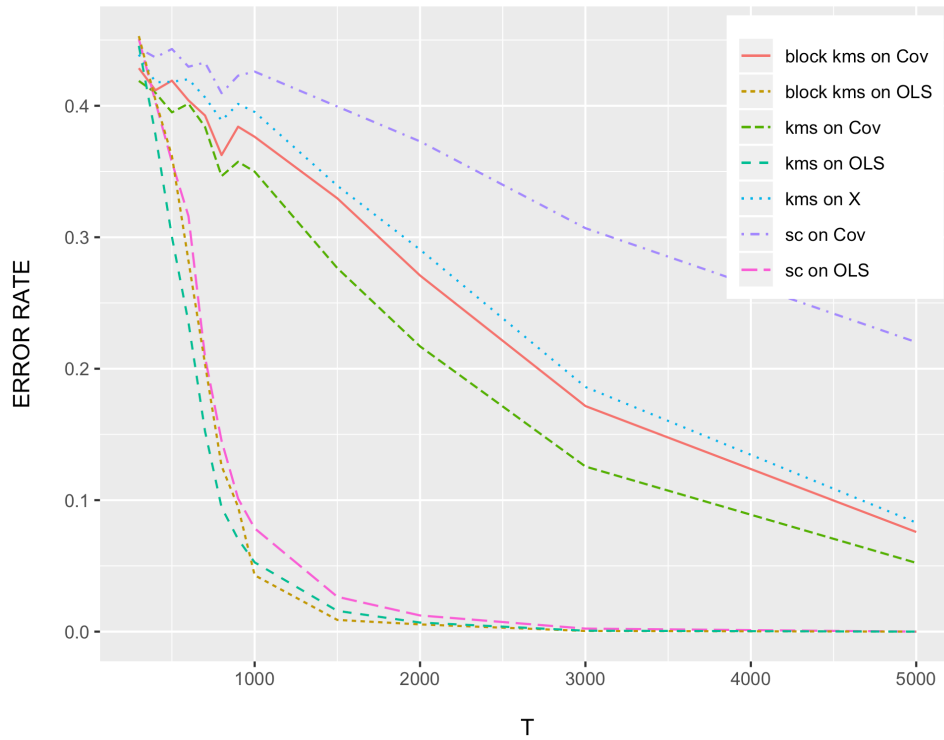
Figure 3.4: Clustering error rate for SBM-driven VAR model. kms stands for k-means and sc for spectral clustering.

3.6.5   Summary

Overall, k-means++ (applied directly to the data) performs reasonably well, but is outperformed by k-means++ applied to the sample covariance matrix. However, in cases where the processes follow an underlying VAR model with coefficient matrices specified by community structure, clustering algorithms should be applied to estimate of coefficient matrices instead.

CHAPTER 4
Change point detection in SBM-driven VAR model

In time series analysis, parameters governing the underlying model are usually assumed to be constant across time. However, in many applications these parameters are subject to abrupt changes at unknown time points. Examples include quality control and fault detection in industrial processes, and automatic segmentation of a signal into stationary segments in signal processing. Change point detection addresses this problem by allowing model parameters to vary across time segments, and aims to estimate the time points where changes occur.

In this chapter, we consider change point detection in the SBM-driven VAR model described in Chapter 3. Specifically, let $\mathbf{X}(t) = (X_1(t), X_2(t), ..., X_n(t))'$ be a n-dimensional vector, with dynamics given by the following VAR(p) model:

$$\mathbf{X}(t) = \sum_{l=1}^{p} \Phi_l^{(j)} \mathbf{X}(t-l) + \varepsilon(t), \quad t = T_{j-1}^0 + 1, ..., T_j^0, \tag{4.1}$$

where $\Phi_l^{(j)}$ ($j = 1, ..., m$ and $l = 1, ..., p$) are $n \times n$ coefficient matrices, and $1 = T_0^0 < T_1^0 < ... < T_m^0 < T_{m+1}^0 = T$ are the $m \geq 1$ change points. Moreover, $\Phi_l^{(j)} = \phi_l^{(j)} A_l^{(j)}/n$ where $A_l^{(j)}$ is an adjacency matrix generated from a dense unweighted stochastic block model without vertex-specific weights, independent of $\{\varepsilon(t)\}$, and $\phi_l^{(j)} > 0$ is called the stationarity parameter. Recall that factor n is required here to ensure stationarity of the process, since degrees are of order n in dense graphs. Parameters of the underlying stochastic block model, including both the block partition and community-specific edge probabilities, are allowed to change across time segments, but are the same for different l within each segment. Our goal is to estimate the change points $T_1^0 < ... < T_m^0$, given the observations $\mathbf{X}(t)$ ($t = 1, ..., T$) and the number of change points m. More precisely, we want to understand what is the minimum number of observations T needed to detect change points in underlying networks of size n.

Closely related to our problem is the change point detection in general VAR model with changing coefficients, also known as structural breaks in the econometrics literature. However, dimension of the process n is usually treated as a constant, and asymptotic analysis mostly focus

90

on consistency of estimated change points as T goes to infinity. While it is true that changes in a single coefficient can be hard to detect, things should be easier when groups of coefficients are changing in a structural way as in our case. To exploit that community structure, we need to either know the communities explicitly, or extract that information from observations, *e.g.* through community detection.

There has also been some work done in the temporal network literature on dynamic communities. Most work in this area studies a deterministic sequence of evolving networks [3][101], with a focus on social networks [84][99][63][104]. Temporal networks with underlying probabilistic model are less explored, and typically assume a static network and independent snapshots. Specifically, static network has a fixed size so its evolution across time is restricted compared to dynamic network where size of the network is allowed to change across time, *e.g.* a growing social network. Moreover, observations at different time points, called snapshots, are taken to be independent, which makes the analysis even simpler. In most cases, statistics are computed for each snapshot, and traditional change point detection techniques are applied. As an example, consider the sequence of adjacency matrices in our model, which are independently generated from stochastic block model with changing communities and edge probabilities. Recall that communities can be perfectly recovered, and model parameters consistently estimated, from just one realization of a dense stochastic block model. Based on that, one can show that both change points and model parameters within each time segment can be consistently estimated as well, using the standard least squares criterion [16]. On the other hand, very little work has been done to model dynamic networks under a probabilistic framework. For an example, see [15] for a growing linear preferential attachment graph with changing parameter. In our model, the underlying network is static and independently generated across different time segments. However, we do not has access to the underlying network directly, and estimates of the adjacency matrix are noisy and correlated across time, which makes the two problems connected but different.

The rest of this chapter is organized as follows: Section 4.1 gives an overview on related literature. Section 4.2 connects our model to structural breaks when communities are fixed and known. Section 4.3 introduces our proposed algorithm. And Section 4.4 contains numerical experiments and discussion.

4.1 Literature review

4.1.1 Change point detection

We shall not dwell on the general change point detection problem, as any overview of this vast literature, even a brief one, will amount to a chapter on its own. Instead, we refer the interested readers to [14] for an exhaustive overview of classic literature, and [20][97][5][102][8] for more recent results. Nevertheless, we will give a short note on the typical consistency results that can be found in asymptotic analysis of change point detection, *e.g.* [28][25][87].

Write $1 < T_1 < ... < T_m < T$ for the m change points, and $\hat{T}_1 < ... < \hat{T}_m$ their estimates. The usual notion of consistency as T goes to infinity would be $\hat{T}_j \xrightarrow{P} T_j$ for $1 \leq j \leq m$. Since both $\hat{T}_j$ and T_j are integer-valued, this is equivalent to $\lim_{T \to \infty} \mathbb{P}(\hat{T}_j = T_j) = 1$. As can be expected, this is not achievable in most nontrivial cases, and consistency in estimation is defined in terms of T_j/T instead, that is, the relative location of change points. Alternatively, write $T_j = \lfloor T\alpha_j \rfloor$ and consider consistency of α_j. In many cases, however, one can hope for a stronger result of the form $\hat{T}_j - T_j = O(1)$, that is, for any $\varepsilon > 0$, there exists $M > 0$ such that $\mathbb{P}(|\hat{T}_j - T_j| > M) < \varepsilon$ for large T.

4.1.2 Structural breaks

In this section we give a brief introduction to structural breaks in multivariate regression, for which vector regressive model is a special case. Our focus here is on key assumptions needed for consistent estimation of break dates. Interested readers are directed to [86] for a more detailed review from the viewpoint of econometric applications, which notably addresses the interplay between structural breaks, unit root and long range dependence.

For simplicity we present the framework in univariate case only, and note that the same result holds in multivariate case, with properly extended assumptions. Consider the following multiple linear regression with $m \geq 1$ breaks:

$$y_t = x_t'\beta + z_t'\lambda_j + \varepsilon_t, \quad t = T_{j-1}^0 + 1, ..., T_j^0.$$

where $1 = T_0^0 < T_1^0 < ... < T_m^0 < T_{m+1}^0 = T$ are the break dates. Here y_t denote the observed dependent variable at time t, x_t and z_t are covariates of dimension p and q, β and λ_j are the vectors of corresponding coefficients, and ε_t is the noise. For a given partition of time into

segments $1 = T_0 < T_1 < ... < T_m < T_{m+1} = T$, denote by $\hat{\beta}_j$ and $\hat{\lambda}_j$ the least squares estimates of β and λ_j based on observations in $t \in (T_{j-1}, T_j]$. Define the least squares loss function:

$$l(T_1, ..., T_m) := \sum_{j=1}^{m+1} \sum_{t=T_{j-1}+1}^{T_j} (y_t - x_t'\hat{\beta}_j - z_t'\hat{\lambda}_j)^2.$$

The estimated break dates are given by

$$(\hat{T}_1, ..., \hat{T}_m) = \underset{1 < T_1 < ... < T_m < T}{\operatorname{argmin}} l(T_1, ..., T_m). \tag{4.2}$$

Then under suitable assumptions $\hat{T}_j - T_j^0 = O(1)$ as $T \to \infty$.

While some assumptions like invertibility are technical, others are crucial as they limit the range of models the results can be applied to. Next we proceed to examine these key assumptions, as stated in [86], which are based on [12] and [90].

Assumption 4.1. *Denote $u_t = (x_t', z_t')'$. For each $1 \le j \le m+1$ and $l_j = T_j^0 - T_{j-1}^0$, $(1/l_j) \sum_{t=T_{j-1}^0+1}^{T_{j-1}^0 + \lfloor l_j v \rfloor} \xrightarrow{P} Q_j(v)$ uniformly in $v \in [0,1]$, where $Q_j(v)$ is a deterministic positive definite matrix.*

This assumption requires the covariates to be covariance-stationary processes, *e.g.* as in a stationary VAR model.

