Business Intelligence & Process Modelling

Frank Takes

Universiteit Leiden

Lecture 6 — Network Analytics

Recap

- **Business Intelligence**: anything that aims at providing actionable information that can be used to support business decision making
  - Business Intelligence
  - Visual Analytics
  - Descriptive Analytics & Predictive Analytics
    - **Network Analytics**
- Process Modelling (April and May)

Overview

Universiteit Leiden

- Context
- Network science
- Real-world networks
- Centrality
- Case: criminal networks
- Community detection
- Case: corporate networks
- Network flow: offshore financial centers

Universiteit Leiden

Context

Context: Data

- **Data**: facts, measurements or text collected for reference or analysis (Oxford dictionary)
  - Unstructured data: data that does not fit a certain data structure (text, images, audio, video, a list of numeric measurements)
  - Structured data: data that fits a certain data structure (table, **graph/network**, tree, etc.)

Context: Data

- **Data**: facts, measurements or text collected for reference or analysis (Oxford dictionary)
  - Unstructured data: data that does not fit a certain data structure (text, images, audio, video, a list of numeric measurements)
  - Structured data: data that fits a certain data structure (table, **graph/network**, tree, etc.)
- **Data mining**: extracting information/knowledge from data

Data evolution

- Census data (60s)
- Transaction data (80s)
- Micro event data (00s)
- Social data (10s)

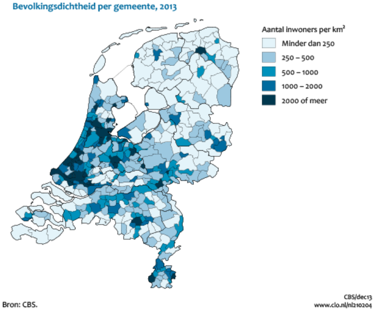
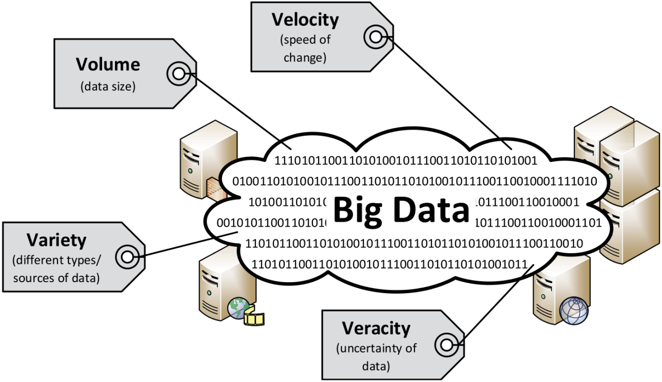
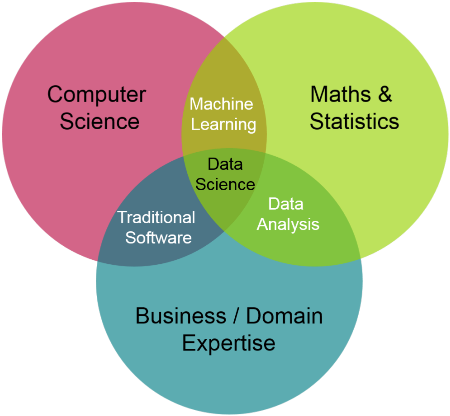


Figure : Census data

Context: Big data



Context: Data science



Source: Forbes, A Very Short History of Data Science, May 28, 2013.

Context: Social media

Social media mining

- Social media **platforms**: Facebook, Twitter, LinkedIn, Reddit, YouTube, Blogger, . . .
- Platforms generate enormous amounts of (un)structured data
- **Social Media Mining**: analyzing this data in order to get insight in user(s), trends, usage patterns, the platform itself, . . .
  - Sentiment Mining
  - Trend Analysis
  - **Social Network Analysis**

Network Science

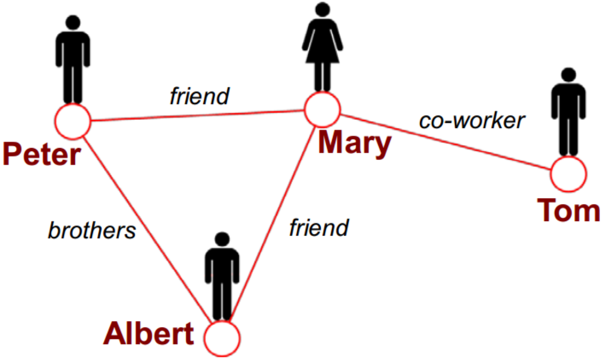
Networks

- **Objects**/entities/nodes/vertices
- **Relationships**/ties/links/edges
- **Network**/graph: objects and relationships between objects
- Data attributes are annotations on the nodes and the edges
- Examples:
  - Online social networks
  - Scientific citation and collaboration networks
  - Webgraphs
  - Biological networks
  - Communication networks
  - Financial networks
  - Corporate networks

Notation

Concept	**Symbol**
■ Network (graph)	$G = (V, E)$
■ Objects (nodes/vertices)	V
■ Relations (links/edges)	E
■ Directed — $E \subseteq V \times V$	
■ Undirected	
■ Number of nodes — $\lvert V \rvert$	n
■ Number of edges — $\lvert E \rvert$	m

Networks



Source: `http://web.stanford.edu/class/cs224w`

Types of graphs

- Directed vs undirected graphs
  - Reciprocity/Symmetry: extend to which directed links are mutual
- Weighted vs. unweighted graphs
  - Unweighted: weight of 1 for computational reasons
  - Signed networks: positive and negative weights
- Labeled (annotated) vs. unlabeled graphs
  - Labels on nodes and/or edges
- One-mode (homogenic) vs. two-mode networks
  Or: multi-mode (heterogenic) networks
- Static vs. dynamic (temporal) networks
  - Timestamps on nodes and/or edges

Representation

- (Undirected) **Edge List**
  1 3
  2 3
  2 6
  3 4
  3 5
  3 6
- Commonly used as an input format
- $O(2m)$ memory

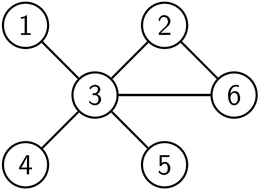


Figure : $n = 6$ and $m = 6$

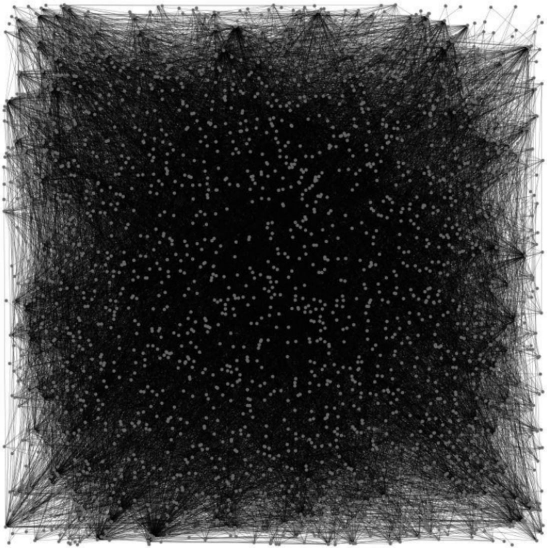
Network science

- **Network science**: understanding data by investigating interactions and relationships between individual data objects as a network
- **Networks** are the central model of computation

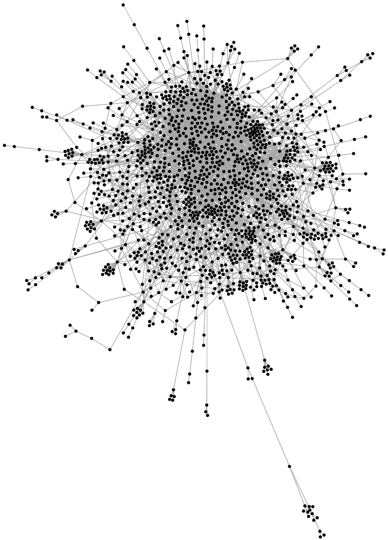
Network science

- **Network science**: understanding data by investigating interactions and relationships between individual data objects as a network
- **Networks** are the central model of computation
- Branch of data science focusing on network data
- Method in complexity research
- Complex systems approach: the behavior emerging from the network reveals patterns not visible when studying the individuals

