Data Science for Massive (Dynamic) Networks

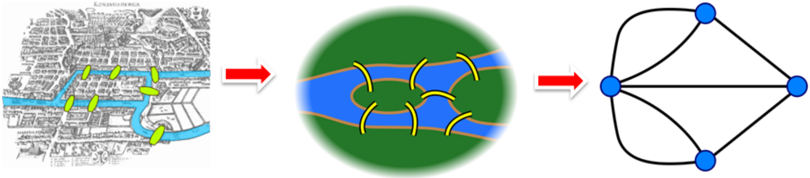
Panos M. Pardalos

Distinguished Professor
Center For Applied Optimization,
Industrial and Systems Engineering, University of Florida,
Florida, USA.
www.ise.ufl.edu/pardalos
National Research University Higher School of Economics,
Laboratory of Algorithms and Technologies for Network
Analysis, Nizhny Novgorod, Russia
http://nnov.hse.ru/en/latna/

Network Analysis History

Network analysis originated many years ago.

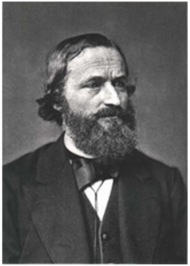
- In the 18th century Euler solved the famous Königsberg bridge problem.
- Euler's solution is considered to be the first theorem of network analysis and graph theory.

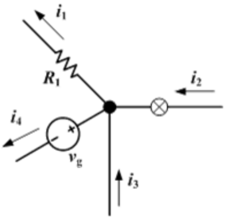


Network Analysis History

Network analysis originated many years ago.

- In the 19th century Gustav Kirchhoff initiated the theory of electrical networks.
- Kirchhoff was the first person who defined the flow conservation equations, one of the milestones of network flow theory.
- After the invention of the telephone by Alexander Graham Bell in the 19th century the resulting applications gave the network analysis a great stimulus.





Network Analysis History

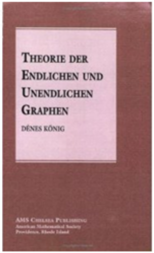
Network analysis originated many years ago.

- In the 19th century Gustav Kirchhoff initiated the theory of electrical networks.
- Kirchhoff was the first person who defined the flow conservation equations, one of the milestones of network flow theory.
- After the invention of the telephone by Alexander Graham Bell in the 19th century the resulting applications gave the network analysis a great stimulus.

Network Analysis History

The field evolved dramatically after the 19th century.

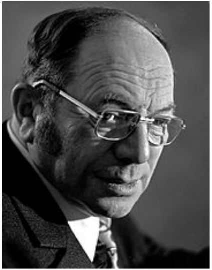
- The first graph theory book was written by Dénes König in 1936.
- As in many other fields, WWII played a crucial role in the development of the field.
- Many algorithms and techniques were developed to solve logistic problems from the military.
- After the war, these technological advances were applied successfully in other fields.



The field evolved dramatically after the 19th century.

- The first graph theory book was written by Dénes König in 1936.
- As in many other fields, WWII played a crucial role in the development of the field.
- Many algorithms and techniques were developed to solve logistic problems from the military.
- After the war, these technological advances were applied successfully in other fields.

The field evolved dramatically after the 19th century.

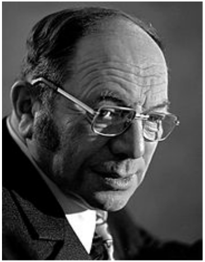


- The earliest linear programming model was developed by Leonid Kantorovich in 1939 during World War II, to plan expenditures to reduce the costs of the army.
- In 1940, also during World War II, Tjalling Koopmans formulated also linear optimization models to select shipping routes to send commodities from America, to Specific destinations in England.
- For their work in the theory of optimum allocation of resources, these two researchers were awarded with the Nobel price in Economics in 1975.

Network Analysis History

The field evolved dramatically after the 19th century.

- The earliest linear programming model was developed by Leonid Kantorovich in 1939 during World War II, to plan expenditures to reduce the costs of the army.
- In 1940, also during World War II, Tjalling Koopmans formulated also linear optimization models to select shipping routes to send commodities from America, to Specific destinations in England.
- For their work in the theory of optimum allocation of resources, these two researchers were awarded with the Nobel price in Economics in 1975.

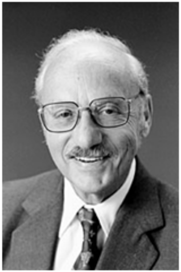


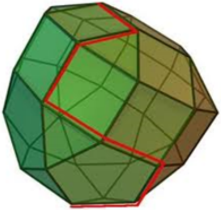
The field evolved dramatically after the 19th century.

- The first complete algorithm for solving the transportation problem was proposed by Frank Lauren Hitchcock in 1941.
- With the development of the Simplex Method for solving linear programs by George B. Dantzig in 1957, a new framework for solving several network problems became available.
- The network simplex algorithm is still in practice one of the most efficient algorithms for solving network flow problems.

The field evolved dramatically after the 19th century.

- The first complete algorithm for solving the transportation problem was proposed by Frank Lauren Hitchcock in 1941.
- With the development of the Simplex Method for solving linear programs by George B. Dantzig in 1957, a new framework for solving several network problems became available.
- The network simplex algorithm is still in practice one of the most efficient algorithms for solving network flow problems.

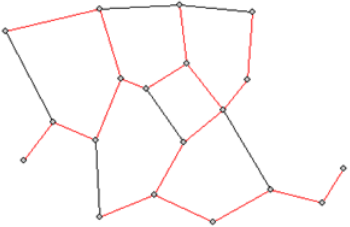




Network Analysis History

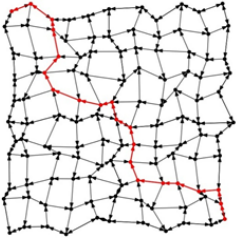
The field evolved dramatically after the 19th century.

- Many other authors proposed efficient algorithms for solving different network problems.
- Joseph Kruskal in 1956 and Robert C. Prim in 1957 developed algorithms for solving the minimum spanning tree problem.
- In 1956 Edsger W. Dijkstra developed his algorithm for solving the shortest path problem, one of the most recognized algorithms in network analysis.



The field evolved dramatically after the 19th century.

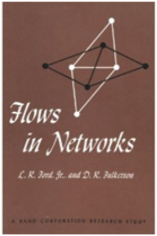
- Many other authors proposed efficient algorithms for solving different network problems.
- Joseph Kruskal in 1956 and Robert C. Prim in 1957 developed algorithms for solving the minimum spanning tree problem.
- In 1956 Edsger W. Dijkstra developed his algorithm for solving the shortest path problem, one of the most recognized algorithms in network analysis.



The field evolved dramatically after the 19th century.

