Interactive Visualization of the Stock Market Graph

Presented by Camilo Rostoker
rostokec@cs.ubc.ca
Department of Computer Science
University of British Columbia

Overview

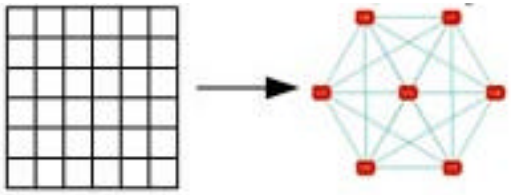
1. Introduction
2. The Market Graph
3. Motivation
4. Visualization Goals
5. Solutions & Methods
6. Future Work
7. Conclusion
8. Demo

Stock Market Data

- Huge amounts of accessible data on a daily basis
- Consists of a variety of fields such as price, volume, change
- Stock price interactions form a complex system
- Want to understand these interactions of the subsystems

Constructing the Market Graph

1. From a dataset, compute the correlation matrix

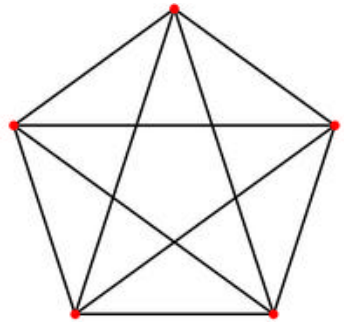


2. Convert correlation matrix to a graph, where
   - Vertices represent stocks
   - Edges represents a relationships between two stocks

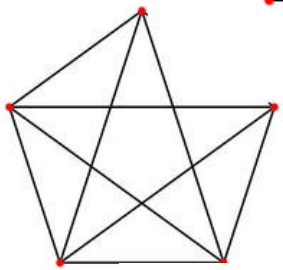
correlation(stock1,stock2) > THRESHOLD

What Are We Visualizing?

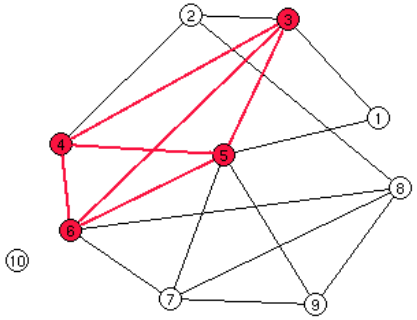
- Find clusters/groups of stocks that exhibit certain trading patterns



- Maximum Cliques
  - Highly positively/negatively correlated subsets of stocks
- Independent Sets
  - Completely diversified stocks



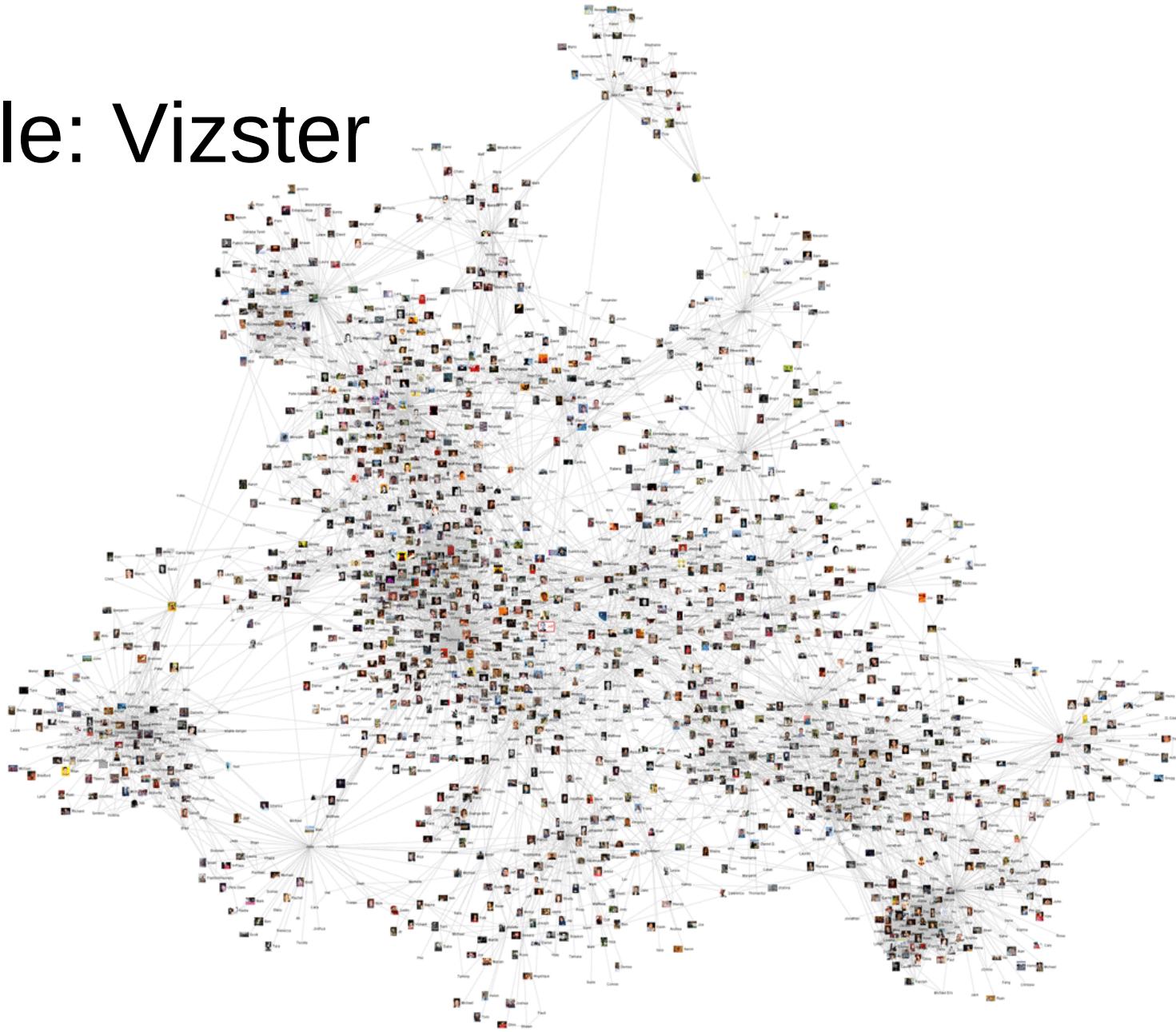
- Quasi-Cliques/Independent Sets
  - Generalizations → allow for near matches
- Clusters ? Cliques/IS interchangeably



