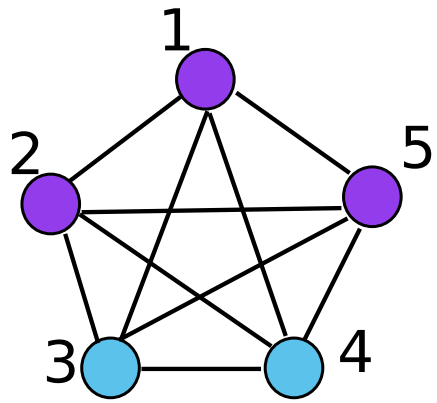
Signal Processing Techniques for Interpolation in Graph Structured Data

Reporter: Feng Zhao
Department of Electronic Engineering

April 23, 2018

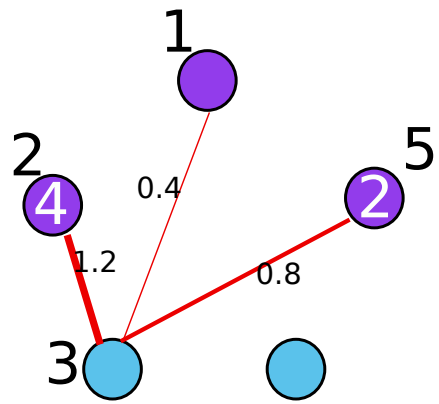
An example of recommendation system

- Five movies
- a 5×5 similarity matrix
- a user rating on movie 1,2,5
- predict his rating on movie 3,4



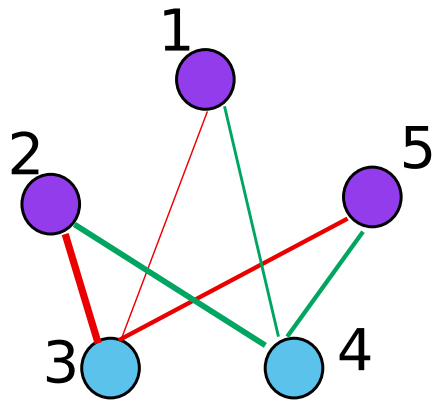
kNN Method to predict user rating on movies

- predict on movie 3
- suppose k=2
- movie 2 and 5 are more similar to 3 than 1
- predicted rating on movie 3: $\frac{1.2\times 4+0.8\times 2}{1.2+0.8}$



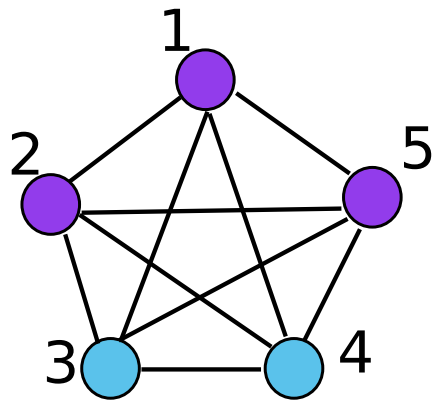
issues of kNN

- discard mutual information between movie 1,2,5
- predict rating 3,4 respectively
- jointly prediction with all information used will improve the *accuracy* but also increase the *complexity*.



Problem Reformulation

- $f(1), f(2), f(5)$ are known
- *interpolate* $f(3), f(4)$
- $\boldsymbol{f} = [f(1), f(2), f(3), f(4), f(5)]^T$
- $\boldsymbol{f} = f(1)\boldsymbol{\Theta}_1 + f(2)\boldsymbol{\Theta}_2 + f(5)\boldsymbol{\Theta}_5$



What is the space spanned by $\boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2, \boldsymbol{\Theta}_5$

band limited graph signal reconstruction

Nyquist Shannon sampling theorem

If a continuous signal is band-limited, it can be reconstructed by discrete sampling without loss.

Pesenson, Isaac

If a graph signal is band-limited, it is uniquely determined by their values on some sets of vertices.

Reconstruction from basis

Critial Frequency

If the graph signal is band-limited, its spectrum decomposition has only components less than critical frequency ω_S^*.

$$\boldsymbol{f} = \sum_i x_i \boldsymbol{v}_i \tag{1}$$

where $\boldsymbol{v}_i$ is the eigenvector with eigenvalues less than ω_S^*.

Least Square

$\boldsymbol{f}$ known only partially, least square techinque can be used to solve x_i

Numerical Results

Dataset

- Student body test data with 488 records and 37 features.
- Use other features(excluding age) to predict the gender
- 80% training set, 20% test set, 5 fold cross validation

Algorithm

- **Linear Discrimant Analysis** with average error rate at 0.4%
- **Graph Interpolation** with average error rate at 3%

Problems for Graph Interpolation

- Edge weight from data features
- Heuristic critical frequency

Acknowledgement

Thanks for listening