LEHRSTUHL FÜR DATENBANKSYSTEME UND DATA MINING

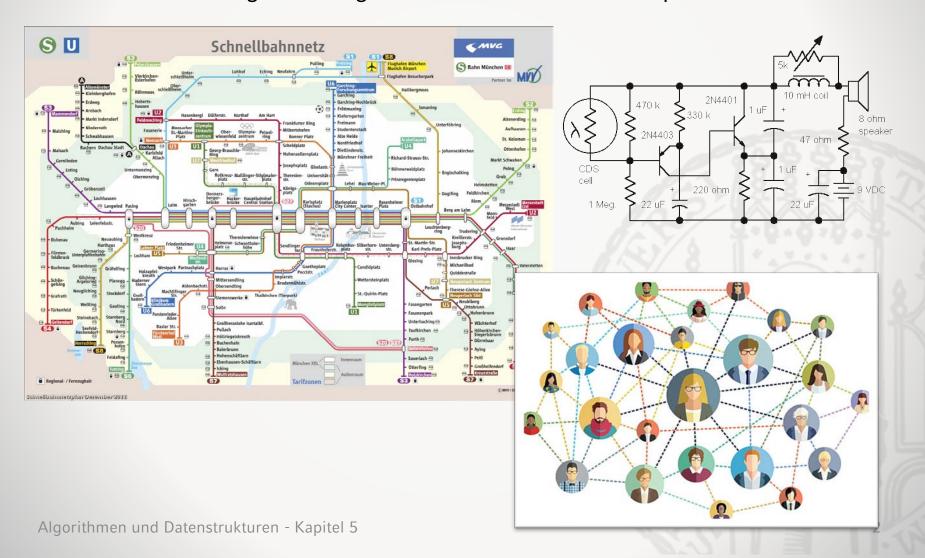
Kapitel 5: Graphen

Graph-Repräsentationen
Kürzeste Wege
Minimale Spannbäume
Flussnetzwerke



Motivation zu Graphen

Viele reale Fragestellungen lassen sich durch Graphen darstellen



Motivation zu Graphen

- Bezogen auf einen Graphen ergeben sich unterschiedliche Fragen:
 - Existiert eine Verbindung zwischen zwei Knoten A und B?
 - Existiert eine zweite Verbindung, falls die erste blockiert ist?
 - Wie lautet die kürzeste Verbindung von A nach B?
 - Wie sieht ein minimaler Spannbaum zu einem Graphen aus?
 - Wie plane ich eine optimale Rundreise?
 (Traveling Salesman Problem)

Gerichteter Graph

Ein gerichteter Graph G = (V, E) (engl. digraph = "directed graph") ist ein Tupel einer Menge V, deren Elemente Knoten (nodes, vertices) heißen, und einer Menge $E \subseteq V \times V$, deren Elemente Kanten (edges, arcs) heißen.

Bemerkungen:

- |V| = Knotenanzahl
- Wir betrachten hier nur endliche Graphen: $0 \le |V| \le \infty$
- $|E| \le |V|^2 = \text{Kantenanzahl}$
- Meist werden die Knoten durchnummeriert: i = 0, 1, 2, ..., |V| 1

Gerichteter Graph

Graphische Darstellung einer Kante von v nach w:

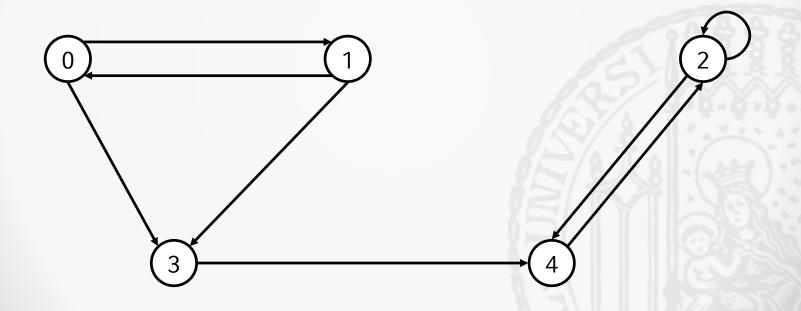


Begriffe:

- v ist Vorgänger von w
- w ist Nachfolger von v
- v und w sind Nachbarn und heißen adjazent

Gerichteter Graph: Beispiel

- $V = \{0,1,2,3,4\}$
- $E = \{(0,1), (0,3), (1,0), (1,3), (2,2), (2,4), (3,4), (4,2)\}$



Gerichteter Graph: Definitionen

- Grad eines Knotens := Anzahl der ein- und ausgehenden Kanten
- Ein *Pfad* ist eine Folge von Knoten $v_0, ..., v_{n-1}$ mit $(v_i, v_{i+1}) \in E$ für $0 \le i \le n-1$, also eine Folge "zusammenhängender" Kanten.
- Länge eines Pfades := Anzahl der Kanten auf dem Pfad
- Ein Pfad heißt einfach, wenn alle Knoten auf dem Pfad paarweise verschieden sind.
- Ein Zyklus ist ein Pfad mit $v_0 = v_{n-1}$ und Länge $n \ge 2$.
- Ein Teilgraph G' = (V', E') eines Graphen G = (V, E) ist ein Graph mit $V' \subseteq V$ und $E' \subseteq E \cap (V' \times V')$.

Gerichteter Graph: Markierungen

- Man kann Markierungen oder Beschriftungen für Kanten und Knoten einführen.
- Häufig verwendet: Kostenfunktionen für Kanten
- Notation:
 - -c[v,w] oder cost(v,w), c(v,w)
- Bedeutung:
 - Entfernung zwischen v und w
 - Reisezeit
 - Reisekosten
 - **–** ...

Ungerichteter Graph

Ein ungerichteter Graph ist ein gerichteter Graph, in dem die Relation *E* symmetrisch ist:

$$(v, w) \in E \Rightarrow (w, v) \in E$$

Die graphische Darstellung wählt man ohne Pfeil:



Bemerkung:

- Die eingeführten Begriffe (Grad eines Knoten, Pfad, ...) definiert man analog zu denen für gerichtete Graphen.
- Bisweilen sind Modifikationen erforderlich, z.B. muss ein Zyklus hier mindestens drei Knoten haben.

Repräsentationen von Graphen

- Man kann Graphen je nach Zielsetzung knoten- oder kantenorientiert abspeichern.
- Die knotenorientierte Darstellungsform ist gebräuchlicher und existiert in verschiedenen Variationen.

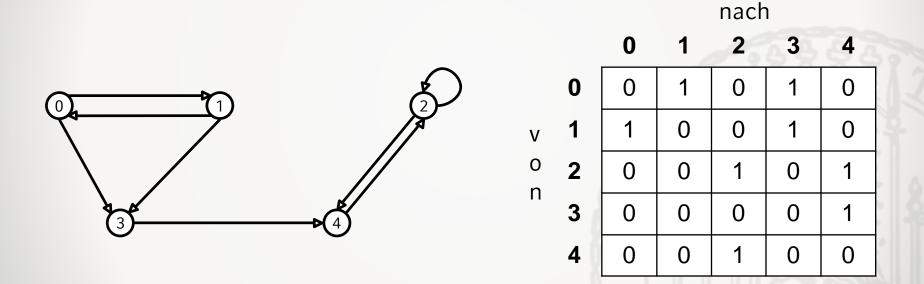
Die Adjazenzmatrix A ist eine boolesche Matrix mit:

$$A_{ij} = \begin{cases} true & \text{falls}(v_i, v_j) \in E \\ false & \text{sonst} \end{cases}$$

Eine solche Matrix $[A_{ij}]$ lässt sich als Array A[i][j] darstellen.

Boolesche Adjazenzmatrix – Beispiel

Für den Beispielgraph G_1 ergibt sich folgende Adjazenzmatrix mit der Konvention true = 1, false = 0:

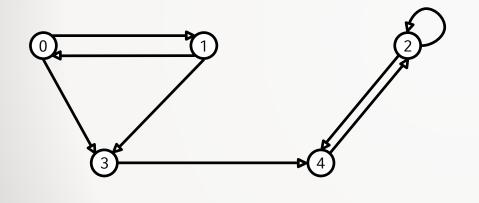


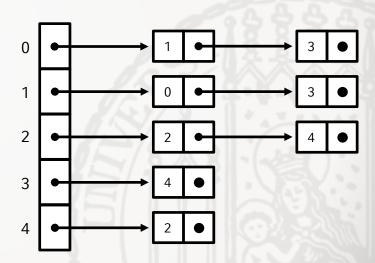
Adjazenzmatrix

- Vorteile
 - Entscheidung, ob $(i,j) \in E$ geht in Zeit O(1)
- Nachteile
 - Platzbedarf stets $O(|V|^2)$, ineffizient falls $|E| \ll |V|^2$
 - Initialisierung benötigt Zeit $O(|V|^2)$
- Kantenbeschriftung
 - anstelle von booleschen Werten lassen sich auch Zusatzinformationen in der Matrix speichern
 - Bsp: Kosten; Weglängen
 - Definition Kosten-Adjazenzmatrix: $A_{ij} = \begin{cases} c(v_i, v_j) & (v_i, v_j) \in E \\ \infty & sonst \end{cases}$ zwischen v_i und v_i
- Achtung: Bei boolescher Adjazenzmatrix bedeutet A(i,j) = 0, dass keine Kante besteht; bei Kostenadjazenzmatrix, dass die Kosten $c(v_i, v_i) = 0$ sind.

