Social Media Analytics: Graph Essentials

Lecture:8

Garima Jindal

BITS Pilani
Pilani Campus

Recap

Homophily
Communities and Applications
Types of communities
disjoint overlaping hierarchical, local
Node centric community detection cliques, K-cliques, Kclan, K-club, K-plex, K-core
Cut : ratio cut, normalized cut
Grivan newman - edge betweenness based
Modularity

Community Detection: Modularity

- Node-centric methods discussed so far are not very useful when the network is large
- Modularity comes from the word 'module'
- a network-centric metric to determine the quality of a community structure
- Based on the principle of comparison between
  - the actual number of edges in a subgraph and its expected number of edges
  - the expected number of edges is calculated by assuming a null model
- In the null model,
  - each vertex is randomly connected to other vertices irrespective of the community structure
  - However, some of the structural properties are preserved
  - One popular structural property is the degree distribution

$$Q = \frac{1}{2m} \sum_{ij} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \delta(c_i, c_j)$$

Where:

- A_{ij}: adjacency matrix
- k_i, k_j: degrees of nodes i and j
- m: total number of edges
- $\delta(c_i, c_j)$: 1 if nodes are in same community, else 0

what modularity value means

Modularity (Q)	Interpretation
~0	No community structure (random)
0.3 – 0.5	Moderate structure
> 0.5	Strong community structure

$$Q = \sum_{c} \left[\frac{l_c}{m} - \left(\frac{d_c}{2m} \right)^2 \right]$$

Where:

- l_c = number of **edges inside** community c
- d_c = **sum of degrees** of nodes in community c
- m = total edges

question

Single Community Contribution

-
  - Total edges: $m = 10$
  - A community has:
    - Internal edges: $l_c = 4$
    - Sum of degrees: $d_c = 8$

$$\boxed{Q = \sum_c \left[\frac{l_c}{m} - \left(\frac{d_c}{2m} \right)^2 \right]}$$

$$Q_c = \frac{l_c}{m} - \left(\frac{d_c}{2m}\right)^2$$

$$Q_c = \frac{4}{10} - \left(\frac{8}{20}\right)^2$$

$$Q_c = 0.4 - (0.4)^2$$

$$Q_c = 0.4 - 0.16 = \boxed{0.24}$$

Question 2

A network has:

- Total edges: $m = 14$

$$Q = \sum_c \left[\frac{l_c}{m} - \left(\frac{d_c}{2m} \right)^2 \right]$$

Community A:

- Internal edges: $l_A = 5$
- Sum of degrees: $d_A = 12$

Community B:

