Limit Order Book in High Frequency Trading (HFT) A Network Approach

Chonkit Pun
Tianchi Chen

Outline

- Introduction to limit order book
- Motivations
- Data and reconstruction
- Result and analysis (Minimum spanning tree)
- Conclusion

Intro to Limit Order Book (LOB)

- A book keeping track of outstanding orders at different price levels

- Types of order (basic):
  - Buy (bid) limit order
  - Sell (ask) limit order
  - Buy market order
  - Sell market order

- Best ask – best bid = bid-ask spread

PRICE	BID	ASK
120.242		670
120.234		787
120.227		815
120.219		479
120.211	748	
120.203	1089	
120.195	890	
120.188	901	
120.180	801	
120.172	818	

Bloomberg business

Intro to LOB (cont.)

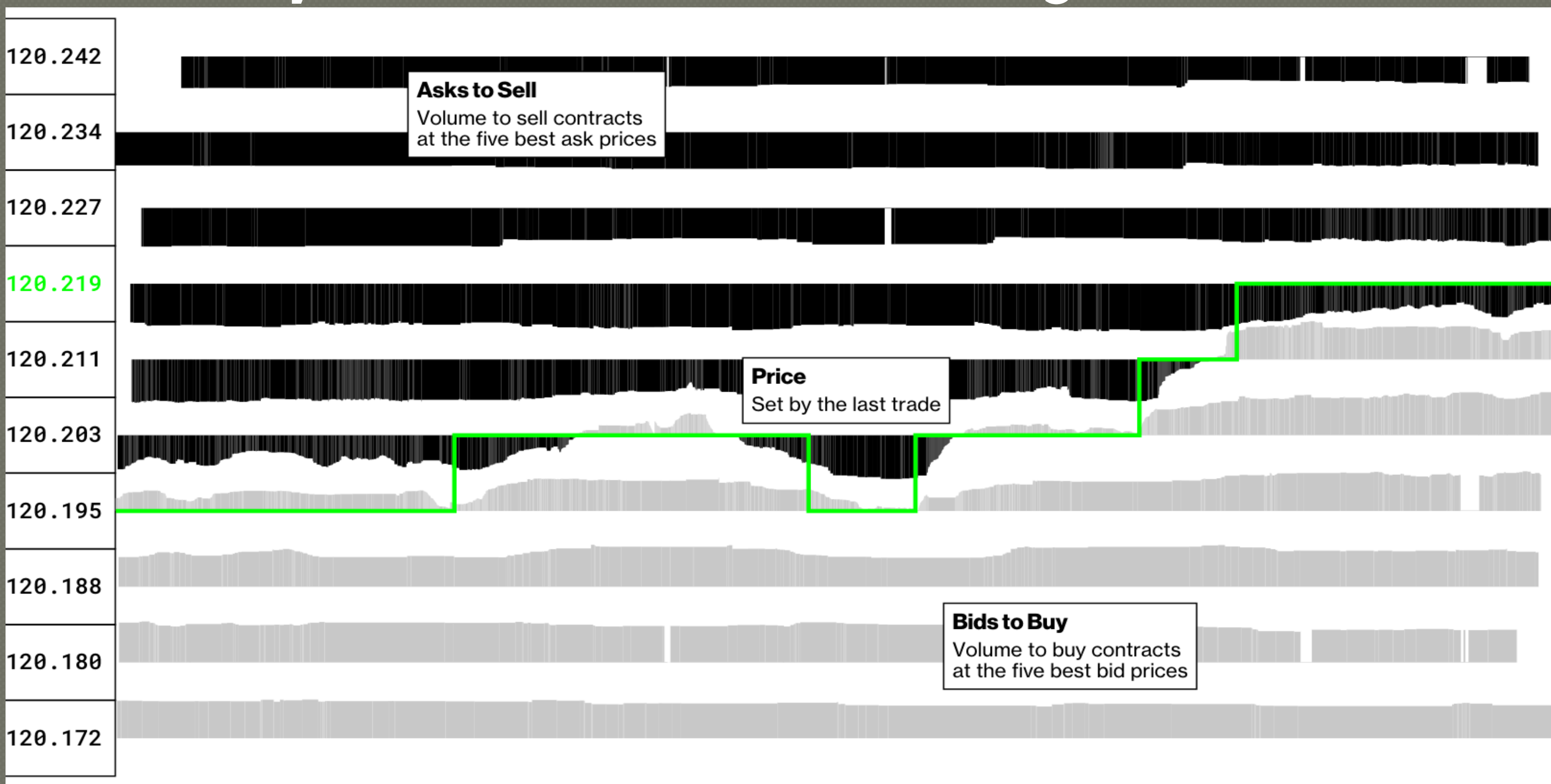
- Shares in the LOB are also removed when they are being executed (by market order of the other side)
- Price moves when all outstanding shares in the best price level (ask/bid) are depleted by market orders of the opposite side of the market
- Can also cancel in part or in total the placed order, in which orders are removed
- LOB gives investors information about the supply and demand in the market

