Social Network Analysis for Computer Scientists

Frank Takes

LIACS, Leiden University

https://liacs.leidenuniv.nl/~takesfw/SNACS

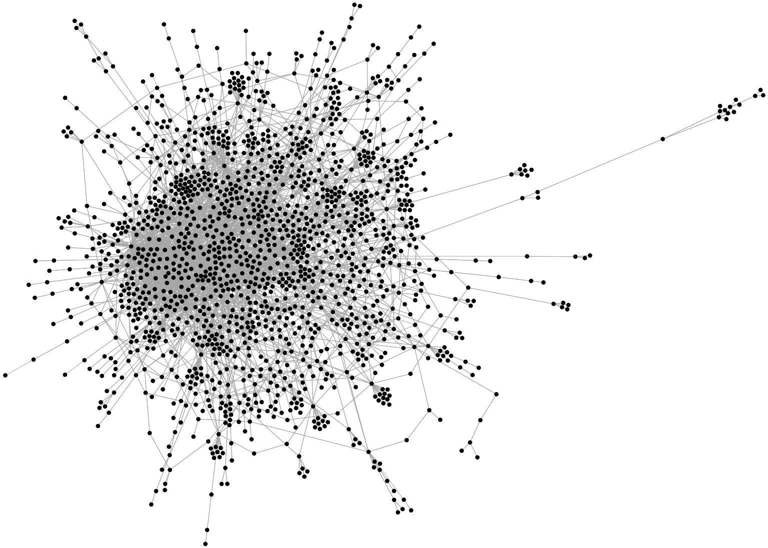
Lecture 3 — Applied network concepts and Network projection

Today

- Recap
- Power law exponent
- Applied concepts
- Network projection
- Case study: criminal networks

Recap

Networks



Notation

Concept	Symbol
■ Network (graph)	$G = (V, E)$
■ Nodes (objects, vertices, ...)	V
■ Links (ties, relationships, ...)	E
■ Directed — $E \subseteq V \times V$ — “links”	
■ Undirected — “edges”	
■ Number of nodes — $\|V\|$	n
■ Number of edges — $\|E\|$	m
■ Degree of node u	$deg(u)$
■ Distance from node u to v	$d(u, v)$

Real-world networks

1. Sparse networks — density
2. Fat-tailed power-law degree distribution — degree
3. Giant component — components
4. Low pairwise node-to-node distances — distance
5. Many triangles — clustering coefficient

- Many examples: social networks, communication networks, citation networks, collaboration networks, protein interaction networks, information networks, webgraphs, financial networks, . . .

Advanced concepts

- Reciprocity
- Complete graph, complement of a graph
- Planar graph, lattice, regular graph
- Tree and dag
- Subgraphs
- Ego network
- Spanning trees
- Diameter
- Bridges

Centrality measures

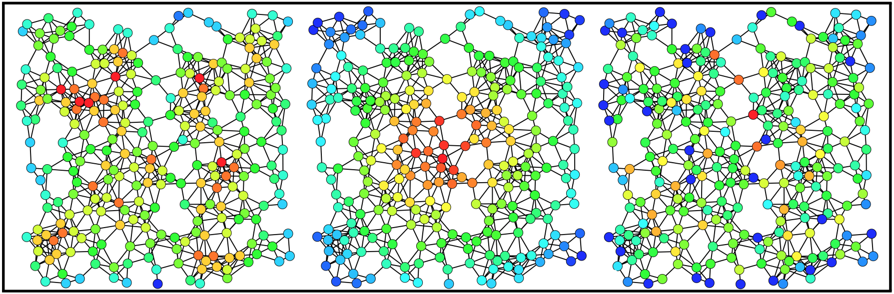


Figure: Degree, closeness and betweenness centrality

Source: “Centrality” by Claudio Rocchini, Wikipedia `File:Centrality.svg`

Degree distributions

Power law degree distribution?

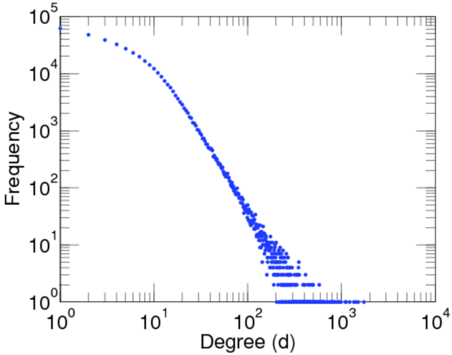
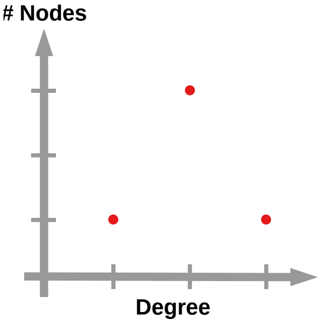


Figure: This is "just" degree vs. frequency, also called a degree scatter plot.

Source: `http://konect.cc/networks/citeseer/`

Degree scatter plot: degree vs. frequency



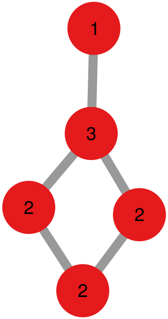
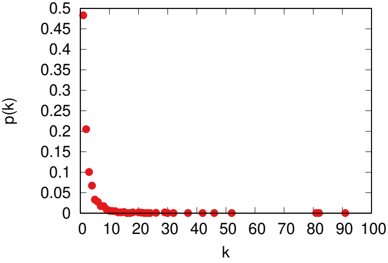


Figure: The degree scatter plot (left) of the graph on the right.

Degree distribution



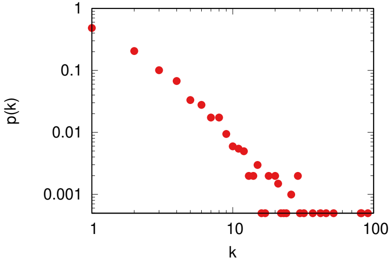


Figure: Degree distribution (degree k vs. its probability $p(k)$) of a protein–protein interaction network, on a linear scale (left) and a log-log scale (right).

Power law exponent in undirected networks

- The probability $p(k)$ of a node having degree k depends on the **power law exponent** α:

$$p(k) \sim k^{-\alpha}$$

- This means that

$$\log p(k) \sim -\alpha \log k$$

  And as such, the straight line in log-log scale plots is observed.
- In real-world networks, α has a value of around 2 to 3
- Useful to compare between similar networks

From degree distribution to CCDF

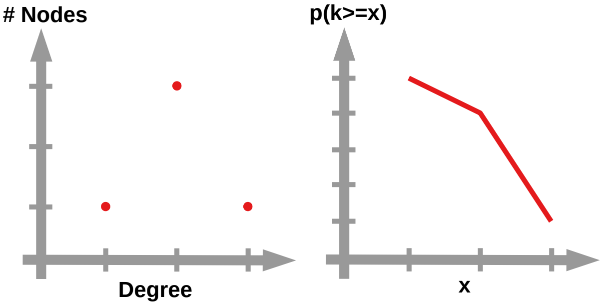
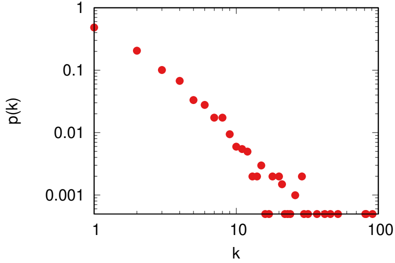


Figure: The degree scatter plot (left) and its complement of the cumulative distribution (CCDF, right): the probability of finding a node of degree k or higher.