- As it happened in other fields, computer science and networks influenced each other in many aspects.
- In 1963 the book by Lester R. Ford and Delbert R. Fulkerson introduced new developments in data structure techniques and computational complexity into the field of networks.

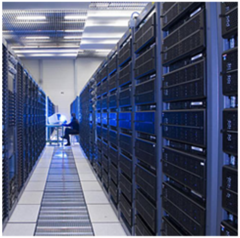




Network Analysis History

In recent years the evolution of computers have changed the field. We are now able to solve large-scale network problems.

- Parallel computing
- Grid computing
- Cloud computing
- Quantum computing

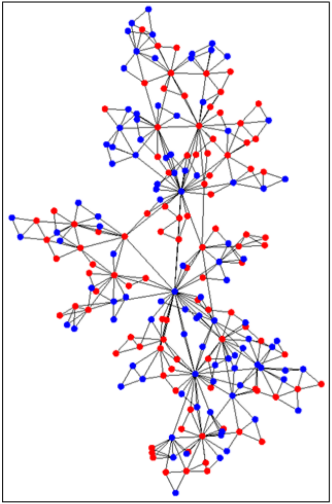


Network Analysis History

Network Analysis has become a major research topic over the last years.

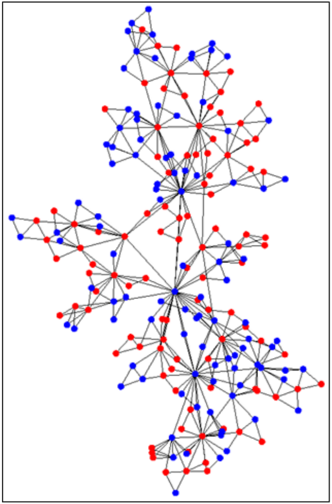
The broad range of applications that can be described and analyzed by means of a network is bringing together researches from numerous fields:

- Operations Research
- Computer Science
- Transportation
- Biomedicine
- Energy
- Social Sciences
- Computational Neuroscience
- Others.

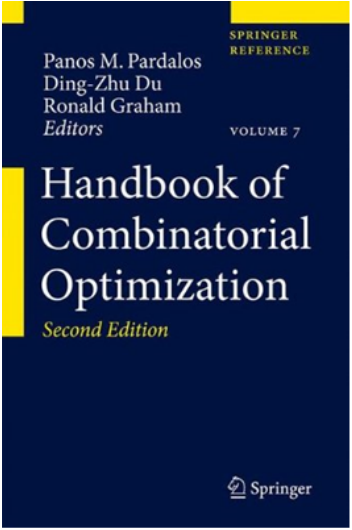


Network Analysis History

- **This remarkable diversity of the fields that take advantage of Network Analysis makes the endeavor of gathering up-to-date material a very useful task.**
- **Since 2011 we organize each every year an international conference on network analysis. The objective of this conference is to initiate joint research among different groups in particular CAO and LATNA.**

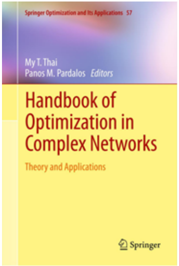


- Handbook of Combinatorial Optimization (2nd Edition). **2013, 4930p.** Panos M. Pardalos, Ding-Zhu Du, and Ronald L. Graham (Eds.) ISBN 978-1-4419-7996-4.

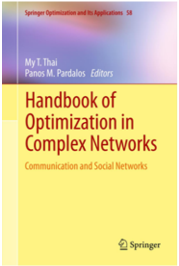


Handbook of Optimization in Complex Networks

- Theory and Applications. My T. Thai and Panos M. Pardalos (Eds.) Springer. Series: Springer Optimization and Its Applications , 2012. Vol. 57. ISBN 978-1-4614-0753-9.

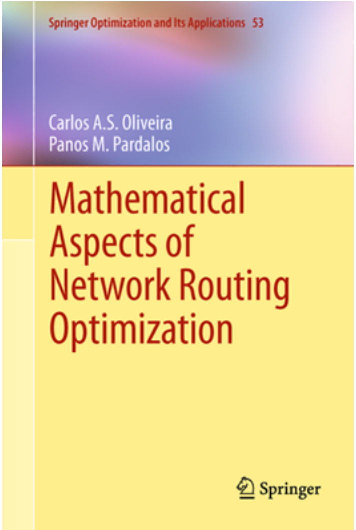


- Communications and Social Networks. My T. Thai and Panos M. Pardalos (Eds.) Springer. Series: Springer Optimization and Its Applications , 2012. Vol. 58. ISBN 978-1-4614-0856-7.



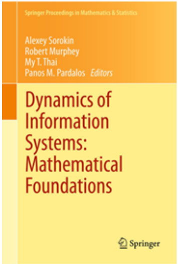
Handbook of Optimization in Complex Networks

- Mathematical Aspects of Network Routing Optimization. Carlos Oliveira and Panos M. Pardalos. Springer. Series: Springer Optimization and Its Applications , 2011. Vol. 53. ISBN 978-1-4614-0310-4.

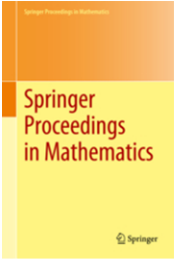


Handbook of Optimization in Complex Networks

- Dynamics of Information Systems: Mathematical Foundations. Alexey Sorokin, Robert Murphey, My T. Thai, and Panos M. Pardalos (Eds.) Springer. Springer Proceedings in Mathematics & Statistics, 2012. Vol. 20. ISBN 978-1-4614-3905-9.

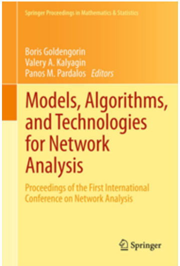


- Dynamics of Information Systems: Algorithmic Approaches. Alexey Sorokin, Robert Murphey, My T. Thai, and Panos M. Pardalos (Eds.) Springer. Springer Proceedings in Mathematics & Statistics., 2013. Vol. 51. ISBN 978-1-4614-3905-9.

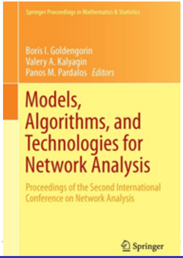


Models, Algorithms, and Technologies for Network Analysis

- Proceedings of the First International Conference on Network Analysis.Boris Goldengorin Valery A. Kalyagin Panos M. Pardalos(Eds.) Springer. Springer Proceedings in Mathematics & Statistics, 2013. Vol. 32. ISBN 978-1-4614-5574-5.

