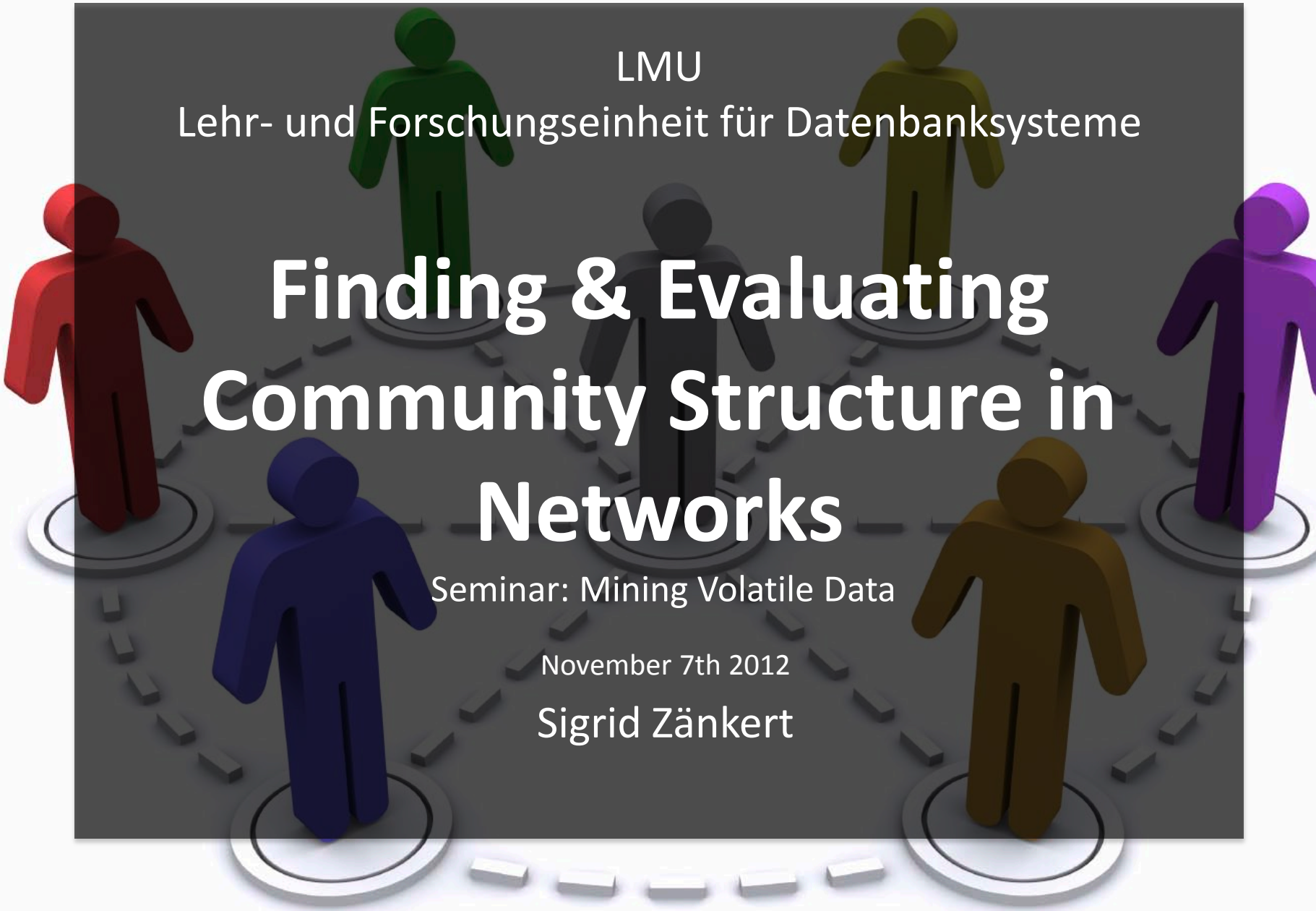


LMU
Lehr- und Forschungseinheit für Datenbanksysteme

Finding & Evaluating Community Structure in Networks

Seminar: Mining Volatile Data

November 7th 2012

Sigrid Zänkert


Executive Summary
Motivation
Methodology
Results
Evaluation


Executive Summary
Motivation
Methodology
Results
Evaluation

Executive Summary

„Finding and evaluating community structure in networks“

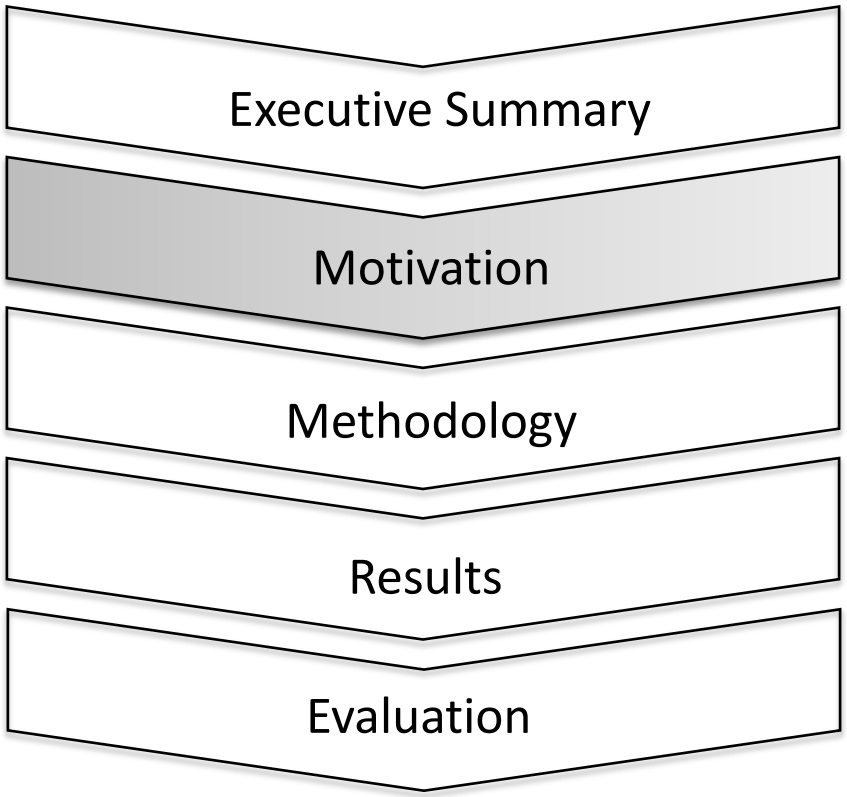
Mark E.J. Newman and Michelle Girvan

The American Physical Society 2004

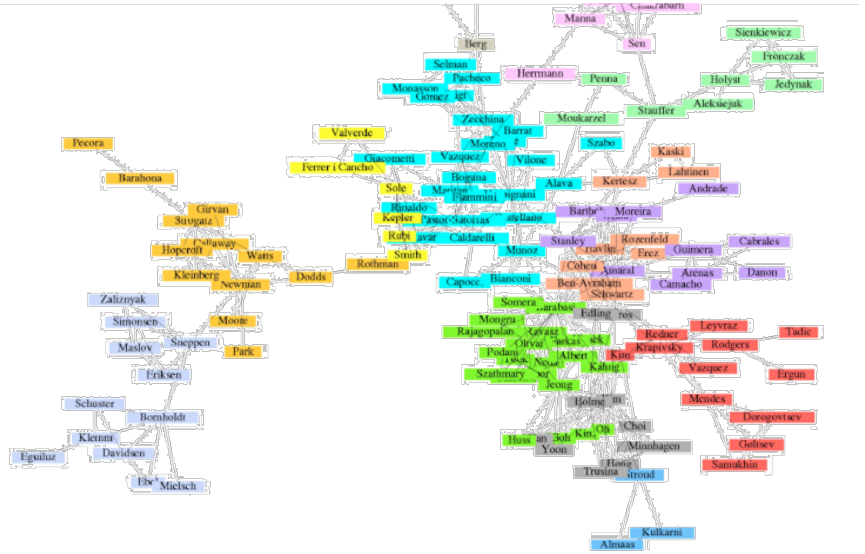
Set of algorithms for discovering community structure in networks

Quality measure for the detected community structure

Demonstration of effectiveness with both real-world and computer generated network data


Executive Summary
Motivation
Methodology