PRICE	BID	ASK
120.242		670
120.234		787
120.227		815
120.219		479
120.211	748	
120.203	1089	
120.195	890	
120.188	901	
120.180	801	
120.172	818	

Bloomberg business

Intro to LOB (cont.)

- Evolution of the LOB for US government's 10-year Treasury bond, traded on the Chicago Board of Trade



Bloomberg business: '*How to Catch a Spoofer?*'

Motivation

- Is there a correlation in the changes of number of shares between different price level?
- What is the structure of the price levels in terms of the correlation between the price levels?
- Can the structure be explained by investors' psychology?

Motivation: Building a network

- We treat different price levels / ticks as the nodes in the network
- The correlation in the changes of number of shares in different price levels determine the links between the nodes
- Threshold correlation = $\frac{1}{\sqrt{N}}$
- To simplify the network, we do not distinguish between positive and negative correlations.

$$\tilde{S}_i = \frac{(S_i - \langle S \rangle)}{\sqrt{T}\sigma_S}$$

$$Corr_{ij} = \tilde{S}_i \tilde{S}_j$$

Data and Reconstruction

- Data from LOBSTER
  http://lobsterdata.com

LOBSTER
Limit Order Book System – The Efficient Reconstructor
academic data.

- Data sample:
  - AAPL (Apple) on NASDAQ
  - high frequency data on 2012/6/21 from 9:30 – 16:00
  - 400391 events (e.g. place, cancel, delete ,execute orders)
  - Does not include hidden events (e.g. orders placed in dark pools)
  - Data up to level 10 (first 10 best ask/bid), <u>not necessarily</u> the first 10 prices by the smallest ($0.01) increment

Data reconstruction (cont.)

- 400391 events = 224811 asks and 175580 bids
- Data resolution down to **10 μs !**
- Minimum price level increment (a tick)= **$0.01**

⇒ expect large fluctuations in the change in number of shares in a price level

- → Time binning: bin size = **500 events** (several seconds to half a minutes) ~ **5000 events** (2.5 to 5 minutes)

- → Time window: **4811th ~ 224811th events** (excluding events close to the opening of market), 9:36am ~4:00pm

- → Binning and time window is chosen to reduce fluctuations due to high frequency data and to generate enough statistics

- Look at correlations between the change in outstanding shares in the first 50 best prices (each differ by a tick $0.01)

Data reconstruction

time	Event	Size (shares)	Limit order price	Market price (best price)
34200.02555	place	200	585.91	585.91
34200.02558	place	200	585.92	585.91
34200.02561	cancel	(−)100	585.93	585.90
34200.20152	execution	(−) 150	585.92	585.92

:

$\Delta t = 2$

time	P_1	P_2	P_3	P_4	……
1	**200**	**200**	**0**	**0**	
2	**−150**	**0**	**0**	**−100**	
3	……				
4	……				

P_1 = best price (at time t)
P_2 = best price + \$0.01
P_3 = best price + \$0.02
….

