

Fortgeschrittene Netzwerk- und Graph-Algorithmen

Prof. Dr. Hanjo Täubig

Lehrstuhl für Effiziente Algorithmen
(Prof. Dr. Ernst W. Mayr)
Institut für Informatik
Technische Universität München

Wintersemester 2010/11



Übersicht

- 1 Algorithmen für Zentralitätsindizes
 - Berechnung der Betweenness-Zentralitäten
 - Das Absolute 1-Center Problem

Betweenness Centrality

$$c_B(v) = \sum_{s \in V \setminus \{v\}} \sum_{t \in V \setminus \{v\}} \delta_{st}(v) = \sum_{s \in V \setminus \{v\}} \sum_{t \in V \setminus \{v\}} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

Mögliche Berechnung:

- 1 Berechne Länge und Anzahl kürzester Pfade zwischen allen Knotenpaaren
- 2 Betrachte zu jedem Knoten v alle möglichen Paare s, t und berechne den Anteil kürzester Pfade durch v

Bedingung(Bellman-Kriterium):

$$d(s, t) = d(s, v) + d(v, t)$$

Betweenness Centrality

1 Modifiziere BFS bzw. Dijkstras Algorithmus

- ▶ Ersetze einzelne Vorgängerknoten durch eine Vorgängermenge
 $\text{pred}(s, v) = \{u \in V : \{u, v\} \in E, d(s, v) = d(s, u) + w(u, v)\}$
- ▶ Es gilt dann

$$\sigma_{sv} = \sum_{u \in \text{pred}(s, v)} \sigma_{su}$$

- ⇒ Für einen Startknoten $s \in V$ kann die Anzahl kürzester Pfade zu jedem anderen Knoten in $\mathcal{O}(m + n \log n)$ für **gewichtete** und in $\mathcal{O}(m)$ für **ungewichtete** Graphen berechnet werden.
- ⇒ Die Berechnung von σ_{st} für alle Paare $s, t \in V$ kann in $\mathcal{O}(mn + n^2 \log n)$ bzw. $\mathcal{O}(mn)$ berechnet werden.

Betweenness Centrality

- 2 Im Fall $d(s, t) = d(s, v) + d(v, t)$ gilt

$$\sigma_{st}(v) = \sigma_{sv} \cdot \sigma_{vt}$$

ansonsten ($d(s, t) < d(s, v) + d(v, t)$) ist $\sigma_{st}(v) = 0$

⇒ naive Berechnung:

$\mathcal{O}(n^2)$ pro Knoten v (Summation über alle $s \neq v \neq t$),
also insgesamt **Zeit $\mathcal{O}(n^3)$** und **Platz $\mathcal{O}(n^2)$** für Speicherung aller
 $d(s, t)$ und σ_{st}

Betweenness Centrality (Brandes)

Abhängigkeit eines Paares (s, t) von v :

$$\delta_{st}(v) = \frac{\sigma_{st}(v)}{\sigma_{st}}$$

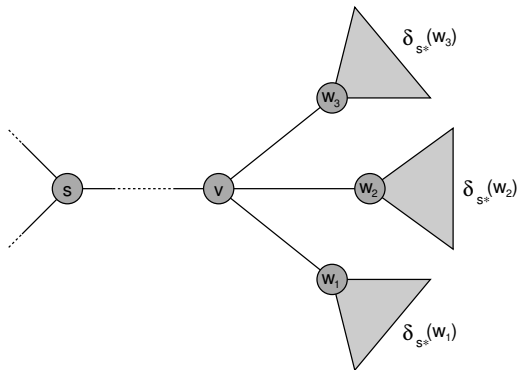
Abhängigkeit eines Startknotens s von v :

$$\delta_{s^*}(v) = \sum_{t \in V \setminus \{v\}} \delta_{st}(v)$$

Betweenness von v :

$$\Rightarrow c_B(v) = \sum_{s \in V \setminus \{v\}} \delta_{s^*}(v)$$

Betweenness Centrality (Brandes)



Betweenness Centrality (Brandes)

Lemma

Für den Fall, dass es von $s \in V$ **genau einen** kürzesten Pfad zu jedem $t \in V$ gibt, gilt

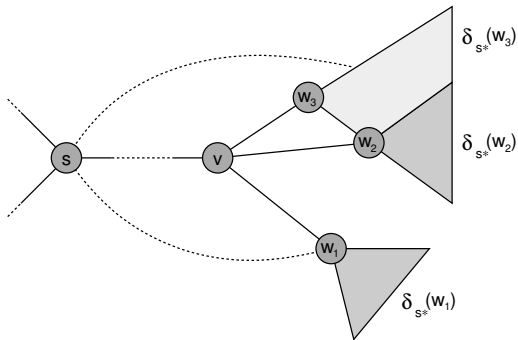
$$\delta_{s^*}(v) = \sum_{w: v \in \text{pred}(s,w)} (1 + \delta_{s^*}(w))$$

Beweis.