Results
Evaluation

What is a Community?



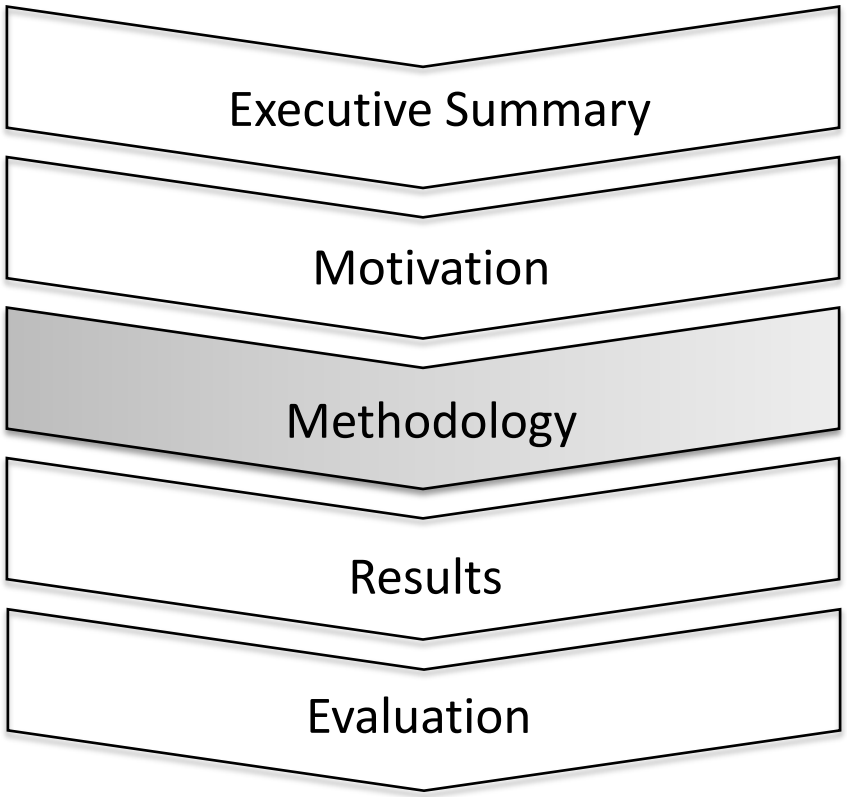
- Community structure is the natural division of network nodes into densely connected subgroups
- Communities are densely connected within but there is only sparse connection between the different communities

Community Structure in Networks

- Detecting community structures within networks: important research topic in statistical physics, applied mathematics, computer science and sociology

- Versatile application of „network ideas“: World Wide Web, epidemiology, citation, collaboration and many more

- Challenge: Find algorithms that generate valid results within reasonable computing time


Executive Summary
Motivation
Methodology
Results
Evaluation

Hierarchical Clustering

Hierarchical Clustering

Agglomerative
(addition of edges)

