Network analysis of a stock market

Project description:

How is information incorporated into financial markets and asset prices? This is a general and important question that is receiving a lot of attention in the research field of Financial Economics. Network theory offers a novel approach to analyzing this question. In a market where investors share private information through a network, specific predictions arise, e.g., about the distribution of trading profits, the correlation of trades and the trading volume.

The goal of the project is to analyze a real market from a network perspective. Specifically, a dataset is available, containing all trades in a stock market (at the account level) over an extended time period. Under the assumption that investors who are informationally connected in a network trade in similar ways, an empirical information network (EIN) can be constructed from the data. The EIN can then be used to analyze several open questions:

1. The EIN can be constructed by identifying agents as connected, who trade in the same stock close in time. Although the approach is intuitively clear, the specific details are not. For example, what is the correct “time window” of trades to use when identifying two investors as connected?

2. What is the distribution of connectedness - i.e., the degree distribution - in the network? Oftentimes the distribution is well approximated by a power law in the tails. Is this the case for the EIN? If so, what is the so-called tail-exponent of the degree distribution?

3. What is the relationship between degree, centrality and trading volume for individual agents in the EIN?

4. How persistent are these measures over time?

5. How are these measures related to realized trading profits and trade correlations between individual investors? If information diffusion is indeed an important factor, then the centrality should be superior to other measures.

The analysis is theoretically straightforward, but the dataset is extensive (about 500,000 traders, 200,000 trades per day and a year of data). A careful implementation is needed to keep the analysis computationally feasible. An initial C++ program that uses sparse matrices (from the Boost library) to construct the EIS has already been implemented in Linux (gcc compiler).

Deliverables:

The objective of the project is to extend the C++ program and analysis to answer the five questions above. Deliverables:

1. The C++ program

2. The data analysis

3. A report that summarizes the results

The specific details of later parts of the project will depend on the initial results. Students will therefore interact with the Professor (who is on sabbatical leave in Stockholm this semester) on a regular basis.

Prerequisites:

- Extensive experience of C++ programming
- A good understanding of numerical methods in linear algebra (e.g., eigenvector decomposition/spectral analysis of matrices)
- Some experience with network/graph theory is a plus
- Some knowledge about finance is also a plus

Contact:

Students who are interested and/or have questions can contact:

Johan Walden
University of California at Berkeley
e-mail: walden@haas.berkeley.edu

Please provide some information about yourself in the e-mail (e.g., about relevant courses and experience)