Δ t = time binning

→ Price ticks are defined relative to the **instantaneous** market price

Result

Result

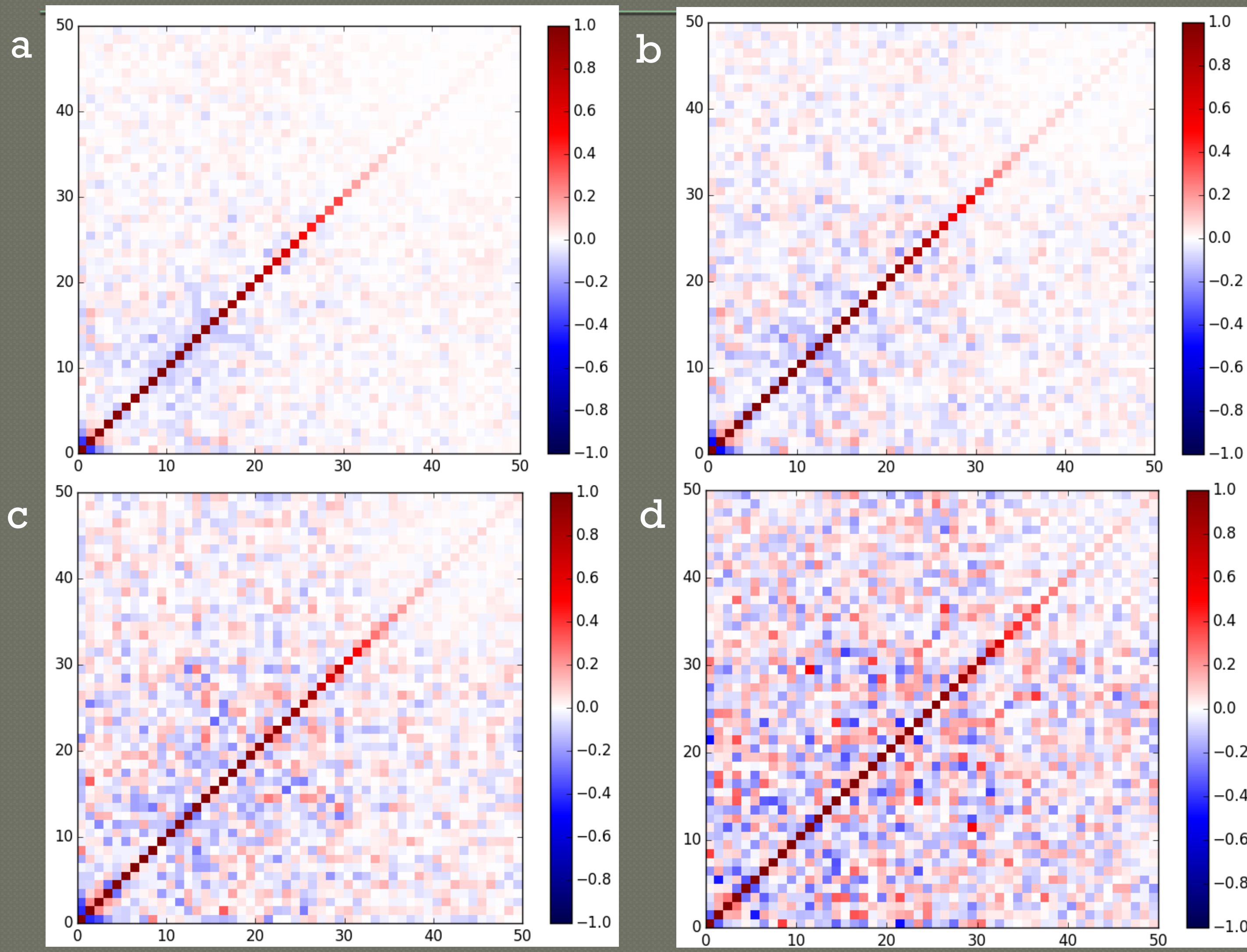
- Build a network with different time binnings
- Correlations less than the threshold are discard

Time bin size Δt	Number of bins N	Threshold $1/\sqrt{N}$
500	440	0.048
1000	220	0.067
2000	110	0.095
4000	55	0.13
5000	44	0.15

