IRTG 1792 Discussion Paper 2019-009

Dynamic Network Perspective of Cryptocurrencies

Li Guo [*]
Yubo Tao [*]
Wolfgang K. Härdle [*2]

[*] Singapore Management University
[*2] Humboldt-Universität zu Berlin

This research was supported by the Deutsche Forschungsgesellschaft through the International Research Training Group 1792 "High Dimensional Nonstationary Time Series".

http://irtg1792.hu-berlin.de
ISSN 2568-5619

A Dynamic Network Perspective of Cryptocurrencies*

Li Guo†
Lee Kong Chian School of Business, Singapore Management University

Yubo Tao‡
School of Economics, Singapore Management University

Wolfgang Karl Härdle§
Center for Applied Statistics and Economics, Humboldt-Universität zu Berlin
Sim Kee Boon Institute for Financial Economics, Singapore Management University
Wang Yanan Institute for Studies in Economics, Xiamen University
Department of Mathematics and Physics, Charles University

This Draft: March 2019

Abstract

Cryptocurrencies are becoming an attractive asset class and are the focus of recent quantitative research. The joint dynamics of the cryptocurrency market yields information on network risk. Utilizing the adaptive LASSO approach, we build a dynamic network of cryptocurrencies and model the latent communities with a dynamic stochastic blockmodel. We develop a dynamic covariate-assisted spectral clustering method to uniformly estimate the latent group membership of cryptocurrencies consistently. We show that return inter-predictability and crypto characteristics, including hashing algorithms and proof types, jointly determine the crypto market segmentation. Based on this classification result, it is natural to employ eigenvector centrality to identify a cryptocurrency's idiosyncratic risk. An asset pricing analysis finds that a cross-sectional portfolio with a higher centrality earns a higher risk premium. Further tests confirm that centrality serves as a risk factor well and delivers valuable information content on cryptocurrency markets.

Keywords: Community Detection, Dynamic Stochastic Blockmodel, Spectral Clustering, Node Covariate, Return Predictability, Portfolio Management.

*Li Guo gratefully acknowledges all participants who attended the workshop "Crypto-Currencies in a Digital Economy" at the Humboldt-Universität zu Berlin for their helpful discussion and comments. Yubo Tao would like to thank Xiaoyi Han, Shuyang Sheng, Yichong Zhang, and the participants of Econometric Session 16 at the 2018 China Meeting of the Econometric Society for their insightful suggestions. Wolfgang K. Härdle acknowledges financial support from IRTG 1792 "High Dimensional Non-stationary Time Series", Humboldt-Universität zu Berlin and that of the Czech Science Foundation under grant no. 19-28231X.

†Address: 50 Stamford Rd, Singapore 178899. Email: liguo.2014@pbs.smu.edu.sg.

‡Address: 90 Stamford Rd, Singapore 178903. Email: yubo.tao.2014@phdecons.smu.edu.sg.

§Correspondence author. Address: Unter den Linden 6 10099 Berlin, Germany. Email: haerdle@hu-berlin.de.

1 Introduction

The invention of Bitcoin (Nakamoto, 2008) spurred the creation of many cryptocurrencies (cryptos hereafter) commonly known as *Altcoins*. As of December 31, 2018, more than 1500 cryptos are actively traded worldwide, with a market capitalization of more than 200 billion USD. The growing number of Altcoins led investors to investigate interrelationships between Altcoins to make a profit. Unlike stocks that we can group into different industries by GIC or SIC, there are no stringent criteria to classify cryptos. By virtue of network analysis, we develop a covariate-assisted spectral clustering (CASC) method that accommodates important network features such as connection sparsity, degree heterogeneity, and relation asymmetry, to study the interrelationships between cryptos systematically. We thereby provide a novel angle to study the market segmentation problem of cryptos and other financial instruments.

The crypto market is distinct from the equity market in various aspects, which hinders the application of traditional classification methodology. Given that both cryptos and stocks are traded at high frequency, return information is particularly important as it serves as timely information to understand the dynamics of the market structure. According to market efficiency, the covariance between the prices of speculative assets cannot exceed the covariance between their fundamental information. Consequently, in the equity market, return co-movement is frequently adopted to project the fundamental similarity between stocks. However, excess return co-movement has been widely documented in the literature (see, e.g., Kumar and Lee, 2006; Boyer, 2011) and it is more significant in the crypto market given the strong behavioral bias of market participants and high information uncertainty of its future cash flows. Inspired by Hoberg and Phillips (2016), who ameliorate industry classification by studying a set of dynamic industry structures generated from product differentiation and competition, we use crypto's contract information to help identify the fundamental similarities between cryptos. In particular, we extract the most fundamental characteristics of each mining contract, that is, the cryptographic algorithm and proof types, as additional input for clustering analysis. As we expected, our method shows superior classification accuracy over state-of-the-art methods available in the literature. In particular, cryptos in the same group show stronger return co-movement than the cross-group return co-movement across all empirical settings. Moreover, within-group cryptos show stronger connections in algorithms and proof types than cross-group cryptos do.

To understand the economic meaning of the latent group structure, we conduct several tests to verify the asset pricing implication of the grouping results. Acemoglu et al. (2012) proposed a theoretical framework to model spillover effects through sector-level shocks. The model suggests that if the linkages in the inter-sectoral network are sufficiently asymmetric, then sectoral shocks might not cancel out through diversification, but aggregate into macroeconomic fluctuations. Ahern (2013) further pointed out that the stocks with higher incoming linkages tend to receive more shocks from linked stocks and thus require a higher risk premium. Motivated by these results, we construct a cross-sectional portfolio by sorting on group centrality and show that high-centrality cryptos require a higher risk premium than the low-centrality ones. Next, we investigate whether other factors such as liquidity (Amihud and Mendelson, 1986), investor attention (Liu and Tsyvinski, 2018), and macro uncertainty (Baker et al., 2016) could possibly explain this augmented risk premium. Our results suggest that the return predictability of centrality survives after controlling for all of these factors. Hence, it provides an important empirical implication for both academic studies and participants in the crypto market.

This paper develops statistical theory for dynamic networks and thereby makes several important contributions to classical finance as well as FinTech. First, we offer a network angle to study the crypto market by connecting cryptos according to their inter-predictive relationship estimated by adaptive LASSO. Second, we provide a new set of quantitative tools to study crypto market segmentation that can be applied to a wide variety of assets. Specifically, we extend the static spectral clustering methods (Binkiewicz et al., 2017; Zhang et al., 2018, among others,) to identify communities in dynamic networks with both time-evolving membership and node covariates. To make full use of the relevant information, we address the challenges of the features of the real data, namely, time dependency, degree heterogeneity, sparsity, and node covariates. Our proposed community detection method can resolve the aforementioned data issues. The methodology we present can also be extended to cover more asset-specific characteristics to achieve higher classification accuracy.

In addition, we deepen the understanding of the crypto market in terms of both market segmentation and portfolio management. Intensive research in this area considers asset pricing inferences from different angles, but there is limited work that shows the economic link between crypto fundamentals and its performance. Härdle et al. (2019) suggest crypto dynamics as an extraordinary research opportunity for academia and provide some insights

into the mechanics of this market. Härdle and Trimborn (2015) construct the CRIX (thecrix.de), a market index consisting of a selection of cryptos representative of the whole crypto market. Given the low liquidity in the current Altcoin market compared to traditional assets, Trimborn et al. (2019) propose a Liquidity Bounded Risk-return Optimization (LIBRO) approach that accounts for liquidity issues by studying the Markowitz framework under liquidity constraints. Chen et al. (2018) propose an option pricing technique for cryptos based on a stochastic volatility model with correlated jumps. Lee et al. (2018) compare cryptos with traditional asset classes and find that cryptos provide additional diversification to mainstream assets, hence improving the portfolio performance. Petukhina et al. (2018) characterize the effects of adding cryptos to the set of traditional eligible assets in portfolio management and find that cryptos can significantly improve the risk-return profile of mainstream asset portfolios. Our results provide new insights into the fundamentals of the crypto market structure by dividing them into different groups. We find that cryptos' fundamentals have very different features from those of traditional assets, and these features indeed affect a crypto's price evolution.

The remainder of the paper is organized as follows. In section 2, we introduce the model and method to estimate the dynamic group structure and demonstrate the effectiveness of our method via simulation. In section 3, we employ our method to identify the latent group structure of cryptos and provide its economic interpretation. Then, in section 4, we check the time series and cross-sectional return predictability and demonstrate its portfolio implications. We conclude in section 5. All proofs and technical details are provided in the supplement. R codes to implement the algorithms are available at QuantNet (quantlet.de) by searching the keyword "CASC."

2 Models and Methodology

In the equity market, network structures are powerful in revealing risk percolations in assets such as firms, industries, and financial instruments (Cohen and Frazzini, 2008; Aobdia et al., 2014; Acemoglu et al., 2015; Chen et al., 2019, see, e.g.,). The latest study, Herskovic (2018), constructs a sector level network based on the Bureau of Economic Analysis (BEA) Input-Output Accounts. Here, we borrow the network idea to model the interdependencies in between cryptos, such as technological similarities and return co-movements. However,

just applying a network view on cryptos will not give us any insights into the dominant elements of the market. We therefore represent the adjacency matrices stochastically via a block structure to identify the latent communities. To build such a stochastic blockmodel with time-varying communities, we need to establish a more advanced methodology to identify group memberships. Based on adaptive LASSO in a 60-day rolling window, we generate a time series of adjacency matrices. By imposing an assumption on the switch in group memberships, we can uniformly identify communities consistently. We base our numerical implementation of this procedure on spectral clustering. Binkiewicz et al. (2017) show that the classification accuracy of the spectral clustering method can be improved by introducing covariate assistance. Here, we present an extension of the static covariate-assisted spectral clustering (CASC) algorithm to deal with the dynamic stochastic blockmodel and co-blockmodel. The theoretical justification and simulations also demonstrate the consistency of this method.

2.1 Dynamic network model with covariates

2.1.1 Undirected network

Consider a dynamic network defined as a sequence of random undirected graphs with N nodes, $G_{N,t}$, $t = 1, \cdots, T$, on the vertex set $V_N = \{v_1, v_2, \cdots, v_N\}$, which does not change over horizons. For each period, we model the unipartite network structure with the *spectral-contextualized degree-corrected stochastic blockmodel* (SC-DCBM) introduced by Zhang et al. (2018). Specifically, we generate the adjacency matrices A_t by

$$A_t(i,j) = \begin{cases} \text{Bernoulli}\{P_t(i,j)\}, & \text{if } i < j \\ 0, & \text{if } i = j \\ A_t(j,i), & \text{if } i > j \end{cases} \tag{1}$$

where $P_t(i,j) = \Pr\{A_t(i,j) = 1\}$. To reflect the group structure, the probabilities of a connection $P_t(i,j)$ at period t are blocked. In particular, denote $z_{i,t}$ as the group label of node i at time t; then, if $z_{i,t} = k$ and $z_{j,t} = k'$, then $P_t(i,j) = B_t(z_{i,t}, z_{j,t}) = B_t(k,k')$. Hence, for any $t = 1, \cdots, T$, we can obtain the population adjacency matrix

$$\mathcal{A}_t \overset{\text{def}}{=} \mathsf{E}(A_t) = Z_t B_t Z_t^\top, \tag{2}$$

where $Z_t \in \{0,1\}^{N\times K}$ is the *clustering matrix* such that there is only one 1 in each row and at least one 1 in each column.

Since the conventional stochastic blockmodel presumes that each node in the same group should have the same expected degrees, following Karrer and Newman (2011), we introduce the *degree parameters* $\psi = (\psi_1, \cdots, \psi_N)$ to capture the degree heterogeneity of the groups. In particular, the edge probability between node i and j at time t is

$$P_t(i,j) = \psi_i \psi_j B_t(z_{i,t}, z_{j,t}), \tag{3}$$

with the identifiability restriction

$$\sum_{i\in\mathcal{G}_k} \psi_i = 1, \quad \forall k \in \{1,2,\cdots,K\}. \tag{4}$$

where $\mathcal{G}_k$ is the set of nodes that belongs to the kth group. Denote Diag(ψ) by Ψ. The population adjacency matrices for the dynamic SC-DCBM is then:

$$\mathcal{A}_t = \Psi Z_t B_t Z_t^\top \Psi, \tag{5}$$

Define the regularized graph Laplacian as

$$L_{\tau,t} = D_{\tau,t}^{-1/2} A_t D_{\tau,t}^{-1/2}, \tag{6}$$

where $D_{\tau,t} = D_t + \tau_t I$ and D_t is a diagonal matrix with $D_t(i,i) = \sum_{j=1}^N A_t(i,j)$. As Chaudhuri et al. (2012) shows, regularization improves the spectral clustering performance, especially for sparse networks. We fix τ_t as the value of average node degree, that is, $\tau_t = N^{-1}\sum_{i=1}^N D_t(i,i)$.

Recent developments suggest that using node features or covariates can greatly improve classification accuracy. For example, Binkiewicz et al. (2017) add the covariance $XX^\top$, with $X \in [-J,J]^R$ being the node covariate matrix, to the regularized graph Laplacian and perform the spectral clustering on the static *similarity matrix*. We extend the static similarity matrix to cover the dynamic case below:

$$S_t = L_{\tau,t} + \alpha_t C. \tag{7}$$

where $C = XX^\top$ and $\alpha_t \in [0, \infty)$ is a tuning parameter that controls the informational balance between $L_{\tau,t}$ and X in the leading eigenspace of S_t. As a generalization of the model, Zhang et al. (2018) refines this by replacing C with $C_w = XWX^\top$, where W is some weight matrix. Finally, we substitute C with the new covariate-assisted component $C_t^w = XW_tX^\top$, and the population similarity matrix now becomes

$$\mathcal{S}_t = \mathcal{L}_{\tau,t} + \alpha_t \mathcal{C}_t^w, \tag{8}$$

where $\mathcal{L}_{\tau,t} = \mathcal{D}_{\tau,t}^{-1/2}\mathcal{A}_t\mathcal{D}_{\tau,t}^{-1/2}$ and $\mathcal{C}_t^w = \mathcal{X}\mathcal{W}_t\mathcal{X}$.

The setup in (8) addresses several extensions of existing methods. First, W_t creates a time-varying interaction between different covariates. For instance, we may think of different refined algorithms that stem from the same origins. Such inheritance relationships will potentially lead to an interaction between the cryptos. In addition, over time, some algorithms may become more popular while the others may near extinction. Thus, this interaction would also change over time. These interactions are not included in C.

Second, we can easily select covariates by setting certain elements of W_t to zero. This is necessary as it helps us to model the evolution of technologies. At some point in time, some cryptographic technology may be eliminated due to upgrades or cracking. Therefore, W_t offers us the flexibility to exclude covariates, which we cannot do easily with C.

Lastly, the role of C is to link similarity in covariates to a high probability of node connection. However, this is questionable in crypto networks. Due to the open source nature of the blockchain, crypto developers can easily copy and paste the source code and launch a new coin without any costs. Consequently, we observe a high degree of homogeneity in the crypto market. However, this homogeneity does not necessarily result in a co-movement of prices: some cryptos are negatively correlated. In this case, we may set $W_t(i,i)$ to be negative and C_t^w will eventually bring the cryptos with different technologies closer in the similarity matrix.

2.1.2 Directed network

To model the dynamic block structure in a directed network, we employ the dyanmic spectral-contextualized degree-corrected stochastic co-blockmodel (SC-DCcBM). For a di-

rected network, the adjacency matrix A_t is not necessarily symmetric; that is,

$$A_t(i,j) = \begin{cases} \text{Bernoulli}\{P_t(i,j)\}, & \text{if } i \neq j \\ 0, & \text{if } i = j \end{cases} \tag{9}$$

Similarly, define the regularized graph Laplacian $L_{\tau,t} \in \mathbb{R}^{N \times N}$ for the directed network as

$$L_{\tau,t} = D_{R,t}^{-1/2} A_t D_{C,t}^{-1/2}, \tag{10}$$

where $D_{R,t}$ and $D_{C,t}$ are diagonal matrices with $D_{R,t}(i,i) = \sum_{j=1}^{N} A_t(i,j) + \tau_{R,t}$ and $D_{C,t}(i,i) = \sum_{j=1}^{N} A_t(j,i) + \tau_{C,t}$, where $\tau_{R,t}$ and $\tau_{C,t}$ are set to be the average row and column degrees at each period, respectively.

We now include the node covariates by constructing a similarity matrix from regularized graph Laplacian $L_{\tau,t}$ and covariate matrix X in the same way as in an undirected network; that is, for each $t = 1, \cdots, T$,

$$S_t = L_{\tau,t} + \alpha_t X W_t X^\top = D_{R,t}^{-1/2} A_t D_{C,t}^{-1/2} + \alpha_t X W_t X^\top, \tag{11}$$

where $\alpha_t \in [0, \infty)$ is the tuning parameter. Then, let $Z_{R,t} \in \{0,1\}^{N_R \times K_R}$ and $Z_{C,t} \in \{0,1\}^{N_C \times K_C}$, such that there is only one 1 in each row and at least one 1 in each column. Let the block probability matrix in each period be $B_t \in [0,1]^{K_R \times K_C}$ with rank $K = \min\{K_R, K_C\}$. Then, the population adjacency matrix is

$$\mathcal{A}_t = \mathsf{E}(A_t) = Z_{R,t} B_t Z_{C,t}^\top, \tag{12}$$

and the population regularized graph Laplacian is

$$\mathcal{L}_{\tau,t} = \mathcal{D}_{R,t}^{-1/2} \mathcal{A}_t \mathcal{D}_{C,t}^{-1/2}. \tag{13}$$

Therefore, the population similarity matrix is

$$\mathcal{S}_t = \mathcal{L}_{\tau,t} + \alpha_t \mathcal{X} \mathcal{W}_t \mathcal{X}^\top. \tag{14}$$

By construction, we know $D_{R,t}(i,i) = \sum_{j=1}^{N} \mathbb{1}_{\{i \to j\}}^{(t)} + \tau_{R,t}$, which controls for the number

of the parents of node j, and $D_{C,t}(i,i) = \sum_{i=1}^{N} \mathbb{1}^{(t)}_{\{j\to i\}} + \tau_{C,t}$, which controls the number of the offspring of node j. To analyze the asymmetric adjacency matrix A_t caused by directional information, Rohe et al. (2016) propose using the *singular value decomposition* instead of eigen-decomposition for the regularized graph Laplacian. The intuition behind this methodology is to use both the eigenvectors of $L_{\tau,t}^{\top}L_{\tau,t}$ and $L_{\tau,t}L_{\tau,t}^{\top}$, which contains information about "the number of common parents" and "the number of common offspring"; that is, for each $t = 1, \cdots, T$,

$$(L_{\tau,t}^{\top}L_{\tau,t})_{ab} = \sum_{i=1}^{N} L_{\tau,t}(i,a)L_{\tau,t}(i,b) = \frac{1}{\sqrt{D_{C,t}(a,a)D_{C,t}(b,b)}} \sum_{i=1}^{N} \frac{\mathbb{1}^{(t)}_{\{i\to a \text{ and } i\to b\}}}{D_{R,t}(i,i)},$$
$$(L_{\tau,t}L_{\tau,t}^{\top})_{ab} = \sum_{i=1}^{N} L_{\tau,t}(a,i)L_{\tau,t}(b,i) = \frac{1}{\sqrt{D_{R,t}(a,a)D_{R,t}(b,b)}} \sum_{i=1}^{N} \frac{\mathbb{1}^{(t)}_{\{a\to i \text{ and } b\to i\}}}{D_{C,t}(i,i)}.$$

2.2 Dynamic CASC

To set up a dynamic CASC, we face two major difficulties: (i) defining W_t and (ii) estimating the similarity matrix with dynamic network information. For the first issue, we follow Zhang et al. (2018) by setting $W_t = X^{\top}L_{\tau,t}X$, which measures the correlation between covariates along the graph. For the second issue, we follow Pensky and Zhang (2017) by constructing the estimator of $\mathcal{S}_t$ with a discrete kernel to bring in historical network information. Klochkov et al. (2019) present a similar idea. Specifically, we first pick an integer $r \geq 0$, obtain two sets of integers

$$\mathcal{F}_r = \{-r, \cdots, 0\}, \quad \mathcal{D}_r = \{T - r + 1, \cdots, T\},$$

and assume that $|W_{r,l}(i)| \leq W_{\max}$, where $W_{\max}$ is independent of r and i, and satisfies

$$\frac{1}{|\mathcal{F}_r|} \sum_{i\in\mathcal{F}_r} i^k W_{r,l}(i) = \begin{cases} 1, & \text{if } k = 0, \\ 0, & \text{if } k = 1, 2, \cdots, l. \end{cases} \tag{15}$$

Obviously, the $W_{r,l}$ is a discretized version of the continuous boundary kernel that weighs only the historical observations. This kernel assigns more recent similarity matrices higher scores. To choose an optimal bandwidth r, Pensky and Zhang (2017) propose an adaptive estimation procedure using Lepski et al. (1997)'s method. Here, we also employ

their method and construct the estimator for edge connection matrices:

$$\widehat{\mathcal{S}}_{t,r} = \frac{1}{|\mathcal{F}_r|} \sum_{i \in \mathcal{F}_r} W_{r,l}(i) S_{t+i}. \tag{16}$$

Once we obtain $\widehat{\mathcal{S}}_{t,r}$, we create an eigen-decomposition of $\widehat{S}_{t,r} = \widehat{U}_t \widehat{\Lambda}_t \widehat{U}_t^\top$ for each $t = 1, 2, \cdots, T$. As Lei and Rinaldo (2015) discuss, the matrix $\widehat{U}_t$ may now have more than K distinct rows due to the degree correction, whereas the rows of $\widehat{U}_t$ still only point to at most K directions. Therefore, we apply the spherical clustering algorithm to find a cluster structure among the rows of the normalized matrix $\widehat{U}_t^+$ with $\widehat{U}_t^+(i,*) = \widehat{U}_t(i,*)/\|\widehat{U}_t(i,*)\|$. More specifically, we consider the following spherical k-means spectral clustering:

$$\left\| \widehat{Z}_t^+ \widehat{Y}_t - \widehat{U}_t^+ \right\|_F^2 \leq (1+\varepsilon) \min_{\substack{Z_t^+ \in \mathcal{M}_{N_+,K} \\ Y_t \in \mathbb{R}^{K \times K}}} \left\| Z_t^+ Y_t - \widehat{U}_t^+ \right\|_F^2 \tag{17}$$

where Y_t is some rotation matrix. In the last step, we extend $\widehat{Z}_t^+$ to obtain $\widehat{Z}_t$ by adding $N - N_+$ canonical unit row vectors at the end. $\widehat{Z}_t$ is the estimate of Z_t from this method. We summarize the algorithm in detail below.