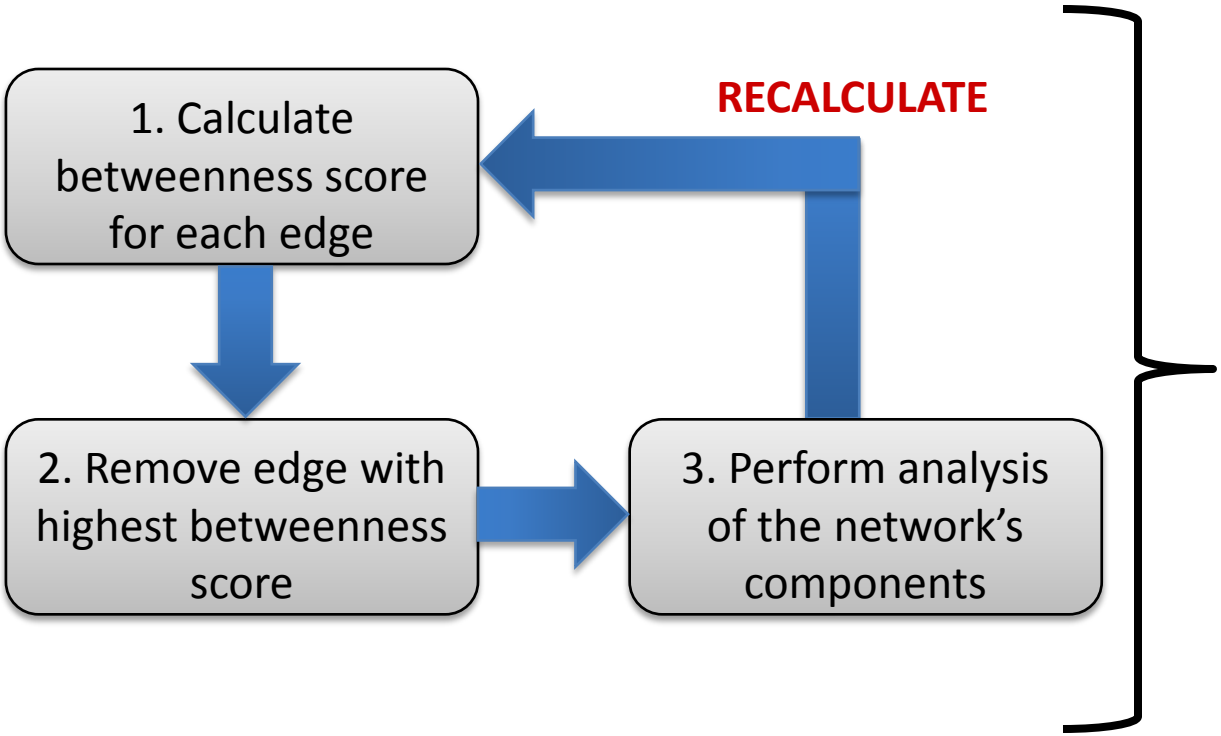
- Similarity between vertex pairs
- Edges are added based on that similarity
- Algorithms often find only the core communities or even wrong community structures

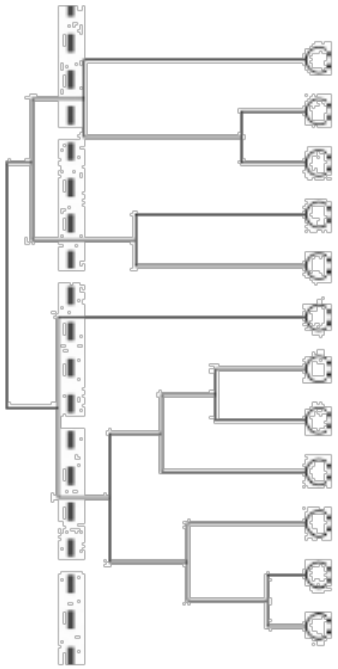
Divisive
(removal of edges)

- Start with the complete network of interest
- Repeated removal of the edges between the least similar connected pairs of vertices

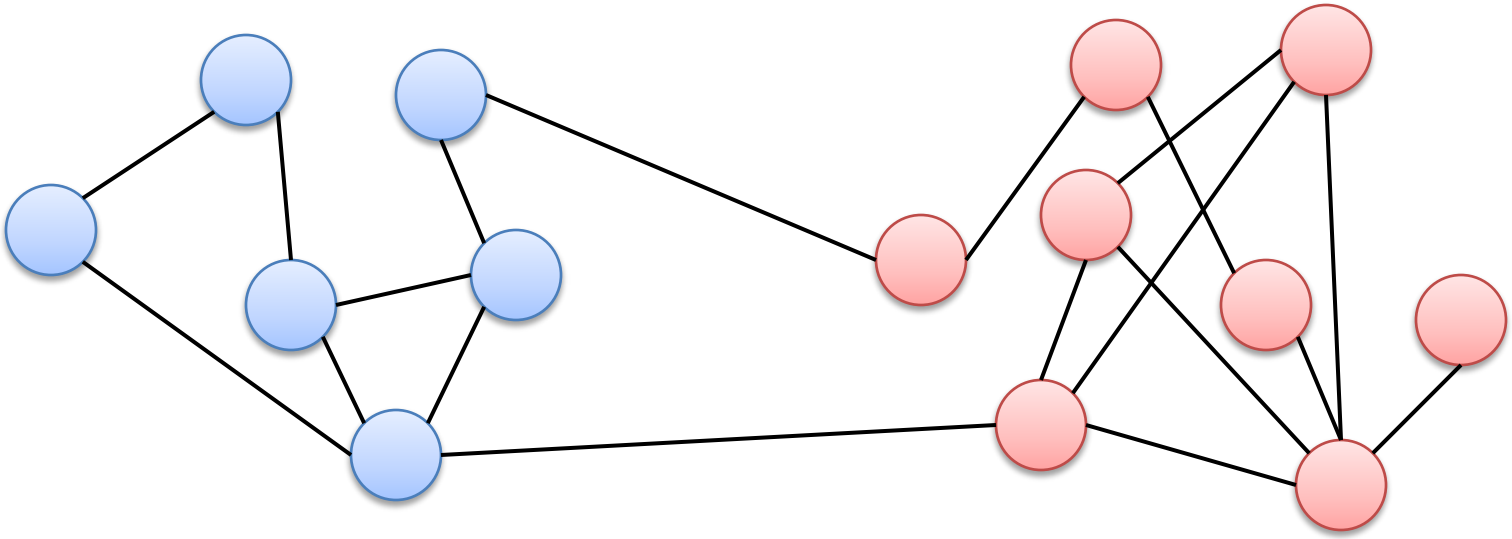
An Iterative and Divisive Approach

Remove edges that are most „between“ because they connect the different communities.





