Durham University

Social Network Analysis

Computational Social Science

Lecture 6

Overview

Networks consider cases in the context of their relationship with other cases

Not just a method, but an entire discipline

- Social science and mathematical traditions, terminology, questions

How network data is collected and represented

Centrality: Exerting influence within networks

Structures within networks

Dynamic processes on networks: changing attributes of cases

Dynamic networks: changing relationships between cases

Two traditions of network analysis

Social Sciences

Social Network Analysis

- Originally sociometry, about 1920s

Focus on specific networks

- Meaning of pattern of relationship
- How influence is exerted

Terminology of: actors, ties, links, arcs

Mathematical Sciences

Mathematics: Graph Theory (1950s)

Other: Network Science (1980s)

Focus on properties of networks

Terminology of: vertices, nodes, edges

Different questions so different methods

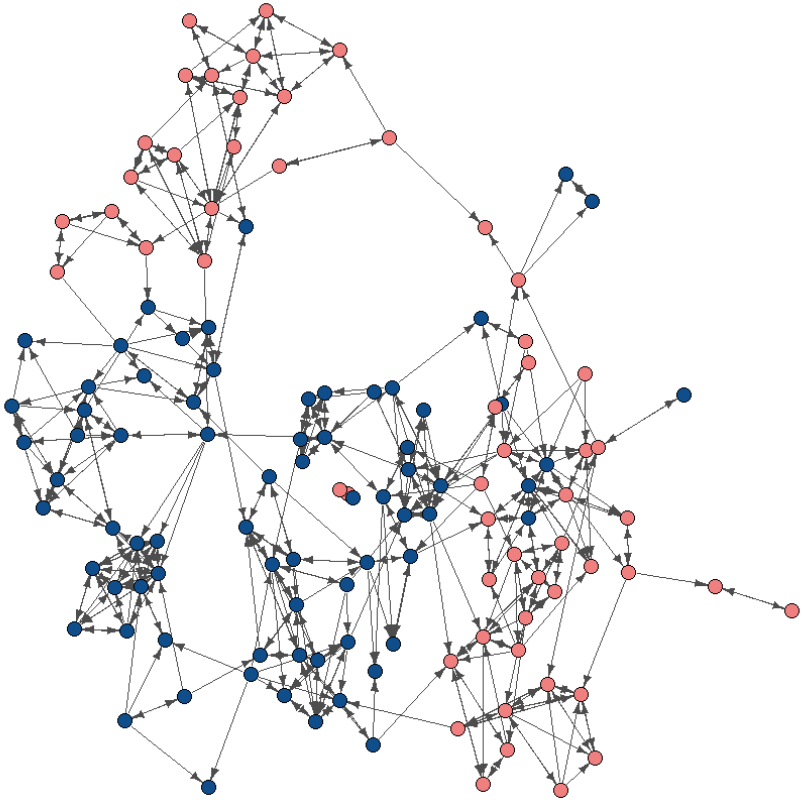
Research in the social sciences is often concerned with the reason behind the network connections rather than the properties of the network structure itself.

Keeling & Eames 2005 – mathematical epidemiologists

Borgatti et al 2009 – social scientists

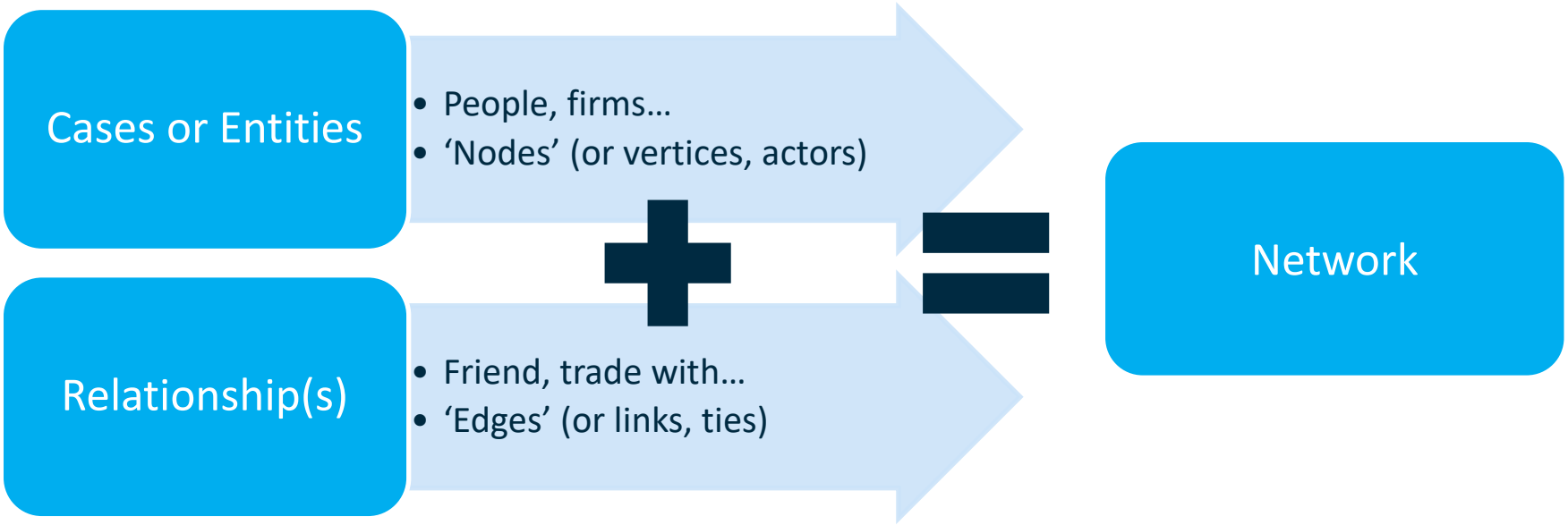
From a social scientist's point of view, network research in the physical sciences can seem alarmingly simplistic and coarse-grained. And, no doubt, from a physical scientist's point of view, network research in the social sciences must appear oddly mired in the minute and the particular, using tiny data sets and treating every context as different.

What is a network?



Durham University

What is a network?



Many types of relationships

Social Relations:

- Perceptual (familiar)
- Friends
- Affectual (like, dislike)
- Family, kinship

Co-occurrence:

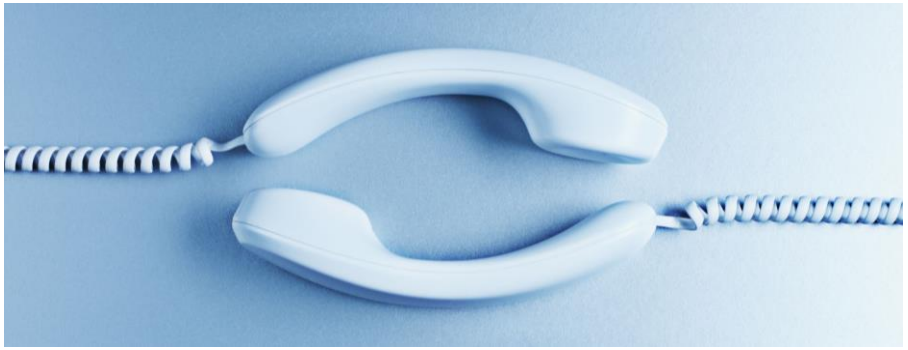
- Common group membership
- Attendance of the same event
- Close geographic proximity
- Collaboration (music, research)

Many types of relationships

Flows:

- Information
- Capital, goods, trade
- Transport routes

https://commons.wikimedia.org/wiki/File:World_airline_routes.png



Interactions:

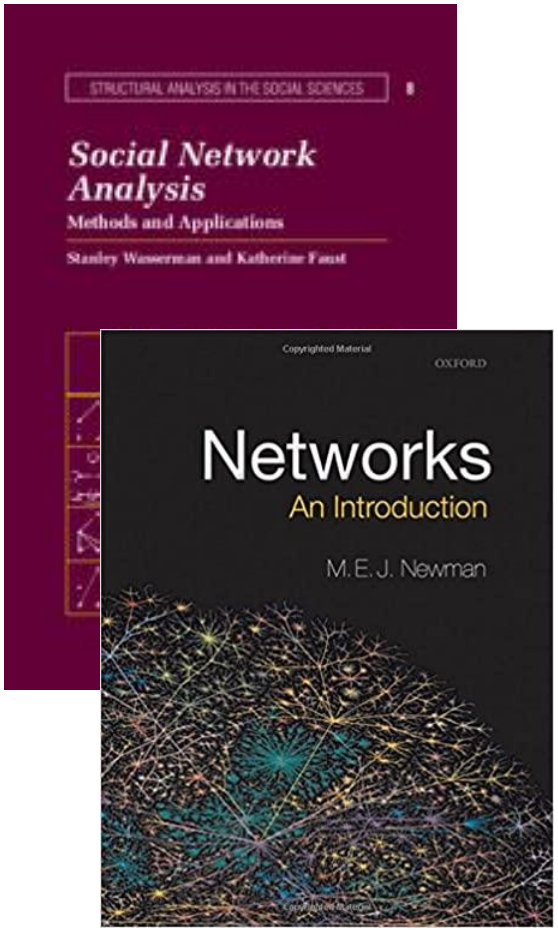
- Influence (advice)
- Activities (visiting)
- Communication (email)

Why social network analysis (or network science)?

Standard social science analysis methods do not effectively model relational structure and dynamics

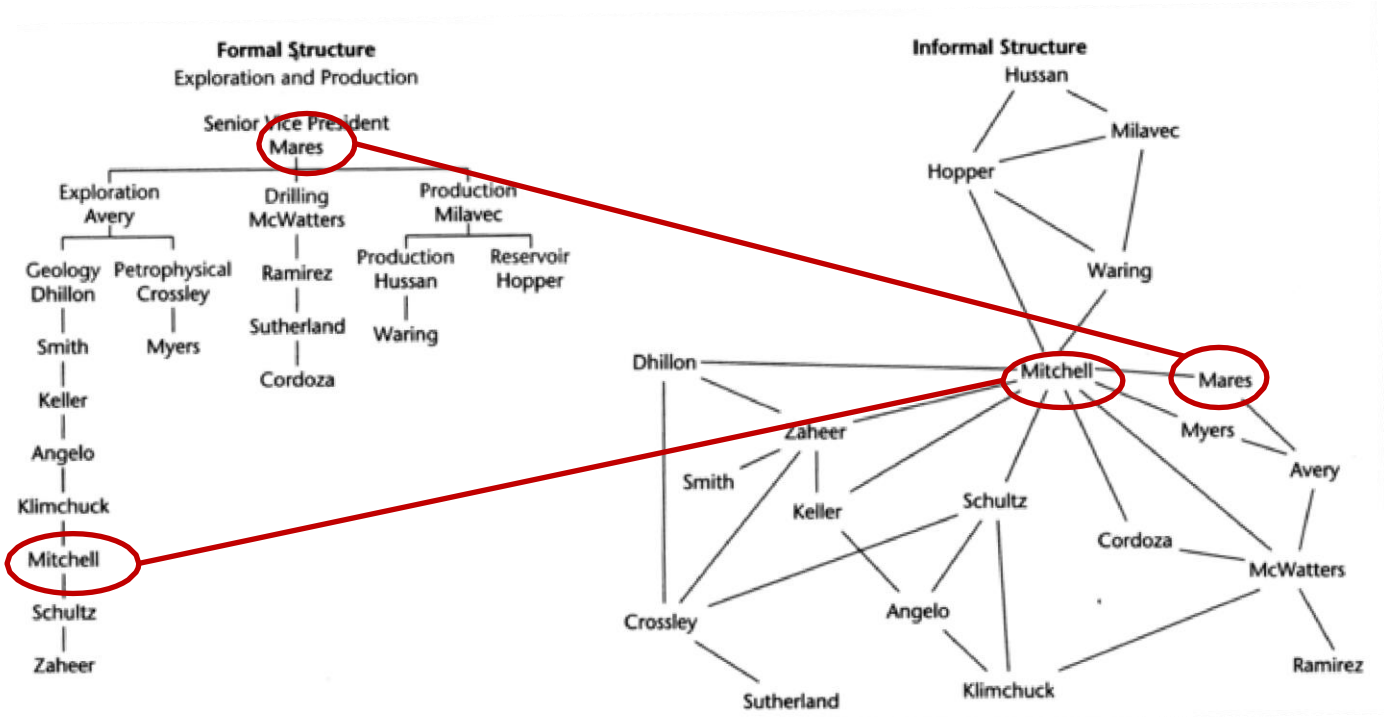
Network analysis provides a set of tools to extend our intuition of the patterns that construct social structure

- Visualisation to highlight patterns and structure obscured by the complexity of relationships
- Develop theories about the way in which relationships influence social phenomena and processes
- Empirically test such theories



Durham University

Network analysis provides tools for understanding structure
Mismatch between relationships and formal hierarchy



