

Drama Mining

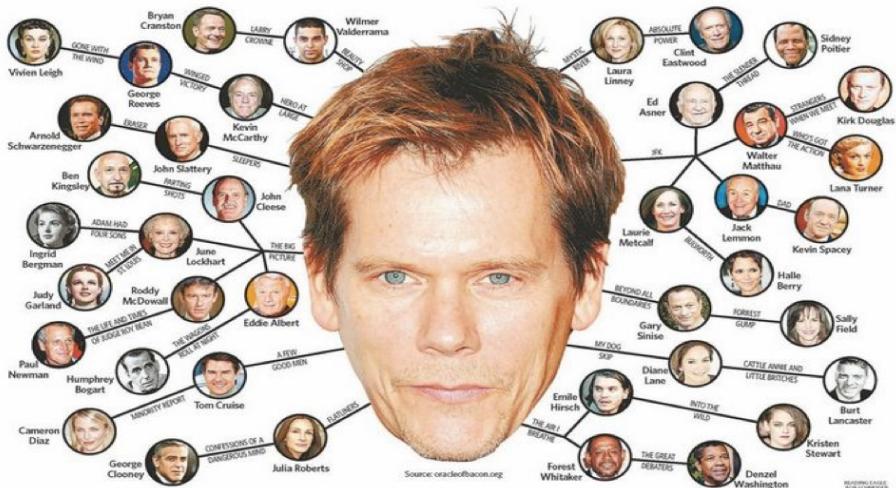
Social Network Analysis / Netzwerkanalyse I

Dr. Jochen Tiepmar

Institut für Informatik
Abteilung Computational Humanities
Universität Leipzig

13. Mai 2020

[Letzte Aktualisierung: 06/05/2020, 18:49]



6 Degrees of Separation

- 2 beliebige Menschen sind in max. 5 Schritten verbunden
- Soziologisches Phänomen (Vor Internet)
- Empirisch beobachtet aber nicht gesichert (und auch leicht widerlegbar)
- Gesellschaftsspiel 6 Degrees of Kevin Bacon
- Bacon-Zahl, Kevin-Bacon-Zahl (KBZ): Länge der kürzesten Verbindungskette von SchauspielerInnen zu Kevin Bacon
- Ähnlich für Ko-Autorenschaft: Erdős-Zahl

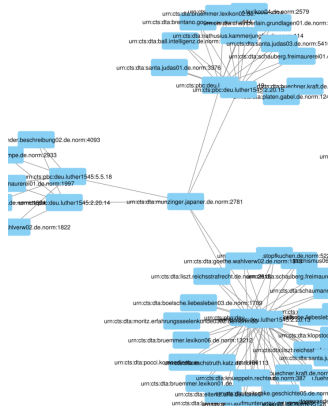
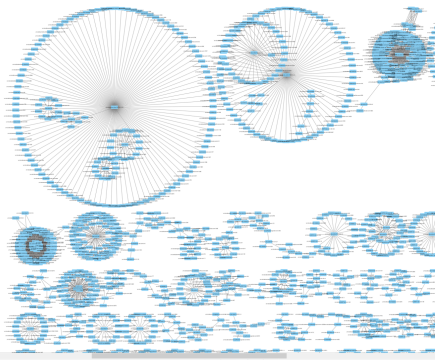


Facebook Relationships

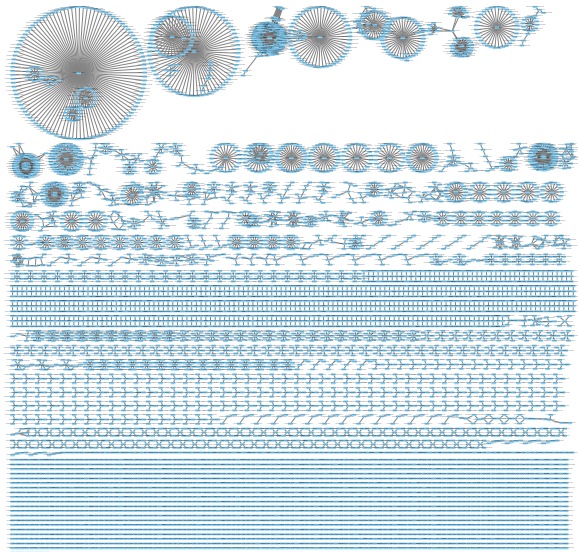


- Verbindung zwischen Geolocations zweier Social Media Freunde
- Indikator für freien Internetzugang

