Study of Dynamic Networks

Rohan Shinde (MD2219)

M.Stat Second Year Project

Supervised by Dr. Arindam Chatterjee, ISI Delhi

Abstract

This report explores models and methods for analyzing dynamic networks, with a focus on the dynamic stochastic block model (Dynamic SBM). Theoretical results are derived for the edge density and its conditional evolution over time under a particular Dynamic SBM. Simulations are conducted to verify the theoretical findings and investigate the distribution of edge densities at large time points. Potential extensions to other network statistics and distributional properties are discussed, highlighting avenues for further research in characterizing the dynamics of evolving network structures.

1 Introduction

Networks are a mathematical representation of pairwise relationships or interactions among a set of objects, wherein the objects are depicted as vertices (nodes) and edges (links) connect vertices that share specific relationships. Examples of networks pervade various domains, such as the Internet, social networks, and biological systems. Networks arise in natural, social, and information sciences and in many other disciplines. The study of network data in Statistics has been a leading area of research for the last decade and more (cf. Kolaczyk (2009)). Networks can be viewed from both a static and as well as a dynamic viewpoint. In a static approach, a single snapshot of a network is the object of interest, and in a dynamic approach, the changes in a network over time are taken into consideration. The term 'dynamic' is commonly employed in the context of networks to describe systems where edges among vertices, and sometimes the vertices themselves, change over time. While static methodologies emerged in the 1960s, particularly in sociology, the literature on dynamic models is a more recent phenomenon. Comprehensive references on statistical modelling of random graphs include Kolaczyk (2009), Goldenberg et al. (2010), and Snijders (2011).

A dynamic network can be conceptualized as a time-indexed graph $G(t) = (V(t), E(t))$, where time t varies discretely or continuously, and $V(t)$ represents the set of vertices at time t, while $E(t)$ denotes the presence or absence of edges between vertices (Kolaczyk (2009)) e.g. see Fig. (1) which show real-world tortoise interaction network data set.

Consider a friendship network among school students in a class, with individuals as nodes, and connections as edges. It would be more intuitive to look at the behaviour of this network over a period of time, and observe how the friendship structure changes over time, as possibly some students join or leave the class. The change in the network structure provides a lot of information about

the edge-evolution processes, changes in communities/clusters in the network (if any), changes in the sparsity of the network, etc. Examining dynamic networks facilitates comprehension of latent graph evolution in specific problems, addressing realistic questions. However, this advantage is counterbalanced by the need for more complex modelling compared to static networks.

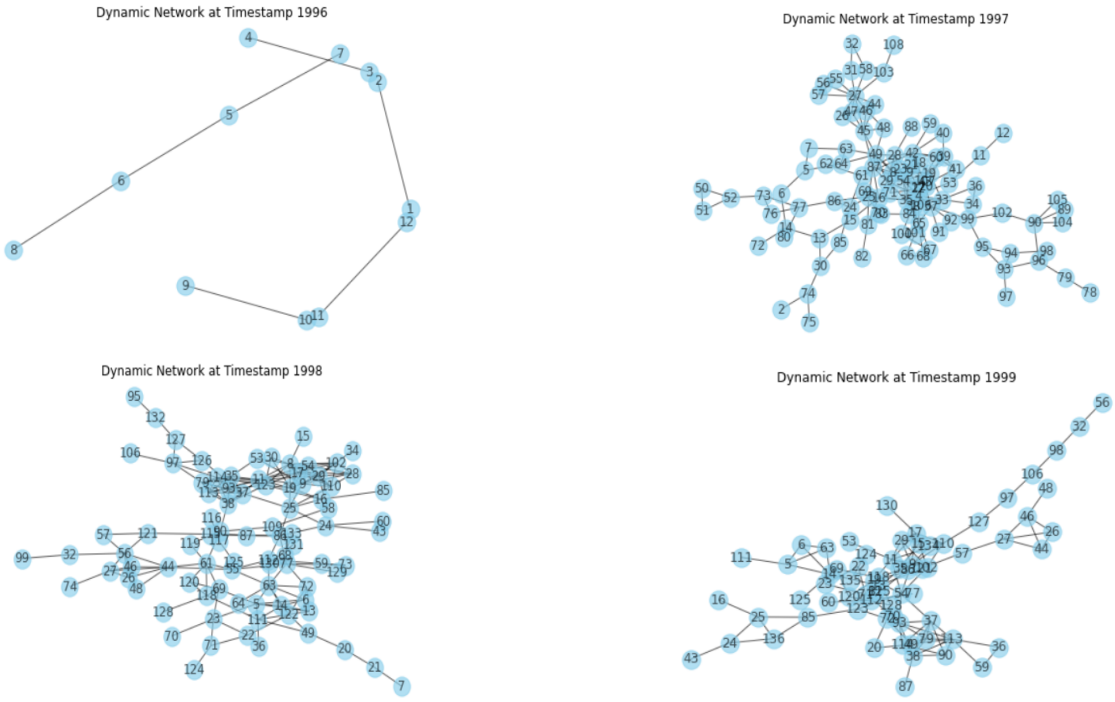

Figure 1: Networks of burrow use to infer social associations in desert tortoise (Rossi and Ahmed (2015))

In this context, this report will explore common approaches for modelling dynamic networks. The extensive literature in this domain offers diverse techniques for capturing dynamic aspects. The project delves into selected modeling methods, aiming to comprehend their dynamic characteristics. In Section 2.1, the Dynamic Erdos Renyi model is introduced. Section 2.2 explores community detection and statistical inferences using Dynamic Stochastic Block Models. Section 2.3 presents a time series modeling approach, while Section 2.4 summarizes models for graph structure estimation using Markov Random Fields. In section 3 we introduce descriptive summary statistics for networks alongwith a few examples. Focusing on a particular summary statistic, section 4 presents edge density of dynamic networks. Further on, section 4.1 presents few theoretical results of the edge density based on the model introduced in 2.2. Finally, section 5 conducts few simualtions to verify the theoretical results and pave a path for further exploration in the topic.

2 Dynamic Network Models

Various techniques are available for incorporating the time-varying information into dynamic networks, and the selection of a method is influenced by factors such as data interpretation, accuracy, the nature of the time-varying process of the network (discrete or continuous), and more. The chosen method reflects the researcher's perception of their system, an aspect that is often insufficiently justified in their research works even though it is deeply associated with underlying assumptions of the data. In particular, a researcher may majorly focus on the *dynamic* (the "time-varying" aspect) or the *networks* aspect in a model for dynamic networks. Such aspects of dynamic networks and the assumptions on the form of data, the time-varying process and its evolutionary behaviour are crucial to making statistical inferences about dynamic networks. In this section, we will explore some common approaches for modelling dynamic networks.

2.1 Dynamic Erdos-Renyi Random Graph Model

The simplest model in the literature for networks is the Erdos-Renyi Random Graph Model (Erdős and Rényi (1959)). In the Erdos-Renyi random graph $G(N, p)$ model, a graph is constructed by connecting N labelled nodes randomly. Each edge is included in the graph with probability $p \in (0, 1)$ independently from every other edge.

A very natural extension to the time-varying networks would be to let each edge of the network evolve in a Markovian manner. The actual formalisation of this idea is proposed in Zhang et al. (2017) wherein, each of the edges undergoes independent evolution. Specifically, an edge exists for an exponential time with a parameter μ (referred to as the 'up-rate') and disappears for an exponential time with a parameter λ (referred to as the 'down-rate'). This model particularly focuses on the estimation of rate at which edges appear and disappear in the network. If we denote by $p_1(t)$ and $p_0(t)$ respectively the probabilities that there is and is not an edge between our nodes at time t, then Zhang et al. (2017) prove via solving a differential equation that p_1 can be given by

$$p_1(t) = \frac{\lambda}{\lambda + \mu} - c\, \mathrm{e}^{-(\mu+\lambda)t}$$

where c is an integration constant, λ is the rate (in continuous time) at which an edge appears between two nodes where previously there was none, and μ is the rate at which an existing edge disappears.

To address the limitation of independent edge evolution, modified Dynamic Erdős-Rényi graphs were introduced in Mandjes et al. (2019) where the authors consider two variations in which the edges evolve dependently. A basic underlying assumption here is that the number of nodes remains fixed over time and the evolution in time is characterized in terms of the changes in edge-structure. The authors manage to uniquely characterize their transient and stationary behaviour using the probability generating function (PGF). In the continuous time model, the authors prove a functional

central limit theorem for $Y(t)$ which denotes the number of edges in the network at time t using a process they call "regime-switching". The weak limit of the appropriately scaled $Y(t)$ is given as a solution to a stochastic differential equation. The authors also prove the central limit theorem for their discrete-time model by embedding it into a specific explicit continuous-time model. The authors also discuss several diffusion results under scaling and large deviation results for $Y(t)$ for both of their models.

2.2 Dynamic stochastic block models

The homogeneity of Erdos-Renyi is a drawback when modelling real-world networks. Node-specific edge probabilities can be used to take into account the natural heterogeneity of the data and summarise the data through node classification. Such structure allows to study the underlying connection structure of the network. Proposed by Frank and Harary (1982), another popular modelling approach in the case of static networks, particularly very useful for community detection and clustering, is the Stochastic Block Model (SBM). The major idea behind the SBM is that the probability of the existence of an edge between any two nodes of the network (with a fixed set of nodes) depends only on the communities in which these nodes belong.