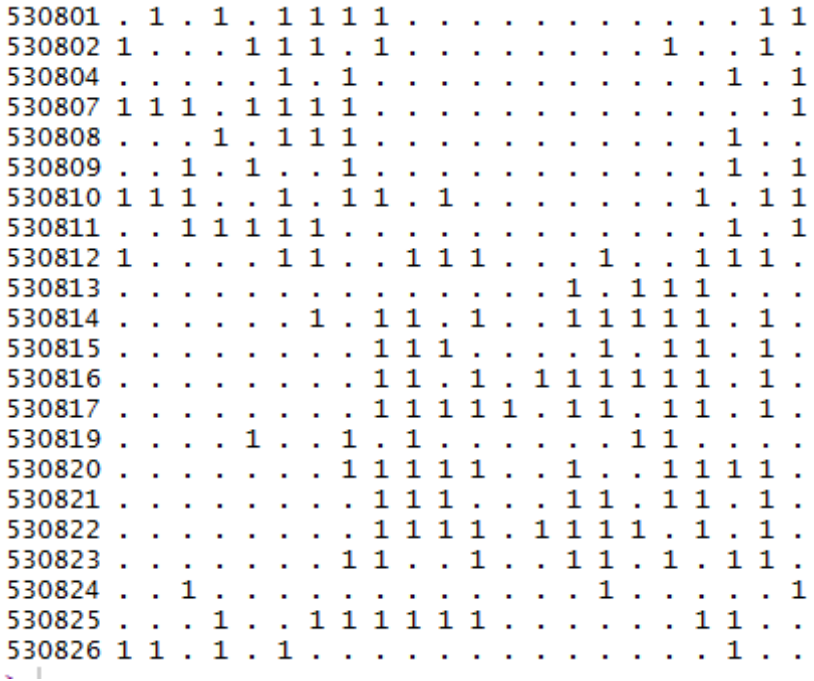
Cross (2010). Improving leadership effectiveness through personal network analysis and development, in The Organizational Network Fieldbook (Fig 19.1)

Networks are a natural level at which to study society

…the survey is a sociological meatgrinder, tearing the individual from his social context… If our aim is to understand people's behavior rather than to simply record it, we want to know about primary groups, neighborhoods, organizations, social circles, and communities; about interaction, communication, role expectations, and social control.

Allen H. Barton (1968)
Professor of Sociology, Columbia University
Director, Bureau of Applied Social Research

Collecting and Representing Network Data



Durham University

Primary data collection: name generator

We are interested in patterns of behaviour among friends. The next questions ask about your ***friends***; friends that you have both in school and outside of school.

Q10 **Please name up to 10 of your closest friends in your school form class?**

Only list those friends who are in your school form class. *You do not need to use all the boxes if you do not want to. Do not worry if you are unsure of the spelling. Just try to spell their name as best you can. Please write their full name (i.e. first name and surname), and put a mark (*) beside your best friend's name.*

1. _____
2. _____
3. _____
4. _____
5. _____
6. _____
7. _____
8. _____
9. _____
10. _____

☐ Or, my friends are not in my school form class (please tick)

Respondent and their nominees processed to convert all names to unique identifiers

Question from Wellbeing in Schools 2016

Durham University

Secondary network data

Social media data

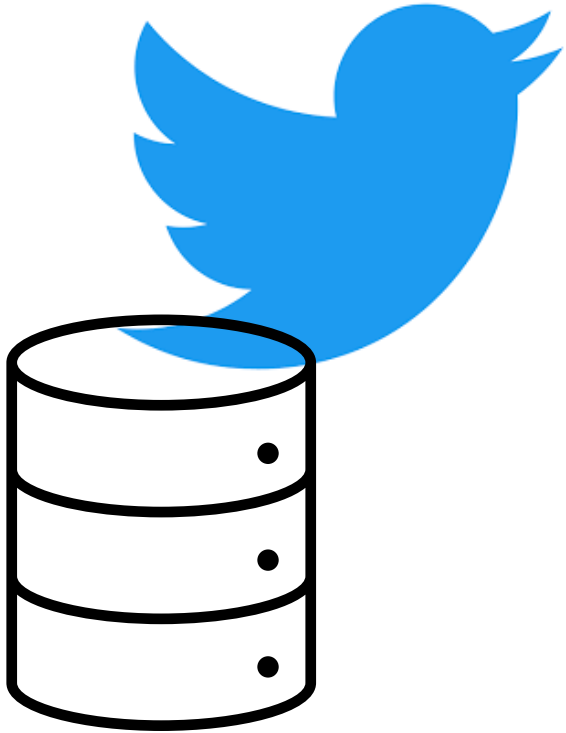
- Posts and replies, following

Administrative data

- Membership
- Trade flows
- Transport

Scraping

- Collaborations



Characteristics of relationships (edges)

Directed vs undirected

- One way (eg nominate friend)
- Bidirectional (eg marriage)

Weighted have a value on the edge

- Ordinal: best friend vs friend vs acquaintance
- Scale: amount of trade (annual) between firms

What if there's more than one relationship over the same set of cases?

- separate networks OR
- multiplex

Representation: in datasets

Nodelist: looks like the survey data

StudyID	Name1	Name2	Name3	Nam
530801	530802	530810	530809	530
530802	530801	530810	530808	530
530804	530811	530824	1	530
530807	530808	1	530809	530
530808	1	530807	530811	530
530809	530811	530804	530824	530
530810	530814	530801	530802	530
530811	530824	530804	1	530
530812	530825	530823	530820	530
530813	530819	530822	530821	530
530814	530820	530810	530823	530
530815	530823	530814	530812	530
530816	530817	530813	530812	530
530817	530820	530813	530822	530
530810	530813	530821	530822	530

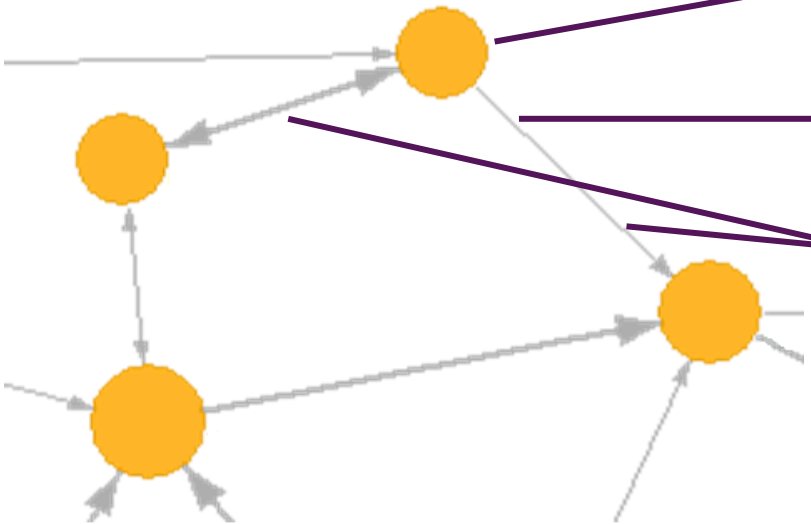
Edgelist: allows edge attributes

StudyID	toStudyID
530801	530802
530801	530807
530801	530809
530801	530810
530801	530811
530801	530812
530801	530825
530801	530826
530802	530801
530802	530808
530802	530809
530802	530810
530802	530812
530802	530822
530802	530825
530804	530809
530804	530811

Adjacency matrix: mathematically convenient

```
530801 . 1 . 1 . 1 1 1 1 . . . . . . . . . . . . 1 1
530802 1 . . . 1 1 1 . 1 . . . . . . . . . 1 . . 1 .
530804 . . . . . 1 . 1 . . . . . . . . . . . 1 . 1
530807 1 1 1 . 1 1 1 1 . . . . . . . . . . . . . 1
530808 . . . 1 . 1 1 1 . . . . . . . . . . . 1 . .
530809 . . 1 . 1 . . 1 . . . . . . . . . . . 1 . 1
530810 1 1 1 . . 1 . 1 1 . 1 . . . . . . . 1 . 1 1
530811 . . 1 1 1 1 1 . . . . . . . . . . . . 1 . 1
530812 1 . . . . 1 1 . . 1 1 1 . . . 1 . . 1 1 1 .
530813 . . . . . . . . . . . . . 1 . 1 1 1 . . .
530814 . . . . . 1 . 1 1 . 1 . . 1 1 1 1 1 . 1 .
530815 . . . . . . . . 1 1 1 . . . . 1 . 1 1 . 1 .
530816 . . . . . . . . 1 1 . 1 . 1 1 1 1 1 1 . 1 .
530817 . . . . . . . . 1 1 1 1 1 . 1 1 . 1 1 . 1 .
530819 . . . . 1 . . 1 . 1 . . . . . . 1 1 . . . .
530820 . . . . . . . 1 1 1 1 1 . . . 1 . . 1 1 1 1 .
530821 . . . . . . . . 1 1 1 . . . 1 1 . 1 1 . 1 .
530822 . . . . . . . . 1 1 1 1 . 1 1 1 1 . 1 . 1 .
530823 . . . . . . . 1 1 . . 1 . . 1 1 . 1 . 1 1 .
530824 . . 1 . . . . . . . . . . 1 . . . . . . 1
530825 . . . 1 . . 1 1 1 1 1 . . . . . . 1 1 . . .
530826 1 1 . 1 . 1 . . . . . . . . . . . . 1 . . .
```

Representation: visual



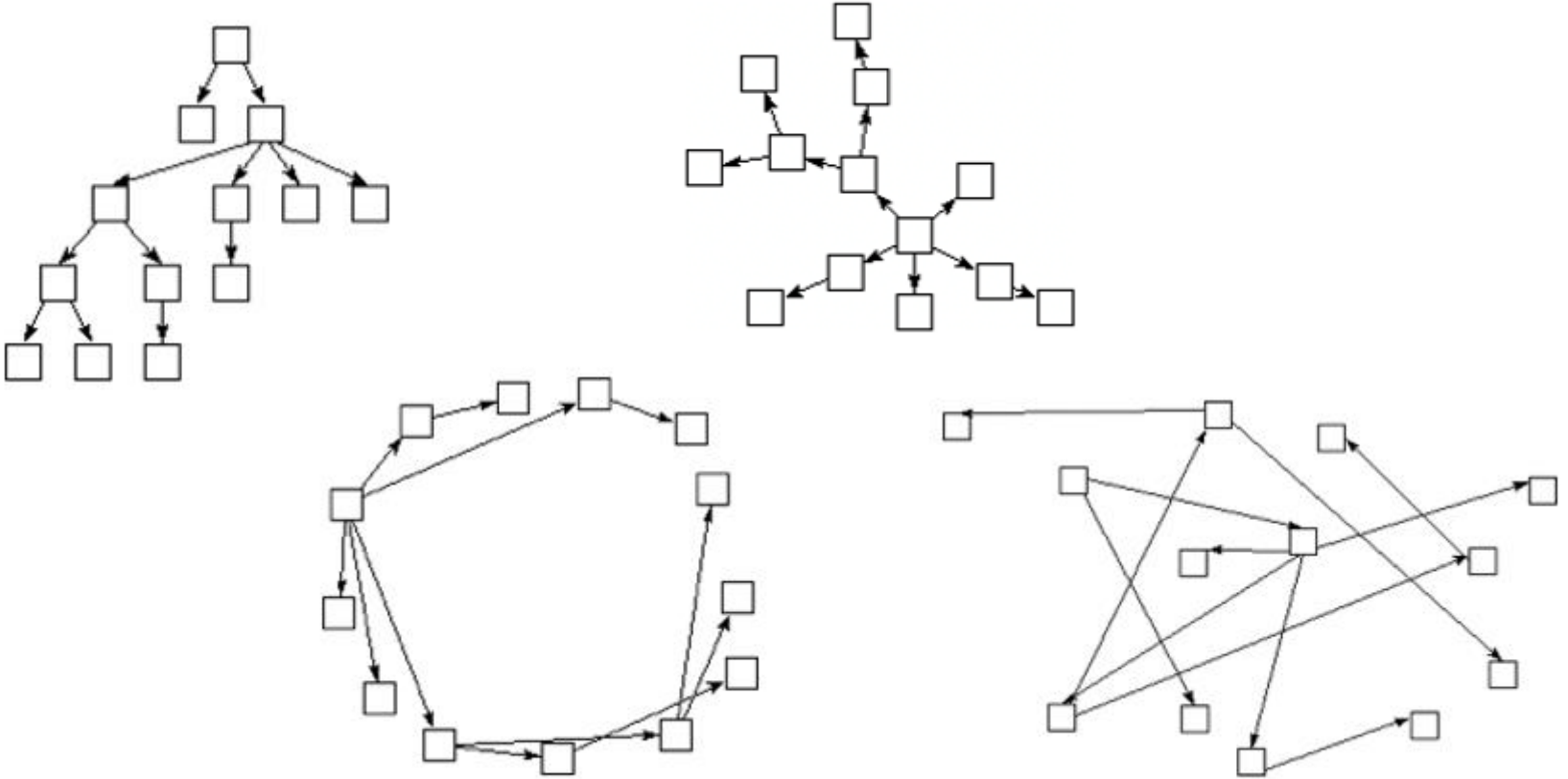
Each case is represented as a node (vertex)

Each relationship between cases are represented as lines (edge)

