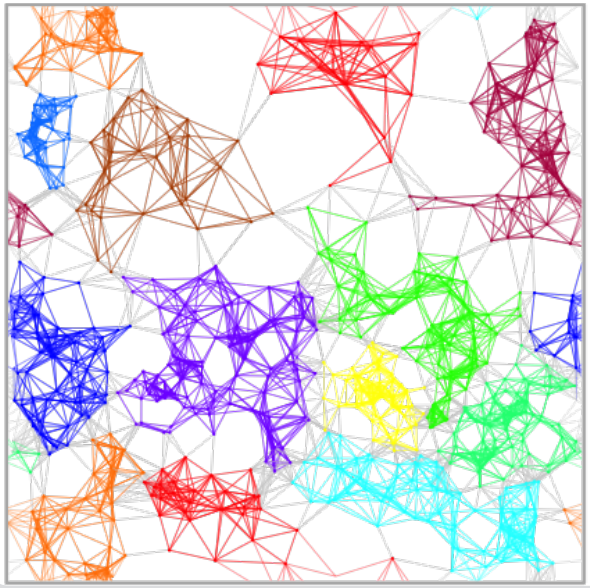
Karlsruher Institut für Technologie

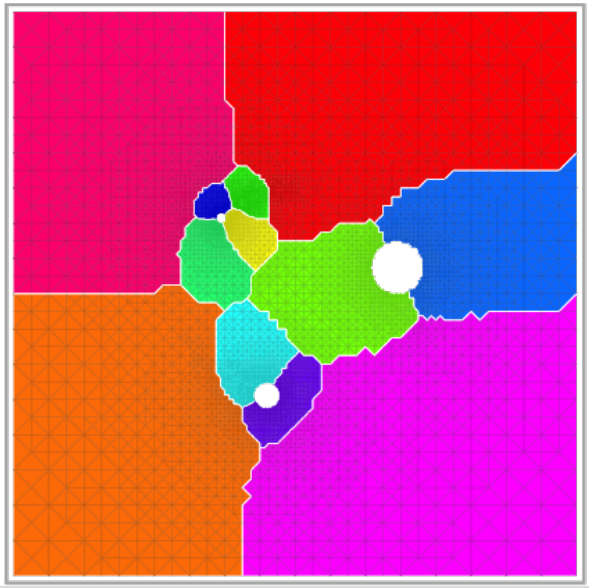
FINCA: Fast Inexact Combinatorial and Algebraic Solvers for Massive Networks

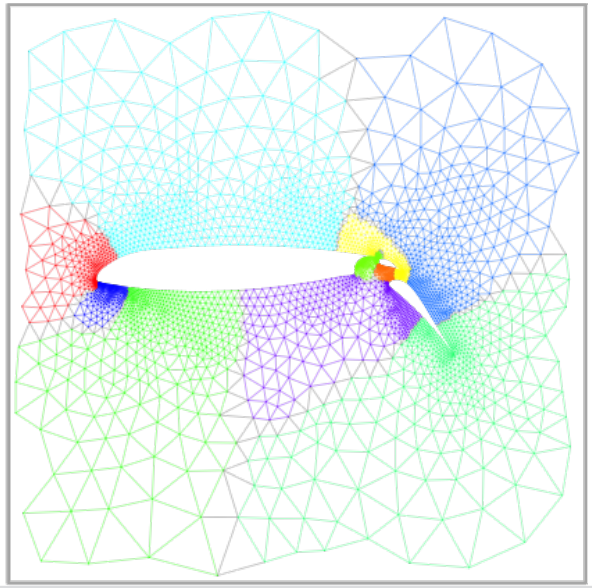
Henning Meyerhenke

Theoretical Informatics and Parallel Computing
Karlsruhe Institute of Technology (KIT)