An extension to the static SBM using Markovian evolution of the node groups was proposed in Matias and Miele (2017). Unlike earlier works that either combined SBMs with Markov structures (e.g., Yang et al. (2011)) or utilized state space models (e.g., Xu and Hero (2014)), this model uniquely separates these elements and revolutionizes node classification in evolving networks by incorporating a model selection criterion for determining the optimal number of clusters. In this model, the authors consider a set of fixed N vertices/individuals divided into Q latent groups which may vary over time. This community membership is assumed to be a stationary, irreducible and aperiodic Markov Chain which further determines the weights of each edge in the network depending on the communities to which the vertices forming the edge are in. A brief mathematical formulation for the model is given below,

In Matias and Miele (2017), the data matrices $Y = (Y^t)_{1\leq t\leq T}$ represent weighted interactions among N individuals observed over time, where T is the number of time points. Random variables $\mathbf{Z} = (Z_i^t)_{1\leq t\leq T, 1\leq i\leq N}$ with values in $\mathcal{Q}^{NT} := \{1, ..., Q\}^{NT}$ encode the community to which each of the N nodes belong to at each time instant. Each Z_i is an i.i.d. random variable, and $Z_i = (Z_i^t)_{1\leq t\leq T}$ is an irreducible, aperiodic stationary Markov chain with transition matrix $\boldsymbol{\pi} = (\pi_{qq'})_{1\leq q,q'\leq Q}$ and initial stationary distribution $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_Q)$.Given latent groups $\mathbf{Z}$, $\mathbf{Y} = (Y^t)_{1\leq t\leq T}$ are independent, and the conditional distribution of each Y^t depends only on Z^t. Thus for each fixed t, the random graph Y^t follows a Stochastic Block Model (SBM), where $(Y_{ij}^t)_{1\leq i<j\leq N}$ are independent given Z^t and the distribution of each Y_{ij}^t depends on only Z_i^t and Z_j^t. The model is then given by

the distribution:

$$Y_{ij}^t|\{Z_{iq}^t Z_{jl}^t = 1\} \sim (1-\beta_{ql}^t)\delta_0(\cdot) + \beta_{ql}^t F(\cdot, \gamma_{ql}^t) \tag{1}$$

where $\{F(\cdot,\gamma), \gamma \in \Gamma\}$ is a parametric family of distributions, and $f(\cdot,\gamma)$ represents their densities (Dirac mass at 0 (δ_0) is introduced to accommodate sparse weighted graphs e.g. see Ambroise and Matias (2012)). The model is parameterized by

$$\theta = (\boldsymbol{\pi}, \boldsymbol{\beta}, \boldsymbol{\gamma}) = (\boldsymbol{\pi}, \{\beta^t, \gamma^t\}_{1\le t\le T}) = (\{\pi_{qq'}\}_{1\le q,q'\le Q}, \{\beta_{ql}^t, \gamma_{ql}^t\}_{1\le t\le T, 1\le q\le l\le Q}) \in \Theta$$

and the probability distribution is denoted as $\mathbb{P}_\theta$ on the space $\mathcal{Q}^{\mathbb{N}} \times \mathbb{R}^{\mathbb{N}}$. The authors provide weak and strong identifiability conditions for the model parameters to tackle non-identifiability in the model and label switching issues.

Since the conditional distribution $\mathbf{Z}|\mathbf{Y}$ does not have a factored form, the Variational Expectation Maximation (VEM) algorithm (see e.g. Jordan et al. (1999), Gunawardana and Byrne (2005), Celisse et al. (2012))is then used for estimation of model parameters. The model selection on number of groups Q is done by maximising the integrated classification likelihood (ICL) criterion given by

$$\mathrm{ICL}(Q) = \log\{\mathbb{P}_{\hat{\theta}_Q}(\mathbf{Y}, \hat{\mathbf{Z}})\} - \frac{1}{2}Q(Q-1)\log\{N(T-1)\} - \mathrm{pen}(N, T, \boldsymbol{\beta}, \boldsymbol{\gamma})$$

where for any number of groups $Q \ge 1$, $\hat{\theta}_Q$ is the estimated parameter value with Q groups, $\hat{\mathbf{Z}}$ the corresponding maximum a posteriori classification at $\hat{\theta}_Q$, the first penalization term accounts for transition matrix $\boldsymbol{\pi}$ and $\mathrm{pen}(N, T, \boldsymbol{\beta}, \boldsymbol{\gamma})$ is a penalizing term for the connectivity parameters $(\boldsymbol{\beta}, \boldsymbol{\gamma})$. The initial number of groups for the VEM procedure are chosen via k-means clustering applied to the rows of a concatenated data matrix comprising all adjacency time step matrices Y^t stacked in consecutive column blocks. This initialization strategy excels when group memberships exhibit limited variation across time, and for larger values of T, the model initialization described in the paper may yield inaccurate inferences. Matias and Miele (2017) applied their dynamic Stochastic Block Model (SBM) to analyze various dynamic contact networks. They studied face-to-face interactions among high school students (Fournet and Barrat (2014); $T = 4$), sparrow interactions (Shizuka et al. (2014); $T = 3$), and Onager associations (Rubenstein et al. (2015); $T = 4$). The model effectively captured network dynamics in diverse datasets, demonstrating its potential for broader applications in dynamic network analysis.

Prior dynamic Stochastic Block Models (SBM) like those in Matias and Miele (2017), Fu et al. (2009), Yang et al. (2011), and Xu and Hero (2014) assume discrete-time dynamics for both community membership and network evolution, with independent SBMs at each time step. Ludkin et al. (2018) proposed the autoregressive SBM (ARSBM), a continuous-time extension that relaxes this independence assumption and allows modelling networks with unequal observation intervals. This enables the handling of irregularly observed or incomplete data more effectively compared to discrete-time models.

In the context of ARSBM, the node set V with $|V| = N$ is partitioned into a fixed number of communities Q. A continuous-time Markov chain (CTMC) $C_i(\cdot)$, taking values in $1, 2, \ldots, Q$ is defined as: $C_i(t) = k$ if individual i belongs to community k at time t. It is assumed that irrespective of the current community, a node spends exponential Exp(λ) time in this community before transitioning uniformly at random to a new community (this assumption is subject to relaxation). Additionally, all edges connecting nodes across different communities share similar dynamics (this assumption can also be relaxed). The network dynamics involve $Q+1$ processes: one for each community k, governing edges (i, j) where $C_i(t) = C_j(t) = k$, and one for edges between communities, governing edges (i, j) where $C_i(t) \neq C_j(t)$. The community membership process $C_{ij}(\cdot)$ is defined as $C_{ij}(t) = C_i(t)$ if $C_i(t) = C_j(t)$ and 0 otherwise. The edge status process $E_{ij}(\cdot)$ is defined as $E_{ij}(t) = 1$ if an edge exists between nodes i and j at time t, and 0 otherwise. The edge process follows a piecewise time-homogeneous CTMC, with the generator matrix given by $\begin{pmatrix} -\alpha_k & \alpha_k \\ \delta_k & -\delta_k \end{pmatrix}$, where α_k (appearance rates) governs edge appearance in community k, and δ_k (deletion rates) governs the transition from state 1 to 0. Observations are collected as network snapshots of N nodes at time points $\mathbf{t} = (t_0, t_1, \ldots, t_T)$ in the interval $[t_0, t_T]$. The states $E_{ij}(t_s)$ are observed for $s = 0, 1, \ldots, T$ and $i \neq j; i, j \in 1, \ldots, N$. For brevity, let $e_{ij}^s = E_{ij}(t_s)$, $c_i^s = C_i(t_s)$ (latent variable), and $\Delta s = t_s - t_{s-1}$. Define $\mathbf{e}(\mathbf{t}) = e_{ij}^s | 1 \leq i < j \leq N, s = 0, 1, \ldots, T$ as the set of all network snapshot data. Similarly, let $\mathbf{c}_i(\mathbf{t}) = c_i^s | s = 0, 1, \ldots, T$ represent the community membership of node i at each observation time, with $\mathbf{c}(\mathbf{t}) = \mathbf{c}_i(\mathbf{t}) | i = 1, 2, \ldots, N$ as the set of all community memberships. Define $\pi_k = \alpha_k/(\alpha_k + \delta_k)$, $\rho_k = \alpha_k + \delta_k$, $\boldsymbol{\pi} = (\pi_0, \pi_1, \ldots, \pi_K)$, and $\boldsymbol{\rho} = (\rho_0, \rho_1, \ldots, \rho_K)$. The focus of the model is on the joint posterior distribution $\pi(\boldsymbol{\theta}, \mathbf{c}(\mathbf{t}) | \mathbf{e}(\mathbf{t}))$, where $\boldsymbol{\theta} = (\lambda, \boldsymbol{\pi}, \boldsymbol{\rho})$ and $\mathbf{c}(t_0) = (c_1^0, c_2^0, \ldots, c_N^0)$.

The community membership process is modelled as piecewise constant, requiring knowledge of changepoints $\boldsymbol{\tau}_i = (\tau_i^1, \cdots, \tau_i^{M_i})$ and community memberships at these points, $\mathbf{c}_i(\boldsymbol{\tau}_i)$, with $c(\boldsymbol{\tau}) = (\mathbf{c}_1(\boldsymbol{\tau}_1), \cdots, \mathbf{c}_N(\boldsymbol{\tau}_N))$. Since the number of changepoints is assumed to be unknown, employing Reversible Jump MCMC (RJMCMC) with data augmentation, samples are drawn from the joint distribution of $\boldsymbol{\theta} = (\lambda, \boldsymbol{\pi}, \boldsymbol{\rho})$ and $\mathbf{c}(\mathbf{t})$ given $\mathbf{e}(\mathbf{t})$ and augmented data $(\boldsymbol{\tau}, \mathbf{c}(\boldsymbol{\tau}), \mathbf{e}(\boldsymbol{\tau}))$.

A critical distinction between ARSBM and dynSBM lies in their suitability for different types of dynamic network data. dynSBM excels at modeling networks where edges represent transient "interactions" (e.g., e-mail, text message, phone call, etc.), while ARSBM is better suited for "relationships" that persist over time (e.g., friendships, collaborations). This distinction aligns with their temporal characteristics: ARSBM operates in continuous time, capturing gradual relationship evolution, while dynSBM adopts discrete time steps, ideal for frequent interaction events. Furthermore, dynSBM follows a parametric approach, assuming edge weights come from specific probability distributions. In contrast, ARSBM utilizes a more general probabilistic generative modelling approach. Despite these differences, both models share key features: fixed node count

(N) and independent edge evolution. This renders them valuable tools for statistical analysis of dynamic networks, with the optimal choice depending on the specific data and research question.

