EPJB_BCKC

1. **Introduction**
    - The article talks about the topological system of networks and the factors which affect it. First of all there is an analysis if we should consider a longer time period or a shorter one. A longer builds a better network because of more data points but suffers from potential mixing of turbulent and non turbulent periods. A short data frame is able to handle with the first issue to separate turbulent and non turbulent periods but is not able to handle the build a topologically strong network because of the less number of data points
    - The research paper aims to establish a connection between the networks and the stock market. This is primarily due to 2 big reasons One, reliable high-frequency financial data are easily available. Two, it is relatively easy to differentiate between periods of financial turbulence and calmness.
    - Turbulence or calmness is an important factor but is not the sole factor which affects the topology of a network. 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.
    - Building a financial network is able to infer a lot of interesting findings such as:
        i. A network is able to tell us the current state of the financial market
        ii. A network can predict futures up and downs
        iii. Correlation between various sectors
    - The network's stability is evaluated using an algorithm named page rank, this is used to capture changes in the network properties both across different days and within the same day at various frequencies. The network stability is influenced by factors such as sample size and turbulence and the time dependent characteristics of returns. So one must take caution while interpreting the changes in the networks and drawing conclusions because a change in the network does not indicate a financial crisis and might be due to some other factor.
2. **Comovement networks from High frequency data**
    - *Intraday price dyanamics*
        i. There are certain terminologies given which are used in the model. The first section talks abt a financial model/network built on intraday price movement.
        ii. The model is as follows: Instead of using price p, it uses log p. Based on these price, the below equation is used in the model.. $X_{t+\tau}$ represents the log-prices of all p stocks at time τ, $\mu_{t+\tau}$ is the drift process.

Then the covariance matrix is computed for all stock returns.

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

- *Estimator of integrated covariance matrix Σt*
  i. Another widely used method for estimating covariance is RCV which computes the difference in processes and then draws interpretation
- *Filtering microstructure noise from SRC*
  i. In real world pricing data, there is an accumulation of a large amount of microstructure noise which arises due to bid-ask spreads, trading delays,etc
  ii. So the method to filter this noise is the Two Scales Realized Volatility(TSRV) estimator which modifies the covariance matrix in order to filter out the noise, it is based on the principle of average subsample covariance.

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

  iii. TSRV is estimated using the below formulae
  iv. There are also other indicators like pre-averaging indicator, realized kernel estimator,etc which almost yield similar results to TSRV, however still TSRV is used because it is easy to implement and retains useful asymptotic properties making it suitable for filtering noise in intraday price data
- **Network construction**
  i. After the construction of the covariance data, one converts this into the correlation matrix by scaling the assets to understand how closely are they both related to each other
  ii. In order to construct networks, each asset represents a node and the edges weight is calculated using the distance formulae given below.

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

  iii. Hence we obtain a weighted graph, where the weight on the edges represent the strength of the graph, short distances implies stronger edges and vice versa.
- **Network characteristics and stability**

i. Now the network system has also been established, the next step is to analyze the network characteristics and stability.
ii. The 1st method is to study the variability of the edges weight, this can be done by studying the difference in each weight's across different networks.
iii. The 2nd method is PageRank which computes the adjacency matrix of the network, the formulae is given below:

$$Z_{it} = \frac{1-\alpha}{p} + \alpha \sum_{j} E_{ijt} Z_{jt},$$

iv. To quantify the stability we can use a correlation measure(pRCV), it assesses the correlation between Pagerank centralities with different frequencies, high value indicates higher network stability and low value suggests instability. The average value of pRCV captures how the network changes with time.

