Dynamics of the Financial Market

Panos M. Pardalos

Center for Applied Optimization

Dept. of Industrial & Systems Engineering

Affiliated Faculty of:
Computer & Information Science & Engineering Department
Biomedical Engineering Program, McKnight Brain Institute

Massive Datasets

The proliferation of **massive datasets** brings with it a series of special computational challenges. This **data avalanche** arises in a wide range of scientific and commercial applications. With advances in computer and information technologies, many of these challenges are beginning to be addressed.

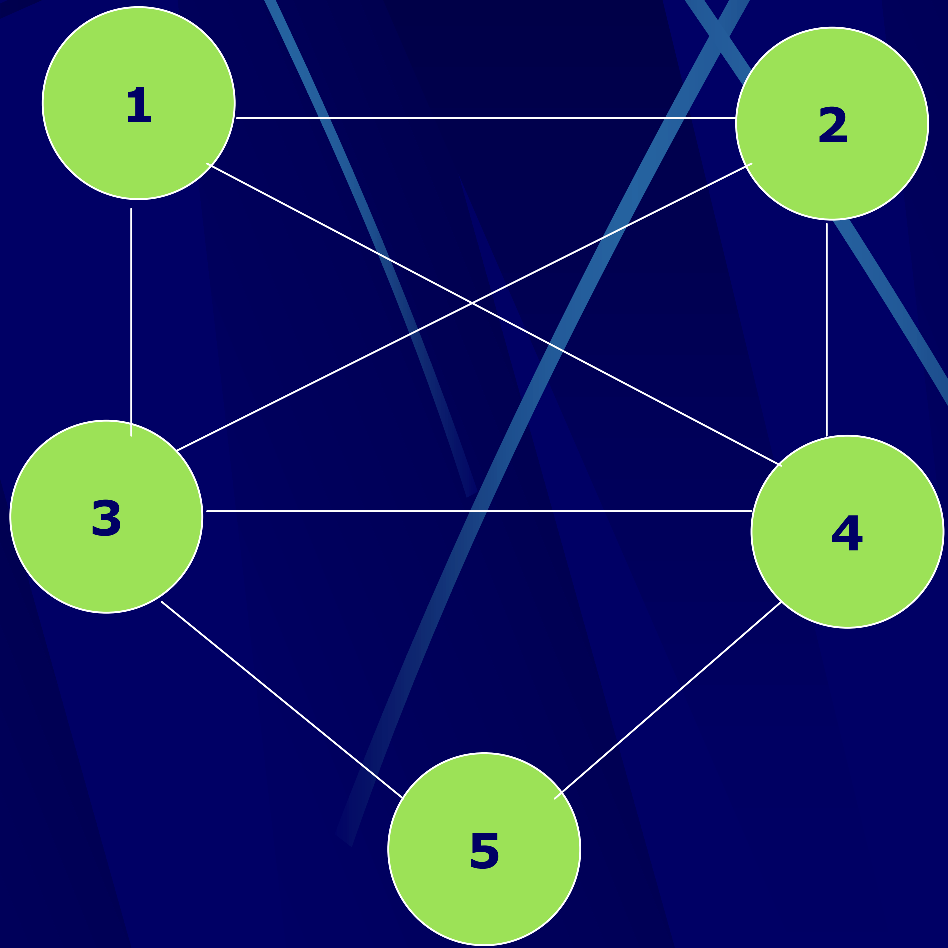
Graph Representation of Massive Datasets

- In many cases, it is convenient to represent a dataset as a ***graph*** (***network***) with certain attributes associated with its vertices and edges
- Studying the properties of these graphs often provides useful information about the internal structure of the datasets they represent

Important Concepts

- A graph $G = (V, E)$, V = set of vertices, E = set of edges
- Degrees of the vertices, degree distribution
- Size of connected components
- Edge density
- *Cliques* and *independent sets*

Example of a graph



Examples of Real-Life Massive Graphs

- ***Web graph*** (links between websites)
- ***Call graph*** (telephone traffic data)
- ***Market graph*** (stock prices data)
- ***Brain networks*** (neurons and connections between them)

