ECONSTOR
Make Your Publications Visible.

A Service of

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Guo, Li; Härdle, Wolfgang; Tao, Yubo

Working Paper

A time-varying network for cryptocurrencies

IRTG 1792 Discussion Paper, No. 2021-016

Provided in Cooperation with:
Humboldt University Berlin, International Research Training Group 1792 "High Dimensional Nonstationary Time Series"

Suggested Citation: Guo, Li; Härdle, Wolfgang; Tao, Yubo (2021) : A time-varying network for cryptocurrencies, IRTG 1792 Discussion Paper, No. 2021-016, Humboldt-Universität zu Berlin, International Research Training Group 1792 "High Dimensional Nonstationary Time Series", Berlin

This Version is available at:
https://hdl.handle.net/10419/241273

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

IRTG 1792 Discussion Paper 2021-016

A Time-Varying Network for Cryptocurrencies

Li Guo [*]
Wolfgang Karl Härdle [*2] [*3] [*4] [*5] [*6]
Yubo Tao [*4]

[*] Fudan University, China
[*2] Humboldt-Universität zu Berlin, Germany
[*3] Xiamen University, China
[*4] Singapore Management University, Singapore
[*5] Charles University, Czech Republic
[*6] National Chiao Tung University, Taiwan

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

International Research Training Group 1792

A Time-Varying Network for Cryptocurrencies

Li Guo*
Fudan University
Shanghai Institute of International Finance and Economics

Wolfgang Karl Härdle†
Humboldt-Universität zu Berlin, Singapore Management University
Xiamen University, Charles University

Yubo Tao‡
Singapore Management University

August 26, 2021

Abstract

Cryptocurrencies return cross-predictability and technological similarity yield information on risk propagation and market segmentation. To investigate these effects, we build a time-varying network for cryptocurrencies, based on the evolution of return cross-predictability and technological similarities. We develop a dynamic covariate-assisted spectral clustering method to consistently estimate the latent community structure of cryptocurrencies network that accounts for both sets of information. We demonstrate that investors can achieve better risk diversification by investing in cryptocurrencies from different communities. A cross-sectional portfolio that implements an inter-crypto momentum trading strategy earns a 1.08% daily return. By dissecting the portfolio returns on behavioral factors, we confirm that our results are not driven by behavioral mechanisms.

Keywords: Community detection, Dynamic stochastic blockmodel, Covariates, Co-clustering, Network risk, Momentum.

*600 Guoquan Rd, Shanghai 200433. Email: guo_li@fudan.edu.cn.
†Unter den Linden 6 10099 Berlin, Germany. Email: haerdle@hu-berlin.de.
‡Corresponding author: 90 Stamford Rd, Singapore 178903. Email: ybtao@smu.edu.sg.

1 Introduction

The invention of Bitcoin has spurred the creation of many cryptocurrencies (cryptos hereafter), commonly known as *Altcoins*. According to the statistics available at *coinlore.com*, as of July 2021, a total of 5,982 cryptos (coins and tokens) are actively traded worldwide, with a total market capitalization of over 1.2 trillion dollars. The growing number of cryptos provides investors with more alternatives to form crypto portfolios that deliver potentially higher Sharpe ratios (Markowitz, 1952; Roy, 1952; Sharpe, 1966). Therefore, it is essential to understand the inter-relationships between the cryptos for achieving optimal investment decisions.

For traditional assets, a natural starting point is to compare their fundamental similarities. For example, we can rely on industrial classifications based on business operations, e.g., Standard Industrial Classification (SIC) and the Global Industry Classification Standard (GICS), or product differentiation (Hoberg and Phillips, 2016) to infer the similarity between stocks. Unlike stocks and many other traditional assets, the fundamentals of cryptos are difficult to determine as they neither deliver dividends nor release accounting information. In addition, many investors invest in cryptocurrencies because they believe in the blockchain technology embodied in these digital coins (Liu, Sheng, and Wang, 2021). However, because of the open-source nature of blockchain technology, technological similarities could be uninformative since many *Clonecoins* are using the same technology but have utterly different price trajectories.

To resolve the dilemma, we introduce a novel method to classify cryptos by adding the return cross-predictability to technological similarities. The benefits are twofold. First, return cross-predictability provides timely information to understand the dynamics of market structure and the information propagation across financial assets (Menzly and Ozbas, 2010). Second, the returns embody investors' beliefs, which is crucial in classifying cryptos or other hard-to-value financial assets (Detzel, Liu, Strauss, Zhou, and Zhu, 2021). We naturally employ a directed time-varying network to model the evolution of return cross-predictability and contextualize the cryptos with the technological characteristics. Using network analysis, we develop a dynamic covariate-assisted spectral clustering (CASC) method, which accommodates important network features, such as degree heterogeneity, directionality, and time-varying membership, to systematically study the inter-relationships between cryptos. We also provide a theory and conduct extensive simulations to illustrate the classification consistency of our method.

We carry out three empirical tests to illustrate the economic contribution of the crypto communities. First, we demonstrate that the investors can achieve better risk diversification by investing in cross-community cryptos. Classical portfolio theory (Markowitz, 1952; Roy, 1952) suggests that investors can achieve the expected portfolio return with a lower risk (variance) by investing in less positively correlated assets. Therefore, we calculate and compare the within- and cross-community average pairwise return correlations using 7-day and 30-day future crypto returns. It shows that the return correlations between cryptos within the same community are, on average, significantly higher than those across communities. The pattern is robust to different evaluation time horizons and only valid for dynamic CASC-estimated communities.

Second, to formally test the information propagation effects within crypto communities, we form a cross-sectional portfolio that implements an inter-crypto momentum trading strategy. Prior literature has documented that, due to information propagation, assets that have fundamental similarities will have momentum spillovers, wherein past return of one asset predicts the returns of assets linked to it (see, e.g., Moskowitz and Grinblatt, 1999; Cohen and Frazzini, 2008; Menzly and Ozbas, 2010; Lee, Sun, Wang, and Zhang, 2019; Ali and Hirshleifer, 2020; Parsons, Sabbatucci, and Titman, 2020). Based on this argument, we construct a trading signal for each crypto using the average returns of the rest of cryptos in the same community. We sort the cryptos into quartiles according to the trading signals and implement a momentum trading strategy through simultaneously buying the cryptos in the top quartile and selling the cryptos in the bottom quartile. It shows that the trading strategy generates a one-day-ahead average daily return of 1.08%, and the return does not revert in a week.

Finally, we explore the behavioral mechanisms, such as limit-to-arbitrage, investor attention, and macro uncertainty, as alternative explanations to momentum spillover effects. Specifically, we split the whole sample into high and low episodes of the behavioral factors by cutting off at their medians. It shows that the momentum portfolio returns exist regardless of these factors, indicating that behavioral mechanisms could not be the drivers of momentum spillover effects.

This paper makes several fundamental contributions to the empirical understanding of the cryptos market. To the best of our knowledge, our paper is the first to establish a crypto network and study the cryptos market segmentation problem using community detection. In addition, we are among the first to empirically justify the importance of crypto technologies in determining

crypto prices. By creatively combining the return cross-predictability and technological similarities, we identify the community pattern between cryptos and document an inter-crypto momentum spillover effect, which is direct evidence for network effects studied in Sockin and Xiong (2020) and Cong, Li, and Wang (2021b). Our paper also contributes to community detection methods in network analysis by extending the spectral clustering method to identify communities in a time-varying network in the presence of both time-varying memberships and node covariates.

This paper is closely related to a rapidly growing literature on the empirical asset pricing of cryptocurrencies. Yermack (2017) is the first paper to study the economic implications of blockchain in the context of corporate governance. Makarov and Schoar (2020) investigated cross-exchange crypto arbitrage and document a sizable price difference between exchanges. Griffin and Shams (2020) analyzed the wallet-level blockchain data and examine the price manipulation in Bitcoin. Liu and Tsyvinski (2021) provided a comprehensive analysis of the risk-return tradeoff of cryptos. Liu, Tsyvinski, and Wu (2021) examined the cross-section of crypto and construct a three-factor model. Liu, Marsh, Mazza, and Petitjean (2021) explored the factor structure in both crypto returns and volatility and establish nine stylized facts in the cryptos market. Our paper provides new insights into the fundamentals of the cryptos market structure by dividing cryptos into different communities and document an inter-crypto momentum spillover effect.

Our paper contributes to the literature on community detection. Namely, community detection is one of the central problems in network analysis, and various algorithms have been proposed to solve this problem. For global approaches, there are spectral clustering (Rohe, Chatterjee, and Yu, 2011; Lei and Rinaldo, 2015; Jin, 2015; Joseph and Yu, 2016; Rohe, Qin, and Yu, 2016; Gao, Ma, Zhang, and Zhou, 2018; Pensky and Zhang, 2019; Zhou and Amini, 2019) and convex relaxations via semidefinite programs (SDPs) (Cai and Li, 2015; Chen and Xu, 2016; Hajek, Wu, and Xu, 2016a,b; Amini and Levina, 2018; Li, Chen, and Xu, 2021). For local approaches, there are acyclic belief propagation (ABP) (Decelle, Krzakala, Moore, and Zdeborová, 2011), Bayesian MCMC and variational Bayes (Nowicki and Snijders, 2001; Celisse, Daudin, and Pierre, 2012; Bickel, Choi, Chang, and Zhang, 2013; Matias and Miele, 2017; Wang and Bickel, 2017a), profile likelihood (Bickel and Chen, 2009; Zhao, Levina, and Zhu, 2012), and pseudo-likelihood maximization (Amini, Chen, Bickel, and Levina, 2013; Zhou and Amini, 2020). Our method modifies the spectral clustering algorithm to incorporate historical network structures, time-varying mem-

bership, and node covariates. The simulation confirms its superior classification accuracy over state-of-the-art methods.

In addition, our paper echos the literature that utilizes auxiliary information for network analysis (contextual network analysis). One strand of literature has tried to understand the network dependencies with the assistance of covariates (Zou, Lan, Wang, and Tsai, 2017; Lan, Fang, Wang, and Tsai, 2018; Zou, Lan, Li, and Tsai, 2021; Zhao, Liu, Wang, and Leng, 2021; Zhu, Cai, and Ma, 2021). Another strand of literature has studied the community detection under the contextual setting (Zhang, Levina, and Zhu, 2016; Weng and Feng, 2016; Binkiewicz, Vogelstein, and Rohe, 2017; Ma and Ma, 2017; Zhang, Poux-Berthe, Wells, Koc-Michalska, and Rohe, 2018; Deshpande, Sen, Montanari, and Mossel, 2018; Mu, Mele, Hao, Cape, Athreya, and Priebe, 2020; Abbe, Fan, and Wang, 2020; Yan and Sarkar, 2020; Esmaeili, Saad, and Nosratinia, 2021). Our paper is closely related to Binkiewicz et al. (2017) and Zhang et al. (2018), where we extend the covariate-assisted spectral clustering method to estimate a time-varying community structure.

The remainder of the paper is organized as follows. In Section 2, we construct a time-varying network contextualized by crypto fundamentals to model the information propagation in the cryptos market. In Section 3, we develop the dynamic covariate-assisted spectral clustering algorithm to estimate the crypto communities and demonstrate its effectiveness via both theory and simulations. In Section 4, we explore the economic implications of the crypto communities. We present our conclusions in Section 5. All proofs and technical details are provided in the supplementary appendix. The R codes to implement the algorithms are available at QuantNet (search keywords "CASC").

2 Construction of the Cryptocurrencies Network

2.1 Network of information propagation

Traditional asset pricing theory states that the cross-predictability of asset returns reflects the correlations in the fundamental determinants caused by common information of the future cash flows or discount rates (Menzly and Ozbas, 2010). Extensive literature has applied network analysis to the equity market to gain insights about information propagation and spillover of behavioral bias across the firms (see, e.g., Cohen and Frazzini, 2008; Aobdia, Caskey, and Ozel, 2014; Acemoglu,

Ozdaglar, and Tahbaz-Salehi, 2015; Herskovic, 2018; Chen, Härdle, and Okhrin, 2019). Most recently, Sockin and Xiong (2020), Cong, Li, and Wang (2021a), and Cong et al. (2021b) develop models for platform tokens and study their network effects caused by decentralization. Based on the above literature, we find it essential to adopt a network view on the crypto market to study how price information or shock transfers from one crypto to another.

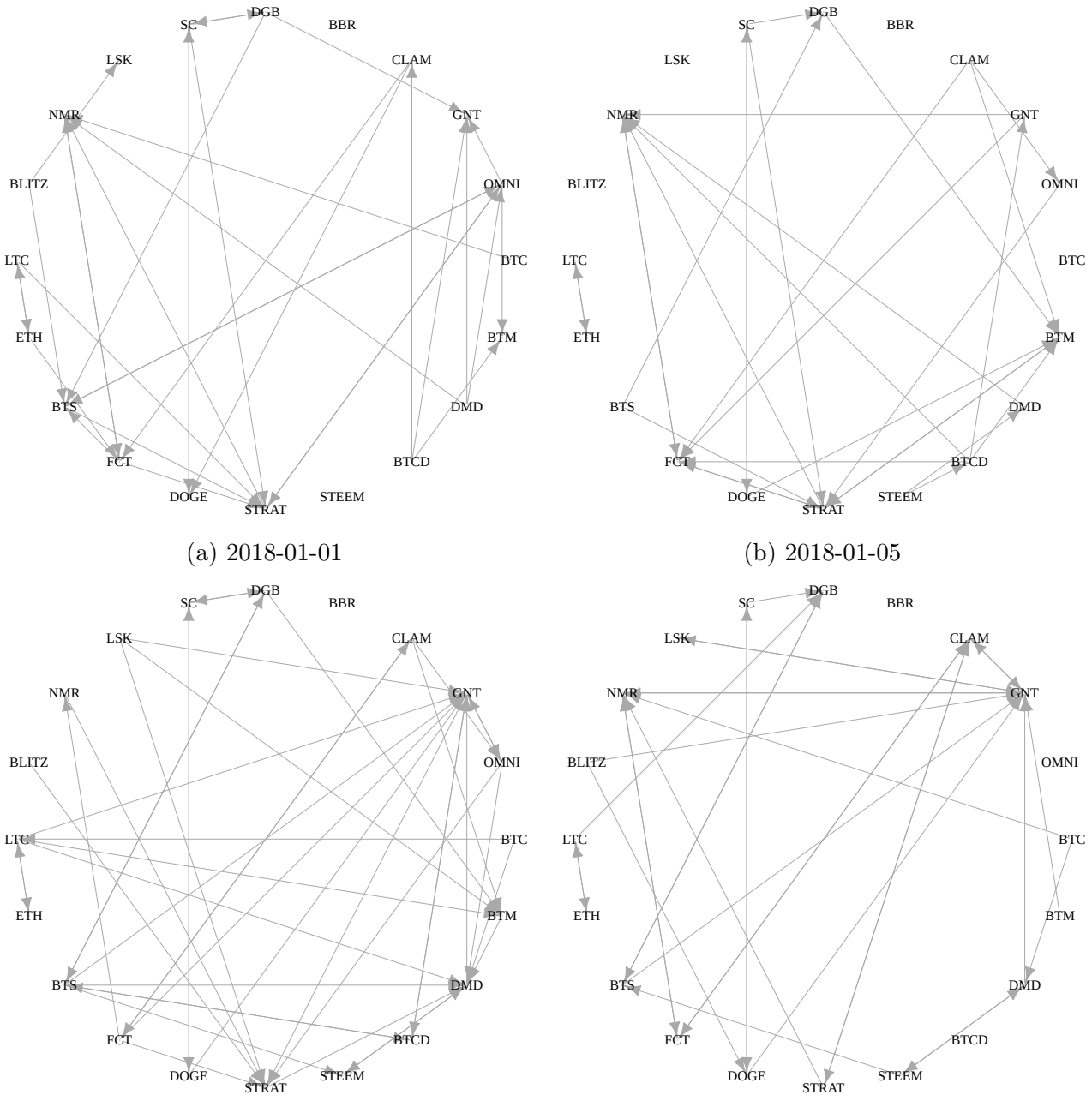
Following Menzly and Ozbas (2010), we model the information propagation in the cryptos network using the return cross-predictability. We collect the historical daily prices, trading volumes, and contract attributes (e.g., algorithm, proof type, age, total coins, etc.) of the top 400 cryptos in market capitalization as of December 31, 2018 from an interactive platform (*Cryptocompare.com*) with free API access. The sampling period is from January 1, 2016 to December 31, 2018. After excluding cryptos under 1 year old or with incomplete price and contract information, we obtain a sample of 182 cryptos.

