

Recommendation System Based on Statistically Validation of Bipartite Structure Networks on IMDB Database

Onur Poyraz

Boğaziçi University
CMPE 556 - Complex Networks
Final Presentation

June 3, 2017

1 Introduction

2 Methods

3 Results and Discussion

- Example Data From Tumminello et al
- IMDB Movie Lens Data

What is Bipartite Complex Systems

Composition of two different sets of nodes such that every link connects a node of the first set with a node of the second set.

Key Terms and Specifications

- Projected Networks
- Heterogeneity
- Preferential Links
- Null Hypothesis

Statistically Validated Networks

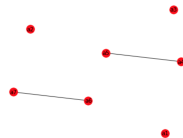
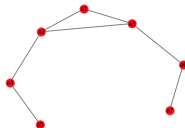
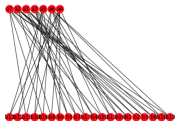
A method that allowing to statistically validate whether a given link in the projected network is consistent or not with a null hypothesis of random connectivity between elements of the bipartite network.

- This is unsupervised process.

The Roadmap

- Introducing Null Hypothesis
- Associating p-value
- Statistical Significance Level

Application of Method in Graphs



- Determine bipartite system S and divide it set A and set B .
- Project Network on one of these subdomain
- Determine Adjacency Projected Network
- Statistically Validate Each Link against null hypothesis that takes into account the degree heterogeneity of elements of both set A and set B
- Decompose bipartite system into subsystems

$$H(X|N_B^k, N_i^k, N_j^k) = \frac{\binom{N_i^k}{X} \binom{N_B^k - N_i^k}{N_j^k - X}}{\binom{N_B^k}{N_j^k}}.$$

$$p(N_{ij}^k) = 1 - \sum_{X=0}^{N_{ij}^k - 1} H(X|N_B^k, N_i^k, N_j^k).$$

This formulas has shown how to associate a p-value with the link between each pair of elements i and j of the projected network for each subsystem S .

Statistical Significance

The parameter which takes into account the fact that we are performing multiple hypothesis testing.

- Determine N_t
- Determine s
- Compare each p-value with s . If $p < s$ then we validate the link.
- Adjust weight as a total number of subsystems which are statistically validated.
- If weight becomes 0 then remove link.
- The resulted weighted network is statistically validated.

Tools Used

- Numpy
- Pandas
- Networkx
- Matplotlib
- Scipy

1 Introduction

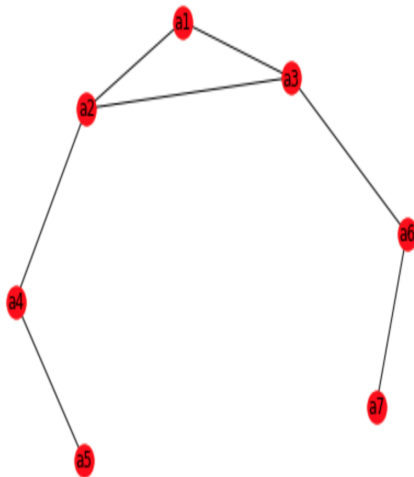
2 Methods

3 Results and Discussion

- Example Data From Tumminello et al
- IMDB Movie Lens Data

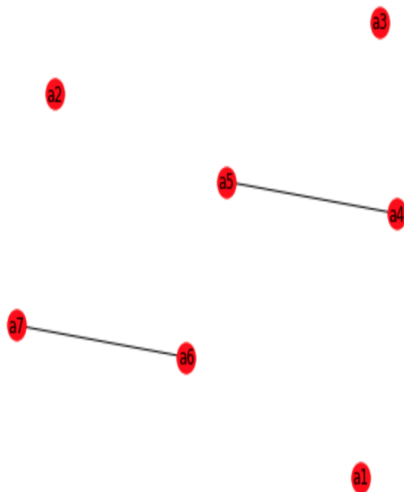
Example Data From Tumminello et al

The projected network is given below;



Example Data From Tumminello et al

After applied statistical validation



- Well suited example data
- Shows heterogeneity
- Clear Results

1 Introduction

2 Methods

3 Results and Discussion

- Example Data From Tumminello et al
- IMDB Movie Lens Data

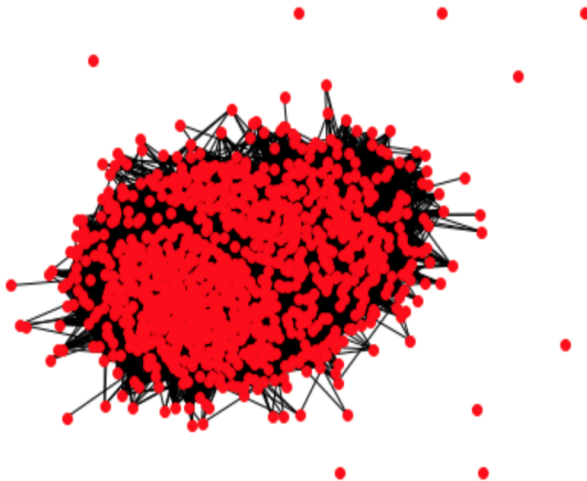
IMDB User-Movie-Rating

The projected network is given below;



IMDB User-Movie-Rating

After applied statistically validation



Recommendation

- Create statistically validated network
- Strong connections remains the others gone
- Make recommendation according to linkages.

- This method allows to validate links describing preferential relationships among the heterogeneous elements of bipartite complex systems.
- From that point we can make recommendation system.
- There could be further complex clustering algorithms applied on that networks which can give better results.

Thank you ..