Most of the time-dependent networks we have seen until now are usually handled via generative models that assume some probabilistic mechanism which governs the evolution of the network in time. Moreover only a few papers (e.g. see Han et al. (2015), Gao et al. (2015), Gao et al. (2016), Klopp et al. (2017)) discuss estimation precision or minimax lower bounds for the risk of estimation of the matrix of connection probabilities.

Pensky (2019) presented the first non-parametric-regression-based Dynamic SBM (DSBM) model based on the smooth evolution of the tensor $\mathbf{G}$ of connection probabilities between communities. The Dynamic Stochastic Block Model (DSBM) tackles the challenge of estimating connection probabilities in time-varying networks. It achieves this through efficient data representation, redundancy removal, and a penalized least squares approach. The paper considers a dynamic network defined as an undirected graph with N nodes, grouped into m classes $\Omega_1, \cdots, \Omega_m$, with connection probabilities changing in time. For simplicity, it's assumed that time instants $0 < t_1 < \cdots < t_L = T$ are equispaced and the time interval is scaled to one, i.e. $t_l = l/L$. For all $i, j \in \{1, \cdots, N\}$, $1 \leq l \leq L$, define $\mathbf{B}_{i,j,l}$ as:

$$\mathbf{B}_{i,j,l} = \begin{cases} 1, & \text{if a connection between nodes } i \text{ and } j \text{ is observed at time } t_l \\ 0, & \text{otherwise} \end{cases}$$

with $\mathbf{B}_{i,i,l} = 0, \quad \text{and} \quad \mathbf{B}_{i,j,l} = \mathbf{B}_{j,i,l}$. The paper derives penalized least squares estimators of $\mathbf{\Lambda}$ (where $\mathbf{\Lambda}_{i,j,l} = \mathbb{P}(\mathbf{B}_{i,j,l} = 1), \mathbf{\Lambda}_{i,i,l} = 0$). Moreover, under the assumption that only at most n_0 nodes can change their memberships between two consecutive time points, the author derives minimax lower bounds for the risk of an estimator of $\mathbf{\Lambda}$. The paper makes a reasonable assumption that the connection probabilities do not change dramatically from one time instant to another. Specifically, it assumes that the connection probabilities are values of some smooth function evaluated at time instant $t_l = l/L \in (0, 1]$ which is a very common assumption in functional data analysis, the analysis of data providing information about curves, surfaces or anything else varying over a continuum. The paper allows group membership switching and enables one to exploit stability in the group memberships over time. Moreover, the estimators constructed in the paper adapt to any number of blocks inherent in the dynamic network data, since the estimation of the number of blocks happens simultaneously using the penalty in the optimization function.

The core data structure is the tensor $\mathbf{\Lambda}$, which stores connection probabilities for all nodes and time points. The DSBM decomposes this tensor using the clustering matrix $\tilde{\mathbf{Z}}^{(l)}$ and a low-complexity tensor $\mathbf{G}_{*,*,l}$ at each time point. This decomposition leverages the inherent structure of the data, where nodes belong to certain groups with specific connection patterns. However, the raw representation still contains redundancies. The model exploits symmetries and sparsity in $\mathbf{\Lambda}$ to remove redundant information from vectors and matrices representing connection probabilities.

This significantly reduces computational complexity and improves estimation efficiency.

Furthermore, the DSBM captures the dynamic nature of the network by assuming smooth changes in connection probabilities over time. These changes are modelled using orthogonal basis functions, such as Fourier or wavelet transforms, allowing for an efficient representation of dynamic patterns with a sparse vector $\mathbf{d}$. Finally, the model formulates an optimization problem using penalized least squares to estimate key parameters. These include the number of clusters, the active entries (non-zero entries) in d, and the block-diagonal matrix C. $\mathbf{C} \in \{0,1\}^{NL \times ML}$ is a block diagonal matrix with blocks $\mathbf{C}^{(l)}, l = 1, \ldots, L$, on the diagonal where $\mathbf{C}^{(l)}$ is obtained by removing the rows in $(\tilde{\mathbf{Z}}^{(l)} \otimes \tilde{\mathbf{Z}}^{(l)})$ corresponding to $(\tilde{\mathbf{Z}}^{(l)}_{i_1,*} \otimes \tilde{\mathbf{Z}}^{(l)}_{i_2,*})$ with $i_1 \geq i_2$. The penalty term in the optimization balances model fits with sparsity in the active entries of $\mathbf{d}$, ensuring both accurate estimation and efficient representation of the dynamic network. It is shown that the correct penalty consists of two parts: the portion which accounts for the complexity of estimation and the portion which accounts for the complexity of clustering and is proportional to the logarithm of the cardinality of the set of clustering matrices. The latter is a novel result and it is obtained by using the innovative Packing lemma (Lemma 4 of the paper Pensky (2019)) which can be viewed as a version of the Varshamov-Gilbert lemma for clustering matrices In summary, the DSBM effectively models and estimates connection probabilities in time-varying networks by exploiting data structure, removing redundancies, and incorporating dynamic changes through efficient sparse representations. This approach offers a computationally efficient and statistically sound solution for analyzing dynamic network data.

The paper also discusses possible extensions that can be made to the modelling approach to incorporate more levels of complexity. It is assumed in the paper that probabilities of connections are spatially homogeneous and are represented by smooth functions of time that belong to the same Sobolev class. The author claims that this can however be generalized to have inhomogeneous or non-smooth connection probabilities as well. The author also presents a brief idea of how dynamic networks with a time-dependent number of nodes can be handled by extending the proposed model.

2.3 Models as time series of networks

Existing approaches to modeling dynamic networks often incorporate temporal dynamics into predefined network structures. Alternatively, one can model the network itself as a time series. Node-based time series models exist, such as vector autoregression, where a node's value depends on its own past and connected neighbours' past values (Zhu et al. (2017)). GNAR models (Knight et al. (2020)) extend this by including larger neighbourhood effects. These methods, however, assume fixed network structures. Both studies assume a fixed network structure across time. Few other studies (e.g. see Kang et al. (2021), Zhu et al. (2019), Chen et al. (2023)) extend/discover new variations of modelling the dynamic networks as time series models over the nodes. In con-

trast, GNAR-edge (Mantziou et al. (2023)) focuses on time series observed on edges, motivated by industry transaction data represented as time-varying edge weights between SIC codes. This approach treats network relationships themselves as dynamic entities, offering a distinct perspective for modelling evolving network structures.

In the multiple time series setting, we observe time series of some fixed length T for different variables $i = 1, ..., K$. Let X_i^t denote the value of variable i at time $t = 1, ..., T$. The VAR(L) model is an autoregressive model with maximum lag L, that has the following linear form,

$$X_i^t = v_i + \alpha_{i_1,1} X_1^{t-1} + \alpha_{i_2,1} X_2^{t-1} + ... + \alpha_{i_K,1} X_K^{t-1} + ...+ \\ \alpha_{i_1,L} X_1^{t-L} + ... + \alpha_{i_{K,L}} X_K^{t-L} + u_i^t, \;\; i = 1, ..., K \tag{2}$$

with intercept v_i, coefficients $\alpha_{i_{j,l}}, j = 1, ..., K, l = 1, ..., L$, and the innovations $(u_1^t, ..., u_K^t)t = 1.....T.$ being a K-dimensional white noise. Equation 2 can be written compactly in vector form.

The GNAR-edge model addresses directed networks on N nodes with possible self-loops, represented by the adjacency matrix $\mathbf{A}$. GNAR-edge model assumes a fixed graph G and a time-varying process on E through the matrix-valued process $V_t, t \geq 0$ of non-negative weights. The time-varying weighted adjacency matrix $\mathbf{X}^t = A \odot V_t$ is defined, with $\odot$ denoting the element-wise product, or Hadamard product, between matrices A and V and X_{ij}^t representing the weight of edge $\{i, j\}$ at time t. Define the set of 1-stage neighbouring edges for some edge $\{i, j\}$ as the set of all edges which are incident to at least one of the nodes i, j; formally, $\mathcal{N}^1(\{ij\}) = \{\{k, l\} \in \delta^+(i) \cup \delta^-(i) \cup \delta^+(j) \cup \delta^-(j) : \{k, l\} \neq \{i, j\}\}$ with $\delta^+(\cdot), \delta^-(\cdot)$ denoting the sets of outgoing and incoming edges of a node, respectively. For $r \geq 2$ we define the set of r-stage neighbouring edges of the edge $\{i, j\}$ as the set $\mathcal{N}^r(\{i, j\}) = \mathcal{N}\{\mathcal{N}^{r-1}(\{i, j\})\} \setminus [\{\{i, j\}\} \cup \{\cup_{q=1}^{r-1} \mathcal{N}^q(\{i, j\})\}]$. As the network is fixed, these sets do not depend on time. It is assumed that the weight X_{ij}^t of an edge between nodes i, j at time t depends not only on its past values but also on the past values of its neighboring edges. The model, denoted as GNAR-edge(L, $[R_1, \ldots, R_L]$), considering a maximum lag L and R_l denoting the maximum stage neighbours for lag l, expresses X_{ij}^t as a function of its past values and past values of its neighbouring edges:

$$X_{ij}^t = \sum_{l=1}^{L} \left(\alpha_{ij,l} X_{ij}^{t-1} + \sum_{r=1}^{R_l} \beta_{l,r} \sum_{m,n:\{m,n\} \in \mathcal{N}^r(\{i,j\})} w_{ij,mn} X_{mn}^{t-1} \right) + u_{ij}^t$$

where $\alpha_{ij,l}$ denotes the standard autoregressive parameters at lag l for edge $\{i, j\}$, $\beta_{l,r}$ denotes the parameters for the effect of r-stage neighboring edges at lag l, $W_{ij,mn} = |\mathcal{N}^r(\{i, j\})|^{-1}$ denoting the normalizing weight for X_{mn}^{t-1} which equally weights all neighboring edges of edge $\{i, j\}$ at lag l, and $u_{ij,t}$ denoting white noise with mean 0 and variance σ^2. The model is formulated as a linear matrix model, and OLS estimators are calculated. The authors provide a sufficient condition for the model to be stationary and assume $\alpha_{ij,l} = \alpha_l \; \forall \; i, j \; \in V$ to reduce model complexity for large networks.