To form the linkages between cryptos, we regress a crypto's current return on the other cryptos' lagged returns in a 360-day (1-year) rolling window. We employ the adaptive Lasso (Zou, 2006) to select the significant regressors and to estimate their coefficients simultaneously:

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

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 $\hat{w}_{i,j} = 1/|\hat{b}_{i,j}^{LS}|^\gamma$ with some $\gamma > 0$, where $\hat{b}_{i,j}^{LS}$ denotes the coefficient estimated by ordinary least squares. The Lasso technique yields an active set that has "parental" influence on the focal crypto. Thus, we obtain an adjacency matrix for each period, A_t, $t = 1, \cdots, T$. Figure 1 presents a sub-network of 20 cryptos with linkages estimated by regression (1) on selected dates in January 2018.

Three critical features can be observed from the return-inferred network structure. First, the return cross-predictability is changing over time. For instance, the linkage between BTC and NMR on January 1 vanishes on January 5, and the predictive relationship between BTS and DGB on January 1 reverts on January 5. These observations indicate that information propagation is not always mono-directional or continues to exist between cryptos.



(a) 2018-01-01

(b) 2018-01-05

(c) 2018-01-15

(d) 2018-01-31

Figure 1: Networks Formed by Return Cross-predictability. This figure presents the return-based cryptos network 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. 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. Only the cryptos selected by adaptive Lasso will be linked to crypto i. An arrow from crypto A to crypto B means crypto A's return can predict the crypto B's return.

Second, the return cross-predictability induces a sparse network structure, that is, many isolated cryptos with no linkages in the network. The examples include BBR and STEEM on January 1; BBR, BLITS, BTC, and LSK on January 5; BBR on January 15; BBR, OMNI, and BTCD on January 31. The sparsity is an important feature that may create many challenges for conducting an in-depth analysis of the network structure, such as community detection, network formation modeling, network differential tests, and so forth. To resolve this issue, we need to explore new information sources to identify additional linkages between cryptos or apply advanced statistical methods designed for analyzing sparse networks. In fact, we use both strategies in the later analyses.

Finally, a larger market capitalization does not imply a leading position in return cross-predictability. For example, the largest three cryptos (BTC, LTC, and ETH) predict fewer crypto returns than CLAM on January 5 and LSK on January 15. These observations suggest that return cross-predictability is a noisy estimator for uncovering the authentic connections between cryptos, which motivates us to seek additional information sources to form crypto linkages so that the noises could cancel out with each other when we conduct a combinatorial analysis for the cryptos network.

2.2 Contextualization with technology

As pointed out by Liu et al. (2021), although cryptos do not distribute dividends or release accounting information, many investors invest in them simply because they believe blockchain technology is an important innovation, and cryptos are assets that represent a stake in the future of this technology. Specifically, there are two building blocks for blockchain technology: hashing algorithms and consensus mechanism (or proof types).

Algorithm, which is short for the *hashing algorithm*, plays a central role in determining the crypto's security. 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 outmine others. Thus, Scrypt-based Altcoins require on average more computing effort

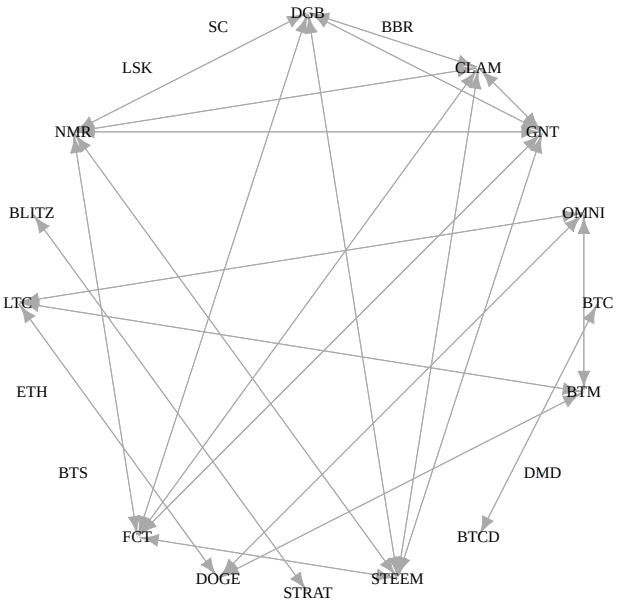
per unit than the equivalent coin using SHA256. The relative difficulty of the algorithm confers a relative value.

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), whereas 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.

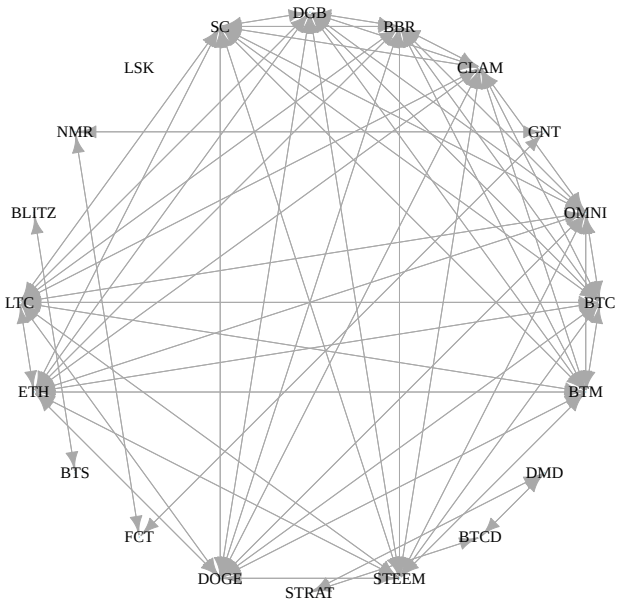
Recent literature has tried to model the mining process and different proof types under a theoretical framework. Many papers (see, e.g., Biais, Bisiere, Bouvard, and Casamatta, 2019; Easley, O'Hara, and Basu, 2019; Cong, He, and Li, 2021) have studied mining activities and transaction costs and their implications for crypto prices. Among them, Pagnotta (2020) is the first to build an equilibrium model embedding a security channel that leads to multiple equilibria. Abadi and Brunnermeier (2018), Budish (2018), and Hinzen, John, and Saleh (2019) studied the limitations and the pricing implications of the PoW mechanism, while Fanti, Kogan, and Viswanath (2019) and Saleh (2021) explored the PoS mechanism. Based on these models, we conjecture that the technological similarities could be vital for understanding the inter-relationships between cryptos.

We collect the algorithm and proof type information for each crypto in our sample and contextualize the network induced by return cross-predictability with the technological similarities between cryptos. For illustration, we plot the sub-network in Figure 2 using the same cryptos as in Figure 1, where the linkages are formed if the cryptos share at least one fundamental characteristic. For example, LTC and DOGE are linked because they both adopt the Scrypt algorithm. Evidently, due to the limited choices of algorithms and proof types, the cryptos are more likely to connect with each other when using technological similarities to form linkages. Intuitively, the network will be denser if we naïvely link cryptos either by return cross-predictability or technological similarity,

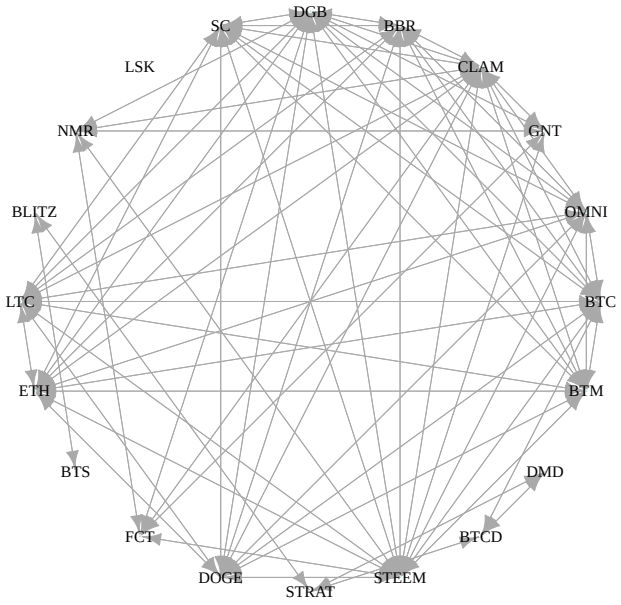
such as adding linkages in subfigure (c) of Figure 2 to each subfigure of Figure 1. In the following section, we will discuss how to combine the two sets of information in a more sophisticated manner.



(a) Algorithm



(b) Proof Types



(c) Combined Fundamentals

Figure 2: Networks Formed by Technological Similarities. This figure depicts the contract-based cryptos network. We connect two cryptos with double-sided arrows if they have the same algorithm or proof type.

In summary, we have built a time-varying cryptos network that characterizes both return cross-predictability and technological similarity between cryptos. The price shocks (e.g., policy changes, macro events, etc.) will travel through return-based linkages, while technological shocks (e.g., innovation, security breaches, etc.) will travel through technology-based connections. In the following sections, we will investigate deeper on the inter-relationships between cryptos based on this network.

3 Communities in the Cryptocurrencies Network

Classical empirical studies on asset pricing have pointed out the importance of market segmentation in generating heterogeneous prices of risk and profitable trading strategies. However, most of them use *ex-ante* information, such as industry, listing exchange, and geography, to conjecture market segments (see, e.g., Fama and French, 1993; Fama, French, Booth, and Sinquefield, 1993; Foerster and Karolyi, 1999; Griffin, 2002). Recently, Patton and Weller (2019) proposed a data-driven method (EM algorithm) to cluster stocks with similar risk prices and provide strong evidence of market segmentation across all portfolios, factor pricing models, and time periods. Based on these findings, we deduct market segmentation also plays an important role in understanding the risks and returns of cryptos and would like to cluster cryptos under the network framework.

A standard framework for studying clustering effect, that is, community detection, in a network setting is the stochastic blockmodel (SBM) proposed by Holland, Laskey, and Leinhardt (1983). Using this framework, we can represent the adjacency matrices stochastically via a block structure and adopt spectral clustering to identify latent communities. Binkiewicz et al. (2017) and Zhang et al. (2018) showed that, by introducing covariate assistance, they could improve the classification accuracy of the spectral clustering method under static network settings. This section presents an extension of the covariate-assisted spectral clustering (CASC) algorithm to deal with the time-varying directed networks. Theoretical justifications and simulations demonstrate the consistency of this method.

3.1 Dynamic contextual stochastic blockmodel

As return cross-predictability between any two cryptos could be mono-directional, we consider a time-varying network defined as a sequence of random directed 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 time. For each period, we model the directed network structure with an adjacency matrix A_t which is mostly asymmetric, that is,

$$A_t(i,j) = \begin{cases} \text{Bernoulli}\{P_t(i,j)\}, & \text{if } i \neq j \\ 0, & \text{if } i = j \end{cases}, \tag{2}$$

where $P_t(i,j) = \Pr\{A_t(i,j) = 1\}$. The adjacency matrix A_t records the pattern of edges that are induced by crypto return cross-predictability in network at time t; if there is an edge starting from node $i \in \{1, \cdots, N\}$ and ending at node $j \in \{1, \cdots, N\}$ (i.e., $i \to j$), then $A_t(i,j) = 1$; otherwise, $A_t(i,j) = 0$. Thus, the ith row of A_t records how crypto i predicts other cryptos (sending edges), and the jth column of A_t records how crypto j is predicted by other cryptos (receiving edges).

To model the community structure, the probabilities of a connection $P_t(i,j)$ at period t are partitioned into blocks. Let the block probability matrix in each period be $B_t \in [0,1]^{K_R \times K_C}$ with rank $K = \min\{K_R, K_C\}$. In particular, let us denote $z_{R,it}$ as the row community label of node i at time t and $z_{C,it}$ as the column community label of node j at time t, respectively; then, if $z_{R,it} = k_R$ and $z_{C,jt} = k_C$, then $P_t(i,j) = B_t(z_{R,it}, z_{C,jt}) = B_t(k_R, k_C)$. Hence, for any $t = 1, \cdots, T$, we can obtain the population adjacency matrix

$$\mathcal{A}_t \overset{\text{def}}{=} \mathsf{E}(A_t) = Z_{R,t} B_t Z_{C,t}^\top, \tag{3}$$

where $Z_{R,t} \in \{0,1\}^{N \times K_R}$ and $Z_{C,t} \in \{0,1\}^{N \times K_C}$ are the *row-clustering matrices* and *column-clustering matrices*, respectively. Notably, there is only one 1 in each row and at least one 1 in each column of each clustering matrix. The co-clustering of the adjacency matrix yields two partitions of the same set of nodes. The row communities contain nodes with similar sending patterns, and the column communities contain nodes with similar receiving patterns.

Qin and Rohe (2013) and Joseph and Yu (2016) showed that the regularization of adjacency matrix can substantially improve the spectral clustering performance, especially for sparse networks. Therefore, we define the regularized graph Laplacian $L_{\tau,t} \in \mathbb{R}^{N \times N}$ for the directed network

as

$$L_{\tau,t} = D_{R,t}^{-1/2} A_t D_{C,t}^{-1/2}, \tag{4}$$

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.

To contextualize the network with the technological similarities between cryptos, we introduce the covariate matrix $X \in \{0,1\}^{N \times R}$, where each row of X denotes a crypto, and each column denotes a type of technology. Thus, $X(i,j)$ equals 1 if crypto i adopts technology j and is 0 otherwise. Following Binkiewicz et al. (2017), we assume the covariate X also reflects the row- and column-cluster's information, that is, let $M_{R,t} \in [0,1]^{K_R \times R}$ and $M_{C,t} \in [0,1]^{K_C \times R}$ denote the covariate expectation matrix for row- and column-clusters in each period t, respectively, then the population covariate matrix is

$$\mathcal{X} = \mathsf{E}(X) = Z_{R,t} M_{R,t} = Z_{C,t} M_{C,t}. \tag{5}$$

Then, we combine return cross-predictability and technological similarities between cryptos by constructing a similarity matrix from regularized graph Laplacian $L_{\tau,t}$ and covariate matrix X, that is, for each $t = 1, \cdots, T$,

$$S_t = L_{\tau,t} + \alpha_t C_t^w = D_{R,t}^{-1/2} A_t D_{C,t}^{-1/2} + \alpha_t X W_t X^\top, \tag{6}$$

where $\alpha_t \in [0, \infty)$ is the tuning parameter that controls for the informational balance between return cross-predictability and technological similarities, and W_t is a time-varying weight matrix that controls for the degree of technological similarities.

The setup in Equation (6) 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 have the same origin. Such inheritance relationships will potentially lead to an interaction between the cryptos. In addition, as shown in Figure 3, some algorithms may become more popular over time, while the others may be near extinction. Thus, this interaction would also change over time. Second, we can easily select covariates by setting certain elements

of W_t to 0. This is necessary as it helps us to model the evolution of technologies. At some point in time, some technology may disappear due to upgrades or cracking. Therefore, W_t offers us the flexibility to exclude covariates. Finally, 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 cryptos market. However, this homogeneity does not necessarily result in a cross-predictability 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.