$$N = \frac{t_f - t_i}{\Delta t}$$

$$t_f - t_i = 224811 - 4811 = 220{,}000$$

AAPL on 2012/6/21 9:36 ~ 4:00



Δ t

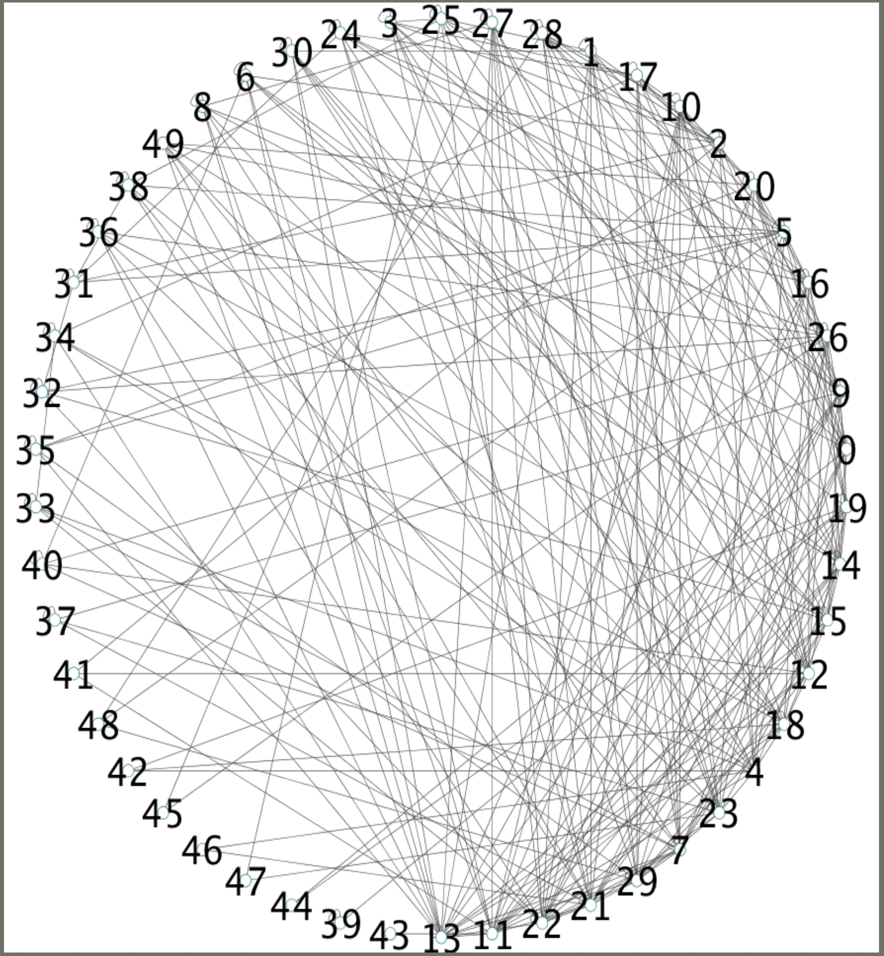
a: 500

b: 1000

c: 2000

d: 5000

- Trend: Increasing bin size increases the number of correlations among the nodes (after threshold correction)
- Not easy to see any structure from the correlation matrix
- → try using minimum spanning tree

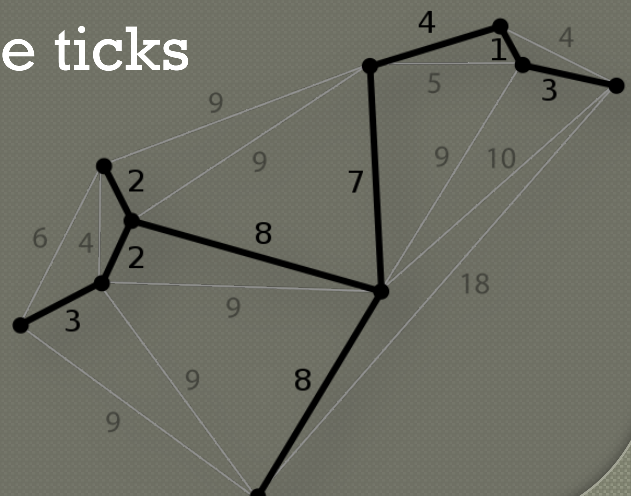


Correlation network of the first 50 best ask (bin size = 2000)