Degree Distribution: Power Law

- **Degree distribution** of a graph characterizes ***global statistical patterns*** underlying the dataset this graph represents
- Interestingly, the degree distribution of all considered real-life graphs has a well-defined ***power-law*** structure:

The probability that a vertex has a degree *k* (i.e., *k* neighbors) is

$P(k) \propto k^{-\gamma}.$ or $\log P \propto -\gamma \log k$

("Self-organized" networks)

Cliques and Independent Sets

- A ***clique*** is a subgraph of G that has *all possible edges*
- Cliques represent dense clusters of ***"similar" objects***
- An independent set is a subgraph of G with *no edges*.
- Independent sets represent groups of ***"different" objects***

Maximum Clique and Independent Set Problems

- The subject of a special interest is to find the *maximum* clique and independent set in the graph
- Maximum clique and maximum independent set problems can be transformed to each other, using the notion of *complementary graph*
- These problems are NP-hard

Finding cliques and independent sets

- ***Heuristic algorithms*** (no guarantee to find an optimal solution)
- ***Exact algorithms*** (finding *maximum* clique or independent set)

Clique Partitioning

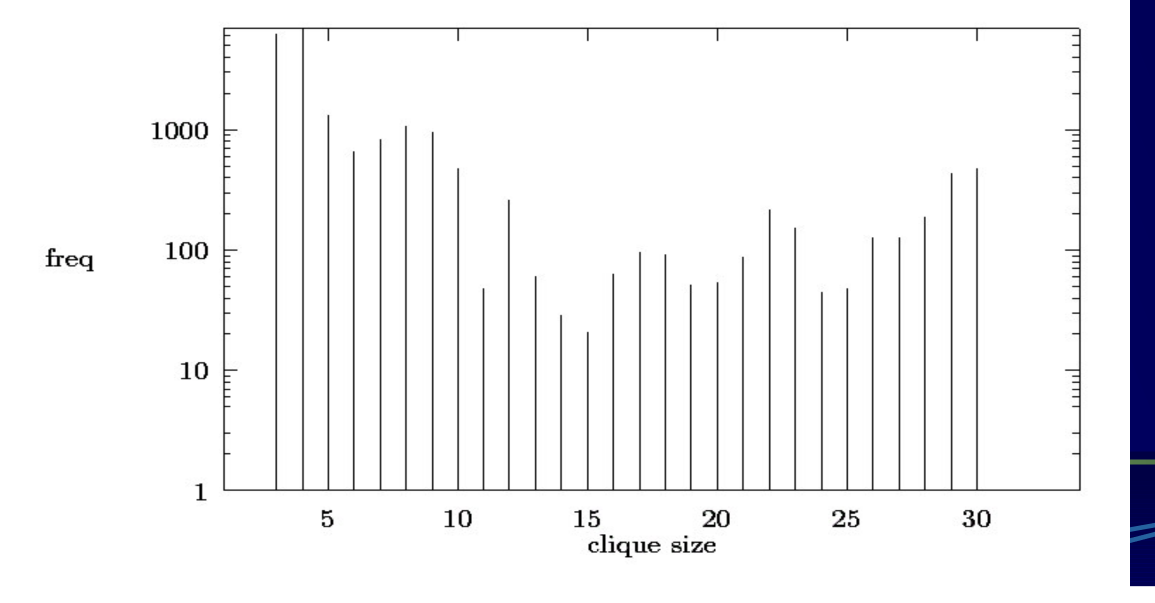
- ***Minimum clique partition***: dividing the graph into a minimum number of distinct cliques
- This provides a natural way of partitioning a dataset represented by a graph into a number of clusters of “similar” objects (*clustering problem*), where the number of clusters is the minimum number of cliques in the graph

