Overlapping Community Detection in Social Networks

1

Overlapping Community Detection: Link Partition

❑Uses links (edges) in the networks to detect communities

❑Two major approaches
  ❑Create a link network and apply a node partitioning algorithm or disjoint community detection algorithm to find the community
  ❑Use similarity measures on the edges to find the communities directly by creating the dendrogram

❑Jaccard coefficient might be good choice for similarity measure

❑For two edges e_{ik} and e_{jk} connected to node k,

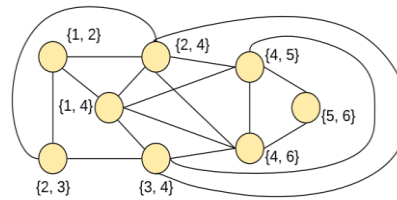$$Sim(e_{ik},e_{jk}) = \frac{|N_i \cap N_j|}{|N_i \cup N_j|}$$

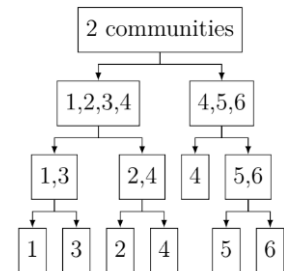Where N_i and N_j are the neighbours of the nodes i and j

2

Link Partition Method for Overlapping Community Detection: Illustration
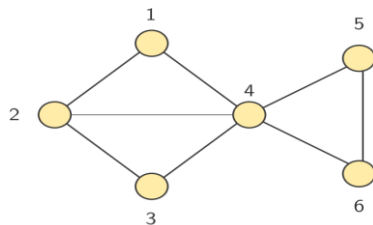
Example Network

Link Network for the example network

Dendrogram for the example network obtained using similarity measures

3

Overlapping Community Detection: BigClam

❑Proposed by Yang and Leskovec in 2013

❑Based on a generative modeling approach
  ▪ Define models that can generate the required network
  ▪ Find a model that generates a network that fits best

❑Affiliation Graphical Model (AGM) parameterized as $G(V, C, M, P_c)$ to generate a network
  ▪ V is the set of nodes of the network
  ▪ C is the set of communities
  ▪ M is the association between nodes and communities
  ▪ P_C is the probability of community C

❑Probability of edges between a pair of nodes x and y

$$P(x, y) = 1 - \prod_{c \in M_x \cap M_y} (1 - P_c)$$

where M_x and M_y are communities x and y belongs to

4

Overlapping Community Detection: BigClam
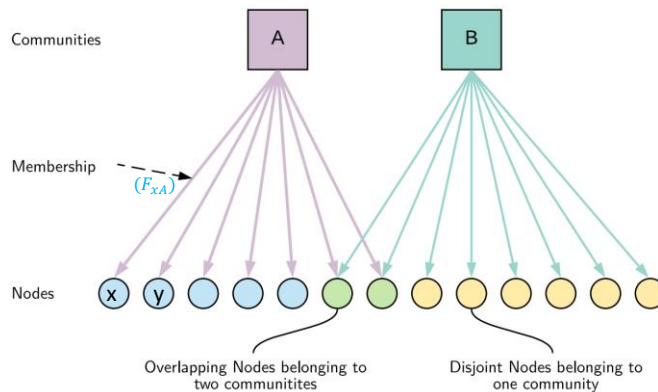
- F_{xA}: Membership strength of node x to community A

- $F_{xA} = 0$: No Membership

- Each community links nodes independently

- Probability of two edges in one community
  $$P_A(x, y) = 1 - \exp(-F_{xA} \cdot F_{yA})$$

- Overall probability of an edge between two nodes
  $$P(x, y) = 1 - \exp(-F_x \cdot F_y^T)$$

5

Overlapping Community Detection: BigClam

❑Given the above, we need to find values for matrix F

❑Estimate F such that probability of network G can be maximized.

❑Above can be solved just by applying gradient descent method

6