

# Social Network Analysis Part 1: Allgemeine Einleitung und Metriken

Sonja Wang, Technische Universität Berlin, 17.11.2022

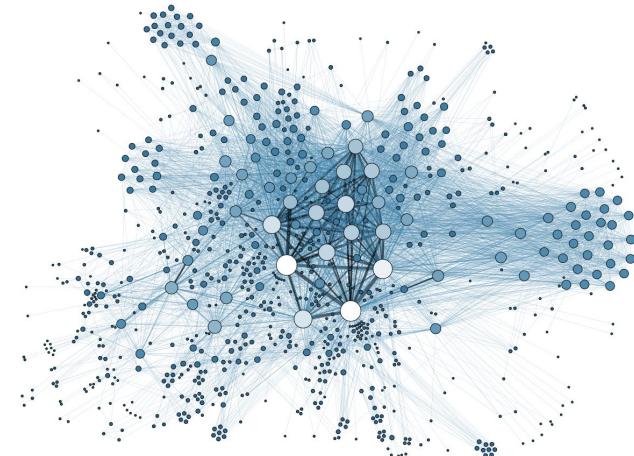
---

# Was wir heute machen...

- **Wie findet man in einem sozialen Netzwerk...**
  - Personen mit vielen sozialen Verbindungen?
  - Wer sich im Zentrum des sozialen Netzwerks befindet?
  - Personen, die soziale Kreise verbinden?
  - einflussreiche Personen im Netzwerk?

# Inhalt

- Begriffe
- Grundlagen der Graphentheorie
- Arten von Netzwerken
- Schreibweisen von Netzwerken
- Metriken (lokale, meso und globale Perspektive)
- Visualisierung von Netzwerken



[Grandjean 2014](#)

# Begriffe

- **Graph:** theoretischer Kontext, in Graphentheorie abstrakte Struktur, die eine Menge von Objekten zusammen mit den zwischen diesen Objekten bestehenden Verbindungen repräsentiert
- **Netzwerk:** reale Strukturen aus der Praxis, wo Knoten Akteure oder Gegenstände sind und die Kanten die Beziehungen zwischen diesen
- **Soziales Netzwerk (Soziologie):** besteht aus einer Gruppe von Individuen, die in Interaktion miteinander stehen, seien es individuelle oder geschäftliche Kontakte oder Bindungen über Familie oder Heirat
- **Social Network Analysis (soziale Netzwerkanalyse):** Untersuchung von sozialen Phänomenen durch Netzwerke und Graphentheorie

# Grundlagen: Graphentheorie

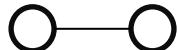
Ein Graph besteht aus Knoten V und Kanten E, die jeweils zwei Knoten verbinden. Es gibt kein alleinstehendes Kantenende und keine alleinstehende Kante.



Knoten



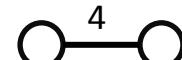
Kante



Zwei Knoten, die durch eine  
ungerichtete Kante  
verbunden sind



Gerichtete Kante



Gewichtete Kante

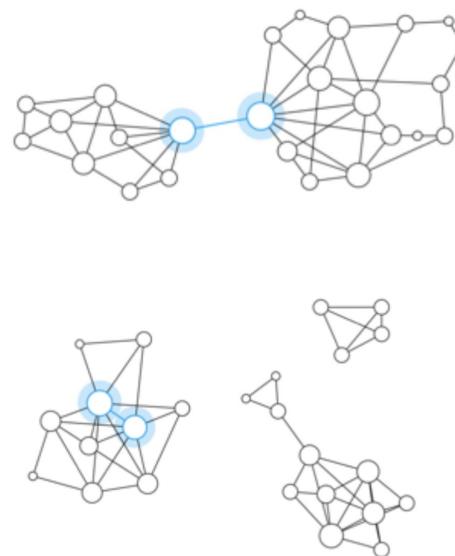
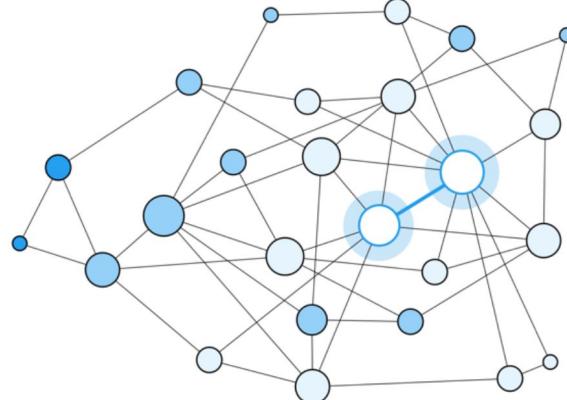


Schlinge (Kante mit gleichem  
Anfangs- und Endknoten)

# Grundlagen: Netzwerke

Alle drei Netzwerke haben 25 Knoten und 50 Kanten.

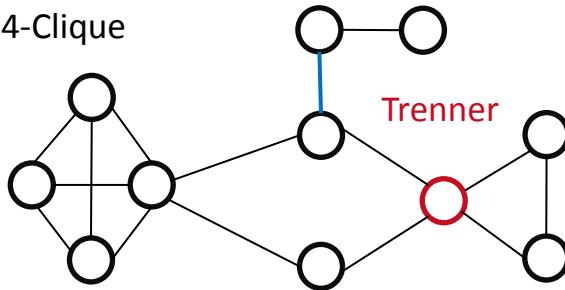
Dieselbe Beziehung in unterschiedlichen Kontexten:



# Grundlagen: Graphentheorie

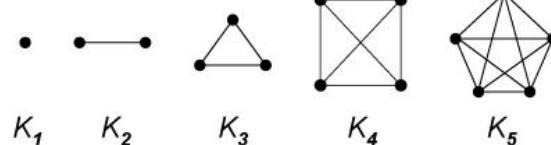
Brücke

4-Clique

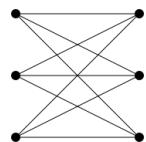


Nicht-zusammenhängender Graph

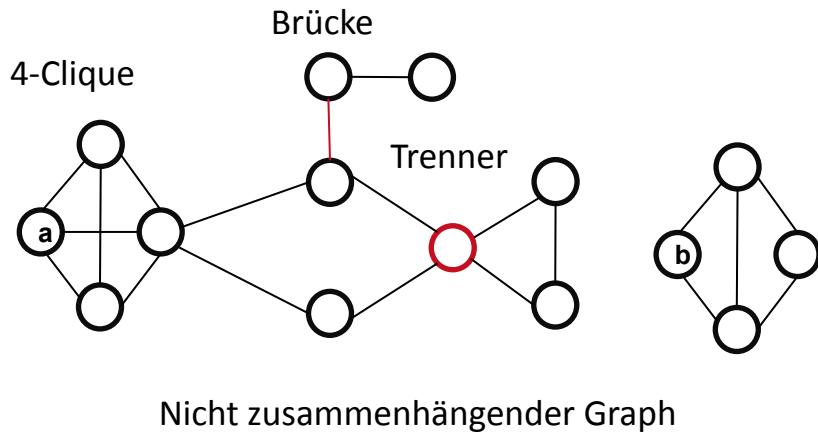
Vollständiger Graph



Bipartiter Graph



# Grundlagen: Graphentheorie

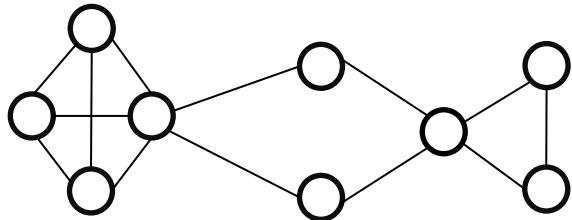


**Umfang (Graph):**  $\text{diam}(G)$  ist die Länge des längsten Weges, für nicht-zusammenhängende Graphen gilt  $\text{diam}(G) = \infty$

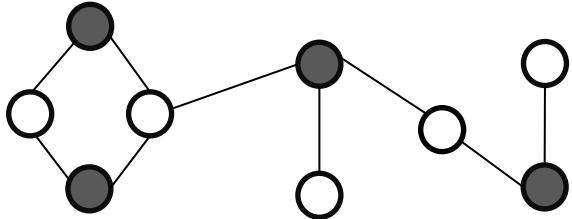
**Distanz/Abstand:**  $d(v,w)$  ist die Länge eines kürzesten Pfades, der von  $v$  nach  $w$  führt, falls nicht existent ist der Abstand unendlich ( $\infty$ ), der Abstand von einem Knoten zu sich selbst ist null.

$d(a,b) = \infty$

# Grundlagen: Arten von Netzwerken



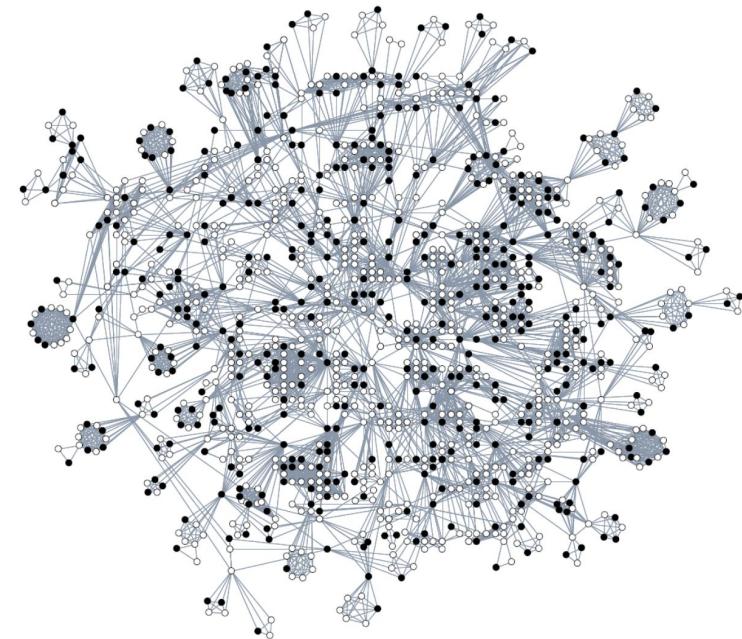
1-mode network



2-mode network



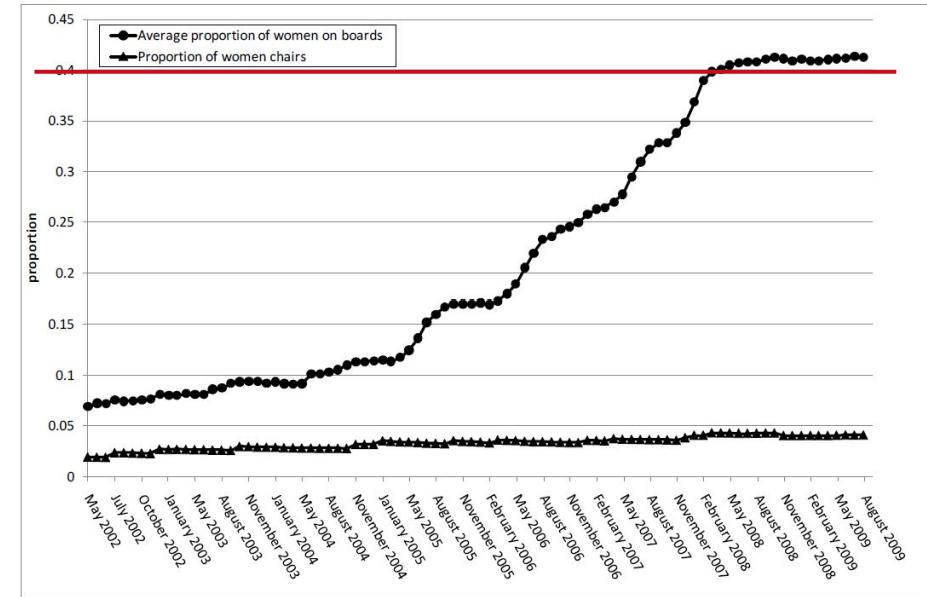
projection



Schwarz gefärbte Knoten: Männer  
Weiß gefärbte Knoten: Frauen

# Two-mode networks: Vorstände in Norwegen

- In Norwegen wurde ein Gesetz über die Repräsentation der Geschlechter in Aufsichtsräten von börsen-notierten Aktiengesellschaften verabschiedet welches 2006 in Kraft trat
- besagt, dass beide Geschlechter mit jeweils mindestens 40% vertreten sein müssen
- Seierstad und Opsahl untersuchten die Effekte dieses Gesetzes