Degree distribution and CCDF



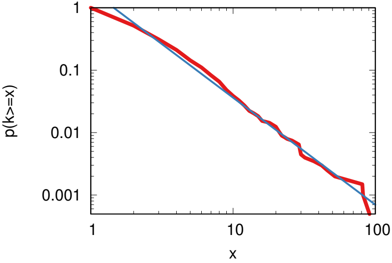


Figure: The same protein–protein interaction network in log-log scale: degree histogram (left) and CCDF with the best fit in blue (right).

Power law

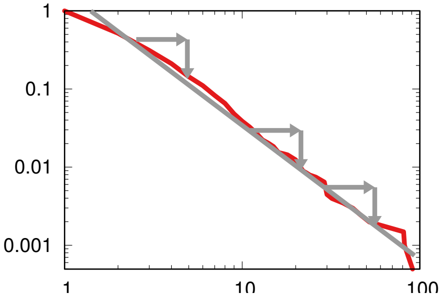


Figure: Every step of the same size along the horizontal axis gives the same proportional drop along the vertical axis, in the head, the middle and the tail.

Comparing power law exponents

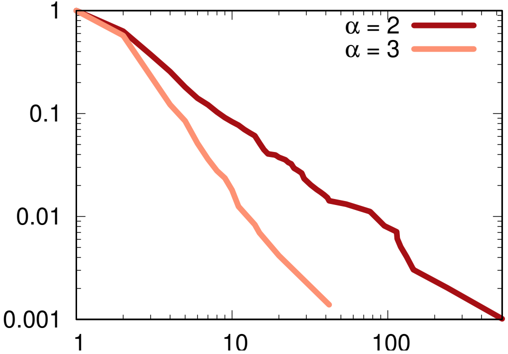


Figure: CCDF degree distributions of two random networks with a different power law exponent; a larger exponent gives a steeper slope and thus smaller hubs.

Power law exponent in directed networks

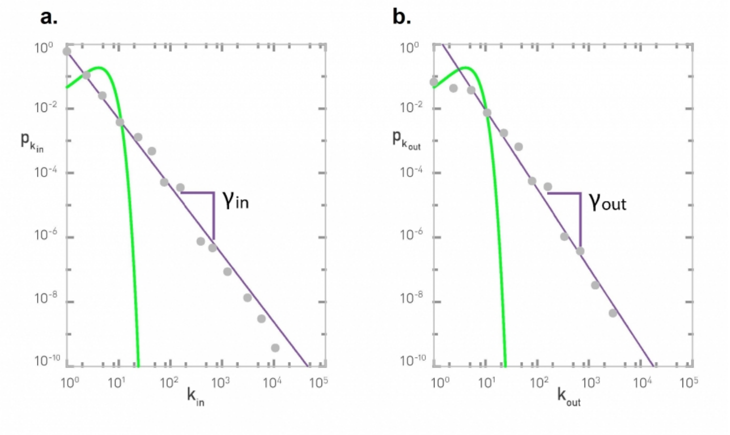


Figure: Instead of α, this figure uses γ.

Source: A. Barabasi, Network Science, 2016.

Applied network concepts

Applied network concepts

- **Operationalization**: linking real-world (applied) concepts to concrete quantitative measures

Applied network concepts

- **Operationalization**: linking real-world (applied) concepts to concrete quantitative measures
- Applied concepts:
  - Cohesion and fragmentation — clustering
  - Strength of weak ties — neighborhood
  - Social capital: bonding and bridging — neighborhood
  - Homophily, segregation — assortativity

Applied network concepts

- **Operationalization**: linking real-world (applied) concepts to concrete quantitative measures
- Applied concepts:
  - Cohesion and fragmentation — clustering
  - Strength of weak ties — neighborhood
  - Social capital: bonding and bridging — neighborhood
  - Homophily, segregation — assortativity
- Examples often from **computational social science**, an interdisciplinary field at the intersection of computer science and social science

Remember the clustering coefficient?

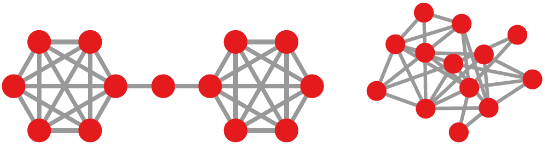


Figure: High clustering (left) and lower clustering (right).

- **Local clustering coefficient** $CC(v)$ is the extent to which a node v forms triangles with its neighbors. For a node $v \in V$:

$$CC(v) = \frac{|\{(u,w) \in E : (u,v) \in E \wedge (v,w) \in E\}|}{\frac{1}{2} deg(v) \cdot (deg(v) - 1)}$$

$$CC(v) = \frac{\text{edges between neighbors of } v}{\text{maximum number of such edges}}$$

Social cohesion

- **Social circle**: your connections (network neighbors)

Social cohesion

- **Social circle**: your connections (network neighbors)
- In your social circles, are you alone, or surrounded by familiar faces?

Social cohesion

- **Social circle**: your connections (network neighbors)
- In your social circles, are you alone, or surrounded by familiar faces?
- **Social cohesion**: strength of local social connectedness

Social cohesion

- **Social circle**: your connections (network neighbors)
- In your social circles, are you alone, or surrounded by familiar faces?
- **Social cohesion**: strength of local social connectedness
- Captures **embeddedness** within tightly connected groups
- Operationalized by **triadic closure**
- Measured by the **clustering coefficient**

Social cohesion

- **Social circle**: your connections (network neighbors)
- In your social circles, are you alone, or surrounded by familiar faces?
- **Social cohesion**: strength of local social connectedness
- Captures **embeddedness** within tightly connected groups
- Operationalized by **triadic closure**
- Measured by the **clustering coefficient**
- Good: it fosters **trust**, **shared norms**, and **collective identity**

Social cohesion

- **Social circle**: your connections (network neighbors)
- In your social circles, are you alone, or surrounded by familiar faces?
- **Social cohesion**: strength of local social connectedness
- Captures **embeddedness** within tightly connected groups
- Operationalized by **triadic closure**
- Measured by the **clustering coefficient**
- Good: it fosters **trust**, **shared norms**, and **collective identity**
- But potentially also **echo chambers**; will you still know different others?

Looking for a job?

- Imagine you are looking for a job. Who do you ask?

Looking for a job?

- Imagine you are looking for a job. Who do you ask?
  - Your parents?

Looking for a job?

