Evaluating Structural Connectivity in Multiple Networks

Michele Coscia

Harvard University

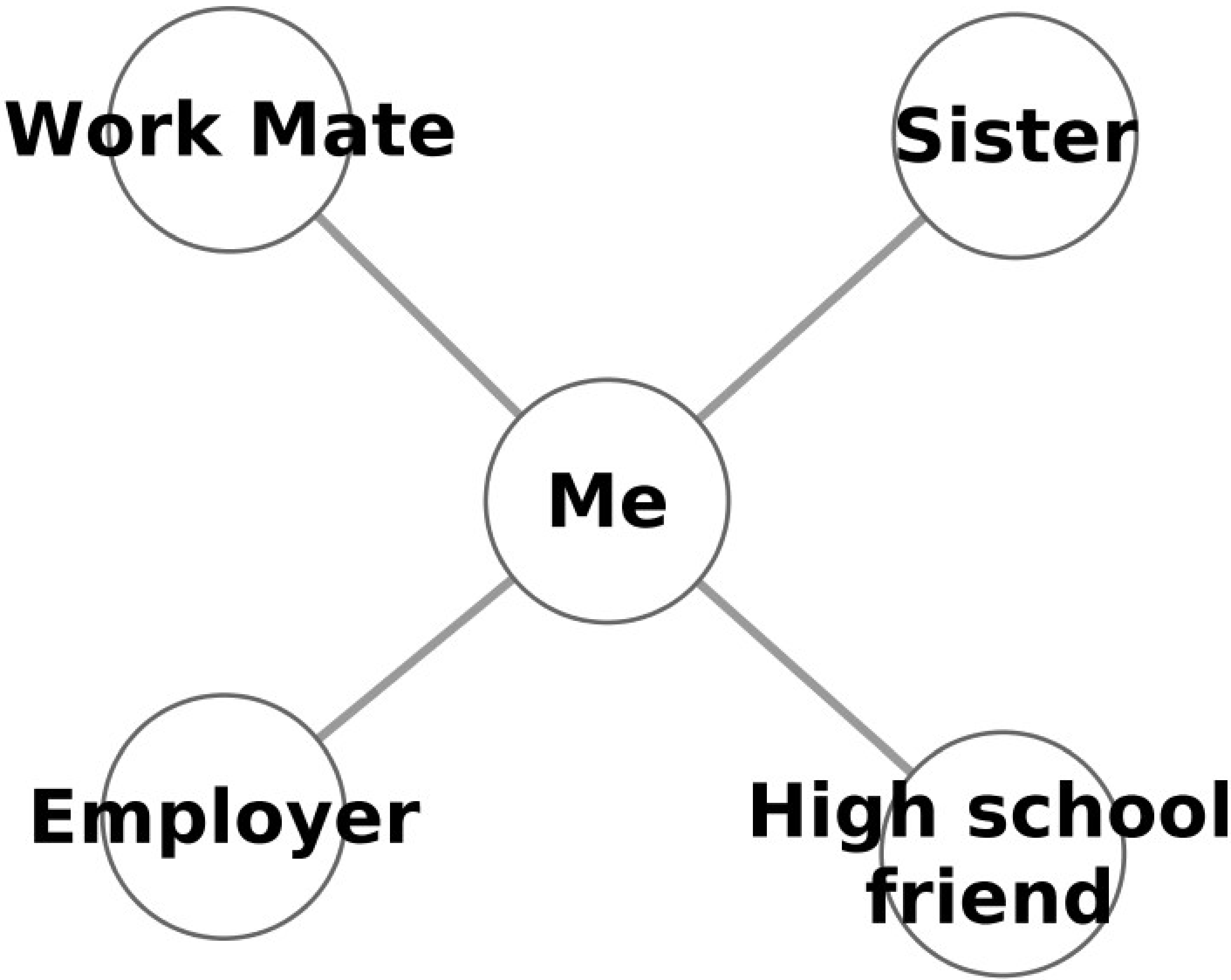
NetSci - June 3rd, 2013

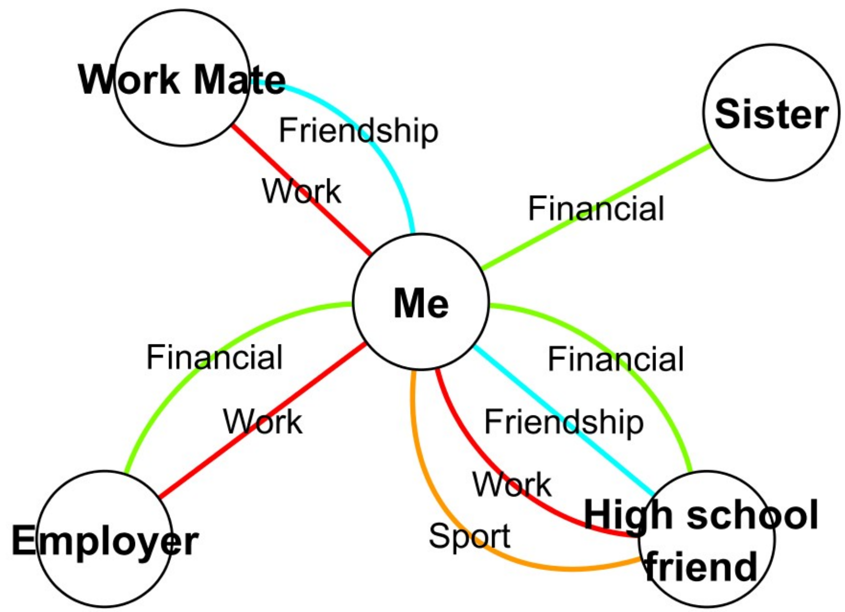
Structural Connectivity

Degree-related measures are very useful in complex networks

Does it change significantly in multiple networks?

