Eur. Phys. J. B (2022) 95:71
https://doi.org/10.1140/epjb/s10051-022-00332-x

THE EUROPEAN PHYSICAL JOURNAL B

Regular Article - Statistical and Nonlinear Physics

Instability of networks: effects of sampling frequency and extreme fluctuations in financial data

Jalshayin Bhachech[1,a], Arnab Chakrabarti[2,b], Taisei Kaizoji[3,c], and Anindya S. Chakrabarti[4,d]

[1] Indian Institute of Management Ahmedabad, Ahmedabad, Gujarat 380015, India
[2] MCFME and CDSA, Indian Institute of Management Ahmedabad, Ahmedabad, Gujarat 380015, India
[3] Division of Arts and Sciences, International Christian University, Mitaka, Tokyo 181-8585, Japan
[4] Economics Area, MCFME and CDSA, Indian Institute of Management Ahmedabad, Ahmedabad, Gujarat 380015, India

Received 19 January 2022 / Accepted 11 April 2022 / Published online 25 April 2022

Abstract. What determines the stability of networks inferred from dynamical behavior of a system? Internal and external shocks in a system can destabilize the topological properties of comovement networks. In real-world data, this creates a trade-off between identification of turbulent periods and the problem of high dimensionality. Longer time-series reduces the problem of high dimensionality, but suffers from mixing turbulent and non-turbulent periods. Shorter time-series can identify periods of turbulence more accurately, but introduces the problem of high dimensionality, so that the underlying linkages cannot be estimated precisely. In this paper, we exploit high-frequency multivariate financial data to analyze the origin of instability in the inferred networks during periods free from external disturbances. We show that the topological properties captured via centrality ordering is highly unstable even during such non-turbulent periods. Simulation results with multivariate Gaussian and fat-tailed stochastic process calibrated to financial data show that both sampling frequencies and the presence of outliers cause instability in the inferred network. We conclude that instability of network properties do not necessarily indicate systemic instability.

1 Introduction

Networks provide a granular view of complex systems characterized by interactions across constituent components [34]. Such interactions are often dynamic in nature leading to emergence of intricate comovement structures [41]. Empirically, periods of external and internal perturbations are often associated with changes in the network topology inferred from the comovement structures. However, there is a trade-off between identification of turbulent periods and the problem of high dimensionality for the network inference problem. Networks constructed from longer time-series circumvent the problem of high dimensionality, but suffers from potential mixing of turbulent and non-turbulent periods. On the other hand, networks inferred from shorter time-series can reduce the mixing of turbulent and non-turbulent periods more accurately, but they suffer from the problem of high dimensionality, such that the underlying linkages cannot be estimated accurately.

In this paper, we show that the observed instability in network topology can go beyond the external sources of turbulence [5] and may, in fact, stem from the data-generating process itself even during non-turbulent periods. We utilize financial markets, a standard example of complex systems [33], as our ground of exploration. This is useful for two reasons. One, reliable high-frequency financial data are easily available. Two, it is relatively easy to differentiate between periods of financial turbulence and calmness. Previous studies have noted that periods of turbulence may change the topological characteristics of financial networks. We show that even during periods of relative calm, topological characteristics of financial networks can display instability. Our empirical and simulation analysis shows that there are at least two major sources of the instability. First, sampling frequency influences the estimated comovement structure and, hence, potentially leads to variation in the network properties. Second, the presence of extreme fluctuations also influences the estimated comovement structure and substantially impacts the relative centrality of the nodes in the network. Since financial asset returns often exhibit extreme fluctuations even during times of relative calm (see [32] in this context which develops an idea about quasi-stationarity), our results indicate that network estimation on asset returns would be susceptible to such behavior. A corollary of these two results is that instability of the network topology does not necessarily imply periods of turbulence.

[a] e-mail: jalshayin@iima.ac.in
[b] e-mail: arnab_c@zohomail.in
[c] e-mail: kaizoji@icu.ac.jp
[d] e-mail: anindyac@iima.ac.in (corresponding author)

The literature on asset price networks has focused on two major aspects. First, the time-series properties of assets and inference of networks arising out of them. Second, the literature has attempted to find connections between the evolution of the networks and external economic events like market crash for example. The long-term analysis of the collective dynamics of the stocks can offer significant insight into the recurrent observable patterns in the market. Using 90 years of stock-level NYSE data, Ref. [21] shows that systemic crisis periods are associated with emergence of frustration in the interaction networks. Ref. [19] propose a methodological framework to quantify and evaluate the interdependencies in the global financial network. The methodology aims to function as a "financial seismograph" of the network that is able to detect early sign of a global financial crisis. There is a stream of work that focuses on how the composition of the network changes over time. Ref. [4] for example suggests a consistent decrease in centrality of the financial sector in the network over the previous 2 decades. Inter-linkages have also been studied through a time-series perspective. Reference [18] uses meta-correlations to study the relation between mean correlation of the entire portfolio and the changes in the index return over time. By analyzing the time-series of daily returns, Ref. [29] shows that the average correlation between returns has predictive components for quarterly excess returns. The analysis on stationary data has also been extended to non-stationary data. Reference [28] proposes detrended cross correlation analysis to analyze non-stationary time series, which complements the existing literature.

All the works mentioned above have leveraged the daily stock price data. In the context of financial networks, relatively fewer studies have dealt with intraday high-frequency data. Our work has a direct connection to Ref. [8] which shows that the sampling of stock return series affects the hierarchical organization of the networks. In a similar vein, Ref. [9] studies the emergence of a structure in the correlation matrix of asset returns for different time horizons. Reference [35] showed that sampling time horizon relates to the hierarchical network structure of stocks. More recently, Ref. [26] show that technology and innovation influence the nature of the financial network by changing the mode of interactions between traders. In this work, we do not directly make inference about traders and focus instead on the emerging linkages across traded assets via their time-series behavior. Finally, we note that there is a literature on econometric approach to graph construction and analyzing shock propagation [14,20]. In the context of this paper, we will not directly relate to that approach. However, our work shows that the shock propagation mechanism may be affected by sampling frequencies and presence of extreme fluctuations.

We characterize the intraday networks arising out of high-frequency data on returns of assets. In particular, we consider the 50 constituent firms in the NIFTY-50 index from the Indian financial market. This index reflects market determined asset prices of the registered firms. The data are available at intervals of 10 s. From the baseline data, we construct returns at the intervals of 30 s, 1, 2, 5, 10, 30 min and finally, 1 h. We note that while higher frequency data would potentially exhibit finer dynamic patterns, market liquidity may affect such data. Specifically, some stocks may not register trades at every second indicating low liquidity. However, that creates a problem of missing data. Therefore, we take 10 s as an optimal level of sampling the data where the properties of high-frequency returns will persist and the market would be liquid enough to register trades. From the realized returns, we construct covariance matrices, and then correspondingly, we construct the networks.

Our main proposition is that even during times of relative calm in the market, the network topology may not remain stable. To capture the stability of the resulting matrix, we utilize PageRank measure as a summary statistics of the network topology and we characterize the volatility in evolution of the PageRank across days or within day across sampling frequencies, and to characterize the degree of instability. While PageRank is not the only measure for capturing the properties of a network, it has the benefit of capturing the spectral properties of the network, it is a global measure, and finally, it is readily comparable across networks of the same size. Additionally, one can consider clusters as well in the form of minimum spanning tree for example. However, they are known to be even less stable across networks. Therefore, the degree of instability characterized by the PageRank would be lower bound for instability.

We empirically observe that for networks constructed both across days with the same sampling frequencies and across sampling frequencies within days, the correlation goes down sharply. We conjecture that the decay in network stability can arise out of three sources, viz., sample size, extreme fluctuations, and finally time-dependent properties of return in higher moments. First, we employ a null model of independently and identically distributed multivariate Gaussian noise to capture the portion of the decay that can be attributed to simply changes in sampling frequency. In the next step, we employ a independent multivariate t-distribution to capture the additional effect arising out of extreme fluctuations. Here, we exploit the fact that t-distribution interpolates between normal distribution in one limit and displays fat tails in the other limit. However, while t-distribution captures the extreme fluctuations, it falls short of modeling the time-varying properties of returns. The residual of the decay was attribute to the third possible source of time-dependent properties.

Our work represents a cautionary note on the statistical inference on asset price networks. One approach to the networks literature treats the realized networks as given and builds toolkits to analyze them. However, if we see that realized network as one of the possible outcomes arising from an underlying data-generating process, then each realized network needs to be treated as a sample, and therefore, the resulting properties are contingent on the realization of a specific sample. While this point has been recognized and utilized

in developing network filters through statistical validation, our work essentially shows that network properties are affected by sampling frequencies as well as presence of outliers among others. Thus, on a complementary note, changes in the network properties would not generally imply existence of external shocks, i.e., such inferences would potentially indicate *false positives*. However, our work is fully consistent with the idea that external shocks can induce changes in the network properties.