Edges can vary in intensity (weight), represented by line thickness

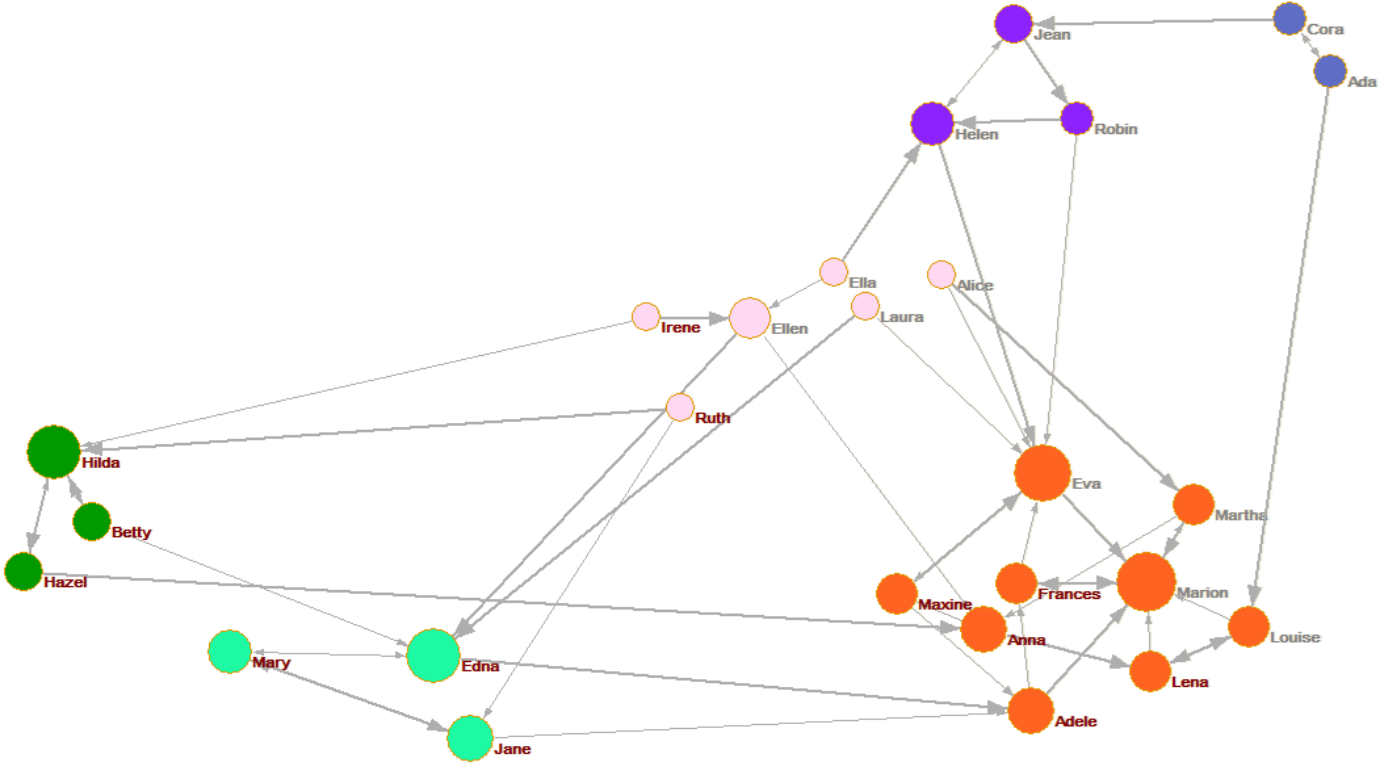
Edges can be directional (arrow) or undirected (line, or double arrow)

No standard layout: these are the same network



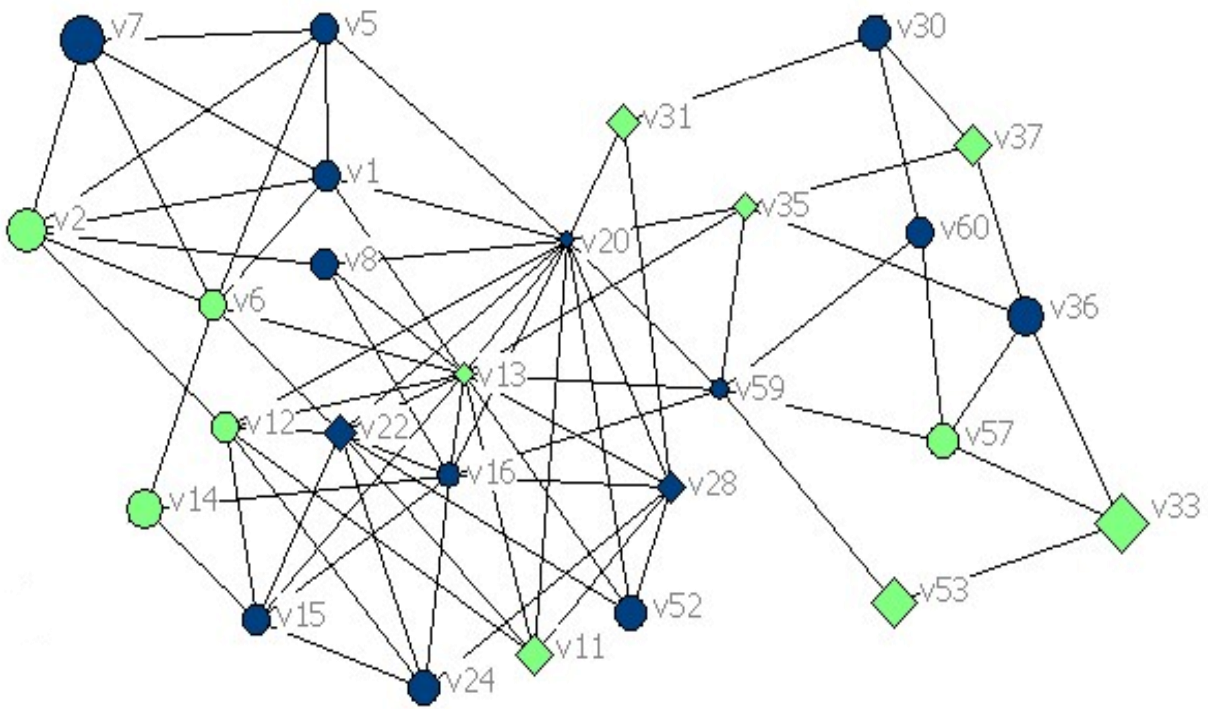
Durham University

Layouts instead used to emphasise structures



Durham University

Colour, size, shape of nodes for different attributes



Durham University

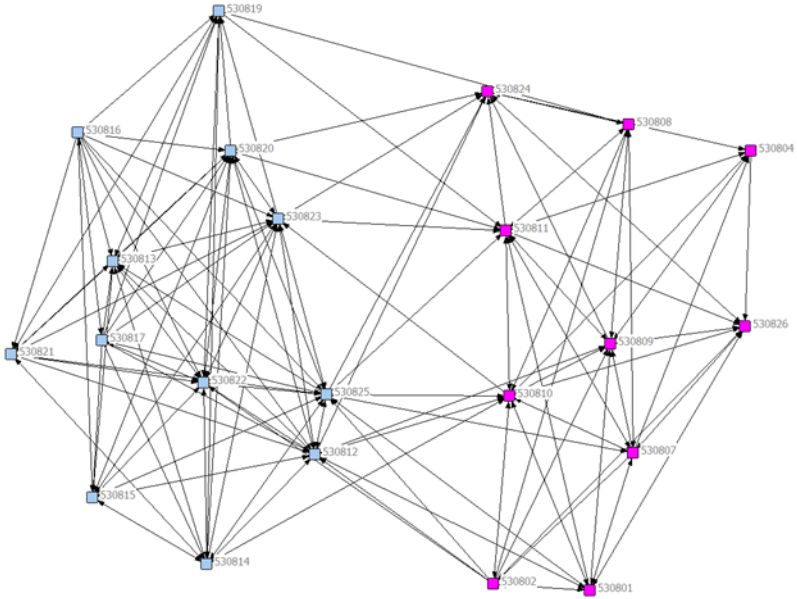
Appropriate visualisation reveals network properties

Secondary school friendships

- Teenage Friends and Lifestyle Study (Glasgow)

Nodes coloured by gender

- Notice that same gender edges are relatively dense
- Many same gender friendships
- Few mixed gender friendships



https://www.stats.ox.ac.uk/~snijders/siena/Glasgow_data.htm

Durham University

Sources of error are particularly problematic in social network analysis

Missing cases/attribute data

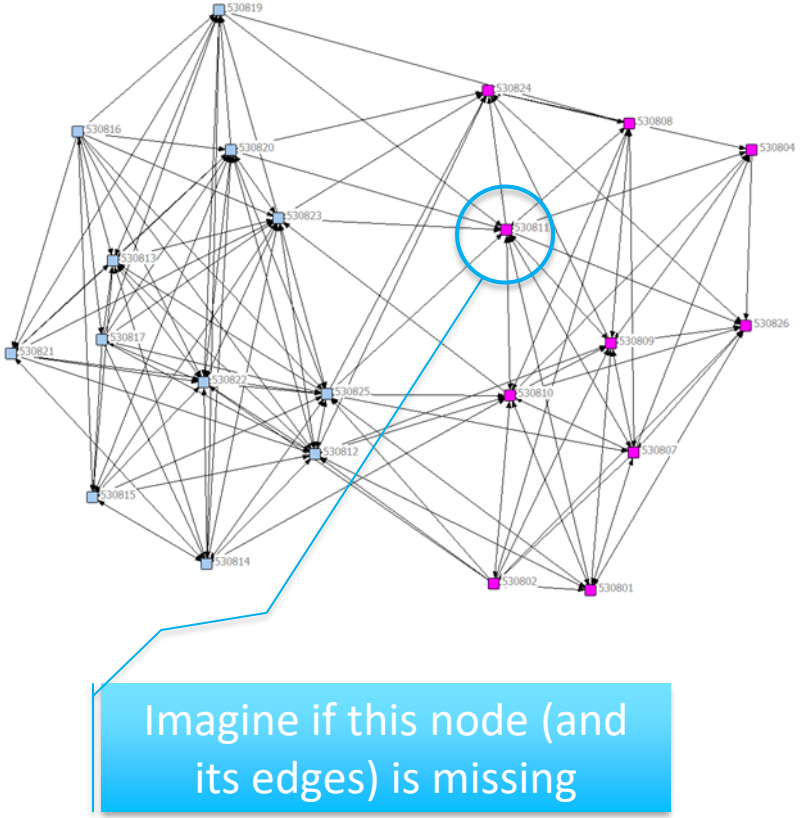
- Lack of data can undermine especially a sociocentric design as it renders the network incomplete and analysis circumspect

Misattribution of information/ties

- Relying on subjective accounts of relations or information about other people may not be reliable and valid

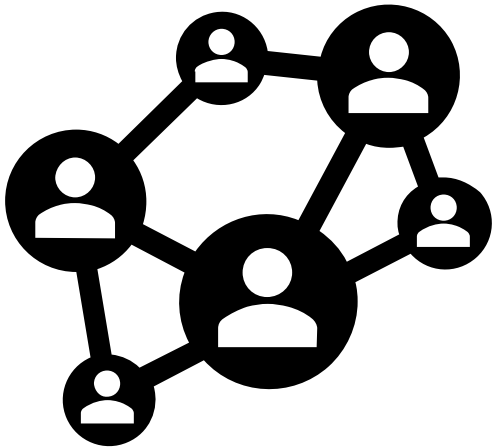
Retrospection error

- Requiring participants to recall interactions/ties might result in poor quality or inaccurate data



Durham University

Ego networks



Durham University

Ego-networks: relationships between a central case (egos) and the others cases (alters) they connect to

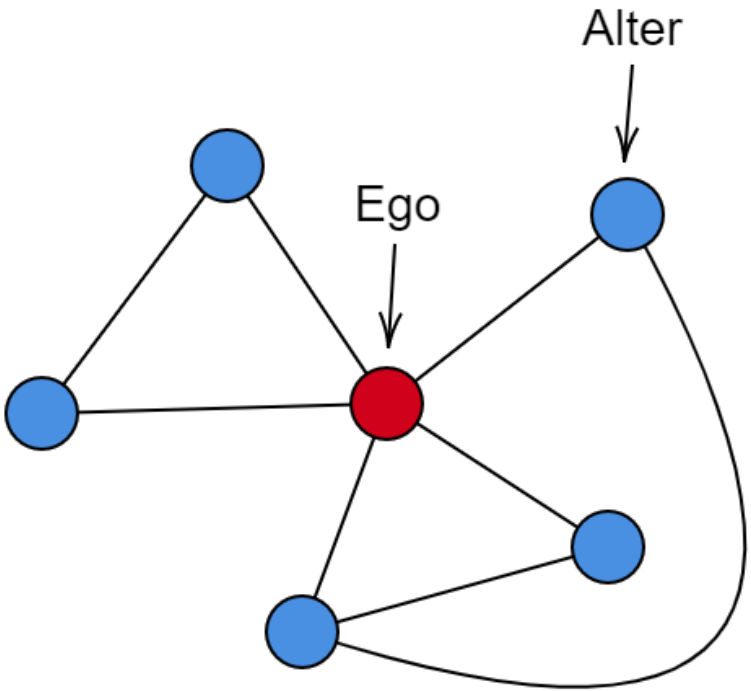


Image source: https://commons.wikimedia.org/wiki/File:Egocentric_Network.png

Analysis of ego networks

Used to sample from a larger group or network:

- If the full network is too large to analyse as a whole
- If the full network is not sufficiently connected and so cannot easily be traced

Popular in studies of social integration such as migration

Podolny and Barren (1997) studied career progression

- Collected workplace ego-networks for 658 employees at the same firm
- People with large, sparse networks of informal ties promoted more often and to higher levels

Joel M. Podolny and James N. Baron “Resources and Relationships: Social Networks and Mobility in the Workplace,” American Sociological Review, 62(5), pp 673-693

Relationships between alters critical for ego-networks

Primary data collection uses Name Interpreters

- Understand how the people named connect with each other

For example: “Please think about the relations between the people you just mentioned. Some of them may be total strangers in the sense they wouldn’t recognise each other if they bumped into each other on the street. Others may be especially close, as close or closer as they are to you. First, think about NAME1 and NAME2. Are they total strangers? Are they especially close?”