Do we add a new degree of freedom for new degree-related measures?


Work Mate
Sister
Me
Employer
High school friend



My Degree: 9

My Neighbors: 4

…

Neighbors in at least 3 networks: 1

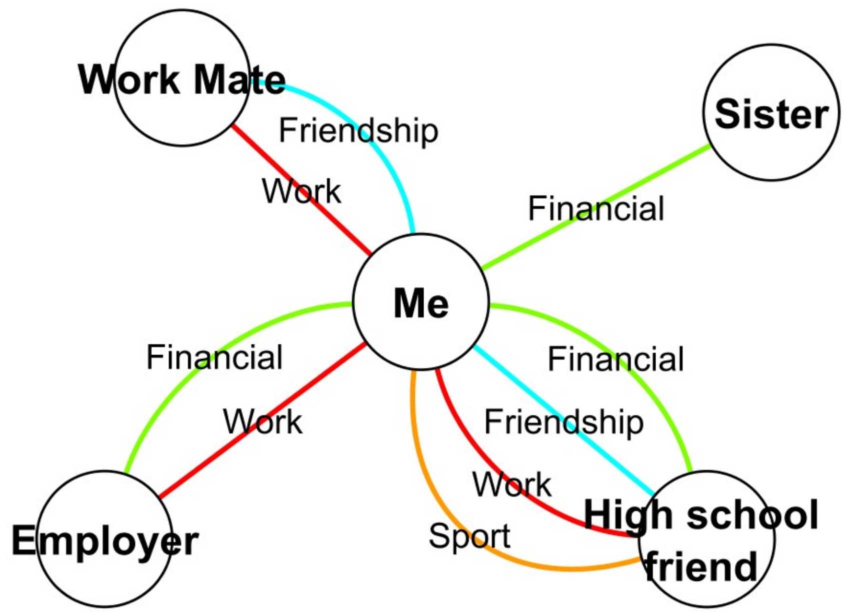
…

Neighbors in the “Friendship” network: 2

…

Question #1

How many neighbors will I lose if we remove one network?



If “Financial” goes down, we lose one neighbor

If “Work” goes down, we lose three edges but no neighbors

“Financial” looks very important, even if it had only that one edge

Formally:

$$Neighbors_{XOR}(v, D) =$$

$$|\{u \in V|\ \underline{\exists d \in D : (u, v, d) \in E} \wedge \underline{\nexists d' \notin D : (u, v, d') \in E}\}|$$

Counting the edges appearing in that network

Discarding the ones also appearing in other networks

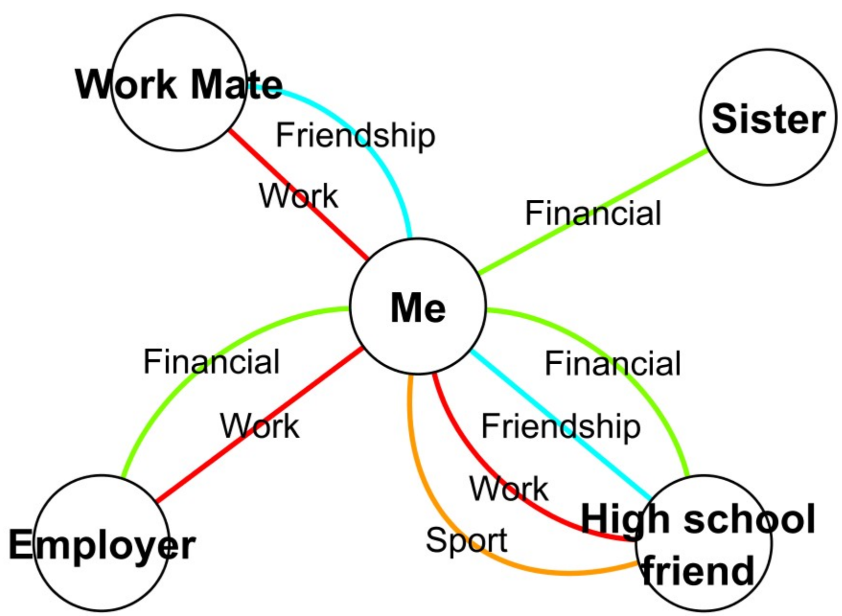
Question #2

What is the most relevant network for a node (given different relevance criteria)?

