

Fortgeschrittene Netzwerk- und Graph-Algorithmen

Dr. Hanjo Täubig

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

Wintersemester 2009/10



Übersicht

- 1 Zentralitätsindizes
 - Grad- und Distanz-basierte Zentralitäten
 - Kürzeste Pfade und Zentralität
 - Abgeleitete Kantenzentralitäten
 - Vitalität

Zentroid-Wert

Konkurrenzsituation:

- Knoten repräsentieren Kunden, die beim nächstgelegenen Geschäft einkaufen
- Annahme: zwei Anbieter, der zweite Anbieter berücksichtigt bei der Standortwahl den Standort des ersten Anbieters
- Fragen:
 - Welchen Standort muss der erste Anbieter wählen, damit er durch den zweiten Anbieter möglichst wenig Kunden verliert?
 - Ist es vorteilhaft als Erster auswählen zu können?

Zentroid-Wert

geg.: ungerichteter zusammenhängender Graph G

Für zwei Knoten u und v sei $\gamma_u(v)$ die Anzahl der Knoten, deren Distanz zu u kleiner ist als zu v :

$$\gamma_u(v) = |\{w \in V : d(u, w) < d(v, w)\}|$$

Wähle Knoten u , Gegenspieler wählt Knoten v

Resultierende Anzahl Kunden:

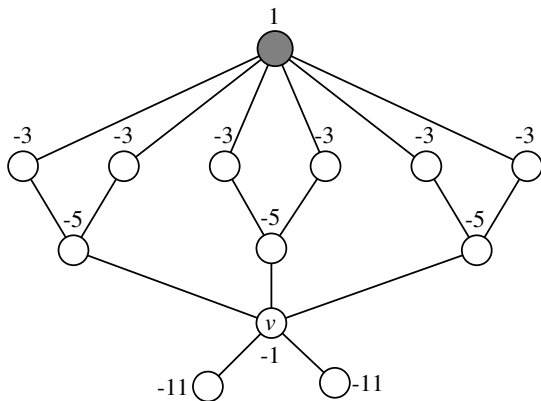
$$\Rightarrow \gamma_u(v) + \frac{n - \gamma_u(v) - \gamma_v(u)}{2} = \frac{n + \gamma_u(v) - \gamma_v(u)}{2}$$

\Rightarrow Gegenspieler minimiert $f(u, v) = \gamma_u(v) - \gamma_v(u)$

Zentralitätsmaß:

$$c_F(u) = \min_{v \in (V \setminus u)} \{f(u, v)\}$$

Zentroid-Wert / Beispiel 1

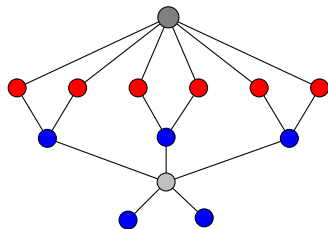
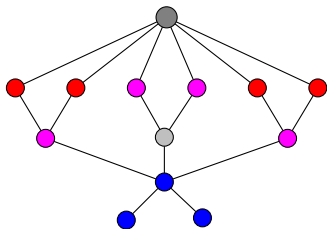


(Quelle: Brandes/Erlebach (Eds.): Network Analysis)

Der graue Knoten hat maximalen Zentroid-Wert,
aber v ist der Knoten mit maximaler Closeness.

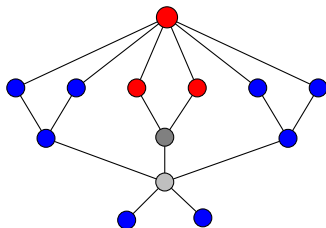
Zentroid-Wert / Beispiel 1

Beispiel-Lösungen:
optimale Auswahl des 1. Knotens (dunkelgrau)

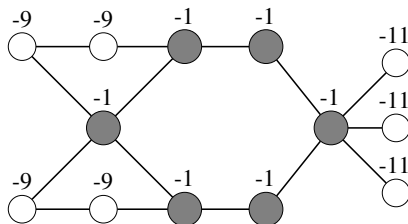


Zentroid-Wert / Beispiel 1

Beispiel-Lösungen:
nicht-optimale Auswahl des 1. Knotens (dunkelgrau)



Zentroid-Wert / Beispiel 2



(Quelle: Brandes/Erlebach (Eds.): Network Analysis)

Die grauen Knoten haben maximalen Zentroid-Wert,
aber selbst sie haben einen negativen Wert.
⇒ Es ist hier vorteilhaft, erst als Zweiter zu wählen.