(General Social Survey, 1985; cited in Knoke and Yang, 2008: 21)

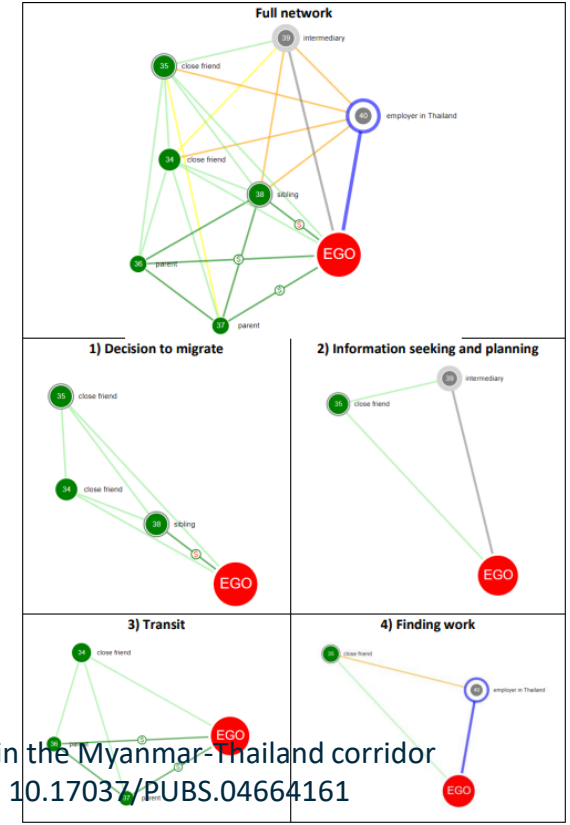
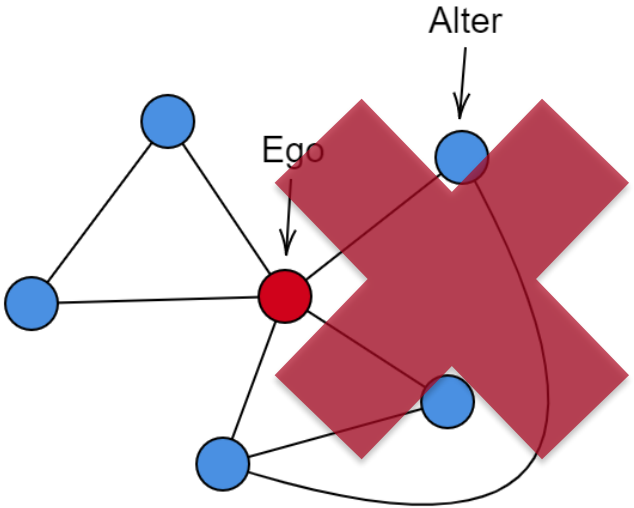


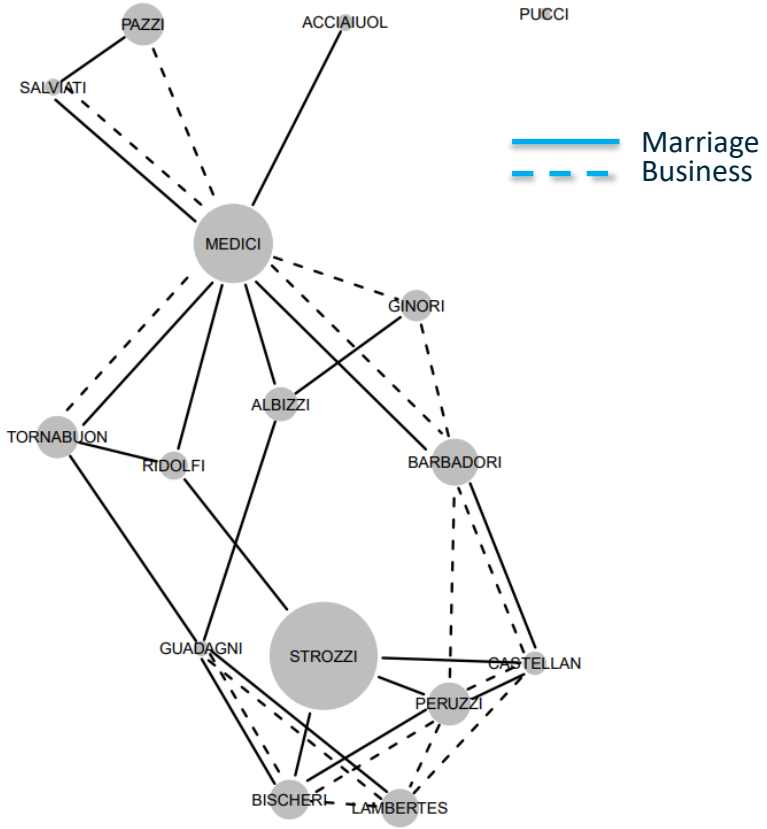
Figure 50. Example of ego-net dynamics over migration stages

McAlpine, AM (2021). Mediated labour migration in the Myanmar-Thailand corridor and precarious outcomes (PhD thesis LSHTM). doi: 10.17037/PUBS.04664161

Durham University

Our focus: whole network (socio-centric) analysis

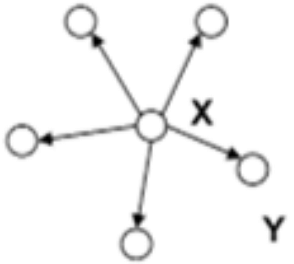




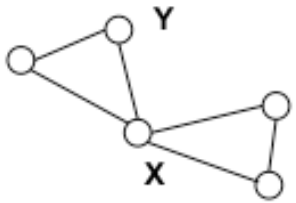
Durham University

Centrality: Influence within network

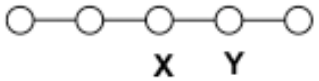
indegree



outdegree



betweenness



closeness

Durham University

Centrality

Centrality is a NODE level attribute

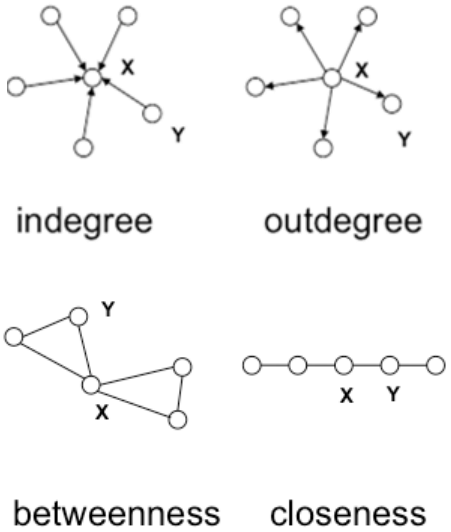
Concerns each node's influence within a network

- common measures are degree, betweenness and closeness
- MANY other measures used

'Best' measure depends on meaning of influence

- eigenvector (and PageRank) accounts for the influence of the connections as well

Centrality X > centrality Y for all networks shown



Properties of nodes

Centrality measures concern different types of importance in the network

- Degree (popularity): number of edges
- Closeness: how far to other nodes
- Betweenness: paths in the network pass through the node
- Spectral (eg eigenvector, PageRank): importance of nodes it is connected to

Typical analysis:

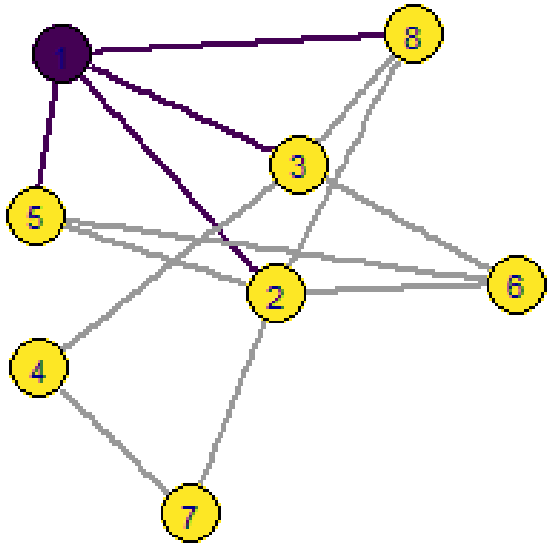
- Identifying nodes with high (low) values and considering their influence
- Comparing node network properties with social capital concepts such as support

Node property: degree

Number of edges attached (incident) to node

- If directed: in-degree and out-degree

Example: purple node has degree = 4



Quiz question

In an undirected network with 7 nodes, what is the maximum number of edges possible?

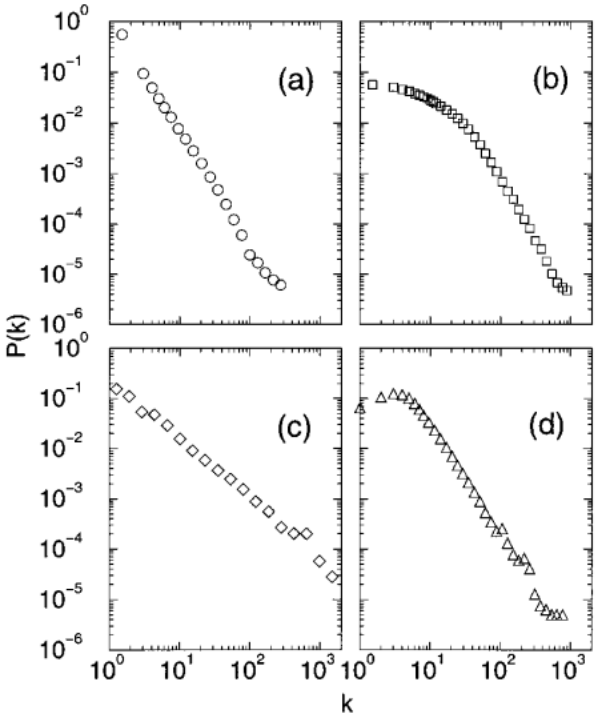
- HINT: Think about how many ‘other’ nodes each node can be connected to
- 21
- 7 nodes can each connect with the 6 other nodes (7 x 6 = 42)…
- …and the edge AB is also BA for undirected (42 / 2= 21)

Scale-free degree distribution

Many real-world networks have highly skewed degree distributions

- (a) internet
- (b) movie actor collaboration
- (c) co-authorship in high-energy physics
- (d) co-authorship in neuroscience

Social networks that involve real world interactions do not have extremely high degree



Albert and Barabási (2002). Statistical mechanics of complex networks. doi: 10.1103/RevModPhys.74.47

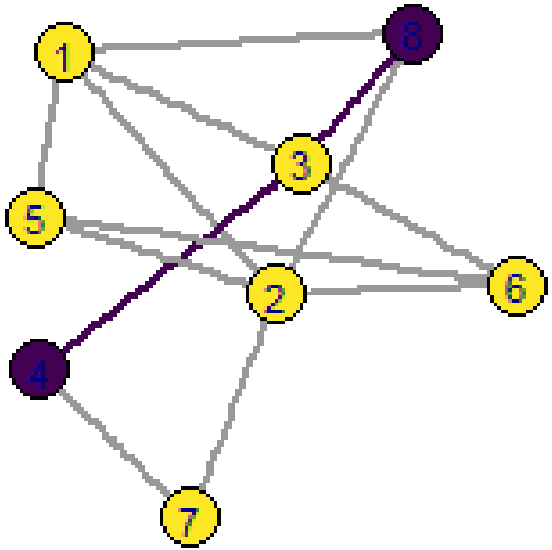
Shortest path (geodesic)

Shortest path is a property of pairs of nodes

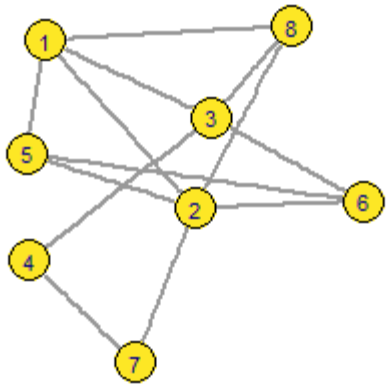
- Calculated for pairs of nodes
- How many edges to get from one to the other

[4,8] has SP of 2

- shortest path is 238
- other paths include 4728, 4318, 436218, 4726318 or …



Node property: Betweenness



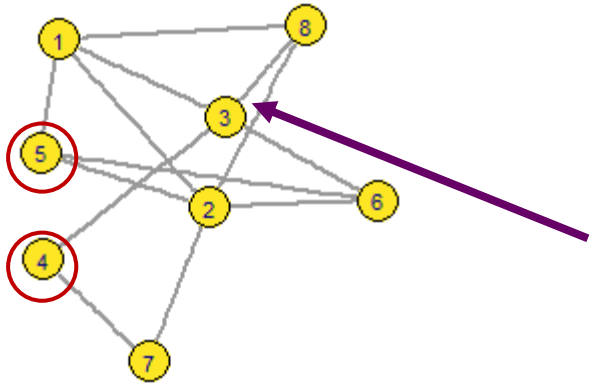
The more paths that must be traced through a node, the more central it is

- Controls information that passes between sections of the network

Betweenness is the number of shortest paths passing through the node

- Node 3 on shortest paths 48 and 46 and others
- Node 3 is on some paths between 17 or 18, but not on shortest path

Quiz questions



What is the length of the longest geodesic in the network?