- Die kürzesten Pfade von s formen einen Baum.
- Also liegt v auf *allen* kürzesten Pfaden von s zu einem t oder auf *keinem*, d.h. $\delta_{st}(v)$ ist 1 oder 0.
- v liegt auf allen kürzesten Pfaden zu den Nachfolgern, sowie auf allen kürzesten Pfaden, auf denen diese Nachfolger liegen.



Betweenness Centrality (Brandes)



Im allgemeinen Fall werden die Anteile der Abhängigkeiten von den Nachfolgern über die Kanten des Kürzeste-Pfade-DAGs propagiert.

Betweenness Centrality (Brandes)

Satz

Für die Abhängigkeit $\delta_{s^*}(v)$ eines Startknotens $s \in V$ von den anderen Knoten $v \in V$ gilt:

$$\delta_{s^*}(v) = \sum_{w: v \in \text{pred}(s,w)} \frac{\sigma_{sv}}{\sigma_{sw}} (1 + \delta_{s^*}(w))$$

Betweenness Centrality (Brandes)

Beweis.

- $\delta_{st}(v) > 0$ gilt nur für die $t \in V \setminus \{s\}$, für die v auf mindestens einem kürzesten Pfad von s nach t liegt.
- Jeder solche Pfad hat exakt eine Kante $\{v, w\}$ mit $v \in \text{pred}(s, w)$ (siehe Abbildung).

Betweenness Centrality (Brandes)

Beweis.

- Definiere $\sigma_{st}(v, e)$: Anzahl kürzester s - t -Pfade, die sowohl Knoten v als auch Kante e enthalten
- Definiere entsprechend

$$\delta_{st}(v, e) = \frac{\sigma_{st}(v, e)}{\sigma_{st}}$$

$$\begin{aligned} \delta_{s^*}(v) &= \sum_{t \in V \setminus \{v\}} \delta_{st}(v) = \sum_{t \in V \setminus \{v\}} \left(\sum_{w: v \in \text{pred}(s, w)} \delta_{st}(v, \{v, w\}) \right) \\ &= \sum_{w: v \in \text{pred}(s, w)} \left(\sum_{t \in V \setminus \{v\}} \delta_{st}(v, \{v, w\}) \right) \end{aligned}$$

Betweenness Centrality (Brandes)

Beweis.

- Betrachte Knoten w , so dass $v \in \text{pred}(s, w)$
- σ_{sw} kürzeste Pfade von s nach w ,
davon σ_{sv} von s nach v gefolgt von Kante $\{v, w\}$
- Anteil σ_{sv}/σ_{sw} der Anzahl kürzester Pfade von s nach t über w benutzt auch die Kante $\{v, w\}$:

$$\delta_{st}(v, \{v, w\}) = \begin{cases} \frac{\sigma_{sv}}{\sigma_{sw}} & \text{falls } t = w \\ \frac{\sigma_{sv}}{\sigma_{sw}} \cdot \frac{\sigma_{st}(w)}{\sigma_{st}} & \text{falls } t \neq w \end{cases}$$

Betweenness Centrality (Brandes)

Beweis.

$$\begin{aligned}
 \delta_{s^*}(v) &= \sum_{w: v \in \text{pred}(s,w)} \left(\sum_{t \in V \setminus \{v\}} \delta_{st}(v, \{v, w\}) \right) \\
 &= \sum_{w: v \in \text{pred}(s,w)} \left(\frac{\sigma_{sv}}{\sigma_{sw}} + \sum_{t \in V \setminus \{v,w\}} \frac{\sigma_{sv}}{\sigma_{sw}} \cdot \frac{\sigma_{st}(w)}{\sigma_{st}} \right) \\
 &= \sum_{w: v \in \text{pred}(s,w)} \left(\frac{\sigma_{sv}}{\sigma_{sw}} + \frac{\sigma_{sv}}{\sigma_{sw}} \sum_{t \in V \setminus \{v,w\}} \frac{\sigma_{st}(w)}{\sigma_{st}} \right) \\
 &= \sum_{w: v \in \text{pred}(s,w)} \frac{\sigma_{sv}}{\sigma_{sw}} (1 + \delta_{s^*}(w))
 \end{aligned}$$

Betweenness Centrality (Brandes)

Beweis.