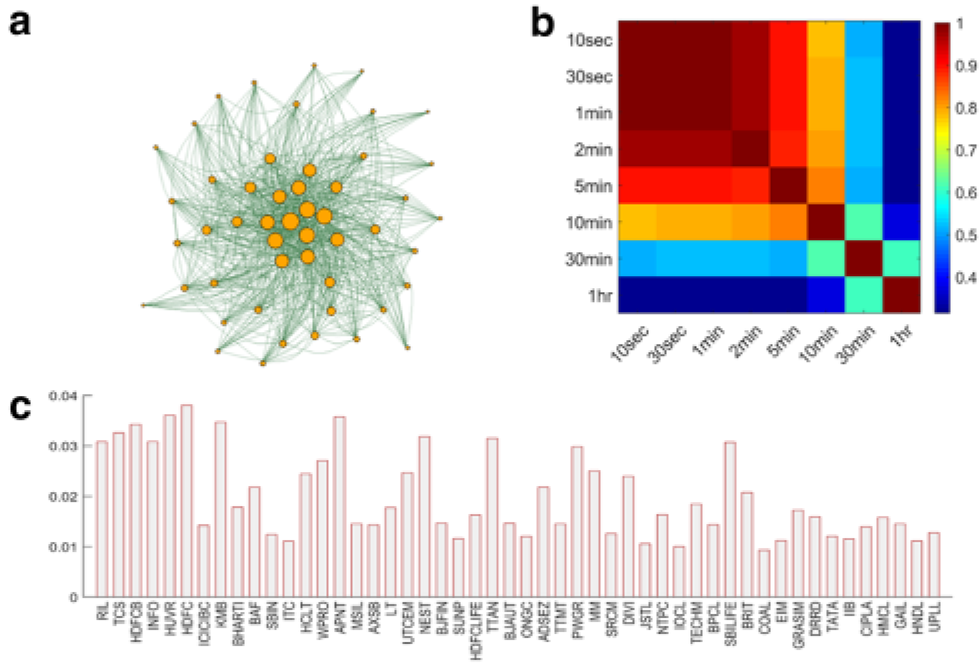
$$\rho_t^{RCV}(f_i, f_j) = Cor(Z(\Gamma_{f_i,t}), Z(\Gamma_{f_j,t})),$$

v. There is another approach focusing on the centrality measures arising from the first order neighbourhood of each node, they construct node wise degree centrality within the weighted network.

3. Characteristics of instability

- *Construction of the emperical network*

i. The below shows examples of correlation network obtained at a 10 sec interval from the research paper along with a heatmap and distribution of normalized PageRank. They are analyzed on the NIFTY 50 equities

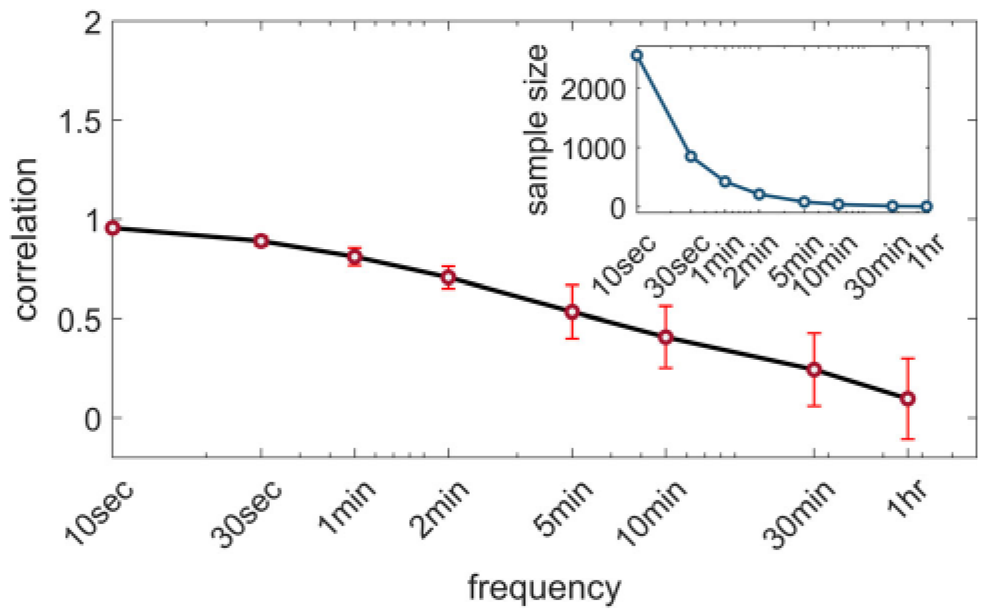


- *<u>Instability of rank-ordering</u>*
    i. The table computes the average value of the PageRank in order to check the correlation values. While analyzing Lower frequency data the PageRank value decreases, this could be due to the sampling size reduction, but that is not the sole factor.
    ii. This also shows through real life data that the correlation values of the original matrix(containg noise) RCV is much higher than those of the noise corrected covariance matrix

4. **Robustness**
    - The research paper also cites that the above method of building a network through correlation matrices was also again applied and it generated similar results as the data evaluated in the previous paras. It showed that sampling size(frequency) does have a strong effect on the correlation values hence being consistent with the study proving the method as robust.

