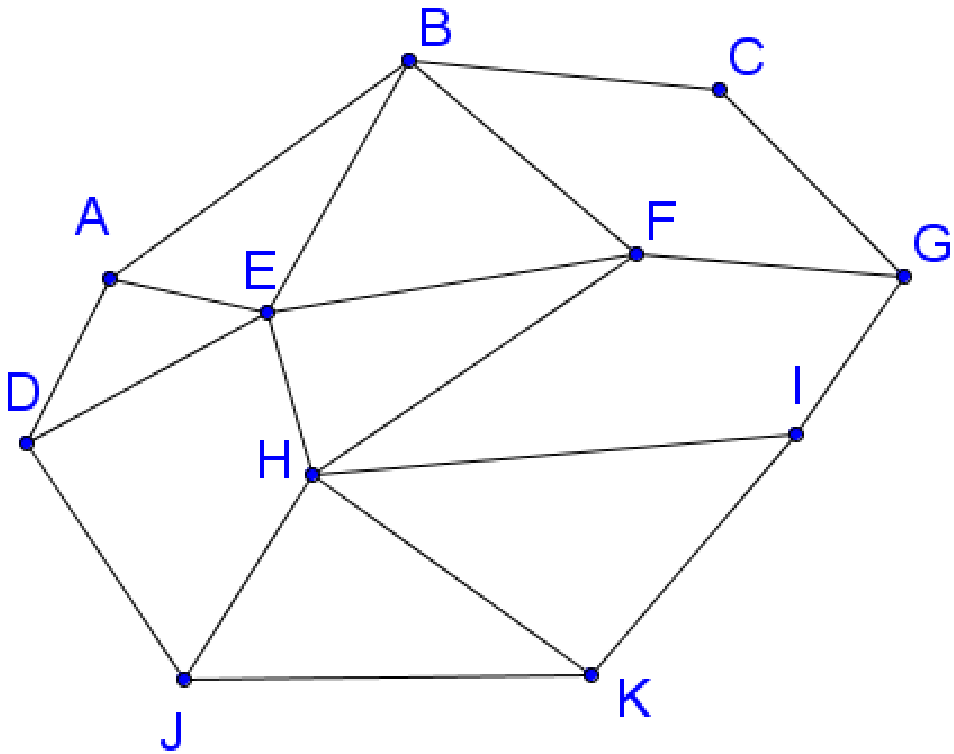
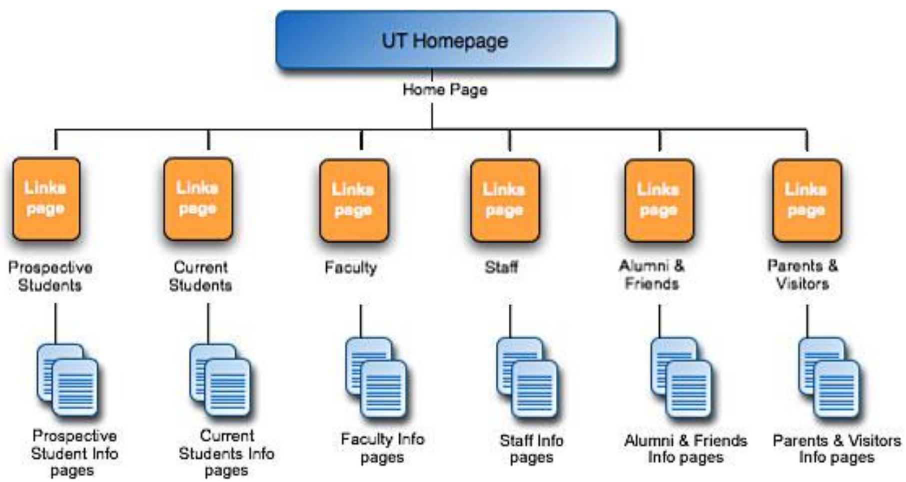
Erläuterungen zum Selbststudium



Hypertext-Struktur

- HTML-Seiten enthalten Hyperlinks zu anderen HTML-Seiten oder sich selbst
- Hyperlinks drücken eine Relation aus (Quelle -> Ziel)
- Sitemaps geben diese Relationen/Struktur grafisch wieder



?