Zusammenfassung

- Exzentrizität, Closeness und Zentroid-Wert sind strukturelle Indizes.
- Die Knoten maximaler Zentralität unterscheiden sich bei den verschiedenen Maßen.
- Im Gegensatz zu Exzentrizität und Closeness kann der Zentroid-Wert negativ sein.

Knoten maximaler Zentralität

Definition

Für ein Zentralitätsmaß c sei die Menge der Knoten mit maximaler Zentralität in einem Graphen G definiert als

$$S_c(G) = \{v \in V : \forall w \in V \ c(v) \geq c(w)\}$$

Eigenschaft des Baum-Zentrums

Satz (Jordan, 1869)

Für jeden Baum T gilt, dass sein Zentrum aus höchstens zwei Knoten besteht, die miteinander benachbart sind.

Beweis.

Baum aus höchstens 2 Knoten \Rightarrow trivial

Für jeden Knoten v von T kann nur ein Blatt exzentrisch sein.

Knoten v ist dann exzentrisch zu Knoten w , wenn $d(v, w) = e(w)$

Betrachte nun den Baum T' , der durch Entfernen aller Blätter aus T entsteht. Die Exzentrizität jedes Knotens in T' ist um genau Eins kleiner als in T . Deshalb haben beide Bäume das gleiche Zentrum (falls T' nicht leer ist). Die Fortsetzung dieses Verfahrens führt zwangsläufig zu einem Baum, der aus genau einem oder genau zwei adjazenten Knoten besteht. □

Berechnung des Baum-Zentrums

Der vorangegangene Beweis impliziert einen einfachen Algorithmus für die Berechnung des Zentrums eines Baums, der nicht einmal die Berechnung der Exzentrizität der einzelnen Knoten erfordert.

Eigenschaft des Graph-Zentrums

Satz

*Sei G ein zusammenhängender ungerichteter Graph.
Dann existiert ein Block (zweifach zusammenhängender Teilgraph
oder Brücke oder Artikulationsknoten) in G , der alle Knoten des
Zentrums $\mathcal{C}(G)$ enthält.*

*Oder anders formuliert:
Die Knoten des Graph-Zentrums befinden sich alle in einem Block.*

Eigenschaft des Graph-Zentrums

Beweis.

- Annahme: Es gibt keine Zweifachzusammenhangskomponente in G , die alle Knoten des Zentrums $\mathcal{C}(G)$ enthält.
- ⇒ G enthält einen Artikulationsknoten u , so dass $G - u$ in nicht verbundene Teilgraphen G_1 und G_2 zerfällt, die jeweils mindestens einen Knoten ($\neq u$) aus $\mathcal{C}(G)$ enthalten (oder $|\mathcal{C}(G)| < 3 \dots$).
- Sei v ein exzentrischer Knoten von u und P ein entsprechender kürzester Pfad (der Länge $e(u)$) zwischen u und v .
 v liegt in G_1 oder G_2 , nehmen wir o.B.d.A. an $v \in G_2$.
- ⇒ \exists Knoten $w \in \mathcal{C}(G)$ in G_1 ($w \neq u$). w liegt nicht auf P .
- ⇒ $e(w) \geq d(w, u) + d(u, v) \geq 1 + e(u)$
- ⇒ wegen $e(w) > e(u)$ gehört w nicht zu $\mathcal{C}(G)$ (Widerspruch)

Graph-Median

$$s(G) = \min_{v \in V} \left\{ \sum_{w \in V} d(v, w) \right\}$$

Median:

$$\mathcal{M}(G) = \left\{ v \in V : \sum_{w \in V} d(v, w) = s(G) \right\}$$

Eigenschaften des Graph-Medians

Satz

*Sei G ein zusammenhängender ungerichteter Graph.
Dann existiert ein Block in G , der alle Knoten des Medians $\mathcal{M}(G)$ enthält.*

Oder anders formuliert:

Die Knoten des Graph-Medians befinden sich alle in einem Block.

