ABSTRACT

ZHAO, WEI. Estimation of the Parameters in a Class of Dynamic Network Models. (Under the direction of Soumendra Lahiri).

Models have been proposed for static networks, such as stochastic block model, exponential network models. However, under many situations, the networks are time-varying. In this paper, we introduce three kinds of dynamic network models. For each model, we give the estimators of parameters based on the Maximum Likelihood method. Then, the asymptotic distributions are derived for the estimators using Martingale's Central Limit Theorem. The first model introduced is Markov Dynamic Network base model, in which edges are placed independently with the same probabilities conditioned on their former status, and the network size is kept unchanged. The second model is called Growing Size Markov Dynamic Network, where the first model is generalized by allowing fixed/random number of nodes becoming active/inactive at each time. The last model introduced is Multi-Classes Markov Dynamic Network. In this model, nodes are assigned to multiple classes in considering both the former status of the network and the nodes covariates. Simulations are conducted for each model to analyze the properties and behaviors of the estimators. The MathOverflow network data set and MovieLens Ratings network data set are introduced as applications of the second and third model.

Estimation of the Parameters in a Class of Dynamic Network Models

by
Wei Zhao

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Statistics

Raleigh, North Carolina

2019

APPROVED BY:

Wenbin Lu	Rui Song
Denis Pelletier	Soumendra Lahiri Chair of Advisory Committee

DEDICATION

This work is in honor of

Yu Yi (Mama), and Yinbao Zhao (Papa)

Juzi (kitty)

This work is also in memory of

Alba (kitty)

BIOGRAPHY

The author was born and raised in an industrial town called Dexing Copper Mine, Jiangxi, China in January 1992. She graduated from The Attached Middle School of Jiangxi Normal University in June 2009, and was admitted to Beijing Normal University to study Information Management afterwards. However, after one year of learning, she found her great interest in quantitative fields and changed her major to Statistics without any hesitation. Ranking No.1 among over 30 students, she got her Bachelor's degree in Statistics in 2014, and was offered a full Teaching Assistantship in North Carolina State University to pursue her PhD in Statistics. During the time in R aleigh, she joined the Sunny Dance Club and performed in Triangle Chinese New Year Festivals in 2016, 2017 and 2018.

She has been studying the dynamic network models under the advisory of Prof. Soumendra Lahiri since 2016. She will complete her PhD study in August, 2019.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Soumendra Lahiri for his guidance and help. His gift and proficiency in statistics and researches taught me a lot, and guided me all through my work of this thesis. Moreover, Dr. Lahiri, thank you so much for giving us such good lectures in the courses Advanced Probability and Advanced Time Series that I found my interest and finished my first project of dynamic network. In addition, I feel so lucky to have Dr. Lahiri as my advisor who is so generous to support me with Research Assistantship during my unfunded time. At last, I want to say to Dr. Lahiri, without your support, I wouldn't overcome the hard time when I was still looking for a job. Your encouragement and understanding gave me much more confidence in myself.

Here I also want to express my sincere thankfulness to all my committee members, Dr. Rui Song, Dr. Denis Pelletier and Dr. Wenbin Lu for their advice on this thesis. Thank you for giving me so much good advice during my Oral Exam and pointing out the problems of my work. In addition, I should say that it was their understanding so that I could re-schedule my graduation plans to have enough time to prepare for the job interviews and finally got my dream job. Dr. Song, thank you so much for being my advisor during my first year in NCSU when I was new in the department and knew little about the life in U.S.. Dr. Pelletier, thanks a lot for giving me such impressive lectures of Econometrics and being both my academic committee and GSR. Dr. Lu, thank you for your precious advice on my third dynamic network project. Also thanks so much for giving me a lot of advice during my graduate study when you are the DGP.

Besides my advisor and committees, I also want to thank my summer intern managers Dr. Eric Song from Seattle Genetics and Dr. Lawrence Wang from Google. It was their recognition of my abilities that gave me the chances to be an intern in their companies. Their guidance during the internship helped me gain experiences in industry and apply my knowledge in statistics to practice. Those experiences helped me succeed in the interviews during my full-time job searches. The time, both in Seattle and in the Bay Area, is so unforgettable that it will be engraved in my mind forever.

Last but not the least, I am eternally grateful to my parents for their generous love and support both in my study and my life. They kept encouraging me whenever I was upset. Although they are both in China, oceans and hours apart, I always feel that they are here by my side. I also want to thank my two lovely roommates, Wanying Ma and Yuan Tian. With them, I learned to cook, to raise

a cat, and to be a better one.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER

1

INTRODUCTION

Network data analysis has become an important branch of statistical science. It originally comes from graphs in mathematics modeling pairwise relations between objects. A *graph*, which is denoted by $G = (V, E)$, is basically consist of *vertices* and *edges*, where an edge is a pair of distinct vertices $i, j \in V$ who are linked. The study of graph model was brought to the statisticians eyes in recent years as there has been an explosion of graph data, and statistical tools are becoming more and more important in analyzing graphs. However, it is more often referred to as *Network Data* in statistical analysis of instead of graph, and *nodes* instead of vertices.

There are a broad applications of network data analysis in multiple disciplines such as biological, physical, computational, and social sciences. Studying the network model can help us know deeper about user behaviors, spread of diseases, then adapt the services or products from a personal aspect, while considering the network effect. For example, if a game company attempts to understand the underlying drivers and root causes of customers spending more hours with their portfolio of games, the network effect is never negligible. Game players can be largely affected by their friends or online social medias, especially for group games. In this online game networks, players are considered as different nodes, and the interactions or associations among players are edges. Another application of network model is recommending system. Online merchants or music apps can benefit a lot by building a social network model for their products. In this kind of networks, the products are treated as nodes, the co-buying links among products are edges.

Statistician's interests on networks mostly lie in analyzing distributions of network metrics, sampling methods of network data, network modeling, network topology inferences and predictions.

Network Characteristics

There has been a great effort on studying the network structure (or graph topology) in mathematics or computer science fields. Such are descriptive tasks that require studying tools much different from statistical tools. Analyzing the distributions of metrics that characterizing network structure looks more statistical. However, different metrics attracts attentions of researchers in different fields. For example, people who study Social Networks may be interested in triangles (small relation circles) of the network. [KC14] breaks the topic of featuring networks into two areas, one focuses on characterizing nodes and edges properties, another focuses on characterizing network cohesion.

An important node-edge features is *degree*, which counts the number of edges that connect to a node. Statistical analysis mainly focuses on the degree distribution, which is the distribution of degrees of all nodes in the network, ranging from 0 to total nodes -1. Another frequently referred node-edge feature is *centrality*, which characterizes how important a node is in the network. There are multiple way of measuring centrality. One example is the *Closeness Centrality*, where the node being similar to the most other nodes are consider to be the center.

The network cohesion features includes network density, connectivity, groups, assortativity, and so on. The *density* describes how dense the network edges are among a group of nodes or the entire network. A complete network, or a *clique*, which is connective between each pair of nodes, is considered to be the densest. Researchers are often interested in the number of n-cliques a network has. The *connectivity* measures how connective a network is in a global perspective. A connected part is a sub-network where every node is reachable to the other. The networks *groups* is another important feature characterizing network cohesion. It forms an extensive discussion about the network partition methods including hierarchical clustering and spectral analysis.

Sampling Methods

When it is impossible or too expensive to observe the entire network, sampling is necessary. However, due to the special structure of network data, it differs a lot from traditional statistical sampling. An intuitive way is random sampling pure nodes or edges first, then the related edges among sampled nodes or the nodes connected to the sampled edges are then taken directly from the network. Such ways re referred to as *Induced or Incident Sampling*. Another sampling method is *Star Sampling*, which just as its name implies, samples the nodes and its induced edges, together with the nodes connected to the edges. An extension to the star sampling is *Snowball Sampling*. It samples the nodes, their induced edges, then the incident nodes and their induced edges continuously like rolling a snowball until the Kth stage.

When sampled network data is available, the traditional statistical inference theories are used to estimate the network properties.

Network Models

Much of statistical work on network data is focusing on modeling. This topic of work falls in a number of important classes. Some is network motifs. Among those is the fundamental model, *Random Graph Model* originally proposed by [ER59], where edges are placed between pairs of nodes identically and independently with a fixed probability p. Later models such as *Stochastic Blockmodel* and *Latent Space Model* are based on the random graph model. The *Exponential Random Graph Model* is also a generalization of the random graph models. However, it generalizes the independence among edges to local dependence, and assumes an exponential family of probability distributions.

Another class of models addresses the generative mechanism, such as the *Network Growth*

Models. This kind of models are mostly used on dynamic networks, which address the changing of network structure in time. Such models are first proposed in describing the growth of World Wide Web. They are also widely used in biological sciences. The *Preferential Attachment Models* and *Copying Models* are typical network growth models.

Inferences and Predictions

Once the data and model are constructed, the key work falls on estimation of model parameters and prediction on unknown network structures. Our work in this paper mainly focuses on modeling and inferring on dynamic networks. Then real data are applied and predictions on the data are made based on proposed model and estimations.

1.1 Notations

Typically, we use $G(V,E)$ to refer to a network graph in statistical modeling, where $V = \{1,...,n\}$ stands for the vertices(nodes) set, and E stands for the edges set. The superscript t is added when studying dynamic network graph models. That is, $G(V^t, E^t)$ stands for the dynamic network status at time t. The link structure of a network is represented by an adjacency matrix:

$$\boldsymbol{X} = \begin{bmatrix} 0 & X_{12} & \dots & X_{1n} \\ x_{11} & 0 & \dots & X_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & 0 \end{bmatrix}.$$

In this paper, we only talk about undirected, unweighted network, where the adjacency matrix is symmetric with 0's on the diagonal, and 0 or 1 off diagonal. If $X_{ij} = 1$, it stands that nodes i and j are connected. Otherwise, they are not connected.

In the statistical inference on link prediction of network model, the studies are focused on the adjacency matrix. In dynamic network, $\{X^t\}_{1\le t\le T}$ is a series of adjacency matrix at each time snapshot t.

The node degree $d_i = \sum_{j \in V} X_{ij}$ counts the number of edges that are connected to nodes i. In an directed network, d_i can be divided to d_i^{in} and d_i^{out}, where in stands for edges pointing toward i, and out stands for edges pointing from i towards other nodes. The citation network is an example of directed network, where the edges point from the paper to the cited one. Our models discussed in later chapters are undirected networks. Such includes friendship network, where the node degree is simply the number of friends of a person i. The degree distribution of a network is the distribution of the series $\{d_1, ..., d_n\}$. It describes how connective among nodes the network is.

The Network Growth Model is one kind of models which bases the evolution mechanism on the degree distribution. The Preferred Attachment mechanism adds new node with degree $d \geq 1$. The Copying mechanism adds a new node to the existed network with randomly chosen edges of a randomly chosen existed node. Both mechanism grow the network to a power-law degree, which suggests that the degree distribution

$$f(d) \propto d^{-\alpha}.$$

The network density, which is denoted by

$$\rho = \frac{|E|}{|V|(|V|-1)/2},$$

is basically the total edges divided by total potential edges, describes how dense the network edges are. A visualization of two toy networks with different densities are show in Fig.1.1.

1.2 Organizations

The paper is organized as follows. Chapter 2 are literature reviews on existing famous network models and existing estimation methods of those models.

Chapter 3 introduces a first modeling of dynamic network called Markov Dynamic Network. It is a generalization of the random graph model bases on the idea that from the time scale, the linkages of nodes follow a Markov processes. For any given nodes pair, if there is no edge in between

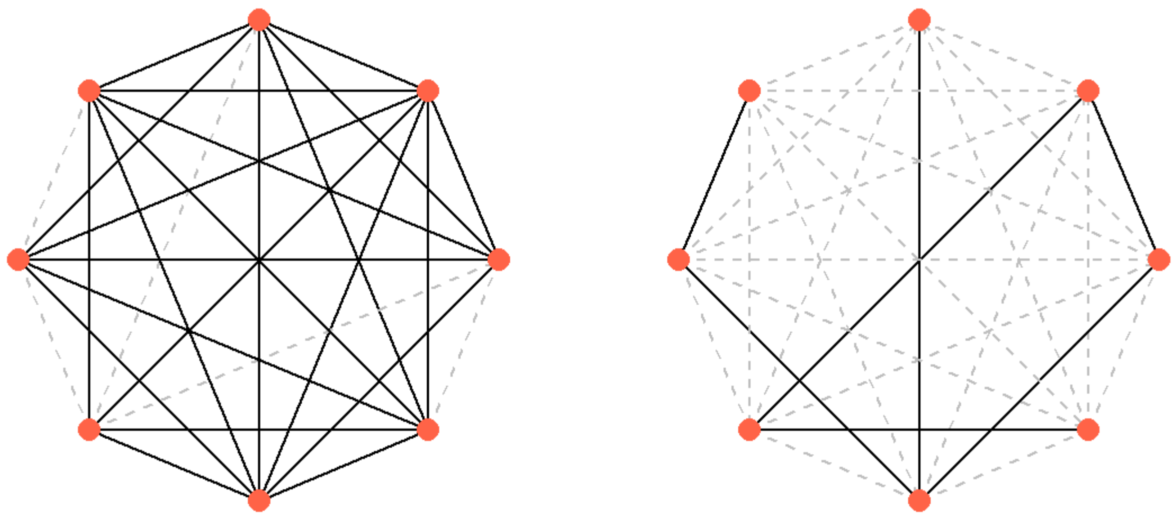


Figure 1.1 Visualization of Two Toy Networks. The solid black line are real edges, while the dash grey line are potential edges. Both networks have 8 nodes in total. The left one is an example of dense network with 23 edges in total. Its nodes' degrees are $\{6,5,4,7,6,5,7,6\}$. Its density is $\frac{23}{28} = 0.82$. The right one is a sparse network with 7 edges. Its nodes' degrees are $\{1,2,2,3,1,2,2,1\}$. Its density is $\frac{7}{28} = 0.25$.

before, then the edge variable follows Bernoulli(p), otherwise, Bernoulli(q). This model assumes that all edges follow the same pattern across the whole network, which is usually not the case in reality. Then, we give Maximum Likelihood Estimators for the parameters and analyze the limiting distribution of the estimators by applying Martingale Central Limit Theorem.

In Chapter 4, We look into a more general case where the network size is growing over time. We provide two models: one assumes that nodes coming in a constant rate, the other is a more generalized that assume node come and leave in Poisson distributions with coming rate larger than leaving rate. Consequently, The network size grows to infinity as time goes to infinity. Asymptotic distributions are provided for the MLE of the models. Simulation results are then provided for the estimation of the Growing Size Models.

The third model (Multi-Classes Markov Dynamic Network) is proposed in chapter 5. We borrow the idea of Stochastic Block Model, but with some modification that each node can belong to multiple blocks (classes). This corresponds with the fact that in some fields like sociology, people can belong to multiple organizations, or join different groups. In addition, different from the static cases, in dynamic cases, the group memberships vary in time. We model the group membership with a random class vector for each nodes as the function of linkages at the former time. Thus forms a Markov Chain.

1.3 Summary of Results

We have shown that for the base model, Markov Dynamic Network:

1. The network tends to be dynamic stable;

2. The Markov chain formed by the edges is irreducible and ergodic. The status 0 and 1 are recurrent;

3. The asymptotic distribution of the maximum likelihood estimator normal with a constant variance that is the function of the keeping probability p and appearing probability q. The convergence rate is T.

For the second model, Fixed rate Markov Dynamic Network, and its generalized model, Random Process Markov Dynamic Network, we have shown that:

1. The network density goes in probability to a constant function of the parameters. To be specific, the limiting network density is $\frac{q}{1-p+q}$ under both models. The limiting density is independent of the how node come into or leave the network;

2. The asymptotic distribution of MLE for keeping probability p and appearing probability q are normal with a constant variance that is the function of p and q. That for the forming probability a is normal with a constant variance of function of a only. The convergence rate is T^3;

3. The application to the MathOverflow interaction network shows that the generalized model fits well, and shows us that new registered users and those haven't interact before, are much less likely to interact with each other.

Finally, for the third model, Multi-Classes Markov Dynamic Network, we get the following results:

1. The network tends to be dynamic stable. The Markov Chain formed by the vectorized adjacency matrix has an unique stationary distribution;
2. The MLE is asymptotically normal with convergence rate of T;
3. The application to the MovieLens ratings network shows that the multi-classes model fits well. Specially, the time-varying class model outperforms the fixed class model in that it better captures the latent features of the movies. The class connection probability matrix and the mean class vectors are both generated. The results show that movies in class 1 has higher scores in adventure/action/IMAX genres. They are more likely to be co-rated, which means people who watched these kind of movies tend to watch similar movies. However, movies in class 4 has higher scores in horror and documentary than others. But this class has lower within-class connection probability than between-classes probabilities.

CHAPTER

2

LITERATURE REVIEWS

2.1 Existing Network Models

Network models has been widely studied by researchers for dealing with relational data in sociology, technology and biology. The Random Graph Model for static networks addresses the randomness of edges, but ignores the diversity of nodes. To solve this issue, [Hol83] proposed a new model called Stochastic Block model (SBM), where nodes are assigned into different blocks, and those from the same block are considered statistically the same. More specifically:

Definition 2.1.1. ***(Stochastic Blockmodel)*** *Suppose* $X \in \{0,1\}^{n\times n}$ *is the adjacency matrix of a random network with* n *nodes. Define* $C \in \{0,1\}^{n\times k}$ *be the fixed class matrix such that it has exactly one 1 in each row and at least one 1 in each column. (i.e. each node belongs to exactly one class and there is at* *least one node in each class). Then*

$$P(X_{ij} = 1 | c_i, c_j) = \mathrm{E}(X_{ij} | c_i, c_j),$$

where c_i *is the* i*th row of class matrix* C*. Then, the Stochastic Blockmodel satisfies that*

$$\mathrm{E}(X|C) = CBC^T,$$

where $B \in [0,1]^{k \times k}$ *is the class connection probability matrix which is full rank and symmetric.*

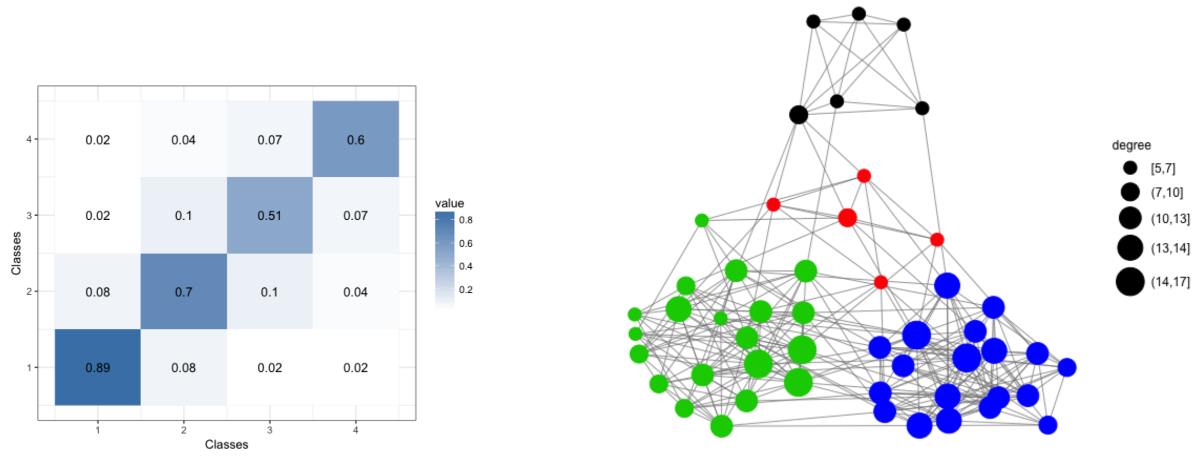
Figure 2.1 shows an example network generated from the Stochastic Blockmodel with 4 blocks (classes). The class connection probability matrix is shown in the heat-map in (a), where we can tell that the within-group connection probability is much higher than the between-group connection probability. The network generated is shown in (b). The color of the nodes stands for the group the nodes belong to. It can be seen easily that nodes are closely connected if they belong to the same group. The estimation of the SBM mainly includes two steps: 1. grouping of nodes using the adjacency matrix; 2. Estimation of the class connection matrix. The first step shows its importance in the estimation of SBM and we will discuss it later.

A more general model, Latent Space Model (LSM), is proposed by [Hof02] to describe social network, where the edge probability depends on the "positions" of nodes in an unobserved "social space".

Definition 2.1.2. ***(Latent Space Model)*** *Suppose* $X \in \{0,1\}^{n \times n}$ *is the adjacency matrix of a random network with* n *nodes. Define* $Z \in \mathbb{R}^{n \times k}$ *is the latent space matrix such that the row vectors are independent in* $\mathbb{R}^k$*. Assume that a probability distribution on* X *has conditional independence relationship*

$$P(X|Z) = \prod_{i<j} P(X_{ij} | z_i, z_j),$$

$$\textit{and } P(X_{ii} = 0)\ \forall i,$$



(a) (b)

Figure 2.1 Visualization of SBM. (a): Heat map of the classes connection probability matrix. (b): Example of a network generated by SBM using the class connection matrix of the left.

where z_i is the ith row of latent space matrix Z. Then, it is an undirected Latent Space Model. The Stochastic Blockmodel is a special case of Latent Space Model with the latent space matrix $Z = C$.

To be specific, the probability of an edge in a LSM are determined by the latent distance between two nodes, which is determined by the latent vectors Z_i. Likelihood-based method can be used to estimate the latent distances and thus the latent vectors once the distance structure is defined.

It should be brought into consideration that under many situations, the networks are time-varying. In recent years, dynamic networks has drawn more and more attention of researchers. [Han15] view dynamic network as multi-layer graphs, and the multi-graph SBM is analyzed with an exploration of the asymptotic properties of spectral clustering estimator and the maximum likelihood estimate (MLE).

Definition 2.1.3. ***(Multi-layer Blockmodel)*** *Suppose $X^t \in \{0,1\}^{n\times n}$ is the adjacency matrix of a random network with n nodes at time layer t. Define $C \in \{0,1\}^{n\times k}$ be the fixed class matrix the same as in Stochastic Block Model. Then, the Multi-layer Blockmodel is that*

$$\mathrm{E}(X^t|C) = CB^tC^T,$$

where $B^t \in [0,1]^{k\times k}$ is the class connection probability matrix at time layer t which is full rank and symmetric.

However, this multi-layer model ignores the time scale relationships between the "layers". Another limitation is the assumption that block structure is fixed overtime. This is not true in many real situations when nodes may switch classes. In addition, The class connection probability matrix is changing overtime, making the number of parameters grows linearly as time evolves. This may add to the computation time greatly.

Later works regarding dynamic stochastic block model mainly focus on block detection and clustering of the nodes. [MM17] proposed a clustering method on estimating time related Stochastic Blockmodel, which allows class switching across time while assuming that most of the nodes do not change classes across two different time steps. Furthermore, they assume the class vector $C_i = \{C_i^t\}_{0\le t\le T}$ to be an irreducible, aperiodic stationary Markov chain with a stationary distribution. Conditional on C^t, the distribution of adjacency matrix X^t follows a Stochastic Blockmodel with the class connection probability matrix varying across time.

[Zha17a] addresses the time scale interaction of networks, and generalizes some standard network models with the assumption that the presence and absence of edges are governed by Markov processes. A blockmodel based dynamic network is also introduced, where one assumes that the appearance and disappearance of edges are governed by continuous-time Markov processes. Those probabilities are depend on an rate parameters that can depend on properties of the nodes. The parameters can be estimated from the adjacency matrix using likelihood based methods. They provide some algorithm for achieving the MLE, but the analysis of the asymptotic properties of the estimators is not included. One of our tasks is to analyze the long time behaviors of the networks under different dynamic models, and then give an asymptotic distribution of MLE.

Another issue that achieves less attention so far is the network size in dynamic network. Most of the dynamic network researches are dealing with fixed size networks as the models discussed above. In other words, the total number of nodes are fixed. However, it is usually an unrealistic assumption in practical. For example, in the early stage of a social network, new users continuously join the network and form relationships with the existing ones. Another example is when studying the long-time interaction of students in a college, the entering of new students and leaving of graduating students should also be considered. We give an discussion about a generalization of the Markov Dynamic Network by allowing the continuous coming of new nodes, with the assumption that existed nodes do not disappear at all.

2.2 Existing Model Estimation Methods

Non-parametric graphon estimations has been proposed by [WO13] for estimating kernel-based network models. The graphon is non-negative symmetric, measurable and bounded function that represents a discrete network as an infinite-dimensional analytic object. It can be viewed as a heat map of the expected adjacency matrix independently of the network size. Usually a graphon $f(x,y)$ is defined on $(0,1)^2$. Each node i will be mapped uniformly onto a position on $\xi_i \in (0,1)$ identically. Then the edge probability p_{ij} can be specified by

$$p_{ij} = \rho_n f(\xi_i, \xi_j),\ \{\xi_1, ..., \xi_n\} \overset{i.i.d.}{\sim} \text{Uniform}(0,1),\ \int\int f(x,y)dxdy = 1.$$

It is stated that this graphon function is unique in the exchangeable sense. In estimating the graphon, the basic method is approximating the function by class connection probability matrix in Stochastic Blockmodel. Once the nodes are grouped, they are permuted such that those in the same group are in neighbour positions. The last step is fitting a block model, and the estimated class connection probability matrix is a discretized approximation of the graphon function.

A famous method in estimating the node classes is using spectral clustering of the adjacency matrix proposed by [Roh11]. For a given network adjacency matrix $X_{n\times n}$ where the diagonals are zeros, define matrix $L_{n\times n}$ as

$$L = D^{-1/2} X D^{-1/2},$$

where $D_{n\times n}$ is a diagonal matrix with d_{ii} be the row sums of X, i.e. the degree of node i. The spectral clustering of the nodes are done in the following two steps:

a) Find out the eigenvectors corresponding to the first k largest absolute eigenvalues of matrix L. Use the k vectors to form an eigenmatrix $E_{n\times k}$.

b) Treat each row i of the matrix E as position of node i in the space $\mathbb{R}^k$. Use k-means clustering to group the rows. Output the class of row i as the class of node i.

Such clustering is efficient and has wide usage with little assumptions. The idea behind this method is that the largest k eigenvectors expressed the k main directions of the whole adjacency matrix in a $n\times n$ space. The nodes points to the similar directions are more likely to be grouped into the same class.

[Zha17b] proposed a neighbourhood smoothing method on estimating the class connection probability matrix. They define a nodeâĂŹs neighborhood to consists of nodes with similar rows in the adjacency matrix. However, this method requires much more computation than the spectral method, though it shows comparable estimation results.

2.3 Convergence Theorems for Markov Chains

[AL06] gives the definition of Markov Chain in countable state space and some of important properties as follows.