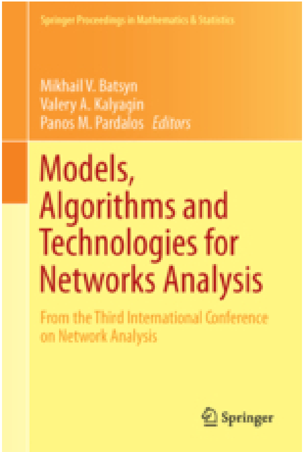


- Proceedings of the Second International Conference on Network Analysis. Boris Goldengorin Valery A. Kalyagin Panos M. Pardalos(Eds.) Springer. Springer Proceedings in Mathematics & Statistics, 2013. Vol. 59. ISBN 978-1-4614-8588-9.

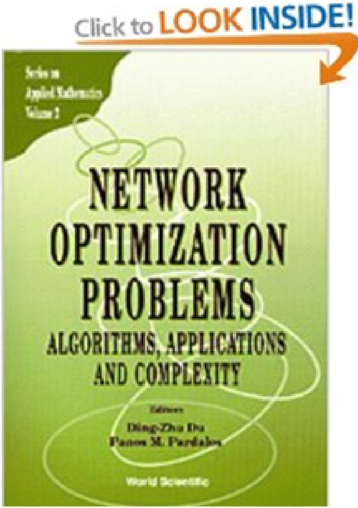


Models, Algorithms, and Technologies for Network Analysis

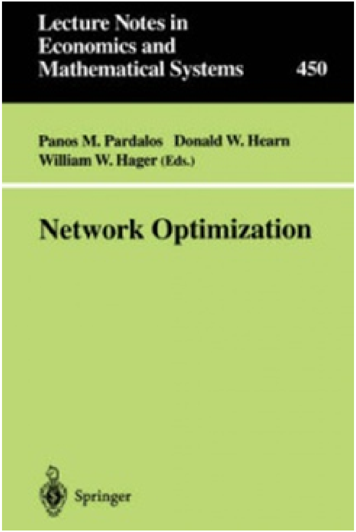
- Proceedings of the Third International Conference on Network Analysis.Batsyn, Mikhail V., Kalyagin, Valery A., Pardalos, Panos M. (Eds.) Springer. Springer Proceedings in Mathematics & Statistics, 2014. Vol. 104 ISBN 978-3-319-09757-2.



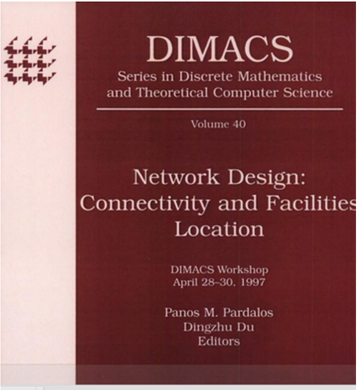
- Network Optimization Problems: Algorithms, Applications and Complexity. Ding-Zhu Du and Panos M. Pardalos(Eds.) WSPC 1993. Vol. 2. ISBN 978-981-02-1277-3.



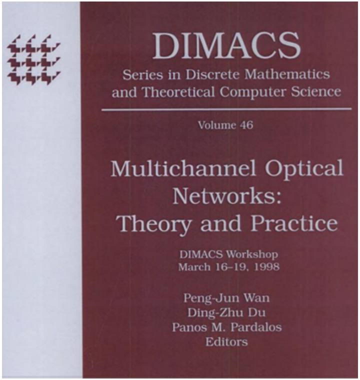
- Network Optimization (Lecture Notes in Economics and Mathematical Systems). Panos M. Pardalos, Donald W. Hearn, and William W. Hager (Eds.) Springer. 1997 ISBN 978-354-06-2541-4.



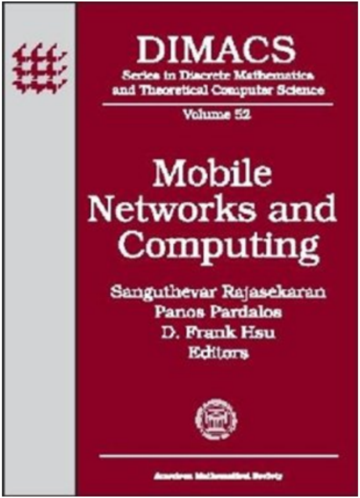
- Network Design (Connectivity and Facilities Location). Panos M. Pardalos and Ding-Zhu Du (Eds.) DIMACS Workshop. 1997 ISBN 0-8218-0834-6.



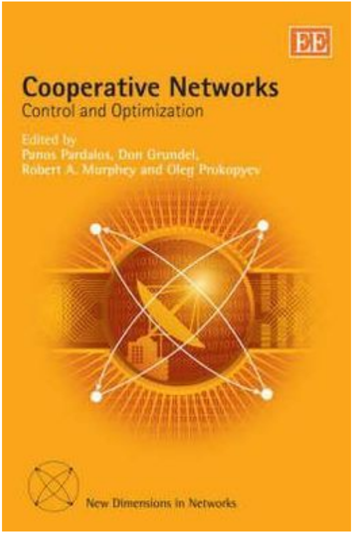
- Multichannel Optical Networks Theory and Practice. Peng-Jun Wan, Panos M. Pardalos, and Ding-Zhu Du (Eds.) DIMACS Workshop. 1998 ISBN 978-082-18-1004-0.



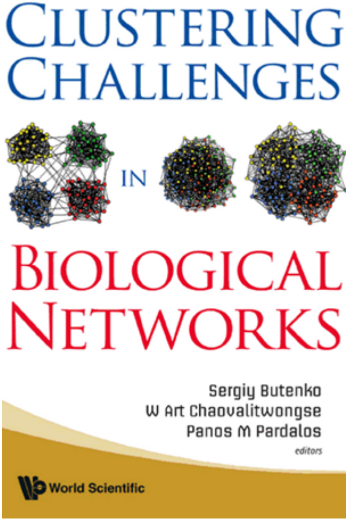
- Mobile Networks and Computing. Sanguthevar Rajasekaran, Panos M. Pardalos, and D. Frank Hsu (Eds.) DIMACS Workshop. 1999. 0-8218-1-5474.



- Cooperative Networks: Control and Optimization. Panos M. Pardalos, Don Grundel, Robert A. Murphey, and Oleg Prokopyev (Eds.) Elgar, Edward Publishing, Inc. 2008. ISBN 9781-8472-04-530.

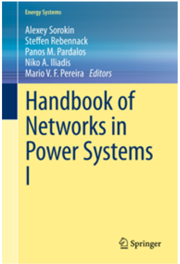


- Clustering Challenges In Biological Networks. Sergiy Butenko, W. Art Chaovalitwongse, and Panos M. Pardalos (Eds.) World Scientific. 2009. ISBN 978-981-2771-6-50.

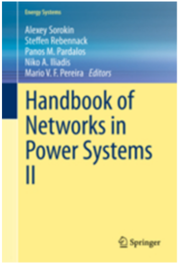


Handbook of Networks in Power Systems

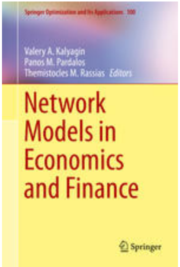
- Handbook of Networks in Power Systems I. Alexey Sorokin, Steffen Rebennack, Panos M. Pardalos, Niko A. Iliadis, and Mario V. F. Pereira (Eds.) Springer. 2012. ISBN 978-3-642-23193-3.