Welche Beschränkung hat diese grafische Darstellung?

Definition von Graphen

Ein Graph G = (V, E) ist eine Datenstruktur mit folgenden Eigenschaften:

- V <> 0 ist eine endliche, nichtleere Menge, die "Knotenmenge".
  Die Elemente $v \in V$ werden als Knoten bezeichnet.
  V steht für vertices, englisch für "Eckpunkt, Knoten".
- $E \subseteq V$ x V ist eine Menge von Knotenpaaren, die "Kantenmenge".
  Die Elemente $(u,v) \in E$ werden als Kante bezeichnet; Elemente $(v,v) \in E$ als Schlinge oder (direkter) Zyklus.
  E steht für edges, also "Kanten"

Gerichteter und ungerichteter Graph

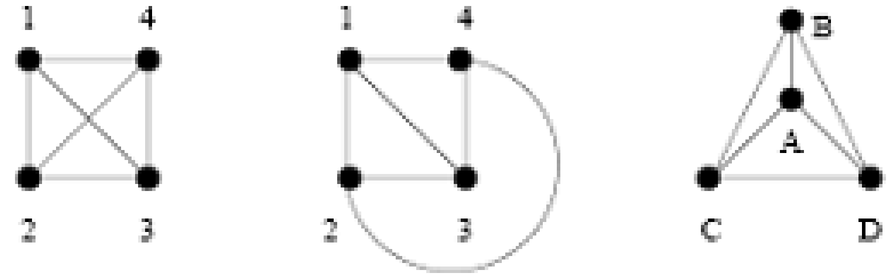
- In einem ungerichteten Graphen ist die Anordnung der Knoten innerhalb eines Knotenpaars irrelevant: (u,v,) = (v,u) . Beim ungerichteten Graph gelten Knoten u, v als benachbart (adjazent), wenn $(u,v) \in E \vee (v,u) \in E$ gilt.
- In einem gerichteten Graph hingegen gibt (u, v) an, dass eine Kante von u nach v existiert.
  u wird dann als die Quelle oder Startknoten, v als Endknoten der Kante (u,v) bezeichnet und auch als $u \rightarrow v$ notiert.
- Teilen sich zwei Kanten einen Endknoten, so sind diese inzident
- Der Grad deg(a) eines Knotens a ist die Anzahl der Kanten, die inzident mit dem Knoten sind (Anzahl der Kanten, die vom Knoten "ausgehen")
- Einen Knoten, dessen Grad gleich 0 ist, bezeichnet man als isoliert

Darstellung von Graphen

$G = (V, E)$ mit $V_G = \{1, 2, 3, 4\}$, $E_G = \{(1, 2), (1, 3), (1, 4), (2, 3), (2, 4), (3, 4)\}$

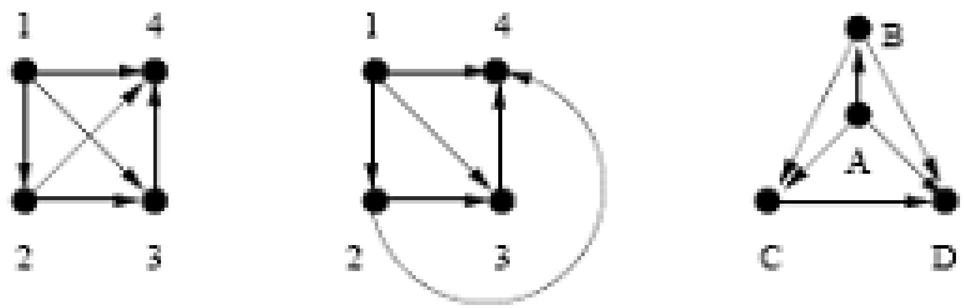
$H = (V, E)$ mit $V_H = \{A, B, C, D\}$, $E_H = \{(A, B), (B, C), (C, D), (A, D), (A, C), (B, D)\}$

Bei Interpretation als ungerichtete Graphen:



Darstellungen von G   Darstellung von H

Bei Interpretation als gerichtete Graphen:



Darstellungen von G   Darstellung von H

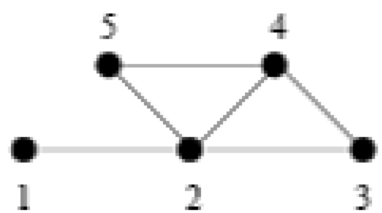
Kantenfolge, Kantenzug und Weg

Sei G = (V,E) Graph k = $\{v_0, v_1, ..., v_n\}$ eine Folge von n + 1 Knoten von G.

