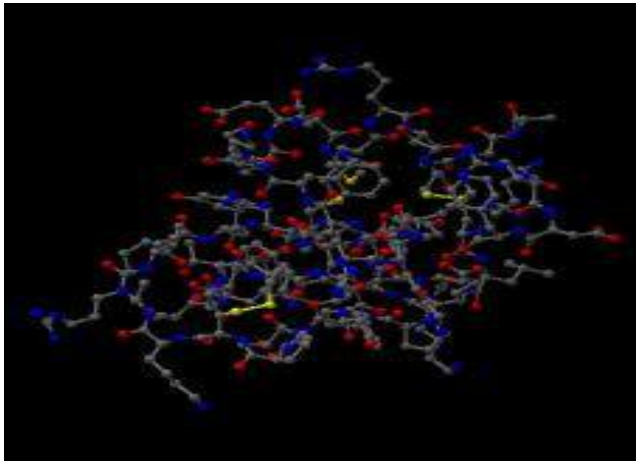
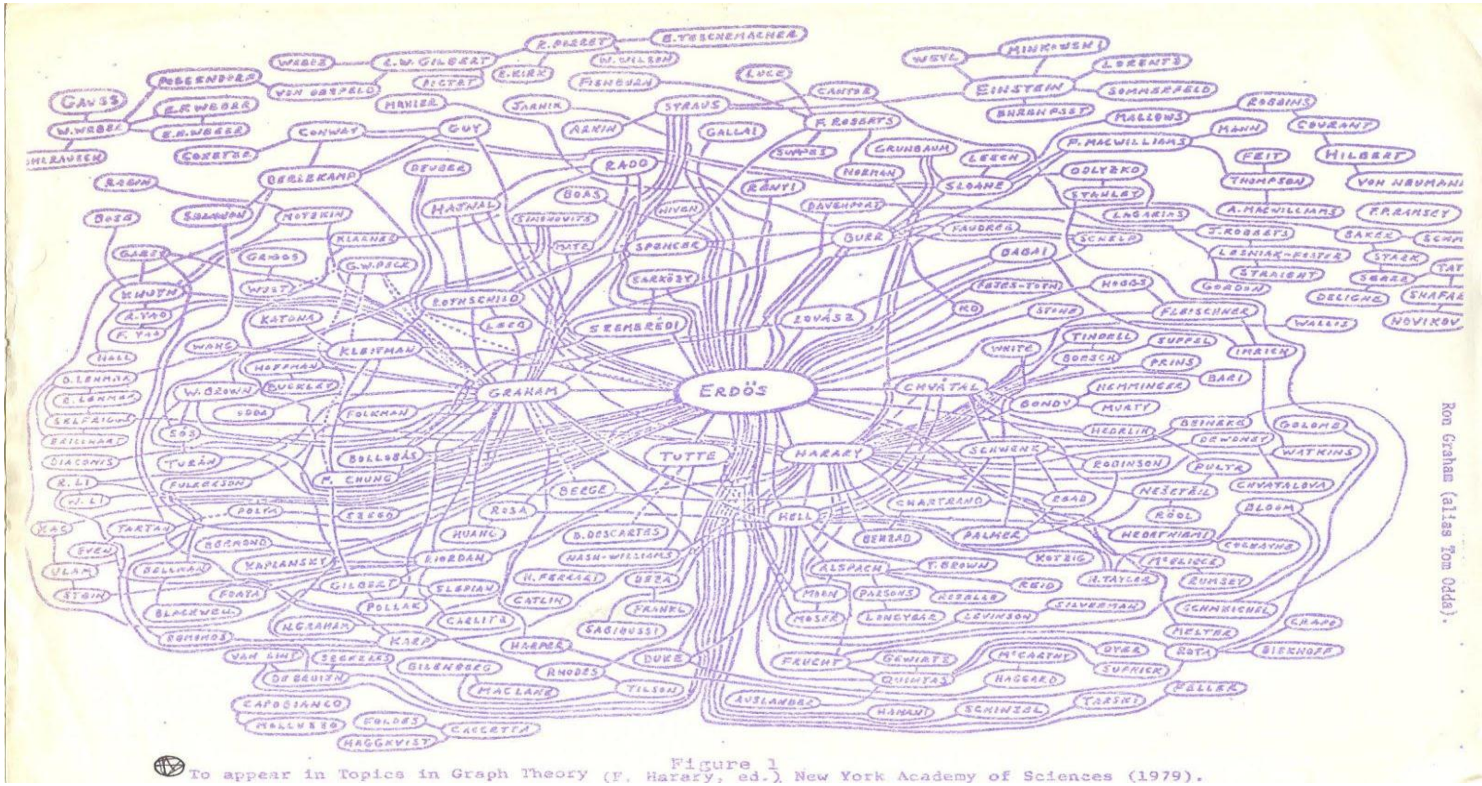
GRAPH KERNELS