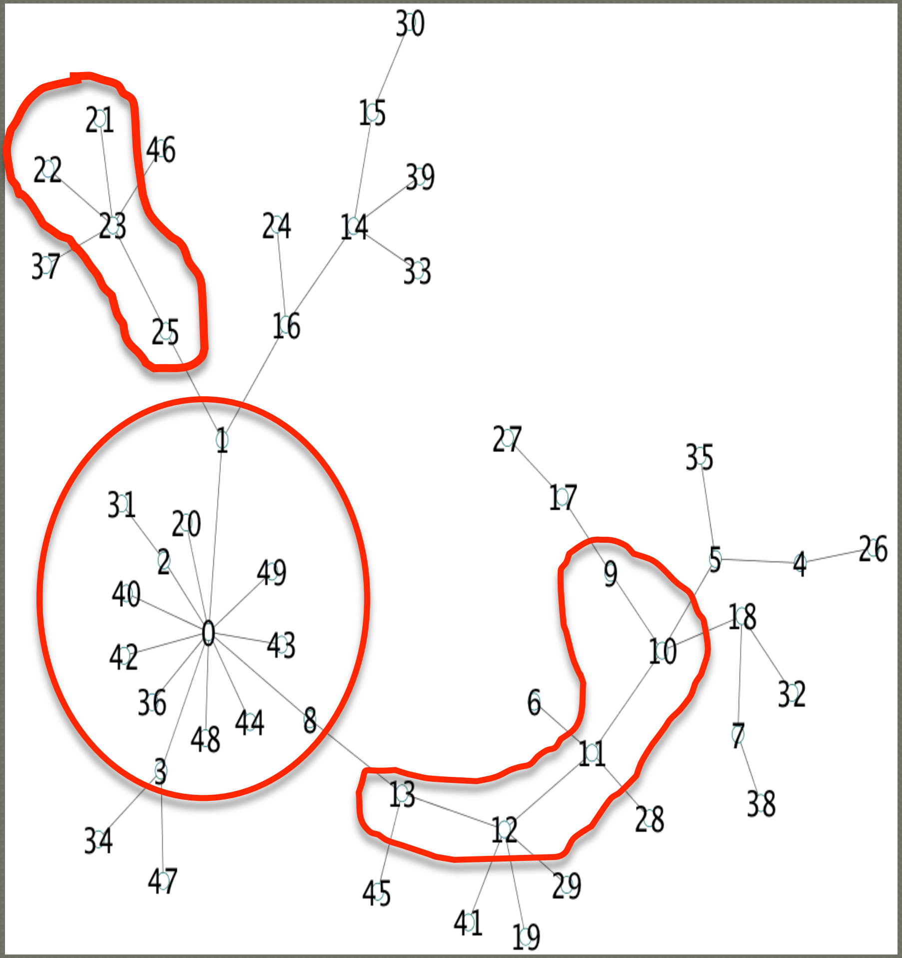
Minimum Spanning Tree (MST)

- **A minimum spanning tree** is a spanning tree of a connected, undirected graph. It connects all the vertices together with the minimal total weighting for its edges.
- In general a graph does not have a unique MST
- In our case, nodes = best (50) price ticks,
  links = magnitude of correlation between the number of orders of the price ticks

$$d_{ij} = \sqrt{2(1 - \rho_{ij})}$$

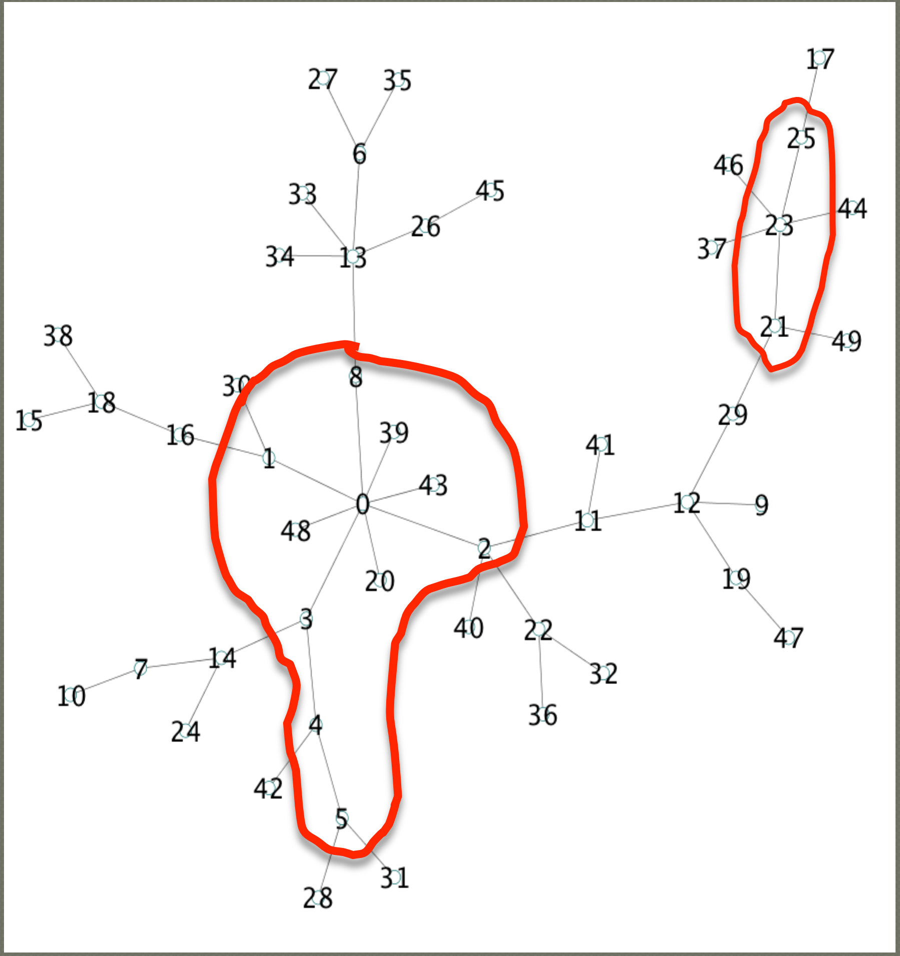


MST of the best 50 price ticks (bin=500)