- Internal edges: $l_B = 6$
- Sum of degrees: $d_B = 16$

Step 1: Modularity of Community A

$$Q_A = \frac{5}{14} - \left(\frac{12}{28}\right)^2$$

$$Q_A = 0.357 - 0.184 = 0.173$$

Step 2: Modularity of Community B

$$Q_B = \frac{6}{14} - \left(\frac{16}{28}\right)^2$$

$$Q_B = 0.429 - 0.327 = 0.102$$

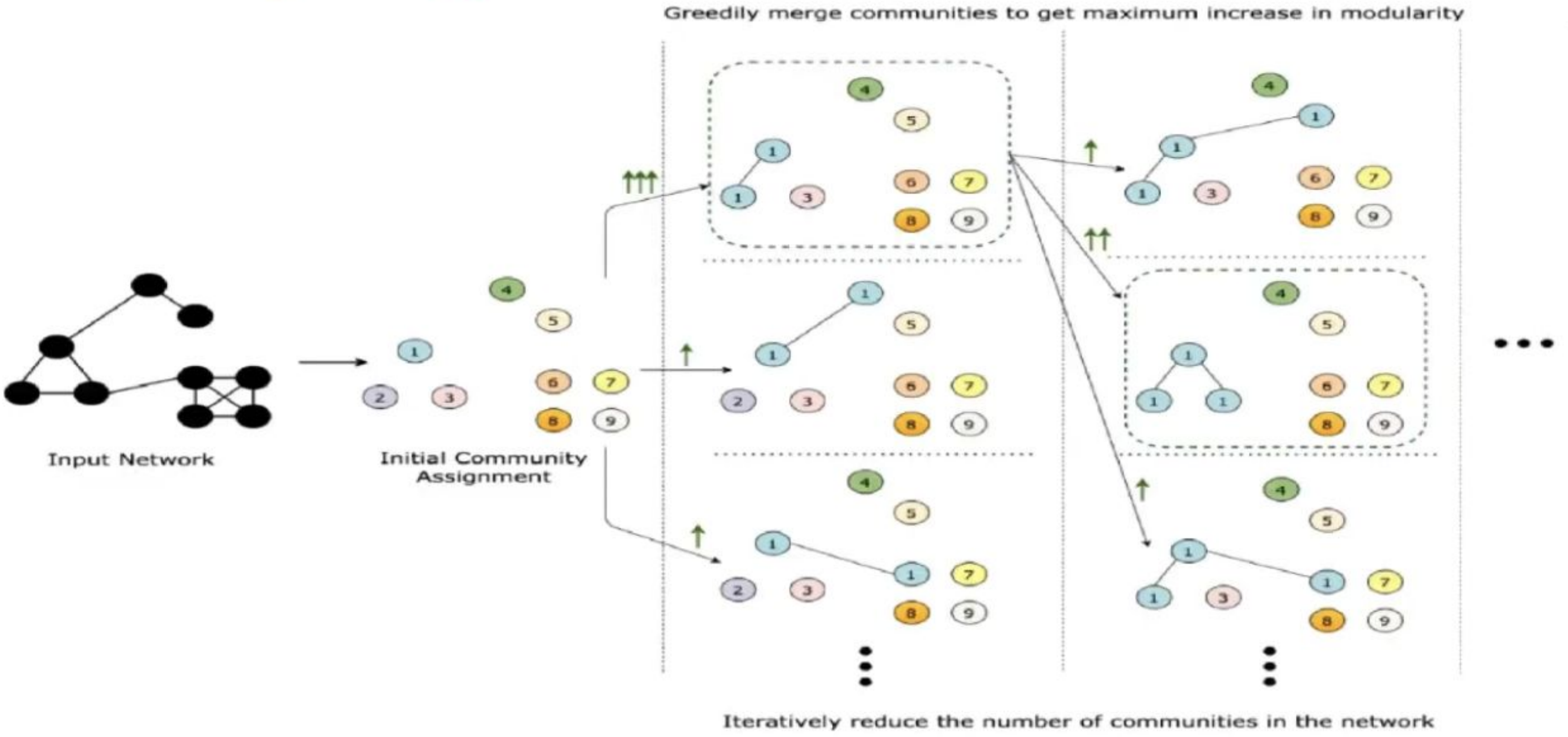
Step 3: Total Modularity

$$Q = Q_A + Q_B = 0.173 + 0.102 = \boxed{0.275}$$

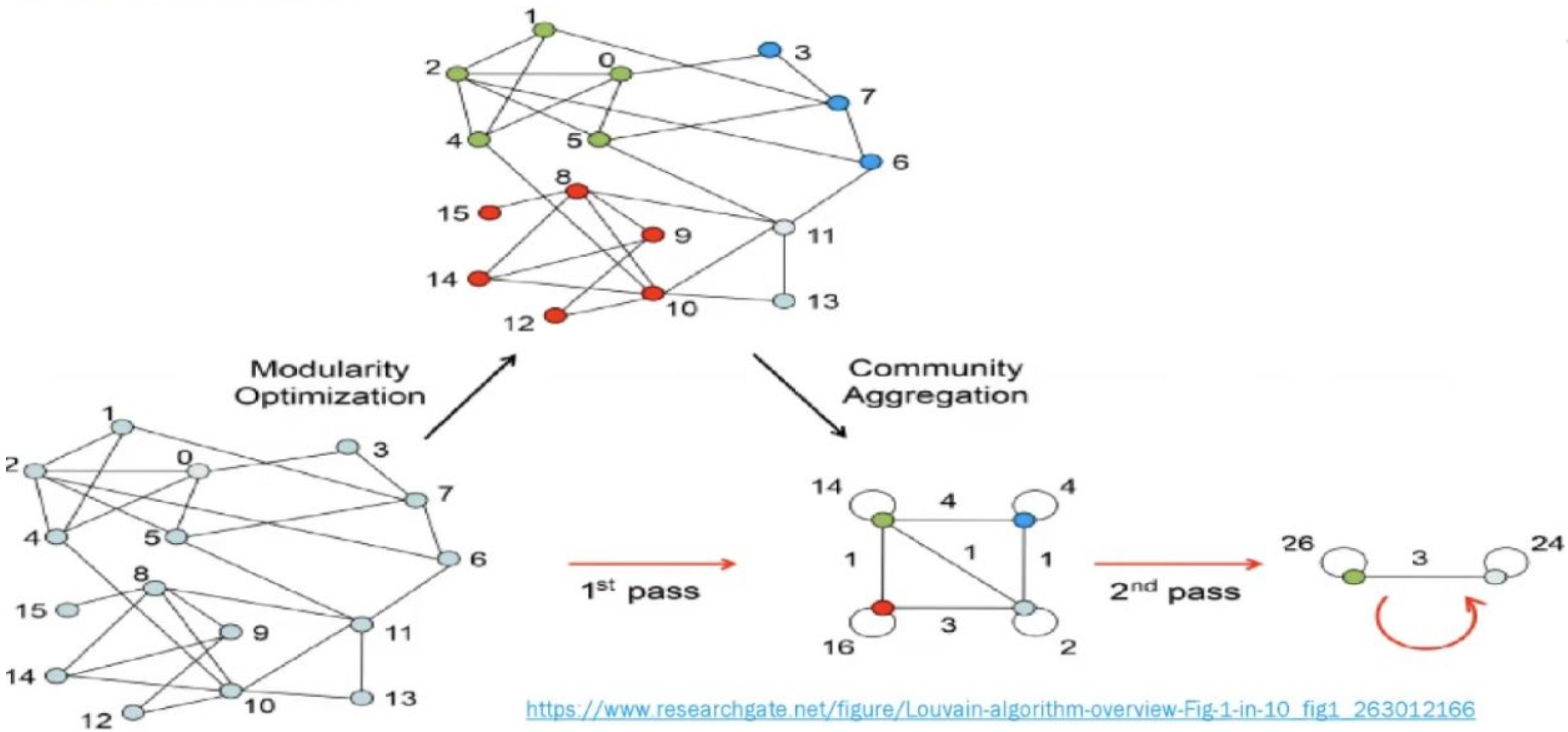
Community Detection: Modularity Maximization