Graph Coloring

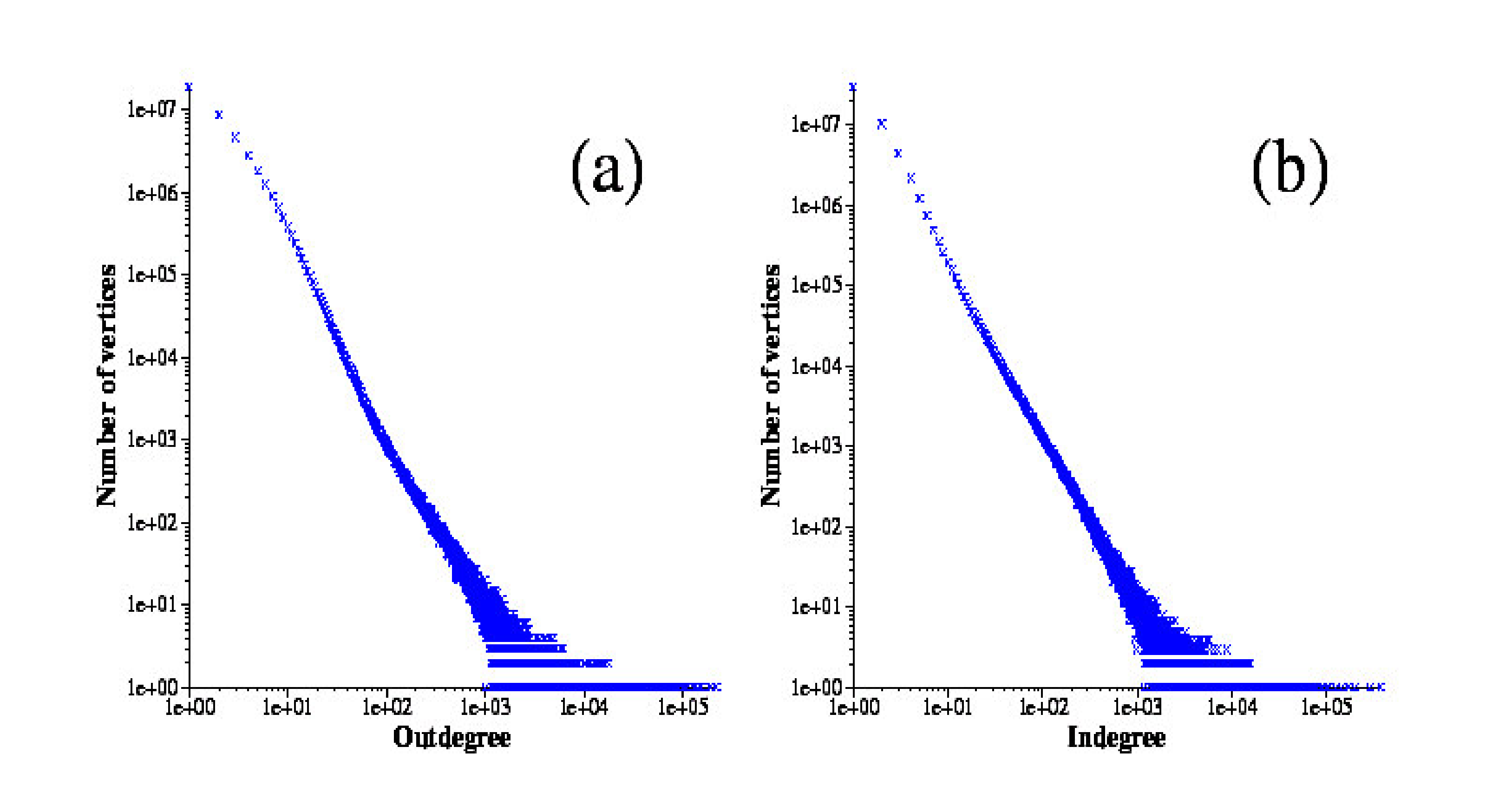
- Coloring essentially represents the partitioning of the graph into a minimum number of *independent sets*
- Partitioning a dataset represented by a graph into a number of clusters of “different” objects

Call graph

- 53,000,000 vertices, 170,000,000 edges
- 100,000 iterations of ***GRASP***
- largest clique size = 32 (Abello, Pardalos & Resende, 1999)