Assumption 4.2. *Denote by $||X||_r = (\sum_i \sum_j E|X_{ij}|^r)^{1/r}$ the L_r $(r \ge 1)$ norm of a random matrix X. With some increasing σ-field $\{\mathcal{F}_t\}$, assume that $\{u_t \varepsilon_t, \mathcal{F}_t\}$ forms a L^r-mixingale sequence with $r = 4 + \delta$ for some $\delta > 0$. That is, there exist nonnegative constants $\{c_t : t \ge 1\}$ and $\{\psi_j : j \ge 0\}$ with $\psi_j \downarrow 0$ as $j \to \infty$, such that for all $t \ge 1$ and $j \ge 0$: (a) $||\mathbb{E}(u_t\varepsilon_t|\mathcal{F}_{t-j})||_r \le c_t\psi_j$, (b) $||u_t\varepsilon_t - \mathbb{E}(u_t\varepsilon_t|\mathcal{F}_{t+j})||_r \le c_t\psi_{j+1}$, (c) $\sup_t c_t < \infty$, (d) $\sum_{j=0}^{\infty} j^{1+\kappa}\psi_j < \infty$ for some $\kappa > 0$, (e) $\sup_t ||z_t||_r < \infty$ and $\sup_t ||\varepsilon_t||_r < \infty$.*

This set of assumptions is needed to establish certain functional central limit theorem that plays an important role in the proofs. It allows for a wide range of correlation (*e.g.* lagged dependent variables) and heterogeneity in covariates and errors.

Assumption 4.3. *$T_j^0 = \lfloor T\alpha_j^0 \rfloor$ where $0 = \alpha_0 < \alpha_1^0 < ... < \alpha_m^0 < \alpha_{m+1}^0 = 1$ and $\alpha_{j+1}^0 - \alpha_j^0 \ge \varepsilon$ for some (small) $\varepsilon > 0$.*

This assumption specifies the minimum separation between break dates, which is standard for offline change point detection. Note that it does not make sense for online detection where

more and more observations are collected chronologically so that only the last time segment is increasing.

4.1.3 Factor model

In this section we relate our SBM-driven VAR model to dynamic factor model (DFM) and show that they are similar yet different. Let $\mathbf{X}(t) = (X_1(t), X_2(t), ..., X_n(t))'$ be a n-dimensional stationary process, with dynamics given by the following factor model:

$$\mathbf{X}(t) = \Lambda \mathbf{F}(t) + \eta(t), \tag{4.3}$$

where $\mathbf{F}(t)$ is a r-dimensional stationary process representing latent common factors, Λ is a $n \times r$ matrix consisting of factor loadings, and $\eta(t)$ is a n-dimensional stationary process corresponding to the idiosyncratic component. For identifiability assume $\mathbb{E}\,\mathbf{F}(t)\mathbf{F}(t)' = I_r$.

Recall that our SBM-driven VAR model (without change points) is:

$$\mathbf{X}(t) = \phi \frac{A}{n}\mathbf{X}(t-1) + \varepsilon(t). \tag{4.4}$$

where A is an adjacency matrix generated from a stochastic block model. Note that $\mathbb{E}(A) = ZPZ'$ where Z is the $n \times K$ membership matrix with entries $Z_{ia} = 1$ when vertex i belongs to community a and 0 otherwise, and P is the $K \times K$ edge probability matrix with entries P_{ab} specifying the probability of an edge between two vertices in community a and b. Then (4.4) can be rewritten as:

$$\mathbf{X}(t) = \frac{\phi ZPZ'}{n}\mathbf{X}(t-1) + \frac{\phi(A - ZPZ')}{n}\mathbf{X}(t-1) + \varepsilon(t).$$

Write $Y_a(t) = \sum_{i \in C_a} \mathbf{X}(t-1)/\sqrt{|C_a|}$ where $\{1, 2, ..., n\} = \cup_{a=1}^K C_a$ is the community partition. Write $\mathbf{Y}(t) = (Y_1(t), ...Y_K(t))'$. Then by stationarity of $\mathbf{X}(t)$ we have that $\mathbf{Y}(t)$ is also stationary with fixed covariance $\Sigma_{\mathbf{Y}}$ that does not depend on n. Write $\mathbf{F}(t) = \Sigma_{\mathbf{Y}}^{-1/2}\mathbf{Y}(t)$ and $C = \text{diag}\{|C_1|, ..., |C_K|\}$. It follows that $\mathbb{E}\,\mathbf{F}(t)\mathbf{F}(t)' = I_K$ and $Z'\mathbf{X}(t-1) = C^{1/2}\mathbf{Y}(t) = C^{1/2}\Sigma_{\mathbf{Y}}^{1/2}\mathbf{F}(t)$. Let $r = K$ and write $\Lambda = \phi ZPC^{1/2}\Sigma_{\mathbf{Y}}^{1/2}/n$ and $\eta(t) = \phi(A - ZPZ')\mathbf{X}(t-1)/n + \varepsilon(t)$. Then (4.4) can be formulated into a DFM of the form in (4.3).

However, in asymptotic analysis of dynamic factor models (cf. [36]) eigenvalues of $\Lambda'\Lambda$ are assumed to be of order n. In our case, entries of $\Lambda'\Lambda = \phi^2 \Sigma_{\mathbf{Y}}^{1/2} C^{1/2} PZ'ZPC^{1/2}\Sigma_{\mathbf{Y}}^{1/2}/n^2 =$

$\phi^2 \Sigma_{\mathbf{Y}}^{1/2} C^{1/2} P C P C^{1/2} \Sigma_{\mathbf{Y}}^{1/2} / n^2$ are of constant order, assuming that community sizes are of order n. Intuitively what this means is that the signal, in terms of factor loadings, is diminishing as n increases. It is not clear if this weak signal case can be handled within the factor model framework.

As a side note, there is work done in the dynamic factor model literature on change point detection [50], where factor loadings are allowed to vary across time. Since the algorithm proposed does not seem to work very well in numerical experiments under our model, we will not include it in Section 4.4.

4.2 Structural breaks with fixed communities

For simplicity we consider a VAR(1) model, but the results have obvious extension to VAR(p) model. Recall that our SBM-driven VAR(1) model can be written as:

$$X(t) = \phi^{(j)} \frac{A^{(j)}}{n} X(t-1) + \varepsilon(t), \quad t = T_{j-1}^0 + 1, ..., T_j^0. \tag{4.5}$$

Here $A^{(j)}$ is an adjacency matrix generated from stochastic block model with K fixed and known communities $\{1, 2, ..., n\} = \cup_{a=1}^K C_a$, and edge probability matrix $P^{(j)}$. The superscript 0 here denotes true change points. We make the following assumptions:

Assumption 4.4. $\{\varepsilon(t), \mathcal{F}_t\}$ is a martingale difference sequence, where $\mathcal{F}_t = \sigma\{\varepsilon(1), ..., \varepsilon(t)\}$ (i.e. $\mathbb{E}(\varepsilon(t)|\mathcal{F}_{t-1}) = 0$). In addition, $\mathbb{E}(\varepsilon(t)\varepsilon(t)'|\mathcal{F}_{t-1}) = I$ and $\sup_t E(||\varepsilon(t)||_2^{4+\delta}) < \infty$ for some $\delta > 0$.

Assumption 4.5. $P^{(j)} \neq P^{(j+1)}$ and each pair of $(\phi^{(j)}, A^{(j)})$ gives rise to a stationary VAR model in (4.5).

Assumption 4.6. $T_j^0 = \lfloor T\alpha_j^0 \rfloor$ where $0 = \alpha_0 < \alpha_1^0 < ... < \alpha_m^0 < \alpha_{m+1}^0 = 1$ and $\alpha_{j+1}^0 - \alpha_j^0 \geq \varepsilon$ for some (small) $\varepsilon > 0$.

By recursion we have that

$$X(t) = \sum_{s=0}^{l-1} (\phi^{(j)} \frac{A^{(j)}}{n})^s \varepsilon(t-s) + (\phi^{(j)} \frac{A^{(j)}}{n})^l X(t-l) \tag{4.6}$$

for $T_{j-1}^0 \leq t - l < t \leq T_j^0$.

Write $n_a := |C_a|$ for size of community a. Define $Y_a(t) = \sum_{i \in C_a} X_i(t)/n_a$ and $Y(t) = (Y_1(t), ..., Y_K(t))'$. Denote by $a(i)$ the community for which vertex i belongs to. We have

$$Y(t) = B^{(j)} Y(t-1) + \eta(t), \quad t = T_{j-1}^0 + 1, ..., T_j^0.$$

Here $B_{ab}^{(j)} = \phi^{(j)} P_{ab}^{(j)} n_a/n$ and $\eta(t) = (\eta_1(t), ..., \eta_K(t))'$ where

$$\eta_a(t) = \frac{\sum_{i \in C_a} \varepsilon_i(t)}{n_a} + \frac{\sum_{i=1}^n \sum_{k=1}^n \left(A_{i,k}^{(j)} - P_{a(i),a(k)}^{(j)} \right) X_k(t-1)}{n n_a}. \tag{4.7}$$

This falls into the structural break framework described in Section 4.1.2. For the change point detection method based on least squares loss there to work, we need to verify a collection of assumption. Assumption 4.4 is the same as in [11] and ensures the technical assumptions in [90] to hold. Assumption 4.5 ensures stationarity of the process, and corresponds to Assumption 4.1. Assumption 4.6 is identical to Assumption 4.3. Finally we need to verify the multivariate counterpart of Assumption 4.2:

Assumption 4.7. *With some increasing σ-field $\{\mathcal{F}_t\}$, $\{Y(t-1)'\eta(t), \mathcal{F}_t\}$ forms a L^r-mixingale sequence with $r = 4 + \delta$ for some $\delta > 0$. That is, there exist nonnegative constants $\{c_t : t \geq 1\}$ and $\{\psi_j : j \geq 0\}$ with $\psi_j \downarrow 0$ as $j \to \infty$, such that for all $t \geq 1$ and $j \geq 0$: (a) $||\mathbb{E}(Y(t-1)'\eta(t)|\mathcal{F}_{t-j})||_r \leq c_t \psi_j$, (b) $||Y(t-1)'\eta(t) - \mathbb{E}(Y(t-1)'\eta(t)|\mathcal{F}_{t+j})||_r \leq c_t \psi_{j+1}$, (c) $\sup_t c_t < \infty$, (d) $\sum_{j=0}^{\infty} j^{1+\kappa} \psi_j < \infty$ for some $\kappa > 0$, (e) $\sup_t ||Y(t)||_r < \infty$ and $\sup_t ||\eta(t)||_r < \infty$.*

Let $\mathcal{F}_t = \sigma\{\varepsilon_1, ..., \varepsilon_t, A^{(1)}, ..., A^{(m+1)}\}$. From (4.7) we have that

$$Y(t-1)'\eta(t) = \sum_{a=1}^K \left(\frac{\sum_{i \in C_a} X_i(t-1)}{n_a} \right)' \left(\frac{\sum_{i \in C_a} \varepsilon_i(t)}{n_a} + \frac{\sum_{i=1}^n \sum_{k=1}^n \left(A_{i,k}^{(j)} - P_{a(i),a(k)}^{(j)} \right) X_k(t-1)}{n n_a} \right).$$

Plug in (4.6) and note that from Assumption 4.4 we have $\mathbb{E}(\varepsilon(t_1)'\varepsilon(t_2)|\mathcal{F}_t) = 0$ for all $t_1, t_2 > t$. Also, $A_{i,k}^{(j)} - P_{a(i),a(k)}^{(j)}$ is a centered Bernoulli random variable with bounded moments, independent of all $\varepsilon(t)$. It follows that $||\mathbb{E}(Y(t-1)'\eta(t)|\mathcal{F}_{t-j})||_r \leq C\lambda^j$, where $\lambda < 1$ is the largest eigenvalue among all $A^{(j)}$, and $C > 0$ is some constant. Let $r = 4 + \delta$, $c_t = C$ and $\psi_j = \lambda^j$, where $\delta > 0$ is the constant in Assumption 4.4. Then both (c) and (d) are satisfied. Since $Y(t-1)'\eta(t)$ is $\mathcal{F}_t$-measurable, (b) is satisfied as well. Finally, (e) follows from (4.6) and Assumption 4.4.