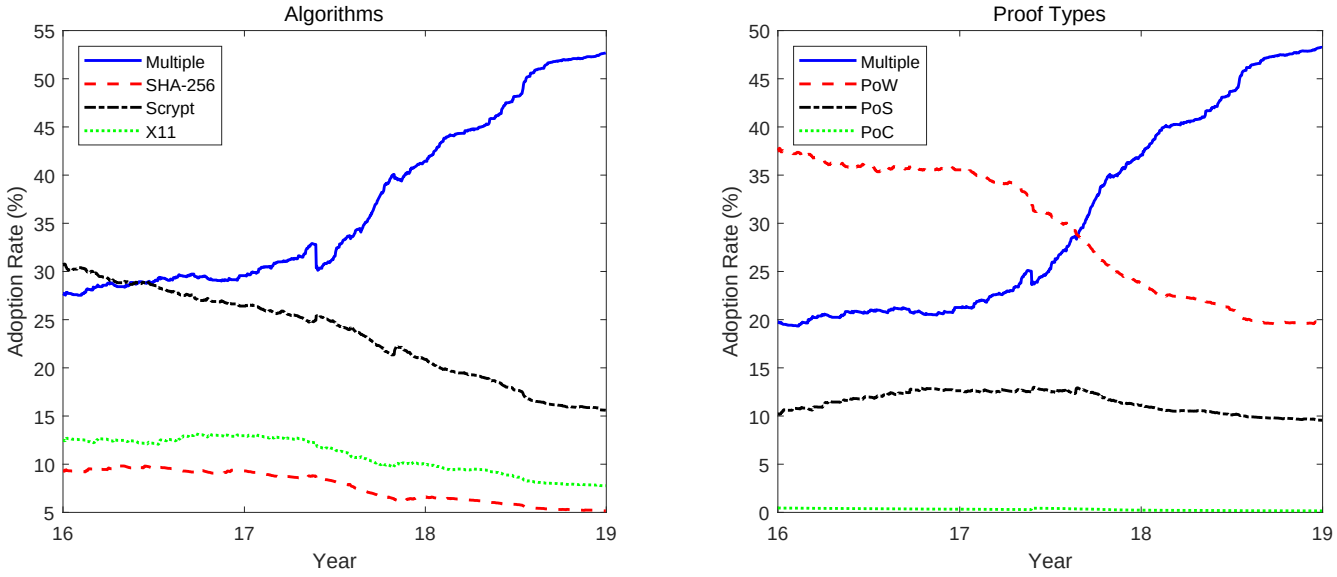


Figure 3: The Evolution of Technology Adoption Rates. This figure presents the adoption rates (or the popularity) of the major technologies in cryptos market from January 1, 2016 to December 31, 2018.

Following Karrer and Newman (2011), we employ degree correction to allow for heterogeneous degrees. Specifically, we introduce two sets of degree parameters $\psi^R = (\psi_1^R, \psi_2^R, \cdots, \psi_N^R)$ and $\psi^C = (\psi_1^C, \psi_2^C, \cdots, \psi_N^C)$ so that the edge probability are given by

$$P_t(i,j) = \psi_i^R \psi_j^C B_t(i,j)$$

under the restriction that

$$\sum_{i\in\mathcal{G}_{t,k}^R}\psi_i^R=1,\forall k\in\{1,2,\cdots,K_R\},\quad \sum_{i\in\mathcal{G}_{t,k}^C}\psi_i^C=1,\forall k\in\{1,2,\cdots,K_C\}.$$

Then, the population adjacency matrix for dynamic degree-corrected contextual stochastic blockmodel (DCCBM) is

$$\mathcal{A}_t=\mathsf{E}(A_t)=\Psi^R Z_{R,t}B_t Z_{C,t}^\top\Psi^C, \tag{7}$$

where $\Psi^s=\mathrm{Diag}(\psi^s), s\in\{R,C\}$, and the population regularized graph Laplacian is

$$\mathcal{L}_{\tau,t}=\mathcal{D}_{R,t}^{-1/2}\mathcal{A}_t\mathcal{D}_{C,t}^{-1/2}. \tag{8}$$

Therefore, the population similarity matrix is

$$\mathcal{S}_t=\mathcal{L}_{\tau,t}+\alpha_t\mathcal{C}_t^w, \tag{9}$$

where $\mathcal{C}_t^w=\mathcal{X}\mathcal{W}_t\mathcal{X}^\top$.

By construction, we know that $D_{R,t}(i,i)=\sum_{j=1}^N\mathbf{1}_{\{i\to j\}}^{(t)}+\tau_{R,t}$, which controls for the number of the parents of node j, and that $D_{C,t}(j,j)=\sum_{i=1}^N\mathbf{1}_{\{j\to i\}}^{(t)}+\tau_{C,t}$, 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) proposed using the *singular value decomposition* instead of eigendecomposition 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 offsprings"; that is, for each $t=1,\cdots,T$:

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

Therefore, the column communities consist of cryptos that receive similar price and technology shocks (common parents), and the row communities contain cryptos transmitting shocks to similar

cryptos (common offspring).

3.2 Dynamic covariate-assisted spectral clustering

When setting up a dynamic covariate-assisted spectral clustering algorithm, we face two major difficulties: (i) determining the interactive weights W_t between covariates and (ii) estimating the similarity matrix with time-varying network information. For the first issue, we use the adoption rates to present the popularity of the crypto technology. Specifically, for any two technologies (algorithms or proof types) a and b at time t, we have

$$W_t(a,b) = \frac{N_{a,t}}{N_t} \times \frac{N_{b,t}}{N_t}, \tag{10}$$

where N is the total number of cryptos in the market, N_a and N_b are the numbers of cryptos that adopt technology a and b, respectively. Intuitively, the weight is greater when the technologies are more popular among the cryptos, indicating a higher strength of technological similarity.

For the second issue, we follow Pensky and Zhang (2019) by constructing the estimator of $\mathcal{S}_t$ with a discrete kernel to bring in historical linkage information. Specifically, we first pick an integer $r \geq 0$, obtain a set of integers $\mathcal{F}_r = \{-r, \cdots, 0\}$ and assume that $\left\|W_{r,\ell}(i)\right\| \leq W_{\max}$, where $W_{\max}$ is independent of r and i, and satisfies

$$\frac{1}{|\mathcal{F}_r|} \sum_{i \in \mathcal{F}_r} i^k W_{r,\ell}(i) = \begin{cases} 1, & \text{if } k = 0, \\ 0, & \text{if } k = 1, 2, \cdots, \ell - 1. \end{cases} \tag{11}$$

By construction, the $W_{r,\ell}$ is a discretized version of the order ℓ optimal continuous boundary kernel that weighs only the historical observations (Gasser and Müller, 1979; Altman, 1990). As shown in Figure 4, the kernel assigns more recent similarity matrices higher scores. To choose an optimal bandwidth r, Pensky and Zhang (2019) proposed an adaptive estimation procedure using Lepski, Mammen, and Spokoiny (1997). We employ the same method (see supplement appendix 6) and construct the estimator for the smoothed similarity matrices:

$$\widehat{\mathcal{S}}_{t,r} = \frac{1}{|\mathcal{F}_r|} \sum_{i \in \mathcal{F}_r} W_{r,\ell}(i) S_{t+i}. \tag{12}$$

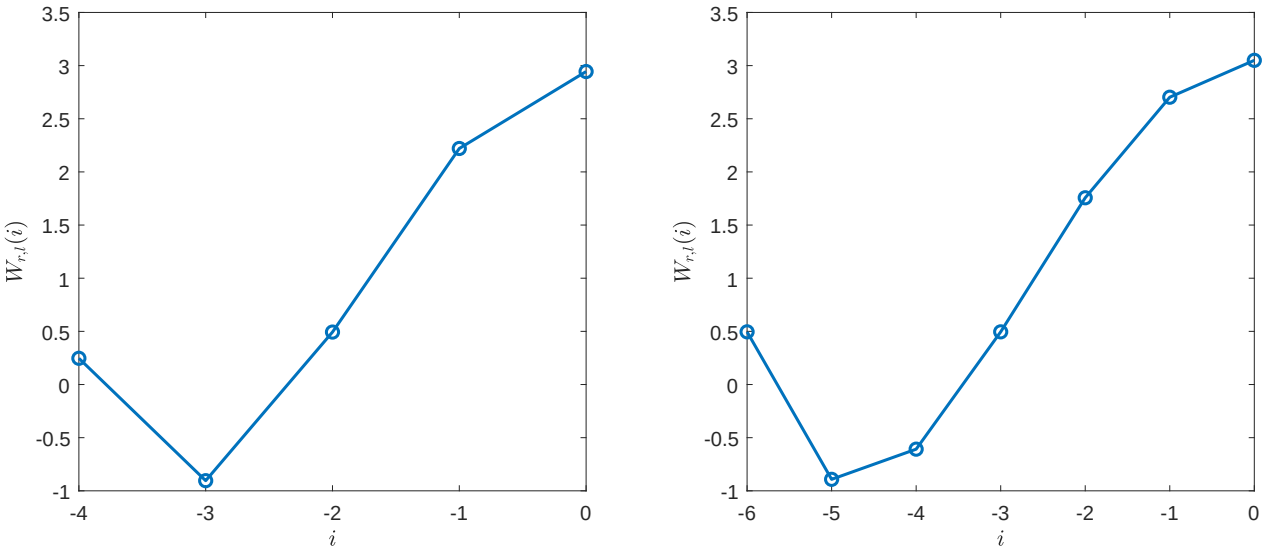


Figure 4: Examples of Discrete Kernel. The first figure plots the value of discrete kernel $W_{r,\ell}(i)$ with bandwidth $r = 4$, smoothing parameter $\ell = 4$, and $i \in \mathcal{F}_r$. The second figure plots $W_{r,\ell}(i)$ with bandwidth $r = 6$, smoothing parameter $\ell = 4$, and $i \in \mathcal{F}_r$.

Remark 1. *Different from Pensky and Zhang (2019), we only apply a boundary kernel on the historical similarity matrices for community detection. The changes are to avoid the "look-ahead bias" (Kan and Kirikos, 1995) in asset pricing inference in subsequent sections.*

Once we obtain $\widehat{\mathcal{S}}_{t,r}$, we create a singular value decomposition of $\widehat{S}_{t,r} = \widehat{U}_t \widehat{\Sigma}_t \widehat{V}_t^\top$ for each $t = 1, 2, \cdots, T$. As Lei and Rinaldo (2015) discussed, 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 community 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-medians spectral clustering:

$$\left\|\widehat{Z}_{R,t}^+ \widehat{Y}_t - \widehat{U}_t^+\right\|_F^2 \leq (1+\varepsilon) \min_{\substack{Z_{R,t}^+ \in \mathcal{M}_{N_+^R,K} \\ Y_t \in \mathbb{R}^{K\times K}}} \left\|Z_{R,t}^+ Y_t - \widehat{U}_t^+\right\|_F^2 \tag{13}$$

where Y_t is some rotation matrix. In the last step, we extend $\widehat{Z}_{R,t}^+$ to obtain $\widehat{Z}_{R,t}$ by adding $N - N_+^R$ canonical unit row vectors at the end. $\widehat{Z}_{R,t}$ is the estimate of $Z_{R,t}$ from this method. Same techniques are applied to $\widehat{V}_t$ to obtain column communities. We summarize the detailed steps in

Algorithm 1.

Algorithm 1: CASC in the Dynamic DCCBM

Input : Adjacency matrices A_t for $t = 1, \cdots, T$;
Covariates matrix X;
Number of row communities K_R and number of column communities 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 (12).

3 Compute the singular value decomposition of $\widehat{\mathcal{S}}_{t,r} = \widehat{U}_t \widehat{\Sigma}_t \widehat{V}_t^\top$ 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 $\widehat{U}_t^K \in \mathbb{R}^{N \times K}$ and $\widehat{V}_t^K \in \mathbb{R}^{N \times K}$.

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

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

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

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

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

3.3 Classification consistency

In the subsequent analysis, we illustrate that the dynamic CASC achieves classification consistency uniformly over time. We first make some assumptions on the graph that generates the dynamic network. The major assumption required here is *assortativity*, which ensures that the nodes within the same community are more likely to share an edge than nodes in two different communities.

Assumption 1. *The time-varying network is composed of a series of assortative graphs that are generated under the stochastic blockmodel with covariates, where the block probability* $B(k,k) \geq B(k,k')$, $\forall k' \neq k$.

Intuitively, the more frequent the community membership changes, the less stable the network will be. Consequently, it becomes harder to utilize 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 bind the dynamic behavior of clustering matrices ($Z_{R,t+r} - Z_{R,t}$ or $Z_{C,t+r} - Z_{C,t}$) 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* $1 \leq k \leq k' \leq K$, *let* $\varsigma_t = t/T$ *denote the time instance, 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*

$$|f^{(\ell)}(x) - f^{(\ell)}(x')| \leq L|x - x'|^{\beta - \ell}, \text{ for any } x, x' \in [0,1], \tag{14}$$

with ℓ *being the largest integer smaller than* β.

Assumption 3 states that neither the connection probabilities nor the community memberships change drastically over the horizons. Finally, 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. *Define* $\underline{\delta} = \inf_t\{\min\{\min_i \mathcal{D}_{R,t}(i,i), \min_i \mathcal{D}_{C,t}(i,i)\}\}$ *as the minimum average degree. Then, for any* $\epsilon \in (0,1)$, *assume:*

(i) $\underline{\delta} > 3\ln(16NT/\epsilon)$;

(ii) $\alpha_{\max} = \sup_t \alpha_t \leq (NR)^{-1}a$, *where* $a = \sqrt{\frac{3\ln(16NT/\epsilon)}{\underline{\delta}}}$.

Remark 2. *Assumption 4 embodies several important conditions for classification consistency. Condition (i) indicates the graph should be at least semi-dense. Condition (ii) ensures the population eigen-gap to be greater than the maximum of the absolute difference between the sum of covariate variances within a block and the mean of the sums across all blocks, which means covariate similarities will not dominate the clustering results.*

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

$$\mathbb{M}_t^p = \left\{ i\colon \left\| C_{i,t}^p - \mathcal{C}_{i,t}^p \mathcal{O}_t^p \right\| > \left\| C_{i,t}^p - \mathcal{C}_{j,t}^p \mathcal{O}_t^p \right\|, \text{ for any } j \neq i \right\}, \quad p \in \{R, C\}, \tag{15}$$

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

Theorem 1. *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 $\mathcal{S}_t$. 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 1-4 and $K_R \leq K_C$, the misclustering rate satisfies*

$$\sup_t \frac{\left|\mathbb{M}_t^R\right|}{N} \leq \frac{c_2(\varepsilon) K W_{\max}^2}{m_r^2 N \lambda_{K,\max}^2} \left\{ (6+c_1)\frac{b}{\underline{\delta}^{1/2}} + \frac{2K_C}{\underline{\delta}}(\sqrt{2P_{\max} rs} + 2P_{\max}) + \frac{P_{\max} L}{\underline{\delta} \cdot \ell!} \left(\frac{r}{T}\right)^\beta \right\}^2,$$
$$\sup_t \frac{\left|\mathbb{M}_t^C\right|}{N} \leq \frac{c_3(\varepsilon) K W_{\max}^2}{m_c^2 N \gamma_c^2 \lambda_{K,\max}^2} \left\{ (6+c_1)\frac{b}{\underline{\delta}^{1/2}} + \frac{2K_C}{\underline{\delta}}(\sqrt{2P_{\max} rs} + 2P_{\max}) + \frac{P_{\max} L}{\underline{\delta} \cdot \ell!} \left(\frac{r}{T}\right)^\beta \right\}^2,$$

with a probability of at least $1-\epsilon$, where $\lambda_{K,\max} = \max_t\{\lambda_{K,t}\}$, $c_2(\varepsilon) = 2^6(2+\varepsilon)^2$, $c_3(\varepsilon) = 2^7(2+\varepsilon)^2$, $b = \sqrt{3\ln(16NT/\epsilon)}$. m_r, m_c, and γ_c are defined by equation (21), (36), and (34) in the supplementary appendix, respectively.