- Imagine you are looking for a job. Who do you ask?
  - Your parents?
  - Your best friend?

Looking for a job?

- Imagine you are looking for a job. Who do you ask?
  - Your parents?
  - Your best friend?
  - A distant friend?
- **Strong tie**: a relationship between individuals whose social circles **overlap substantially**; embedded within the same community.
- **Weak tie**: a relationship between individuals whose social circles **overlap little**; often connecting otherwise separate communities, providing access to novel information.
- Note: here, "community" is a somewhat loose term for a group of nodes only used here; later we give a different definition

Strength of weak ties

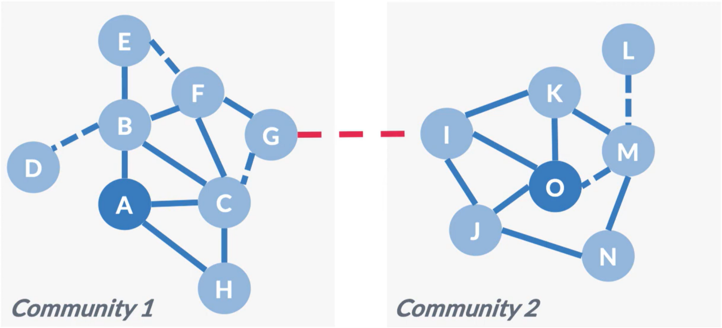


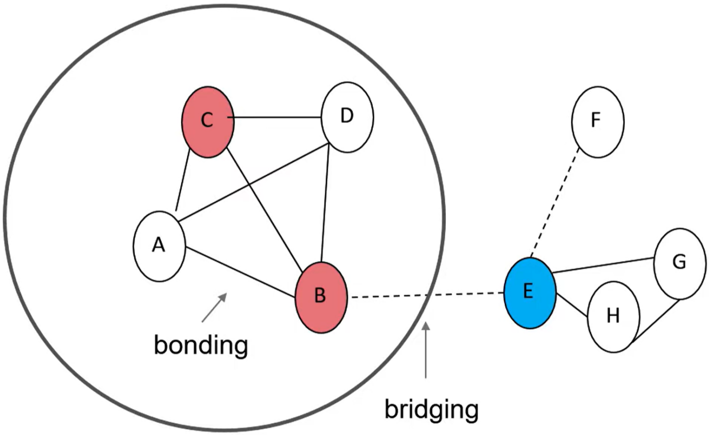
Figure: The red tie is a weak tie.

Source: F. van Tubergen, Utrecht University

Social capital

- **Social capital** refers to the resources and benefits individuals can access through their social relationships and network position.
- Includes access to **information**, **support**, **trust**, and **opportunities**
- Depends not only on who you know, but also on **how your contacts are connected**
- **Bonding social capital**: resources from dense, cohesive groups
- **Bridging social capital**: resources from ties connecting different groups

Bonding and bridging



Source: F. van Tubergen, Utrecht University

Assortativity

- **Assortativity**: extent to which "similar" nodes attract each other
  Value close to -1 if dissimilar nodes more often attract each other
  Value close to 1 if similar nodes more often attract each other

Assortativity

- **Assortativity**: extent to which "similar" nodes attract each other
  Value close to -1 if dissimilar nodes more often attract each other
  Value close to 1 if similar nodes more often attract each other
- **Degree assortativity**: nodes with a similar degree connect more frequently

Assortativity

- **Assortativity**: extent to which "similar" nodes attract each other
  Value close to -1 if dissimilar nodes more often attract each other
  Value close to 1 if similar nodes more often attract each other
- **Degree assortativity**: nodes with a similar degree connect more frequently
- **Attribute assortativity**: nodes with similar attributes attract each other

Assortativity

- **Assortativity**: extent to which "similar" nodes attract each other
  Value close to -1 if dissimilar nodes more often attract each other
  Value close to 1 if similar nodes more often attract each other
- **Degree assortativity**: nodes with a similar degree connect more frequently
- **Attribute assortativity**: nodes with similar attributes attract each other (also called categorical homophily)

Assortativity

- **Assortativity**: extent to which "similar" nodes attract each other
  Value close to -1 if dissimilar nodes more often attract each other
  Value close to 1 if similar nodes more often attract each other
- **Degree assortativity**: nodes with a similar degree connect more frequently
- **Attribute assortativity**: nodes with similar attributes attract each other (also called categorical homophily)
- Social networks terminology: **homophily**
- Complex networks terminology: **mixing patterns**

Attribute assortativity

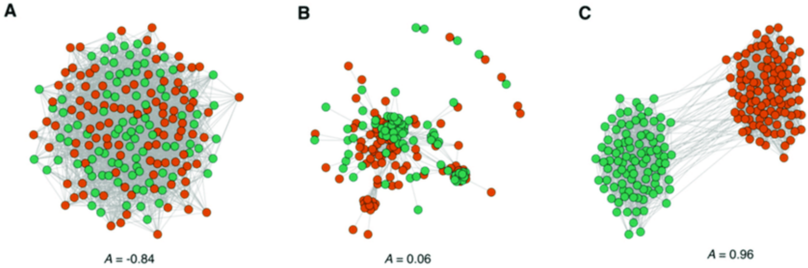


Figure: Attribute assortativity

Image: Moya-García, A. et al. Identification of New Toxicity Mechanisms ... *Genes*, 13(7), 1292, 2022.

Computing assortativity

- **Attribute assortativity coefficient** (Newman)

$$r = \frac{\sum_i e_{ii} - \sum_i a_i b_i}{1 - \sum_i a_i b_i}$$

- Value of -1 indicates perfect disassortativity and 1 is perfect assortativity (each attribute value a separate component)
- e_{ii} = probability that an edge connects two nodes that *both* have attribute value i
- a_i = probability that an edge has a node with attribute value i as origin
- b_i = probability that an edge has a node with attribute value i as destination
- In undirected networks $a_i = b_i$

Computing attribute assortativity

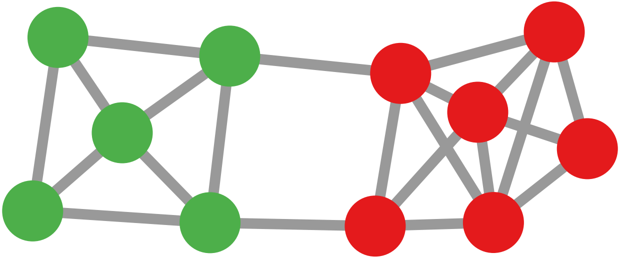


Figure: A rather assortative network

Computing attribute assortativity

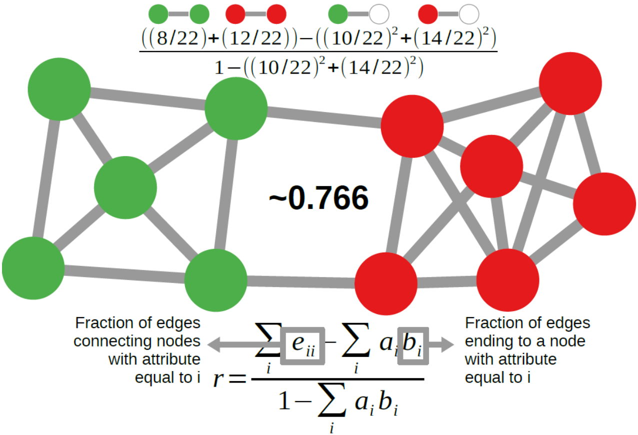


Figure: Computing r for that same network: $r \approx 0.766$.