Definition 2.3.1. ***(Markov Chain in countable State Space)*** *Suppose* $\{X_t\}_{t=0}^T$ *is a series of random variables on probability space* $(\Omega, \mathscr{F}, \mathscr{P})$. *The state space* $\mathscr{S} = \{i_1, ..., i_K\}, K < \infty$ *is countable. If*

- $P(X_0 = i_k) = \mu_k, \forall\ k;$
- $P(X_{t+1} = i_b | X_t = i_a, X_{t-1} = i_{a_{t-1}}, ..., X_0 = i_{a_0}) = p_{ab} \forall\ i_a, i_b, i_{a_{t-1}}, ..., i_{a_0} \in \mathscr{S}$ *and* $t = 0, 1, 2, ..., T$.

Then $\{X_t\}_{t=1}^T$ *is Markov Chain with stationary probability* $\mathbf{P} = (p_{ab})$, *initial distribution* μ, *and state space* $\mathscr{S}$.

Definition 2.3.2. ***(Irreducibility)*** *The transition probability* $\mathbf{P} = (p_{ab})$ *of a Markov Chain is called irreducible if any pair of states* $i, j \in \mathscr{S}$ *are communicate. Here communicate means the states* i *and* j *and go to each other in finite steps with positive probabilities.*

Definition 2.3.3. ***(Recurrence)*** *The state i of a Markov Chain is called recurrent if $P(T_i < \infty) = 1$, where T_i is the hitting time of state i, which is the first time after 0 that the chain enters i. Furthermore, a recurrent state i is called positive recurrent if $\mathrm{E}_i(T_i) < \infty$.*

The basic ergodic theorem and its proof is then given in the [AL06].

Theorem 2.3.1. ***(Basic Ergodic Theorem)*** *Suppose a Markov Chain with countable state space $\mathscr{S}$ has transition probability $\mathbf{P}$. The transition probability is irreducible, let one state to be positive recurrent, then*

1. *All states are positive recurrent;*
2. *There exists a stationary probability π;*
3. *For any initial probability, and any state $i \in \mathscr{S}$,*
    - $\frac{1}{T+1}\sum_{t=0}^{T} P(X_t = j) \to \pi_j$;
    - $L_t(j) = \frac{1}{T+1}\sum_{t=0}^{T} \delta(X_t = j) \to^p \pi_j$, *where $L_t(j)$ is the empirical distribution at t.*

2.3.1 Martingale CLT

[AL06] introduces Martingale CLT that is useful in the analysis of estimators in our models. First, the concept of Martingale Difference Array (MDA) is introduced.

Definition 2.3.4. *Let $(\Omega, \mathscr{F}, \mathscr{P})$ be a probability space. Let $\{X_i\}_{i=1}^n$ be r.v.s on $(\Omega, \mathscr{F}, \mathscr{P})$, and $\{\mathscr{F}_i\}_{i=1}^n$ be sub σ-fields of $\mathscr{F}$.*

1. *If $\mathscr{F}_i \subset \mathscr{F}_{i+1}$, $\forall i = 1, 2, ..., n$, then $\mathscr{F}_i$ is called a filtration.*
2. *$\{X_i\}_{i=1}^n$ is called a Martingale Difference Array if*
    - *X_i is $\mathscr{F}_i$-measurable, $\forall i = 1, 2, ..., n$.*
    - $\mathrm{E}(X_i|\mathscr{F}_{i-1}) = 0$, $\forall i = 1, 2, ..., n$, *where $\mathscr{F}_0 = \{\varnothing, \Omega\}$.*

Then, it is easy to introduce Martingale Central Limit Theorem. We will use the Martingale CLT in Chapter 3-5 to prove the consistency and asymptotic normality of the estimators.

Theorem 2.3.2. ***(Martingale CLT)*** *For* $n \geq 1$*, let* $\{X_i\}_{i=1}^n$ *be an MDA with respect to* $\{\mathscr{F}_i\}_{i=1}^n$ *such that*

$$\frac{1}{n}\sum_{i=1}^{n} \mathrm{E}(X_i^2|\mathscr{F}_{i-1}) \xrightarrow{p} \sigma_\infty^2 \in (0,\infty),$$

and one of the following

- *Lindeberg's Condition:* $\forall \epsilon > 0$*,*

$$\frac{1}{n}\sum_{i=1}^{n} \mathrm{E}(X_i^2 \mathbb{1}(|X_i| > \epsilon)|\mathscr{F}_{i-1}) \xrightarrow{p} 0 \text{ as } n \to \infty;$$

- *Lyapunov's Condition:*

$$\frac{1}{n}\sum_{i=1}^{n} \mathrm{E}(X_i^4|\mathscr{F}_{i-1}) \xrightarrow{p} 0 \text{ as } n \to \infty,$$

is satisfied. Then,

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p} N(0,\sigma_\infty^2).$$

CHAPTER

3

MARKOV DYNAMIC NETWORK

3.1 Model Settings

[Zha17a] proposed 3 dynamic network models, including dynamic random graph model, the dynamic random graph with arbitrary expected degrees and the dynamic block model. The first model proposed was defined after the classical random graph $G(n, p)$ studied by [ER59]. This model describes the varying of the network by adding an edge to the previously disconnected nodes with probability α, or not with probability $1-\alpha$. Similarly, the model deletes the edge between the previously connected nodes with the probability β, or not with probability $1-\beta$. And at the starting time point $t = 0$, it is a random graph $G(n, p)$.

In this chapter, we further study the dynamic random graph model (hereinafter Markov Dynamic Network model or MDN model). Suppose X_t is an $n \times n$ adjacency matrix for network graph G_t,

$t=1,2,...,T$. Then

$$X_{ij}^t = X_{ij}^{t-1} \circ B^t + (1-X_{ij}^{t-1}) \circ C^t, \tag{3.1}$$

where $B^t \sim \text{Bern}(p)$ and $C^t \sim \text{Bern}(q)$ are Bernoulli random variables at time t (Assume i.i.d.). We have assumed that the initial distribution is nonrandom. In addition, we assume that X_{ij}^t are independent with respect to the index $i < j$, i.e., the edges are independent with each other for different nodes pairs.

Although [Zha17a] has given the Maximum Likelihood Estimator, the properties of the estimators have not been studied yet. Our goal in this chapter is mainly extending their work by studying the behavior of the MLE as $T \to \infty$ of the model.

3.2 Markov Properties

3.2.1 Irreducibility, Ergodicity and Recurrence

From equation 3.1, it is easy to have that $\{X_{ij}^t : t \geq 0\}$ is a Markov Chain on a state space of $\mathbb{S} = \{0,1\}$, with the transition probability matrix

$$P = \begin{bmatrix} 1-q & q \\ 1-p & p \end{bmatrix}, \quad \lim_{T\to\infty} P^T = \begin{bmatrix} \frac{1-p}{1-p+q} & \frac{q}{1-p+q} \\ \frac{1-p}{1-p+q} & \frac{q}{1-p+q} \end{bmatrix},$$

Thus the invariant probability $\pi = (\frac{1-p}{1-p+q}, \frac{q}{1-p+q})$. In addition, the eigenvalues of matrix P are 1 and $p-q$, and the corresponding eigenvectors are $(1,1)'$ and $(-q, 1-p)'$. Using the eigen-decomposition, we have

$$P = U \begin{bmatrix} 1 & 0 \\ 0 & p-q \end{bmatrix} U^{-1} \Rightarrow \; P^n = U \begin{bmatrix} 1 & 0 \\ 0 & (p-q)^n \end{bmatrix} U^{-1},$$

where,

$$U = \begin{bmatrix} 1 & -q \\ 1 & 1-p \end{bmatrix}, \quad U^{-1} = \begin{bmatrix} \frac{1-p}{1-p+q} & \frac{q}{1-p+q} \\ \frac{-1}{1-p+q} & \frac{1}{1-p+q} \end{bmatrix}.$$

First, since the 1-step transition probabilities of $0 \to 1$, and $1 \to 0$ are $p_{00} = 1-q > 0$ and $p_{11} = p > 0$ respectively, the chain is *irreducible*, and thus *ergodic*. Second, the n-step transition probabilities of $0 \to 1$, and $1 \to 0$ are

$$p_{00}^{(n)} = \frac{1-p}{1-p+q} + \frac{q}{1-p+q}(p-q)^n,$$
$$p_{11}^{(n)} = \frac{q}{1-p+q} + \frac{1-p}{1-p+q}(p-q)^n.$$

Hence, states 0 and 1 are *recurrent*, as

$$\lim_{n\to\infty} \mathrm{E}_0\, N_n(0) = \sum_{n=1}^{\infty} p_{00}^{(n)} = \sum_{n=1}^{\infty}\Big[\frac{1-p}{1-p+q} + \frac{q}{1-p+q}(p-q)^n\Big] = \infty,$$
$$\lim_{n\to\infty} \mathrm{E}_1\, N_n(1) = \sum_{n=1}^{\infty} p_{11}^{(n)} = \sum_{n=1}^{\infty}\Big[\frac{q}{1-p+q} + \frac{1-p}{1-p+q}(p-q)^n\Big] = \infty.$$

Furthermore, they are *positive recurrent* since

$$\lim_{n\to\infty} \mathrm{E}_0\, L_n(0) = \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} p_{00}^{(k)} = \frac{1-p}{1-p+q} > 0,$$
$$\lim_{n\to\infty} \mathrm{E}_1\, L_n(1) = \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} p_{11}^{(k)} = \frac{q}{1-p+q} > 0.$$

From [AL06] Theorem 14.1.16, $(\mathbb{S}, \mathbf{P})$ is irreducible and state 0 and 1 are positive recurrent. Thus,

$$\frac{1}{T}\sum_{t=1}^{T}(1-X^t) = L_T(0) = \frac{1}{T}\sum_{t=1}^{T}\delta\{X^t = 0\} \xrightarrow{p} \pi_0 = \frac{1-p}{1-p+q}, \tag{3.2}$$
$$\frac{1}{T}\sum_{t=1}^{T}X^t = L_T(1) = \frac{1}{T}\sum_{t=1}^{T}\delta\{X^t = 1\} \xrightarrow{p} \pi_1 = \frac{q}{1-p+q}. \tag{3.3}$$

3.3 Model Estimation

3.3.1 Maximum Likelihood Estimators

The likelihood of p and q given X_t and X_{t-1} is

$$L_t(p,q|X_t,X_{t-1}) = \prod_{i<j} \left[p^{x_{ij}^t}(1-p)^{1-x_{ij}^t}\right]^{x_{ij}^{t-1}} \left[q^{x_{ij}^t}(1-q)^{1-x_{ij}^t}\right]^{1-x_{ij}^{t-1}}.$$

Then the likelihood of p and q given $\{X_0, X_1, ..., X_T\}$ is

$$L(p,q|X_0,...,X_T) = \prod_{t=1}^{T}\prod_{i<j} \left[p^{x_{ij}^t}(1-p)^{1-x_{ij}^t}\right]^{x_{ij}^{t-1}} \left[q^{x_{ij}^t}(1-q)^{1-x_{ij}^t}\right]^{1-x_{ij}^{t-1}}.$$

Take the log and we get

$$\begin{aligned} l(p,q|X_0,...,X_T) &= \sum_{t=1}^{T}\sum_{i<j}\Big[x_{ij}^{t-1}x_{ij}^t \ln p + x_{ij}^{t-1}(1-x_{ij}^t)\ln(1-p) \\ &\quad + (1-x_{ij}^{t-1})x_{ij}^t \ln q + (1-x_{ij}^{t-1})x_{ij}^t \ln(1-q)\Big], \\ \frac{\partial l}{\partial p} &= \frac{1}{p}\sum_{t=1}^{T}\sum_{i<j} x_{ij}^{t-1}x_{ij}^t - \frac{1}{1-p}\sum_{t=1}^{T}\sum_{i<j} x_{ij}^{t-1}(1-x_{ij}^t), \\ \frac{\partial l}{\partial q} &= \frac{1}{q}\sum_{t=1}^{T}\sum_{i<j}(1-x_{ij}^{t-1})x_{ij}^t - \frac{1}{1-q}\sum_{t=1}^{T}\sum_{i<j}(1-x_{ij}^{t-1})(1-x_{ij}^t). \end{aligned}$$

Set the partial derivative to 0, we get the maximum likelihood estimator for p and q as

$$\hat{p}_{MLE} = \frac{\sum_{t=1}^{T}\sum_{i<j} x_{ij}^{t-1}x_{ij}^t}{\sum_{t=1}^{T}\sum_{i<j} x_{ij}^{t-1}} = \frac{\sum_{i<j}\sum_{t=1}^{T} x_{ij}^{t-1}x_{ij}^t}{\sum_{i<j}\sum_{t=1}^{T} x^{t-1}}, \tag{3.4}$$

$$\hat{q}_{MLE} = \frac{\sum_{t=1}^{T}\sum_{i<j}(1-x_{ij}^{t-1})x_{ij}^t}{\sum_{t=1}^{T}\sum_{i<j}(1-x_{ij}^{t-1})} = \frac{\sum_{i<j}\sum_{t=1}^{T}(1-x_{ij}^{t-1})x_{ij}^t}{\sum_{i<j}\sum_{t=1}^{T}(1-x_{ij}^{t-1})}. \tag{3.5}$$

3.3.2 Asymptotic Distribution of MLE

Theorem 3.3.1. *Let $\{X_{ij}^t\}_{t>0}$ defined on $(\Omega, \mathscr{A}, \mathscr{P})$ follows the Markov Dynamic Network model, assuming that the initial network is nonrandom, then the MLE of keeping probability p is consistent and the asymptotic distribution*

$$\sqrt{\frac{Tn(n-1)}{2}}(\hat{p}_{MLE} - p) \xrightarrow{d} N\Big(0, \frac{p(1-p)(1-p+q)}{q}\Big). \tag{3.6}$$

The MLE of appearance probability q is consistent and follows

$$\sqrt{\frac{Tn(n-1)}{2}}(\hat{q}_{MLE} - q) \xrightarrow{d} N\Big(0, \frac{q(1-q)(1-p+q)}{1-p}\Big). \tag{3.7}$$

As the proof of form (3.7) is quite similar to the proof of form (3.6), we will only show the proof of form (3.6), the proof of the other is omitted.

Proof. From the large number result we got in form (3.3), we have

$$\frac{1}{T}\sum_{t=1}^{T} X_{ij}^t \xrightarrow{p} \pi_1 = \frac{q}{1-p+q}.$$

As the edges are i.i.d.,

$$\frac{2}{Tn(n-1)}\sum_{i<j}\sum_{t=1}^{T} X_{ij}^t \xrightarrow{p} \pi_1 = \frac{q}{1-p+q}.$$

Now let

$$\begin{aligned} Y_{ij}^t &= X_{ij}^{t-1} X_{ij}^t \\ &= X_{ij}^{t-1}(X_{ij}^t - p) + pX_{ij}^{t-1} \\ &= Z_{ij}^t + pX_{ij}^{t-1}. \end{aligned}$$

Then

$$\hat{p}_{MLE} = \frac{\sum_{i<j}\sum_{t=1}^{T} x_{ij}^{t-1} x_{ij}^{t}}{\sum_{i<j}\sum_{t=1}^{T} x_{ij}^{t-1}} = \frac{\sum_{i<j}\sum_{t=1}^{T} z_{ij}^{t}}{\sum_{i<j}\sum_{t=1}^{T} x_{ij}^{t-1}} + p.$$

Let $\mathscr{F}_a = \sigma\langle X^t : t \le a, t \in \mathbb{Z}\rangle, -\infty \le a \le b \le \infty$. As X_{ij}^t are i.i.d. for each i and j, we leave out the subscript i, j for simplicity.

Now we show that Z^t is MDA. First, it is easy to see that Z^t is $\mathscr{F}_t$-measurable, $\forall t = 1, 2, ..., T$. Then, for all t,

$$\begin{aligned}
\mathrm{E}(Z^t|\mathscr{F}_{t-1}) &= \mathrm{E}\left[X^{t-1}(X^t - p)\middle|X^{t-1}, X^{t-2}, ..., X^0\right] \\
&= X^{t-1}\,\mathrm{E}\left[X^t - p\middle|X^{t-1}, X^{t-2}, ..., X^0\right] \\
&= X^{t-1}\left[X^{t-1}p + (1 - X^{t-1})q - p\right] \\
&= \left[(X^{t-1})^2 - X^{t-1}\right](p - q) \\
&= 0.
\end{aligned}$$

Thus Z^t is MDA. In addition, the conditional variance of Z^t is

$$\begin{aligned}
\sigma_t^2 &= \mathrm{E}\left[(Z^t)^2\middle|\mathscr{F}_{t-1}\right] \\
&= \mathrm{E}\left\{\left[X^{t-1}(X^t - p)\right]^2\middle|X^{t-1}, X^{t-2}, ..., X^0\right\} \\
&= (X^{t-1})^2\,\mathrm{E}\left[(X^t)^2 - 2pX^t + p^2\middle|X^{t-1}, X^{t-2}, ..., X^0\right] \\
&= X^{t-1}\,\mathrm{E}\left[(1 - 2p)X^t + p^2\middle|X^{t-1}, X^{t-2}, ..., X^0\right] \\
&= X^{t-1}\left[(1 - 2p)(X^{t-1}p + (1 - X^{t-1})q) + p^2\right] \\
&= p(1 - p)X^{t-1}.
\end{aligned}$$

From form (3.3), we get

$$\frac{1}{T}\sum_{t=1}^{T}\sigma_t^2 = \frac{1}{T}\sum_{t=1}^{T} p(1 - p)X^{t-1} \xrightarrow{p} \frac{pq(1 - p)}{1 - p + q}.$$

At last,

$$\frac{1}{T}\sum_{t=1}^{T}\mathrm{E}\Big[(Z^t)^2\mathbb{1}(|Z^t|>\epsilon\sqrt{T})\Big|\mathscr{F}_{t-1}\Big]\leq\frac{1}{T}\sum_{t=1}^{T}\mathrm{E}\Big[\mathbb{1}(|Z^t|>\epsilon\sqrt{T})\Big|\mathscr{F}_{t-1}\Big]$$
$$\to 0 \text{ as } T\to\infty.$$

Therefore, we have the limiting distribution of Z_{ij}^t as

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}Z_{ij}^t \xrightarrow{d} N\Big(0,\frac{pq(1-p)}{1-p+q}\Big).$$

As $\{Z_{ij}^t\}$ are i.i.d for each i, j, and by Slutsky's Theorem,

$$\sqrt{\frac{Tn(n-1)}{2}}(\hat{p}_{MLE}-p)=\sqrt{\frac{Tn(n-1)}{2}}\frac{\sum_{i<j}\sum_{t=1}^{T}Z_{ij}^t}{\sum_{i<j}\sum_{t=1}^{T}X_{ij}^t}\xrightarrow{d}N\Big(0,\frac{p(1-p)(1-p+q)}{q}\Big). \tag{3.8}$$

□

3.4 Simulation Study

3.4.1 Settings

In our simulation studies for MDN model, the networks are generated with sizes $N \in (20,100)$, and initial edge probability of 1/3. The total time we run for the dynamic network is 500. The keeping probability p, appearing probability q are set to be combinations of $p \in (0.9,0.5,0.2)$, $q \in (0.4,0.05)$. Figure 3.1 shows the change in network density over time for each settings.

3.4.2 Simulation Results

The simulation results are summarized in Table 3.1. From the simulation results, we can see that the MLEs are quite close to the true values of p. It is possible to get a good estimation with an MSE as small as 10^{-5} when $T = 50$. The variances of the estimators are also at the level of 10^{-5}, which is

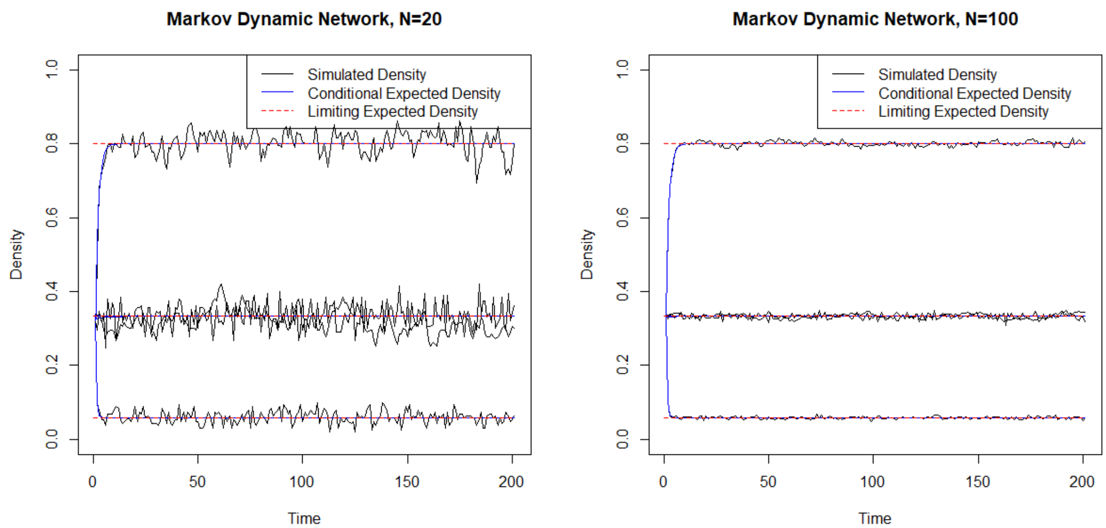


Figure 3.1 Plots of Network Density from MDN Models with sizes $N = 20$ or $N = 100$. We can tell from the graphs that the conditional density converge to its limit value shortly. The network density varies around the limit value with larger variation for smaller network size.

relatively small compared to the true parameter.

By looking at the results of coverage rates of the constructed confidence intervals, we found that the 95% theoretical CIs has a higher than expected coverage rate when T is small. This means that when T is small, the theoretical CIs tend to be conservative. However, as T goes larger, the coverage rates tend to be quite near 95%. This convergence tends to be faster when p and q are both larger. The Bootstrap method also gives good estimation on confidence intervals. Especially when T is small, it tends to have better estimation than the theoretical one. However, the Bootstrap method tends to be excessive with slightly lower coverage rate than 95%. In addition, the Bootstrap CI requires much heavier computation than theoretical method because of the repeated sampling.

Table 3.1 Simulation Results for Markov Dynamic Network. For each setting, the dynamic network is simulated for $t : 0 \to 200$, and Bootstrapped for 200 times to get the 95% CIs. The estimation results, coverage, variance and the MSE are then calculated from 500 repeated simulations.

Settings	T	$\hat{p}$	Cvr.T	Cvr.B	Var	MSE	$\hat{q}$	Cvr.T	Cvr.B	Var	MSE
$n = 20$	50	0.9001018	0.960	0.960	2.548e-05	2.544e-05	0.0501229	0.954	0.930	7.129e-06	7.130e-06
$p = .9$	100	0.9001643	0.940	0.934	1.473e-05	1.473e-05	0.0500193	0.934	0.918	3.917e-06	3.910e-06
$q = .05$	200	0.9000983	0.940	0.928	7.307e-06	7.302e-06	0.0500747	0.956	0.950	1.833e-06	1.835e-06
$n = 20$	50	0.8997754	0.928	0.936	1.234e-05	1.237e-05	0.4001240	0.964	0.938	1.167e-04	1.164e-04
$p = .9$	100	0.8998784	0.928	0.924	6.955e-06	6.955e-06	0.3999425	0.968	0.932	5.850e-05	5.838e-05
$q = .4$	200	0.8999314	0.946	0.946	2.892e-06	2.891e-06	0.3999924	0.932	0.928	3.188e-05	3.181e-05
$n = 20$	50	0.2000149	0.952	0.934	2.624e-04	2.618e-04	0.0499268	0.952	0.942	4.925e-06	4.920e-06
$p = .2$	100	0.1999041	0.964	0.950	1.246e-04	1.243e-04	0.0500245	0.968	0.950	2.288e-06	2.284e-06
$q = .05$	200	0.1997725	0.960	0.956	6.373e-05	6.365e-05	0.0499820	0.956	0.946	1.227e-06	1.225e-06
$n = 20$	50	0.1997693	0.932	0.920	5.321e-05	5.316e-05	0.4004581	0.962	0.940	3.804e-05	3.817e-05
$p = .2$	100	0.1997739	0.958	0.938	2.786e-05	2.786e-05	0.4001641	0.948	0.928	1.846e-05	1.845e-05
$q = .4$	200	0.1999220	0.950	0.942	1.349e-05	1.347e-05	0.4001875	0.946	0.936	8.795e-06	8.812e-06
$n = 100$	50	0.9000052	0.954	0.948	1.079e-06	1.077e-06	0.0500517	0.938	0.924	3.215e-07	3.235e-07
$p = .9$	100	0.9000285	0.948	0.940	5.257e-07	5.254e-07	0.0500333	0.926	0.920	1.624e-07	1.632e-07
$q = .05$	200	0.9000315	0.954	0.950	2.600e-07	2.605e-07	0.0500164	0.966	0.952	7.125e-08	7.138e-08
$n = 100$	50	0.8999716	0.946	0.938	4.866e-07	4.864e-07	0.3999573	0.956	0.948	4.524e-06	4.516e-06
$p = .9$	100	0.9000006	0.944	0.930	2.356e-07	2.352e-07	0.4000725	0.954	0.938	2.423e-06	2.424e-06
$q = .4$	200	0.9000088	0.950	0.932	1.193e-07	1.191e-07	0.3999904	0.952	0.934	1.281e-06	1.279e-06
$n = 100$	50	0.1998470	0.952	0.940	1.076e-05	1.076e-05	0.0500290	0.926	0.922	2.257e-07	2.261e-07
$p = .2$	100	0.1998468	0.956	0.946	5.276e-06	5.289e-06	0.0500214	0.954	0.942	1.034e-07	1.036e-07
$q = .05$	200	0.1998848	0.956	0.940	2.528e-06	2.536e-06	0.0500173	0.948	0.934	5.169e-08	5.189e-08
$n = 100$	50	0.1999284	0.956	0.948	1.955e-06	1.956e-06	0.3999697	0.964	0.948	1.348e-06	1.346e-06
$p = .2$	100	0.1999579	0.952	0.950	9.949e-07	9.946e-07	0.3999965	0.950	0.936	7.242e-07	7.228e-07
$q = .4$	200	0.1999474	0.956	0.940	5.038e-07	5.055e-07	0.3999960	0.940	0.926	3.877e-07	3.870e-07

CHAPTER

4

GROWING SIZE MARKOV DYNAMIC NETWORK

In this chapter, we introduce the Growing Size Dynamic Network Model (GSMDN). In this model, we no longer assume that the total number of nodes stays the same over time. Instead, we allow the network size to be changeable.