Remark 3. *The upper bounds on misclustering rates obtained in the Theorem 1 are non-asymptotic. To justify the asymptotic behavior of the misclustering rates, following Rohe et al. (2016), we first have $m_r = O(\sqrt{K/N})$ and $m_c = O(\sqrt{K/N})$. Then, following Fan, Liao, and Mincheva (2013), we have $\lambda_{K,\max}$ diverging at rate $O(N)$. Based on these two results, we can easily find the convergence rate of the component outside of the bracket in error bound being $O(1/N^2)$. Taking each component*

inside the bracket diverges slower than $O(N)$, we conclude that the error bound will converge to 0 as $N, T \to \infty$.

Remark 4. *The uniform upper bound is analogous to the static upper bounds in Binkiewicz et al. (2017) and Zhang et al. (2018). Abbe et al. (2020) pointed out that these bounds are not tight enough to inform the information gain in combining the network and node attributes. Deshpande et al. (2018), Abbe et al. (2020), and Lu and Sen (2020) tried to derive the information threshold for exact recovery under static network settings. It is definitely worth another theoretical paper to derive the uniform information bound for dynamic networks, but it is beyond the scope of the current paper. For empirical analysis, we only need consistency to guarantee the performance of our method. In subsequent sections, we will illustrate the improvement of classification accuracy by combining two sets of information through simulations and empirical tests.*

3.4 Monte Carlo simulations

To intuitively illustrate the theoretical results intuitively, we carry out several simulation studies using our algorithm and existing clustering methods under different model setups. Our benchmark algorithms for the directed networks are the degree-corrected DI-SIM (DI-SIM-DC) by Rohe et al. (2016) and the covariate-assisted DI-SIM for static networks (CA-DI-SIM-Stc) by Zhang et al. (2018). Since both methods are intended for static networks, we apply them to the adjacency matrix period by period and report the average misclustering rate over time.

We set the block probability matrix B_t as

$$B_t = \begin{bmatrix} 0.60 & 0.30 & 0.20 & 0.10 \\ 0.30 & 0.50 & 0.20 & 0.10 \\ 0.20 & 0.20 & 0.40 & 0.10 \\ 0.10 & 0.10 & 0.10 & 0.30 \end{bmatrix} \times \frac{T+2t}{2T}, \text{ with } 1 \le t \le T, \tag{16}$$

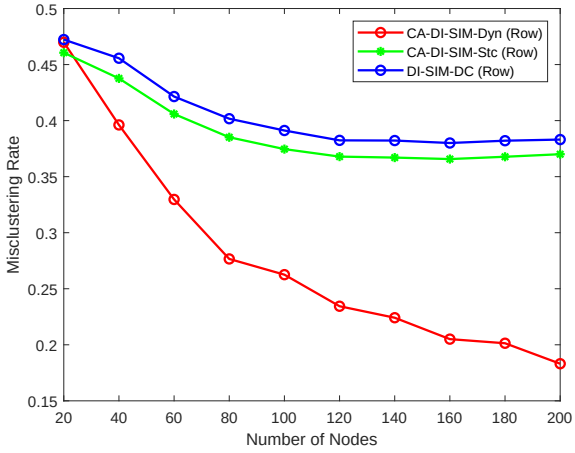
and set the order of the polynomials for kernel construction at $L = 4$ for all simulations. Then, we generate the initial clustering matrices $Z_{R,1}$ and $Z_{C,1}$ by randomly choosing one entry in each row and assign it to 1. For $t = 2, \cdots, T$, we fix the last $N - s$ rows of $Z_{R,t-1}$ and $Z_{C,t-1}$ and re-assign 1s in the first s rows of $Z_{R,t-1}$ and $Z_{C,t-1}$, respectively, to mimic the community membership

changes. Lastly, we assume that the number of communities $K_R = K_C = 4$ for directed network is known throughout the simulations. The time-invariant node covariates are $R = \lfloor \ln(NT) \rfloor$ dimensional with values $X \sim U(0, 10)$. We replicate all experiments 1000 times, and the misclustering rate we report is the temporal average of the misclustering rates; that is, $T^{-1} \sum_{t=1}^{T} |\mathbb{M}_t^R|/N$ and $T^{-1} \sum_{t=1}^{T} |\mathbb{M}_t^C|/N$.

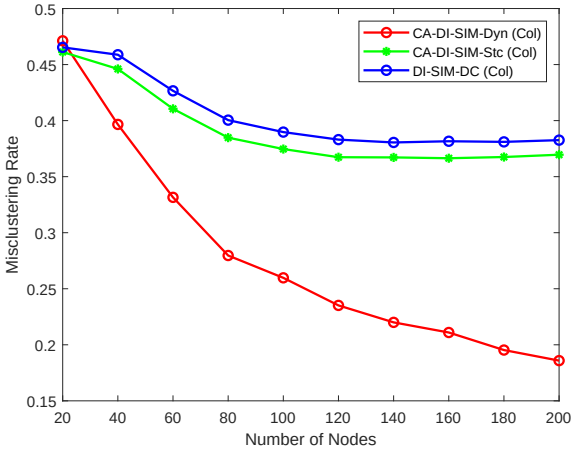
We first examine the clustering performance with a growing network size. The number of vertices in the network varies from 20 to 200 with step size 20. The time span is $T = 10$ and the number of membership changes is $s = 10$. We summarize the results in Figure 5 (a) and (b). Evidently, as the size of the network increases, the misclustering rates of the CA-DI-SIM-Dyn decrease sharply and dominate both CA-DI-SIM-Stc and DI-SIM-DC throughout the simulation. The CA-DI-SIM-Stc dominates DI-SIM-DC in classification accuracy, which is consistent with the argument that some similarity is sufficient for the covariates to assist in the discovery of the graph structure (Binkiewicz et al., 2017). The dominating performance of CA-DI-SIM-Dyn over CA-DI-SIM-Stc proves that historical linkage information is crucial for improving classification accuracy.

Next, we check the relative performance for a growing maximum number of community membership changes. Here, we fix the total number of vertices at 200 and vary the community membership changes for each period, s, from 20 to 100 with step size 20. The total number of horizons is $T = 10$. We summarize the results in Figure 5 (c) and (d). Evidently, the misclustering rate of our method increases with the number of community membership changes. The benchmark methods are insensitive to the change of community memberships because they estimate communities using a one-period adjacency matrix. Despite the result, our method still achieves the lowest misclustering rate among all methods when the community memberships are relatively stable ($s < N/2$).

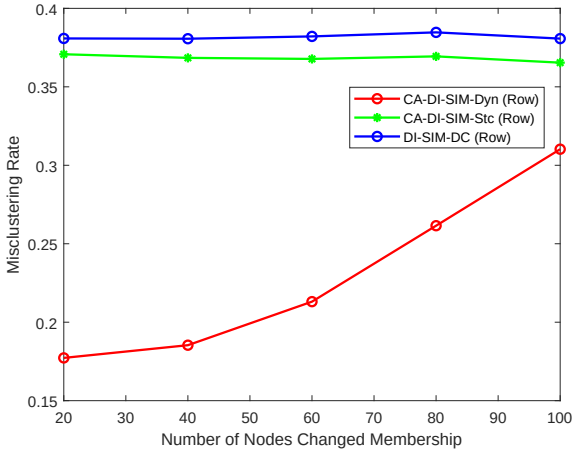
To perceptually illustrate the supreme accuracy of the dynamic CASC over the benchmark methods, we visualize the block structure of the dynamic network in Figure 6 by drawing the average adjacency matrices over a certain period and sort the rows and columns according to their community labels. Specifically, we simulate a dynamic network of 200 nodes with four blocks across 10 periods according to the block probability matrix B_t in Equation (16) by setting parameter values $s = 10$, $R = \lfloor \ln(NT) \rfloor$, and $X \sim U(0, 10)$. Then, we calculate the average adjacency matrix over six periods, $1/6 \sum_{t=1}^{6} A_t$, and sort the rows and columns jointly according to their community labels. We choose six periods because the optimal bandwidth r is chosen as 5 in the simulation.



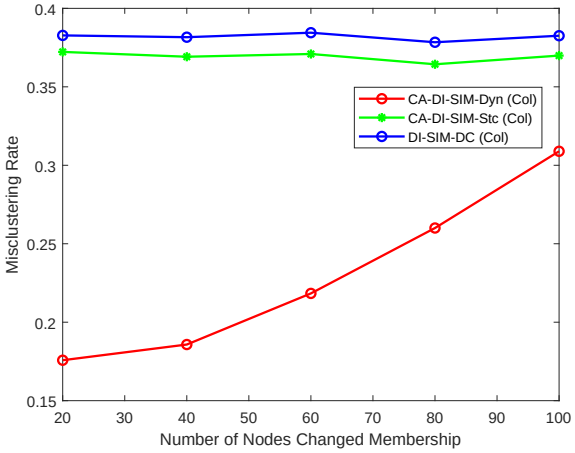
(a) Directed Network (Row Community)



(b) Directed Network (Column Community)



(c) Directed Network (Row Community)



(d) Directed Network (Column Community)

Figure 5: Misclustering Rates. This figure reports misclustering rates of different spectral clustering algorithms for networks with a growing number of vertices or with a growing number of membership changes. The CA-DI-SIM-Dyn represents Algorithm 1. The 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 subfigures (a) and (b), the number of nodes (N) grows from 20 to 200 with a step size of 20, and the number of membership changes is fixed at $s = 10$. In subfigures (c) and (d), the number of nodes is fixed at 200, and the number of membership changes (s) grows from 20 to 100 with stepsize of 20. The number of periods is $T = 10$, and the number of blocks is $K = 4$, with the block probability matrix given in equation (16). The covariates are simulated with $R = \lfloor \ln(NT) \rfloor$ and $X \sim U(0, 10)$. The smoothness parameter for kernel construction is $L = 4$. All simulations are conducted 1000 times.

Thus, using six periods will lose no information in the kernel. Taking the average adjacency matrix as a 200×200 grid, we paint the square with a darker color when it has a larger value.

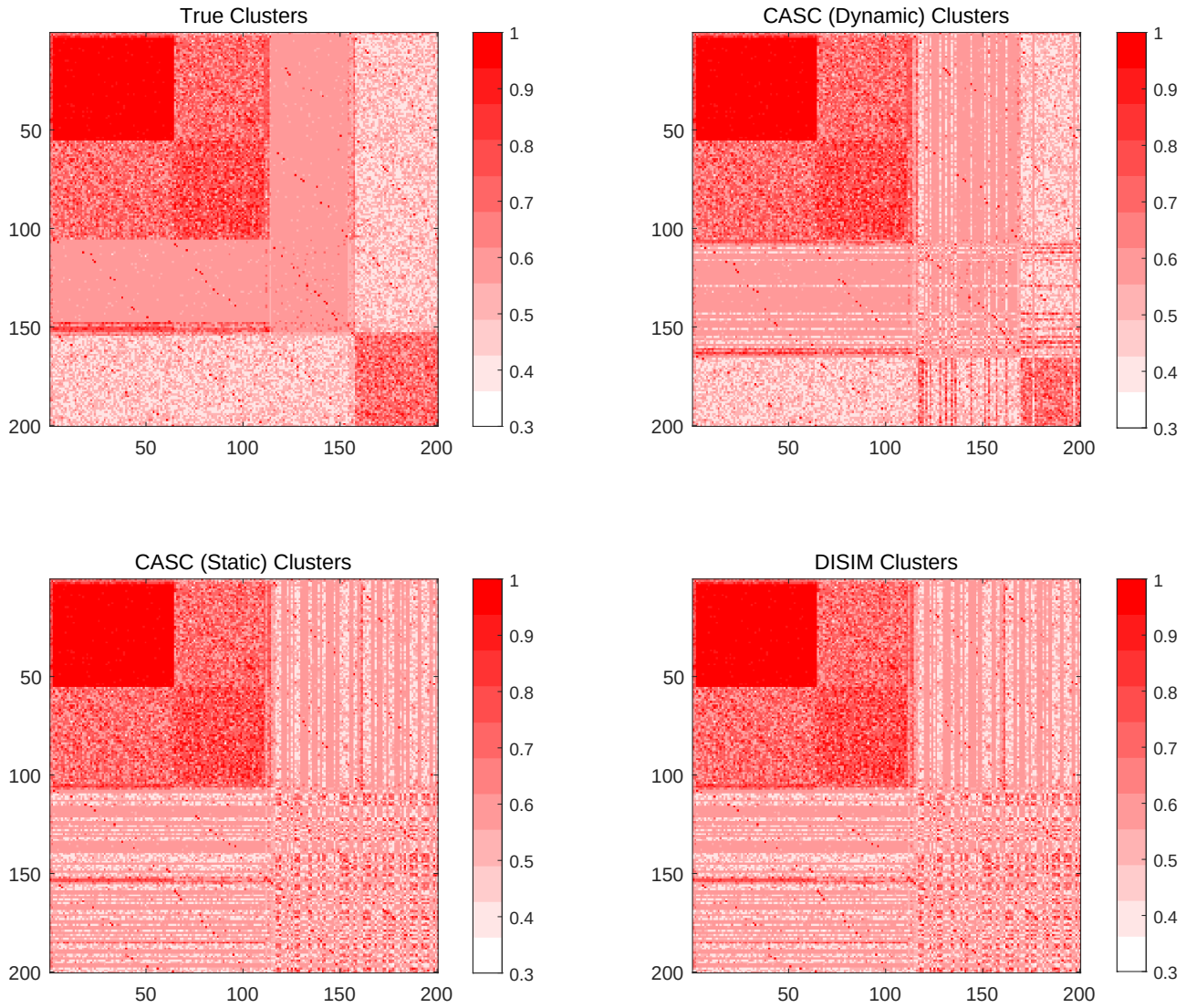


Figure 6: The True and the Estimated Blocks of the Network. This figure presents the block structure of the averaged adjacency matrices according to the true communities (simulation setting) and the estimated communities using different clustering methods. The averaged adjacency matrix reported is $1/6 \sum_{t=1}^{6} A_t$. The adjacency matrices have $N = 200$ nodes with $K = 4$ communities. The number of periods is $T = 10$, and the block probability matrix is given in Equation (16). The covariates are simulated with $R = \lfloor \ln(NT) \rfloor$ and $X \sim U(0, 10)$. The number of community member changes is fixed at $s = 20$, and the smoothness parameter for kernel construction is $L = 4$. All simulations are conducted 1000 times.

Evidently, the graph with true communities exhibits a clear block structure where each block is

associated with reasonable connection frequencies abide by the block probability matrix. All three methods successfully recover the dense blocks (blocks in the northwest corner), whereas only the dynamic CASC can recover the less dense blocks (blocks in the southeast corner). The simulation results serve as remarkable evidence to confirm the advantages of the dynamic CASC in dealing with dense networks.