Example



Calculation of Edge-Betweenness

1. Shortest-path betweenness
   ⇨ amount of shortest paths running along one edge
2. Random-walk betweenness
   ⇨ net number of times a random walk signal passes one edge
3. Current-flow betweenness
   ⇨ value of current flow from source to sink along one edge

Calculation of Edge-Betweenness

1. Shortest-path betweenness
   ⇨ amount of shortest paths running along one edge
2. Random-walk betweenness
   ⇨ net number of times a random walk signal passes one edge
3. Current-flow betweenness
   ⇨ value of current flow from source to sink along one edge

- Produce the exact same output
- Only practical for small graphs due to bad performance

Calculation of Edge-Betweenness

1. Shortest-path betweenness
   ⇨ amount of shortest paths running along one edge
2. Random-walk betweenness
   ⇨ net number of times a random walk signal passes one edge
3. Current-flow betweenness
   ⇨ value of current flow from source to sink along one edge

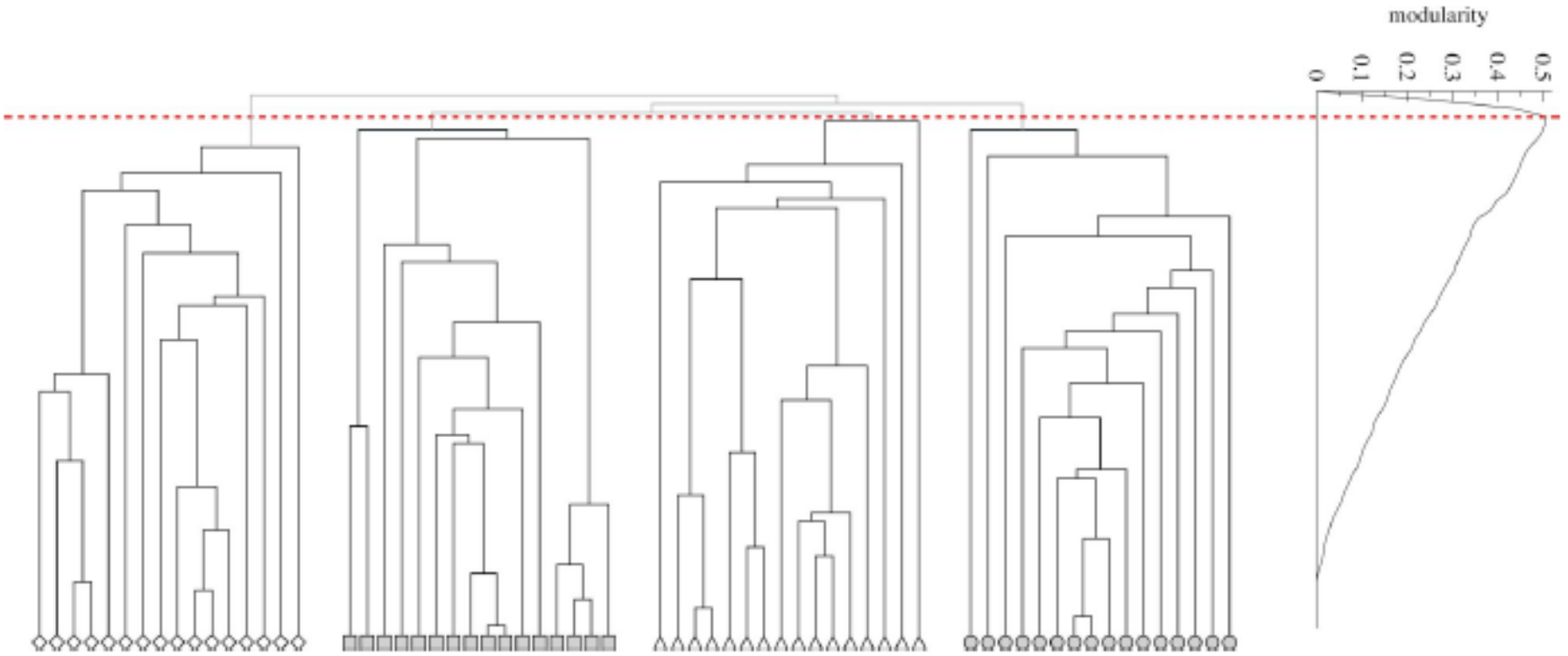
- Recommended by the authors
- Operates in $O(n^3)$
- Usable for networks of up to 10.000 vertices