Aashay Harlalka
Ankit Agrawal

Social Network



Protein Molecule



Erdos Graph

Common Tasks

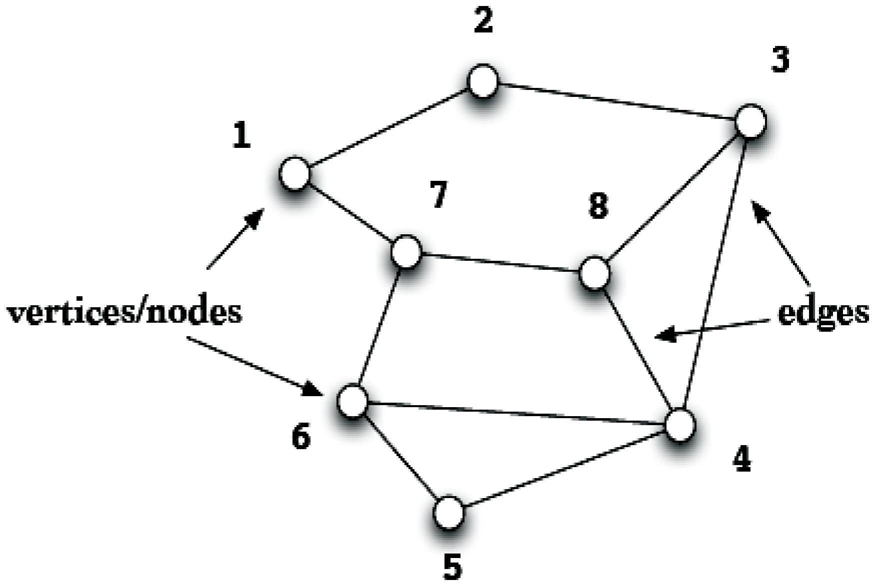
- Comparing two Graphs

  How similar are two graphs?

- Comparing two nodes in a Graph

  How similar are two nodes in a graph?

FINDING NODE SIMILARITY



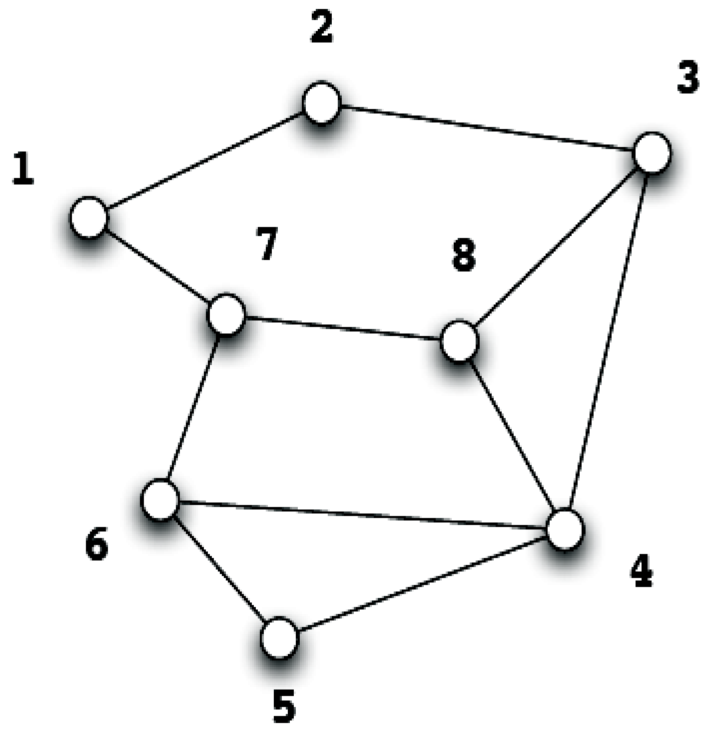
Undirected Graph G(V, E)

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \end{bmatrix}$$

sub-matrix of A = a subgraph of G

- A is Adjacency Matrix representation of an undirected graph.

Finding Node Similarity



$$A^2 = \begin{bmatrix} 2 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 2 & 0 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 3 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 4 & 1 & 1 & 2 & 1 \\ 0 & 0 & 1 & 1 & 2 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 & 3 & 0 & 2 \\ 0 & 1 & 1 & 2 & 1 & 0 & 3 & 0 \\ 1 & 1 & 1 & 1 & 1 & 2 & 0 & 3 \end{bmatrix}$$

A^2 represents no. of random walks of length 2.

$[\text{stochastic}(A)]^2$ will give the probability of random walks of length 2.

Finding Node Similarity

- Count number of walks between two nodes.
- Two nodes are similar if they are connected by many walks.
- Length of random walks will be infinite in a cycled graph. Hence limit it according to your application.
- Discount contribution of longer walks.

Finding Node Similarity

- Discounting Factor :