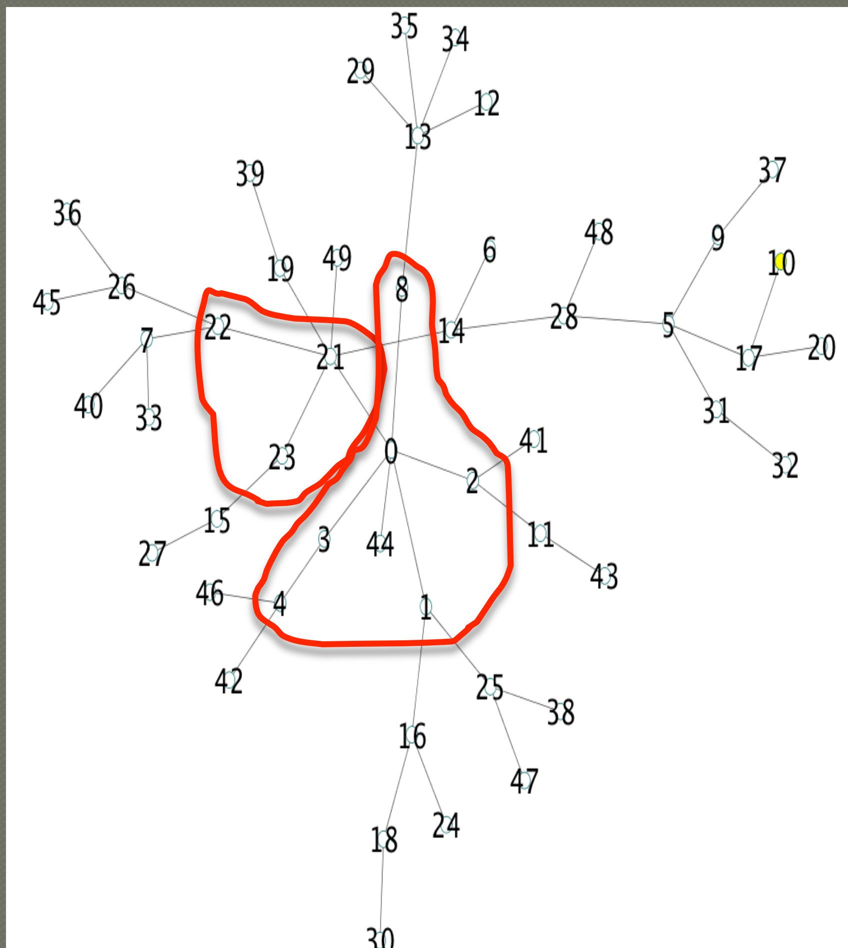
- Major clusters:
  - (0,1,2,3,…+high prices)
  - (21,22,23,25)
  - (9,10,11,12,13)
- First few prices are clustered together:
  - 0,1,2,3

MST of the best 50 price ticks (bin=1000)