Algorithm 1: CASC in the Dynamic SC-DCBM

Input : Adjacency matrices A_t for $t = 1, \cdots, T$;
Covariates matrix X;
Number of communities K;
Approximation parameter ε.

Output: Membership matrices Z_t for any $t = 1, \cdots, T$.

1 Calculate regularized graph Laplacian $L_{\tau,t}$ and weight matrix W_t.
2 Estimate $\mathcal{S}_t$ by $\widehat{\mathcal{S}}_{t,r}$ as in (16).
3 Let $\widehat{U}_t \in \mathbb{R}^{N \times K}$ be a matrix representing the first K eigenvectors of $\widehat{\mathcal{S}}_{t,r}$.
4 Let N_+ be the number of nonzero rows of $\widehat{U}_t$. Then, obtain $\widehat{U}^+ \in \mathbb{R}^{N_+ \times K}$ consisting of normalized nonzero rows of $\widehat{U}_t$; that is, $\widehat{U}_t^+(i,*) = \widehat{U}_t(i,*) / \left\| \widehat{U}_t(i,*) \right\|$ for i such that $\left\| \widehat{U}_t(i,*) \right\| > 0$.
5 Apply the $(1+\varepsilon)$-approximate k-means algorithm to the row vectors of $\widehat{U}_t^+$ to obtain $\widehat{Z}_t^+ \in \mathcal{M}_{N_+,K}$.
6 Extend $\widehat{Z}_t^+$ to obtain $\widehat{Z}_t$ by arbitrarily adding $N - N_+$ canonical unit row vectors at the end, such as $\widehat{Z}_t(i) = (1, 0, \cdots, 0)$ for i such that $\left\| \widehat{U}_t(i,*) \right\| = 0$.
7 Output $\widehat{Z}_t$.

Similar to the dynamic SC-DCBM case, we estimate the block structure of the dy-

namic SC-DCcBM by analyzing the normalized singular vectors on both sides. Then, using the spherical k-means analysis, we can also obtain the clustering matrices. The spectral clustering algorithm for the dynamic SC-DCcBM is below.

Algorithm 2: CASC in the Dynamic SC-DCcBM

Input : Adjacency matrices A_t for $t = 1, \cdots, T$;
Covariates matrix X;
Number of row clusters K_R and number of column clusters K_C;
Approximation parameter ε.

Output: Membership matrices of rows and columns $Z_{R,t}$ and $Z_{C,t}$ for $t = 1, \cdots, T$.

1 Calculate regularized graph Laplacian $L_{\tau,t}$.

2 Estimate $\mathcal{S}_t$ by $\widehat{\mathcal{S}}_{t,r}$ as in (16).

3 Compute the singular value decomposition of $\widehat{\mathcal{S}}_{t,r} = U_t \Sigma_t V_t^\top$ for $t = 1, \cdots, T$.

4 Extract the first K columns of U_t and V_t that correspond to the K largest singular values in Σ_t, where $K = \min\{K_R, K_C\}$. Denote the resulting matrices $U_t^K \in \mathbb{R}^{N \times K}$ and $V_t^K \in \mathbb{R}^{N \times K}$.

5 Let N_+^R be the number of nonzero rows of U_t^K; then, obtain $U_{t+}^K \in \mathbb{R}^{N_+^R \times K}$ consisting of normalized nonzero rows of U_{t+}^K; that is, $U_{t+}^K(i,*) = U_t^K(i,*) / \left\|U_t^K(i,*)\right\|$ for i such that $\left\|U_t^K(i,*)\right\| > 0$.

6 Similarly, let N_+^C be the number of nonzero rows of V_t^K; then, obtain $V_{t+}^K \in \mathbb{R}^{N_+^C \times K}$ consisting of normalized nonzero rows of V_{t+}^K; that is, $V_{t+}^K(i,*) = V_t^K(i,*) / \left\|V_t^K(i,*)\right\|$ for i such that $\left\|V_t^K(i,*)\right\| > 0$.

7 Apply the $(1+\varepsilon)$-approximate k-means algorithm to cluster the rows (columns) of $\widehat{\mathcal{S}}_t$ into K_R (K_C) clusters by treating each row of U_{t+}^K (V_{t+}^K) as a point in $\mathbb{R}^K$ to obtain $\widehat{Z}_{R,t}^+$ ($\widehat{Z}_{C,t}^+$).

8 Extend $\widehat{Z}_{R,t}^+$ ($\widehat{Z}_{C,t}^+$) to obtain $\widehat{Z}_{R,t}$ ($\widehat{Z}_{C,t}$) by arbitrarily adding $N - N_+^R$ ($N - N_+^C$) canonical unit row vectors at the end, such as $\widehat{Z}_{R,t}(i) = (1, 0, \cdots, 0)$ ($\widehat{Z}_{C,t}(i) = (1, 0, \cdots, 0)$) for i such that $\|U_t(i,*)\| = 0$ ($\|V_t(i,*)\| = 0$).

9 Output $\widehat{Z}_{R,t}$ and $\widehat{Z}_{C,t}$.

2.3 Uniform consistency

2.3.1 Undirected case

In the subsequent analysis, we illustrate that the dynamic CASC is uniformly consistent over time for both undirected and directed networks. We first make some assumptions on the graph that generates the dynamic network. The major assumption we need here is *assortativity*, which ensures that the nodes within the same cluster are more likely to share an edge than nodes in two different clusters.

Assumption 1. *The dynamic network is composed of a series of assortative graphs that are generated under the stochastic blockmodel with covariates whose block probability matrix B_t is positive definite for all $t = 1, \cdots, T$.*

Intuitively, the more frequent the group membership changes, the less stable the network will be. Consequently, it becomes harder to make use of the information from the historical and future network structures to detect the communities in the present network structure. In Assumption 2, we restrict the maximum number of nodes that switch memberships (s) to some finite number. Based on this assumption, the proportion of nodes that switch their memberships shrinks to 0 as the size of the network grows to infinity. Additionally, we can easily bound the dynamic behavior of clustering matrices ($Z_{t+r} - Z_t$) by noting that there are at most rs nonzero rows in the differenced matrix.

Assumption 2. *At most, $s < \infty$ number of nodes can switch their memberships between any consecutive time instances.*

Assumption 3. *For $1 \leq k \leq k' \leq K$, there exists a function $f(\cdot; k, k')$ such that $B_t(k, k') = f(\varsigma_t; k, k')$ and $f(\cdot; k, k') \in \Sigma(\beta, L)$, where $\Sigma(\beta, L)$ is a Hölder class of functions $f(\cdot)$ on $[0, 1]$ such that $f(\cdot)$ are ℓ times differentiable and*

$$|f^{(\ell)}(x) - f^{(\ell)}(x')| \leq L|x - x'|^{\beta-\ell}, \text{ for any } x, x' \in [0, 1], \tag{18}$$

with ℓ being the largest integer smaller than β.

Assumption 3 states that neither the connection probabilities nor the cluster memberships change drastically over the horizons. Lastly, to guarantee the performance of our clustering method, we impose some conditions to regularize the behavior of the covariate matrix and the eigenvalues of the similarity matrices.

Assumption 4. *Let $\lambda_{1,t} \geq \lambda_{2,t} \geq \cdots \geq \lambda_{K,t} > 0$ be the K largest eigenvalues of $\mathcal{S}_t$ for each $t = 1, \cdots, T$. In addition, assume that*

$$\underline{\delta} = \inf_t \{\min_i \mathcal{D}_{\tau,t}(i, i)\} > 3\log(8NT/\epsilon) \quad \text{and} \quad \alpha_{\max} = \sup_t \alpha_t \leq \frac{a}{NRJ^2\xi},$$

with

$$a = \sqrt{\frac{3\log(8NT/\epsilon)}{\underline{\delta}}} \quad \text{and} \quad \xi = \max(\sigma^2\|L_\tau\|_F\sqrt{\log(TR)}, \sigma^2\|L_\tau\|\log(TR), NRJ^2/\underline{\delta}),$$

where $\sigma = \max_{i,j} \|X_{ij} - \mathcal{X}_{ij}\|_{\phi_2}$, $L_\tau = \sup_t L_{\tau,t}$.

To establish the consistency of the CASC for the dynamic SC-DCBM, we need to determine the upper bounds for the misclustering rates. Following Binkiewicz et al. (2017), we denote $C_{i,t}$ and $\mathcal{C}_{i,t}$ as the cluster centroids of the ith node at time t generated using k-means clustering on the sample eigenvector U_t and the population $\mathcal{U}_t$, respectively. Then, we define the set of mis-clustered nodes at each period as

$$\mathbb{M}_t = \left\{ i\colon \left\| C_{i,t}\mathcal{O}_t^\top - \mathcal{C}_{i,t} \right\| > \left\| C_{i,t}\mathcal{O}_t^\top - \mathcal{C}_{j,t} \right\|, \text{ for any } j \neq i \right\}, \tag{19}$$

where $\mathcal{O}_t$ is a rotation matrix that minimizes $\|U_t\mathcal{O}_t^\top - \mathcal{U}_t\|_F$ for each $t = 1, \cdots, T$.

The misclustering error in $\mathbb{M}_t$ has two sources: the estimation error of $\mathcal{S}_t$ using the discrete kernel estimator and from spectral clustering. In Theorem 1, we provide the uniform upper bound of the misclustering rate for the undirected and directed networks separately.

Theorem 1. *Let clustering proceed according to Algorithm 1 based on the estimator* $\widehat{\mathcal{S}}_{t,r}$ *of* $\mathcal{S}_t$. *Let* $Z_t \in \mathcal{M}_{N,K}$ *and* $P_{\max} = \max_{i,t}(Z_t^\top Z_t)_{ii}$ *denote the size of the largest block over the horizons. Then, under Assumptions 1-4, the misclustering rate satisfies*

$$\sup_t \frac{|\mathbb{M}_t|}{N} \leq \frac{c_1(\varepsilon)KW_{\max}^2}{m_z^2 N\lambda_{K,\max}^2}\left\{(6+c_w)\frac{b}{\underline{\delta}^{1/2}} + \frac{2K}{\underline{\delta}}(\sqrt{2P_{\max}rs} + 2P_{\max}) + \frac{NL}{\underline{\delta}\cdot l!}\left(\frac{r}{T}\right)^\beta\right\}^2.$$

with a probability of at least $1-\epsilon$, *where* $c_1(\varepsilon) = 2^9(2+\varepsilon)^2$, $b = \sqrt{3\log(8NT/\epsilon)}$, *and* $\lambda_{K,\max} = \max_t\{\lambda_{K,t}\}$ *with* $\lambda_{K,t}$ *being the* K*th largest absolute eigenvalue of* $\mathcal{S}_t$.

2.3.2 Directed case

Analogous to the undirected case, we modify Assumption 4 to accommodate the stochastic co-blockmodel setup.

Assumption 4'. *Let* $\lambda_{1,t} \geq \lambda_{2,t} \geq \cdots \geq \lambda_{K,t} > 0$ *be the* $K = \min K_R, K_C$ *largest singular values of* $\mathcal{S}_t$ *for each* $t = 1, \cdots, T$. *In addition, assume that*

$$\underline{\delta}' = \inf_t\{\min\{\min_i \mathcal{D}_{R,t}(i,i), \min_i \mathcal{D}_{C,t}(i,i)\}\} > 3\log(16NT/\epsilon)$$

and

$$\alpha_{\max} = \sup_t \alpha_t \leq \frac{a}{NRJ^2\xi},$$

with $a = \sqrt{\frac{3\log(16NT/\epsilon)}{\underline{\delta}'}}$ *and* $\xi = \max(\sigma^2\|L_\tau\|_F\sqrt{\log(TR)}, \sigma^2\|L_\tau\|\log(TR), NRJ^2/\underline{\delta}')$, *where* $\sigma = \max_{i,j}\|X_{ij} - \mathcal{X}_{ij}\|_{\phi_2}$, $L_\tau = \sup_t L_{\tau,t}$.

Following Rohe et al. (2016), we define the "R-mis-clustered" and "C-mis-clustered" vertices as

$$\mathbb{M}_t^p = \left\{i\text{: } \left\|C_{i,t}^p - \mathcal{C}_{i,t}^p\mathcal{O}_t^p\right\| > \left\|C_{i,t}^p - \mathcal{C}_{j,t}^p\mathcal{O}_t^p\right\|, \text{ for any } j \neq i\right\}, \quad p \in \{R, C\}, \tag{20}$$

where $C_{i,t}^p$ and $\mathcal{C}_{i,t}^p$ for $p \in \{R, C\}$ are the cluster centroids of the ith node at time t generated using the k-means clustering on the left/right singular vectors and the population left/right singular vectors, respectively.

Theorem 2. *Assuming* $K_R \leq K_C$, *let* $Z_{R,t} \in \mathcal{M}_{N,K_R}$, $Z_{C,t} \in \mathcal{M}_{N,K_C}$, *and* $P_{\max} = \max\{\max_{i,t}(Z_{R,t}^\top Z_{R,t})_{ii}, \max_{i,t}(Z_{C,t}^\top Z_{C,t})_{ii}\}$ *denote the size of the largest block over the horizons. Then, under Assumptions 1-3 and 4', the misclustering rate satisfies*

$$\sup_t \frac{\left|\mathbb{M}_t^R\right|}{N} \leq \frac{c_2(\varepsilon)KW_{\max}^2}{m_r^2 N\lambda_{K,\max}^2}\left\{(6+c_w')\frac{b'}{\underline{\delta}'^{1/2}} + \frac{2K_C}{\underline{\delta}'}(\sqrt{2P_{\max}rs} + 2P_{\max}) + \frac{NL}{\underline{\delta}'\cdot\ell!}\left(\frac{r}{T}\right)^\beta\right\}^2,$$

$$\sup_t \frac{\left|\mathbb{M}_t^C\right|}{N} \leq \frac{c_3(\varepsilon)KW_{\max}^2}{m_c^2 N\gamma_c^2\lambda_{K,\max}^2}\left\{(6+c_w')\frac{b'}{\underline{\delta}'^{1/2}} + \frac{2K_C}{\underline{\delta}'}(\sqrt{2P_{\max}rs} + 2P_{\max}) + \frac{NL}{\underline{\delta}'\cdot\ell!}\left(\frac{r}{T}\right)^\beta\right\}^2,$$

with a probability of at least $1-\epsilon$, *where* $c_2(\varepsilon) = 2^6(2+\varepsilon)^2$, $c_3(\varepsilon) = 2^7(2+\varepsilon)^2$, $b' = \sqrt{3\log(16NT/\epsilon)}$, γ_c *are defined in supplement equation (44), and* $\lambda_{K,\max} = \max_t\{\lambda_{K,t}\}$ *with* $\lambda_{K,t}$ *being the* K*th largest absolute singular value of* $\mathcal{S}_t$.

2.4 Choice of tuning parameters

Obviously, we must choose the tuning parameters r, α, and K carefully. For the choice of r, we first need to determine the upper bound of the variance proportion of the estimation error $\|\widehat{\mathcal{S}}_{t,r} - \mathcal{S}_t\|$, which is $\|\widehat{\mathcal{S}}_{t,r} - \mathcal{S}_{t,r}\|$. In the following lemma, we derive a sharp probabilistic upper bound on $\|\widehat{\mathcal{S}}_{t,r} - \mathcal{S}_{t,r}\|$ using the device provided in Lei and Rinaldo (2015).

Lemma 1. *Let* $d = rN\|\mathcal{S}_t\|_\infty$ *and* $\eta \in (0,1)$. *Then,*

$$\|\widehat{\mathcal{S}}_{t,r} - \mathcal{S}_{t,r}\| \leq (1-\eta)^{-2}\frac{W_{\max}\sqrt{d}}{r \vee 1},$$

with probability $1-\epsilon$*, where* $\epsilon = N^{\left(\frac{3}{16\|\mathcal{S}_t\|_\infty}-2\log\left(\frac{7}{\eta}\right)\right)}$.

From Lemma 1 and the proofs of the previous theorems, we can see that $\|\widehat{\mathcal{S}}_{t,r}-\mathcal{S}_{t,r}\|$ is decreasing, while $\|\mathcal{S}_{t,r}-\mathcal{S}_t\|$ is increasing in r. Therefore, there exists an optimal r^* that achieves the best bias-variance balance; that is,

$$r^* = \arg\min_{0\le r\le T/2}\left((1-\eta)^{-2}\frac{W_{\max}\sqrt{d}}{r\vee 1}+\|\mathcal{S}_{t,r}-\mathcal{S}_t\|\right). \tag{21}$$

Then, we can apply Lepski's method (Lepski et al., 1997) to construct the adaptive estimator for r^*. Without loss of generality, we choose $\eta = 1/2$. The, we define the adaptive estimator as

$$\widehat{r} = \max\left\{0\le r\le T/2 : \left\|\widehat{\mathcal{S}}_{t,r}-\widehat{\mathcal{S}}_{t,\rho}\right\| \le 4W_{\max}\sqrt{\frac{N\|\mathcal{S}_t\|_\infty}{\rho\vee 1}},\ \text{for any } \rho<r\right\}. \tag{22}$$

Next, for the choice of α_t, we select α_t to achieve a balance between $L_{\tau,t}$ and C_t^w:

$$\alpha_t = \frac{\lambda_K(L_{\tau,t})-\lambda_{K+1}(L_{\tau,t})}{\lambda_1(C_t^w)}. \tag{23}$$

Lastly, to determine K, we have several choices. Wang and Bickel (2017) implement a pseudo likelihood approach to choose the number of clusters in a stochastic blockmodel without covariates. Chen and Lei (2017) propose a network cross-validation procedure to estimate the number of clusters by utilizing adjacency information. Li et al. (2016) refine the network cross-validation approach by proposing an edge sampling algorithm. In our case, we apply the network cross-validation approach directly by inputting the similarity matrix instead of the adjacency matrix because the covariate matrix C_t^w behaves just like an adjacency matrix when we use dummy variables to indicate different technology attributes. Therefore, the network cross-validation applies to the similarity matrix in our study.

2.5 Monte Carlo simulations

In this section, we carry out several simulation studies using our algorithm and existing clustering methods under different model setups. Our benchmark algorithms for undirected networks are the dynamic degree-corrected spectral clustering for the sum of the squared adjacency matrix (DSC-DC) by Bhattacharyya and Chatterjee (2018) and the dy-

namic spectral clustering method (DSC-PZ) by Pensky and Zhang (2017). For the directed networks, as we do not have a fair competitor for a dynamic model, we choose several algorithms designed for a static model. In particular, we compete with the degree-corrected DI-SIM (DI-SIM-DC) by Rohe et al. (2016) and the covariate-assisted DI-SIM (CA-DI-SIM-St) method by Zhang et al. (2018) for the adjacency matrix in each period.