Adjazenzliste

- Für jeden Knoten wird eine Liste der Nachbarknoten angelegt.
- Für G₁ ergibt sich folgende Adjazenzliste:





Adjazenzliste

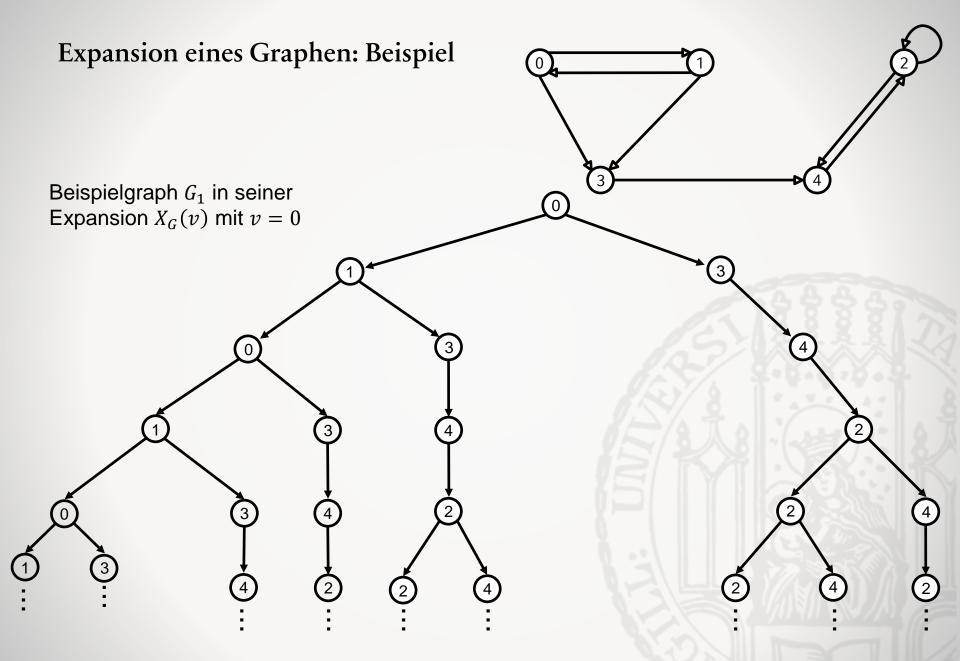
- Vorteile
 - geringer Platzbedarf von O(|V| + |E|)
 - Initialisierung in Zeit O(|V| + |E|)

- Nachteile
 - Entscheidung, ob $(i,j) \in E$ benötigt $O\left(\frac{|E|}{|V|}\right)$ Zeit im Average Case
- Kantenbeschriftung
 - als Zusatzinformation bei Listenelementen

Expansion eines Graphen

Die Expansion $X_G(v)$ eines Graphen G in einem Knoten v ist ein Baum, der wie folgt definiert ist:

- Falls v keine Nachfolger hat, ist $X_G(v)$ nur der Knoten v.
- Falls $v_1, ..., v_k$ die Nachfolger von v sind, ist $X_G(v)$ der Baum mit der Wurzel v und den Teilbäumen $X_G(v_1), ..., X_G(v_k)$.



Expansion eines Graphen: Anmerkungen

- Die Knoten des Graphen können mehrfach im Baum vorkommen.
- Ein Baum ist unendlich, falls der Graph Zyklen hat.
- Der Baum $X_G(v)$ stellt die Menge aller Pfade dar, die von v ausgehen.

Graph-Durchlauf

Entspricht Baum-Durchlauf durch Expansion (ggf. mit Abschneiden)

- Tiefendurchlauf: preorder traversal (depth first)
- Breitendurchlauf: level order traversal

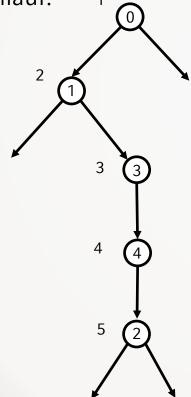
Wichtige Modifikation:

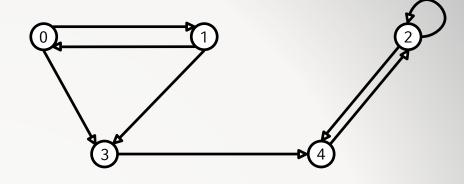
- 1. Schon besuchte Knoten müssen markiert werden, weil Graphknoten im Durchlauf mehrfach vorkommen können. (Zyklen!)
- 2. Abbruch des Durchlaufs bei schon besuchten Knoten.

Graph-Durchlauf: Beispiel

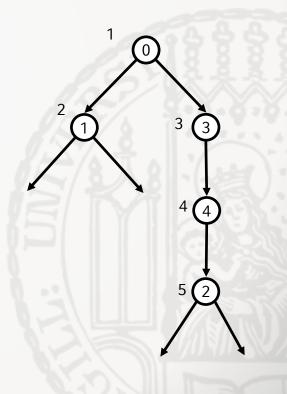
 G_1 mit Startknoten: v = 0

Tiefendurchlauf:





Breitendurchlauf:



Ansatz für Graph-Durchlauf

- 1. Initialisierung: markiere alle Knoten als "not visited"
- Abarbeiten der Knotenif (node "not visited") then
 - bearbeite
 - markiere: "visited"
 - weitergehen zu Nachfolger

Für die Markierung "visited" reicht der Typ boolean. Für andere Berechnungen auf Graphen benötigt man aber auch mehr als die zwei Werte "true" and "false".

Markierungen beim Durchlauf

Während des Graph-Durchlaufs werden folgende Markierungen für die Graph-Knoten verwendet:

- Ungesehene Knoten (unseen vertices): Knoten, die noch nicht erreicht worden sind: val[v] = 0
- Baum-Knoten (tree vertices): Knoten, die schon besucht und abgearbeitet sind. Diese Knoten ergeben die Expansion: val[v] = id > 0
- Rand-Knoten (fringe vertices), aktive Knoten: Knoten, die über eine Kante mit einem Baum-Knoten verbunden sind: val[v] = -1

Beliebiges Auswahlkriterium

 Start: Markiere den Startknoten als Rand-Knoten und alle anderen Knoten als ungesehene Knoten.

Schleife: repeat

- Wähle einen Rand-Knoten x mittels eines Auswahlkriteriums (depth first, breadth first, priority first).
- Dazu: Priority Queue,
- Markiere x als Baum-Knoten und bearbeite x.
- Markiere alle ungesehenen Nachbar-Knoten von x als Rand-Knoten.

... until (alle Knoten abgearbeitet)

Kürzeste Wege

- Problemstellung: Suche kürzesten Weg
 - Von einem Knoten zu allen anderen: "Single Source Shortest Path"
 - 2. Von allen Knoten zu einem Ziel: "Single Destination Shortest Path"
 - 3. Von allen Knoten zu allen anderen: "All Pairs Shortest Path"
- Geg.: Gerichteter Graph G mit Kostenfunktion (= Adjazenzmatrix)

$$c[v,w] \begin{cases} \geq 0 & \text{, falls eine Kante von } v \text{ nach } w \text{ existiert} \\ = \infty & \text{, falls keine Kante von } v \text{ nach } w \text{ existiert} \\ = 0 & \text{, für } w = v \end{cases}$$

Startknoten v_0 , Endknoten w

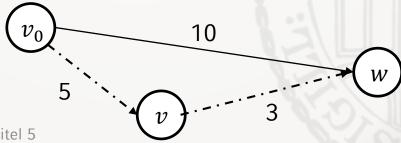
• Ges.: Pfad von v_0 zu jedem Knoten w mit minimalen Gesamtkosten

Eigenschaften von Pfaden

- Pfadkosten können durch Erweiterung eines Pfades nur wachsen, da Kantengewichte stets positiv sind.
- Falls beste Pfade von v_0 zu allen anderen Knoten $V \{v_0\}$ höhere Kosten haben, ist der kürzeste Pfad bereits gefunden.
- Ein kürzester Pfad hat keinen Zyklus.
- Ein kürzester Pfad hat max. (|V| 1) Kanten.
- Notation:
 - $-S_k$: Menge von k Knoten v mit k besten Pfaden von v_0 nach v
 - $-D_k(v)$: Kosten/Distanz des besten Pfads von v_0 über maximal k Knoten in S_k nach v

Dijkstra-Algorithmus

- Edsger Wybe Dijkstra (1930-2002):
 niederländischer Informatiker & Turingpreisträger
- Ziel:
 Berechnung des kürzesten Weges von einem Startknoten zu allen anderen Knoten im Graphen
- Idee:
 - Wir speichern im Array D für jeden Knoten v die aktuell gültige Kostenschätzung.
 - Als Initialisierung verwenden wir die Kosten der direkten Pfade aus der Adjazenzmatrix c[v, w].
 - In jedem Schritt versuchen wir alle Pfade zu verbessern, indem wir mögliche Zwischenknoten untersuchen, die den Pfad eventuell kürzer machen.



