

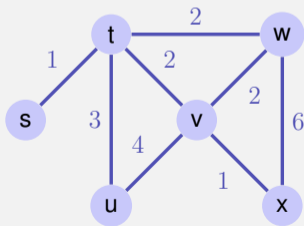
15. Minimale Spann­b­äume

Motivation, Greedy, Algorithmus von Kruskal, Allgemeine Regeln, Union-Find Struktur, Algorithmus von Jarnik, Prim, Dijkstra, [Ottman/Widmayer, Kap. 9.6, 6.2, 6.1, Cormen et al, Kap. 23, 19]

Problem

Gegeben: Ungerichteter, zusammenhängender, gewichteter Graph $G = (V, E, c)$.

Gesucht: Minimaler Spannbaum $T = (V, E')$: zusammenhängender, zyklenfreier Teilgraph $E' \subset E$, so dass $\sum_{e \in E'} c(e)$ minimal.



Beispiele von Anwendungen

- Netzwerk-Design: finde das billigste / kürzeste Netz oder Leitungssystem, welches alle Knoten miteinander verbindet.
- Approximation einer Lösung des Travelling-Salesman Problems: finde einen möglichst kurzen Rundweg, welcher jeden Knoten einmal besucht. ²⁵

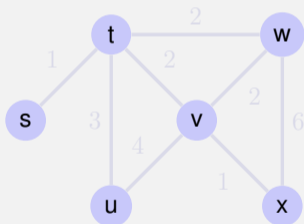
²⁵Der beste bekannte Algorithmus zur exakten Lösung des TS Problems hat exponentielle Laufzeit

Greedy Verfahren

- Gierige Verfahren berechnen eine Lösung schrittweise, indem lokal beste Lösungen gewählt werden.
- Die meisten Probleme sind nicht mit einer greedy Strategie lösbar.
- Das Problem des Minimalen Spannbaumes kann mit einem gierigen Verfahren effizient gelöst werden.

Greedy Idee (Kruskal, 1956)

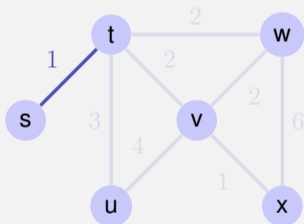
Konstruiere T indem immer die billigste Kante hinzugefügt wird, welche keinen Zyklus erzeugt.



(Lösung ist nicht eindeutig.)

Greedy Idee (Kruskal, 1956)

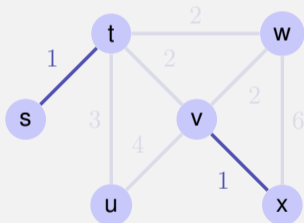
Konstruiere T indem immer die billigste Kante hinzugefügt wird, welche keinen Zyklus erzeugt.



(Lösung ist nicht eindeutig.)

Greedy Idee (Kruskal, 1956)

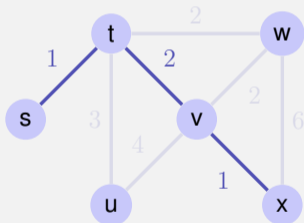
Konstruiere T indem immer die billigste Kante hinzugefügt wird, welche keinen Zyklus erzeugt.



(Lösung ist nicht eindeutig.)

Greedy Idee (Kruskal, 1956)

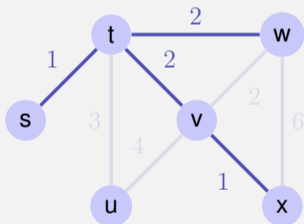
Konstruiere T indem immer die billigste Kante hinzugefügt wird, welche keinen Zyklus erzeugt.



(Lösung ist nicht eindeutig.)

Greedy Idee (Kruskal, 1956)

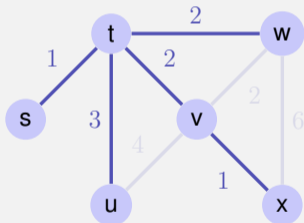
Konstruiere T indem immer die billigste Kante hinzugefügt wird, welche keinen Zyklus erzeugt.



(Lösung ist nicht eindeutig.)

Greedy Idee (Kruskal, 1956)

Konstruiere T indem immer die billigste Kante hinzugefügt wird, welche keinen Zyklus erzeugt.



(Lösung ist nicht eindeutig.)

Algorithmus MST-Kruskal(G)

Input: Gewichteter Graph $G = (V, E, c)$

Output: Minimaler Spannbaum mit Kanten A .