4.1 Base Model: Fixed Rate Model

4.1.1 Model Settings

Firstly, we consider the situation that the network size grows in a fixed rate, that is, the number of coming nodes is fixed. For simplicity, the Fixed Rate Growing Size Dynamic Network model hereinafter is referred as FR model. Without loss of generality, we assume one node comes at a time.

Suppose the original network size is n_0. Then the network size at time t is $n_t = n_0 + t$.

It should be noticed that the appearing probability between old nodes is different from the one related with new coming node. For example, in social network, People are more likely to keep their old relationship than forming a friendship with new comer. In addition, people tend to trust more on the users who entered in earlier than on the new comers. Now let a be the probability with respect to the new node, which is called the forming probability in this paper. Now, in the FR model, the keeping probability, appearing probability and forming probability are

$$
\begin{aligned}
&P(X_{ij}^t = 1 | X_{ij}^{t-1} = 1) = p, \text{ if } i, j \in V_{t-1},\\
&P(X_{ij}^t = 1 | X_{ij}^{t-1} = 0) = q, \text{ if } i, j \in V_{t-1},\\
&P(X_{ij}^t = 1 | X_{ij}^{t-1} = 0) = a, \text{ if } i \text{ is the new coming node}, j \in V_{t-1}.
\end{aligned}
$$

4.1.2 Behavior of the Network Density

Let L_t be the total edges at time t. The expected edges at time t is $\mathrm{E}(L_t) = l_t$. Then suppose at time $t+1$, the expected edges that still keep should be $L_t p$. Expected number of new edges form where there is no edge at time t is $[n_t(n_t-1)/2 - L_t]q$. As for the new coming node, there are n_t new pair of nodes that could potentially form an edge in between. Thus expected new edges coming along with the new node is $n_t a$. The expected total edges at time $t+1$ is

$$
\begin{aligned}
l_{t+1} &= \mathrm{E}(L_{t+1})\\
&= \mathrm{E}\big[\mathrm{E}(L_{t+1}|L_t)\big]\\
&= \mathrm{E}\Big\{L_t p + \Big[\frac{n_t(n_t-1)}{2} - L_t\Big]q + n_t a\Big\}\\
&= \mathrm{E}(L_t)(p-q) + \frac{n_t(n_t-1)}{2}q + n_t a\\
&= l_t(p-q) + \frac{n_t(n_t-1)}{2}q + n_t a.
\end{aligned}
$$

Note that $\frac{n_t(n_t-1)}{2}$ is the potential edge number at time t. According to the definition of network density $\varrho_t = \frac{2L_t}{n_t(n_t-1)}$, we have the expected network density $\rho_t = \mathrm{E}(\varrho_t)$ as below:

$$n_t(n_t+1)\rho_{t+1} = n_t(n_t-1)\rho_t(p-q) + n_t(n_t-1)q + 2n_t a. \tag{4.1}$$

Equation (4.1) can be seen as a recursion formula for calculating the expected density. Now let

$$b_t = (n_t-1)n_t\rho_t + c_0 n_t^2 + c_1 n_t + c_2.$$

We should find some constants c_0, c_1, c_2 to satisfy that

$$b_t = b_{t-1}(p-q).$$

Solve for b_t we have

$$b_t = b_0(p-q)^t,$$

which can also be written as

$$(n_t-1)n_t\rho_t + c_0 n_t^2 + c_1 n_t + c_2 = \left[(n_0-1)n_0\rho_0 + c_0 n_0^2 + c_1 n_0 + c_2\right](p-q)^t.$$

Solve for ρ_t we have

$$\rho_t = \frac{(p-q)^t}{(n_t-1)n_t}\left[(n_0-1)n_0\rho_0 + c_0 n_0^2 + c_1 n_0 + c_2\right] - \frac{c_0 n_t}{n_t-1} - \frac{c_1}{n_t-1} - \frac{c_2}{(n_t-1)n_t}.$$

From the expression for ρ_t, we can get

$$\lim_{t\to\infty}\rho_t = -c_0.$$

Now we solve for c_0, c_1, c_2. Compare

$$n_t(n_t+1)\rho_{t+1}+c_0(n_t+1)^2+c_1(n_t+1)+c_2=[(n_t-1)n_t\rho_t+c_0n_t^2+c_1n_t+c_2](p-q).$$

with equation (4.1), we have

$$(p-q-1)c_0=q,$$
$$(p-q-1)c_1-2c_0=2a-q,$$
$$(p-q-1)c_2-c_0-c_1=0.$$

Thus we get

$$c_0=\frac{q}{p-q-1},\ c_1=\frac{2c_0+2a-q}{p-q-1},\ c_2=\frac{c_0+c_1}{p-q-1}, \tag{4.2}$$

which leads

$$\lim_{t\to\infty}\rho_t=\frac{q}{1-p+q}.$$

Now we look at the limit variance of ϱ_t. First, we have

$$\begin{aligned}\mathrm{Var}(\varrho_{t+1})&=\frac{4}{n_t^2(n_t+1)^2}\mathrm{Var}(L_{t+1})\\&=\frac{4}{n_t^2(n_t+1)^2}\Big[\mathrm{E}(\mathrm{Var}(L_{t+1}|L_t))+\mathrm{Var}(\mathrm{E}(L_{t+1}|L_t))\Big],\end{aligned}$$

and

$$\begin{aligned}\mathrm{E}(\mathrm{Var}(L_{t+1}|L_t))&=\mathrm{E}\Big[p(1-p)L_t+q(1-q)\Big(\frac{n_t(n_t-1)}{2}-L_t\Big)+a(1-a)n_t\Big]\\&=\mathrm{E}(L_t)\big[p(1-p)-q(1-q)\big]+q(1-q)\frac{n_t(n_t-1)}{2}+a(1-a)n_t,\\ \mathrm{Var}(E(L_{t+1}|L_t))&=\mathrm{Var}\Big[pL_t+q\Big(\frac{n_t(n_t-1)}{2}-L_t\Big)+an_t\Big]\\&=\mathrm{Var}(L_t)(p-q)^2.\end{aligned}$$

Thus,

$$\begin{aligned}\mathrm{Var}(\varrho_{t+1}) =& \frac{4}{n_t^2(n_t+1)^2}\Big\{\mathrm{E}(L_t)\big[p(1-p)-q(1-q)\big]+q(1-q)\frac{n_t(n_t-1)}{2} \\ &+a(1-a)n_t+\mathrm{Var}(L_t)(p-q)^2\Big\} \\ =& \frac{2(n_t-1)}{n_t(n_t+1)^2}\mathrm{E}(\varrho_t)\big[p(1-p)-q(1-q)\big]+\frac{2(n_t-1)}{n_t(n_t+1)^2}q(1-q) \\ &+\frac{4}{n_t(n_t+1)^2}a(1-a)+\frac{(n_t-1)^2}{(n_t+1)^2}\mathrm{Var}(\varrho_t)(p-q)^2.\end{aligned}$$

Take the limit on both sides, we have

$$\lim_{t\to\infty}\mathrm{Var}(\varrho_{t+1})=(p-q)^2\lim_{t\to\infty}\mathrm{Var}(\varrho_t).$$

As

$$\left|\mathrm{Var}(\varrho_t)\right|=\left|\mathrm{E}(\varrho_t^2)-\mathrm{E}^2(\varrho_t)\right|<\left|\mathrm{E}(\varrho_t^2)\right|+\left|\mathrm{E}^2(\varrho_t)\right|<2<\infty,$$

we get

$$\lim_{t\to\infty}\mathrm{Var}(\varrho_t)=0.$$

Thus we can easily have the following lemma about ϱ_t.

Lemma 4.1.1. *Suppose the FR model. Let L_t be the total edges at time t, and $\varrho_t=\frac{2L_t}{n_t(n_t-1)}$ be the network density at time t, where $n_t=n_0+t$ is the network size at time t. The expected network density is $\rho_t=\mathrm{E}(\varrho_t)$, then we have*

$$\lim_{t\to\infty}\mathrm{E}(\varrho_t)=\frac{q}{1-p+q},\ \lim_{t\to\infty}\mathrm{Var}(\varrho_t)=0.$$

Thus

$$\varrho_t\xrightarrow{p}\frac{q}{1-p+q}.$$

Furthermore,

$$\frac{6}{T^3}\sum_{t=1}^{T} L_t = \frac{6}{T^3}\sum_{t=1}^{T}\frac{n_t(n_t-1)\varrho_t}{2} \xrightarrow{p} \frac{q}{1-p+q}. \tag{4.3}$$

Proof. The proof of the lemma is straightforward from the above analysis of network density. Then as $n_t = n_0 + t$, the proof of form (4.3) comes directly from law of large numbers. □

Note that the limit of ρ_t does not depend on the appearing probability a from new coming node. Actually, as long as the number of nodes coming at a time is finite, the limit of ρ_t will not depend on a.

4.1.3 Analysis of MLE

The Likelihood function is:

$$L(\lambda,\theta|X^0,\dots,X^T) = \prod_{t=1}^{T}\Big\{\prod_{i,j\in V_{t-1},i<j}\big[p^{x_{ij}^t}(1-p)^{1-x_{ij}^t}\big]^{x_{ij}^{t-1}}\prod_{i,j\in V_{t-1},i<j}\big[q^{x_{ij}^t}(1-q)^{1-x_{ij}^t}\big]^{1-x_{ij}^{t-1}}$$
$$\times\prod_{i,j\in\mathscr{V}_t}\big[a^{x_{ij}^t}(1-a)^{1-x_{ij}^t}\big]\Big\},$$

where V_t stands for the vertex (nodes) set at time t, and $X^t = \{X_{ij}^t, i<j, i,j\in V_t\}$. $\mathscr{V}_t = \{i\in V_t\backslash V_{t-1}, j\in V_{t-1}\}\cup\{j\in V_t\backslash V_{t-1}, i\in V_{t-1}\}\cup\{i,j\in V_t\backslash V_{t-1}, i<j\}$ denotes the set that either i or j is the new comers at time t. Take the log of the likelihood and let the first derivative to 0, we can get the MLE as below:

$$\hat{p}_{MLE} = \frac{\sum_{t=1}^{T}\sum_{i<j,i,j\in V_{t-1}} X_{ij}^{t-1}X_{ij}^t}{\sum_{t=1}^{T}\sum_{i<j,i,j\in V_{t-1}} X_{ij}^{t-1}}, \tag{4.4}$$

$$\hat{q}_{MLE} = \frac{\sum_{t=1}^{T}\sum_{i<j,i,j\in V_{t-1}}(1-X_{ij}^{t-1})X_{ij}^t}{\sum_{t=1}^{T}\sum_{i<j,i,j\in V_{t-1}}(1-X_{ij}^{t-1})}, \tag{4.5}$$

$$\hat{a}_{MLE} = \frac{\sum_{t=1}^{T}\sum_{i,j\in\mathscr{V}_t} X_{ij}^t}{\sum_{t=1}^{T}\sum_{i,j\in\mathscr{V}_t} 1}. \tag{4.6}$$

Theorem 4.1.2. *Let $\{X_{ij}^t\}_{i,j}$ defined on $(\Omega,\mathscr{A},\mathscr{P})$ follows the GSMDN model with fixed number of coming nodes. Assume that the initial state of the network is nonrandom, then the MLEs of p, q and* *a are consistent and have asymptotic distributions as*

$$\sqrt{\frac{T^3}{6}}(\hat{p}_{MLE}-p)\xrightarrow{d}N\Big(0,\frac{p(1-p)(1-p+q)}{q}\Big), \tag{4.7}$$

$$\sqrt{\frac{T^3}{6}}(\hat{q}_{MLE}-q)\xrightarrow{d}N\Big(0,\frac{q(1-q)(1-p+q)}{1-p}\Big), \tag{4.8}$$

$$\sqrt{\frac{T^2}{2}}(\hat{a}_{MLE}-a)\xrightarrow{d}N(0,a(1-a)). \tag{4.9}$$

Proof. Since the proofs of (4.8) and (4.9) are quite similar to that of (4.7), we only prove for (4.7) here. Let

$$Z_t=\sum_{i<j,i,j\in V_{t-1}}X_{ij}^{t-1}(X_{ij}^t-p).$$

Then

$$\hat{p}_{MLE}=\frac{\sum_{t=1}^T\sum_{i<j,i,j\in V_{t-1}}X_{ij}^{t-1}X_{ij}^t}{\sum_{t=1}^T\sum_{i<j,i,j\in V_{t-1}}X_{ij}^{t-1}}=\frac{\sum_{t=1}^T Z_t}{\sum_{t=1}^T L_{t-1}}+p.$$

First, for the denominator

$$D=\sum_{t=1}^T L_{t-1},$$

we have proved in Lemma. 4.1.1 that

$$\frac{6}{T^3}\sum_{t=1}^T L_{t-1}\xrightarrow{p}\frac{q}{1-p+q}.$$

Now let $\mathscr{F}_a = \sigma\langle X^t : t \le a, t \in \mathbb{Z}\rangle, -\infty \le a \le b \le \infty$, we show Z_t is MDA.

$$\begin{aligned}
\mathrm{E}(Z_t|\mathscr{F}_{t-1}) &= \mathrm{E}\Big(\sum_{i<j,i,j\in V_{t-1}} X_{ij}^{t-1}(X_{ij}^t - p)\Big|X^{t-1}\Big)\\
&= \sum_{i<j,i,j\in V_{t-1}} X_{ij}^{t-1}\,\mathrm{E}\big(X_{ij}^t - p\big|X^{t-1}\big)\\
&= \sum_{i<j,i,j\in V_{t-1}} X_{ij}^{t-1}\big(X_{ij}^{t-1}p + (1-X_{ij}^{t-1})q - p\big)\\
&= \sum_{i<j,i,j\in V_{t-1}} X_{ij}^{t-1}\big(X_{ij}^{t-1}p + (X_{ij}^{t-1}-X_{ij}^{t-1})q - X_{ij}^{t-1}p\big)\\
&= 0,
\end{aligned}$$

where X^t stands for $\{X_{ij}^t, i<j, i,j\in V_t\}$. As X_{ij}^{t-1} are 0 or 1, we have $X_{ij}^{t-1} = (X_{ij}^{t-1})^2$. In addition,

$$\begin{aligned}
\mathrm{E}(Z_t^2|\mathscr{F}_{t-1}) &= \mathrm{E}\Big(\Big[\sum_{i<j,i,j\in V_{t-1}} X_{ij}^{t-1}(X_{ij}^t - p)\Big]^2\Big|X^{t-1}\Big)\\
&= \sum_{i<j,i,j\in V_{t-1}} (X_{ij}^{t-1})^2\,\mathrm{E}\big[(X_{ij}^t - p)^2\big|X^{t-1}\big]\\
&\quad + \sum_{i<j,i,j\in V_{t-1}} \sum_{k<l,k,l\in V_{t-1},(k,l)\neq(i,j)} X_{ij}^{t-1}X_{kl}^{t-1}\,\mathrm{E}\big[(X_{ij}^t - p)(X_{kl}^t - p)\big|X^{t-1}\big]\\
&= \mathrm{I} + \mathrm{II}.
\end{aligned}$$

First,

$$\begin{aligned}
\mathrm{I} &= \sum_{i<j,i,j\in V_{t-1}} (X_{ij}^{t-1})^2\,\mathrm{E}\big[(X_{ij}^t)^2 - 2pX_{ij}^t + p^2\big|X^{t-1}\big]\\
&= \sum_{i<j,i,j\in V_{t-1}} X_{ij}^{t-1}\,\mathrm{E}\big[(1-2p)X_{ij}^t + p^2\big|X^{t-1}\big]\\
&= \sum_{i<j,i,j\in V_{t-1}} X_{ij}^{t-1}\big[(1-2p)(X_{ij}^{t-1}p + (1-X_{ij}^{t-1})q) + p^2\big]\\
&= \sum_{i<j,i,j\in V_{t-1}} X_{ij}^{t-1}(1-p)p\\
&= (1-p)pL_{t-1},
\end{aligned}$$

where L_t is the number of edges at time t. Second,

$$
\begin{aligned}
\mathrm{II} &= \sum_{i<j,i,j\in V_{t-1}} \sum_{k<l,k,l\in V_{t-1},(k,l)\neq(i,j)} X_{ij}^{t-1} X_{kl}^{t-1}\,\mathrm{E}(X_{ij}^{t}-p|X^{t-1})\,\mathrm{E}(X_{kl}^{t}-p|X^{t-1}) \\
&= \sum\sum X_{ij}^{t-1}\big[X_{ij}^{t-1}p+(1-X_{ij}^{t-1})q-p\big]X_{kl}^{t-1}\big[X_{kl}^{t-1}p+(1-X_{kl}^{t-1})q-p\big] \\
&= \sum\sum \big[X_{ij}^{t-1}p+(X_{ij}^{t-1}-X_{ij}^{t-1})q-X_{ij}^{t-1}p\big]\big[X_{kl}^{t-1}p+(X_{kl}^{t-1}-X_{kl}^{t-1})q-X_{kl}^{t-1}p\big] \\
&= 0.
\end{aligned}
$$

Thus,

$$
\mathrm{E}(Z_t^2|\mathscr{F}_{t-1}) = (1-p)pL_{t-1}.
$$

Again, use Lemma. 4.1.1, we have

$$
\frac{6}{T^3}\,\mathrm{E}(Z_t^2|\mathscr{F}_{t-1}) \xrightarrow{p} \frac{(1-p)pq}{1-p+q}.
$$

Then, we verify the Lyapunov's condition:

$$
\begin{aligned}
&\frac{1}{T^6}\sum_{t=1}^{T}\mathrm{E}[Z_t^4|\mathscr{F}_{t-1}] \\
&= \frac{1}{T^6}\sum_{t=1}^{T}\mathrm{E}\Big(\Big[\sum_{i<j,i,j\in V_{t-1}} X_{ij}^{t-1}(X_{ij}^{t}-p)\Big]^4\Big|X^{t-1}\Big) \\
&= \frac{1}{T^6}\sum_{t=1}^{T}\Big\{\sum_{i<j,i,j\in V_{t-1}} (X_{ij}^{t-1})^4\,\mathrm{E}\big[(X_{ij}^{t}-p)^4\big|X^{t-1}\big] \\
&+8\sum_{i<j,i,j\in V_{t-1}}\sum_{k<l,k,l\in V_{t-1},(k,l)\neq(i,j)} (X_{ij}^{t-1})^3X_{kl}^{t-1}\,\mathrm{E}\big[(X_{ij}^{t}-p)^3(X_{kl}^{t}-p)\big|X^{t-1}\big] \\
&+6\sum_{i<j,i,j\in V_{t-1}}\sum_{k<l,k,l\in V_{t-1},(k,l)\neq(i,j)} (X_{ij}^{t-1})^2(X_{kl}^{t-1})^2\,\mathrm{E}\big[(X_{ij}^{t}-p)^2(X_{kl}^{t}-p)^2\big|X^{t-1}\big] \\
&+\sum_{(i,j)}\sum_{(k,l)}\sum_{(m,n)}\sum_{(r,s)} X_{ij}^{t-1}X_{kl}^{t-1}X_{mn}^{t-1}X_{rs}^{t-1}\,\mathrm{E}\big[(X_{ij}^{t}-p)(X_{kl}^{t}-p)(X_{mn}^{t}-p)(X_{rs}^{t}-p)\big|X^{t-1}\big]\Big\} \\
&= \frac{1}{T^6}\sum_{t=1}^{T}(\mathrm{I}+8\mathrm{II}+6\mathrm{III}+\mathrm{IV}),
\end{aligned}
$$

where part IV is the summation for different pairs of nodes at time $t-1$. Due to the independence of the edges and the conditional expectation that $X_{ij}^{t-1}\,\mathrm{E}[(X_{ij}^{t}-p)|X^{t-1}]=0$, we can easily have $\mathrm{II}=0, \mathrm{IV}=0$.

Under the FR model where we assume 1 node comes at a time, $n_t = n_0 + t$. Now for part I, using the technique of enlarging and reducing, we have

$$
\begin{aligned}
\mathrm{I} &= \frac{1}{T^6}\sum_{t=1}^{T}\sum_{i<j,i,j\in V_{t-1}} (X_{ij}^{t-1})^4\,\mathrm{E}[(X_{ij}^{t}-p)^4|X^{t-1}] \\
&= \frac{1}{T^6}\sum_{t=1}^{T}\sum_{i<j,i,j\in V_{t-1}} X_{ij}^{t-1}\,\mathrm{E}\big[X_{ij}^{t}-3pX_{ij}^{t}+4p^2X_{ij}^{t}-3p^3X_{ij}^{t}+p^4\big|X^{t-1}\big] \\
&= \frac{1}{T^6}\sum_{t=1}^{T}\sum_{i<j,i,j\in V_{t-1}} X_{ij}^{t-1}(p-3p^2+4p^3-3p^4+p^4) \\
&\le \frac{1}{T^6}\sum_{t=1}^{T}\frac{(n_0+t-1)(n_0+t-2)}{2}(p-3p^2+4p^3-2p^4) \\
&= \frac{1}{T^6}O(T^3)(p-3p^2+4p^3-2p^4) \\
&\to 0, \text{ as } T\to\infty.
\end{aligned}
$$

For part III, using similar technique, we have

$$
\begin{aligned}
\mathrm{III} &= \frac{1}{T^6}\sum_{t=1}^{T}\sum_{i<j,i,j\in V_{t-1}}\sum_{k<l,k,l\in V_{t-1},(k,l)\neq(i,j)} (X_{ij}^{t-1})^2(X_{kl}^{t-1})^2\,\mathrm{E}\big[(X_{ij}^{t}-p)^2(X_{kl}^{t}-p)^2\big|X^{t-1}\big] \\
&= \frac{1}{T^6}\sum_{t=1}^{T}\sum_{i<j,i,j\in V_{t-1}}\sum_{k<l,k,l\in V_{t-1},(k,l)\neq(i,j)} X_{ij}^{t-1}X_{kl}^{t-1}\,\mathrm{E}\big[(X_{ij}^{t}-p)^2\big|X^{t-1}\big]\mathrm{E}\big[(X_{kl}^{t}-p)^2\big|X^{t-1}\big] \\
&= \frac{1}{T^6}\sum_{t=1}^{T}\sum_{i<j,i,j\in V_{t-1}}\sum_{k<l,k,l\in V_{t-1},(k,l)\neq(i,j)} X_{ij}^{t-1}p(1-p)X_{kl}^{t-1}p(1-p) \\
&< p^2(1-p)^2\frac{1}{T^6}\sum_{t=1}^{T}\Big[\frac{(n_0+t-1)(n_0+t-2)}{2}\Big]^2 \\
&= p^2(1-p)^2\frac{1}{T^6}O(T^5) \\
&\to 0, \text{ as } T\to\infty.
\end{aligned}
$$

Thus

$$\frac{1}{T^6}\sum_{t=1}^{T}\mathrm{E}[Z_t^4|\mathscr{F}_{t-1}] \to 0, \text{ as } T \to \infty.$$

Therefore, the Lyapunov's condition is verified. Hence, by Martingale CLT, we have

$$\sqrt{\frac{6}{T^3}}\sum_{t=1}^{T} Z_t \xrightarrow{d} N\Big(0, \frac{p(1-p)q}{1-p+q}\Big).$$

Then, by the Slutsky's theorem,

$$\sqrt{\frac{T^3}{6}}(\hat{p}_{MLE} - p) \xrightarrow{d} N\Big(0, \frac{p(1-p)(1-p+q)}{q}\Big).$$

□

4.2 Generalized Model: Arrive/Leave in Random Process

4.2.1 Model Settings

In real life, it is rare to see a network growing in constant rate. More often, we see nodes arrive or leave the network randomly. In this section, we introduce a generalized model of the Growing Size Markov Dynamic Network which is called the Random Process (RP) model. In this generalized model, we assume that the number of new nodes follows

$$M_t^1 \overset{\text{i.i.d}}{\sim} \text{Poisson}(\lambda_1),$$

and the number of nodes become inactive (leave the network) follows

$$M_t^2 \overset{\text{i.i.d}}{\sim} \text{Poisson}(\lambda_2).$$

Then, the total number of nodes at time t is $N_t = n_0 + \sum_{\tau=1}^{t}(M_\tau^1 - M_\tau^2)$. It is easy to see that

$$\sum_{\tau=1}^{t} M_\tau^1 \sim \text{Poisson}(t\lambda_1), \sum_{\tau=1}^{t} M_\tau^2 \sim \text{Poisson}(t\lambda_2).$$

Denote $M_t = M_t^1 - M_t^2$. Furthermore, assume the nodes leave and arrive independently. In addition, the network is growing, which suggests that $\lambda_1 - \lambda_2 > 0$, that is $\mathrm{E}(M_t) > 0$. The other assumptions are the same with the FR model.

Before we start our analysis of the RP model, a simple lemma about the moments of N_t is brought up here for reference:

Lemma 4.2.1. *Suppose $M_t^1 \overset{i.i.d}{\sim}$ Poisson(λ_1) and $M_t^2 \overset{i.i.d}{\sim}$ Poisson(λ_2) with $\lambda = \lambda_1 - \lambda_2 > 0$, where M_t^1 and M_t^2 are independent processes. Let $N_t = n_0 + \sum_{\tau=1}^{t}(M_\tau^1 - M_\tau^2)$. We have the results about N_t as below:*

1. $\mathrm{E}(N_t) = O(t)$
2. $\mathrm{E}\left(\frac{N_t(N_t-1)}{2}\right) = O(t^2)$;
3. $\mathrm{E}\left[\left(\frac{N_t(N_t-1)}{2}\right)^2\right] = O(t^4)$;
4. $\mathrm{Var}\left(\frac{N_t(N_t-1)}{2}\right) = O(t^3)$.