- 3 (between nodes 4 and 5)
- Note: also called diameter

How many geodesics pass through node 3?

- don’t count paths to node 3
- 5 (16, 48, 14, 45, 68)
- Note: betweenness of 4 (45 and 68 are both 1 of 2)

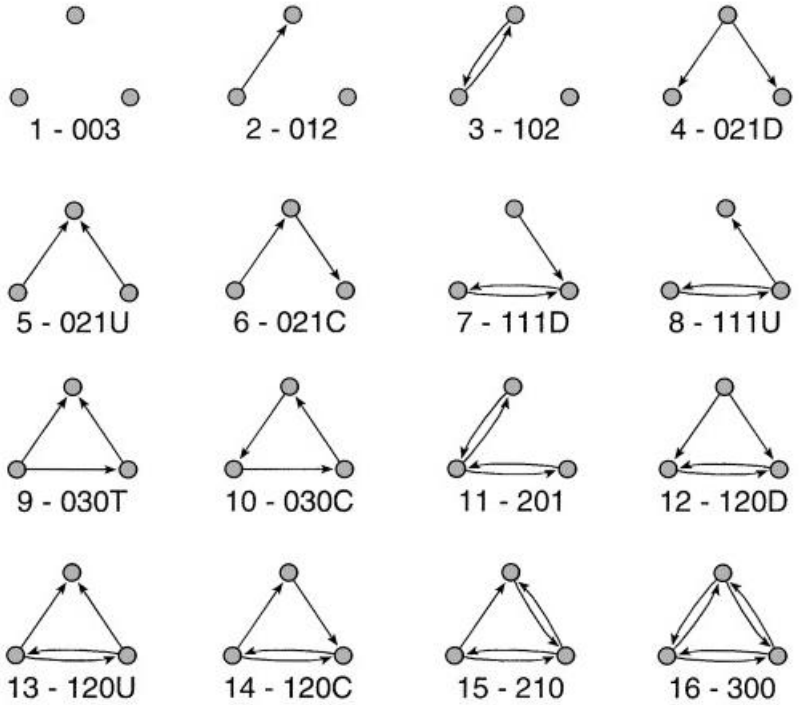
Triads: three nodes and their edges

Numbers indicate MAN reference (directed)

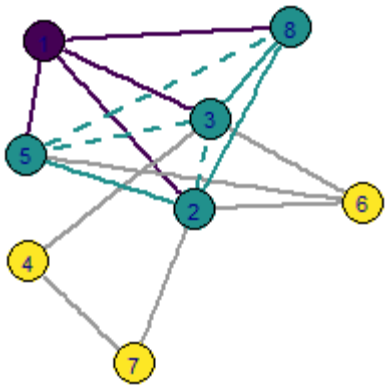
- M is mutual
- A is asymmetric
- N is null (absent)

Final letter distinguishes those with same MAN

- Cycle
- Transitive
- Up
- Down



Node property: clustering coefficient



Conceptually if two people have a friend in common, likely to be friends themselves

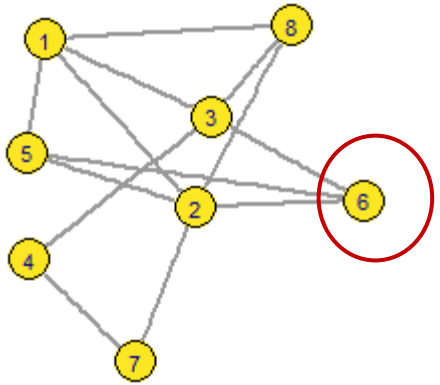
Proportion of potential edges between neighbours that exist

- OR Triangles centred on node

Example: purple node has cc = 0.5

- 4 neighbours
- 6 potential edges
- 3 actual edges

Quiz questions



Degree of node 6?

- 3

Clustering coefficient of node 6?

- HINT: How many possible edges?
- 1/3

Strength of weak ties: theoretical consideration of triads

Strong relationships (close friends) are important in network structure

Weak ties access other areas of the network and have an important role in accessing new information

- New ideas
- Opportunities (such as job vacancies)

The Strength of Weak Ties[1]

Mark S. Granovetter
Johns Hopkins University

Analysis of social networks is suggested as a tool for linking micro and macro levels of sociological theory. The procedure is illustrated by elaboration of the macro implications of one aspect of small-scale interaction: the strength of dyadic ties. It is argued that the degree of overlap of two individuals' friendship networks varies directly with the strength of their tie to one another. The impact of this principle on diffusion of influence and information, mobility opportunity, and community organization is explored. Stress is laid on the cohesive power of weak ties. Most network models deal, implicitly, with strong ties, thus confining their applicability to small, well-defined groups. Emphasis on weak ties lends itself to discussion of relations *between* groups and to analysis of segments of social structure not easily defined in terms of primary groups.

Other theories focussing on triads

Balance Theory

Heider (1958) argues that triads are balanced (stable) if an even number (or none) of negative relations

- My friend's friend is my friend
  My friend's enemy is my enemy
  My enemy's friend is my enemy
  My enemy's enemy is my friend

Hummer & Doreian (2003). Some dynamics of social balance processes: bringing Heider back into balance theory. doi: 10.1016/S0378-8733(02)00019-9

Creativity from Forbidden Triads

Forbidden triads have high weight edges and an open potential edge

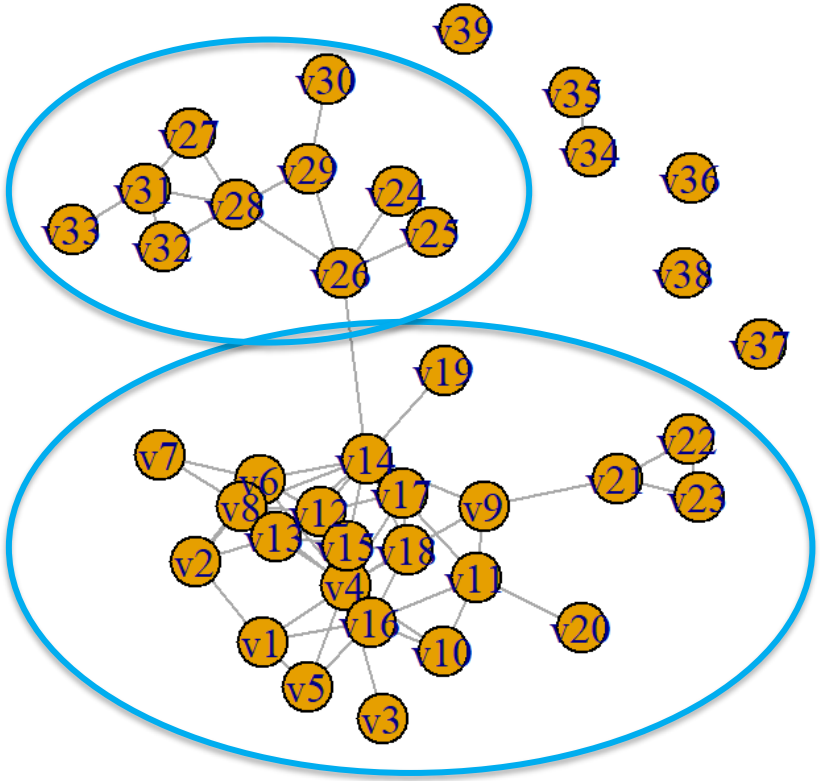
- Expected that edge would close

Vedres (2017) argues that a high density of forbidden triads is associated with creative success

- Intersection of separate communities

Vedres (2017). Forbidden triads and creative success in jazz: the Miles Davis factor. doi: 10.1007/s41109-017-0051-2

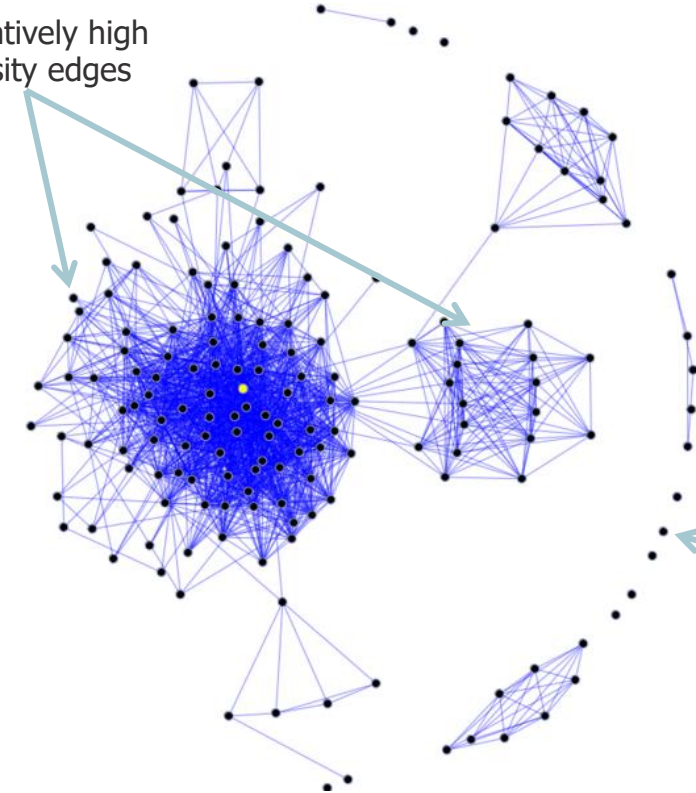
Meso-structure



Durham University

Larger structures

Community (or Cluster): area of relatively high edge density, separated by low density edges



Component: disconnected subnetwork

Isolate: disconnected node

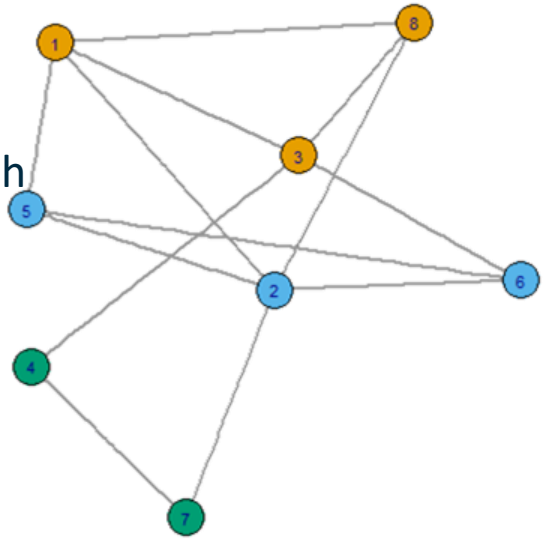
Image source: https://commons.wikimedia.org/wiki/File:Sna_large.png

What is a community?