It then follows from the structural break framework that:

Theorem 4.8. $\hat{T}_j - T_j^0 = O_p(1)$ *where $\hat{T}_j$ $(j = 1, ..., m)$ are as in (4.2).*

To summarize, in this case the known communities are used to reduce dimension of the problem from n to K: the n-dimensional observations $\mathbf{X}(t)$ are replaced by averages among observations within K communities, and the $n \times n$ coefficient matrix $\Phi^{(j)}$ is approximated by its rank-K mean matrix.

4.3   Algorithms

We start with the algorithm described in Section 4.1.2 that applies to general VAR model with structural breaks. Write $\mathbf{R}(t) = (\mathbf{X}(t-1)', ..., \mathbf{X}(t-p)')'$ for the lagged observations. Then the OLS estimate $\hat{\Phi} = (\hat{\Phi}_1, ..., \hat{\Phi}_p)$ is given by

$$\hat{\Phi} = \Big(\sum_{t=p+1}^{T} \mathbf{X}(t)\mathbf{R}(t)' \Big) \Big(\sum_{t=p+1}^{T} \mathbf{R}(t)\mathbf{R}(t)' \Big)^{-1}$$

given that $\sum_{t=p+1}^{T} \mathbf{R}(t)\mathbf{R}(t)'$ is invertible.

Recall from Section 4.1.2 that the estimated break dates are given by $(\hat{T}_1, ..., \hat{T}_m) = \operatorname{argmin}_{1 < T_1 < ... < T_m < T} l_1(T_1, ..., T_m)$ with loss function

$$l_1(T_1, ..., T_m) := \sum_{j=1}^{m+1} \sum_{t=T_{j-1}+1}^{T_j} ||\mathbf{X}(t) - \sum_{l=1}^{p} \hat{\Phi}_l^{(j)} \mathbf{X}(t-l)||_2^2,$$

where $\hat{\Phi}_l^{(j)}$ is the OLS estimate of $\Phi_l^{(j)}$ based on observations in $t \in (T_{j-1}, T_j]$.

For our model, however, we want to utilize the latent block structure. Therefore we propose the following loss function instead:

$$l_2(T_1, ..., T_m) := \sum_{j=1}^{m+1} \sum_{t=T_{j-1}+1}^{T_j} ||\mathbf{X}(t) - \sum_{l=1}^{p} \hat{Z}^{(j)} \hat{B}_l^{(j)} (\hat{Z}^{(j)})' \mathbf{X}(t-l)||_2^2,$$

where $\hat{Z}^{(j)}$ is the membership matrix given by certain community detection algorithm applied to observations in $t \in (T_{j-1}, T_j]$ (e.g. the ones described in Chapter 3), and $\hat{B}_l^{(j)}$ is the OLS estimate with the constraint that $\Phi_l^{(j)} = \hat{Z}^{(j)} B_l^{(j)} (\hat{Z}^{(j)})'$ for some $K \times K$ matrix $B_l^{(j)}$. Change points are then estimated by minimizing the loss function: $(\hat{T}_1, ..., \hat{T}_m) = \operatorname{argmin}_{1 < T_1 < ... < T_m < T} l_2(T_1, ..., T_m)$.

4.4 Simulation

In this section we present numerical results for three algorithms: the algorithm for general VAR model with structural break and the algorithm proposed for our model as described in the previous section, with communities estimated by either spectral clustering or kmeans++ on OLS estimate as in Section 3.6.3. We shall refer to them as "OLS method", "restricted OLS with spectral clustering" and "restricted OLS with kmeans++" respectively. All cases considered here are SBM-driven VAR(1) model as defined in (4.1), with single change point.

4.4.1 Fixed communities

We first consider cases where community structure is fixed and only edge probabilities are changing across time. We start with a simple case with $n = 100$ and $K = 2$ equal-sized communities where edge probabilities change from strictly assortative $P^{(1)} = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix} \right)$ to strictly disassortative $P^{(2)} = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix} \right)$. The true change point is set to $T_1 = 0.5T$ and stationarity parameter is $\phi = 1.5$. Figure 4.1 shows standard error of estimated change point for a range of T. Here standard error is computed from just 10 independent simulation runs so the curves shown, especially for that of OLS method, are not very smooth. We can see that restricted OLS methods perform much better than unrestricted OLS when T is moderately large. Moreover, there seems to be a threshold around $T = 500$, above which estimation of change point becomes much more accurate.

Next we examine the performance of algorithms for different K, *i.e.* number of communities. Table 4.1 shows standard error of estimated change point for a range of K computed from 10 independent simulation runs. Here $n = 100$ and $T = 1200$, with equal-sized communities and edge probabilities changing from strictly assortative to strictly disassortative as in the previous case. The true change point is set to $T_1 = 0.3T$, and stationarity parameter is $\phi^{(j)} = 0.8K/||P^{(j)}||$, where $|| \cdot ||$ denotes spectral norm of a matrix. We see no substantial effect of K on accuracy of estimated change point, and restricted OLS consistently performs better than unrestricted OLS for different K.

Finally we consider a case with $n = 100$ and $K = 2$ equal-sized communities where edge probabilities change from $P^{(1)} = \left(\begin{smallmatrix} .7 & .3 \\ .3 & .7 \end{smallmatrix} \right)$ to $P^{(2)} = \left(\begin{smallmatrix} .9 & .1 \\ .1 & .9 \end{smallmatrix} \right)$. The true change point is set to $T_1 = 0.5T$ where $T = 2000$, and stationarity parameter is $\phi^{(j)} = 0.8K/||P^{(j)}||$ as before. Figure 4.2 shows the averaged loss function computed from 10 independent simulation runs, where
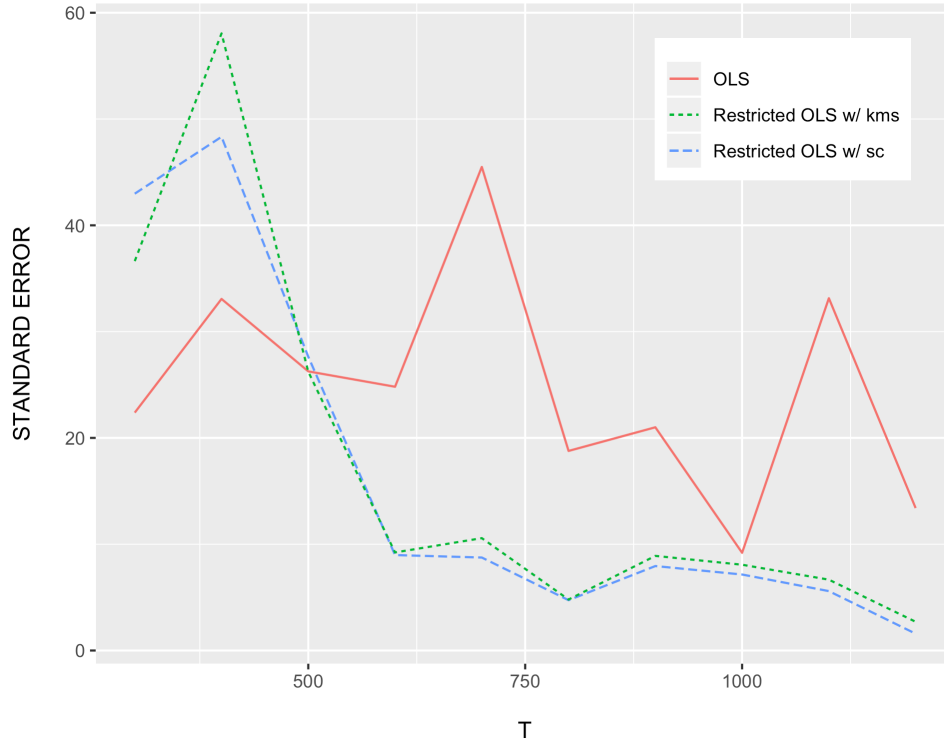
Figure 4.1: Standard error of estimated change point as T increases, with fixed communities

K	OLS	Restricted OLS w/ sc	Restricted OLS w/ kms
2	16.6	13.3	7.4
3	22.5	14.1	7.2
4	21.2	11.7	8.8
5	29.8	7.5	9.6

Table 4.1: Standard error of estimated change point as K increases, with fixed communities
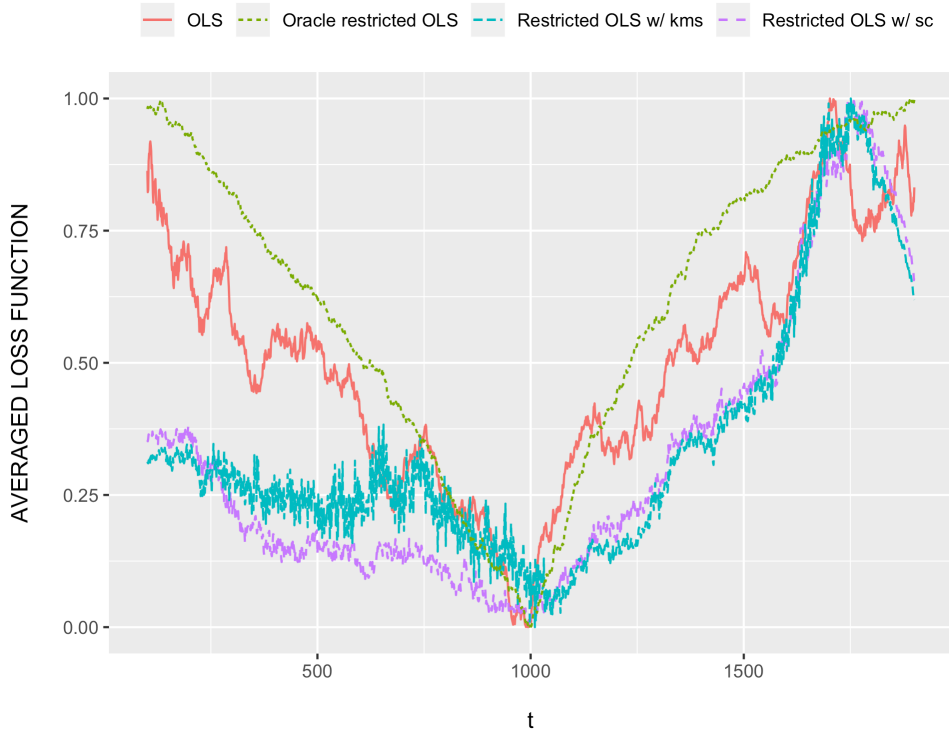
99

Figure 4.2: Averaged loss function, with fixed communities

"oracle restricted OLS" stands for restricted OLS method using the true community partition as constraint. The curves are normalized such that the range is always $[0, 1]$. We can see that the averaged loss function for both OLS and oracle restricted OLS methods shows a clear minimum around $t = 1400$, the true change point. For the other two restricted OLS algorithms, however, the minimum is not so clear. One possible cause is that error rate for recovered communities is not ignorable in this case (cf. Section 3.6.4). Figure 4.3 shows the combined error rate, computed as the sum before and after each time t. We can see that the combined error rate is approximately 0.37 around the true change point. Note that this does not imply that OLS method performs better than restricted OLS, since the loss function curves shown here are averages. In fact, while the estimated change point from oracle restricted OLS lies in the interval $[950, 1050]$ 10 times out of 10 simulation runs, but only 2,2,4 times respectively for OLS, spectral clustering and kmeans++.