Proof. First, it is easy to see that $M_t = \sum_{\tau=1}^{t}(M_\tau^1 - M_\tau^2) \sim \text{Poisson}(t\lambda)$. Thus,

$$\mathrm{E}(M_t) = t\lambda, \mathrm{E}(M_t)^2 = (t\lambda)^2 + t\lambda, \mathrm{E}(M_t)^3 = (t\lambda)^3 + 3(t\lambda)^2 + (t\lambda),$$
$$\mathrm{E}(M_t)^4 = (t\lambda)^4 + 6(t\lambda)^3 + 7(t\lambda)^2 + (t\lambda).$$

As $N_t = n_0 + \sum_{\tau=1}^{t}(M_\tau^1 - M_\tau^2) = n_0 + M_t$,

$$\mathrm{E}(N_t) = \mathrm{E}(n_0 + M_t) = n_0 + t\lambda \sim t.$$

Thus, 1 is proved. Then, for 2, we have

$$\begin{aligned}
\mathrm{E}(N_t^2) &= \mathrm{E}\left[(n_0^2 + 2n_0 M_t + (M_t)^2\right] \\
&= n_0^2 + 2n_0 t\lambda + (t\lambda)^2 + t\lambda \\
&= (t\lambda)^2 + (2n_0+1)(t\lambda) + n_0^2.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathrm{E}\left[\frac{N_t(N_t-1)}{2}\right] &= \frac{1}{2}\,\mathrm{E}(N_t^2 - N_t) \\
&= \frac{1}{2}\left[(t\lambda)^2 + (2n_0+1)(t\lambda) + n_0^2 - n_0 - t\lambda\right] \\
&= \frac{1}{2}[(t\lambda)^2 + 2n_0(t\lambda) + n_0^2 - n_0] \sim t^2
\end{aligned}$$

Similarly, since

$$\begin{aligned}
\mathrm{E}(N_t^3) &= \mathrm{E}\left[n_0^3 + 3n_0^2 M_t + 3n_0(M_t)^2 + (M_t)^3\right] \\
&= n_0^3 + 3n_0^2 t + 3n_0(t^2+t) + t^3 + 3t^2 + t \\
&= (t\lambda)^3 + 3(n_0^2+1)(t\lambda)^2 + (3n_0^2 + 3n_0 + 1)(t\lambda) + n_0^3,
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{E}(N_t^4) &= \mathrm{E}\left[n_0^4 + 4n_0^3 M_t + 6n_0^2(M_t)^2 + 4n_0(M_t)^3 + (M_t)^4\right] \\
&= n_0^4 + 4n_0^3 t + 6n_0^2(t^2+t) + 4n_0(t^3+3t^2+t) + t^4 + 6t^3 + 7t^2 + t \\
&= (t\lambda)^4 + (4n_0+6)(t\lambda)^3 + (6n_0^2 + 12n_0 + 7)(t\lambda)^2 + (4n_0^3 + 6n_0^2 + 4n_0 + 1)(t\lambda) + n_0^4,
\end{aligned}$$

we can easily have 3 proved as

$$\begin{aligned}\mathrm{E}\Big[\Big(\frac{N_t(N_t-1)}{2}\Big)^2\Big] &= \frac{1}{4}\,\mathrm{E}(N_t^4-2N_t^3+N_t^2)\\ &= \frac{1}{4}\big[(t\lambda)^4+4(n_0+1)(t\lambda)^3+2(6n_0+1)(t\lambda)^2+4n_0^3(t\lambda)+n_0^4-2n_0^3+n_0^2\big]\\ &\sim t^4.\end{aligned}$$

Lastly, we have

$$\begin{aligned}&\mathrm{Var}\Big[\frac{N_t(N_t-1)}{2}\Big]\\ =&\mathrm{E}\Big[\Big(\frac{N_t(N_t-1)}{2}\Big)^2\Big]-\mathrm{E}\Big[\frac{N_t(N_t-1)}{2}\Big]^2\\ =&\frac{1}{4}\big[(t\lambda)^4+4(n_0+1)(t\lambda)^3+2(6n_0+1)(t\lambda)^2+4n_0^3(t\lambda)+n_0^4-2n_0^3+n_0^2\big]\\ &-\frac{1}{4}\big[(t\lambda)^2+2n_0(t\lambda)+n_0^2-n_0\big]^2\\ \sim& t^3.\end{aligned}$$

Thus 4 is proved. □

4.2.2 Analysis of Network Density

In order to analyze the behavior of the dynamic network, we will firstly analyze the behavior of expected network density. Let $\rho_t=\mathrm{E}(\varrho_t)$ denote the expected network density at time t, then

$$\begin{aligned}&\mathrm{E}\Big[\frac{N_{t+1}(N_{t+1}-1)\varrho_{t+1}}{2}\Big]\\ =&\mathrm{E}(L_{t+1})=\mathrm{E}\big[\,\mathrm{E}(L_{t+1}|L_t,N_t)\big]\\ =&\mathrm{E}\Big\{L_t p+\Big[\frac{N_t(N_t-1)}{2}-L_t\Big]q+\Big[N_t M_{t+1}+\frac{M_{t+1}(M_{t+1}-1)}{2}\Big]a\Big\}\\ =&\mathrm{E}\Big[\frac{N_t(N_t-1)\varrho_t}{2}\Big](p-q)+\mathrm{E}\Big[\frac{N_t(N_t-1)}{2}\Big]q+\mathrm{E}\Big[N_t M_{t+1}+\frac{M_{t+1}(M_{t+1}-1)}{2}\Big]a.\end{aligned}$$

Note that $\frac{N_t(N_t-1)}{2}$ is the potential edge number at time t, and that the network density $\varrho_t = \frac{2L_t}{N_t(N_t-1)}$. From the Markov property and the independence of network size and density, we have the expected network density $\rho_t = \mathrm{E}(\varrho_t)$ as below:

$$\begin{aligned}&[(n_0+t\lambda+1)^2-n_0-1]\rho_{t+1}\\=&[(n_0+t\lambda)^2-n_0]\rho_t(p-q)-[(n_0+t\lambda)^2-n_0]q+[2(n_0+t\lambda)\lambda+(t\lambda)^2]a.\end{aligned} \tag{4.10}$$

(4.10) can be seen as a recursion formula for calculating the expected density. Now let

$$b_t = [(n_0+t\lambda)^2-n_0]\rho_t + c_0(n_0+t\lambda)^2 + c_1(n_0+t\lambda)+c_2.$$

We could find some constants c_0, c_1, c_2 satisfying

$$b_t = b_{t-1}(p-q).$$

Solve for b_t we have

$$b_t = b_0(p-q)^{t\lambda}.$$

Substitute the expression of b_t, we have

$$[(n_0+t\lambda)^2-n_0]\rho_t + c_0(n_0+t\lambda)^2 + c_1(n_0+t\lambda)+c_2 = [(n_0-1)n_0\rho_0 + c_0 n_0^2 + c_1 n_0 + c_2](p-q)^{t\lambda}.$$

Solve for ρ_t we have

$$\begin{aligned}\rho_t =& \frac{(p-q)^{t\lambda}}{(n_0+t\lambda)^2-n_0}\left[(n_0-1)n_0\rho_0 + c_0 n_0^2 + c_1 n_0 + c_2\right] - \frac{c_0(n_0+t\lambda)^2}{(n_0+t\lambda)^2-n_0}\\ &-\frac{c_1(n_0+t\lambda)}{(n_0+t\lambda)^2-n_0} - \frac{c_2}{(n_0+t\lambda)^2-n_0}.\end{aligned}$$

From the expression for ρ_t, we can get

$$\lim_{t\to\infty} \rho_t = -c_0.$$

Using similar method from that in FR model, we solve for c_0 and get

$$c_0 = \frac{q}{p-q-1}, \tag{4.11}$$

which leads to

$$\lim_{t\to\infty} \rho_t = \frac{q}{1-p+q}.$$

Now the variance limit of ϱ_t is needed in order to prove convergence in probability. Using the variance decomposition formula, we have

$$\begin{aligned} &\mathrm{Var}\Big(\frac{N_{t+1}(N_{t+1}-1)}{2}\varrho_{t+1}\Big) = \mathrm{Var}(L_{t+1}) \\ =&\mathrm{E}\big[\mathrm{Var}(L_{t+1}\big|L_t, M_1, ..., M_t)\big] + \mathrm{Var}\big[\mathrm{E}(L_{t+1}\big|L_t, M_1, ..., M_t)\big]. \end{aligned}$$

On one hand,

$$\begin{aligned} &\mathrm{E}\big[\mathrm{Var}(L_{t+1}\big|L_t, M_1, ..., M_t)\big] \\ =&\mathrm{E}\Big[L_t p(1-p) + q(1-q)\Big(\frac{N_t(N_t-1)}{2} - L_t\Big) + a(1-a)\Big(N_t M_{t+1} + \frac{M_{t+1}(M_{t+1}-1)}{2}\Big)\Big] \\ =&\mathrm{E}(\varrho_t)[p(1-p) - q(1-q)]\,\mathrm{E}\Big(\frac{N_t(N_t-1)}{2}\Big) + q(1-q)\,\mathrm{E}\Big(\frac{N_t(N_t-1)}{2}\Big) \\ &+ a(1-a)\,\mathrm{E}\Big(N_t M_{t+1} + \frac{M_{t+1}(M_{t+1}-1)}{2}\Big), \end{aligned}$$

and

$$
\begin{aligned}
&\mathrm{Var}\Big[\mathrm{E}(L_{t+1}\Big|L_t, M_1, ..., M_t)\Big]\\
=&\mathrm{Var}\Big[L_t p + \Big(\frac{N_t(N_t-1)}{2} - L_t\Big)q + \Big(N_t M_{t+1} + \frac{M_{t+1}(M_{t+1}-1)}{2}\Big)a\Big]\\
=&\mathrm{Var}\Big(\varrho_t \frac{N_t(N_t-1)}{2}\Big)(p-q) + \mathrm{Var}\Big(\frac{N_t(N_t-1)}{2}\Big)q + \mathrm{Var}\Big(N_t M_{t+1} + \frac{M_{t+1}(M_{t+1}-1)}{2}\Big)a.
\end{aligned}
$$

On the other hand,

$$
\begin{aligned}
&\mathrm{Var}\Big(\frac{N_{t+1}(N_{t+1}-1)}{2}\varrho_{t+1}\Big)\\
=&\mathrm{E}\Big[\Big(\frac{N_{t+1}(N_{t+1}-1)}{2}\Big)^2\Big]\mathrm{Var}(\varrho_{t+1}) + \Big[\mathrm{E}(\varrho_{t+1})\Big]^2 \mathrm{Var}\Big(\frac{N_{t+1}(N_{t+1}-1)}{2}\Big),
\end{aligned}
$$

since the appearing and disappearing of edges are independent of the coming of new nodes, N_t and ϱ_t, are independent. To summarize,

$$
\begin{aligned}
&\mathrm{Var}\Big(\frac{N_{t+1}(N_{t+1}-1)}{2}\varrho_{t+1}\Big)\\
=&\mathrm{E}(\varrho_t)\big[p(1-p)-q(1-q)\big]\mathrm{E}\Big(\frac{N_t(N_t-1)}{2}\Big) + q(1-q)\mathrm{E}\Big(\frac{N_t(N_t-1)}{2}\Big)\\
&+a(1-a)\mathrm{E}\Big(N_t M_{t+1} + \frac{M_{t+1}(M_{t+1}-1)}{2}\Big) + \mathrm{E}\Big[\Big(\frac{N_t(N_t-1)}{2}\Big)^2\Big]\mathrm{Var}(\varrho_t)(p-q)\\
&+[\mathrm{E}(\varrho_t)]^2\,\mathrm{Var}\Big(\frac{N_t(N_t-1)}{2}\Big) + \mathrm{Var}\Big(\frac{N_t(N_t-1)}{2}\Big)q + \mathrm{Var}\Big(N_t M_{t+1} + \frac{M_{t+1}(M_{t+1}-1)}{2}\Big)a.
\end{aligned}
$$

Finally,

$$
\begin{aligned}
&\mathrm{Var}(\varrho_{t+1})\\
=&\Big\{\mathrm{E}(\varrho_t)\big[p(1-p)-q(1-q)\big]\mathrm{E}\Big(\frac{N_t(N_t-1)}{2}\Big) + q(1-q)\mathrm{E}\Big(\frac{N_t(N_t-1)}{2}\Big)\\
&+a(1-a)\mathrm{E}\Big(N_t M_{t+1} + \frac{M_{t+1}(M_{t+1}-1)}{2}\Big) + \mathrm{E}\Big[\Big(\frac{N_t(N_t-1)}{2}\Big)^2\Big]\mathrm{Var}(\varrho_t)(p-q)\\
&+\Big[\mathrm{E}(\varrho_t)\Big]^2\mathrm{Var}\Big(\frac{N_t(N_t-1)}{2}\Big) + \mathrm{Var}\Big(\frac{N_t(N_t-1)}{2}\Big)q + \mathrm{Var}\Big(N_t M_{t+1} + \frac{M_{t+1}(M_{t+1}-1)}{2}\Big)a\\
&-\Big[\mathrm{E}(\varrho_{t+1})\Big]^2\mathrm{Var}\Big(\frac{N_{t+1}(N_{t+1}-1)}{2}\Big)\Big\}\Big/\mathrm{E}\Big[\Big(\frac{N_{t+1}(N_{t+1}-1)}{2}\Big)^2\Big].
\end{aligned}
$$

Use Lemma 4.2.1 and take the limit on both sides, we have

$$\lim_{t\to\infty} \mathrm{Var}(\varrho_{t+1}) = (p-q)^2 \lim_{t\to\infty} \mathrm{Var}(\varrho_t).$$

In addition,

$$\left|\mathrm{Var}(\varrho_t)\right| = \left|\mathrm{E}(\varrho_t^2) - \mathrm{E}^2(\varrho_t)\right| < \left|\mathrm{E}(\varrho_t^2)\right| + \left|\mathrm{E}^2(\varrho_t)\right| < 2 < \infty.$$

Thus

$$\lim_{t\to\infty} \mathrm{Var}(\varrho_t) = 0.$$

Now, for the RP model, we also have a similar lemma to the FR model about the network density as below:

Lemma 4.2.2. *Suppose the RP model with nodes arrive process* $M_t^1 \overset{i.i.d}{\sim}$ *Poisson*(λ_1) *and leave process* $M_t^2 \overset{i.i.d}{\sim}$ *Poisson*(λ_2) *with* $\lambda = \lambda_1 - \lambda_2 > 0$*, where* M_t^1 *and* M_t^2 *are independent processes. Let* L_t *be the total edges at time t,* $\varrho_t = \frac{2L_t}{N_t(N_t-1)}$ *be the network density at time t and the expected network density* $\rho_t = \mathrm{E}(\varrho_t)$*. Then, we have*

$$\lim_{t\to\infty} \mathrm{E}(\varrho_t) = \frac{q}{1-p+q}, \lim_{t\to\infty} \mathrm{Var}(\varrho_t) = 0.$$

Thus

$$\varrho_t \xrightarrow{p} \frac{q}{1-p+q}.$$

Furthermore,

$$\frac{6}{(T\lambda)^3} \sum_{t=1}^{T} L_t = \frac{6}{(T\lambda)^3} \sum_{t=1}^{T} \frac{N_t(N_t-1)\varrho_t}{2} \xrightarrow{p} \frac{q}{1-p+q}. \tag{4.12}$$

Proof. The first part is straightforward from the above analysis of the network density, we only need to prove (4.12). Rewrite the left part of (4.12), we have

$$\frac{6}{(T\lambda)^3}\sum_{t=1}^{T}\frac{N_t(N_t-1)\varrho_t}{2} = \frac{6}{(T\lambda)^3}\sum_{t=1}^{T}\Big(\frac{N_t(N_t-1)}{2}-\frac{(n_0+t\lambda)^2-n_0}{2}\Big)\varrho_t + \frac{6}{(T\lambda)^3}\sum_{t=1}^{T}\frac{(n_0+t\lambda)^2-n_0}{2}\varrho_t.$$

First,

$$\frac{6}{(T\lambda)^3}\sum_{t=1}^{T}\frac{(n_0+t\lambda)^2-n_0}{2}\varrho_t \xrightarrow{p} \frac{q}{1-p+q}.$$

Second,

$$\mathrm{E}\Big\{\Big(\frac{N_t(N_t-1)}{2T^2}-\frac{(n_0+t\lambda)^2-n_0}{2T^2}\Big)\varrho_t\Big\} = \mathrm{E}\Big(\frac{N_t(N_t-1)}{2}-\frac{(n_0+t\lambda)^2-n_0}{2}\Big)\mathrm{E}(\varrho_t) = 0.$$

In addition,

$$\begin{aligned}
&\sum_{t=1}^{T}\frac{1}{t}\mathrm{Var}\Big\{\Big(\frac{N_t(N_t-1)}{2T^2}-\frac{(n_0+t\lambda)^2-n_0}{2T^2}\Big)\varrho_t\Big\} \\
&\le \sum_{t=1}^{T}\frac{1}{t}\mathrm{Var}\Big\{\Big(\frac{N_t(N_t-1)}{2t^2}-\frac{(n_0+t\lambda)^2-n_0}{2t^2}\Big)\varrho_t\Big\} \\
&= \sum_{t=1}^{T}\frac{1}{t}\Big\{\mathrm{Var}\Big(\frac{N_t(N_t-1)}{2t^2}-\frac{(n_0+t\lambda)^2-n_0}{2t^2}\Big)\big[\mathrm{E}(\varrho_t)\big]^2 \\
&\quad + \mathrm{E}\Big(\frac{N_t(N_t-1)}{2t^2}-\frac{(n_0+t\lambda)^2-n_0}{2t^2}\Big)^2\mathrm{Var}(\varrho_t)\Big\} < \infty.
\end{aligned}$$

Thus from Law of Large Numbers, we have

$$\frac{6}{(T\lambda)^3}\sum_{t=1}^{T}\Big(\frac{N_t(N_t-1)}{2}-\frac{(n_0+t\lambda)^2-n_0}{2}\Big)\varrho_t \xrightarrow{a.s.} 0.$$

Therefore,

$$\frac{6}{(T\lambda)^3}\sum_{t=1}^{T}L_t \xrightarrow{p} \frac{q}{1-p+q}.$$

□

4.2.3 Analysis of MLE

The Likelihood function is constructed the same as fixed coming nodes model. The different part is the node set V_t, which contains random nodes instead. As the expressions remain the same, we still have the same estimators. In addition, we have the following theorem that is similar to the fixed coming nodes model:

Theorem 4.2.3. *Let $\{X_{ij}^t\}_{i,j}$ defined on $(\Omega, \mathscr{A}, \mathscr{P})$ follows the RP model. Assume that the initial state of the network is nonrandom, then the MLEs of p, q and a given in form (4.4), (4.5) and (4.6) are consistent and have asymptotic distributions as*

$$\sqrt{\frac{(T\lambda)^3}{6}}(\hat{p}_{MLE} - p) \xrightarrow{d} N\left(0, \frac{p(1-p)(1-p+q)}{q}\right), \tag{4.13}$$

$$\sqrt{\frac{(T\lambda)^3}{6}}(\hat{q}_{MLE} - q) \xrightarrow{d} N\left(0, \frac{q(1-q)(1-p+q)}{1-p}\right), \tag{4.14}$$

$$\sqrt{\frac{(T\lambda)^2}{2}}(\hat{a}_{MLE} - a) \xrightarrow{d} N(0, a(1-a)). \tag{4.15}$$

The proof of Theorem.4.2.3 shares a big part with the proof of Theorem.4.1.2. We are able to use Martingale CLT by creating Z_t which can then be shown to be a MDA. They differ in the part of verifying Lyapunov's condition. Our task here is to verify the Lyapunov's condition under the RP model.

Before doing this, we need to prove the following lemma regarding convergence of the summation series.

Lemma 4.2.4. *Suppose $M_t^1 \overset{i.i.d}{\sim}$ Poisson(λ_1) and $M_t^2 \overset{i.i.d}{\sim}$ Poisson(λ_2) with $\lambda = \lambda_1 - \lambda_2 > 0$, where M_t^1 and M_t^2 are independent processes. Let $N_t = n_0 + \sum_{\tau=1}^t (M_\tau^1 - M_\tau^2)$. We have that the summation series below converge almost surely to some constants:*

1. $\sum_{t=1}^T N_{t-1}/T^2 \xrightarrow{a.s.} \lambda/2$;
2. $\sum_{t=1}^T N_{t-1}^2/T^3 \xrightarrow{a.s.} \lambda^2/3$;

3. $\sum_{t=1}^{T} N_{t-1}^3/T^4 \xrightarrow{a.s.} \lambda^3/4$;

4. $\sum_{t=1}^{T} N_{t-1}^4/T^5 \xrightarrow{a.s.} \lambda^4/5$.

Proof. Let $M_i = M_i^1 - M_i^2$. First, its easy to see that

$$\sum_{t=1}^{T}\sum_{i=1}^{t-1} M_i = \sum_{i=1}^{T-1}\sum_{t=i+1}^{T} M_i = \sum_{i=1}^{T-1}(T-i)M_i,$$

and

$$\sum_{t=1}^{T}\sum_{i=t+1}^{T} M_i = \sum_{i=1}^{T}\sum_{t=1}^{i-1} M_i = \sum_{i=1}^{T}(i-1)M_i.$$

Since M_i is i.i.d for $i = 1,2,3,...$, we have

$$M_i = M_{T-i+1} \text{ in distribution.}$$

Thus

$$\sum_{i=1}^{T-1}(T-i)M_i = \sum_{i=1}^{T-1}(T-i)M_{T-i+1} = \sum_{j=T-1}^{1} jM_{j+1} = \sum_{i=1}^{T}(i-1)M_i \text{ in distribution}$$

That is

$$\sum_{t=1}^{T}\sum_{i=1}^{t-1} M_i = \sum_{t=1}^{T}\sum_{i=t+1}^{T} M_i \text{ in distribution.} \tag{4.16}$$

From the Strong Law of Large Numbers for M_i, we have, on one hand,

$$\frac{1}{T^2}\Big(\sum_{t=1}^{T}\sum_{i=1}^{t-1} M_i + \sum_{i=1}^{T} M_i + \sum_{t=1}^{T}\sum_{i=t+1}^{T} M_i\Big) = \frac{1}{T^2}\sum_{t=1}^{T}\sum_{i=1}^{T} M_i = \frac{1}{T}\sum_{t=1}^{T} M_i \xrightarrow{a.s.} \mathrm{E}(M_1) = \lambda, \tag{4.17}$$

on the other hand,

$$\frac{1}{T^2}\sum_{i=1}^{T} M_i \xrightarrow{a.s.} 0. \tag{4.18}$$

From (4.16), (4.17) and (4.18) we have

$$\frac{1}{T^2}\sum_{t=1}^{T}\sum_{i=1}^{t-1} M_i \xrightarrow{a.s.} \frac{1}{2}\,\mathrm{E}(M_1) = \frac{\lambda}{2}. \tag{4.19}$$

Therefore, we can prove 1:

$$\frac{1}{T^2}\sum_{t=1}^{T} N_{t-1} = \frac{1}{T^2}\sum_{t=1}^{T}\Big(n_0 + \sum_{i=1}^{t-1} M_i\Big) = \frac{n_0}{T} + \frac{1}{T^2}\sum_{t=1}^{T}\sum_{i=1}^{t-1} M_i \xrightarrow{a.s.} \frac{\lambda}{2}$$

Similarly, we have

$$\begin{aligned}\frac{1}{T^2}\sum_{t=1}^{T}\sum_{i=1}^{t-1} M_i^2 &\xrightarrow{a.s.} \frac{1}{2}\,\mathrm{E}(M_1^2) = \frac{\lambda^3+3\lambda^2+\lambda}{2},\\ \frac{1}{T^3}\sum_{t=1}^{T}\sum_{i=1}^{t-1}\sum_{j\neq i} M_i M_j &\xrightarrow{a.s.} \frac{1}{3}\,\mathrm{E}(M_1)\,\mathrm{E}(M_1) = \frac{\lambda^2}{3}.\end{aligned}$$

Therefore,

$$\begin{aligned}\frac{1}{T^3}\sum_{t=1}^{T} N_{t-1}^2 &= \frac{1}{T^3}\sum_{t=1}^{T}\Big(n_0 + \sum_{i=1}^{t-1} M_i\Big)^2\\ &= \frac{1}{T^3}\Big[n_0^2 T + 2n_0\sum_{t=1}^{T}\sum_{i=1}^{t-1} M_i + \sum_{t=1}^{T}\Big(\sum_{i=1}^{t-1} M_i\Big)^2\Big]\\ &\xrightarrow{a.s.} \frac{\lambda^2}{3}.\end{aligned}$$

Thus 2 is proved. In addition,

$$\begin{aligned}\frac{1}{T^2}\sum_{t=1}^{T}\sum_{i=1}^{t-1} M_i^3 &\xrightarrow{a.s.} \frac{1}{2}\,\mathrm{E}(M_1^3),\\ \frac{1}{T^3}\sum_{t=1}^{T}\sum_{i=1}^{t-1}\sum_{j\neq i} M_i^2 M_j &\xrightarrow{a.s.} \frac{1}{3}\,\mathrm{E}(M_1^2)\,\mathrm{E}(M_1),\\ \frac{1}{T^4}\sum_{t=1}^{T}\sum_{i=1}^{t-1}\sum_{j\neq i}\sum_{k\neq i,j} M_i M_j M_k &\xrightarrow{a.s.} \frac{1}{4}\,\mathrm{E}(M_1)^3 = \frac{\lambda^3}{4}.\end{aligned}$$

Then we can prove 3:

$$
\begin{aligned}
\frac{1}{T^4}\sum_{t=1}^{T} N_{t-1}^3 &= \frac{1}{T^4}\sum_{t=1}^{T}\Big(n_0+\sum_{i=1}^{t-1} M_i\Big)^3 \\
&= n_0^3 T + 3n_0^2 \sum_{t=1}^{T}\sum_{i=1}^{t-1} M_i + 3n_0 \sum_{t=1}^{T}\Big(\sum_{i=1}^{t-1} M_i\Big)^2 + \sum_{t=1}^{T}\Big(\sum_{i=1}^{t-1} M_i\Big)^3 \\
&\xrightarrow{a.s.} \frac{\lambda^3}{4}.
\end{aligned}
$$

At Last, we prove for 4. Use the same technique, we have

$$
\begin{aligned}
\frac{1}{T^2}\sum_{t=1}^{T}\sum_{i=1}^{t-1} M_i^4 &\xrightarrow{a.s.} \frac{1}{2}\mathrm{E}(M_1^4), \\
\frac{1}{T^3}\sum_{t=1}^{T}\sum_{i=1}^{t-1}\sum_{j\neq i} M_i^3 M_j &\xrightarrow{a.s.} \frac{1}{3}\mathrm{E}(M_1^3)\mathrm{E}(M_1), \\
\frac{1}{T^3}\sum_{t=1}^{T}\sum_{i=1}^{t-1}\sum_{j\neq i} M_i^2 M_j^2 &\xrightarrow{a.s.} \frac{1}{3}\mathrm{E}(M_1^2)\mathrm{E}(M_1^2), \\
\frac{1}{T^4}\sum_{t=1}^{T}\sum_{i=1}^{t-1}\sum_{j\neq i}\sum_{k\neq i,j} M_i^2 M_j M_k &\xrightarrow{a.s.} \frac{1}{4}\mathrm{E}(M_1^2)\mathrm{E}(M_1)\mathrm{E}(M_1), \\
\frac{1}{T^5}\sum_{t=1}^{T}\sum_{i=1}^{t-1}\sum_{j\neq i}\sum_{k\neq i,j}\sum_{l\neq i,j,k} M_i M_j M_k M_l &\xrightarrow{a.s.} \frac{1}{5}\mathrm{E}(M_1)^4 = \frac{\lambda^4}{5}.
\end{aligned}
$$

Then 4 is proved as:

$$
\begin{aligned}
\frac{1}{T^4}\sum_{t=1}^{T} N_{t-1}^4 &= \frac{1}{T^4}\sum_{t=1}^{T}\Big(n_0+\sum_{i=1}^{t-1} M_i\Big)^4 \\
&= n_0^4 T + 4n_0^3 \sum_{t=1}^{T}\sum_{i=1}^{t-1} M_i + 6n_0^2 \sum_{t=1}^{T}\Big(\sum_{i=1}^{t-1} M_i\Big)^2 + 4\sum_{t=1}^{T}\Big(\sum_{i=1}^{t-1} M_i\Big)^3 + \sum_{t=1}^{T}\Big(\sum_{i=1}^{t-1} M_i\Big)^4 \\
&\xrightarrow{a.s.} \frac{\lambda^4}{5}.
\end{aligned}
$$

□

Now we prove for Theorem 4.2.3.