Beweis: Übungsaufgabe!

Folgerung

Der Median eines Baums besteht entweder aus einem einzelnen Knoten oder aus zwei adjazenten Knoten.

Graph-Zentroid

$$f(G) = \max_{v \in V} \{c_F(v)\}$$

Zentroid:

$$\mathcal{Z}(G) = \{v \in V : c_F(v) = f(G)\}$$

Eigenschaften des Graph-Zentroids

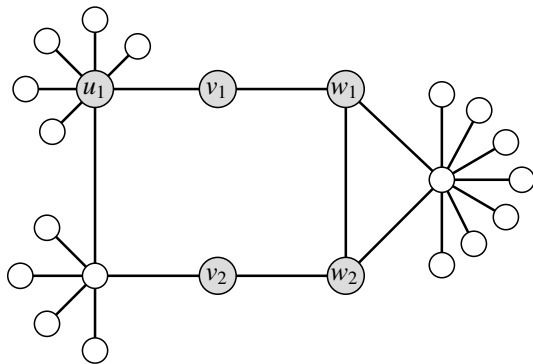
Satz (Slater)

Für jeden Baum sind Median und Zentroid identisch.

Satz (Smart & Slater)

In jedem zusammenhängenden ungerichteten Graphen liegen Median und Zentroid im gleichen Block.

Beispiel



$$\mathcal{C}(G) = \{v_1, v_2\},$$

$$\mathcal{M}(G) = \{u_1\},$$

$$\mathcal{Z}(G) = \{w_1, w_2\}$$

Unterschiedlichkeit der Maße

Satz

Für drei beliebige zusammenhängende ungerichtete Graphen H_1 , H_2 und H_3 und eine beliebige natürliche Zahl $k \geq 4$ existiert ein ungerichteter zusammenhängender Graph G , so dass

- $G[\mathcal{C}(G)] \simeq H_1$,
- $G[\mathcal{M}(G)] \simeq H_2$,
- $G[\mathcal{Z}(G)] \simeq H_3$, und
- *die Distanz zwischen je zwei der zentralen Mengen ist mindestens k .*

Stress-Zentralität

- Anzahl kürzester Pfade zwischen Knoten s und t , die $v \in V$ bzw. $e \in E$ enthalten: $\sigma_{st}(v)$ bzw. $\sigma_{st}(e)$
- Stress-Zentralität:

$$c_S(v) = \sum_{s \neq v \in V} \sum_{t \neq v \in V} \sigma_{st}(v)$$

$$c_S(e) = \sum_{s \in V} \sum_{t \in V} \sigma_{st}(e)$$

- Intuition: Kommunikationsfluss durch Knoten bzw. Kanten auf (allen) kürzesten Wegen zwischen allen Knotenpaaren

Knoten- und Kanten-Stresszentralität

Lemma

In einem gerichteten Graphen $G = (V, E)$ sind Knoten- und Kanten-Stresszentralität wie folgt miteinander verknüpft:

$$c_S(v) = \frac{1}{2} \sum_{e \in \Gamma(v)} c_S(e) - \sum_{v \neq s \in V} \sigma_{sv} - \sum_{v \neq t \in V} \sigma_{vt}$$

Knoten- und Kanten-Stresszentralität

Beweis.

Betrachte einen kürzesten Pfad zwischen $s \neq t \in V$.

\Rightarrow trägt genau 1 zum Stress jedes Knotens und jeder Kante bei.

Wenn man über den Beitrag dieses Pfades über alle (1 oder 2) inzidenten Kanten zu einem Knoten summiert, erhält man

- den doppelten Beitrag (2) zum Knoten selbst, falls der Pfad nicht an dem Knoten anfängt oder endet ($v \in V \setminus \{s, t\}$) und
- den einfachen Beitrag (1) zum Knoten selbst, falls der Pfad an dem Knoten anfängt oder endet ($v \in \{s, t\}$)

Die Größe

$$\sum_{v \neq s \in V} \sigma_{sv} + \sum_{v \neq t \in V} \sigma_{vt}$$

gibt an, wie oft der Knoten v der Anfangs- bzw. Endknoten eines kürzesten Pfades ist. □