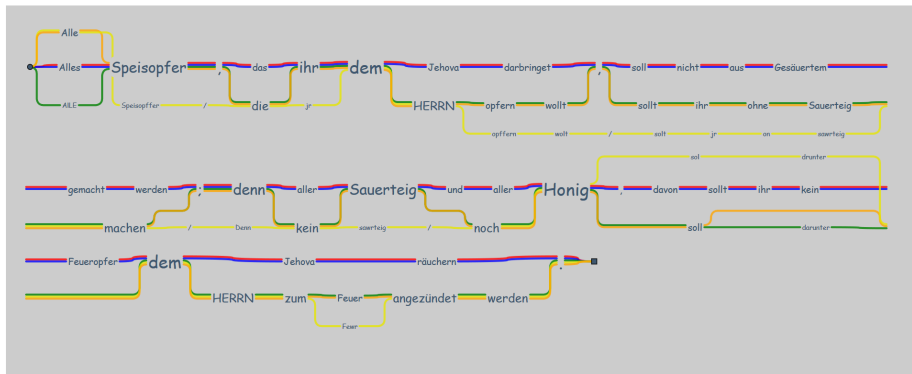
Zitatgraphen



Zitatgraphen



Textvarianzgraphen (TRAViz)



Überblick

- 1 Graphentheorie
- 2 Visualisierung
- 3 Graphlayout
- 4 Netzwerkanalyse
- 5 Netzwerkanwendungen

Überblick

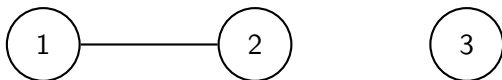
- 1 Graphentheorie
- 2 Visualisierung
- 3 Graphlayout
- 4 Netzwerkanalyse
- 5 Netzwerkanwendungen

Arten von Graphen

$G = (V, E)$ ist ein Graph wenn:

- $V =$ eine endliche Menge von Elementen (Knoten)
- $E =$ eine Menge von Paaren aus Elementen von V (Kanten)

Beispiel $V = \{1, 2, 3\}, E = \{\{1, 2\}\}$



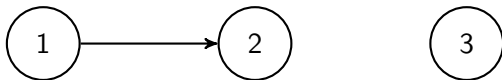
Arten von Graphen

$G = (V, E)$ ist ein gerichteter Graph wenn:

- V = eine endliche Menge von Elementen (Knoten)
- E = eine Menge von geordneten Paaren aus Elementen von V (Kanten)

Genau genommen ist ein ungerichteter Graph und ein bidirektional gerichteter äquivalent

Beispiel $V = \{1, 2, 3\}, E = \{(1, 2)\}$

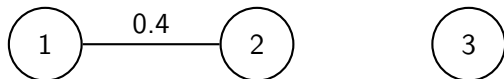


Arten von Graphen

$G = (V, E, w)$ ist ein gewichteter Graph wenn:

- V = eine endliche Menge von Elementen (Knoten)
- E = eine Menge von Paaren aus Elementen von V (Kanten)
- $w = E \rightarrow \mathbb{R}$ eine Zuordnung von Kanten zu Zahlen (Gewicht, Länge)

Beispiel $V = \{1, 2, 3\}$, $E = \{\{1, 2\}\}$, $w(\{1, 2\}) = 0.4$

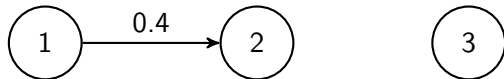


Arten von Graphen

$G = (V, E, w)$ ist ein gewichteter gerichteter Graph wenn:

- V = eine endliche Menge von Elementen (Knoten)
- E = eine Menge von geordneten Paaren aus Elementen von V (Kanten)
- $w = E \rightarrow \mathbb{R}$ eine Zuordnung von Kanten zu Zahlen (Gewicht, Länge)

Beispiel $V = \{1, 2, 3\}$, $E = \{(1, 2)\}$, $w((1, 2)) = 0.4$



Arten von Graphen

$G' = (V', E')$ ist ein Teilgraph von $G = (V, E)$ wenn

- $V' \subseteq V$ (Alle Knoten aus G' sind in G enthalten)
- $E' \subseteq E$ (Alle Kanten aus G' sind in G enthalten)

Beispiel $V = \{1, 2, 3\}$, $E = \{(1, 2), (2, 3)\}$, $V' = \{1, 2\}$, $E' = \{(1, 2)\}$



Arten von Graphen

Jeder maximal zusammenhängende (verbundene) Teilgraph eines Graphen G ist eine Komponente.