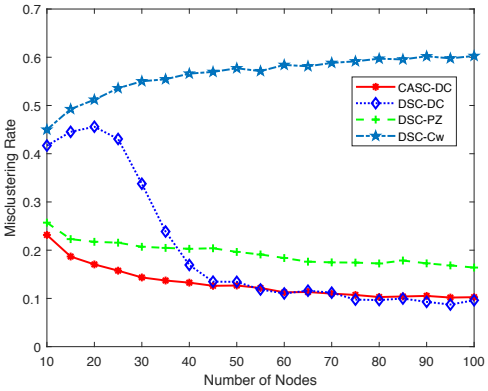
First, we set the block probability matrix B_t as

$$B_t = \frac{t}{T}\begin{bmatrix} 0.9 & 0.6 & 0.3 \\ 0.6 & 0.3 & 0.4 \\ 0.3 & 0.4 & 0.8 \end{bmatrix}, \text{ with } 1 \leq t \leq T.$$

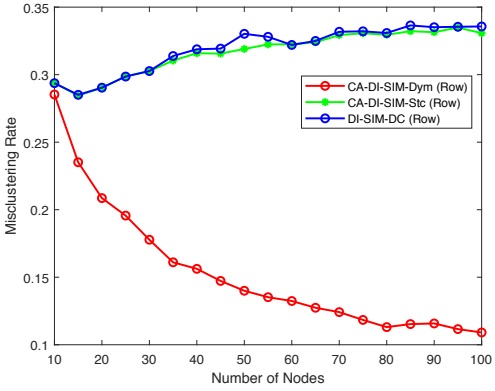
and set the order of the polynomials for kernel construction at $L = 4$ for all simulations. In the next step, for the undirected network, we simulate the first period's clustering matrix Z_1 by randomly choosing one entry in each row and assign it to 1 to generate clustering matrices (Z_t). Then, for $t = 2, \cdots, T$, we fix the last $N - s$ rows of Z_{t-1} and re-assign 1s in the first s rows of Z_1 to mimic the group membership change behaviors. Similarly, for the directed network, we generate each period's row/column clustering matrix ($Z_{R,t}$ or $Z_{C,t}$) in the same way, separately. Lastly, we assume that the number of communities $K = 3$ (or K_R and K_C for directed network) is known throughout the simulations. The time-invariant node covariates are $R = \lfloor \log(N) \rfloor$ dimensional with values $X \sim U(0, 10)$. We replicate all experiments 100 times and the misclustering rate we report is the temporal average of the misclustering rates; that is, $T^{-1} \sum_{t=1}^{T} |\mathbb{M}_t|/N$ (or $T^{-1} \sum_{t=1}^{T} |\mathbb{M}_t^R|/N$ and $T^{-1} \sum_{t=1}^{T} |\mathbb{M}_t^C|/N$ for the directed network).

We first examine the clustering performance with a growing network size. The number of vertices in the network varies from 10 to 100 with step size 5. The time span is $T = 10$. We summarize the results in Figure 1. Evidently, as the size of the undirected network becomes larger (panel (a)), the misclustering rates of the CASC-DC decrease sharply and dominate DSC-PZ in all cases. DSC-DC only performs as well as CASC-DC when the network is large, while CASC-DC retains an acceptable misclustering rate in small networks. It also shows that although using the covariate per se for clustering (DSC-Cw) is unsatisfactory, we can still add covariates to the adjacency matrix for better grouping.

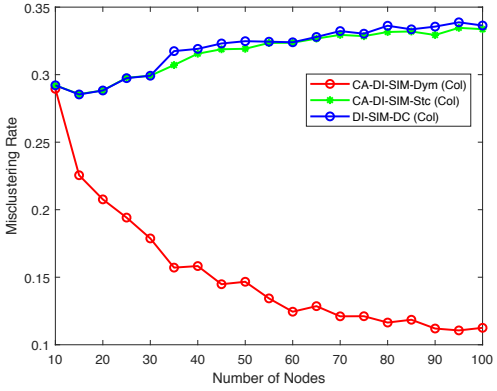
Next, we check the relative performance for a growing maximal number of group membership changes. Here, we fix the total number of vertices at 100 and we vary the group



(a) Undirected Network



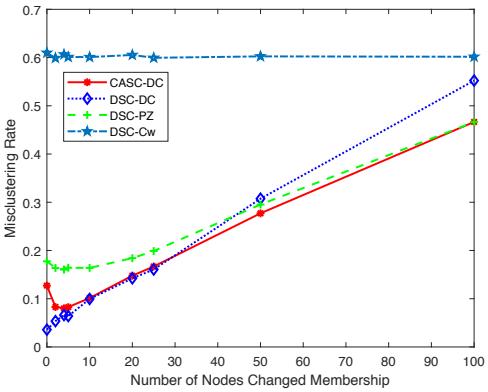
(b) Directed Network (Row Cluster)



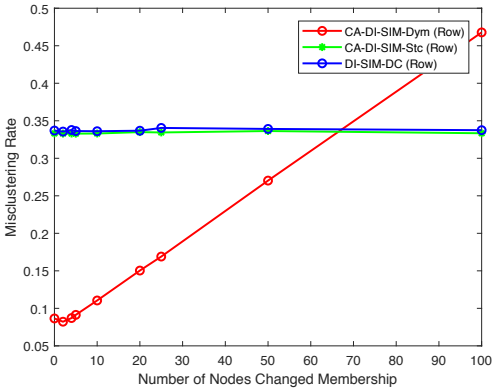
(c) Directed Network (Column Cluster)

Figure 1: This figure reports the misclustering rate of different spectral clustering algorithms for networks with a growing number of vertices. Panel (a) reports the results for undirected networks, while Panels (b) and (c) report the results for directed networks. CASC-DC represents Algorithm 1. DSC-DC denotes the dynamic spectral clustering in Bhattacharyya and Chatterjee (2018). DSC-PZ denotes the dynamic spectral clustering methods in Pensky and Zhang (2017). DSC-Cw is the spectral clustering based on only covariates. CA-DI-SIM-Dym represents Algorithm 2. DI-SIM-DC is the degree-corrected DI-SIM in Rohe et al. (2016) and CA-DI-SIM-Stc is the static covariate-assisted DI-SIM method in Zhang et al. (2018). In all cases, the number of nodes varies from 10 to 100, and the number of membership changes is fixed at $s = N^{1/2}$. The horizon $T = 10$ and all simulations are repeated 100 times.

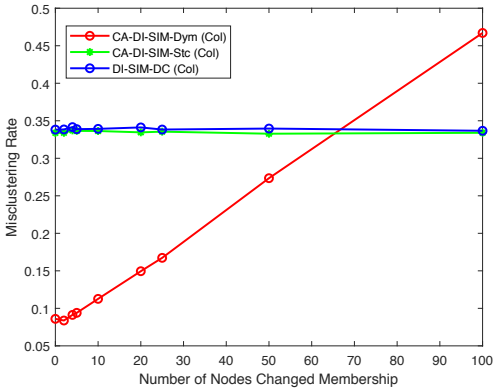
membership changes for each period, s, in $\{0, N/50, N/25, N/20, N/10, N/5, N/4, N/2, N\}$. The total number of horizons is $T = 10$. We summarize the results in Figure 2. Obviously, our methods are sensitive to the total number of group membership changes. In other words, the more unstable the group membership is, the higher the misclustering rate will be. Despite the result, our method still achieves the lowest misclustering rate amongst all methods when the group memberships are relatively stable ($s \leq N/2$).



(a) Undirected Network



(b) Directed Network (Row Cluster)



(c) Directed Network (Column Cluster)

Figure 2: This figure reports the misclustering rate of different spectral clustering algorithms for networks with a growing number of membership changes. Panel (a) reports the results for undirected networks, while Panels (b) and (c) report the results directed networks. CASC-DC represents Algorithm 1. DSC-DC denotes the dynamic spectral clustering in Bhattacharyya and Chatterjee (2018). DSC-PZ denotes the dynamic spectral clustering methods in Pensky and Zhang (2017). DSC-Cw is the spectral clustering based on only covariates. CA-DI-SIM-Dym represents Algorithm 2. DI-SIM-DC is the degree-corrected DI-SIM in Rohe et al. (2016) and CA-DI-SIM-Stc is the static covariate-assisted DI-SIM method in Zhang et al. (2018). In all cases, the network size is fixed at 100, and the number of membership changes varies in $\{0, N/50, N/25, N/20, N/10, N/5, N/4, N/2, N\}$. The horizon is $T = 10$ and all simulations are repeated 100 times.

3 Crypto Networks and Clusters

In this section, we illustrate how we construct a dynamic network structure using crypto returns and its contract information. Specifically, we first form a return-based network using the inter-predictive relations between cryptos. In addition, we add linkages between the cryptos that adopt similar cryptography techniques. We then perform clustering with our new algorithm.

3.1 Data and variables

We collected data on the historical daily prices, trading volumes, and contract attributes of the top 200 cryptos by market capitalization from an interactive platform (*Cryptocompare.com*) with free API access. After excluding cryptos with incomplete contract information, we obtain a sample of 199 cryptos. The sample covers August 31, 2015 to March 31, 2018, and we used an in-sample period for community detection from August 31, 2015 to December 31, 2017and an out-of-sample period of three months (2018-01-01 to March 31, 2018) for return predictability tests and portfolio construction. In term of the time-invariant attributes, we mainly collected algorithm and proof types from each crypto's contract:

Algorithm, which is short for the *hashing algorithm*, plays a central role in determining the security of the crypto. For each crypto, there is a hash function in mining; for example, Bitcoin (BTC) uses double SHA-256 and Litecoin (LTC) uses Scrypt. As security is one of the most important features of cryptos, the hashing algorithm naturally–in terms of trust–determines the intrinsic value of a crypto. In the example above, the Scrypt system was used with cryptos to improve upon the SHA256 protocol. The SHA256 preceded the Scrypt system and was the basis for BTC. Specifically, Scrypt was employed as a solution to prevent specialized hardware from brute-force efforts to out-mine others. Thus, Scrypt-based Altcoins require more computing effort per unit, on average, than the equivalent coin using SHA256. The relative difficulty of the algorithm confers a relative value.

Proof Types, or proof system/protocol, is an economic measure to deter denial of service attacks and other service abuses such as spam on a network by requiring some work from the service requester, usually the equivalent to processing time by a computer. For each crypto, at least one of the protocols will be chosen as a transaction verification method; for example, BTC and Ethereum (ETH) currently use the Proof-of-Work (PoW), and Diamond

(DMD) and Blackcoin use the Proof-of-Stake (PoS). PoW-based cryptos such as BTC use mining–the solving of computationally intensive puzzles–to validate transactions and create new blocks. In PoS-based cryptos, the creator of the next block is chosen through various combinations of random selection and wealth (in terms of crypto) or age (i.e., the stake). In summary, the proof protocol determines the reliability, security, and effectiveness of the transactions.

3.2 Crypto network construction

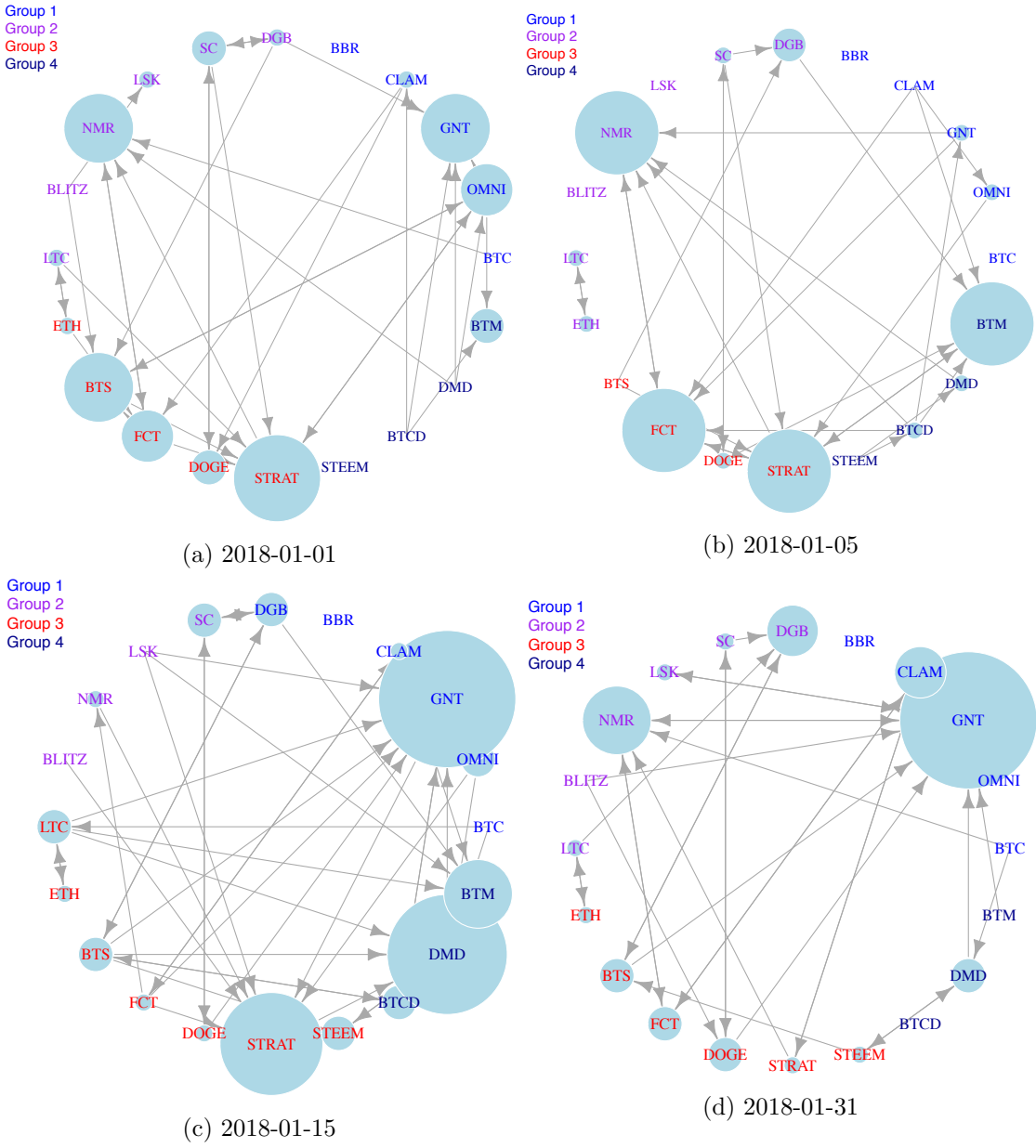
To study how risk or information propagates through the network, we construct it from the interrelations between the crypto returns. More precisely, we focus on one crypto and regresses its returns on the other cryptos' lagged returns in a 60-day estimation window. We employ adaptive LASSO (Zou, 2006) to estimate the regression coefficient; that is,

$$\hat{b}_i^* = \arg\min \left\{ \left\| r_{i,t+1}^s - \alpha_i - \sum_{j \neq i} b_{i,j} r_{j,t}^s \right\|^2 + \lambda_i \sum_{j \neq i} \hat{w_{i,j}} |b_{i,j}| \right\}, \tag{24}$$

where $r_{j,t}^s$ is the standardized return for crypto j, $\hat{b}_i^* = (\hat{b}_{i,1}^*, \cdots, \hat{b}_{i,N}^*)^\top$ is the adaptive LASSO estimate, λ_i are non-negative regularization parameters, and $\hat{w}_{i,j}$ are the weights corresponding to $|b_{i,j}|$ for $j = 1, \cdots, N$ in the penalty term. Conventionally, one defines $\hat{w}_{i,j} = 1/|\hat{b}_{i,j}^{ols}|^\gamma$ with some $\gamma > 0$. The LASSO technique yields an active set that has "parental" influence on the focal crypto. Thus, we obtain an adjacency matrix for each period, $\mathcal{A}_t$, $t = 1, \cdots, T$.

In Figure 3, we visualize a subgroup of 20 cryptos on selected dates to illustrate the structural features revealed by (24). The node color indicates the estimated group membership and the node size denotes its degree centrality from the receiver's perspective. Evidently, the predictive relations between cryptos are highly asymmetrical (rare double-sided arrows). Acemoglu et al. (2012) also observe this feature, which will later help us argue that sectoral shocks might not cancel out through diversification, but aggregate into a systematic fluctuation. Therefore, determining the centered cryptos and the group structure is crucial for understanding how information or shocks propagate in the crypto market.

As Figure 3 shows, the return-inferred network is time-varying and sparse in general. Taking subfigures (a) and (d) as an example, the interrelation between BTC and DMD



(a) 2018-01-01

(b) 2018-01-05

(c) 2018-01-15

(d) 2018-01-31

Figure 3: This figure presents the return-based network structure on selected dates in January 2018. We selected 20 cryptos, including BTC, ETH, LTC and other top cryptos by market capitalization as of December 31, 2017 within each group estimated by dynamic CASC. We obtained the connections from the predictive regression $r^s_{i,t+1} = \alpha_i + \sum_{j=1,j\neq i}^{N-1} b_{i,j} r^s_{j,t} + \epsilon_{i,t}$, where $r^s_{i,t}$ is the standardized daily return on crypto i and N is the total number of cryptos. Adaptive LASSO is employed to estimate the regression above and only the cryptos selected by adaptive LASSO will be linked to crypto i.

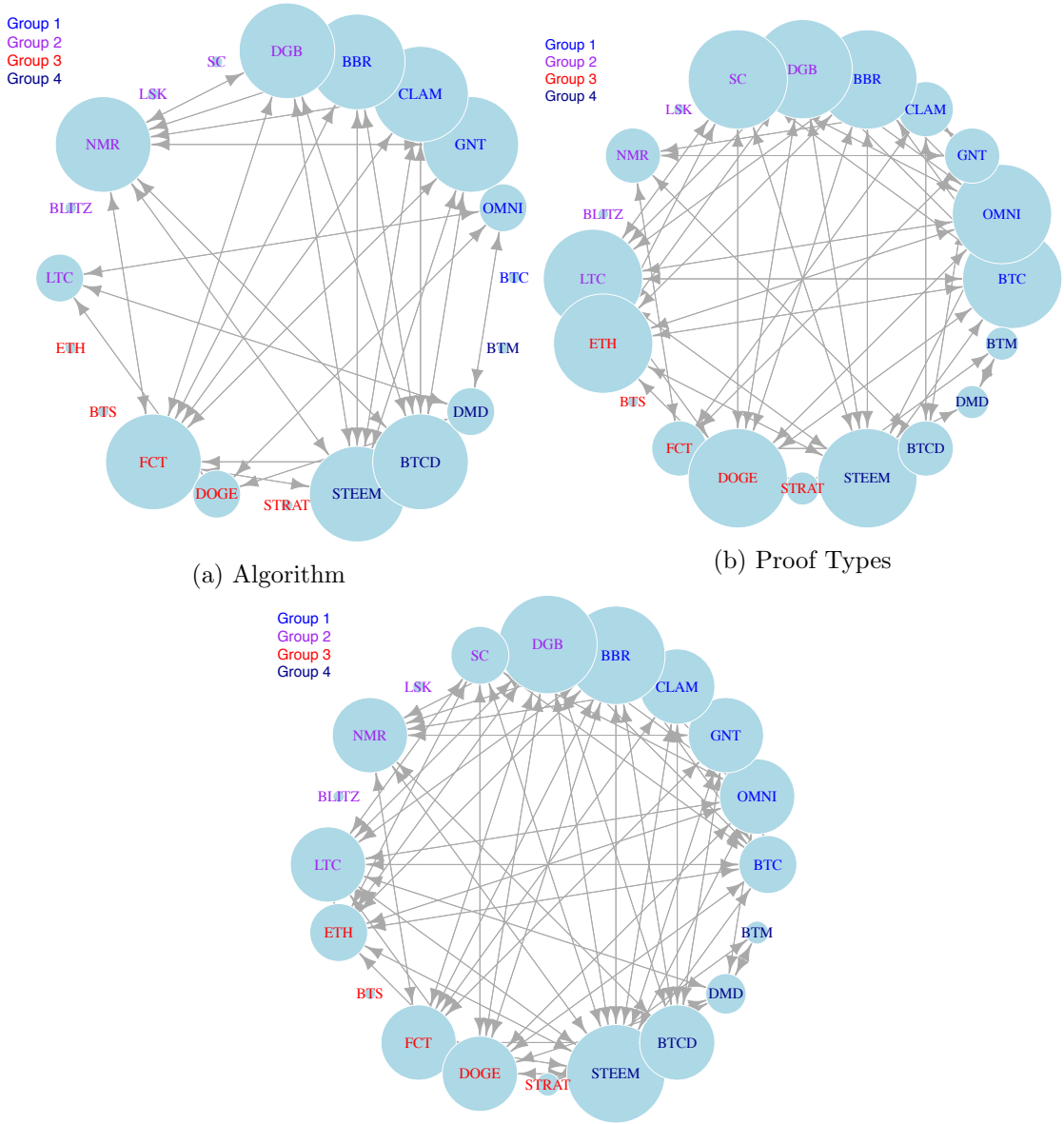
vanishes on January 1, 2018, and the connections on 2018-01-01 are sparser than those on January 15, 2018 are. This observation requires a more refined clustering and the use of node attributions. To demonstrate how node attribution assists classification, we replot the network with the same cryptos in Figure 3 and link the cryptos that share at least one fundamental characteristic to obtain Figure 4. Both LTC and DOGE adopt the Scrypt algorithm; hence, these two cryptos are fundamentally connected.