Existing Approaches to Visualizing Graph Structures

1. Determine target structures (i.e. clusters) *a priori* and use a standard layout algorithm to show the results
2. Use a layout algorithm optimized to visually differentiate target structures
3. Our approach: combine the two
   - Find target structures first, but include additional nodes and edges for context
   - Then use force-directed layout algorithm to effectively visualize the results

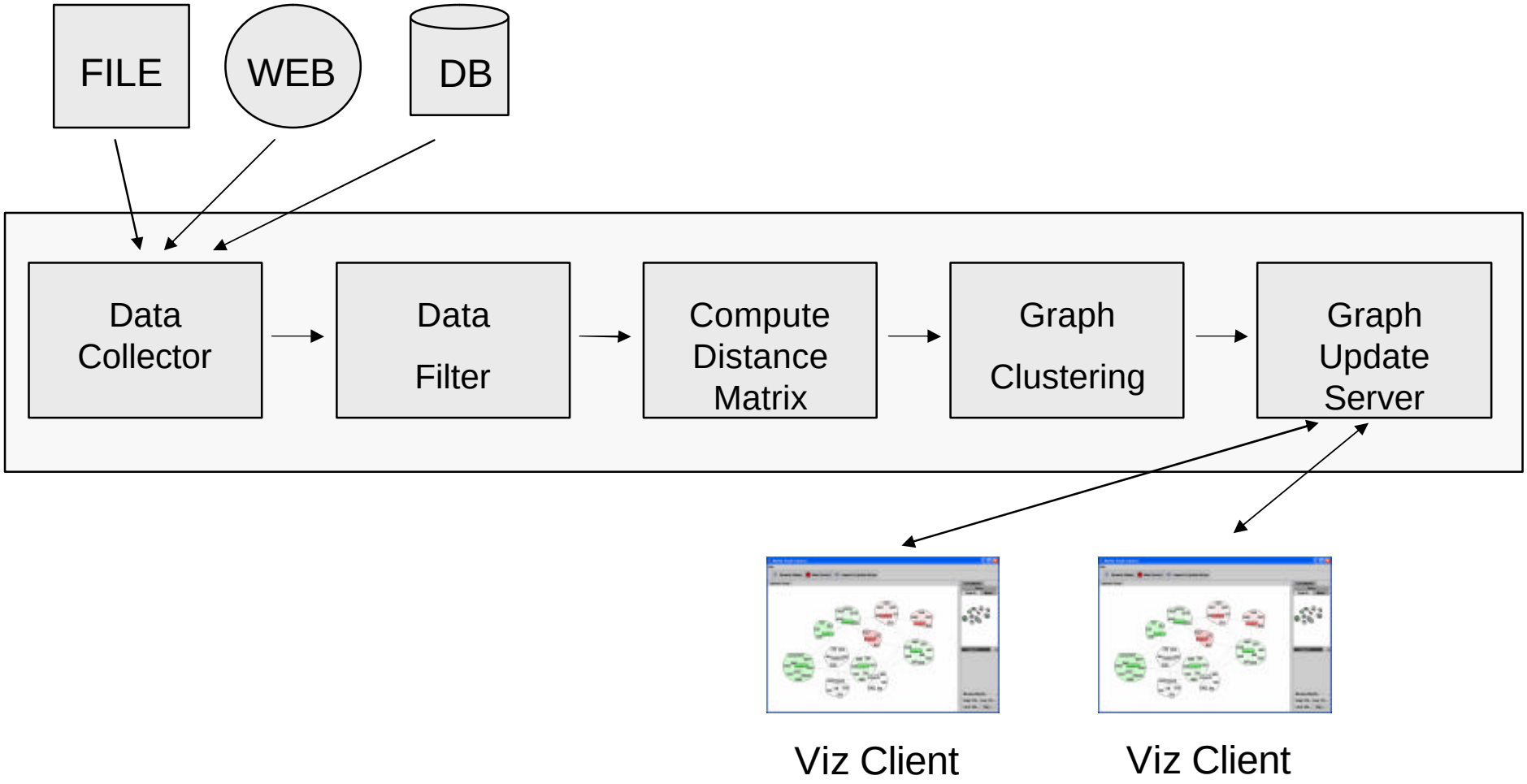
Example: Vizster



Motivation: Usage Scenarios

- Portfolio management (static)
- Real-time market analysis (dynamic)
- Exploratory analysis of trading data to gain new insights, spot patterns/trends, etc (static)

Motivation: Visualizing Results from a Real-time Data-mining Pipeline