A disassortative network

A disassortative network

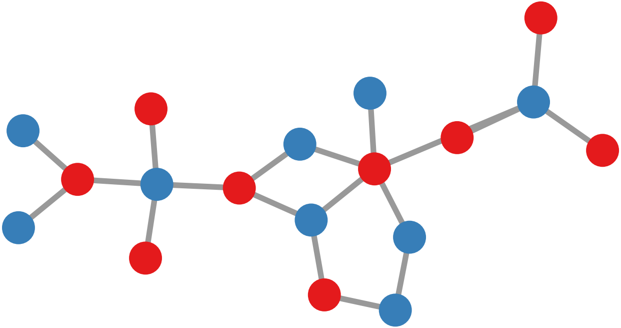


Figure: Dating network (red = female, blue = male)

An assortative network

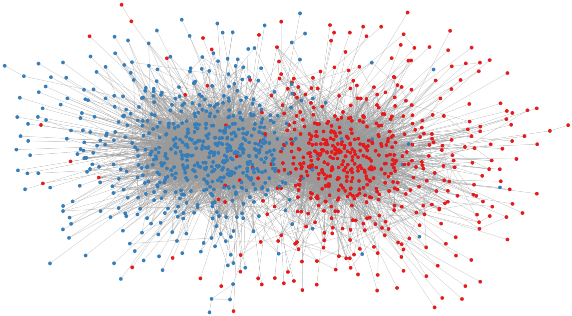


Figure: Links between political blogs (blue = democrat, red = republican)

An assortative network

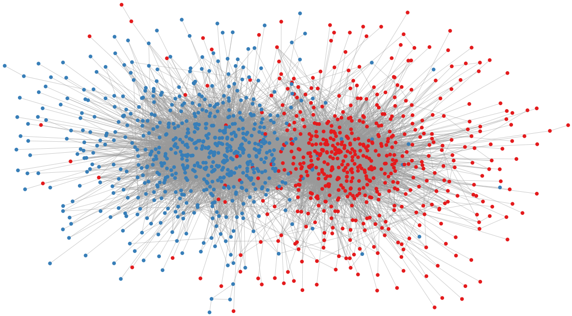


Figure: Links between political blogs (blue = democrat, red = republican)

- Assortativity fosters the formation of **echo chambers**.

An assortative network

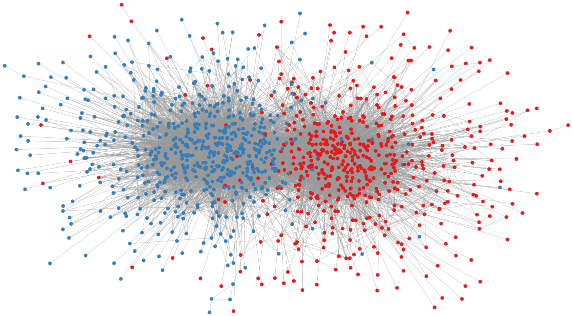


Figure: Links between political blogs (blue = democrat, red = republican)

- Assortativity fosters the formation of **echo chambers**.
- (Social) **segregation** can be measured through assortativity, we speak of **integration** if there is disassortative mixing

Degree assortativity

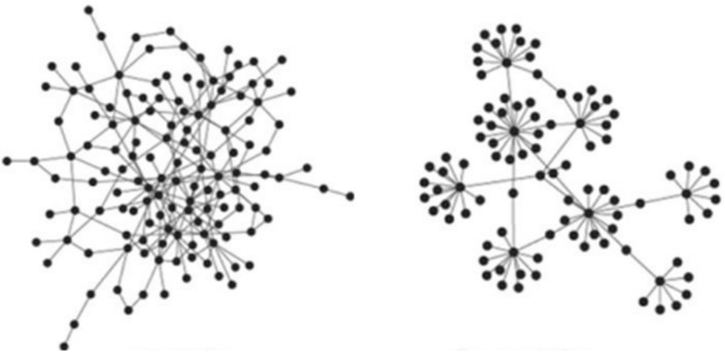


Figure: Degree **assortativity** (left) and degree **disassortativity** (right)

Image: Estrada et al., Clumpiness mixing in complex networks, J. Stat. Mech. Theor. Exp. P03008 (2008).

Visualizing degree assortativity

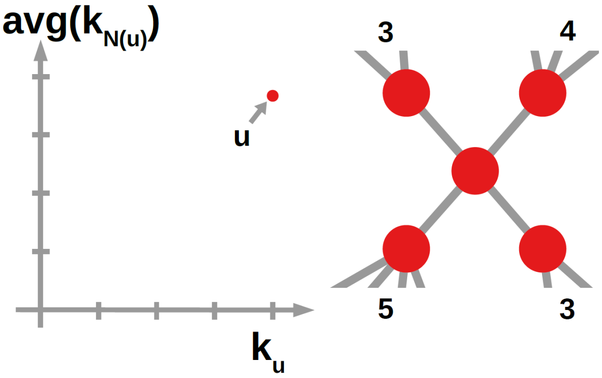


Figure: Assortativity plot: degree vs. average neighbor degree.

Visualizing degree assortativity

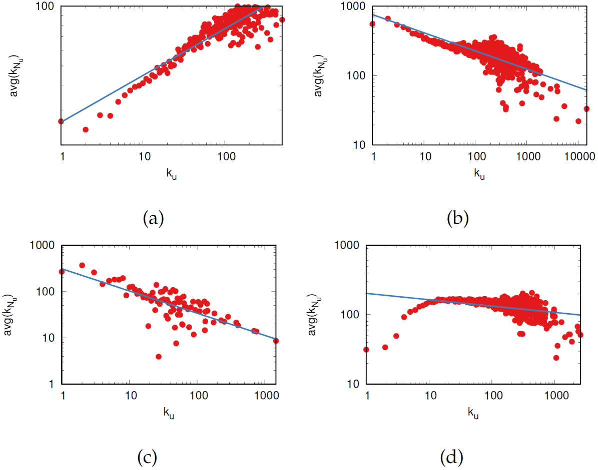


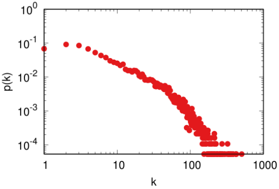
Figure: Assortativity plots of four real-world networks.

Friendship paradox

- Do someone's friends in a social network typically have fewer, an equal number, or more friends?