Dijkstra-Algorithmus: Implementierung

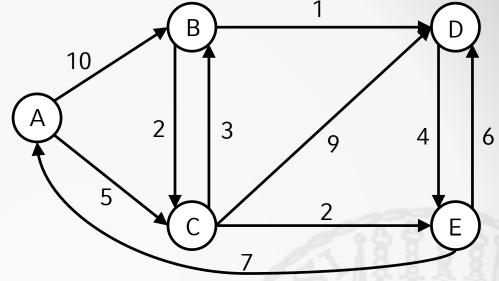
G: Graph, bestehend aus der Adjazenzmatrix c über Knotenmenge V

S: Menge der bereits abgearbeiteten Knoten

D: aktuelle Kostenschätzung für den Pfad von v_0 zu allen anderen Knoten

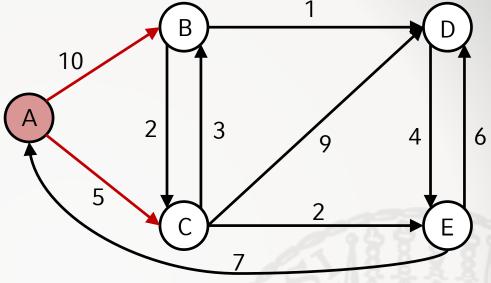
```
Dijkstra(G, v_0){
 S \leftarrow \{v_0\}
 forall v \in V \{ D(v) \leftarrow c[v_0, v] \} // Initialisierung: übernimm Kanten
 while (V - S) \neq \emptyset {
   w_{min} \leftarrow \operatorname{argmin}_{w \in V-S} D(w)
   S \leftarrow S \cup \{w_{min}\}
   for each v \in V - S {
     D(v) = \min(D[v], D[w_{min}] + c[w_{min}, v])
```

Adjazenzmatrix								
0	10	5	∞	∞				
∞	0	2	1	∞				
∞	3	0	9	2				
∞	∞	∞	0	4				
7	∞	∞	6	0				



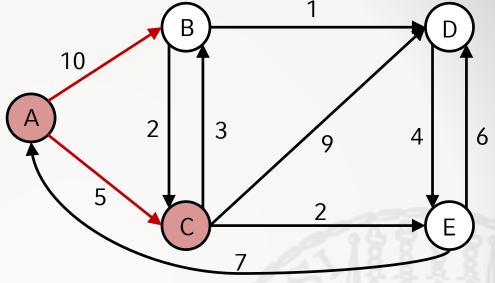
\boldsymbol{k}	w_k	${\mathcal S}_k$	$D_k(B)$	$D_k(C)$	$D_k(D)$	$D_k(E)$

Adjazenzmatrix								
0	10	5	∞	∞				
∞	0	2	1	∞				
∞	3	0	9	2				
∞	∞	∞	0	4				
7	∞	∞	6	0				



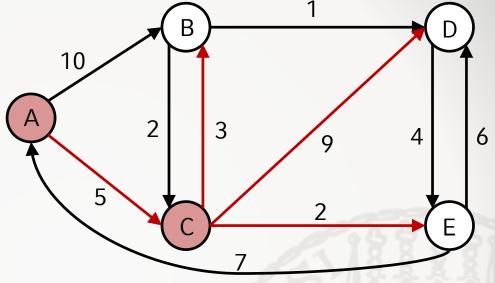
k	w_k	${\mathcal S}_k$	$D_k(B)$	$D_k(C)$	$D_k(D)$	$D_k(E)$
0	_	{ <i>A</i> }	10	5	∞	∞

Adjazenzmatrix								
0	10	5	∞	∞				
∞	0	2	1	∞				
∞	3	0	9	2				
∞	∞	∞	0	4				
7	∞	∞	6	0				



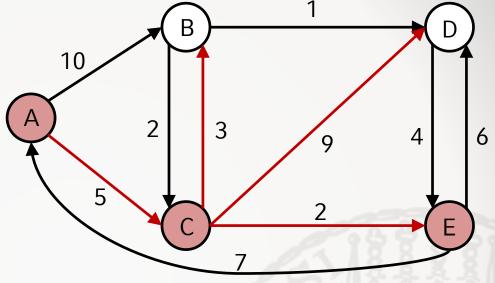
\boldsymbol{k}	w_k	${\mathcal S}_k$	$D_k(B)$	$D_k(C)$	$D_k(D)$	$D_k(E)$
0	_	{ <i>A</i> }	10	5	∞	∞

Adjazenzmatrix								
0	10	5	∞	∞				
∞	0	2	1	∞				
∞	3	0	9	2				
∞	∞	∞	0	4				
7	∞	∞	6	0				



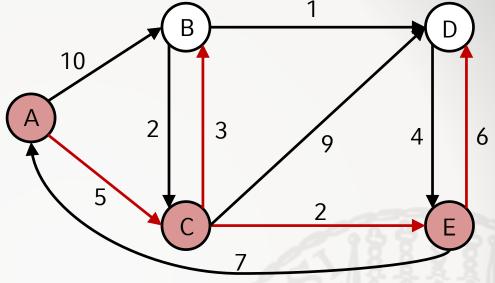
\boldsymbol{k}	w_k	${\mathcal S}_k$	$D_k(B)$	$D_k(C)$	$D_k(D)$	$D_k(E)$
0	_	<i>{A}</i>	10	5	∞	∞
1	С	{ <i>A</i> , <i>C</i> }	8	5	14	7

Adjazenzmatrix								
0	10	5	∞	∞				
∞	0	2	1	∞				
∞	3	0	9	2				
∞	∞	∞	0	4				
7	∞	∞	6	0				



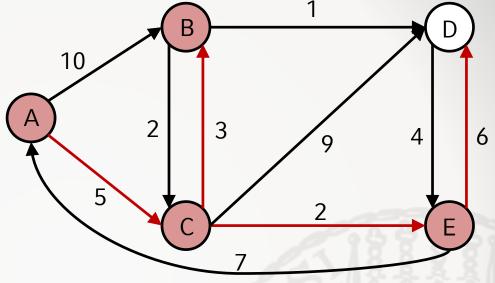
\boldsymbol{k}	w_k	${\mathcal S}_k$	$D_k(B)$	$D_k(C)$	$D_k(D)$	$D_k(E)$
0	_	<i>{A}</i>	10	5	∞	∞
1	С	{ <i>A</i> , <i>C</i> }	8	5	14	7

Adjazenzmatrix								
0	10	5	∞	∞				
∞	0	2	1	∞				
∞	3	0	9	2				
∞	∞	∞	0	4				
7	∞	∞	6	0				



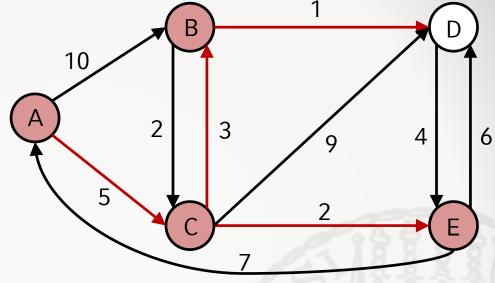
k	w_k	${\mathcal S}_k$	$D_k(B)$	$D_k(C)$	$D_k(D)$	$D_k(E)$
0	_	<i>{A}</i>	10	5	∞	∞
1	С	{ <i>A</i> , <i>C</i> }	8	5	14	7
2	E	$\{A,C,E\}$	8	5	13	7

Adjazenzmatrix						
0	10	5	∞	∞		
∞	0	2	1	∞		
∞	3	0	9	2		
∞	∞	∞	0	4		
7	∞	∞	6	0		



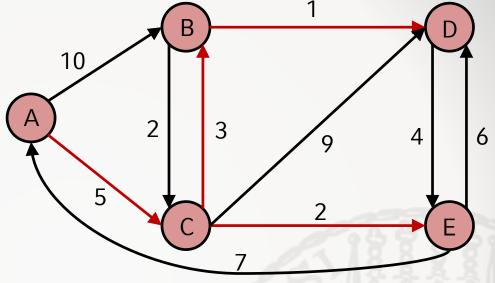
k	w_k	${\mathcal S}_k$	$D_k(B)$	$D_k(C)$	$D_k(D)$	$D_k(E)$
0	_	<i>{A}</i>	10	5	∞	∞
1	С	{ <i>A</i> , <i>C</i> }	8	5	14	7
2	E	$\{A,C,E\}$	8	5	13	7

Adjazenzmatrix						
0	10	5	∞	∞		
∞	0	2	1	∞		
∞	3	0	9	2		
∞	∞	∞	0	4		
7	∞	∞	6	0		



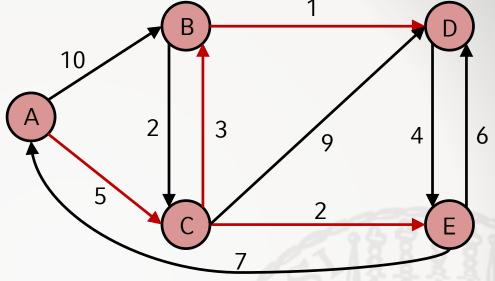
k	w_k	${\mathcal S}_k$	$D_k(B)$	$D_k(C)$	$D_k(D)$	$D_k(E)$
0	_	<i>{A}</i>	10	5	∞	∞
1	С	{ <i>A</i> , <i>C</i> }	8	5	14	7
2	E	$\{A,C,E\}$	8	5	13	7
3	В	$\{A,C,E,B\}$	8	5	9	7