Visualization Goals

1. Visualize different graph structures representing various *patterns* and *trends*
   - (quasi-)cliques and (quasi-)independent sets
   - positively and negatively correlations
2. Represent inter-cluster relationships
3. Dynamic graph capabilities
4. Interaction for efficient data exploration
5. Information integration

Force-Directed Graph Layout Model

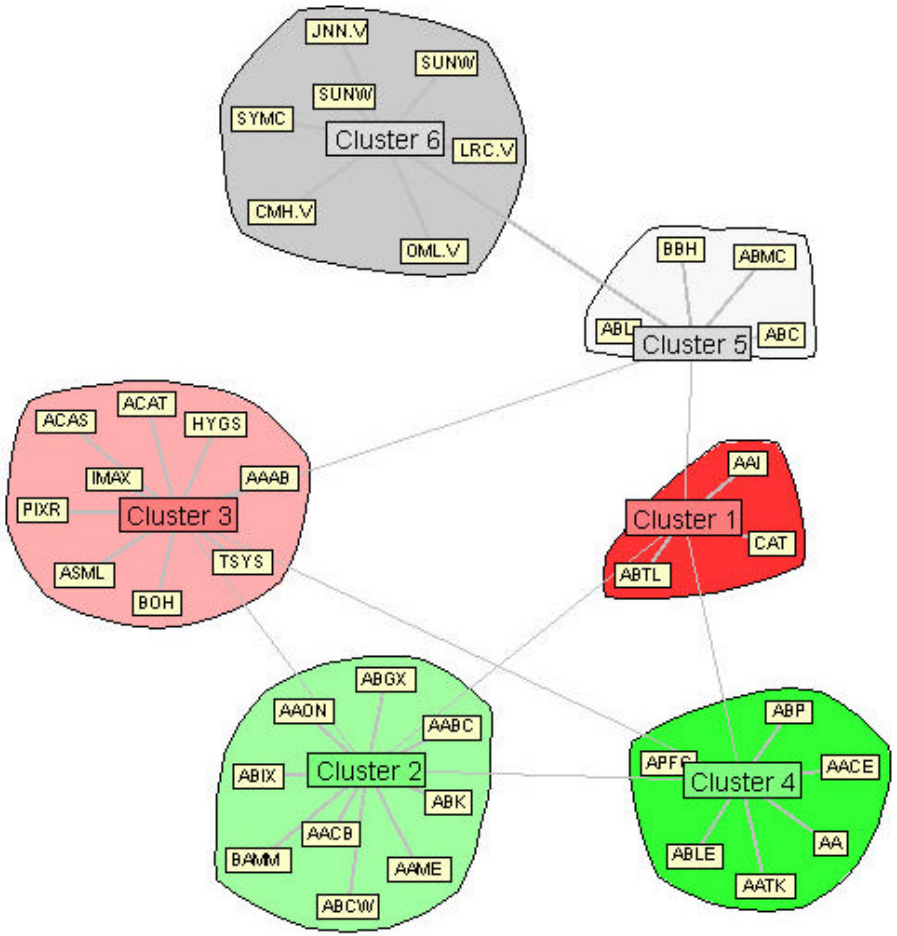
- Create “summaries” of the graph using the clusters and their induced subgraphs
- Force model: spring-embedded layout
- Spring lengths and tensions parameterized to optimize layout
  - Highly related clusters should be close
  - Independent clusters and minimally related clusters should be further apart