Clearly, due to the limited choices of algorithms and other attributes, the cryptos are more likely to connect with each other when using attribute commonality to form linkages. However, using contract information alone is enough to identify the group structure, as crypto returns carry information on investors' beliefs, which is particularly important for the crypto market. In addition, the relationship between a crypto's fundamental characteristics to its value is more complicated than is a firm's fundamental to its equity. It is possible that a new algorithm does not add any valuable features to the existing algorithms. In fact, many developers simply copy and paste the blockchain source code with minor modifications on the parameters to launch a new coin for speculative purposes through an initial coin offering (ICO). Although these Altcoins may show little differences between their fundamental characteristics, their abilities to generate future cash flows vary considerably. A good example is IXCoin, the first BTC *clonecoin*. While IXCoin copied every detail from Bitcoin, IXCoin was unable to replicate the success of BTC. The developers stopped working on IXCoin for months after its ICO. This example shows that a clonecoin could be more risky than its protocoin for speculation reasons. www.deadcoins.com provides other similar cases.

To address the issues raised above fully, we combine the return-based network and the contract-based network using a similarity matrix. Figure 5 illustrates the combined network for selected dates. Compared to the network based on a single information set, the combined network is denser and the degrees of the cryptos are distributed more evenly. Consequently, the similarity matrix will most likely improve classification accuracy.

3.3 Clusters in crypto networks

The combined network structure and application of the CASC created four groups. Table 2 summarizes the grouping results for one example. The table indicates that as of 2017-12-31, the largest cryptos (BTC, ETH, and LTC) in terms of market capitalization are not

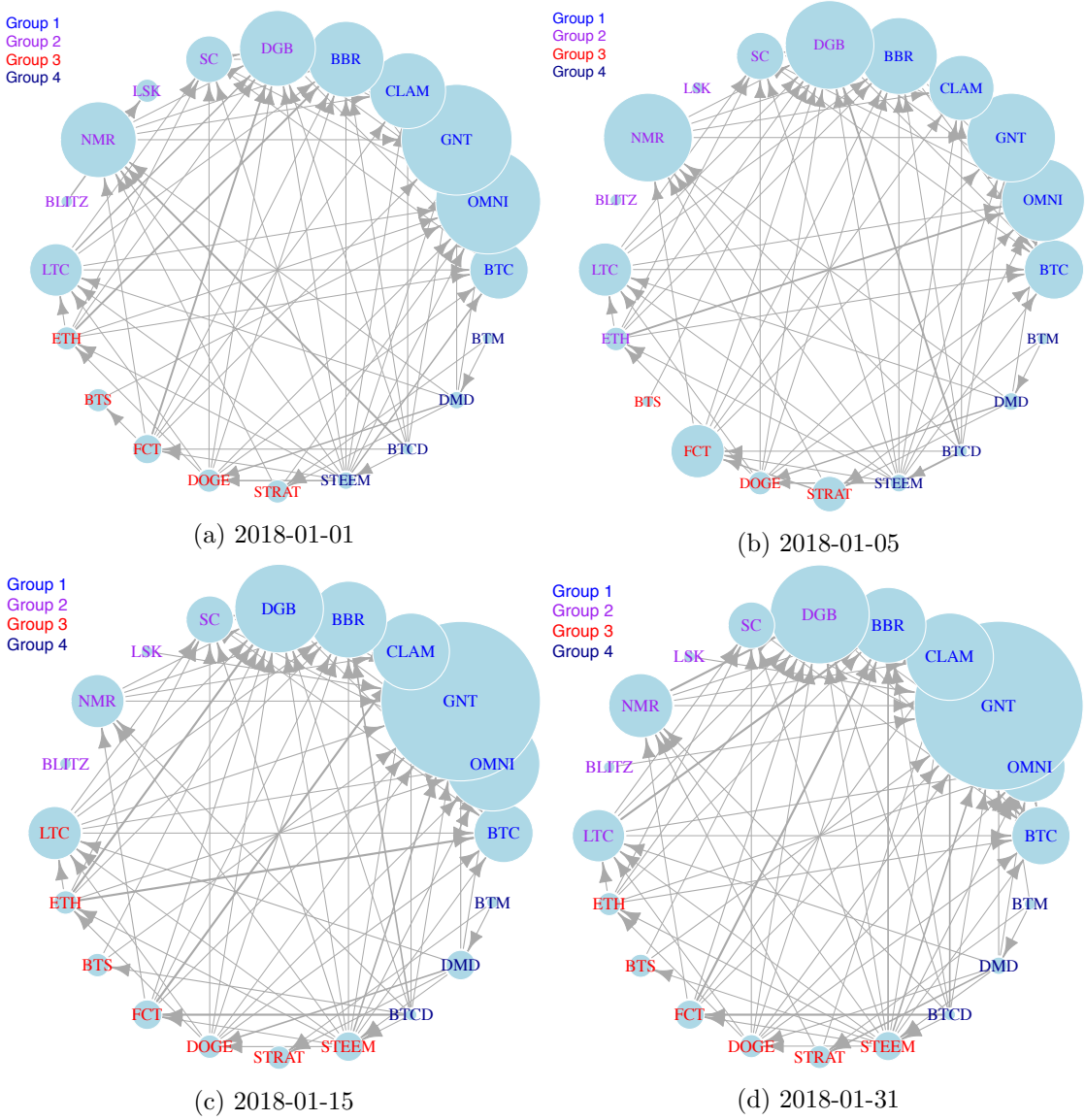


(a) Algorithm

(b) Proof Types

(c) Combined Fundamentals

Figure 4: This figure depicts the contract-based network structure. We link two cryptos if they share the same fundamental technology, that is, algorithm and proof types. Node size denotes the degree centrality of the crypto.



(a) 2018-01-01

(b) 2018-01-05

(c) 2018-01-15

(d) 2018-01-31

Figure 5: This figure depicts the dynamic combined network structure based on a similarity matrix, which combines return information and contract information simultaneously. The color of the node labels indicates the group estimated by dynamic CASC and the node size denotes the degree centrality of the crypto.

necessarily categorized into the same group. Take LTC and BTC as an example. Although their return patterns are closely related, the fundamental attributes between them are rather different: BTC employs SHA256 while LTC uses Scrypt. As a comparison, we also show the grouping results for the same 20 cryptos under DISIM from Rohe et al. (2016) in Table 2.

Table 1: Representative Cryptos of Groups Estimated by the Dynamic CASC.

Group ID	Group 1	Group 2	Group 3	Group 4
Cryptocurrencies	BBR	BLITZ	BTS	BTCD
	BTC	DGB	DOGE	BTM
	CLAM	LSK	*ETH*	DMD
	GNT	NMR	FCT	STEEM
	OMNI	SC	*LTC*	STRAT

Table 2: Representative Cryptos of Groups Estimated by DISIM from Rohe et al. (2016).

Group ID	Group 1	Group 2	Group 3	Group 4
Cryptocurrencies	BBR	BTC	BLITZ	BTCD
	LSK	DGB	STEEM	CLAM
	DOGE	LTC	SC	GNT
	ETH	NMR	BTS	
	OMNI	DMD		
	BTM	STRAT		

To illustrate the performance of our method, we check the differences between the within- and cross-group connections of each group, defined as

$$\textit{Within-Group Connection}_i = \frac{\#\text{ of Degrees of Coins within Group } i}{4N_i},$$
$$\textit{Cross-Group Connection}_i = \frac{\#\text{ of Degrees of Coins between Group } i \text{ and other Groups}}{4\bar{N}_i},$$

where N_i is the number of cryptos in group i and $\bar{N}_i$ is the number of cryptos not in group i. Intuitively, if the clustering method correctly classifies all cryptos, then the within-group connections should be stronger than the cross-group connections; that is, the difference between them should be positive. Table 3 summarizes the within- and cross-group connections of different information sets based on DISIM from Rohe et al. (2016) and dynamic CASC, respectively. Panel A reports the average return-based connection over the sample period. Panels B and C report the algorithm-inferred connections and proof-types-inferred

connections, respectively. The differences between the within- and cross-group connection (W-C difference) are reported in the last column of each panel.

Table 3: Within- and Cross-group Connections using DISIM and Dynamic CASC

Panel A reports the average return-based connection across the sample period. Panels B and C report the algorithm-inferred connections and proof-type-inferred connections, respectively. Statistical significanc indicated by 1% 5% 10% for the positive signs and 1% 5% 10% for the negative signs.

	Panel A: Return			Panel B: Algorithm			Panel C: Proof Types		
	Within	Cross	Diff.	Within	Cross	Diff.	Within	Cross	Diff.
DISIM by Rohe et al. (2016)									
Group 1	0.033	0.051	−0.018	0.252	0.204	0.048	0.251	0.229	0.021
Group 2	0.084	0.074	0.010	0.216	0.198	0.018	0.277	0.238	0.039
Group 3	0.086	0.075	0.011	0.215	0.196	0.019	0.279	0.238	0.041
Group 4	0.084	0.073	0.011	0.216	0.197	0.019	0.278	0.238	0.040
Overall	0.072	0.068	0.004	0.225	0.199	0.026	0.271	0.236	0.035
Dynamic CASC									
Group 1	0.029	0.026	0.003	0.232	0.202	0.030	0.266	0.232	0.033
Group 2	0.029	0.025	0.004	0.243	0.203	0.041	0.272	0.232	0.040
Group 3	0.031	0.025	0.005	0.240	0.202	0.038	0.274	0.233	0.041
Group 4	0.031	0.026	0.006	0.240	0.202	0.038	0.273	0.233	0.040
Overall	0.030	0.025	0.005	0.239	0.202	0.037	0.271	0.233	0.039

Evidently, the dynamic CASC method has superior classification efficiency than DISIM does given that it delivers higher overall differences in both return-inferred connections and contract-inferred connections. For example, the overall W-C difference of DISIM is 0.004, 0.026, and 0.035, while that of dynamic CASC is 0.005, 0.037, and 0.039, respectively. Indeed, dynamic CASC utilizes fundamental information better in the sense that the contract-inferred network structure (Panels B and C) generates a higher W-C difference without discounting the grouping information from the return-inferred network. These facts indicate that fundamental information introduces an extra dimension of commonality for classifying cryptos, and improves the information extraction from return dynamics by emphasizing the return co-movement induced by fundamental commonality.

4 Asset Pricing Inference

In this section, we apply the classifications we obtained to asset pricing. We first study whether the group structure achieves good risk diversification. Then, we sort the cryptos into 4 quartiles according to eigenvector centrality and construct a portfolio that goes long on the high-centrality cryptos and short on the low-centrality cryptos. Lastly, we conduct

several robustness tests to exclude alternative explanations of the centrality measure.

4.1 Risk diversification

Risk diversification is one of the most important issues in portfolio management. Portfolio managers seek to achieve a target return with the smallest variance possible. Therefore, it is crucial to invest in different assets or equity sectors that are not highly correlated with each other. We calculate the correlation coefficients of cryptos within the same group and those of the cryptos across groups. Table 4 summarizes the results.

Table 4: Within- and Cross-group Cryptos' Average Return Correlations by Dynamic CASC.

This table reports the within- and cross-group average return correlation based on dynamic CASC. Each trading day, we balance the portfolio according to the clustering results and calculate the within- and cross-group correlations. The number in brackets below are the t-statistics, which are adjusted by the Newey-West lags(4) method. Statistical significance is indicated by ■ 1% ■ 5% ■ 10% for the positive signs and ■ 1% ■ 5% ■ 10% for the negative signs. The sample period spans from August 31, 2015 to March 31, 2018.

	Within Group	Cross Group	Diff.
Group 1	0.169	0.154	0.014
	(7.626)	(7.423)	(6.856)
Group 2	0.179	0.154	0.021
	(8.077)	(7.423)	(6.077)
Group 3	0.181	0.157	0.021
	(8.191)	(7.506)	(10.374)
Group 4	0.188	0.157	0.027
	(8.114)	(7.416)	(5.607)
Overall	0.188	0.157	0.021
	(7.697)	(7.381)	(6.331)

In Table 4, we compare the average pair-wise correlations between two groups. For the within-group portfolio, we randomly pick 10 cryptos from the same group, and for the cross-group portfolio, we randomly pick 5 cryptos in one group and pick the remaining 5 cryptos from other groups. Then, for each trading day, we balance the portfolio according to the clustering results and calculate the within- and cross-group correlations. Table 4 demonstrates that the correlations between cryptos within the same group are on average significantly higher than those across groups are. Indeed, the average correlation coefficient within a group is 0.18, while it is 0.15 across groups. In economic terms, this result indicates a 17% reduction in return co-movement when investing in cross-group cryptos. The difference is statistically significant at the 1% level with a Newey-West adjusted t-statistic of

6.33. The result suggests that investment practitioners can find attractive upside and diversification possibility through allocating portfolio weights on cryptos from different groups. As buying all cryptos is costly, the findings provide portfolio managers the opportunity to select group representatives with a significant diversification effect.

4.2 Centrality and crypto return

One major advantage of jointly modelling cryptos with a dynamic network is its convenience for studying how risk and trading information propagates from one crypto to another. Acemoglu et al. (2012) propose a theoretical model to explain the spillover effects through sector-level shocks. The model suggests that if the linkages in the inter-sectoral network are sufficiently asymmetric, then sectoral shocks might not cancel out through diversification, but aggregate into macroeconomic fluctuations. Ahern (2013) also finds that idiosyncratic shocks could travel between linked stocks following the direction of the linkages. Therefore, stocks with more "receive linkages" tend to bear more risks in the network and thus require a higher risk premium. Similarly, we would expect that cryptos in a more central position in the network require a higher risk premium.

Centrality, as the key measure describing the importance of the nodes in the network, best proxies the concentration of risks or trading information. There are several measures of centrality, such as degree, closeness, betweenness, and eigenvector centrality. Among them, eigenvector centrality is the most appropriate measure for an asymmetric network for two reasons. First, shocks that transmit across the crypto market do not have final recipients and are unlikely to follow the shortest path between nodes. Therefore, we cannot use closeness and betweenness centrality to describe market shocks as they implicitly assume that traffic follows geodesic paths (Borgatti, 2005). Second, cross-asset shocks are likely to have feedback effects evidenced by the two-way connections between paired cryptos in Figure 3. Thus, using degree centrality tends to overestimate the importance of cryptos with more asymmetric linkages. Eigenvector centrality is calculated via the principal eigenvector of the network's adjacency matrix (Bonacich, 1972). Nodes are more central if they are connected to other nodes that are themselves more central. Figure 6 plots the average return of each group portfolio, labelled as high-, median- (2 groups in the middle), and low-centrality groups. Based on the thoughts on portfolio performance above, we find that the group with a higher centrality wins the horse race.

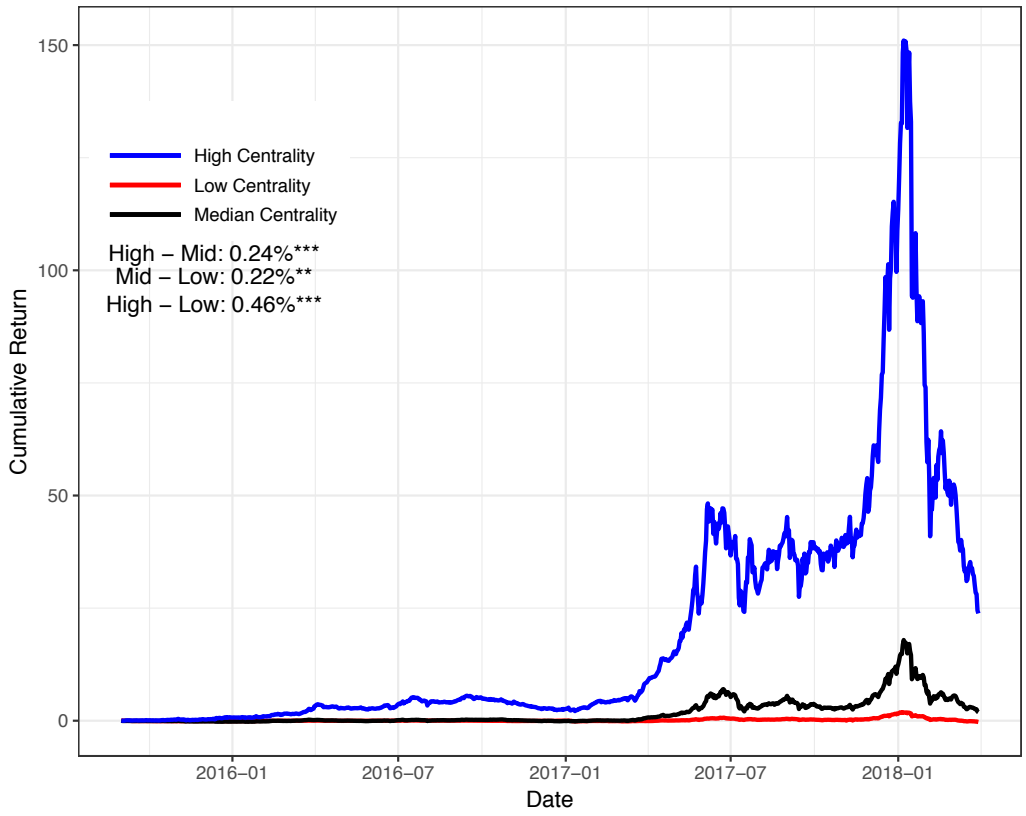


Figure 6: This figure depicts the cumulative portfolio return of the high-, median-, and low-centrality groups. Centrality is based on the similarity matrix, which combines return information and contract information simultaneously. The sample period spans from August 31, 2015 to March 31, 2018.

Next, we formally test this discovery by studying cross-sectional portfolio returns. We first sort cryptos into quartile portfolios based on the eigenvector centrality calculated from the similarity matrix on each trading day. We then look at each portfolio's average future returns. Next, we test the statistical significance of the difference in average future return between the high and low portfolios. To show the informativeness of our centrality measure, we construct the portfolio for several formation periods, ranging from day $t+1$ to $t+7$ days. Table 5 reports the results.

Table 5: Average Future Returns of the Cross-sectional Portfolios by Centrality Sorting.

This table reports the average future return for quartile portfolios sorted by the centrality measure. Each trading day, we balance the portfolio according to the centrality score of the previous trading day and calculate the average portfolio returns for both short and long legs. Statistical significance is indicated by 1% 5% 10% for the positive signs and 1% 5% 10% for the negative signs. The t-statistics in parentheses are computed based on standard errors with a Newey-West lags(4) adjustment. The sample period spans from August 31, 2015 to March 31, 2018.

Centrality	Ret_{t+1}	Ret_{t+2}	Ret_{t+3}	Ret_{t+4}	Ret_{t+5}	Ret_{t+6}	Ret_{t+7}
Low	0.00%	0.00%	-0.03%	-0.01%	0.03%	0.02%	0.06%
2	0.15%	0.18%	0.18%	0.19%	0.16%	0.18%	0.16%
3	0.34%	0.34%	0.28%	0.36%	0.38%	0.28%	0.29%
High	0.40%	0.36%	0.48%	0.38%	0.34%	0.42%	0.39%
High - Low	0.40%	0.36%	0.51%	0.39%	0.32%	0.40%	0.33%
t-statistic	(3.53)	(3.10)	(4.24)	(3.33)	(2.74)	(3.44)	(2.85)

In line with the observations from Figure 6, the cryptos with a higher quartile of centrality receive a higher portfolio return. Particularly, the average portfolio return is 39.78 bps for the highest-centrality group, while it is -0.01 bps for the lowest-centrality group. The difference is statistically significant at the 1% level. We find similar results across different portfolio formation periods. The result provides strong evidence that an informational channel, such as risk and liquidity, should be applied to interpret the eigenvector centrality measure.