Friendship paradox

- Do someone's friends in a social network typically have fewer, an equal number, or more friends?



Friendship paradox

- Do someone's friends in a social network typically have fewer, an equal number, or more friends?

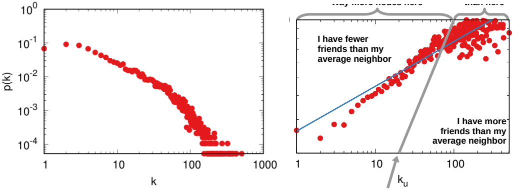


Figure: Degree distribution (left) and degree assortativity plot (right) of the same scientific co-authorship network.

Friendship paradox

- Do someone's friends in a social network typically have fewer, an equal number, or more friends?

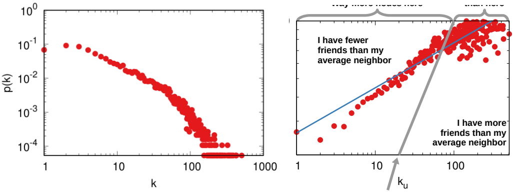


Figure: Degree distribution (left) and degree assortativity plot (right) of the same scientific co-authorship network.

- When we average over people's friends, nodes are not counted equally. A node of degree k appears in k different friend lists, so **high-degree nodes are counted more often**. Relative to a typical node, whose degree is $\langle k \rangle$, a degree-k node is weighted by $k/\langle k \rangle$.

Network projection

Bipartite graphs

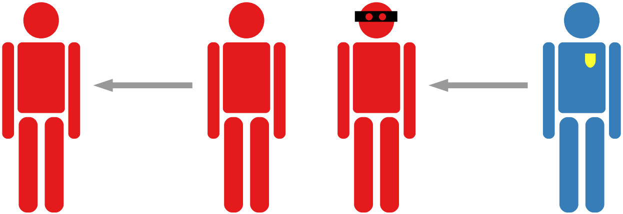


Figure: **Unipartite graph** (left) and **bipartite graph** (right) /
One-mode network (left) and two-mode network (right)

Bipartite and tripartite graphs

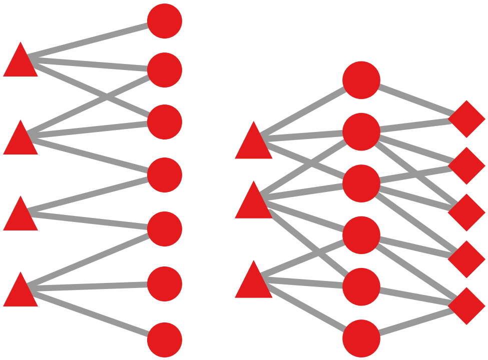
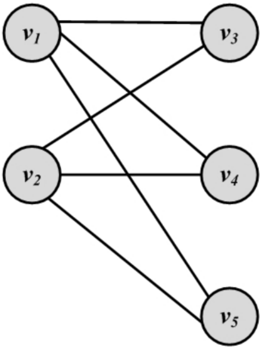


Figure: Bipartite graph (left) and tripartite graph (right)

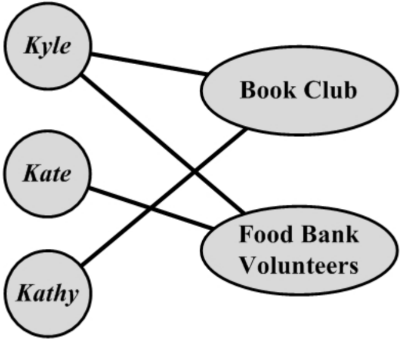
Bipartite graphs

- In **bipartite graphs** the set of nodes V can be split into two node sets V_L and V_R such that all edges in E have their endpoints in different node sets. Specifically:
  - $V = V_L \cup V_R$
  - $V_L \cap V_R = \emptyset$
  - $E \subseteq V_L \times V_R$
- Also called **two-mode networks** or heterogenic networks (as opposed to **one-mode networks** or homogenic networks)
- Called **affiliation networks** in a social network context
- So, two different types of nodes ...

Bipartite graphs



(a) Bipartite Graph



(b) Affiliation Network

Image: Zafarani et al., Social Media Mining, 2014.

Representing bipartite graphs

0	0	0	0	0	0	0	0	0	1
0	0	0	0	0	0	0	1	1	1
0	1	0	0	1	1	1	1	0	0
1	0	0	0	0	0	0	0	0	0
0	0	1	0	0	0	1	0	0	0
0	0	0	0	0	1	0	0	1	1

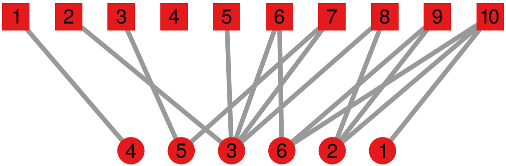


Figure: Bipartite adjacency matrix and corresponding graph

Projecting networks

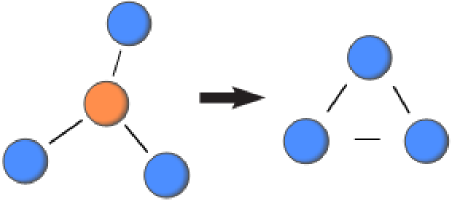
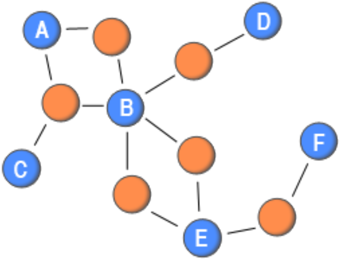


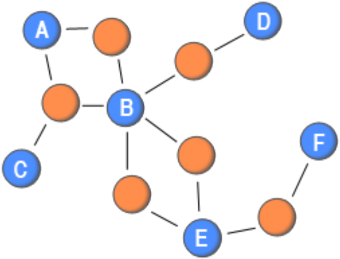
Image: http://toreopsahl.com/

Weighted projection



$\rightarrow$

Weighted projection



→

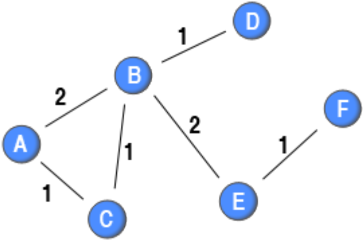
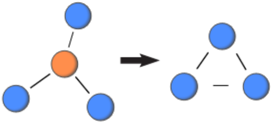


Image: `http://toreopsahl.com`

Projection algorithm



- Given a bipartite graph $G = (V_L \cup V_R, E)$ with $E \subseteq V_L \times V_R$, generate the projected graph $G' = (V_L, E')$
  - Initialize $G' = (V_L, E')$ with $E' = \emptyset$
  - For each node $v \in V_R$, determine its neighborhood $N(v) \subseteq V_L$
    - For each distinct node pair $v_i, v_j \in N(v)$, add the edge (v_i, v_j) to E'
    - Optionally, assign a weight to edge (v_i, v_j) based on how often it occurs
- Analogously, the projection from $G = (V_L \cup V_R, E)$ to $G'' = (V_R, E'')$ can be made

Network analysis on (almost) any dataset

- Different data objects typically have **attributes with identical values**
- The unique object identifier and the common attribute value are the two node types in a two-mode network representing the data
- The two-mode network can be converted into a one-mode network based on the common attribute
- Many projections of a dataset to a network are possible
- Consider a two-mode network in which actors and movies are connected if the actor plays in the movie . . .