Frankfurt, 11th June 2014 · Meeting of DFG SPP 1736 *Algorithms for Big Data*



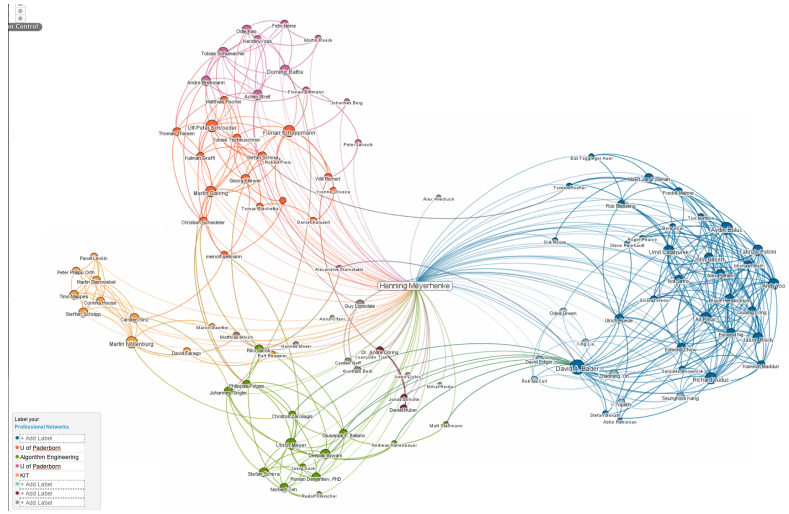




Context

Massive complex networks
are part of big data:

- Social networks



Context

Massive complex networks
are part of big data:

- Social networks
- Climate networks

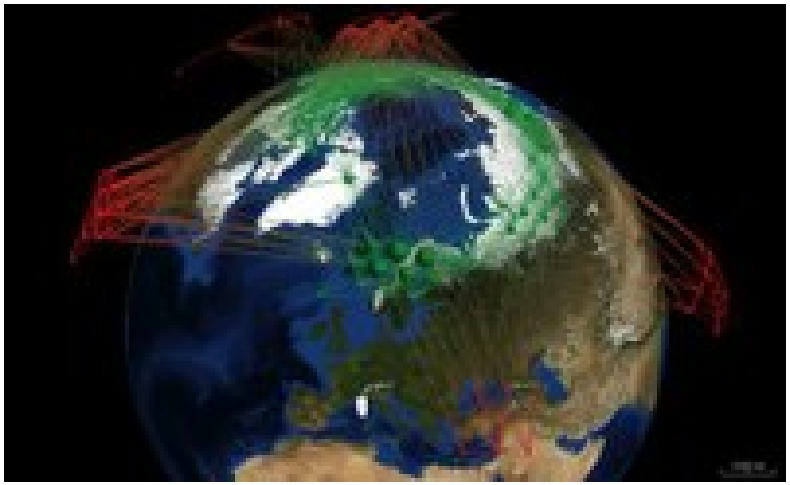


Image source: PIK Potsdam

Context

Massive complex networks
are part of big data:

- Social networks
- Climate networks
- Biological networks
- ...

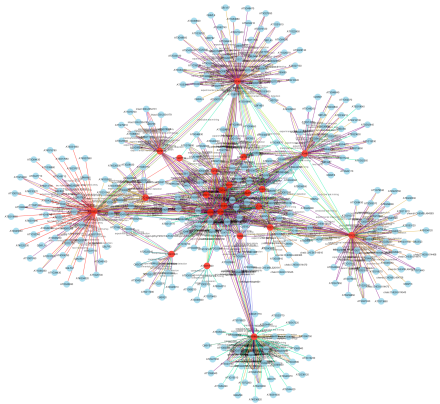


Image source: shgmo.org

Context

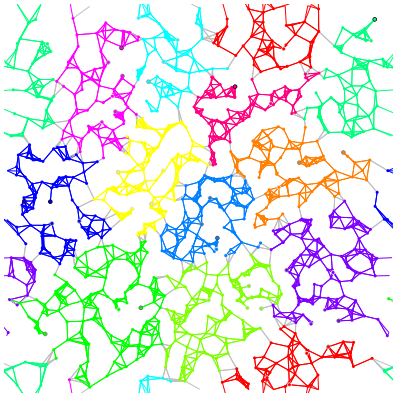
Massive complex networks are part of big data:

- Social networks
- Climate networks
- Biological networks
- ...