If one increase the number of observations, however, restricted OLS starts to work and Figure 4.4 shows an example when T increases to 4000. The standard error computed from 10 runs is 98.9, 42.8 and 32.1 respectively for OLS, spectral clustering and kmeans++.

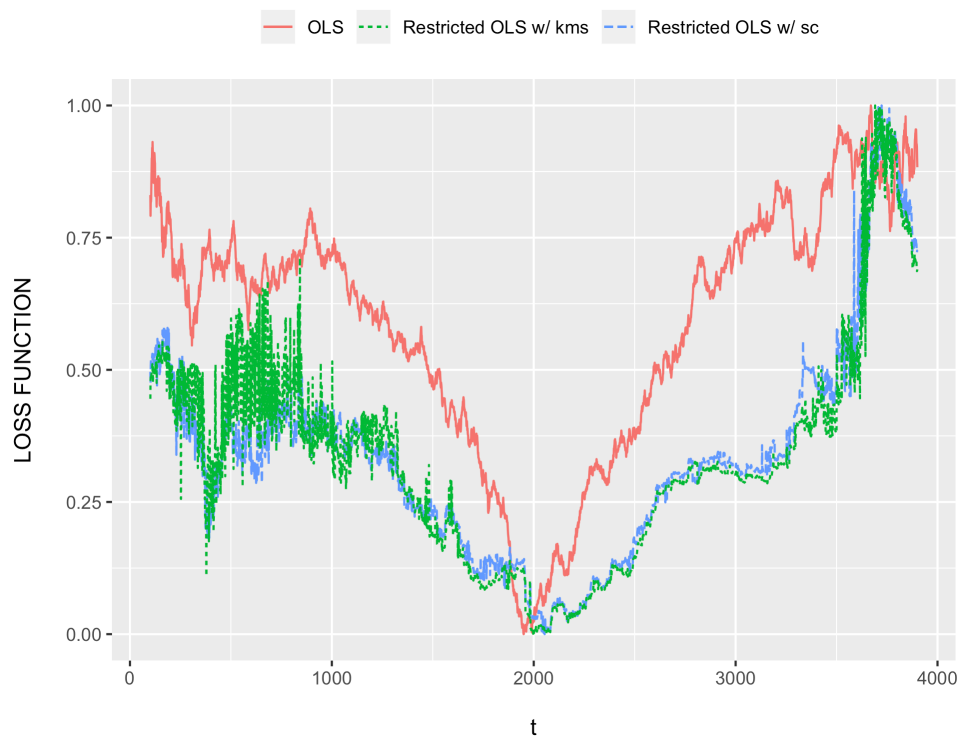**Figure 4.3:** Combined error rate, with fixed communities
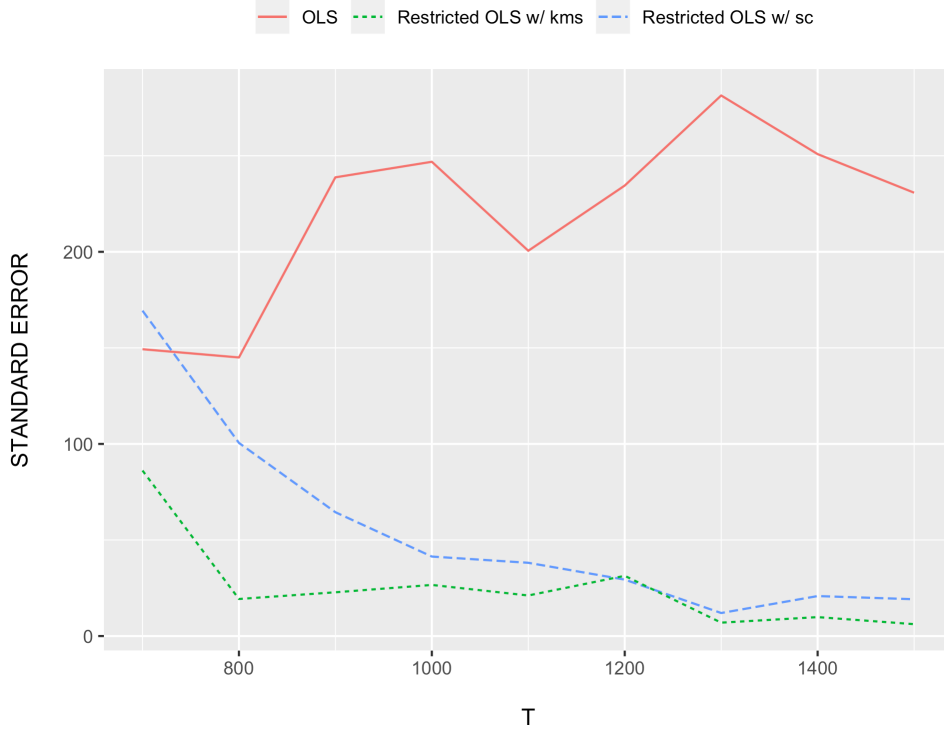

Figure 4.4: Loss function, with fixed communities

Figure 4.5: Standard error of estimated change point as T increases, with changing communities

4.4.2 Changing communities

We start with a simple case with $n = 100$ and $K = 2$ communities where edge probabilities stay strictly assortative $P^{(1)} = P^{(2)} = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix} \right)$ across time, and the block partition changes from $G_1^{(1)} = \{1, ..., 60\}$, $G_2^{(1)} = \{61, ..., 100\}$ to $G_1^{(2)} = \{1, ..., 40\}$, $G_2^{(2)} = \{41, ..., 100\}$. The true change point is set to $T_1 = 0.5T$ and stationarity parameter is $\phi^{(j)} = 0.8n/(||P^{(j)}|| \max_a n_a^{(j)})$, where $n_a^{(j)} = |G_a^{(j)}|$. Figure 4.5 shows standard error of estimated change point for a range of T. We can see that OLS method does not work so well in this case, indicating that the algorithm is not very sensitive to changes in community structure. On the other hand, restricted OLS with kmeans++ performs better than that with spectral clustering, which is consistent with the fact that restricted OLS with kmeans++ has higher recover rate (cf. Section 3.6.4).

APPENDIX A
DISCUSSION ON HYPOTHESIS 2.20

This appendix is devoted to the discussion of Hypothesis 2.20. The first key observation is that coin flips do not seem to matter. Note that there are two kinds of coin flips here, one determining the types, and the other determining the type of vertices a new vertex connects to. More precisely, for $n > 2$, let $I(n) \in \{A, B\}$ denotes the type of vertex n. Also, let $J(n) = 1$ if vertex n connects to a vertex of same type, otherwise let $J(n) = -1$.

For simplicity we treat x, y in $U(x, x, y, y)$ as fixed and suppressed them in notations throughout. Define $U_A(n) := N_A^X(n)/N_A(n)$ and $U_B(n) := N_B^X(n)/N_B(n)$. We condition on the entire sequences of $I(n)$ and $J(n)$, and note that if $I(n+1) = A$ and $J(n+1) = 1$, by the recursive construction we have (again the conditioning is suppressed for simplicity)

$$
\begin{aligned}
\mathbb{E}\, U_A^k(n+1)U_B^l(n+1) &= \frac{N_A^k(n)}{(1+N_A(n))^k}\, \mathbb{E}\, U_A^k(n)U_B^l(n) \\
&+ \frac{1}{(1+N_A(n))^k} \sum_{i=0}^{k-1} \binom{k}{i} N_A(n)^i\, \mathbb{E}\, U_A^{i+1}(n)U_B^l(n).
\end{aligned}
\tag{A.1}
$$

However if $I(n+1) = A$ and $J(n+1) = -1$, we have

$$
\begin{aligned}
\mathbb{E}\, U_A^k(n+1)U_B^l(n+1) &= \frac{N_A^k(n)}{(1+N_A(n))^k}\, \mathbb{E}\, U_A^k(n)U_B^l(n) \\
&+ \frac{1}{(1+N_A(n))^k} \sum_{i=0}^{k-1} \binom{k}{i} N_A(n)^i U_A^i(n)U_B^{l+1}(n)
\end{aligned}
\tag{A.2}
$$

instead. When $I(n+1) = B$ we have similar recursive formulas.

Conditioning on $N_A(n)$ and $N_B(n)$, we suspect that the largest moments occur in one of the two extreme cases:

- $I(m) = A$ and $J(m) = 1$ for $2(x+y) < m \leq N_A(n)+x+y$, $I(m) = B$ and $J(m) = -1$ for $N_A(n)+x+y < m \leq n$. We denote by $g_A(N_A(n), N_B(n), k, l)$ the value of $\mathbb{E}\, U_A^k(n)U_B^l(n)$ corresponding to this case.

- $I(m) = B$ and $J(m) = 1$ for $2(x+y) < m \leq x+y+N_B(n)$, $I(m) = A$ and $J(m) = -1$ for $x+y+N_B(n) < m \leq n$. We denote by $g_B(N_A(n), N_B(n), k, l)$ the value of $\mathbb{E}\, U_A^k(n)U_B^l(n)$ corresponding to this case.

What this essentially says is that the extreme cases are either a string of type A vertices connecting to type A vertices, followed by a string of type B vertices also connecting to type A vertices, or the other way around.

Let $M = x + y$. To better understand this claim we consider a reference sequence $f(n, k)$ $(n \geq M, k \geq 1)$ which corresponds to the moments of classic Pólya urn. To be precise, the sequence is defined by the following recursive formula:

$$f(M, k) = (\frac{x}{M})^k,$$

$$f(n + 1, k) = \frac{n^k}{(n+1)^k} f(n, k) + \sum_{i=0}^{k-1} \binom{k}{i} \frac{n^i}{(n+1)^k} f(n, i+1), \quad n \geq M.$$

It is easy to see that $f(n, k)$ gives the k-th (raw) moment of the ratio of red balls in a classic two-type Pólya urn that starts with x red ball and y blue balls, stopped when the model reaches n balls in total.

In fact, we can calculate the explicit formulas for $f(n, k)$, $g_A(a, b, k, l)$ and $g_B(a, b, k, l)$ using method of undetermined coefficients. Specifically, $f(n, k)$ is of the form $\sum_{i=0}^{k-1} C_i / n^i$. C_i satisfy a system of linear equations, $k - 1$ of which come from equating the coefficients in recursive formulas (the one corresponding to the leading degree is always equal), and the last one is determined by the initial values. In general, given any initial values of $f(M, k)$, one can derive the explicit formulas by solving this linear system.