4.3 Alternative Interpretation

We showed that the centrality measure is economically meaningful as a risk factor. However, it does not rule out other explanations. We therefore conduct several tests to seek other possibilities to link the centrality measure to economic theory. In particular, we test if limit-to-arbitrage, investor attention, and macroeconomic uncertainty can deliver meaningful explanatory power of the anomaly.

The first typical explanation for asset return anomaly is the limit-to-arbitrage. According to Shleifer and Vishny (1997), sophisticated investors would quickly eliminate any return predictability arising from anomalies in a liquid market without impediments to arbitrage. Therefore, when cryptos are illiquid, an arbitrage opportunity is more likely to exist between central and non-central cryptos. As a formal test, we proxy liquidity with trading volume and first sort the cryptos into two groups (high and low) according to their previous day's trading volume. Then, for each group, we sort cryptos by their eigenvector centrality as in the previous sections, and report the corresponding portfolio returns in the first two columns of Table 6.

We find that the centrality portfolio return (High–Low) remains significantly positive for both high- and low-volume cryptos. For example, in the low-volume group, the portfolio return is 5 bps for the low-centrality group, while it increases to 28 bps for the high-centrality group. The significantly positive portfolio returns in both groups indicate that the limit-to-arbitrage does not fully explain the centrality measure.

The recent study of Liu and Tsyvinski (2018) provides an alternative explanation. The authors find that investor attention is a powerful predictor of crypto returns. Barber and Odean (2008) point out that excess attention usually drives investors to overreact to information and thus causes mispricing. Guo et al. (2018) show that investor attention could spill over along the network linkages. Hence, cryptos in a high-investor-attention period are more likely to be mispriced. Following Liu and Tsyvinski (2018), we proxy investor attention by constructing the deviation of Google searches for the word "crypto" on a given day compared to the average of those in the preceding four weeks. We split the sample into two periods (high and low) and test for the existence of the anomaly in each period. We summarize the results in the middle columns of Table 6.

In general, the proposed centrality measure—under both high- and low-attention periods—is a better choice. The effect seems to be stronger in high-attention periods. For example, the centrality portfolio achieves a 0.45% daily return during a high-attention period, while it retains a 0.35% return, if not higher, for the low-attention period. However, we can observe that the results are not fully explained by investor attention, as our centrality measure shows significant cross-sectional return predictability.

Last, observing that government policy and crypto price movement has a strong synchronization (Demir et al., 2018), we must check whether the centrality measure relates

to underlying economic uncertainty. Naturally, when macroeconomic conditions become uncertain, investing in a certain asset is more risky and investors will require a higher risk premium (Brogaard and Detzel, 2015). We employ Baker et al. (2016) policy uncertainty index, which is constructed from three types of underlying components: media news; the Congressional Budget Office (CBO), which compiles lists of temporary federal tax code provisions; and the Federal Reserve Bank of Philadelphia's Survey of Professional Forecasters. Similarly, we divide the sample into two parts, high- and low-uncertainty periods, and test the existence of abnormal returns in each period. The last two columns of Table 6 report the results.

Evidently, the centrality portfolio return remains significantly positive under both high- and low-economic-uncertainty periods. Specifically, in a high-period, the portfolio return is 1 bps for the low-centrality group and 49 bps for the high-centrality group, which reveals a difference of 48 bps with a Newey-West adjusted t-statistic of 2.71. The results are a bit weaker in the low-uncertainty period, but the overall pattern remains. In this case, the centrality measure cannot be fully explained by economic uncertainty.

In summary, the proposed centrality measure is not driven by the pricing factors listed above. Although we did not exhaust all possibilities, the facts suggest that the centrality measure serves well as an idiosyncratic risk factor to predict future crypto returns.

Table 6: Portfolio Returns: Trading Volume, Investor Attention, and Macro Uncertainty

This table reports the quartile portfolio returns sorted by the centrality measure for cryptos with high and low trading volume, in high- and low-investor-attention periods, or under high- and low-macro-uncertainty circumstances. Statistical significance is indicated by ■ 1% ■ 5% ■ 10% for the positive signs and ■ 1% ■ 5% ■ 10% for the negative signs. t-statistics in parentheses are computed based on standard errors with Newey-West lags(4) adjustment. The sample period spans from August 31, 2015 to March 31, 2018.

Centrality	Trading Volume		Investor Attention		Macro Uncertainty	
	Low	High	Low	High	Low	High
Low	0.05%	-0.04%	0.04%	0.01%	0.01%	0.01%
2	0.16%	0.27%	0.11%	0.21%	0.02%	0.26%
3	0.38%	0.12%	0.46%	0.32%	0.22%	0.54%
High	0.56%	0.28%	0.39%	0.46%	0.32%	0.49%
High - Low	0.51%	0.33%	0.35%	0.45%	0.31%	0.48%
t-statistic	(3.62)	(2.73)	(2.27)	(3.06)	(2.23)	(2.71)

5 Conclusion

This study examined the market segmentation problem in the crypto market. To solve the problem, we constructed a dynamic network of cryptos using return inter-predictive relationship and proposed a dynamic CASC method to make full use of the dynamic linkage information, as well as the node attributions, to improve classification accuracy. Based on the fitted crypto network and in the spirit of Ahern (2013), we proposed using eigenvector centrality as the idiosyncratic risk factor for predicting future returns. We find that the cross-sectional portfolio constructed from eigenvector centrality sorting can deliver a persistent 40 bps daily return.

References

Acemoglu, D., V. M. Carvalho, A. Ozdaglar, and A. Tahbaz-Salehi (2012): "The Network Origins of Aggregate Fluctuations," *Econometrica*, 80, 1977–2016.

Acemoglu, D., A. Ozdaglar, and A. Tahbaz-Salehi (2015): "Systemic Risk and Stability in Financial Networks," *American Economic Review*, 105, 564–608.

Ahern, K. R. (2013): "Network Centrality and the Cross Section of Stock Returns," *SSRN Electronic Journal, DOI: 10.2139/ssrn.2197370*.

Amihud, Y. and H. Mendelson (1986): "Asset Pricing and the Bid-ask Spread," *Journal of Financial Economics*, 17, 223–249.

Aobdia, D., J. Caskey, and N. B. Ozel (2014): "Inter-industry Network Structure and the Cross-predictability of Earnings and Stock Returns," *Review of Accounting Studies*, 19, 1191–1224.

Baker, S. R., N. Bloom, and S. J. Davis (2016): "Measuring Economic Policy Uncertainty," *The Quarterly Journal of Economics*, 131, 1593–1636.

Barber, B. M. and T. Odean (2008): "All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors," *Review of Financial Studies*, 21, 785–818.

Bhattacharyya, S. and S. Chatterjee (2018): "Spectral Clustering for Multiple Sparse Networks: I," *arXiv preprint arXiv:1805.10594*.

Binkiewicz, N., J. T. Vogelstein, and K. Rohe (2017): "Covariate-assisted Spectral Clustering," *Biometrika*, 104, 361–377.

Bonacich, P. (1972): "Technique for Analyzing Overlapping Memberships," *Sociological Methodology*, 4, 176–185.

Borgatti, S. P. (2005): "Centrality and Network Flow," *Social Networks*, 27, 55–71.

Boyer, B. H. (2011): "Style-Related Comovement: Fundamentals or Labels?" *The Journal of Finance*, 66, 307–332.

Brogaard, J. and A. Detzel (2015): "The Asset-pricing Implications of Government Economic Policy Uncertainty," *Management Science*, 61, 3–18.

Chaudhuri, K., F. Chung, and A. Tsiatas (2012): "Spectral Clustering of Graphs with General Degrees in the Extended Planted Partition Model," in *Conference on Learning Theory*, 35–1.

Chen, C. Y.-H., W. K. Härdle, A. J. Hou, and W. Wang (2018): "Pricing Cryptocurrency Options: The Case of CRIX and Bitcoin," *SSRN Electronic Journal, DOI: 10.2139/ssrn.3159130*.

Chen, C. Y.-H., W. K. Hrdle, and Y. Okhrin (2019): "Tail Event Driven Networks of SIFIs," *Journal of Econometrics*, 208, 282–298.

Chen, K. and J. Lei (2017): "Network Cross-validation for Determining the Number of Communities in Network Data," *Journal of the American Statistical Association*, 1–11.

Cohen, L. and A. Frazzini (2008): "Economic links and predictable returns," *The Journal of Finance*, 63, 1977–2011.

Demir, E., G. Gozgor, C. K. M. Lau, and S. A. Vigne (2018): "Does Economic Policy Uncertainty Predict the Bitcoin Returns? an Empirical Investigation," *Finance Research Letters*, 26, 145–149.

Guo, L., L. Peng, Y. Tao, and J. Tu (2018): "News Co-Occurrence, Attention Spillover and Return Predictability," *SSRN Electronic Journal, DOI: 10.2139/ssrn.2927561*.

Härdle, W., C. Harvey, and R. Reule (2019): "Understanding Cryptocurrencies," *Journal of Financial Econometrics, Forthcoming*.

Härdle, W. K. and S. Trimborn (2015): "CRIX or Evaluating Blockchain Based Currencies," Tech. Rep. Oberwolfach Report No. 42/2015, The Mathematics and Statistics of Quantitative Risk.

Herskovic, B. (2018): "Networks in Production: Asset Pricing Implications," *The Journal of Finance*, 73, 1785–1818.

Hoberg, G. and G. Phillips (2016): "Text-based Network Industries and Endogenous Product Differentiation," *Journal of Political Economy*, 124, 1423–1465.

Karrer, B. and M. E. J. Newman (2011): "Stochastic Blockmodels and Community Structure in Networks," *Physical Review E*, 83, 016107.

Klochkov, Y., W. Härdle, and X. Xiu (2019): "Localising Multivariate CAViaR," *IRTG 1792 Discussion Paper, under review in Journal of Econometrics*.

Kumar, A. and C. M. Lee (2006): "Retail Investor Sentiment and Return Comovements," *The Journal of Finance*, 61, 2451–2486.

Lee, D. K. C., L. Guo, and Y. Wang (2018): "Cryptocurrency: A New Investment Opportunity?" *The Journal of Alternative Investments*, 20, 16–40.

Lei, J. and A. Rinaldo (2015): "Consistency of Spectral Clustering in Stochastic Block Models," *The Annals of Statistics*, 43, 215–237.

Lepski, O. V., E. Mammen, and V. G. Spokoiny (1997): "Optimal Spatial Adaptation to Inhomogeneous Smoothness: An Approach Based on Kernel Estimates with Variable Bandwidth Selectors," *The Annals of Statistics*, 929–947.

Li, T., E. Levina, and J. Zhu (2016): "Network Cross-validation by Edge Sampling," *arXiv preprint arXiv:1612.04717*.

Liu, Y. and A. Tsyvinski (2018): "Risks and Returns of Cryptocurrency," *NBER Working Paper*.

Nakamoto, S. (2008): "Bitcoin: A Peer-to-peer Electronic Cash System," *Working Paper*.

Pensky, M. and T. Zhang (2017): "Spectral Clustering in the Dynamic Stochastic Block Model," *Arxiv Preprint Arxiv:1705.01204*.

Petukhina, A., S. Trimborn, W. K. Härdle, and H. Elendner (2018): "Investing With Cryptocurrencies – Evaluating the Potential of Portfolio Allocation Strategies," *IRTG 1792 Discussion Paper 2018-058*.

Rohe, K., T. Qin, and B. Yu (2016): "Co-clustering Directed Graphs to Discover Asymmetries and Directional Communities," *Proceedings of the National Academy of Sciences*, 113, 12679–12684.

Shleifer, A. and R. W. Vishny (1997): "The Limits of Arbitrage," *The Journal of Finance*, 52, 35–55.

Trimborn, S., M. Li, and W. K. Härdle (2019): "Investing with Cryptocurrencies - A Liquidity Constrained Investment Approach," *Journal of Financial Econometrics, Forthcoming*.

Wang, Y. X. R. and P. J. Bickel (2017): "Likelihood-based Model Selection for Stochastic Block Models," *The Annals of Statistics*, 45, 500–528.

Zhang, Y., M. Poux-Berthe, C. Wells, K. Koc-Michalska, and K. Rohe (2018): "Discovering Political Topics in Facebook Discussion Threads with Graph Contextualization," *Annals of Applied Statistics*, 12, 1096–1123.

Zou, H. (2006): "The Adaptive Lasso and Its Oracle Properties," *Journal of the American Statistical Association*, 101, 1418–1429.

Supplementary Appendix to "A Dynamic Network Perspective on Cryptocurrencies"

March 26, 2019

This appendix provides the proofs of some technical lemmas used in the above paper.

The notations that have been frequently used in the proofs are as follows: $[n] \stackrel{\text{def}}{=} \{1, 2, \cdots, n\}$ for any positive integer n, $\mathcal{M}_{m,n}$ be the set of all $m \times n$ matrices which have exactly one 1 and $n-1$ 0's in each row. $\mathbb{R}^{m\times n}$ denotes the set of all $m \times n$ real matrices. $\|\cdot\|$ is used to denote Euclidean ℓ_2-norm for vectors in $\mathbb{R}^{m\times 1}$ and the spectral norm for matrices on $\mathbb{R}^{m\times n}$. $\|\cdot\|_\infty$ denotes the largest element of the matrix in absolute value. $\|\cdot\|_F$ is the Frobenius norm on $\mathbb{R}^{m\times n}$, namely $\|M\|_F \stackrel{\text{def}}{=} \sqrt{\text{tr}(M^\top M)}$. $\|\cdot\|_{\phi_2}$ is the sub-Gaussian norm such that for any random variable x, there is $\|x\|_{\phi_2} \stackrel{\text{def}}{=} \sup_{\kappa\geq 1} \kappa^{-1/2}(\mathsf{E}\,|x|^\kappa)^{1/\kappa}$. $\mathbf{1}_{m,n} \in \mathbb{R}^{m\times n}$ consists of all 1's, ι_n denotes the column vector with n elements of all 1's. $\mathbb{1}_A$ denotes the indicator function of the event A.

1 Preliminary Lemmas

Lemma 1. *Suppose A_t and X are the adjacency matrix and the node covariate matrices sampled from the SC-DCBM/SC-DCcBM. Recall W_t and $\mathcal{W}_t$ are empirical and population weight matrices. Then, we have*

$$\sup_t \|W_t - \mathcal{W}_t\|_\infty = \mathcal{O}_p(\xi),$$

where $\xi = \max(\sigma^2\|L_\tau\|_F\sqrt{\log(TR)}, \sigma^2\|L_\tau\|\log(TR), NRJ^2/\underline{\delta})$ *and* $\underline{\delta} = \inf_t\{\min_i \mathcal{D}_{\tau,t}(i,i)\}$.

Proof. Define $\mathcal{I}_t = \mathcal{X}L_{\tau,t}\mathcal{X}$. Then we have

$$\sup_t \|W_t - \mathcal{W}_t\|_\infty \leq \sup_t \|W_t - \mathcal{I}_t\|_\infty + \sup_t \|\mathcal{I}_t - \mathcal{W}_t\|_\infty.$$

For the first part, define $L_\tau = \sup_t L_{\tau,t}$ and $\zeta = \max(\sigma^2\|L_\tau\|_F\sqrt{\log(TR)}, \sigma^2\|L_\tau\|\log(TR))$, then by Hansen-Wright inequality (c.f., Theorem 1.1 of Rudelson and Vershynin (2013)),

we have

$$\begin{aligned}\Pr(\sup_t \|X^\top L_{\tau,t} X - \mathcal{X}^\top L_{\tau,t}\mathcal{X}\| > \zeta) &\le \sum_{t=1}^{T} \Pr(\|X^\top L_\tau X - \mathcal{X}^\top L_\tau \mathcal{X}\| > \zeta) \\ &\le 2T \exp\left\{-c \min\left(\frac{\zeta^2}{\sigma^4 \|L_\tau\|_F^2}, \frac{\zeta}{\sigma^2 \|L_\tau\|}\right)\right\} \\ &= \mathcal{O}(1/R).\end{aligned}$$

Next, denote $\mathcal{C}_t = \mathcal{D}_{\tau,t}^{-1/2} A_t \mathcal{D}_{\tau,t}^{-1/2}$, then we can decompose the second part into two parts:

$$\sup_t \|\mathcal{I}_t - \mathcal{W}_t\|_\infty = \sup_t \|\mathcal{X}(L_{\tau,t} - \mathcal{L}_{\tau,t})\mathcal{X}\|_\infty \le \sup_t \|\mathcal{X}(L_{\tau,t} - \mathcal{C}_t)\mathcal{X}\|_\infty + \sup_t \|\mathcal{X}(\mathcal{C}_t - \mathcal{L}_{\tau,t})\mathcal{X}\|_\infty.$$

Then, for part one, we have

$$\begin{aligned}\sup_t \|\mathcal{X}(L_{\tau,t} - \mathcal{C}_t)\mathcal{X}\|_\infty &= \sup_t \max_{s,r} \left| \sum_{i,j} \mathcal{X}_{is}\mathcal{X}_{jr} \frac{A_t(i,j)}{\sqrt{\mathcal{D}_{\tau,t}(i,i)\mathcal{D}_{\tau,t}(j,j)}} \left(\frac{\sqrt{\mathcal{D}_{\tau,t}(i,i)\mathcal{D}_{\tau,t}(j,j)}}{\sqrt{D_{\tau,t}(i,i)D_{\tau,t}(j,j)}} - 1 \right) \right| \\ &\le \frac{1}{\delta} \max_{s,r} \sum_{i,j} |\mathcal{X}_{is}\mathcal{X}_{jr}| \sup_t \left\{ \max\left(\left| \frac{\mathcal{D}_{\tau,t}(i,i)}{D_{\tau,t}(i,i)} - 1 \right|, \left| \frac{\mathcal{D}_{\tau,t}(j,j)}{D_{\tau,t}(j,j)} - 1 \right| \right) \right\} \\ &= \max_{s,r} \sum_{i,j} |\mathcal{X}_{is}\mathcal{X}_{jr}| \mathcal{O}_p(\underline{\delta}^{-3/2} \log(TR)) \\ &= \mathcal{O}_p\left(\frac{NRJ^2}{\underline{\delta}^{3/2}} \log(TR) \right),\end{aligned}$$

where the second to the last equality comes from the following proof. For any $i \in \{1, \cdots, N\}$ and $\varsigma = \underline{\delta}^{-1/2} \log(TR)$, from Bernstein inequality,

$$\begin{aligned}\Pr\left(\sup_t \left| \frac{D_{\tau,t}(i,i)}{\mathcal{D}_{\tau,t}(i,i)} - 1 \right| > \varsigma \right) &\le \sum_{t=1}^{T} \Pr\left(\left| \frac{D_{\tau,t}(i,i)}{\mathcal{D}_{\tau,t}(i,i)} - 1 \right| > \varsigma \right) \\ &\le 2T \exp\left\{ -\frac{\varsigma^2 \mathcal{D}_{\tau,t}(i,i)}{2 + \frac{2}{3}\varsigma} \right\} \\ &\le 2T \exp\left\{ -\frac{\varsigma^2 \underline{\delta}}{2 + \frac{2}{3}\varsigma} \right\} \\ &= \mathcal{O}(1/R).\end{aligned}$$

For part two, similarly, we have

$$\begin{aligned}\sup_t \|\mathcal{X}(\mathcal{C}_t - \mathcal{L}_{\tau,t})\mathcal{X}\|_\infty &= \sup_t \max_{s,r} \left| \sum_{i,j} \mathcal{X}_{is}\mathcal{X}_{jr} \frac{A_t(i,j) - \mathcal{A}_t(i,j)}{\sqrt{\mathcal{D}_{\tau,t}(i,i)\mathcal{D}_{\tau,t}(j,j)}} \right| \\ &\le \max_{s,r} \left| \sum_{i,j} \mathcal{X}_{is}\mathcal{X}_{jr} \right| \sup_t \max_{i,j} \left| \frac{A_t(i,j) - \mathcal{A}_t(i,j)}{\sqrt{\mathcal{D}_{\tau,t}(i,i)\mathcal{D}_{\tau,t}(j,j)}} \right| \\ &= \mathcal{O}_p\left(\frac{NRJ^2}{\underline{\delta}} \right).\end{aligned}$$

Note that $\varsigma \to 0$ as $\underline{\delta}, R \to \infty$, we then know

$$\sup_t \|\mathcal{I}_t - \mathcal{W}_t\|_\infty = \mathcal{O}_p\left(\frac{NRJ^2}{\underline{\delta}}\right).$$

Thus, by union bounds, we obtain

$$\sup_t \|W_t - \mathcal{W}_t\|_\infty = \mathcal{O}_p\left(\zeta + \frac{NRJ^2}{\underline{\delta}}\right) = \mathcal{O}_p(\xi).$$

□

Lemma 2. *Under Assumption 4, for any $\epsilon > 0$, we have*

$$\sup_t \|S_t - \mathcal{S}_t\| \leq (4 + c_w)\left\{\frac{3\log(8NT/\epsilon)}{\underline{\delta}}\right\}^{1/2}, \tag{1}$$

with probability at least $1 - \epsilon$.