Criterion #1

The ratio of neighbors connected in that network

We call it Dimension Relevance:



$$DimRelevance(v, D) = \frac{Neighbors(v, D)}{Neighbors(v, L)}$$

DimRelevance(Me, Financial) = 3 / 4

DimRelevance(Me, Work) = 3 / 4

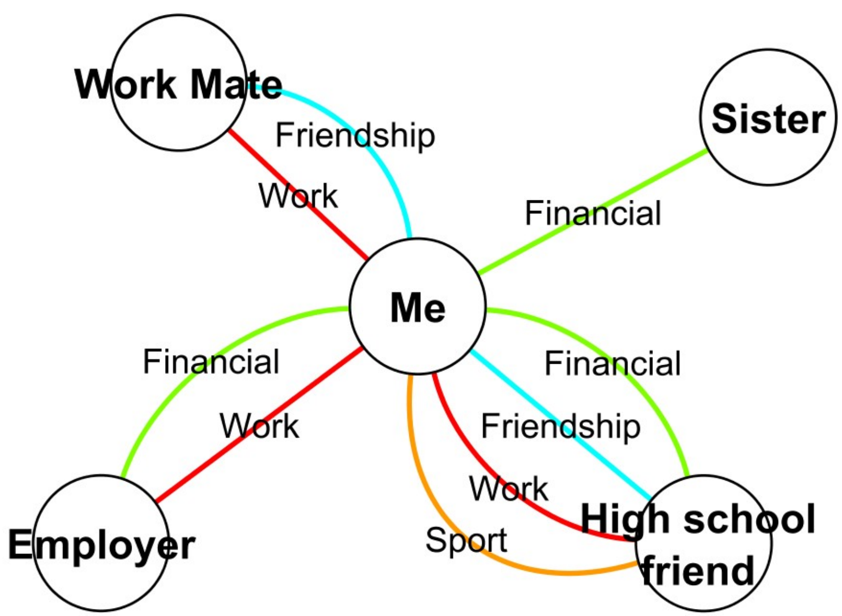
DimRelevance(Me, Friendship) = 2 / 4

DimRelevance(Me, Sport) = 1 / 4

Criterion #2

Weight the # of neighbors on the number of other networks in which I can connect to them

$$DimRelevance_W(v, D) = \frac{\sum_{u \in NeighborSet(v,D)} \frac{n_{uvd}}{n_{uv}}}{Neighbors(v, L)}$$



DimRelevance(Me, Financial) = 1.75 / 4

DimRelevanceW(Me, Work) = 1.25 / 4

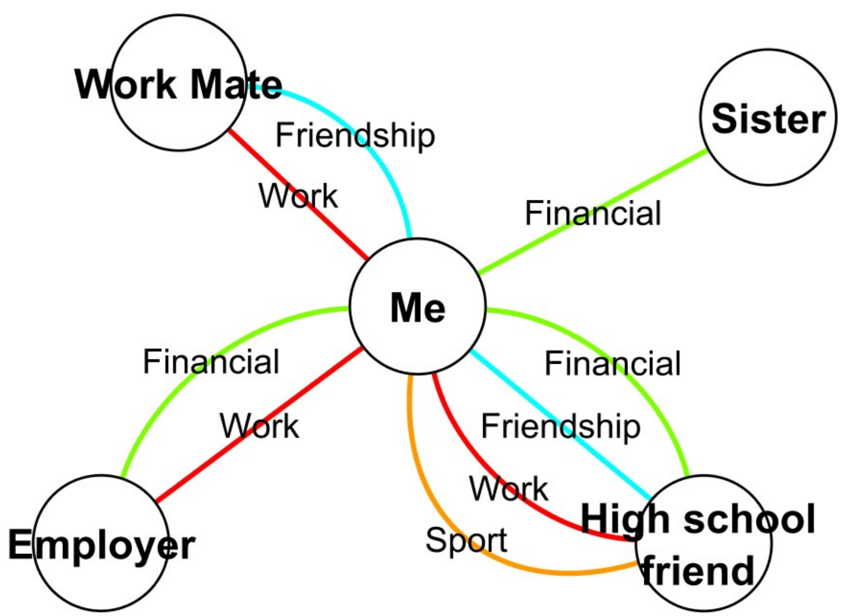
DimRelevanceW(Me, Friendship) = 0.75 / 4

DimRelevanceW(Me, Sport) = 0.25 / 4

Criterion #3

The ratio of neighbors lost if the network disappears

We call it Dimension Relevance XOR:



$$DimRelevance_{XOR}(v, D) = \frac{Neighbors_{XOR}(v, D)}{Neighbors(v, L)}$$

DimRelevance(Me, Financial) = 1 / 4

DimRelevanceW(Me, Work) = 0 / 4

DimRelevanceW(Me, Friendship) = 0 / 4

DimRelevanceW(Me, Sport) = 0 / 4

A Nice Property

DimRelevance

>=

DimRelevanceW

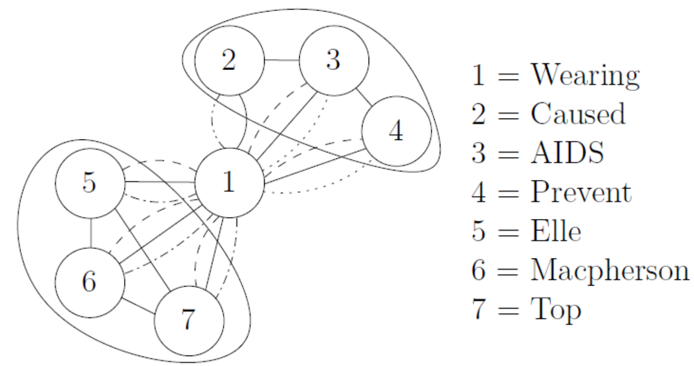
>=

DimRelevanceXOR

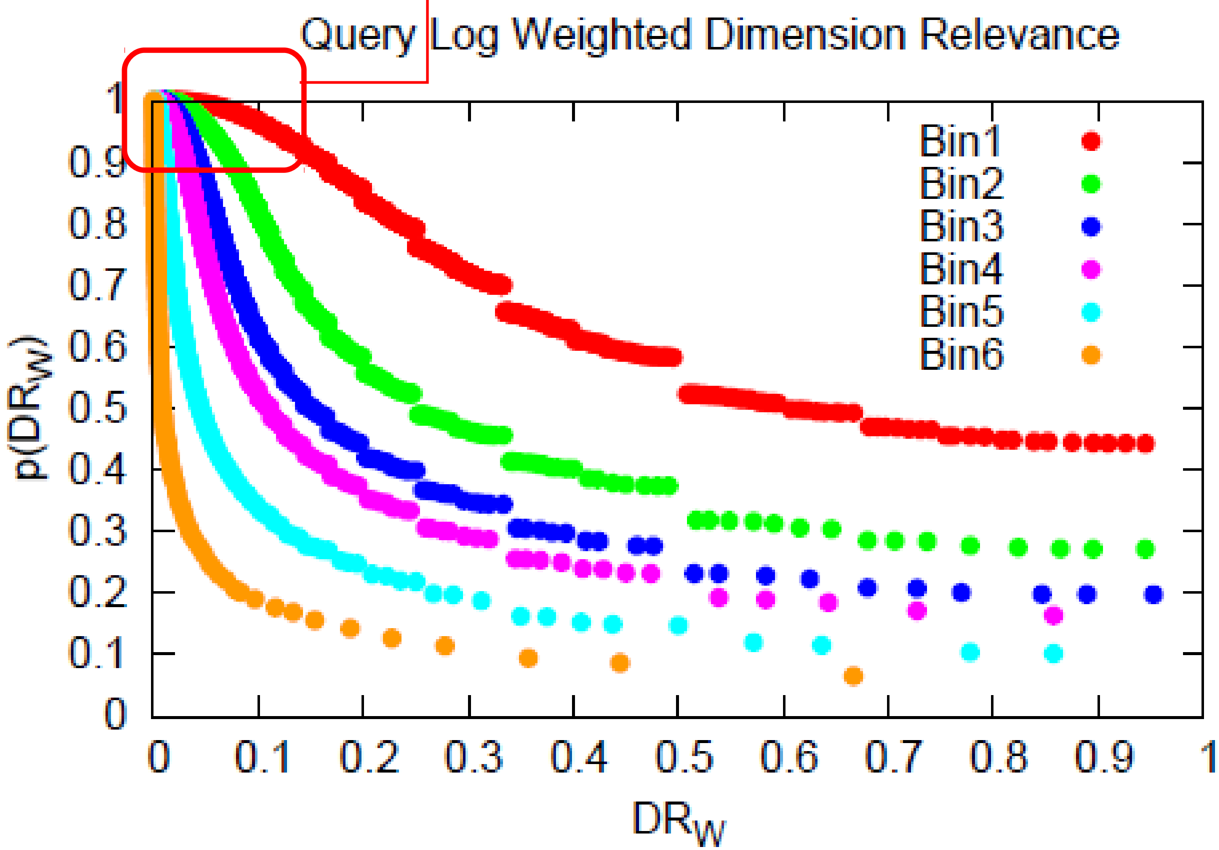
Why do we care?

- Query contextualization
  - Term-Term network from AOL query logs
  - Networks: Rank of the URL clicked by the user
- A-Social User Behavior
  - User-User network from Flickr
  - Networks: Friendship, Comment, Favorite, Tag
- Science Jumpers
  - Co-authorship network from DBLP
  - Networks: Year of the collaboration

