Graph link prediction

Dmitry Denisenko, Aliya Janabekova, Ilia Nasedkin, Ilya Osokin, Artem Pimkin
Skoltech 2018

Plan

- Introduction, problem
- Naive approach
- Spectral clustering
- Results
- Discussion

Problem

Statement:

Predict probability of connection between two nodes, using prior knowledge about the graph structure and/or statistical analysis

Dataset:

- Adjacency matrix
- Sparse
- Zeros and ones

Applications:

- Graph restoration
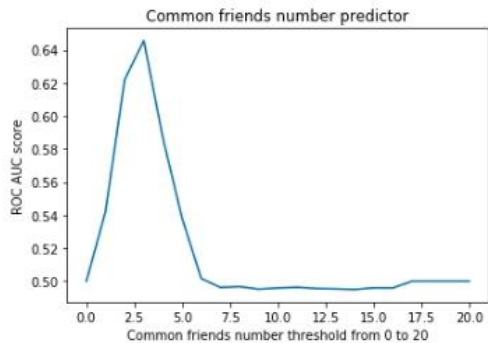- Advising new friends in social networks

Assumptions

- Common Friends:

$$|\mathcal{N}(u) \cap \mathcal{N}(v)|$$

- Jaccard's Coefficient:

$$\frac{|\mathcal{N}(u) \cap \mathcal{N}(v)|}{|\mathcal{N}(u) \cup \mathcal{N}(v)|}$$

- Adamic-Adar Score:

$$\sum_{t \in \mathcal{N}(u) \cap \mathcal{N}(v)} \frac{1}{\log |\mathcal{N}(t)|}$$
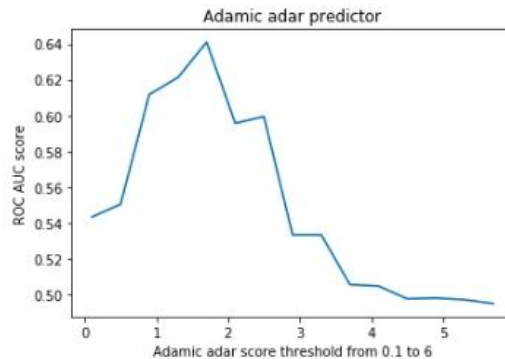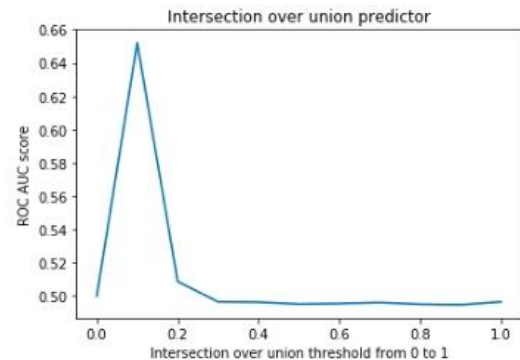
Results

Best score: 0.646 +- 0.044

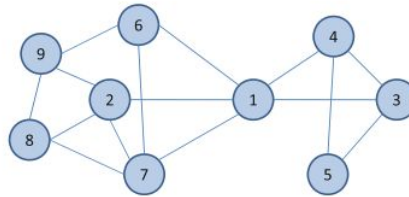Common friends number predictor

Best score: 0.641 +- 0.052

Adamic adar predictor

Best score: 0.652 +- 0.047

Intersection over union predictor

Spectral graph partition

Node 1's local network

Different affiliations

Affiliation-1

Affiliation-2
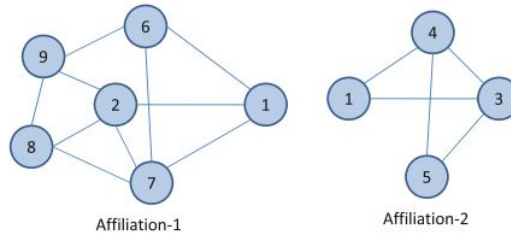
Relaxation

$$\min\, cut(C_1, C_2, \ldots, C_k) = \sum_{i=1}^{k} cut(C_i, V/C_i)$$

$$\implies$$

$$\min_{S}\; Tr(S^T \tilde{L} S)$$

$$s.t.\; S^T S = I$$

$$\tilde{L} = D^{-1/2} L D^{-1/2} = I - D^{-1/2} A D^{-1/2}$$

Leveraging social media networks for classification:
https://link.springer.com/content/pdf/10.1007%2Fs10618-010-0210-x.pdf

Random walk&word2vec

- Random paths are sentences
- Apply word2vec as it is
- Did not work out :(
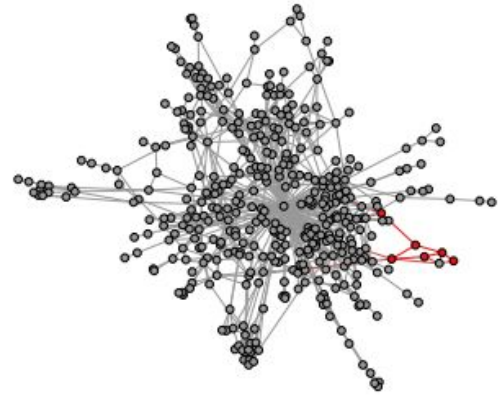
(a) Random walk generation.

Inspired by DeepWalk: Online Learning of Social Representations:
http://www.perozzi.net/publications/14_kdd_deepwalk.pdf

Results comparison

- ROC AUC for basic approach: 0.652 ± 0.047
- ROC AUC for spectral partitioning: 0.887 ± 0.023

Discussion

- Spectral approach is better than the heuristical
- But for bigger graphs online algorithms (node2vec, etc.) are better
- Random walk&word2vec did not succeed

Thank you for your attention!