- k heißt Kantenfolge der Länge n von v_0 nach v_n, wenn für $0 <= i <= n - 1$gilt: $(v_i, v_{i+1}) \in E$.
  Falls G gerichtet ist, ist v_0 der Startknoten und v_n der Endknoten, andernfalls sind beide Endknoten der Kantenfolge k.
- k ist ein Kantenzug der Länge n von v_0 nach v_n, falls k eine Kantenfolge der Länge n von v_0 nach v_n ist und für $0 <= i,j <=n - 1$ gilt: $(v_i, v_{i+1}) \neq (v_j, v_{j+1})$.
  Jede Kante darf in einem Kantenzug also maximal einmal auftreten.
- k ist ein Weg der Länge n von v_0 nach v_n, wenn k eine Kantenfolge der Länge n von v_0 nach v_n ist und für $0 <= i,j <=n$ gilt: $i \neq j$ gilt $v_i \neq v_j$.
  In einem Weg darf jeder Knoten also nur maximal einmal auftreten.
- Ein Graph ist zusammenhängend, wenn es für jedes Paar $i, j \in V$ einen Weg von i nach j gibt.

Zyklus, Länge und Beispiel

- k heißt Zyklus der Länge n, wenn k eine geschlossene Kantenfolge der Länge v0 von nach v_n ist (d. h. $v_0 = v_n$) und k' = (v_0,..., v_{n-1}) ein Weg ist.
  Graphen ohne Zyklen heißen azyklisch.
- |k| gibt die Länge von k an.
- Ist k eine Kantenfolge von v nach w, so gibt es immer auch einen Weg von w nach v



(1, 2, 3, 4, 5, 2, 3) ist *Kantenfolge* der Länge 6 von 1 nach 3

(1, 2, 5, 4, 2, 3) ist *Kantenzug* der Länge 5 von 1 nach 3

(1, 2, 5, 4, 3) ist *Weg* der Länge 4 von 1 nach 3

(2, 3, 4, 5, 2) ist *Zyklus* der Länge 4

Speicherung von Graphen

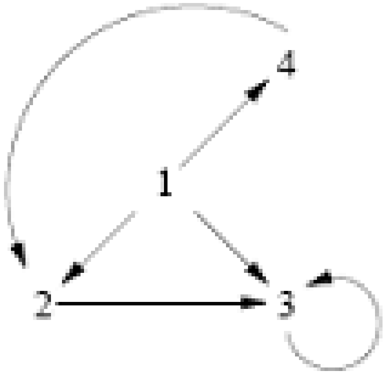
Ein Graph G = (V, E) mit |V| = n wird üblicherweise auf eine der folgenden Arten gespeichert:

- Speicherung in einer Adjazenzmatrix
- Speicherung in einer Adjazenzliste
- Speicherung in einer doppelt verketteten Kantenliste

Speicherung in einer Adjazenzmatrix

- Die Matrix hat die Größe n x n. Der Eintrag an Position a_{ij} ist dabei genau dann true bzw. 1, wenn die Kante (i, j) in E enthalten ist.
- Der Speicherplatzbedarf einer Adjazenzmatrix hängt nur von der Anzahl Knoten ab, nicht von der Anzahl Kanten.
- Es ergibt sich immer ein Platzbedarf von n^2.
- Bei ungerichteten Graphen ist die Matrix symmetrisch, d.h. die Werte an Position (i, j) und (j, i) sind identisch.
  - Statt einer quadratischen Matrix könnte also auch eine (untere) Dreiecksmatrix verwendet werden n^2 / 2 Einträgen.