AOL



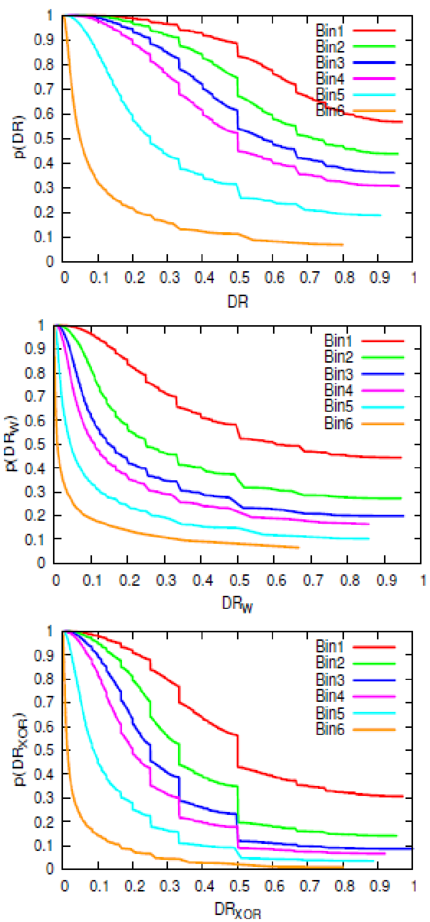
Generally bad results



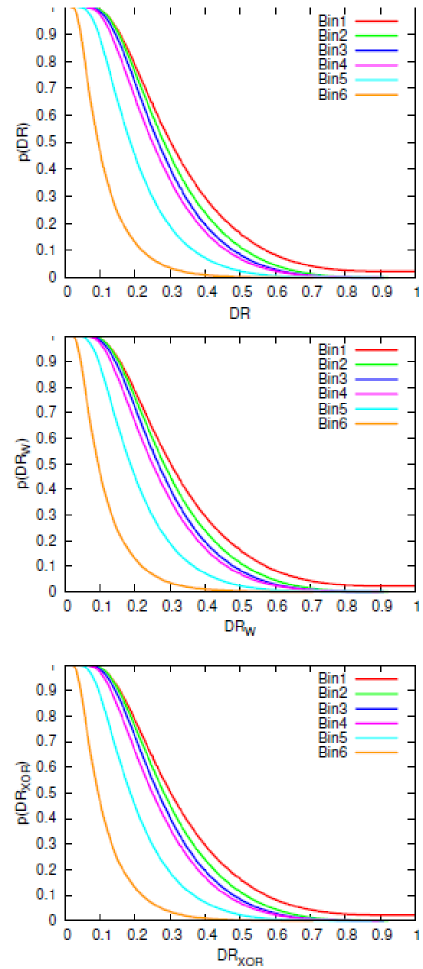
In Flickr:
Users appearing only in the “Friendship” networks

In DBLP:
Authors changing their neighborhood from one year to the other

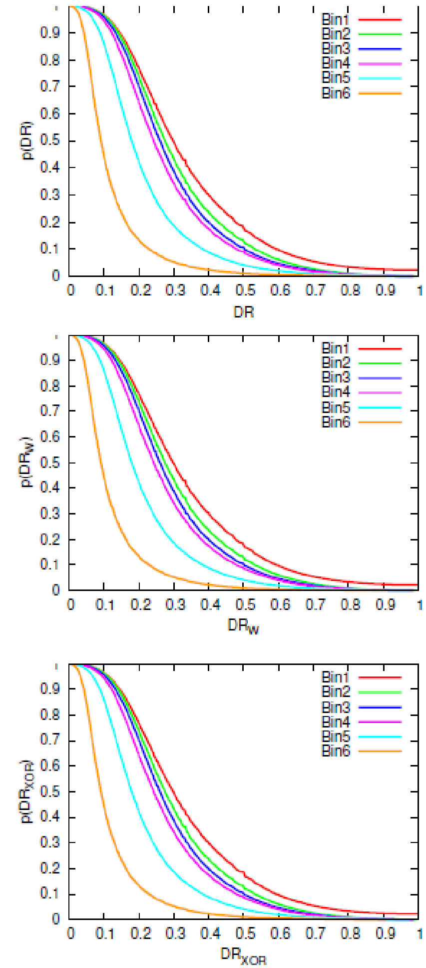
Capturing network interplay



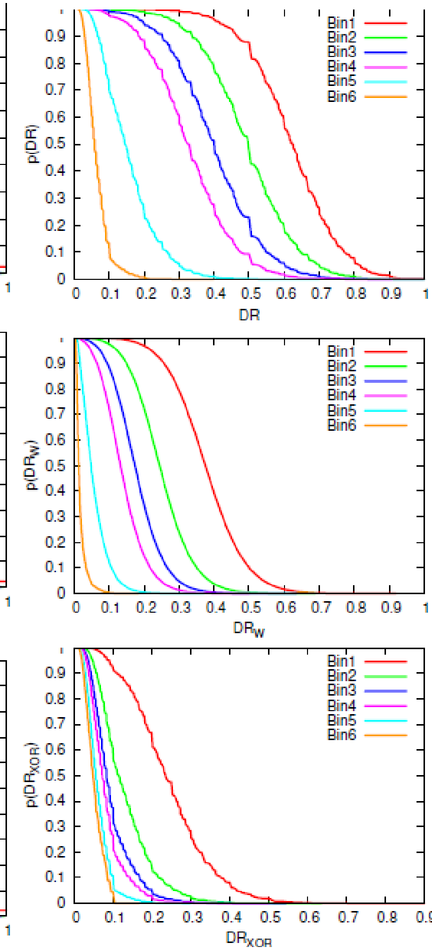
DR Distributions in the original network



