Large-scale structure of complex networks (Part 2)

Snehal M. Shekatkar

Centre for modeling and simulation,
S.P. Pune University, Pune

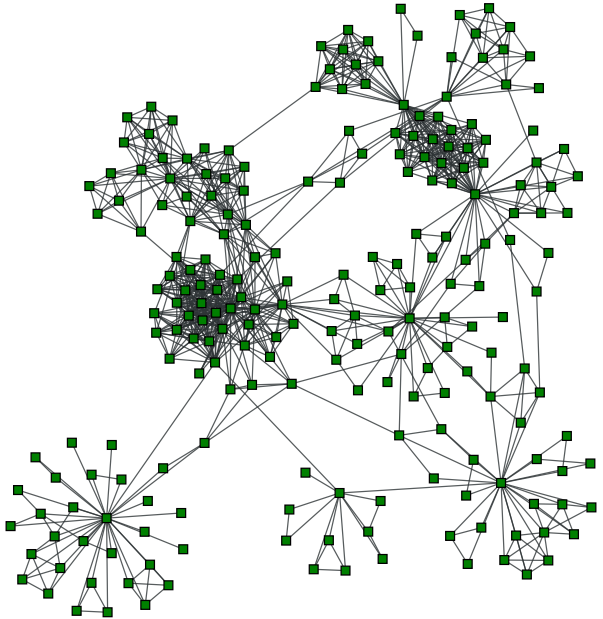
Large-scale structure of complex networks (Part 2)

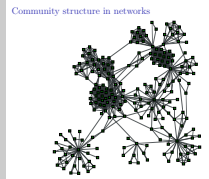
Snehal M. Shekatkar

Centre for modeling and simulation,
S.P. Pune University, Pune

Hello

Community structure in networks





Network of coauthorships in a university department

Community structure in networks

What are communities?

- **Traditional definition**: Groups of nodes with a high internal link density
- **Modern definition**: Nodes with similar connection probabilities to the rest of the network

Community structure in networks
What are communities?
- Traditional definition: Groups of nodes with a high internal link density
- Modern definition: Nodes with similar connection probabilities to the rest of the network

Communities in the real-world networks

- **Social networks**:
  - Friend-circles
  - Research communities
  - Co-workers
- **World Wide Web**:
  - Pages with similar contents
  - Webpages under the same domain (e.g. Wikipedia)
- **Biological network**:
  - Proteins with similar roles in protein interaction networks
  - Chemicals together taking part in chemical reactions in metabolic networks
  - Communities in neuronal networks

Communities in the real-world networks
- Social networks:
  - Friend-circles
  - Research communities
  - Co-workers
- World Wide Web:
  - Pages with similar contents
  - Webpages under the same domain (e.g. Wikipedia)
- Biological network:
  - Proteins with similar roles in protein interaction networks
  - Chemicals together taking part in chemical reactions in metabolic networks
  - Communities in neuronal networks

Community detection

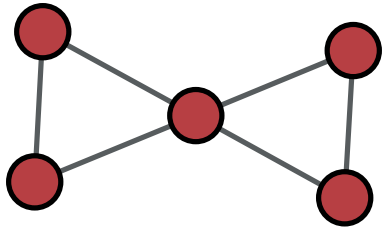
Detecting communities is important!

- Communities are building blocks of networks
- Communities allow us to see “the big picture”
- Functional/Autonomous units
- Non-trivial effects on the processes on networks

Community detection
Detecting communities is important!
- Communities are building blocks of networks
- Communities allow us to see “the big picture”
- Functional/Autonomous units
- Non-trivial effects on the processes on networks

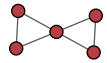
Graph partitioning

Problem of dividing a graph in a given number of groups of given sizes such that the number of links between the groups is minimized



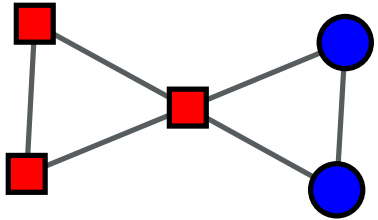
Graph partitioning

Problem of dividing a graph in a given number of groups of given sizes such that the number of links between the groups is minimized

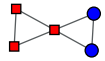


Graph partitioning

Problem of dividing a graph in a given number of groups of given sizes such that the number of links between the groups is minimized



Graph partitioning
Problem of dividing a graph in a given number of groups of given sizes such that the number of links between the groups is minimized



Partitioning is hard!

- Graph with n vertices
- Find two groups with sizes n_1 and n_2 such that the cut size is minimum
- Number of ways: $\frac{n!}{n_1!n_2!} \approx \frac{2^{n+1}}{\sqrt{n}}$

Partitioning is hard!
- Graph with n vertices
- Find two groups with sizes n_1 and n_2 such that the cut size is minimum
- Number of ways: $\frac{n!}{n_1!n_2!} \approx \frac{2^{n+1}}{\sqrt{n}}$

Community detection is harder!

- **Graph partitioning**
  - well defined
  - Number of groups is fixed
  - Sizes of the groups are fixed
  - Divide even if no good division exists
- **Community detection**
  - ill-defined
  - Number of groups depends on the structure of the network
  - Sizes of the groups depend on the structure of the network
  - Discover natural fault lines

Community detection is harder!
Graph partitioning
well defined
Number of groups is fixed
Sizes of the groups are fixed
Divide even if no good division exists
Community detection
ill-defined
Number of groups depends on the structure of the network
Sizes of the groups depend on the structure of the network
Discover natural fault lines

Too many algorithms

- Girvan-Newman algorithm
- Modularity maximization
- Spectral decomposition
- Clique-percolation
- Radom walk methods
- Statistical inference
- Label propagation
- Hierarchical clustering

Too many algorithms
- Girvan-Newman algorithm
- Modularity maximization
- Spectral decomposition
- Clique-percolation
- Radom walk methods
- Statistical inference
- Label propagation
- Hierarchical clustering

“The” simplest community detection problem

- Bisecting a graph with n nodes
- Group sizes are not fixed
- Minimum cut size?

“The” simplest community detection problem
- Bisecting a graph with n nodes
- Group sizes are not fixed
- Minimum cut size?

Quantification of community structure