2 Comovement networks from high-frequency data

Let us first fix notations that will be used throughout the paper.

1. Dimension of the inter-day data: N (days) and p (stocks).
2. Frequency of intraday data: f.
3. Integrated covariance matrix for the tth day: Σ_t.
4. Realized covariance matrix on the tth day based on data of frequency f: $S^{RCV}_{t,f}$.
5. Average subsample covariance matrix on the t-th day day based on data of frequency f: $\bar{S}_{t,f}$
6. Page rank centrality of a network Γ: $Z(\Gamma)$.
7. Intraday network based on RCV matrix and AVS matrix from day t as Γ^{RCV} and $\Gamma^{\bar{S}}$.
8. The sample size of intraday data with a particular frequency f: $n(f)$.

In this section, we describe the mathematical model of intraday price movement and define the daily network.

2.1 Intraday price dynamics

Suppose, we have p stocks and the log-price of ith stock at time τ is denoted by X^i_τ. Due to several obvious merits in context to asset pricing, the stock price process is modeled as a continuous stochastic process [40]. In practice, financial data are recorded at several different frequencies. In many studies, involving financial network analysis, the data consist of daily closing prices over a number of days. This is because the daily closing prices are most easily available. In such studies, the daily network is constructed from the covariance matrix of the return vector.

Let us denote the daily return vector by R_t. Then, the corresponding covariance matrix can be written as $cov(R_t) = \Sigma$, where $R_t = (X^1_t - X^1_{t-1}, ..., X^p_t - X^p_{t-1})$. The typical assumption for such studies is that the underlying true network does not change during the course of these N days.

However, given intraday data, we can infer the daily covariance matrix Σ_t which can vary across days [10], using the following widely used parametric model:

$$\mathrm{d}\mathbf{X}_{t+\tau} = \mu_{t+\tau}d\tau + \sigma_{t+\tau}\mathrm{d}\mathbf{W}_{t+\tau};\ \tau \in [0,1], \quad (1)$$

where $\mathbf{X}_\tau = (X^1_\tau, ..., X^p_\tau)^T$, $\boldsymbol{\mu}_\tau = (\mu^1_\tau, ..., \mu^p_\tau)^T$ is a p-dimensional *drift process*; σ_τ is a $p \times p$ matrix, called instantaneous *covolatility process* and $\boldsymbol{W}_\tau$ is a p-dimensional Brownian motion, for any τ. This model is useful in context of intraday data precisely, because it takes into account time-varying volatility over the time-span of a day.

According to this model, the daily return distribution has mean $\mu_t = \int_0^1 \mu_{t+\tau-1}\mathrm{d}\tau$ and covariance matrix $\Sigma_t = \int_0^1 \sigma_{t+\tau-1}\sigma^T_{t+\tau-1}\mathrm{d}\tau$. As Σ_t is obtained by integrating the instantaneous $\sigma_{t+\tau-1}\sigma^T_{t+\tau-1}$ over $0 < \tau \leq 1$, it is called the *integrated covariance matrix*. Thus, the daily network can be obtained from the integrated covariance matrix. We denote the network obtained from a given covariance matrix Σ as $\Gamma(\Sigma)$—the time-varying analogue being $\Gamma(\Sigma_t)$. Each node of this network corresponds to a stock and a nonzero value of the (i, j)-th entry of Σ indicates the existence of an edge between the ith and jth nodes. This results in a undirected weighted network where the weight of an edge is equal to the distance (see Eq. 5) of the corresponding integrated covariance.

However, from the model (Eq. 1), it is evident that the integrated covariance matrix is our parameter of interest and not directly observed. We need to estimate $\Gamma(\Sigma_t)$ by first estimating Σ_t.

2.2 Estimator of integrated covariance matrix Σ_t

A widely used estimator of the integrated covariance matrix is the so-called *Realized CoVariance* (RCV) matrix. For a given day t, if we observe the processes $X^j_{t+\tau}$ synchronously, at time points (possibly random) $\tau_0, \tau_1, \tau_2, \tau_3, ..., \tau_n$, then the RCV matrix is defined as

$$S^{RCV} = \sum_{l=1}^{n} \Delta X_l (\Delta X_l)^T, \quad (2)$$

where for $l = 1, 2, ..., n$ (we are omitting t to keep the notation simple), we have

$$\Delta X_l = \begin{Bmatrix} \Delta X^1_l \\ \Delta X^2_l \\ . \\ . \\ \Delta X^p_l \end{Bmatrix} = \begin{Bmatrix} X^1_{\tau_l} - X^1_{\tau_{l-1}} \\ X^2_{\tau_l} - X^2_{\tau_{l-1}} \\ . \\ . \\ X^p_{\tau_l} - X^p_{\tau_{l-1}} \end{Bmatrix}.$$

Under the model given by Eq. 1, the realized covariance matrix converges to the integrated covariance matrix [17]. We note that in case of realized covariance, we are effectively summing the intraday covariances and not averaging it out over the span of the trading day. Therefore, the resulting estimator may display bias if the price increments captured in Eq. 2 involve even

small amount of noise. In the following, we correct this bias by filtering out such noise.

2.3 Filtering microstructure noise from S_p^{RCV}

In observed data, intraday prices are often contaminated by market microstructure noise. Instead of observing the true (log) price $X_{t+\tau}$, we may observe the price mixed with noise. The following additive model represents the observed data with noise:

$$Y_{t+\tau_l} = X_{t+\tau_l} + \epsilon_{t+\tau_l}, \quad l = 1, 2, 3, ..., n, \tag{3}$$

where $Y_{\tau_l} = (Y_l^1, ..., Y_l^p)$ is the observed log-price, and $\epsilon_l = (\epsilon_l^1, ..., \epsilon_l^p)$ is an independent and identically distributed noise vector with mean 0 and covariance matrix Σ_ϵ. The noise is also independent of X_{τ_l}.

To filter out this noise, we use *Two Scales Realized Volatility* (TSRV) estimator [40] which modifies the realized variances. We call the resulting covariance matrix *average subsample covariance estimator* (AVS) and denote it by $\bar{S}$.

In this approach, the realized variances are updated in three consecutive steps. In the first step, the lower frequency returns are constructed by sampling from high-frequency data at an optimal lower frequency and the realized variances are calculated. Then, the same is calculated across a number of grids and the resulting variances are averaged. In the last step, the bias is corrected and the final estimator is obtained. Mathematically

$$TSRV = RV_T^{avg} - \frac{\bar{n}}{n} RV_T^{all}, \tag{4}$$

where RCV_T^{all} is the realized variance computed with all the data and RCV_T^{avg} is the average over calculated realized variances from K subsamples of the original grid of observation times; $\bar{n} = n/K$. One can also use other classes of estimators like *pre-averaging estimator* [12], *realized kernel estimator* [6], *quasi-maximum-likelihood estimator* [2], and *two time-scale covariance estimator* [39]. Many of these estimators have good statistical properties and often display very similar asymptotic behavior. A comparative discussion on the realized kernel estimator and the TSRV estimator can be found in [37]. For our purpose, we adopt TSRV estimator as it is easy to implement and retains useful asymptotic properties.

2.4 Network construction

For notational convenience, we are omitting t (the identifier of the trading day) in this paragraph. We convert the covariance matrix $\bar{S}$ to the corresponding correlation matrix. A network is often constructed from a correlation matrix through an intermediate step of transformation to distances. As correlation coefficient cannot be considered as a metric (correlations can be negative, whereas a distance measure is always nonnegative), the distance between two stocks is defined as follows (see [23]):

$$dist_{i,j} = \sqrt{2(1 - r_{i,j})}, \tag{5}$$

where $r_{i,j}$ is the (i, j)th element of the correlation matrix. The edges of the network are having weights equal to the corresponding distance. We construct the daily network $\Gamma(S^{RCV})$ and $\Gamma(\bar{S})$ from the realized covariance matrix (S^{RCV}) and average subsample covariance matrix $(\bar{S})$, respectively. Both of them are weighted networks.

2.5 Network characteristics and stability

In this paper, we study the stability of an observed network chiefly in the framework of two different but complementary methods. First, we capture it through the variability of the estimated edge weights of the network. In other words, we look at the stability or instability of the elements of the adjacency matrix by looking at the distribution of differences of the edge weights across networks. In the second approach, instead of looking at the edge weights, we concentrate on global network characteristics measured via centrality and estimates the stability of the centrality measures to capture the stability of the network itself. The first approach is statistical, while the second approach is network-theoretic in nature.