- Modularity can be positive, negative, and zero
  - Positive modularity shows presence of strong community structure
- Networks with high modularity have dense connections between the nodes within modules but sparse connections between nodes in different modules.
- Different community assignments can lead to different values of modularity
- an assignment that maximizes the modularity of the overall network often finds the communities in the network
  - Fast Greedy Algorithm [Clauset et al. (2004)]
  - Louvain Method

Community Detection: Fast Greedy Algorithm



Community Detection: Louvain Method



https://www.researchgate.net/figure/Louvain-algorithm-overview-Fig-1-in-10_fig1_263012166

Community Detection through Modularity Maximization: Limitations

1) Resolution limit:
   - Well-connected smaller communities tend to get merged with larger communities even if the resultant communities are not that dense
   - Fails to detect those communities which are well-separated with densely connected intra-community nodes but only a single inter-community edge with the rest of the network

2) Degeneracy of solutions:
   - The case when there is an exponential number of community structures with same (maximum) modularity value

Modularity and community structure in networks

M. E. J. Newman*

Department of Physics and Center for the Study of Complex Systems, University of Michigan, Ann Arbor, MI 48109

Edited by Brian Skyrms, University of California, Irvine, CA, and approved April 19, 2006 (received for review February 26, 2006)

Many networks of interest in the sciences, including social networks, computer networks, and metabolic and regulatory networks, are found to divide naturally into communities or modules. The problem of detecting and characterizing this community structure is one of the outstanding issues in the study of networked systems. One highly effective approach is the optimization of the quality function known as "modularity" over the possible divisions of a network. Here I show that the modularity can be expressed in terms of the eigenvectors of a characteristic matrix for the network, which I call the modularity matrix, and that this expression leads to a spectral algorithm for community detection that returns results of demonstrably higher quality than competing methods in shorter running times. I illustrate the method with applications to several published network data sets.

clustering | partitioning | modules | metabolic network | social network

Many systems of scientific interest can be represented as networks, sets of nodes or vertices joined in pairs by lines or edges. Examples include the internet and the worldwide web, metabolic networks, food webs, neural networks, communication and distribution networks, and social networks. The study of

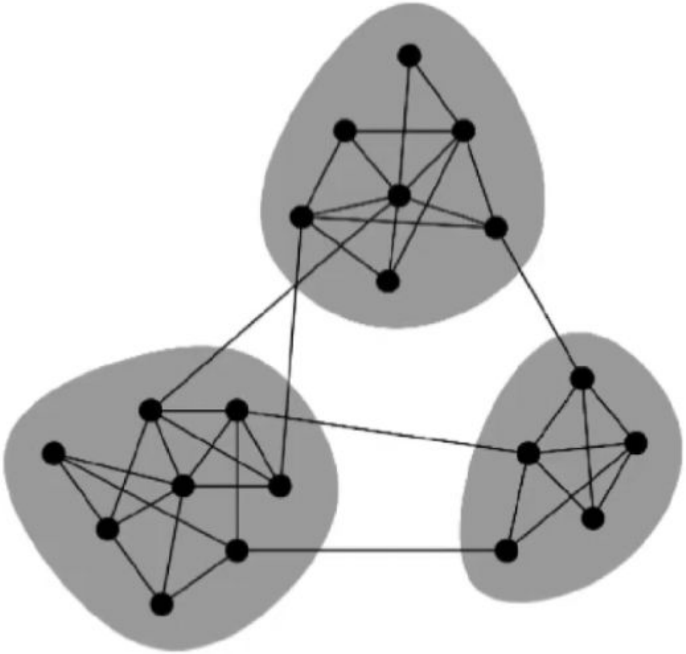


Fig. 1. The vertices in many networks fall naturally into groups or communities, sets of vertices (shaded) within which there are many edges, with only a smaller number of edges between vertices of different groups.

BITS Pilani
Pilani Campus

Questions?