As for $g_A(a, b, k, l)$, it is easy to see that $g_A(a, M, k, l) = f(a, k) f(M, 1)^l = (\frac{x}{M})^l f(a, k)$. Then $g_A(a, b, k, l)$ can be calculated by applying formula (A.2) $b - (x + y)$ times. Similar to the calculation of $f(n, k)$, $g_A(a, b, k, l)$ is of the form $\sum_{i=0}^{l} C_i / b^i$, where C_i satisfy a system of linear equations, one of which is determined by the initial values $(\frac{x}{M})^l f(a, k)$.

Unfortunately, there does not seem to be an obvious pattern for these coefficients and formulas are hard to generalize to arbitrary order.

We gather some useful properties of $f(n, k)$ in the following proposition.

Proposition A.1. *(i) $f(n, k)$ is strictly increasing in n and converges to the k-th moment of $Beta(x, y)$ distribution, which is $\prod_{i=0}^{k-1} \frac{i+x}{i+M}$. (ii) The following equality holds:*

$$\frac{\sum_{i=0}^{k-1} C_i n^i f(n, i+1)}{\prod_{i=0}^{k-1} (n+i)} = \mu_k := \prod_{i=0}^{k-1} \frac{i+x}{i+M}, \tag{A.3}$$

where C_i is the coefficient of n^i in $\prod_{i=0}^{k-1}(n+i)$.

The first statement follows from the convergence of the ratio, as a martingale, to a $Beta(x,y)$ r.v.. Thus its k-th power is a submartingale and has increasing expectations.

The second statement can be proved by induction on n and k.

Now we are ready to state the following hypothesis.

Hypothesis A.2. *Conditioning on I and J,*

$$\mathbb{E}\, U_A^k(n)U_B^l(n) \leq \max(g_A(N_A(n), N_B(n), k, l), g_B(N_A(n), N_B(n), k, l))$$

for all n. Also,

$$\lim_{b \to \infty} \max(g_A(a, b, k, l), g_B(a, b, k, l)) = \lim_{b \to \infty} g_B(a, b, k, l) = \frac{\sum_{i=0}^{l-1} C_i a^i f(a, k-l+i+1)}{\prod_{i=1}^{l-1}(a+i)}.$$

The second equality here is not part of the hypothesis and can be proved by induction on a. When $l = 1$ or 0, both limits are defined as $f(a, k)$.

This hypothesis can be verified for any finite n by computer program, where at each time step $\mathbb{E}\, U_A^k(n)U_B^l(n)$ is replaced by $\max(g_A(N_A(n), N_B(n), k, l), g_B(N_A(n), N_B(n), k, l))$.

Remark A.1. When $l = 2$, the limit is

$$\frac{af(a, k) + f(a, k-1)}{a+1}.$$

When $l = 3$, the limit is

$$\frac{a^2 f(a, k) + 3af(a, k-1) + 2f(a, k-2)}{(a+1)(a+2)}.$$

And so on and so forth. Note that this is consistent with (A.3). In fact, the limits are calculated by first sending b to infinity and then apply (A.2) a times.

Explicit formula for $g_A(a, b, k, 1)$ and $g_B(a, b, 1, k)$ are easily calculated to be

$$g_A(a, b, k, 1) = \frac{M}{b}\mu_1 f(a, k) + \frac{b-M}{b}f(a, k+1)$$

$$g_B(a, b, 1, k) = \frac{M}{a}\mu_1 f(b, k) + \frac{a-M}{a}f(b, k+1)$$

Explicit formula for $g_A(a, b, k, 2)$ are more complicated:

$$g_A(a, b, k, 2) = \frac{M^2}{b^2} \mu_1^2 f(a, k) + \frac{2M(b-M)}{b^2} \mu_1 f(a, k+1)$$
$$+ \frac{(b-M)(b-M-1)}{b^2} f(a, k+2) + \frac{b-M}{b^2} f(a, k+1).$$

Formulas for more complicated moments again seem to have no obvious pattern.

For some special cases we know which one of $g_A(N_A(n), N_B(n), k, l)$ and $g_B(N_A(n), N_B(n), k, l)$ is larger: $\max(g_A(a, b, k, 0), g_B(a, b, k, 0)) = g_A(a, b, k, 0) = f(a, k)$ and $\max(g_A(a, b, k, 1), g_B(a, b, k, 1)) = g_A(a, b, k, 1) = \frac{M}{b} \mu_1 f(a, k) + \frac{b-M}{b} f(a, k+1)$. And from simulations it seems that for $l > 1$, $\max(g_A(a, b, k, l), g_B(a, b, k, l)) = g_B(a, b, k, l)$ would hold for large enough b.

Next we present some actual proofs for the lower moments.

Proof of Hypothesis A.2 for $k + l = 1$: Trivial, since $\mathbb{E} U_A(n) = \mathbb{E} U_B(n) = x/(x+y)$.

Proof of Hypothesis A.2 for $k + l = 2$: Surprisingly, we have $g_A(a, b, 1, 1) = g_B(a, b, 1, 1)$ so type sequence doesn't matter at all and since $g_A(a, b, 1, 1) \le f(a, 2)$ and $g_B(a, b, 1, 1) \le f(b, 2)$ we are done.

Proof of Hypothesis A.2 for $k + l = 3$: By symmetry we only have to proof that if $\mathbb{E} U_A^2(n) U_B(n) = g_A(U_A(n), U_B(n), 2, 1)$ and $\mathbb{E} U_B^2(n) U_A(n) = g_B(U_A(n), U_B(n), 1, 2)$ then $\mathbb{E} U_A^2(n+1) U_B(n+1) \le g_A(U_A(n)+1, U_B(n), 2, 1)$ if $I(n+1) = A$. Turns out that this requires

$$g_B(a, b, 1, 2) \le \frac{M}{b} \mu_1 f(a, 2) + \frac{b-M}{b} \frac{af(a, 3) + f(a, 2)}{a+1},$$

which further requires that

$$g_B(a, b, 0, 3) = f(b, 3) \le \frac{M}{b} \mu_1 \frac{af(a, 2) + f(a, 1)}{a+1} + \frac{b-M}{b} \frac{a^2 f(a, 3) + 3af(a, 2) + 2f(a, 1)}{(a+1)(a+2)}.$$

But

$$\frac{M}{b} \mu_1 \frac{af(a, 2) + f(a, 1)}{a+1} + \frac{b-M}{b} \frac{a^2 f(a, 3) + 3af(a, 2) + 2f(a, 1)}{(a+1)(a+2)} = \frac{M}{b} \mu_1 \mu_2 + \frac{b-M}{b} \mu_3$$
$$= \lim_{a \to \infty} g_A(a, b, 1, 2)$$
$$= \frac{bf(b, 2) + f(b, 1)}{b+1}$$

so the requirement is satisfied.

Proof of Hypothesis A.2 for $k + l = 4$**:** In this case $g_A(a, b, 2, 2) > g_B(a, b, 2, 2)$ when $a > b$ and $g_A(a, b, 2, 2) \leq g_B(a, b, 2, 2)$ when $a \leq b$. However the proof is more involved, since in proving the third moment we used something like:

$$f(n + 1, k + 1) = \frac{n^k}{(n + 1)^k} f(n, k + 1) + \sum_{i=0}^{k-1} \binom{k}{i} \frac{n^i}{(n + 1)^k} \frac{nf(n, i + 2) + f(n, i + 1)}{n + 1},$$

$$\frac{(n + 1)f(n, k + 2) + f(n, k + 1)}{n + 2} = \frac{n^k}{(n + 1)^k} \frac{nf(n, k + 2) + f(n, k + 1)}{n + 1}$$
$$+ \sum_{i=0}^{k-1} \binom{k}{i} \frac{n^i}{(n + 1)^k} \frac{n^2 f(n, i + 3) + 3nf(n, i + 2) + 2f(n, i + 1)}{(n + 1)(n + 2)}$$

(so on and so forth also holds). But for forth moment if we want to do the same thing we would need something like

$$\frac{n}{n + 1} f(n, 3) + \frac{1}{n + 1} \frac{n^2 f(n, 3) + 3nf(n, 2) + 2f(n, 1)}{(n + 1)(n + 2)} \leq f(n + 1, 3)$$

and things like these don't seem to hold in general. Still one can brute force prove it and the fact that $g_A(a, b, 2, 2) > g_B(a, b, 2, 2)$ when $a > b$ is useful.

Proof of Hypothesis A.2 for $k + l > 4$**:** In general there does not seem to be any obvious way to determine whether $g_A(a, b, k, l) > g_B(a, b, k, l)$, and this prohibits generalization of our proofs.

APPENDIX B
PROOF OF THEOREM 3.10

The proof of Theorem 3.10 essentially follows that of Theorem 3.15. In particular, Step 2 of the proof remains the same. For step 1, we need to consider the perturbation term $\xi_{ij}(t_1, t_2)$. In particular, for the matrices in (3.24), we have that covariances between scaled block means are

$$
\begin{aligned}
\Sigma_{bb}^G = \mathrm{Var}(\mathbf{Y}_b) &= \frac{1}{n_b^G} \mathrm{Var}(\sum_{i \in G_b} \mathbf{X}_i) = \frac{1}{n_b^G}(n_b^G I_T + \sum_{i,j \in G_b} P(i,j)) \\
&= \frac{1}{n_b^G}(n_b^G I_T + (n^G)_b^2 P_{bb} + \sum_{i,j \in G_b} \Xi_{ij}) = I_T + n_b^G P_{bb} + \frac{1}{n_b^G} \sum_{i,j \in G_b} \Xi_{ij}
\end{aligned}
\tag{B.1}
$$

and similarly

$$
\Sigma_{bd}^G = \sqrt{n_b^G n_d^G} P_{bd} + \frac{1}{\sqrt{n_b^G n_d^G}} \sum_{i \in G_b} \sum_{j \in G_d} \Xi_{ij}
\tag{B.2}
$$

when $b \neq d$.

Covariances between scaled cluster means are

$$
\begin{aligned}
\Sigma_{aa}^C = \mathrm{Var}(\mathbf{Z}_a) &= \frac{1}{n_a^C} \mathrm{Var}(\sum_{i \in C_a} \mathbf{X}_i) = I_T + \frac{1}{n_a^C} \sum_{i,j \in C_a} P(i,j) \\
&= I_T + \frac{1}{n_a^C} \sum_{1 \leq b,d \leq K} n_{ab} n_{ad} P_{bd} + \frac{1}{n_a^C} \sum_{i,j \in C_a} \Xi_{ij}
\end{aligned}
\tag{B.3}
$$

and similarly

$$
\Sigma_{ac}^C = \frac{1}{\sqrt{n_a^C n_c^C}} \sum_{1 \leq b,d \leq K} n_{ab} n_{cd} P_{bd} + \frac{1}{\sqrt{n_a^C n_c^C}} \sum_{i \in C_a} \sum_{j \in C_c} \Xi_{ij}
\tag{B.4}
$$

when $a \neq c$.