Proof. Note that the estimator can be formed as

$$\hat{p}_{MLE} = \frac{\sum_{t=1}^{T} \sum_{i<j,i,j\in V_{t-1}} X_{ij}^{t-1} X_{ij}^{t}}{\sum_{t=1}^{T} \sum_{i<j,i,j\in V_{t-1}} X_{ij}^{t-1}} = \frac{\sum_{t=1}^{T} Z_t}{\sum_{t=1}^{T} L_{t-1}} + p,$$

where

$$Z_t = \sum_{i<j,i,j\in V_{t-1}} X_{ij}^{t-1}(X_{ij}^{t} - p).$$

Then Z_t is MDA, follows from the proof of Theorem 4.1.2. In order to use the Martingale CLT, We only need to verify the Lyapunov's condition. Here we also have

$$\begin{aligned}
&\frac{1}{T^6}\sum_{t=1}^{T} \mathrm{E}[Z_t^4|\mathscr{F}_{t-1}] \\
=&\frac{1}{T^6}\sum_{t=1}^{T} \mathrm{E}\Big(\Big[\sum_{i<j,i,j\in V_{t-1}} X_{ij}^{t-1}(X_{ij}^{t} - p)\Big]^4 \Big| X^{t-1}\Big) \\
=&\frac{1}{T^6}\sum_{t=1}^{T}\Big\{\sum_{i<j,i,j\in V_{t-1}} (X_{ij}^{t-1})^4\,\mathrm{E}\big[(X_{ij}^{t}-p)^4\big|X^{t-1}\big] \\
&+8\sum_{i<j,i,j\in V_{t-1}}\ \sum_{k<l,k,l\in V_{t-1},(k,l)\neq(i,j)} (X_{ij}^{t-1})^3 X_{kl}^{t-1}\,\mathrm{E}\big[(X_{ij}^{t}-p)^3(X_{kl}^{t}-p)\big|X^{t-1}\big] \\
&+6\sum_{i<j,i,j\in V_{t-1}}\ \sum_{k<l,k,l\in V_{t-1},(k,l)\neq(i,j)} (X_{ij}^{t-1})^2 (X_{kl}^{t-1})^2\,\mathrm{E}\big[(X_{ij}^{t}-p)^2(X_{kl}^{t}-p)^2\big|X^{t-1}\big] \\
&+\sum_{(i,j)}\sum_{(k,l)}\sum_{(m,n)}\sum_{(r,s)} X_{ij}^{t-1}X_{kl}^{t-1}X_{mn}^{t-1}X_{rs}^{t-1}\,\mathrm{E}\big[(X_{ij}^{t}-p)(X_{kl}^{t}-p)(X_{mn}^{t}-p)(X_{rs}^{t}-p)\big|X^{t-1}\big]\Big\} \\
=&\frac{1}{T^6}\sum_{t=1}^{T}(\mathrm{I}+8\mathrm{II}+6\mathrm{III}+\mathrm{IV}),
\end{aligned}$$

where $\mathrm{II}=0, \mathrm{IV}=0$ based on the independence of the edges, and that the conditional expectation

$X_{ij}^{t-1}\,\mathrm{E}[(X_{ij}^{t}-p)|X^{t-1}]=0$. Now for part I, by applying Lemma 4.2.4

$$
\begin{aligned}
\mathrm{I} &= \frac{1}{T^6}\sum_{t=1}^{T}\mathrm{E}\Big\{\sum_{i<j,i,j\in V_{t-1}}(X_{ij}^{t-1})^4(X_{ij}^{t}-p)^4\Big|X^{t-1},M_1,...,M_{t-1}\Big\}\\
&= \frac{1}{T^6}\sum_{t=1}^{T}\mathrm{E}\Big\{\sum_{i<j,i,j\in V_{t-1}}X_{ij}^{t-1}(X_{ij}^{t}-3pX_{ij}^{t}+4p^2X_{ij}^{t}-3p^3X_{ij}^{t}+p^4)\Big|X^{t-1},M_1,...,M_{t-1}\Big\}\\
&= \frac{1}{T^6}\sum_{t=1}^{T}\sum_{i<j,i,j\in V_{t-1}}X_{ij}^{t-1}(p-3p^2+4p^3-3p^4+p^4)\\
&\le \frac{1}{T^6}\sum_{t=1}^{T}\frac{N_{t-1}(N_{t-1}-1)}{2}(p-3p^2+4p^3-2p^4)\\
&= (p-3p^2+4p^3-2p^4)\frac{1}{T^6}O(T^3)\\
&\to 0, \text{ as } T\to\infty.
\end{aligned}
$$

For part III,

$$
\begin{aligned}
\mathrm{III} =& \frac{1}{T^6}\sum_{t=1}^{T}\mathrm{E}\Big\{\sum_{i<j,i,j\in V_{t-1}}\ \sum_{k<l,k,l\in V_{t-1},(k,l)\neq(i,j)}(X_{ij}^{t-1})^2(X_{kl}^{t-1})^2(X_{ij}^{t}-p)^2(X_{kl}^{t}-p)^2\\
&\Big|X^{t-1},M_1,...,M_{t-1}\Big\}\\
=&\frac{1}{T^6}\sum_{t=1}^{T}\sum_{i<j,i,j\in V_{t-1}}\ \sum_{k<l,k,l\in V_{t-1},(k,l)\neq(i,j)}X_{ij}^{t-1}X_{kl}^{t-1}\,\mathrm{E}\big[(X_{ij}^{t}-p)^2\big|X^{t-1}\big]\mathrm{E}\big[X_{kl}^{t}-p)^2\big|X^{t-1}\big]\\
=&\frac{1}{T^6}\sum_{t=1}^{T}\sum_{i<j,i,j\in V_{t-1}}\ \sum_{k<l,k,l\in V_{t-1},(k,l)\neq(i,j)}X_{ij}^{t-1}p(1-p)X_{kl}^{t-1}p(1-p)\\
<&p^2(1-p)^2\frac{1}{T^6}\sum_{t=1}^{T}\Big[\frac{N_{t-1}(N_{t-1}-1)}{2}\Big]^2\\
=&p^2(1-p)^2\frac{1}{T^6}O(T^5)\\
\to& 0, \text{ as } T\to\infty.
\end{aligned}
$$

Thus proved. □

4.3 Simulation Study

4.3.1 Settings

In this section, we summarized our simulation studies for GSMDN models: the base (FR) model and the generalized (RP) model. The networks are initialized with total nodes of 10 and edge probability of 1/3. There is 1 new coming nodes at each time slot for the FR model, and for the RP model, the number of new coming nodes each time follows a Poisson(1) distribution. Here we assume no nodes become inactive. The total time we run for the dynamic network is 500. For the keeping probability p, appearing probability q and forming probability a are set to be combinations of the sets: $p \in (0.9, 0.5, 0.2)$, $q \in (0.4, 0.05)$, and $a = 0.2$. Figure 4.1 shows the network structure in sampled time for the RP model from one of the simulations. Figure 4.2 shows the change in network density over time for each settings.

First, we give the estimated keeping probability using the form in (4.4). Then we calculated the 95% theoretical confidence interval using the form in (4.7). In addition, for each simulated dynamic network, we use Bootstrap method to resample the nodes and edges and get the 95% Bootstrap confidence intervals. At last, coverage rates are given for both theoretical and Bootstrap confidence intervals.

Here we compare the theoretical estimation of the variation of parameters with two different ways of Bootstrapping the dynamic network. The first one is Parametric Bootstrap, where we only resample the initial network. Then, we use the estimated parameters to generate future network. The second method Bootstraps the nodes and edges in every time slots. The details are in the Appendix A.

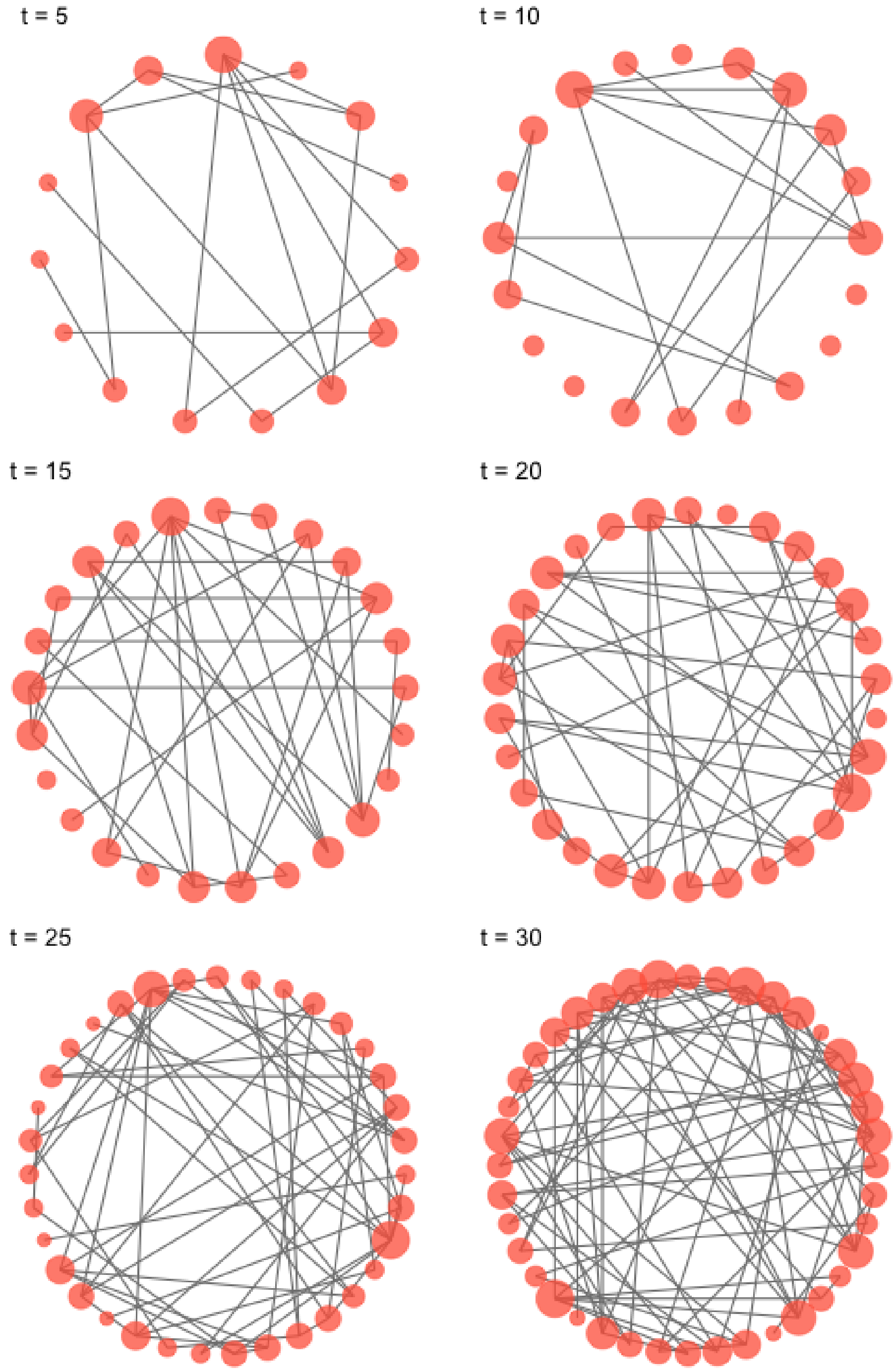
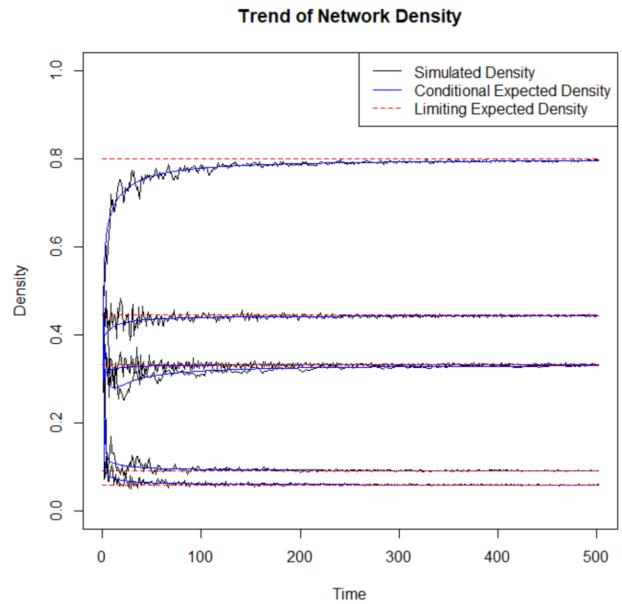


Figure 4.1 Simulated Networks from the RP Model. The size of the node reflects the number of connections it has.



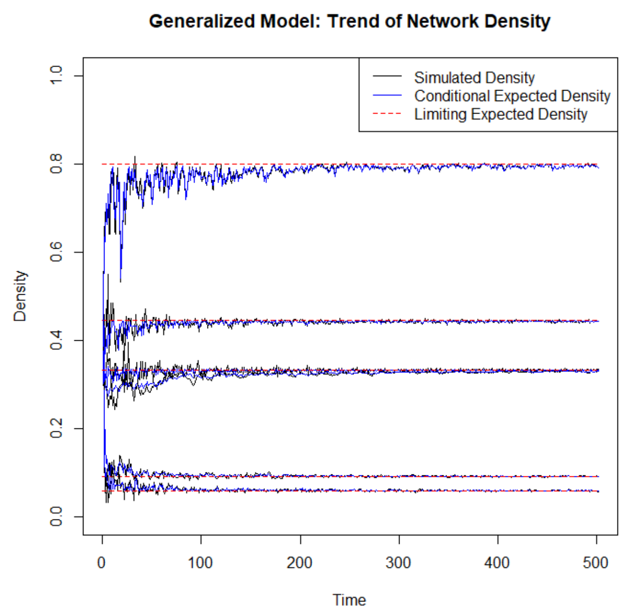


Figure 4.2 Behaviors of Network Density for FR Model (above) and RP Model (below). The probabilities are set to be combinations of the sets: $p \in (0.9, 0.5, 0.2)$, $q \in (0.4, 0.05)$, and $a = 0.2$, which corresponds to the limiting density $\rho = q/(1 - p + q) \in \{0.33, 0.80, 0.09, 0.44, 0.06, 0.33\}$. The initial networks are set to be the same with a density of 0.33. From the plots, we can see that the density of the FR model tends to stable much faster than the RP model.

4.3.2 Simulation Results

We summarized the simulation results in the following two tables: Table 4.1 shows the estimations for keeping probabilities p of the FR model; Table 4.2 shows those of the RP model. From the simulation results, we can see that for both models, the MLEs are quite near the true values of p. It is possible to get a good estimation with an MSE at the level 10^{-5} when $T = 50$. The variances of the estimators are also at the level of 10^{-5}, which is relatively small compared to the true parameters.

However, by looking at the results of coverage rate of the constructed confidence intervals, we found that the 95% theoretical CI has a higher rate than expected when T is small. This means that when T is small, the theoretical CI is conservative. However, when T goes larger, the coverage rate tends to be quite near 95%. This convergence tends to be faster when p and q are both larger. Both the Parametric Bootstrap and All-time Bootstrap methods give good estimation on confidence intervals. Especially when T is small, both tends to have better estimation than the theoretical one. However, they tend to be excessive with coverage rates slightly lower than 95%. In addition, the Bootstrap CI requires much heavier computation than theoretical method because of the repeated sampling.

Table 4.1 Simulation Results for FR Model. For each setting, the dynamic network is simulated for $t: 0 \to 500$, and Bootstrapped for 200 times to get the 95% CIs. The estimation results, coverage, variance and the MSE are then calculated from 500 repeated simulations. Note that the coverage tends to be consistently lower than 95%, which is mainly due to the bias from the network density.

Parms	T	$\hat{p}$	Cvr.Theo	Cvr.ParBoot	Cvr.Boot	Var	MSE
$p = .9$	50	0.8999655	0.986	0.954	0.954	8.348e-06	8.332e-06
$q = .05$	100	0.9000035	0.968	0.942	0.944	1.288e-06	1.286e-06
$a = .2$	200	0.9000009	0.972	0.948	0.950	1.752e-07	1.748e-07
	500	0.9000047	0.952	0.936	0.948	1.273e-08	1.273e-08
$p = .9$	50	0.8999602	0.990	0.926	0.928	3.692e-06	3.686e-06
$q = .4$	100	0.9000239	0.956	0.936	0.926	5.827e-07	5.821e-07
$a = .2$	200	0.9000136	0.956	0.928	0.930	8.293e-08	8.295e-08
	500	0.9000034	0.958	0.940	0.932	5.276e-09	5.278e-09
$p = .5$	50	0.4999480	0.988	0.940	0.928	7.422e-05	7.408e-05
$q = .05$	100	0.5001162	0.972	0.922	0.938	1.309e-05	1.308e-05
$a = .2$	200	0.5000635	0.976	0.960	0.936	1.567e-06	1.567e-06
	500	0.5000037	0.938	0.930	0.928	1.317e-07	1.315e-07
$p = .5$	50	0.4999043	0.988	0.930	0.942	1.726e-05	1.723e-05
$q = .4$	100	0.4999616	0.980	0.938	0.938	2.760e-06	2.756e-06
$a = .2$	200	0.4999777	0.952	0.936	0.956	4.097e-07	4.094e-07
	500	0.4999826	0.954	0.936	0.938	2.548e-08	2.574e-08
$p = .2$	50	0.1994812	0.994	0.952	0.958	6.761e-05	6.774e-05
$q = .05$	100	0.1999452	0.980	0.948	0.936	1.130e-05	1.128e-05
$a = .2$	200	0.2000394	0.972	0.950	0.922	1.618e-06	1.616e-06
	500	0.1999981	0.944	0.932	0.936	1.305e-07	1.302e-07
$p = .2$	50	0.2003509	0.986	0.940	0.934	1.413e-05	1.422e-05
$q = .4$	100	0.2000628	0.968	0.950	0.934	2.210e-06	2.209e-06
$a = .2$	200	0.2000296	0.952	0.932	0.928	3.421e-07	3.423e-07
	500	0.2000028	0.952	0.940	0.932	2.280e-08	2.276e-08

Table 4.2 Simulation Results for RP Model. For each setting, the dynamic network is simulated for $t: 0 \rightarrow 500$, and Bootstrapped for 200 times to get the 95% CIs. The estimation results, coverage, variance and the MSE are then calculated from 500 repeated simulations. Note that the coverage tends to be consistently lower than 95%, which is mainly due to the bias from the network density.

Parms	T	$\hat{p}$	Cvr.Theo	Cvr.ParBoot	Cvr.ABoot	Var	MSE
$p = .9$	50	0.8998509	0.972	0.936	0.946	9.566e-06	9.569e-06
$q = .05$	100	0.8998697	0.964	0.934	0.932	1.440e-06	1.454e-06
$a = .2$	200	0.8999769	0.956	0.948	0.918	1.847e-07	1.848e-07
	500	0.9000010	0.938	0.920	0.924	1.358e-08	1.355e-08
$p = .9$	50	0.8999829	0.978	0.940	0.948	3.826e-06	3.818e-06
$q = .4$	100	0.8999845	0.966	0.942	0.930	5.911e-07	5.902e-07
$a = .2$	200	0.8999770	0.968	0.948	0.928	7.436e-08	7.474e-08
	500	0.9000007	0.952	0.940	0.948	4.984e-09	4.974e-09
$p = .5$	50	0.5000568	0.986	0.956	0.926	7.521e-05	7.506e-05
$q = .05$	100	0.5001102	0.980	0.940	0.948	1.185e-05	1.184e-05
$a = .2$	200	0.5000066	0.972	0.946	0.948	1.776e-06	1.772e-06
	500	0.4999983	0.946	0.940	0.936	1.262e-07	1.260e-07
$p = .5$	50	0.4998834	0.980	0.942	0.952	1.664e-05	1.662e-05
$q = .4$	100	0.5001018	0.974	0.942	0.940	2.599e-06	2.604e-06
$a = .2$	200	0.5000400	0.972	0.948	0.944	3.700e-07	3.708e-07
	500	0.5000118	0.958	0.944	0.954	2.641e-08	2.649e-08
$p = .2$	50	0.1996755	0.984	0.946	0.946	7.713e-05	7.708e-05
$q = .05$	100	0.1998261	0.984	0.936	0.940	1.177e-05	1.178e-05
$a = .2$	200	0.1999804	0.964	0.936	0.950	1.823e-06	1.820e-06
	500	0.1999801	0.956	0.934	0.950	1.285e-07	1.286e-07
$p = .2$	50	0.1997783	0.984	0.942	0.942	1.483e-05	1.485e-05
$q = .4$	100	0.2000487	0.964	0.940	0.938	2.276e-06	2.274e-06
$a = .2$	200	0.2000167	0.962	0.938	0.950	3.007e-07	3.004e-07
	500	0.1999929	0.954	0.944	0.946	2.136e-08	2.137e-08

4.4 Real Data: MathOverflow User Interaction Dynamic Network

In this section, we will introduce the MathOverflow dynamic network data set downloaded from [LK14], and originally collected by [Par17]. MathOverflow is an online community for mathematicians to post mathematical questions and receive answers from other users. People may also comment on the posted question or the answers to it. The total time span of the data is 2350 days.

4.4.1 Data Description

MathOverflow dynamic network is a directed social network, where nodes stand for users, and edges stand for interactions of one of the followings:

- Answer to a posted question;
- Comment on a posted question;
- Comment on a posted answer.

However, we neglect the directions of the network and treat the three interactions as the same in order to fit our model. The data set has three columns: i, j, *ts*, where nodes pair (i, j) stands for edge between user i and j at UNIX timestamp *ts*. Firstly, we transform the variable *ts* and update the network every 30 days as we are interested in monthly interactions. In this manipulating step, edges within one time slot are recorded with no weight and direction. That is, if there are two or more interactions between user i and j, we still count it as one during that time.

Finally, we get a data set with 24759 unique users and 244779 total edges for the 30-day dynamic network. This data set also has three columns: i, j, t, where t stands for the start date of the 30-day time interval. Table 4.3 summarized the basic information of the MathOverflow dynamic network before and after cleaning.

In addition to the 30-day dynamic network, we also create 7-day and 60-day dynamic networks to analyze the weekly and 60-day's interactions among users. The results of all three different networks will be shown and compared in later sections.

Table 4.3 MathOverflow Dynamic Network

Data	MathOverflow (Raw)	MathOverflow (Cleaned)
Time	Timestamp, 389952 time points	Monthly, 79 months in total
Node	24759	24759
Edge	390441	244779
Avg Density	-	0.008
Avg Nodes	-	1132
Avg edges	-	6197
Description	Edge represents 3 possible interactions, points from the actor to the poster.	Edge represents at least one interaction between two users.

4.4.2 Data Visualization

Figure 4.3 shows the network structure constructed among users with ID 1-100 on selected dates. We can see that the network is denser at the beginning when the website first came out than later. A large part of them are inactive with no edges at all.

Figure 4.4 shows the monthly network features over time. From the top plot, we can see that the total active users of the website grows rapidly at the first few months. Then, the number fluctuates with a slightly growing trend. From the middle plot, we can tell that unique interactions grows rapidly first. Then we see an exponential drop with a flat tail. From the bottom plot, we see that the network density drops over time. But it tends to be stable at some low density in the end. It is easy to understand the behavior of network like this. In Question-and-Answer website like MathOverflow, new users enters the network rapidly at the beginning. In addition, people are extremely active when they first register an account in the website. This is usually referred to as the novelty effect. The users and interactions tends to be stable after the website runs for some time.

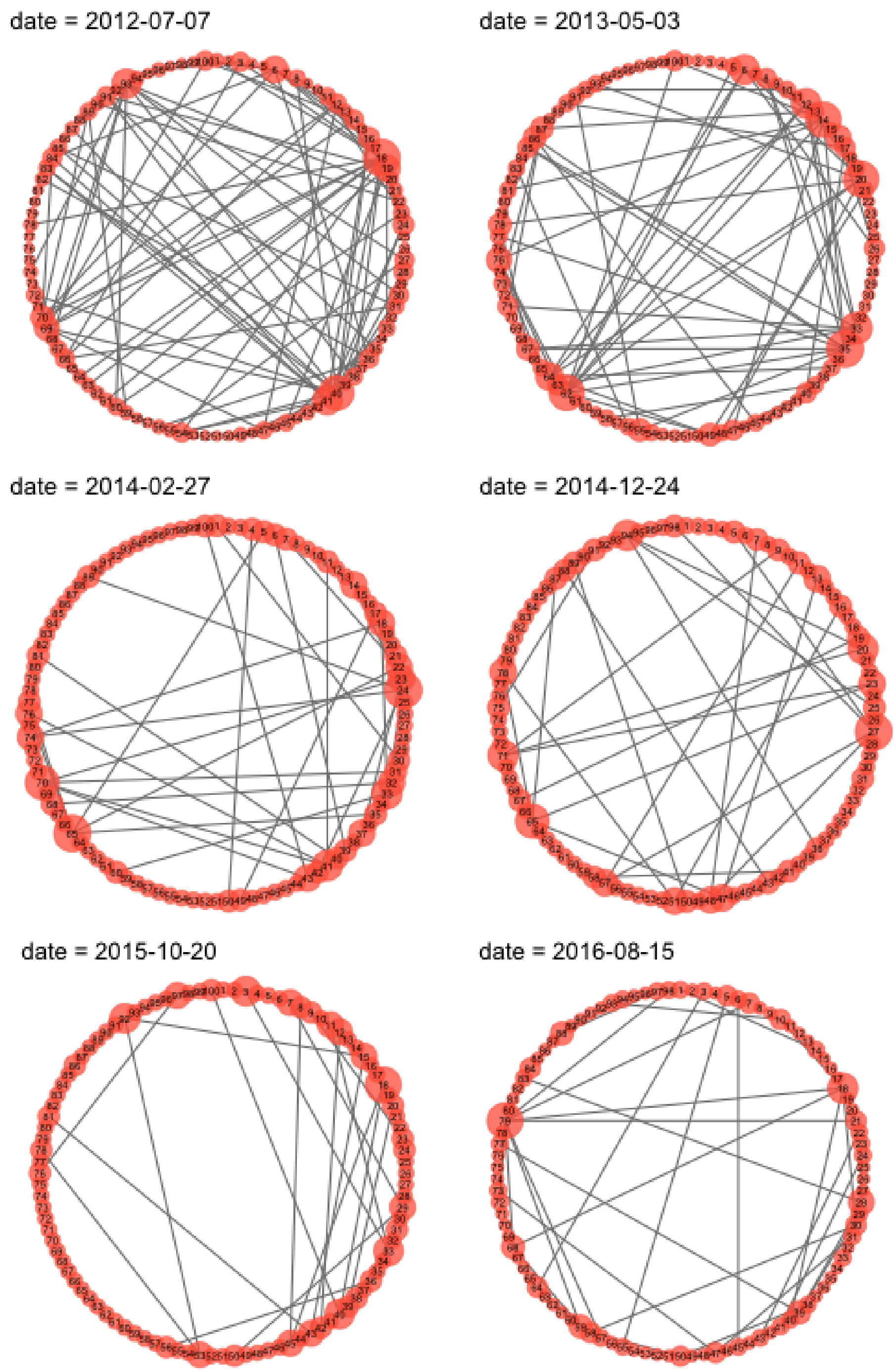


Figure 4.3 MathOverflow Dynamic Network in selected dates. The nodes are placed in clockwise order from user ID 1 to 100. The size of the node reflects the number of connections it has.