MultidimRandom (Only the # of Nodes/Edges are preserved)



Multidim Preferential Attachment (rich-get-richer with same # of Nodes/Edges)

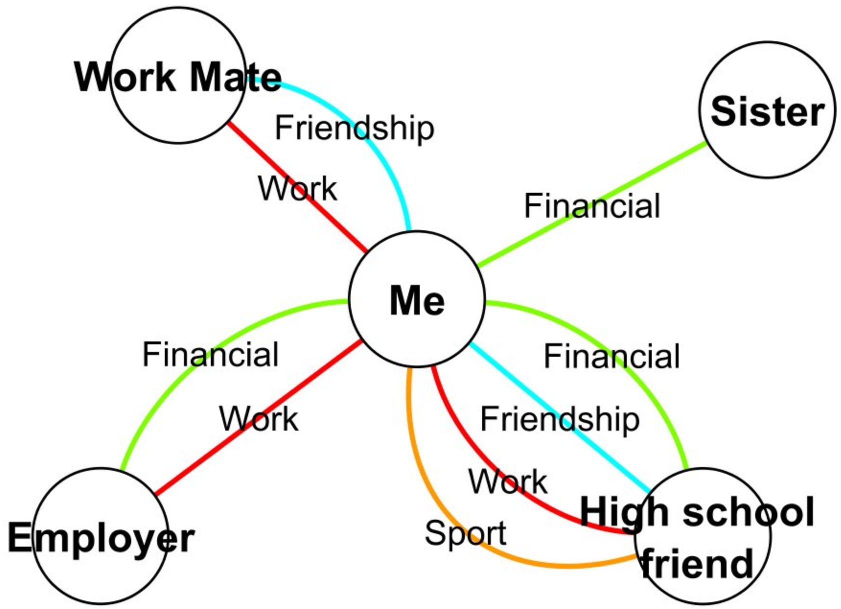


Multidim Jaccard (correlations between dimensions preserved)

Question #3

Are there networks that shows the very same patterns of connections?