[Seierstadt & Opsahl, 2009](#)

# Two-mode networks: Vorstände in Norwegen

- Analyse fand auf Daten statt, die 2009 über die Website des norwegischen Firmenregisters verfügbar waren, dabei wurden 384 Firmen betrachtet
  - “Public Limited” Firmen wurden ausgesucht, da diese von dem Gesetz betroffen waren
- Interpretation der Autoren war, dass das Gesetz zwar die Repräsentation von Frauen in Firmenvorständen erhöht hat, aber eine kleine Elite von weiblichen Firmenvorständen geschaffen hat, welche einen großen Einfluss haben

NORWEGEN

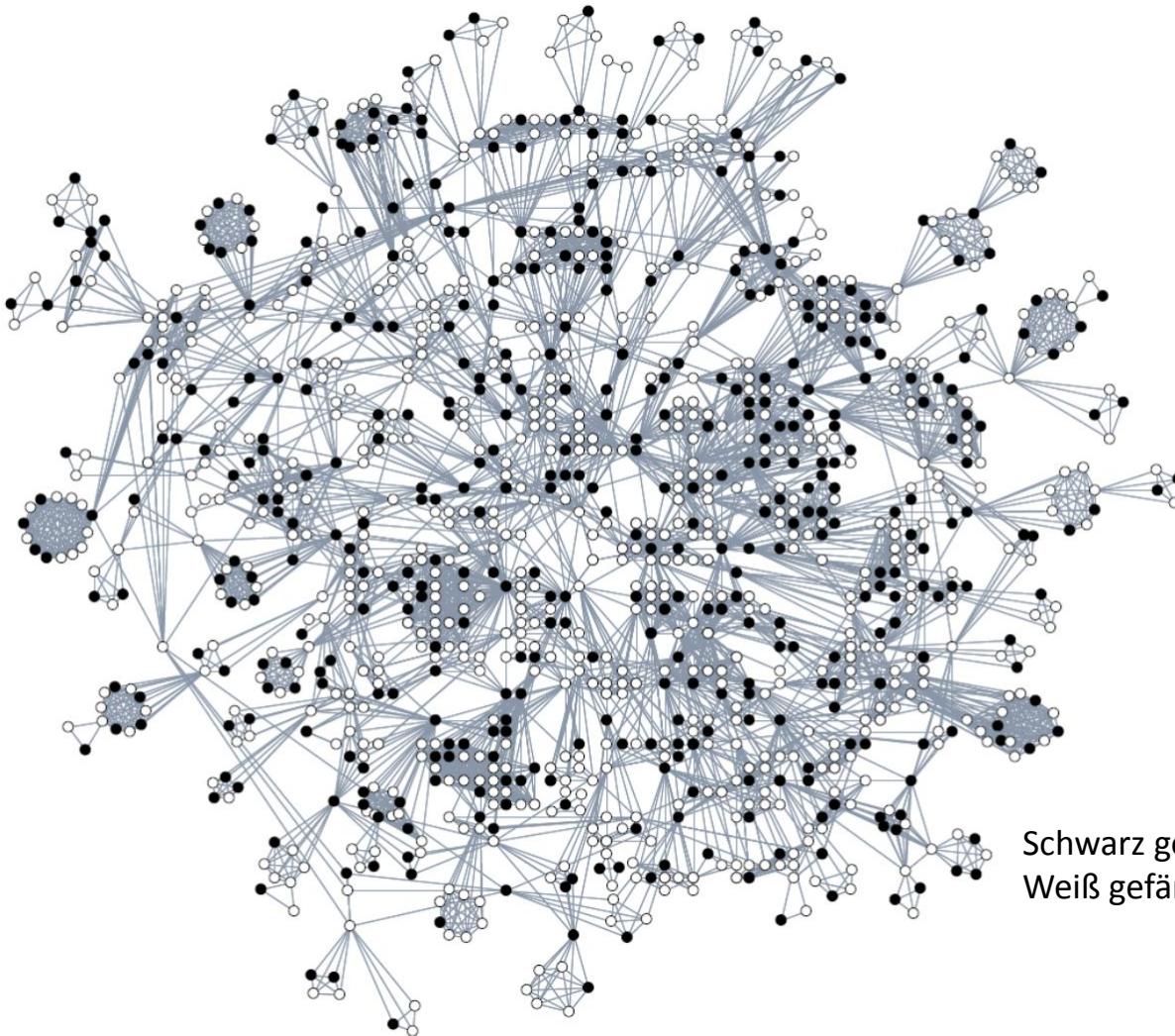
## S+ »Goldene Röcke«

Als erstes Land der Welt hat sich die Ölnation eine Frauenquote für Konzerne verordnet. Sie hat der Wirtschaft nicht geschadet, den Frauen aber auch noch nicht sehr geholfen.

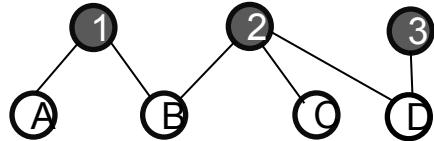
[Traufetter, 2011](#)

Von **Gerald Traufetter**

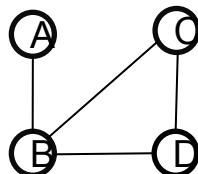
13.02.2011, 13.00 Uhr • aus **DER SPIEGEL 7/2011**



# Grundlagen: Arten von Netzwerken



Anderes Beispiel: Twitter Posts  
(1,2,3) und deren Retweets durch  
Personen (A,B,C,D)

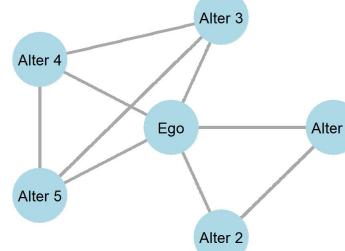


... anstatt der Beziehung der  
Personen untereinander (z.B.  
Follower)

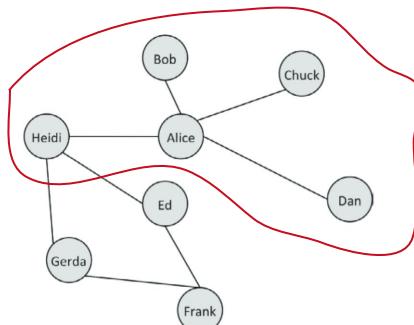
# Grundlagen: Arten von Netzwerken

## Ego-Centric Networks

- Soziales Netzwerk um ein Individuum
- Nur Knoten enthalten, mit denen eine direkte Verbindung besteht
- Beispiel: Das Ego-Centric Network von Alice und ihren Facebook-Freunden würde aus dem Knoten für Alice bestehen (Ego) und den Kanten zwischen ihr und ihren Freunden (Alter)

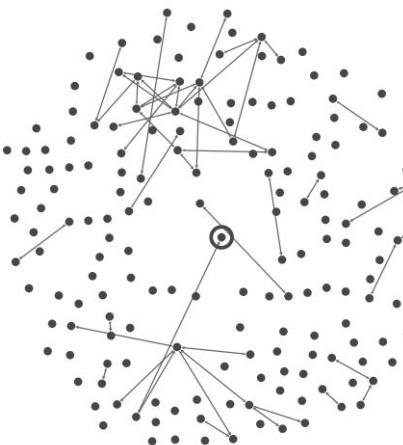


Ego-Centric Network mit 4 Alters  
[Bookdown, n.d.](#); [Golbeck 2015](#)



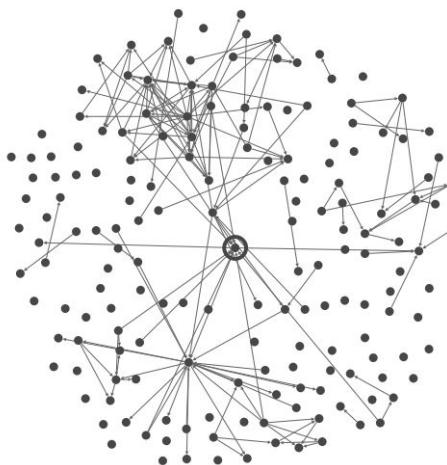
Ego-Centric Network für den Knoten  
 Alice

# Facebook-Kommunikation



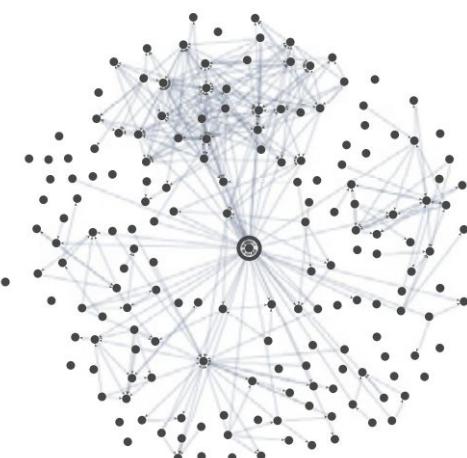
Reziproke Kommunikation

... falls beide Knoten der Kante sich gegenseitig Nachrichten im untersuchten Zeitraum zusanden



Einweg-Kommunikation

... falls der Nutzer mehr Nachrichten an den "Freund" sandte, als er selbst von diesem empfing



Status Beziehung

... falls der Nutzer lediglich Informationen über den "Freund" abrief

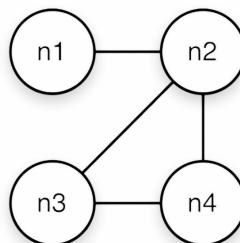
[Krischke, 2014](#)

# Grundlagen: Schreibweisen

- Adjacency Matrices

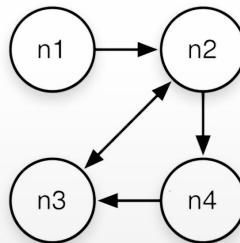
Simple network

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



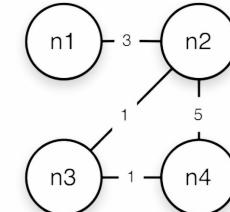
Directed network

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



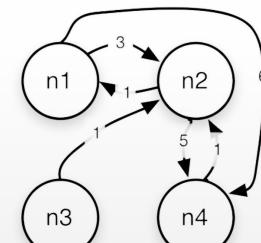
Weighted network

$$A = \begin{pmatrix} 0 & 3 & 0 & 0 \\ 3 & 0 & 1 & 5 \\ 0 & 1 & 0 & 1 \\ 0 & 5 & 1 & 0 \end{pmatrix}$$