Finally covariances between scaled block and cluster means are

$$
\begin{aligned}
\tilde{\Sigma}_{ab} = \mathrm{Cov}(\mathbf{Y}_b, \mathbf{Z}_a) &= \frac{1}{\sqrt{n_a^C n_b^G}} \mathrm{Cov}(\sum_{i \in C_a} \mathbf{X}_i, \sum_{j \in G_b} \mathbf{X}_j) = \frac{1}{\sqrt{n_a^C n_b^G}}(n_{ab} I_T + \sum_{i \in C_a} \sum_{j \in G_b} P(i,j)) \\
&= \frac{n_{ab}}{\sqrt{n_a^C n_b^G}} I_T + \sqrt{\frac{n_b^G}{n_a^C}} \sum_{d=1}^K n_{ad} P_{bd} + \frac{1}{\sqrt{n_a^C n_b^G}} \sum_{i \in C_a} \sum_{j \in G_b} \Xi_{ij}.
\end{aligned}
\tag{B.5}
$$

Note that all of these expressions have a leading part with the blockwise defined P_{bd}, and a perturbation part with the node-wise defined Ξ_{ij}. The two parts will be analyzed separately,

108

so for convenience we introduce the following notation:

$$\Sigma_{bd}^{G(1)} = \delta(b,d)I_T + \sqrt{n_b^G n_d^G} P_{bd},$$

$$\Sigma_{bd}^{G(2)} = \frac{1}{\sqrt{n_b^G n_d^G}} \sum_{i \in G_b} \sum_{j \in G_d} \Xi_{ij},$$

$$\Sigma_{ac}^{C(1)} = \delta(a,c)I_T + \frac{1}{\sqrt{n_a^C n_c^C}} \sum_{1 \le b,d \le K} n_{ab} n_{cd} P_{bd},$$

$$\Sigma_{ac}^{C(2)} = \frac{1}{\sqrt{n_a^C n_c^C}} \sum_{i \in C_a} \sum_{j \in C_c} \Xi_{ij}, \tag{B.6}$$

$$\tilde{\Sigma}_{ab}^{(1)} = \frac{n_{ab}}{\sqrt{n_a^C n_b^G}} I_T + \sqrt{\frac{n_b^G}{n_a^C}} \sum_{d=1}^{K} n_{ad} P_{bd},$$

$$\tilde{\Sigma}_{ab}^{(2)} = \frac{1}{\sqrt{n_a^C n_b^G}} \sum_{i \in C_a} \sum_{j \in G_b} \Xi_{ij}.$$

The following lemma summarizes some of the useful inequalities involving P_{bd} and Ξ_{ij} that we shall need along the proof.

Lemma B.1. *For all $1 \le i,j \le n$ and $1 \le a,b,c,d \le K$,*

$$||\Xi_{ij}||_{max} \le \frac{r\gamma}{4K^3 n^2}, \tag{B.7}$$

$$||\Xi_{ij}||_{\infty} \le \min\{\frac{M}{n}, \frac{r\gamma}{4K^3 n^2}\} \tag{B.8}$$

$$\max\{||\Sigma_{bd}^{G(1)}||_{\infty}, ||\Sigma_{ac}^{C(1)}||_{\infty}, ||\tilde{\Sigma}_{ab}^{(1)}||_{\infty}\} \le 1 + M, \tag{B.9}$$

$$\max\{||\Sigma_{bd}^{G(2)}||_{\infty}, ||\Sigma_{ac}^{C(2)}||_{\infty}, ||\tilde{\Sigma}_{ab}^{(2)}||_{\infty}\} \le \min\{M, \frac{r\gamma}{4K^3 n}\}, \tag{B.10}$$

$$\left| ||\Sigma_{bd}^{G}||_F^2 - ||\Sigma_{bd}^{G(1)}||_F^2 \right| \le \frac{2M+1}{K} \frac{r\gamma T}{2K^2 n} \tag{B.11}$$

$$\left| ||\Sigma_{ac}^{C}||_F^2 - ||\Sigma_{ac}^{C(1)}||_F^2 \right| \le \frac{2M+1}{K} \frac{r\gamma T}{2K^2 n} \tag{B.12}$$

$$\left| ||\tilde{\Sigma}_{ab}||_F^2 - ||\tilde{\Sigma}_{ab}^{(1)}||_F^2 \right| \le \frac{2M+1}{K} \frac{r\gamma T}{2K^2 n} \tag{B.13}$$

Next we turn to $\mathbb{E}\,\Delta$ in (3.31) and note that we still have

$$\mathbb{E}\,\Delta = \text{Tr}(\Omega\Sigma) = \sum_{b=1}^{K} \text{Tr}(\Sigma_{bb}^{G}) - \sum_{a=1}^{K} \text{Tr}(\Sigma_{aa}^{C}).$$

Using (B.1) and (B.3) we have

$$\text{Tr}(\Sigma_{bb}^G) = T + n_b^G \sum_{t=1}^T \rho_{bb}(t,t) + \frac{1}{n_b^G} \sum_{t=1}^T \sum_{i,j \in G_b} \xi_{ij}(t,t)$$

and

$$\text{Tr}(\Sigma_{aa}^C) = T + \frac{1}{n_a^C} \sum_{t=1}^T \sum_{1 \le b,d \le K} n_{ab} n_{ad} \rho_{bd}(t,t) + \frac{1}{n_a^C} \sum_{t=1}^T \sum_{i,j \in C_a} \xi_{ij}(t,t).$$

The first term will cancel out and we can write $\mathbb{E}\,\Delta = \delta_1 + \delta_2$, where

$$\delta_1 := \sum_{t=1}^T \sum_{b=1}^K n_b^G \rho_{bb}(t,t) - \sum_{t=1}^T \sum_{a=1}^K \frac{1}{n_a^C} \sum_{1 \le b,d \le K} n_{ab} n_{ad} \rho_{bd}(t,t)$$

and

$$\delta_2 := \sum_{t=1}^T \sum_{b=1}^K \frac{1}{n_b^G} \sum_{i,j \in G_b} \xi_{ij}(t,t) - \sum_{t=1}^T \sum_{a=1}^K \frac{1}{n_a^C} \sum_{i,j \in C_a} \xi_{ij}(t,t).$$

For the leading part δ_1 we have

$$\delta_1 \ge \left(\sum_{a=1}^K \sum_{\substack{b,d=1 \\ b \ne d}}^K \frac{n_{ab} n_{ad}}{n_a^C} \right) \frac{\gamma T}{n} \tag{B.14}$$

as in (3.31). And in place of Lemma 3.16 we have the same lower bound on δ_1:

Lemma B.2.

$$\delta_1 \ge \frac{rs\gamma}{2} \frac{T}{K^2 n}.$$

The proof of this lemma is identical to that of Lemma 3.16, with δ replaced by δ_1.

For the perturbation part δ_2 we have

$$
\begin{aligned}
|\delta_2| &\le \sum_{t=1}^T \sum_{b=1}^K \frac{1}{n_b^G} \sum_{i,j \in G_b} |\xi_{ij}(t,t)| + \sum_{t=1}^T \sum_{a=1}^K \frac{1}{n_a^C} \sum_{i,j \in C_a} |\xi_{ij}(t,t)| \\
&\le \sum_{t=1}^T \sum_{b=1}^K \frac{1}{n_b^G} \sum_{i,j \in G_b} \|\Xi_{ij}\|_{max} + \sum_{t=1}^T \sum_{a=1}^K \frac{1}{n_a^C} \sum_{i,j \in C_a} \|\Xi_{ij}\|_{max} \\
&\le \sum_{t=1}^T \left(\sum_{b=1}^K \frac{(n_b^G)^2}{n_b^G} + \sum_{a=1}^K \frac{(n_a^C)^2}{n_a^C} \right) \frac{r\gamma}{4K^3 n^2} = \frac{2Tnr\gamma}{4K^3 n^2} = \frac{r\gamma}{2} \frac{T}{K^3 n}.
\end{aligned}
$$

The last inequality here follows from (B.7) in Lemma B.1. Since $K \ge 2$ and $s \ge 1$ we have $|\delta_2| \le \delta_1/2$ by Lemma B.2.

In place of Lemma 3.18 and 3.19 we have the following lemmas:

Lemma B.3. $||\Sigma^{1/2}\Omega\Sigma^{1/2}||_{op} \le 2K(2M+1)$, *where M is the constant in* (3.10) *of Assumption 3.9.*

Lemma B.4. $||\Sigma^{1/2}\Omega\Sigma^{1/2}||_F^2 \le (5 + \frac{2}{r\gamma} + \frac{2M^2}{\gamma} + 8M)K\delta_1$.

The rest of the proof stays the same.

B.1   Proof of Lemma B.1

The first two inequalities are straightforward. (B.7) follows from $||\Xi_{ij}||_{max} \le ||\Xi_{ij}||_\infty$ and (3.12). Since (3.9) implies that $||\Xi_{ij}||_\infty \le ||P_{bd}||_\infty$ for $i \in G_b$ and $j \in G_d$, (B.8) follows from (3.10) and (3.12).

For (B.9), using (3.10) in Assumption 3.9 we have

$$||\Sigma_{bd}^{G(1)}||_\infty \le 1 + \sqrt{n_b^G n_d^G}||P_{bd}||_\infty \le 1 + n||P_{bd}||_\infty \le 1 + M,$$

$$
\begin{aligned}
||\Sigma_{ac}^{C(1)}||_\infty &\le 1 + \frac{1}{\sqrt{n_a^C n_c^C}} \sum_{1 \le b,d \le K} n_{ab}n_{cd}||P_{bd}||_\infty \\
&\le 1 + \frac{n}{n_a^C n_c^C} \sum_{1 \le b,d \le K} n_{ab}n_{cd}||P_{bd}||_\infty \\
&\quad (\text{using } n_a^C, n_c^C \le n) \\
&= 1 + \frac{M}{n_a^C n_c^C} \sum_{1 \le b,d \le K} n_{ab}n_{cd} \\
&= 1 + M
\end{aligned}
$$

and

$$||\tilde{\Sigma}_{ab}^{(1)}||_\infty \le \frac{n_{ab}}{\sqrt{n_a^C n_b^G}} + \sqrt{\frac{n_b^G}{n_a^C}}\sum_{d=1}^{K} n_{ad}||P_{bd}||_\infty \le 1 + n\sum_{d=1}^{K}||P_{bd}||_\infty \le 1 + M.$$

For (B.10), note that

$$||\Sigma_{bd}^{G(2)}||_\infty \le \frac{1}{\sqrt{n_b^G n_d^G}} \sum_{i \in G_b} \sum_{j \in G_d} ||\Xi_{ij}||_\infty \le \frac{n_b^G n_d^G}{\sqrt{n_b^G n_d^G}} \min\{\frac{M}{n}, \frac{r\gamma}{4K^3 n^2}\} \le \min\{M, \frac{r\gamma}{4K^3 n}\}.$$

Similarly we have $\max\{||\Sigma_{ac}^{C(2)}||_\infty, ||\tilde{\Sigma}_{ab}^{(2)}||_\infty\} \le \min\{M, \frac{r\gamma}{4K^3 n}\}$.