Shortest-Path Betweenness

- eine Art relative Stress-Zentralität
- Anzahl kürzester Pfade zwischen $s, t \in V$: σ_{st}
- Anteil der kürzesten Wege zwischen s und t , die v enthalten:

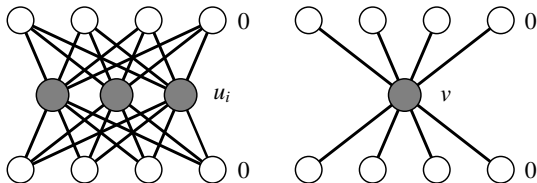
$$\delta_{st}(v) = \frac{\sigma_{st}(v)}{\sigma_{st}}$$

⇒ interpretiert als Anteil oder Wahrscheinlichkeit der Kommunikation

- Betweenness-Zentralität:

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

Stress und Betweenness



- Vorteil gegenüber Closeness: funktioniert auch bei nicht zusammenhängenden Graphen
- Normierung: Teilen durch Anzahl Paare $((n - 1)(n - 2))$

Betweenness für Kanten

- Anteil der Kante e am Informationsfluss (auf allen kürzesten Pfaden) zwischen den Knoten s und t

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

- Betweenness-Zentralität der Kante e :

$$c_B(e) = \sum_{s \in V} \sum_{t \in V} \delta_{st}(e)$$

Knoten- und Kanten-Betweenness-Zentralität

Lemma

In einem gerichteten Graphen $G = (V, E)$ sind Knoten- und Kanten-Betweenness-Zentralität (für kürzeste Pfade) wie folgt miteinander verknüpft:

$$c_B(v) = \sum_{e \in \Gamma^+(v)} c_B(e) - (n - 1) = \sum_{e \in \Gamma^-(v)} c_B(e) - (n - 1)$$

Kantengraph

Definition

Der **Kantengraph** des Graphen $G = (V, E)$ ist definiert als $G' = (E, K)$, wobei K die Menge der Kanten $e = ((x, y), (y, z))$ mit $(x, y) \in E$ und $(y, z) \in E$ ist.

Zwei Knoten im Kantengraph sind also benachbart, wenn die entsprechenden Kanten im ursprünglichen Graphen einen Knoten gemeinsam haben (im gerichteten Fall als Zielknoten der einen Kante und Startknoten der anderen).

⇒ Wende Knotenzentralität auf den Kantengraph an
Nachteile:

- Größe des Kantengraphen kann quadratisch in der Größe des Graphen sein,
- keine natürliche Interpretation / Generalisierung

Inzidenzgraph

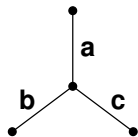
Definition

Der **Inzidenzgraph** des Graphen $G = (V, E)$ ist definiert als $G'' = (V'', E'')$, wobei $V'' = V \cup E$ und $E'' = \{(v, e) | \exists w : e = (v, w) \in E\} \cup \{(e, w) | \exists v : e = (v, w) \in E\}$.

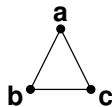
Im Inzidenzgraphen sind also ein 'echter Knoten' und ein 'Kantenknoten' benachbart, wenn der entsprechende Knoten und die entsprechende Kante im ursprünglichen Graphen inzident sind.

- ⇒ Wende Knotenzentralität auf den Inzidenzgraph an, wobei nur die Pfade zwischen 'echten Knoten' als relevant betrachtet werden

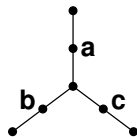
Abgeleitete Kantenzentralitäten



Graph



Kantengraph



Inzidenzgraph

Vitalität

Vitalitätsmaße bewerten die Wichtigkeit eines Knotens oder einer Kante anhand eines Qualitätsverlusts durch das Löschen des Knotens bzw. der Kante.

Definition (Vitalitätsindex)

Sei \mathcal{G} die Menge aller einfachen ungerichteten ungewichteten Graphen $G = (V, E)$ und $f : \mathcal{G} \rightarrow \mathbb{R}$ eine reellwertige Funktion auf $G \in \mathcal{G}$.