The GNAR-edge model's performance was evaluated through simulations on various network structures (Erdos-Renyi, SBM, RDP graphs) and under different network time series scenarios (heavy-tailed noise, correlated innovations, misspecified connectivity). Parameter estimation accuracy remained consistent across network types, showcasing the model's robustness. Further, model misspecification effects were explored for different network sizes. Additionally, a real-world application on industry transaction data demonstrated a good fit to the GNAR-edge model, with limitations in capturing larger peaks. Sparsification via lead-lag analysis and trend removal were employed before model fitting.

2.4 Other modelling approach

Motivated by challenges in understanding complex systems like gene regulatory networks and stock markets, where real-time network data is often unavailable, Kolar et al. (2010) propose leveraging Markov Random Fields (MRFs) to infer these time-varying networks. The key hurdle lies in the absence of serial snapshots of the underlying networks, making traditional methods inapplicable e.g. in the context of gene regulatory networks, the task is to reconstruct the dynamic interactions between genes based on microarray measurements of their expression levels across various developmental stages. The focus of Kolar et al. (2010) is to estimate dynamic network structure from a time series of entity attributes.

Consider a graph $G = (V, E)$ where $V = \{1, \cdots, N\}$ represents entities (e.g., stocks, genes), and E denotes relationships (e.g., correlation, friendship). Each node in V corresponds to an element of a random vector $X = (X_1, \cdots, X_N)$ with a probability distribution indexed by $\theta \in \Theta$. In the Markov Random Field (MRF) framework, nodal states are discrete, i.e., $\mathbf{X} \in \mathcal{X}^N \equiv \{s_1, \cdots, s_k\}^N$ and the edge set $E \subset V \times V$ encodes conditional independence assumptions of whether X_u is conditionally independent of X_v given the rest of the variables if $(u, v) \notin E$. Kolar et al. (2010) analyzes a special kind of MRF known as the Ising model (with $\mathcal{X} \in \{-1, 1\}$), under which: $\mathbb{P}_\theta(\mathbf{x}) = \frac{1}{Z} \exp\{\sum_{u<v} \theta_{uv} x_u x_v\}$, where Z denotes the partition function, and the pairwise potentials θ_{uv} represent interactions between nodes for all $(u, v) \in E$ and $\theta_{uv} = 0$ otherwise. The main objective in Kolar et al. (2010) is to estimate the time-varying graph structures of MRFs from a time series of nodal states $\{x_t\}_{t \in \mathcal{T}_n}$, with $\mathcal{T}_n = \{1/n, 2/n, \cdots, 1\}$ being the time index set, that are independent (but not identically distributed) samples from a series of time-evolving MRFs $\{\mathbb{P}_{\theta^t}(\cdot)\}_{t \in \mathcal{T}_n}$ which is a much more challenging and more realistic scenario than the one that assumes that the nodal states are sampled i.i.d. from a time-invariant MRF (Bresler et al. (2018),Ravikumar et al. (2010)). Let $\mathcal{D}_n = \{\mathbf{x}^t \sim \mathbb{P}_{\boldsymbol{\theta}^t} | t \in \mathcal{T}_n\}$ be an independent sample of n observations. It is assumed that $\mathbf{X}_t$ is a N-dimensional random variable with values in $\{-1, 1\}^N$ following a distribution given by $\mathbb{P}_{\boldsymbol{\theta}^t}(\mathbf{x}) = \frac{1}{Z(\boldsymbol{\theta}^t)} \exp\{\sum_{(u,v) \in E^t} \theta^t_{uv} x_u x_v\}$, where $Z(\boldsymbol{\theta}^t)$ is the partition

function, $\boldsymbol{\theta}^t \in \mathbb{R}^{\binom{N}{2}}$ is the parameter vector, and $G^t = (V, E^t)$ is an undirected graph representing conditional independence assumptions. Kolar et al. (2010) addresses the problem of graph structure estimation from observational data: given any time point $\tau \in [0, 1]$ estimate the graph structure associated with $\mathbb{P}_{\boldsymbol{\theta}^t}$, given the observations $\mathcal{D}_n$, by focusing on estimating the non-zero pattern of the vector $\boldsymbol{\theta}^\tau$. Define $\boldsymbol{\theta}_u^\tau := \{\theta_{uv}^\tau | v \in V \setminus u\}$ and $S^\tau(u) := \{(u, v) \in V \times V | \theta_{uv}^\tau \neq 0\}$. The paper focuses on obtaining node-wise estimators $\hat{\boldsymbol{\theta}}_u^\tau$ of the non-zero pattern of the subvector $\boldsymbol{\theta}_u^\tau$. Due to possible asymmetry in the estimated non-zero pattern ($\hat{\theta}_{uv}^\tau = 0$, but $\hat{\theta}_{vu}^\tau \neq 0$), two symmetrization methods are suggested to arrive at a combined estimator $\tilde{\boldsymbol{\theta}}^\tau$: *min symmetrization* ($\tilde{\theta}_{uv} = \hat{\theta}_{uv}$ if $|\hat{\theta}_{uv}| < |\hat{\theta}_{vu}|$, else $\hat{\theta}_{vu}$) and *max symmetrization* ($\tilde{\theta}_{uv} = \hat{\theta}_{uv}$ if $|\hat{\theta}_{uv}| > |\hat{\theta}_{vu}|$, else $\hat{\theta}_{vu}$)..

The estimator $\hat{\boldsymbol{\theta}}^\tau$ is obtained by maximizing the pseudo-likelihood based on the conditional distribution of X_u^τ given other variables $\mathbf{X}_{\setminus u}^\tau = \{X_v^\tau | v \in V \setminus u\}$, wherein the distribution is given by

$$\mathbb{P}_{\boldsymbol{\theta}_u^\tau}(x_u^\tau | \mathbf{X}_{\setminus u}^\tau = \mathbf{x}_{\setminus u}^\tau) = \frac{\exp(x_u^\tau \langle \boldsymbol{\theta}_u^\tau, \mathbf{x}_{\setminus u}^\tau \rangle)}{\exp(x_u^\tau \langle \boldsymbol{\theta}_u^\tau, \mathbf{x}_{\setminus u}^\tau \rangle) + \exp(-x_u^\tau \langle \boldsymbol{\theta}_u^\tau, \mathbf{x}_{\setminus u}^\tau \rangle)}$$

In the absence of assumptions about $\boldsymbol{\theta}^t$, the estimation problem becomes ill-posed when the number of observations is small compared to the dimensionality. To address this, assuming sparse graphs $\{G_t\}_{t \in \mathcal{T}_n}$ is a common strategy. Furthermore, the paper explores two ways to constrain the parameter vector $\boldsymbol{\theta}^t$ over time: *Smooth changes in parameters* and *piecewise constant with abrupt structural changes in parameters*. The authors discuss in detail the procedure of estimation of the parameters in both these cases. Parameter tuning is done via BIC score defined in the paper where the degrees of freedom in the model are approximated appropriately.

Simulation studies on Erdos Renyi graphs revealed that MIN symmetrization tends to be more conservative in edge inclusion and is sensitive to noise. The model's application extended to Senate voting data from the 109th Congress and a gene expression dataset of Drosophila melanogaster's life cycle. The paper establishes, following Kolar. and Xing (2009), that asymptotic recovery of the graph is possible with an appropriate regularization parameter, given moderate model dimensions, maximum node degrees, and a non-rapid decline of the minimum parameter value to zero. Notably, Kolar et al. (2010) assumed independence of observations at different time points, suggesting future research avenues in graph structure estimation from dependent time series data. An extension to accommodate multi-category data is also suggested (Ravikumar et al. (2010)).

3 Network Summary Statistics

Networks or graphs are used to model and study complex systems across many domains. While visually inspecting a network can provide some intuition, but as soon as the number of nodes and edges increase beyod a certain amount, it becomes impossible to analyse the network visually. Numerical summary statistics allow us to quantitatively capture key topological properties in a

concise manner. These statistics serve several important purposes:

1. **Summarizing Structure**: Networks can be large and intricate, making them difficult to comprehend visually. Summary statistics distill the vast structural information into interpretable numeric values, enabling compact representations and comparisons.

2. **Identifying Patterns**: By quantifying specific features like centrality, degree distributions, or subgraph counts, summary statistics help identify and measure structural patterns that may reveal underlying organizational principles or dynamics of the system.

3. **Testing Hypotheses**: Summary statistics can be compared against theoretical models or benchmarks to test hypotheses about the processes governing a network's formation, evolution, or behavior.

4. **Characterizing Networks**: These statistics provide a quantitative basis for characterizing and contrasting the topologies of different networks or snapshots of an evolving network over time.

5. **Guiding Modeling**: When developing generative network models, summary statistics from real-world networks indicate which topological properties the models should reproduce accurately.

However, some drawbacks exist:

1. **Information Reduction**: Summarizing a complex structure risks losing potentially important details or outliers relevant to analysis.

2. **Instability**: Certain statistics may be highly sensitive to minor perturbations in the network data.

3. **Context Dependence**: The utility of a statistic depends on the network type and research context studied.

4. **Descriptive Nature**: Traditionally, these statistics describe but do not directly support inferential conclusions about processes underlying an observed network.

Despite limitations, descriptive summary statistics provide an invaluable quantitative lens into network topology, underpinning further analysis, hypothesis testing, and model development. A key statistic explored in this thesis is the edge density.

3.1 Examples of network summary statistics

- **Degree distribution**: The degree distribution is a fundamental summary statistic in network analysis that characterizes the connectivity patterns within a network. It represents the probability distribution of the degrees of nodes, where the degree of a node is the number of connections or edges it has to other nodes. For static networks, the degree distribution provides a compact representation of the overall level of connectivity and the variability in the number of connections across nodes. In dynamic networks that evolve over time, the degree distribution becomes a time-varying quantity, capturing the temporal characteristics of the network's connectivity patterns.