Erklärung:

- Zuerst wird die Summe in die beiden Fälle $t = w$ und $t \neq w$ aufgeteilt.
- Dann wird ausgeklammert.
- Es gilt

$$\sum_{t \in V \setminus \{v, w\}} \frac{\sigma_{st}(w)}{\sigma_{st}} = \sum_{t \in V \setminus \{w\}} \frac{\sigma_{st}(w)}{\sigma_{st}} = \delta_{s^*}(w)$$

weil aufgrund der Annahme, dass v ein Vorgänger von w bezüglich Startknoten s ist, für den Fall $t = v$ gilt, dass $\sigma_{st}(w) = \sigma_{sv}(w) = 0$



Betweenness Centrality (Brandes)

- Berechne n kürzeste-Pfade-DAGs (einen für jeden Startknoten $s \in V$)
- Berechne nacheinander für alle $s \in V$ aus dem kürzeste-Pfade-DAG von s die Abhängigkeiten $\delta_{s*}(v)$ für alle anderen Knoten $v \in V$

Vorgehen: rückwärts, von den Blättern im kürzeste-Pfade-Baum bzw. von der entferntesten Schicht im kürzeste-Pfade-DAG zum Startknoten hin

- Summiere die einzelnen Abhängigkeiten (kann schon parallel während der Berechnung aufsummiert werden, um nicht $\mathcal{O}(n^2)$ Platz zu verbrauchen)

Betweenness Centrality (Brandes)

Satz

Die Betweenness-Zentralität $c_B(v)$ für alle Knoten $v \in V$ kann

- für *ungewichtete* Graphen in $\mathcal{O}(nm)$
- für *gewichtete* Graphen in Zeit $\mathcal{O}(n(m + n \log n)) = \mathcal{O}(nm + n^2 \log n)$

berechnet werden.

Der Algorithmus benötigt dabei nur $\mathcal{O}(n + m)$ Speicherplatz.

Bemerkung:

Die anderen auf kürzesten Pfaden basierenden Zentralitäten kann man relativ einfach mit SSSP-Traversierung berechnen.

Absolute p -Center Problem

gegeben:

- ungerichteter Graph $G = (V, E)$
- nichtnegative Gewichtsfunktion $w(v)$ für Knoten $v \in V$
- positive Länge $\ell(e)$ für Kante $e \in E$

Sei $X_p = \{x_1, x_2, \dots, x_p\}$ eine Menge von p Punkten auf G (dürfen auf Knoten oder an beliebigen Kantenpositionen liegen).

$d(v, x_i)$: Länge eines kürzesten Pfades zwischen v und x_i

Distanz eines Knotens zur Positionsmenge:

$$d(v, X_p) = \min_{1 \leq i \leq p} \{d(v, x_i)\}$$

Definition: Absolute p -Center

Weiteste gewichtete Distanz eines Knotens $v \in V$ zu X_p :

$$F(X_p) = \max_{v \in V} \{w(v) \cdot d(v, X_p)\}$$

X_p^* : eine optimale Menge aus p Punkten:

$$F(X_p^*) = \min_{X_p \text{ auf } G} \{F(X_p)\}$$

Man bezeichnet X_p^* als [absolute] p -center
und $r_p = F(X_p^*)$ als [absolute] p -radius von G .

Falls X_p beschränkt ist auf (echte) Knoten des Graphen, dann spricht man vom **vertex p -center** bzw. **vertex p -radius**.

Annahmen

- Alle Knoten haben gleiches Gewicht $w(v) = c$,
o.B.d.A. $c = 1$ (ungewichteter Fall).
- Graph enthält keine Schleifen oder Multikanten.
- Länge jeder Kante $e = (v_r, v_s)$ ist gleich der Distanz der inzidenten Knoten, also $\ell(e) = d(v_r, v_s)$
(ansonsten könnte man e einfach eliminieren, ohne dass sich der p -Radius ändert)
- Distanzen zwischen allen Knotenpaaren sind bekannt.
(Falls diese erst berechnet werden müssen, erhöht sich die Komplexität um den Aufwand zur Lösung des APSP-Problems.)

Hinweis: Domination Number / Set als inverses Problem

gegeben:

- $G = (V, E)$
- natürliche Zahl r

gesucht:

- kleinste natürliche Zahl p , so dass der p -Radius von G nicht größer als r ist.