Degree Distribution of the Call Graph (data by AT&T)



Market Graph

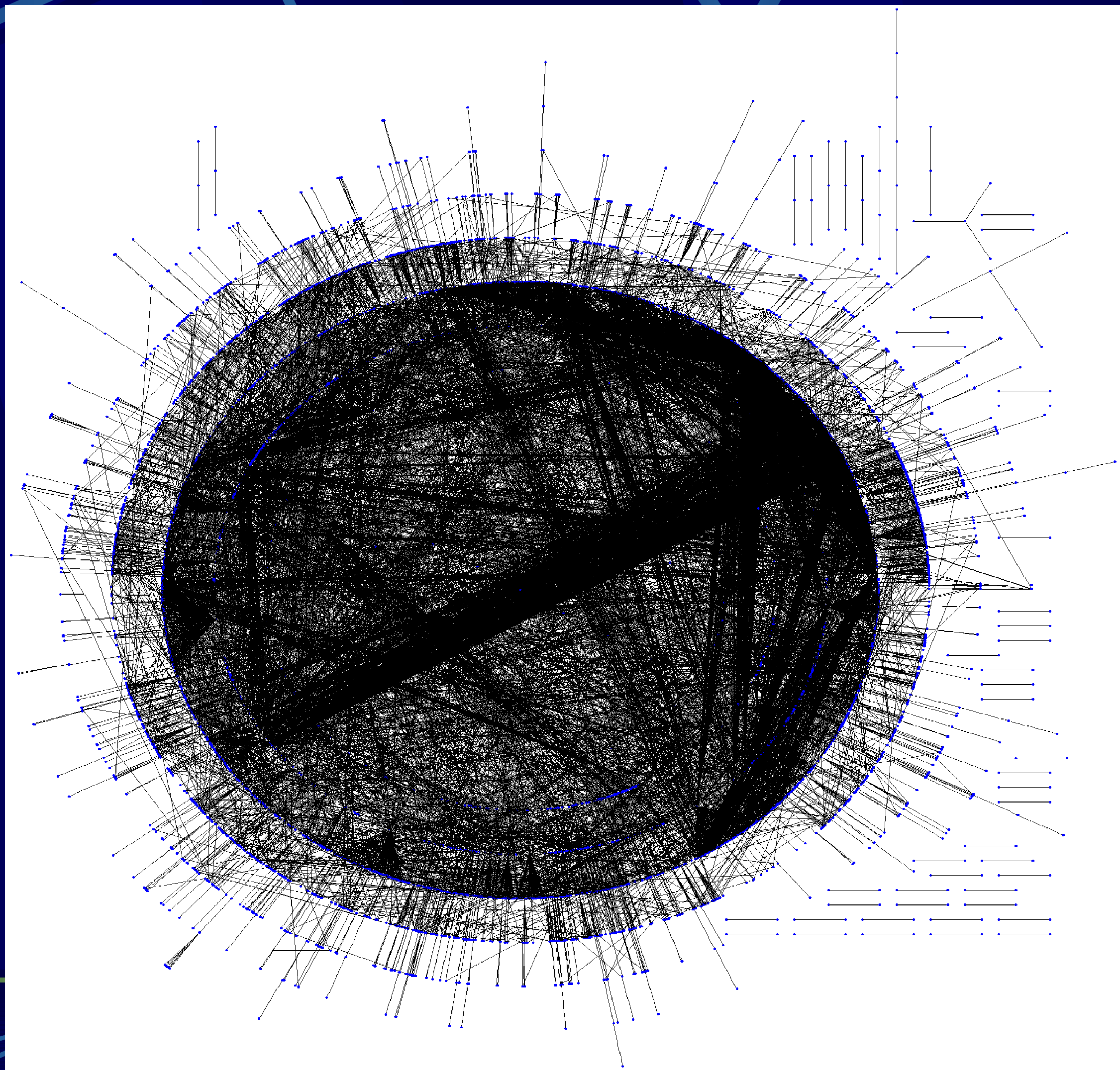
- Vertices are stocks, and an edge connects two stocks if the correlation between their price fluctuations over a certain period is greater than a specified threshold
- ~6000 vertices (stocks)