Weighted directed network

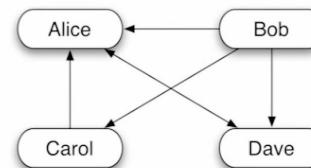
$$A = \begin{pmatrix} 0 & 3 & 0 & 6 \\ 1 & 0 & 0 & 5 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$



# Grundlagen: Schreibweisen

- Weitere Möglichkeiten sind
  - Edge List/Kantenliste (Attribute in weiteren Spalten möglich)
  - Sociograms (visuelle Darstellung von sozialen Beziehungen)

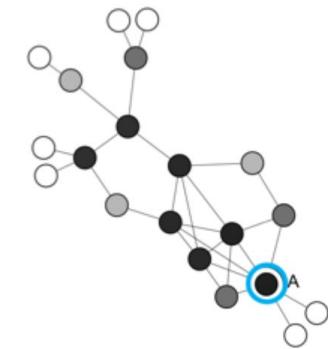
	Alice	Bob	Carol	Dave
Alice	-	0	0	1
Bob	1	-	1	1
Carol	1	0	-	0
Dave	1	0	0	-



From	To
Alice	Dave
Bob	Alice
Bob	Carol
Bob	Dave
Carol	Alice
Dave	Alice
Dave	Bob
Dave	Carol

# Metriken: Degree Centrality

- **Konzept:**
  - Wert für Wichtigkeit eines Knotens in einem Netzwerk basierend auf direkten Verbindungen des Knotens
- **Eigenschaften:**
  - sagt aus, wie viele direkte (ein Hop) Verbindungen der Knoten zu anderen Knoten im Netzwerk hat (je mehr, desto wichtiger)
- **Wann benutzt man es:**
  - Um sehr vernetzte Personen zu finden, beliebte Personen und Personen, die sich schnell vernetzen können (Popularitätsmaß)

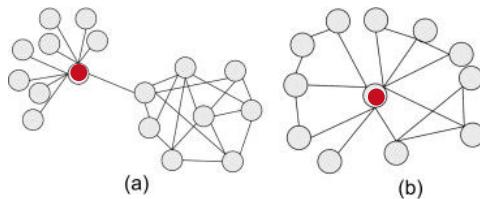


A hat einen Grad von 7  
 (die meisten Verbindungen)

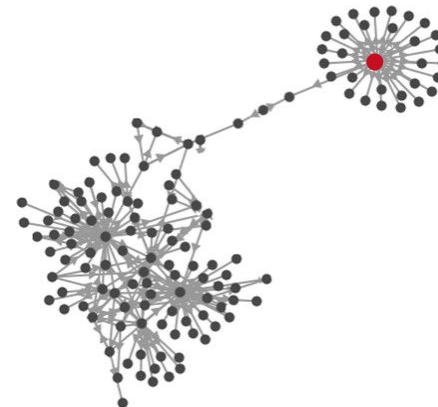
$$C_D(v) = \deg(v)$$

# Metriken: Degree Centrality

- Zeigt zwar, wie viele Verbindungen eine Person hat, aber sie muss sich nicht im Zentrum des Netzwerks befinden, sondern kann in der Peripherie sein

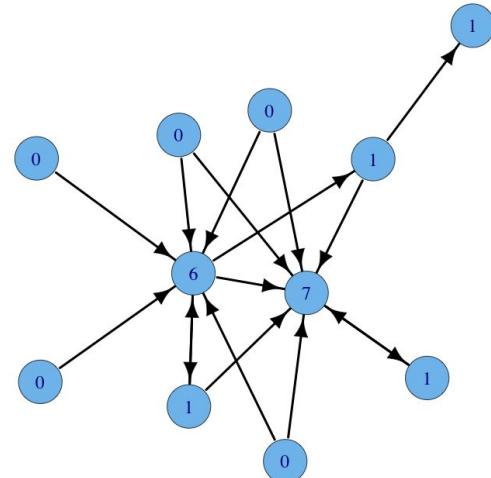


- a) Der Knoten hat einen Grad von 9, aber ist ein Trenner
- b) Der Knoten hat Kanten zu vielen anderen und kurze Pfade um diese zu erreichen



# Metriken: Degree Centrality

- In- und out-degree bei gerichteten Kanten
- **In-degree:** Anzahl an Kanten, die zu einem Knoten zeigen
- **Out-degree:** Anzahl an Verbindungen die am Knoten starten und zu anderen Knoten gerichtet sind
- Beispiel: @Hans erwähnt 10 andere User:innen in seinen Tweets. Wenn man ein Netzwerk betrachtet, in dem dargestellt wird, wie oft Personen sich erwähnen, dann wäre sein Out-Degree Centrality 10.

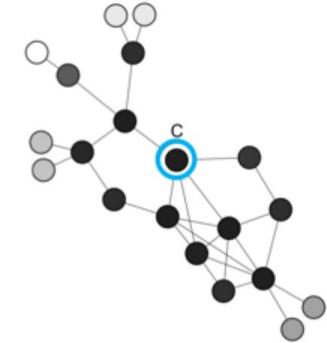


Netzwerk mit Zahlen, die Out-Degree Centrality angeben für jeden Knoten

[Franceschet, n.d.](#)

# Metriken - Closeness Centrality

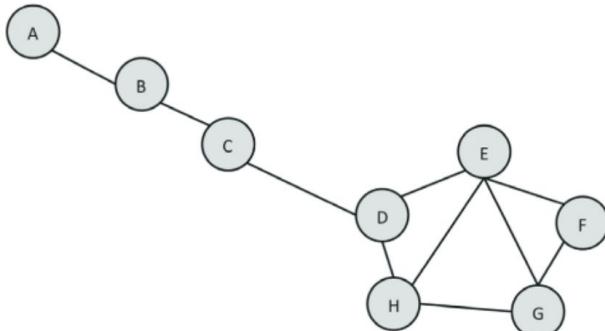
- **Konzept:**
  - Suche nach dem Knoten, der am nächsten an allen anderen ist mit durchschnittlicher Länge aller kürzesten Pfade von diesem Knoten zu jedem anderen im Netzwerk
- **Eigenschaften:**
  - Gibt an, wie nah ein Knoten zu allen anderen Knoten in einem Netzwerk ist
- **Wann benutzt man es:**
  - Die Person ist in einer guten Position um von den meisten seiner Freunde zu hören (gute Informationsquelle, da einfach zu erreichen)



[Grandjean 2021](#)

# Metriken - Closeness Centrality

Ursprünglich als Kehrwert zu Entfernung definiert,  
 in der Regel in normalisierter Form verwendet:  
 (N die Anzahl der Knoten)



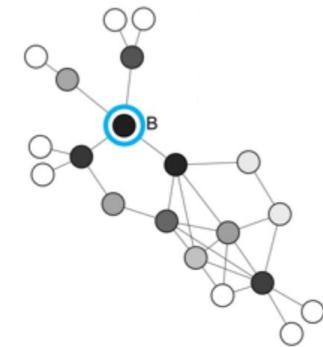
$$C(x) = \frac{N - 1}{\sum_y d(y, x)}.$$

Node	Shortest path from D
A	3
B	2
C	1
E	1
F	2
G	2
H	1

Die Summe der Distanzen zu d ist 12. Die Anzahl der Knoten ist 8. Eingesetzt in die Formel ergibt das  $7/12 = 0,583$

# Metriken - Betweenness Centrality

- **Konzept:**
  - Wie oft der Knoten auf dem kürzesten Pfad zwischen anderen Knoten liegt
- **Eigenschaften:**
  - Zeigt welche Knoten als “Brücken” fungieren zwischen anderen Knoten im Netzwerk
- **Wann benutzt man es:**
  - Um Personen zu finden, die Informationsflüsse in einem Netzwerk (zu einem anderen) beeinflussen
  - Diese Personen haben also die Möglichkeit Informationen zu verändern oder für sich zu behalten



B hat eine hohe Betweenness Centrality, da er sich auf vielen der kürzesten Pfade befindet

[Grandjean 2021](#)

# Metriken - Betweenness Centrality

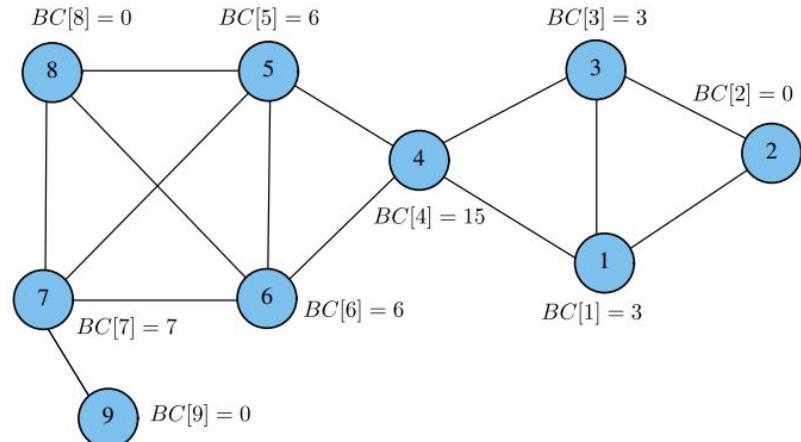
- Bei hoher Betweenness Centrality hat Person Autorität über verschiedene Cluster oder befindet sich an der Peripherie von beiden Clustern oder hat Informationen über mehrere soziale Kreise
- Betweenness Centrality ist etwas komplexer zu berechnen und es gibt verschiedene Algorithmen diese Metrik zu berechnen (z.B. Brandes's Algorithmus)

$$BC(v) = \sum_{s \neq t \neq v} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

Formel zur Berechnung der Betweenness Centrality

# Metriken - Betweenness Centrality

- Knoten 4 ist der einzige Knoten, der auf den Pfaden von Knoten von links nach rechts liegt (also Knoten 5-9 zu Knoten 1-3)
- Knoten 4 hat somit einen hohen BC-Wert
- Knoten 9 dagegen liegt auf keinem Pfad zwischen zwei Knoten und hat damit einen BC-Wert von 0.