5. **Instability of networks: simulations**
    - *<u>5.1 Effects of sampling frequency</u>*
        i. So what we are doing is analysing a gaussian distribution with constant covariance matrix.The return distribution is assumed to be Gaussian, and the mean and covariance matrix for this 50-dimensional Gaussian distribution are set to match the sample mean and sample covariance obtained from the real 10-second data.

ii. The average correlation value decreases as the size decreases. Even considering a simple model the analysis suggest that to retain 90% of true PageRank we would need a data of abt 6 years which challenges the unchanged nature of the network for such an extend period of time.



iii. The Findings also indicate that a larger difference in sample sizes implies a lower correlation, those with small difference are better correlated.

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

iv. Hence a data with a very low sampling frequency can perhaps be misleading and drive us to false conclusions/trends. There is a substantial difference in the real data with the simulated data, this is due to the fat tailed nature of the distribution.

v. The weight of the edge is computed using the covariance matrix. However the errors in estimating the weight of edges are huge in case of low frequency data, the variance of errors handles as well as the average error also increases.

vi. So hence now the research paper considers the U shaped volatility pattern in the financial markets, this is composed of steps, c1 for volatile at beginning of day, c2 for less volatile during mid of day, c3 for volatile period at end of day. Then the returns are generated considering the volatility in each part of the trading day and then this is plotted for different frequencies just like the previous one.

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

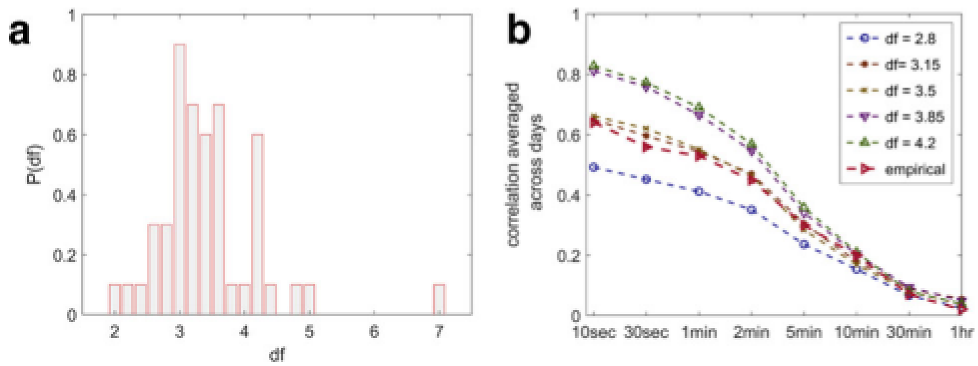
vii. But to the surprise even after changing the model the values did not have a much severe impact and still had low correlation values for less sampling size/frequency.

viii. This analysis suggest that choice of data frequency is perhaps the most crucial factor while evaluating network stability compated to the specific details off the time varying volatility process. Hence low frequency is accompanied with a larger network instability.

- *Effect of extreme fluctuations*

  i. This passage evaluates a different modelling from the gaussian distribution which is the student's T distribution. A heavy tail implies that extreme events are more likely to occur than under a gaussian distribution.

  ii. The Student's T model has a parameter called degree of freedom(df) which governs the heaviness of tails. Higher df is similar to gaussian while lower df is for heavier tails.



iii. The analysis as given below computes the df of the 50 stocks in 1 minute frequency and finds the most of the df is concentrated between 2.5 and 4 and hence the average is around 3.5.
iv. The analysis point that correlation is inversely proportional to the heaviness of tails.
v. Hence the correlation is both inversely proportional to the sampling size as well as the heaviness of tails

6. **Summary**
   - In summary, the real world networks display a wide variety in network stability, and these are affected by a variety of factors.
   - Constructing a network from the time series data is challenging as taking a large length data gives us a lot of data but is accompanied with a lot of noise along with mixing periods of turbulence and calm.
   - Constructing using a short length data is also not appropriate as even though it distinguishes between turbulence and calm they do not take account a large number of nodes and hence may not be accurate.
   - The data use is a high frequency data or a low frequency intra day data in relatively calm periods in the indian market.
   - The major factors which affect the network stability are the sampling frequency and the presence of extreme fluctuations. The paper also deals with the statistical ways as to how to remove noise and use them to build the correlation matrix.