To characterize the topology of the network, we focus on the PageRank centrality. There are two reasons for that. One, PageRank centrality is global in nature and accounts for the complete adjacency matrix of the network. Two, it is a measure to the influence of nodes in the financial networks which directly relates to dynamics attributed to the market mode through the dominant eigenvector.

Let us define the measure as follows. PageRank $(Z_{i,t})$ of the i-th stock at day t is defined as the solution of the following recursive equation:

$$Z_{it} = \frac{1 - \alpha}{p} + \alpha \sum_j E_{ijt} Z_{jt}, \tag{6}$$

where α is a tunable damping factor, p is the number of nodes, and E_{ijt} is the normalized adjacency matrix corresponding to day t. The linkages E_{ijt} are constructed from the time series of returns of assets i and j in day t. It may also be noted that by construction, the PageRank is a generalization of *eigenvector centrality*, a widely used measure of the importance of the nodes based on their overall connectivity.

Recall that we have denoted the realized covariance matrix calculated from the data with frequency f on day t as $S_{f,t}^{RCV}$ and the corresponding network as $\Gamma_{f,t} = \Gamma(S_{f,t}^{RCV})$. To quantify the stability of the constructed network with frequency (or sample size), we define

$$\rho_t^{RCV}(f_i, f_j) = Cor(Z(\Gamma_{f_i,t}), Z(\Gamma_{f_j,t})), \tag{7}$$

where the PageRank centrality of $\Gamma_{f,t}$ is denoted as $Z(\Gamma_{f,t})$. A high value of this correlation indicates high level of network stability (similarly, the interpretation holds for low values as well).

In the simulation studies (Sects. 5.1 and 5.2), where we know the true underlying network, the following quantity is also calculated:

$$\bar{\rho}_t^{RCV}(f_i) = Cor(Z(\Gamma_{TRUE}), Z(\Gamma_{f_j,t})). \tag{8}$$

The average correlation of PageRanks across days but for a fixed frequency and same underlying data-generating process is captured by

$$\gamma_f^{RCV}(t_i, t_j) = Cor(Z(\Gamma_{f,t_i}), Z(\Gamma_{f,t_j})), \tag{9}$$

where t_i and t_j denote the ith and jth days, respectively. This measures how the constructed network changes over days. Note that these changes can be attributed to random fluctuation. The average fluctuation over N days is measured by $\bar{\gamma}_f^{RCV}(t_i, t_j) = \frac{2}{N(N-1)} \sum_{t_i \neq t_j} \gamma_f^{RCV}(t_i, t_j)$.

As in case of simulated data analysis, following the Itô process in Eq. 1, there is no microstructure noise, we only use RCV matrix for simulated data analysis. However, for real data analysis, we calculate the same quantities described above also for $\bar{S}$.

Additionally, there is a third approach we follow. PageRank is a global measure based on the spectrum of the network. We also compute the centrality arising from the first-order neighborhood of each nodes by constructing the node-wise degree centrality of the weighted network.

3 Characterization of instability

3.1 Construction of the empirical network

For our baseline analysis, we collect data from Bloomberg Terminal at intraday frequency of 10sec from 22-12-2020 to 03-02-2021 (30 trading days) of equities listed in NIFTY-50. Later on, we extend our analysis till 22-11-2021 (covering one full year in the whole sample) to check for robustness of the findings. For each intraday frequency, timing of observations recorded is from 9:30 AM to 3:30 PM for synchronous data (i.e., throughout the period of active trading when the market remains open). We create daily synchronous snapshots of price and log returns for 30 s, 1, 2, 5, 10, 30 min and 1 h frequencies for analysis arising out of the data with the base frequency of 10 s, i.e., each of these dataset is created from 10 s data by aggregating them over the targeted frequencies. In case there are missing values, we fill them by the previous ticks' observations. At the outset, we note that this period was relatively stable in the Indian market. The effects of the panic due to COVID-19 had subsided within a few months of march, 2020.

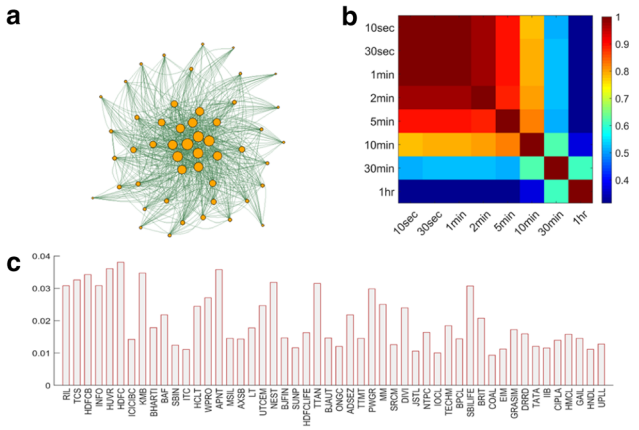

Fig. 1 Properties of financial networks. We have utilized intraday data for all of the NIFTY-50 stocks listed in Table 4. Panel (a): Depiction of the correlation network created from return data on 22.12.2020 sampled at 10 s interval. The network has been thresholded for visual clarity. Panel (b): Heatmap of the correlation matrix for the PageRank vectors across the inferred networks constructed with different sampling frequencies, viz. 10, 30 s, 1, 2, 5, 10, 30 min and 1 hr. Panel (c): Distribution of normalized PageRank across the stocks shown in panel (a)

Figure 1 depicts an instance of the inferred network along with its properties. In panel (a), we show a network constructed from 10 s return data of 22-12-2020. The corresponding PageRanks across stocks are shown in panel (c). Panel (b) shows the heatmap of the correlation matrix (Table 6) with the values of $\rho_t^{\bar{S}(f_i,f_j)}$ for different values of the frequencies f_i and f_j.

3.2 Instability of rank-ordering

In Table 6, we present the average $\rho_t^{\bar{S}}(f_i, f_j)$ over baseline period along with the standard deviations. The corresponding heatmap of the correlation table is presented in panel (c) of Fig. 1. It is evident from the table (and the heatmap) that the larger is the difference between the two frequencies, lower the correlation between PageRanks.

We can see that the correlations near the diagonal line are higher in magnitude.

Table 1 shows the $\gamma_f^{RCV}(t_i, t_j)$ and $\gamma_f^{\bar{S}}(t_i, t_j)$ for different frequencies along with standard deviations, averaged over all pairs of days t_i and t_j in the sample. The values obtained using $\bar{S}$ (average subsample estimator of the Integrated Covariance) are much lower than what we get using RCV. We see that the correlation values decrease as we choose lower frequency data which also suggest the decreasing stability of the realized network structure. This decreasing pattern is a consequence of the sample size (see Sect. 5.1). However, the question is the following: are the values obtained in Table 1 only a consequence of sample size or a result of a more convoluted effect of several variables of which sample size is

Table 1 Correlation between the PageRanks across days averaged over all possible pairs of days in the sample, corresponding to different frequencies (f). The average correlation decreases for lower frequencies. The correlation values for RCV are higher than the noise-corrected covariance matrix $\bar{S}$

Frequency	Average correlation (RCV)	Average correlation ($\bar{S}$)
f=10sec	0.64 (± 0.13)	0.38 (± 0.16)
f=30sec	0.56 (± 0.14)	0.37 (± 0.16)
f=1min	0.53 (± 0.14)	0.37 (± 0.16)
f=2min	0.45 (± 0.15)	0.37 (± 0.16)
f=5min	0.30 (± 0.16)	0.34 (± 0.15)
f=10min	0.20 (± 0.19)	0.20 (± 0.17)
f=30min	0.07 (± 0.16)	0.09 (± 0.16)
f=1hr	0.02 (± 0.15)	0.02 (± 0.16)

one of the most important one? We aim to find plausible answer to this question in the next sections.

4 Robustness

To check the robustness of the reported tables, we perform identical analysis on additional dataset consisting the stock prices of 138 trading days ranging from 03-05-2021 to 22-11-2021. The set of stocks considered is identical to that of the baseline data. We find that the outcomes are perfectly synced with the earlier results obtained from the baseline dataset. The results on additional dataset are presented in the appendix. In Table 9, we present the average $\rho_t^{\bar{S}}(f_i, f_j)$ along with the standard deviations calculated from the additional dataset. The values are similar to what we reported in Table 6. Similarly, the average $\gamma_f^{RCV}(t_i, t_j)$ on additional data are reported in Table 10 along with standard deviations. The values not only exhibit the same pattern as in Table 1, but also they are sufficiently close. This ensures that the reported results are robust. Additionally, we implement the centrality measure based on rank-order of first-order connectivity and report the corresponding correlations in Table 11. Similar to the estimates with PageRank in Table 6, the correlations with rank-ordering of degree centrality indicate a monotonic decay. To summarize all the above findings, sampling frequency matters for the stability of the matrix.