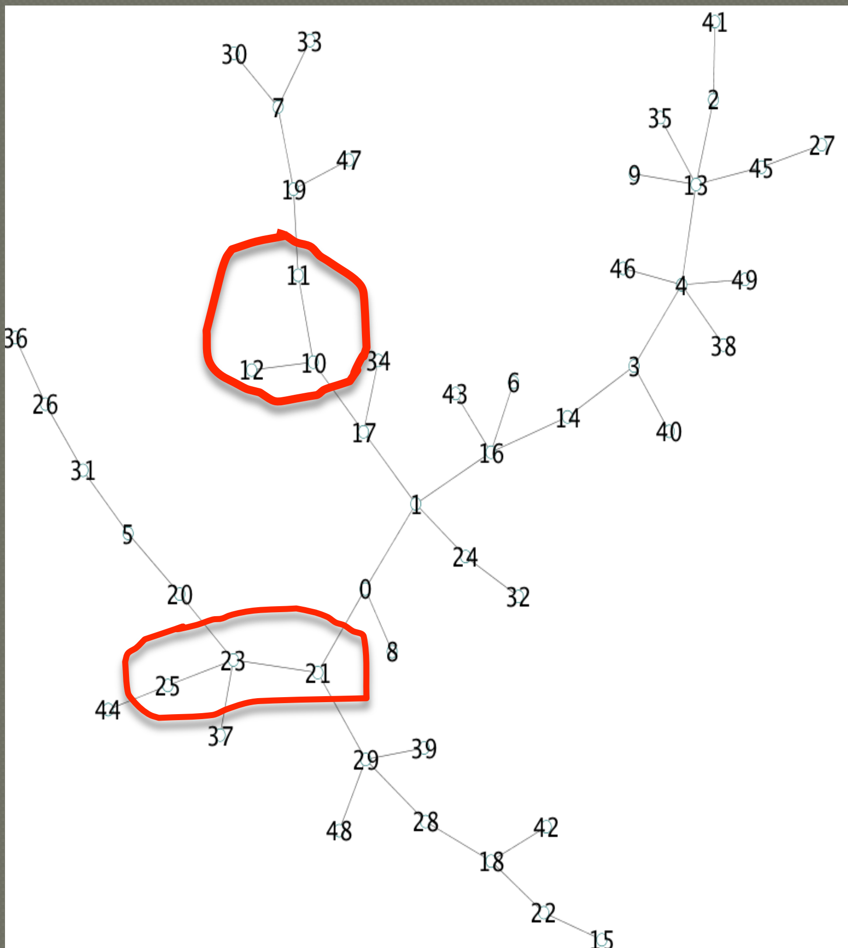
- Major clusters:
  - (0,1,2,3,4,5…+high prices)
  - (21,23,25)
- 0,1,2,3 are still clustered together

MST of the best 50 price ticks (bin=2000)



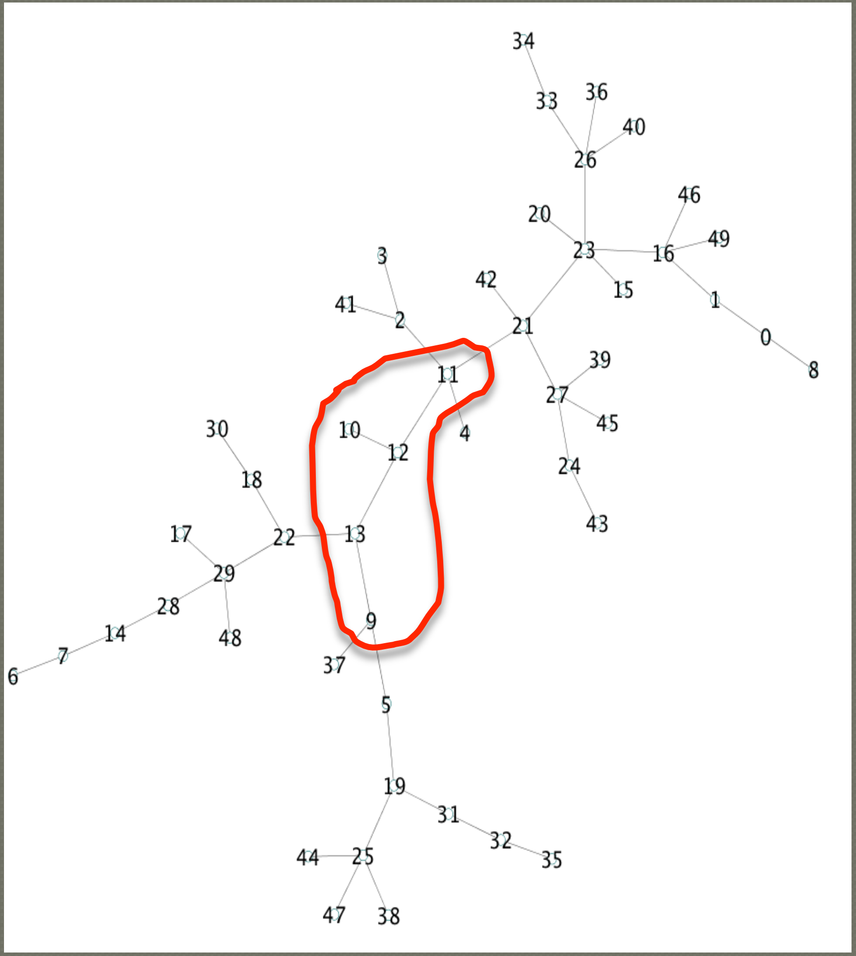
- Major clusters:
  - (0,1,2,3,4,…)
  - (21,22,23)
- 0,1,2,3 are still clustered together

