Project Phase 1

Community Computation

|  |  |  |
| --- | --- | --- |
| Method | What’s that | Algorithm |
| Attribute-Based Clustering | This approach involves clustering nodes based on their attribute similarity. | k-means  hierarchical clustering  DBSCAN |
| Modularity Optimization | Modularity is a widely used measure to evaluate the quality of community partitions in networks. | Louvain method  Infomap algorithm |
| Network Embedding | Transform the network into a low-dimensional vector space where nodes with similar attributes or structural roles are close to each other. | node2vec  GraphSAGE |
| Graph Convolutional Networks (GCNs) | GCNs are a type of neural network that can be applied to attributed networks for community detection. |  |
| Tensor Factorization |  | CANDECOMP/PARAFAC (CP) decomposition |

Membership Score Calculation

Modularity-based Methods:

Modularity optimization algorithms, such as the Louvain method or the Infomap algorithm, can be used to identify communities in a network. Once communities are detected, nodes can be assigned membership scores indicating their belongingness to each community. These scores are often based on the fraction of intra-community connections compared to total connections.

Fuzzy Clustering:

Fuzzy clustering techniques, like Fuzzy C-Means (FCM), assign membership scores to nodes on a continuous scale, indicating the degree of association with each community. FCM, for example, assigns membership values to each cluster, allowing nodes to belong partially to multiple communities.

Probabilistic Models:

Some algorithms model community membership as a probabilistic distribution over communities. For example, the Stochastic Block Model (SBM) estimates the probability of a node belonging to each community, providing a probabilistic view of community membership.

Node2Vec and Graph Embeddings:

Node embedding techniques, like Node2Vec and GraphSAGE, can be used to learn low-dimensional representations of nodes based on their attributes and network structure. These embeddings can then be used to calculate membership scores by measuring the similarity between a node's embedding and the embeddings of community representatives.

Attribute-based Methods:

Membership scores can be calculated by comparing the attributes of nodes to the average or centroid attributes of communities. Nodes that are more similar in terms of attributes to a community's centroid are assigned higher membership scores.

Spectral Clustering:

Spectral clustering methods can be used to find communities in the network, and membership scores can be calculated based on the eigenvectors corresponding to the community structure. Nodes with higher values in these eigenvectors have stronger membership in the corresponding community.

Deep Learning-based Methods:

Deep learning models, such as graph convolutional networks (GCNs) or graph attention networks (GATs), can be employed to predict community membership scores. These models can capture complex relationships between attributes and network structure.

Community Detection and Attribute Agreement:

Algorithms that aim to find a balance between community structure and attribute similarity can be used to calculate membership scores. These methods optimize the agreement between communities and attributes.

Tensor Factorization:

Tensor factorization techniques, such as CANDECOMP/PARAFAC (CP) decomposition, can be used to simultaneously factorize the network's attribute tensor and the community membership tensor to estimate membership scores.

Iterative Ranking Update  
  
PageRank with Attribute Propagation (PAP):

PageRank is a classic algorithm for ranking nodes in a network. PAP extends PageRank by incorporating node attributes into the ranking process. It iteratively updates the ranking scores by considering both the network structure and attribute similarities between nodes.

Attributed PageRank:

Like PAP, this algorithm combines network structure and node attributes to rank nodes. It iteratively updates node rankings based on a combination of their attribute similarity and the PageRank scores of their neighbors.

Personalized PageRank:

Personalized PageRank focuses on finding nodes that are important to a specific seed node. In attributed networks, personalized PageRank can be adapted to consider attribute similarity as well as structural connectivity in the iterative ranking process.

Random Walk with Restart (RWR):

RWR is a variant of personalized PageRank where random walks are performed from a seed node, and at each step, there is a probability of restarting the walk from the seed node. Attribute information can be integrated into RWR to update rankings iteratively.

Heat Kernel Ranking (HKR):

HKR is based on the heat diffusion process on the network. It assigns scores to nodes based on how fast heat diffuses from a source node to other nodes. Attribute information can be incorporated into HKR to influence the heat diffusion process.

Node Similarity-based Ranking:

This approach uses node similarity measures, which consider both attributes and network structure, to iteratively update node rankings. Nodes that are more similar in terms of attributes and network connections receive higher rankings.

Community Detection-based Ranking:

In attributed networks, community detection algorithms can be used to identify communities of nodes with similar attributes. Node rankings can then be updated iteratively within each community, considering both attributes and community structure.

Deep Learning-based Ranking:

Deep learning models, such as graph neural networks (GNNs), can be used for iterative ranking updates. GNNs can learn node representations by aggregating information from attributes and neighboring nodes and update node rankings iteratively.

Matrix Factorization-based Ranking:

Matrix factorization techniques can be applied to attributed networks to factorize the network's adjacency matrix and attribute matrix. The factors can be used to update node rankings iteratively based on both network structure and attributes.