For (B.11), we have

$$||\Sigma_{bd}^G||_F^2 = \text{Tr}((\Sigma_{bd}^G)'\Sigma_{bd}^G) = ||\Sigma_{bd}^{G(1)}||_F^2 + ||\Sigma_{bd}^{G(2)}||_F^2 + 2\text{Tr}((\Sigma_{bd}^{G(2)})'\Sigma_{bd}^{G(1)}).$$

And by (B.9) and (B.10) we know that

$$||\Sigma_{bd}^{G(2)}||_F^2 \leq T||\Sigma_{bd}^{G(2)}||_\infty^2 \leq T\frac{r^2\gamma^2}{16K^6n^2} \leq T\frac{r^2\gamma^2}{2K^3n}$$

and

$$|\operatorname{Tr}((\Sigma_{bd}^{G(2)})'\Sigma_{bd}^{G(1)})| \leq T||\Sigma_{bd}^{G(2)}||_\infty||\Sigma_{bd}^{G(1)}||_\infty \leq T\frac{r\gamma}{4K^3n}(M+1).$$

Now, recall that $r \in (0,1)$ and $\gamma \leq M$. Therefore we have

$$\left|||\Sigma_{bd}^G||_F^2 - ||\Sigma_{bd}^{G(1)}||_F^2\right| \leq T\frac{r\gamma M}{2K^3n} + 2T\frac{r\gamma}{4K^3n}(M+1) = \frac{2M+1}{K}\frac{r\gamma T}{2K^2n}.$$

Similarly we have (B.12) and (B.13). This completes the proof of Lemma B.1.

■

B.2 Proof of Lemma B.3

Same as in the proof of Lemma 3.18 we have

$$||\Sigma^{1/2}\Omega\Sigma^{1/2}||_{op} \leq ||\Sigma^{1/2}||_{op}||\Omega||_{op}||\Sigma^{1/2}||_{op} = ||\Sigma||_{op}.$$

Next, note that we can upper-bound the operator norm by the L_∞ norm. Therefore from (3.24), (B.9) and (B.10) we know that

$$\begin{aligned}
||\Sigma||_{op} &\leq 2K \max_{1\leq a,b,c,d\leq K}\{||\Sigma_{bd}^G||_\infty, ||\Sigma_{ac}^C||_\infty, ||\tilde{\Sigma}_{ab}||_\infty\} \\
&\leq 2K(1 + M + M) \\
&= 2K(2M+1).
\end{aligned}$$

This proves the lemma.

■

B.3 Proof of Lemma B.4

Same as in the proof of Lemma 3.19 we have

$$||\Sigma^{1/2}\Omega\Sigma^{1/2}||_F^2 = \operatorname{Tr}(\Sigma^{1/2}\Omega\Sigma\Omega\Sigma^{1/2}) = \operatorname{Tr}(\Omega\Sigma\Omega\Sigma) \tag{B.15}$$

by properties of traces. And from (3.30) we can compute the trace as

$$\mathrm{Tr}(\Omega\Sigma\Omega\Sigma) = \sum_{1\le b,d\le K} \mathrm{Tr}((\Sigma_{bd}^G)'\Sigma_{bd}^G) + \sum_{1\le a,c\le K} \mathrm{Tr}((\Sigma_{ac}^C)'\Sigma_{ac}^C) - 2\sum_{1\le a,b\le K} \mathrm{Tr}(\tilde\Sigma_{ab}'\tilde\Sigma_{ab})$$

$$= \sum_{1\le b,d\le K} ||\Sigma_{bd}^G||_F^2 + \sum_{1\le a,c\le K} ||\Sigma_{ac}^C||_F^2 - 2\sum_{1\le a,b\le K} ||\tilde\Sigma_{ab}||_F^2.$$

Divide this trace into three parts and write $\mathrm{Tr}(\Omega\Sigma\Omega\Sigma) = I_1 + I_2 + I_3$, where

$$I_1 := \sum_{1\le b,d\le K} ||\Sigma_{bd}^{G(1)}||_F^2 - \sum_{1\le a,b\le K} ||\tilde\Sigma_{ab}^{(1)}||_F^2,$$

$$I_2 := \sum_{1\le a,c\le K} ||\Sigma_{ac}^{C(1)}||_F^2 - \sum_{1\le a,b\le K} ||\tilde\Sigma_{ab}^{(1)}||_F^2$$

and

$$I_3 := \sum_{1\le b,d\le K} (||\Sigma_{bd}^G||_F^2 - ||\Sigma_{bd}^{G(1)}||_F^2) + \sum_{1\le a,c\le K} (||\Sigma_{ac}^C||_F^2 - ||\Sigma_{ac}^{C(1)}||_F^2) - 2\sum_{1\le a,b\le K} (||\tilde\Sigma_{ab}||_F^2 - ||\tilde\Sigma_{ab}^{(1)}||_F^2).$$

Since I_1 and I_2 are identical to those defined in the proof of Lemma 3.19, we use the same bounds to get

$$I_1 + I_2 \le K(1 + \frac{2}{r\gamma} + \frac{2M^2}{\gamma})\delta_1.$$

For I_3, we have

$$I_3 \le 4K^2 \frac{2M+1}{K} \frac{r\gamma T}{2K^2 n} = 4K(2M+1)\frac{r\gamma T}{2K^2 n}$$

by (B.11),(B.12) and (B.13). Using Lemma B.2 we get $I_3 \le 4K(2M+1)\delta_1$.

It follows from the above inequalities that

$$\mathrm{Tr}(\Omega\Sigma\Omega\Sigma) = I_1 + I_2 + I_3 \le K(1 + \frac{2}{r\gamma} + \frac{2M^2}{\gamma})\delta_1 + 4K(2M+1)\delta_1 \le (5 + \frac{2}{r\gamma} + \frac{2M^2}{\gamma} + 8M)K\delta_1.$$

Lemma B.4 then follows from (B.15).

∎

BIBLIOGRAPHY

[1] Emmanuel Abbe, Afonso S Bandeira, and Georgina Hall. Exact recovery in the stochastic block model. *IEEE Transactions on Information Theory*, 62(1):471–487, 2015.

[2] Emmanuel Abbe and Colin Sandon. Community detection in general stochastic block models: Fundamental limits and efficient algorithms for recovery. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*, pages 670–688. IEEE, 2015.

[3] Charu Aggarwal and Karthik Subbian. Evolutionary network analysis: A survey. *ACM Computing Surveys (CSUR)*, 47(1):1–36, 2014.

[4] Arash A Amini, Elizaveta Levina, et al. On semidefinite relaxations for the block model. *The Annals of Statistics*, 46(1):149–179, 2018.

[5] Samaneh Aminikhanghahi and Diane J Cook. A survey of methods for time series change point detection. *Knowledge and information systems*, 51(2):339–367, 2017.

[6] David Arthur and Sergei Vassilvitskii. k-means++: The advantages of careful seeding. Technical report, Stanford, 2006.

[7] Krishna B Athreya and Samuel Karlin. Embedding of urn schemes into continuous time markov branching processes and related limit theorems. *The Annals of Mathematical Statistics*, 39(6):1801–1817, 1968.

[8] Alexander Aue and Lajos Horváth. Structural breaks in time series. *Journal of Time Series Analysis*, 34(1):1–16, 2013.

[9] Pranjal Awasthi, Afonso S Bandeira, Moses Charikar, Ravishankar Krishnaswamy, Soledad Villar, and Rachel Ward. Relax, no need to round: Integrality of clustering formulations. In *Proceedings of the 2015 Conference on Innovations in Theoretical Computer Science*, pages 191–200. ACM, 2015.

[10] Pranjal Awasthi and Or Sheffet. Improved spectral-norm bounds for clustering. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 37–49. Springer, 2012.

[11] Jushan Bai et al. Vector autoregressive models with structural changes in regression coefficients and in variance-covariance matrices. Technical report, China Economics and Management Academy, Central University of Finance and , 2000.

[12] Jushan Bai and Pierre Perron. Estimating and testing linear models with multiple structural changes. *Econometrica*, pages 47–78, 1998.

[13] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.

[14] Michèle Basseville, Igor V Nikiforov, et al. *Detection of abrupt changes: theory and application*, volume 104. prentice Hall Englewood Cliffs, 1993.

[15] Shankar Bhamidi, Jimmy Jin, and Andrew Nobel. Change point detection in network models: Preferential attachment and long range dependence. *arXiv preprint arXiv:1508.02043*, 2015.

[16] Monika Bhattacharjee, Moulinath Banerjee, and George Michailidis. Change point estimation in a dynamic stochastic block model. *arXiv preprint arXiv:1812.03090*, 2018.

[17] JD Biggins. The first-and last-birth problems for a multitype age-dependent branching process. *Advances in Applied Probability*, 8(3):446–459, 1976.

[18] B´ela Bollobás, Oliver Riordan, Joel Spencer, and Gábor Tusnády. The degree sequence of a scale-free random graph process. *Random Structures & Algorithms*, 18(3):279–290, 2001.

[19] Charles Bordenave, Marc Lelarge, and Laurent Massoulié. Non-backtracking spectrum of random graphs: community detection and non-regular ramanujan graphs. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*, pages 1347–1357. IEEE, 2015.

[20] E Brodsky and Boris S Darkhovsky. *Nonparametric methods in change point problems*, volume 243. Springer Science & Business Media, 2013.

[21] Sébastien Bubeck, Luc Devroye, and Gábor Lugosi. Finding adam in random growing trees. *Random Structures & Algorithms*, 50(2):158–172, 2017.

[22] Florentina Bunea, Christophe Giraud, and Xi Luo. Minimax optimal variable clustering in g-models via cord. *arXiv preprint arXiv:1508.01939*, 2015.

[23] Florentina Bunea, Christophe Giraud, Xi Luo, Martin Royer, and Nicolas Verzelen. Model assisted variable clustering: minimax-optimal recovery and algorithms. *arXiv preprint arXiv:1508.01939*, 2015.

[24] Florentina Bunea, Christophe Giraud, Martin Royer, and Nicolas Verzelen. Pecok: a convex optimization approach to variable clustering. *arXiv preprint arXiv:1606.05100*, 2016.

[25] Edward G Carlstein, Hans-Georg Müller, and David Siegmund. Change-point problems. IMS, 1994.

[26] James Samuel Coleman et al. Introduction to mathematical sociology. *Introduction to mathematical sociology.*, 1964.

[27] Kenny S Crump and Charles J Mode. A general age-dependent branching process. i. *Journal of Mathematical Analysis and Applications*, 24(3):494–508, 1968.

[28] Miklos Csorgo and Lajos Horváth. *Limit theorems in change-point analysis*. John Wiley & Sons Chichester, 1997.

[29] Aurelien Decelle, Florent Krzakala, Cristopher Moore, and Lenka Zdeborová. Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Physical Review E*, 84(6):066106, 2011.

[30] Maria Deijfen and Robert Fitzner. Birds of a feather or opposites attract-effects in network modelling. *arXiv preprint arXiv:1612.03127*, 2016.

[31] Luc Devroye. Branching processes in the analysis of the heights of trees. *Acta Informatica*, 24(3):277–298, 1987.

[32] Luc Devroye. Applications of the theory of records in the study of random trees. *Acta Informatica*, 26(1-2):123, 1988.

[33] Luc Devroye. Branching processes and their applications in the analysis of tree structures and tree algorithms. In *Probabilistic methods for algorithmic discrete mathematics*, pages 249–314. Springer, 1998.

[34] Luc Devroye, Omar Fawzi, and Nicolas Fraiman. Depth properties of scaled attachment random recursive trees. *Random Structures & Algorithms*, 41(1):66–98, 2012.

[35] Luc Devroye and Jiang Lu. The strong convergence of maximal degrees in uniform random recursive trees and dags. *Random Structures & Algorithms*, 7(1):1–14, 1995.

[36] Catherine Doz, Domenico Giannone, and Lucrezia Reichlin. A two-step estimator for large approximate dynamic factor models based on kalman filtering. *Journal of Econometrics*, 164(1):188–205, 2011.