Dann ist ein **Vitalitätsindex** $\mathcal{V}(G, x)$ definiert als die Differenz der Werte von f auf G und $G \setminus \{x\}$:

$$\mathcal{V}(G, x) = f(G) - f(G \setminus \{x\})$$

Flow Betweenness Vitality

- ähnlich zu Shortest Path Betweenness
- Motivation: Information in einem Kommunikationsnetzwerk muss sich nicht unbedingt auf kürzesten Pfaden bewegen.
- Maß: Abhängigkeit des maximalen Flusses zwischen zwei Knoten von der Existenz des betrachteten Knotens

Flow Betweenness Vitality

- f_{st} : Maximum-Fluss zwischen Knoten s und t unter Berücksichtigung der Kantenkapazitäten
- $f_{st}(v)$: Fluss zwischen Knoten s und t , der bei Maximum-Fluss von s nach t durch v gehen **muss**
- $\tilde{f}_{st}(v)$: Maximum-Fluss von s nach t in $G - v$

$$c_{mf}(v) = \sum_{\substack{s, t \in V \\ v \neq s, v \neq t \\ f_{st} > 0}} \frac{f_{st}(v)}{f_{st}} = \sum_{\substack{s, t \in V \\ v \neq s, v \neq t \\ f_{st} > 0}} \frac{f_{st} - \tilde{f}_{st}(v)}{f_{st}}$$

Closeness Vitality

- Wiener Index:

$$I_W(G) = \sum_{v \in V} \sum_{w \in V} d(v, w)$$

- oder in Abhängigkeit der Closeness-Zentralitäten:

$$I_W(G) = \sum_{v \in V} \frac{1}{c_C(v)}$$

- Closeness Vitality für Knoten/Kante x :

$$c_{CV}(x) = I_W(G) - I_W(G - x)$$

- Interpretation: Um wieviel steigen die Gesamtkommunikationskosten, wenn jeder Knoten mit jedem kommuniziert?

Closeness Vitality - Durchschnitt

- Durchschnittliche Distanz:

$$\bar{d}(G) = \frac{l_W(G)}{n(n-1)}$$

- Das Vitalitätsmaß auf der Basis $f(G) = \bar{d}(G)$ misst die durchschnittliche Verschlechterung der Entfernung zwischen zwei Knoten beim Entfernen eines Knotens / einer Kante x .
- Vorsicht! Beim Entfernen eines Artikulationsknotens (cut vertex) oder einer Brücke (cut edge) x ist $c_{CV}(x) = -\infty$.

Shortcut-Werte

- kein echter Vitalitätsindex im Sinne der Definition
 - maximale Erhöhung eines Distanzwerts, wenn Kante $e = (v, w)$ entfernt wird
- ⇒ nur relevant für Knoten, bei denen alle kürzesten Pfade über e laufen
- maximale Erhöhung tritt direkt zwischen Knoten v und w auf
 - alternativ: maximale relative Erhöhung
 - Berechnung: mit $m = |E|$ Single Source Shortest Path Instanzen (und Vergleich mit der jeweiligen Kante)
 - später: mit $n = |V|$ SSSP-Bäumen
 - Anwendung auch auf Knotenlöschungen möglich

Stress-Zentralität als Vitalitätsindex

- Stress-Zentralität zählt die Anzahl der kürzesten Pfade, an denen ein Knoten oder eine Kante beteiligt ist
- ⇒ kann als Vitalitätsmaß betrachtet werden
-
- Aber: Anzahl der kürzesten Pfade kann sich durch Löschung erhöhen (wenn sich die Distanz zwischen zwei Knoten erhöht)
- ⇒ Längere kürzeste Pfade müssen ignoriert werden

Stress-Zentralität als Vitalitätsindex

- $f(G \setminus \{v\})$ muss ersetzt werden durch

$$\sum_{s \in V} \sum_{t \in V} \sigma_{st} [d_G(s, t) = d_{G \setminus \{v\}}(s, t)]$$

- Ausdruck in Klammern ist 1 (wahr) oder 0 (falsch)
- $f(G \setminus \{e\})$ analog:

$$\sum_{s \in V} \sum_{t \in V} \sigma_{st} [d_G(s, t) = d_{G \setminus \{e\}}(s, t)]$$

- Also: $f(G) - f(G - x) = \sum_{s \in V} \sum_{t \in V} \sigma_{st}(x)$
- kein echter Vitalitätsindex im Sinne der Definition, weil der Index-Wert nach der Löschung von der Distanz vor der Löschung abhängt