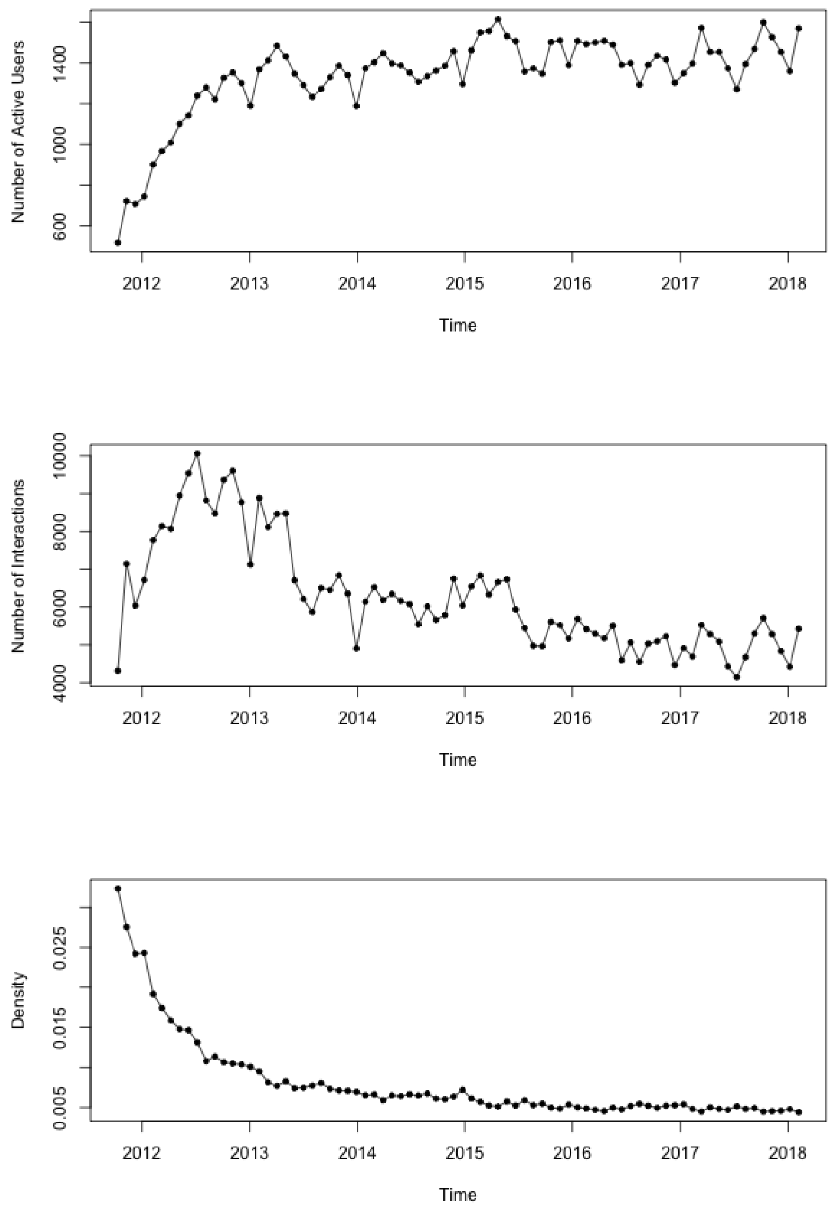


Figure 4.4 MathOverflow Dynamic Network Monthly Features: The top plot shows that the number of active users during each time slot. The middle one shows the total interactions (answers to a posted question, comment on a posted question, comment on a posted answer) among users. The bottom plot presents us the trend of network density over time. We can tell clearly that the network grows fast at the beginning, and then tends to stable after about 1 year.

4.4.3 Real Data Results

Table 4.4 shows the results of estimated parameters. From the results, we can see that the keeping probability p is approximately 0.066, which is significantly higher than the appearing probability $q \approx 0.001$. By looking at the keeping probability and the appearing probability of the MathOverflow Interaction Network, users are more likely to communicate (answers to a posted questions, comment on a posted questions or comment on a posted answer) with those who already had some communications with them than those who had not. The first reason is that people are more likely to reply with a post he had answered. Another reason is probably because of the recommendation systems of website. MathOverflow might recommend the posts of a user to another more often if they had communications before.

In addition, by looking at the forming probability, we can conclude that it is significantly lower than both forming probability and appearing probability. A possible reason is that new comers are less familiar with the website, so that they interact less with others than the old users. As the dynamic network we construct here is 30-days based network, we might also guess that new comers tends to look for answers that already exist in the website instead of post questions or answer others' questions. However, more data and deeper analysis are required to better understand the reason. Therefore, it is possible to make suggestions on how to improve the website. For example, engineers of the website should focus more on the new comers and modify its recommendation algorithms to them, so that new comers could get more involved in the network.

While comparing the results from the 7-day updated network, 30-day updated network and the 60-day updated network, we found that as the time interval gets larger, the keeping probability gets higher, while both the appearing probability and the forming probability gets lower. The results are straightforward, since when we lengthen the updating time interval, the users that communicate and re-communicate in longer period are considered to "keep their connection" instead of "lose connection then re-connect".

In conclusion, by fitting the GSMDN on the MathOverflow Interaction Network, we are able to get a good sense about the monthly behavior of both old users and new users of the website. Thus, we can get good advice on an which part to improve for the website.

Table 4.4 Results: MathOverflow Interaction Network

Time Interval	Parm	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\theta}$	Var	95% CI
7 days	p	297.86	296.12	0.063056143	1.059e-06	(0.063054069, 0.063058218)
	q			0.001610795	4.954e-11	(0.001610795, 0.001610795)
	a			0.002389932	1.417e-08	(0.002389904, 0.002389959)
30 days	p	725.66	711.99	0.066089403	2.110e-07	(0.066088989, 0.066089816)
	q			0.001406097	7.226e-12	(0.001406097, 0.001406097)
	a			0.001309537	2.359e-09	(0.001309533, 0.001309542)
60 days	p	1139.71	1103.00	0.073744256	1.158e-07	(0.073744029, 0.073744483)
	q			0.001208759	2.672e-12	(0.001208759, 0.001208759)
	a			0.000989454	1.016e-09	(0.000989452, 0.000989456)

CHAPTER

5

MULTI-CLASSES MARKOV DYNAMIC NETWORK

In the usual settings of Stochastic Block model, nodes are classified into different classes. There is no overlap among classes. The characteristics of the nodes are determined by the classes they belong to. This is sometimes not true in practical. Take social network as an example, a user can belong to a university as well as a company where he is working as a intern. He might have a lot of interactions with people from both groups, or share similar characteristics with people from both groups. In addition, he might change his groups as he moves or changes company. The Multi-Classes Markov Dynamic Network (MCMDN) model we introduce here can better characterize this feature by using the Markov class vectors.

5.1 Model Settings

Suppose there are K different classes. Let $c_i^t = (c_{i,1}^t, ..., c_{i,K}^t)^T \in \{0,1\}^K$ stands for the class vector of node i at time t. To be specific, each node can belong to multiple classes, and

$$c_{i,k}^t = \begin{cases} 1, & i \text{ is in class } k \\ 0, & i \text{ is not in class } k \end{cases}$$

In addition, the class vector at time t is a function of the network edges at time $t-1$. That is:

$$c_{i,k}^t = \begin{cases} \mathbb{1}\big(\zeta(\eta_i, X_i^{t-1}) \geq b_k\big), & \text{if } k < K \\ \prod_{k=2}^K \mathbb{1}\big(\zeta(\eta_i, X_i^{t-1}) < b_k\big), & \text{if } k = K \end{cases} \tag{5.1}$$

The class K is the class of nodes that doesn't satisfy the conditions of any classes $k < K$ at time $t-1$. b_k is the class threshold. η_i are covariates of node i. As in the classical Stochastic Block model, there is also a class connection matrix $\boldsymbol{\theta}_{K\times K}$ in MCMDN model characterizing the features among the classes. Let

$$\boldsymbol{\theta}_{K\times K} = \begin{pmatrix} \theta_{11} & \theta_{12} & \dots & \theta_{1K} \\ \theta_{21} & \theta_{22} & \dots & \theta_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{K1} & \theta_{K2} & \dots & \theta_{KK} \end{pmatrix}.$$

Then the probability of an edge forming between two nodes at time t is

$$p_{ij}^t = P(X_{ij}^t = 1 | X^{t-1}) = H(c_i^t, c_j^t | \boldsymbol{\theta}).$$

It is the function of class vectors of the nodes pair based on the kernel matrix. A reasonable kind of function is linear function of the elements of $\boldsymbol{\theta}$ as follows

$$p_{ij}^t = \sum_{i,j} h(c_i^t, c_j^t)\theta_{ij} = \frac{\sum_{m=1}^K \sum_{l=1}^K c_{i,l}^t c_{j,m}^t \cdot \theta_{lm}}{\sum_{m=1}^K \sum_{l=1}^K c_{i,l}^t c_{j,m}^t} = \frac{c_i^{tT} \boldsymbol{\theta} c_j^t}{c_i^{tT} \mathbf{1} c_j^t}, \tag{5.2}$$

where

$$\mathbf{1}_{K\times K} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{pmatrix}.$$

The MCMDN model shows its advantage over the SBM in many aspects. Firstly, the Markov properties allows short time dependence of edges, thus makes it easier for future prediction. The process is dynamic but still stable as it is long time independent.

Secondly, the model allows for much more flexibility in assigning node's classes. Unlike traditional classification problems where each node is strictly restricted to one class, the MCMDN model allows each node being assigned to multiple classes. Since the SBM separates the nodes into disjoint communities, the nodes are either alike or dislike in between. It is hard to characterize the closeness. However, by allowing multiple classes for an individual node, it is possible that nodes can be alike in part of the communities but different in some others. This is more intuitive, and the model can be applied to real network (especially social network) more easily.

Lastly, the MCMDN model allows more flexibility of the network sizes. Instead of fixed size, it can also be dynamic. In another word, nodes can be inactive and reactivated at each time. However, there is an upper bound on the total number of nodes over all times. Here, we denote the maximum total nodes to be N.

5.2 Markov Properties

In order to study the long time behavior of the network, we vectorize the upper triangle of the adjacency matrix $\{X_{ij}^t\}_{1\le i<j\le N}$. Let

$$\mathbf{X}^t = \begin{pmatrix} X_{12}^t \\ X_{1N}^t \\ \vdots \\ X_{(N-1),N}^t \end{pmatrix},$$

where N is the maximum total nodes of the network. Thus $\{\mathbf{X}^t\}_{t\ge 0}$ is a Markov Chain with state space $\mathbb{S}$ and transition probability $\mathbf{P}$. Since N is fixed and finite, the state space $\mathbb{S}$ is finite. Thus, to show the existence of invariant probability measure, it suffices to show the irreducibility of the transition probability matrix.

First, for any given status $A, B \in \mathbb{S}$, A and B are communicate, as the probability of $A \to B$ in one time step is

$$P_{AB} = P(A \to B) = \prod_{i<j} \left(p_{ij}(A)\right)^{B_{ij}} \left(1 - p_{ij}(A)\right)^{(1-B_{ij})} > 0$$

where $p_{ij}(A)$ stands for the probability of linkage between i and j based on vectorized adjacency matrix A. Furthermore, we can conclude that the transition probability matrix is irreducible. Therefore, there exists an unique stationary distribution π of the Markov Chain $(\mathbb{S}, \mathbf{P})$. For any initial distribution μ, the distribution at time T converges to π in average sense, and the Law of Large Numbers (LLN) holds.

5.3 Analysis of MLE

We can write down the likelihood as below:

$$L(\boldsymbol{\theta}) = \prod_{t=1}^{T} \left\{ \prod_{i<j} (p_{ij}^t)^{x_{ij}^t} (1 - p_{ij}^t)^{1-x_{ij}^t} \right\} = \prod_{t=1}^{T} f(X^t, C^t | \boldsymbol{\theta}) = \prod_{t=1}^{T} f(X^t | X^{t-1}, \boldsymbol{\theta}).$$

Since C^t is the function of X^{t-1}, there is short time dependency. The log-likelihood is

$$\begin{aligned}
l(\boldsymbol{\theta}) &= \sum_{t=1}^{T}\Big\{\sum_{i<j}\big[x_{ij}^t \log p_{ij}^t + (1-x_{ij}^t)\log(1-p_{ij}^t)\big]\Big\} \\
&= \sum_{t=1}^{T}\Big\{\sum_{i<j}\big[x_{ij}^t \log p_{ij}^t(\boldsymbol{\theta}) + (1-x_{ij}^t)\log(1-p_{ij}^t(\boldsymbol{\theta}))\big]\Big\} \\
&= \sum_{t=1}^{T}\Big\{\sum_{i<j}\Big[x_{ij}^t \log\Big(\frac{c_i^{tT}\boldsymbol{\theta} c_j^t}{c_i^{tT}\mathbf{1} c_j^t}\Big) + (1-x_{ij}^t)\log\Big(1-\frac{c_i^{tT}\boldsymbol{\theta} c_j^t}{c_i^{tT}\mathbf{1} c_j^t}\Big)\Big]\Big\} \\
&= \sum_{t=1}^{T}\log f(X^t, C^t|\boldsymbol{\theta}) \\
&= \sum_{t=1}^{T}\log f(X^t|X^{t-1}, \boldsymbol{\theta}).
\end{aligned}$$

The mathematical solution is not available here. We should seek some numerical method like Gradient Descent or Newton's Method. The details is discussed in the Appendix B.

5.3.1 Asymptotic Distribution of MLE

We have the following theorem regarding the MLE of the MCMDN model.

Theorem 5.3.1. *Under the MCMDN Model, the estimator given by the Maximum Likelihood method follows the asymptotic distribution that*

$$\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \xrightarrow{d} N(0, \mathscr{I}(\boldsymbol{\theta}_0)^{-1}), \tag{5.3}$$

where $\mathscr{I}(\boldsymbol{\theta}_0) = \mathrm{E}_{\boldsymbol{\theta}_0}\big(-\frac{\partial^2 \log f(X^t|X^{t-1},\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\big) = \mathrm{E}_{\boldsymbol{\theta}_0}\big(\frac{\partial \log f(X^t|X^{t-1},\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\frac{\partial \log f(X^t|X^{t-1},\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}'}\big)$*, and* $\boldsymbol{\theta}_0$ *is the underlying true parameter of the model.*

In order to show Theorem 5.3.1, we need to show some lemmas first. Then, by using Taylor expansion and Slutsky's Theorem. we can easily prove the form (5.3).

Lemma 5.3.2. *Suppose* $f(X^t|X^{t-1}, \boldsymbol{\theta}_0)$ *is the joint probability density function of the network edges*

X^t based on its former status X^{t-1} under the MCMDN Model, then,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial \log f(X^t|X^{t-1},\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \xrightarrow{d} N\big(0,\mathscr{I}(\boldsymbol{\theta}_0)\big),$$

where $\mathscr{I}(\boldsymbol{\theta}_0)=\mathrm{E}_{\boldsymbol{\theta}_0}\big(-\frac{\partial^2 \log f(X^t|X^{t-1},\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\big)=\mathrm{E}_{\boldsymbol{\theta}_0}\big(\frac{\partial \log f(X^t|X^{t-1},\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\frac{\partial \log f(X^t|X^{t-1},\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}'}\big)$, and $\boldsymbol{\theta}_0$ is the underlying true parameter of the model.

Proof. We will use Martingale CLT to prove the lemma. Let

$$Z_t=\frac{\partial \log f(X^t|X^{t-1},\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}},$$

where

$$\log f(X^t|X^{t-1},\boldsymbol{\theta}_0)=\sum_{i<j}\Big\{X_{ij}^t\log p_{ij}^t-(1-X_{ij}^t)\log(1-p_{ij}^t)\Big\},$$

and

$$\begin{aligned}\frac{\partial \log f(X^t|X^{t-1},\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}&=\sum_{i<j}\Big\{X_{ij}^t\frac{c_j^t c_i^{tT}}{c_i^{tT}\boldsymbol{\theta}_0 c_j^t}-(1-X_{ij}^t)\frac{c_j^t c_i^{tT}}{c_i^{tT}\mathbf{1}c_j^t-c_i^{tT}\boldsymbol{\theta}_0 c_j^t}\Big\}\\ &=\sum_{i<j}\Big\{X_{ij}^t G(X^{t-1},\boldsymbol{\theta}_0)-(1-X_{ij}^t)H(X^{t-1},\boldsymbol{\theta}_0)\Big\}.\end{aligned}$$

Then let $\mathscr{F}_a=\sigma\langle X^t: t\le a, t\in\mathbb{Z}\rangle, -\infty\le a\le b\le\infty$. Now we show that Z_t is MDA. Firstly,

$$\mathrm{E}_{\boldsymbol{\theta}_0}(X_{ij}^t|\mathscr{F}_{t-1})=p_{ij}^t=\frac{c_i^{tT}\boldsymbol{\theta}_0 c_j^t}{c_i^{tT}\mathbf{1}c_j^t},$$

and,

$$
\begin{aligned}
\mathrm{E}_{\boldsymbol{\theta}_0}(Z_t|\mathscr{F}_{t-1}) = \mathrm{E}_{\boldsymbol{\theta}_0}\Big\{\sum_{i<j}\Big(X_{ij}^t G(X^{t-1},\boldsymbol{\theta}_0) - (1-X_{ij}^t)H(X^{t-1},\boldsymbol{\theta}_0)\Big)\Big|\mathscr{F}_{t-1}\Big\} \\
= \sum_{i<j}\Big\{\mathrm{E}_{\boldsymbol{\theta}_0}(X_{ij}^t|\mathscr{F}_{t-1})G(X^{t-1},\boldsymbol{\theta}_0) - \big(1-\mathrm{E}_{\boldsymbol{\theta}_0}(X_{ij}^t|\mathscr{F}_{t-1})\big)H(X^{t-1},\boldsymbol{\theta}_0)\Big\} \\
= \sum_{i<j}\Big\{\frac{c_i^{tT}\boldsymbol{\theta}_0 c_j^t}{c_i^{tT}\mathbf{1}c_j^t}\frac{c_j^t c_i^{tT}}{c_i^{tT}\boldsymbol{\theta}_0 c_j^t} - \Big(1-\frac{c_i^{tT}\boldsymbol{\theta}_0 c_j^t}{c_i^{tT}\mathbf{1}c_j^t}\Big)\frac{c_j^t c_i^{tT}}{c_i^{tT}\mathbf{1}c_j^t - c_i^{tT}\boldsymbol{\theta}_0 c_j^t}\Big\} = 0.
\end{aligned}
$$

Therefore, Z_t is MDA. Secondly, we show that conditions for Martingale CLT are satisfied. For the second order expectation, we have

$$
\begin{aligned}
&\mathrm{E}_{\boldsymbol{\theta}_0}\Big(\frac{\partial^2 \log f(X^t|X^{t-1},\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\Big) \\
=&\sum_{X^t}\frac{\partial^2 \log f(X^t|X^{t-1},\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}f(X^t|X^{t-1},\boldsymbol{\theta}_0) \\
=&\sum_{X^t}\frac{\partial}{\partial\boldsymbol{\theta}}\Big(\frac{\partial \log f(X^t|X^{t-1},\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}'}\Big)f(X^t|X^{t-1},\boldsymbol{\theta}_0) \\
=&\sum_{X^t}\frac{\partial}{\partial\boldsymbol{\theta}}\Big(\frac{1}{f(X^t|X^{t-1},\boldsymbol{\theta}_0)}\frac{\partial f(X^t|X^{t-1},\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}'}\Big)f(X^t|X^{t-1},\boldsymbol{\theta}_0) \\
=&\sum_{X^t}\Big(\frac{-1}{f^2(X^t|X^{t-1},\boldsymbol{\theta}_0)}\frac{\partial f(X^t|X^{t-1},\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\frac{\partial f(X^t|X^{t-1},\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}'}f(X^t|X^{t-1},\boldsymbol{\theta}_0) + \frac{\partial^2 f(X^t|X^{t-1},\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\Big) \\
=&\sum_{X^t}\Big(-\frac{\partial \log f(X^t|X^{t-1},\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\frac{\partial \log f(X^t|X^{t-1},\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}'}f(X^t|X^{t-1},\boldsymbol{\theta}_0) + \frac{\partial^2 f(X^t|X^{t-1},\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\Big) \\
=&\mathrm{E}_{\boldsymbol{\theta}_0}\Big(-\frac{\partial \log f(X^t|X^{t-1},\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\frac{\partial \log f(X^t|X^{t-1},\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}'}\Big) + \sum_{X^t}\frac{\partial^2 f(X^t|X^{t-1},\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} \\
=&\mathrm{E}_{\boldsymbol{\theta}_0}\Big(-\frac{\partial \log f(X^t|X^{t-1},\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\frac{\partial \log f(X^t|X^{t-1},\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}'}\Big),
\end{aligned}
$$

which follows from the fact that

$$
\sum_{X^t}\frac{\partial^2 f(X^t|X^{t-1},\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} = \frac{\partial^2 \sum_{X^t} f(X^t|X^{t-1},\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} \equiv 0.
$$

Now,

$$
\begin{aligned}
\mathrm{E}_{\boldsymbol{\theta}_0}(Z_t Z_t^T|\mathscr{F}_{t-1}) &= \mathrm{E}_{\boldsymbol{\theta}_0}\left(\frac{\partial \log f(X^t|X^{t-1},\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\frac{\partial \log f(X^t|X^{t-1},\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'}\middle|\mathscr{F}_{t-1}\right)\\
&= \mathrm{E}_{\boldsymbol{\theta}_0}\left(-\frac{\partial^2 \log f(X^t|X^{t-1},\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\middle|\mathscr{F}_{t-1}\right)\\
&= \sum_{i<j}\frac{c_j^t c_i^{tT}\bigotimes c_i^t c_j^{tT}}{c_i^{tT}\mathbf{1}c_j^t(c_i^{tT}\mathbf{1}c_j^t - c_i^{tT}\boldsymbol{\theta}_0 c_j^t)} = \mathscr{I}_t(\boldsymbol{\theta}_0).
\end{aligned}
$$

In addition, from the Markov properties of the network, there exists a unique stationary distribution of $\mathbf{X}^t$. Using the Weak Law of Large Numbers, we have

$$
\frac{1}{T}\sum_{t=1}^{T}\mathrm{E}_{\boldsymbol{\theta}_0}(Z_t Z_t^T|\mathscr{F}_{t-1}) = \frac{1}{T}\sum_{t=1}^{T}\mathscr{I}_t(\boldsymbol{\theta}_0)\xrightarrow{p}\mathscr{I}(\boldsymbol{\theta}_0).
$$

At last, we need to verify the Lindeberg's condition. On one hand,

$$
\begin{aligned}
\|Z_t\| &= \left\|\sum_{i<j}\left(X_{ij}^t\frac{c_j^t c_i^{tT}}{c_i^{tT}\boldsymbol{\theta}_0 c_j^t} - (1-X_{ij}^t)\frac{c_j^t c_i^{tT}}{c_i^{tT}\mathbf{1}c_j^t - c_i^{tT}\boldsymbol{\theta}_0 c_j^t}\right)\right\|\\
&= \left\|\sum_{i<j}\frac{X_{ij}^t c_i^{tT}\mathbf{1}c_j^t - c_i^{tT}\boldsymbol{\theta}_0 c_j^t}{c_i^{tT}\boldsymbol{\theta}_0 c_j^t(c_i^{tT}\mathbf{1}c_j^t - c_i^{tT}\boldsymbol{\theta}_0 c_j^t)}\cdot c_j^t c_i^{tT}\right\|\\
&= \left\|\sum_{i<j}\frac{X_{ij}^t - p_{ij}^t}{p_{ij}^t(1-p_{ij}^t)c_i^{tT}\mathbf{1}c_j^t}\cdot c_j^t c_i^{tT}\right\|\\
&\le \left\{\sum_{i<j}\left(\frac{1}{p_{ij}^t(1-p_{ij}^t)c_i^{tT}\mathbf{1}c_j^t}\right)^2\cdot K^2\right\}^{1/2}\\
&\le \left\{\sum_{i<j}\left(\frac{1}{p_{ij}^t(1-p_{ij}^t)}\right)^2\cdot K^2\right\}^{1/2},
\end{aligned}
$$

where K is the number of classes, $c_j^t c_i^{tT}$ are 0/1 matrices and $c_i^{tT}\mathbf{1}c_j^t \ge 1$. On the other hand, under the model assumptions,

$$
0 < a \le p_{ij}^t \le b < 1,
$$

where $a = \min_{i,j}\{\theta_{ij}\}$ and $b = \max_{i,j}\{\theta_{ij}\} < 1$. Thus,

$$p_{ij}^t(1-p_{ij}^t) \geq \min\{a(1-a), b(1-b)\} = c.$$

Therefore,

$$\|Z_t\| \leq \left\{\sum_{i<j}\left(\frac{K}{c}\right)^2\right\}^{1/2} = \left\{\frac{K(K-1)}{2}\left(\frac{K}{c}\right)^2\right\}^{1/2} = \left(\frac{K(K-1)}{2}\right)^{1/2}\frac{K}{c} = C.$$