  Discount a k length walk by $c^k/k!$ for $0 < k <= 1$

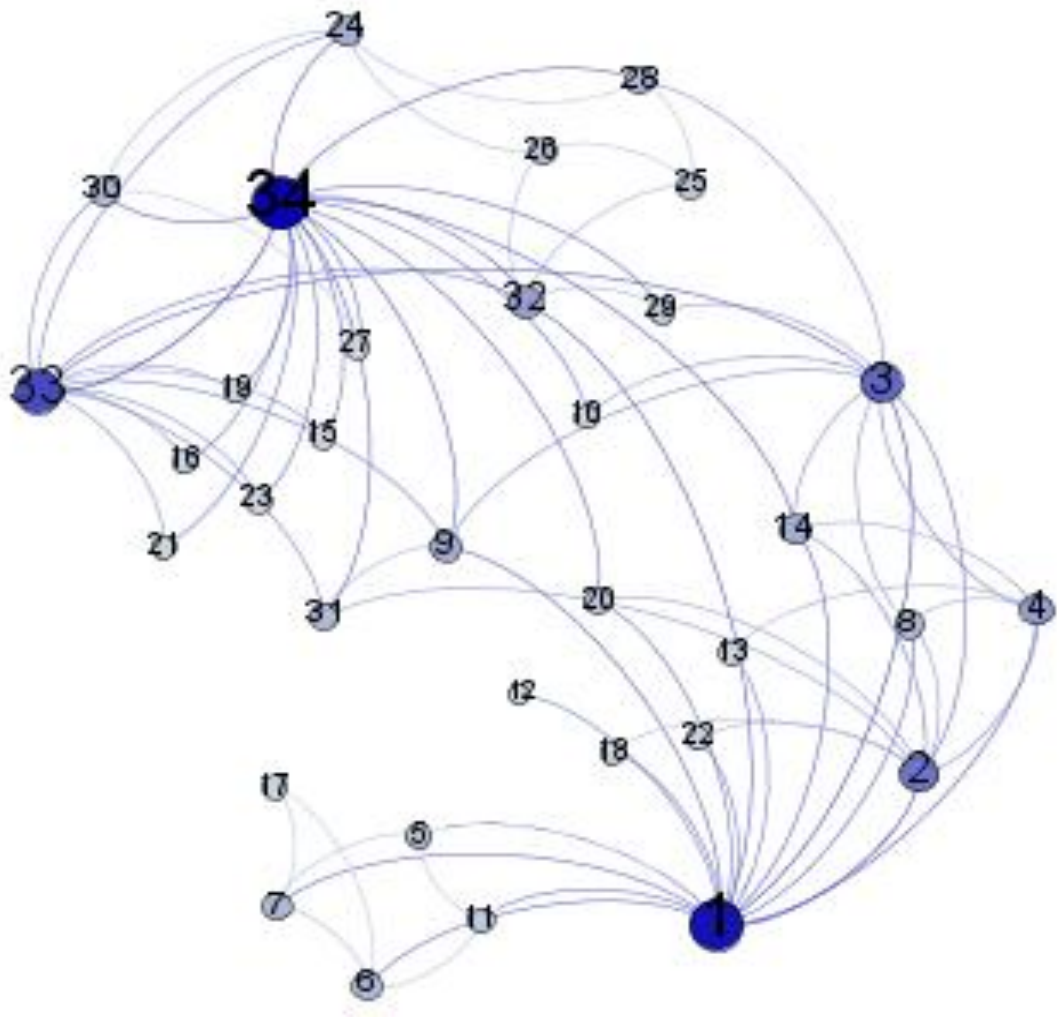
- Similarity Kernel :

$$k(i,j) = \left[\sum_k \frac{\lambda^k}{k!} A^k\right]_{ij} = [\exp(\lambda A)]_{ij}$$

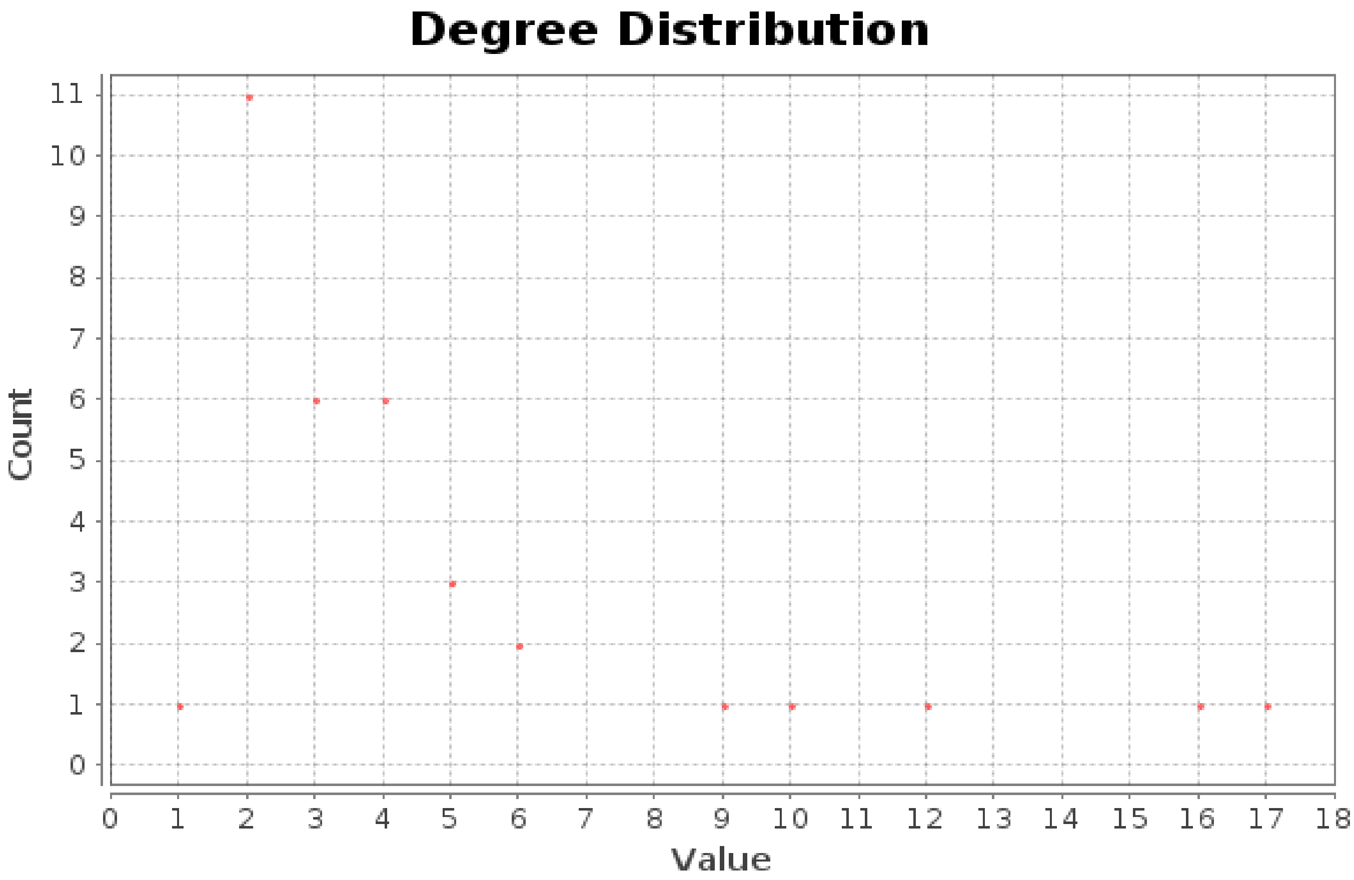
Data Description (Graph Z1)

Data is collected from the members of a university karate club by Wayne Zachary. It represents the presence or absence of ties among the members of the club. Data is collected from 34 members.