Network Correlation

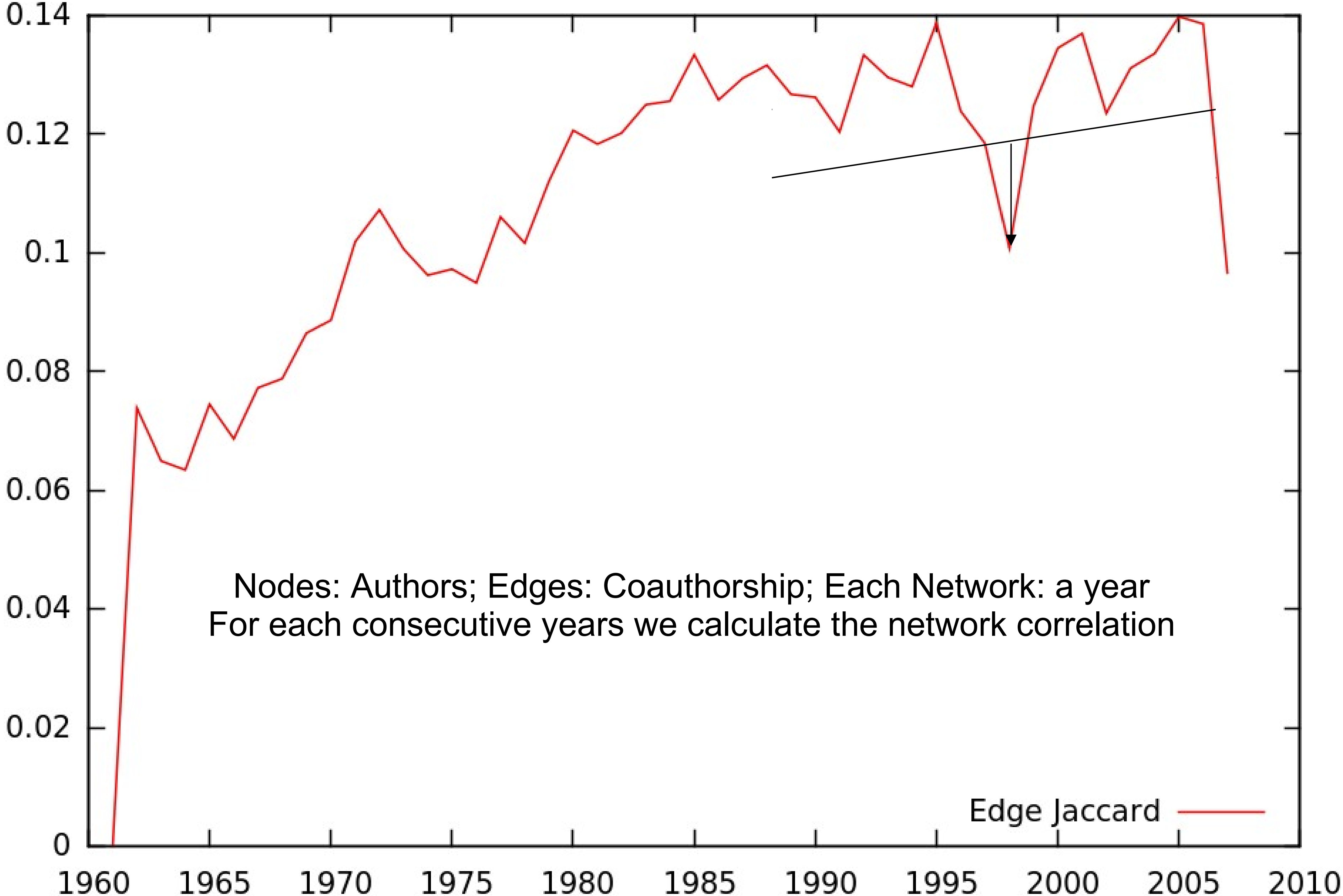


$$\rho_{node}(d_1, d_2) = \frac{|V_{d_1} \cap V_{d_2}|}{|V_{d_1} \cup V_{d_2}|}$$

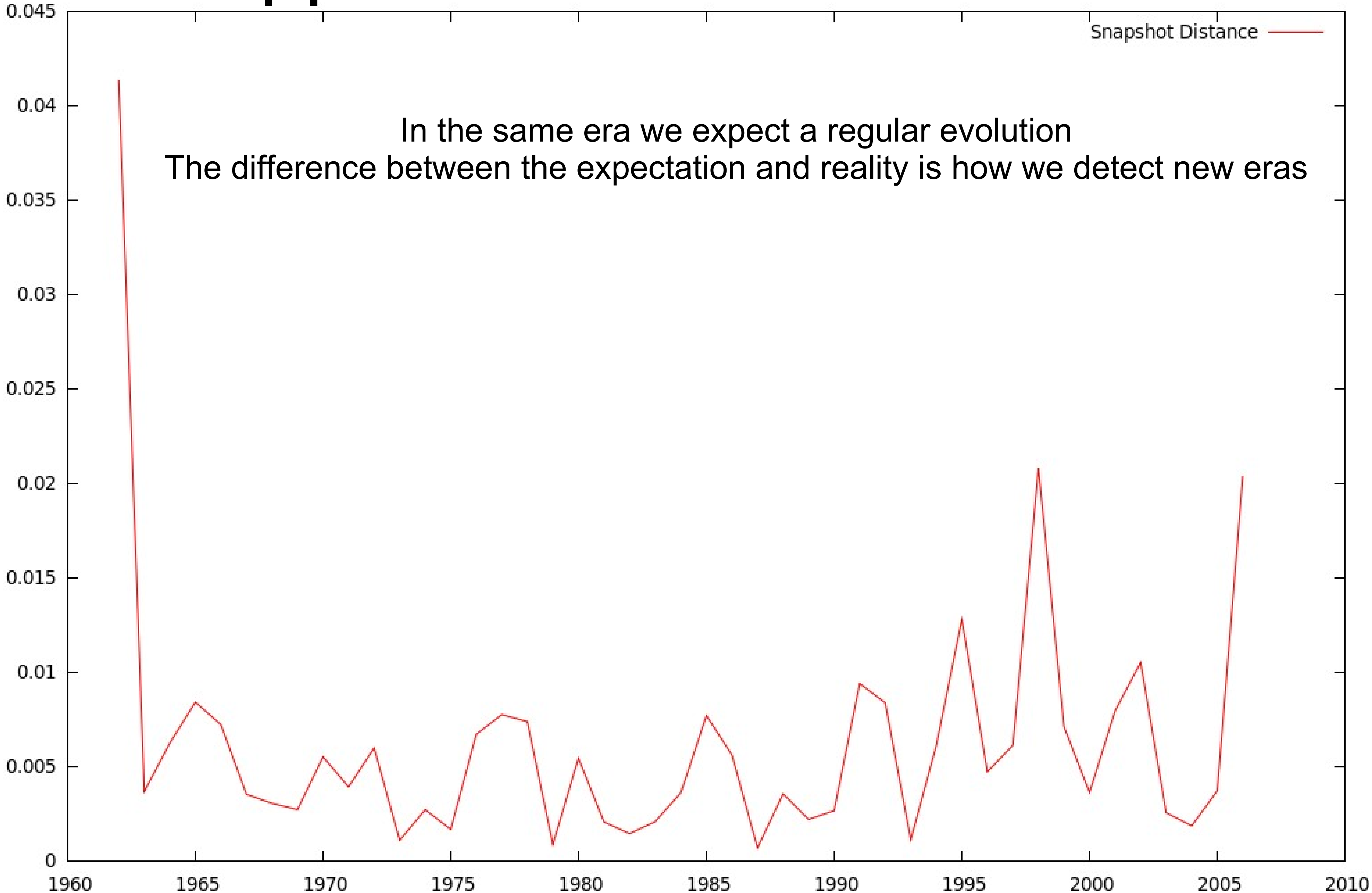
$$\rho_{edge}(d_1, d_2) = \frac{|E_{d_1} \cap E_{d_2}|}{|E_{d_1} \cup E_{d_2}|}$$