Proof. Note by triangular inequality, we have

$$\sup_t \|S_t - \mathcal{S}_t\| \leq \sup_t \left\|\alpha_t X W_t X^\top - \alpha_t \mathcal{X}\mathcal{W}_t\mathcal{X}^\top\right\| \tag{2}$$

$$+ \sup_t \left\|\mathcal{D}_{\tau,t}^{-1/2} A_t \mathcal{D}_{\tau,t}^{-1/2} - \mathcal{D}_{\tau,t}^{-1/2}\mathcal{A}_t\mathcal{D}_{\tau,t}^{-1/2}\right\| \tag{3}$$

$$+ \sup_t \left\|D_{\tau,t}^{-1/2} A_t D_{\tau,t}^{-1/2} - \mathcal{D}_{\tau,t}^{-1/2} A_t \mathcal{D}_{\tau,t}^{-1/2}\right\|. \tag{4}$$

For equation (2), we have,

$$\begin{aligned}\sup_t \left\|\alpha_t X W_t X^\top - \alpha_t \mathcal{X}\mathcal{W}_t\mathcal{X}^\top\right\| &= \sup_t \left\|\alpha_t X(W_t - \mathcal{W}_t)X^\top\right\| + \sup_t \left\|\alpha_t X\mathcal{W}_t X^\top - \alpha_t\mathcal{X}\mathcal{W}_t\mathcal{X}^\top\right\| \\ &\leq \alpha_{\max} NRJ^2 \sup_t \|W_t - \mathcal{W}_t\| + 2\alpha_{\max} NRJ^2 \sup_t \|\mathcal{W}_t\| \\ &= \mathcal{O}_p(\alpha_{\max} NRJ^2\xi).\end{aligned}$$

So, by Assumption 4 we know, for large enough N, with probability at least $1 - \epsilon/2$,

$$\sup_t \left\|\alpha_t X W_t X^\top - \alpha_t \mathcal{X}\mathcal{W}_t\mathcal{X}^\top\right\| \leq c_w a$$

For equation (3), let $Y_t(i,j) = \mathcal{D}_{\tau,t}^{-1/2}[(A_t(i,j) - p_t(i,j))E_{ij}]\mathcal{D}_{\tau,t}^{-1/2}$ with $E_{ij} \in \mathbb{R}^{N\times N}$ being the matrix with 1 in ij and ji'th positions and 0 everywhere else. Then we know

$$\sup_t \|Y_t(i,j)\| \leq \sup_t \sqrt{\mathcal{D}_{\tau,t}(i,i)\mathcal{D}_{\tau,t}(j,j)} \leq \frac{1}{\underline{\delta}}, \quad v^2 = \sup_t \|\sum \mathsf{E}(Y_t^2(i,j))\| \leq \frac{1}{\underline{\delta}}.$$

So, denote $a = \left\{ \frac{3\log(8NT/\epsilon)}{\underline{\delta}} \right\}^{1/2}$, which is smaller than 1 by assumption, and by matrix Bernstein inequality, we have

$$\begin{aligned}
&\Pr(\sup_t \|\mathcal{D}_{\tau,t}^{-1/2}[A_t(i,j) - \mathcal{A}_t(i,j)]\mathcal{D}_{\tau,t}^{-1/2}\| > a) \\
&\leq \sum_{t=1}^{T} \Pr(\|\mathcal{D}_{\tau,t}^{-1/2}[A_t(i,j) - \mathcal{A}_t(i,j)]\mathcal{D}_{\tau,t}^{-1/2}\| > a) \\
&\leq 2NT \exp\left(-\frac{a^2}{2/\underline{\delta} + 2a/3\underline{\delta}} \right) \\
&\leq 2NT \exp\left(-\frac{3\log(8NT/\epsilon)}{3} \right) \\
&= \epsilon/4.
\end{aligned}$$

Hence, with probability at least $1 - \epsilon/4$,

$$\sup_t \|\mathcal{D}_{\tau,t}^{-1/2} A_t \mathcal{D}_{\tau,t}^{-1/2} - \mathcal{D}_{\tau,t}^{-1/2} \mathcal{A}_t \mathcal{D}_{\tau,t}^{-1/2}\| \leq a \tag{5}$$

Lastly, for equation (4), by Qin and Rohe (2013) and setting $\lambda = a\mathcal{D}_{\tau,t}(i,i)$ we have

$$\begin{aligned}
\Pr(|D_{\tau,t}(i,i) - \mathcal{D}_{\tau,t}(i,i)| \geq \lambda) &\leq \exp\left\{ -\frac{\lambda^2}{2\mathcal{D}_{\tau,t}(i,i)} \right\} + \exp\left\{ -\frac{\lambda^2}{2\mathcal{D}_{\tau,t}(i,i) + \frac{2}{3}\lambda} \right\} \\
&\leq 2\exp\left\{ -\frac{\lambda^2}{2\mathcal{D}_{\tau,t}(i,i) + \frac{2}{3}\lambda} \right\} \\
&= 2\exp\left\{ -\frac{a^2 \mathcal{D}_{\tau,t}(i,i)}{2 + \frac{2}{3}a} \right\} \\
&\leq 2\exp\left\{ -\log(8NT/\epsilon) \times \frac{\mathcal{D}_{\tau,t}(i,i)}{\underline{\delta}} \right\} \\
&\leq \frac{\epsilon}{4NT}.
\end{aligned}$$

Further note that

$$\begin{aligned}
\Pr\left(\sup_t \|\mathcal{D}_{\tau,t}^{-1/2} D_{\tau,t}^{1/2} - I\| \geq a \right) &\leq \sum_{t=1}^{T} \Pr\left(\|\mathcal{D}_{\tau,t}^{-1/2} D_{\tau,t}^{1/2} - I\| \geq a \right) \\
&\leq \sum_{t=1}^{T} \Pr\left(\max_i \left| \frac{D_{\tau,t}(i,i)}{\mathcal{D}_{\tau,t}(i,i)} - 1 \right| \geq a \right) \\
&\leq \sum_{t=1}^{T} \sum_{i=1}^{N} \Pr\left(|D_{\tau,t}(i,i) - \mathcal{D}_{\tau,t}(i,i)| \geq a\mathcal{D}_{\tau,t}(i,i) \right) \\
&\leq NT \times \frac{\epsilon}{4NT} \\
&= \epsilon/4.
\end{aligned}$$

Therefore, with probability at least $1-\epsilon/4$, we have

$$\begin{aligned}
&\sup_t \|D_{\tau,t}^{-1/2} A_t D_{\tau,t}^{-1/2} - \mathcal{D}_{\tau,t}^{-1/2} A_t \mathcal{D}_{\tau,t}^{-1/2}\| \\
&\quad = \sup_t \|L_{\tau,t} - \mathcal{D}_{\tau,t}^{-1/2} D_{\tau,t}^{1/2} L_{\tau,t} D_{\tau,t}^{1/2} \mathcal{D}_{\tau,t}^{-1/2}\| \\
&\quad = \sup_t \|(I - \mathcal{D}_{\tau,t}^{-1/2} D_{\tau,t}^{1/2}) L_{\tau,t} D_{\tau,t}^{1/2} \mathcal{D}_{\tau,t}^{-1/2} + L_{\tau,t}(I - D_{\tau,t}^{1/2} \mathcal{D}_{\tau,t}^{-1/2})\| \\
&\quad \le \sup_t \|\mathcal{D}_{\tau,t}^{-1/2} D_{\tau,t}^{1/2} - I\| \sup_t \|\mathcal{D}_{\tau,t}^{-1/2} D_{\tau,t}^{1/2}\| + \sup_t \|\mathcal{D}_{\tau,t}^{-1/2} D_{\tau,t}^{1/2} - I\| \\
&\quad \le a^2 + 2a
\end{aligned}$$

where the second last inequality comes from the fact that $\sup_t \|L_{\tau,t}\| \le 1$.

Therefore, joining the results for these three equations, we have, with probability at least $1-\epsilon$,

$$\sup_t \|S_t - \mathcal{S}_t\| \le a^2 + 3a + c_w a \le (4 + c_w) a = (4 + c_w) \left\{ \frac{3 \log(8NT/\epsilon)}{\underline{\delta}} \right\}^{1/2}. \tag{6}$$

□

Lemma 3. *Under the dynamic SC-DCBM with K blocks, define $\Gamma_{\tau,t} \in \mathbb{R}^{N \times K}$ with columns containing the top K eigenvectors of $\mathcal{S}_t$. Then, under Assumption 4, there exists an orthogonal matrix U_t depending on τ_t for each $t = 1, \cdots, T$, such that for any $i, j = 1, \cdots, N$,*

$$\Gamma_{\tau,t} = \Psi_{\tau,t}^{1/2} Z_t (Z_t^\top \Psi_{\tau,t} Z_t)^{-1/2} U_t \;\; \textit{and} \;\; \Gamma_{\tau,t}^*(i,*) = \Gamma_{\tau,t}^*(j,*) \Longleftrightarrow Z_t(i,*) = Z_t(j,*),$$

where $\Gamma_{\tau,t}^*(i,*) = \Gamma_{\tau,t}(i,*)/\|\Gamma_{\tau,t}(i,*)\|$.

Proof. Denote $D_{B,t}$ as a diagonal matrix with entries $D_{B,t}(i,i) = \sum_{j=1}^K B_t(i,j)$, and $\Psi_{\tau,t} = \mathrm{Diag}(\psi_{\tau,t})$ with $\psi_{\tau,t}(i) = \psi_t \frac{\mathcal{D}_t(i,i)}{\mathcal{D}_{\tau,t}(i,i)}$. Then, Under the dynamic SC-DCBM, we have the decomposition below

$$\mathcal{L}_{\tau,t} = \mathcal{D}_{\tau,t}^{-1/2} \mathcal{A}_t \mathcal{D}_{\tau,t}^{-1/2} = \Psi_{\tau,t}^{1/2} Z_t B_{L,t} Z_t^\top \Psi_{\tau,t}^{1/2},$$

where $B_{L,t} = D_{B,t}^{-1/2} B_t D_{B,t}^{-1/2}$.

Define M_t such that $\mathcal{X} = \mathsf{E}(X) = \Psi_{\tau,t}^{1/2} Z_t M_t$, and $\Omega_t = B_{L,t} + \alpha_t M_t \mathcal{W}_t M_t^\top$, then we know

$$\mathcal{S}_t = \Psi_{\tau,t}^{1/2} Z_t \Omega_t Z_t^\top \Psi_{\tau,t}^{1/2}. \tag{7}$$

Now, denote $Y_{\tau,t} = Z_t^\top \Psi_{\tau,t} Z_t$, and let $H_{\tau,t} = Y_{\tau,t}^{1/2} \Omega_t Y_{\tau,t}^{1/2}$. Then, by eigen-decomposition, we have $H_{\tau,t} = U_t \Lambda_t U_t^\top$. Define $\Gamma_{\tau,t} = \Psi_{\tau,t}^{1/2} Z_t Y_{\tau,t}^{-1/2} U_t$, then

$$\begin{aligned}
\Gamma_{\tau,t}^\top \Gamma_{\tau,t} &= U_t^\top Y_{\tau,t}^{-1/2} Z_t^\top \Psi_{\tau,t}^{1/2} \Psi_{\tau,t}^{1/2} Z_t Y_{\tau,t}^{-1/2} U_t \\
&= U_t^\top Y_{\tau,t}^{-1/2} Y_{\tau,t} Y_{\tau,t}^{-1/2} U_t \\
&= U_t^\top U_t = I,
\end{aligned}$$

and we have

$$\begin{aligned}
\mathcal{S}_t \mathit{\Gamma}_{\tau,t} &= (\Psi_{\tau,t}^{1/2} Z_t \Omega_t Z_t^\top \Psi_{\tau,t}^{1/2}) \Psi_{\tau,t}^{1/2} Z_t (Z_t^\top \Psi_{\tau,t} Z_t)^{-1/2} U_t \\
&= \Psi_{\tau,t}^{1/2} Z_t \Omega_t Y_{\tau,t}^{1/2} U_t \\
&= \left\{ \Psi_{\tau,t}^{1/2} Z_t Y_{\tau,t}^{-1/2} \left(Y_{\tau,t}^{1/2} \Omega_t Y_{\tau,t}^{1/2} \right) \right\} U_t \\
&= \Psi_{\tau,t}^{1/2} Z_t Y_{\tau,t}^{-1/2} (U_t \Lambda_t U_t^\top) U_t \\
&= \mathit{\Gamma}_{\tau,t} \Lambda_t.
\end{aligned}$$

Following Qin and Rohe (2013), it is obvious that

$$\mathit{\Gamma}_{\tau,t}^*(i,*) = \frac{\mathit{\Gamma}_{\tau,t}(i,*)}{\|\mathit{\Gamma}_{\tau,t}(i,*)\|} = Z_{i,t} U_t.$$

Then, by directly applying the Lemma 1 in Binkiewicz et al. (2017), we complete the proof. □

Lemma 4. *Under Assumption 4', for any $\epsilon > 0$, we have*

$$\sup_t \|sym\,(S_t - \mathcal{S}_t)\| \leq \delta_{\max} \{3 \log(16NT/\epsilon)\}^{1/2}, \tag{8}$$

with probability at least $1 - \epsilon$

Proof. By triangular inequality, we have

$$\begin{aligned}
\sup_t \|\mathrm{sym}\,(S_t - \mathcal{S}_t)\| \leq & \sup_t \left\|\mathrm{sym}\left(\alpha_t X W_t X^\top - \alpha_t \mathcal{X} \mathcal{W}_t \mathcal{X}^\top\right)\right\| && (9) \\
& + \sup_t \left\|\mathrm{sym}\left(\mathcal{D}_{R,t}^{-1/2} A_t \mathcal{D}_{C,t}^{-1/2} - \mathcal{D}_{R,t}^{-1/2} \mathcal{A}_t \mathcal{D}_{C,t}^{-1/2}\right)\right\| && (10) \\
& + \sup_t \left\|\mathrm{sym}\left(D_{R,t}^{-1/2} A_t D_{C,t}^{-1/2} - \mathcal{D}_{R,t}^{-1/2} A_t \mathcal{D}_{C,t}^{-1/2}\right)\right\|. && (11)
\end{aligned}$$

For equation (9), by similar results in proof of Lemma 2, the spectral norm of the symmetrized $\alpha_t X W_t X^\top - \alpha_t \mathcal{X} \mathcal{W}_t \mathcal{X}^\top$ is bounded by

$$\begin{aligned}
&\sup_t \left\|\mathrm{sym}\left(\alpha_t X W_t X^\top - \alpha_t \mathcal{X} \mathcal{W}_t \mathcal{X}^\top\right)\right\| \\
&\quad = \alpha_{\max} \sup_t \left\|\mathrm{sym}\left(X(W_t - \mathcal{W}_t) X^\top\right)\right\| + \alpha_{\max} \sup_t \left\|\mathrm{sym}\left(X \mathcal{W}_t X^\top - \mathcal{X} \mathcal{W}_t \mathcal{X}^\top\right)\right\| \\
&\quad \leq \alpha_{\max} NRJ^2 \sup_t \|\mathrm{sym}\,(W_t - \mathcal{W}_t)\| + 2\alpha_{\max} NRJ^2 \sup_t \|\mathrm{sym}\,(\mathcal{W}_t)\| \\
&\quad = \mathcal{O}_p(\alpha_{\max} NRJ^2 \xi).
\end{aligned}$$

So, by Assumption 4', we know that for large enough N, with probability at least $1 - \epsilon/2$,

$$\sup_t \left\|\mathrm{sym}\left(\alpha_t X W_t X^\top - \alpha_t \mathcal{X} \mathcal{W}_t \mathcal{X}^\top\right)\right\| \leq c'_w a.$$

For equation (10), by Assumption 4' and matrix Bernstein inequality, we know under assumption $\underline{\delta}' > 3\log(16NT/\epsilon)$, $a < 1$. Therefore, similar to proof of Lemma 2, we have

$$\begin{aligned}
&\Pr\left(\sup_t \left\|\text{sym}\left(\mathcal{D}_{R,t}^{-1/2} A_t \mathcal{D}_{C,t}^{-1/2} - \mathcal{D}_{R,t}^{-1/2} \mathcal{A}_t \mathcal{D}_{C,t}^{-1/2}\right)\right\| > a\right) \\
&\leq \sum_{t=1}^{T} \Pr\left(\left\|\text{sym}\left(\mathcal{D}_{R,t}^{-1/2} A_t \mathcal{D}_{C,t}^{-1/2} - \mathcal{D}_{R,t}^{-1/2} \mathcal{A}_t \mathcal{D}_{C,t}^{-1/2}\right)\right\| > a\right) \\
&\leq 4NT \exp\left(-\frac{3\log(16NT/\epsilon)/\underline{\delta}'}{2/\underline{\delta}' + 2a/(3\underline{\delta}')}\right) \\
&\leq 4NT \exp\left(-\log(16NT/\epsilon)\right) \\
&= \epsilon/4.
\end{aligned}$$

Lastly, for equation (11), by Rohe et al. (2016), we know with probability at least $1 - \epsilon/2$,

$$\begin{aligned}
&\sup_t \left\|\text{sym}\left(D_{R,t}^{-1/2} A_t D_{C,t}^{-1/2} - \mathcal{D}_{R,t}^{-1/2} A_t \mathcal{D}_{C,t}^{-1/2}\right)\right\| \\
&= \sup_t \left\|\text{sym}\left(L_{\tau,t} - \mathcal{D}_{\tau,t}^{-1/2} D_{\tau,t}^{1/2} L_{\tau,t} D_{\tau,t}^{1/2} \mathcal{D}_{\tau,t}^{-1/2}\right)\right\| \\
&= \sup_t \left\|\text{sym}\left((I - \mathcal{D}_{\tau,t}^{-1/2} D_{\tau,t}^{1/2}) L_{\tau,t} D_{\tau,t}^{1/2} \mathcal{D}_{\tau,t}^{-1/2} + L_{\tau,t}(I - D_{\tau,t}^{1/2} \mathcal{D}_{\tau,t}^{-1/2})\right)\right\| \\
&\leq \sup_t \left\|\text{sym}\left(\mathcal{D}_{\tau,t}^{-1/2} D_{\tau,t}^{1/2} - I\right)\right\| \sup_t \left\|\text{sym}\left(\mathcal{D}_{\tau,t}^{-1/2} D_{\tau,t}^{1/2}\right)\right\| + \sup_t \left\|\text{sym}\left(\mathcal{D}_{\tau,t}^{-1/2} D_{\tau,t}^{1/2} - I\right)\right\| \\
&\leq a^2 + 2a
\end{aligned}$$

Therefore, combine the results above, we obtain the upper bound for $\|\text{sym}\,(S_t - \mathcal{S}_t)\|$, i.e., with probability at least $1 - \epsilon$,

$$\sup_t \|\text{sym}\,(S_t - \mathcal{S}_t)\| \leq a^2 + 3a + c'_w a \leq (4 + c'_w)a = (4 + c'_w)\left\{\frac{3\log(16N/\epsilon)}{\underline{\delta}'}\right\}^{1/2}. \quad (12)$$

□

Lemma 5. *Under the dynamic SC-DCcBM with K_R row blocks and K_C column blocks, define $\Gamma_{R,t} \in \mathbb{R}^{N \times K_R}$ with columns containing the top K_R left singular vectors of $\mathcal{S}_t$ and $\Gamma_{C,t} \in \mathbb{R}^{N \times K_C}$ with columns containing the top K_C right singular vectors of $\mathcal{S}_t$. Then, under Assumption 4', there exist orthogonal matrices $U_{R,t}$ and $U_{C,t}$ depending on τ_t for each $t = 1, \cdots, T$, such that for any $i, j = 1, \cdots, N$,*

$$\Gamma_{p,t} = {\Psi_{\tau,t}^{p}}^{1/2} Z_{p,t} (Z_{p,t}^{\top} {\Psi_{\tau,t}^{p}}^{1/2} Z_{p,t})^{-1/2} U_{p,t}$$

and

$$\Gamma_{p,t}^{*}(i,*) = \Gamma_{p,t}^{*}(j,*) \iff Z_{p,t}(i,*) = Z_{p,t}(j,*).$$

where $\Gamma_{p,t}^{*}(i,*) = \Gamma_{p,t}(i,*)/\|\Gamma_{p,t}(i,*)\|$ with $p \in \{R, C\}$.