Choosing the right representation

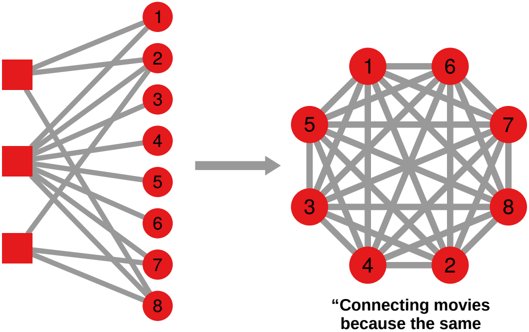


Figure: ...actor played in it. **Unweighted bipartite projection**.

Choosing the right representation

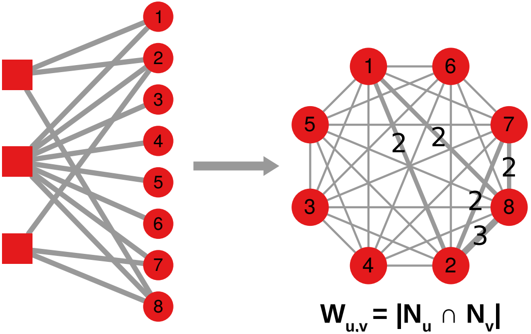
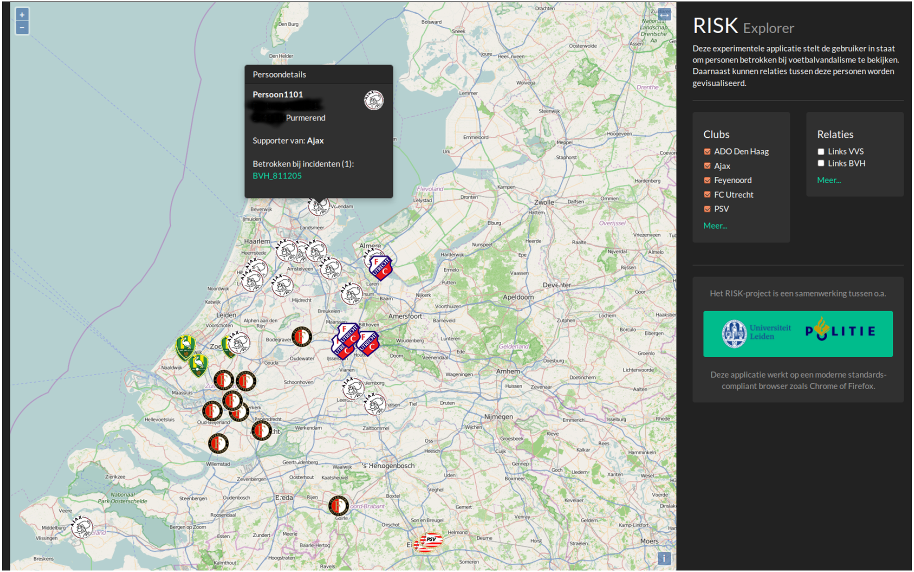


Figure: **Simple weighted bipartite projection**. No weight is weight 1.

Criminal networks

Example: Criminal networks

- Data science project with Dutch National Police
- Gain insight in social networks of soccer fans, group formation and organization
- Dataset: all entries in police systems of law violations of a particular group of people involved in soccer violence


Persoondetails
Persoon1101
[redacted]
[redacted] Purmerend
Supporter van: Ajax
Betrokken bij incidenten (1):
BVH_811205
RISK Explorer
Deze experimentele applicatie stelt de gebruiker in staat om personen betrokken bij voetbalvandalisme te bekijken. Daarnaast kunnen relaties tussen deze personen worden gevisualiseerd.
Clubs
ADO Den Haag
Ajax
Feyenoord
FC Utrecht
PSV
Meer...
Relaties
Links VVS
Links BVH
Meer...
Het RISK-project is een samenwerking tussen o.a.
Universiteit Leiden
POLITIE
Deze applicatie werkt op een moderne standards-compliant browser zoals Chrome of Firefox.

Criminal networks

Person ID	Incident ID	Incident Type
P000001	X00011	Straatroof/diefstal
P000001	X00014	Eenv. Mishandeling
P000002	X00011	Straatroof/diefstal
P000002	X00012	Eenv. Mishandeling
P000003	X00012	Eenv. Mishandeling
P000003	X00016	Bedreiging
P000004	X00012	Eenv. Mishandeling
P000004	X00017	Eenv. Mishandeling
P000005	X00013	Bedreiging
P000005	X00014	Eenv. Mishandeling
P000005	X00015	Straatroof/diefstal
P000006	X00013	Bedreiging
P000007	X00013	Bedreiging
P000008	X00013	Bedreiging
P000009	X00015	Straatroof/diefstal
P000010	X00016	Bedreiging
P000010	X00017	Eenv. Mishandeling
P000011	X00016	Bedreiging

Table: Data on suspects involved in incidents

Network formation

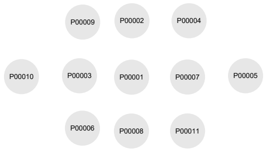
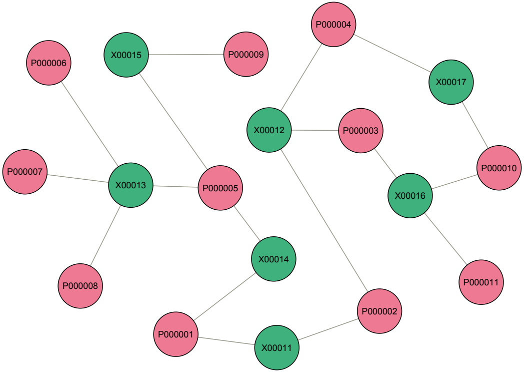