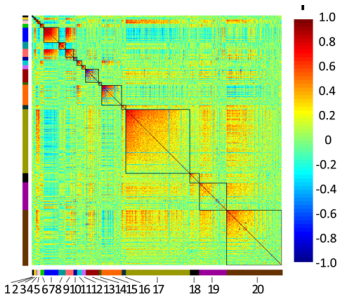
Network analysis

- **What:** Extract info from network structure and/or data stored within
- **Why:** Analysts/scientists need **information** for decisions!
- **How:** FINCA: Parallelism, approximation algorithms, heuristics

Objectives

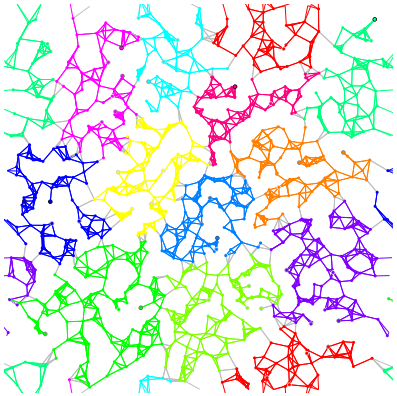


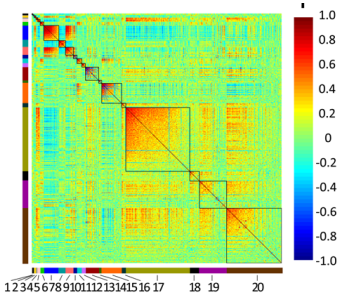
- Make recent theoretical results usable in practice:
  - Sparsification of dense graphs
  - Low-rank approximation of matrices
  - Approximate maximum network flow



Img. source: E. A. Stone, J. F. Ayroles

Objectives

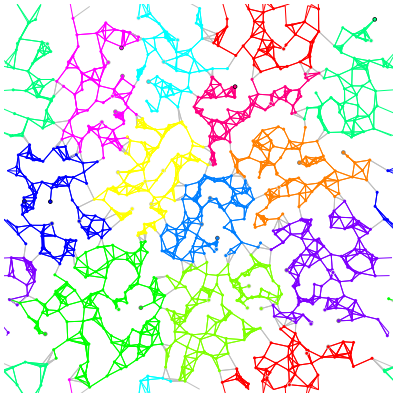


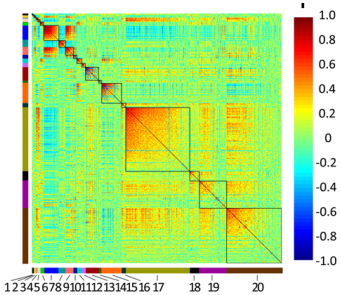


Img. source: E. A. Stone, J. F. Ayroles

- Make recent theoretical results usable in practice:
  - Sparsification of dense graphs
  - Low-rank approximation of matrices
  - Approximate maximum network flow
- Improve combinatorial “solvers”:
  - Clustering dense graphs derived from correlation data
  - Graph drawing for massive complex networks
  - Network flow
  - Generation of massive dynamic complex networks

Objectives





Img. source: E. A. Stone, J. F. Ayroles

- Make recent theoretical results usable in practice:
  - Sparsification of dense graphs
  - Low-rank approximation of matrices
  - Approximate maximum network flow
- Improve combinatorial “solvers”:
  - Clustering dense graphs derived from correlation data
  - Graph drawing for massive complex networks
  - Network flow
  - Generation of massive dynamic complex networks
- Improve applications from bioinformatics:
  - Gene expression profiling
  - Molecular distance geometry
  - Image segmentation tasks in biology

Methods

Clustering and drawing graphs

Clustering dense networks

- Sparsen dense graphs based on different edge importance measures
- Approximation/estimation of edge importance measures
- Integration into multilevel clustering algorithm
- Clustering dynamic/streaming networks