Modularity

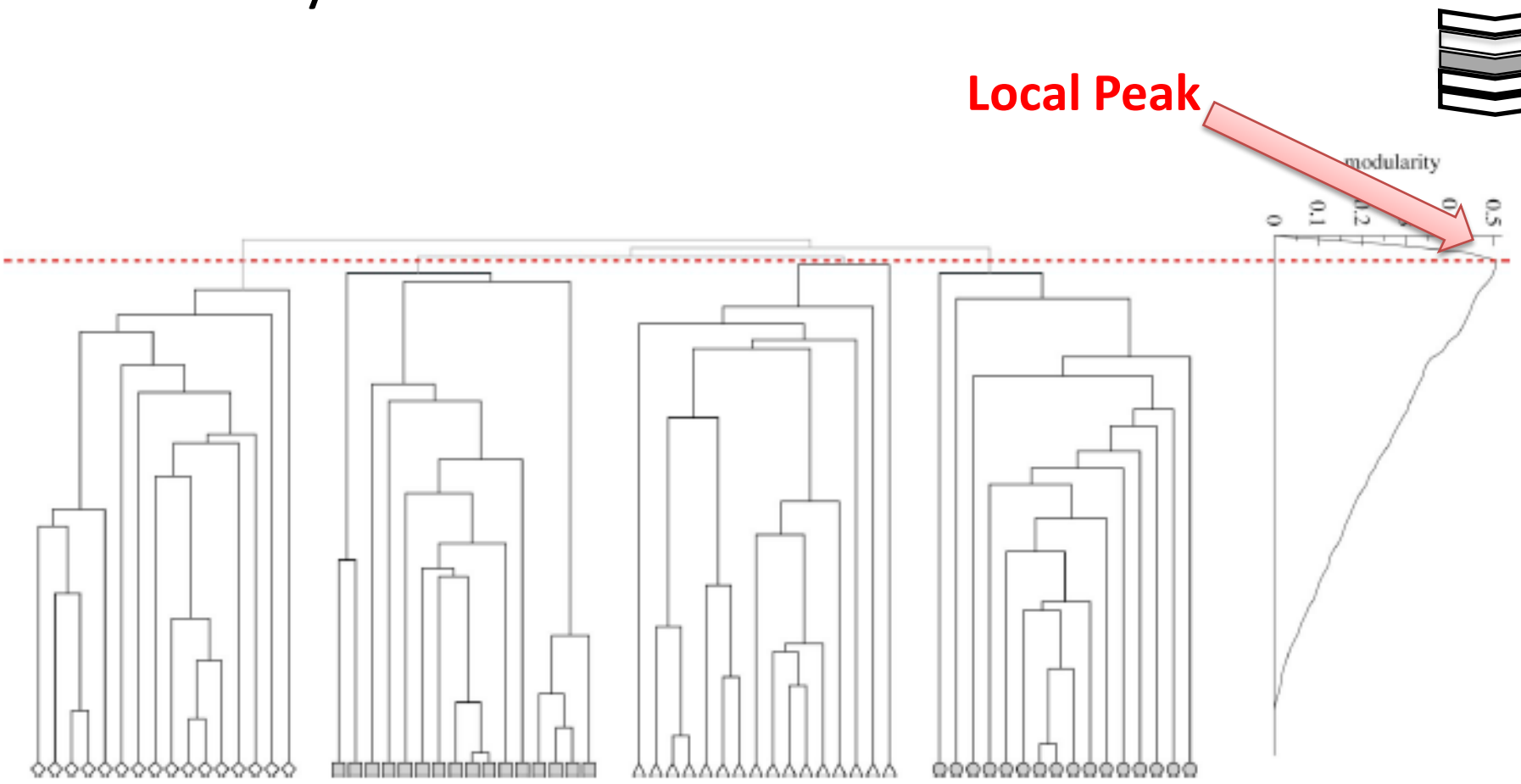
- Modularity Q is an objective quality measure for each split of a network into communities
- Measures the degree of correlation between the probability of having an edge joining two sites and the fact that the sites belong to the same community
- Calculated for each split while moving down a dendrogram to detect local peaks