Large network data

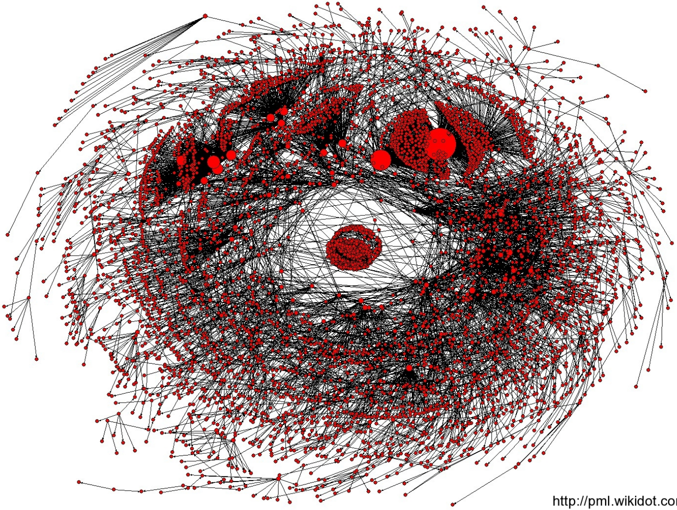


Visualization algorithms



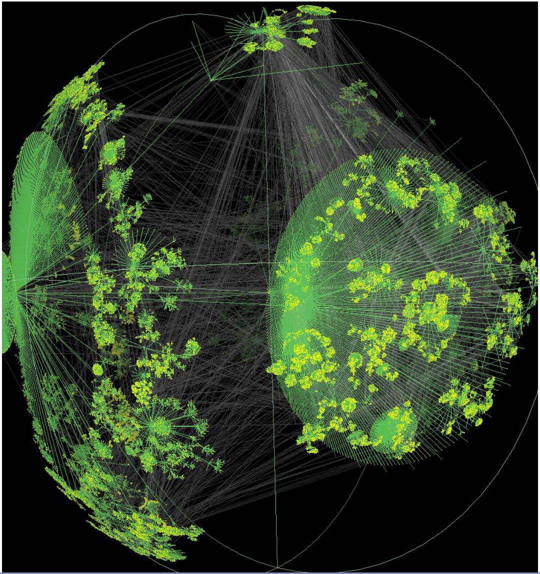
Protein interaction network

Collaboration network: 10,000 nodes

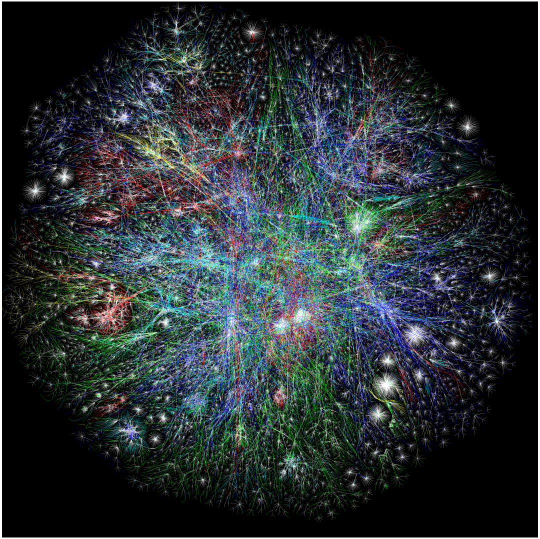


http://pml.wikidot.com

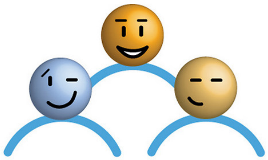
Webgraph: 500,000 nodes



Webgraph: 500,000 nodes



HYVES: 8,000,000 nodes

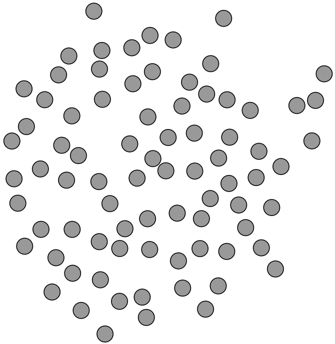


- Online Social Network
- Dutch & pre-Facebook
- Full snapshot
- $n = 8,000,000$ (8 million)
- $m = 1,000,000,000$ (1 billion)

Facebook: 1,000,000,000 nodes

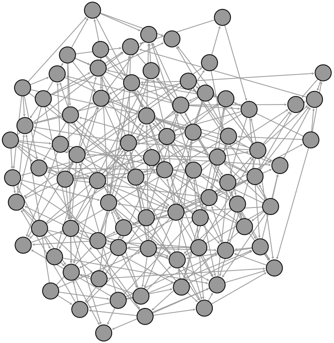
Graph properties

- We have seen:
  - From 6 to $1,000,000,000$ (1 billion) nodes
  - From 8 to $120,000,000,000$ (120 billion) edges
- Measuring only number of nodes and edges is too simple



Graph properties

- We have seen:
  - From 6 to $1,000,000,000$ (1 billion) nodes
  - From 8 to $120,000,000,000$ (120 billion) edges
- Measuring only number of nodes and edges is too simple



Graph properties

- We have seen:
  - From 6 to $1,000,000,000$ (1 billion) nodes
  - From 8 to $120,000,000,000$ (120 billion) edges
- Measuring only number of nodes and edges is too simple
- Other interesting measures:
  1. Density
  2. Degree
  3. Components
  4. Distance
  5. Clustering coefficient

Sparse networks

- Maximum number of edges m_{max}
  - $m_{\text{max}} = n(n-1)$ for directed graphs
  - $m_{\text{max}} = \frac{1}{2}n(n-1)$ for undirected graphs
- Sparse graph if $m \ll m_{\text{max}}$
- Measure sparseness using **density**: $\frac{m}{n(n-1)}$
- HYVES: $8 \cdot 10^6$ nodes, at most $64 \cdot 10^{12}$ edges.
  But network has "only" $1 \cdot 10^9$ edges, so density 0.0000156.
- Density is relevant when comparing networks

Degree

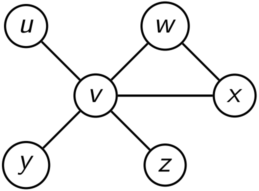


Figure : Undirected graph

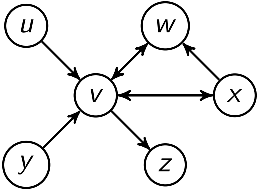


Figure : Directed graph

- Undirected graphs: degree $deg(v) = 5$

Degree

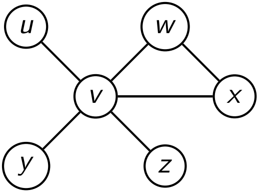


Figure : Undirected graph

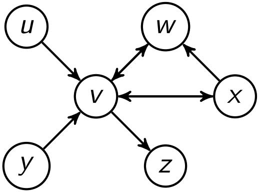


Figure : Directed graph

- Undirected graphs: degree $deg(v) = 5$
- Directed graphs
  - Indegree $indeg(v) = 4$
  - Outdegree $outdeg(v) = 3$

Degree

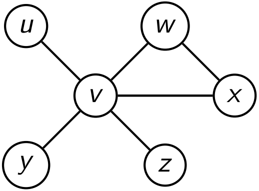


Figure : Undirected graph

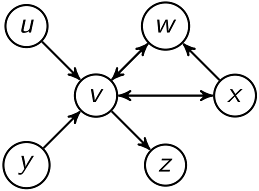
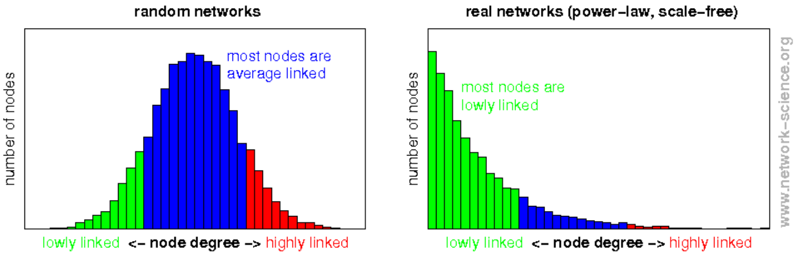


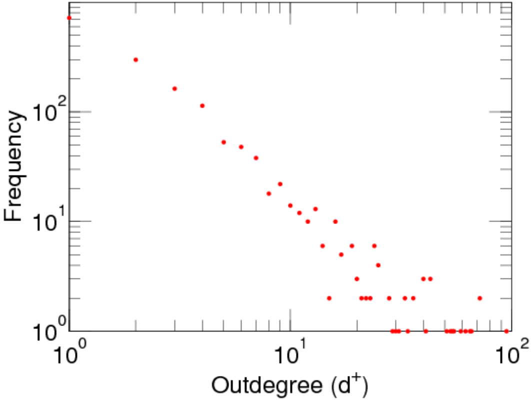
Figure : Directed graph

- Undirected graphs: degree $deg(v) = 5$
- Directed graphs
  - Indegree $indeg(v) = 4$
  - Outdegree $outdeg(v) = 3$
- **Degree distribution**: frequency of each degree value.
  Lognormal or power law distribution with "fat tail"