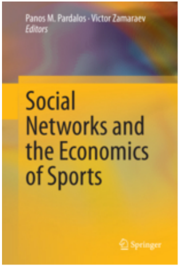
- Handbook of Networks in Power Systems II. Alexey Sorokin, Steffen Rebennack, Panos M. Pardalos, Niko A. Iliadis, and Mario V. F. Pereira (Eds.) Springer. 2012. ISBN 978-3-642-23406-4.



- Network Models in Economics and Finance. Kalyagin A. Valery, Panos M. Pardalos, Themistocles M. Rassias (Eds.). Springer-Verlag, 2014, 312 p. ISBN 978-3-319-09682-7.



- Social Networks and the Economics of Sports. Panos M. Pardalos, Victor A. Zamaraev (Eds.). Springer-Verlag, 2014, 169 p. ISBN 978-3-319-08439-8.

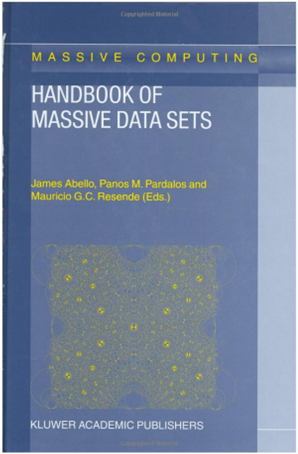


What is Big Data?

- 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 rapid advances in computer and information technologies, many of these challenges are beginning to be addressed.

Handbook of Massive Data Sets

- J. Abello, P.M. Pardalos and M. Resende,
  Handbook of Massive Data Sets,
  Kluwer Academic Publishers, 2002



Knowledge Discovery in Databases

KDD is the process of identifying valid, novel, potentially useful, and ultimately understandable structure (models and patterns) in the data

- Understand the application domain
- Create a target dataset
- Remove (or correct) corrupted data
- Apply data-reduction algorithms
- Apply data mining algorithms
- Interpret the mined patterns

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

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 is:

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

("Self-organized" networks)

Cliques

- A set of vertices S is called a **clique** if the subgraph $G(S)$ induced by S is complete; i.e. there is an edge between any two vertices in $G(S)$.
- A **maximal clique** is a clique which is not a proper subset of another clique.
- A **maximum clique** is a clique of the maximum cardinality.

The Maximum Clique Problem

- The maximum clique problem (MCP) is to find a maximum clique in a given graph G.
- We will denote the cardinality of the maximum clique in graph G by $\omega(G)$.
- The MCP is one of the classical problems in graph theory with many applications in many fields including project selection, classification, fault tolerance, coding, computer vision, economics, information retrieval, signal transmission, and alignment of DNA with protein sequences.

The Maximum Independent Set Problem

- A set of nodes S in a graph G is an independent set (stable set) if any two vertices in S are not adjacent.
- The maximum independent set problem is to find the independent set of the maximum cardinality.
- We denote the cardinality of this maximum independent set by $\alpha(G)$.

Finding cliques and independent sets

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

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

Optimization on Very Large Graphs

- The graphs we have to deal with in some applications are very massive. Examples are the WWW graph and the **call graph**.
- The various gigantic graphs that have lately attracted notice share some properties:
  - They tend to be **sparse**: The graphs have relatively few edges, considering their vast numbers of vertices.
  - They tend to be **clustered**. In the World Wide Web, two pages that are linked to the same page have an elevated probability of including links to one another.

Optimization on Very Large Graphs (cont.)

- They tend to have a **small diameter**. The diameter of a graph is the longest shortest path across it.
- Graphs with the three properties of **sparseness**, **clustering** and **small diameter** have been termed **"small-world" graphs**.

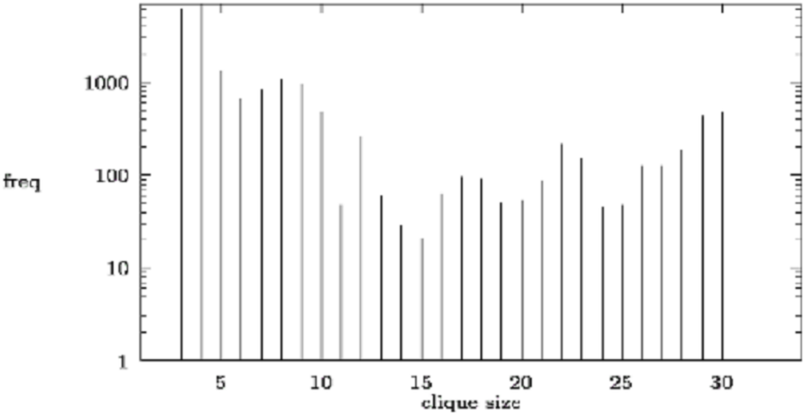
The Call Graph

- In the **call graph**, the vertices are telephone numbers, and two vertices are connected by an edge if a call was made from one number to another.
- A **call graph** was constructed with data from AT&T telephone billing records. Based on one 20-day period it had 290 million vertices and 4 billion edges.
- The analyzed one-day **call graph** had 53,767,087 vertices and over 170 millions of edges

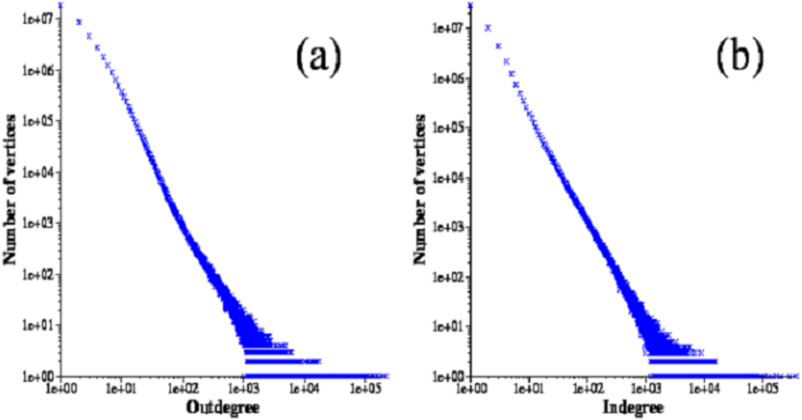
The Call Graph (cont.)

- This graph appeared to have 3,667,448 connected components, most of them tiny.
- A **giant connected component** with 44,989,297 vertices (more than 80 percent of the total) was computed.
- The distribution of the degrees of the vertices follows the **power-law distribution**.

The Call Graph (cont.)