3.5 Empirical communities

Applying the dynamic CASC algorithm to the cryptos network, we classify the crypto network into four column communities. We present three representative cryptos in each estimated community in Table 1. It shows that as of December 31, 2017, the largest cryptos (BTC, ETH, and LTC) in terms of market capitalization are not necessarily categorized into the same community. For example, although the return patterns of BTC and ETH are closely related, their fundamental attributes are somewhat different: BTC employs SHA256, while ETH uses KECCAK-256. For comparison, we also report the representative cryptos of each community estimated by DI-SIM (Rohe et al., 2016). It shows that the representative cryptos of each community are pretty different when choosing different classification algorithms, indicating that the DI-SIM-estimated communities could have distinctive features compared to dynamic CASC-estimated communities.

Table 1: Representative Cryptos of Communities Estimated by Dynamic CASC and by DI-SIM.

	Dynamic CASC			DI-SIM		
	1	2	3	1	2	3
Community 1	Bitcoin (BTC)	Litecoin (LTC)	Monero (XMR)	Bitcoin (BTC)	Bytecoin (BCN)	Dogecoin (DOGE)
Community 2	NEM (XEM)	Dash (DASH)	Syscoin (SYS)	Dash (DASH)	BitcoinDark (BTCD)	BitBay (BAY)
Community 3	ReddCoin (RDD)	Einsteinium (EMC2)	Crown (CRW)	ReddCoin (RDD)	Aeon (AEON)	OKCash (OK)
Community 4	Ripple (XRP)	Ethereum (ETH)	Augur (REP)	Ripple (XRP)	Ethereum (ETH)	Litecoin (LTC)

NOTE: This table lists the top three cryptos in market capitalization of each community estimated by dynamic CASC and by DI-SIM on December 31, 2017. The crypto abbreviations are provided in the brackets below their names.

To evaluate the classification efficiency under empirical settings, we define two important mea-

sures, the average within- and cross-community degrees:

$$\textit{Within-Community Degrees} = \frac{1}{T}\sum_{t=1}^{T}\left[\frac{1}{N_C^2}\sum_{i\in\mathcal{C}}\sum_{j\in\mathcal{C}}A_t(i,j)\right],$$

$$\textit{Cross-Community Degrees} = \frac{1}{T}\sum_{t=1}^{T}\left\{\frac{1}{2N_C(N-N_C)}\sum_{i\in\mathcal{C}}\sum_{j\notin\mathcal{C}}[A_t(i,j)+A_t(j,i)]\right\},$$

where N_C is the number of cryptos in community $\mathcal{C}$, and N is the total number of cryptos. Intuitively, if the clustering method correctly classifies all cryptos, then the within-community degrees should be larger than the cross-community degrees, that is, the difference between the two measures should be positive. Table 2 summarizes the average within- and cross-community degrees of network linkages induced by the return cross-predictability, the technological similarities, and the joint similarities based on the dynamic CASC and the DI-SIM, respectively.

Table 2: Average Within- and Cross-Community Degrees

	Return				Covariates				Combined			
	Within	Cross	Diff.	t-stat.	Within	Cross	Diff.	t-stat.	Within	Cross	Diff.	t-stat.
Panel A: Dynamic CASC												
Community 1	3.24	3.31	−0.07	−0.97	47.6	43.7	3.97	10.38	76.5	76.1	0.39	1.61
Community 2	3.40	3.38	0.03	0.30	47.4	43.3	4.03	11.60	76.7	76.3	0.39	1.51
Community 3	3.42	3.33	0.10	1.24	48.0	43.5	4.53	13.70	76.9	76.3	0.58	2.31
Community 4	3.61	3.42	0.19	2.55	48.0	43.4	4.53	9.83	75.6	75.9	−0.32	−1.24
Panel B: DI-SIM												
Community 1	2.58	3.04	−0.46	−3.74	44.4	44.2	0.20	0.01	72.0	76.1	−4.09	−6.26
Community 2	2.38	2.83	−0.45	−6.16	52.4	47.2	5.20	14.30	88.9	81.0	7.96	9.63
Community 3	2.27	2.62	−0.35	−2.50	38.3	41.2	−2.00	−3.09	74.1	75.5	−1.46	−3.84
Community 4	5.82	4.21	1.61	6.36	53.5	45.9	7.60	41.24	78.7	77.4	1.30	5.73

NOTE: This table reports the average within- and cross-community degrees (in %) for each community. Panel A reports the results based on the communities estimated by dynamic CASC and Panel B presents the results based on the communities estimated by DI-SIM. The first four columns report the results for the network linkages based on return cross-predictability. The middle four columns report results for the network linkages based on technological similarity. The last four columns present the results for the network linkages based on the combined information. The t-statistics are based on the Newey–West robust standard errors with 4 lags. The sample period ranges from January 1, 2016 to December 31, 2018.

Consistent with our simulations, the dynamic CASC method shows superior classification effi-

ciency than that of DI-SIM. It delivers more positive and significant differences in average degrees for all three types of network linkages. More interestingly, even though the DI-SIM only uses the return information to classify crypto, its performance for classifying return linkages is worse than that of the dynamic CASC, which needs to balance the return and technology information. This striking result proves that technological similarity assists extracting community information from return dynamics by emphasizing the return cross-predictability cause by common technological risk, and thus improves the classification.

4 Economic Implications

In this section, we conduct three empirical tests to assess the economic contributions of crypto communities to the cryptos market. First, we examine whether the investors can diversify risks better by investing cryptos across different communities than those within the same community. Second, we explore the asset pricing implication of information propagation effects on within-community cryptos by constructing a cross-sectional portfolio that implements an inter-crypto momentum trading strategy. Finally, we dissect the portfolio return on several behavioral factors as falsification tests to support our informational mechanism.

4.1 Risk diversification

Diversification is one of the major components of investment decision making under risk and uncertainty. Its underlying philosophy is to minimize both the probability of portfolio loss and its severity through multilateral insurance, in which each asset is insured by the remaining assets. Seminal papers by Markowitz (1952) and Roy (1952) mathematically formalized the idea of diversification of investment and showed that diversification can reduce risk without changing the expected portfolio returns. According to Rubinstein (2002), the key ingredient of risk diversification is the return covariance between risky assets that to be selected into the portfolio. The less positively correlated the assets, the lower the probability that the assets suffer losses simultaneously in the same proportion, and the better the protection offered by this multilateral insurance, namely diversification (Koumou, 2020).

To testify that crypto communities facilitate risk diversification, we compare the out-of-sample

average pairwise return correlations of cryptos within and across the communities. Specifically, we first estimate the crypto communities on each trading day and compute the pairwise Pearson correlation coefficients using the next 7- (1-week) and 30-day (1-month) crypto returns. Then, we calculate the average within-community correlations for each community by averaging the correlation coefficients of the crypto pairs whose members both belong to the focal community. We also obtain the average cross-community correlations for each community by averaging the correlation coefficients of the crypto pairs where only one of the members belongs to the focal community. Ideally, the average within-community correlations should be significantly higher than the average cross-community correlations for all communities because the cryptos within the same community share common sources of price and technological shocks. For comparison, we also present the results with the communities being estimated by DI-SIM.

Table 3: Within- and Cross-Community Average Return Correlations

	7-Day Return Correlations				30-Day Return Correlations			
	Within	Cross	Diff.	t-stat.	Within	Cross	Diff.	t-stat.
Panel A: Dynamic CASC								
Community 1	0.366	0.337	0.029	7.55	0.345	0.317	0.028	9.00
Community 2	0.369	0.337	0.032	8.10	0.350	0.317	0.033	9.87
Community 3	0.381	0.341	0.040	9.46	0.358	0.318	0.040	10.68
Community 4	0.381	0.340	0.041	10.99	0.362	0.319	0.042	11.64
Panel B: DI-SIM								
Community 1	0.336	0.328	0.008	1.24	0.322	0.309	0.014	0.85
Community 2	0.295	0.307	−0.012	−1.02	0.268	0.284	−0.016	−0.43
Community 3	0.562	0.388	0.173	10.15	0.550	0.369	0.181	4.34
Community 4	0.296	0.310	−0.014	−1.78	0.266	0.284	−0.018	−1.48

NOTE: This table reports the out-of-sample average return correlations of the cryptos within and across communities. Panel A presents the results for communities estimated by dynamic CASC and Panel B reports the results for communities estimated by DI-SIM. We first calculate the pairwise return correlations between cryptos in each period using the 7- and 30-day future returns. Then, we compute the within-community (cross-community) average return correlations using the correlation coefficients between the cryptos from the same community (different communities). The t-statistics reported are based on the Newey–West robust standard errors with 4 lags. The sample period spans from January 1, 2016 to December 31, 2018.

Table 3 summarizes the results, and three findings are obtained. First, the return correlations between cryptos within the same community are, on average, significantly higher than those across

communities. For example, in Panel A, the average within- and cross-community 30-day-ahead return correlations for Community 1 are 34.5 % and 31.7%, respectively. The correlation difference is 2.8%, which is significant at the 1% level with a Newey–West adjusted t-statistic of 9.0. Economically, the substantial reduction in return correlations suggests that investors can find attractive upside and diversification possibilities through allocating portfolio weights on cryptos from different communities.

Second, the dynamic CASC-estimated communities exhibit better risk diversification performance than the DI-SIM-estimated communities. In particular, all the dynamic CASC-estimated communities possess significantly higher within-community return correlations than cross-community return correlations. By contrast, only one DI-SIM community (Community 3) shows significantly positive differences between the within- and cross-community return correlations. The consistent correlation patterns across the dynamic CASC-estimated communities make it more convenient for investors to form a diversified crypto portfolio following the cross-community principle. This is also empirical evidence that supports the validity of the dynamic CASC algorithm for community detection in time-varying networks.

Finally, the differences between the within- and cross-community return correlations are robust to the length of the evaluation window. Specifically, the changes between the 7- and 30-day return correlation differences for dynamic CASC-estimated communities are around 0.1%. From practitioners' perspective, a stable return correlation pattern allows the investors to make flexible investment decisions on diversifying short- or long-term risks.

In summary, the dynamic CASC-estimated communities have meaningful implications for investors' portfolio choices and risk diversification through delivering lower cross-community return correlations robustly across all communities and over different horizons.

4.2 Momentum spillover

Since information diffuses gradually across investors (Hong and Stein, 1999; Hong, Lim, and Stein, 2000), the price of an asset may react sluggishly to the arrival of relevant news/information about other assets. This suggests that assets having fundamental similarities will have momentum spillovers, wherein past return of one asset predicts the returns of assets linked to it. Numerous papers have verified such spillovers in the equity market by using a variety of proxies for inter-firm

linkages (see, e.g., Moskowitz and Grinblatt, 1999; Cohen and Frazzini, 2008; Menzly and Ozbas, 2010; Lee et al., 2019; Ali and Hirshleifer, 2020; Parsons et al., 2020). Similarly, the estimated communities in the cryptos market are designed to optimally capture the information propagation and technological similarities between cryptos. Therefore, we expect the momentum spillover effects to exist within each community in the cryptos market.

We construct a long-short portfolio that implements an inter-crypto momentum trading strategy to formally test the momentum spillover effects and justify the information propagation mechanism between similar cryptos in the network. The portfolio is constructed in three steps. First, for each crypto i on day t, we compute the average return of all the other cryptos within the same community (R_{it}) to form a trading signal:

$$R_{it} = \frac{1}{|\mathcal{C}_{-i}|} \sum_{j \in \mathcal{C}_{-i}} r_{jt},$$

where $\mathcal{C}_{-i}$ denotes the set of other cryptos that are in the same community as crypto i. Second, we sort cryptos into four quartiles according to the trading signal and label the cryptos in the top/bottom quartile as the winner/loser cryptos. Finally, we buy the winner cryptos and sell the loser cryptos with equal positions at the end of day t and rebalance the trading positions at the end of next trading day.

We examine the portfolio return up to 7 days after the signal formation period. The hypothesis is that the information propagation mechanism should predict a drifting return pattern instead of a return reversal. Ideally, the winner/loser cryptos will generate higher/lower returns as the positive/negative shocks transmit through the network linkages between similar cryptos.

Table 4 reports the results. Consistent with our hypothesis, the long-short portfolio generates a one-day-ahead average daily return of 1.08% (t-stat. 12.24), statistically significant at the 1% level. The results exhibit a strong monotonic relationship between past connected crypto returns and future crypto returns. More importantly, the positive portfolio return does not revert over the horizons. The portfolio analysis confirms that our network model and community structure effectively capture the information propagation between cryptos and provide economically significant investment opportunities.

Table 4: Average Returns of the Inter-Crypto Momentum Portfolio over the Horizons

	Ret_{t+1}	Ret_{t+2}	Ret_{t+3}	Ret_{t+4}	Ret_{t+5}	Ret_{t+6}	Ret_{t+7}
Loser	0.06	0.55	0.55	0.51	0.69	0.58	0.61
2	0.38	0.53	0.58	0.53	0.58	0.49	0.51
3	0.75	0.63	0.50	0.60	0.45	0.49	0.63
Winner	1.14	0.59	0.67	0.65	0.56	0.73	0.52
Winner−Loser	1.08	0.05	0.12	0.14	−0.13	0.16	−0.09
t-statistic	12.24	0.55	1.30	1.56	−1.37	1.89	−1.05

NOTE: This table reports the average future returns (in %) of the inter-crypto momentum portfolio. The t-statistics are computed based on Newey–West robust standard errors with 4 lags. The sample period spans from January 1, 2016 to December 31, 2018.

4.3 Behavioral interpretation

The momentum spillover effect evidences strongly that return cross-predictability and technological similarities identify fundamental relationships between cryptos. However, the information propagation mechanism is not the only possibility that could generate the momentum spillover effect. Investors may be slow to carry information across cryptos due to a behavioral bias or constraint. Therefore, we explore three classical behavioral mechanisms commonly used to explain cross-sectional return anomalies (i.e., limit-to-arbitrage, investor attention, and economic policy uncertainty) to see if they can provide significant explanatory power for the momentum spillover effects in the cryptos market.

The first behavioral explanation is the limit-to-arbitrage. Cryptos market exhibits periods of large, recurrent arbitrage opportunities across exchanges (Makarov and Schoar, 2020). According to Shleifer and Vishny (1997), sophisticated investors would quickly eliminate any return predictability arising from anomalies in a frictionless market without impediments to arbitrage. Hou and Moskowitz (2005) documented that the asset prices will delay to respond to information due to the impact of market frictions. Therefore, when the cryptos market is more frictional, the information travels slower in the cryptos network, and delayed cryptos will not exhibit price momentum. Analogous to market friction measures in the equity market (Brunnermeier, Nagel, and Pedersen, 2008; Brunnermeier and Pedersen, 2009), we employ the cryptos market volatility index (VCRIX) constructed by Kim, Trimborn, and Härdle (2019) and TED spread (TED) as the market-friction indicators. Specifically, the volatility index affects the market liquidity, and TED determines the

investors' funding constraints. We split the sample into two subperiods (high and low) by cutting at the median of the market friction indicators. Then, we calculate and report the average portfolio returns within each subperiods in Table 5.

Table 5: Market Frictions and the Inter-Crypto Momentum Portfolio Returns

	VCRIX			TED		
	Low	High	High−Low	Low	High	High−Low
Loser	−0.03	0.10	0.13	0.17	−0.39	−0.56
2	0.43	0.36	−0.06	0.58	−0.05	−0.63
3	0.81	0.73	−0.08	1.03	−0.02	−1.05
Winner	1.27	1.08	−0.19	1.42	0.46	−0.96
Winner−Loser	1.30	0.98	−0.32	1.24	0.85	−0.39
t-statistic	8.98	8.86	−1.59	9.65	5.12	−1.20

NOTE: This table reports the inter-crypto momentum portfolio returns (in %) in subperiods of high and low market frictions. t-statistics are computed based on Newey–West robust standard errors with 4 lags. The sample period spans from January 1, 2016 to December 31, 2018.