[37] Raissa M D'Souza, Paul L Krapivsky, and Cristopher Moore. The power of choice in growing trees. *The European Physical Journal B*, 59(4):535–543, 2007.

[38] Vladimir Estivill-Castro. Why so many clustering algorithms: a position paper. *SIGKDD explorations*, 4(1):65–75, 2002.

[39] Yingjie Fei and Yudong Chen. Hidden integrality of sdp relaxations for sub-gaussian mixture models. In *Conference On Learning Theory*, pages 1931–1965, 2018.

[40] Gary William Flake, Steve Lawrence, C Lee Giles, and Frans M Coetzee. Self-organization and identification of web communities. *Computer*, (3):66–71, 2002.

[41] Santo Fortunato. Community detection in graphs. *Physics reports*, 486(3-5):75–174, 2010.

[42] Linton Freeman. The development of social network analysis. *A Study in the Sociology of Science*, 1, 2004.

[43] Joseph L Gastwirth. A probability model of a pyramid scheme. *The American Statistician*, 31(2):79–82, 1977.

[44] Joseph L Gastwirth and PK Bhattacharya. Two probability models of pyramid or chain letter schemes demonstrating that their promotional claims are unreliable. *Operations Research*, 32(3):527–536, 1984.

[45] Christophe Giraud and Nicolas Verzelen. Partial recovery bounds for clustering with the relaxed k-means. *Mathematical Statistics and Learning*, 1(3):317–374, 2019.

[46] Michelle Girvan and Mark EJ Newman. Community structure in social and biological networks. *Proceedings of the national academy of sciences*, 99(12):7821–7826, 2002.

[47] Gudmundur Gudmundsson. Community detection in large vector autoregressions. *Available at SSRN 3072985*, 2019.

[48] Olivier Guédon and Roman Vershynin. Community detection in sparse networks via grothendiecks inequality. *Probability Theory and Related Fields*, 165(3-4):1025–1049, 2016.

[49] Bruce Hajek, Yihong Wu, and Jiaming Xu. Achieving exact cluster recovery threshold via semidefinite programming. *IEEE Transactions on Information Theory*, 62(5):2788–2797, 2016.

[50] Xu Han and Atsushi Inoue. Tests for parameter instability in dynamic factor models. *Econometric Theory*, 31(5):1117–1152, 2015.

[51] Takayuki Iguchi, Dustin G Mixon, Jesse Peterson, and Soledad Villar. On the tightness of an sdp relaxation of k-means. *arXiv preprint arXiv:1505.04778*, 2015.

[52] Peter Jagers and Olle Nerman. The asymptotic composition of supercritial, multitype branching populations. *LECTURE NOTES IN MATHEMATICS-SPRINGER VERLAG-*, pages 40–54, 1996.

[53] Svante Janson. Functional limit theorems for multitype branching processes and generalized pólya urns. *Stochastic Processes and their Applications*, 110(2):177–245, 2004.

[54] Svante Janson. Asymptotic degree distribution in random recursive trees. *Random Structures & Algorithms*, 26(1-2):69–83, 2005.

[55] Stephen C Johnson. Hierarchical clustering schemes. *Psychometrika*, 32(3):241–254, 1967.

[56] Ioannis Karatzas and Steven Shreve. *Brownian motion and stochastic calculus*, volume 113. Springer Science & Business Media, 2012.

[57] John Frank Charles Kingman et al. The first birth problem for an age-dependent branching process. *The Annals of Probability*, 3(5):790–801, 1975.

[58] Ann E Krause, Kenneth A Frank, Doran M Mason, Robert E Ulanowicz, and William W Taylor. Compartments revealed in food-web structure. *Nature*, 426(6964):282, 2003.

[59] Florent Krzakala, Cristopher Moore, Elchanan Mossel, Joe Neeman, Allan Sly, Lenka Zdeborová, and Pan Zhang. Spectral redemption in clustering sparse networks. *Proceedings of the National Academy of Sciences*, 110(52):20935–20940, 2013.

[60] Amit Kumar and Ravindran Kannan. Clustering with spectral norm and the k-means algorithm. In *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*, pages 299–308. IEEE, 2010.

[61] Jing Lei, Alessandro Rinaldo, et al. Consistency of spectral clustering in stochastic block models. *The Annals of Statistics*, 43(1):215–237, 2015.

[62] Xiaodong Li, Yang Li, Shuyang Ling, Thomas Strohmer, and Ke Wei. When do birds of a feather flock together? k-means, proximity, and conic programming. *Mathematical Programming*, pages 1–47.

[63] Yu-Ru Lin, Yun Chi, Shenghuo Zhu, Hari Sundaram, and Belle L Tseng. Analyzing communities and their evolutions in dynamic social networks. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 3(2):1–31, 2009.

[64] Robert Liptser and Albert Nikolaevich Shiryayev. *Theory of Martingales*, volume 49. Springer Science & Business Media, 2012.

[65] Stuart Lloyd. Least squares quantization in pcm. *IEEE transactions on information theory*, 28(2):129–137, 1982.

[66] Yu Lu and Harrison H Zhou. Statistical and computational guarantees of lloyd's algorithm and its variants. *arXiv preprint arXiv:1612.02099*, 2016.

[67] Hosam M Mahmoud. Limiting distributions for path lengths in recursive trees. *Probability in the Engineering and Informational Sciences*, 5(1):53–59, 1991.

[68] Hosam M Mahmoud. The power of choice in the construction of recursive trees. *Methodology and Computing in Applied Probability*, 12(4):763–773, 2010.

[69] Hosam M Mahmoud and Robert T Smythe. On the distribution of leaves in rooted subtrees of recursive trees. *The Annals of Applied Probability*, pages 406–418, 1991.

[70] Hosam M Mahmoud and Robert T. Smythe. Asymptotic joint normality of outdegrees of nodes in random recursive trees. *Random Structures & Algorithms*, 3(3):255–266, 1992.

[71] Laurent Massoulié. Community detection thresholds and the weak ramanujan property. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, pages 694–703. ACM, 2014.

[72] Florence Merlevède, Magda Peligrad, and Emmanuel Rio. A bernstein type inequality and moderate deviations for weakly dependent sequences. *Probability Theory and Related Fields*, 151(3-4):435–474, 2011.

[73] Carl D Meyer. *Matrix analysis and applied linear algebra*, volume 71. Siam, 2000.

[74] JW Moon. The distance between nodes in recursive trees. *London Mathematical Society Lecture Notes*, 13:125–132, 1974.

[75] Tamás F Móri. The maximum degree of the barabási–albert random tree. *Combinatorics, Probability and Computing*, 14(3):339–348, 2005.

[76] Elchanan Mossel, Joe Neeman, and Allan Sly. Stochastic block models and reconstruction. *arXiv preprint arXiv:1202.1499*, 2012.

[77] Elchanan Mossel, Joe Neeman, Allan Sly, et al. Belief propagation, robust reconstruction and optimal recovery of block models. *The Annals of Applied Probability*, 26(4):2211–2256, 2016.

[78] Elchanan Mossel, Yuval Peres, et al. Information flow on trees. *The Annals of Applied Probability*, 13(3):817–844, 2003.

[79] Hwa Sung Na and Anatol Rapoport. Distribution of nodes of a tree by degree. 1970.

[80] Dietmar Najock and CC Heyde. On the number of terminal vertices in certain random trees with an application to stemma construction in philology. *Journal of Applied Probability*, 19(3):675–680, 1982.

[81] Mark EJ Newman. Detecting community structure in networks. *The European Physical Journal B*, 38(2):321–330, 2004.

[82] Mark EJ Newman and Michelle Girvan. Finding and evaluating community structure in networks. *Physical review E*, 69(2):026113, 2004.

[83] Bernt Øksendal. Stochastic differential equations. In *Stochastic differential equations*, pages 65–84. Springer, 2003.

[84] Symeon Papadopoulos, Yiannis Kompatsiaris, Athena Vakali, and Ploutarchos Spyridonos. Community detection in social media. *Data Mining and Knowledge Discovery*, 24(3):515–554, 2012.

[85] Jiming Peng and Yu Wei. Approximating k-means-type clustering via semidefinite programming. *SIAM journal on optimization*, 18(1):186–205, 2007.

[86] Pierre Perron et al. Dealing with structural breaks. *Palgrave handbook of econometrics*, 1(2):278–352, 2006.

[87] Pierre Perron and Xiaokang Zhu. Structural breaks with deterministic and stochastic trends. *Journal of Econometrics*, 129(1-2):65–119, 2005.

[88] Boris Pittel. Note on the heights of random recursive trees and random m-ary search trees. *Random Structures & Algorithms*, 5(2):337–347, 1994.

[89] Tai Qin and Karl Rohe. Regularized spectral clustering under the degree-corrected stochastic blockmodel. In *Advances in Neural Information Processing Systems*, pages 3120–3128, 2013.

[90] Zhongjun Qu and Pierre Perron. Estimating and testing structural changes in multivariate regressions. *Econometrica*, 75(2):459–502, 2007.

[91] Stuart A Rice. The identification of blocs in small political bodies. *American Political Science Review*, 21(3):619–627, 1927.

[92] Karl Rohe, Sourav Chatterjee, Bin Yu, et al. Spectral clustering and the high-dimensional stochastic blockmodel. *The Annals of Statistics*, 39(4):1878–1915, 2011.

[93] Sebastian Rosengren. A multi-type preferential attachment tree. *arXiv preprint arXiv:1704.03256*, 2017.

[94] Sheldon M Ross. *Introduction to probability models*. Academic press, 2014.

[95] Martin Royer. Adaptive clustering through semidefinite programming. In *Advances in Neural Information Processing Systems*, pages 1795–1803, 2017.

[96] Mark Rudelson and Roman Vershynin. Hanson-wright inequality and sub-gaussian concentration. *Electronic Communications in Probability*, 18, 2013.

[97] David Siegmund. *Sequential analysis: tests and confidence intervals*. Springer Science & Business Media, 2013.

[98] Robert T Smythe and Hosam M Mahmoud. A survey of recursive trees. *Theory of Probability and Mathematical Statistics*, (51):1–28, 1995.

[99] Myra Spiliopoulou. Evolution in social networks: A survey. In *Social network data analytics*, pages 149–175. Springer, 2011.

[100] Victor Spirin and Leonid A Mirny. Protein complexes and functional modules in molecular networks. *Proceedings of the National Academy of Sciences*, 100(21):12123–12128, 2003.

[101] Jimeng Sun, Christos Faloutsos, Spiros Papadimitriou, and Philip S Yu. Graphscope: parameter-free mining of large time-evolving graphs. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 687–696, 2007.

[102] Charles Truong, Laurent Oudre, and Nicolas Vayatis. Selective review of offline change point detection methods. *Signal Processing*, 167:107299, 2020.

[103] Lei Xu and Michael I Jordan. On convergence properties of the em algorithm for gaussian mixtures. *Neural computation*, 8(1):129–151, 1996.

[104] Tianbao Yang, Yun Chi, Shenghuo Zhu, Yihong Gong, and Rong Jin. Detecting communities and their evolutions in dynamic social networksa bayesian approach. *Machine learning*, 82(2):157–189, 2011.