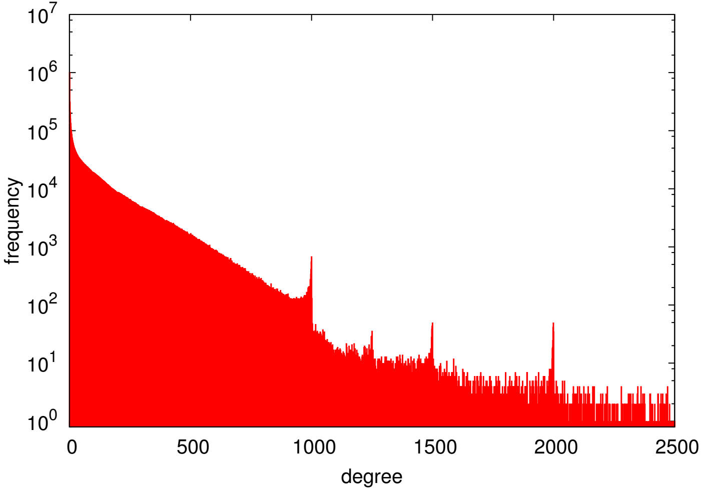
Degree distribution



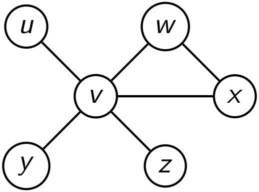
Degree distribution



HYVES degree distribution



Paths



Concept	**Example**
Path	$p = (u, v, z, v, w, x)$
Path length	$\|p\| - 1 = 5$
Simple path: no repeated vertices	$p' = (u, v, w, x)$
Shortest path: path of minimal length	$sp = (u, v, x)$
Distance: length of shortest path	$d(u, x) = \|sp\| - 1 = 2$

Connected components

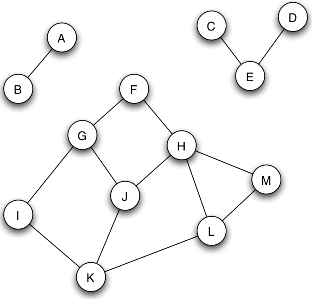
- What if $d(a, c) = \infty$? (so, no path between nodes a and c)

Connected components

- What if $d(a, c) = \infty$? (so, no path between nodes a and c)
- Undirected Graphs
  - **Connected component**: subset of nodes (maximal in size) in which each node can form a path to each other node in the set

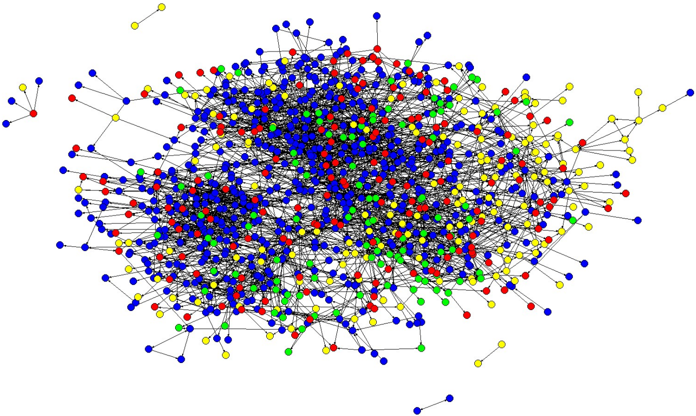
Connected components

- What if $d(a, c) = \infty$? (so, no path between nodes a and c)
- Undirected Graphs
  - **Connected component**: subset of nodes (maximal in size) in which each node can form a path to each other node in the set
- Directed Graphs
  - **Weakly** connected component: subgraph in which there is a path between any pair of nodes, ignoring link direction
  - **Strongly** connected component: subgraph in which there is a directed path between any pair of nodes



Connected components
Image source: D. Easley and J. Kleinberg, "Networks, Crowds, and Markets", 2010

Giant component



Small world experiment

- Stanley Milgram
- Picked 96 random people in Omaha
- Ask them to get a letter to a stock-broker in Boston by passing it through to closer acquaintance.
- How many steps did it take?

Small world experiment

- Stanley Milgram
- Picked 96 random people in Omaha
- Ask them to get a letter to a stock-broker in Boston by passing it through to closer acquaintance.
- How many steps did it take?
- Letters arrived after on average 5.9 steps
- Total of 18 chains completed

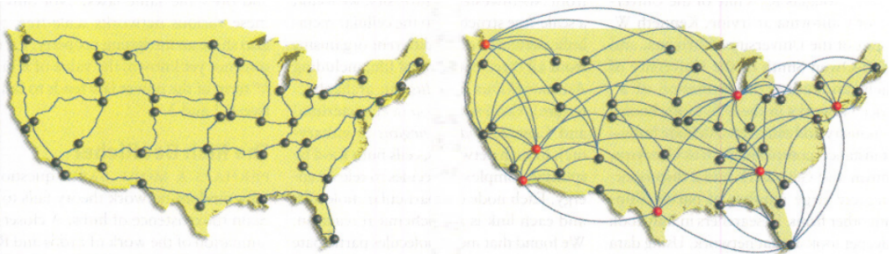
J. Travers and S. Milgram, "An Experimental Study of the Small World Problem", Sociometry 32(4): 425-443, 1969

Yahoo small world experiment

Universiteit Leiden

Small world

- Dense **core** containing many hubs
- **Periphery** with many nodes with a small distance to the core



Distance

- Average distance $\overline{d} = \frac{1}{n(n-1)} \sum_{v,w \in V} d(v,w)$
- Distribution with a "fat tail"
- **Diameter**: maximal distance value

Dataset	Nodes	Links	Average degree	Average distance
AstroPhys	17,903	396K	21	4.15
Enron	33,696	362K	10	4.07
Web	855,802	8.64M	10	6.30
YouTube	1,134,890	5.98M	5.3	5.32
Flickr	1,624,992	30.9M	18	5.38
Skitter	1,696,415	22.2M	13	5.08
Wikipedia	2,213,236	23.5M	11	4.81
Orkut	3,072,441	234M	76	4.16
LiveJournal	5,189,809	97.4M	19	5.48
Hyves	8,057,981	871M	112	4.75

F.W. Takes and W.A. Kosters, Determining the Diameter of Small World Networks, In CIKM, pp. 1191-1196, 2011.

Erdös number

- Scientific collaboration network
- Edges between scientists who wrote a paper together
- Erdös number: the distance of a scientist (node) to Erdös
- `http://www.ams.org/mathscinet/collaborationDistance.html`

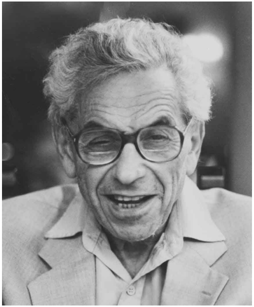
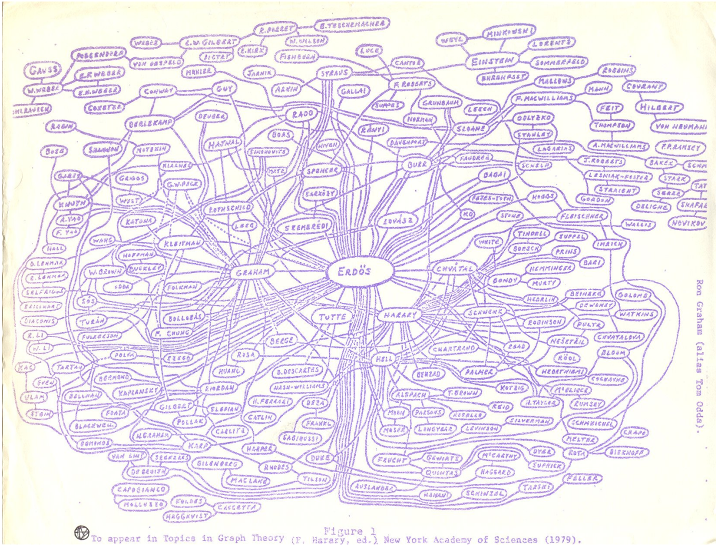
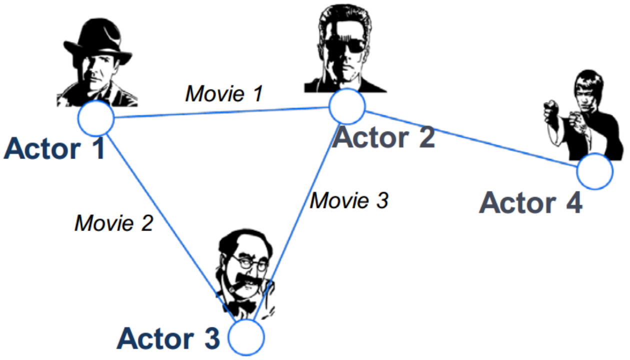


Figure : Paul Erdös (1913-1996)

Erdös


ERDÖS
GRAHAM
HARARY
TUTTE
CHVÁTAL
HELL
EINSTEIN
MINKOWSKI
LORENTZ
SOMMERFELD
GAUSS
W. WEBER
CONWAY
BERLEKAMP
RADO
STRAUS
SZEKERES
SZEMERÉDI
LOVÁSZ
ROTHSCHILD
KLEITMAN
KNUTH
HAJNAL
BOLLOBÁS
F. CHUNG
BERGE
TURÁN
F. ROBERTS
SLOANE
ODLYZKO
HILBERT
VON NEUMANN
F.P. RAMSEY
COURANT
ULAM
STEIN
KAC
KARP
DUKE
MAC LANE
ROTA
SPENCER
SÓS
POLLAK
SHANNON
BLOOM
WATKINS
NEŠETŘIL
Ron Graham (alias Tom Odda).
Figure 1
To appear in Topics in Graph Theory (F. Harary, ed.) New York Academy of Sciences (1979).