Evidently, the average portfolio returns do not exhibit significant differences across the high and low episodes of market frictions. The average portfolio returns are generally more positive, albeit insignificant, when market frictions are lower, which is consistent with Hou and Moskowitz (2005). In summary, the momentum spillover effect exists regardless of market frictions, indicating that the limit-to-arbitrage mechanism does not explain our crypto anomalies.

The second behavioral explanation is investor attention. Peng and Xiong (2006) pointed out that investors have limited attention and are subject to overconfidence. Therefore, they need to process information to infer the value of cryptos and overestimate the precision of the acquired information due to overconfidence. Guo, Peng, Tao, and Tu (2021) show that investor attention could spill over along the network linkages. Together, the limited attention, attention spillover, and overconfidence lead to an overreaction-induced momentum effect in the cryptos network. Following Liu and Tsyvinski (2021), we proxy the market-wide investor attention by using the daily Google search volume for the words "crypto" and "cryptocurrency," subtracting its 1-month historical average. Considering the leading position of Bitcoin in the cryptos market (Griffin and Shams, 2020), we also proxy the market-wide investor attention with the investor attention paid to Bitcoin, that is, the Google search volume for the word "Bitcoin." Then, we repeat the test in Table 5 by

splitting the sample into high and low attention subperiods.

Table 6: Investor Attention and the Inter-Crypto Momentum Portfolio Returns

	Bitcoin Attention			Cryptos Attention		
	Low	High	High−Low	Low	High	High−Low
Loser	−0.02	0.20	0.22	0.33	−0.44	−0.77
2	0.24	0.66	0.42	0.71	−0.20	−0.90
3	0.48	1.26	0.78	0.95	0.41	−0.53
Winner	1.02	1.36	0.33	1.43	0.63	−0.80
Winner−Loser	1.04	1.15	0.11	1.09	1.06	−0.03
t-statistic	9.61	7.54	0.55	9.58	7.70	−0.16

NOTE: This table reports the inter-crypto momentum portfolio returns (in %) in subperiods of high and low investor attention. t-statistics are computed based on Newey–West robust standard errors with 4 lags. The sample period spans from January 1, 2016 to December 31, 2018.

The results are presented in Table 6. Clearly, the inter-crypto momentum portfolio returns do not exhibit significant differences in high and low episodes of investor attention. For example, the portfolio generates an average daily return of 1.15% during low Bitcoin attention episodes while retaining a 1.04% average daily return for the high ones. The difference is merely 0.11% and statistically insignificant. The portfolio return patterns under high and low cryptos attention episodes are similar to that of Bitcoin attention. These results imply that investor attention is not the mechanism that drives the momentum spillover effects in the cryptos market.

Finally, observing that government policy and crypto prices have a strong synchronization (Demir, Gozgor, Lau, and Vigne, 2018; Cheng and Yen, 2020), it is natural to wonder whether the portfolio returns are related to underlying policy uncertainty. By representing uncertainty with investor belief dispersion (disagreement), Atmaz and Basak (2018) theoretically showed that higher disagreement leads to higher average bias and more overvaluation, thereby implying a negative disagreement–return relationship, as empirically documented in Brogaard and Detzel (2015). Based on Daniel, Hirshleifer, and Subrahmanyam (1998), when there are misinformed investors who underreact to firm-specific news, the price momentum will occur. Therefore, more uncertainty leads to price momentum. To examine this hypothesis, we employ the U.S. and Global economic policy uncertainty indices (EPU and GEPU) introduced by Baker, Bloom, and Davis (2016) to examine the portfolio performance in high and low uncertainty subperiods, respectively.

Table 7: Economic Policy Uncertainty and the Inter-Crypto Momentum Portfolio Returns

	EPU			GEPU		
	Low	High	High−Low	Low	High	High−Low
Loser	−0.30	0.38	0.68	−2.01	0.16	2.17
2	−0.10	0.83	0.93	−1.76	0.49	2.25
3	0.51	0.98	0.46	−1.61	0.87	2.49
Winner	0.71	1.53	0.83	−1.07	1.25	2.32
Winner−Loser	1.01	1.15	0.14	0.94	1.09	0.14
t-statistic	7.70	8.88	0.77	3.07	11.91	0.33

NOTE: This table reports the inter-crypto momentum portfolio returns (in %) in subperiods of high and low U.S. economic policy uncertainty (EPU) and global currency policy uncertainty (GEPU-Currency). t-statistics are computed based on Newey–West robust standard errors with 4 lags. The sample period spans from January 1, 2016 to December 31, 2018.

Table 7 summarizes the results. It is evident that the portfolio returns are statistically indifferent across high and low uncertainty subperiods. Specifically, in high EPU episodes, the portfolio return is 1.15% bps, whereas it is 1.01% in low EPU episodes. The difference between the portfolio returns is 0.14% with a Newey–West adjusted t-statistic of 0.77. When replacing the EPU with the GEPU indicator, the difference in portfolio returns remains the same magnitude and statistically insignificant. Both results indicate that policy uncertainty or investor disagreement has no impact on the momentum spillover effects in the cryptos market.

In summary, by dissecting the portfolio returns with limit-to-arbitrage, investor attention, and economic policy uncertainty factors, we falsify the potential behavioral mechanisms that would generate momentum spillover effects in the cryptos market. The joint evidence in the current section and Section 4.2 strongly supports that the momentum spillover effects result from information propagation between fundamentally similar cryptos.

5 Conclusions

In this paper, we examine the market segmentation problem and uncover the information propagation mechanism in the cryptos market by constructing a time-varying network of cryptos that combines return cross-predictability and technological similarities. To analyze this network, we propose a dynamic CASC algorithm for community detection and prove its classification consis-

tency. We demonstrate the performance of our methodology on simulated networks and show that it generally performs better than other state-of-the-art methods. Empirically, we apply dynamic CASC to the cryptos network and obtain four latent communities. Based on the estimated communities, we show that investors can better diversify risk through investing in cross-community cryptos and make a daily profit of 1.08% through an inter-crypto momentum trading strategy within the communities. Finally, we dissect the portfolio returns on several behavioral mechanisms, namely limit-to-arbitrage, investor attention, and economic policy uncertainty, but find no explanatory power. These results confirm that information propagation rather than behavioral bias drives our findings.

This paper mainly focuses on the uniform consistency of the classification method. Recent literature has developed inspiring framework for statistical inference on membership profiles as well as on estimation of low-rank matrices (e.g., Fan, Fan, Han, and Lv, 2019; Chen, Fan, Ma, and Yan, 2019). It would be an interesting and critical research direction to investigate how to accommodate dynamic networks and time-varying membership in these frameworks. We will leave them to future research.

Supplementary materials. The supplementary material contains details about tuning parameter choice, proofs of Theorem 1 in Section 3.

Acknowledgements. The authors sincerely thank the editor, Professor Jianqing Fan, an associate editor, and two anonymous referees for their helpful and insightful comments that have significantly improved the article. We are also grateful to Ying Chen, Dashan Huang, Oliver Linton, Peter C. B. Phillips, Shuyang Sheng, Vladimir Spokoiny, Liangjun Su, Jun Yu, Yichong Zhang, and all participants of the 2017 "Crypto-Currencies in a Digital Economy" workshop at Humboldt-Universität zu Berlin, the 2018 China Meeting of the Econometric Society, the 2019 SH3 Conference on Econometrics, the 2019 STAT of ML Conference, and the 2019 SoFiE 12th Annual Conference for their helpful discussions and comments.

Funding. Financial support of the European Union's Horizon 2020 research and innovation program "FIN- TECH: A Financial supervision and Technology compliance training programme" under the grant agreement No 825215 (Topic: ICT-35-2018, Type of action: CSA), the European

Cooperation in Science & Technology COST Action grant CA19130 - Fintech and Artificial Intelligence in Finance - Towards a transparent financial industry, the Deutsche Forschungsgemeinschaft's IRTG 1792 grant, the Yushan Scholar Program of Taiwan and the Czech Science Foundation's grant no. 19-28231X / CAS: XDA 23020303 are greatly acknowledged.

References

Abadi, J. and M. Brunnermeier (2018). Blockchain economics. *NBER Working Paper 25407*.

Abbe, E., J. Fan, and K. Wang (2020). An ℓ_p theory of PCA and spectral clustering. *arXiv preprint arXiv:2006.14062*.

Acemoglu, D., A. Ozdaglar, and A. Tahbaz-Salehi (2015). Systemic risk and stability in financial networks. *American Economic Review 105*(2), 564–608.

Ali, U. and D. Hirshleifer (2020). Shared analyst coverage: unifying momentum spillover effects. *Journal of Financial Economics 136*(3), 649–675.

Altman, N. S. (1990). Kernel smoothing of data with correlated errors. *Journal of the American Statistical Association 85*(411), 749–759.

Amini, A. A., A. Chen, P. J. Bickel, and E. Levina (2013). Pseudo-likelihood methods for community detection in large sparse networks. *Annals of Statistics 41*(4), 2097–2122.

Amini, A. A. and E. Levina (2018). On semidefinite relaxations for the block model. *Annals of Statistics 46*(1), 149–179.

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*(3), 1191–1224.

Atmaz, A. and S. Basak (2018). Belief dispersion in the stock market. *Journal of Finance 73*(3), 1225–1279.

Baker, S. R., N. Bloom, and S. J. Davis (2016). Measuring economic policy uncertainty. *Quarterly Journal of Economics 131*(4), 1593–1636.

Biais, B., C. Bisiere, M. Bouvard, and C. Casamatta (2019). The blockchain folk theorem. *Review of Financial Studies 32*(5), 1662–1715.

Bickel, P., D. Choi, X. Chang, and H. Zhang (2013). Asymptotic normality of maximum likelihood

and its variational approximation for stochastic blockmodels. *Annals of Statistics 41*(4), 1922–1943.

Bickel, P. J. and A. Chen (2009). A nonparametric view of network models and newman–girvan and other modularities. *Proceedings of the National Academy of Sciences 106*(50), 21068–21073.

Binkiewicz, N., J. T. Vogelstein, and K. Rohe (2017). Covariate-assisted spectral clustering. *Biometrika 104*(2), 361–377.

Brogaard, J. and A. Detzel (2015). The asset-pricing implications of government economic policy uncertainty. *Management Science 61*(1), 3–18.

Brunnermeier, M. K., S. Nagel, and L. H. Pedersen (2008). Carry trades and currency crashes. *NBER Macroeconomics Annual 23*(1), 313–348.

Brunnermeier, M. K. and L. H. Pedersen (2009). Market liquidity and funding liquidity. *Review of Financial Studies 22*(6), 2201–2238.

Budish, E. (2018). The economic limits of bitcoin and the blockchain. *NBER Working Paper 24717*.

Cai, T. T. and X. Li (2015). Robust and computationally feasible community detection in the presence of arbitrary outlier nodes. *Annals of Statistics 43*(3), 1027–1059.

Celisse, A., J.-J. Daudin, and L. Pierre (2012). Consistency of maximum-likelihood and variational estimators in the stochastic block model. *Electronic Journal of Statistics 6*, 1847–1899.

Chen, C. Y.-H., W. K. Härdle, and Y. Okhrin (2019). Tail event driven networks of SIFIs. *Journal of Econometrics 208*(1), 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.

Chen, Y., J. Fan, C. Ma, and Y. Yan (2019). Inference and uncertainty quantification for noisy matrix completion. *Proceedings of the National Academy of Sciences 116*(46), 22931–22937.

Chen, Y. and J. Xu (2016). Statistical-computational tradeoffs in planted problems and submatrix localization with a growing number of clusters and submatrices. *Journal of Machine Learning Research 17*(1), 882–938.

Cheng, H.-P. and K.-C. Yen (2020). The relationship between the economic policy uncertainty and the cryptocurrency market. *Finance Research Letters 35*, 101308.

Cohen, L. and A. Frazzini (2008). Economic links and predictable returns. *Journal of Fi-*

nance 63(4), 1977–2011.

Cong, L. W., Z. He, and J. Li (2021). Decentralized mining in centralized pools. *Review of Financial Studies 34*(3), 1191–1235.

Cong, L. W., Y. Li, and N. Wang (2021a). Token-based platform finance. *Journal of Financial Economics, forthcoming*.

Cong, L. W., Y. Li, and N. Wang (2021b). Tokenomics: dynamic adoption and valuation. *Review of Financial Studies 34*(3), 1105–1155.

Daniel, K., D. Hirshleifer, and A. Subrahmanyam (1998). Investor psychology and security market under-and overreactions. *Journal of Finance 53*(6), 1839–1885.

Decelle, A., F. Krzakala, C. Moore, and L. Zdeborová (2011). Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Physical Review E 84*(6), 066106.

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.

Deshpande, Y., S. Sen, A. Montanari, and E. Mossel (2018). Contextual stochastic block models. In *Advances in Neural Information Processing Systems*, Volume 31.

Detzel, A., H. Liu, J. Strauss, G. Zhou, and Y. Zhu (2021). Learning and predictability via technical analysis: evidence from bitcoin and stocks with hard-to-value fundamentals. *Financial Management 50*(1), 107–137.

Easley, D., M. O'Hara, and S. Basu (2019). From mining to markets: the evolution of bitcoin transaction fees. *Journal of Financial Economics 134*(1), 91–109.

Esmaeili, M., H. M. Saad, and A. Nosratinia (2021). Semidefinite programming for community detection with side information. *IEEE Transactions on Network Science and Engineering*.

Fama, E. F. and K. R. French (1993). Common risk factors in stock and bond returns. *Journal of Financial Economics 33*(1), 3–56.

Fama, E. F., K. R. French, D. G. Booth, and R. Sinquefield (1993). Differences in the risks and returns of NYSE and NASD stocks. *Financial Analysts Journal 49*(1), 37–41.

Fan, J., Y. Fan, X. Han, and J. Lv (2019). SIMPLE: Statistical inference on membership profiles in large networks. *arXiv preprint arXiv:1910.01734*.

Fan, J., Y. Liao, and M. Mincheva (2013). Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society. Series B, Statistical methodology 75*(4).

Fanti, G., L. Kogan, and P. Viswanath (2019). Economics of Proof-of-Stake payment systems. *Working paper, MIT*.

Foerster, S. R. and G. A. Karolyi (1999). The effects of market segmentation and investor recognition on asset prices: evidence from foreign stocks listing in the united states. *Journal of Finance 54*(3), 981–1013.

Gao, C., Z. Ma, A. Y. Zhang, and H. H. Zhou (2018). Community detection in degree-corrected block models. *Annals of Statistics 46*(5), 2153–2185.

Gasser, T. and H.-G. Müller (1979). Kernel estimation of regression functions. In *Smoothing techniques for curve estimation*, pp. 23–68. Springer.

Griffin, J. M. (2002). Are the Fama and French factors global or country specific? *The Review of Financial Studies 15*(3), 783–803.

Griffin, J. M. and A. Shams (2020). Is Bitcoin really un-Tethered? *Journal of Finance 75*(4), 1913–1964.

Guo, L., L. Peng, Y. Tao, and J. Tu (2021). Joint news, attention spillover, and market returns. *Available at SSRN 2927561*.

Hajek, B., Y. Wu, and J. Xu (2016a). Achieving exact cluster recovery threshold via semidefinite programming. *IEEE Transactions on Information Theory 62*(5), 2788–2797.

Hajek, B., Y. Wu, and J. Xu (2016b). Achieving exact cluster recovery threshold via semidefinite programming: Extensions. *IEEE Transactions on Information Theory 62*(10), 5918–5937.

Herskovic, B. (2018). Networks in production: asset pricing implications. *Journal of Finance 73*(4), 1785–1818.

Hinzen, F. J., K. John, and F. Saleh (2019). Proof-of-Work's limited adoption problem. *Working Paper, NYU Stern School of Business*.