5 Instability of networks: simulations

We explore the sources of the empirically documented decay in the correlations across networks. We construct a hierarchical model to explain the decay. Three properties of financial time-series network are as follows: sampling frequency, extreme fluctuations, and time dependence of higher order moments. In the baseline null model, we simulate independent and identically distributed multivariate Gaussian noise with parameters calibrated to the data to explore how sampling frequency affects network stability. We extend the model to simulate time-varying covariance matrices, as well. Finally, we extend the model to incorporate extreme fluctuations by simulating independent and identically distributed multivariate t-distributed stochastic variables with parameters calibrated to the data. We quantify the explanatory power arising out of extreme fluctuations.

5.1 Effects of sampling frequency

In this section through simulated data analysis, we attempt to see the impact of sampling frequency (or sample size) on the stability of network $\Gamma(S^{RCV})$ obtained from the realized covariance matrix. We assume the model described by Eq. 1 which implies that the return distribution is Gaussian. The parameters of the return distribution are chosen to be the estimated mean and covariance of the real data (described in Sect. 3).

5.1.1 Gaussian distribution with constant covariance matrix

Equation 1 implies that the underlying return distribution is Gaussian. Therefore, we generate the return dataset from multivariate Gaussian distribution. The mean and the covariance matrix of this 50-dimensional Gaussian distribution are taken to be the sample mean and sample covariance obtained from the real 10 s data. We construct return data with the same number of days as in the baseline dataset, with 2550 observations per day, from this distribution. This constitutes the simulated 10sec data over a period of 30 consecutive days. From this dataset, we can also create synchronous data of other lower frequencies. Thus we perform the analysis with 10 s, 30 s, 1, 2, 5, 10, 30 min and 1 h frequencies.

5.1.2 Network characteristics

In Fig. 2, we see that the average correlation $(1/N \times \sum_t \bar{\rho}_t^{RCV})$ of page-ranks between the true network and the sample network, obtained from the simulated data, is decreasing with sample size (n_f). Note that the sample size is a function of the corresponding frequency f.

In the baseline sample, the number of daily returns is 30 (corresponding to 30 days) which is less than the number of return data corresponding to 10 min intraday data (36; spanning over 6 h in a given trading day) and more than that of 30min data (12; spanning over 6 h in a given trading day). If we calculate simple correlation using daily data then the $\bar{\rho}_t^{RCV}(daily)$ turns out to be 0.328 which is less than the average correlation (0.408) for 10 min data and more than the average correlation (0.244) for 30 min data (see Table 8 in Appendix). This also gives us an idea about the instability of a network

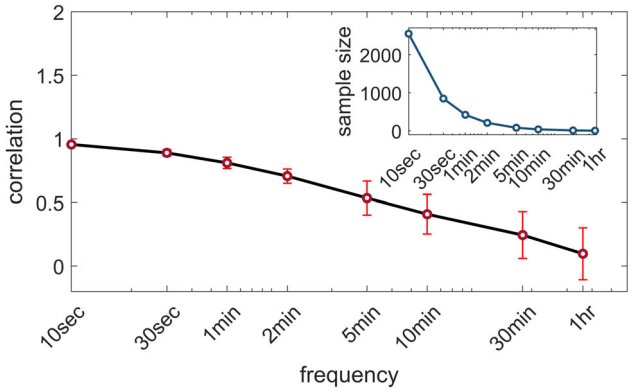

Fig. 2 Instability of simulated networks induced by sampling frequencies. *Main panel*: The average correlations of PageRanks between the network with known covariance matrix and realized sample networks. We simulate sample networks by Eq. 1 calibrating the covariance matrix and we generate simulated data of same size as in baseline dataset at different frequencies. Along with the point estimates of the average correlations between PageRanks, we show the corresponding standard deviation in red bars. *Inset*: The sample size corresponding to each frequency is shown. Sample size decreases for lower frequencies, leading to increase in the increase in the standard deviation in the left panel

constructed from daily financial data. Even under the simple model as in Eq. 1, if we want to retain 90% of the true PageRank, then we should construct the network from at least 2000 daily return data, which is approximately 6 years of time. The assumption that the underlying network remains unchanged for such a long time period would be questionable.

In Table 7, we present the average $\rho_t^{RCV}(f_i, f_j)$ over 30 days along with the standard deviations. It is evident from the table that the larger is the difference between the two frequencies lower the correlation between the pageranks. We can see that the correlations near the diagonal line are higher in magnitude. It is simply a consequence of the sample size. The difference of the sample sizes of, say, 10 s-data and 1 h-data is so large the networks constructed from these two datasets have little in common. On the other hand, network constructed from 10 min data and 30 min data are relatively alike, even though it does not retain much of the key features of the true underlying network. If we compare Table 7 (from simulated data) and Table 6 (from real data), we see that the correlations that are far from the diagonal line are smaller in Table 7. For example, $\rho_t^{RCV}(10\text{sec}, 1\text{hr}) = 0.19$ for 7, whereas 0.32 for 6.

For the same simulated data, Table 2 shows the average correlation of PageRanks (Eq. 9) across days but for a fixed frequency and same underlying data generating process. This captures how the constructed network changes over days due to random fluctuations.

The $\bar{\gamma}_f^{RCV}(t_i, t_j)$ for each frequency, presented in Table 2, shows that when the sample size is low, the constructed network is highly unstable and does not reflect on the true underlying network. As a consequence, inference based on the network from very low frequency data (more precisely from low sample size data) would be highly misleading. For the same reason, network constructed from daily data would be highly unstable unless we consider very large number of days. A noteworthy feature is that the values of the correlations indicating stability of the inferred networks are substantially higher than those found in the real data (Table 1). We will show below that this discrepancy can be explained to a large extent by considering the fat-tailed nature of the return distribution (Sect. 5.2).

Table 2 Stability of networks generated from simulated multivariate Gaussian processes (identical to the simulation in Table 7). Similarity of the networks constructed with different frequencies over the span of the data. Networks constructed from higher frequency data are relatively more stable across days. The networks constructed from lower frequency data on the other hand shows much less stability. This simulation exercise assumed identical stochastic data-generating process, and therefore, the decay in network stability can be attributed to only sampling frequencies

Frequency	Average correlation	Sample size
f=10 s	0.96 (± 0.01)	2550
f=30 s	0.88 (± 0.02)	849
f=1 min	0.78 (± 0.04)	424
f=2 min	0.63 (± 0.08)	212
f=5 min	0.37 (± 0.13)	84
f=10 min	0.22 (± 0.14)	42
f=30 min	0.07 (± 0.14)	14
f=1 h	0.03 (± 0.13)	7

5.1.3 Instability of the edge weights of the network

The weight of an edge of the network is given by the absolute value of the entries of the covariance matrix (Σ). Here, we want to see how close each edge weight is approximated for different frequencies. Figure 3 illustrates that the errors of estimating the true edges increase in magnitude when we move from high- to low-dimensional data. The box plot indicates that not only the variance of error distribution increases as we decrease n_f, but also the average error is increasing in magnitude. The error does contribute to the instability in the estimation of network characteristics like PageRank and the associated risk-measures.

5.1.4 Gaussian distribution with time-varying covariance matrix

In the previous section, we have considered constant covolatility process σ (see Eq. 1). However, it is well known that the covolatility process is typically not constant. The market is observed to be more volatile in the beginning and the end of the trading day, resulting in a U-shaped covolatility process [1,3]. Therefore, we need to consider time-varying covolatility process. In other

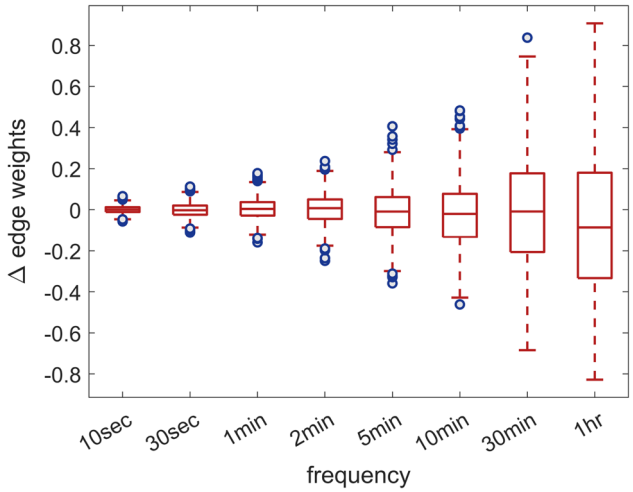

Fig. 3 Box-plot displaying error distribution across edges of simulated networks for known covariance matrices, across four different frequencies (10 s, 30 s, 1 min, and 2 min, respectively). The simulation data are used from the same dataset used in Table 7. The error for a given edge is defined as the difference between simulated weight and the underlying true weight. The variance of the error distribution increases for lower frequencies. Additionally, the median shifts further from zero as the frequency decreases. Since the underlying process is Gaussian, the monotonic variation in the dispersion of the error distribution can be attributed to changes in the sampling frequency

words, σ_τ in Eq. 1 has to be dependent on τ in practice. Here, following Refs. [42] and [38], we assume that $\sigma_\tau = \gamma_\tau \Lambda$, where Λ is a non-time-varying $p \times p$ matrix with its transpose defined as Λ^T, and γ_τ is a càdlàg function from $[0,1]$ to $\mathbb{R}$. Therefore, the corresponding integrated covariance matrix for the tth day is

$$ICV_t = \int_0^1 \gamma^2_{t+\tau} d\tau \bar{\Sigma}, \text{ where } \bar{\Sigma} = \Lambda\Lambda^T. \tag{10}$$