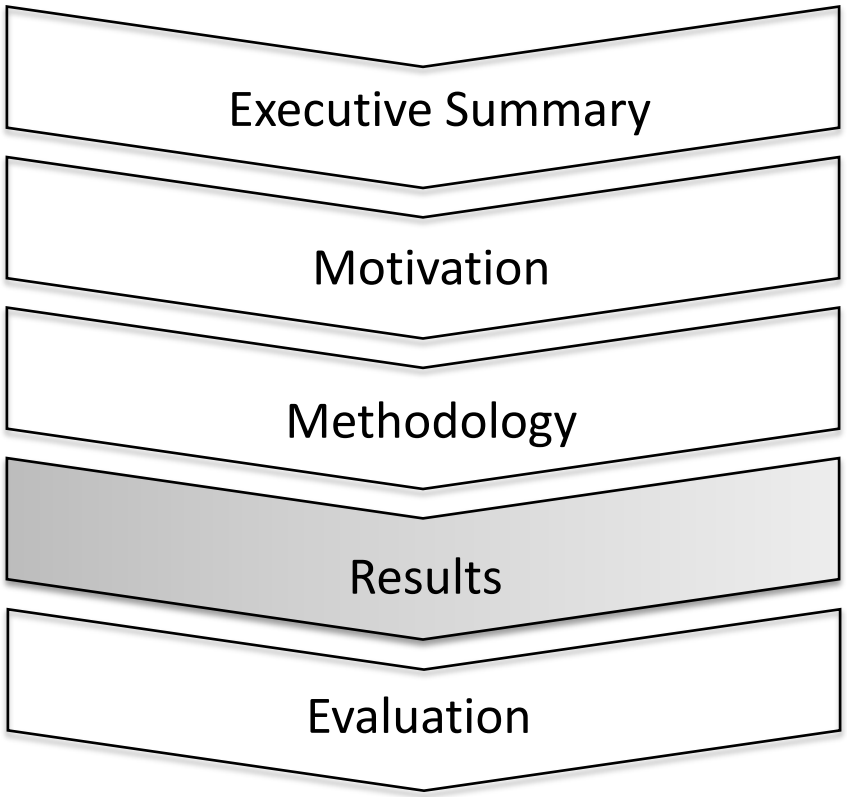
Modularity

Q = 1 strong community structures – in reality 0.3 - 0.7



Modularity



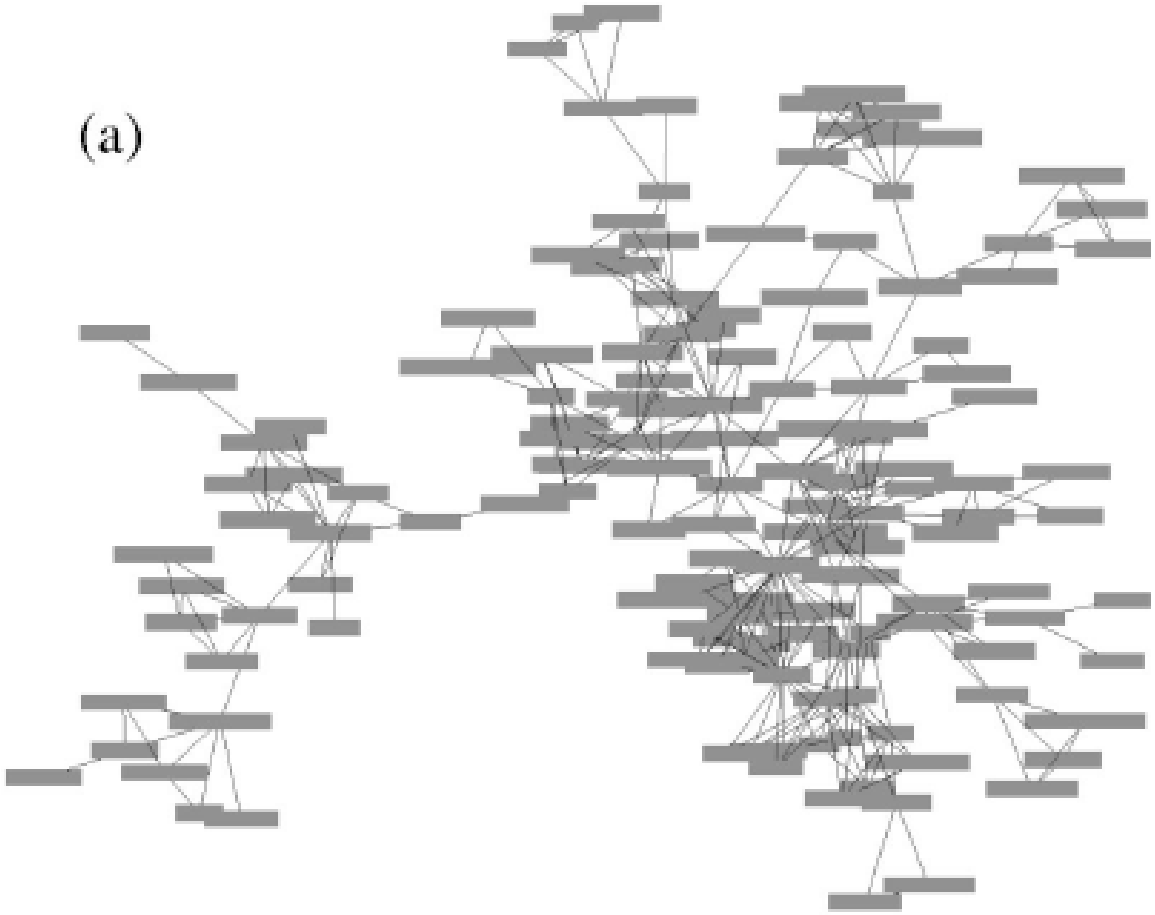

Executive Summary
Motivation
Methodology
Results
Evaluation

Results

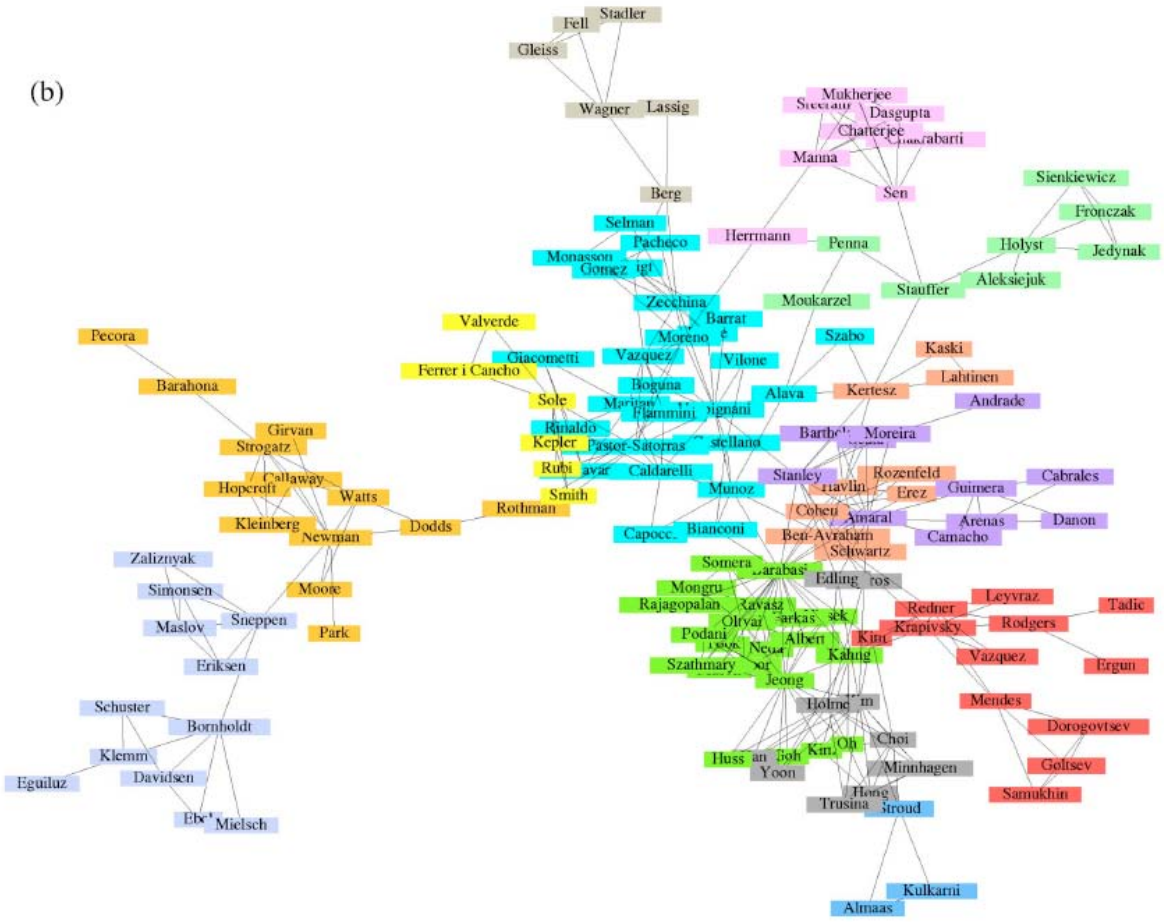
- Shortest-path algorithm delivers quality results in computer-generated and real-life network data
- Application: detection of community structures, analysis and visualization of difficult and not-obvious network structures