Beispiel $V = \{1, 2, 3\}$, $E = \{(1, 2)\}$ hat 2 Komponenten K_1 und K_2 :

- $K_1 = \{1, 2\}$, $E = \{(1, 2)\}$
- $K_2 = \{3\}$, $E = \{ \}$

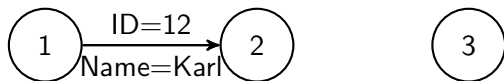


Arten von Graphen

Ein gerichteter Graph ist ein Propertygraph wenn

- V = eine endliche Menge von Elementen (Knoten)
- E = eine Menge von geordneten Paaren aus Elementen von V (Kanten)
- P_E = eine Menge von Zuordnungen von Key-Value Mappings zu Kanten
- (P_V = eine Menge von Zuordnungen von Key-Value Mappings zu Knoten)

Beispiel $V = \{1, 2, 3\}$, $E = \{(1, 2)\}$ $P_E((1, 2)) = \{ID = 12, Name = Karl\}$

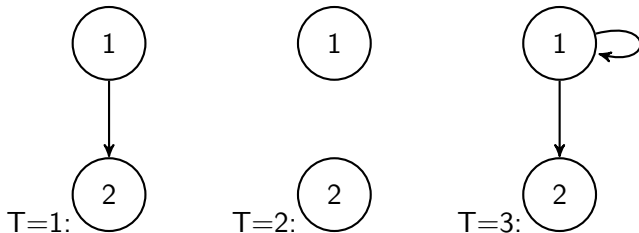


Arten von Graphen

Ein (un)gerichteter Graph ist ein Temporaler Graph wenn

- V = eine endliche Menge von Elementen (Knoten)
- E = eine Menge von Tripeln aus möglicherweise redundanten Knotenpaaren aus V und Zeitpunkten $T \in \mathbb{N}$ (oder bei Bedarf auch $T \in \mathbb{R}$)

Beispiel $V = \{1, 2\}$, $E = \{(1, 2, 1), (1, 2, 3), (1, 1, 3)\}$

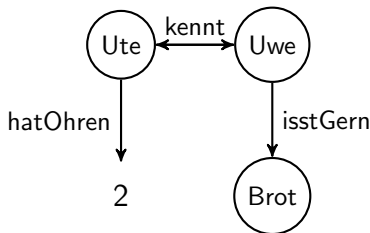


Arten von Graphen

Ein gerichteter Graph ist ein semantischer Graph wenn

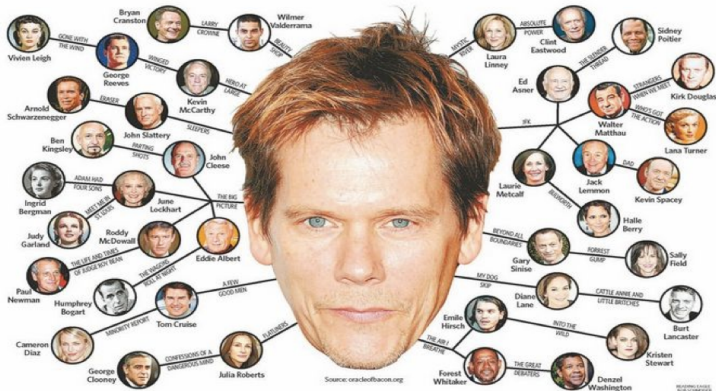
- S = eine endliche Menge von Entitäten (Subjekten)
- O = eine endliche Menge von Entitäten und Literalen (Objekten)
- R = eine Menge von Relationen
- E = eine Menge von Tripeln aus je einem Element aus S , O und R

Beispiel $S = \{Uwe, Ute\}$, $O = \{Eis, Brot\}$, $R = \{isstGern, hatOhren, kennt\}$,
 $E =$
 $\{(Ute, hatOhren, "2"), (Uwe, isstGern, Brot), (Ute, kennt, Uwe), (Uwe, kennt, Ute)\}$