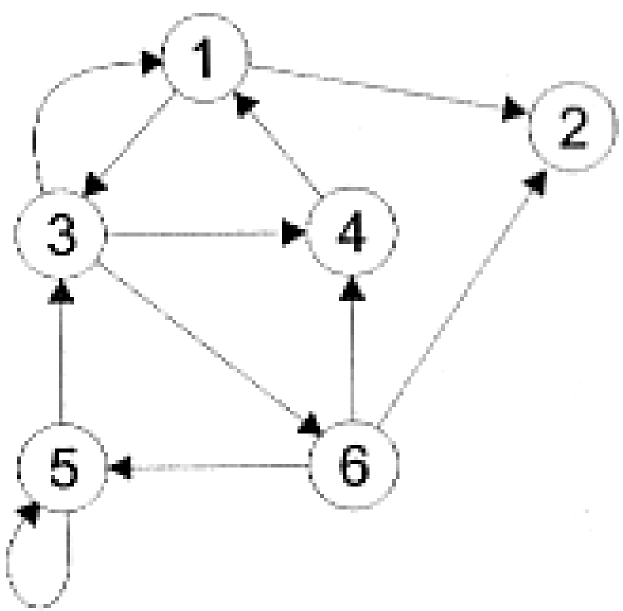
Beispiel zur Speicherung



	1	2	3	4	5	6
1	0	1	1	1	0	0
2	0	0	1	0	0	0
3	0	0	1	0	0	0
4	0	1	0	0	0	0
5	0	0	0	0	0	0
6	0	0	0	0	1	0

Graph G

Adjazenzmatrix von G



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

Gewichtung von Kanten

Bislang war die Anordnung der Knoten beliebig wählbar. In einigen Fällen ist dies aber nicht zulässig, z.B. in den folgenden Fällen:

- Modellierung von geometrischen Punkten als Koordinaten
  Hier wird jedem Punkt eine feste Koordinate zugeordnet; die Anordnung der Knoten ist damit fest vorgegeben; es kann höchstens eine Skalierung der Knoten vorgenommen werden.
  Beispiel: Darstellung von CAD-Zeichnungen als Graph.
- Gewichtete Kanten
  Hierbei geht es nicht nur um die Frage "existiert eine Kante", sondern um das Gewicht der Kante (in der Regel > 0).
  Beispiel: Kostenfunktionen für die Verfolgung einer Kante; Verzögerungszeiten

Gewichtung der Kanten, Adjazenzmatrix

Internationale Entfernungstabelle

Die Entfernungstabelle Deutschland (gelb) und Europa (blau) können nur wie angegeben miteinander verknüpft werden.