Figure: Suspects are nodes

Two-mode criminal network



Network formation

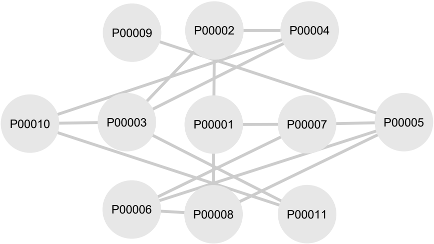


Figure: Edges are based on common involvement as a suspect

Network visualization

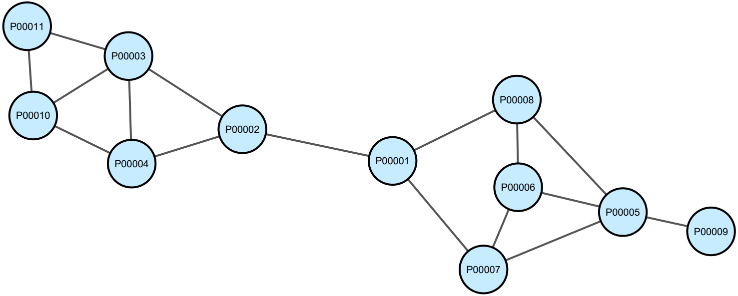


Figure: Force-directed visualization algorithm reveals structure

Network analysis: Centrality

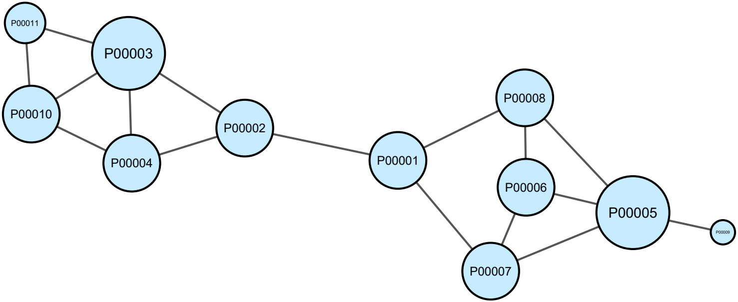


Figure: Degree centrality finds locally important nodes

Network analysis: Centrality

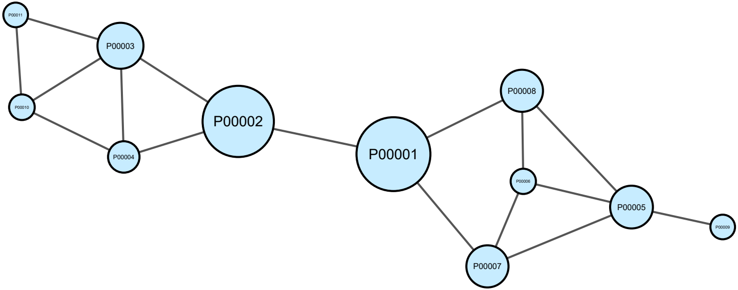


Figure: Betweenness centrality reveals globally important nodes

Network analysis: Community detection

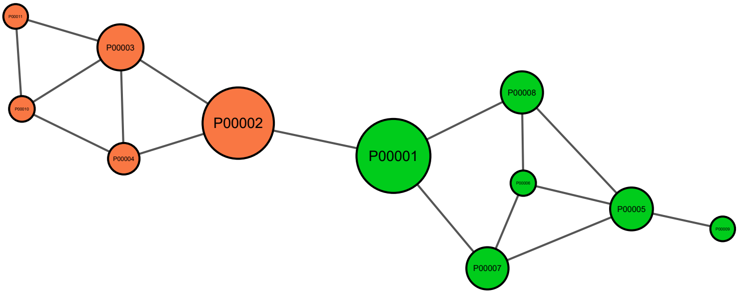
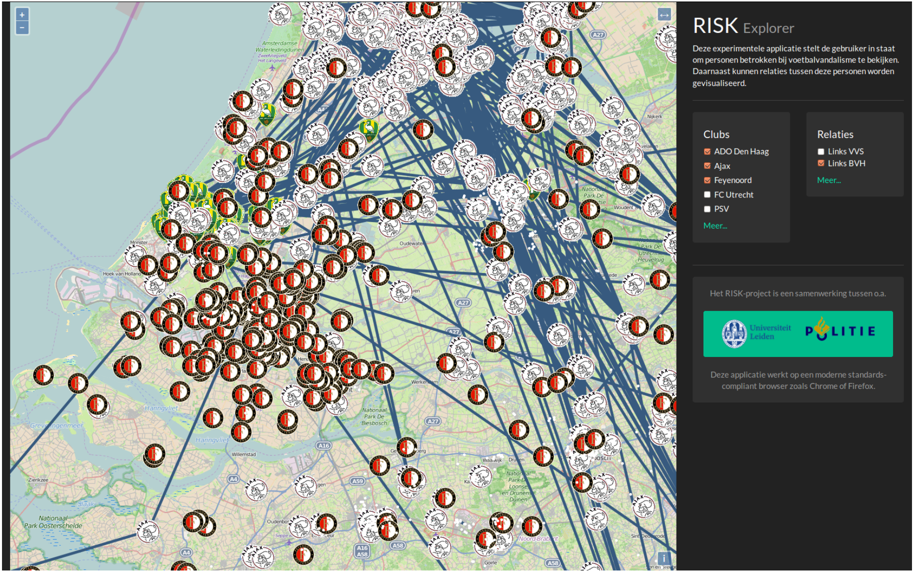
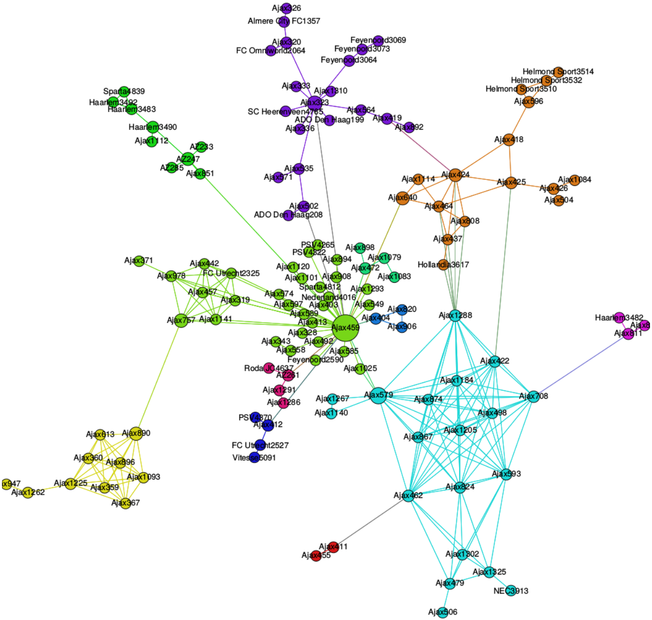
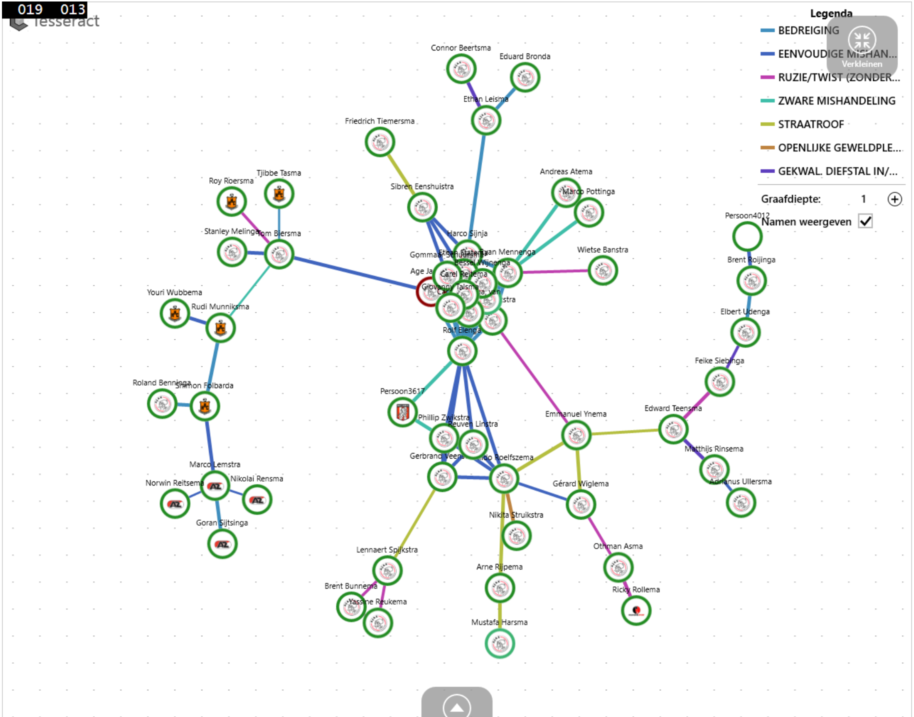