Market Graph

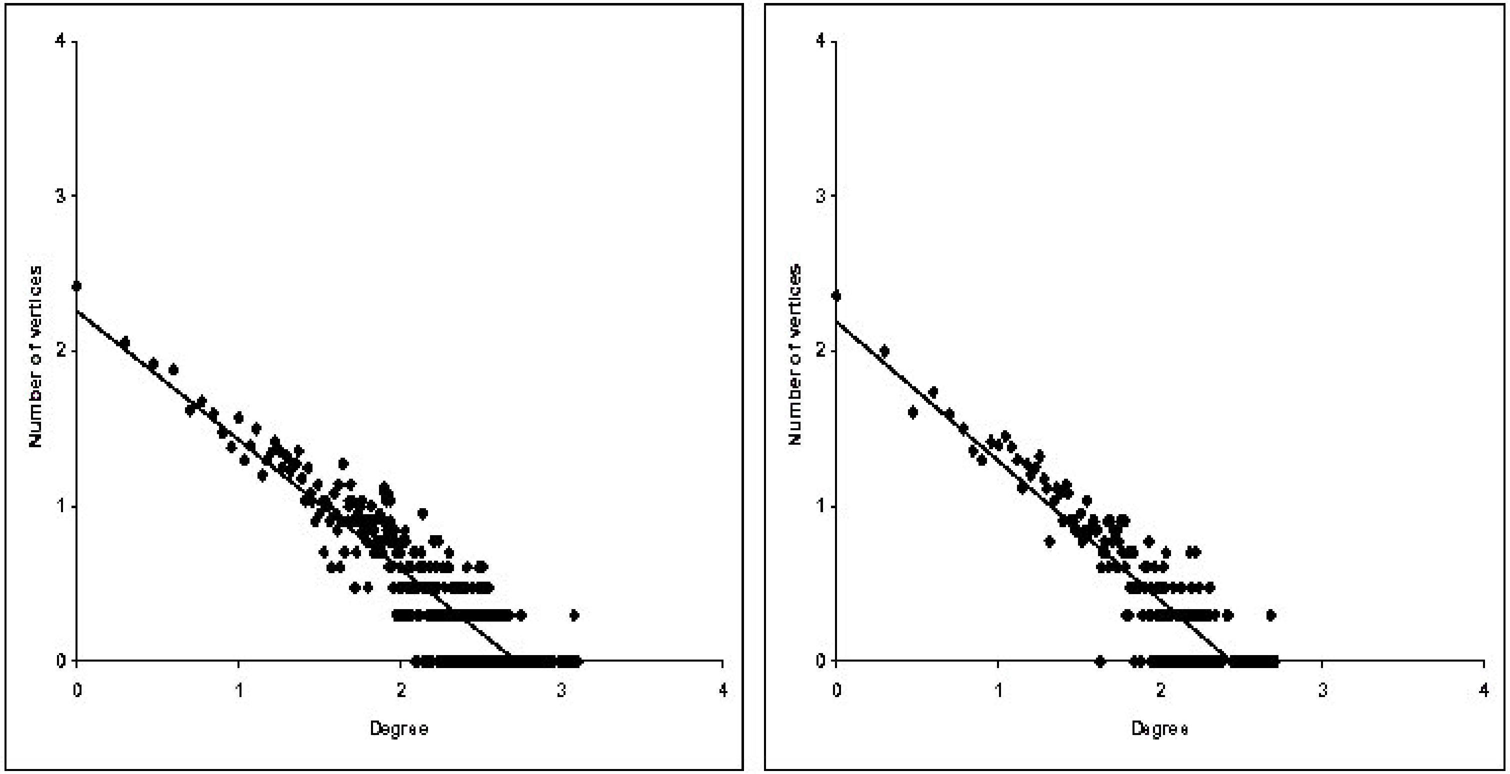
- Market graph (all the considered instances for different correlation thresholds) follows the *power-law* model
- Using the combination of heuristic and exact algorithms, the exact solution of the maximum clique problem was found