  The degree distribution serves several important purposes in network analysis:

  - Identifying structural patterns and features, such as the presence of hubs or hierarchical organization.
  - Comparing observed networks against theoretical models or random graph ensembles.
  - Characterizing the generative processes and mechanisms underlying network formation and evolution.
  - Detecting changes, trends, or regime shifts in the connectivity dynamics of temporal networks.

  As a fundamental statistic, the degree distribution provides a quantitative basis for understanding the topological properties of networks and their implications for processes occurring on or driven by the network structure, such as diffusion, influence propagation, or resilience to disruptions.

- **Degree centrality**: In network analysis, degree centrality is a measure that quantifies the importance or influence of a node based on the number of connections it has within the network. For static networks, degree centrality is simply the degree (number of edges) of the node. However, in dynamic networks where connections are continuously formed and dissolved over time, the degree centrality of a node becomes a time-varying quantity.

  The time-varying degree centrality captures how the importance or connectivity of a node evolves as the network topology changes dynamically. Tracking the degree centrality of nodes over time enables the analysis of phenomena such as:

  - Identification of influential or prominent nodes whose centrality exhibits significant changes or distinct temporal patterns.
  - Characterization of node-level connectivity dynamics and their relation to local network processes or external factors.

  - Detection of critical events, shifts, or anomalies in the network structure based on sudden changes in node centralities.
  - Modeling and prediction of future centrality trajectories, which can inform applications like influence maximization or network control.

  As a fundamental centrality measure, the time-varying degree centrality provides insights into the dynamic roles and positions of nodes within an evolving network topology, enabling the analysis of processes driven by connectivity patterns, such as information diffusion, disease spread, or the robustness of the network to targeted attacks or failures.

- **Motif Counts**: Motif counts quantify the occurrences of specific subgraph patterns, or motifs, within a network. In dynamic networks where the topology evolves over time, motif counts become time-varying quantities that capture the dynamics of these recurring structural patterns.

  Tracking the temporal evolution of motif counts enables the analysis of:

  - Identifying characteristic motif signatures and their changes, which can reveal underlying generative mechanisms or constraints governing the network dynamics.
  - Detecting temporal correlations, periodicity, or regime shifts in the formation and dissolution of specific motif types, potentially linked to external factors or critical events.
  - Comparing motif distributions across different dynamic networks or time periods to uncover similarities or differences in their organizational principles or evolution patterns.
  - Modeling and predicting future motif counts, which can inform applications such as network growth forecasting or the design of dynamic processes constrained by specific motif patterns.

  As informative descriptors of local connectivity structures, time-varying motif counts provide a lens into the co-evolution of network topology and dynamics, enabling the analysis of processes governed by or embedded within these recurring building blocks of complex networks.

In the rest of the report we will be discussing about another particular network summary statistic called the edge density.

4 Edge Density

Edge density stands as a fundamental metric in network analysis, providing a concise measure of connectivity within a network. The properties like size and density of a network are conceptually similar to the mass and composition of matter—they just tell us how much stuff is in it, but

they don't tell us anything about how the matter is organized internally. Nonetheless, they are still the most fundamental characteristics, which are particularly important when you want to compare multiple networks. Analogous to assessing traffic flow on a congested highway, edge density quantifies the proportion of actual connections present to the total possible connections within a network.

The density of a network is the fraction between 0 and 1 that tells us what portion of all possible edges are realized in the network. For a network G made of n nodes and m edges, the density $\rho(G)$ is given by

$$\rho(G) = \frac{m}{\frac{n(n-1)}{2}} = \frac{2m}{n(n-1)}$$

for an undirected network, or

$$\rho(G) = \frac{m}{n(n-1)}$$

for a directed network. The density of a network in graph theory or network science refers to the measure of how closely connected the nodes are in a graph. It quantifies the number of edges present in the graph relative to the total number of possible edges. A dense network has a high density, indicating that a large proportion of possible connections between nodes are present. Dense networks are often observed in various domains such as on-line social networks, recommendation networks, and the brain. The density of a graph can be used as a topological descriptor to partition the nodes into regions of uniform density, providing insights into the structure of the network. The concept of density is also used in the context of finding disjoint spanning trees and subgraphs in a graph. The density of a network can influence its spectral density, which is important for understanding the behavior of models defined on graphs

Note that the size and density of a network don't specify much about the network's actual topology(i.e. shape). There are many networks with different topologies that have the same size and density. But there are some things the size and density can still predict about networks. One such example is network percolation, i.e., whether or not the nodes are sufficiently connected so that they form a giant component that is visible at macroscopic scales. A giant component is a connected component whose size is on the same order of magnitude as the size of the whole network.

4.1 Results on edge density of Matias and Miele (2017)'s model

Among several different models available for dynamic networks, we discussed a handful of them earlier. To better understand the evolution of edge density we will focus on one of the above-discussed particular models namely the one by Matias and Miele (2017). To better understand the model first, we find the dependence of each edge evolution over time as in (4.1.1).

4.1.1 Conditional Distribution $Y_{ij}^{(t)}|Y_{ij}^{(t-1)}$

Consider the dynSBM model with N nodes, Q communities and T discrete time steps. The community memberships $Z_i^{(t)}$ follow an irreducible, aperiodic, and ergodic Markov chain. Let the transition matrix for these community memberships be η. Also, assume that the different nodes have independent and identically distributed (i.i.d.) community memberships for all time steps t. The initial state of any node is sampled from an initial distribution $\begin{pmatrix}\alpha_1 & \alpha_2 \cdots & \alpha_Q\end{pmatrix}$. We wish to find the conditional distribution $Y_{ij}^{(t)}|Y_{ij}^{(t-1)} = 1$ and $Y_{ij}^{(t)}|Y_{ij}^{(t-1)} = 0$.
First consider finding $\mathbb{P}(Y_{ij}^{(t)} = 1|Y_{ij}^{(t-1)} = 1)$

Note that

$$\mathbb{P}(Y_{ij}^{(t)} = 1|Y_{ij}^{(t-1)} = 1) = \sum_{q,l=1}^{Q} \mathbb{P}(Y_{ij}^{(t)} = 1|Y_{ij}^{(t-1)} = 1, Z_i^{(t-1)} = q, Z_j^{(t-1)} = l) \\ \mathbb{P}(Z_i^{(t-1)} = q, Z_j^{(t-1)} = l|Y_{ij}^{(t-1)} = 1)$$

First, consider

$$\mathbb{P}(Y_{ij}^{(t)} = 1|Y_{ij}^{(t-1)} = 1, Z_i^{(t-1)} = q, Z_j^{(t-1)} = l) \\ = \sum_{r,s=1}^{Q} \mathbb{P}(Y_{ij}^{(t)} = 1|Y_{ij}^{(t-1)} = 1, Z_i^{(t-1)} = q, Z_j^{(t-1)} = l, Z_i^{(t)} = r, Z_j^{(t)} = s) \\ \mathbb{P}(Z_i^{(t)} = r, Z_j^{(t)} = s|Y_{ij}^{(t-1)} = 1, Z_i^{(t-1)} = q, Z_j^{(t-1)} = l)$$

But, given the current community memberships at time t of both the nodes, the distribution of $Y_{ij}^{(t)}$ is completely determined and independent of $Y_{ij}^{(t-1)}$

$$\mathbb{P}(Y_{ij}^{(t)} = 1|Y_{ij}^{(t-1)} = 1, Z_i^{(t-1)} = q, Z_j^{(t-1)} = l, Z_i^{(t)} = r, Z_j^{(t)} = s) \\ = \mathbb{P}(Y_{ij}^{(t)} = 1|Z_i^{(t)} = r, Z_j^{(t)} = s) \\ = \beta_{rs}^{(t)}$$

Thus $\mathbb{P}(Y_{ij}^{(t)} = 1|Y_{ij}^{(t-1)} = 1, Z_i^{(t-1)} = q, Z_j^{(t-1)} = l, Z_i^{(t)} = r, Z_j^{(t)} = s) = \beta_{rs}^{(t)}$.
And,

$$\mathbb{P}(Z_i^{(t)} = r, Z_j^{(t)} = s|Y_{ij}^{(t-1)} = 1, Z_i^{(t-1)} = q, Z_j^{(t-1)} = l) \\ = \mathbb{P}(Z_i^{(t)} = r|Z_i^{(t-1)} = q)\mathbb{P}(Z_j^{(t)} = s|Z_i^{(t-1)} = l) \\ = \eta_{qr}\eta_{ls}$$

$$\therefore \quad \mathbb{P}(Y_{ij}^{(t)} = 1 | Y_{ij}^{(t-1)} = 1, Z_i^{(t-1)} = q, Z_j^{(t-1)} = l) = \sum_{r,s=1}^{Q} \beta_{rs}^{(t)} \eta_{qr} \eta_{ls}$$
$$= (\eta \Lambda_t \eta')_{ql}$$

where $1 \leq t \leq T$, $\Lambda_t = \left(\left(\beta_{ql}^{(t)}\right)\right)_{q,l=1}^{Q}$