GRAPHICAL REPRESENTATION (GRAPH Z1)



DEGREE DISTRIBUTION OF NODES



Results

- node 33 is 82% similar to node 34
- node 1 and node 2 are 67% similar
- node 1 and node 3 are 70% similar
- node 17 is less than 5% similar to most of the nodes

Similarity of Two Graphs

Similarity Kernel :

$$k(G1, G2) = \sum_{i=1}^{n} u(\hat{i}).v(\hat{i})$$

where $u(\hat{i})$ and $v(\hat{i})$ are normalized vectors of length = max(#nodes(G1), #nodes(G2)) representing the degree distribution of G1 and G2 respectively.

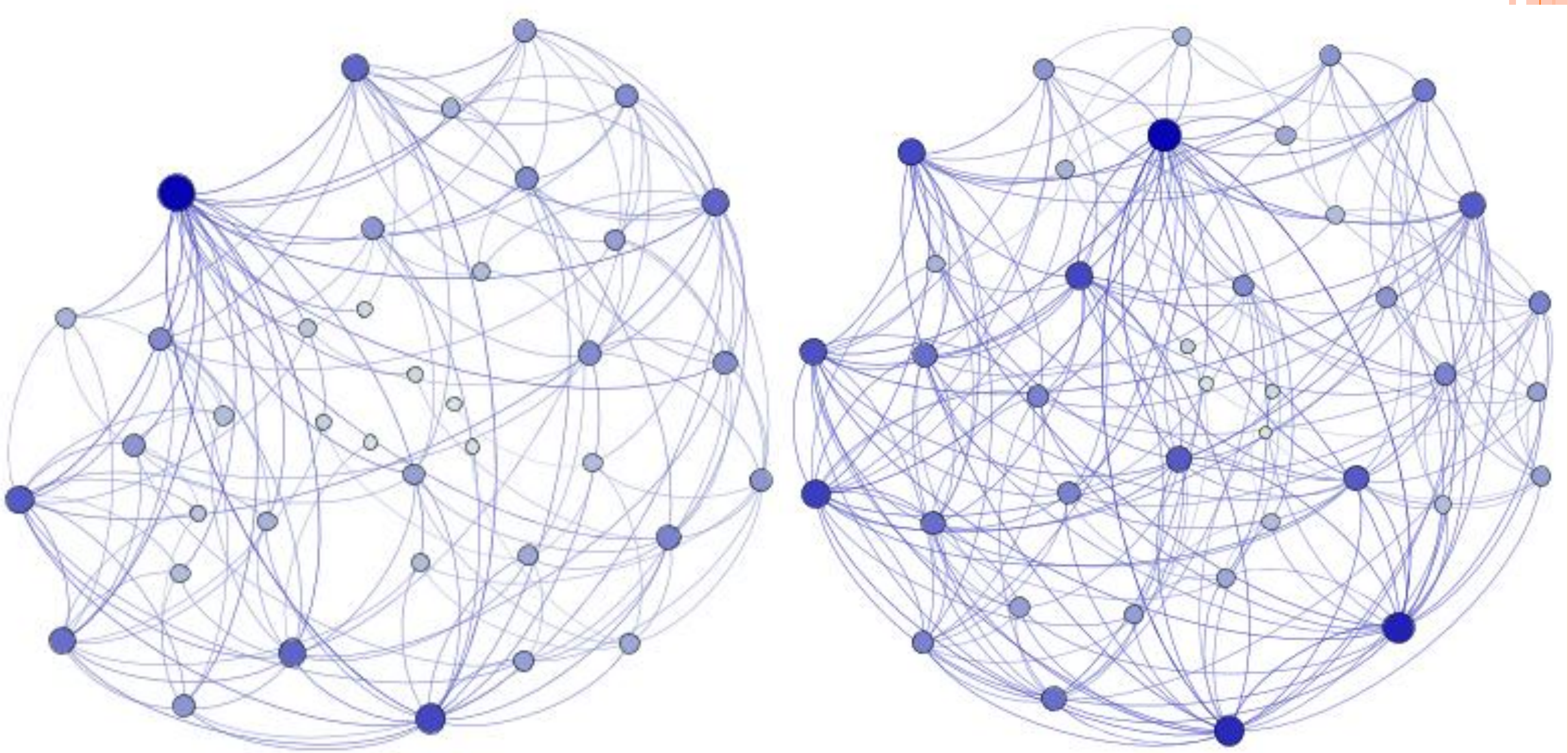
n = Maximum order degree being considered

Experimentation & Data Description (Graph TS1 and TS2)