Proof. Define $D^R_{B,t}$ and $D^C_{B,t}$ are diagonal matrices with entries $D^R_{B,t}(i,i) = \sum_{j=1}^K B_t(i,j)$ and $D^C_{B,t}(i,i) = \sum_j^K B_t(j,i)$, and $\Psi^p_{\tau,t} = \mathrm{Diag}(\psi^p_{\tau,t})$ with $\psi^p_{\tau,t}(i) = \psi^p_i \dfrac{\mathcal{D}_{p,t}(i,i)}{\mathcal{D}_{p,t}(i,i) + \tau_{p,t}}$ for $p \in \{R, C\}$. Then under dynamic SC-DCcBM, we have the decomposition below,

$$\mathcal{L}_{\tau,t} = \mathcal{D}^{-1/2}_{R,t} \mathcal{A}_t \mathcal{D}^{-1/2}_{C,t} = {\Psi^R_{\tau,t}}^{1/2} Z_{R,t} B_{L,t} Z^\top_{C,t} {\Psi^C_{\tau,t}}^{1/2},$$

where $B_{L,t} = \left(D^R_{B,t}\right)^{-1/2} B_t \left(D^C_{B,t}\right)^{-1/2}$.

Define $M_{R,t}$ and $M_{C,t}$ such that $\mathcal{X} = \mathsf{E}(X) = {\Psi^R_{\tau,t}}^{1/2} Z_{R,t} M_{R,t} = {\Psi^C_{\tau,t}}^{1/2} Z_{C,t} M_{C,t}$, and $\Omega_t = B_{L,t} + \alpha_t M_{R,t} \mathcal{W}_t M^\top_{C,t}$, then we know

$$\mathcal{S}_t = {\Psi^R_{\tau,t}}^{1/2} Z_{R,t} \Omega_t Z^\top_{C,t} {\Psi^C_{\tau,t}}^{1/2}. \tag{13}$$

Now, denote $Y_{R,t} = Z^\top_{R,t} \Psi^R_{\tau,t} Z_{R,t}$ and $Y_{C,t} = Z^\top_{C,t} \Psi^C_{\tau,t} Z_{C,t}$, and let $H_{\tau,t} = Y^{1/2}_{R,t} \Omega_t Y^{1/2}_{C,t}$. Then, by singular value decomposition, we have $H_{\tau,t} = U_{R,t} \Lambda_t U^\top_{C,t}$. Define $\Gamma_{R,t} = {\Psi^R_{\tau,t}}^{1/2} Z_{R,t} Y^{-1/2}_{R,t} U_{R,t}$ and $\Gamma_{C,t} = {\Psi^C_{\tau,t}}^{1/2} Z_{C,t} Y^{-1/2}_{C,t} U_{C,t}$, then, for $p \in \{R, C\}$,

$$\begin{aligned}
\Gamma^\top_{p,t} \Gamma_{p,t} &= U^\top_{p,t} Y^{-1/2}_{p,t} Z^\top_{p,t} {\Psi^p_{\tau,t}}^{1/2} {\Psi^p_{\tau,t}}^{1/2} Z_{p,t} Y^{-1/2}_{p,t} U_{p,t} \\
&= U^\top_{p,t} Y^{-1/2}_{p,t} Y_{p,t} Y^{-1/2}_{p,t} U_{p,t} \\
&= U^\top_{p,t} U_{p,t} = I,
\end{aligned}$$

and we have

$$\begin{aligned}
&\Gamma_{R,t} \Lambda_t \Gamma_{C,t} = {\Psi^R_{\tau,t}}^{1/2} Z_{R,t} Y^{-1/2}_{R,t} U_{R,t} \Lambda_t U^\top_{C,t} Y^{-1/2}_{C,t} Z^\top_{C,t} {\Psi^C_{\tau,t}}^{1/2} \\
&= {\Psi^R_{\tau,t}}^{1/2} Z_{R,t} Y^{-1/2}_{R,t} H_{\tau,t} Y^{-1/2}_{C,t} Z^\top_{C,t} {\Psi^C_{\tau,t}}^{1/2} \\
&= {\Psi^R_{\tau,t}}^{1/2} Z_{R,t} Y^{-1/2}_{R,t} \left(Y^{1/2}_{R,t} \Omega_t Y^{1/2}_{C,t} \right) Y^{-1/2}_{C,t} Z^\top_{C,t} {\Psi^C_{\tau,t}}^{1/2} \\
&= {\Psi^R_{\tau,t}}^{1/2} Z_{R,t} \Omega_t Z^\top_{C,t} {\Psi^C_{\tau,t}}^{1/2} = \mathcal{S}_t
\end{aligned}$$

Following Rohe et al. (2016), it is obvious that

$$\Gamma^*_{p,t}(i,*) = \frac{\Gamma_{p,t}(i,*)}{\|\Gamma_{p,t}(i,*)\|} = Z_{p,t}(i,*) U_{p,t}, \text{ for } p \in \{R, C\},$$

which completes the proof. □

2 Proof of Theorem 1

Proof. By Binkiewicz et al. (2017) and the solution of $(1+\varepsilon)$-approximate k-means method, we know for each period $t = 1, 2, \cdots, T$, we have

$$\frac{|\mathbb{M}_t|}{N} \leq \frac{2(2+\varepsilon)^2}{m_z^2 N} \|U_t - \mathcal{U}_t \mathcal{O}_t\|^2_F \tag{14}$$

and

$$\|U_t - \mathcal{U}_t\mathcal{O}_t\|_F \leq \frac{8K^{1/2}}{\lambda_{K,t}} \left\|\widehat{\mathcal{S}}_{t,r} - \mathcal{S}_t\right\|, \tag{15}$$

where $m_z \stackrel{\text{def}}{=} \min_{i,t}\{\min\{\|\Gamma_{\tau,t}(i,*)\|, \|\mathit{\Gamma}_{\tau,t}(i,*)\|\}\}$ with $\Gamma_{\tau,t}$ and $\mathit{\Gamma}_{\tau,t}$ being defined in Lemma 3.

Then, we have

$$\sup_t \frac{|\mathbb{M}_t|}{N} \leq \frac{2^9(2+\varepsilon)^2K}{m_z^2 N\lambda_{K,\max}^2} \sup_t \left\|\widehat{\mathcal{S}}_{t,r} - \mathcal{S}_t\right\|^2. \tag{16}$$

Then, for $\mathcal{S}_t$, we have the following representation:

$$\mathcal{S}_t = \mathcal{D}_{\tau,t}^{-1/2}\Psi Z_t B_t Z_t^\top \Psi \mathcal{D}_{\tau,t}^{-1/2} + \alpha_t \mathcal{X}\mathcal{W}_t\mathcal{X}^\top, \tag{17}$$

To figure out the upper bound of the estimation error, we have to evaluate the error bound $\sup_t \left\|\widehat{\mathcal{S}}_{t,r} - \mathcal{S}_t\right\|$. Define

$$\mathcal{S}_{t,r} = \frac{1}{|\mathcal{F}_r|}\sum_{i\in\mathcal{F}_r} W_{r,l}(i)\mathcal{S}_{t+i}, \tag{18}$$

then by triangle inequality, we have

$$\Delta(r) = \sup_t \left\|\widehat{\mathcal{S}}_{t,r} - \mathcal{S}_t\right\| \leq \sup_t \left\|\widehat{\mathcal{S}}_{t,r} - \mathcal{S}_{t,r}\right\| + \sup_t \|\mathcal{S}_{t,r} - \mathcal{S}_t\| = \Delta_1(r) + \Delta_2(r). \tag{19}$$

For $\Delta_1(r)$, by Lemma 2, we have

$$\begin{aligned}
\Delta_1(r) &= \frac{1}{|\mathcal{F}_r|}\sum_{i\in\mathcal{F}_r} W_{r,l}(i) \sup_t \|S_{t+i} - \mathcal{S}_{t+i}\| \\
&\leq \frac{1}{|\mathcal{F}_r|}\sum_{i\in\mathcal{F}_r} W_{r,l}(i) \left\{(4+c_w)\left[\frac{3\log(8NT/\epsilon)}{\underline{\delta}}\right]^{1/2}\right\} \\
&\leq W_{\max}(4+c_w)\left\{\frac{3\log(8NT/\epsilon)}{\underline{\delta}}\right\}^{1/2}.
\end{aligned} \tag{20}$$

For $\Delta_2(r)$, we have the following decomposition

$$\Delta_2(r) = \sup_t \|\mathcal{S}_{t,r} - \mathcal{S}_t\| \leq \sup_t \left\|\mathcal{S}_{t,r} - \widetilde{\mathcal{S}}_{t,r}\right\| + \sup_t \left\|\widetilde{\mathcal{S}}_{t,r} - \mathcal{S}_t\right\| = \Delta_{21}(r) + \Delta_{22}(r), \tag{21}$$

where

$$\widetilde{\mathcal{S}}_{t,r} = \frac{1}{|\mathcal{F}_r|}\sum_{i\in\mathcal{F}_r} W_{r,l}(i)\left(\mathcal{D}_{\tau,t}^{-1/2}\Psi Z_t B_{t+i} Z_t^\top \Psi \mathcal{D}_{\tau,t}^{-1/2} + \alpha_{t+i}\mathcal{X}\mathcal{W}_{t+i}\mathcal{X}^\top\right). \tag{22}$$

Then for Δ_{21}, we have

$$\begin{aligned}
\Delta_{21}(r) &\leq W_{\max}\frac{1}{|\mathcal{F}_r|}\sum_{i\in\mathcal{F}_r}\sup_t \left\|\mathcal{D}_{\tau,t+i}^{-1/2}\Psi Z_{t+i}B_{t+i}Z_{t+i}^\top\Psi\mathcal{D}_{\tau,t+i}^{-1/2} - \mathcal{D}_{\tau,t}^{-1/2}\Psi Z_t B_{t+i}Z_t^\top\Psi\mathcal{D}_{\tau,t}^{-1/2}\right\| \\
&\leq W_{\max}\frac{1}{|\mathcal{F}_r|}\sum_{i\in\mathcal{F}_r}\sup_t \left\{\left(\left\|\mathcal{D}_{\tau,t+i}^{-1/2}\Psi Z_{t+i}\right\| + \left\|\mathcal{D}_{\tau,t}^{-1/2}\Psi Z_t\right\|\right)\|B_{t+i}\|\left\|\mathcal{D}_{\tau,t+i}^{-1/2}\Psi Z_{t+i} - \mathcal{D}_{\tau,t}^{-1/2}\Psi Z_t\right\|\right\} \\
&\leq W_{\max}\frac{1}{|\mathcal{F}_r|}\sum_{i\in\mathcal{F}_r}\sup_t \left\{\left(\left\|\mathcal{D}_{\tau,t}^{-1/2}\right\|\|Z_{t+i}\| + \left\|\mathcal{D}_{\tau,t}^{-1/2}\right\|\|Z_t\|\right)\|B_{t+i}\|\left\|\mathcal{D}_{\tau,t+i}^{-1/2}\Psi Z_{t+i} - \mathcal{D}_{\tau,t}^{-1/2}\Psi Z_t\right\|\right\},
\end{aligned}$$

where the last inequality comes from the fact that $\|\Psi\| = \max_i \left|\sqrt{\psi_i}\right| \leq 1$.

Then, observe that $\sup_t \left\|\mathcal{D}_{\tau,t}^{-1/2}\right\| \leq \underline{\delta}^{-1/2}$, $\sup_t \|Z_t\| \leq P_{\max}^{1/2}$, $\sup_t \|B_t\| \leq K$, we then have

$$\sup_t \left\{ \left\|\mathcal{D}_{\tau,t}^{-1/2}\right\| \|Z_{t+i}\| + \left\|\mathcal{D}_{\tau,t}^{-1/2}\right\| \|Z_t\| \right\} \leq 2\underline{\delta}^{-1/2} P_{\max}^{1/2}. \tag{23}$$

Further, note that

$$\begin{aligned}
&\sup_t \left\|\mathcal{D}_{\tau,t+i}^{-1/2}\Psi Z_{t+i} - \mathcal{D}_{\tau,t}^{-1/2}\Psi Z_t\right\| \qquad (24)\\
&\quad \leq \sup_t \left\{ \left\|\mathcal{D}_{\tau,t+i}^{-1/2}\Psi Z_{t+i} - \mathcal{D}_{\tau,t+i}^{-1/2}\Psi Z_t\right\| + \left\|\mathcal{D}_{\tau,t+i}^{-1/2}\Psi Z_t - \mathcal{D}_{\tau,t}^{-1/2}\Psi Z_t\right\| \right\}\\
&\quad \leq \sup_t \left\{ \left\|\mathcal{D}_{\tau,t+i}^{-1/2}\right\| \|\Psi\| \|Z_{t+i} - Z_t\| + \left(\left\|\mathcal{D}_{\tau,t+i}^{-1/2}\right\| \|\Psi\| + \left\|\mathcal{D}_{\tau,t}^{-1/2}\right\| \|\Psi\| \right) \|Z_t\| \right\}\\
&\quad \leq \sqrt{\frac{2|r|s}{\underline{\delta}}} + \sqrt{\frac{4P_{\max}}{\underline{\delta}}}.
\end{aligned}$$

Then, combine the results above with the assumption $\underline{\delta} > 3\log(8NT/\epsilon)$ in Lemma 2, we have

$$\Delta_{21}(r) \leq \frac{2W_{\max}K}{\underline{\delta}}(\sqrt{2P_{\max}rs} + 2P_{\max}). \tag{25}$$

Lastly, for $\Delta_{22}(r)$, for notational simplicity, denote $Y_{\tau,t} \stackrel{\text{def}}{=} \mathcal{D}_{\tau,t}^{-1/2}\Psi Z_t$. Then, apply the results in Pensky and Zhang (2017) and proof of Lemma 2, we obtain

$$\begin{aligned}
\Delta_{22}(r) &= \sup_t \left\|\widetilde{\mathcal{S}}_{t,r} - \mathcal{S}_t\right\|\\
&= \frac{1}{|\mathcal{F}_r|}\sum_{i\in\mathcal{F}_r} W_{r,l}(i) \sup_t \left(Y_{\tau,t}\|B_{t+i} - B_t\| Y_{\tau,t}^\top + \left\|\alpha_{t+i}\mathcal{X}\mathcal{W}_{t+i}\mathcal{X}^\top - \alpha_t\mathcal{X}\mathcal{W}_t\mathcal{X}^\top\right\| \right)\\
&\leq \sup_t \left\{ \max_{1\leq j'\leq N} \sum_{j=1}^N \left|(Y_{\tau,t}Q_{r,t}Y_{\tau,t}^\top)(j,j')\right| \right\} + 2\alpha_{\max}W_{\max}NRJ^2 \sup_t \|\mathcal{W}_t\|\\
&\leq \sup_t \left\{ \max_{k,k'} |Q_{r,t}| \max_{1\leq j'\leq N} \sum_{k=1}^K \sum_{k'=1}^K \left[\sum_{j\in\mathcal{G}_{t,k}} Y_{\tau,t}(j,k) \right] Y_{\tau,t}(j',k') \right\} + 2W_{\max}\left\{ \frac{3\log(8NT/\epsilon)}{\underline{\delta}} \right\}^{1/2}\\
&\leq \frac{NLW_{\max}}{\underline{\delta}\cdot l!}\left(\frac{r}{T}\right)^\beta + 2W_{\max}\left\{ \frac{3\log(8NT/\epsilon)}{\underline{\delta}} \right\}^{1/2} \qquad (26)
\end{aligned}$$

where the second last inequality comes from Assumption 4 and the last inequality come from the fact that $\max_i \psi_i \leq 1$.

Now, combine the results provided by equation (16), (20), (25), and (26), we derive the upper bound for misclustering rate of dynamic DCBM: with probability at least $1-\epsilon$,

$$\sup_t \frac{|\mathbb{M}_t|}{N} \leq \frac{c_1(\varepsilon)KW_{\max}^2}{m_z^2 N\lambda_{K,\max}^2}\left\{ (6+c_w)\frac{b}{\underline{\delta}^{1/2}} + \frac{2K}{\underline{\delta}}(\sqrt{2P_{\max}rs} + 2P_{\max}) + \frac{NL}{\underline{\delta}\cdot l!}\left(\frac{r}{T}\right)^\beta \right\}^2.$$

where $b = \sqrt{3\log(8NT/\epsilon)}$, $\lambda_{K,\max} = \max_t\{\lambda_{K,t}\}$ and $c_1(\varepsilon) = 2^9(2+\varepsilon)^2$. □

3 Proof of Lemma 1

Proof. Firstly, by Lemma B.1 in Supplementary material of Lei and Rinaldo (2015), fix $\eta \in (0,1)$, we have

$$\left\|\widehat{\mathcal{S}}_{t,r} - \mathcal{S}_{t,r}\right\| \leq (1-\eta)^{-2} \sup_{x,y\in\mathcal{T}} \left|x^\top(\widehat{\mathcal{S}}_{t,r} - \mathcal{S}_{t,r})y\right|, \tag{27}$$

where $\mathcal{T} = \{x = (x_1, \cdots, x_N) \in \mathbb{R}^N, \|x\| = 1, \sqrt{N}x_i/\eta \in \mathbb{Z}, \forall i\}$. Then, let $d = rN\|\mathcal{S}_t\|_\infty$ with $r \geq 1$, we can split the pairs (x_i, y_j) into light pairs

$$\mathscr{L} = \mathscr{L}(x,y) \stackrel{\text{def}}{=} \{(i,j) : |x_i y_j| \leq \sqrt{d}/N\},$$

and into heavy pairs

$$\bar{\mathscr{L}} = \bar{\mathscr{L}}(x,y) \stackrel{\text{def}}{=} \{(i,j) : |x_i y_j| > \sqrt{d}/N\}.$$

For the light pair, first denote

$$u_{ij} = x_i y_j \mathbb{1}_{\{|x_i y_j| \leq \sqrt{d}/N\}} + x_j y_i \mathbb{1}_{\{|x_j y_i| \leq \sqrt{d}/N\}},$$

then we have

$$\begin{aligned}
&\sum_{(i,j)\in\mathscr{L}(x,y)} x_i y_j (\widehat{\mathcal{S}}_{t,r}(i,j) - \mathcal{S}_{t,r}(i,j)) \\
&\qquad = \frac{1}{|\mathcal{F}_r|} \sum_{1\leq i\leq j\leq N} \sum_{k\in\mathcal{F}_r} u_{ij} W_{r,\ell}(k) \left[S_{t+k}(i,j) - \mathcal{S}_{t+k}(i,j)\right].
\end{aligned}$$

Denote $w_{ij} = |\mathcal{F}_r|^{-1} \sum_{k\in\mathcal{F}_r} W_{r,\ell}(k) \left[S_{t+k}(i,j) - \mathcal{S}_{t+k}(i,j)\right]$ and $\xi_{ij} = w_{ij}u_{ij}$, then we have $|w_{ij}| \leq W_{\max}\|\mathcal{S}_t\|_\infty$, and by Pensky and Zhang (2017), it is known that ξ_{ij} is a independent random variable with zero mean and absolute values bounded by $|\xi_{ij}| \leq 2W_{\max}\sqrt{r\|\mathcal{S}_t\|_\infty^3/N}$, using the fact that $|u_{ij}| \leq 2\sqrt{d}/N$.