Communities are subnetworks where edges are between members more than connecting to other subnetworks

Different approaches to identify communities

- Calculate betweenness of all edges, delete the edge with highest value, iterate
- Random local walks
- Brute force: trial and error to maximise modularity



Properties of the whole network

Density

- Proportion of potential edges that are realised

Diameter

- Length of the longest geodesic
- Maximum distance to traverse

Centralisation

- Distribution of node centrality

Homophily of an attribute

- Measured as the correlation coefficient of the attribute over all pairs of nodes connected by an edge
  - Tendency for edges where the node pair have common attribute
- Also termed assortativity, particularly if that property is degree (popularity)

Network property: Centralisation

Centralisation is a NETWORK level attribute

Concerns how much a single node dominates, or how evenly distributed edges are across the network

- other nodes are not considered

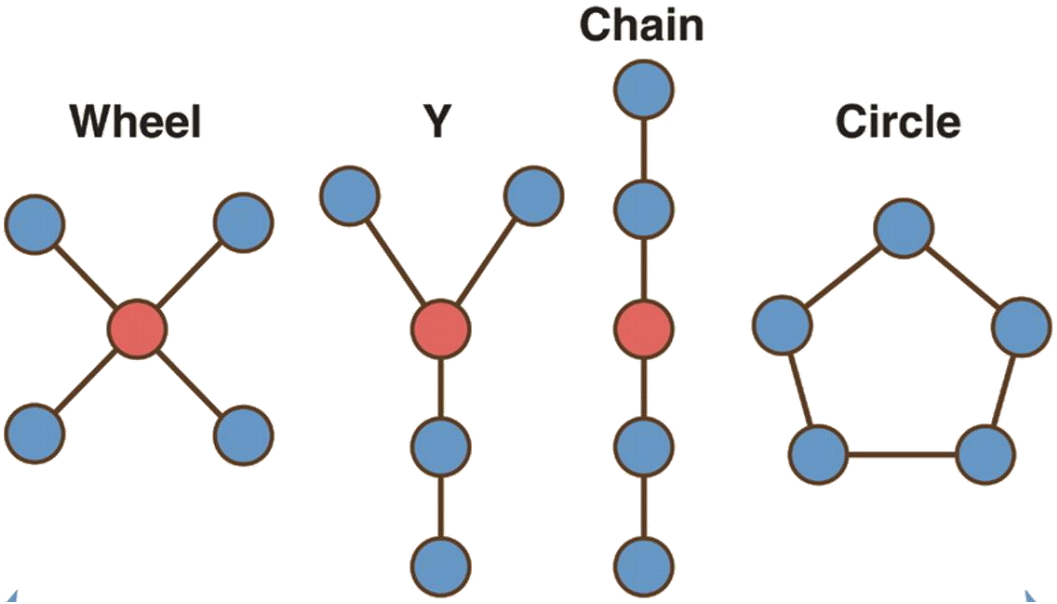
Calculated as:

Σ (centrality of most central node – centrality of each node)

theoretical maximum of such Σ for networks with the same number of nodes

Think of it as the difference between the maximum degree [or other centrality property] and the mean degree in a network

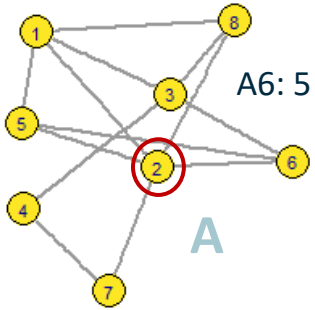
Network property: Centralisation



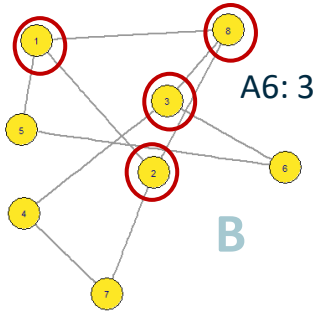
Durham University

Quiz questions

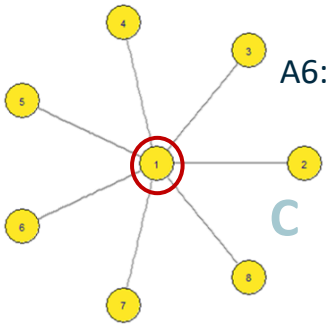
Q6 Maximum degree for each network?



A6: 5



A6: 3



A6: 7

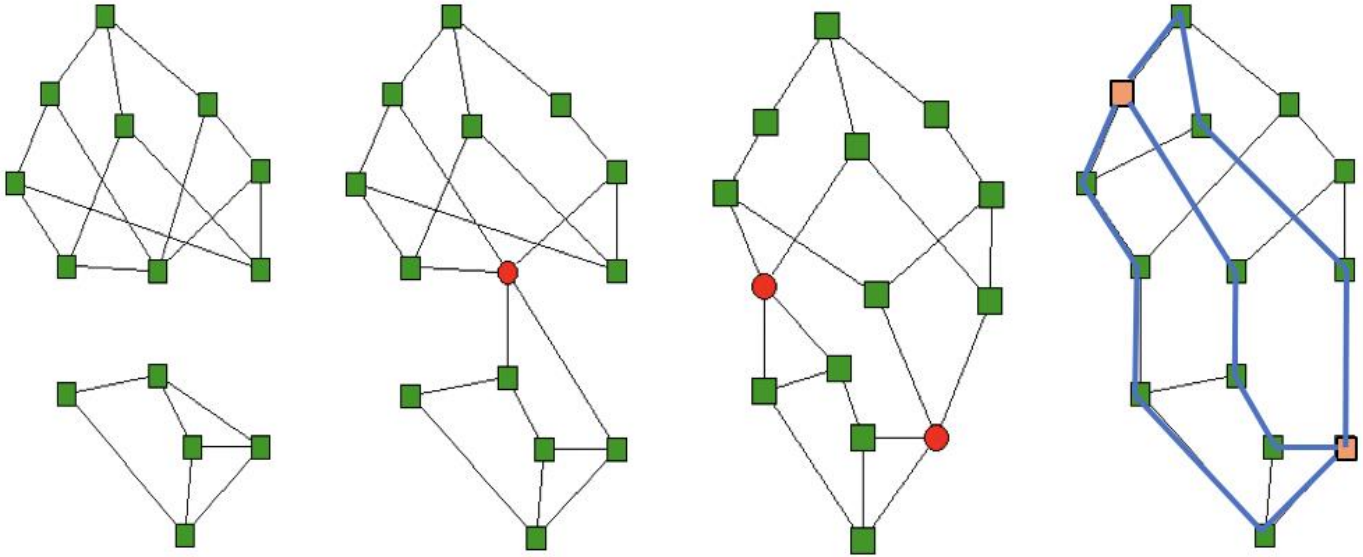
Q7 Rank the networks from most to least (degree) centralised

- Hint: Don't calculate, just compare the maximum degree to the typical degree

A7: C, A, B

Durham University

Cohesive networks remain connected even if nodes removed
Multiple paths between nodes



Node Connectivity

Small world: social networks have high clustering (triangles) but also short paths

Apparently inconsistent because triangles are local, so paths should be long

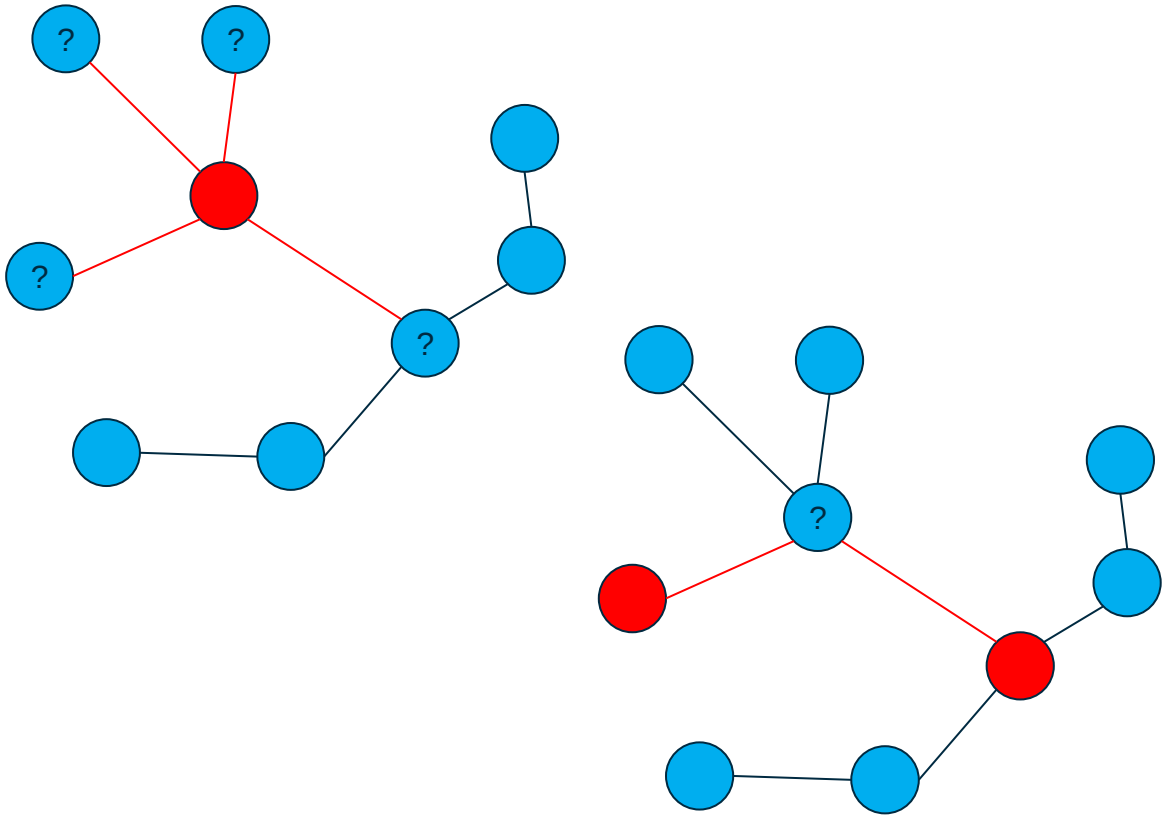
- Mathematical models generated graphs with one or the other, but not both
- Milgram experiments found average 6 steps (for completed chains)

Resolved by Watts and Strogatz

- A small number of random edges substantially reduces average path length

DJ Watts, SH Strogatz (1998), Collective dynamics of 'small-world' networks, Nature 393 440:442

Network Dynamics

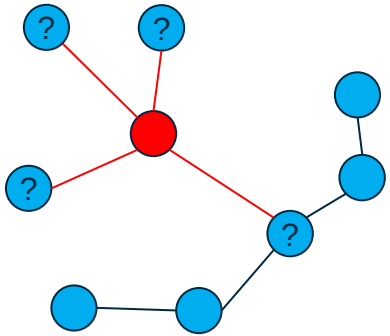


Durham University

Two modes of diffusion through networks

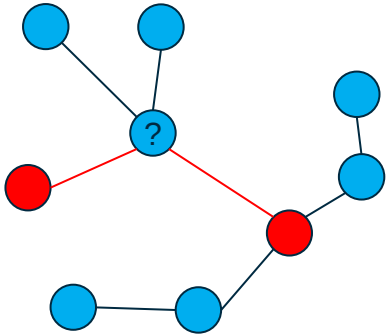
Simple contagion

- Epidemic spread, probabilistic each period of exposure
- Idealised information diffusion



Complex contagion

- Spread depends on the proportion of network neighbours already adopted
- Idealised behaviour diffusion



Durham University

Centola, D., Macy, M. (2007), "Complex contagion and the weakness of long ties" in *American Journal of Sociology* 113(3), 702–734

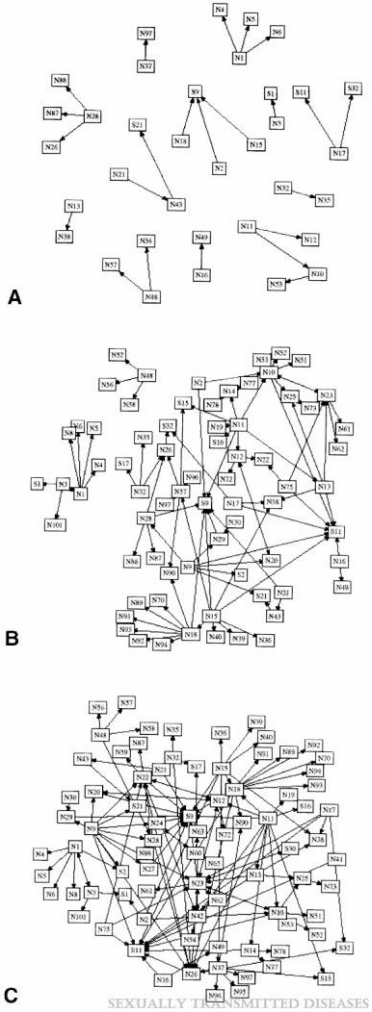
Communicable diseases spread via networks

Example study of a 1996 syphilis outbreak near Atlanta (US)

- 99 young people in sex party group over more than a year
- 10 eventually diagnosed with syphilis

Rothenberg et al (1998) compared traditional epidemiological tools to network based tools

- Constructed sexual history networks at time points
- Network visualisation highlighted the importance of specific girls who connected two separate groups of boys



Rothenberg et al (1998). Using Social Network and Ethnographic Tools to Evaluate Syphilis Transmission. Fig 2
https://journals.lww.com/stdjournal/Fulltext/1998/03000/Using_Social_Network_and_Ethnographic_Tools_to.9.aspx

Durham University

Homophily: network neighbours have similar attributes

Homophily measured as the correlation of attribute values of the pairs of nodes at the ends of edges

- Also referred to as assortativity, particular for degree
- Social networks have positive assortativity, popular people are disproportionally friends with popular people

Similarity arises in two ways

- Selection: choose similar friends
  - Dynamics of networks
- Influence: change behaviour to be similar to friends
  - Dynamics over networks