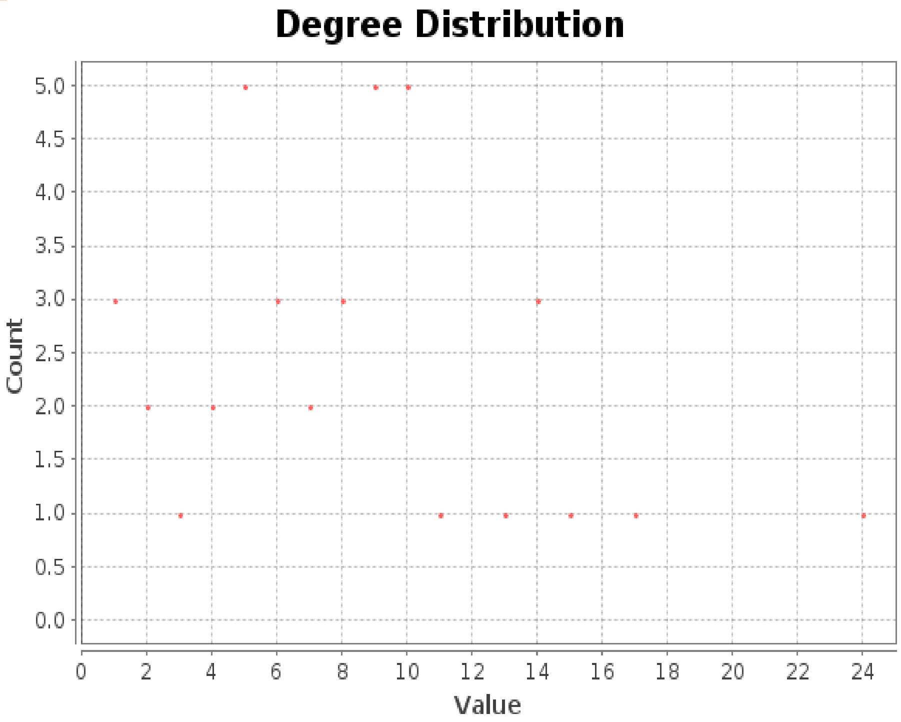
Bruce Kapferer (1972) observed interactions in a tailor shop in Zambia (then Northern Rhodesia) over a period of ten months. His focus was the changing patterns of alliance among workers during extended negotiations for higher wages.The matrices represent two different types of interaction, recorded at two different times (seven months apart) over a period of one month. TS1 and TS2 the "sociational" (friendship, socioemotional) interactions.

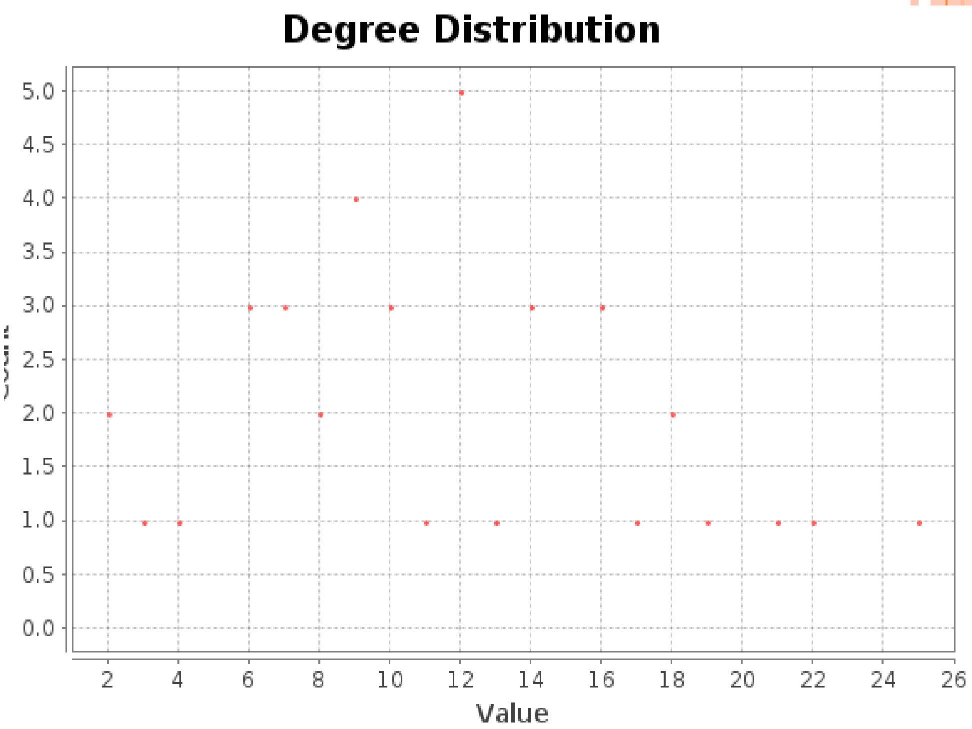
The data are particularly interesting since an abortive strike occurred after the first set of observations, and a successful strike took place after the second.