Now, applying Bernstein inequality, for any $c > 0$, we have

$$\begin{aligned}
&\Pr\left(\sup_{x,y\in\mathcal{T}} \left|\sum_{1\leq i\leq j\leq N} \xi_{ij}\right| \geq \frac{c\sqrt{d}}{r}\right) \\
&\leq 2\exp\left(-\frac{\frac{c^2 d}{2r}}{\sum_{1\leq i\leq j\leq N} \mathbb{E}\left(\xi_{ij}^2\right) + \frac{2W_{\max}}{3}\sqrt{\frac{r\|\mathcal{S}_t\|_\infty^3}{N}} \times \frac{c\sqrt{d}}{r}}\right) \\
&\leq 2\exp\left(-\frac{\frac{c^2 d}{2r}}{\left(\sum_{1\leq i\leq j\leq N} u_{ij}^2\right) W_{\max}^2\|\mathcal{S}_t\|_\infty^2 + \frac{2W_{\max}}{3}\sqrt{\frac{r\|\mathcal{S}_t\|_\infty^3}{N}} \times \frac{c\sqrt{d}}{r}}\right) \\
&\leq 2\exp\left(-\frac{3c^2 N}{12W_{\max}^2\|\mathcal{S}_t\|_\infty + 4cW_{\max}\|\mathcal{S}_t\|_\infty}\right).
\end{aligned}$$

Then, by a standard volume argument, we have the cardinality of $\mathcal{T} \leq \exp(N \log(7/\eta))$, and this ensures

$$\begin{aligned}
&\Pr\left(\sup_{x,y\in\mathcal{T}}\left|\sum_{(i,j)\in\mathscr{L}(x,y)} x_i y_j(\widehat{\mathcal{S}}_{t,r}(i,j) - \mathcal{S}_{t,r}(i,j))\right| \geq \frac{c\sqrt{d}}{r}\right) \\
&\quad\leq \exp\left\{-\left(\frac{3c^2}{12W_{\max}^2\|\mathcal{S}_t\|_\infty + 4cW_{\max}\|\mathcal{S}_t\|_\infty} - 2\log\left(\frac{7}{\eta}\right)\right)N\right\}. \qquad (28)
\end{aligned}$$

For the heavy pairs, we know

$$\begin{aligned}
&\left|\sum_{(i,j)\in\bar{\mathscr{L}}(x,y)} x_i y_j(\widehat{\mathcal{S}}_{t,r}(i,j) - \mathcal{S}_{t,r}(i,j))\right| \\
&= \left|\frac{1}{|\mathcal{F}_r|}\sum_{(i,j)\in\bar{\mathscr{L}}(x,y)} x_i y_j \sum_{k\in\mathcal{F}_r} W_{r,\ell}(k)(S_{t+k}(i,j) - \mathcal{S}_{t+k}(i,j))\right| \\
&\leq \left|\frac{1}{|\mathcal{F}_r|}\sum_{(i,j)\in\bar{\mathscr{L}}(x,y)} \frac{x_i^2 y_j^2}{|x_i y_j|}\sum_{k\in\mathcal{F}_r} W_{r,\ell}(k)(S_{t+k}(i,j) - \mathcal{S}_{t+k}(i,j))\right| \\
&\leq \frac{N}{\sqrt{d}} W_{\max}\|\mathcal{S}_t\|_\infty \sum_{(i,j)\in\bar{\mathscr{L}}(x,y)} x_i^2 y_j^2 \\
&= \frac{W_{\max}}{r}\sqrt{d}\sum_{(i,j)\in\bar{\mathscr{L}}(x,y)} x_i^2 y_j^2 \\
&\leq \frac{W_{\max}}{r}\sqrt{d}.
\end{aligned}$$

Therefore, choosing $c = W_{\max}$ in equation (28), we have

$$\Pr\left(\sup_{x,y\in\mathcal{T}}\left|\sum_{1\leq i\leq j\leq N} x_i y_j(\widehat{\mathcal{S}}_{t,r}(i,j) - \mathcal{S}_{t,r}(i,j))\right| \leq \frac{W_{\max}\sqrt{d}}{r}\right) \geq 1-\epsilon \qquad (29)$$

where $\epsilon = N^{\left(\frac{3}{16\|\mathcal{S}_t\|_\infty} - 2\log\left(\frac{7}{\eta}\right)\right)}$.

In the end, by equation (27) and (29), we obtain, with probability $1-\epsilon$,

$$\left\|\widehat{\mathcal{S}}_{t,r} - \mathcal{S}_{t,r}\right\| \leq (1-\eta)^{-2}\sup_{x,y\in\mathcal{T}}\left|x^\top(\widehat{\mathcal{S}}_{t,r} - \mathcal{S}_{t,r})y\right| \leq (1-\eta)^{-2}\frac{W_{\max}\sqrt{d}}{r}.$$

□

4 Proof of Theorem 2

Proof. In this proof, we deal with the clustering of left singular vector and the right singular vectors separately.

(1) Clustering for $Z_{R,t}$. First, by Rohe et al. (2016) and solution of $(1+\varepsilon)$-approximate k-means clustering, for each period $t = 1, \cdots, T$, we have

$$\frac{\left|\mathbb{M}_t^R\right|}{N} \leq \frac{8(2+\varepsilon)^2}{m_r^2 N} \left\|U_t - \mathcal{U}_t\mathcal{O}_t\right\|_F^2, \tag{30}$$

where denote $m_r \overset{\text{def}}{=} \min_{i,t}\{\min\{\|\Gamma_{R,t}(i,*)\|, \|\mathit{\Gamma}_{R,t}(i,*)\|\}\}$, and by improved version of Davis-Kahn theorem from Lei and Rinaldo (2015), we have

$$\left\|U_t - \mathcal{U}_t\mathcal{O}_t\right\|_F \leq \frac{2\sqrt{2K_R}}{\lambda_{K_R,t}} \left\|\operatorname{sym}\left(\mathcal{S}_{t,r} - \mathcal{S}_t\right)\right\|, \tag{31}$$

as $K_R \leq K_C$.

Then, base on equation (30) and (31), we have

$$\sup_t \frac{\left|\mathbb{M}_t^R\right|}{N} \leq \frac{2^6(2+\varepsilon)^2 K_R}{m_r^2 N \lambda_{K_R,\max}^2} \sup_t \left\|\operatorname{sym}\left(\mathcal{S}_{t,r} - \mathcal{S}_t\right)\right\|^2. \tag{32}$$

Then, for $\mathcal{S}_t$, we have the following representation:

$$\mathcal{S}_t = \mathcal{D}_{R,t}^{-1/2} \mathit{\Psi}^R Z_{R,t} B_t Z_{C,t}^\top \mathit{\Psi}^C \mathcal{D}_{C,t}^{-1/2} + \alpha_t \mathcal{X}\mathcal{W}_t\mathcal{X}^\top, \tag{33}$$

where $\mathit{\Psi}^p = \operatorname{Diag}(\psi^p)$ with $p \in \{R, C\}$. Then, by definition of $\mathcal{S}_{t,r} \overset{\text{def}}{=} |\mathcal{F}_r|^{-1} \sum_{i\in\mathcal{F}_r} W_{r,\ell}(i)\mathcal{S}_{t+i}$, we have the decomposition

$$\Delta(r) = \sup_t \left\|\operatorname{sym}\left(\widehat{\mathcal{S}}_{t,r} - \mathcal{S}_t\right)\right\| \leq \sup_t \left\|\operatorname{sym}\left(\widehat{\mathcal{S}}_{t,r} - \mathcal{S}_{t,r}\right)\right\| + \sup_t \left\|\operatorname{sym}\left(\mathcal{S}_{t,r} - \mathcal{S}_t\right)\right\| = \Delta_1(r) + \Delta_2(r). \tag{34}$$

Now, we evaluate $\Delta_1(r)$ and $\Delta_2(r)$ respectively. For $\Delta_1(r)$, by Lemma 4, we have

$$\begin{aligned} \sup_t \left\|\operatorname{sym}\left(\widehat{\mathcal{S}}_{t,r} - \mathcal{S}_{t,r}\right)\right\| &= \frac{1}{|\mathcal{F}_r|} \sum_{i\in\mathcal{F}_r} W_{r,\ell}(i) \sup_t \left\|\operatorname{sym}\left(S_{t+i} - \mathcal{S}_{t+i}\right)\right\| \qquad (35) \\ &\leq W_{\max}(4 + c_w') \left\{\frac{3\log(16N/\epsilon)}{\underline{\delta}'}\right\}^{1/2}. \end{aligned}$$

For $\Delta_2(r)$, we first define

$$\widetilde{\mathcal{S}}_{t,r} = \frac{1}{|\mathcal{F}_r|} \sum_{i\in\mathcal{F}_r} W_{r,\ell}(i) \left(Y_{R,t} B_{t+i} Y_{C,t}^\top + \alpha_{t+i}\mathcal{X}\mathcal{W}_{t+i}\mathcal{X}\right). \tag{36}$$

where $Y_{R,t} \overset{\text{def}}{=} \mathcal{D}_{R,t}^{-1/2}\mathit{\Psi}^R Z_{R,t}$ and $Y_{C,t} \overset{\text{def}}{=} \mathcal{D}_{C,t}^{-1/2}\mathit{\Psi}^C Z_{C,t}$.

Then, we decompose $\Delta_2(r)$ as

$$\sup_t \left\|\operatorname{sym}\left(\mathcal{S}_{t,r} - \mathcal{S}_t\right)\right\| \leq \sup_t \left\|\operatorname{sym}\left(\mathcal{S}_{t,r} - \widetilde{\mathcal{S}}_{t,r}\right)\right\| + \sup_t \left\|\operatorname{sym}\left(\widetilde{\mathcal{S}}_{t,r} - \mathcal{S}_t\right)\right\| = \Delta_{21}(r) + \Delta_{22}(r). \tag{37}$$

For notation simplicity, we define $Y_{p,t} \stackrel{\text{def}}{=} \mathcal{D}_{p,t}^{-1/2} \Psi^p Z_{p,t}$ for $p \in \{R, C\}$, and the block diagonal matrix $\mathcal{Y}_t$ such that

$$\mathcal{Y}_t \stackrel{\text{def}}{=} \begin{bmatrix} Y_{R,t} & 0_{N \times K_C} \\ 0_{N \times K_R} & Y_{C,t} \end{bmatrix}. \tag{38}$$

Note that

$$\begin{aligned} \left\| \text{sym}\left(\mathcal{S}_{t,r} - \widetilde{\mathcal{S}}_{t,r} \right) \right\| &\leq W_{\max} \max_{|i| \leq r} \left\| \text{sym}\left(Y_{R,t+i} B_{t+i} Y_{C,t+i}^\top - Y_{R,t} B_{t+i} Y_{C,t}^\top \right) \right\| \\ &= W_{\max} \max_{|i| \leq r} \left\| \mathcal{Y}_{t+i} \text{sym}(B_{t+i}) \mathcal{Y}_{t+i}^\top - \mathcal{Y}_t \text{sym}(B_{t+i}) \mathcal{Y}_t^\top \right\| \\ &\leq W_{\max} \max_{|i| \leq r} \left(\|\mathcal{Y}_{t+i}\| + \|\mathcal{Y}_t\| \right) \|\text{sym}(B_{t+i})\| \, \|\mathcal{Y}_{t+i} - \mathcal{Y}_t\|, \end{aligned}$$

and $\|\text{sym}(B_{t+i})\| \leq K_C$ and $\|\Psi^p\| \leq 1$ for $p \in \{R, C\}$, we then have

$$\begin{aligned} \|\mathcal{Y}_t\| &= \max\{\|\mathcal{D}_{R,t}^{-1/2} \Psi^R Z_{R,t}\|, \|\mathcal{D}_{C,t}^{-1/2} \Psi^C Z_{C,t}\|\} \\ &\leq \max\{\|\mathcal{D}_{R,t}^{-1/2}\| \|\Psi^R\| \|Z_{R,t}\|, \|\mathcal{D}_{C,t}^{-1/2}\| \|\Psi^C\| \|Z_{C,t}\|\} \\ &\leq \underline{\delta}'^{-1/2} P_{\max}^{1/2}, \end{aligned}$$

and

$$\begin{aligned} \|\mathcal{Y}_{t+i} - \mathcal{Y}_t\| &= \max\left\{ \|Y_{R,t+i} - Y_{R,t}\|, \|Y_{C,t+i} - Y_{C,t}\| \right\} \\ &\leq \max\left\{ \left\| \mathcal{D}_{R,t+i}^{-1/2} Z_{R,t+i} - \mathcal{D}_{R,t}^{-1/2} Z_{C,t} \right\|, \left\| \mathcal{D}_{C,t+i}^{-1/2} Z_{C,t+i} - \mathcal{D}_{C,t}^{-1/2} Z_{C,t} \right\| \right\} \\ &\leq \sqrt{\frac{2|r|s}{\underline{\delta}'}} + \sqrt{\frac{4P_{\max}}{\underline{\delta}'}}, \end{aligned}$$

where the last inequality comes from the same derivation as equation (24).

Therefore, we have

$$\begin{aligned} \Delta_{21}(r) &= \sup_t \left\| \text{sym}\left(\mathcal{S}_{t,r} - \widetilde{\mathcal{S}}_{t,r} \right) \right\| \\ &\leq W_{\max} \times (2\underline{\delta}'^{-1/2} P_{\max}^{1/2}) \times K_C \times \left(\sqrt{\frac{2|r|s}{\underline{\delta}'}} + \sqrt{\frac{4P_{\max}}{\underline{\delta}'}} \right) \\ &= \frac{2W_{\max} K_C}{\underline{\delta}'} \left(\sqrt{2P_{\max} r s} + 2P_{\max} \right). \end{aligned} \tag{39}$$

Lastly, for $\Delta_{22}(r)$, define

$$\mathcal{Z}_t \stackrel{\text{def}}{=} \begin{bmatrix} \Psi_{\tau,t}^{R\,1/2} Z_{R,t} & 0_{N \times K_C} \\ 0_{N \times K_R} & \Psi_{\tau,t}^{C\,1/2} Z_{C,t} \end{bmatrix}. \tag{40}$$

Then, by Assumption 4', we have

$$\begin{aligned}\Delta_{22}(r) &= \sup_t \left\| \mathrm{sym}\left(\widetilde{\mathcal{S}}_{t,r} - \mathcal{S}_t\right)\right\| \\ &\le \frac{1}{|\mathcal{F}_r|}\sum_{i\in\mathcal{F}_r} W_{r,\ell}(i) \sup_t \left(\left\| \mathrm{sym}\left(Y_{R,t}(B_{t+i}-B_t)Y_{C,t}^\top\right)\right\| + \left\|\mathrm{sym}\left(\alpha_{t+i}\mathcal{X}\mathcal{W}_{t+i}\mathcal{X} - \alpha_t\mathcal{X}\mathcal{W}_t\mathcal{X}\right)\right\|\right) \\ &\le \frac{1}{|\mathcal{F}_{r,j}|}\sum_{i\in\mathcal{F}_{r,j}} W_{r,\ell}^j(i) \left\|\mathcal{Y}_t \mathrm{sym}(B_{t+i}-B_t)\mathcal{Y}_t^\top\right\| + 2\alpha_{\max}W_{\max}NRJ^2 \sup_t \|\mathrm{sym}(\mathcal{W}_t)\| \\ &\le \Delta_{22}^{(1)} + 2W_{\max}\left\{\frac{3\log(16N/\epsilon)}{\underline{\delta}'}\right\}^{1/2}.\end{aligned}$$

For $\Delta_{22}^{(1)}$, apply the same argument in previous proof for ScBM, we know

$$\begin{aligned}\Delta_{22}^{(1)} &\le \max_{k,k'}|Q_{r,t}(k,k')| \max_{1\le j'\le 2N}\sum_{k=1}^{K_R+K_C}\sum_{k'=1}^{K_R+K_C}\left[\sum_{j\in\mathcal{G}_{t,k}}\mathcal{Y}_t(j,k)\right]\mathcal{Y}_t(j',k') \\ &\le W_{\max}\frac{NL}{\underline{\delta}'\cdot\ell!}\left(\frac{r}{T}\right)^\beta. \end{aligned} \tag{41}$$

Therefore, combine equation (32), (34), (35), (39), and (41), we obtain

$$\frac{\left|\mathbb{M}_t^R\right|}{N} \le \frac{c_2(\varepsilon)K_RW_{\max}^2}{m_r^2N\lambda_{K_R,\max}^2}\left\{(6+c_w')\frac{b'}{\underline{\delta}'^{1/2}} + \frac{2K_C}{\underline{\delta}'}(\sqrt{2P_{\max}rs}+2P_{\max}) + \frac{NL}{\underline{\delta}'\cdot\ell!}\left(\frac{r}{T}\right)^\beta\right\}^2,$$

where $c_2(\varepsilon) = 2^6(2+\varepsilon)^2$, $b' = \{3\log(16NT/\epsilon)\}^{1/2}$ and $\lambda_{K_R,\max} = \max_t\{\lambda_{K_R,t}\}$.

(2) Clustering for $Z_{C,t}$.

As shown in equation (13), the population regularized graph Laplacian of dynamic DCcBM has following decomposition

$$\mathcal{S}_t = {\Psi_{\tau,t}^R}^{1/2} Z_{R,t}\Omega_t Z_{C,t}^\top {\Psi_{\tau,t}^C}^{1/2} \tag{42}$$

Then, let $Y_{R,t} = Z_{R,t}^\top\Psi_{\tau,t}^R Z_{R,t}$ and $Y_{C,t} = Z_{C,t}^\top\Psi_{\tau,t}^C Z_{C,t}$, and

$$H_{\tau,t} = Y_{R,t}^{1/2}\Omega_t Y_{C,t}^{1/2}. \tag{43}$$

Now, following Rohe et al. (2016), we can define

$$\gamma_c \overset{\text{def}}{=} \min_t\{\min_{i\ne j}\|H_t(*,i) - H_t(*,j)\|\}, \tag{44}$$

and thus

$$\frac{\left|\mathbb{M}_t^C\right|}{N} \le \frac{16(2+\varepsilon)^2}{m_c^2N\gamma_c^2}\left\|U_t - \mathcal{U}_t\mathcal{O}_t\right\|_F^2. \tag{45}$$

where $m_c \overset{\text{def}}{=} \min_{i,t}\{\min\{\|\Gamma_{C,t}(i,*)\|, \|\mathit{\Gamma}_{C,t}(i,*)\|\}\}$.

Now, combining equation (45) with equations (31), (34), (35), (39), and (41), we obtain

$$\sup_t\frac{\left|\mathbb{M}_t^C\right|}{N} \le \frac{c_3(\varepsilon)K_RW_{\max}^2}{m_c^2N\gamma_c^2\lambda_{K_R,\max}^2}\left\{(6+c_w')\frac{b'}{\underline{\delta}'^{1/2}} + \frac{2K_C}{\underline{\delta}'}(\sqrt{2P_{\max}rs}+2P_{\max}) + \frac{NL}{\underline{\delta}'\cdot\ell!}\left(\frac{r}{T}\right)^\beta\right\}^2$$

where $c_3(\varepsilon) = 2^7(2+\varepsilon)^2$, $b' = \{3\log(16N/\epsilon)\}^{1/2}$ and $\lambda_{K_R,\max} = \max_t\{\lambda_{K_R,t}\}$. □

References

Binkiewicz, N., J. T. Vogelstein, and K. Rohe (2017): "Covariate-assisted Spectral Clustering," *Biometrika*, 104, 361–377.

Lei, J. and A. Rinaldo (2015): "Consistency of Spectral Clustering in Stochastic Block Models," *The Annals of Statistics*, 43, 215–237.

Pensky, M. and T. Zhang (2017): "Spectral Clustering in the Dynamic Stochastic Block Model," *Arxiv Preprint Arxiv:1705.01204*.

Qin, T. and K. Rohe (2013): "Regularized Spectral Clustering under the Degree-corrected Stochastic Blockmodel," in *Advances in Neural Information Processing Systems*, 3120–3128.

Rohe, K., T. Qin, and B. Yu (2016): "Co-clustering Directed Graphs to Discover Asymmetries and Directional Communities," *Proceedings of the National Academy of Sciences*, 113, 12679–12684.

Rudelson, M. and R. Vershynin (2013): "Hanson-wright Inequality and Sub-gaussian Concentration," *Electronic Communications in Probability*, 18.

IRTG 1792 Discussion Paper Series 2019

For a complete list of Discussion Papers published, please visit http://irtg1792.hu-berlin.de.

001 "Cooling Measures and Housing Wealth: Evidence from Singapore" by Wolfgang Karl Härdle, Rainer Schulz, Taojun Xie, January 2019.

002 "Information Arrival, News Sentiment, Volatilities and Jumps of Intraday Returns" by Ya Qian, Jun Tu, Wolfgang Karl Härdle, January 2019.

003 "Estimating low sampling frequency risk measure by high-frequency data" by Niels Wesselhöfft, Wolfgang K. Härdle, January 2019.

004 "Constrained Kelly portfolios under alpha-stable laws" by Niels Wesselhöfft, Wolfgang K. Härdle, January 2019.

005 "Usage Continuance in Software-as-a-Service" by Elias Baumann, Jana Kern, Stefan Lessmann, February 2019.

006 "Adaptive Nonparametric Community Detection" by Larisa Adamyan, Kirill Efimov, Vladimir Spokoiny, February 2019.

007 "Localizing Multivariate CAViaR" by Yegor Klochkov, Wolfgang K. Härdle, Xiu Xu, March 2019.

008 "Forex Exchange Rate Forecasting Using Deep Recurrent Neural Networks" by Alexander J. Dautel, Wolfgang K. Härdle, Stefan Lessmann, Hsin-Vonn Seow, March 2019.

009 "Dynamic Network Perspective of Cryptocurrencies" by Li Guo, Yubo Tao, Wolfgang K. Härdle, April 2019.

IRTG 1792, Spandauer Strasse 1, D-10178 Berlin
http://irtg1792.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the IRTG 1792.