Hoberg, G. and G. Phillips (2016). Text-based network industries and endogenous product differentiation. *Journal of Political Economy 124*(5), 1423–1465.

Holland, P. W., K. B. Laskey, and S. Leinhardt (1983). Stochastic blockmodels: first steps. *Social*

Networks 5(2), 109–137.

Hong, H., T. Lim, and J. C. Stein (2000). Bad news travels slowly: Size, analyst coverage, and the profitability of momentum strategies. *Journal of Finance 55*(1), 265–295.

Hong, H. and J. C. Stein (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets. *Journal of Finance 54*(6), 2143–2184.

Hou, K. and T. J. Moskowitz (2005). Market frictions, price delay, and the cross-section of expected returns. *Review of Financial Studies 18*(3), 981–1020.

Jin, J. (2015). Fast community detection by score. *Annals of Statistics 43*(1), 57–89.

Joseph, A. and B. Yu (2016). Impact of regularization on spectral clustering. *Annals of Statistics 44*(4), 1765–1791.

Kan, R. and G. Kirikos (1995). Biases in evaluating trading strategies. *Working Paper, University of Toronto*.

Karrer, B. and M. E. J. Newman (2011). Stochastic blockmodels and community structure in networks. *Physical Review E 83*(1), 016107.

Kim, A., S. Trimborn, and W. K. Härdle (2019). VCRIX – a volatility index for cryptocurrencies. *Available at SSRN 3480348*.

Koumou, G. B. (2020). Diversification and portfolio theory: a review. *Financial Markets and Portfolio Management 34*, 267–312.

Lan, W., Z. Fang, H. Wang, and C.-L. Tsai (2018). Covariance matrix estimation via network structure. *Journal of Business & Economic Statistics 36*(2), 359–369.

Lee, C. M., S. T. Sun, R. Wang, and R. Zhang (2019). Technological links and predictable returns. *Journal of Financial Economics 132*(3), 76–96.

Lei, J. and A. Rinaldo (2015). Consistency of spectral clustering in stochastic block models. *Annals of Statistics 43*(1), 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. *Annals of Statistics 25*(3), 929–947.

Li, T., E. Levina, and J. Zhu (2020). Network cross-validation by edge sampling. *Biometrika 107*(2), 257–276.

Li, X., Y. Chen, and J. Xu (2021). Convex relaxation methods for community detection. *Statistical Science 36*(1), 2–15.

Liu, J., I. W. Marsh, P. Mazza, and M. Petitjean (2021). Factor structure in cryptocurrency returns and volatility. *Available at SSRN 3389152*.

Liu, Y., J. Sheng, and W. Wang (2021). Technology and cryptocurrency valuation: evidence from machine learning. *Available at SSRN 3577208*.

Liu, Y. and A. Tsyvinski (2021). Risks and returns of cryptocurrency. *Review of Financial Studies 34*(6), 2689–2727.

Liu, Y., A. Tsyvinski, and X. Wu (2021). Common risk factors in cryptocurrency. *Journal of Finance, forthcoming*.

Lu, C. and S. Sen (2020). Contextual stochastic block model: Sharp thresholds and contiguity. *arXiv preprint arXiv:2011.09841*.

Ma, Z. and Z. Ma (2017). Exploration of large networks with covariates via fast and universal latent space model fitting. *arXiv preprint arXiv:1705.02372*.

Makarov, I. and A. Schoar (2020). Trading and arbitrage in cryptocurrency markets. *Journal of Financial Economics 135*(2), 293–319.

Markowitz, H. (1952). The utility of wealth. *Journal of Political Economy 60*(2), 151–158.

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

Menzly, L. and O. Ozbas (2010). Market segmentation and cross-predictability of returns. *Journal of Finance 65*(4), 1555–1580.

Moskowitz, T. J. and M. Grinblatt (1999). Do industries explain momentum? *Journal of Finance 54*(4), 1249–1290.

Mu, C., A. Mele, L. Hao, J. Cape, A. Athreya, and C. E. Priebe (2020). On spectral algorithms for community detection in stochastic blockmodel graphs with vertex covariates. *arXiv preprint arXiv:2007.02156*.

Nowicki, K. and T. A. B. Snijders (2001). Estimation and prediction for stochastic blockstructures. *Journal of the American statistical association 96*(455), 1077–1087.

Pagnotta, E. (2020). Decentralizing money: Bitcoin prices and blockchain security. *Review of Financial Studies, forthcoming.*

Parsons, C. A., R. Sabbatucci, and S. Titman (2020). Geographic lead-lag effects. *Review of Financial Studies 33*(10), 4721–4770.

Patton, A. J. and B. Weller (2019). Risk price variation: The missing half of empirical asset pricing. *Economic Research Initiatives at Duke (ERID) Working Paper* (274).

Peng, L. and W. Xiong (2006). Investor attention, overconfidence and category learning. *Journal of Financial Economics 80*(3), 563–602.

Pensky, M. and T. Zhang (2019). Spectral clustering in the dynamic stochastic block model. *Electronic Journal of Statistics 13*(1), 678–709.

Qin, T. and K. Rohe (2013). Regularized spectral clustering under the degree-corrected stochastic blockmodel. In *Advances in Neural Information Processing Systems*, Volume 26, pp. 3120–3128.

Rohe, K., S. Chatterjee, and B. Yu (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *Annals of Statistics 39*(4), 1878–1915.

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*(45), 12679–12684.

Roy, A. D. (1952). Safety first and the holding of assets. *Econometrica*, 431–449.

Rubinstein, M. (2002). Markowitz's "portfolio selection": a fifty-year retrospective. *Journal of Finance 57*(3), 1041–1045.

Saleh, F. (2021). Blockchain without waste: Proof-of-Stake. *Review of Financial Studies 34*(3), 1156–1190.

Sharpe, W. F. (1966). Mutual fund performance. *Journal of Business 39*(1), 119–138.

Shleifer, A. and R. W. Vishny (1997). The limits of arbitrage. *Journal of Finance 52*(1), 35–55.

Sockin, M. and W. Xiong (2020). A model of cryptocurrencies. *NBER Working Paper 26816.*

Wang, Y. R. and P. J. Bickel (2017a). Likelihood-based model selection for stochastic block models. *Annals of Statistics 45*(2), 500–528.

Wang, Y. X. R. and P. J. Bickel (2017b). Likelihood-based model selection for stochastic block models. *Annals of Statistics 45*(2), 500–528.

Weng, H. and Y. Feng (2016). Community detection with nodal information. *arXiv preprint*

arXiv:1610.09735.

Yan, B. and P. Sarkar (2020). Covariate regularized community detection in sparse graphs. *Journal of the American Statistical Association*, 1–12.

Yermack, D. (2017). Corporate governance and blockchains. *Review of Finance 21*(1), 7–31.

Zhang, Y., E. Levina, and J. Zhu (2016). Community detection in networks with node features. *Electronic Journal of Statistics 10*(2), 3153–3178.

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*(2), 1096–1123.

Zhao, J., X. Liu, H. Wang, and C. Leng (2021). Dimension reduction for covariates in network data. *Biometrika, forthcoming*.

Zhao, Y., E. Levina, and J. Zhu (2012). Consistency of community detection in networks under degree-corrected stochastic block models. *Annals of Statistics 40*(4), 2266–2292.

Zhou, Z. and A. A. Amini (2019). Analysis of spectral clustering algorithms for community detection: the general bipartite setting. *Journal of Machine Learning Research 20*(1), 1774–1820.

Zhou, Z. and A. A. Amini (2020). Optimal bipartite network clustering. *Journal of Machine Learning Research 21*, 1–68.

Zhu, X., Z. Cai, and Y. Ma (2021). Network functional varying coefficient model. *Journal of the American Statistical Association*, 1–12.

Zou, H. (2006). The adaptive Lasso and its oracle properties. *Journal of the American Statistical Association 101*(476), 1418–1429.

Zou, T., W. Lan, R. Li, and C.-L. Tsai (2021). Inference on covariance-mean regression. *Journal of Econometrics*.

Zou, T., W. Lan, H. Wang, and C.-L. Tsai (2017). Covariance regression analysis. *Journal of the American Statistical Association 112*(517), 266–281.

Supplementary Appendix to "A Time-varying Network for Cryptocurrencies"

Li Guo*

Fudan University

Shanghai Institute of International Finance and Economics

Wolfgang Karl Härdle†

Humboldt-Universität zu Berlin, Singapore Management University

Xiamen University, Charles University

Yubo Tao‡

Singapore Management University

August 26, 2021

Supplementary Appendix to "A Time-varying Network for Cryptocurrencies"

This appendix comprises three sections. In section 6, we present how to choose the tuning parameters. In section 7, we provides the proofs of the main theorems in the above paper. In section 8, we present the necessary technical lemmas for proving the main results.

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. $\mathbf{1}_A$ denotes the indicator function of the event A.

6 Choice of Tuning Parameters

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).

*600 Guoquan Rd, Shanghai 200433. Email: guo_li@fudan.edu.cn.

†Unter den Linden 6 10099 Berlin, Germany. Email: haerdle@hu-berlin.de.

‡Corresponding author: 90 Stamford Rd, Singapore 178903. Email: ybtao@smu.edu.sg.

*Address: 600 Guoquan Rd, Shanghai 200433. Email: guo_li@fudan.edu.cn.

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

‡Correspondence author. Address: 90 Stamford Rd, Singapore 178903. Email: ybtao@smu.edu.sg.

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

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,

$$r^* = \arg\min_{0\leq r\leq T/2}\left((1-\eta)^{-2}\frac{W_{\max}\sqrt{d}}{r \vee 1} + \|\mathcal{S}_{t,r} - \mathcal{S}_t\|\right). \tag{17}$$

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

$$\widehat{r} = \max\left\{0 \leq r \leq T/2 : \left\|\widehat{\mathcal{S}}_{t,r} - \widehat{\mathcal{S}}_{t,\rho}\right\| \leq 4W_{\max}\sqrt{\frac{N\|\mathcal{S}_t\|_\infty}{\rho \vee 1}},\ \text{for any } \rho < r\right\}. \tag{18}$$

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

$$\alpha_t = \frac{\lambda_K(L_{\tau,t}) - \lambda_{K+1}(L_{\tau,t})}{\lambda_1(C_t^w)}. \tag{19}$$

Lastly, to determine K, we have several choices. Wang and Bickel (2017b) 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, Levina, and Zhu (2020) refine the network cross-validation approach by proposing an edge sampling algorithm. In our case, we apply the network cross-validation approach directly to the similarity matrix instead of the adjacency matrix. This is 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.

7 Proof of Main Results

7.1 Proof of Theorem 1

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

$$\frac{\left|\mathbb{M}_t^R\right|}{N} \leq \frac{8(2+\varepsilon)^2}{m_r^2 N} \left\|U_t - \mathcal{U}_t \mathcal{O}_t\right\|_F^2, \tag{20}$$

where

$$m_r \stackrel{\text{def}}{=} \min_{i,t}\{\min\{\|\Gamma_{R,t}(i,*)\|, \|\mathit{\Gamma}_{R,t}(i,*)\|\}\}, \tag{21}$$

and $\Gamma_{R,t}$ and $\mathit{\Gamma}_{R,t}$ are defined in Lemma 3.

Then, by improved version of Davis-Kahn theorem from Lei and Rinaldo (2015), we have

$$\left\|U_t - \mathcal{U}_t \mathcal{O}_t\right\|_F \leq \frac{2\sqrt{2K}}{\lambda_{K,t}} \left\|\operatorname{sym}\left(\mathcal{S}_{t,r} - \mathcal{S}_t\right)\right\|, \tag{22}$$

as $K_R \leq K_C$ and $K \stackrel{\text{def}}{=} \min\{K_R, K_C\}$. Base on equation (20) and (22), we have

$$\sup_t \frac{\left|\mathbb{M}_t^R\right|}{N} \leq \frac{2^6(2+\varepsilon)^2 K}{m_r^2 N \lambda_{K,\max}^2} \sup_t \left\|\operatorname{sym}\left(\mathcal{S}_{t,r} - \mathcal{S}_t\right)\right\|^2. \tag{23}$$

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

$$\mathcal{S}_t = \mathcal{D}_{R,t}^{-1/2} \Psi^R Z_{R,t} B_t Z_{C,t}^\top \Psi^C \mathcal{D}_{C,t}^{-1/2} + \alpha_t \mathcal{X} \mathcal{W}_t \mathcal{X}^\top, \tag{24}$$

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

$$\Delta(r) = \sup_t \left\|\operatorname{sym}\left(\widehat{\mathcal{S}}_{t,r} - \mathcal{S}_t\right)\right\| \leq \sup_t \left\|\operatorname{sym}\left(\widehat{\mathcal{S}}_{t,r} - \mathcal{S}_{t,r}\right)\right\| + \sup_t \left\|\operatorname{sym}\left(\mathcal{S}_{t,r} - \mathcal{S}_t\right)\right\| = \Delta_1(r) + \Delta_2(r). \tag{25}$$

Now, we evaluate $\Delta_1(r)$ and $\Delta_2(r)$ respectively. For $\Delta_1(r)$, by Lemma 2, we have

$$\sup_t \left\| \text{sym}\left(\widehat{\mathcal{S}}_{t,r} - \mathcal{S}_{t,r}\right)\right\| = \frac{1}{|\mathcal{F}_r|}\sum_{i\in\mathcal{F}_r} W_{r,\ell}(i) \sup_t \|\text{sym}\left(S_{t+i} - \mathcal{S}_{t+i}\right)\| \tag{26}$$
$$\leq W_{\max}(4+c_1)\left\{\frac{3\log(16NT/\epsilon)}{\underline{\delta}}\right\}^{1/2}.$$

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

$$\widetilde{\mathcal{S}}_{t,r} = \frac{1}{|\mathcal{F}_r|}\sum_{i\in\mathcal{F}_r} W_{r,\ell}(i)\left(Y_{R,t}B_{t+i}Y_{C,t}^\top + \alpha_{t+i}\mathcal{X}\mathcal{W}_{t+i}\mathcal{X}\right). \tag{27}$$

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

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

$$\sup_t \|\text{sym}\left(\mathcal{S}_{t,r} - \mathcal{S}_t\right)\| \leq \sup_t \left\|\text{sym}\left(\mathcal{S}_{t,r} - \widetilde{\mathcal{S}}_{t,r}\right)\right\| + \sup_t \left\|\text{sym}\left(\widetilde{\mathcal{S}}_{t,r} - \mathcal{S}_t\right)\right\| = \Delta_{21}(r) + \Delta_{22}(r). \tag{28}$$

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

$$\mathcal{Y}_t \stackrel{\text{def}}{=} \begin{bmatrix} Y_{R,t} & 0_{N\times K_C} \\ 0_{N\times K_R} & Y_{C,t}\end{bmatrix}. \tag{29}$$