Now, by Bayes rule,

$$\begin{aligned}
&\mathbb{P}(Z_i^{(t-1)} = q, Z_j^{(t-1)} = l | Y_{ij}^{(t-1)} = 1) \\
&= \frac{\mathbb{P}(Y_{ij}^{(t-1)} = 1 | Z_i^{(t-1)} = q, Z_j^{(t-1)} = l) \mathbb{P}(Z_i^{(t-1)} = q, Z_j^{(t-1)} = l)}{\sum_{k,m=1}^{Q} \mathbb{P}(Y_{ij}^{(t-1)} = 1 | Z_i^{(t-1)} = k, Z_j^{(t-1)} = m) \mathbb{P}(Z_i^{(t-1)} = k, Z_j^{(t-1)} = m)} \\
&= \frac{\beta_{ql}^{(t)} \mathbb{P}(Z_i^{(t-1)} = q) \mathbb{P}(Z_j^{(t-1)} = l)}{\sum_{k,m=1}^{Q} \beta_{km}^{(t)} \mathbb{P}(Z_i^{(t-1)} = k) \mathbb{P}(Z_j^{(t-1)} = m)} \\
&= \frac{\beta_{ql}^{(t)} \mathbb{P}(Z_i^{(t-1)} = q) \mathbb{P}(Z_j^{(t-1)} = l)}{\left(\mathbb{P}(Z_i^{(t-1)} = 1) \quad \mathbb{P}(Z_i^{(t-1)} = 2) \quad \cdots \quad \mathbb{P}(Z_i^{(t-1)} = Q)\right) \Lambda_{t-1} \begin{pmatrix} \mathbb{P}(Z_i^{(t-1)} = 1) \\ \mathbb{P}(Z_i^{(t-1)} = 2) \\ \vdots \\ \mathbb{P}(Z_i^{(t-1)} = Q) \end{pmatrix}}
\end{aligned}$$

But,

$$\begin{aligned}
\left(\mathbb{P}(Z_i^{(t-1)} = 1) \quad \cdots \quad \mathbb{P}(Z_i^{(t-1)} = Q)\right) &= \left(\mathbb{P}(Z_i^{(t-2)} = 1) \quad \cdots \quad \mathbb{P}(Z_i^{(t-2)} = Q)\right) \eta \\
&= \left(\mathbb{P}(Z_i^{(t-3)} = 1) \quad \cdots \quad \mathbb{P}(Z_i^{(t-3)} = Q)\right) \eta^2 \\
&\cdot \\
&\cdot \\
&\cdot \\
&= \left(\mathbb{P}(Z_i^{(1)} = 1) \quad \cdots \quad \mathbb{P}(Z_i^{(1)} = Q)\right) \eta^{t-2} \\
&= \left(\alpha_1 \quad \alpha_2 \quad \cdots \quad \alpha_Q\right) \eta^{t-2}
\end{aligned}$$

where $\boldsymbol{\alpha} = \left(\alpha_1 \quad \cdots \quad \alpha_Q\right)'$ denotes the initial distribution of the community membership of any node Similarly, $\left(\mathbb{P}(Z_j^{(t-1)} = 1) \quad \mathbb{P}(Z_j^{(t-1)} = 2) \cdots \quad \mathbb{P}(Z_j^{(1)} = Q)\right) = \left(\alpha_1 \quad \alpha_2 \quad \cdots \quad \alpha_Q\right) \eta^{t-2}$. Let $\boldsymbol{\alpha} = \left(\alpha_1 \quad \alpha_2 \quad \cdots \quad \alpha_Q\right)^T$. Thus we have that

$$\sum_{k,m=1}^{Q} \beta_{km}^{(t)} \mathbb{P}(Z_i^{(t-1)} = k)\mathbb{P}(Z_j^{(t-1)} = m) = \boldsymbol{\alpha}^T \eta^{t-2} \Lambda_{t-1} (\eta^{t-2})' \boldsymbol{\alpha}$$

$$\therefore \;\; \mathbb{P}(Z_i^{(t-1)} = q, Z_j^{(t-1)} = l | Y_{ij}^{(t-1)} = 1) = \frac{\beta_{ql}^{(t)} \; \mathbb{P}(Z_i^{(t-1)} = q) \; \mathbb{P}(Z_j^{(t-1)} = l)}{\boldsymbol{\alpha}' \; \eta^{t-2} \; \Lambda_{t-1} \; (\eta^{t-2})' \; \boldsymbol{\alpha}}$$

$$\begin{aligned}
\therefore \;\; \mathbb{P}(Y_{ij}^{(t)} = 1 | Y_{ij}^{(t-1)} = 1) &= \sum_{q,l=1}^{Q} \left\{ (\eta \Lambda_t \eta')_{ql} \;\; \left(\frac{\beta_{ql}^{(t)} \; \mathbb{P}(Z_i^{(t-1)} = q) \; \mathbb{P}(Z_j^{(t-1)} = l)}{\boldsymbol{\alpha}' \; \eta^{t-2} \; \Lambda_{t-1} \; (\eta^{t-2})' \; \boldsymbol{\alpha}} \right) \right\} \\
&= \frac{\sum_{q,l=1}^{Q} \mathbb{P}(Z_i^{(t-1)} = q) \; \beta_{ql}^{(t)} \; (\eta \Lambda_t \eta')_{ql} \; \mathbb{P}(Z_j^{(t-1)} = l)}{\boldsymbol{\alpha}' \; \eta^{t-2} \; \Lambda_{t-1} \; (\eta^{t-2})' \; \boldsymbol{\alpha}} \\
&= \frac{\boldsymbol{\alpha}' \; \eta^{t-2} \; [\Lambda_{t-1} \odot (\eta \Lambda_t \eta')] \; (\eta^{t-2})' \; \boldsymbol{\alpha}}{\boldsymbol{\alpha}' \; \eta^{t-2} \; \Lambda_{t-1} \; (\eta^{t-2})' \; \boldsymbol{\alpha}}
\end{aligned}$$

where $\odot$ represents the Hadamard product.

Similarly, with J representing the matrix with all ones, we have that

$$\mathbb{P}(Y_{ij}^{(t)} = 0 | Y_{ij}^{(t-1)} = 1) = \frac{\boldsymbol{\alpha}' \; \eta^{t-2} \; [(J - \Lambda_{t-1}) \odot (\eta \Lambda_t \eta')] \; (\eta^{t-2})' \; \boldsymbol{\alpha}}{\boldsymbol{\alpha}' \; \eta^{t-2} \; (J - \Lambda_{t-1}) \; (\eta^{t-2})' \; \boldsymbol{\alpha}}$$

4.1.2 Expected Edge Density

After having a look at the conditional dependence of edge behaviour, we are also interested in exploring the expected edge density of the dynamic networks i.e. if $Y_{ij}^{(t)}$ denotes the presence or absence of an edge between nodes i and j at time t ($Y_{ij}^{(t)} = 1$ if an edge exists between nodes i and j at time t, else 0), then the expected edge density at time t is given by

$$\mathbb{E}\left[\frac{1}{\binom{n}{2}} \sum_{i<j} Y_{ij}^{(t)} \right]$$

where $Y_{ij}^{(t)} | \{ Z_i^{(t)} = q, Z_j^{(t)} = l \} \sim \text{Ber}(\beta_{ql}^{(t)})$ where $\beta_{ql}^{(t)}$ are the parameters of the model and $(Z_i^{(t)})_{1 \le i \le N, 1 \le t \le T}$ denote the community memberships of each node at a particular time instant. Let $\Lambda_t = \left(\left(\beta_{ql}^{(t)} \right) \right)_{q,l=1}^{Q}$

Now note that

$$\begin{aligned}
\text{Expected Edge density at time } t &= \mathbb{E}\left[\frac{1}{\binom{n}{2}} \sum_{i<j} Y_{ij}^{(t)} \right] \\
&= \frac{2}{n(n-1)} \sum_{i<j} \mathbb{E}\left[Y_{ij}^{(t)} \right]
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}\left[Y_{ij}^{(t)}\right] &= \mathbb{E}\left[\mathbb{E}\left[Y_{ij}^{(t)}|Z_i^{(t)}, Z_j^{(t)}\right]\right] \\
&= \sum_{q,l=1}^{Q} \beta_{ql}^{(t)} \mathbb{P}(Z_i^{(t)} = q, Z_j^{(t)} = l) \\
&= \sum_{q,l=1}^{Q} \beta_{ql}^{(t)} \mathbb{P}(Z_i^{(t)} = q)\mathbb{P}(Z_j^{(t)} = l) \\
&= \begin{pmatrix} \mathbb{P}(Z_i^{(t)} = 1) & \mathbb{P}(Z_i^{(t)} = 2) & \cdots & \mathbb{P}(Z_i^{(t)} = Q) \end{pmatrix} \Lambda_t \begin{pmatrix} \mathbb{P}(Z_i^{(t)} = 1) \\ \mathbb{P}(Z_i^{(t)} = 2) \\ \vdots \\ \mathbb{P}(Z_i^{(t)} = Q) \end{pmatrix}
\end{aligned}$$

From section (4.1.1) we know that (following the same notations as in(4.1.1))

$$\begin{pmatrix} \mathbb{P}(Z_i^{(t)} = 1) & \mathbb{P}(Z_i^{(t)} = 2) & \cdots & \mathbb{P}(Z_i^{(t)} = Q) \end{pmatrix} = \begin{pmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_Q \end{pmatrix} \eta^{t-1} = \boldsymbol{\alpha}' \eta^{t-1}$$

Thus we have that

$$\mathbb{E}\left[Y_{ij}^{(t)}\right] = \begin{pmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_Q \end{pmatrix} \eta^{t-1} \Lambda_t (\eta^{t-1})' \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_Q \end{pmatrix} = \boldsymbol{\alpha}' \eta^{t-1} \Lambda_t (\eta^{t-1})' \boldsymbol{\alpha}$$

$$\begin{aligned}
\therefore \text{ Expected Edge Density at time } t &= \mathbb{E}\left[\frac{1}{\binom{n}{2}} \sum_{i<j} Y_{ij}^{(t)}\right] \\
&= \frac{2}{n(n-1)} \sum_{i<j} \mathbb{E}\left[Y_{ij}^{(t)}\right] \\
&= \frac{2}{n(n-1)} \sum_{i<j} \boldsymbol{\alpha}' \eta^{t-1} \Lambda_t (\eta^{t-1})' \boldsymbol{\alpha} \\
&= \boldsymbol{\alpha}' \eta^{t-1} \Lambda_t (\eta^{t-1})' \boldsymbol{\alpha}
\end{aligned}$$