We consider γ_τ to be a step function taking three distinct values

$$\gamma_t = \begin{cases} c_1 & \text{where } \tau \in [0, \tau_1) \\ c_2 & \text{where } \tau \in [\tau_1, \tau_2) \\ c_3 & \text{where } \tau \in [\tau_2, \tau_3). \end{cases} \tag{11}$$

The first interval $[0, \tau_1)$ stands for the volatile period in the beginning of a day. This has been taken to be the first 1 h of the whole trading period in a given day (9:30 AM to 3:30 PM). $[\tau_1, \tau_2)$ and $[\tau_2, \tau_3)$ stand for less-volatile period in the middle of a day and the volatile period at the end of the day, respectively. We divide the timeline within a day into these three parts and denote the corresponding number of observations as $n_1 = 360$, $n_2 = 1440$, and $n_3 = 360$, respectively, to capture return evolution at 10 s intervals (e.g., n_1 being equal to 360 implies that we are modeling return evolution over $360 \times 10 = 3600$ s, i.e., 1 h) to cover the whole trading period in a day.

We simulate n_1 return data for the time period $[0, \tau_1)$ from a Gaussian distribution $N_p(0, c_1\hat{\Sigma})$, where $\hat{\Sigma}$ refers to the estimated covariance matrix, and $c_1 > 1$. Similarly n_2 and n_3 observations are obtained from $N(0, c_2\hat{\Sigma})$ and $N(0, c_3\hat{\Sigma})$, respectively. In this exercise, we normalize c_2 to 1 to capture the idea that during the middle of the day, the market remains relatively less volatile. The values of c_1 and c_3 are obtained from real data and they are expected to be more than 1. To estimate these parameters, we partition each day's data in the baseline sample, into three parts with n_1, n_2, and n_3. The average ratio of the traces of covariance matrices in first and second periods of a day gives us the c_1. Similarly, c_3 is obtained as the average ratio of the traces of covariance matrices in third and second periods. The ICV of such process is $n_1 c_1 \hat{\Sigma} + n_2 c_2 \hat{\Sigma} + n_3 c_3 \hat{\Sigma}$. The correlations of the PageRanks between the RCV and ICV matrix are presented in Table 3. We see that the pattern of decay of the correlation with sample size is similar to that of Table 2, indicating that the time-varying volatility does not seem to have a prominent effect.

Table 3 Stability of networks generated from simulated multivariate Gaussian processes with time-varying volatility (using stochastic processes described by Eqs. 10 and 11). Correlation of PageRanks between the true network (from $n_1 c_1 \hat{\Sigma} + n_2 c_2 \hat{\Sigma} + n_3 c_3 \hat{\Sigma}$) and the realized network constructed from RCV matrix. Following the process described in the text, we estimate the values of c_1, c_2 and c_3 as 3.33, 1 and 1.2 respectively

Frequency	Average correlation	Sample size
f=10sec	0.94 (±0.01)	2160
f=30sec	0.86 (±0.03)	720
f=1min	0.77 (±0.06)	360
f=2min	0.66 (±0.06)	182
f=5min	0.47 (±0.15)	72
f=10min	0.39 (±0.14)	36
f=30min	0.19 (±0.14)	12
f=1hr	0.15 (±0.19)	6

5.2 Effect of extreme fluctuations

Till now, Eq. 1 formed the core of our estimation, using a multivariate model of Brownian motion. This is consistent with the stylized fact of *aggregational Gaussianity* in stock returns [13], although in very high frequency data, deviations from Gaussianity has been well documented and heavier tailed distributions are found to be more appropriate. Ref. [27] studied the distribution of log-returns of a variety of diversified world stock indices using the maximum-likelihood ratio test on the large class of generalized hyperbolic distributions. They found that the Student's t-distribution with about 4 degrees of freedom is a good candidate to model

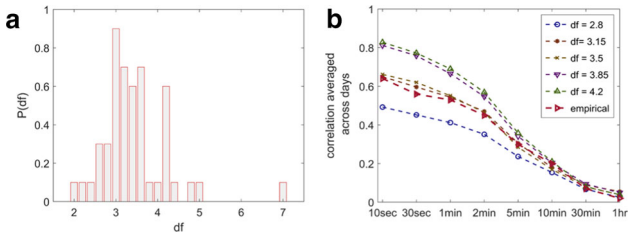

Fig. 4 Instability of networks induced by extreme fluctuations. *Left panel*: Histogram of degrees of freedom (df) of t-distributions calibrated to 1 min return data for all 50 stocks in our dataset. We see that bulk of the distribution is concentrated between degrees of freedom 2.5 to 4. *Right panel*: The values of correlation of the PageRanks between networks across days averaged for each frequency considered. The black line with stars corresponds to the empirical data. The other lines with different identifiers are obtained from simulated data from a multivariate t-distribution with degrees of freedom ranging from 2.8 to 4.2. Smaller degrees of freedom lead to lower values of correlation. The overall pattern of decay in correlation is visible for all the cases considered. In particular, the empirical feature seems to match with t-distributed multivariate stochastic process with degrees of freedom around 3.5

the return distribution. Refs. [24,25] studied log-return data of the S&P500 stock market indices and reported that within a rich family of distributions, Student t distribution with about 4.5 degrees of freedom best fits the data. Many other independent studies showed Student's t distribution is a good choice for modeling financial return data [7,16,30]. It is well known that lower the degrees of freedom, the heavier the tail is. As the degree of freedom tends to ∞, the Student's t-distribution tends to Gaussian distribution. In the other limit, when the degrees of freedom is 1, the distribution would correspond to Cauchy distribution.

We fit the distribution to 1min data. While it is possible to fit the distribution at higher frequencies, we do not do that for two reasons. One, for some stocks, liquidity can be low and that may impact the estimation. Two, more importantly, the returns at higher frequency like 10 s, are often not independent and exhibit significant autocorrelation. For example, Fig. 5 shows the ACF and PACF functions of TCS stock returns at 10sec and 1min. The figure clearly suggests a first-order serial-dependence for 10sec data but not for 1min. Therefore, we calibrate the model at 1min return for all stocks.

The left panel in Fig. 4 shows the histogram of degrees of freedom of the fitted t-distribution of 50 stocks estimated on the baseline dataset. We see that most of the mass of the distribution is concentrated between 2.5 and 4 with an average around 3.5. This range of degrees of freedom of the fitted t-distribution remains very similar even in the full dataset.

Table 5 shows the correlation of page-ranks between true and sample network for several frequencies with different heavy-tailed Student's t-distribution. We can see that the correlation is inversely related with tail concentration which suggests that not only the sample size but also the tail concentration contributes significantly to the estimation error.

To see whether heavy-tail affects the estimation of PageRanks, we perform a short simulation and match the output with the obtained results of real data analysis as presented in Table 1. We take multivariate t-distribution with varying degrees of freedom suggesting increasing heavy-tail property. For each of such distribution, we calculate the average correlation for all frequencies more than 1 min data (second column of Table 1). This excludes 10 s and 30 s data as there can be significant autocorrelation in the returns.

In Fig. 4, the black line with stars corresponds to the correlation of PageRanks of the empirical networks across days averaged for each frequency. The same quantities are calculated for simulated data with different degrees of freedom (shown in different colors) with the t-distribution. As the values of correlations are smaller for lower degrees of freedom, the plot suggests a heavy-tail underlying distribution. However, it is evident from the figure that the observed correlations are decreasing at a rate similar to a t-distributed multivariate stochastic process with degrees of freedom around 3.5.

6 Summary

Real-world dynamic networks display time-varying degrees of stability [31]. Often, stability of the network is directly influenced by external perturbations [36]. However, the origin of network instability may also stem from the inherent dynamical properties of the underlying system. This makes the inference of the source of instability problematic from a statistical point of view.