- Examples:
  1. Collaboration network of scientists
  2. Bottlenose dolphins

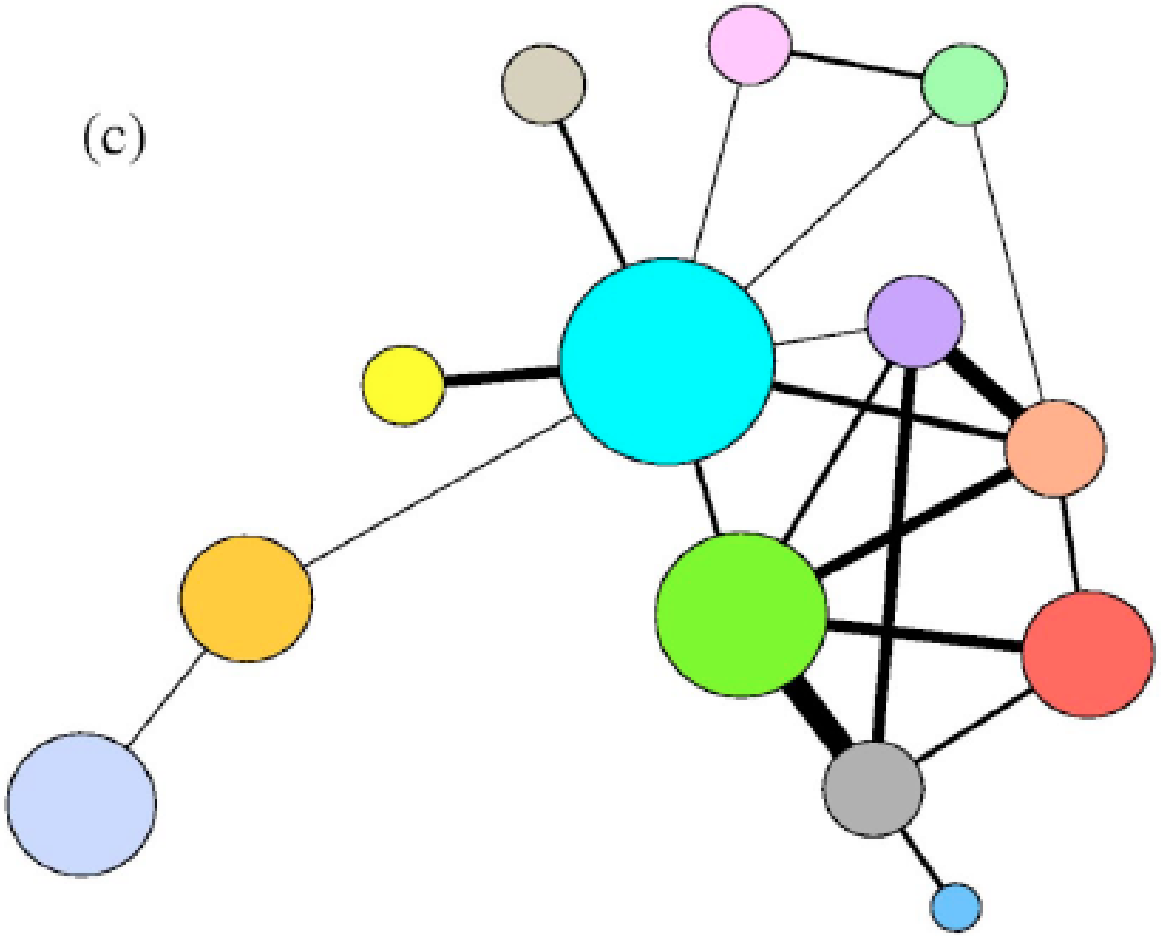
Collaboration Network of Scientists



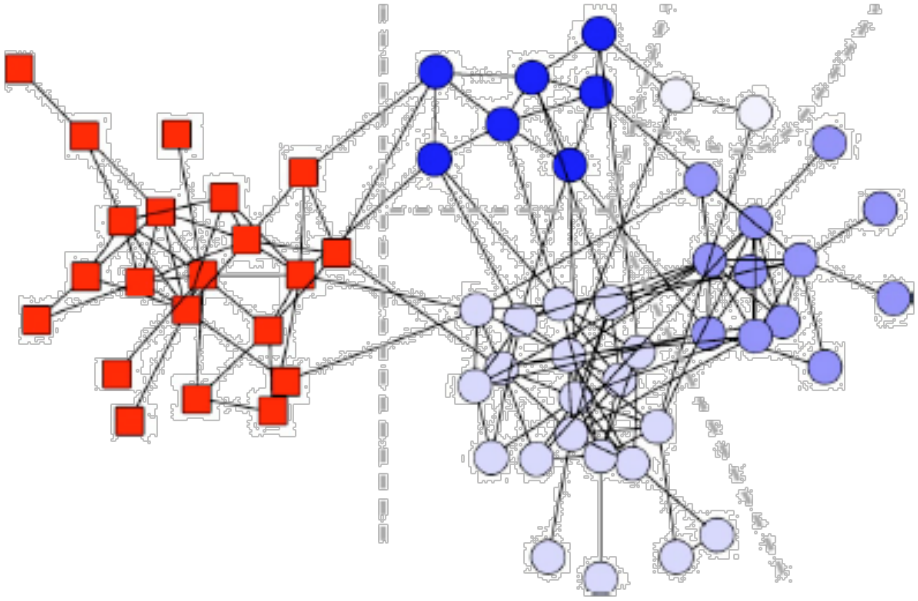
Collaboration Network of Scientists



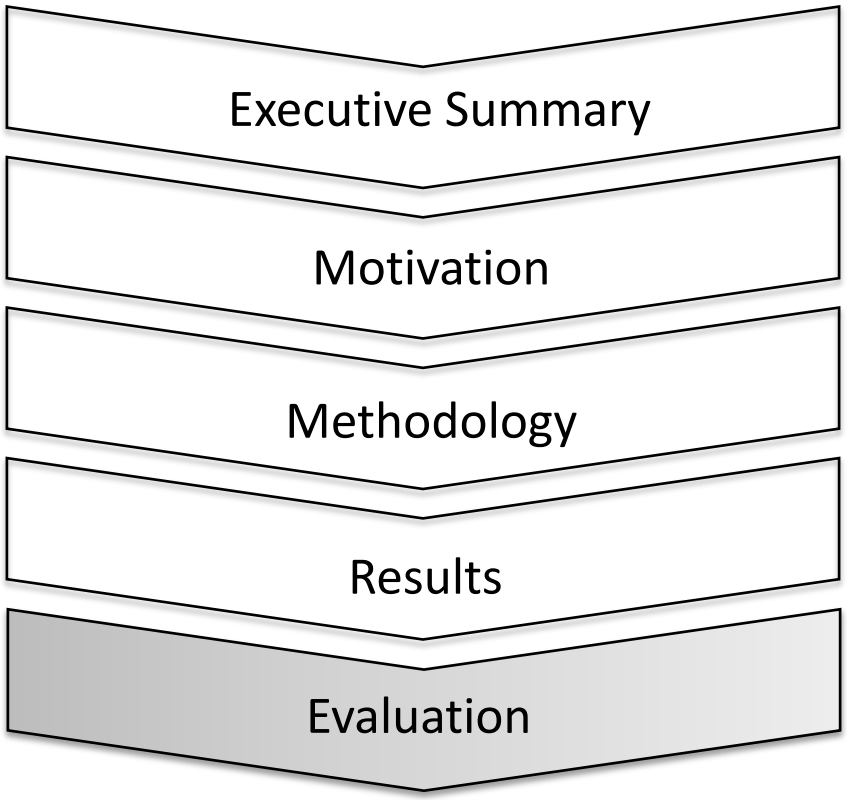
Collaboration Network of Scientists



Bottlenose Dolphins



- Split corresponds to a known division of the dolphin community
- Q = 0.52
- Closely related to evolution of community in human social networks


Executive Summary
Motivation
Methodology
Results
Evaluation

Evaluation

- Newman & Girvan present the first satisfactory and feasible solution for the task of finding community structures in networks
- Foundation for many similar approaches and variations of the presented set of algorithms

Problem	Possible Solutions
• Computing time $O(n^3)$ only reasonable for networks up to 10.000 vertices • In reality the networks of interest are much bigger	• Parallelization • Betweenness calculation for certain subsets • Greedy optimization of modularity • Stochastic blockmodels (2011)

Thank you very much!

…Questions?

"Harris, when I said 'any questions' I was using only a figure of speech."

Bildquellen