[Bader, 2015](#)

# Metriken - Eigenvector Centrality

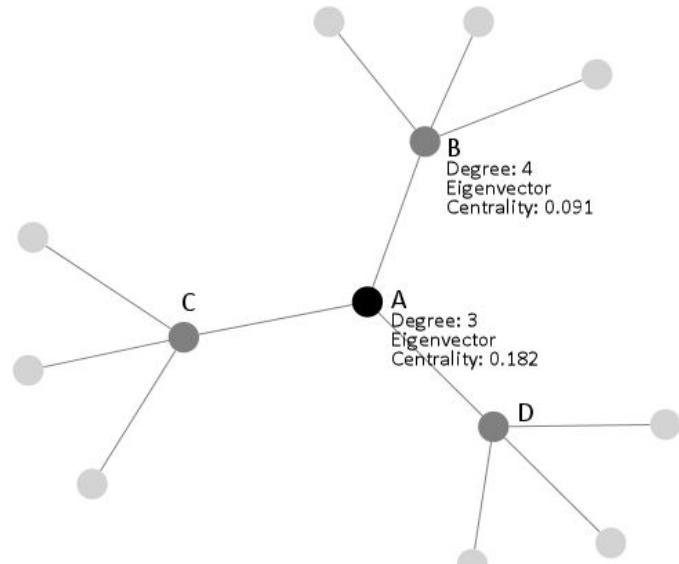
- **Konzept:** Indikator für “Prestige” oder “Einfluss”
- **Eigenschaften:**
  - Misst Einfluss eines Knotens im Netzwerk durch Einbeziehung der anderen Knoten
- **Wann benutzt man es:**
  - Ein Knoten kann geringe Degree-Centrality haben und auch schwache Closeness-Centrality und Betweenness-Centrality, aber trotzdem einflussreich sein



[Grandjean 2021](#)

# Metriken - Eigenvector Centrality

- Eigenvector Centrality “expandiert” die Idee von Degree Centrality
- Es werden die Alters von einem Knoten einbezogen
- Knoten A hat einen Grad von drei und Knoten B von vier, also höher
- B ist beliebter im Netzwerk, wenn wir uns nur die direkten Nachbarn anschauen
- Aber A ist mit Knoten verbunden, die zu vielen anderen Knoten Verbindungen haben



# Metriken - Eigenvector Centrality

- Eigenvector Centrality hängt ab von der Anzahl und der Qualität der Verbindungen (diese Knoten haben wiederum viele Verbindungen zu anderen)
- (Nach dem Perron-Frobenius Theorem muss Lambda der größte Eigenwert der Adjazenzmatrix A und dem korrespondierenden Eigenvektor x sein)
- PageRank-Algorithmus von Google ist eine Variante von Eigenvector Centrality, hauptsächlich genutzt in gerichteten Netzwerken

$$x_i = \frac{1}{\lambda} \sum_{j=1}^n A_{ij}x_j,$$



$$\lambda x = A \cdot x,$$

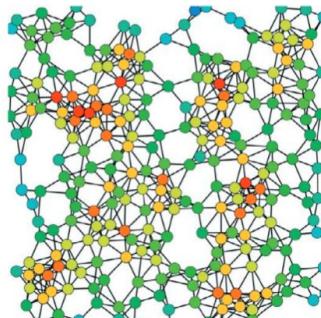
[Newman 2006](#)

# Zusammenfassung

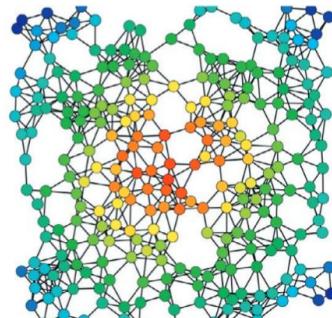
- **Degree Centrality** zeigt Personen mit vielen sozialen Verbindungen - Anzahl der Verbindungen eines Knotens
- **Closeness Centrality** zeigt an, wer sich in der Mitte des sozialen Netzwerks befindet - durchschnittliche Distanz zu allen anderen Knoten
- **Betweenness Centrality** beschreibt Personen, die soziale Kreise verbinden - Position auf den kürzesten Pfaden
- **Eigenvector Centrality** ist hoch bei einflussreichen Personen im Netzwerk -

# Zusammenfassung

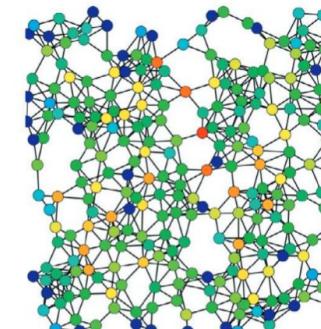
- Dasselbe Netzwerk vier Mal dargestellt, aber mit verschiedenen Metriken:
  - a) Degree Centrality
  - b) Closeness Centrality
  - c) Betweenness Centrality
  - d) Eigenvector Centrality



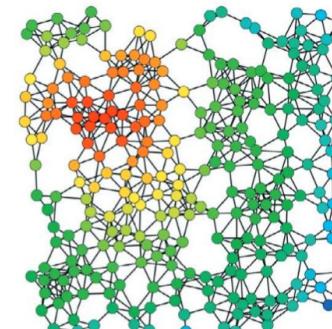
(a)



(b)



(c)



(d)

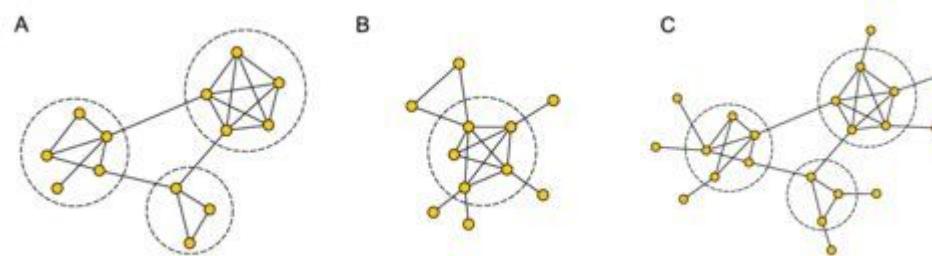
# Social Network Analysis: Clustering Coefficient und Homophilie

Sonja Wang, 17.11.2023

---

# Perspektiven

- Lokale: letztes Mal gelernt (Centrality Metriken)
- Globale: betreffen das ganze Netzwerk
- Intermediate/Meso-Scale: betrachtet einen Teil des Netzwerks
  - Core-Periphery
  - Community
  - ...



Beispiele für Community, Core-Periphery und Multicore-Periphery Strukturen ([Yan & Luo, 2019](#))

# Metriken: Clustering Coefficient

- Cluster: Gruppe von “ähnlichen” Objekten, die sich ähnlicher sind, als solche aus anderen Gruppen (ähnlich kann graphentheoretisch, partitionierend, ... sein)
- Koeffizient für Maß von Cliquenbildung
- Es gibt einen lokalen und globalen Clustering Coefficient
- Der (globale) Clustering Coefficient unterscheidet sich von Centrality Metriken, da es sich mehr um eine Metrik für die Dichte des gesamten Netzwerk handelt, der lokale Clustering Coefficient betrachtet einen Knoten

# Metriken: Local Clustering Coefficient

- Verhältnis von Anzahl von Verbindungen zwischen Knoten in der Nachbarschaft geteilt durch Anzahl aller Verbindungen, die möglich sind
- Angenommen ein Knoten  $v$  hat  $k_i$  Nachbarn, dann können höchstens  $k_i(k_i - 1)$  Kanten zwischen diesen bestehen (wenn alle Nachbarn untereinander verbunden sind)
- $n$  ist die Anzahl der Verbindungen zwischen den Nachbarn
- $C_i = 0 \rightarrow$  sternförmig,  $C_i = 1 \rightarrow$  Clique

$$C_i = \frac{2n}{k_i(k_i - 1)}.$$

für ungerichtete Graphen  
 (nach [Watts und Strogatz, 1998](#))



Knoten 1 hat einen Clustering Coefficient von 0 (Nachbarn nicht verbunden), Knoten 2 hat einen von  $\frac{2}{3}$  ( $n=2$ ,  $k=3$ )  
[Hauff et al., 2005](#)

# Metriken: Global Clustering Coefficient

- Mittel aller lokalen Clustering Coefficients
- Problem: Bei Projektion von two-mode-networks kommt es zu hohem Clustering, welches verzerrt ist

$$\bar{C} = \frac{1}{n} \sum_{i=1}^n C_i.$$

## Für Ego-centric Networks:

- Wenn die Freunde (Alters) sich alle gegenseitig kennen, ist der Clustering Coefficient hoch
- Wenn die Freunde (Alters) sich nicht kennen, dann ist der Clustering Coefficient niedrig

Mittel aller lokalen Clustering  
Coefficients

(nach [Watts und Strogatz, 1998](#))

# The Strength of Weak Ties

## The strength of weak ties

[PDF] usp.br

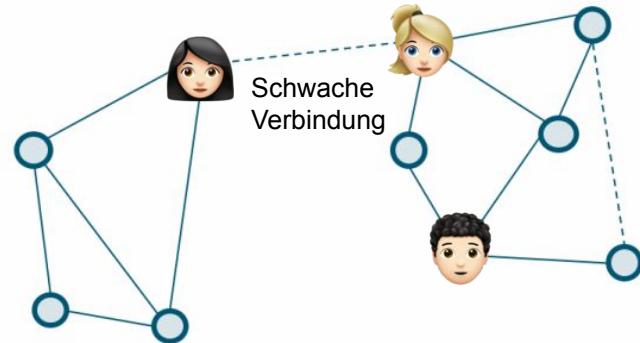
[MS Granovetter - American journal of sociology, 1973 - journals.uchicago.edu](#)

... 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 ...

[☆ Save](#) [✉ Cite](#) Cited by 67603 Related articles All 107 versions [»»](#)

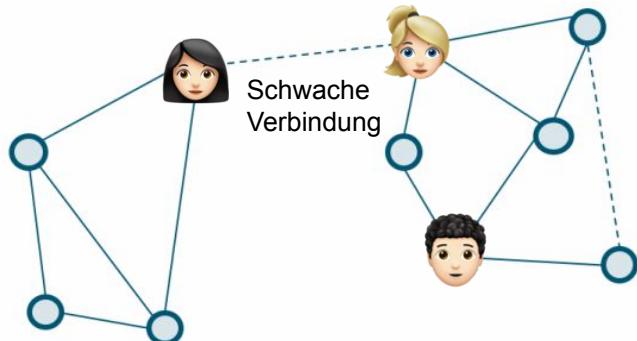
# The Strength of Weak Ties

- Wegweisendes Paper von [Granovetter \(1973\)](#) in Forschung zu sozialen Netzwerken
- Untersuchung von Bedeutung der Verbindungen zu anderen Menschen (u.a. in der Arbeitssuche)
- **Definition von „Tie“ (Verbindung) und Stärke der Verbindung:**  
„a combination of the amount of time, the emotional intensity, the intimacy (mutual confiding), and the reciprocal services which characterize the tie.“



# The Strength of Weak Ties

- Beziehungen als symmetrisch und positiv angenommen
- **Dyadische Beziehung** (Soziologie): intensive soziale Beziehung zwischen zwei Personen, analog triadisch für drei Personen
- Wenn A und B verbunden sind, A und C verbunden sind und B und C verbunden sind  
→ nennt sich Triade in sozialen Netzwerken



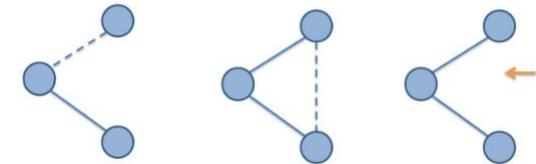
# Gruppenarbeit

**Lesen Sie “The strength of weak ties” Seiten 1361-1366 und beantworten Sie folgende Fragen:**

- 1) Wie legt der Autor fest, dass es sich um strong, weak oder absent ties handelt?
- 2) Wie stützt der Autor die Hypothese, dass sich Triaden in sozialen Beziehungen bilden?
- 3) Was ist ein “Forbidden Triad”?
- 4) Was ist die Bedeutung von Brücken?