Adjazenzmatrix						
0	10	5	∞	∞		
∞	0	2	1	∞		
∞	3	0	9	2		
∞	∞	∞	0	4		
7	∞	∞	6	0		



k	w_k	${\mathcal S}_k$	$D_k(B)$	$D_k(C)$	$D_k(D)$	$D_k(E)$
0	_	<i>{A}</i>	10	5	∞	∞
1	С	{ <i>A</i> , <i>C</i> }	8	5	14	7
2	E	$\{A,C,E\}$	8	5	13	7
3	В	$\{A,C,E,B\}$	8	5	9	7

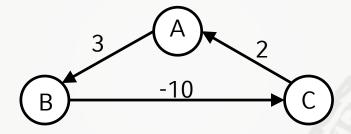
Adjazenzmatrix						
0	10	5	∞	∞		
∞	0	2	1	∞		
∞	3	0	9	2		
∞	∞	∞	0	4		
7	∞	∞	6	0		



k	w_k	${\mathcal S}_k$	$D_k(B)$	$D_k(C)$	$D_k(D)$	$D_k(E)$
0	_	<i>{A}</i>	10	5	∞	∞
1	С	{ <i>A</i> , <i>C</i> }	8	5	14	7
2	E	$\{A,C,E\}$	8	5	13	7
3	В	$\{A,C,E,B\}$	8	5	9	7
4	D	$\{A,C,E,B,D\}$	8	5	9	7

Dijkstra-Algorithmus: Analyse

- Liefert optimale Lösung, nicht nur Näherung
- Falls Zyklen mit negativen Kosten zugelassen wären, gäbe es keinen eindeutigen Pfad mit minimalen Kosten mehr



- Komplexität:
 - Falls G zusammenhängend, mit Adjazenzmatrix $O(|V|^2)$
 - Einsatz als "All Pairs Shortest Path" prinzipiell möglich, ergibt Zeitkomplexität $O(|V| \cdot |V|^2) = O(|V|^3)$.

Floyd-Algorithmus

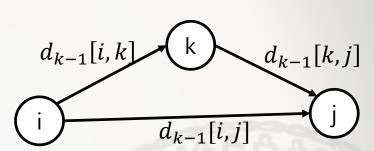
- Robert W Floyd (1936 2001):
 Amerikanischer Informatiker & Turingpreisträger
- "All Pairs Shortest Path"
- Gegeben: Gerichteter Graph G mit Kostenfunktion

$$c[v,w] \begin{cases} \geq 0 & \text{falls eine Kante } v \text{ nach } w \text{ existiert} \\ = \infty & \text{falls keine Kante } v \text{ nach } w \text{ existiert} \\ = 0 & \text{falls } v = w \end{cases}$$

- Gesucht: Pfad von jedem Knoten v zu jedem Knoten w mit minimalen Gesamtkosten
- Idee:
 - Existierende Kanten im Graphen zugrunde legen
 - Versuche sukzessive, zwei Knoten über einen Zwischenknoten günstiger zu verbinden als bisher
 - Lösung: Dynamische Programmierung

Floyd-Algorithmus: Dynamische Programmierung

- Grundidee
 - Betrachte alle Knoten der Reihe nach als mögliche Zwischenknoten k
 - Speicherung in einer Matrix d[i,j],
 die in jedem Schritt aktualisiert wird



- Initialisierung durch direkte Kanten des gegebenen Graphen
 - Für alle $i, j \in \{1, ..., |V|\}$: setze $d_0[i, j] = c[i, j]$
 - Entspricht Lösung der Elementarprobleme
- Aktualisiere Matrix d[i,j] für Zwischenknoten k:
 - Sind Wege über Knoten k günstiger als bisherige Wege?
 - $d_k[i,j] = \min\{d_{k-1}[i,j], d_{k-1}[i,k] + d_{k-1}[k,j]\}$
 - Entspricht Zusammensetzen der Teilergebnisse zur Gesamtlösung

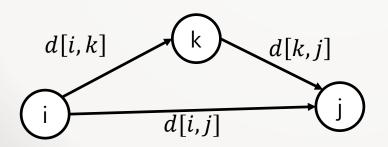
Floyd-Algorithmus: Pseudo-Code

Gegeben: Kosten $c[v, w] \in \mathbb{R}_0^+ \cup \{\infty\}$ von Knoten v nach w

Für alle Knotenpaare
$$i, j \in \{0, ..., |V| - 1\}$$

 $d[i][j] \leftarrow c[i, j]$

Für alle $k \in \{0, ..., |V| - 1\}$ Für alle $i \in \{0, ..., |V| - 1\}$ Für alle $j \in \{0, ..., |V| - 1\}$ $d[i][j] \leftarrow \min(d[i][j], d[i][k] + d[k][j])$



Initialisierung von Matrix d:

- Jeder Knoten hat Distanz0 zu sich selbst
- Sonst übernehmen wir erst einmal die direkten (schon bekannten)
 Verbindungen

Falls Weg über *k* besser / kürzer als bisher bester Weg, ist dieser Weg nun der Favorit.

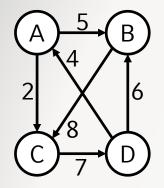
Floyd-Algorithmus: Weginformationen

Gegeben: Kosten $c[v, w] \in \mathbb{R} \cup \{\infty\}$ von Knoten v nach w

```
Für alle Knotenpaare i,j \in \{0,...,|V|-1\} d[i][j] \leftarrow c[i,j] P[i][j] \leftarrow j \text{ (sonst undef.)} \begin{bmatrix} P_{11} & \cdots & P_{1j} & \cdots & P_{1n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ P_{i1} & \cdots & P_{ij} & \cdots & P_{in} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ P_{n1} & \cdots & P_{nj} & \cdots & P_{nn} \end{bmatrix} Für alle k \in \{0,...,|V|-1\} Der kürzeste Weg von i nach j verläuft über P_{ij}. Für alle j \in \{0,...,|V|-1\} falls d[i][j] > d[i][k] + d[k][j] // d.h. neuer Weg ist kürzer d[i][j] \leftarrow d[i][k] + d[k][j] P[i][j] \leftarrow k
```

 P speichert für zwei Knoten i, j einen zentralen Knoten auf dem optimalen Pfad; die Pfadrekonstruktion läuft dann rekursiv.

Floyd-Algorithmus: Beispiel



Initialisierung:

$$\begin{pmatrix} 0 & 5 & 2 & \infty \\ \infty & 0 & 8 & \infty \\ \infty & \infty & 0 & 7 \\ 4 & 6 & \infty & 0 \end{pmatrix}$$

Floyd-Algorithmus: Beispiel

Initialisierung:
$$\begin{pmatrix}
0 & 5 & 2 & \infty \\
\infty & 0 & 8 & \infty \\
\infty & \infty & 0 & 7 \\
4 & 6 & \infty & 0
\end{pmatrix}$$

Initialisierung:
$$P_0$$
:
$$\begin{pmatrix}
0 & 5 & 2 & \infty \\
\infty & 0 & 8 & \infty \\
\infty & \infty & 0 & 7 \\
4 & 6 & \infty & 0
\end{pmatrix}$$

$$\begin{pmatrix}
A & B & C & - \\
- & B & C & - \\
- & - & C & D \\
A & B & - & D
\end{pmatrix}$$

$$\begin{array}{cccc}
\kappa = A \\
0 & 5 & 2 & \infty \\
\infty & 0 & 8 & \infty \\
\infty & \infty & 0 & 7 \\
4 & 6 & 6 & 0
\end{array}$$

$$\begin{pmatrix}
0 & 5 & 2 & \infty \\
\infty & 0 & 8 & \infty \\
\infty & \infty & 0 & 7 \\
4 & 6 & 6 & 0
\end{pmatrix}$$

$$\begin{pmatrix} 0 & 5 & 2 & \infty \\ \infty & 0 & 8 & \infty \\ \infty & \infty & 0 & 7 \\ 4 & 6 & 6 & 0 \end{pmatrix} \begin{pmatrix} 0 & 5 & 2 & \infty \\ \infty & 0 & 8 & \infty \\ \infty & \infty & 0 & 7 \\ 4 & 6 & 6 & 0 \end{pmatrix} \begin{pmatrix} 0 & 5 & 2 & 9 \\ \infty & 0 & 8 & 15 \\ \infty & \infty & 0 & 7 \\ 4 & 6 & 6 & 0 \end{pmatrix} \begin{pmatrix} 0 & 5 & 2 & 9 \\ \infty & 0 & 8 & 15 \\ \infty & \infty & 0 & 7 \\ 4 & 6 & 6 & 0 \end{pmatrix}$$

$$\begin{pmatrix}
0 & 5 & 2 & 9 \\
19 & 0 & 8 & 15 \\
11 & 13 & 0 & 7 \\
4 & 6 & 6 & 0
\end{pmatrix}$$

$$P_{A}$$
:
 $\begin{pmatrix} A & B & C & - \\ - & B & C & - \\ - & - & C & D \\ A & B & A & D \end{pmatrix}$

$$\begin{pmatrix}
A & B & C & - \\
- & B & C & - \\
- & - & C & D \\
A & B & A & D
\end{pmatrix}$$

$$P_{D}$$
:
$$\begin{pmatrix}
A & B & C & C \\
D & B & C & C \\
D & D & C & D \\
A & B & A & D
\end{pmatrix}$$

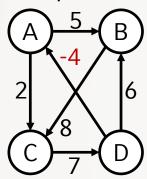
Floyd-Algorithmus: Komplexität

- 3 geschachtelte Schleifen mit k, i, j über V
- Zeitkomplexität: $O(|V|^3)$
- Platzkomplexität: $O(|V|^2)$
- Warshall-Algorithmus
 - Aus demselben Jahr (1962) stammt ein Algorithmus von Warshall, der statt Kosten nur die Existenz von Verbindungen betrachtet (transitive Hülle).
 - Innere Schleife läuft auf booleschen Werten: if not d[i][j] $d[i][j] = d[i][k] \wedge d[k][j]$

Floyd-Algorithmus: Negative Kanten

• Wir haben negative Kanten nicht ausgeschlossen: $c[v,w] \in \mathbb{R} \cup \{\infty\}$

Was passiert mit einer negativen Kante?