Methods

Clustering and drawing graphs

Clustering dense networks

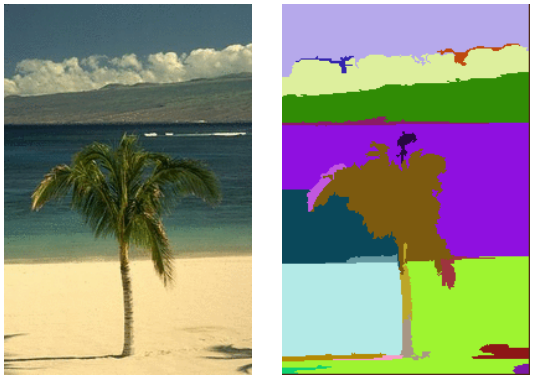
- Sparsen dense graphs based on different edge importance measures
- Approximation/estimation of edge importance measures
- Integration into multilevel clustering algorithm
- Clustering dynamic/streaming networks

Drawing massive complex networks

- Entropy-stress model for GD
- Parallel implementation (ongoing work with C. Schulz and M. Nöllenburg)
- Integration of clustering concepts and edge importance measures
- Application to molecular distance geometry

Methods

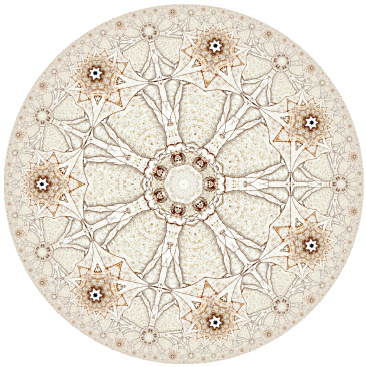
Approximate network flow and network generation



Img. source: Felzenszwalb, Huttenlocher

Approximate network flow

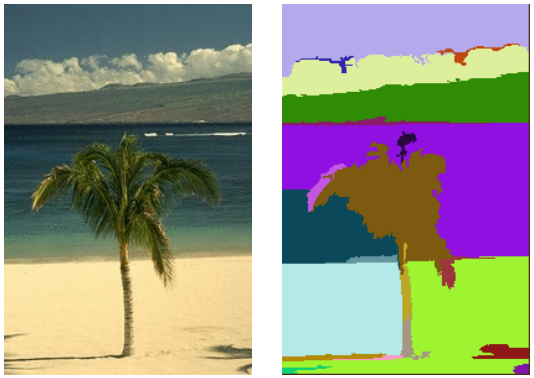
- Sparsification
- Low-rank approximation
- Combinatorial linear solvers
- Possible application: Image segmentation



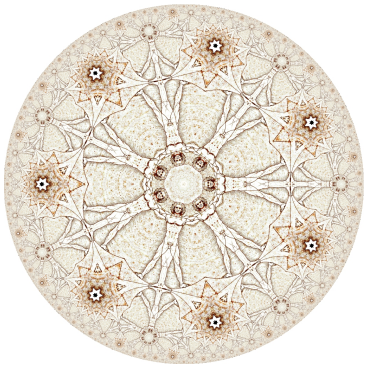
Img. source: Krioukov et al.

Methods

Approximate network flow and network generation



Img. source: Felzenszwalb, Huttenlocher



Img. source: Krioukov et al.

Approximate network flow

- Sparsification
- Low-rank approximation
- Combinatorial linear solvers
- Possible application: Image segmentation

Massive dynamic complex networks

- Hyperbolic geometry
- Space-partitioning data structures
- Dynamics

Nucleation points for cooperation

NetworKit tool suite

- Fast interactive network analysis (C++ with OpenMP, Python)
- Focus on massive networks
- Open source (MIT license)
- Used by several national and international projects
- Info, talk, docu, code: `http://www.network-analysis.info`

Nucleation points for cooperation

NetworKit tool suite

- Fast interactive network analysis (C++ with OpenMP, Python)
- Focus on massive networks
- Open source (MIT license)
- Used by several national and international projects
- Info, talk, docu, code: `http://www.network-analysis.info`

FINCA connections to other projects

- Graph generator will be made available for all projects
- **Brandes/Wagner:** Backbones / sparsification
- **Koch/Mutzel:** Motifs / clustering, drawing
- **Zweig:** Centrality, clustering, motifs