In this paper, we analyze this question in the context of financial comovement networks. A well-known observation in this literature is that changes in the topology of financial networks tend to correlate with periods of turbulence. However, we show that changes in network properties are not a reliable indicator of external perturbations, and in fact, can be caused by sampling frequencies and extreme fluctuations even during non-turbulent periods.

The network construction from time-series data of asset returns has known problem. Longer time-series allows construction of the comovement network via credible estimation of pairwise linkages. However, that also introduces the problem of mixing of periods of turbulence and relative calm (see, for example, [11]). Shorter time horizons on the other hand can segregate the periods of turbulence and relative calm, but creates the problem of high dimensionality, so that the linkages may not be estimated accurately.

We employ high-frequency data at the interval of 10 s along with lower frequency intraday data specifically from periods which are known to be relatively calm in the Indian financial market. We show that even during that period, the sampling frequency and the presence

of extreme fluctuations change the topological properties of the inferred networks. Via simulation exercises, we decompose the effect arising out of purely statistical noise and extreme fluctuations. The residual of the effect can be potentially attributed to time-varying properties of asset returns.

Our work directly relates to the networks and complex systems literature. While the context is asset price networks, the core message would carry through for other types of time-series networks, as well [43]. In particular, the temporal dependence of the networks may also destabilize mesoscopic organizations within them, e.g., communities [22]. Beyond the current literature on complex systems, our work brings forth the question of what is the optimal sampling frequency for inferring networks from dynamical behavior of a system, especially the ones that exhibit extreme fluctuations [15].

Acknowledgements TK thanks JSPS KAKENHI C17K01270, C 20K01752 for their financial assistance. ASC thanks R&P grant from IIM Ahmedabad and research support from UTI AMC chair in macroeconomics. We would like to thank Vikram Sarabhai library at IIMA for providing the data. All calculations were carried out at the HPC lab at IIMA. All remaining errors are ours.

Author contributions

JB and AC performed the numerical, computational and statistical analysis. AC, ASC, and TK contributed to the writing of the draft. ASC and TK conceptualized the problem statement.

Data availability The data used in this work can be purchased from Bloomberg (bloomberg.com).

Appendix

Table 4 List of equities in NIFTY-50 dataset

Ticker	Name
RIL	RELIANCE INDUSTRIES LTD
TCS	TATA CONSULTANCY SVCS LTD
HDFCB	HDFC BANK LIMITED
INFO	INFOSYS LTD
HUVR	HINDUSTAN UNILEVER LTD
HDFC	HOUSING DEVELOPMENT FINANCE
ICICIBC	ICICI BANK LTD
KMB	KOTAK MAHINDRA BANK LTD
BHARTI	BHARTI AIRTEL LTD
BAF	BAJAJ FINANCE LTD
SBIN	STATE BANK OF INDIA
ITC	ITC LTD
HCLT	HCL TECHNOLOGIES LTD
WPRO	WIPRO LTD
APNT	ASIAN PAINTS LTD
MSIL	MARUTI SUZUKI INDIA LTD
AXSB	AXIS BANK LTD
LT	LARSEN & TOUBRO LTD
UTCEM	ULTRATECH CEMENT LTD
NEST	NESTLE INDIA LTD
BJFIN	BAJAJ FINSERV LTD
SUNP	SUN PHARMACEUTICAL INDUS
HDFCLIFE	HDFC LIFE INSURANCE CO LTD
TTAN	TITAN CO LTD
BJAUT	BAJAJ AUTO LTD
ONGC	OIL & NATURAL GAS CORP LTD
ADSEZ	ADANI PORTS AND SPECIAL ECON
TTMT	TATA MOTORS LTD
PWGR	POWER GRID CORP OF INDIA LTD
MM	MAHINDRA & MAHINDRA LTD
SRCM	SHREE CEMENT LTD
DIVI	DIVI'S LABORATORIES LTD
JSTL	JSW STEEL LTD
NTPC	NTPC LTD
IOCL	INDIAN OIL CORP LTD
TECHM	TECH MAHINDRA LTD
BPCL	BHARAT PETROLEUM CORP LTD
SBILIFE	SBI LIFE INSURANCE CO LTD
BRIT	BRITANNIA INDUSTRIES LTD
COAL	COAL INDIA LTD
EIM	EICHER MOTORS LTD
GRASIM	GRASIM INDUSTRIES LTD
DRRD	DR. REDDY'S LABORATORIES
TATA	TATA STEEL LTD
IIB	INDUSIND BANK LTD
CIPLA	CIPLA LTD
HMCL	HERO MOTOCORP LTD
GAIL	GAIL INDIA LTD
HNDL	HINDALCO INDUSTRIES LTD
UPLL	UPL LTD

Table 5 Simulation exercises to capture the effects of extreme fluctuations on similarity between true network with known covariance matrix and realized sample networks. This table provides average correlations of PageRanks between the true network and the sample networks for different frequencies and different degrees of freedom. Low frequencies are associated with less similarities. This model captures the effects of extreme fluctuations as well as sampling frequencies

	10 s	30 s	1 min	2 min	5 min	10 min	30 min	1 h
TRUE network	0.83	0.79	0.71	0.61	0.45	0.34	0.19	0.09
(df=4)	(±0.08)	(±0.08)	(±0.09)	(±0.10)	(±0.17)	(±0.17)	(±0.16)	(±0.18)
TRUE network	0.92	0.85	0.77	0.66	0.47	0.36	0.20	0.10
(df=6)	(±0.02)	(±0.02)	(±0.05)	(±0.09)	(±0.14)	(±0.14)	(±0.16)	(±0.17)
TRUE network	0.45	0.44	0.42	0.38	0.33	0.27	0.16	0.10
(df=3)	(±0.24)	(±0.23)	(±0.21)	(±0.19)	(±0.18)	(±0.19)	(±0.15)	(±0.12)
Sample size	2550	849	424	212	84	42	14	7

Table 6 Correlation table of $\rho_t^{\bar{S}}(f_i, f_j)$ for different values of frequencies f_i and f_j corresponding to average subsample correlation matrix $\bar{S}$. The mean correlation values are reported and their corresponding standard deviations are also reported in the brackets. Larger differences in frequencies are associated with lower correlations between $Z(\Gamma_{f_i,t})$ and $Z(\Gamma_{f_j,t})$

Frequency	10 s	30 s	1 min	2 min	5 min	10 min	30 min	1 h
10sec	1.0	0.99 (±0.00)	0.99 (±0.00)	0.97 (±0.01)	0.89 (±0.04)	0.78 (±0.10)	0.51 (±.18)	0.32 (±0.19)
30sec	0.99 (±0.00)	1.00	0.99 (±0.00)	0.97 (±0.01)	0.89 (±0.04)	0.78 (±0.10)	0.52 (±0.18)	0.32 (±0.19)
1min	0.99 (±0.00)	0.99 (±0.00)	1.00	0.97 (±0.01)	0.89 (±0.05)	0.79 (±0.10)	0.52 (±.18)	0.32 (±0.20)
2min	0.97 (±0.01)	0.97 (±0.01)	0.97 (±0.01)	1.00	0.89 (±0.05)	0.80 (±0.11)	0.52 (±.18)	0.31 (±0.20)
5min	0.89 (±0.04)	0.89 (±0.04)	0.89 (±0.05)	0.89 (±0.05)	1.00	0.81 (±0.09)	0.51 (±0.20)	0.31 (±0.21)
10min	0.78 (±0.10)	0.78 (±0.10)	0.79 (±0.10)	0.80 (±0.11)	0.81 (±0.09)	1.00	0.62 (±0.16)	0.31 (±0.18)
30min	0.51 (±.18)	0.52 (±0.18)	0.52 (±0.18)	0.52 (±.18)	0.51 (±0.20)	0.62 (±0.16)	1.00	0.60 (±0.14)
1hr	0.32 (±0.19)	0.32 (±0.19)	0.32 (±0.20)	0.31 (±0.20)	0.31 (±0.21)	0.37 (±0.18)	0.60 (±0.14)	1.00

Table 7 Stability of networks generated from simulated multivariate Gaussian processes. The table shows of $\rho_t^{RCV}(f_i, f_j)$ for different values of frequencies f_i and f_j corresponding to the average covariance matrix RCV obtained from simulation. Standard deviations corresponding to the point estimates of the correlations across PageRanks are given in the brackets. Larger differences in frequencies are associated with lower correlations between the resulting networks. Since the underlying stochastic process is Gaussian, the decay in correlations can be attributed to the differences in sample sizes across frequencies