F.D. Model Parameterizations

- Edge Length
  - Cluster-Cluster edges (CC)
    - # intra-cluster edges (shows “connectedness” of clusters)
  - Cluster-Member edges (CM)
    - Quasi-cliques → # intra-cluster edges (“clique contribution”)
    - Cliques: cluster sizes (more space to larger clusters)
- Tension
  - CM edges use constant “tight” tension
  - CC edge tension proportional to # of inter-cluster links

Differentiating cluster types

- Correlation Metrics: positive negative, independent
  - Color encoded
- Cluster types: (quasi-) Cliques and (quasi-) Independent sets
  - Transparency-encoding for cluster summary
  - Individual members edge length encodes “clique contribution”



Interaction & Information Integration

- Interaction Features
  - Geometric pan/zoom
  - Display/hide cluster outlines
  - Symbol search for quick navigation
  - Overview display for global context
- Node context menus provide stock quotes and news:
  - Stock news from various sources integrated via RSS feeds
  - Online quote details and Google search for provided by opening an external web browser

Dynamic graph capabilities

- Receive remote graph updates via socket connection to a “graph update server”
- Nodes/edges can be added, removed or replaced
- Event-based architecture allows for automatic processing of new updates
- Force-model allows for efficient incremental layouts when new nodes/edges placed “intelligently”

Future Work & Improvements

- Handle overlapping clusters
- Encode other variables
  - i.e. node size could encode trade volume
- Ability to view underlying edge weights
- Ability to optionally view complete underlying graph
  - especially the intra-cluster edges

Future Work & Improvements (2)

- Interactively adding/removing nodes and edges
- Semantic zoom
- Focus+Context
- Other clustering methods besides partitioning via (quasi-)cliques and independent sets