Sortiere Kanten nach Gewicht $c(e_1) \leq \dots \leq c(e_m)$

$A \leftarrow \emptyset$

for $k = 1$ **to** $|E|$ **do**

if $(V, A \cup \{e_k\})$ kreisfrei **then**
 $A \leftarrow A \cup \{e_k\}$

return (V, A, c)

[Korrektheit]

Zu jedem Zeitpunkt ist (V, A) ein Wald, eine Menge von Bäumen.

MST-Kruskal betrachtet jede Kante e_k einmal und wählt e_k oder verwirft e_k

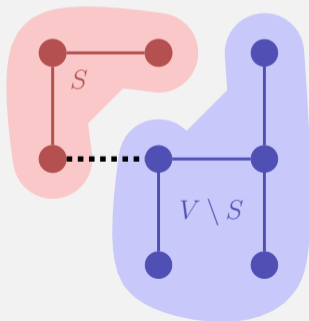
Notation (Momentaufnahme im Algorithmus)

- A : Menge gewählte Kanten
- R : Menge verworfener Kanten
- U : Menge der noch unentschiedenen Kanten

[Schnitt]

Ein Schnitt von G ist eine Partition $S, V \setminus S$ von V . ($S \subseteq V$).

Eine Kante kreuzt einen Schnitt, wenn einer Ihrer Endpunkte in S und der andere in $V \setminus S$ liegt.



[Regeln]

- 1 Auswahlregel: Wähle einen Schnitt, den keine gewählte Kante kreuzt. Unter allen unentschiedenen Kanten, welche den Schnitt kreuzen, wähle die mit minimalem Gewicht.
- 2 Verwerfregel: Wähle einen Kreis ohne verworfene Kanten. Unter allen unentschiedenen Kanten im Kreis verwerfe die mit maximalem Gewicht.

[Regeln]

Kruskal wendet beide Regeln an:

- 1 Ein gewähltes e_k verbindet zwei Zusammenhangskomponenten, sonst würde ein Kreis erzeugt werden. e_k ist beim Verbinden minimal, man kann also einen Schnitt wählen, den e_k mit minimalem Gewicht kreuzt.
- 2 Ein verworfenes e_k ist Teil eines Kreises. Innerhalb des Kreises hat e_k maximales Gewicht.

[Korrektheit]

Theorem

Jeder Algorithmus, welcher schrittweise obige Regeln anwendet bis $U = \emptyset$ ist korrekt.

Folgerung: MST-Kruskal ist korrekt.

[Auswahlinvariante]

Invariante: Es gibt stets einen minimalen Spannbaum, der alle gewählten und keine der verworfenen Kanten enthält.

Wenn die beiden Regeln die Invariante erhalten, dann ist der Algorithmus sicher korrekt. Induktion:

- Zu Beginn: $U = E, R = A = \emptyset$. Invariante gilt offensichtlich.
- Invariante bleibt nach jedem Schritt des Algorithmus erhalten.
- Am Ende: $U = \emptyset, R \cup A = E \Rightarrow (V, A)$ ist Spannbaum.

Beweis des Theorems: zeigen nun, dass die beiden Regeln die Invariante erhalten.

[Auswahlregel erhält Invariante]

Es gibt stets einen minimalen Spannbaum T , der alle gewählten und keine der verworfenen Kanten enthält.

Wähle einen Schnitt, den keine gewählte Kante kreuzt. Unter allen unentschiedenen Kanten, welche den Schnitt kreuzen, wähle eine Kante e mit minimalem Gewicht.

- Fall 1: $e \in T$ (fertig)
- Fall 2: $e \notin T$. Dann hat $T \cup \{e\}$ einen Kreis, der e enthält. Kreis muss eine zweite Kante e' enthalten, welche den Schnitt auch kreuzt.²⁶ Da $e' \notin R$ ist $e' \in U$. Somit $c(e) \leq c(e')$ und $T' = T \setminus \{e'\} \cup \{e\}$ ist auch minimaler Spannbaum (und $c(e) = c(e')$).

²⁶Ein solcher Kreis enthält mindestens einen Knoten in S und einen in $V \setminus S$ und damit mindestens zwei Kanten zwischen S und $V \setminus S$.

[Verwerfregel erhält Invariante]

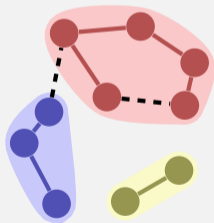
Es gibt stets einen minimalen Spannbaum T , der alle gewählten und keine der verworfenen Kanten enthält.