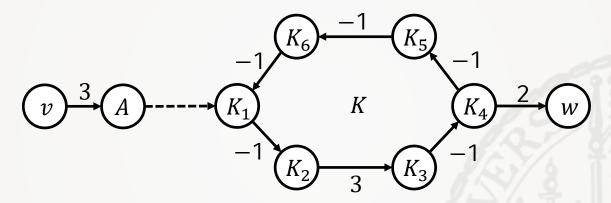
$$d = \begin{pmatrix} 0 & 5 & 2 & 9 \\ \mathbf{11} & 0 & 8 & 15 \\ \mathbf{3} & \mathbf{8} & 0 & 7 \\ -\mathbf{4} & \mathbf{1} & -\mathbf{2} & 0 \end{pmatrix}$$

Scheint ok. Noch ein Beispiel:

$$d = \begin{pmatrix} -1 & 1 & 1 & -6 \\ 5 & 0 & 7 & 0 \\ -3 & -1 & -1 & -8 \\ 3 & 5 & 5 & -2 \end{pmatrix}$$

Negative Kreise

- Einzelne negative Kanten
 - → kein Problem, kürzeste Wege bleiben meist wohldefiniert.
- Negative Kreise



- Falls es einen Pfad von v nach K und einen Pfad von K nach w gibt mit K < 0, dann ist der kürzeste Pfad von v nach w nicht definiert.
- Für $K \ge 0$ gibt es keine Probleme. Der kürzeste Pfad ist wohldefiniert.

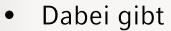
Informierte Suche

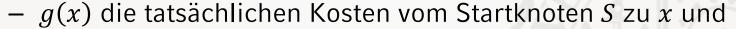
- Dijkstra-Algorithmus:
 - Greedy: Füge Kante sofort hinzu, falls sie geringere Kosten verspricht
 - Kosten zu allen potentiellen Zielen (= Knoten) werden bestimmt
- Verbesserung: Falls das Ziel bekannt ist, können Kanten gezielt ausgewählt werden
- Heuristiken
 - Eine Heuristik bezieht zusätzlichen Wissens ein, um das Auffinden von Lösungen zu beschleunigen.
 - Viele Heuristiken orientieren sich an menschlicher Intuition.
 - Hier: Heuristiken, die Schätzungen von Kosten angeben.

A*-Algorithmus: Idee

- Besuche zuerst Knoten, die wahrscheinlich schnell zum Ziel führen.
- Jeder besuchte Knoten erhält einen Wert f(x), der angibt, wie lange der Pfad vom Start zum Ziel über den Knoten x im günstigsten Fall ist.
- Der Knoten mit dem niedrigsten f-Wert wird als nächstes untersucht:

$$f(x) = g(x) + h(x)$$



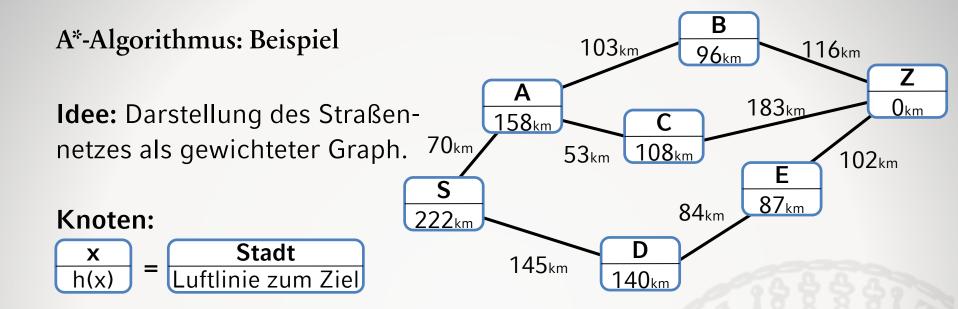




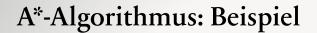
• Die verwendete Heuristik h(x) darf die Kosten für keinen Knoten x überschätzen, da sonst die optimale Lösung vielleicht nicht gefunden wird.

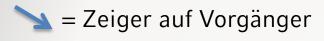
A*-Algorithmus: Datenstrukturen

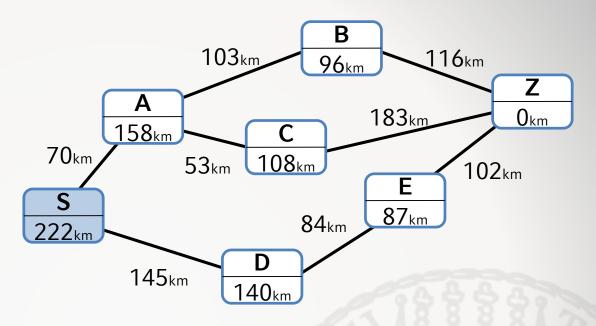
- Jeder Knoten des Graphen kann einer der folgenden Zustände zugeordnet werden. Die Knoten werden dementsprechend in Listen verwaltet:
 - Der Knoten wurde noch nicht verarbeitet und wir kennen noch keinen Weg dorthin.
 - Ein Weg zum Knoten ist bekannt, aber es könnte einen kürzeren Weg geben → OpenList
 - Der kürzeste Weg zum Knoten wurde gefunden → ClosedList



- Jeder Knoten steht für eine Stadt.
- Zwei Knoten sind verbunden wenn es eine direkte
 Straßenverbindung zwischen den entsprechenden Städten gibt.
- Die Kosten der Kanten entsprechen der Länge der Straße zwischen den Städten.
- Gesucht ist der kürzeste Weg von Stadt S nach Stadt Z.
- Als Heuristik h(x) nutzen wir die Luftlinie zwischen der Stadt x und der Zielstadt Z.



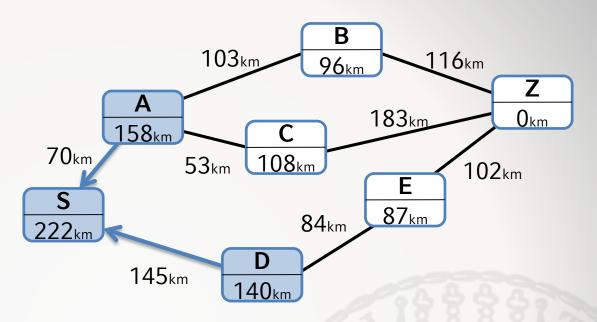




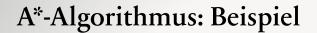
Schritt	OpenList (Stadt, f)	ClosedList (Stadt, Entfernung von S)
0	(S, 222)	



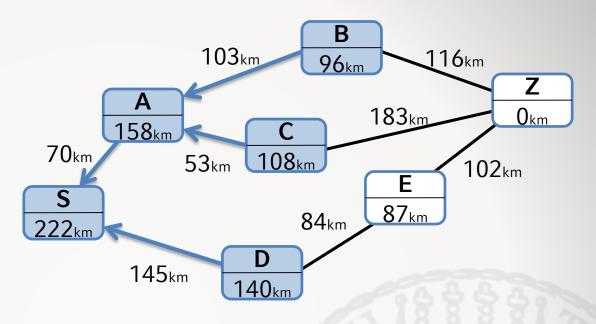
🔪 = Zeiger auf Vorgänger



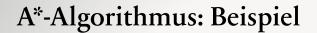
Schritt	OpenList (Stadt, f)	ClosedList (Stadt, Entfernung von S)
0	(S, 222)	
1	(A, 228),(D, 285)	(S, 0)



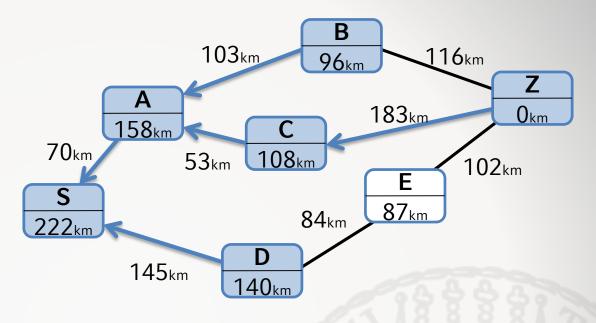
🔪 = Zeiger auf Vorgänger



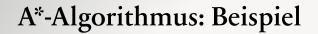
Schritt	OpenList (Stadt, f)	ClosedList (Stadt, Entfernung von S)
0	(S, 222)	
1	(A, 228),(D, 285)	(S, 0)
2	(D, 285), (B, 269), (C, 231)	(S, 0), (A, 70)