Figure: Community detection finds groups of tightly connected nodes


RISK Explorer
Deze experimentele applicatie stelt de gebruiker in staat om personen betrokken bij voetbalvandalisme te bekijken. Daarnaast kunnen relaties tussen deze personen worden gevisualiseerd.
Clubs
ADO Den Haag
Ajax
Feyenoord
FC Utrecht
PSV
Meer...
Relaties
Links VVS
Links BVH
Meer...
Het RISK-project is een samenwerking tussen o.a.
Universiteit Leiden
POLITIE
Deze applicatie werkt op een moderne standards-compliant browser zoals Chrome of Firefox.


Ajax326
Almere City FC1357
Ajax320
FC Omniworld2064
Feyenoord3069
Feyenoord3073
Feyenoord3064
Ajax333
Ajax1310
Ajax323
SC Heerenveen4765
ADO Den Haag199
Ajax564
Ajax419
Ajax892
Ajax336
Ajax535
Ajax571
Ajax502
ADO Den Haag208
Helmond Sport3514
Helmond Sport3532
Helmond Sport3510
Ajax596
Ajax418
Ajax424
Ajax1114
Ajax425
Ajax1084
Ajax426
Ajax504
Ajax640
Ajax464
Ajax808
Ajax437
Hollandia3617
Sparta4839
Haarlem3492
Haarlem3483
Haarlem3490
Ajax1112
AZ233
AZ247
AZ285
Ajax851
PSV4265
PSV4322
Ajax898
Ajax894
Ajax1079
Ajax1120
Ajax472
Ajax1101
Ajax908
Ajax1083
Sparta4812
Ajax1293
Ajax371
Ajax442
Ajax978
FC Utrecht2325
Ajax457
Ajax574
Nederland4016
Ajax319
Ajax597
Ajax403
Ajax549
Ajax589
Ajax820
Ajax757
Ajax1141
Ajax413
Ajax459
Ajax404
Ajax906
Ajax1288
Haarlem3482
Ajax811
Ajax328
Ajax343
Ajax492
Ajax558
Ajax585
Feyenoord2590
Ajax1025
Roda JC4637
AZ261
Ajax422
Ajax1184
Ajax1291
Ajax1267
Ajax579
Ajax874
Ajax708
Ajax1286
Ajax1140
Ajax498
PSV4870
Ajax412
Ajax1205
Ajax867
FC Utrecht2527
Vitesse5091
Ajax613
Ajax890
Ajax360
Ajax896
Ajax1093
Ajax1225
Ajax359
Ajax1262
Ajax367
Ajax593
Ajax824
Ajax462
Ajax411
Ajax455
Ajax1302
Ajax1325
Ajax479
NEC3913
Ajax506

019 013


Tesseract
Legenda
BEDREIGING
EENVOUDIGE MISHAN...
RUZIE/TWIST (ZONDER...
ZWARE MISHANDELING
STRAATROOF
OPENLIJKE GEWELDPLE...
GEKWAL. DIEFSTAL IN/...
Verkleinen
Graafdiepte: 1
Namen weergeven
Connor Beertsma
Eduard Bronda
Ethan Leisma
Friedrich Tiemersma
Andreas Atema
Marco Pottinga
Sibren Eenshuistra
Roy Roersma
Tjibbe Tasma
Harco Sijnja
Stanley Melinga
Tom Biersma
Bessel Wijnenga
Carel Reitema
Giovanny Talsma
Wietse Banstra
Persoon4012
Brent Roijinga
Youri Wubbema
Rudi Munniksma
Rolf Elenga
Elbert Udenga
Feike Siebinga
Roland Benninga
Persoon3617
Phillip Zwikstra
Reuven Linstra
Emmanuel Ynema
Edward Teensma
Gerbrand Veen
Matthijs Rinsema
Adrianus Ullersma
Marco Lemstra
Norwin Reitsema
Nikolai Rensma
Gérard Wiglema
Nikita Struikstra
Goran Sijtsinga
Othman Asma
Lennaert Spijkstra
Arne Rijpema
Ricky Rollema
Brent Bunnema
Yassine Reukema
Mustafa Harsma

Community detection

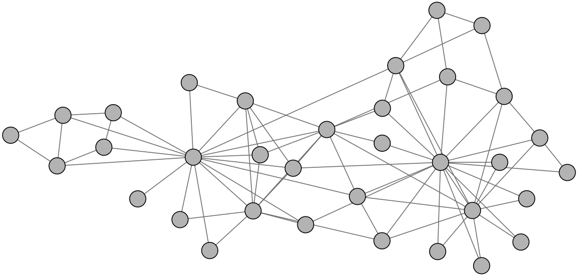


Figure: Communities: node subsets connected more strongly with each other.

Community detection

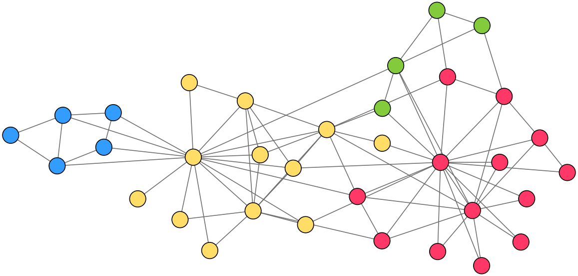


Figure: Communities: node subsets connected more strongly with each other.

Upcoming lab session and "homework"

- Today: practice with theoretical/non-programming questions in ANS
- Goal is to test the system; bear with us should things go wrong
- Recall that these practice sessions are not a graded component, but that they are useful practice for the exam
- Advance substantially with Assignment 1

Credits

- Several images are gratefully reused from
  Michele Coscia, The Atlas for the Aspiring Network Scientist, v3 of
  `https://arxiv.org/abs/2101.00863`