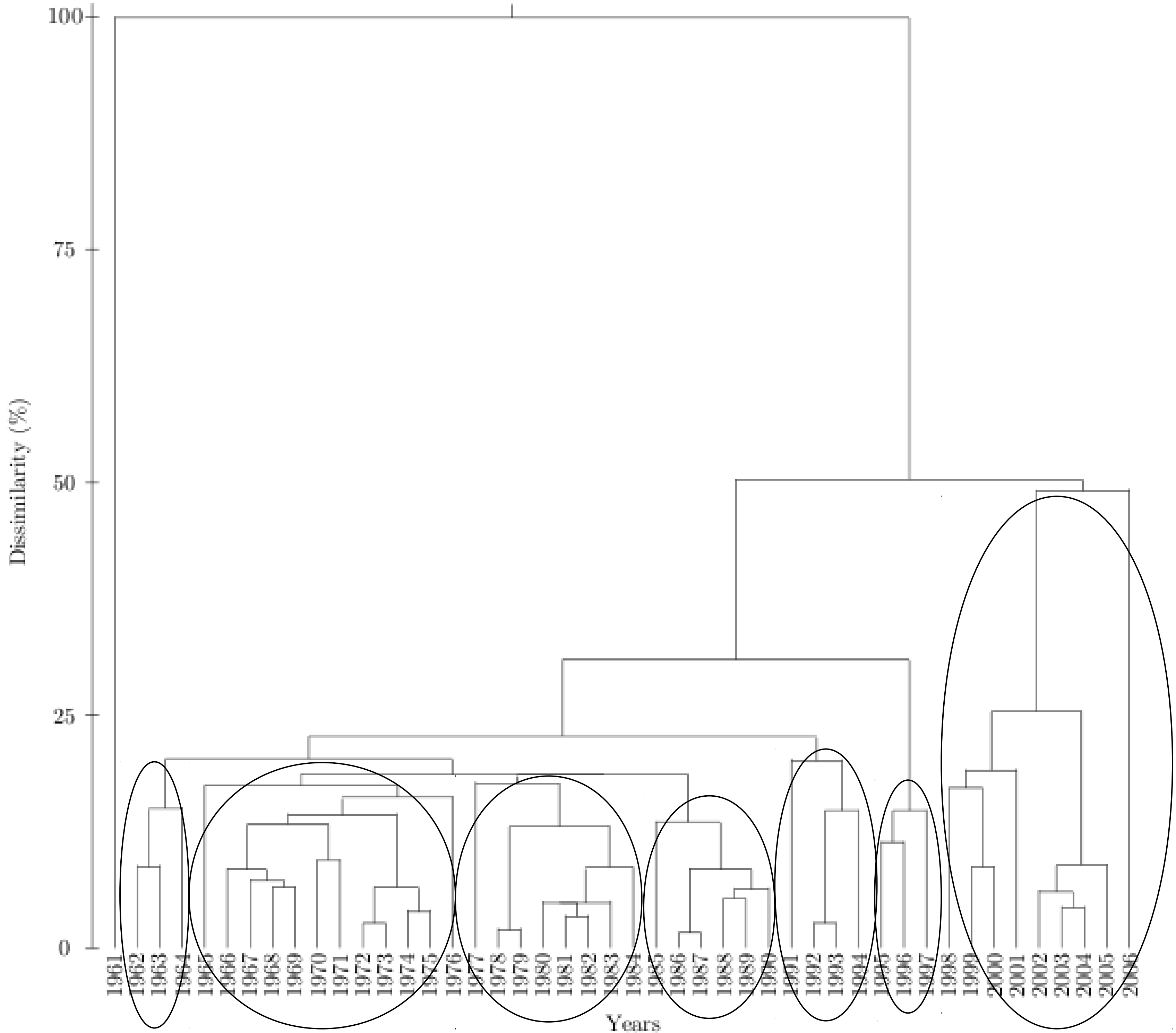
The Jaccard coefficient over the nodes with at least one edge in the network; or over the edges themselves

Application: Network Eras



Application: Network Eras





Dissimilarity (%)
100
75
50
25
0
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
Years

Conclusion

Network connectivity is much different in multiple networks

We have an additional degree of freedom

Advanced analysis can be performed even with measures as simple as the degree

Bibliography

Michele Berlingerio, Michele Coscia, Fosca Giannotti, Anna Monreale, Dino Pedreschi: Foundations of Multidimensional Network Analysis. ASONAM 2011: 485-489

Michele Berlingerio, Michele Coscia, Fosca Giannotti, Anna Monreale, Dino Pedreschi: The pursuit of hubbiness: Analysis of hubs in large multidimensional networks. J. Comput. Science 2(3): 223-237 (2011)

Michele Berlingerio, Michele Coscia, Fosca Giannotti, Anna Monreale, Dino Pedreschi: Evolving networks: Eras and turning points. Intell. Data Anal. 17(1): 27-48 (2013)

M. Schich and M. Coscia: Untangling the Complex Overlap of Subject Themes in Classical Archaeology. MLG @ KDD, (2011)

Thank you

Questions?

NetSci

Center for International Development at Harvard University