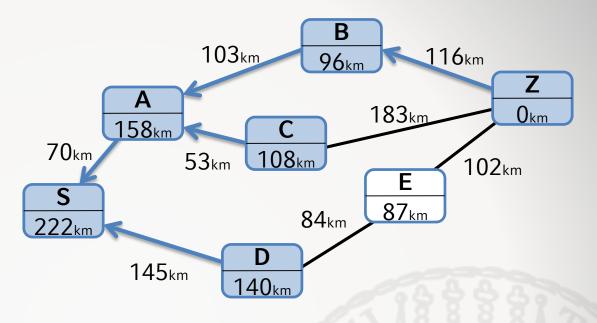
🔌 = Zeiger auf Vorgänger



Schritt	OpenList (Stadt, f)	ClosedList (Stadt, Entfernung von S)
0	(S, 222)	
1	(A, 228),(D, 285)	(S, 0)
2	(D, 285), (B, 269), (C, 231)	(S, 0), (A, 70)
3	(D, 285), (B, 269), (Z, 306)	(S, 0), (A, 70), (C, 123)



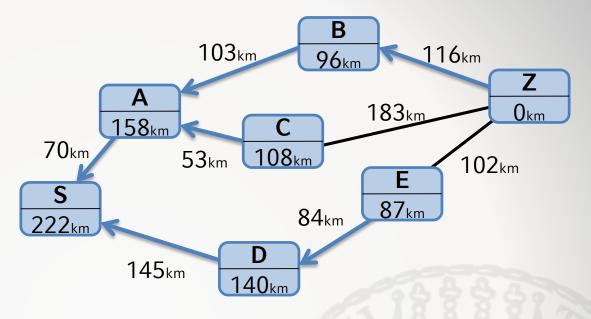
🔪 = Zeiger auf Vorgänger



Schritt	OpenList (Stadt, f)	ClosedList (Stadt, Entfernung von S)
0	(S, 222)	
1	(A, 228),(D, 285)	(S, 0)
2	(D, 285), (B, 269), (C, 231)	(S, 0), (A, 70)
3	(D, 285), (B, 269), (Z, 306)	(S, 0), (A, 70), (C, 123)
4	(D, 285), (Z, 289)	(S, 0), (A, 70), (C, 123), (B, 173)



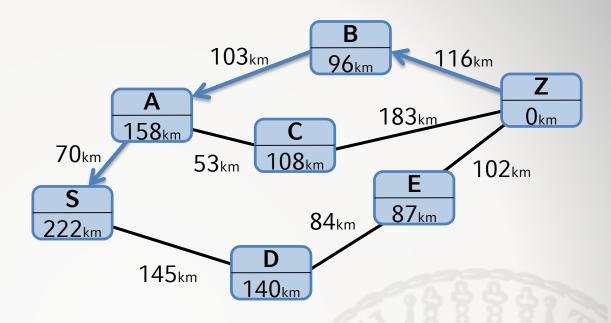




Schritt	OpenList (Stadt, f)	ClosedList (Stadt, Entfernung von S)
0	(S, 222)	
1	(A, 228),(D, 285)	(S, 0)
2	(D, 285), (B, 269), (C, 231)	(S, 0), (A, 70)
3	(D, 285), (B, 269), (Z, 306)	(S, 0), (A, 70), (C, 123)
4	(D, 285), (Z, 289)	(S, 0), (A, 70), (C, 123), (B, 173)
5	(Z, 289),(E, 316)	(S, 0), (A, 70), (C, 123), (B, 173),(D, 145)

A*-Algorithmus: Beispiel

🔌 = Zeiger auf Vorgänger



Schritt	OpenList (Stadt, f)	ClosedList (Stadt, Entfernung von S)
0	(S, 0)	
1	(A, 228),(D, 285)	(S, 0)
2	(D, 285), (B, 269), (C, 231)	(S, 0), (A, 70)
3	(D, 285), (B, 269), (Z, 306)	(S, 0), (A, 70), (C, 123)
4	(D, 285), (Z, 289)	(S, 0), (A, 70), (C, 123), (B, 173)
5	(Z, 289),(E, 316)	(S, 0), (A, 70), (C, 123), (B, 173), (D, 145)
6	(E, 316)	Pfad gefunden: $S \rightarrow A \rightarrow B \rightarrow Z$

A*-Algorithmus: Qualitätseigenschaften

Vollständig

Wenn es eine Lösung gibt, so wird diese auch gefunden.

Optimal

Es wird immer eine optimale Lösung gefunden.

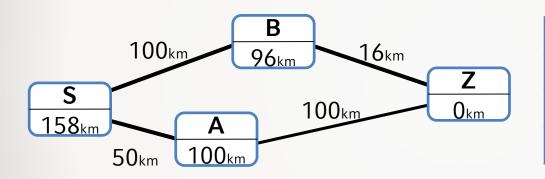
Optimal effizient

Bezogen auf die Laufzeit gibt es keinen Algorithmus, der die gleiche Heuristik verwendet und weniger Knoten besucht.

A*-Algorithmus: Eigenschaft der Heuristik

• Die verwendete Heuristik für h(x) darf die Kosten für keinen Knoten x überschätzen.

Werden die Kosten überschätzt, so ist die *Optimalität* und *Vollständigkeit* des Algorithmus nicht mehr gewährleistet:



OpenList	ClosedList
(S, 0)	
(A, 150),(B,196)	(S, 0)
(B,196), (Z,150)	(S, 0), (A, 150)

- Für die schlechteste Heuristik h(x) = 0 gilt:
 - die geschätzten Kosten für jeden Knoten entsprechen genau den Kosten, um diesen Knoten zu erreichen.
 - Der A*-Algorithmus bildet den Dijkstra-Algorithmus nach.

Minimaler Spannbaum

Gegeben:

Ungerichteter zusammenhängender Graph G = (V, E) mit Kantengewichten $c: E \to \mathbb{R}$.

Gesucht:

Ungerichteter Subgraph G' = (V, E') mit

- \bullet $E' \subseteq E$.
- G' ist zyklenfrei und zusammenhängend, d.h. G' ist ein Spannbaum von V.
- G' ist minimal, d.h. für alle (anderen) Subgraphen G'' = (V, E'') gilt:

$$\sum_{e \in E'} c(e) \le \sum_{e \in E''} c(e)$$



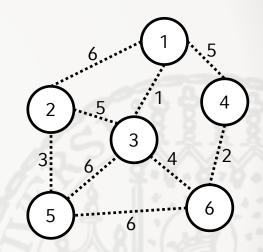
Prim-Algorithmus

- Idee:
 - Beginne mit einem beliebigen Knoten des Graphen.
 - Finde sukzessive die minimale Kante, die den Subgraphen mit einem noch nicht gewählten Knoten verbindet.
 - In jedem Schritt wird der aktuelle Subgraph um jeweils diese Kante und den inzidenten Knoten erweitert.

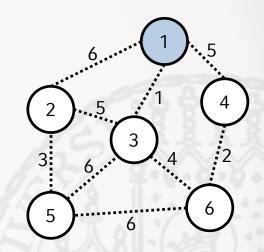
```
V' \leftarrow v_0 beliebig E' \leftarrow \emptyset Solange V \neq V': (u, v) = \underset{e \in V' \times (V - V')}{\operatorname{argmin}} c(e) E' = E' \cup \{(u, v)\} V' = V' \cup v
```

• Komplexität ist $O(|V|^2)$, denn für jeden neu einzufügenden Knoten werden die Kanten zu anderen Knoten überprüft.

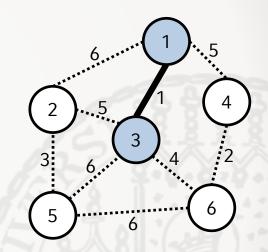
V'	$m{E}'$



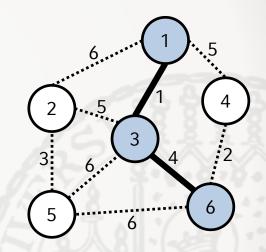
V'	$m{E}'$
{1}	Ø



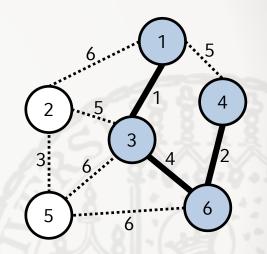
V'	E'
{1}	Ø
{1,3}	{(1,3)}



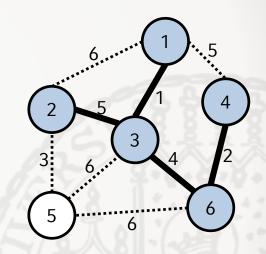
V'	E'
{1}	Ø
{1,3}	{(1,3)}
{1,3,6}	{(1,3), (3,6)}



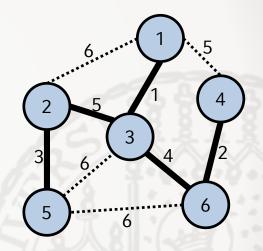
V'	E'
{1}	Ø
{1,3}	{(1,3)}
{1,3,6}	{(1,3), (3,6)}
{1,3,6,4}	{(1,3), (3,6), (6,4)}