(**Boginski, Butenko & Pardalos, 2003**)

Market Graph



Degree distribution of the Market graph



Finding Cliques in the Market Graph

- Applying a heuristic algorithm to find a large clique: let ***N(i)*** be the set of neighbors of the vertex ***i***

$C = \emptyset, G_0 = G;$

do

$$G_0 = \bigcap_{i \in C} N(i) \setminus C;$$

$C = C \bigcup j$, where j is a vertex of largest degree in G_0;

until $G_0 = \emptyset.$

Finding Cliques in the Market graph

- Preprocessing procedure:

  C is the clique found by the heuristic algorithm: recursively remove from the graph all of the vertices which are not in C and whose degree is less than |C|

  Denote the resulting (reduced) graph as G' = (V', E')

Finding Cliques in the Market graph

- Using the IP formulation of the maximum clique problem to find the *exact* solution:

$$\text{maximize} \quad \sum x_i$$

$$s.t.$$

$$x_i + x_j \leq 1, (i,j) \notin E'$$

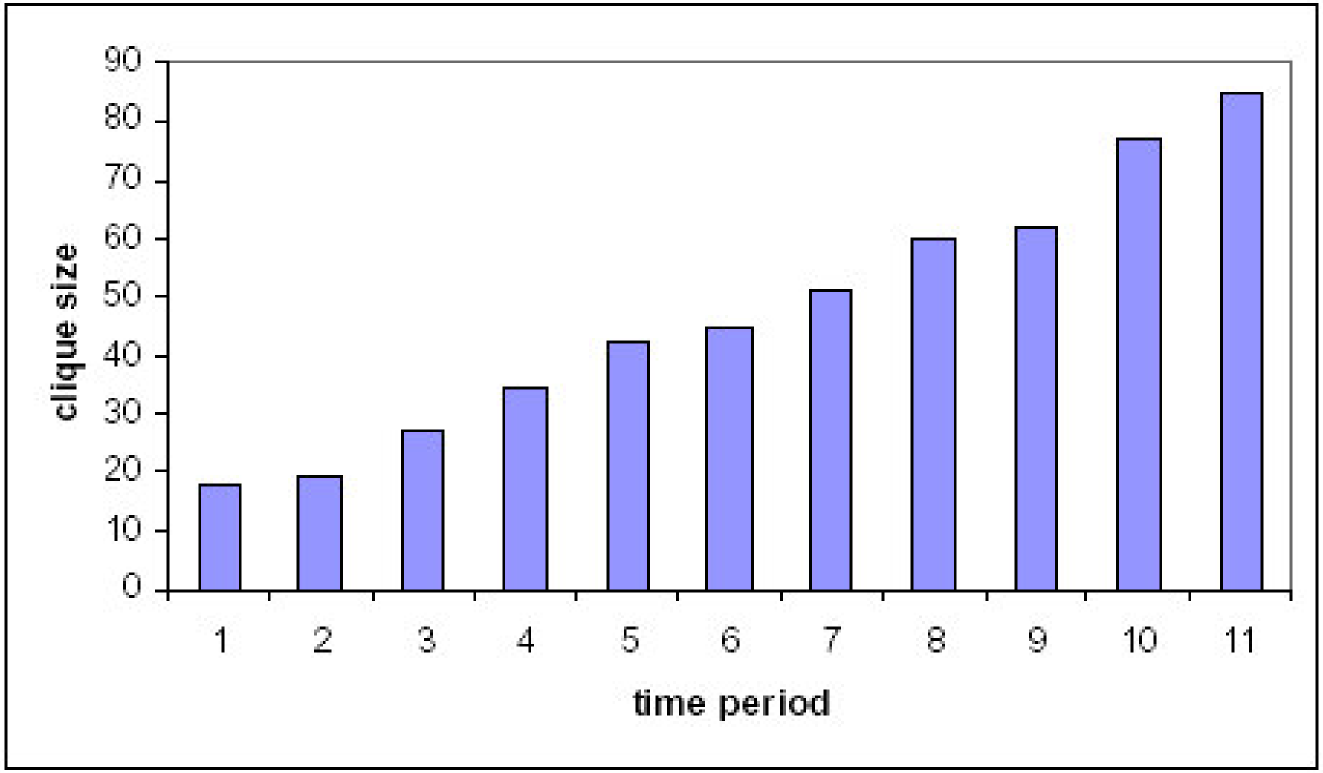
$$x_i \in \{0,1\}$$

Maximum Clique size for different correlation thresholds

- Large cliques despite very low edge density – confirms the idea about the ***"globalization"*** of the market

θ	edge density	clique size
0.35	0.0090	193
0.4	0.0047	144
0.45	0.0024	109
0.5	0.0013	85
0.55	0.0007	63
0.6	0.0004	45
0.65	0.0002	27
0.7	0.0001	22

Maximum Clique Size Trends (4 year data)



Classification of Stocks Using Clique Partitioning

- A clique in the market graph represents a dense cluster of stocks whose prices ***exhibit a similar behavior over time***
- Therefore, dividing the market graph into a set of distinct cliques (***clique partitioning***) is a natural approach to ***classifying stocks*** (dividing the set of stocks into ***clusters of similar objects*** – an approach to solving the clustering problem)

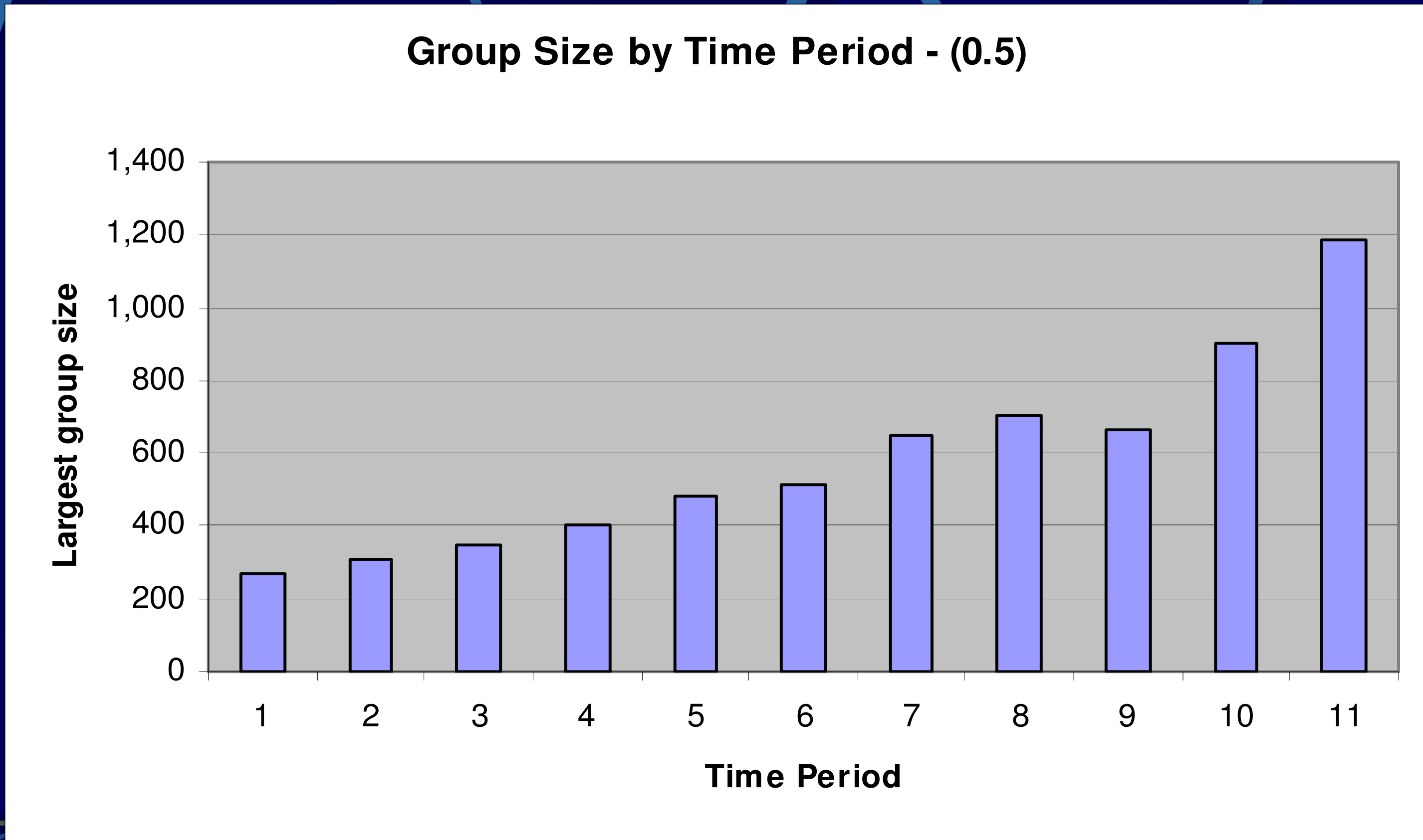
Clustering Stocks Using Connected Components