Wähle einen Kreis ohne verworfene Kanten. Unter allen unentschiedenen Kanten im Kreis verwerfe die Kante e mit maximalem Gewicht.

- Fall 1: $e \notin T$ (fertig)
- Fall 2: $e \in T$. Entferne e von T , Das ergibt einen Schnitt. Diesen Schnitt muss eine weitere Kante e' aus dem Kreis kreuzen. Da $c(e') \leq c(e)$ ist $T' = T \setminus \{e\} \cup \{e'\}$ auch minimal (und $c(e) = c(e')$).

Zur Implementation

Gegeben eine Menge von Mengen $i \equiv A_i \subset V$. Zur Identifikation von Schnitten und Kreisen: Zugehörigkeit der beiden Endpunkte einer Kante zu einer der Mengen.



Zur Implementation

Allgemeines Problem: Partition (Menge von Teilmengen) z.B.
 $\{\{1, 2, 3, 9\}, \{7, 6, 4\}, \{5, 8\}, \{10\}\}$

Benötigt: Abstrakter Datentyp „Union-Find“ mit folgenden Operationen

- $\text{Make-Set}(i)$: Hinzufügen einer neuen Menge i .
- $\text{Find}(e)$: Name i der Menge, welche e enthält.
- $\text{Union}(i, j)$: Vereinigung der Mengen mit Namen i und j .

Union-Find Algorithmus MST-Kruskal(G)

Input: Gewichteter Graph $G = (V, E, c)$

Output: Minimaler Spannbaum mit Kanten A .

Sortiere Kanten nach Gewicht $c(e_1) \leq \dots \leq c(e_m)$

$A \leftarrow \emptyset$

for $k = 1$ **to** $|V|$ **do**

\lfloor MakeSet(k)

for $k = 1$ **to** m **do**

$(u, v) \leftarrow e_k$

if Find(u) \neq Find(v) **then**

 Union(Find(u), Find(v))

$A \leftarrow A \cup e_k$

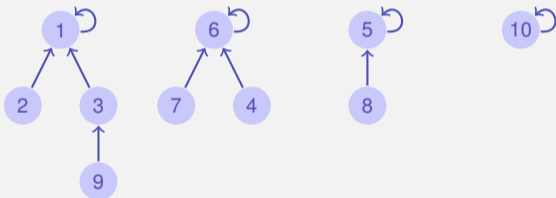
else

// konzeptuell: $R \leftarrow R \cup e_k$

return (V, A, c)

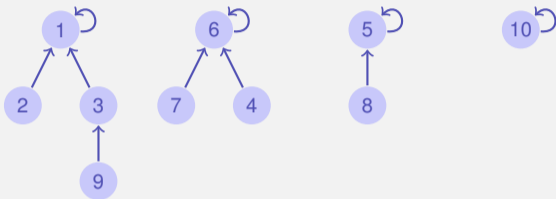
Implementation Union-Find

Idee: Baum für jede Teilmenge in der Partition, z.B.
 $\{\{1, 2, 3, 9\}, \{7, 6, 4\}, \{5, 8\}, \{10\}\}$



Baumwurzeln = Namen (Stellvertreter) der Mengen,
Bäume = Elemente der Mengen

Implementation Union-Find



Repräsentation als Array:

Index	1	2	3	4	5	6	7	8	9	10
Parent	1	1	1	6	5	6	5	5	3	10

Implementation Union-Find

Index	1	2	3	4	5	6	7	8	9	10
Parent	1	1	1	6	5	6	5	5	3	10

Make-Set(i) $p[i] \leftarrow i$; **return** i

Find(i) **while** ($p[i] \neq i$) **do** $i \leftarrow p[i]$
 return i

Union(i, j)²⁷ $p[j] \leftarrow i$;

²⁷ i und j müssen Namen (Wurzeln) der Mengen sein. Andernfalls verwende `Union(Find(i), Find(j))`

Optimierung der Laufzeit für Find

Baum kann entarten: Beispiel Union(8, 7), Union(7, 6), Union(6, 5), ...

Index	1	2	3	4	5	6	7	8	..
Parent	1	1	2	3	4	5	6	7	..

Laufzeit von Find im schlechtesten Fall in $\Theta(n)$.