Movie actor network



Source: `http://web.stanford.edu/class/cs224w`

Six degrees of Kevin Bacon

- Collaboration network based on co-starring actors
- Variant of "Six degrees of Separation"
- Edges between actors indicate they played in a movie together

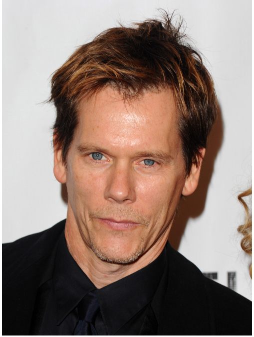


Figure : Kevin Bacon (1958)

The Wiki Game

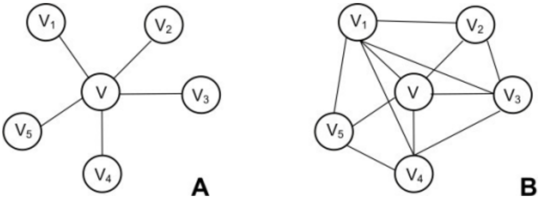
Triangles

Network	Nodes	Edges	Expected	Real	Difference
Facebook (WOSN)	63,731	817,035	2,809	3,500,542	1,246×
Epinions	75,879	508,837	402	162,448	404×
Amazon (TWEB)	403,394	3,387,388	789	398,6507	5,049×
Baidu	415,641	3,284,387	658	14,287,651	21,718×
Youtube links	1,138,499	4,942,297	109	3,049,419	27,957×
Flickr	2,302,925	33,140,017	3,973	837,605,842	210,806×
LiveJournal links	5,204,176	49,174,464	1,125	310,876,909	276,367×
Twitter (MPI)	52,579,682	1,963,263,821	69,410	55,428,217,664	798,565×

Table : Expected vs. real triangle counts in real-world networks.

Expected triangles: $(8m^3/n^6)(n^3/6) = \frac{4}{3}(m/n)^3$

Node clustering coefficient



- Situation A: v has a clustering coefficient of 0
- Situation B: v has a clustering coefficient of $\frac{14}{20} = \frac{7}{10} = 0.7$

$$C(v) = \frac{2 \cdot \text{edges between neighbors of } v}{\text{maximum number of such edges}}$$

Image: G.A. Pavlopoulos et al., "Using graph theory to analyze biological networks", in BioData Mining 4(1), 2011.

Small 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

Small 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 real-world networks: communication networks, citation networks, collaboration networks (Erdös, Kevin Bacon), protein interaction networks, information networks (Wikipedia), webgraphs, financial networks (Bitcoin) ...

Other topics

- Virality, contagion and outbreak detection
- Diffusion of information and innovation
- Link prediction
- Visualization Algorithms
- Privacy, Anonymity and Ethics
- **Centrality**
- **Community Detection**

Centrality

Centrality

- Given a social network, which person is most important?
- What is the most important page on the web?
- Which protein is most vital in a biological network?
- Who is the most respected author in a scientific citation network?
- What is the most crucial router in an internet topology network?

Centrality

- **Node centrality**: the importance of a node with respect to the other nodes based on the structure of the network
- **Centrality measure**: computes the centrality value of all nodes in the graph
- For all $v \in V$ a measure M returns a value $C_M(v) \in [0;1]$
- $C_M(v) > C_M(w)$ means that node v is more important than w


GAME OF THRONES
Watermark

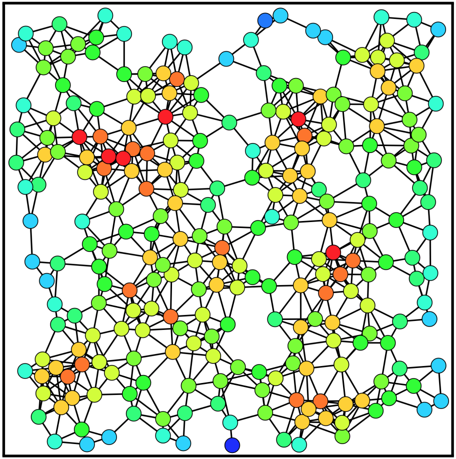
Degree centrality

- Undirected graphs – **degree centrality**: measure the number of adjacent nodes

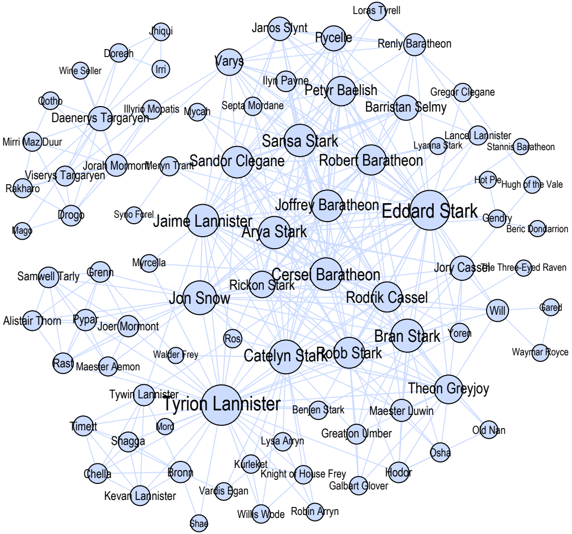
$$C_d(v) = \frac{deg(v)}{n-1}$$

- Directed graphs — indegree centrality and outdegree centrality
- Local measure
- $O(1)$ time to compute

Degree centrality



Degree centrality



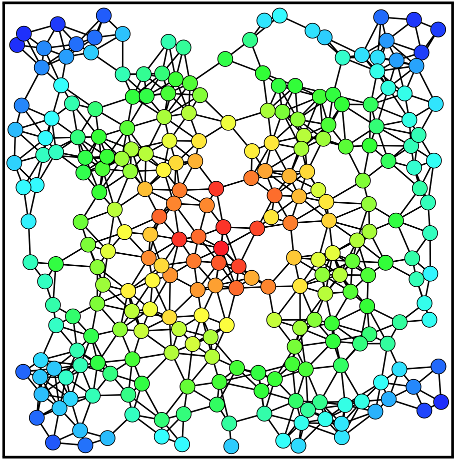
Closeness centrality

- **Closeness centrality**: the average distance to each other node in the network

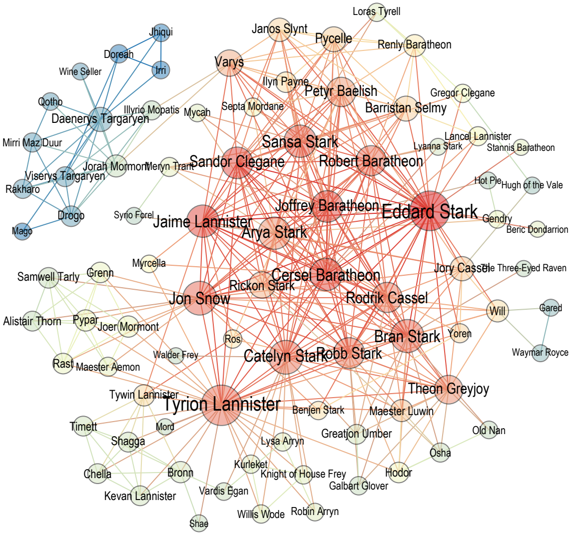
$$C_c(v) = \left(\frac{1}{n-1} \sum_{w \in V} d(v,w) \right)^{-1}$$

- Global distance-based measure
- Connected component(s)...
- $O(mn)$ to compute: one BFS in $O(m)$ for each of the n nodes

Closeness centrality



Degree vs. closeness centrality



Betweenness centrality

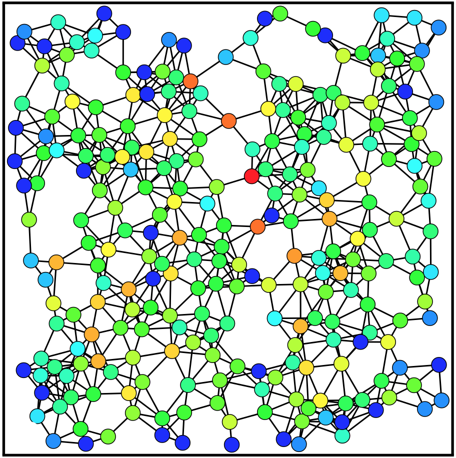
- **Betweenness centrality**: measure the number of shortest paths that run through a node; i.e., its involvement as a broker

$$C_b(u) = \sum_{\substack{v,w \in V \\ v \neq w, u \neq v, u \neq w}} \frac{\sigma_u(v,w)}{\sigma(v,w)}$$

- $\sigma(v,w)$ is the number of shortest paths from v to w
- $\sigma_u(v,w)$ is the number of such shortest paths that run through u
- Divide by largest value to normalize to $[0;1]$
- Global path-based measure
- $O(2mn)$ time to compute (two "BFSes" for each node)