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

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

Noticing that $\left\|\mathcal{D}_{p,t}^{-1/2}\right\| \leq \underline{\delta}^{-1/2}$ and $\|Z_{p,t}\| \leq P_{\max}^{1/2}$ for $p \in \{R, C\}$, and by Assumption 2, $\|Z_{p,t+i} - Z_{p,t}\| \leq rs$, we then have

$$\begin{aligned}
\|\mathcal{Y}_{t+i} - \mathcal{Y}_t\| &= \max\left\{\|Y_{R,t+i} - Y_{R,t}\|, \|Y_{C,t+i} - Y_{C,t}\|\right\} \\
&\leq \max\left\{\left\|\mathcal{D}_{R,t+i}^{-1/2}\Psi^R Z_{R,t+i} - \mathcal{D}_{R,t}^{-1/2}\Psi^R Z_{R,t}\right\|, \left\|\mathcal{D}_{C,t+i}^{-1/2}\Psi^C Z_{C,t+i} - \mathcal{D}_{C,t}^{-1/2}\Psi^C Z_{C,t}\right\|\right\} \\
&\leq \max\left\{\left\|\mathcal{D}_{R,t+i}^{-1/2}\right\| \|\Psi^R\|\|Z_{R,t+i} - Z_{R,t}\| + \left(\left\|\mathcal{D}_{R,t+i}^{-1/2}\right\| \|\Psi^R\| + \left\|\mathcal{D}_{R,t}^{-1/2}\right\| \|\Psi^R\|\right)\|Z_{R,t}\|,\right. \\
&\qquad \left.\left\|\mathcal{D}_{C,t+i}^{-1/2}\right\| \|\Psi^C\|\|Z_{C,t+i} - Z_{C,t}\| + \left(\left\|\mathcal{D}_{C,t+i}^{-1/2}\right\| \|\Psi^C\| + \left\|\mathcal{D}_{C,t}^{-1/2}\right\| \|\Psi^C\|\right)\|Z_{C,t}\|\right\} \\
&\leq \sqrt{\frac{2rs}{\underline{\delta}}} + \sqrt{\frac{4P_{\max}}{\underline{\delta}}}.
\end{aligned}$$

Combining results above, we have

$$\begin{aligned}
\Delta_{21}(r) &= \sup_t \left\|\text{sym}\left(\mathcal{S}_{t,r} - \widetilde{\mathcal{S}}_{t,r}\right)\right\| \\
&\leq W_{\max} \times (2\underline{\delta}^{-1/2} P_{\max}^{1/2}) \times K_C \times \left(\sqrt{\frac{2rs}{\underline{\delta}}} + \sqrt{\frac{4P_{\max}}{\underline{\delta}}}\right) \\
&= \frac{2W_{\max}K_C}{\underline{\delta}}\left(\sqrt{2P_{\max}rs} + 2P_{\max}\right).
\end{aligned} \tag{30}$$

Lastly, for $\Delta_{22}(r)$, by Assumption 4 and the definition of W_t giving $\|\mathrm{sym}(W_t)\| \leq 1$, we have

$$\begin{aligned}
\Delta_{22}(r) &= \sup_t \left\| \mathrm{sym}\left(\widetilde{\mathcal{S}}_{t,r} - \mathcal{S}_t\right) \right\| \\
&\leq \frac{1}{|\mathcal{F}_r|} \sum_{i \in \mathcal{F}_r} W_{r,\ell}(i) \sup_t \left(\left\| \mathrm{sym}\left(Y_{R,t}(B_{t+i} - B_t) Y_{C,t}^\top\right) \right\| + \left\| \mathrm{sym}\left(\alpha_{t+i} \mathcal{X} \mathcal{W}_{t+i} \mathcal{X} - \alpha_t \mathcal{X} \mathcal{W}_t \mathcal{X}\right) \right\| \right) \\
&\leq \frac{1}{|\mathcal{F}_r|} \sum_{i \in \mathcal{F}_r} W_{r,\ell}(i) \left\| \mathcal{Y}_t \mathrm{sym}(B_{t+i} - B_t) \mathcal{Y}_t^\top \right\| + 2\alpha_{\max} W_{\max} NR \sup_t \|\mathrm{sym}(\mathcal{W}_t)\| \\
&\leq \Delta_{22}^{(1)} + 2W_{\max} \left\{ \frac{3\log(16N/\epsilon)}{\underline{\delta}} \right\}^{1/2}.
\end{aligned}$$

Denote $Q_{r,t} = \frac{1}{|\mathcal{F}_r|} \sum_{i \in \mathcal{F}_r} W_{r,\ell}(i)(B_{t+i} - B_t)$, and following Pensky and Zhang (2019), we know for any $k, k' = 1, \cdots, K_C + K_R$,

$$|Q_{r,t}(k,k')| \leq \frac{L W_{\max}}{\ell!} \left(\frac{r}{T}\right)^\beta.$$

Then, for $\Delta_{22}^{(1)}$, we have

$$\begin{aligned}
\Delta_{22}^{(1)} &= \frac{1}{|\mathcal{F}_{r,j}|} \sum_{i \in \mathcal{F}_{r,j}} W_{r,\ell}^j(i) \left\| \mathcal{Y}_t \mathrm{sym}(B_{t+i} - B_t) \mathcal{Y}_t^\top \right\| \\
&\leq \max_{1 \leq j' \leq 2N} \sum_{j=1}^{2N} \left| (\mathcal{Y}_t Q_{r,t} \mathcal{Y}_t^\top)(j,j') \right| \\
&\leq \max_{k,k'} |Q_{r,t}(k,k')| \max_{1 \leq j' \leq 2N} \sum_{k=1}^{K_R+K_C} \sum_{k'=1}^{K_R+K_C} \left[\sum_{j \in \mathcal{G}_{t,k}} \mathcal{Y}_t(j,k) \right] \mathcal{Y}_t(j',k') \\
&\leq W_{\max} \frac{P_{\max} L}{\underline{\delta} \cdot \ell!} \left(\frac{r}{T}\right)^\beta, \qquad (31)
\end{aligned}$$

where the last inequality comes from the fact that the spectral norm of a matrix is dominated by the ℓ_1 norm.

Therefore, combine equation (23), (25), (26), (30), and (31), we obtain

$$\frac{\left|\mathbb{M}_t^R\right|}{N} \leq \frac{c_2(\varepsilon) K_R W_{\max}^2}{m_r^2 N \lambda_{K,\max}^2} \left\{ (6 + c_1) \frac{b}{\underline{\delta}^{1/2}} + \frac{2K_C}{\underline{\delta}} (\sqrt{2P_{\max} rs} + 2P_{\max}) + \frac{P_{\max} L}{\underline{\delta} \cdot \ell!} \left(\frac{r}{T}\right)^\beta \right\}^2,$$

where $c_2(\varepsilon) = 2^6(2+\varepsilon)^2$, $b = \{3\log(16NT/\epsilon)\}^{1/2}$ and $\lambda_{K,\max} = \max_t\{\lambda_{K,t}\}$.

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

As shown in equation (45), the population regularized graph Laplacian of dynamic DCcBM has following decomposition

$$\mathcal{S}_t = {\Psi^R_{\tau,t}}^{1/2} Z_{R,t}\Omega_t Z^\top_{C,t} {\Psi^C_{\tau,t}}^{1/2} \tag{32}$$

Then, let $Y_{R,t} = Z^\top_{R,t}\Psi^R_{\tau,t}Z_{R,t}$ and $Y_{C,t} = Z^\top_{C,t}\Psi^C_{\tau,t}Z_{C,t}$, and

$$H_{\tau,t} = Y^{1/2}_{R,t}\Omega_t Y^{1/2}_{C,t}. \tag{33}$$

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

$$\gamma_c \overset{\text{def}}{=} \min_t\{\min_{i\neq j}\|H_{\tau,t}(*,i) - H_{\tau,t}(*,j)\|\}, \tag{34}$$

and thus

$$\frac{\left|\mathbb{M}^C_t\right|}{N} \leq \frac{16(2+\varepsilon)^2}{m_c^2 N\gamma_c^2}\left\|U_t - \mathcal{U}_t\mathcal{O}_t\right\|_F^2. \tag{35}$$

where

$$m_c \overset{\text{def}}{=} \min_{i,t}\{\min\{\|\Gamma_{C,t}(i,*)\|, \|\mathit{\Gamma}_{C,t}(i,*)\|\}\}. \tag{36}$$

Now, combining equation (35) with equations (22), (25), (26), (30), and (31), we obtain

$$\sup_t \frac{\left|\mathbb{M}^C_t\right|}{N} \leq \frac{c_3(\varepsilon)K_R W^2_{\max}}{m_c^2 N\gamma_c^2\lambda^2_{K_R,\max}}\left\{(6+c_1)\frac{b}{\underline{\delta}^{1/2}} + \frac{2K_C}{\underline{\delta}}(\sqrt{2P_{\max}rs} + 2P_{\max}) + \frac{P_{\max}L}{\underline{\delta}\cdot\ell!}\left(\frac{r}{T}\right)^\beta\right\}^2$$

where $c_3(\varepsilon) = 2^7(2+\varepsilon)^2$, $b = \{3\log(16NT/\epsilon)\}^{1/2}$ and $\lambda_{K_R,\max} = \max_t\{\lambda_{K_R,t}\}$. □

7.2 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

$$\left\|\widehat{\mathcal{S}}_{t,r} - \mathcal{S}_{t,r}\right\| \leq (1-\eta)^{-2}\sup_{x,y\in\mathcal{T}}\left|x^\top(\widehat{\mathcal{S}}_{t,r} - \mathcal{S}_{t,r})y\right|, \tag{37}$$

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

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

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 **?**, 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

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

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

$$\begin{aligned}
&\Pr\left(\sup_{x,y\in\mathcal{T}}\left|\sum_{(i,j)\in\mathscr{L}(x,y)} x_i y_j(\widehat{\mathcal{S}}_{t,r}(i,j) - \mathcal{S}_{t,r}(i,j))\right| \ge \frac{c\sqrt{d}}{r}\right) \\
&\quad\le \exp\left\{-\left(\frac{3c^2}{12W_{\max}^2\|\mathcal{S}_t\|_\infty + 4cW_{\max}\|\mathcal{S}_t\|_\infty} - 2\log\left(\frac{7}{\eta}\right)\right)N\right\}. \qquad (38)
\end{aligned}$$

For the heavy pairs, we know

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

Therefore, choosing $c = W_{\max}$ in equation (38), we have

$$
\Pr\left(\sup_{x,y\in\mathcal{T}}\left|\sum_{1\le i\le j\le N} x_i y_j(\widehat{\mathcal{S}}_{t,r}(i,j)-\mathcal{S}_{t,r}(i,j))\right|\le\frac{W_{\max}\sqrt{d}}{r}\right)\ge 1-\epsilon \tag{39}
$$

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

In the end, by equation (37) and (39), we obtain, with probability $1-\epsilon$,

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

□

8 Technical Lemmas

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

$$\sup_t \|sym\,(S_t - \mathcal{S}_t)\| \leq \delta_{\max}\{3\log(16NT/\epsilon)\}^{1/2}, \tag{40}$$

with probability at least $1-\epsilon$

Proof. By triangular inequality, we have

$$\sup_t \|\mathrm{sym}\,(S_t - \mathcal{S}_t)\| \leq \sup_t \left\|\mathrm{sym}\left(\alpha_t X W_t X^\top - \alpha_t \mathcal{X}\mathcal{W}_t\mathcal{X}^\top\right)\right\| \tag{41}$$

$$+ \sup_t \left\|\mathrm{sym}\left(\mathcal{D}_{R,t}^{-1/2} A_t \mathcal{D}_{C,t}^{-1/2} - \mathcal{D}_{R,t}^{-1/2}\mathcal{A}_t\mathcal{D}_{C,t}^{-1/2}\right)\right\| \tag{42}$$

$$+ \sup_t \left\|\mathrm{sym}\left(D_{R,t}^{-1/2} A_t D_{C,t}^{-1/2} - \mathcal{D}_{R,t}^{-1/2} A_t \mathcal{D}_{C,t}^{-1/2}\right)\right\|. \tag{43}$$

For equation (41), 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

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

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

$$\sup_t \left\|\mathrm{sym}\left(\alpha_t X W_t X^\top - \alpha_t \mathcal{X}\mathcal{W}_t\mathcal{X}^\top\right)\right\| \leq c_1 a,$$

for some constant c_1.

For equation (42), by Assumption 4 and matrix Bernstein inequality, we know under assumption

$\underline{\delta} > 3\log(16NT/\epsilon)$, $a < 1$. Therefore, 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 (43), 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 $1 - \epsilon$,

$$\sup_t \|\text{sym}\,(S_t - \mathcal{S}_t)\| \leq a^2 + 3a + c_1 a \leq (4 + c_1)a = (4 + c_1)\left\{\frac{3\log(16NT/\epsilon)}{\underline{\delta}}\right\}^{1/2}. \tag{44}$$

□

Lemma 3. *Under the dynamic 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*

$i, j = 1, \cdots, N$,

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

and

$$\Gamma^*_{p,t}(i,*) = \Gamma^*_{p,t}(j,*) \Longleftrightarrow Z_{p,t}(i,*) = Z_{p,t}(j,*).$$

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 DCCBM, we have the decomposition below,

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

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 $\Omega_t = B_{L,t} + \alpha_t M_{R,t} \mathcal{W}_t M_{C,t}^\top$, then we know

$$\mathcal{S}_t = {\Psi^R_{\tau,t}}^{1/2} Z_{R,t} \Omega_t Z_{C,t}^\top {\Psi^C_{\tau,t}}^{1/2}. \tag{45}$$

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

$$\begin{aligned}
\Gamma_{p,t}^\top \Gamma_{p,t} &= U_{p,t}^\top Y_{p,t}^{-1/2} Z_{p,t}^\top {\Psi^p_{\tau,t}}^{1/2} {\Psi^p_{\tau,t}}^{1/2} Z_{p,t} Y_{p,t}^{-1/2} U_{p,t} \\
&= U_{p,t}^\top Y_{p,t}^{-1/2} Y_{p,t} Y_{p,t}^{-1/2} U_{p,t} \\
&= U_{p,t}^\top U_{p,t} = I,
\end{aligned}$$

and we have

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

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. □

IRTG 1792 Discussion Paper Series 2021

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

001 "Surrogate Models for Optimization of Dynamical Systems" by Kainat Khowaja, Mykhaylo Shcherbatyy, Wolfgang Karl Härdle, January 2021.

002 "FRM Financial Risk Meter for Emerging Markets" by Souhir Ben Amor, Michael Althof, Wolfgang Karl Härdle, February 2021.

003 "K-expectiles clustering" by Bingling Wang, Yingxing Li, Wolfgang Karl Härdle, March 2021.

004 "Understanding Smart Contracts: Hype or Hope?" by Elizaveta Zinovyev, Raphael C. G. Reule, Wolfgang Karl Härdle, March 2021.

005 "CATE Meets ML: Conditional Average Treatment Effect and Machine Learning" by Daniel Jacob, March 2021.

006 "Coins with benefits: on existence, pricing kernel and risk premium of cryptocurrencies" by Cathy Yi-Hsuan Chen, Dmitri Vinogradov, April 2021.

007 "Rodeo or Ascot: which hat to wear at the crypto race?" by Konstantin Häusler, Wolfgang Karl Härdle, April 2021.

008 "Financial Risk Meter based on Expectiles" by Rui Ren, Meng-Jou Lu, Yingxing Li, Wolfgang Karl Härdle, April 2021.

009 "Von den Mühen der Ebenen und der Berge in den Wissenschaften" by Annette Vogt, April 2021.

010 "A Data-driven Explainable Case-based Reasoning Approach for Financial Risk Detection" by Wei Li, Florentina Paraschiv, Georgios Sermpinis, July 2021.

011 "Valuing cryptocurrencies: Three easy pieces" by Michael C. Burda, July 2021.

012 "Correlation scenarios and correlation stress testing" by Natalie Packham, Fabian Woebbeking, July 2021.

013 "Penalized Weigted Competing Risks Models Based on Quantile Regression" by Erqian Li, Wolfgang Karl Härdle, Xiaowen Dai, Maozai Tian, July 2021.

014 "Indices on Cryptocurrencies: an Evaluation" by Konstantin Häusler, Hongyu Xia, August 2021.

015 "High-dimensional Statistical Learning Techniques for Time-varying Limit Order Book Networks" by Shi Chen, Wolfgang Karl Härdle, Melanie Schienle, August 2021.

016 "A Time-Varying Network for Cryptocurrencies" by Li Guo, Wolfgang Karl Härdle, Yubo Tao, August 2021.

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.