Now, since

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}\mathrm{E}_{\boldsymbol{\theta}_0}\left(\|Z_t\|^2\mathbb{1}(\|Z_t\| > \epsilon\sqrt{T})|\mathscr{F}_{t-1}\right) &\leq \frac{1}{T}\sum_{t=1}^{T}\mathrm{E}_{\boldsymbol{\theta}_0}\left(C^2\mathbb{1}(C > \epsilon\sqrt{T})|\mathscr{F}_{t-1}\right)\\ &= \frac{1}{T}\sum_{t=1}^{T}C^2\mathbb{1}(C > \epsilon\sqrt{T})\\ &\to 0 \text{ as } T \to \infty,\end{aligned}$$

the Lindeberg's condition is satisfied. In conclusion, the lemma is proved. □

Lemma 5.3.3. ***(Kronecker's Lemma)*** *If $\{x_n\}_{n=1}^{\infty}$ is an infinite sequence of real numbers such that*

$$\sum_{m=1}^{\infty} x_m = s$$

exists and is finite, then we have for all $0 < b \leq b_2 \leq b_3 \leq \cdots$ and $b_n \to \infty$ that

$$\lim_{n\to\infty}\frac{1}{b_n}\sum_{k=1}^{n} b_k x_k = 0.$$

Proof. The lemma is given and proved in [AL06]. □

Lemma 5.3.4. *Suppose $l(\boldsymbol{\theta}|X^0, ..., X^T)$ is the log-likelihood function of the estimator under the MCMDN model. $\boldsymbol{\theta}^*$ is the point on the line segment of the true parameter $\boldsymbol{\theta}_0$ and the MLE $\boldsymbol{\theta}_T$, which* *satisfies the Taylor series expansion of the first order condition around* $\boldsymbol{\theta}_0$

$$0 = \frac{\partial l(\hat{\boldsymbol{\theta}}_T|X^0,...,X^T)}{\partial \boldsymbol{\theta}} = \frac{\partial l(\boldsymbol{\theta}_0|X^0,...,X^T)}{\partial \boldsymbol{\theta}} + \frac{\partial^2 l(\boldsymbol{\theta}^*|X^0,...,X^T)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0).$$

Then,

$$-\frac{1}{T}\frac{\partial^2 l(\boldsymbol{\theta}^*|X^0,...,X^T)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'} \xrightarrow{p} \mathscr{I}(\boldsymbol{\theta}_0). \tag{5.4}$$

Proof. Firstly,

$$-\frac{1}{T}\frac{\partial^2 l(\boldsymbol{\theta}^*|X^0,...,X^T)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'} = -\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2 \log f(X^t|X^{t-1},\boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}.$$

From the property of MLE, we have as $T \to \infty$, $\hat{\boldsymbol{\theta}}_T \to^p \boldsymbol{\theta}_0$. Therefore, $\boldsymbol{\theta}^* \to^p \boldsymbol{\theta}_0$. Then

$$-\frac{1}{T}\frac{\partial^2 l(\boldsymbol{\theta}^*|X^0,...,X^T)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'} \xrightarrow{p} -\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2 \log f(X^t|X^{t-1},\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}.$$

Let

$$\begin{aligned} Y_t &= -\frac{\partial^2 \log f(X^t|X^{t-1},\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'} - \mathrm{E}_{\boldsymbol{\theta}_0}\left(-\frac{\partial^2 \log f(X^t|X^{t-1},\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\middle|\mathscr{F}_{n-1}\right) \\ &= -\frac{\partial^2 \log f(X^t|X^{t-1},\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'} - \mathscr{I}_t(\boldsymbol{\theta}_0). \end{aligned}$$

Then,

$$\mathrm{E}_{\boldsymbol{\theta}_0}(Y_t|\mathscr{F}_{t-1}) = 0.$$

Moreover,

$$\begin{aligned} &\mathrm{E}_{\boldsymbol{\theta}_0}\left\{\left\|\frac{\partial^2 \log f(X^t|X^{t-1},\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\left(\frac{\partial^2 \log f(X^t|X^{t-1},\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\right)'\right\|\middle|\mathscr{F}_{t-1}\right\} \\ =&\mathrm{E}_{\boldsymbol{\theta}_0}\left\{\left\|\sum_{i<j}\left(\frac{X_{ij}^t}{(c_i^{tT}\boldsymbol{\theta}_0 c_j^t)^4} + \frac{1-X_{ij}^t}{(c_i^{tT}\mathbf{1}c_j^t - c_i^{tT}\boldsymbol{\theta}_0 c_j^t)^4}\right)Q_t Q_t'\right\|\middle|\mathscr{F}_{t-1}\right\}, \end{aligned} \tag{5.5}$$

where $Q_t = c_j^t c_i^{tT} \bigotimes c_i^t c_j^{tT}$ and $||\cdot||$ is the L_1 norm. Since

$$0 < a \leq p_{ij}^t \leq b < 1,$$

where $a = \min_{i,j}\{\theta_{ij}\}$ and $b = \max_{i,j}\{\theta_{ij}\} < 1$, it is easy to see that form (5.5) is bounded by a constant M for any $t > 0$. Therefore, there exists M' such that

$$\sup_t \mathrm{E}_{\boldsymbol{\theta}_0}(Y_t Y_t' | \mathscr{F}_{t-1}) \leq M' < \infty.$$

Let

$$M_t = \sum_{n=1}^{t} \frac{Y_n}{n}.$$

It is obvious that M_t is a martingale. In addition,

$$\mathrm{E}_{\boldsymbol{\theta}_0}\left(||M_t||\right) \leq \sum_{n=1}^{t} \frac{\mathrm{E}_{\boldsymbol{\theta}_0}||Y_n||}{n} \leq \sum_{n=1}^{t} \frac{\mathrm{E}_{\boldsymbol{\theta}_0}||Y_n||^2}{n^2} < \infty.$$

By the Martingale convergence theorem, we have that M_t converge almost surely. Then, by the Kronecker's Lemma,

$$\frac{1}{T}\sum_{t=1}^{T} Y_t \xrightarrow{a.s.} 0.$$

That is

$$\frac{1}{T}\sum_{t=1}^{T} -\frac{\partial^2 \log f(X^t | X^{t-1}, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} - \frac{1}{T}\sum_{t=1}^{T} \mathscr{I}_t(\boldsymbol{\theta}_0) \xrightarrow{a.s.} 0. \tag{5.6}$$

Combine form (5.6) and equation (5.4), we have

$$\frac{1}{T}\sum_{t=1}^{T} -\frac{\partial^2 \log f(X^t | X^{t-1}, \boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} - \frac{1}{T}\sum_{t=1}^{T} \mathscr{I}_t(\boldsymbol{\theta}_0) \xrightarrow{a.s.} 0. \tag{5.7}$$

At last, from Markov properties of $\{\mathbf{X}^T\}_{t>0}$, we can easily have the large number convergence that $\frac{1}{T}\sum_{t=1}^{T} \mathscr{I}_t(\boldsymbol{\theta}_0) \to^p \mathscr{I}(\boldsymbol{\theta}_0)$. Thus the lemma is proved. □

Now we can easily prove for Theorem 5.3.1.

Proof. Firstly, Taylor series expansion of the first order condition around the true value of $\boldsymbol{\theta}$, $\boldsymbol{\theta}_0$ yields,

$$0 = \frac{\partial l(\hat{\boldsymbol{\theta}}_T|X^0,...,X^T)}{\partial \boldsymbol{\theta}} = \frac{\partial l(\boldsymbol{\theta}_0|X^0,...,X^T)}{\partial \boldsymbol{\theta}} + \frac{\partial^2 l(\boldsymbol{\theta}^*|X^0,...,X^T)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0),$$

where $\boldsymbol{\theta}^*$ is on the line segment between $\hat{\boldsymbol{\theta}}_T$ and $\boldsymbol{\theta}_0$. Thus we have

$$\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) = -\left(\frac{1}{T}\frac{\partial^2 l(\boldsymbol{\theta}^*|X^0,...,X^T)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\right)^{-1}\frac{1}{\sqrt{T}}\frac{\partial l(\boldsymbol{\theta}_0|X^0,...,X^T)}{\partial \boldsymbol{\theta}}.$$

Now by Lemma 5.3.2, Lemma 5.3.4, and Slutsky's Theorem, we have Theorem 5.3.1 proved. □

5.4 Simulation Study

5.4.1 Settings

In this section, we did a simulation study on the MCMDN model. The network has 200 nodes over 50 time slots. The initial network has 40 nodes which are assigned to 4 blocks with probabilities $q = (0.3, 0.3, 0.3, 0)$. The 4th block is the newly active nodes block where we defined as who doesn't have any interaction with other nodes during last time interval. At initial status, we don't assign any nodes to block 4. The initial network is generated using the true underlying class connection probability matrix:

$$\boldsymbol{\theta}_{4\times 4} = \begin{pmatrix} 0.35 & 0.05 & 0.05 & 0.05 \\ 0.05 & 0.25 & 0.05 & 0.05 \\ 0.05 & 0.05 & 0.15 & 0.05 \\ 0.05 & 0.05 & 0.05 & 0.10 \end{pmatrix}.$$

Then, we generate a covariates data set which contains one categorical variable η_1 with 4 levels, and one continuous variable η_2. The covariates are time invariant, where

$$\eta_1 \sim \text{Multinomial}(0.1, 0.2, 0.3, 0.4),$$

$$\eta_2 \sim \text{Beta}(2, 2).$$

The classes for each node are generated in the following mechanic: For node i, suppose its neighbor nodes set is $\mathcal{N}_{t-1}^i$. The first step is calculate the scores for classes $k = 1,2,3$. The score:

$$s_{ik}^t = \begin{cases} \frac{\sum_{j\in\mathcal{N}_{t-1}^i} \mathbb{1}(c_{jk}^{t-1}=1)}{\sum_{j\in V_{t-1}} \mathbb{1}(c_{jk}^{t-1}=1)} + \frac{\sum_{j\in\mathcal{N}_{t-1}^i} \mathbb{1}(\eta_{j1}=\eta_{i1})}{|\mathcal{N}_{t-1}^i|} - \frac{\left|\eta_{i2}-\sum_{j\in\mathcal{N}_{t-1}^i} \eta_{j2}/|\mathcal{N}_{t-1}^i|\right|}{\sqrt{\sum_{j\in\mathcal{N}_{t-1}^i} \eta_{j2}/|\mathcal{N}_{t-1}^i|}}, & \mathcal{N}_{t-1}^i \neq \varnothing \\ 0, & \mathcal{N}_{t-1}^i = \varnothing \end{cases}$$

where the first item is the percentage of nodes from class k which are neighbors of node i. the second term is the percentage of neighbor nodes which are in the same level l of X_1 with node i. The last term stands for the standardized difference between X_2 of node i and the mean X_2 of all its neighbors. $\mathcal{N}_{t-1}^i = \varnothing$ indicates that node i is not active at time $t-1$.

The class vector elements for node i at time t are calculated as follows:

$$c_{ik}^t = \begin{cases} \mathbb{1}\left\{\frac{\exp(s_{ik}^t)}{\sum_k \exp(s_{ik}^t)} > \frac{1}{3}\right\}, & k = 1,2,3. \\ \mathbb{1}\{\sum_{k=1}^3 c_{ik}^t = 0\}, & k = 4. \end{cases} \tag{5.8}$$

The reason that we use a exponential function here is that we need to ensure that the score is at least greater than 0. At time t, we set the nodes to keep active with probability $p_1 = 0.9$, and the inactive nodes to become active with probability $p_2 = 0.1$.

In Figure 5.1, the top left plot shows the change in total active nodes at different time in each simulation. Although the total nodes grows rapidly at first, it never reaches the upper bound within time of interests. The top right plot shows the total nodes from each class. We can see from the plot that the total nodes in class 1-3 grows in a similar trend as the total nodes, but that in class 4 decreases gradually. The bottom plot shows the network density trends. In some of the simulations, it goes to a higher limit than in others.

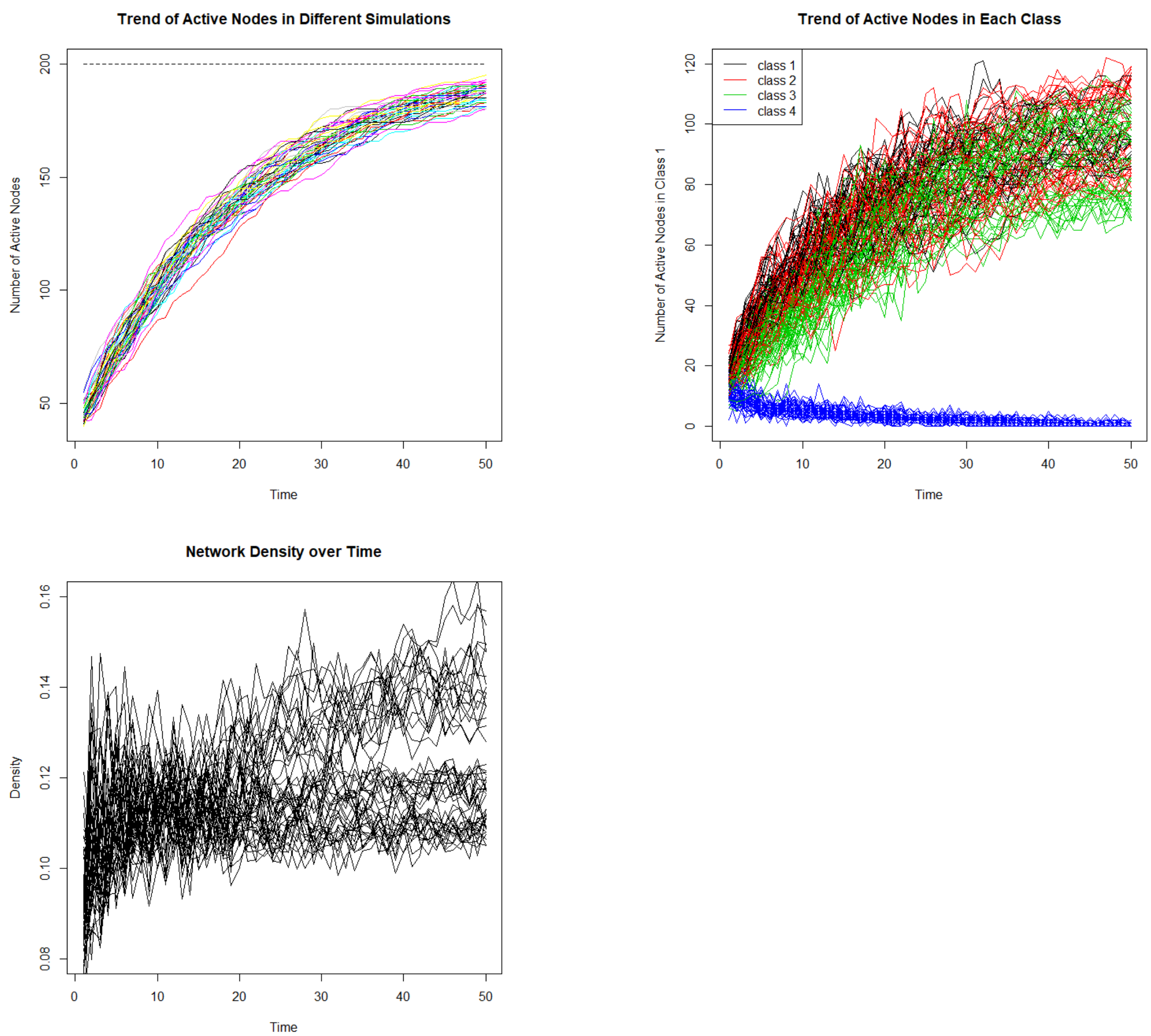


Figure 5.1 MCMDN Simulation: Plots of Total Nodes and Density over Time. In this simulation study, we allow nodes to become active and inactive at each time slots. But the number of active nodes will never exceeds the upper bound of 200.

5.4.2 Simulation Results

The dynamic network is simulated for $t : 0 \rightarrow 100$, and parameters are estimated at time $T = 25, 50, 100$. The network size are restricted by 200 different nodes. The theoretical coverage, variance and MSE are calculated through 500 simulations.The results are shown in Table 5.1. From the variance and MSE, we can see that the point estimator is consistent and gives good estimation with low variance and bias. The coverage rates show that estimated 95% CIs are quite reliable.

Table 5.1 Simulation Results: Multi-Classes Markov Dynamic Network

T	Parms	$\hat{\theta}$	Coverage	Var	MSE
	$\theta_{11} = .35$	0.34998004	0.950	3.565e-05	7.634e-06
	$\theta_{12} = .05$	0.05010068	0.942	1.145e-05	2.589e-06
	$\theta_{22} = .25$	0.24958874	0.946	3.292e-05	5.357e-06
	$\theta_{13} = .05$	0.05027489	0.936	9.254e-06	1.434e-06
25	$\theta_{23} = .05$	0.05011954	0.934	8.540e-06	3.118e-06
	$\theta_{33} = .15$	0.14979175	0.944	1.801e-05	3.295e-06
	$\theta_{14} = .05$	0.05013260	0.960	2.233e-05	6.199e-07
	$\theta_{24} = .05$	0.05064832	0.954	2.297e-05	3.784e-05
	$\theta_{34} = .05$	0.05077824	0.940	2.039e-05	1.712e-09
	$\theta_{44} = .10$	0.10379654	0.952	2.609e-04	7.546e-04
	$\theta_{11} = .35$	0.34997697	0.948	2.708e-05	1.431e-06
	$\theta_{12} = .05$	0.05007239	0.956	9.839e-06	3.336e-09
	$\theta_{22} = .25$	0.24955123	0.936	2.458e-05	9.307e-06
	$\theta_{13} = .05$	0.05022138	0.946	8.017e-06	1.292e-06
50	$\theta_{23} = .05$	0.05006800	0.942	7.209e-06	2.447e-06
	$\theta_{33} = .15$	0.14986834	0.948	1.217e-05	3.141e-06
	$\theta_{14} = .05$	0.05011424	0.946	1.804e-05	1.306e-06
	$\theta_{24} = .05$	0.05039216	0.952	1.784e-05	9.122e-06
	$\theta_{34} = .05$	0.05054726	0.950	1.560e-05	4.332e-07
	$\theta_{44} = .10$	0.10362947	0.950	2.427e-04	8.785e-04
	$\theta_{11} = .35$	0.34977482	0.946	2.618e-05	1.030e-06
	$\theta_{12} = .05$	0.05022217	0.954	1.024e-05	7.771e-08
	$\theta_{22} = .25$	0.24943668	0.934	2.357e-05	4.100e-06
	$\theta_{13} = .05$	0.05019521	0.958	7.769e-06	6.481e-07
100	$\theta_{23} = .05$	0.05008983	0.944	6.965e-06	3.118e-06
	$\theta_{33} = .15$	0.14974733	0.936	1.103e-05	1.926e-06
	$\theta_{14} = .05$	0.05010439	0.938	1.723e-05	1.153e-08
	$\theta_{24} = .05$	0.05038526	0.942	1.686e-05	4.718e-06
	$\theta_{34} = .05$	0.05048703	0.948	1.464e-05	3.125e-06
	$\theta_{44} = .10$	0.10401242	0.948	2.430e-04	8.077e-04

5.5 Real Data: MovieLens Ratings Network

In this section, we introduce the MovieLens 20M data set downloaded on the Grouplens website (http://grouplens.org/datasets/), collected by [HK15]. This data set is famous for its usage on building recommendation system. However, we have found its usage in dynamic network studies after some data manipulation. The original data set contains 27278 movies rated by 138493 users between January 09, 1995 and March 31, 2015 on the MovieLens website (http://movielens.org). In addition, the movies are grouped into different genres.

5.5.1 Data Description

Since the data set is extremely large, we have randomly selected 1000 movies and acquired their related information to make our dynamic network. Those movies without genre listed or no ratings are deleted. First, we transform the rating time variable from UNIX timestamp to the 1st day of each season (01-01, 04-01, 07-01, 10-01) as we are interested in seasonally change of the structure. Then, we treat each movie as a node. An edge is formed if two movies are rated by at least one same user during the time slot. If the movie is not connected to any other movie, it is considered inactive and will be removed from the current network. As a result, we get a seasonally updated dynamic network. The ratings of the movies are averaged over all users during the time interval. The genres are kept unchanged. We also deleted the last time's data since the interval is shorter than one season.

Finally, we get two data sets: 1. The ratings network data set has 896 unique movies. It has three columns: i, j, t , where i, j stands for the movie id's between which there exists an edge, and t stands for the start date of the time interval; 2. The movie features data set has 896 unique movies with columns of ratings and genres. The rating column is the averaged ratings of the movies from different users during the season. The genres contains 19 factorized columns in the following genres:

- *Action, Adventure, Animation, Children, Comedy, Crime, Documentary, Drama, Fantasy, Film Noir, Horror, IMAX, Musical, Mystery, Romance, Science Fiction, Thriller, War, Western*

The MovieLens Ratings Network models the time-varying co-rating structures of online movie rating system. Here co-rating means that people who rates one movie also rates another. People who rates a movie expresses his interests on the movie, no matter positive attitude or not. We are interested in co-rating structure since it shows us some underlying relationships among movies. Although two movies can be co-rated by more than one users, which makes the co-rating edges to be weighted, we are using the unweighted network to fit our model.

5.5.2 Data Visualization

Figure 5.2 shows the seasonally network features over time. From the left plot, we can see that the total active movies of the website grows in linear rate at first. The number reaches its maximum during 2011. From the right plot, we see that the network density drops over time. Like most of the online ratings website, MovieLens grows fast at first. Users were extremely excited at first and were interested in all kinds of movies. As time goes on, people become less active and finally remains at a lower level.

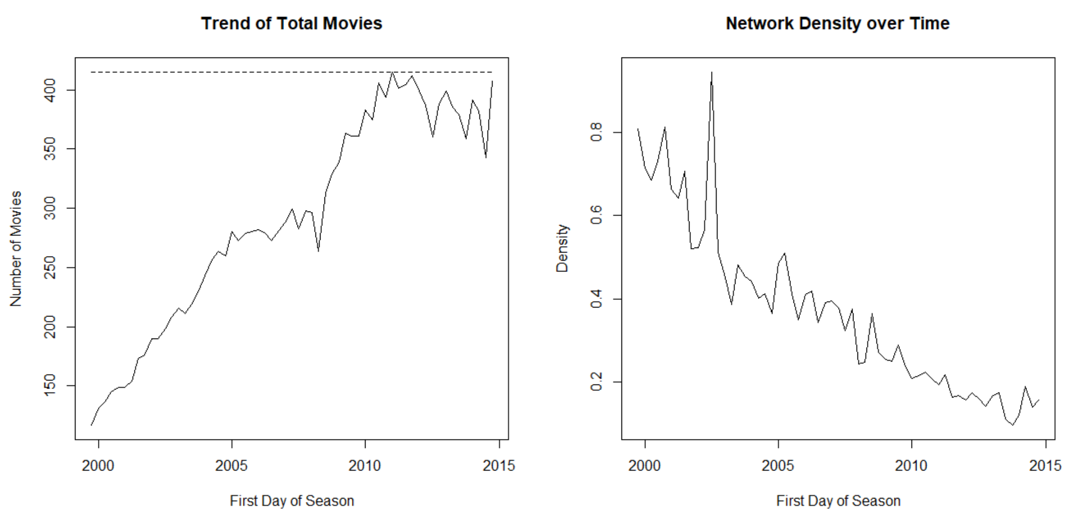
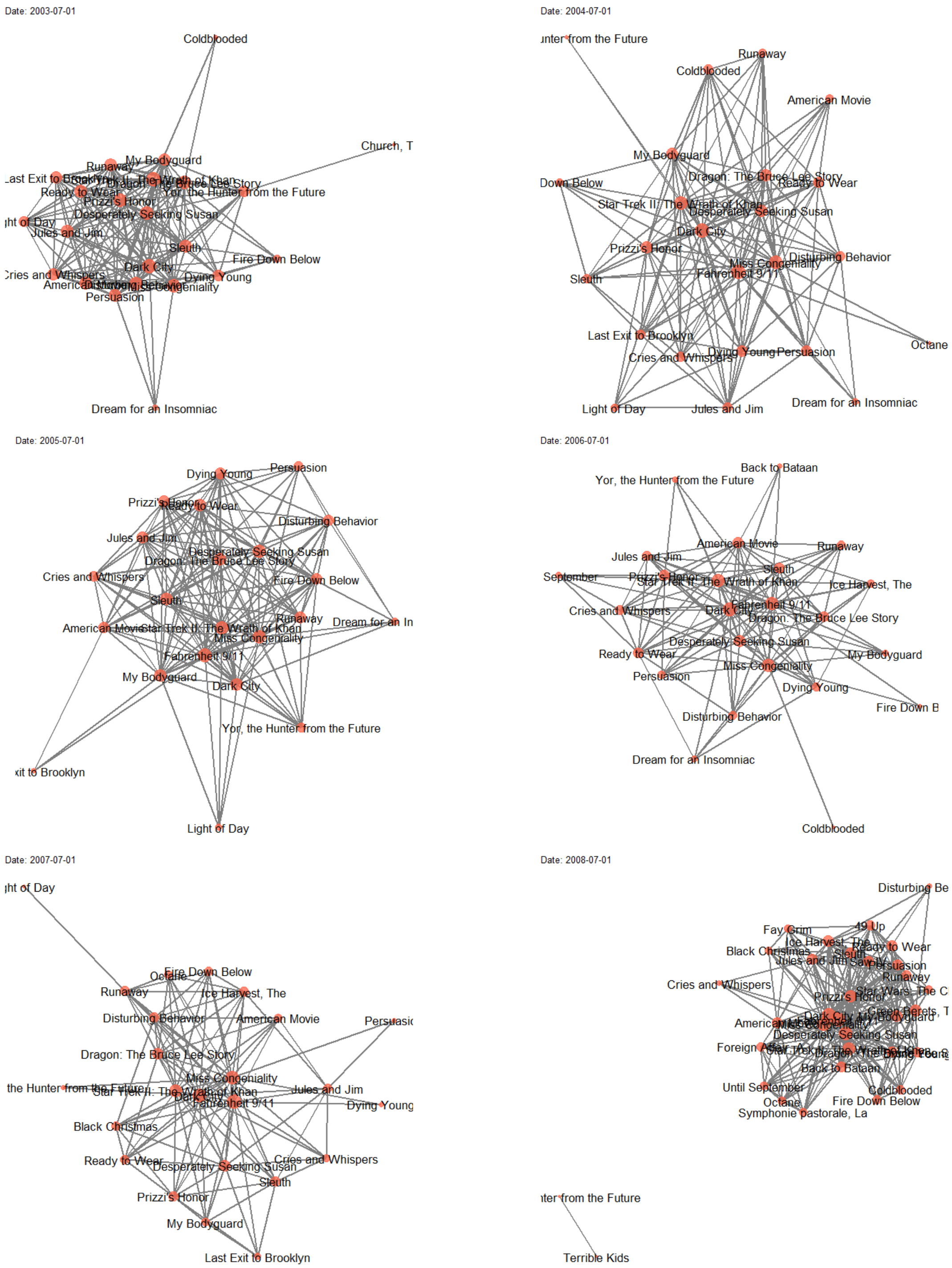


Figure 5.2 Plots of Movie Rating Network Features. The total movies has a growing trend but never exceed 415. The network density is high at first, but decreases gradually to around 0.2 at last.