Die rot markierten Städte ermöglichen die Berechnung von Entfernungen zwischen Orten in Deutschland und Europa (siehe Bsp.)	Amsterdam	Athen	Barcelona	Belgrad	Berlin	Bern	Brüssel	Budapest	Bukarest	Dublin	Frankfurt M.	Hamburg	Helsinki	Istanbul	Köln	Kopenhagen	Lissabon	London	Madrid	Minsk	Moskau	München	Oslo	Paris	Prag	Rom	Sofia	Stockholm	Warschau	Wien	Entfernungen Europa – Orte aus Deutschland-tabelle in rot
Aachen		2950	1600	1800	650	950	200	1400	2220	950	450	450	1700	2700	260	800	2350	500	1800	1800	2500	850	1300	500	900	1700	2100	1450	1200	1150	Amsterdam
Basel	545		3200	1150	2400	2700	2900	1500	1150	3600	2500	2750	3650	1100	2700	2750	4500	3200	3800	2400	2900	2150	3450	2900	2050	900	750	3450	2200	1750	Athen
Berlin	650	875		2050	1900	950	1350	2000	2600	2000	1350	1750	3050	2950	1400	2100	1300	1500	650	2900	3600	1400	2650	1100	1750	1400	2400	2750	2500	1900	Barcelona
Bremen	370	775	400		1300	1450	1750	400	600	2500	1300	1700	2500	950	1500	1700	3350	2050	2670	1400	2200	950	2300	1800	900	1300	400	2300	1100	650	Belgrad
Dortmund	155	555	495	235		950	750	900	1700	1550	550	300	1250	2200	580	350	2900	1050	2350	1150	1850	590	950	1050	400	1550	1650	950	600	650	Berlin
Dresden	645	745	200	490	515		700	1150	1950	1250	400	950	2050	2300	600	1350	2200	1000	1550	2000	2700	450	1800	550	850	1000	1750	1900	1600	1000	Bern
Düsseldorf	90	550	560	285	70	580		1400	2200	800	450	600	1900	2650	250	950	2150	300	1550	1900	2600	800	1550	300	950	1550	2100	1650	1400	1150	Brüssel
Emden	375	845	520	140	305	620	290		800	2150	1000	1200	2250	1350	1150	1200	3300	1700	2600	1200	1900	700	1900	1550	500	1250	800	1900	700	250	Budapest
Erfurt	440	585	300	340	310	215	375	470		3000	1800	2000	2500	700	2000	2050	3950	2500	3200	1300	1800	1500	2850	2300	1350	1950	400	3150	1200	1050	Bukarest
Flensburg	625	980	450	275	490	660	540	400	520		1200	1350	2750	3500	1000	1700	2750	500	2200	2700	3400	1550	2250	900	1700	2350	2950	2500	2100	1900	Dublin
Frankfurt M.	255	335	550	445	225	460	225	520	260	650		500	1600	2250	200	800	2450	700	1900	1650	2350	400	1350	600	500	1300	1700	1500	1050	700	Frankfurt M.
Frankfurt O.	700	940	105	460	560	180	625	590	370	550	610		1050	2600	430	300	2800	900	2200	1450	2200	780	850	950	700	1750	2050	950	850	1100	Hamburg
Garm.-Patenk.	700	370	675	835	700	550	680	930	510	1020	480	740		3250	1600	850	3950	2100	3400	1250	1150	1850	750	2100	1550	2900	2950	200	1500	1850	Helsinki
Görlitz	750	840	220	575	610	105	680	700	320	690	560	170	650		2450	2650	4250	2950	3500	2000	2500	1900	3500	2700	1850	1600	550	3300	1850	1600	Istanbul
Hamburg	480	825	300	130	350	500	400	255	370	160	500	385	860	530		750	2400	550	1800	1550	2250	580	1150	500	700	1500	1900	1300	1150	900	Köln
Hannover	355	680	290	130	215	385	280	260	220	310	350	350	720	470	155		3100	1250	2500	1550	2250	1100	600	1250	750	2050	2100	650	900	1000	Kopenhagen
Kassel	310	525	385	285	165	350	235	390	150	470	200	450	560	450	310	170		2250	650	4000	4700	2700	3650	1850	3000	2700	3700	3800	3550	3050	Lissabon
Koblenz	165	405	600	410	190	510	145	425	310	670	120	660	550	610	520	390	250		1700	2150	2850	1100	1800	400	1200	1800	2400	1900	1650	1450	London
Köln	75	505	580	315	100	570	40	330	370	570	200	640	650	670	430	300	250	105		3500	4200	2050	3100	1250	2350	2050	3000	3200	3000	2450	Madrid
Leipzig	570	710	190	390	440	115	505	520	140	570	385	255	520	215	440	290	280	440	500		700	1600	2000	2150	1200	2450	1700	1400	550	1300	Minsk
Mannheim	290	270	615	515	300	530	280	600	330	725	85	680	410	630	570	430	270	150	250	460		2300	1900	2850	1900	3150	2150	1300	1250	2000	Moskau
München	650	390	590	750	605	460	610	840	420	930	400	650	90	560	780	630	470	490	580	430	350		1600	850	400	950	1350	1600	1050	450	München
Nürnberg	470	450	440	585	440	315	445	675	270	795	230	510	260	420	610	470	310	340	410	280	240	170		1800	1350	2600	2900	550	1400	1950	Oslo
Passau	690	580	630	800	660	465	655	890	460	980	440	700	280	570	820	680	520	550	620	470	440	190	220		1100	1450	2150	1950	1600	1250	Paris
Rostock	650	1000	230	300	515	440	570	425	490	285	740	325	870	470	180	330	560	690	600	380	810	780	630	820		1350	1300	1350	650	300	Prag
Saarbrücken	310	265	725	550	330	640	280	560	440	810	200	790	500	740	660	530	380	180	250	570	140	430	370	570	920		1050	2750	1850	1150	Rom
Salzburg	800	530	735	910	755	605	770	980	580	1080	540	800	180	540	920	780	625	660	720	570	510	150	320	120	940	600		2750	1450	1000	Sofia
Stuttgart	420	270	630	630	420	510	410	710	350	810	210	700	300	610	660	520	360	280	370	470	135	230	210	400	820	220	390		450	1900	Stockholm
Trier	160	325	715	480	260	635	200	500	430	735	190	780	580	730	590	460	340	140	180	560	180	500	420	620	760	100	700	300		700	Warschau
Wiesbaden	230	350	570	430	210	490	200	480	280	680	40	640	500	590	520	380	220	80	170	410	100	430	260	470	760	160	570	220	150		Wien
Entfernungen Deutschland – Orte aus Europatabelle in rot	Aachen	Basel	Berlin	Bremen	Dortmund	Dresden	Düsseldorf	Emden	Erfurt	Flensburg	Frankfurt M.	Frankfurt O.	Garm.-Patenk.	Görlitz	Hamburg	Hannover	Kassel	Koblenz	Köln	Leipzig	Mannheim	München	Nürnberg	Passau	Rostock	Saarbrücken	Salzburg	Stuttgart	Trier	Wiesbaden	Beispiel: Berlin bis Dresden =200 km; Berlin bis Prag =400 km; Dresden bis Prag =400 km – 200 km