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



References

J. Abello and P.M. Pardalos and M. Resende,
On maximum clique problems in very large graphs, External Memory Algorithms, 1999, pp. 119-130

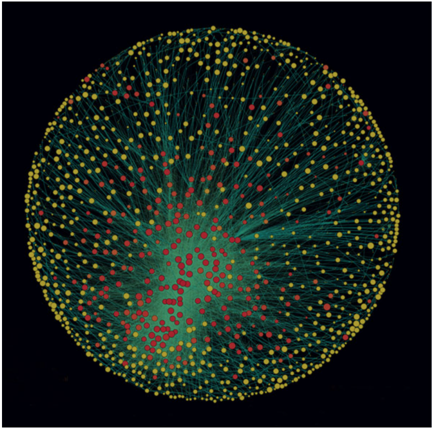
V. Boginski and S. Butenko and P. M. Pardalos,
Modeling and optimization in massive graphs, Novel Approaches to Hard Discrete Optimization, 2003, pp. 17-39

Market Graph

- Each stock corresponds to a **vertex** in the network.
- Link between two stocks (vertices) is represented by a weighted edge, where the weight reflects **degree of similarity** between stocks.
- Market network is a **complete weighted graph**.
- Mining market data: filter the information in complete weighted graph in order to extract the **most valuable information**.
- Remark: Common measure of similarity is **correlation** calculated from time series of observations.

“The capitalist network that runs the world”

The 1318 transnational corporations that form the core of the economy. Super-connected companies are red, very connected companies are yellow. The size of the dot represents revenue (Image: PLoS One)



References

V. Boginski and S. Butenko and P.M. Pardalos, ***On structural properties of the market graph***, in Innovations in financial and economic networks, Edward Elgar Publishers, 2003

V. Boginski and S. Butenko and P. M. Pardalos, ***Statistical analysis of financial networks***, Computational Statistics and Data Analysis, 2005, pp. 431443

V. Boginski and S. Butenko and P. M. Pardalos, ***Mining market data: A network approach***, Computers & Operations Research, 2006, pp. 3171-3184

A. Vizgunov and B. Goldengorin and V. Kalyagin and A. Koldanov and P. Koldanov and P.M. Pardalos, ***Network approach for the Russian stock market***,Computational Management Science January 2014, Volume 11, Issue 1-2, pp 45-55

A. Koldanov and P. Koldanov and V. Kalyagin and P.M. Pardalos, ***Statistical procedures for the market graph construction***, Computational Statistics & Data Analysis, Volume 68, December 2013, Pages 1729

G. Bautin and V. Kalyagin and A. Koldanov and P. Koldanov and P.M. Pardalos, ***Simple measure of similarity for the market graph construction***, Computational Management Science, 2013, pp. 105-124

S. Vitali, J. B. Glattfelder, S. Battiston ***The Network of Global Corporate Control***, PloS ONE, vol. 6, is. 10, e25995,

References (cont.)

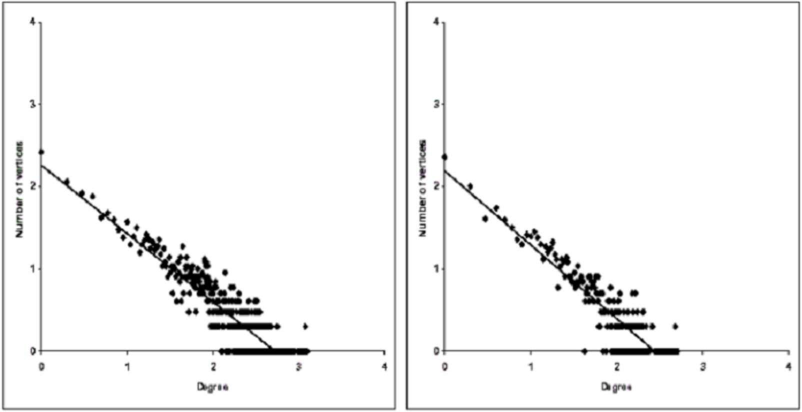
V.A. Kalyagin, A.P. Koldanov, P.A. Koldanov, P.M. Pardalos, V.A. Zamaraev, ***Measures of uncertainty in market network analysis***, Physica A: Statistical Mechanics and its Applications, Volume 413, 1 November 2014, Pages 59-70

Anton Kocheturov, Mikhail Batsyn, Panos M. Pardalos, ***Dynamics of cluster structures in a financial market network***, Physica A: Statistical Mechanics and its Applications, Volume 413, 1 November 2014, Pages 523-533

Market Graph Analysis Tools

- Minimum Spanning Tree.
- Planar Maximally Filtered Graph.
- Maximum cliques and clique partitions.
- Maximum independent sets.

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

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

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

  until $G_0 = \emptyset.$

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

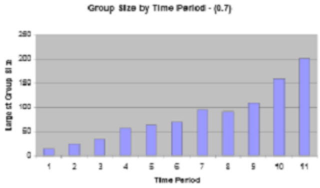
θ	edge density	clique size
0.35	0.0090	193
0.4	0.0047	144
0.45	0.0024	109
0.5	0.0013	85
0.55	0.0007	63
0.6	0.0004	45
0.65	0.0002	27
0.7	0.0001	22

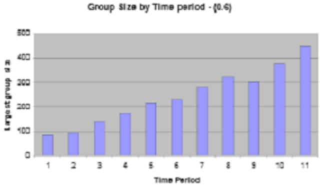
Classification of Stocks Using Clique Partitioning

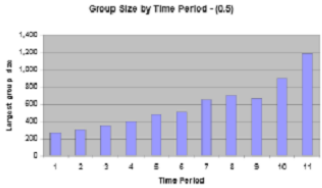
- A clique in the market graph represents a dense cluster of stocks whose prices exhibit a similar behavior over time
- Therefore, dividing the market graph into a set of distinct cliques (clique partitioning) is a natural approach to classifying stocks (dividing the set of stocks into clusters of similar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

Connected Components in Market Graph

- Largest Group size by Time Period







Connected Components in Market Graph

Observations

- The increase in the giant component size from oldest to newest time period indicates the globalization tendency, just as in maximum clique size and edge density
- The giant component includes semiconductor industries and the increase in the size of the giant component corroborates the observation that the number of these industries has been increasing with time

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

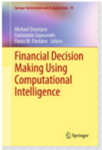
- Finding a perfectly diversified portfolio containing **any given stock**
- For every vertex in the market graph, an independent set that contains this vertex was detected, and the sizes of these independent sets were almost the same, which means that it is possible to find a diversified portfolio containing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

Analysis of the Russian Stock Market

- The market network constructed using **correlation** as a measure of similarity between stocks.
- We considered **11 shifted periods of 500-day each** from September 1, 2007 to September 16, 2011
- Results are surprising:
  - Russian stock market is dominated by a few **highly correlated** stokes with the biggest value.
  - The nodes of the maximum clique for the threshold are 9 most valuable stocks.
  - The stocks in the clique account for 89% of the total value of the market.
  - The most valuable stocks have the strongest connections between their return.