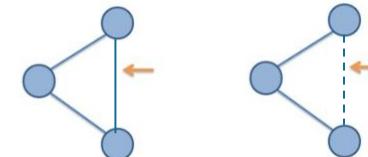
# The Strength of Weak Ties

- „*Forbidden Triad*“: Wenn A und B eine starke Verbindung haben und A und C eine starke Verbindung haben, dann ist es sehr unwahrscheinlich, dass A und C keine Verbindung haben
  - Für starke Verbindungen gilt somit, dass sie keine Brücke sind
  - Alle Brücken weak ties (aber nicht alle weak ties müssen Brücken sein)



Forbidden Triad

Lalone et al. (2018)



# The Strength of Weak Ties

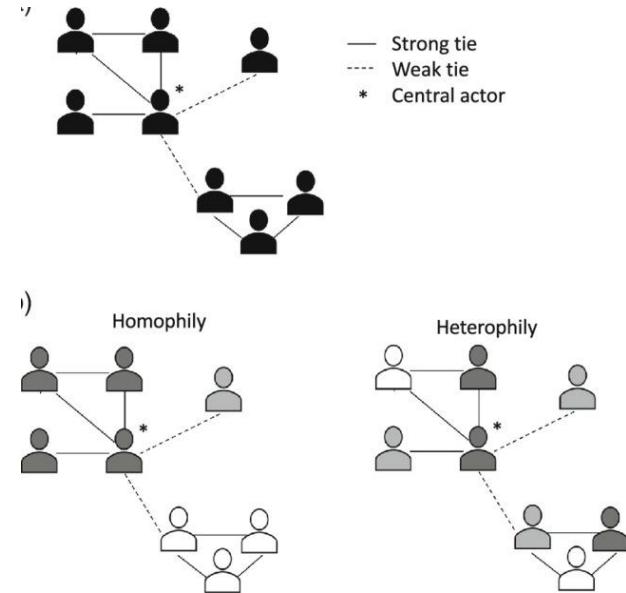
How does policy information flow through Washington 'issue networks'? And how does information flow determine which **lobbyists** get access in policy-making? Drawing upon the 'strength of weak ties' argument, the authors argue that policy information passes more through acquaintances ('weak ties') than through close, trusted, contacts ('strong ties'). They support this argument in an analysis of data on lobbying networks in health-care policy-making in the 1970s and 1980s.

[Carpenter et al., 1998](#)

# Homophilie

**“Birds of a feather flock together”**

- Tendenz in sozialen Gruppen, dass ähnliche Menschen sich verbinden, aber auch im Zusammenhang mit Polarisierung betrachtet
- Kommunikation zwischen ähnlichen Personen häufiger als zwischen nicht-ähnlichen
- Lazarsfeld & Merton (1954):
  - **Status Homophily:** Ähnlichkeit basierend auf informellen, formalen oder zugeschriebenem Status (soziale Position)
  - **Value Homophily:** Ähnlichkeit bei Werten, Ansichten und Glauben



[Lane et al., 2019](#)

# Homophilie

- **EI index:** (External ties-internal ties)/(External+Internal)

- **Individual Homophily/Polarization Score:**

$$h_i(v) = \frac{d_i(v)}{d(v)}$$

v sind Knoten aus  $A_i$ , wobei  $A_i$  eine spezifische Gruppe ist (z.B: Anzahl konservativer Freunde aus Gesamtzahl aller Freunde,  $d_i$  ist Grad des Knotens v

- **Group Homophily/Group Polarization:**

berechnet durch Summe von

$$H_i = \frac{\sum_{v \in A_i} d_i(v)}{\sum_{v \in A_i} d(v)}$$

individual homophilies von Gruppenmitgliedern

- **Network-level Homophily:**

Homophilie eines gesamten Netzwerkes

$$H_i = \frac{\sum_{v \in V} d_s(v)}{\sum_{v \in V} d(v)}$$

(über alle Knoten aus V)

# Homophilie

- Weitere Maße:
  - Inbreeding homophily
  - Assortativity
  - Yule's Q (EI normalisiert)

# Homophilie

- Studien zu Homophilie in Social Media Daten bei Tweets, Follower-Listen...
- Auch im Zusammenhang von Echo Chambers und Filter Bubbles untersucht, vor allem durch Freundschaftsnetzwerke und Follower-Followee-Netzwerke
- Echo chambers untersucht mit unterschiedlichen Konzepten, Homophily ist eines dieser Konzepte, bei dem Echo Chambers als Communities von ähnlich-gesinnten Personen angesehen werden
  - Viele Paper betrachten Anzahl von Social Media Kontakten oder Verbindungen zwischen Gruppen/Organisationen, die bereits stark polarisiert sind (Demokraten vs. Republikaner, Verschwörungstheoretiker vs. Wissenschaftscommunity,...)

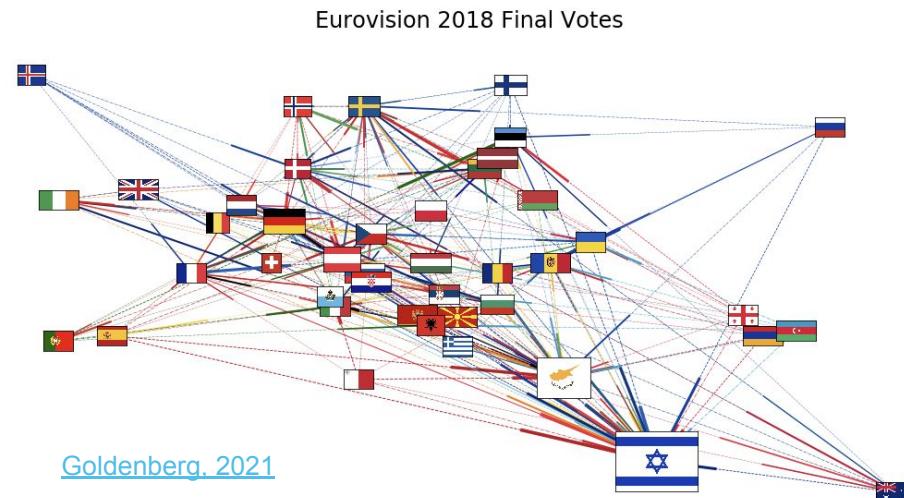
# Visualisierungstools

- [\*\*Gephi\*\*](#) (Open-Source, Java-basiert)
- [\*\*NetworkX\*\*](#) (Python package)
- [\*\*Igraph\*\*](#) (Sammlung von Network Analysis Tools in C geschrieben, auch als Python und R package)
- [\*\*Neo4J\*\*](#) (Graph Datenbank Management System, Java-basiert)
- [\*\*Cytoscape\*\*](#) (Eigentlich für computational biology/bioinformatics)
- [\*\*NodeXL\*\*](#) (Excel Add-in)
- [\*\*GraphViz\*\*](#) (Für Visualisierung von Graphen)

# Visualisierung

## “Visual Variables”:

1. Position: eines Objekts im Bild, kann Realität entsprechen (z.B. geolocation)
2. Size: Größenunterschiede von Objekten (z.B. Knoten mit hohem Grad größer)
3. Shape: geometrische Form, die ein Objekt haben kann (z.B. Kreis, Dreieck, Rechteck, ...)

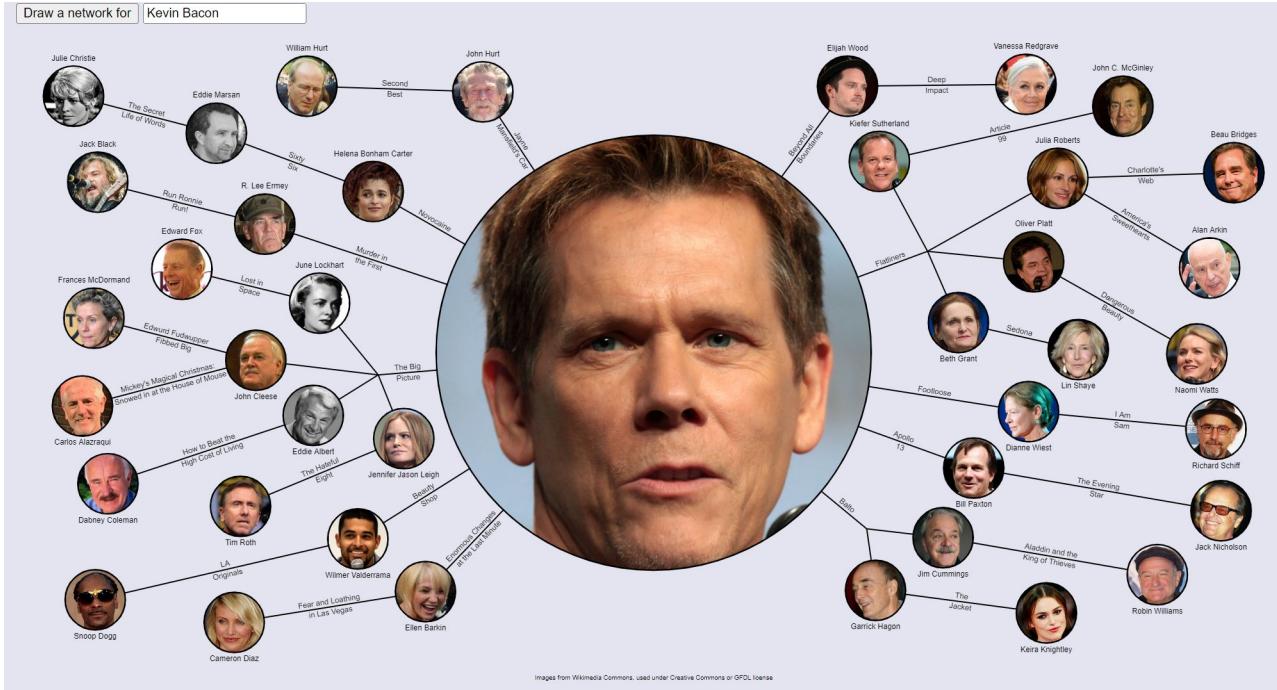


# Visualisierung

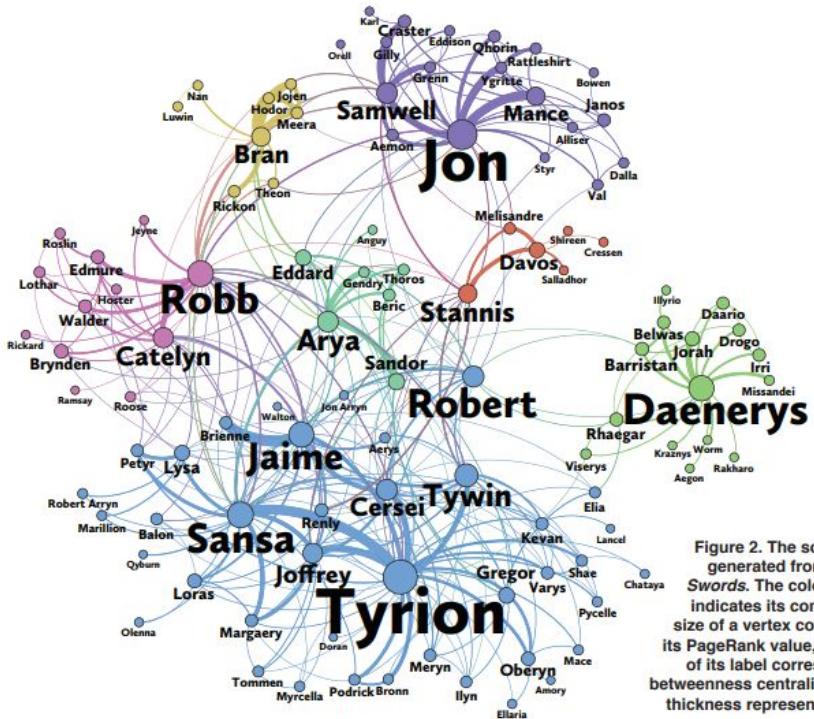
## “Visual Variables”:

1. Orientation: Wohin das Objekt zeigt, zum Beispiel Pfeil für Richtung von Follower/Person der gefolgt wird
2. Color: Farbe eines Objekts, z.B. eine Farbe für eine Gruppe im 2-mode-network
3. Saturation: Hellere oder gesättigtere Farbe, z.B. höhere Centrality mit mehr gesättigter Farbe
4. Texture: Füllung oder Muster eines Objekts, z.B. unterschiedliche Muster für Mitglieder unterschiedlicher Parteien

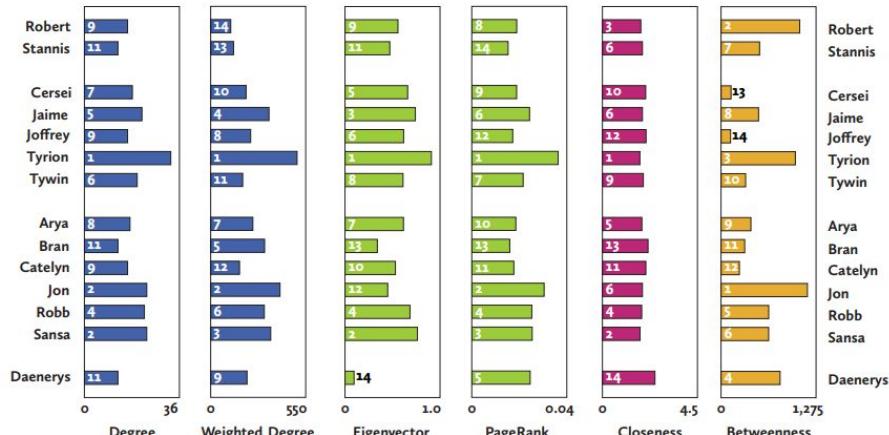
# Visualisierung



# Visualisierung



**Figure 2.** The social network generated from *A Storm of Swords*. The color of a vertex indicates its community. The size of a vertex corresponds to its PageRank value, and the size of its label corresponds to its betweenness centrality. An edge's thickness represents its weight.



**Figure 3.** Centrality measures for the network. Larger values correspond to greater importance, except for closeness centrality, where smaller values are better. Numbers in the bars give the rankings of these characters.

# Visualisierungsalgorithmen

- Visualisierungen sind für das menschliche Gehirn oft einfacher zu verarbeiten als Tabellen oder Berichte
- Verschiedene Algorithmen für Visualisierungen (Anordnung der Knoten)
  - **Force-directed Algorithmen:** in 2-D oder 3-D-Raum Knoten platzieren, sodass die Kanten etwa gleiche Länge haben und es so wenig kreuzende Kanten gibt wie möglich, durch Vergabe von Kräften in der Kantenmenge und Knotenmenge
    - **“Spring-Electrical Models”:** einer der bekanntesten ist der Fruchterman-Reingold Algorithmus
    - **“Stress and Strain Models”:** einer der bekanntesten ist der Kamada-Kawai Algorithmus
- Kombination von mehreren Algorithmen möglich, Einstellungen beachten!

# Visualisierungsalgorithmen

## Kamada-Kawai:

- Arbeiten mit partiellen Differentialgleichungen
- Basierend auf Hookeschen Gesetz (elastische Verformung von Festkörpern)
- Auch Betrachtung von Knoten und Kanten als Stahlringe und Federn

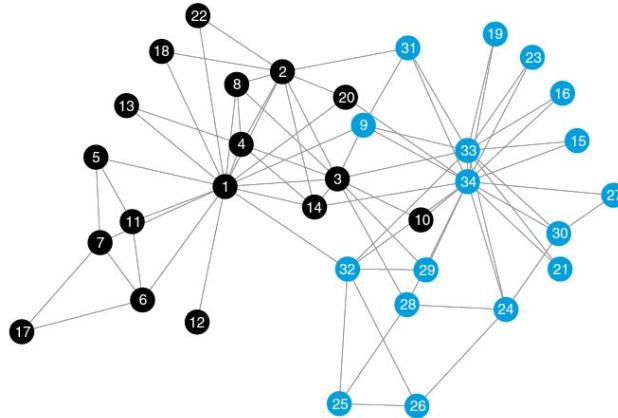


FIG. 4.1. The Zachary Karate Club network [70], which we visualize using the implementation of the Kamada–Kawai algorithm [33] in [60]. The colors represent the two groups into which the club split while it was under study.

[Rombach et al. 2014](#)

# Visualisierungsalgorithmen

## Graph-Drawing Problem

- Gleichmäßige Verteilung von Knoten im Rahmen
- Minimierung von kreuzenden Kanten
- Gleichmäßige Länge von Kanten
- Symmetrie soll graphisch beibehalten werden
- Anpassung an Rahmen

→ **Fruchterman-Reingold:** nicht explizit Streben nach all diesen Zielen, aber gut für Symmetrie wiedergeben, gleichmäßige Verteilung und gleichmäßige Kantenlängen, Ziele für Implementierung waren Geschwindigkeit und Einfachheit

# Visualisierungsalgorithmen

## Fruchterman-Reingold:

- behandelt Knoten im Graphen als atomare Partikel oder himmlische Körper, die anziehende und abstoßende Kräfte gegenüber einander ausüben
  - Inspiriert durch Federn und Gravitation
  - Nur Nachbarn ziehen sich gegenseitig an, alle Knoten stoßen sich gegenseitig ab
- attractive and repulsive forces

# Visualisierungsalgorithmen

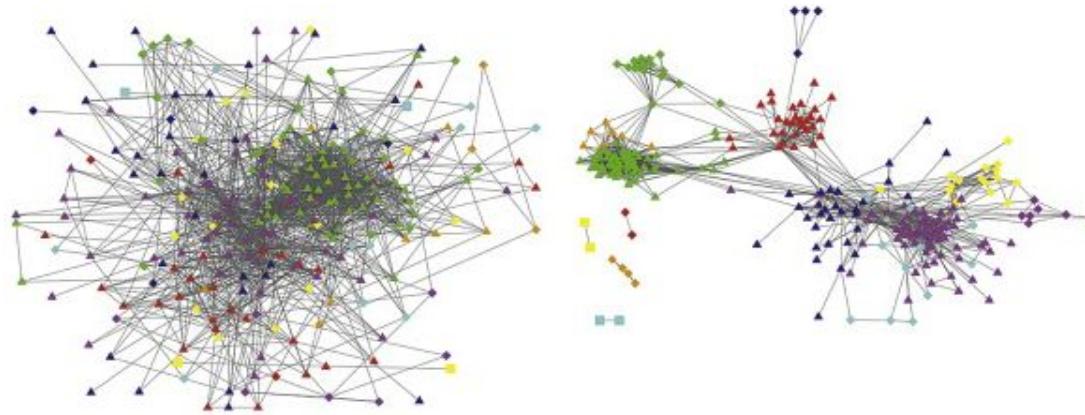
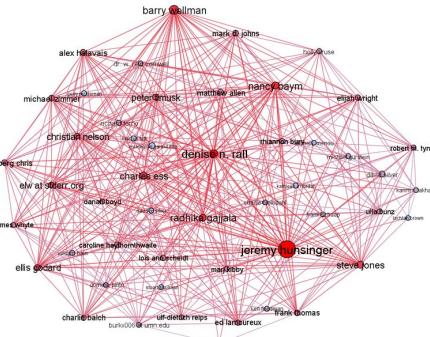
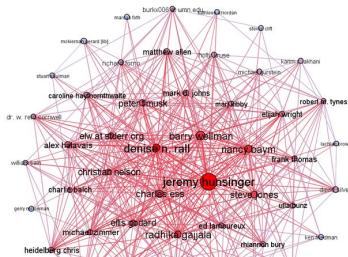


Figure 11.5. Comparing two network layouts of same graph. The first uses the default Fruchterman-Reingold settings and the second uses 100 iterations and a repulsion of 3.3. Clusters and bridges between these clusters are more clearly shown in the second layout.

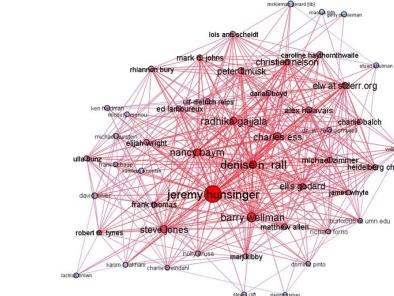
# Visualisierungsalgorithmen



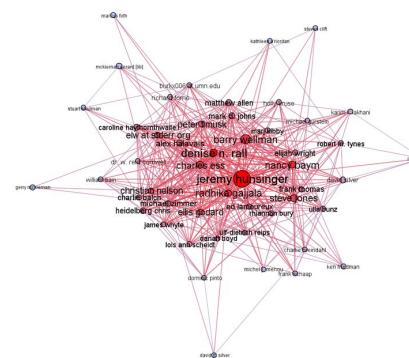
Force Atlas with Attraction Distribution



Fruchterman Reingold



Force Atlas



Yifan Hu Proportional

# Visualisierung

- 107 Knoten (Charaktere wie Dorfbewohner:innen, Ladies, Lords, ...)
- 353 gewichtete Kanten (integer), wobei höhere Gewichte stärkere Beziehung darstellen
- E-book parsing, Inkrementierung von Kantengewicht immer dann, wenn zwei Charaktere (Name oder Spitzname) innerhalb von 15 Wörtern auftraten

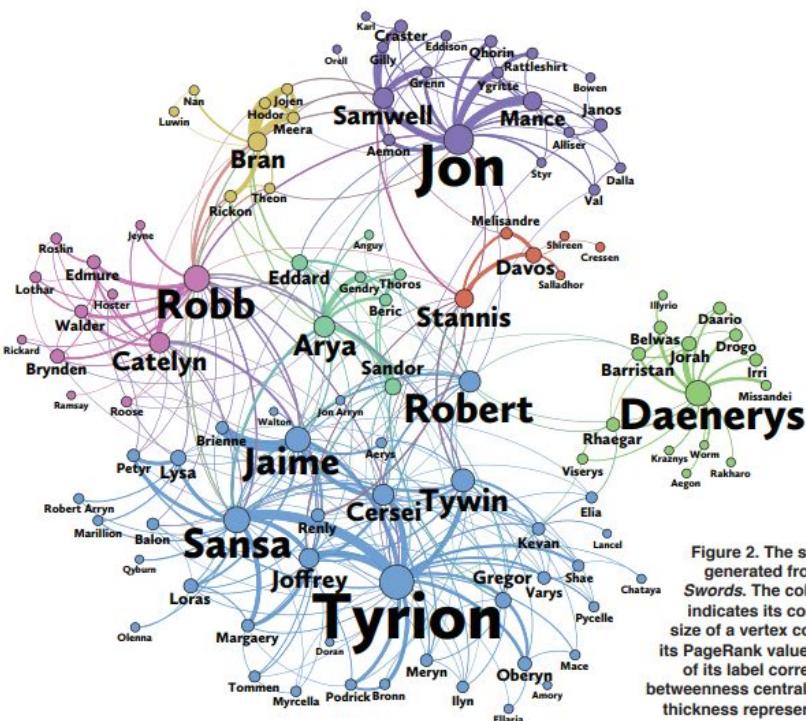


Figure 2. The social network generated from *A Storm of Swords*. The color of a vertex indicates its community. The size of a vertex corresponds to its PageRank value, and the size of its label corresponds to its betweenness centrality. An edge's thickness represents its weight.