- Finding connected components is much more computationally efficient than finding cliques
- Distinct connected components correspond to specific industry sectors

Clustering Stocks Using Connected Components

- ***Giant connected component:***
  - Financial Services / High-Tech sectors + Exchange Traded Funds (Indices)
- Smaller connected components:
  - Utilities/Services
  - Basic materials/Energy
  - Gold ore industries
  - Transportation

Giant Connected Component Size Trends (4 year data)



Independent sets in the Market graph

- Maximum independent set represents the largest "***perfectly diversified***" portfolio
- Solving the maximum clique problem in the complementary graph
- The preprocessing procedure could not reduce the size of the initial graph, the exact solution could not be found
- Large diversified portfolios are hard to find

Independent set sizes for different correlation thresholds

- Relatively small independent sets found by the heuristic algorithm

θ	edge density	indep. set size
0.05	0.4794	36
0.0	0.2001	12
-0.05	0.0431	5
-0.1	0.005	3
-0.15	0.0005	2

Independent Sets in the Market Graph

- Finding a perfectly diversified portfolio containing ***any given stock***
- For every vertex in the market graph, an independent set that contains this vertex was detected, and the sizes of these independent sets were almost the same, which means that it is possible to find a diversified portfolio containing any given stock using the market graph methodology

Social Networks

- Market graph may be considered an example of social network created by interaction of decisions of market participants
- There is a number of other social networks observed with degree distribution satisfying the power law

Social Networks

- Network of mutual acquaintances
- Authors collaboration network
- Movie actors network
- Disease network

Ongoing Research

- Studying the ***long-term*** trends and behavior of the structural characteristics of the market graph (e.g., over a century)
- Relating this behavior to major events and social mood trends

Summary

- There are many mathematical techniques for addressing data mining problems in massive data sets
- Network-based techniques for financial applications is a promising research area allowing us to address many needs of financial data analysis

References

- V. Boginski, S. Butenko, and P.M. Pardalos. On Structural Properties of the Market Graph. In: A. Nagurney (editor), Innovations in Financial and Economic Networks, Edward Elgar Publishers, 28-45 , 2003.
- V. Boginski, S. Butenko and P. M. Pardalos. Statistical Analysis of Financial Networks. Computational Statistics and Data Analysis, 48(2), 431-443, 2005.
- V. Boginski, S. Butenko, and P.M. Pardalos. Mining Market Data: A Network Approach. Computers and Operations Research, 2005, in press.
- A. Arulselvan, V. Boginski, A. Kammerdiner, P.M. Pardalos. Analysis of Stock Market Structure by Identifying Connected Components in the Market Graph. Journal of Financial Decision Making, 1(1), 27-37, 2005.

Information

- **Center for Applied Optimization** (Industrial and Systems Engineering Dept., University of Florida)

  www.ise.ufl.edu/cao