V'	E'
{1}	Ø
{1,3}	{(1,3)}
{1,3,6}	{(1,3), (3,6)}
{1,3,6,4}	{(1,3), (3,6), (6,4)}
{1,3,6,4,2}	$\{(1,3), (3,6), (6,4), (3,2)\}$



$oldsymbol{V}'$	$m{E}'$
{1}	Ø
{1,3}	{(1,3)}
{1,3,6}	{(1,3), (3,6)}
{1,3,6,4}	{(1,3), (3,6), (6,4)}
{1,3,6,4,2}	{(1,3), (3,6), (6,4), (3,2)}
{1,3,6,4,2,5}	$\{(1,3), (3,6), (6,4), (3,2), (2,5)\}$



Algorithmus von Kruskal

- Idee (ähnlich zu Prim):
 - Starte mit leerer Kantenmenge.
 - Füge sukzessive minimale Kanten bezüglich ihrer Kosten hinzu, sodass kein Kreis entsteht.
 - Stoppe, falls keine solche Kante mehr gefunden werden kann (die nächste Kante bildet einen Kreis, alle Knoten erreichbar).

```
E' \leftarrow \emptyset
Sortiere E aufsteigend nach c(E).
Solange E \neq \emptyset
e \leftarrow \min(E)
E = E - \{e\}
Falls G(V, E' \cup \{e\}) kreisfrei
E' = E' \cup \{e\}
```

• Das Sortieren dominiert hier die Laufzeit: $O(|E|\log|E|)$.

Algorithmus von Kruskal: Beispiel

E

$$(1,3) \to 1$$

$$(4,6) \to 2$$

$$(2,5) \to 3$$

$$(3,6) \to 4$$

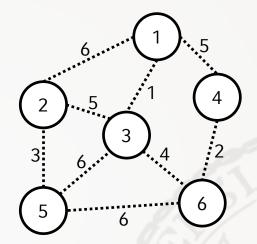
$$(1,4) \rightarrow 5$$

$$(2,3) \to 5$$

$$(1,2) \rightarrow 6$$

$$(3,5) \to 6$$

$$(5,6) \to 6$$



Algorithmus von Kruskal: Beispiel

E

$$(1,3) \rightarrow 1$$

$$(4,6) \to 2$$

$$(2,5) \to 3$$

$$(3,6) \to 4$$

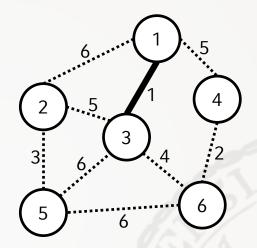
$$(1,4) \rightarrow 5$$

$$(2,3) \to 5$$

$$(1,2) \rightarrow 6$$

$$(3,5) \to 6$$

$$(5,6) \to 6$$



Algorithmus von Kruskal: Beispiel

E

$$(1,3) \rightarrow 1$$

$$(4,6) \to 2$$

$$(2,5) \to 3$$

$$(3,6) \to 4$$

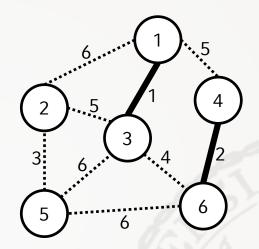
$$(1,4) \rightarrow 5$$

$$(2,3) \to 5$$

$$(1,2) \rightarrow 6$$

$$(3,5) \to 6$$

$$(5,6) \to 6$$



$$(1,3) \to 1$$

$$(4,6) \to 2$$

$$(2,5) \to 3$$

$$(3,6) \to 4$$

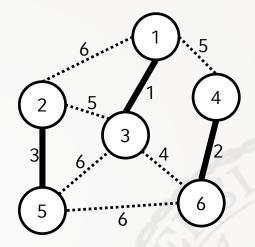
$$(1,4) \rightarrow 5$$

$$(2,3) \to 5$$

$$(1,2) \rightarrow 6$$

$$(3,5) \to 6$$

$$(5,6) \to 6$$



$$(1,3) \to 1$$

$$(4,6) \to 2$$

$$(2,5) \to 3$$

$$(3,6) \to 4$$

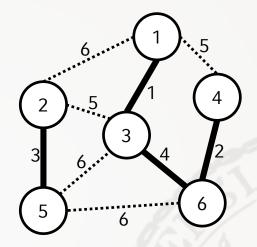
$$(1,4) \rightarrow 5$$

$$(2,3) \to 5$$

$$(1,2) \rightarrow 6$$

$$(3,5) \to 6$$

$$(5,6) \to 6$$



$$(1,3) \to 1$$

$$(4,6) \to 2$$

$$(2,5) \to 3$$

$$(3,6) \to 4$$

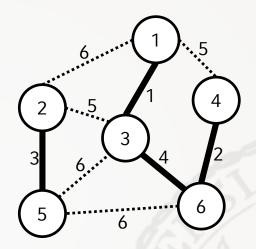
$$(1,4) \rightarrow 5 \rightarrow Kreis$$

$$(2,3) \to 5$$

$$(1,2) \rightarrow 6$$

$$(3,5) \to 6$$

$$(5,6) \to 6$$



$$(1,3) \to 1$$

$$(4,6) \to 2$$

$$(2,5) \to 3$$

$$(3,6) \to 4$$

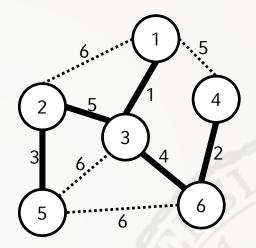
$$(1,4) \rightarrow 5$$

$$(2,3) \to 5$$

$$(1,2) \rightarrow 6$$

$$(3,5) \to 6$$

$$(5,6) \to 6$$



$$(1,3) \to 1$$

$$(4,6) \to 2$$

$$(2,5) \to 3$$

$$(3,6) \to 4$$

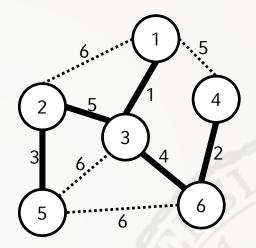
$$(1,4) \rightarrow 5$$

$$(2,3) \to 5$$

$$(1,2) \rightarrow 6 \rightarrow Kreis$$

$$(3,5) \to 6$$

$$(5,6) \to 6$$



\boldsymbol{E}

$$(1,3) \to 1$$

$$(4,6) \to 2$$

$$(2,5) \to 3$$

$$(3,6) \to 4$$

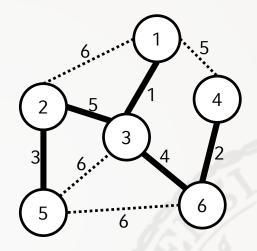
$$(1,4) \rightarrow 5$$

$$(2,3) \to 5$$

$$(1,2) \rightarrow 6$$

$$(3,5) \rightarrow 6 \rightarrow Kreis$$

$$(5,6) \to 6$$



$$(1,3) \to 1$$

$$(4,6) \to 2$$

$$(2,5) \to 3$$

$$(3,6) \to 4$$

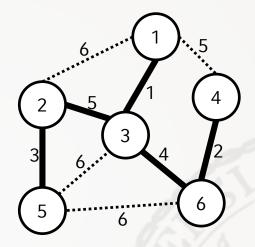
$$(1,4) \rightarrow 5$$

$$(2,3) \to 5$$

$$(1,2) \rightarrow 6$$

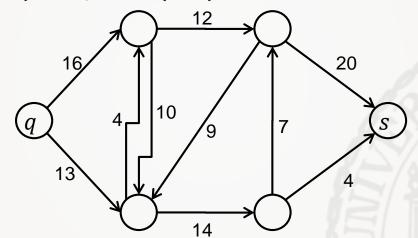
$$(3,5) \to 6$$

$$(5,6) \rightarrow 6 \rightarrow Kreis$$



Flussnetzwerke

- Ein Flussnetzwerk (G,c) ist ein gerichteter Graph G=(V,E), wobei
 - jede Kante $(u, v) \in E$ die Kapazität $c(u, v) \ge 0$ hat, und
 - es eine Quelle $q \in V$ und eine Senke $s \in V$ gibt.
- Wir setzen c(u,v)=0, falls $(u,v) \notin E$



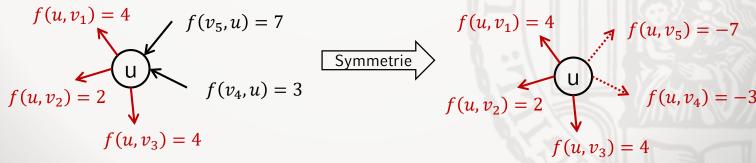
- anschaulich:
 - Wasserleitungen mit unterschiedlichen Kapazitäten
 - Verkehrsströme, Straßenkapazitäten
 - Kommunikationswege mit Bandbreiten

Fluss in einem Flussnetzwerk

- Ein Fluss ist eine Funktion $f: V \times V \to \mathbb{R}$ mit den Eigenschaften
 - Kapazitätsbeschränkung: Für $u, v \in V$ gilt $f(u, v) \leq c(u, v)$

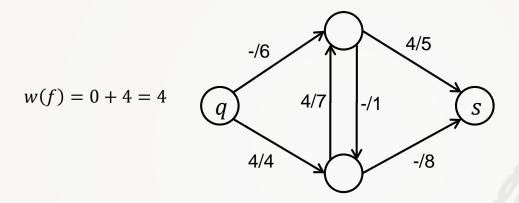
(u)
$$7/12$$
 (v) $f(u,v) = 7 \le 12 = c(u,v)$