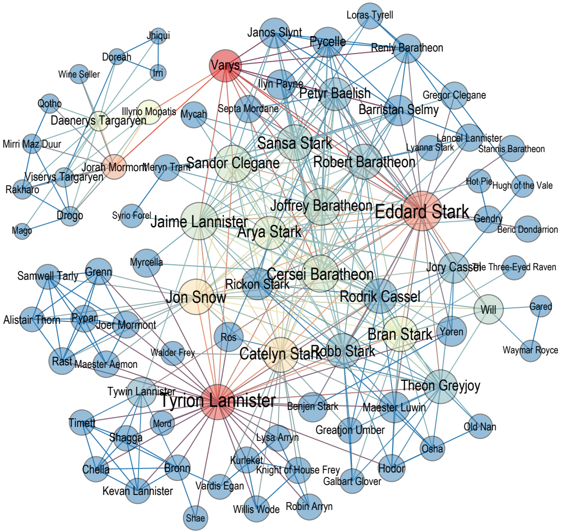
U. Brandes, "A faster algorithm for betweenness centrality", Journal of Mathematical Sociology 25(2): 163–177, 2001

Betweenness centrality



Universiteit Leiden

Degree vs. betweeness centrality



Centrality measures compared

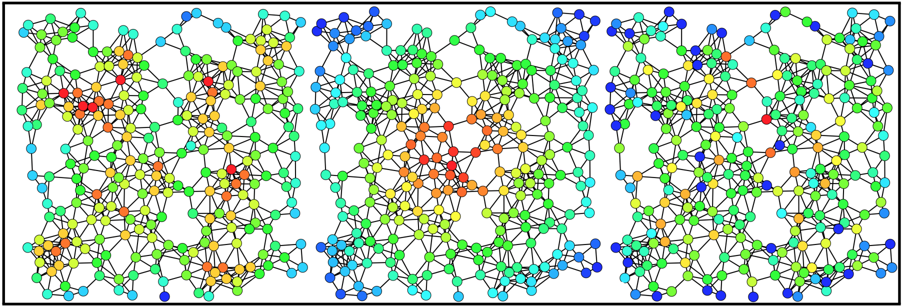


Figure : Degree, closeness and betweenness centrality

Source: "Centrality"' by Claudio Rocchini, Wikipedia `File:Centrality.svg`

Centrality measures compared

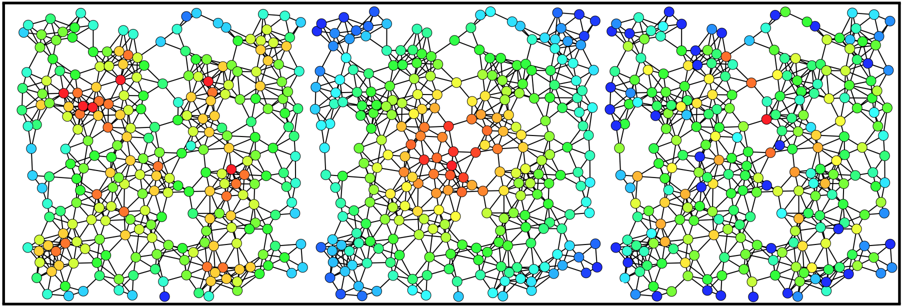
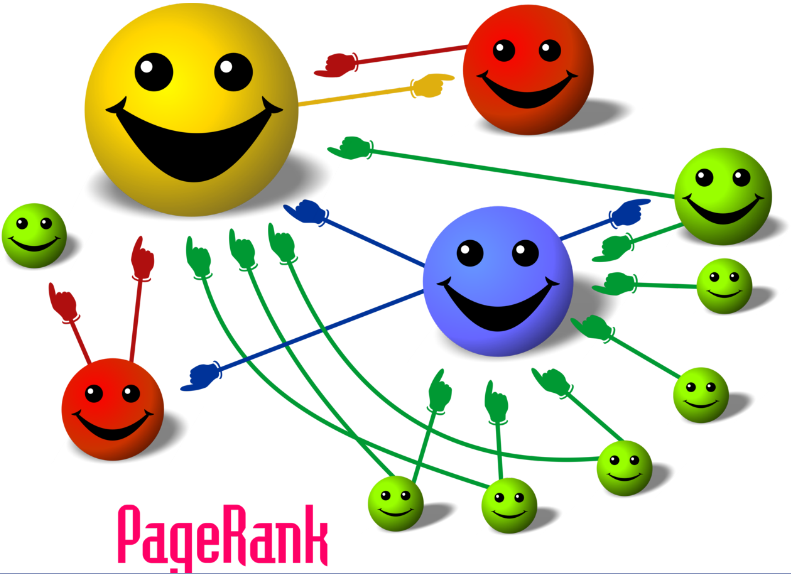


Figure : Degree, closeness and betweenness centrality

Source: "Centrality"' by Claudio Rocchini, Wikipedia `File:Centrality.svg`

Centrality measures (3)



Periodic table of centrality

Universiteit Leiden

Periodic Table of Network Centrality

	1 IA	2 IIA	3 IIIA	4 IVB	5 VB	6 VIB	7 VIIB	8 VIIIB	9 VIIIB	10 VIIIB	11 IB	12 IIB	13 IIIA	14 IVA	15 VA	16 VIA	17 VIIA	18 VIIIA
1	8000 1979 **DC** Degree																	518 1989 **IC** Information C
2	224 1971 **BC** Betweenness	239 2008 **EBC** Endpoint BC											26 1989 **kPC** kPath C.	275 2002 **EGO** Ego	51 2004 **HYPER** Hypergraphs	279 1997 **AFF** Affiliation C.	399 2 001 **α-C** α-Cent.	178 1995 **ECC** Eccentricity
3	942 1966 **CC** Closeness	239 2008 **PBC** Proxy BC											9068 1999 **HITS** Hubs/Authority	573 2006 **g-kPC** geodesic kPath	296 1999 **GROUP** Groups/Classes	80 2006 **HYPSC** Hyperg. SC	34 2010 **t-SC** t-Subgraph	116 1998 **RAD** Radiality
4	1279 1972 **EC** Eigenvector	239 2008 **LSBC** LscaledBC	224 1971 **EBC** Edge BC	53 2009 **CBC** Commun. BC	236 2007 **ΔC** Delta Cent.	5 2010 **MDC** MD Cent.	0 2015 **EYC** Entropy C.	2 2013 **CAC** Comm. Ability	56 2007 **EPTC** Entropy PC	281 1971 **CCoef** Clust. Coef.	42 2012 **PeC** PeC	427 2007 **BN** Bottleneck	43 2009 **EI** Essentiality I.	573 2006 **e-kPC** e-disjoint kPC	573 2006 **v-kPC** v-disjoint kPC	505 2010 **WEIGHT** Weighted C.	17 2013 **TCom** Total Comm.	116 1998 **INT** Integration
5	1306 1953 **KS** Katz Status	239 2008 **DBBC** DBounded BC	979 2005 **RWBC** RWalk BC	477 1991 **TEC** Total Effects	42 2009 **LI** Lobby Index	11 2008 **MC** Mod Cent.	0 2014 **COMCC** Community C.	45 2012 **ECCoef** ECCoef	0 2015 **SMD** Super Mediat.	1 2014 **UCC** United Comp.	4 2012 **WDC** WDC	119 2008 **MNC** MNC	43 2009 **KL** Clique Level	179 2005 **BIP** Bipartivity	426 1988 **GPI** GPI Power	116 1991 **kRPC** Reachability	58 2007 **SCodd** odd Subgraph	586 2004 **RWCC** RWalk CC
6	8053 1999 **PR** Page Rank	239 2008 **DSBC** DScaled BC	291 1953 **σ** Stress	477 1991 **IEC** Immediate Eff.	1 2014 **DM** Degree Mass	10 2012 **LAPC** Laplacian C.	0 2012 **ABC** Attentive BC	1699 2001 **STRC** Straightness C	0 2015 **SNR** Silent Node R.	15 2011 **HPC** Harm. Prot.	26 2011 **LAC** Local Average	119 2008 **DMNC** DMNC	3 2013 **LR** Lurker Rank	2457 1987 **β-C** β Cent.	X X **HYP** Hyperbolic C.	27 2012 **kEPC** k-edge PC	13 2007 **FC** Functional C.	0 2014 **HCC** Hierar. CC
7	484 2005 **SC** Subgraph	613 1991 **FBC** Flow BC	14 2012 **RLBC** RLimited BC	477 1991 **MEC** Mediative Eff.	69 2010 **LEVC** Leverage Cent.	35 2010 **TC** Topological C.	X X **SDC** Sphere Degree	15 2010 **ZC** Zonal Cent.	14 2013 **CI** Collab. Index	11 2013 **CoEWC** CoEWC	45 2012 **NC** NC	108 2010 **MLC** Moduland C.	X X **RSC** Resolvent SC	1 2014 **SWIPD** SWIPD	36 2009 **XXXX** LinComb	0 2014 **BCPR** BCPR	0 2014 **TPC** Tunable PC	0 2015 **EDCC** Effective Dist.