Selected Books

- M. Doumpos, C. Zopounidis and P. Pardalos, ***Financial Decision Making Using Computational Intelligence***, Springer Optimization and Its Applications Vol. 70, Springer, 2012



- C. Zopounidis, M. Doumpos and P. Pardalos, ***Handbook of Financial Engineering***, Springer Optimization and Its Applications Vol. 18, Springer, 2008

Critical Elements Detection

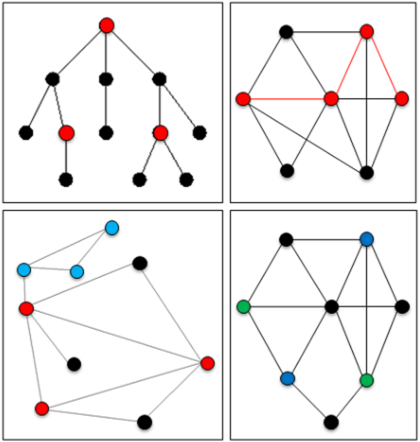
Given a graph $G(V, E)$ and an integer k, find a set of at most k **elements**, whose deletion minimizes the **connectivity** of the residual network.

Elements?

- Nodes (arcs)
- Paths
- Cliques
- Node subsets

Connectivity?

- Max flow
- Number of pairwise connections
- Number of components



Critical Elements Detection

The problem is proven to be NP-hard in the general case for different elements:

- Nodes (Arcs)
- Paths
- Cliques

A. Arulsevan and C. W. Commander and L. Elefteriadou and P. M. Pardalos, ***Detecting Critical Nodes in Sparse Graphs***, Computers and Operations Research, 2009, pp. 2193-2200

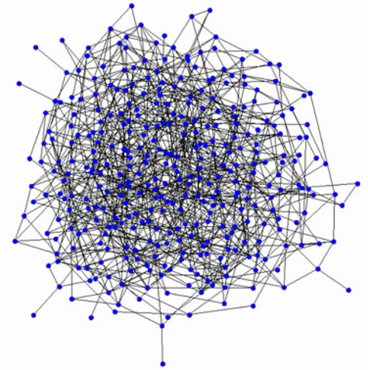
T. N. Dinh and Y. Xuan and M. T. Thai and P. M. Pardalos, ***On New Approaches of Assessing Network Vulnerability: Hardness and Approximation***, IEEE ACM Transactions on Networking, 2012, pp. 609-619

J. Walteros and P. M. Pardalos, ***A Decomposition Approach for Solving Critical Clique Detection Problems***, Experimental Algorithms, Springer, 2012, pp. 393-404

Why should we study this problem?

Disconnecting a network by element removal is not trivial!

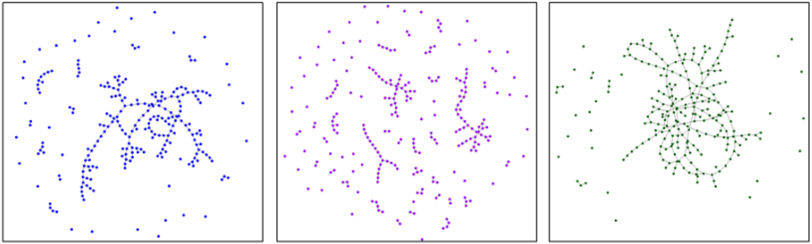
- 350 nodes, 900 arcs
- **Network 1:** U(0,1)
- **Network 2:** greedy construction
- **Network 3:** Power law a=0.44 b=50



Why should we study this problem?

Greedy algorithm: Eliminate the node with the largest degree. (112 nodes removed)

- 350 nodes, 900 arcs
- **Network 1:** 43 components left
- **Network 2:** 87 components left
- **Network 3:** 26 components left



Connectivity Measures

Network Flow Measures

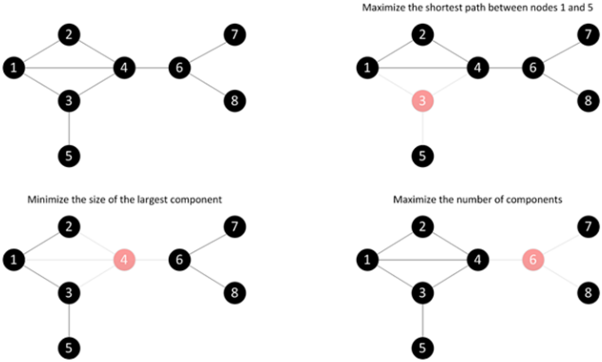
- Single/Multiple commodity shortest path
- Single/Multiple commodity maximum flow
- Single/Multiple commodity minimum cost

Topological Measures

- Pairwise (weighted) connectivity
- Largest component size
- Total number of components

Connectivity Measures: Different results

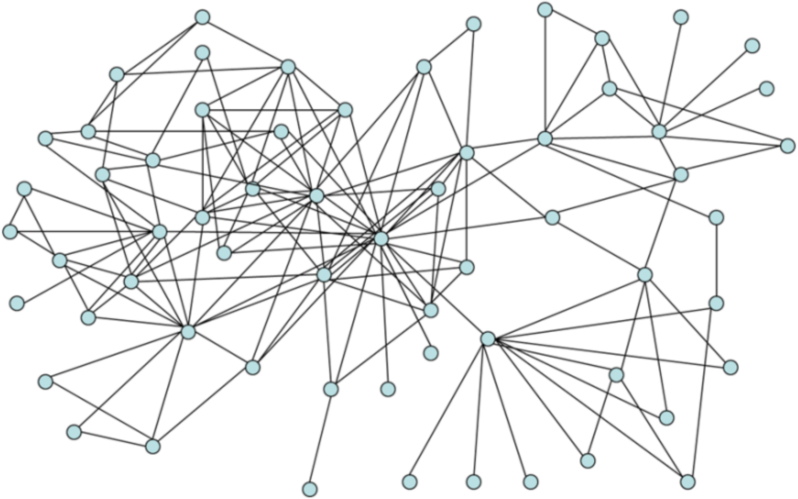
- The selection of the connectivity measure is crucial
- Despite the fact that all these measures account for a disconnection level, using one over the other may lead to different critical elements



