Project proposal

Elements of network science

Shashvath PV, Yuhui, Siddhanth

April 4 2016

What we intend to do?

We intend to pave a new methodology for predicting stock price or behavior based on the investors' behavior by constructing a new network. The edge in the network should be marked, since we need their relation between each other. The investor to investor relation can help us easily understand the cascading when we try to make the network balanced, this kind of relation will give us insights for predicting the price change. The investor to stock relation can help us with knowing investors' behavior and these markings will stand for the sale and purchase behavior, when the investors' decide it. The relation between one stock to another is usually based on if they have cooperation or if they have competition. For constructing this network, we will use the previous date. After we get this new mark network, we will analyze the network to get the prediction of the specific stock for the next few days. Since the stock market behavior is fully based on the investors' behavior, we only have to predict the investors' behavior, the result of which shall be presented after the analysis on the investors. This kind of prediction can guide people remarkably for their investments and give provide new approaches for their financial tools.

Methods planned to used and developed:

For this new method, we shall be using cascading; since people sometimes invest just irrationally, they tend to follow other people. For cascading, we might mainly use the balanced part but we would also consider the weak balanced, since people are fickle-minded and change their decision very fast or they struggle a lot, the weak balanced can easily show the network balanced when people struggle takes place. Also we will use game theory, since every investor wants to maximize their profit and minimize their risks. The trade-off for each behavior they do would be considering tax, interest rate and other reasons. These things will represent in full limelight how game theory takes place. Another point we need consider is about the investor population change, since for one stock, it might just have different investors decide to invest in it or quit it. That is how diffusion occurs here. We will use some diffusion model to check about the diffusion process happening in a specific stock; also we need to consider some other factors which come from the company itself or the financial environment. Since the network will be marked, we will need to find the most positive relation in the highest degree, which can represent a lot of nodes idea when they start investor, since most of these kind nodes should represent some big financial companies, investors' funding or some famous independent investors.

Data to be used and obtaining them:

We shall be sourcing real data sets from Yahoo Finance, wherein we can use it directly on R by pulling the data through the URL; we aim to pull it up for one or more stock on the market for detailed insights. Another website called Quandl is the one we would look into for exporting data from by installing the Quandl R-Package and have the data sets directly accessible from within R. This site just allows 50 calls a day, so we would need an authentication token by registering on their API page for free unlimited usage. The kind of data we are sourcing from these sites for our evaluation includes but is not limited to the following –

Close, High, Low, Open, Volume, Date, Granularity as Daily, Weekly, Monthly, Insider Purchases – Total Insider Shares Held, Net Shares Purchased (Sold), Sales, Insider Holders.

Background work:

With respect to the details provided in first few sections of the document, we intend to have a new method for predicting stock prices by building a new network. In order to have a clear picture about the stocks and financial data, we went through one of the research paper called "A social network approach to examine the role of influential stocks in shaping interdependence structures in global stock markets", published at Advances in Social network analysis and mining(ASONAM), 2011. This paper discusses about the role of influential stocks in shaping the system level interdependence in global stock markets and also method to identify influential stocks using various centrality measures. The study also showed the change in network topology during the collapse of Lehman Brothers.

In order to obtain the data about the stocks (opening, closing, volume, day wise change, etc.) we tend to export from yahoo finance. It also shows the network traffic change during the opening and closing times of the stock market.

"A network approach. Computers and operations research" Volume: 33 (2006), gives the idea behind networks and its behavior. It gives the detailed information about the calculating cross correlations between pair of stocks based on opening prices data over a certain period of time.

These are the certain details that we have captured up to now in order to have a clear picture about the proposed project.

Tentative Plans:

The tentative plans for the successful completion of the project are as follows:

1. Have a detailed study about the stocks opening and closing details and network load created during different timings of a day and user behavior relation with respect to stocks.
2. Prepare the .xsl files and .csv files for different companies' stocks.
3. Gathering the required knowledge to build a network that would be able to serve our purpose. Know more about the the behavior of network during peak times and passive times. How will the network be affected due to high traffic, etc.
4. User behavior to stock market operations and how the other factors like prices, networks, etc. would be affected due to this.
5. Preparation of the prediction model to predict stock price based on all the above characteristics.