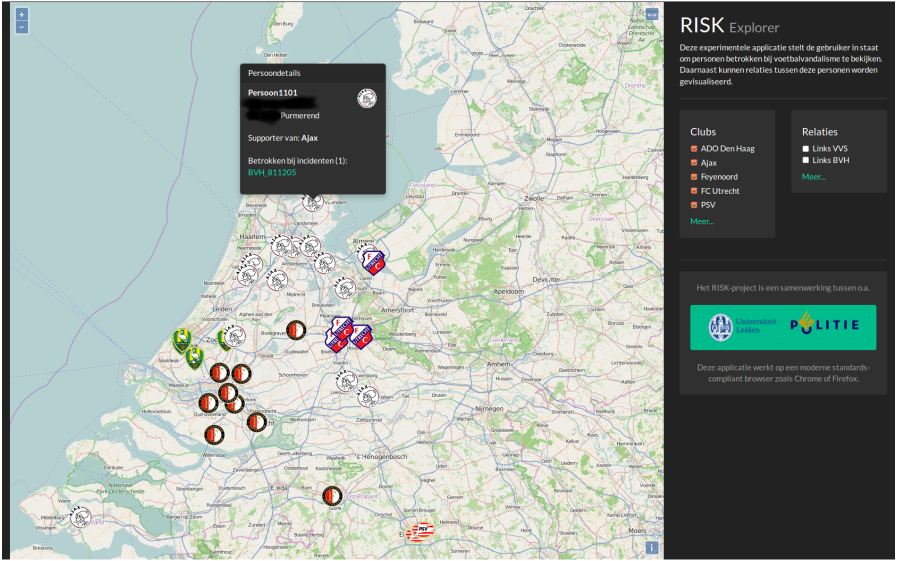
citations year **C** Name

8000 1979 **Freeman** Conceptual	942 1966 **Sabidussi** Axiomatic	573 2006 **Borgatti/Everett** Conceptual	1130 2005 **Borgatti** Conceptual	24 2014 **Boldi/Vigna** Axiomatic	252 1974 **Nieminen** Axiomatic	6 1981 **Kishi** Axiomatic	3 2012 **Kitti** Axiomatic	3 2009 **Garg** Axiomatic
2065 1934 **Moreno** Historic	1546 1950 **Bavelas** Historic	780 1948 **Bavelas** Historic	1475 1951 **Leavitt** Historic	297 1992 **Borgatti/Everett** Conceptual	3649 2001 **Jeong et al.** Empirical	4167 1998 **Tsai/Ghoshal** Empirical	961 1993 **Ibarra** Empirical	71 2008 **Valente** Empirical

- "Traditional"
- Betweenness-like
- Friedkin Measures
- Miscellaneous
- Path-based
- Specific Network Type
- Spectral-based
- Closeness-like

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] 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

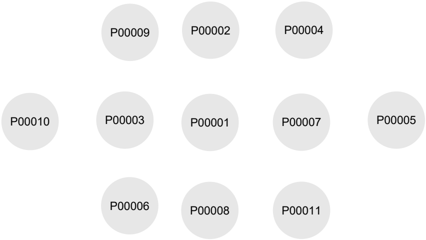


Figure : Suspects are nodes

Network formation

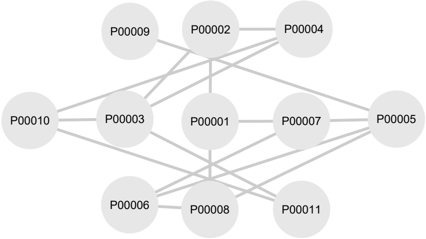


Figure : Edges are based on common involvement as a suspect

Network visualization

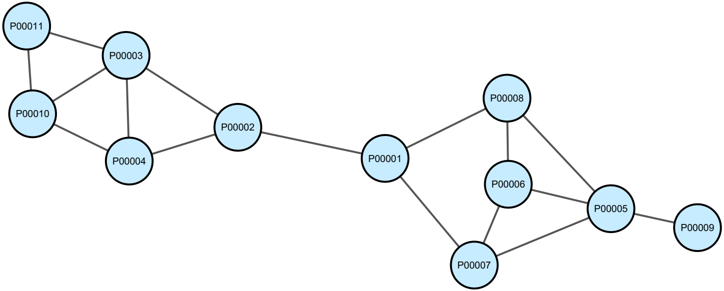


Figure : Force-directed visualization algorithm reveals structure

Network analysis: Centrality

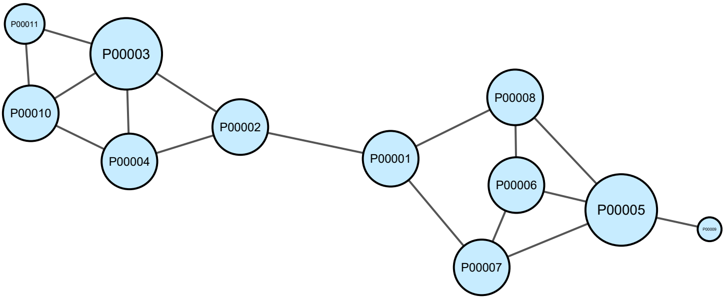


Figure : Degree centrality finds locally important nodes

Network analysis: Centrality

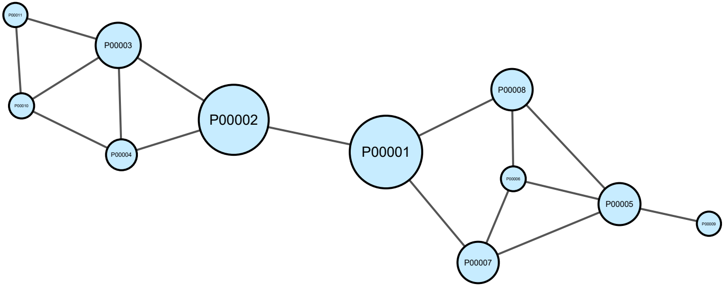


Figure : Betweenness centrality reveals globally important nodes

Network analysis: Community detection

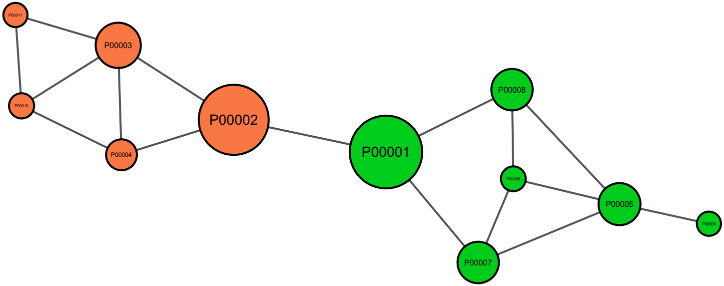
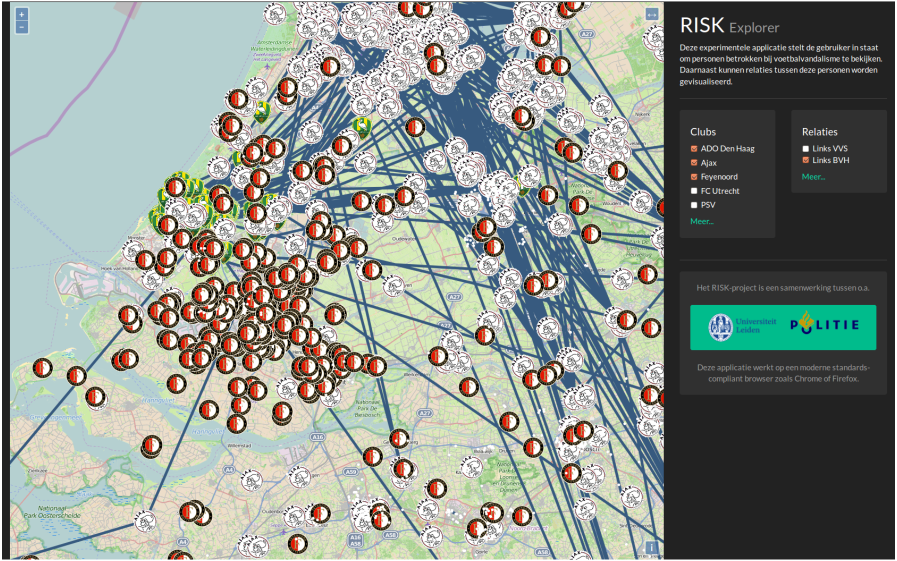
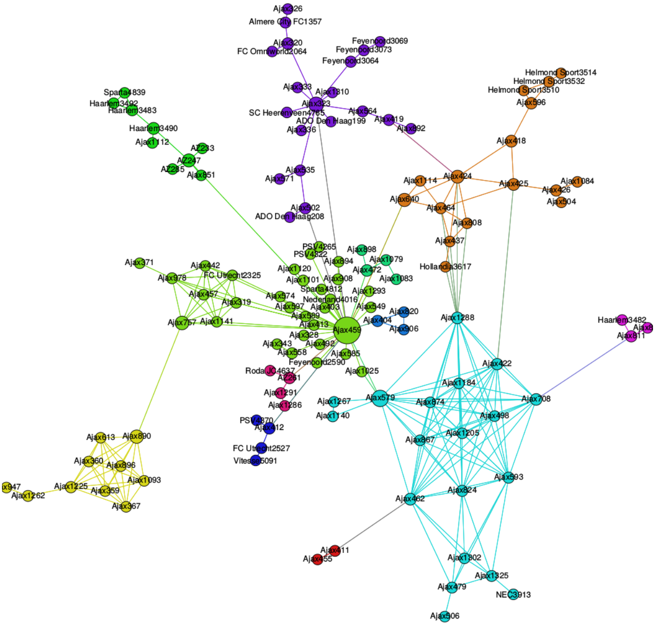


