Unit IV - Social Recommendation

- Group recommendation
  1. Maximizing average satisfaction
  2. Least misery
  3. Most pleasure
- Hybrid recommender

  Pipeline hybridization - puts recommender systems in a line and the output of one system is the input of another.
  - Predictive accuracy
  - Classification accuracy
  - Rank accuracy

  Spearman’s rank correlation

  Kendall’s T
- Link level features
  1. Distance based features
     Shortest path length
  2. Local neighbourhood overlap
  3. Global neighbourhood overlap
- Katz and Simrank
- Spectrum of a graph
  - Eigenvalues of G are the eigenvalues of A where A is the adjacency matrix
  - Characteristic polynomial, phi(g;lambda)
  - Spectrum of G is the eigenvectors xi of a graph ordered by their strength

  Spectral theorem

Any real symmetric n x n matrix has real eigenvalues and n orthonormal eigenvectors

▼ Spectral clustering as an optimization problem for graphs

Spectral clustering algorithm

1. Preprocessing
   Construct a matrix of the graph
2. Decomposition
   Find the eigenvalues and eigenvectors of the graph. Map each point to a lower dimensional representation
3. Grouping
   Assign points to two or more clusters based on new representation.

Partitioning a graph into k - clusters -

1. Recursive bipartitioning
2. Cluster multiple eigenvectors

The most stable clustering is generally given by the value k that maximizes eigengap.

▼ Semi-supervised learning with label propagation

Assumptions in semi supervised learning -

1. Smoothness - if two samples are close, their labels will also be close
2. Low density assumption
3. Manifold assumption - data points on the same low dimensional manifold should have the same label
4. Cluster assumption - data points belonging to the same cluster belong to the same class.

Label propagation algorithm

1. Create a connected graph

2. Determine the weights
3. Perform random walk

Label Spreading Algorithm

Uses symmetric normalized matrix while LPA uses random walk normalized matrix.

LPA uses hard clamping whereas LSA uses soft clamping controlled by hyperparameter alpha.

1. Define a pairwise relation between points called an affinity matrix
2. Use RBF radial basis function to create this matrix
3. Diagonals have weight 0
4. Use matrix multiplication to spread information from labeled points to unlabeled points

▼ Traditional ML

Node features for interconnectivity

1. Weiner index - Sum of shortest paths between each pair of reachable nodes.
2. Closeness vitality - change in weiner index when node is excluded
3. Effective size
4. Small worldness
5. Communicability

Graph ML Downstream tasks

- Node level
- Edge level
- Community level - ETA

- Clustering - disease drug interaction networks

- Visualization
- Reconstruction - use similarity between low dimensional embedding to reconstruct the original graph
- Graph generation - drug discovery
- Graph evolution - how it evolve over time

Evaluation metrics

Classification

1. Accuracy, Precision, Recall, F1 score, F1 Macro, F1 Microweighted
2. Precision at K

Clustering

1. Purity
2. Normalized Mutual Information