Figure 5.3 shows the dynamic network formed by randomly selected movies. We can see that some movies have extremely high degrees. Take *Star Trek II: The Wrath of Khan* as an example, it has links to almost all of the other movies at different time. The reason is simple and intuitional, since the series of Star Trek are always popular among different people.

Figure 5.3 The following two pages show the evolution of MovieLens Ratings Network for selected movies at selected times (A through B). We can tell that the structure changes over time. However, some of the movies like *Star Trek II: The Wrath of Khan* always have more connections than other movies.

(A)


Date: 2003-07-01
Coldblooded
Church, T
My Bodyguard
Runaway
Yor, the Hunter from the Future
Ready to Wear
Prizzi's Honor
Desperately Seeking Susan
Jules and Jim
Sleuth
Dark City
Fire Down Below
Dying Young
Persuasion
Dream for an Insomniac
Date: 2004-07-01
Runaway
Coldblooded
American Movie
My Bodyguard
Down Below
Dragon: The Bruce Lee Story
Ready to Wear
Star Trek II: The Wrath of Khan
Desperately Seeking Susan
Dark City
Prizzi's Honor
Disturbing Behavior
Sleuth
Fahrenheit 9/11
Last Exit to Brooklyn
Cries and Whispers
Octane
Light of Day
Jules and Jim
Dream for an Insomniac
Date: 2005-07-01
Dying Young
Persuasion
Disturbing Behavior
Jules and Jim
Desperately Seeking Susan
Dragon: The Bruce Lee Story
Cries and Whispers
Fire Down Below
Sleuth
Runaway
Dream for an In
Miss Congeniality
Fahrenheit 9/11
My Bodyguard
Dark City
Yor, the Hunter from the Future
xit to Brooklyn
Light of Day
Date: 2006-07-01
Back to Bataan
Yor, the Hunter from the Future
Jules and Jim
American Movie
Runaway
September
Sleuth
Ice Harvest, The
Star Trek II: The Wrath of Khan
Fahrenheit 9/11
Cries and Whispers
Dark City
Dragon: The Bruce Lee Story
Desperately Seeking Susan
Ready to Wear
My Bodyguard
Miss Congeniality
Persuasion
Dying Young
Fire Down B
Disturbing Behavior
Dream for an Insomniac
Coldblooded
Date: 2007-07-01
ght of Day
Fire Down Below
Runaway
Ice Harvest, The
Disturbing Behavior
American Movie
Persuasio
Dragon: The Bruce Lee Story
Miss Congeniality
the Hunter from the Future
Jules and Jim
Fahrenheit 9/11
Dying Young
Black Christmas
Ready to Wear
Desperately Seeking Susan
Cries and Whispers
Sleuth
Prizzi's Honor
My Bodyguard
Last Exit to Brooklyn
Date: 2008-07-01
Disturbing Be
Fay Grim
49 Up
Black Christmas
Ready to Wear
Persuasion
Cries and Whispers
Runaway
Desperately Seeking Susan
Back to Bataan
Until September
Coldblooded
Octane
Fire Down Below
Symphonie pastorale, La
nter from the Future
Terrible Kids

(B)


Date: 2009-07-01
Green Berets, The
k to Bataan
Krabat
Church, The
Fay Grim
Black Christmas
Ice Harvest, The
Star Wars: The Clone Wars
Saw IV
Dark City
49 Up
Miss Congeniality
American Movie
Disturbing Behavior
Ready to Wear
Lost Girl
Fire Down Below
Punisher: War Zone
Seven Years Bi
Dying Young
Date: 2010-07-01
Green Berets, The
My Bodyguard
Persuasion
Prizzi's Honor
Fire Down Below
Dying Young
Runaway
Ice Harvest, The
Until September
American Movie
Sleuth
Dragon: The Bruce Lee Story
Star Trek II: The Wrath of Khan
Last Exit to Br
Jules and Jim
Dark City
Punisher: War Zone
Two Cities, A
Ready to Wear
Cries and Whispers
Black Christmas
Ip Man 2
Disturbing Behavior
Krabat
Yor, the Hunter from the Future
Date: 2011-07-01
Madeinusa
Pope's Toilet, The
Fire Down Below
re: The Search for Christopher Robin
Sleuth
Disturbing Behavior
Cries and Whispers
Prizzi's Honor
Fahrenheit 9/11
Light of Day
My Bodyguard
Saw IV
Dark City
Miss Congeniality
Dragon: The Bruce Lee Story
Krabat
Back to Bataan
or an Insomniac
Ip Man 2
Dying Young
Black Christmas
American Movie
House of Stra
Church, The
Date: 2012-07-01
Fay Grim
Last Exit to Brooklyn
Yor, the Hunter from the Future
American Movie
Cries and Wh
Disturbing Behavior
Dark City
Jules and Jim
Dying Young
Star Trek II: The Wrath of Khan
Harvest, The
Saw IV
Star Wars: The Clone Wars
Persuasion
Runaway
Prizzi's Honor
Dragon: The Bruce Lee Story
49 Up
Punisher: War Zone
Ready to Wear
Fire Down Below
Guilty
Date: 2013-07-01
Disturbing Behavior
Coldblooded
American Movie
Apollo 18
Desperately Seeking Susan
Star Trek II: The Wrath of Khan
Persuasion
49 Up
Dark City
Sleuth
Saw IV
Miss Congeniality
Cries and Whispers
Jules and Jim
Ip Man 2
Church, T
her: War Zone
Dragon: The Bruce Lee Story
Prizzi's Honor
My Bodyguard
Date: 2014-07-01
Desperately Seeking Susan
Punisher: War Zone
Until September
Dragon: The Bruce Lee Story
Miss Congeniality
Saw IV
Star Wars: The Clone Wars
Jules and Jim
Dark City
Persuasion
American Movie
If I Stay
Get on Up
Cries and Whispers
49 Up
Apollo 18
Runaway

5.5.3 Determine Class Vectors

In the real data application of the MCMDN model, it is often the case that the classes are unknown. Luckily, in this MovieLens data set, we have movie genres. It is easy to see that the people tend to favor some specific genres of movies, which results in higher connection probabilities among movies that share similar genres. However, the movie genres are not time varying. In another word, we are not able to catch the latent classes of movies which are indicated by user behaviors at each season.

In this example, we compare two different kinds of class vectors while fitting the MCMDN model. In the first method, the class vectors are just the movie genres. However, since there are totally 19 genres, we put together close genres, then use the first 3 frequent classes, and treat the other as category "Other". Thus, the class vectors indicates the following classes:

- *Drama, Comedy/Musical/Romance, Crime/Film noir/Horror/Mystery/Thriller, Other*

In the second method, we use genres, ratings and year of the movie, as well as the adjacency matrix to determine the classes. In detail, the first step is calculating the scores of classes $k = 1, 2, 3$:

$$s_{ik}^t = \begin{cases} s_I + s_{II} + s_{III} + s_{IV}, & \mathcal{N}_{t-1}^i \neq \varnothing \\ 0, & \mathcal{N}_{t-1}^i = \varnothing \end{cases}$$

where

$$s_I = \frac{\sum_{j \in \mathcal{N}_{t-1}^i} \mathbb{1}(c_{jk}^{t-1} = 1)}{\sum_{j \in V_{t-1}} \mathbb{1}(c_{jk}^{t-1} = 1)}, \qquad s_{II} = \frac{\overrightarrow{genre}_i \overrightarrow{genre}_{\mathcal{N}_{t-1}^i}}{|\overrightarrow{genre}_i||\overrightarrow{genre}_{\mathcal{N}_{t-1}^i}|},$$
$$s_{III} = \frac{-\left|rate_i - \sum_{j \in \mathcal{N}_{t-1}^i} rate_j / |\mathcal{N}_{t-1}^i|\right|}{\sqrt{\sum_{j \in \mathcal{N}_{t-1}^i} rate_j / |\mathcal{N}_{t-1}^i|}}, \qquad s_{IV} = \frac{-\left|year_i - \sum_{j \in \mathcal{N}_{t-1}^i} year_j / |\mathcal{N}_{t-1}^i|\right|}{\sqrt{\sum_{j \in \mathcal{N}_{t-1}^i} year_j / |\mathcal{N}_{t-1}^i|}},$$

where s_I is the percentage of movies from class k which are neighbors of movie i. s_{II} is the cosine distance of genre vector of movie i and the mean genre vector of its neighbors. s_{III} stands for the standardized difference between rate of movie i and mean rate of all its neighbors. The last term stands for the standardized difference between year of movie i and the mean year of all its neighbors.

Then, the class vectors are generated using the same technique with form (5.8). Figure 5.4 shows the number of movies in each class/genre during each season.

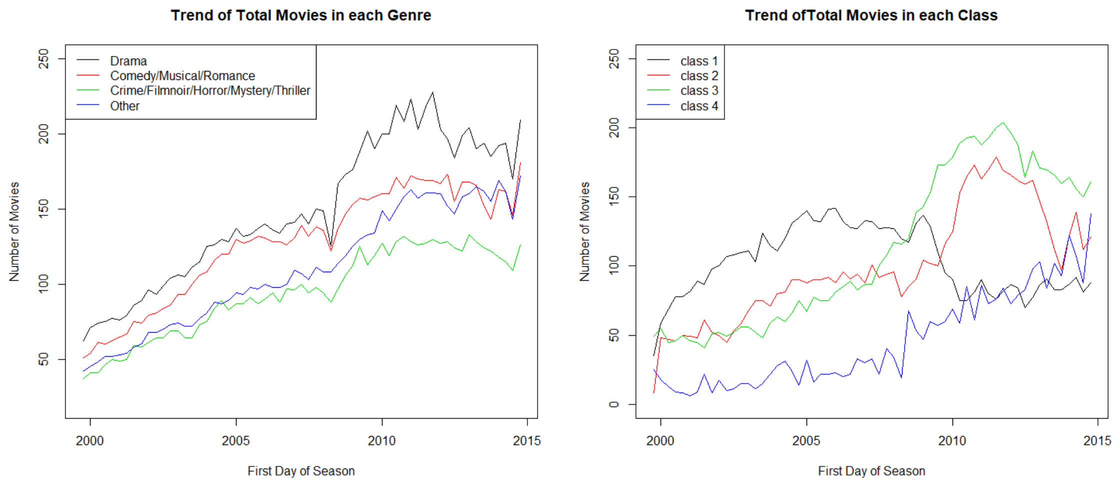


Figure 5.4 Movies Ratings Dynamic Network: Total Movies in Each Class.

5.5.4 Real Data Results

The fitted models are summarized in Table 5.2. Firstly, we can see that the Changing Class method fits the data much better than the Fixed Class method. The log-likelihood is increased by 6%. However, since the classes are generated from both the network connection and the covariates using form (5.8), it is harder to interpret than simply using the genres as classes. Secondly, by looking at the estimated parameters in changing class method, we found that class 1 has the highest within-class connection probability, while class 4 has the lowest. Since class 4 are consists of new movies added to the network, they are less likely to be recommended to the users. It takes some time for users to get familiar with them. In addition, people are more likely to rate movies that are already very popular. Taking the average of the covariates in each class, we can see from Table 5.3 that movies in class 1 has higher scores in adventure/action/IMAX genres. Those movies are more likely to be

Table 5.2 Results: Movies Ratings Dynamic Network

Model	Parms	$\hat{\theta}$	SE	Log Likelihood
	θ_{11}	0.2013274	0.0011028510	
	θ_{12}	0.2279413	0.0008818533	
	θ_{22}	0.3654774	0.0014612087	
	θ_{13}	0.2330079	0.0011452180	
Fixed Class	θ_{23}	0.3511075	0.0013196900	-1642947
	θ_{33}	0.4875066	0.0024304741	
	θ_{14}	0.1922651	0.0010436291	
	θ_{24}	0.2992210	0.0012483195	
	θ_{34}	0.3570110	0.0016442452	
	θ_{44}	0.3250536	0.0021354787	
	θ_{11}	0.4637852	0.0011212642	
	θ_{12}	0.3617330	0.0008775962	
	θ_{22}	0.4079144	0.0013874415	
	θ_{13}	0.3226217	0.0008185696	
Changing Class	θ_{23}	0.3188071	0.0008624000	-1543350
	θ_{33}	0.3220637	0.0011041584	
	θ_{14}	0.1278587	0.0007390964	
	θ_{24}	0.1002277	0.0006715215	
	θ_{34}	0.0979783	0.0005908126	
	θ_{44}	0.0467386	0.0006734104	

co-rated from the estimated class connection matrix, which means people who watched these kind of movies tend to watch similar movies. However, movies in class 4 has higher scores in horror and documentary than others. But this class has lower within-class connection probability than between-classes probabilities.

It should be mentioned here that both model use 4 total classes with 10 class connection probability parameters. This is relatively a simple model that might not catch up the entire information from the network as well as the covariates. More complex models with larger number of classes can have better results.

Table 5.3 Movies Ratings Dynamic Network: Genres, Year and Rates in each class

Class			Covariates			
1	Action	0.2585681	Drama	0.2190107	Romance	0.0963007
	Adventure	0.1278238	Fantasy	0.0603264	Science Fiction	0.1243441
	Animation	0.0269502	Film noir	0.0050504	Thriller	0.2523842
	Children	0.0690941	Horror	0.2422381	War	0.0226874
	Comedy	0.4273991	IMAX	0.0111681	Western	0.0085398
	Crime	0.1229582	Musical	0.0172287	Year	1995.160
	Documentary	0.0337057	Mystery	0.0335337	Rate	2.811219
2	Action	0.0903604	Drama	0.6681967	Romance	0.2366316
	Adventure	0.0769944	Fantasy	0.0443744	Science Fiction	0.0396752
	Animation	0.0291578	Film noir	0.0118860	Thriller	0.1129373
	Children	0.0566877	Horror	0.0486511	War	0.0442330
	Comedy	0.3279448	IMAX	0.0006153	Western	0.0129531
	Crime	0.2013274	Musical	0.0363817	Year	1985.549
	Documentary	0.0375970	Mystery	0.0423103	Rate	3.451727
3	Action	0.0926425	Drama	0.7400653	Romance	0.2299632
	Adventure	0.0908614	Fantasy	0.0442754	Science Fiction	0.0397762
	Animation	0.0281847	Film noir	0.0119927	Thriller	0.1161126
	Children	0.0499735	Horror	0.0476564	War	0.0515547
	Comedy	0.2424609	IMAX	0.0001763	Western	0.0187011
	Crime	0.1294356	Musical	0.0409741	Year	1982.038
	Documentary	0.0354255	Mystery	0.0467314	Rate	3.529264
4	Action	0.0996266	Drama	0.5456714	Romance	0.1288142
	Adventure	0.0494785	Fantasy	0.0391959	Science Fiction	0.0439671
	Animation	0.0208302	Film noir	0.0148189	Thriller	0.1172836
	Children	0.0260664	Horror	0.1113273	War	0.0422997
	Comedy	0.3011945	IMAX	0.0068297	Western	0.0307562
	Crime	0.0999237	Musical	0.0137071	Year	1985.929
	Documentary	0.0524026	Mystery	0.0229598	Rate	3.052763

CHAPTER

6

CONCLUSIONS AND FUTURE DIRECTIONS

In this paper, we have introduced mainly three different dynamic network models: 1. The base model: Markov Dynamic Network; 2. The growing size model: Fixed Rate model and Random Process model; 3. The Multi-classes Markov Dynamic Network model. In all three models, we have discussed on long time behaviors of the networks. In addition, we have given the model estimation methods with the analysis for asymptotic distributions of the estimators.

In conclusion, the three Markov Dynamic Network models have characterized different structure of dynamic networks. Specifically, they are able to catch the short time dependency of the dynamic network. The Growing Size models and Multi-Classes models allow much more flexibility in modeling the network sizes, which have higher applicability in real world data. In addition, the Multi-Classes model further develops the traditional Stochastic Block model such that nodes are characterized by their class vectors, which are also changeable in time.

The simulation studies for all three models show that the theoretical formula given by the MLE are consistent and asymptotically normal, which allows us to calculate the estimators and their confidence intervals simpler and faster without using any resampling techniques.

However, there are still some limitations in our work. Therefore, in future works, we can focus on the following aspects:

Considering directed network. All of the models we have discussed are focused on undirected networks. However, in real applications, e.g. the MathOverflow network, the connections among nodes are actually directed. Modeling directed network is actually quite similar, where the adjacency matrix is no longer symmetric. The probabilities should change to in-probabilities and out-probabilities between each nodes pair. For the Multi-classes Markov Dynamic Network model, the class connection probability matrix should also be asymmetric.

Nodes coming processes. In the growing size model, we made assumptions on constant coming nodes or Poisson coming nodes. In future works, more efforts can be payed on modeling nodes coming processes. For example, in social dynamic networks, at the beginning, the network tends to grow especially. Then, the growing rate decreases and becomes constant. At last, the network size becomes stable. Such makes the network size an "S" shaped growth over time.

Class estimation. In the Multi-classes model, one of the challenging work is to identify the classes of each nodes. In real life, the classes might be already known in some cases. But more likely the case is that we only have nodes features and need to estimate classes. A good estimation of the total number of classes and the class vector of each node are both important and make a lot of sense to the model behavior. In this paper, we didn't compare difference choices of these hyper-parameters and methods on estimating classes. However, future works may focus on how to choose between different numbers of classes, as well as comparing different models for classes estimations. In addition, interpreting the estimated classes are also important in real world. More study can be done in exploring how to better interpret the model.

Threshold for defining edges in MovieLens Ratings Network. In the analysis of the MovieLens data set, we define an edge between two movies as when there exists at least one users who rated both of them during the same time period. However, such definition results in an much denser network as more and more people join the website. We may consider a dynamic threshold for an edge between two movies as certain proportion of total users on the website rated the two movies during the same time period.

BIBLIOGRAPHY

[AL06] Athreya, K. B. & Lahiri, S. N. *Measure theory and probability theory*. Springer Science & Business Media, 2006.

[ER59] Erdös, P & Rényi, A. "On random graphs". *Publ. Math. Debrecen* **6**.290 (1959), pp. 290–297.

[Han15] Han, Q. et al. "Consistent estimation of dynamic and multi-layer block models". *Proceedings of the 32nd International Conference on Machine Learning (ICML-15)*. 2015, pp. 1511–1520.

[HK15] Harper, F. M. & Konstan, J. A. "The MovieLens Datasets: History and Context". *ACM Trans. Interact. Intell. Syst.* **5**.4 (2015), 19:1–19:19.

[Hof02] Hoff, P. D. et al. "Latent space approaches to social network analysis". *Journal of the American Statistical Association* **97**.460 (2002), pp. 1090–1098.

[Hol83] Holland, P. W. et al. "Stochastic blockmodels: First steps". *Social networks* **5**.2 (1983), pp. 109–137.

[KC14] Kolaczyk, E. D. & Csárdi, G. *Statistical analysis of network data with R*. Vol. 65. Springer, 2014.

[LK14] Leskovec, J. & Krevl, A. *SNAP Datasets: Stanford Large Network Dataset Collection*. `http://snap.stanford.edu/data`. 2014.

[MM17] Matias, C. & Miele, V. "Statistical clustering of temporal networks through a dynamic stochastic block model". *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **79**.4 (2017), pp. 1119–1141.

[Par17] Paranjape, A. et al. "Motifs in temporal networks". *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*. ACM. 2017, pp. 601–610.

[Roh11] Rohe, K. et al. "Spectral clustering and the high-dimensional stochastic blockmodel". *The Annals of Statistics* **39**.4 (2011), pp. 1878–1915.

[WO13] Wolfe, P. J. & Olhede, S. C. "Nonparametric graphon estimation". *arXiv preprint arXiv : 1309.5936* (2013).

[Zha17a] Zhang, X. et al. "Random graph models for dynamic networks". *The European Physical Journal B* **90**.10 (2017), p. 200.

[Zha17b] Zhang, Y. et al. "Estimating network edge probabilities by neighbourhood smoothing". *Biometrika* **104**.4 (2017), pp. 771–783.

APPENDICES

APPENDIX

A

SUPPLEMENTARY MATERIALS FOR CHAPTER 4

Bootstrap is usually a first choice of estimating the model parameters while in real data applications. However, Bootstrapping of the network models can be quite different than Bootstrapping traditional data. When it comes to dynamic networks, another problem appears since there exists time dependence. We applied two difference Bootstrap methods in our simulations studies which will be introduced here.

A.1 Parametric Bootstrap of the Dynamic Networks

The Parametric Bootstrap is done in the following 4 steps:

- Step 1: At $t = 1$, Bootstrap the nodes pairs. Record the total edges.
- Step 2: At $t = 2, 3, ...,$ use the estimated parameters to regenerate the network status.
- Step 3: Calculate the Bootstrap estimation $\hat{p}_b$.
- Step 4: For $b = 1, 2, ..., 200$, repeat the above steps.

A.2 All-time Bootstrap of the Dynamic Networks

The algorithm of All-time Bootstrap is summarized as the following 4 steps:

- Step 1: At $t = 1$, Bootstrap the nodes pairs. Record the total edges. Let

$$Numerator = 0,$$

$$Denominator = \sum_{i,j \in B_1} X_{i,j}^1,$$

  where B_t stands for the Bootstrap sample at time t.

- Step 2: At $t = 2, 3, ...,$ record the total remaining edges which are Bootstrapped at $t - 1$. Update

$$Numerator = Numerator + \sum_{i,j \in B_{t-1}} X_{i,j}^t X_{i,j}^{t-1}.$$

  Then, Bootstrap the nodes pairs at time t, record the total edges and update

$$Denominator = Denominator + \sum_{i,j \in B_t} X_{i,j}^t.$$

- Step 3: Calculate the Bootstrap estimation

$$\hat{p}_b = Numerator/Denominator.$$

- Step 4: For $b = 1, 2, ..., 200$, repeat the above steps.

APPENDIX

B

SUPPLEMENTARY MATERIALS FOR CHAPTER 5

B.1 Algorithm: Numerical Solution of MLE

The mathematical solution of log-likelihood function is mostly impossible as $p^t_{ij}(\boldsymbol{\theta})$ is usually complex function of $\boldsymbol{\theta}$. Here we provide an algorithm for a numerical solution in Table B.1.

Under most situations, the conditional MLE of one element of $\boldsymbol{\theta}$ is much easier to get when we know the rest elements, since p^t_{ij} is usually defined as linear function of $\boldsymbol{\theta}$ as follows:

$$p^t_{ij} = \sum_{i,j} h(c^t_i, c^t_j)\theta_{ij}.$$

Table B.1 Algorithm for Solving MLE of MCMDN

Algorithm: MLE for MCMDN
Initialize and vectorize θ as $\theta^{(0)} = (\theta_{11}^{(0)}, \theta_{21}^{(0)}, \theta_{22}^{(0)}, ..., \theta_{KK}^{(0)})^T$;
While $\|\boldsymbol{\theta}^{(m+1)} - \boldsymbol{\theta}^{(m)}\| > \epsilon$, do:
Update $\theta_{11}^{(m+1)} = \theta_{11}^{MLE}(\theta_{21}^{(m)}, \theta_{31}^{(m)}, ..., \theta_{KK}^{(m)})^T)$;
Update $\theta_{21}^{(m+1)} = \theta_{21}^{MLE}(\theta_{11}^{(m)}, \theta_{22}^{(m)}, ..., \theta_{KK}^{(m)})^T)$;
$\cdots$
Update $\theta_{KK}^{(m+1)} = \theta_{KK}^{MLE}(\theta_{21}^{(m)}, \theta_{22}^{(m)}, ..., \theta_{K(K-1)}^{(m)})^T)$;
Return final $\boldsymbol{\theta}$.

Then the log-likelihood becomes:

$$\begin{aligned} l(\boldsymbol{\theta}) &= \sum_{t=1}^{T} \Big\{ \sum_{i<j} \big[x_{ij}^t \log p_{ij}^t + (1 - x_{ij}^t) \log(1 - p_{ij}^t) \big] \Big\} \\ &= \sum_{t=1}^{T} \Big\{ \sum_{i<j} \big[x_{ij}^t \log p_{ij}^t(\boldsymbol{\theta}) + (1 - x_{ij}^t) \log(1 - p_{ij}^t(\boldsymbol{\theta})) \big] \Big\} \\ &= \sum_{t=1}^{T} \Big\{ \sum_{i<j} \Big[x_{ij}^t \log\Big(\sum_{i,j} h(c_i^t, c_j^t)\theta_{ij} \Big) + (1 - x_{ij}^t) \log\Big(1 - \sum_{i,j} h(c_i^t, c_j^t)\theta_{ij} \Big) \Big] \Big\}. \end{aligned}$$

Now based on the m-th iteration result, take the first derivative on θ_{11}

$$\begin{aligned} d_{11} &= \frac{\partial l(\boldsymbol{\theta})}{\theta_{11}} \bigg|_{\boldsymbol{\theta}^{(m)}} = \sum_{t=1}^{T} \Bigg\{ \sum_{i<j} \Bigg[x_{ij}^t \Big(\frac{h_{11}^t}{\sum_{i,j} h_{ij}^t \theta_{ij}^{(m)}} \Big) - (1 - x_{ij}^t) \Big(\frac{h_{11}^t}{1 - \sum_{i,j} h_{ij}^t \theta_{ij}^{(m)}} \Big) \Bigg] \Bigg\}, \\ d_{11}^2 &= \frac{\partial^2 l(\boldsymbol{\theta})}{\theta_{11}} \bigg|_{\boldsymbol{\theta}^{(m)}} = -\sum_{t=1}^{T} \Bigg\{ \sum_{i<j} \Bigg[x_{ij}^t \Big(\frac{h_{11}^t}{(\sum_{i,j} h_{ij}^t \theta_{ij}^{(m)})^2} \Big) - (1 - x_{ij}^t) \Big(\frac{h_{11}^t}{(1 - \sum_{i,j} h_{ij}^t \theta_{ij}^{(m)})^2} \Big) \Bigg] \Bigg\}, \end{aligned}$$

we can get the $(m+1)$-th iteration of MLE for θ_{11} as

$$\theta_{11}^{(m+1)} = \theta_{11}^{(m)} + \alpha \Delta^{(m+1)},$$

where $\Delta^{(m+1)} = d_{11} d_{11}^2$, and α is the step size.