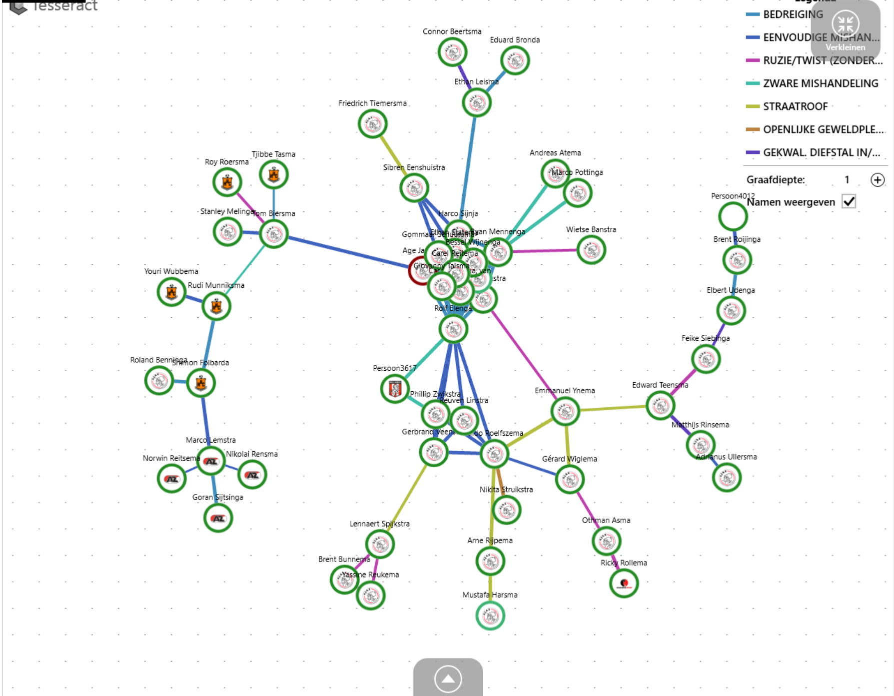
Figure : Community detection finds groups of tightly connected nodes


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 Heerenveen4785
Ajax564
ADO Den Haag199
Ajax419
Ajax336
Ajax892
Sparta4839
Haarlem3492
Haarlem3483
Haarlem3490
Ajax1112
AZ233
AZ247
AZ285
Ajax851
Ajax535
Ajax571
Ajax502
ADO Den Haag208
Helmond Sport3514
Helmond Sport3532
Helmond Sport3510
Ajax596
Ajax418
Ajax424
Ajax1114
Ajax425
Ajax1084
Ajax426
Ajax640
Ajax464
Ajax504
Ajax808
Ajax437
Hollandia3617
PSV4265
PSV4322
Ajax898
Ajax1079
Ajax894
Ajax472
Ajax1120
Ajax908
Ajax1083
Ajax1101
Sparta4812
Ajax1293
Ajax371
Ajax442
FC Utrecht2325
Ajax978
Ajax457
Ajax319
Ajax574
Nederland4016
Ajax597
Ajax403
Ajax549
Ajax820
Ajax589
Ajax404
Ajax757
Ajax1141
Ajax413
Ajax459
Ajax906
Ajax1288
Haarlem3482
Ajax8
Ajax811
Ajax328
Ajax343
Ajax492
Ajax558
Ajax585
Feyenoord2590
Ajax1025
Ajax422
Roda JC4637
AZ261
Ajax1184
Ajax1291
Ajax579
Ajax708
Ajax1267
Ajax874
Ajax1286
Ajax498
Ajax1140
PSV4870
Ajax412
Ajax1205
Ajax867
FC Utrecht2527
Vitesse5091
Ajax613
Ajax890
Ajax360
Ajax896
Ajax1093
Ajax593
Ajax1225
Ajax359
Ajax1262
Ajax367
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
Tjibbe Tasma
Roy Roersma
Sibren Eenshuistra
Harco Sijnja
Stanley Melinga
Tom Biersma
Bessel Wijnenga
Carel Reitema
Giovanny Talsma
Wietse Banstra
Persoon4012
Brent Roijinga
Elbert Udenga
Youri Wubbema
Rudi Munniksma
Rolf Elenga
Feike Siebinga
Roland Benninga
Simon Folbarda
Persoon3617
Phillip Zwikstra
Reuven Linstra
Emmanuel Ynema
Edward Teensma
Matthijs Rinsema
Adrianus Ullersma
Gerbrand Veens
Marco Lemstra
Nikolai Rensma
Norwin Reitsema
Gérard Wiglema
Goran Sijtsinga
Nikita Struikstra
Othman Asma
Lennaert Spijkstra
Arne Rijpema
Ricky Rollema
Brent Bunnema
Yassine Reukema
Mustafa Harsma

Community detection

Community detection

- **Community**: set of nodes connected more strongly with eachother than with the rest of the network
- Community detection algorithms:
  - Clique-based methods
  - Hierarchical clustering
  - Divisive algorithms (centrality-based)
  - **Modularity maximization** algorithms

Community detection

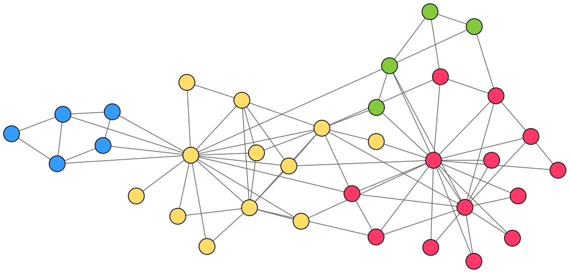


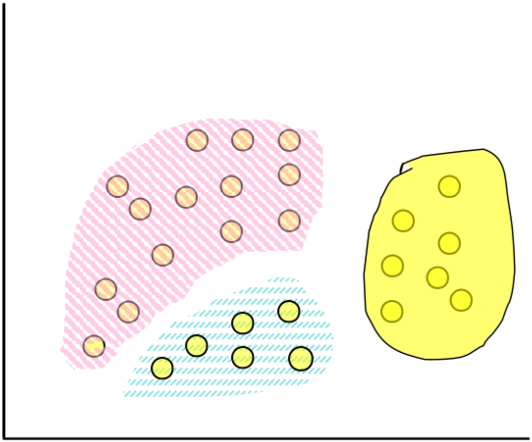
Figure : Communities: node subsets connected more strongly with each other

Modularity

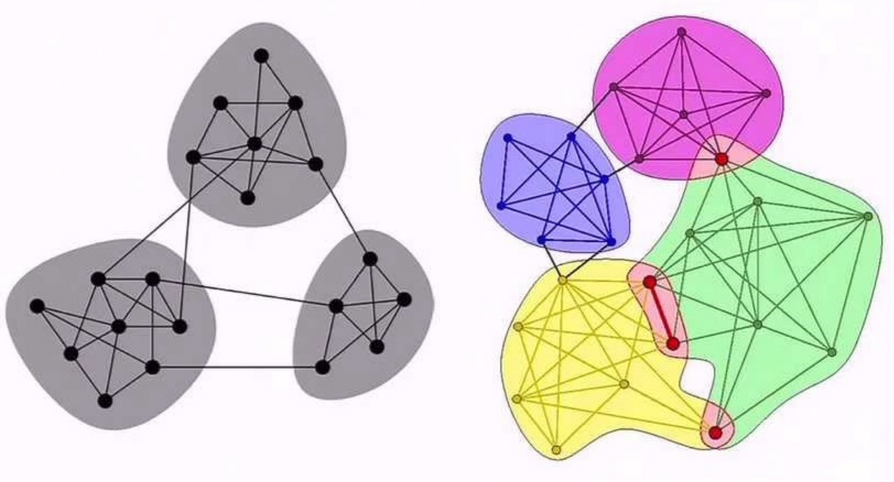
- **Modularity**: numerical value indicating the quality of a division of a network into communities
- **Community**: subset of nodes for which the fraction of links inside the community is higher than expected in a random network
- Modularity $Q \in [0, 1]$
- Resolution parameter r indicating how "tough" the algorithm should look for communities
- Algorithms optimize the modularity score Q given some r (using hill climbing, heuristics, genetic algorithms and many more optimization techniques)

V.D. Blondel, J-L. Guillaume, R. Lambiotte and E. Lefebvre, Fast unfolding of communities in large networks in *Journal of Statistical Mechanics: Theory and Experiment* 10: P10008, 2008.