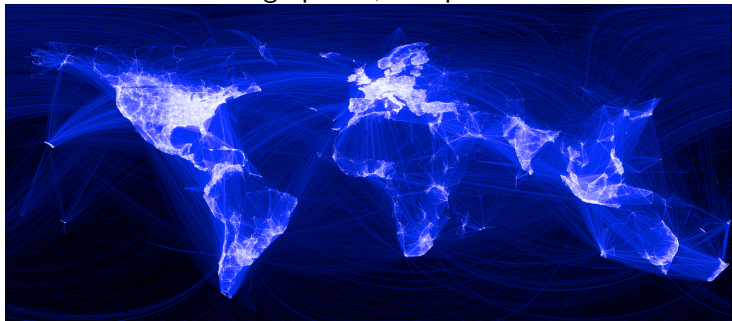
Übung Graphbeschreibung

Gerichtet?, Gewichtet?, Semantisch?, Property?, Komponenten?,
Teilgraphen?, Temporal?



Übung Graphbeschreibung

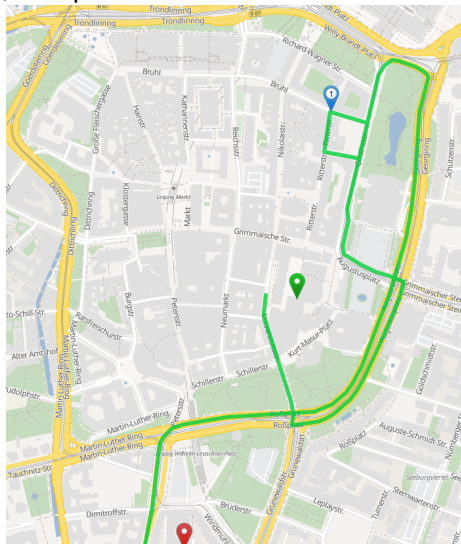
Gerichtet?, Gewichtet?, Semantisch?, Property?, Komponenten?,
Teilgraphen?, Temporal?



Übung Graphbeschreibung

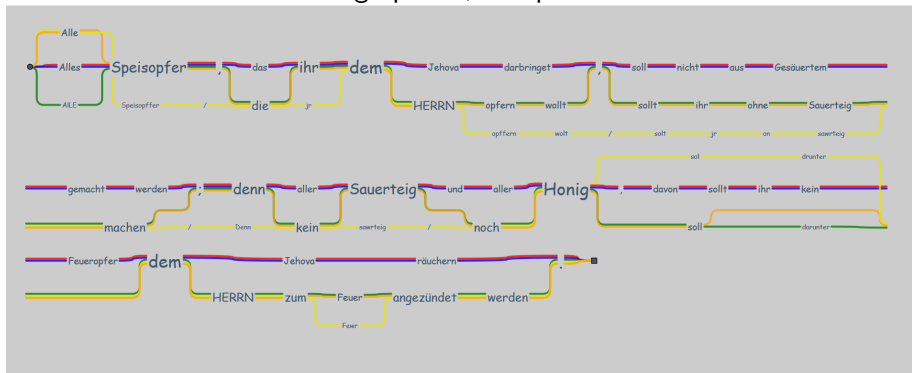
Gerichtet?, Gewichtet?, Semantisch?, Property?, Komponenten?,
Teilgraphen?, Temporal?

- Straßen als Verbindungen zwischen Orten
- Routenfindung als Berechnung der besten Verbindung (kurz, schnell, ...)



Übung Graphbeschreibung

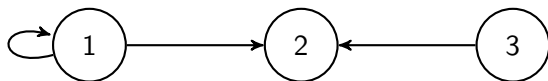
Gerichtet?, Gewichtet?, Semantisch?, Property?, Komponenten?,
Teilgraphen?, Temporal?