# Visualisierung

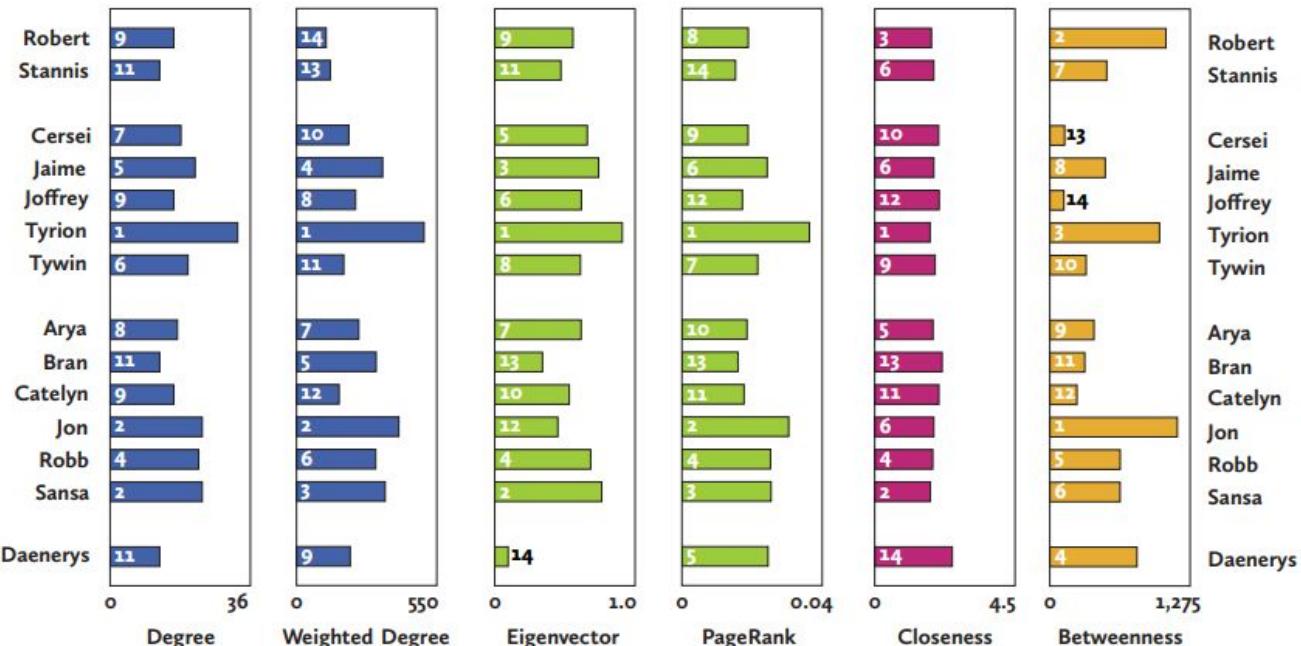


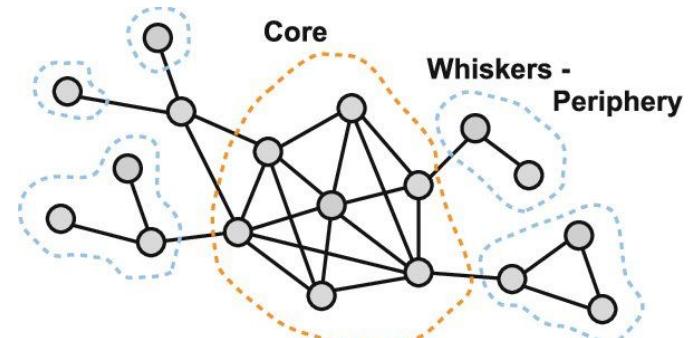
Figure 3. Centrality measures for the network. Larger values correspond to greater importance, except for closeness centrality, where smaller values are better. Numbers in the bars give the rankings of these characters.

# Meso-Scale Perspective

---

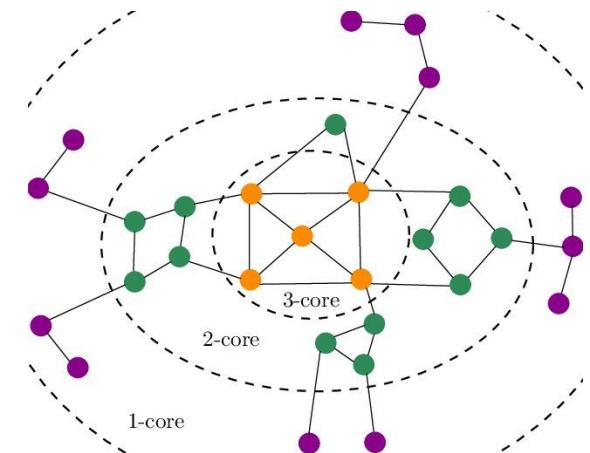
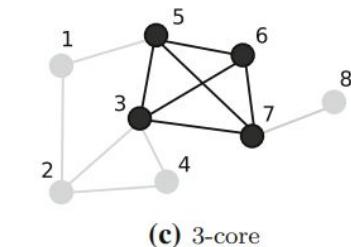
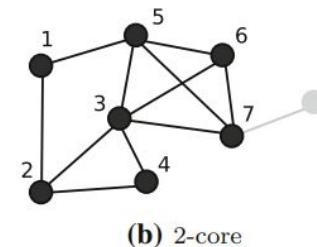
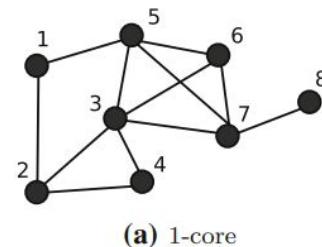
# Zentrum und Peripherie

- Herausfinden, welche Knoten Teil eines dicht-zusammenhängenden Kerns sind (“Core”) und welche zu der wenig zusammenhängenden Peripherie gehören
- Der Kern ist nicht nur dicht-zusammenhängend, sondern auch “zentral” im Netzwerk (in Bezug auf kürzeste Pfade durch das Netzwerk)
- Knoten mit sehr hohem Grad werden manchmal “Hub” genannt
- Core-Knoten sollten gut verbunden sein zu peripheren Knoten, aber periphere Knoten nicht zum Kern oder untereinander



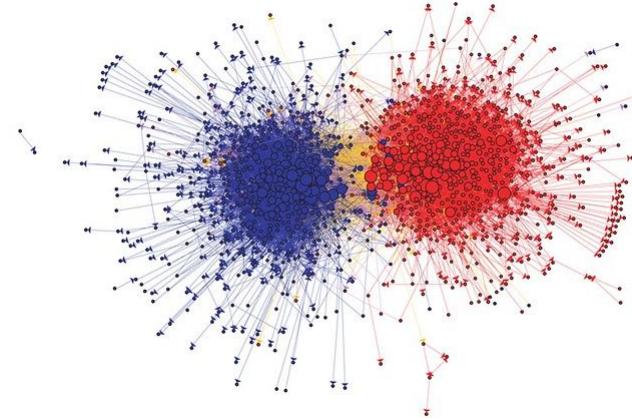
# Zentrum und Peripherie

- Core Decomposition eines Netzwerks:** k-Core eines Graphen  $G$  ist maximal induzierter Untergraph  $G_k$ , bei dem die Anzahl der Nachbarn jedes Knoten  $u$  in  $G_k$  mindestens  $k$  ist. Die Core-Zahl wird mit  $c(u)$  angegeben.
- Beispiel:**  $c(v_8)=1$ ,  $c(v_1)=2$ ,  $c(v_6)=3$ .



# Communities

- Auch manchmal Cluster genannt
- **Eigenschaften von Communities:**
  - Gegenseitigkeit
  - Kompaktheit
  - Dichte Kanten
  - Abgrenzung
- **Definition:** Netzwerk Communities sind Gruppen von Knoten, bei denen die Knoten *in* der Gruppe mit deutlich mehr Kanten verbunden sind, als mit solchen *außerhalb* der Gruppe



Visualisierung der von US political bloggers 2004  
blau: Democrat  
rot: Republican

[Adamic & Glance, 2005](#)

# Communities

- **Community Detection Problem:** Aufteilung von Graphen in Cluster von Knoten basierend auf der Netzwerkstruktur
- Verschiedene Möglichkeiten für Community Detection: random walks, spectral clustering, modularity maximization, statistische Herangehensweisen...
- Problem: Hubs können zu unterschiedlichen Communities, je nach verwendetem Algorithmus, gehören

- Graph density

$$\rho = \frac{m}{n(n - 1)/2}$$

- community internal density

$$\delta_{int}(C) = \frac{m_s}{n_s(n_s - 1)/2}$$

- external edges density

$$\delta_{ext}(C) = \frac{m_{ext}}{n_c(n - n_c)}$$

- community (cluster):  $\delta_{int} > \rho$ ,  $\delta_{ext} < \rho$

Community-Dichte

# Community Detection

## Girvan und Newman (nicht-überlappende Communities):

- Methode basierend auf “Edge Betweenness”  
(bestimmt Anzahl kürzester Pfade durch eine gegebene Kante)
- Durch diese Metrik lassen sich sukzessive Kanten mit hoher Edge Betweenness entfernen (Brücken zwischen zwei Communities)
- Problem: hohe Komplexität und Rechenaufwand für große Netzwerke

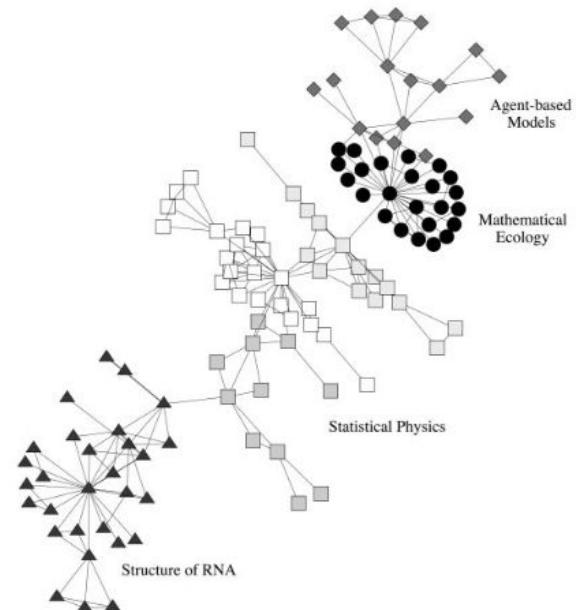


Fig. 6. The largest component of the Santa Fe Institute collaboration network, with the primary divisions detected by our algorithm indicated by different vertex shapes.

# Community Detection

- Es gibt sehr viele Algorithmen für Community Detection
- Bei der Auswahl spielt unter anderem die Laufzeit eine Rolle

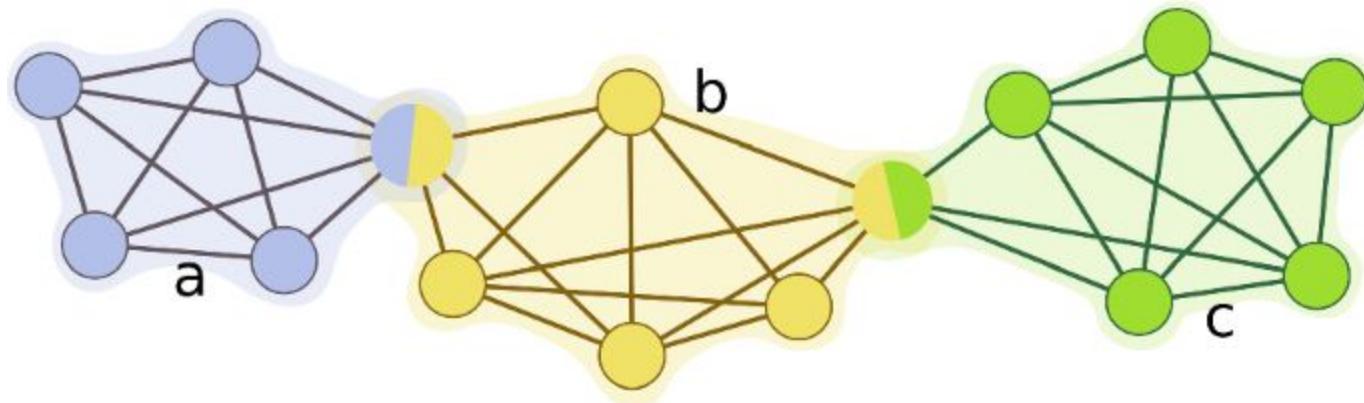
**Table 1**

List of the algorithms used in the comparative analysis of Danon et al. [238]. The first column indicates the names of the algorithm designers, the second the original reference of the work, the third the symbol used to indicate the algorithm and the last the computational complexity of the technique. Adapted from Ref. [238].