Related: clustering



Partitions vs. communities



J. Leskovec, Affiliation Network Models for Densely Overlapping Communities, MMDS 2012.

Evaluating communities and partitions

- **Communities**: groups of nodes that are more connected amongst each other than with the other nodes of the network
- **Partitions**: non-overlapping communities
- Compare with groups of nodes based on common attributes
- Human interpretation by hand can suffer from subjective bias

Corporate networks

- **Nodes** are organizations/firms/companies/corporations

Corporate networks

- **Nodes** are organizations/firms/companies/corporations
- **Links** represent:
  - Trade
  - Loans
  - Ownership
  - Social ties

Corporate networks

- **Nodes** are organizations/firms/companies/corporations
- **Links** represent:
  - Trade
  - Loans
  - Ownership
  - Social ties

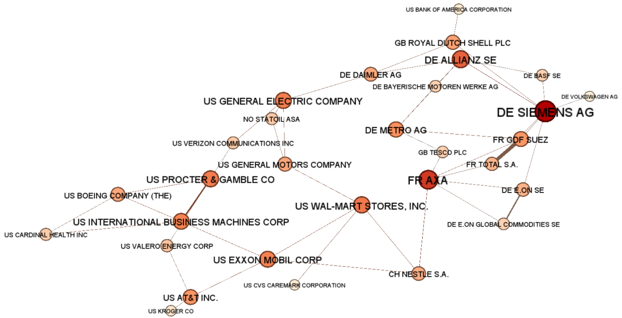


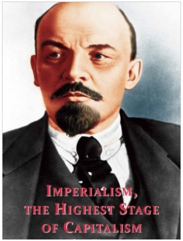
Figure : Board interlock network of 30 firms

Board interlocks

- **Board interlock**: there is a relationship between firms because they share a board member or director

Board interlocks

- **Board interlock**: there is a relationship between firms because they share a board member or director
- Vladimir I. Lenin, *Imperialism, The Highest Stage of Capitalism*, 1916.
- "... a personal union, so to speak, is established between the banks and the biggest industrial and commercial enterprises, the merging of one with another through the acquisition of shares, through the appointment of bank directors to the Supervisory Boards (or Boards of Directors) of industrial and commercial enterprises, and vice versa."



Board interlocks

- **Causes** of interlocks:
  - Collusion
  - Cooptation and monitoring
  - Legitimacy
  - Career advancement
  - Social cohesion
- **Consequences** of interlocks:
  - Corporate control
  - Economic performance
  - Access to resources

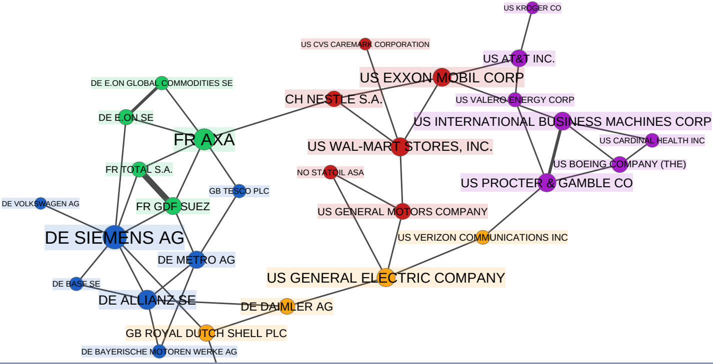
Board interlocks

- **Causes** of interlocks:
  - Collusion
  - Cooptation and monitoring
  - Legitimacy
  - Career advancement
  - Social cohesion
- **Consequences** of interlocks:
  - Corporate control
  - Economic performance
  - Access to resources

M. Mizruchi, What do interlocks do? An analysis, critique, and assessment of research on interlocking directorates, *Annual review of Sociology* 22: 271–298, 1996.

Corporate board interlock networks

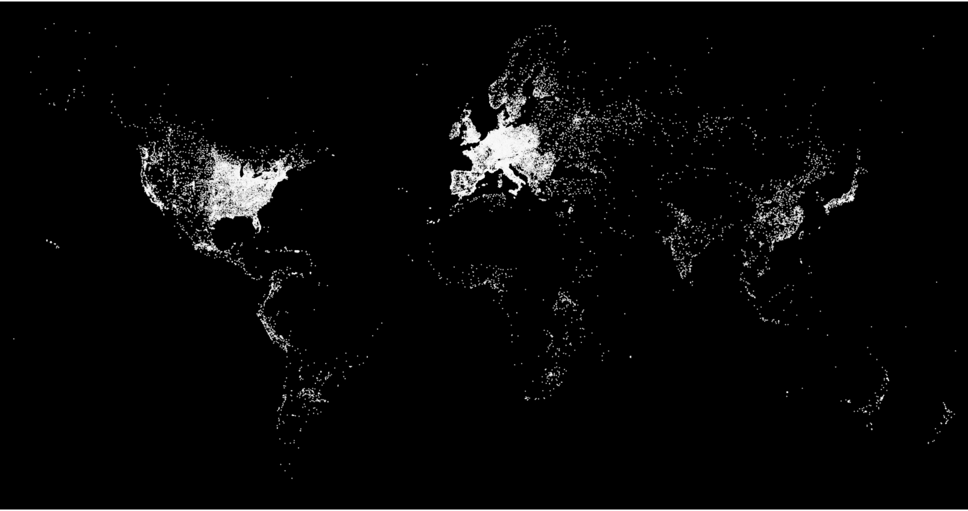
- Nodes are firms, edges represent shared board members
- Power and control
- Globalization
- Influence of countries
- Local or global business groups?



Corporate board interlock networks

- Nodes are **firms** (world-wide)
- Links are social ties between firms (board interlocks)
- **Board interlock**: two firms share a senior level director
- 1,068,409 firms
- 3,262,413 interlocks
- 80 countries
- National subgraphs based on country attribute

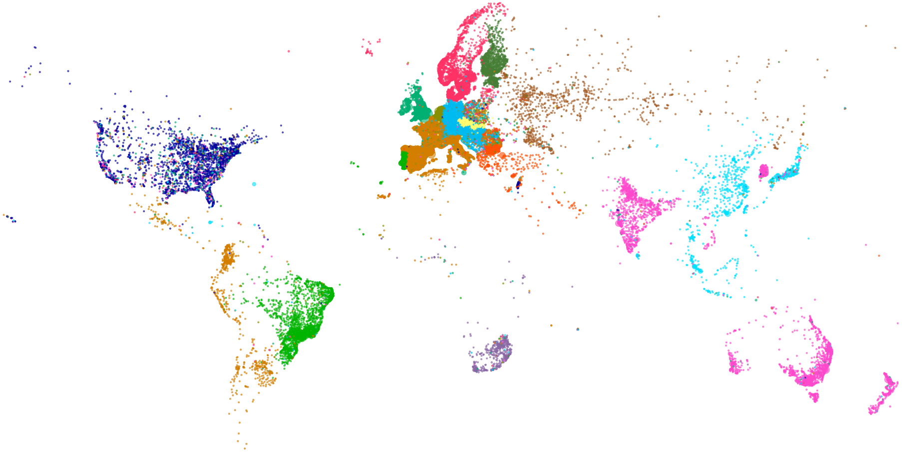
Corporations



Board interlock network

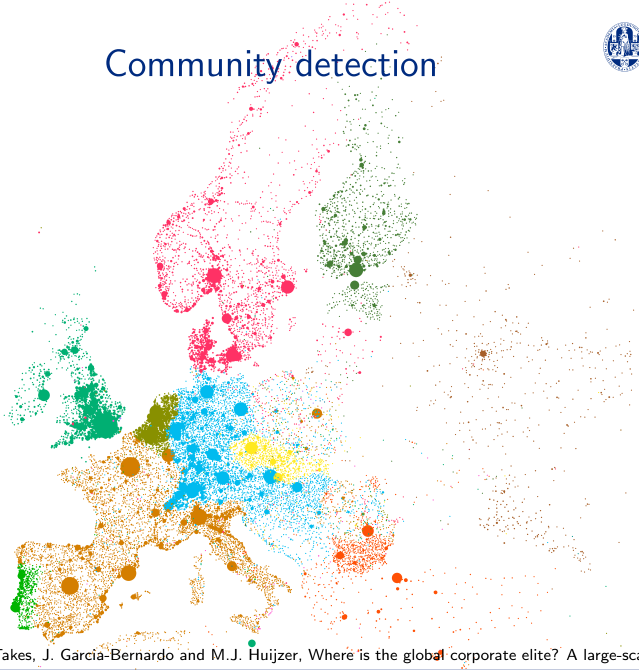
F.W. Takes and E.M. Heemskerk, Centrality in the Global Network of Corporate Control, *Social Network Analysis and Mining* 6(1): 1-18, 2016.

Community detection



E.M. Heemskerk and F.W. Takes, The Corporate Elite Community Structure of Global Capitalism, *New Political Economy* 21(1), 90-118, 2016

Community detection



E.M. Heemskerk, F.W. Takes, J. Garcia-Bernardo and M.J. Huijzer, Where is the global corporate elite? A large-scale network

Lab session March 23

- Finish Python lectures: `https://github.com/jrjohansson/scientific-python-lectures`
- Master Pandas: `http://pandas.pydata.org/pandas-docs/stable/10min.html`
- Replicate metrics from Assignment 1 using this toolstack
- Learn Matplotlib, for example via Lecture 4 of `https://github.com/jrjohansson/scientific-python-lectures` or `http://pandas.pydata.org/pandas-docs/stable/visualization.html`
- Replicate visualizations from Assignment 1 using this toolstack
- Start Assignment 2

Credits

Slides on neural networks based on slides from
"Social Network Analysis for Computer Scientists"
(`https://liacs.leidenuniv.nl/~takesfw/SNACS`)