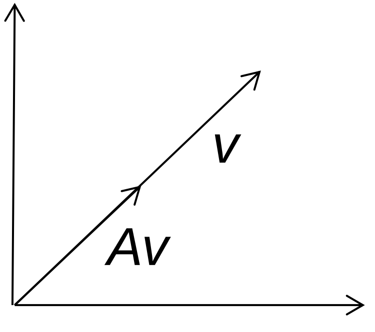
Eigenwerte und Eigenvektoren

Gegeben sei eine n x n-Matrix **A**. Ein Vektor **v ≠ 0** heißt **Eigenvektor** zu **A**, wenn **Av** die gleiche oder die gegengleiche Richtung zu **v** hat. Als Gleichung ausgedrückt: Es gibt eine Zahl

λ, so dass gilt

Av = λ **v**

Die Zahl λ nennt man den **Eigenwert** zu **A**.



Google Page Rank

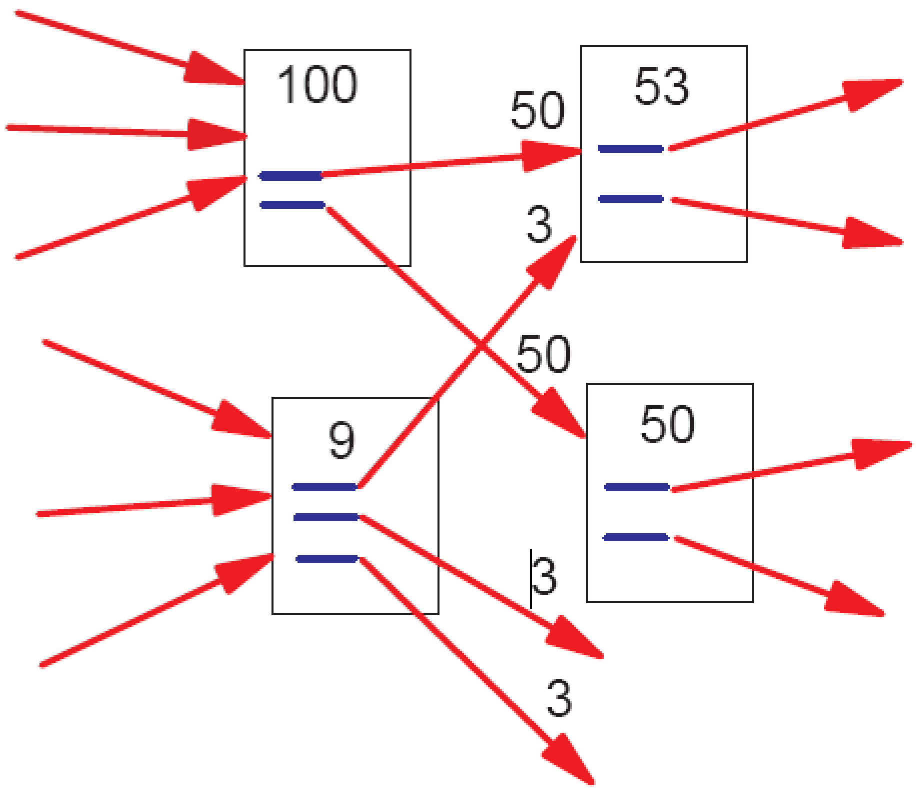
- Google PageRank ermittelt die „Wichtigkeit“ von Hyperlinks und ist patentiert (U.S. Patent 6,285,999)
- Patentinhaber ist die Stanford University
  - PageRank ist Ergebnis des Dissertationsprojekts von Larry Page und Sergej Brin
- Standford University erhielt 1,8 Mio. Google-Aktien für die Nutzung des Patents durch Google
- Sie verkaufte die Google-Aktien im Jahr 2005 für 336 Mio. USD
- Wert am 03.12.2013: 1,4 Mrd. USD