Graph TS1 and TS2

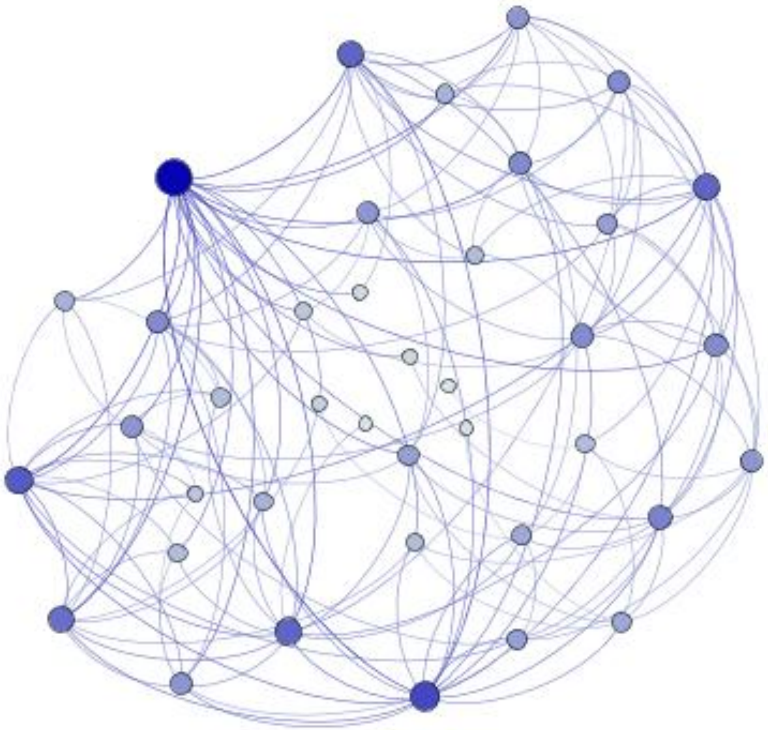


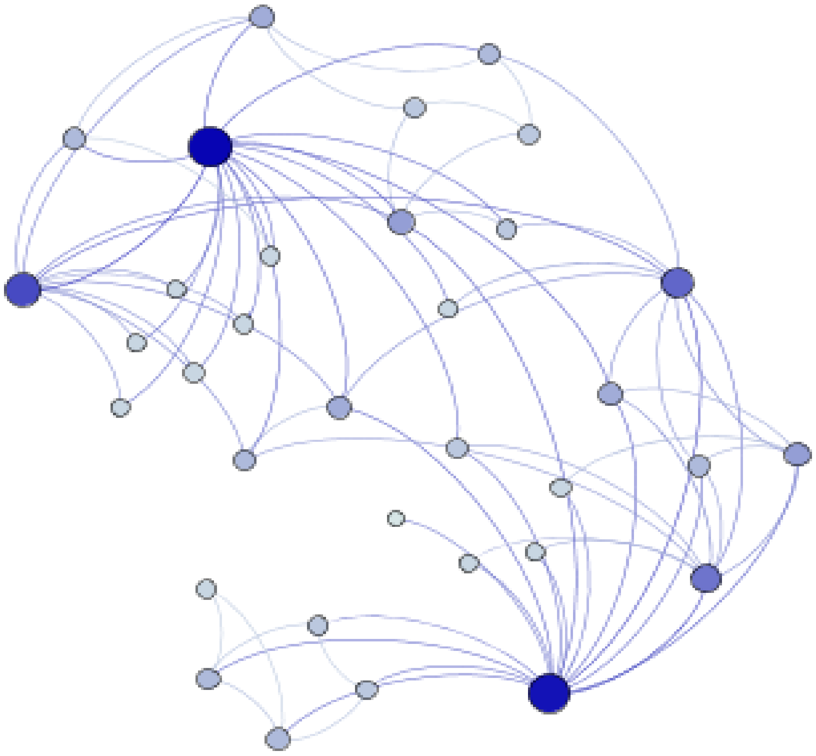
DEGREE DISTRIBUTION OF TS1 & TS2



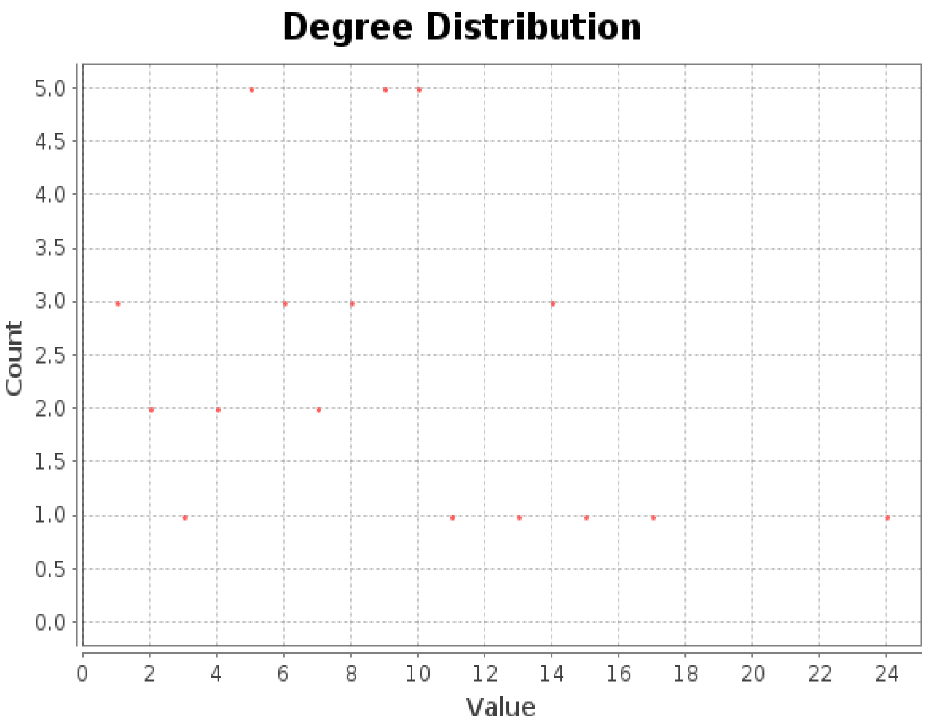