The only way to differentiate is with longitudinal data

Networks associated with changes in behaviour

Health behaviour of adolescents similar to that of their friends

- Substance use (tobacco, alcohol, drugs), physical activity

Preventive Medicine 130 (2020) 105900

Contents lists available at ScienceDirect

Preventive Medicine

journal homepage: www.elsevier.com/locate/ypmed

Review Article

Peer social network processes and adolescent health behaviors: A systematic review

Shannon C. Montgomery[a,*], Michael Donnelly[a], Prachi Bhatnagar[b], Angela Carlin[c], Frank Kee[a], Ruth F. Hunter[a,*]

Opportunity for networks to be used in interventions

Individuals – select specific initial people

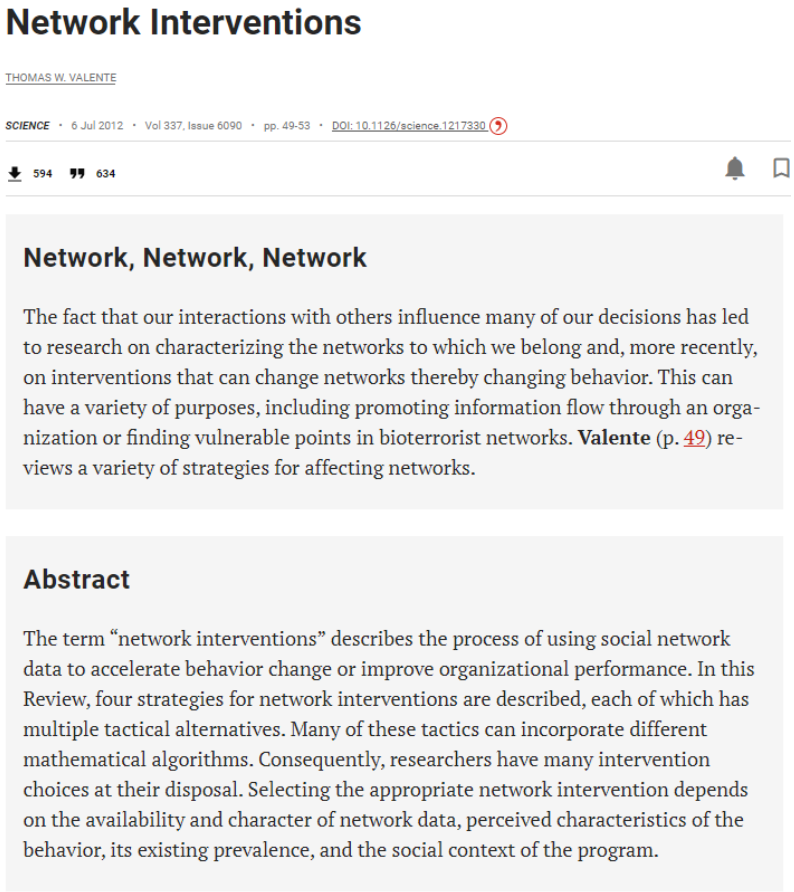
- Leaders, bridges
- Low threshold
- Peripherals

Segmentation – tailor for groups

Induction – encourage extra communication

Alteration – change who interact with

- Add to, or delete from, network



Network Interventions

THOMAS W. VALENTE

SCIENCE · 6 Jul 2012 · Vol 337, Issue 6090 · pp. 49-53 · DOI: 10.1126/science.1217330

Network, Network, Network

The fact that our interactions with others influence many of our decisions has led to research on characterizing the networks to which we belong and, more recently, on interventions that can change networks thereby changing behavior. This can have a variety of purposes, including promoting information flow through an organization or finding vulnerable points in bioterrorist networks. **Valente** (p. 49) reviews a variety of strategies for affecting networks.

Abstract

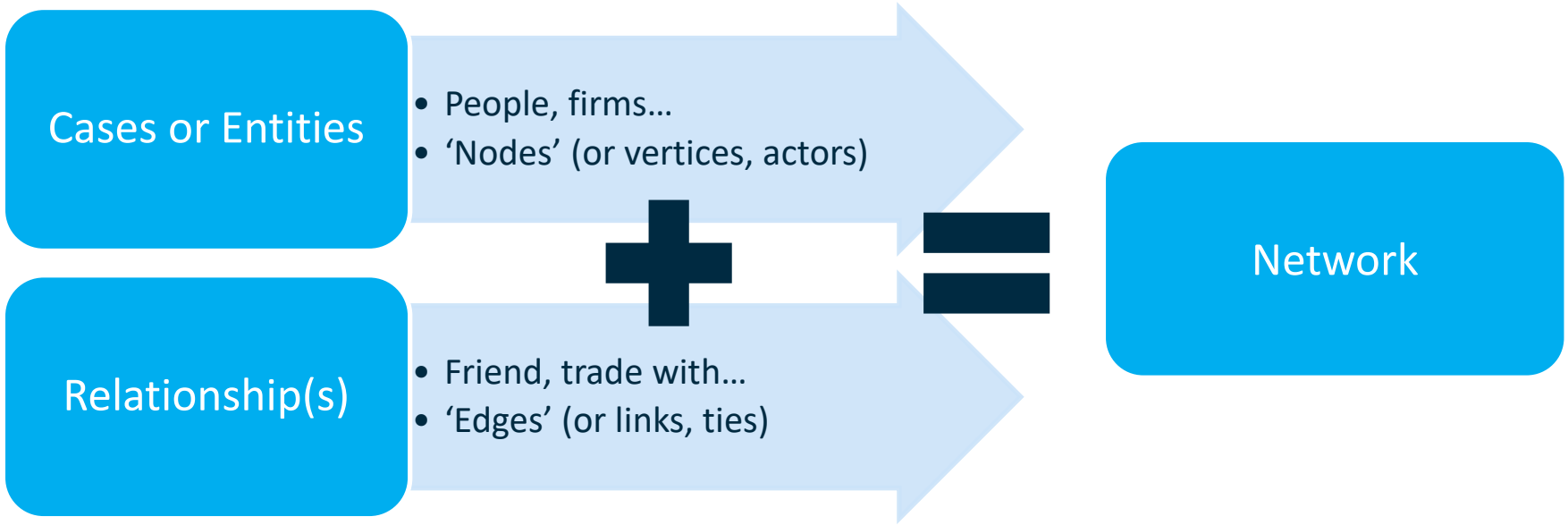
The term "network interventions" describes the process of using social network data to accelerate behavior change or improve organizational performance. In this Review, four strategies for network interventions are described, each of which has multiple tactical alternatives. Many of these tactics can incorporate different mathematical algorithms. Consequently, researchers have many intervention choices at their disposal. Selecting the appropriate network intervention depends on the availability and character of network data, perceived characteristics of the behavior, its existing prevalence, and the social context of the program.

Networks

(wrap up)

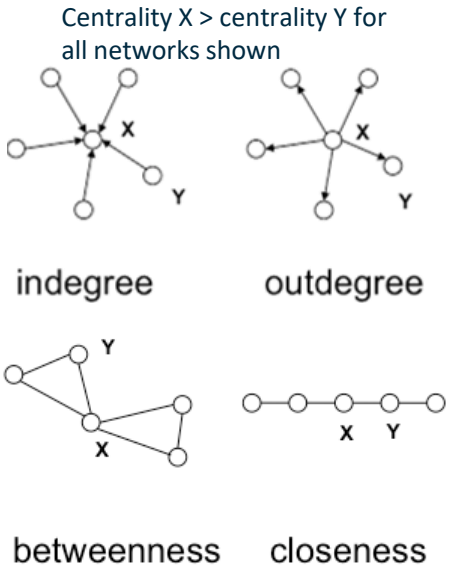
Durham University

Networks comprise BOTH the cases and the relationship

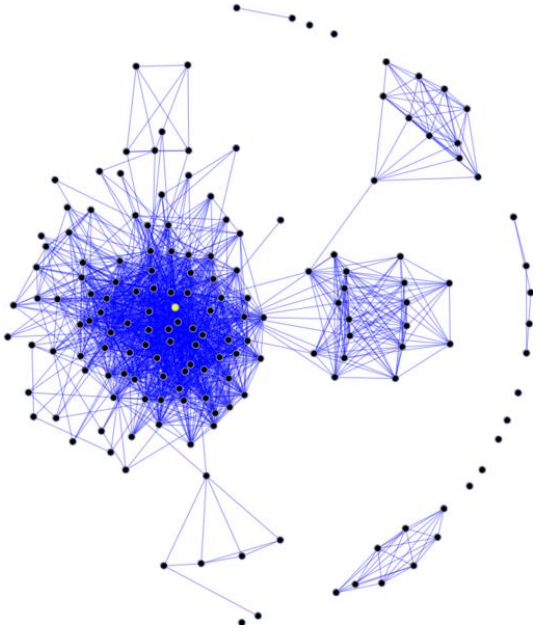


Analysis methods suitable for different questions

Some properties concern the influence of a node within a network



Other properties concern the larger structure of the network



Next session: Social Network Analysis Workshop

Dataset

- Korean villages
- Adoption of family planning

Network measures to compare two villages with different adoption

Workshop will use R throughout

R packages (+ tidyverse)

- igraph

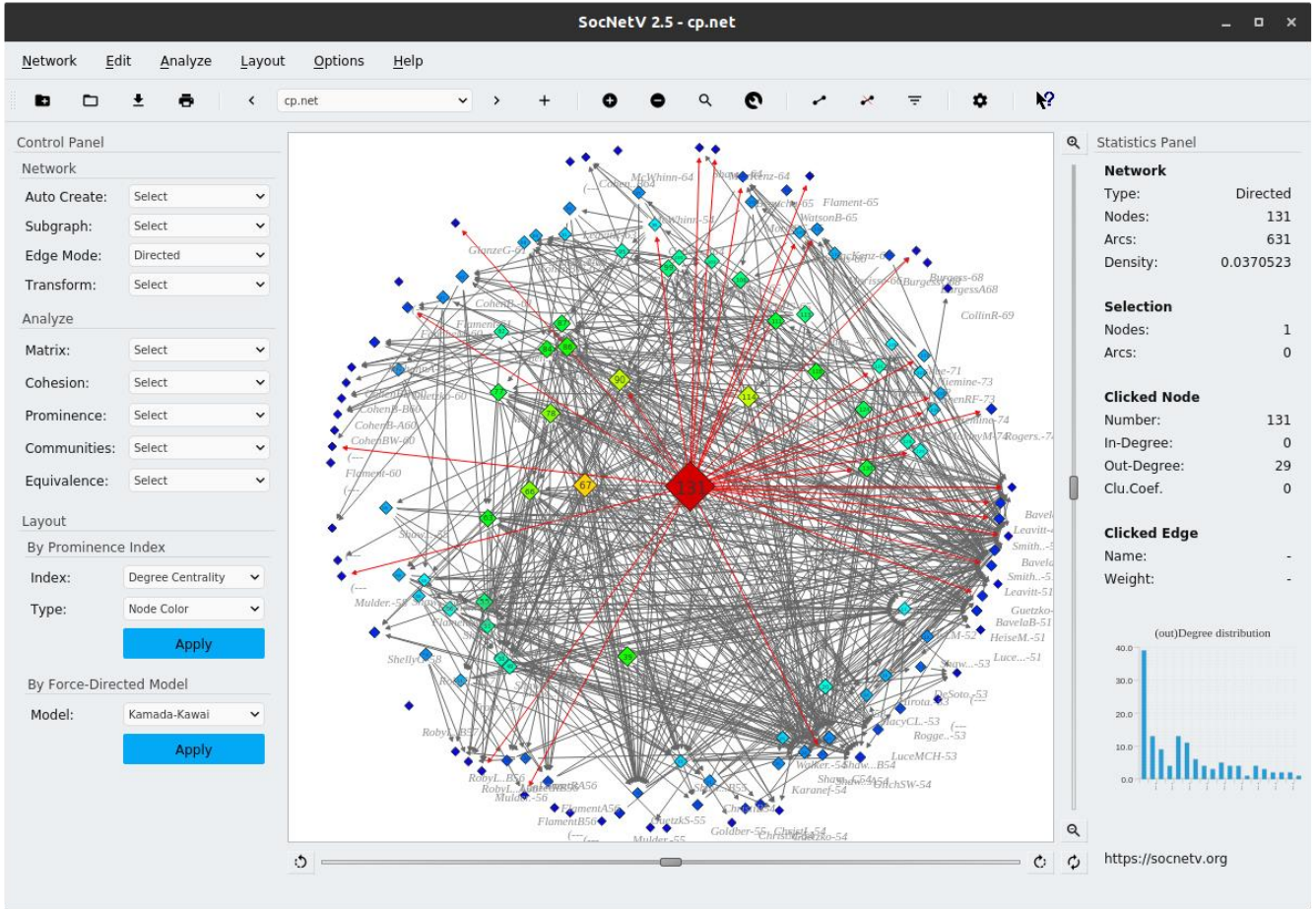
RStudio IDE

- Data analysis scripts
- Interactive web applications
- Documents
- Reports
- Graphs