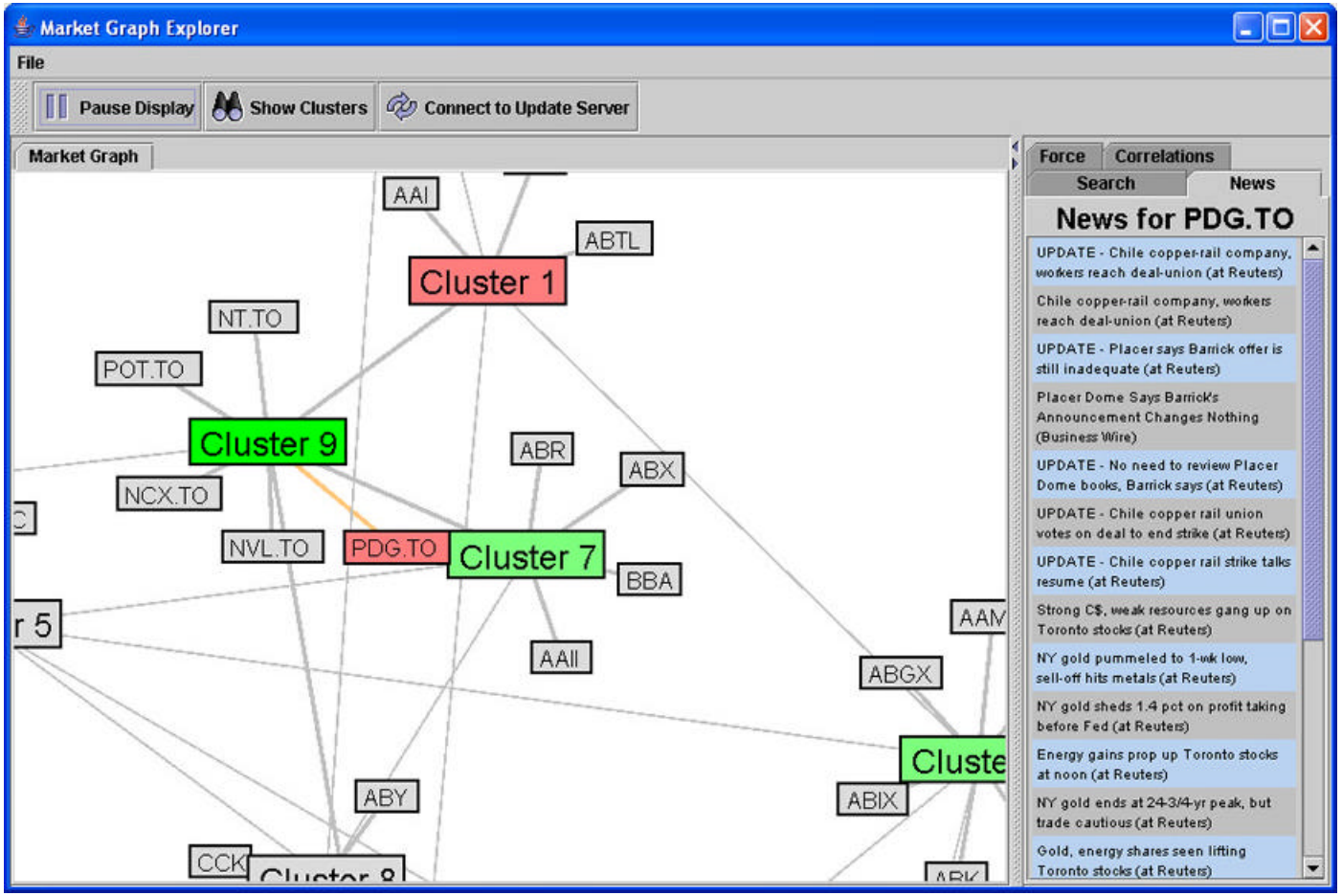
Conclusion

- Implemented basic Visualization tool for exploring the market graph
- Visualizes different cluster types and their attributes
- User interaction for pan/zoom, on-demand details (quotes, news, web search)
- Dynamic graph capability to support a real-time data processing pipeline

References

1. Jeffrey Heer and Danah Boyd. Vizster: **Visualizing online social networks**. *InfoVis 2005 IEEE Symposium on Information Visualization*, 2005.
2. Jeffrey Heer, Stuart K. Card, and James A. Landay. **prefuse: a toolkit for interactive information visualization**. In *CHI '05: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 421–430, New York, NY, USA, 2005. ACM Press.
3. Frank van Ham and Jarke J. van Wijk. **Interactive visualization of small world graphs**. In *Proceedings of the IEEE Symposium on Information Visualization*, pages 199–206, Washington, DC, USA, 2004. IEEE Computer Society.
4. Vladimir Boginski, Sergiy Butenko, and Panos M. Pardalos. **Mining market data**: A network approach.

DEMO


Market Graph Explorer
File
Pause Display
Show Clusters
Connect to Update Server
Market Graph
AAI
ABTL
Cluster 1
NT.TO
POT.TO
Cluster 9
ABR
ABX
NCX.TO
NVL.TO
PDG.TO
Cluster 7
BBA
AAII
ABGX
ABIX
ABY
CCK
Force
Correlations
Search
News
News for PDG.TO
UPDATE - Chile copper-rail company, workers reach deal-union (at Reuters)
Chile copper-rail company, workers reach deal-union (at Reuters)
UPDATE - Placer says Barrick offer is still inadequate (at Reuters)
Placer Dome Says Barrick's Announcement Changes Nothing (Business Wire)
UPDATE - No need to review Placer Dome books, Barrick says (at Reuters)
UPDATE - Chile copper rail union votes on deal to end strike (at Reuters)
UPDATE - Chile copper rail strike talks resume (at Reuters)
Strong C$, weak resources gang up on Toronto stocks (at Reuters)
NY gold pummeled to 1-wk low, sell-off hits metals (at Reuters)
NY gold sheds 1.4 pct on profit taking before Fed (at Reuters)
Energy gains prop up Toronto stocks at noon (at Reuters)
NY gold ends at 24-3/4-yr peak, but trade cautious (at Reuters)
Gold, energy shares seen lifting Toronto stocks (at Reuters)

THE END!

Construct a Similarity Matrix

- Currently, our similarity measure is

$$C_{ij} = \frac{\langle R_i R_j \rangle - \langle R_i \rangle \langle R_j \rangle}{\sqrt{\langle R_i^2 - \langle R_i \rangle^2 \rangle \langle R_j^2 - \langle R_j \rangle^2 \rangle}},$$

where:

$$R_i(t) = \ln P_i(t) / P_i(t-1)$$