Author	Ref.	Label	Order
Eckmann & Moses	[334]	EM	$O(m\langle k^2 \rangle)$
Zhou & Lipowsky	[264]	ZL	$O(n^3)$
Latapy & Pons	[100]	LP	$O(n^3)$
Clauset et al.	[174]	NF	$O(n \log^2 n)$
Newman & Girvan	[54]	NG	$O(nm^2)$
Girvan & Newman	[12]	GN	$O(n^2 m)$
Guimerà et al.	[185,27]	SA	Parameter dependent
Duch & Arenas	[188]	DA	$O(n^2 \log n)$
Fortunato et al.	[171]	FLM	$O(m^2 n)$
Radicchi et al.	[78]	RCCLP	$O(m^4/n^2)$
Donetti & Muñoz	[48,423]	DM/DMN	$O(n^3)$
Bagrow & Bollt	[327]	BB	$O(n^3)$
Capocci et al.	[250]	CSCC	$O(n^2)$
Wu & Huberman	[336]	WH	$O(n + m)$
Palla et al.	[28]	PK	$O(\exp(n))$
Reichardt & Bornholdt	[256]	RB	Parameter dependent

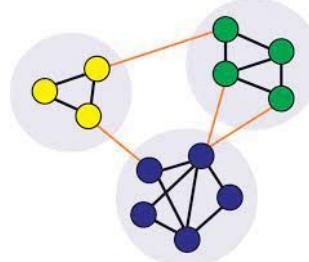
# Community Detection

- Überlappende Communities sind Communities, in denen ein Knoten zu mehr als einer Community gehören kann
- Beispiel: im sozialen Kontext ist eine Person stark vernetzt mit der eigenen Familie, aber auch mit Mitarbeiter:innen oder Mitgliedern des Sportvereins

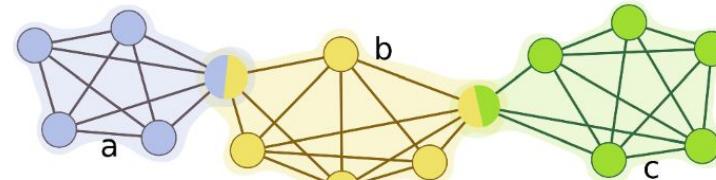


# Communities: Berechnung durch Modularität

- **Modularität** ist ein Maß, welches die Struktur von Netzwerken betrachtet, indem es misst, wie stark Netzwerke in Module (Communities) unterteilt sind
- Modularität nur für **nicht-überlappende** Communities
- Maximierung von Modularität ist generell NP-schwer, daher Approximierung



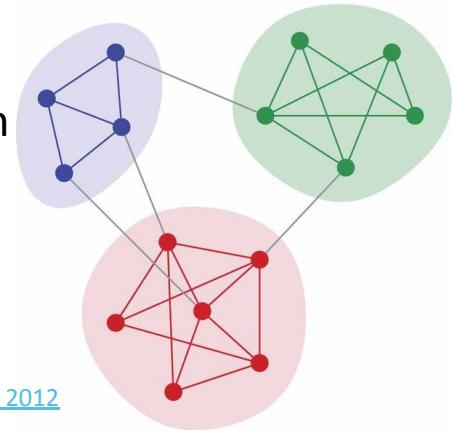
[Linhares et al. 2020](#)



[Gabardo & Beretta, 2020](#)

# Community Detection: Modularität

- Die beobachtete Menge an Kanten in den Communities wird mit der erwarteten Menge an Kanten in Communities eines zufälligen Netzwerk verglichen, wobei die Knoten denselben Grad haben
- Entfernen der Kanten führt zu “Stubs” ( $k_i$  für den Stub von Knoten i bzw Grad von i)
- Jeder Stub wird zufällig mit einem anderen Stub im Netzwerk verbunden
- Erwartete Anzahl solcher zufälligen Kanten in großen Netzwerken  $k_i k_j / 2m$   
(m die Gesamtzahl der Kanten im Netzwerk)



# Community Detection: Modularität

- Modularität = Bruchteil der Kanten, die Knoten desselben Typs verbinden - Bruchteil der Kanten, die zufällig existieren würden
- A die Adjazenzmatrix, m die Anzahl der Kanten im Netzwerk,  $k_i$  der Grad des Knotens i,  $c_i$  und  $c_j$  Communities (z.B. Community blau und rot) und die Kronecker-Delta Funktion, die 1 ist wenn  $x=y$  und 0 sonst. Beispiel:  $c_i = c_j = \text{blau}$ , dann bekommen wir 1.
- Wir schauen uns also nur Kanten an, die **in** der Community liegen
- Summe über alle Knoten in Community C ist größer oder gleich null (wenn es sich nicht um eine Community handelt, dann kann diese Summe auch negativ sein)

$$Q = \frac{1}{2m} \sum_{i,j} \left[ A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j),$$

Formel für gewichtete Netzwerke,

# Modularity

- Es gilt  $-1 \leq Q \leq 1$
- $Q > 0$  gilt, wenn Anzahl Kanten in den Gruppen größer ist als Anzahl Kanten in Gruppen eines randomly rewired Graph
- $Q > 0.3$  bedeutet signifikante Community-Struktur liegt vor (je höher  $Q$ , desto stärkere Community-Struktur)
- $Q = 0$  bedeutet, dass die Community-Struktur nicht stärker ist, als bei dem zufälligen Graphen (“randomly rewired”)

# Communities: Louvain-Methode

- Basierend auf Greedy Modularity Optimization
- Finde Partitionen mit hoher Modularität
- Skalierbar und schnell (Laufzeit von  $O(n \log n)$ ), daher für große Netzwerke geeignet
- Besteht aus zwei Phasen

# Communities: Louvain-Methode

**Start:** jeder Knoten ist in einer eigenen Community ( $n$  Knoten →  $n$  Communities)

## **Phase 1: Modularity Optimization (Optimierung der Formel für Q)**

Nehme den Knoten  $i$ . Für jeden Nachbarn  $j$  von Knoten  $i$ , schaue, ob die Modularität größer wird, wenn  $i$  von seiner Community zu Community von  $j$  verschoben wird. Der Knoten  $i$  wird dann in die Community verschoben, bei der der Zuwachs an  $Q$  maximal ist. Wenn die Änderung kleiner als 0 ist (negativ), dann bleibt  $i$  in seiner derzeitigen Community. Wiederhole so lange, bis alle Knoten eine Änderung von 0 haben. Ein Knoten kann mehrmals verschoben werden.

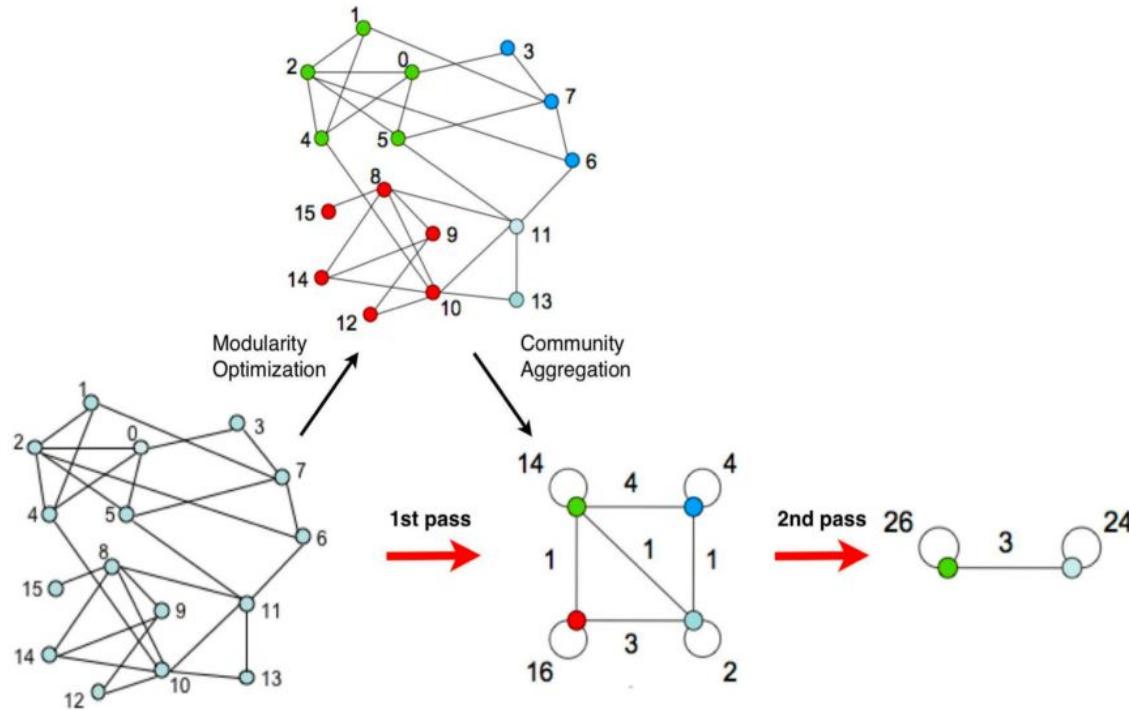
→ Erste Phase fertig, wenn lokales Maximum von Modularity erreicht wurde, also keine einzelne Verschiebung Modularity verbessern kann

# Communities: Louvain-Methode

## Phase 2: Community Aggregation

- Erstelle ein gewichtetes Netzwerk von Communities aus Phase 1.
  - Neue Knoten → Communities aus Phase 1 (d.h. Knoten werden zu einem zusammengefasst)
  - Neue Kantengewichte → Summe von Gewichten von Kanten zwischen Communities
  - Kanten in derselben Community werden zu Schleifen\*<sup>2</sup> im neuen Netzwerk
- Nach Beendigung der zweiten Phase, kann Phase 1 auf das neue Netzwerk angewendet werden. Wiederhole Phase 1 und 2 bis keine Änderung in der Modularität mehr stattfindet.

# Beispiel Louvain Methode



# Communities vs. Core-Periphery Strukturen

- Beide können Einflussbereiche und asymmetrische Interaktionen sichtbar machen
- Communities zeigen gut die Grenzen einer Untergruppe auf
- Core-Periphery Strukturen beziehen sich eher auf die interne Struktur einer Untergruppe

# Ausblick

- Latent Space Models: Um nicht-lineare Beziehungen zu erkennen

Vielen Dank für die Aufmerksamkeit!

Gibt es Fragen?