Frequency	10sec	30sec	1min	2 min	5 min	10 min	30 min	1 h
10sec	1.00	0.96 (±0.00)	0.89 (±0.02)	0.81 (±0.04)	0.62 (±0.08)	0.47 (±0.12)	0.27 (±.14)	0.19 (±0.12)
30sec	0.96 (±0.00)	1.00	0.94 (±0.01)	0.84 (±0.04)	0.65 (±0.08)	0.50 (±0.10)	0.29 (±0.15)	0.20 (±0.11)
1min	0.89 (±0.02)	0.94 (±0.01)	1.00	0.90 (±0.03)	0.70 (±0.07)	0.54 (±0.09)	0.31 (±.14)	0.21 (±0.11)
2min	0.81 (±0.04)	0.84 (±0.04)	0.90 (±0.03)	1.00	0.76 (±0.05)	0.61 (±0.08)	0.36 (±0.14)	0.24 (±0.12)
5min	0.62 (±0.08)	0.65 (±0.08)	0.70 (±0.07)	0.76 (±0.05)	1.00	0.79 (±0.05)	0.47 (±0.14)	0.30 (±0.13)
10min	0.47 (±0.12)	0.50 (±0.10)	0.54 (±0.09)	0.61 (±0.08)	0.79 (±0.05)	1.00	0.60 (±0.12)	0.41 (±0.14)
30min	0.27 (±0.14)	0.29 (±0.15)	0.31 (±0.14)	0.36 (±0.14)	0.47 (±0.14)	0.60 (±0.12)	1.00	0.66 (±0.17)
1hr	0.19 (±0.12)	0.20 (±0.11)	0.21 (±0.11)	0.24 (±0.12)	0.30 (±0.13)	0.41 (±0.14)	0.66 (±0.17)	1.00

Table 8 Simulation exercises with Gaussian process to capture the similarity between true network with known covariance matrix and realized sample networks. This table provides average correlations of PageRanks between the true network and the sample networks for different frequencies. Low frequencies are associated with less similarities. This model captures the effects of sampling frequencies. This can be compared with table 5 which has the effects coming from both extreme fluctuations and sampling frequencies. The effects of extreme fluctuations are evident in the low values of the correlations in Table 5, especially for smaller degrees of freedom indicating fatter tails

	10sec	30sec	1min	2min	5min	10min	30min	1hr
TRUE network	0.95 (±0.01)	0.88 (±0.02)	0.81 (±0.04)	0.70 (±0.05)	0.53 (±0.13)	0.40 (±0.15)	0.24 (±0.18)	0.09 (±0.20)
Sample size	2550	849	424	212	84	42	14	7

Table 9 Robustness check: Correlation table of $\rho_t^{\bar{S}}(f_i, f_j)$ for different values of frequencies f_i and f_j corresponding to average subsample correlation matrix $\bar{S}$ obtained from the data of additional 138 days. The mean correlation values are reported and their corresponding standard deviations are also reported in the brackets. Larger differences in frequencies are associated with lower correlations between $Z(\Gamma_{f_i,t})$ and $Z(\Gamma_{f_j,t})$. The values are quite similar to the values reported in Table 6, and thus, the method discussed in the paper is robust

Frequency	10sec	30sec	1min	2min	5min	10min	30min	1hr
10sec	1	0.99 (±0.00)	0.99 (±0.00)	0.96 (±0.02)	0.88 (±0.05)	0.78 (±0.09)	0.51 (±0.17)	0.36 (±0.19)
30sec	0.99 (±0.00)	1	0.99 (±0.00)	0.97 (±0.02)	0.88 (±0.05)	0.78 (±0.09)	0.51 (±0.17)	0.36 (±0.19)
1min	0.99 (±0.00)	0.99 (±0.00)	1	0.97 (±0.01)	0.88 (±0.05)	0.78 (±0.09)	0.51 (±0.17)	0.37 (±0.19)
2min	0.96 (±0.02)	0.97 (±0.02)	0.97 (±0.02)	1	0.88 (±0.05)	0.79 (±0.08)	0.51 (±0.17)	0.35 (±0.19)
5min	0.88 (±0.05)	0.88 (±0.05)	0.88 (±0.05)	0.88 (±0.05)	1	0.81 (±0.08)	0.49 (±0.17)	0.33 (±0.19)
10min	0.78 (±0.09)	0.78 (±0.09)	0.78 (±0.09)	0.79 (±0.08)	0.80 (±0.08)	1	0.61 (±0.15)	0.41 (±0.20)
30min	0.50 (±0.17)	0.51 (±0.17)	0.51 (±0.17)	0.51 (±0.17)	0.49 (±0.17)	0.62 (±0.15)	1	0.65 (±0.16)
1hr	0.36 (±0.19)	0.36 (±0.19)	0.37 (±0.19)	0.35 (±0.19)	0.33 (±0.19)	0.41 (±0.20)	0.65 (±0.16)	1

Table 10 Robustness check: Correlation between the PageRanks corresponding to different frequencies (f) averaged across all possible pairs of days in the additional dataset of 138 days. The average correlation decreases for lower frequencies. The correlation values for RCV are higher than the noise-corrected covariance matrix $\bar{S}$. The values are close to the values in Table 1, and thus, the results are not vulnerable to sampling fluctuations

Frequency	Average Correlation (RCV)	Average correlation ($\bar{S}$)
f=10sec	0.65 (±0.11)	0.35 (±0.18)
f=30sec	0.54 (±0.15)	0.35 (±0.18)
f=1min	0.47 (±0.16)	0.35 (±0.18)
f=2min	0.38 (±0.16)	0.35 (±0.17)
f=5min	0.26 (±0.17)	0.31 (±0.17)
f=10min	0.18 (±0.17)	0.23 (±0.17)
f=30min	0.07 (±0.17)	0.08 (±0.17)
f=1hr	0.03 (±0.17)	0.03 (±0.17)

Table 11 Robustness check through alternative stability measure: Heatmap of the Spearman correlation matrix of the row-sums of $\bar{S}$. Instead of the PageRank, here, we take the p-dimensional row-sum of the $\bar{S}$ matrix as an lower dimensional representation of the underlying network structure. To check the stability of the rank ordering of the stocks, we calculate the Spearman correlation between the row-sums of $\bar{S}$ matrix corresponding to different sampling frequency. We observe a pattern similar to that we find for PageRank, suggesting that the stability of the network can be significantly affected by sampling frequency and sample size

	10sec	30sec	1min	2min	5min	10min	30min	1hr
10sec	1	0.99 (± 0.00)	0.99 (± 0.00)	0.97 (± 0.01)	0.89 (± 0.05)	0.78 (± 0.11)	0.52 (± 0.18)	0.33 (± 0.19)
30sec	0.99 (± 0.00)	1	0.99 (± 0.00)	0.97 (± 0.01)	0.89 (± 0.05)	0.78 (± 0.11)	0.52 (± 0.18)	0.33 (± 0.19)
1min	0.99 (± 0.00)	0.99 (± 0.00)	1	0.97 (± 0.01)	0.89 (± 0.05)	0.78 (± 0.11)	0.52 (± 0.18)	0.33 (± 0.19)
2min	0.97 (± 0.01)	0.97 (± 0.01)	0.97 (± 0.01)	1	0.88 (± 0.05)	0.80 (± 0.11)	0.52 (± 0.17)	0.32 (± 0.19)
5min	0.89 (± 0.05)	0.89 (± 0.05)	0.89 (± 0.05)	0.88 (± 0.05)	1	0.81 (± 0.09)	0.51 (± 0.19)	0.32 (± 0.21)
10min	0.78 (± 0.10)	0.78 (± 0.11)	0.78 (± 0.11)	0.80 (± 0.11)	0.81 (± 0.09)	1	0.62 (± 0.16)	0.38 (± 0.18)
30min	0.51 (± 0.17)	0.52 (± 0.17)	0.52 (± 0.17)	0.52 (± 0.17)	0.51 (± 0.19)	0.62 (± 0.16)	1	0.61 (± 0.14)
1hr	0.33 (± 0.19)	0.33 (± 0.19)	0.33 (± 0.19)	0.32 (± 0.19)	0.32 (± 0.21)	0.38 (± 0.18)	0.61 (± 0.14)	1