Folie 01+06: http://www.nieuwsmarkt.nl/wp-content/uploads/2011/09/comm.jpg

Folie 06: Newman, M. E. J. and Girvan, M. (2004): Finding and evaluating community structure in networks. In: Physical Review E 69, 026113 (2004), p.11.

Folie 10: Newman, M. E. J. and Girvan, M. (2004): Finding and evaluating community structure in networks. In: Physical Review E 69, 026113 (2004), p.2

Folie 16+17: Newman, M. E. J. and Girvan, M. (2004): Finding and evaluating community structure in networks. In: Physical Review E 69, 026113 (2004), p.8.

Folie 20-22: Newman, M. E. J. and Girvan, M. (2004): Finding and evaluating community structure in networks. In: Physical Review E 69, 026113 (2004), p.11.

Folie 23: http://static.guim.co.uk/sys-images/Guardian/Pix/pictures/2008/04/17/dolphin11a.jpg

Newman, M. E. J. and Girvan, M. (2004): Finding and evaluating community structure in networks. In: Physical Review E 69, 026113 (2004), p.12.

Folie 26: http://www.businesscartoons.co.uk/shop/images/uploads/3803bwc.gif

Folie28 : Newman, M. E. J. and Girvan, M. (2004): Finding and evaluating community structure in networks. In: Physical Review E 69, 026113 (2004), p.10.

Backup – Example Application Without Recalculation