Knoten

Ein Knoten x hat

- Vorgänger = Knoten mit Kanten (\cdot, x) (nur in gerichtetem Graph)
- Nachfolger = Knoten mit Kanten (x, \cdot) (nur in gerichtetem Graph)
- Nachbarn = Nachfolger \cup Vorgänger
- Eingangsgrad = Anzahl der Vorgänger (nur in gerichtetem Graph)
- Ausgangsgrad = Anzahl der Nachfolger (nur in gerichtetem Graph)
- Grad = Eingangsgrad + Ausgangsgrad



Knoten	Eingangsgrad	Ausgangsgrad	Vorgänger	Nachfolger
1	1	2	{1}	{1,2}
2	2	0	{1,3}	{}
3	0	1	{}	{2}

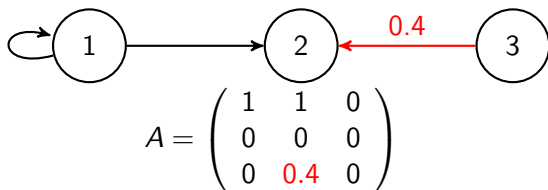
Adjazenzmatrix

Ein Graph $G = (V, E)$ mit n vielen Kanten wird in einer Boole'schen $n \times n$ -Matrix $A = (a_{ij})$ gespeichert, wobei

$$a_{ij} = \begin{cases} 1 & \text{falls } (i, j) \in E \\ 0 & \text{sonst} \end{cases}$$

$$A = \begin{pmatrix} A_{(1,1)} & A_{(1,2)} \\ A_{(2,1)} & A_{(2,2)} \end{pmatrix}$$

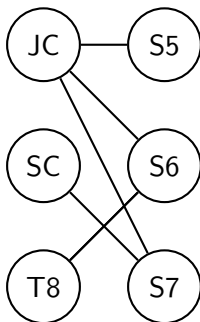
- Ungerichtete Graphen ergeben immer eine symmetrische Adjazenzmatrix
- Statt boolean (1, 0) können auch **Kantengewichte** eingetragen werden



Adjazenzmatrix & Konfigurationsmatrix

- Konfigurationsmatrix ist ein bipartiter ungerichteter Graph, der Personen auf Szenen abbildet (Matching)

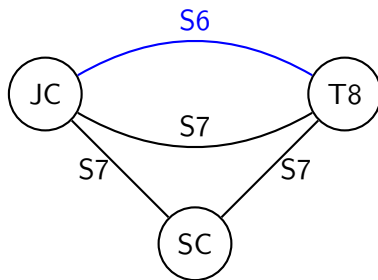
Person	...	Szene 5	Szene 6	Szene 7	...
John Connor		1	1	1	
Sarah Connor		0	0	1	
T-800		0	1	1	



Adjazenzmatrix & Konfigurationsmatrix

- Aus Konfigurationsmatrix lässt sich ein ungerichteter und/oder gewichteter und/oder temporaler Graph ableiten

Person	...	Szene 5	Szene 6	Szene 7	...
John Connor		1	1	1	
Sarah Connor		0	0	1	
T-800		0	1	1	



Adjazenzmatrix & Konfigurationsmatrix

Marcus, S. (1979): *Mathematische Poetik*

- Konfigurationsmatrix ist ein bipartiter ungerichteter Graph, der Personen auf Szenen abbildet (Matching)
- Aus Konfigurationsmatrix lässt sich ein ungerichteter und/oder gewichteter und/oder temporaler Graph ableiten
- Konfigurationsmatrix eröffnet Drama- und Filmanalysen mittels Graphalgorithmen sowie Graphvisualisierungen

Adjazenzmatrix & Konfigurationsmatrix

Marcus, S. (1979): *Mathematische Poetik*

- Konfigurationsmatrix ist ein bipartiter ungerichteter Graph, der Personen auf Szenen abbildet (Matching)
- Aus Konfigurationsmatrix lässt sich ein ungerichteter und/oder gewichteter und/oder temporaler Graph ableiten
- Konfigurationsmatrix eröffnet Drama- und Filmanalysen mittels Graphalgorithmen sowie Graphvisualisierungen

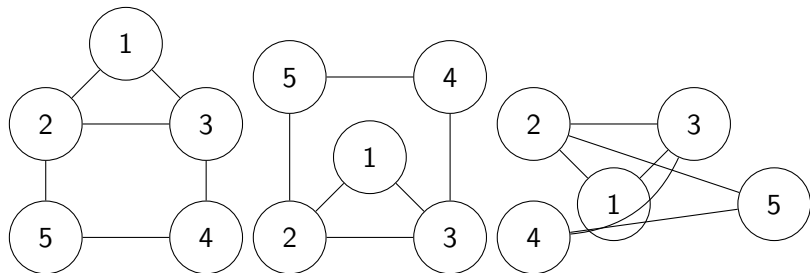
Konfigurationsmatrizen



Visualisierung

Solange nicht explizit anders angegeben ohne inhaltliche Aussagekraft:

- Knotenpositionen und Knotennähe
- Kantenschwünge
- Kantenüberschneidung
- Farbe, Form
- Kantenlänge



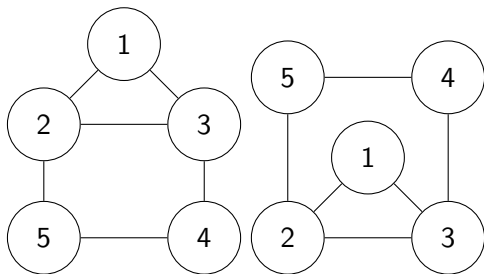
Ästhetikkriterien

- Kantenschnitte: Minimierung der Anzahl der Schnittpunkte von Kanten (Optimalfall Planarität nicht immer erreichbar)
 - bei K4 ja, bei K5 nein)
- Knicke
 - Möglichst wenig geknickte Kanten
 - Möglichst wenig Knicke in Kanten
 - Möglichst geringe Varianz an Knicken/Kante
- Winkelauflösung: Kleinster auftretender Winkel zwischen zwei Kanten wird maximiert
- Seitenverhältnis: Seitenverhältnis des kleinsten umschriebenen Rechtecks soll dem Seitenverhältnis der Zeichenebene entsprechen
- Symmetrie: Symmetrien des Graphen sollen möglichst gut abgebildet werden

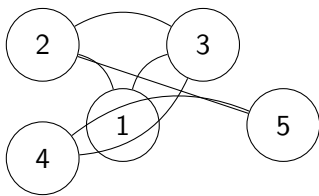
Ästhetikkriterien

- Kantenschnitte: Minimierung der Anzahl der Schnittpunkte von Kanten (Optimalfall Planarität nicht immer erreichbar)
→ bei K4 ja, bei K5 nein)
- Knicke
 - Möglichst wenig geknickte Kanten
 - Möglichst wenig Knicke in Kanten
 - Möglichst geringe Varianz an Knicken/Kante
- Winkelauflösung: Kleinster auftretender Winkel zwischen zwei Kanten wird maximiert
- Seitenverhältnis: Seitenverhältnis des kleinsten umschriebenen Rechtecks soll dem Seitenverhältnis der Zeichenebene entsprechen
- Symmetrie: Symmetrien des Graphen sollen möglichst gut abgebildet werden
- Graph Drawing generell breiter Forschungsbereich (→ Abteilung für Bild- und Signalverarbeitung → Informationsvisualisierung)

Objektiv und wissenschaftlich ästhetisch:



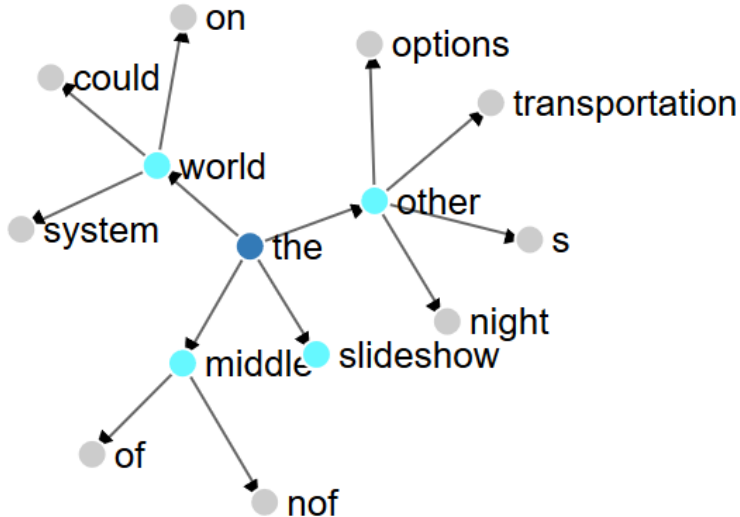
Objektiv und wissenschaftlich unästhetisch:



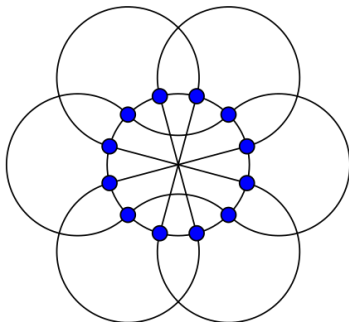
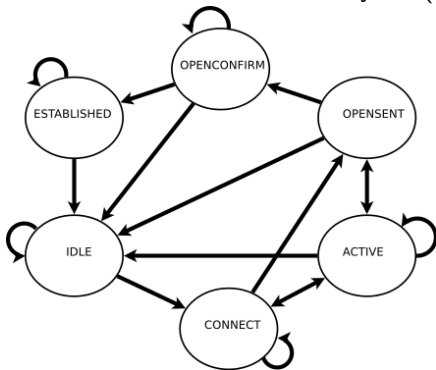
Überblick

- 1 Graphentheorie
- 2 Visualisierung
- 3 Graphlayout**
- 4 Netzwerkanalyse
- 5 Netzwerkanwendungen

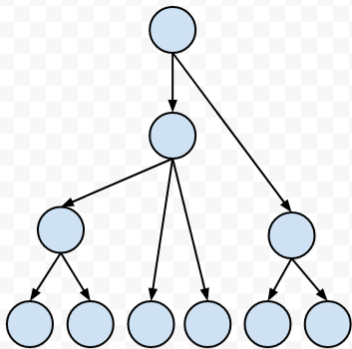
Kräftebasiert (Force Directed)



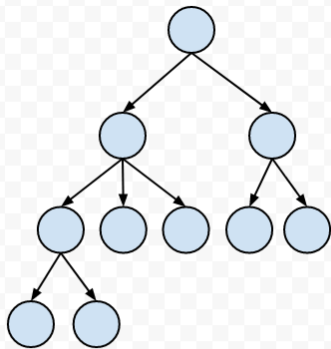
Kreislayout (Circular)



Hierarchisch / Layered

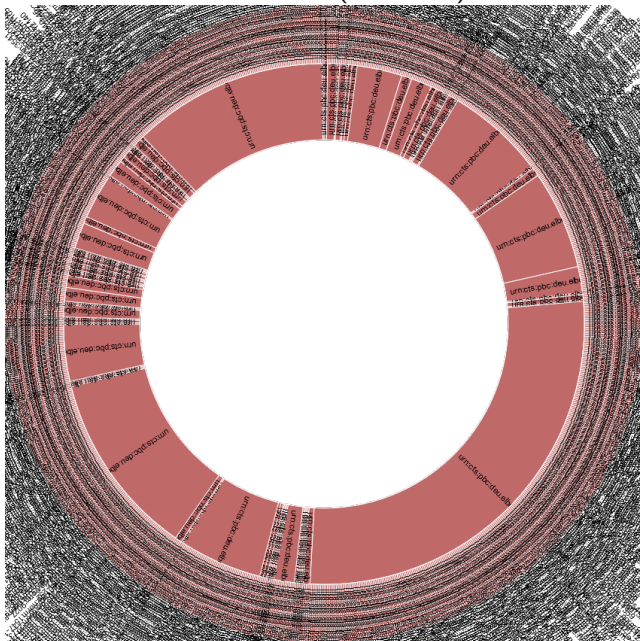


“stacked”

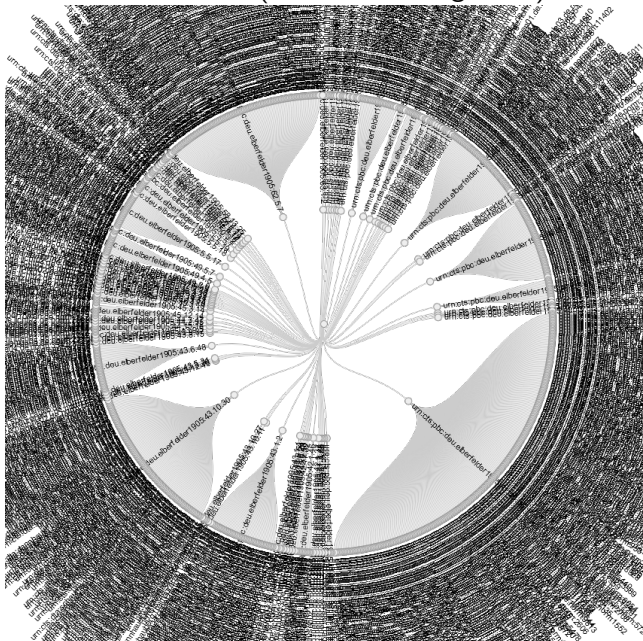


“hanging”