Google Page Rank

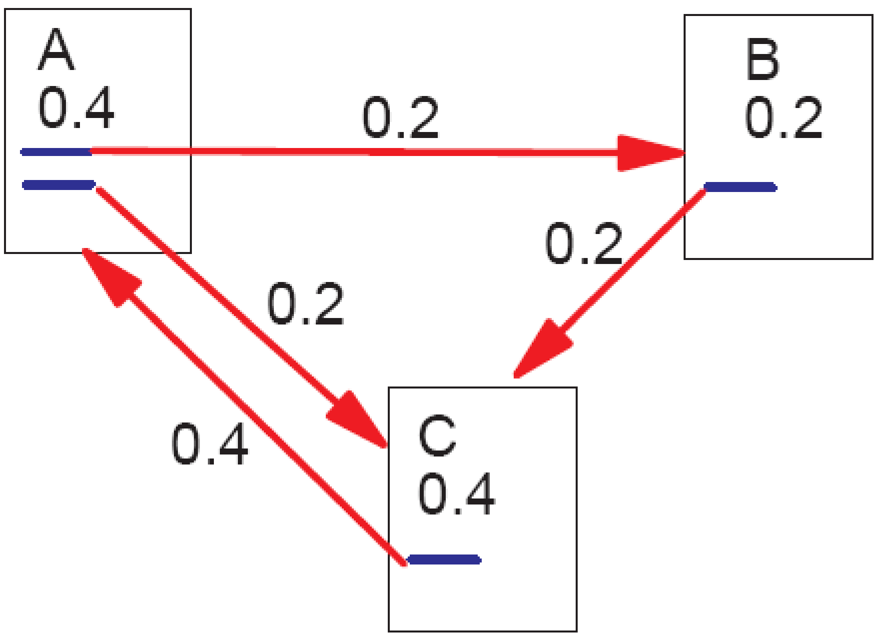
- Grundidee: Die Websites nach *Bedeutung* sortieren
- Was ist "bedeutend"?
  - Eine Website v ist umso bedeutender, je mehr bedeutende Websites u auf sie zeigen
- **Henne-Ei-Problem**: Um die Bedeutung von v zu berechnen, müssen wir bereits die Bedeutung von u kennen
- Formulieren wir genauer (vereinfachter PageRank-Algorithmus):
  - x_u : Bedeutung der Website n
  - N_u : Anzahl der Websites, auf die n zeigt (sog. „**Outlinks**“ von n).

Google Page Rank

- Jede Website u verteilt ihre Bedeutung x_n auf ihre Outlinks, und zwar zu gleichen Teilen:



Bedeutungsübertragung



konvergierter Zustand