This expected edge density can also be calculated using the results from the section 4.1.1 by noting the following identity:

$$\begin{aligned}
\mathbb{E}[Y_{ij}^{(t)}] &= \mathbb{P}(Y_{ij}^{(t)} = 1) \\
&= \mathbb{P}(Y_{ij}^{(t-1)} = 1)\ \mathbb{P}(Y_{ij}^{(t)} = 1 | Y_{ij}^{(t-1)} = 1) + \mathbb{P}(Y_{ij}^{(t-1)} = 0)\ \mathbb{P}(Y_{ij}^{(t)} = 1 | Y_{ij}^{(t-1)} = 0)
\end{aligned}$$

Thus by letting $a_t = \mathbb{P}(Y_{ij}^{(t)} = 1)$ for $1 \leq t \leq T$, we have that

$$a_t = a_{t-1}\ \mathbb{P}(Y_{ij}^{(t)} = 1 | Y_{ij}^{(t-1)} = 1) + (1 - a_{t-1})\ \mathbb{P}(Y_{ij}^{(t)} = 1 | Y_{ij}^{(t-1)} = 0)$$

Substituting the results obtained in section 4.1.1, and solving this above recurrence relation we again get that

$$\mathbb{E}[Y_{ij}^{(t)}] = \boldsymbol{\alpha}' \eta^{t-1} \Lambda_t (\eta^{t-1})' \boldsymbol{\alpha}$$

5 Simulation Study

To empirically validate the theoretical results derived in Sections 4.1.1 and 4.1.2, and further investigate the behaviour of edge density under the dynamic stochastic block model in Matias and Miele (2017), a comprehensive simulation study was undertaken. The study involved generating multiple realizations of dynamic networks and analyzing the evolution of edge density over time.

5.1 Simulation Setup

The simulation setup consisted of the following specifications:

1. Number of nodes: 400
2. Number of time stamps: 500
3. Number of communities: 3

The parameters $\{\beta_{ql}^{(t)}\}_{q,l=1}^{Q}$ governing the connection probabilities between communities at each time point were simulated as follows:

- For all $1 \leq t \leq T$, the diagonal entries of Λ_t were drawn uniformly from the interval $(0.55, 0.9)$, representing within-community connection probabilities.
- The off-diagonal entries of Λ_t, representing between-community connection probabilities, were drawn uniformly from $(0, 1)$ while ensuring symmetry of the matrices.

The transition matrix η for the community membership Markov chain was specified as:

$$\eta = \begin{bmatrix} 0.8 & 0.05 & 0.15 \\ 0.2 & 0.75 & 0.05 \\ 0.25 & 0.2 & 0.55 \end{bmatrix}$$

With the above parameter settings, a total of 850 sequences of dynamic networks were generated, each consisting of $T = 500$ time-indexed networks. The generation procedure for each sequence was as follows:

1. Simulate initial community memberships from a uniform distribution over the three communities.

2. Generate the network at time $t = 1$ using the initial community memberships and Λ_1.

3. Update the community membership distribution for time $t = 1$ using the transition matrix η.

4. Simulate community memberships for each node at time $t = 1$ from the updated distribution and generate the corresponding network.

5. Repeat steps 3 and 4 until the network at time $t = T$ is generated.

5.2 Results and Analysis

Let W_i^t denote the edge density of the t^{th} network in the i^{th} simulation sequence. The evolution of W_i^t across different initial community memberships and parameter sets $\{\beta_{ql}^{(t)}\}_{q,l=1}^{Q}$ is illustrated in Figure (2).

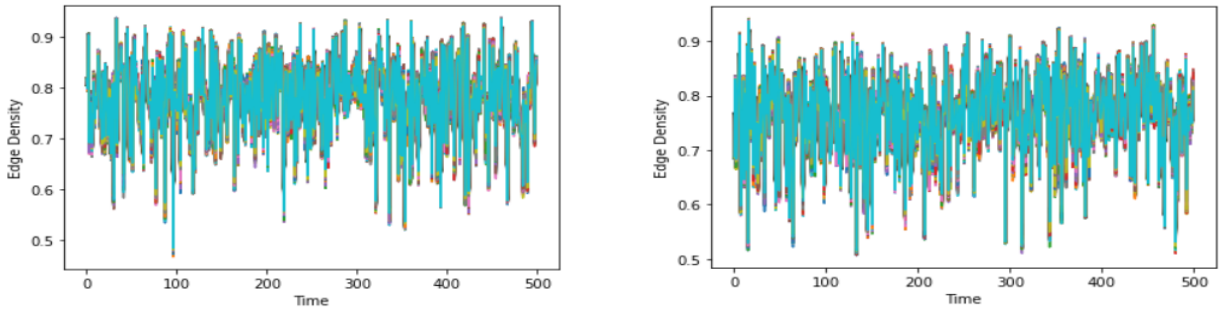

Figure 2: Evolution of W_i^t

Figure 3 displays the adjacency matrix for the network at time $t = T$ for a particular simulation, revealing a community structure consistent with the stationary distribution $\eta^* = \begin{pmatrix} 0.52564 & 0.26923 & 0.20512 \end{pmatrix}$ of the transition matrix η.

To verify the theoretical result in Section 4.1.1, the conditional probability $P(Y_{ij}^{(t)} = 1 | Y_{ij}^{(t-1)} = 1)$ was approximated using Monte Carlo methods. Specifically, the proportion of times the edge $(1, 2)$ existed at time t, given its presence at time $t-1$, was computed across the simulated sequences. Figure 4 compares these simulated probabilities with the theoretical values obtained in Section 4.1.1, demonstrating excellent agreement between the two.

Furthermore, the distribution of edge densities for large values of t was investigated by constructing a histogram of the simulated values $\{W_i^{500}\}_{i=1}^{850}$. Figure 6 presents a centered and scaled version of this histogram. As depicted in Figure 6, the edge densities when centered and scaled as: $T\left(W_i^{500} - \mathbb{E}[W_1^{500}]\right)$, exhibited a non-degenerate distribution, suggesting the potential for constructing large-sample statistical tests based on the edge density statistic.

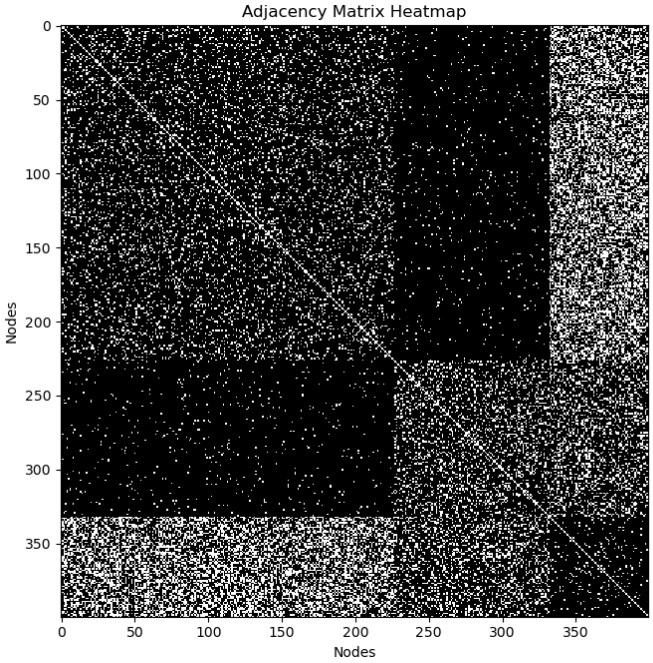

Figure 3: Adjacency matrix for the network at time $t = T$ for a particular simulation

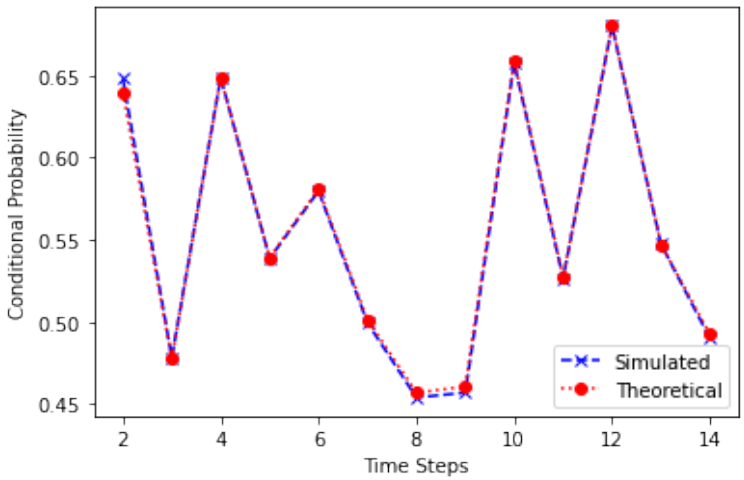

Figure 4: The simulated and theoretical obtained values of $P(Y_{12}^{(t)} = 1 | Y_{12}^{(t-1)} = 1)$

```
Variance of edge density at t = 500:  1.9648870525560727e-05
Mean of edge density at t = 500:  0.8481389945451865
```

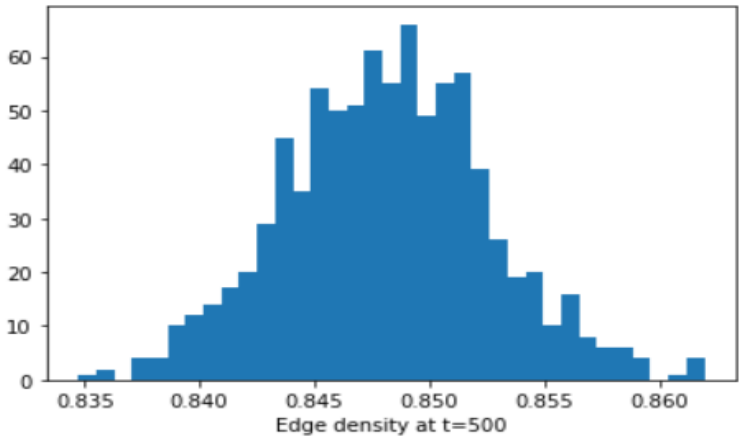

Figure 5: Histogram of edge densities at $t = 500$

```
Variance of centred and scaled edge density at t = 500:  4.912217631390182
Mean of centred and scaled edge density at t = 500:  4.572551487773821e-15
```

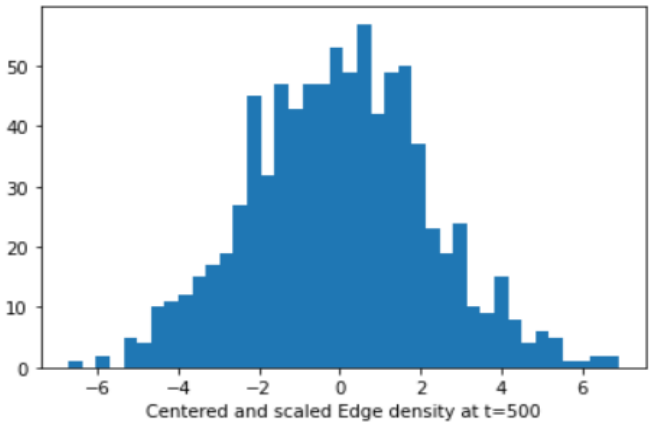

Figure 6: Centered and scaled histogram of the edge densities at $t = 500$

The simulation study validated the theoretical findings and provided insights into the distributional properties of edge density under the dynamic stochastic block model framework. These results lay the groundwork for further exploration and development of statistical inference procedures tailored to the analysis of dynamic network data.