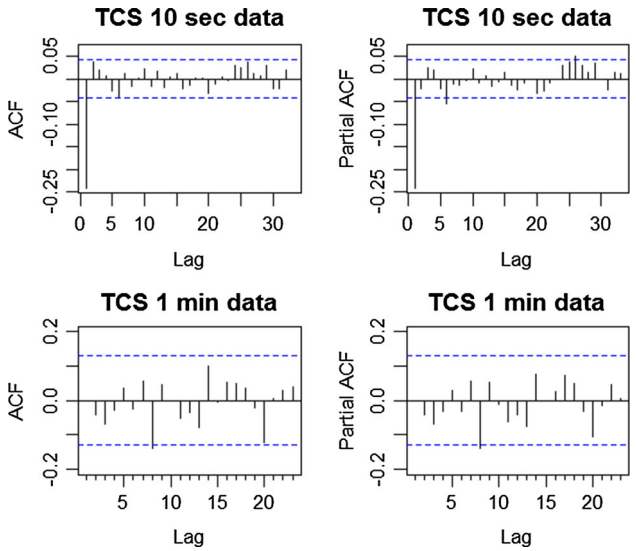

Fig. 5 ACF and PACF obtained from 10 s and 1 min returns data of TCS on a single day. 10 s Returns data show significant time-dependence in lag 1 both in terms of ACF and Partial ACF as opposed to the 1 min returns data

References

1. A.R. Admati, P. Pfleiderer, A theory of intraday patterns: Volume and price variability. The Review of Financial Studies **1**(1), 3–40 (1988)
2. Y. Aït-Sahalia, J. Fan, D. Xiu, High-frequency covariance estimates with noisy and asynchronous financial data. J. Am. Stat. Assoc. **105**(492), 1504–1517 (2010)
3. T.G. Andersen, T. Bollerslev, Intraday periodicity and volatility persistence in financial markets. J. Empir. Financ. **4**(2–3), 115–158 (1997)
4. T. Aste, W. Shaw, T. Di Matteo, Correlation structure and dynamics in volatile markets. New J. Phys. **12**(8), 085009 (2010)
5. M. Bardoscia, P. Barucca, S. Battiston, F. Caccioli, G. Cimini, D. Garlaschelli, F. Saracco, T. Squartini, G. Caldarelli, The physics of financial networks. Nature Reviews Physics, pages 1–18 (2021)
6. O.E. Barndorff-Nielsen, P.R. Hansen, A. Lunde, N. Shephard, Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. Journal of Econometrics **162**(2), 149–169 (2011)
7. R.C. Blattberg, N.J. Gonedes, A comparison of the stable and student distributions as statistical models for stock prices, in *Perspectives on promotion and database marketing: The collected works of Robert C Blattberg*. (World Scientific, 2010), pp. 25–61
8. G. Bonanno, F. Lillo, R.N. Mantegna, High-frequency cross-correlation in a set of stocks. Quantitative Finance **1**(1), 96–104 (2001)
9. C. Borghesi, M. Marsili, S. Micciche, Emergence of time-horizon invariant correlation structure in financial returns by subtraction of the market mode. Phys. Rev. E **76**(2), 026104 (2007)
10. T.T. Cai, J. Hu, Y. Li, X. Zheng, High-dimensional minimum variance portfolio estimation based on high-frequency data. Journal of Econometrics **214**(2), 482–494 (2020)
11. A. Chakraborti, K. Sharma, H.K. Pharasi, K.S. Bakar, S. Das, T.H. Seligman, Emerging spectra characterization of catastrophic instabilities in complex systems. New J. Phys. **22**(6), 063043 (2020)
12. K. Christensen, S. Kinnebrock, M. Podolskij, Pre-averaging estimators of the ex-post covariance matrix

in noisy diffusion models with non-synchronous data. Journal of Econometrics **159**(1), 116–133 (2010)
13. R. Cont, Empirical properties of asset returns: stylized facts and statistical issues. Quantitative finance **1**(2), 223 (2001)
14. F.X. Diebold, K. Yılmaz, *Financial and macroeconomic connectedness: A network approach to measurement and monitoring* (Oxford University Press, USA, 2015)
15. Z. Eisler, I. Bartos, J. Kertész, Fluctuation scaling in complex systems: Taylor's law and beyond. Adv. Phys. **57**(1), 89–142 (2008)
16. S.R. Hurst, E. Platen, The marginal distributions of returns and volatility. Lecture Notes-Monograph Series, pages 301–314 (1997)
17. J. Jacod, P. Protter, Asymptotic error distributions for the euler method for stochastic differential equations. Ann. Probab. **26**(1), 267–307 (1998)
18. D.Y. Kenett, T. Preis, G. Gur-Gershgoren, E. Ben-Jacob, Quantifying meta-correlations in financial markets. EPL (Europhysics Letters) **99**(3), 38001 (2012)
19. D.Y. Kenett, M. Raddant, T. Lux, E. Ben-Jacob, Evolvement of uniformity and volatility in the stressed global financial village. PLoS ONE **7**(2), e31144 (2012)
20. S. Kumar, A. Bansal, A.S. Chakrabarti, Ripples on financial networks. The European Journal of Finance, (2021)
21. C. Kuyyamudi, A.S. Chakrabarti, S. Sinha, Emergence of frustration signals systemic risk. Phys. Rev. E **99**(5), 052306 (2019)
22. M. MacMahon, D. Garlaschelli, Community detection for correlation matrices. arXiv preprint arXiv:1311.1924 (2013)
23. R.N. Mantegna, H.E. Stanley, *Introduction to econophysics: correlations and complexity in finance* (Cambridge University Press, 1999)
24. H.M. Markowitz, N. Usmen, The likelihood of various stock market return distributions, part 1: Principles of inference. J. Risk Uncertain. **13**(3), 207–219 (1996)
25. H.M. Markowitz, N. Usmen, The likelihood of various stock market return distributions, part 2: Empirical results. J. Risk Uncertain. **13**(3), 221–247 (1996)
26. F. Musciotto, J. Piilo, R.N. Mantegna, High-frequency trading and networked markets. Proceedings of the National Academy of Sciences **118**(26), (2021)
27. E. Platen, R. Rendek, Empirical evidence on student-t log-returns of diversified world stock indices. Journal of Statistical Theory and Practice **2**(2), 233–251 (2008)
28. B. Podobnik, H.E. Stanley, Detrended cross-correlation analysis: a new method for analyzing two nonstationary time series. Phys. Rev. Lett. **100**(8), 084102 (2008)
29. J.M. Pollet, M. Wilson, Average correlation and stock market returns. J. Financ. Econ. **96**(3), 364–380 (2010)
30. P.D. Praetz, The distribution of share price changes. Journal of Business, pages 49–55 (1972)
31. A. Rai, A. Bansal, A.S. Chakrabarti, Statistical estimation of time-varying complexity in financial networks. The European Physical Journal B **92**(10), 1–9 (2019)
32. T. Squartini, D. Garlaschelli, Stationarity, non-stationarity and early warning signals in economic networks. Journal of Complex Networks **3**(1), 1–21 (2015)
33. S.K. Stavroglou, A.A. Pantelous, H.E. Stanley, K.M. Zuev, Hidden interactions in financial markets. Proc. Natl. Acad. Sci. **116**(22), 10646–10651 (2019)
34. S.K. Stavroglou, A.A. Pantelous, H.E. Stanley, K.M. Zuev, Unveiling causal interactions in complex systems. Proc. Natl. Acad. Sci. **117**(14), 7599–7605 (2020)
35. M. Tumminello, T. Di Matteo, T. Aste, R.N. Mantegna, Correlation based networks of equity returns sampled at different time horizons. The European Physical Journal B **55**(2), 209–217 (2007)
36. M. Vyas, T. Guhr, T. Seligman, Multivariate analysis of short time series in terms of ensembles of correlation matrices. Sci. Rep. **8**(1), 1–12 (2018)
37. Y. Wang, J. Zou, Vast volatility matrix estimation for high-frequency financial data. Ann. Stat. **38**(2), 943–978 (2010)
38. N. Xia, X. Zheng, On the inference about the spectra of high-dimensional covariance matrix based on noisy observations-with applications to integrated covolatility matrix inference in the presence of microstructure noise. arXiv preprint arXiv:1409.2121 (2014)
39. L. Zhang, Estimating covariation: Epps effect, microstructure noise. Journal of Econometrics **160**(1), 33–47 (2011)
40. L. Zhang, P.A. Mykland, Y. Aït-Sahalia, A tale of two time scales: Determining integrated volatility with noisy high-frequency data. J. Am. Stat. Assoc. **100**(472), 1394–1411 (2005)
41. L. Zhao, G.-J. Wang, M. Wang, W. Bao, W. Li, H.E. Stanley, Stock market as temporal network. Phys. A **506**, 1104–1112 (2018)
42. X. Zheng, Y. Li, On the estimation of integrated covariance matrices of high dimensional diffusion processes. Ann. Stat. **39**(6), 3121–3151 (2011)
43. Y. Zou, R.V. Donner, N. Marwan, J.F. Donges, J. Kurths, Complex network approaches to nonlinear time series analysis. Phys. Rep. **787**, 1–97 (2019)