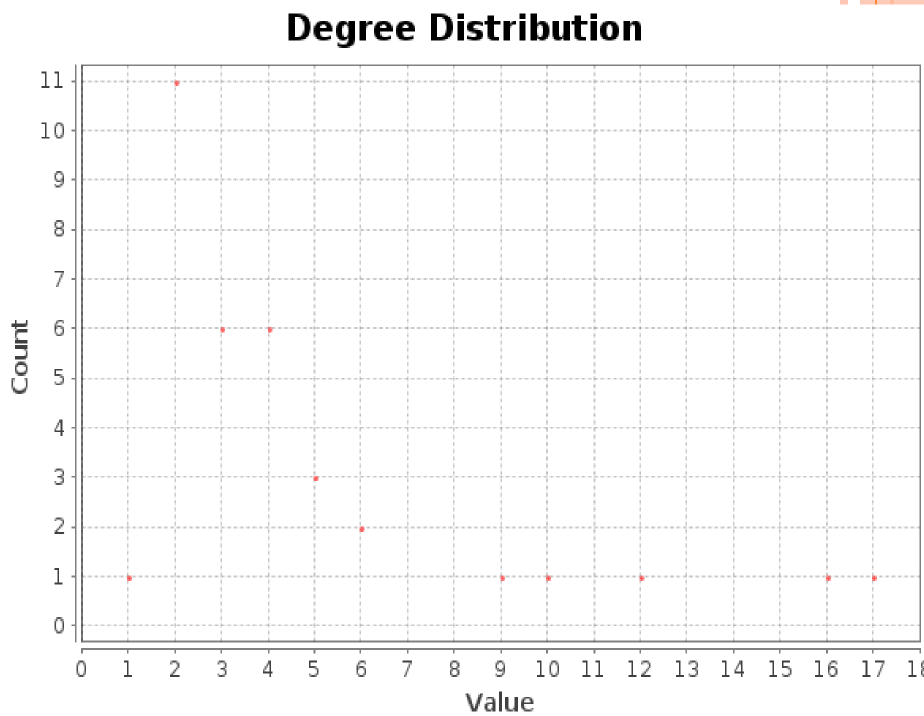
Graph TS1 and Z1



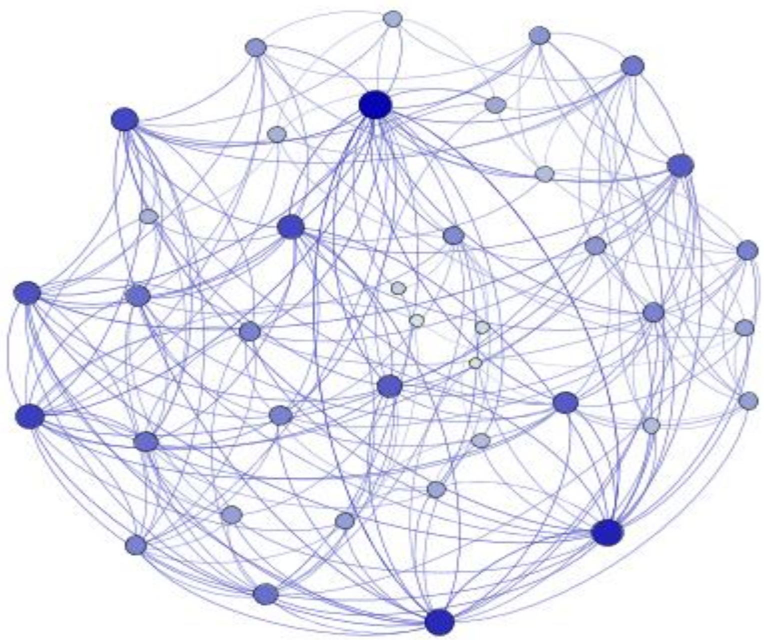


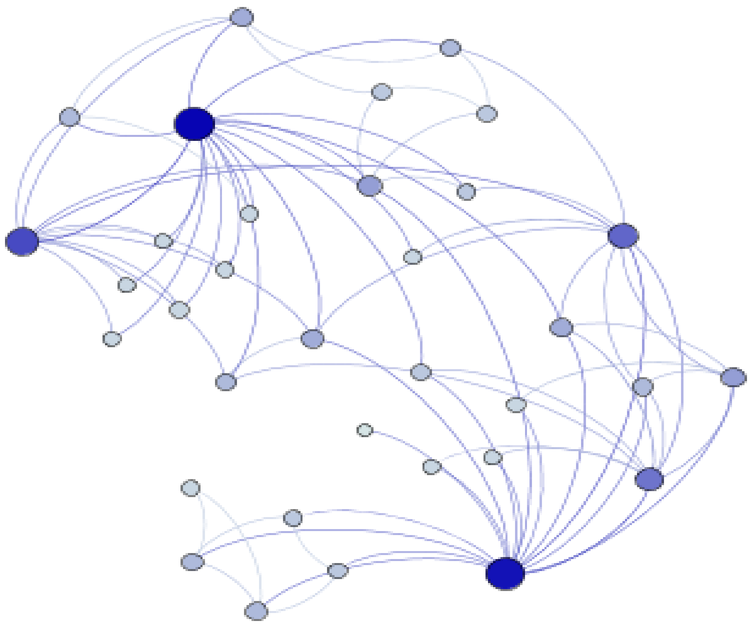
DEGREE DISTRIBUTION OF NODES OF TS1 AND Z1