- Symmetrie: Für $u, v \in V$ gilt f(u, v) = -f(v, u)
 - "u gibt v 7 Einheiten"
 - \rightarrow "v nimmt u 7 Einheiten"
 - \rightarrow "v gibt u-7 Einheiten"
- Flusserhaltung: Für $u \in V \{q, s\}$ gilt $\sum_{v \in V} f(u, v) = 0$



Maximaler Fluss

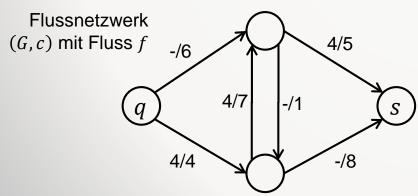
- Der Wert w(f) eines Flusses f ist definiert als $w(f) = \sum_{v \in V} f(q, v)$
 - entspricht Gesamtfluss aus der Quelle q heraus

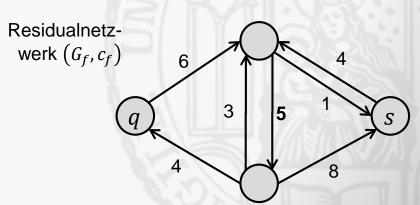


- Problem des maximalen Flusses
 - Gegeben ein Flussnetzwerk (G, c)
 - Gesucht ein Fluss f auf (G,c) mit maximalem Wert w(f)

Residualnetzwerk

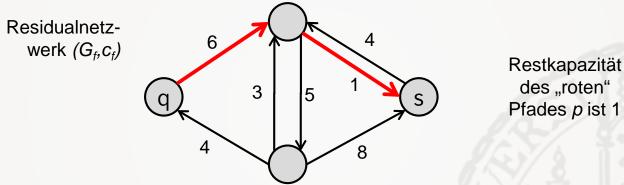
- Restkapazität $c_f(u,v)$ zwischen $u,v\in V$ ist $c_f(u,v)=c(u,v)-f(u,v)$
 - beachte: formal ist f(u, v) < 0 möglich (vgl. Symmetrie)
- Der Restgraph $G_f = (V, E_f)$ bzgl. Flussnetzwerk (G, c) und Fluss f ist definiert durch die Kantenmenge $E_f = \{(u, v) \in V \times V | c_f(u, v) > 0\}$
- (G_f, c_f) ist das sogenannte Residualnetzwerk \rightarrow "Flussnetzwerk minus Fluss = Residualnetzwerk"



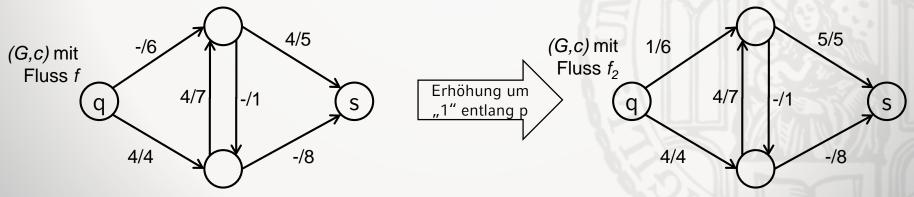


Flussvergrößernder Pfad

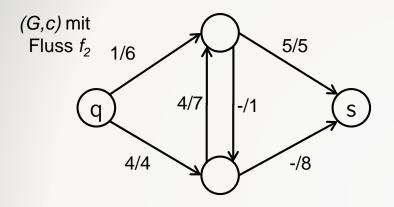
- Ein Pfad p von q nach s im Residualnetzwerk heißt flussvergrößernder oder augmentierender Pfad.
- Die Restkapazität von p ist $c_f(p) = \min\{c_f(u, v) | (u, v) \in p\}$

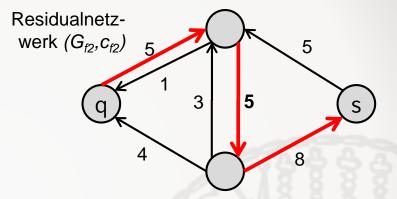


• Fluss f in (G,c) kann entlang des Pfades p um $c_f(p)$ erhöht werden

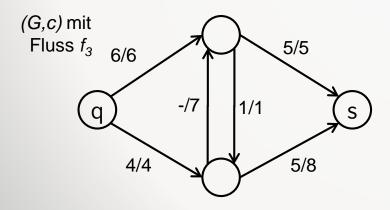


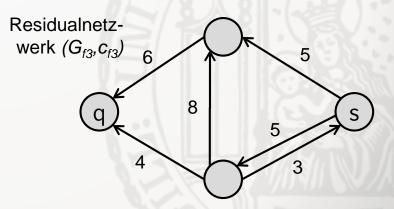
Beispiel (fortgesetzt)





Achtung: "Rückwärtskante" entlang p





kein Pfad von *q* nach *s* möglich → maximalen Fluss gefunden

Min-Cut-Max-Flow-Theorem

- Ein s-t-Schnitt (S,T) in einem Flussnetzwerk ist eine Partition der Knotenmenge in zwei disjunkte Mengen $S \cap T = \emptyset$ mit $s \in S$, $t \in T$.
- Die Kapazität eines Schnitts ist das Gesamtgewicht der Kanten von S nach T:

$$c(S,T) = \sum_{u \in S, v \in T \mid (u,v) \in E} c(u,v)$$

- Die folgenden Aussagen sind äquivalent:
 - − f ist der maximale Fluss in G.
 - Das Residualnetzwerk G_f enthält keinen augmentierenden Pfad.
 - Für mindestens einen Schnitt (S,T) ist der Wert des Flusses gleich der Kapazität des Schnittes: |f| = c(S,T)
- Damit gilt: Der maximale Fluss entspricht der Kapazität des minimalen Schnittes.

Ford-Fulkerson-Methode

Idee:

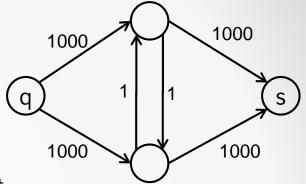
Initialisiere Fluss f mit 0Solange es einen flussvergrößernden Pfad p gibt erhöhe f entlang p

• Formal ist Erhöhung von f entlang p für alle $u, v \in V$ definiert durch

$$f_{neu}(u, v) = f_{alt}(u, v) + \begin{cases} c_f(p) & \text{falls } (u, v) \text{ auf } p \\ -c_f(p) & \text{falls } (v, u) \text{ auf } p \\ 0 & \text{sonst} \end{cases}$$

Laufzeit der Ford-Fulkerson-Methode

- Die Laufzeit kann beliebig schlecht sein
 - im Beispiel: wähle flussvergrößernden Pfad stets über mittlere Kanten
 - → sehr langsame Erhöhung des Gesamtflusses



- Falls alle Kapazitäten ganzzahlig sind, benötigt die Methode O(f*) Iterationen, um das Problem zu lösen (dabei ist f* der Wert des maximalen Flusses)
 - in jeder Iteration wird der Wert des Flusses um $c_f(p) \ge 1$ erhöht
 - zu Beginn 0 und am Ende f*
- Verbesserung:
 - wähle einen <u>kürzesten</u> flussvergrößernden Pfad
 - im Beispiel würden mittlere Kanten vermieden → schnellere Terminierung
 - → Algorithmus von Edmonds und Karp

Edmonds-Karp-Algorithmus

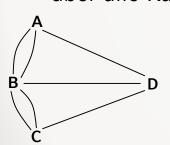
Idee:

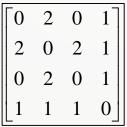
Initialisiere Fluss f mit 0Solange es einen flussvergrößernden Pfad p gibt finde einen kürzesten flussvergrößernden Pfad p erhöhe f entlang p

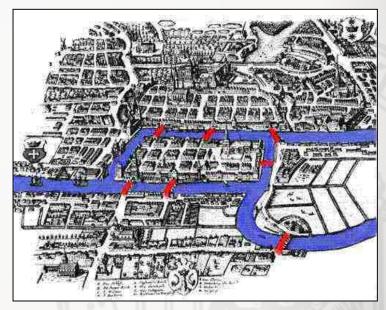
- Laufzeit der Methode ist polynomiell in der Größe des Netzwerks
 - $-O(|V| \cdot |E|^2) = O(|V|^5)$ bei spezieller Implementierung
- Weitere Verbesserungen durch Dinic (1970) führten zu
 - $O(|V|^2 \cdot |E|) = O(|V|^4)$

Graph-Anwendungen: Euler-Tour

- Historisches Problem ("Königsberger Brückenproblem):
 - 1736 lebte der deutsche Mathematiker Leonhard Euler in Königsberg
 - Fluß Pregel bildete dort eine Insel mit mehreren Brücken.
 - Häufige Frage: Ist ein Spaziergang möglich, so dass man
 - schließlich wieder am Ausgangspunkt ankommt und
 - alle Brücken genau einmal überquert?
- Graphentheoretisch:
 - "Geschlossene Euler-Tour"
 - Existiert geschlossener, einfacher Pfad über alle Kanten?







 Eulers Antwort: Genau dann, wenn alle Knoten von geradem Grad sind bzw. die Spalten- / Zeilensummen der Adjazenzmatrix alle gerade sind.