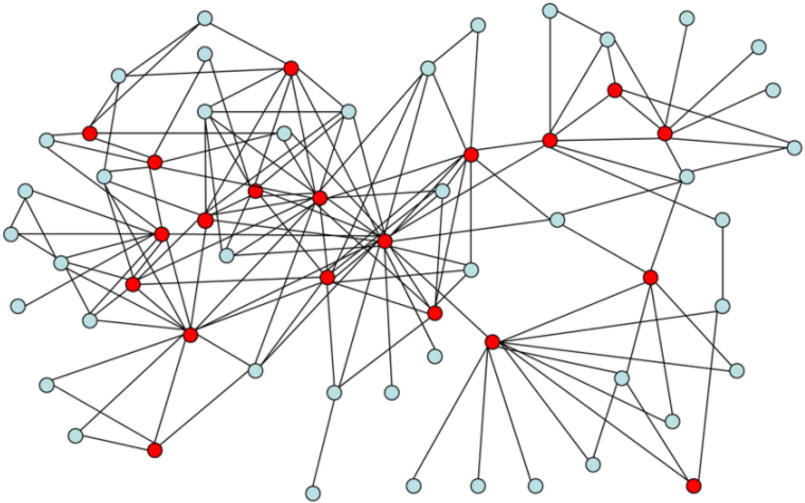
Applications

- Evacuation planning
- Fragmentation of terrorist organizations
- Epidemic contagion analysis and immunization planning
- Social network analysis (Prestige and dominance)
- Transportation (Cross-dock and hub-and-spoke networks)
- Marketing and customer services design
- Biomaterials and drugs design

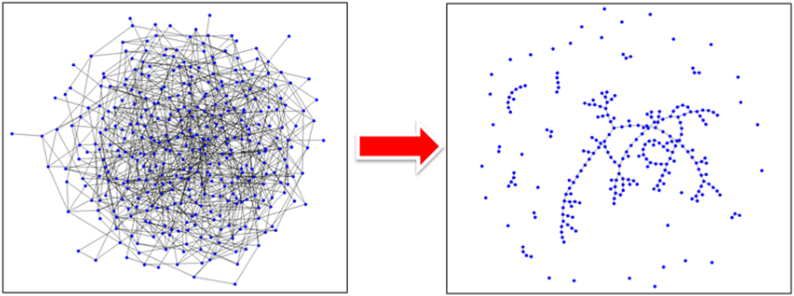
Critical Nodes Detection Problem



Critical Nodes Detection Problem



Given a graph $G(V, E)$ and an integer k, find a set of at most k nodes, whose deletion minimizes the pairwise connections of the residual network.



CNP - Formulation

- $V :=$ Set of vertices
- $E :=$ Set of edges
- $k :=$ Number of critical nodes to identify
- $v_i := \begin{cases} 1 & \text{if node } i \text{ is critical} \\ 0 & \text{otherwise} \end{cases}$
- $u_{ij} := \begin{cases} 1 & \text{if } i \text{ and } j \text{ are in the same component} \\ 0 & \text{otherwise} \end{cases}$

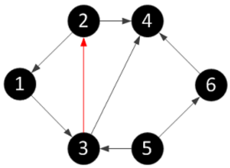
$$
\begin{aligned}
\min \sum_{i,j \in V} u_{ij} & \\
\text{s.t. } u_{ij} + v_i + v_j \geq 1 & \qquad \forall (i,j) \in E \\
u_{ij} + u_{jk} - u_{ki} \leq 1 & \qquad \forall (i,j,k) \in V \\
u_{ij} - u_{jk} + u_{ki} \leq 1 & \qquad \forall (i,j,k) \in V \\
-u_{ij} + u_{jk} + u_{ki} \leq 1 & \qquad \forall (i,j,k) \in V \\
\sum_{i \in V} v_i \leq k & \\
u_{ij} \in \{0,1\} & \qquad \forall (i,j) \in V \\
v_i \in \{0,1\} & \qquad \forall i \in V
\end{aligned}
$$

A. Arulsevan and C. W. Commander and L. Elefteriadou and P. M. Pardalos,
Detecting Critical Nodes in Sparse Graphs, Computers and Operations Research, 2009, pp. 2193-2200

- **β-edge disruptor**
  Given a directed graph $G(V, E)$ and a number $0 \leq \beta < 1$, a subset $S \subset E$ is β-edge disruptor if the overall pairwise connectivity in $G(E/S)$, obtained by removing S from E is no more than

$$\beta \binom{n}{2}$$

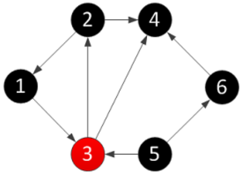
- By minimizing the cost of such edges in $S \subset E$, we have the β-edge disruptor problem.



- **β-vertex disruptor**
  Given a directed graph $G(V, E)$ and a number $0 \leq \beta < 1$, a subset $S \subset V$ is β-vertex disruptor if the overall pairwise connectivity in $G(V/S)$, obtained by removing S from V is no more than

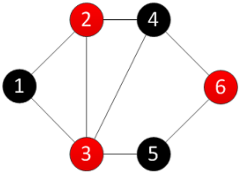
$$\beta \binom{n}{2}$$

- By minimizing the cost of such edges in $S \subset V$, we have the β-vertex disruptor problem.



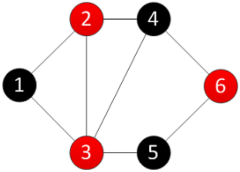
β-vertex disruptor: NP-Completeness

- **β-vertex disruptor can be proven to be NP-Complete by a reduction from the vertex cover problem**
- Note that, if the graph has a vertex cover of size k, it also has a β-vertex disruptor for $\beta = 0$, as the residual graph after deleting the vertex cover is disconnected.



β-vertex disruptor: Inapproximability

- **β-vertex disruptor cannot be approximated within a factor of 1.36**
- Since it is known that the Vertex Cover problem cannot be approximated within a constant factor less than 1.36 unless P=NP, the β-vertex cannot be approximated as the VC reduces to the β-vertex.
- A similar result follows for the β-edge disruptor



β-edge disruptor: Pseudo-approximability

- **The β-edge disruptor can be pseudo-approximated**
- Formally, our algorithm finds in a uniform directed graph a β'-edge disruptor whose cost is at most $O(\log^{3/2} n)\mathrm{OPT}$
- where OPT is the optimal cost and

$$\beta'/4 < \beta < \beta'$$

Algorithm 1: β-Edge Disruptor

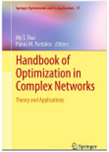
Input: Uniform edges' weight directed graph $G = (V, E)$ and $0 \leq \beta < \beta' < 1$
Output: A β'-edge disruptor of G.