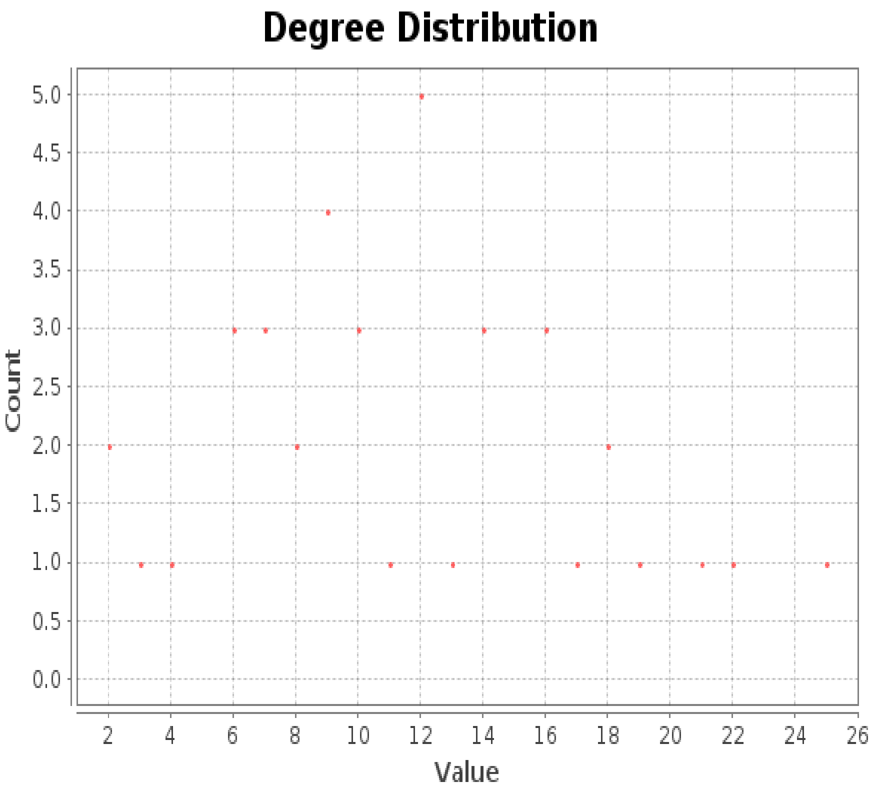


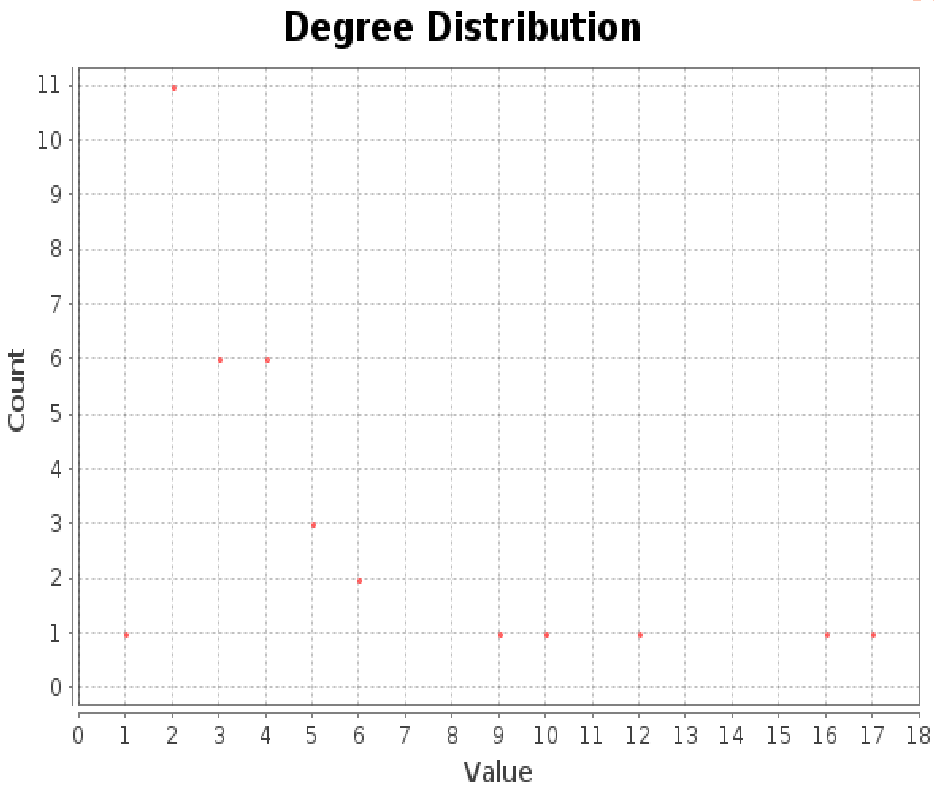
Graphs TS1 and Z1





DEGREE DISTRIBUTION OF NODES OF GRAPH TS2 AND Z1





Results

- Similarity(TS1,TS2) comes out to be 65%
- Similarity(TS1, Z1) comes out to be 24%
- Similarity(TS2, Z1) comes out to be 17%

References and Credits

Research paper referred :

Graph Kernels by S.V.N. Vishwanathan, Karsten Borgwardt, Nic Schraudolph, and Risi Kondor, Journal of Machine Learning Research, 11:1201–1242, April 2010.

Link for the datasets used :

http://vlado.fmf.uni-lj.si/pub/networks/data/ucinet/ucidata.htm

Kapferer B. (1972). Strategy and transaction in an African factory. Manchester: Manchester University Press.

Zachary W. (1977). An information flow model for conflict and fission in small groups. Journal of Anthropological Research, 33, 452-473.

GEPHI, a network visualization software for all the plots and graphs in this presentation.

THANK YOU ☺