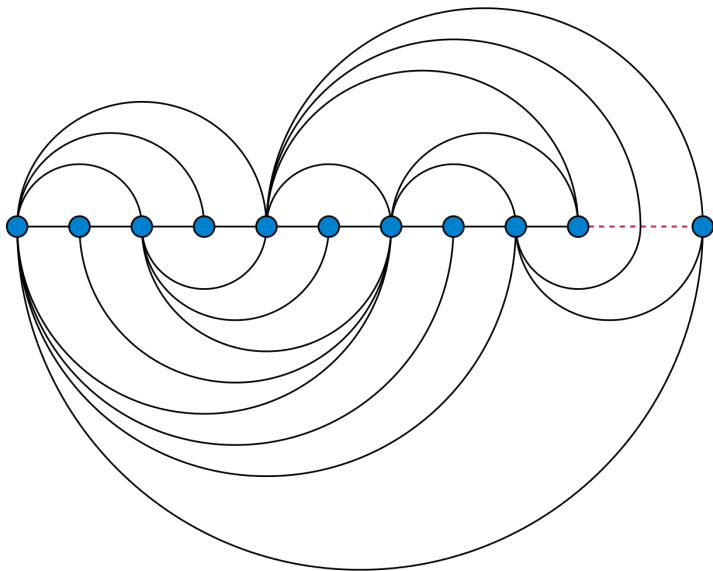
Hierarchisch (Sunburst)



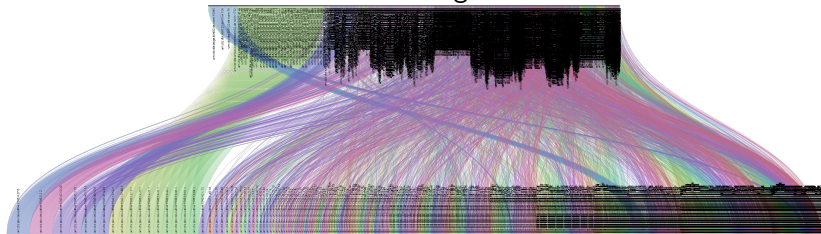
Hierarchisch (Circular Dendrogramm)



Bogenlayout (Arc)



Alluvial Diagram



Überblick

- 1 Graphentheorie
- 2 Visualisierung
- 3 Graphlayout
- 4 Netzwerkanalyse**
- 5 Netzwerkanwendungen

Nächste Woche

- Pagerank
- Clustering
- ...

Überblick

- 1 Graphentheorie
- 2 Visualisierung
- 3 Graphlayout
- 4 Netzwerkanalyse
- 5 Netzwerkanwendungen**

Nächste Woche

- Social Network Analysis
- Zitatanalyse / Textvergleich
- Modellierung von Argumenten / SNA-Feindschaften
- Semantifizierung / Semantic Web

Zusammenfassung

- Arten von Graphen
 - Gerichtet, Gewichtet
 - Teilgraphen, Komponenten
 - Property, Semantisch, Temporal
- Drama Mining → Konfigurationsmatrix → Graphanalyse
- Ästhetik von Graphen, Visualisierung vs. Interpretation
- Graphlayouts

Links

<http://www.markrobinsonwrites.com/the-music-that-makes-me-dance/2018/3/11/movie-morsel-six-degrees-of-kevin-bacon>
<https://twiki.org/p/pub/Blog/BlogEntry201207x2/fb-relationships-full.png>
<https://graphhopper.com/maps/>
<http://ctstm.informatik.uni-leipzig.de:8080/trvis/>
https://www.informatik.uni-leipzig.de/bsv/homepage/sites/default/files/Infovis-4-Graphen_4.pdf
<https://www.mofgym.co.uk/wp-content/uploads/2017/02/thumb-up-terminator-pablo-M-R.jpg>
https://en.wikipedia.org/wiki/Circular_layout#/media/File:Chvatal_Lombardi.svg
https://en.wikipedia.org/wiki/Circular_layout#/media/File:BGP_FSM_3.svg
https://en.wikipedia.org/wiki/Graph_drawing#/media/File:Goldner-Harary-linear.svg
<https://i.stack.imgur.com/J39Hd.png>