/* Construct the decomposition tree */
1. $c \leftarrow 1 - \sqrt{\frac{\beta}{\beta'}}$.
2. $T(V_T, E_T) \leftarrow (\{t_0\}, \phi)$, $V(t_0) \leftarrow V(G)$, $l(t_0) = 1$
3. **while** $\exists$ unvisited t_i with $|V(t_i)| \geq 2$ **do**
4. Mark t_i visited, create new child nodes t_{i1}, t_{i2} of t_i.
5. $V_T \leftarrow V_T \cup \{t_{i1}, t_{i2}\}$
6. $E_T \leftarrow E_T \cup \{(t_i, t_{i1}), (t_i, t_{i2})\}$
7. Separate $G_{[V(t_i)]}$ using directed c-balanced cut.
8. Associate $V(t_{i1}), V(t_{i2})$ with two separated components.
9. $\mathrm{cost}(t_i) \leftarrow$ The cost of the balanced cut

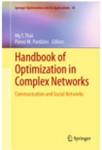
/* Find the minimum cost G-partitionable */
10. Traverse T in post-order, **for** each $t_i \in T$ **do**
11. **for** $p \leftarrow 0$ **to** $\beta'\binom{n}{2}$
12. **if** $\mathcal{P}(G_{[V(t_i)]}) \leq p$ **then** $cost(t_i, p) \leftarrow 0$
13. **else** $\mathrm{cost}(t_i, p) \leftarrow \min\{\mathrm{cost}(t_{i1}, p_1) + \mathrm{cost}(t_{i2}, p_2) + \mathrm{cost}(t_i) \mid p_1 + p_2 = p\}$
14. Find $F_{\beta'}^{\mathrm{opt}}$ associating with $T_{\beta'}^{\mathrm{opt}} = \min_{p \leq \beta'\binom{n}{2}}\{\mathrm{cost}(t_0, p)\}$
15. Return union of c-balanced cuts at $t_i \in \mathcal{A}(F_{\beta'}^{\mathrm{opt}})$.

Selected Books

- M. Thai and P. Pardalos
  Handbook of Optimization in Complex Networks: Theory and Applicatios, Springer, 2011



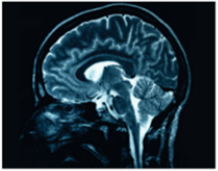
- M. Thai and P. Pardalos
  Handbook of Optimization in Complex Networks: Communication and Social Networks, Springer, 2011

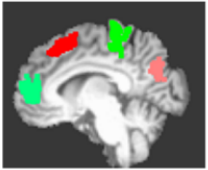


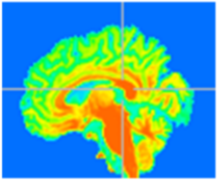
Human Brain Networks

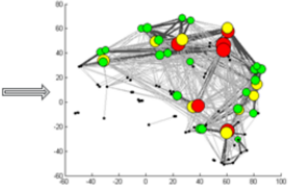
Human Brain Networks

- Network modeling approach to study **complex systems**
- Statistically dependent **neural activity patterns** in distinct brain regions
- **Functional interactions** through neural impulses and information exchange
- Brain networks: nodes represent **brain regions**; connections represent **functional interactions**



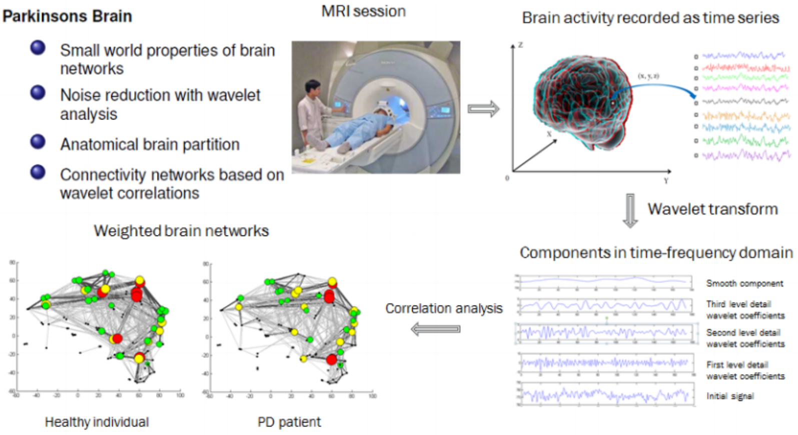






F. Skidmore and D. Korenkevych and Y. Liu and G. He and E. Bullmore and P. M. Pardalos,
Connectivity brain networks based on wavelet correlation analysis in Parkinson fMRI data, Neuroscience Letters, 2011, pp. 47-51

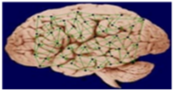
Human Brain Networks (cont.)

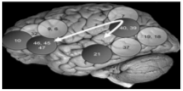


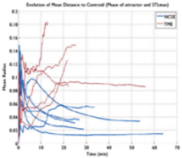
Epilepsy

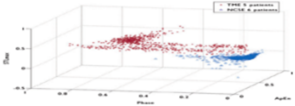
Brain state transition study and epilepsy research

- Nonlinear dynamic modeling
- Patients classification of epilepsy
- Pattern discovery and machine learning
- Epileptic seizure prediction
- Network modeling and optimization
- Epileptic brain state transition study
- Seizure monitoring / warning system









Selected references

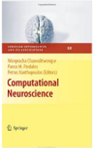
- J. Zhang and P. Xanthopoulos and C. Liu and S. Bearden and B. M Uthman and P. M. Pardalos,
  Real-time differentiation of nonconvulsive status epilepticus from other encephalopathies using quantitative EEG analysis: A pilot study, Epilepsia, 2010, pp. 243-250
- L. Iasemidis and D. Shiau and J. Sackellares and P.M. Pardalos,
  Quadratic binary programming and dynamic system approach to determine the predictability of epileptic seizures, Journal of Combinatorial Optimization, 2001, pp. 9-26
- L. Iasemidis and J. Sackellares and D. Shiau and W. Chaovalitwongse and P. Carney and J. Principe and M. Yang and V. Yatsenko and S. Roper and P.M. Pardalos,
  Seizure Warning Algorithm Based on Optimization and Nonlinear Dynamics, Mathematical Programming, 2004, pp. 365385

The William Pierskalla best paper award for research excellence in health care management science, INFORMS.

Selected Books

- P. Pardalos, P. Xanthopoulos and M. Zervakis
  Data Mining for Biomarker Discovery, Springer, 2012

- W. Chaovalitwongse, P. Pardalos and P. Xanthopoulos
  Computational Neuroscience, Springer, 2010



Concluding remarks

ΧΡΥΣΟΝ ΓΑΡ ΟΙ ΔΙΖΗΜΕΝΟΙ ΓΗΝ ΠΟΛΛΗΝ ΟΡΥΣΣΟΥΣΙ ΚΑΙ ΕΥΡΙΣΚΟΥΣΙΝ ΟΛΙΓΟΝ

"Seekers after gold dig up much earth and find little"

"The lord whose oracle is at Delphi neither speaks nor conceals, but gives signs"

- Heraclitus

I shall be telling this with a sigh
Somewhere ages and ages hence:
Two roads diverged in a wood, and I,
I took the one less traveled by,
And that has made all the difference.

- Robert Frost