Optimierung der Laufzeit für Find

Idee: Immer kleineren Baum unter grösseren Baum hängen.
Benötigt zusätzliche Grösseninformation (Array) g

Make-Set(i) $p[i] \leftarrow i; g[i] \leftarrow 1; \text{return } i$

Union(i, j) **if** $g[j] > g[i]$ **then** $\text{swap}(i, j)$
 $p[j] \leftarrow i$
 if $g[i] = g[j]$ **then** $g[i] \leftarrow g[i] + 1$

\Rightarrow Baumtiefe (und schlechteste Laufzeit für Find) in $\Theta(\log n)$

[Beobachtung]

Theorem

Obiges Verfahren Vereinigung nach Grösse konserviert die folgende Eigenschaft der Bäume: ein Baum mit Höhe h hat mindestens 2^h Knoten.

Unmittelbare Folgerung: Laufzeit Find = $\mathcal{O}(\log n)$.

[Beweis]

Induktion: nach Voraussetzung haben Teilbäume jeweils mindestens 2^{h_i} Knoten. ObdA:
 $h_2 \leq h_1$.

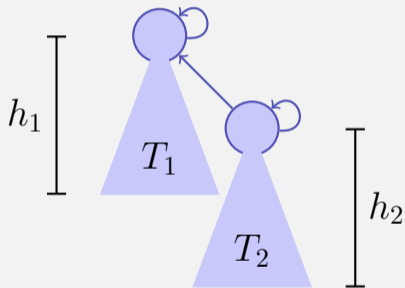
■ $h_2 < h_1$:

$$h(T_1 \oplus T_2) = h_1 \Rightarrow g(T_1 \oplus T_2) \geq 2^{h_1}$$

■ $h_2 = h_1$:

$$g(T_1) \geq g(T_2) \geq 2^{h_2}$$

$$\Rightarrow g(T_1 \oplus T_2) = g(T_1) + g(T_2) \geq 2 \cdot 2^{h_2} = 2^{h_2+1} = 2^{h(T_1 \oplus T_2)}$$



Weitere Verbesserung

Bei jedem Find alle Knoten direkt an den Wurzelknoten hängen.

Find(i):

$j \leftarrow i$

while ($p[i] \neq i$) **do** $i \leftarrow p[i]$

while ($j \neq i$) **do**

$t \leftarrow j$
 $j \leftarrow p[j]$
 $p[t] \leftarrow i$

return i

Laufzeit: amortisiert *fast* konstant (Inverse der Ackermannfunktion).²⁸

²⁸Wird hier nicht vertieft.

Laufzeit des Kruskal Algorithmus

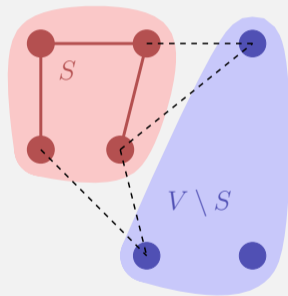
- Sortieren der Kanten: $\Theta(|E| \log |E|) = \Theta(|E| \log |V|)$.²⁹
 - Initialisieren der Union-Find Datenstruktur $\Theta(|V|)$
 - $|E| \times \text{Union}(\text{Find}(x), \text{Find}(y))$: $\mathcal{O}(|E| \log |E|) = \mathcal{O}(|E| \log |V|)$.
- Insgesamt $\Theta(|E| \log |V|)$.

²⁹da G zusammenhängend: $|V| \leq |E| \leq |V|^2$

Algorithmus von Jarnik (1930), Prim, Dijkstra (1959)

Idee: Starte mit einem $v \in V$ und lasse von dort unter Verwendung der Auswahlregel einen Spannbaum wachsen:

```
A ← ∅  
S ← {v0}  
for i ← 1 to |V| do  
  Wähle billigste (u, v) mit u ∈ S, v ∉ S  
  A ← A ∪ {(u, v)}  
  S ← S ∪ {v} // (Färbung)
```



Anmerkung: man benötigt keine Union-Find Datenstruktur. Es genügt, Knoten zu färben, sobald sie zu S hinzugenommen werden.

Laufzeit

Trivial $\mathcal{O}(|V| \cdot |E|)$.

Verbesserung (wie bei Dijkstras Kürzeste Pfade):

■ Mit Min-Heap, Kosten:

- Initialisierung (Knotenfärbung) $\mathcal{O}(|V|)$
- $|V| \times \text{ExtractMin} = \mathcal{O}(|V| \log |V|)$,
- $|E| \times \text{Insert oder DecreaseKey} = \mathcal{O}(|E| \log |V|)$,

$\mathcal{O}(|E| \cdot \log |V|)$