Problem:

- p heißt [absolute] domination number für Radius r .
- Ein entsprechendes p -center heißt [absolute] domination set für Radius r .
- Für ungewichtete Graphen und Radius 1 ergibt sich das bekannte Problem von domination number/set (welches \mathcal{NP} -vollständig ist!)

1-center / local center

- Vereinfachung: einelementige Menge $X_1 = \{x\}$

$$F(x) = \max_{v \in V} \{w(v) \cdot d(v, x)\}$$

- Um ein 1-center zu finden, sucht man nach einem “lokalen” Zentrum auf jeder Kante.
- Ein **local center** auf Kante $e \in E$ ist eine Position $x^*(e)$ auf e , so dass

$$F(x^*(e)) = \min_{x(e) \text{ auf } e} \{F(x(e))\}$$

wobei $r(e) = F(x^*(e))$ **lokaler Radius** von G auf e heißt.

$$r_1 = r(e_j) = \min_{1 \leq i \leq |E|} \{r(e_i)\}$$

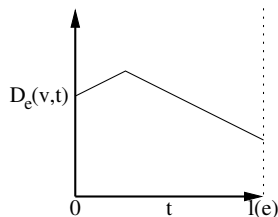
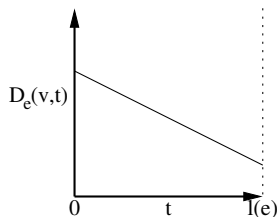
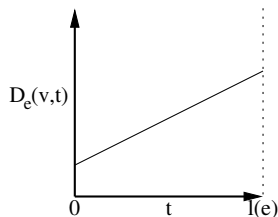
r_1 : 1-radius von G und das entsprechende $x^*(e_j)$ ist ein 1-center.

1-center / local center

- Um ein local center auf Kante $e = (v_r, v_s)$ zu finden, betrachten wir für jeden Knoten $v \in V$ seine (gewichtete) Distanz zu einem Punkt $x(e)$ auf e .
- Sei $t = t(x(e))$ die Distanz von $x(e)$ zu v_r auf Kante e .
Dann ist die gewichtete Distanz von v zu $x(e)$

$$D_e(v, t) = w(v) \cdot \min \left\{ \begin{array}{l} t + d(v_r, v), \\ \ell(e) - t + d(v_s, v) \end{array} \right\}$$

1-center / local center

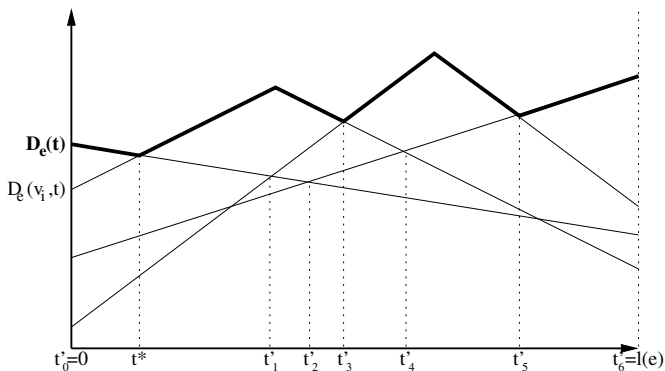


- Jede solche Funktion besteht aus ein oder zwei linearen Stücken mit Anstieg $\pm w(v)$.
- Im Fall von zwei Segmenten ist der Knickpunkt ein **Maximum**.

1-center / local center

Definiere für $0 \leq t \leq \ell(e)$

$$D_e(t) = \max_{v \in V} \{D_e(v, t)\}$$



1-center / local center

- Jeder Punkt $x^*(e)$, an dem $D_e(t)$ sein (absolutes) Minimum annimmt, ist ein local center auf e .
- $t^* = t(x^*(e))$ sei der Wert von t an dieser Stelle $x^*(e)$.
- Der Wert $D_e(t^*)$ ist der local-radius auf e .
- t^* ist ein beliebiger Punkt für den folgendes gilt:
 - 1 $t^* = 0$ oder $t^* = \ell(e)$ oder t^* ist ein Punkt, wo sich zwei Funktionen $D_e(v_i, t)$ und $D_e(v_j, t)$ so schneiden, dass ihre Anstiege im Schnittpunkt gegensätzliche Vorzeichen haben.
 - 2 Wenn T^* die Menge aller Punkte ist, die diese erste Voraussetzung erfüllen, dann ist

$$D_e(t^*) = \min_{t' \in T^*} \{D_e(t')\}$$

⇒ Höchstens $n(n-1)/2 + 2$ Punkte kommen als local-center von e in Frage.