6 Further exploration

Our simulations majorly focused on the verification of theoretical results obtained in sections 4.1.2 and 4.1.1. We also explored the distribution of edge densities for large t. Another area of exploration could have been the distribution of average edge densities from $t = 1$ to $t = T$. Since the edge densities would not be independent for different t (since the underlying dynamic networks had their community memberships evolved via a Markov Chain), it would be interesting to check if the average edge densities have any law of large number type results or results similar to that of the central limit theorem. In this report, we have only focused on a particular descriptive network summary statistic namely the edge density. Summary statistics like the motif counts and centrality of the network would provide more information about the structure of the network hence extending the results (and the simulations) of this report to these metrics may be useful for in-depth analysis of the dynamic networks.

7 Conclusion

This report explored various approaches for modelling and analyzing dynamic networks, with a particular focus on the dynamic stochastic block model proposed by Matias and Miele (2017). Different models were reviewed, including the dynamic Erdos-Renyi model, time series models for networks, and Markov random field approaches. Simulations were conducted to verify theoretical results derived for the edge density and its evolution over time under the dynamic stochastic block model. The simulations validated the theoretical expressions obtained for the conditional probability of edge existence given past states, as well as the expected edge density at any time point. Additionally, the distribution of edge densities for large time points was explored, revealing a non-degenerate limiting distribution after appropriate centering and scaling.

While this report concentrated primarily on edge density, further research could extend the analysis to other topological descriptors, such as motif counts and centrality measures. Investigating the behaviour and distributional properties of these statistics would provide deeper insights into the structural dynamics of evolving networks. Moreover, potential avenues for future work include studying the distribution of average edge densities across multiple time points, accounting for dependencies introduced by the Markovian evolution of community memberships. Establishing limit theorems or laws of large numbers for such averaged quantities could facilitate the development

of statistical inference procedures based on edge density or other summary statistics.

Overall, this project has contributed to the understanding of dynamic network models, particularly the dynamic stochastic block model, and has paved the way for further exploration of descriptive statistics and their analytical and distributional properties in the context of time-varying networks.

References

C. Ambroise and C. Matias. New consistent and asymptotically normal parameter estimates for random graph mixture models. *J. R. Statist. Soc. B*, 74:3–35, 2012.

G. Bresler, E. Mossel, and A. Sly. Reconstruction of markov random fields from samples: Some observations and algorithms. In *Goel, A., Jansen, K., Rolim, J.D.P., Rubinfeld, R. (eds) Approximation, Randomization and Combinatorial Optimization. Algorithms and Techniques. APPROX RANDOM 2008. Lecture Notes in Computer Science*, volume 5171, 2018.

A. Celisse, J.-J. Daudin, and L. Pierre. Consistency of maximum-likelihood and variational estimators in the stochastic block model. *Electron. J. Statist.*, 6:1847–1899, 2012.

E. Y. Chen, J. Fan, and X. Zhu. Community network auto-regression for high-dimensional time series. *Journal of Econometrics*, 235(2):1239–1256, 2023.

P. Erdős and A. Rényi. On random graphs i. *Publicationes Mathematicae Debrecen*, 6:290–297, 1959.

J. Fournet and A. Barrat. Contact patterns among high school students. *PLOS One*, 9, 2014.

O. Frank and F. Harary. Cluster inference by using transitivity indices in empirical graphs. *J. Amer. Statist. Assoc.*, pages 835–840, 1982.

W. Fu, L. Song, and E. Xing. Dynamic mixed membership blockmodel for evolving networks. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 329–336. ACM, 2009.

C. Gao, Y. Lu, and H. Zhou. Rate-optimal graphon estimation. ann. statist. *The Annals of Statistics*, 43:2624–2652, 2015.

C. Gao, Y. Lu, H. Zhou, and Z. Ma. Optimal estimation and completion of matrices with biclustering structures. *Journ. Machine Learn. Research*, 17:1–29, 2016.

A. Goldenberg, A. X. Zheng, S. E. Fienberg, and E. M. Airoldi. A survey of statistical network models. *Found. Trends Mach. Learn.*, 2:129–233, 2010. doi: 10.1561/2200000005.

A. Gunawardana and W. Byrne. Convergence theorems for generalized alternating minimization procedures. *Mach. Learn. Res*, 6:2049–2073, 2005.

Q. Han, K. S. Xu, and E. M. Airoldi. Consistent estimation of dynamic and multi-layer block models. In *Proceedings of the 32nd International Conference on International Conference on Machine Learning*, volume 37 of *ICML'15*, page 1511–1520, 2015.

M. Jordan, Z. Ghahramani, T. Jaakkola, and L. Saul. An introduction to variational methods for graphical models. *Mach. Learn.*, 37:183–233, 1999.

X. Kang, A. Ganguly, and E. Kolaczyk. E.d. *Sankhya A*, 84:218–260, 2021.

O. Klopp, A. Tsybakov, and N. Verzelen. Oracle inequalities for network models and sparse graphon estimation. 45:316–354, 2017.

M. Knight, K. Leeming, G. Nason, and M. Nunes. Generalized network autoregressive processes and the gnar package. *Journal of Statistical Software*, 96:1–36, 2020.

E. D. Kolaczyk. *Statistical Analysis of Network Data: Methods and Models*. Springer New York, NY, 2009.

M. Kolar, L. Song, A. Ahmed, and E. P. Xing. Estimating time-varying networks. *The Annals of Applied Statistics*, 4, 2010.

R. Kolar. and E. Xing. Sparsistent estimation of time-varying discrete markov random fields. *ArXiv e-prints.*, 2009.

M. Ludkin, I. Eckley, and P. Neal. Dynamic stochastic block models: parameter estimation and detection of changes in community structure. *Statistics and Computing*, 28:1201–1213, 2018. doi: https://doi.org/10.1007/s11222-017-9788-9.

M. Mandjes, N. Starreveld, R. Bekker, and P. Spreij. Dynamic erdős-rényi graphs. In *Steffen, B., Woeginger, G. (eds) Computing and Software Science. Lecture Notes in Computer Science*, volume 10000. Springer, 2019. doi: https://doi.org/10.1007/978-3-319-91908-9_8.

A. Mantziou, M. Cucuringu, V. Meirinhos, and G. Reinert. The gnar-edge model: a network autoregressive model for networks with time-varying edge weights. *Journal of Complex Networks*, 11, 2023.

C. Matias and V. Miele. Statistical clustering of temporal networks through a dynamic stochastic block model. *J. R. Stat. Soc. Ser. B (Statistical Methodology)*, 79:1119–1141, 2017.

M. Pensky. Dynamic network models and graphon estimation. *The Annals of Statistics*, 47:2378–2403, 2019.

P. Ravikumar, M. J. Wainwright, and J. D. Lafferty. High-dimensional ising model selection using l1-regularized logistic regression. *Ann. Statist.*, 38:1287 – 1319, 2010.

R. A. Rossi and N. K. Ahmed. The network data repository with interactive graph analytics and visualization. In *AAAI*, 2015. URL https://networkrepository.com.

D. I. Rubenstein, S. R. Sundaresan, I. R. Fischhoff, C. Tantipathananandh, and T. Y. Berger-Wolf. Similar but different: dynamic social network analysis highlights fundamental differences between the fission-fusion societies of two equid species, the onager and grevy's zebra. *PLOS One*, 10, 2015.

D. Shizuka, A. S. Chaine, J. Anderson, O. Johnson, I. M. Laursen, and B. E. Lyon. Across-year social stability shapes network structure in wintering migrant sparrows. *Ecol. Lett.*, 17:998–1007, 2014.

T. A. Snijders. Statistical models for social networks. *A. Rev. Sociol.*, 37:129–151, 2011.

K. Xu and A. Hero. Dynamic stochastic blockmodels for time evolving social networks. *IEEE J. Sel. Top. Signal Process.*, 8:552–562, 2014.

T. Yang, Y. Chi, S. Zhu, Y. Gong, and R. Jin. Detecting communities and their evolutions in dynamic social networks—a bayesian approach. *Mach. Learn.*, 82:157–189, 2011.

X. Zhang, C. Moore, and M. E. J. Newman. Random graph models for dynamic networks. *Eur. Phys. J. B*, 90, 2017. doi: https://doi.org/10.1140/epjb/e2017-80122-8.

X. Zhu, R. Pan, G. Li, Y. Liu, and H. Wang. Network vector autoregression. *The Annals of Statistics*, 45:1096–1123, 2017.

X. Zhu, W. Wang, H. Wang, and K. W. Härdle. Network quantile autoregression. *Journal of Econometrics*, 212:345–358, 2019.