MST of the best 50 price ticks (bin=3000)



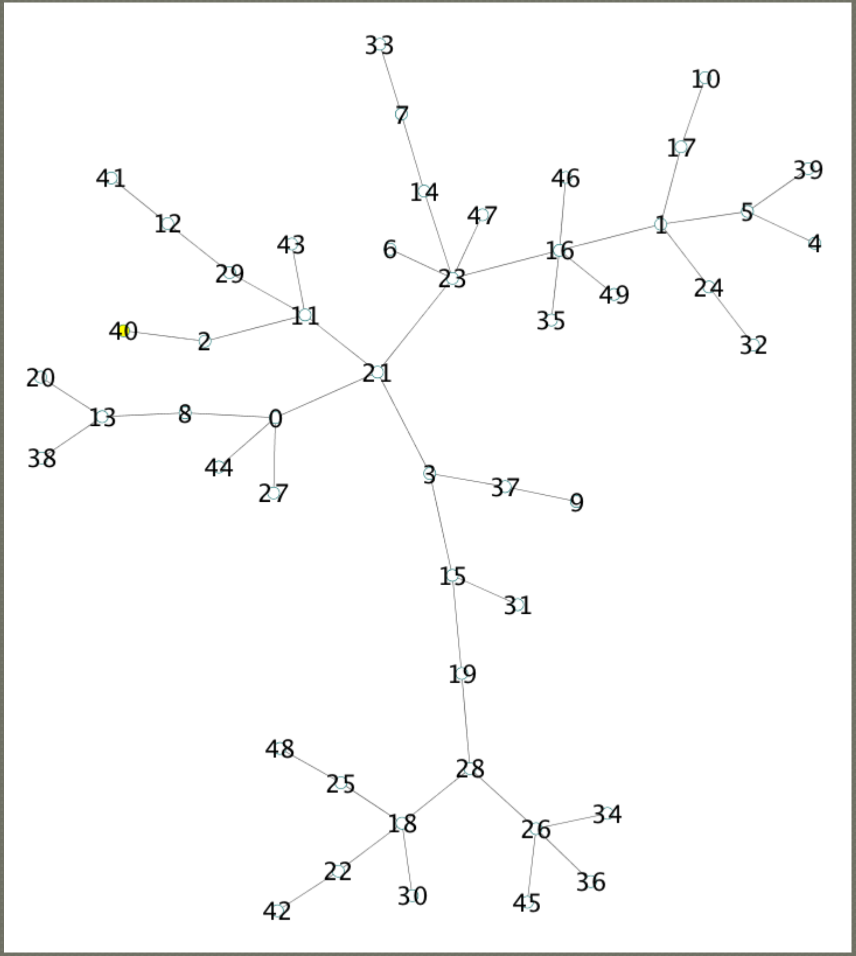
- Major cluster:
  - (21,23,25)
  - (10,11,12)
- First few best prices no long cluster together

MST of the best 50 price ticks (bin=4000)



- Major clusters:
  - (9,10,11,12,13)
- First few best prices not clustering together

MST of the best 50 price ticks (bin=5000)



- No obvious trivial clusters

- Some binning shows that similar prices are clustered together
- High prices are usually the leaves of the MST
- There are some prominent (robust) clusters for different binning.
- Other clusters do not seem to have a robust structure, this may be due to the noise in the data as the nodes in the clusters are not showing genuine correlations

Conclusion

- Using MST, we found clusters of price levels, corresponding to regions in the LOB that shows correlated behavior.
- The clusters are among the first few best prices, and among some regions in higher price levels
- The first few best prices are also seen to be correlated to other high price levels. This may suggest that the first few best prices are more important in the sense that investors may use them to assess the demand and supply in the market and they make decisions (place/cancel orders) based on the the few important prices.
- Interesting questions:

Is the correlation due to day traders (human) or algorithmic traders (computer)? Like human trader, computers are programmed to sense the chance in supply and demand in the market and put/cancel orders accordingly.

Thank you!

Special thanks:
Prof Eugene Stanley
Dr. Jianxi Gao