Durham University

Alternative software: SocNetV

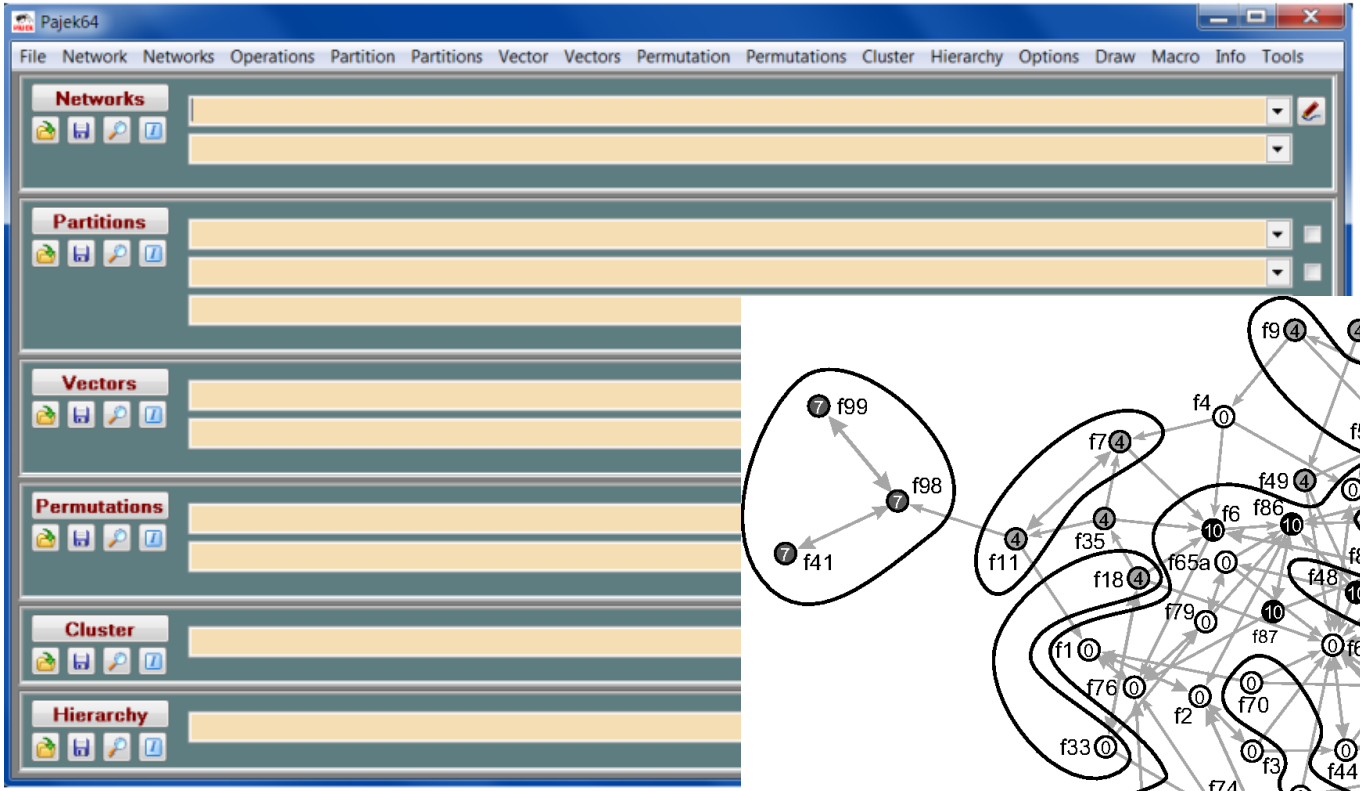
https://socnetv.org/

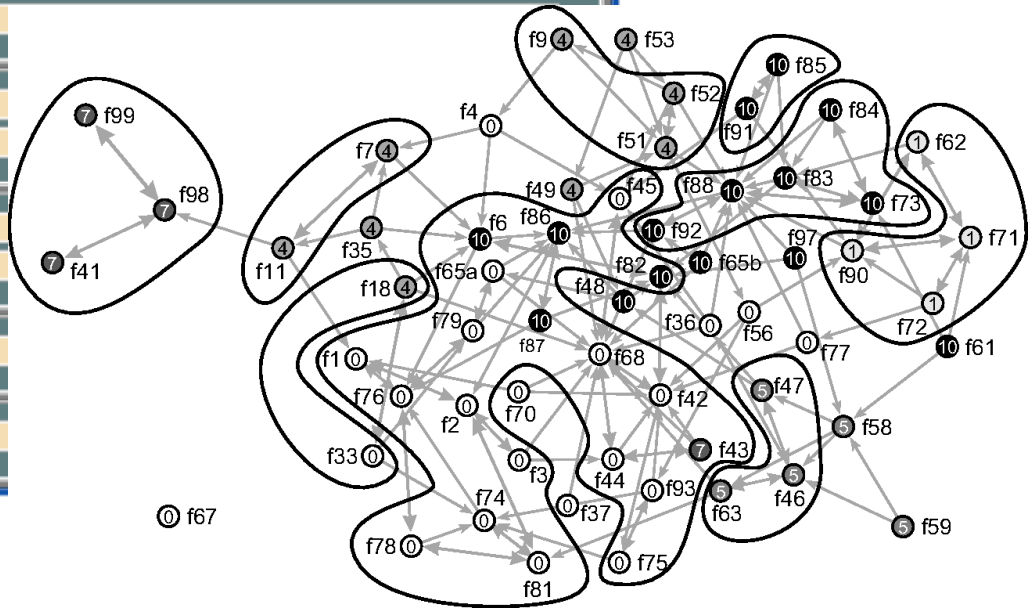


Durham University

Alternative software: Pajek

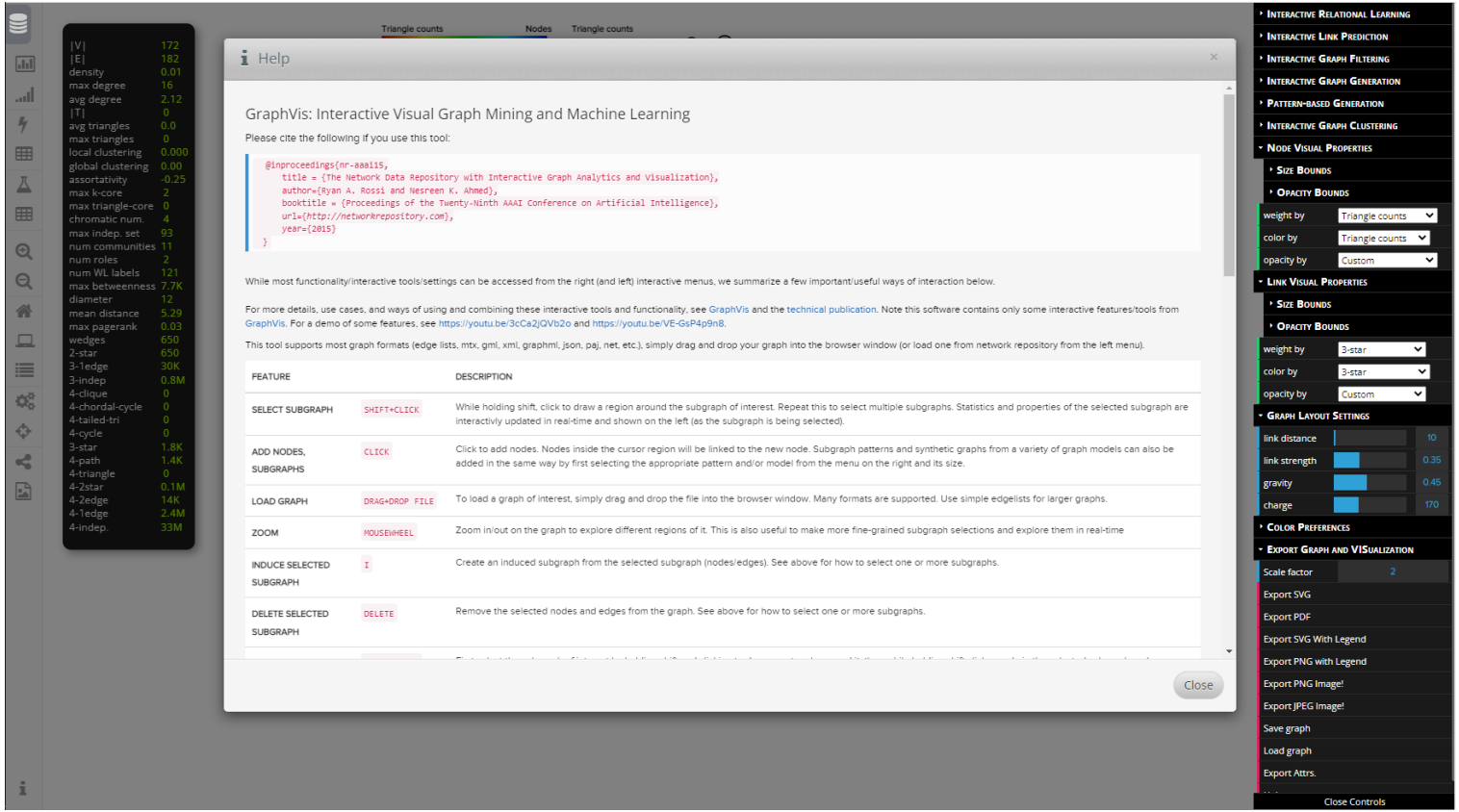
http://mrvar.fdv.uni-lj.si/pajek/





Alternative software: GraphVis

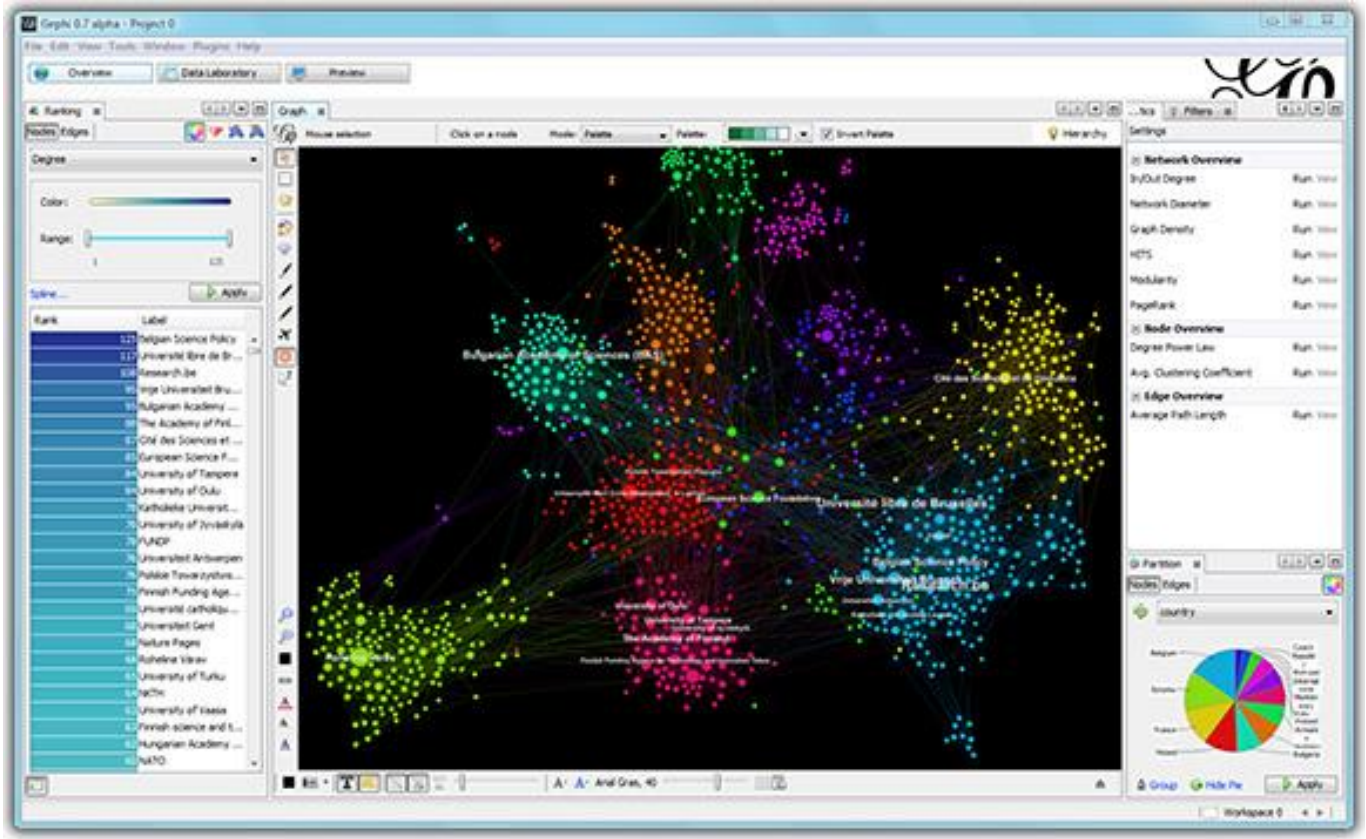
https://networkrepository.com/graphvis.php



Durham University

Alternative software: Gephi

https://gephi.org/



Durham University