Fortgeschrittene Netzwerk- und Graph-Algorithmen

Dr. Hanjo Täubig

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

Wintersemester 2007/08

Übersicht

Zentralitätsindizes

- Manche Knoten / Kanten in Netzwerken sind wichtiger, zentraler oder einflussreicher als andere.
- Zentralitätsmaße bzw. -indizes (kurz Zentralitäten) versuchen, diese Eigenschaften durch skalare Werte zu quantifizieren.
- Es gibt jedoch kein universelles Zentralitätsmaß, das jeder Anwendung gerecht wird.
  Das 'richtige' Maß hängt vom Kontext der Anwendung ab.

Beispiel: Leader election (1)

Bsp.: Wahl eines Klassensprechers

- Knoten entsprechen Personen
- Kante von Knoten A nach B, wenn Person A für Person B stimmt

- Person ist zentraler, je höher die Anzahl der erhaltenen Stimmen ist
  $\Rightarrow$ in-degree centrality

Beispiel: Leader election (2)

Bsp.: Wahl eines Klassensprechers

- Knoten entsprechen Personen
- Kante von Knoten A nach B, wenn Person A Person B überzeugt hat, für seinen/ihren Favoriten zustimmen
- Einfluss-Netzwerk

- Person ist zentraler, je mehr diese Person gebraucht wird, um die Meinung anderer zu transportieren
  $\Rightarrow$ betweenness centrality

Beispiel: Leader election (3)

Bsp.: Wahl eines Klassensprechers

- Knoten entsprechen Personen
- Kante von Knoten A nach B, wenn Person A mit Person B befreundet ist

- Person ist zentraler, je mehr Freunde diese Person hat und je zentraler diese Freunde sind
  $\Rightarrow$ feedback centrality

Kantenzentralität

Genau wie die Zentralität von Knoten kann man auch die Wichtigkeit von Kanten betrachten

Beispiel: Internet

- Backbone: Verbindungen zwischen den Kontinenten gibt es wenige und sie müssen eine große Kapazität haben
- Kantenzentralität
  - Beteiligung einer Kante an kürzesten Wegen usw.
    $\Rightarrow$ betweenness edge centrality
  - Veränderung von Netzwerk-Parametern durch Löschen der Kante
    $\Rightarrow$ (edge) vitality (Bsp. flow betweenness vitality)

Definition: Zentralitätsindex

Abgesehen von der Intuition

- Wichtigkeit,
- Prestige,
- Einfluss,
- Kontrolle,
- Unentbehrlichkeit

gibt es keine allgemeingültige (formale) Definition von Zentralität.

Mindestanforderung:
Das Maß darf nur von der Struktur des Graphen abhängen.

Graph-Isomorphie

Definition (Isomorphie)

Zwei Graphen $G_1 = (V_1, E_1)$ und $G_2 = (V_2, E_2)$ sind isomorph ($G_1 \simeq G_2$), falls es eine Bijektion $\phi : V_1 \to V_2$ gibt, so dass

$$(u, v) \in E_1 \quad \Leftrightarrow \quad (\phi(u), \phi(v)) \in E_2$$

Definition (Struktureller Index)

Sei $G = (V, E)$ ein gewichteter (gerichteter oder ungerichteter) graph und sei X die Knotenmenge (V) oder die Kantenmenge (E). Eine reell-wertige Funktion s heißt *struktureller Index* genau dann, wenn die folgende Bedingung erfüllt ist:

$$\forall x \in X : \quad G \simeq H \Rightarrow s_G(x) = s_H(\phi(x))$$

Grad-Zentralität

Grad-Zentralität (degree centrality):

$$c_D(v) = \deg(v)$$

- lokales Maß,
  hängt nur von der direkten Nachbarschaft eines Knotens ab

Exzentrizität

Anwendungsbeispiel:
Positionierung eines Hospitals oder einer Feuerwehrstation

Ziel: Minimierung der maximal notwendigen Anfahrtzeit

Exzentrizität eines Knotens $v \in V$:

$$e(v) = \max_{w \in V} \{d(v, w)\}$$

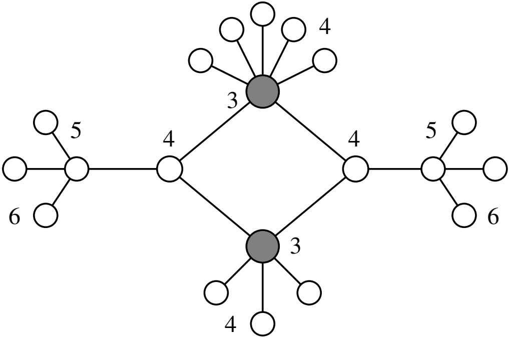
Zentralitätsmaß:

$$c_E(v) = \frac{1}{e(v)}$$

Optimaler Standort:

Knoten v mit minimalem Wert $e(v)$ (*Zentrum* von G)

Exzentrizität / Beispiel



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

Closeness

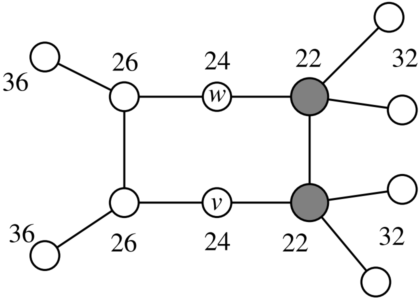
Anwendungsbeispiel:
Positionierung eines Einkaufszentrums
(minisum location / median / service facility location problem)

Ziel:
Minimierung der Summe der Entfernungen zu den anderen Knoten
(und damit auch der durchschnittlichen Entfernung)

Zentralitätsmaß:

$$c_C(v) = \frac{1}{\sum_{w \in V} d(v, w)}$$

Closeness / Beispiel



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

Graue Knoten sind am wichtigsten bezüglich Closeness,
v und w sind zentraler in Hinsicht auf Exzentrizität

Radiality

ähnlich zu Closeness

$$c_R(v) = \frac{\sum_{w \in V} (\Delta_G + 1 - d(v,w))}{n-1}$$

Δ_G: Durchmesser des Graphen
(größte Distanz zweier Knoten, nicht Länge des längsten Pfads!)

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

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

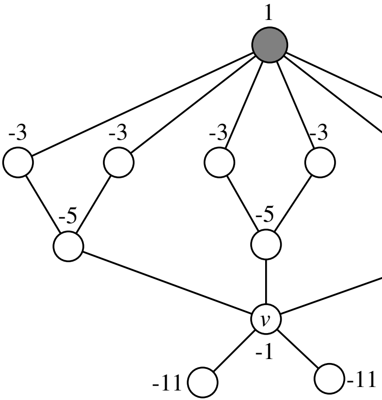
Wähle Knoten u, Gegenspieler wählt Knoten v
Resultierende Anzahl Kunden:

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

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

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

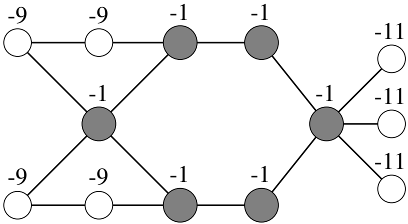
Zentroid-Wert / Beispiel 1



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

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

Zentroid-Wert / Beispiel 2



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

Die grauen Knoten haben maximalen Zentroid-Wert, aber selbst sie haben einen negativen Wert.

$\Rightarrow$ 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. $\square$

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 zweifach zusammenhängender Teilgraph in G, der alle Knoten des Zentrums $\mathcal{C}(G)$ enthält.
Oder anders formuliert:
Die Knoten des Graph-Zentrums befinden sich alle in einer Zweifachzusammenhangskomponente (einem Block).

Eigenschaft des Graph-Zentrums

Beweis.

- Annahme: Es gibt keine Zweifachzusammenhangskomponente in G, die alle Knoten des Zentrums $\mathcal{C}(G)$ enthält.

$\Rightarrow$ G enthält einen Artikulationsknoten, so dass G in nicht verbundene Teilgraphen G_1 und G_2 zerfällt, die jeweils mindestens einen Knoten aus $\mathcal{C}(G)$ enthalten.

- Sei v ein exzentrischer Knoten von u und P ein entsprechender kürzester Pfad (der Länge $e(u)$) zwischen u und v. O.B.d.A. sei v aus G_2.

$\Rightarrow$ $\exists$ Knoten $w \in \mathcal{C}(G)$ in G_1, der nicht zu P gehört

$\Rightarrow$ $e(w) \geq d(w, u) + d(u, v) \geq 1 + e(u)$

$\Rightarrow$ wegen $e(w) > e(u)$ gehört w nicht zu $\mathcal{C}(G)$ (Widerspruch)

$\square$

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 zweifach zusammenhängender Teilgraph in G, der alle Knoten des Medians $\mathcal{M}(G)$ enthält.
Oder anders formuliert:
Die Knoten des Graph-Medians befinden sich alle in einer Zweifachzusammenhangskomponente (einem Block).

Beweis.

(analog zu Beweis für Graph-Zentrum) □

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

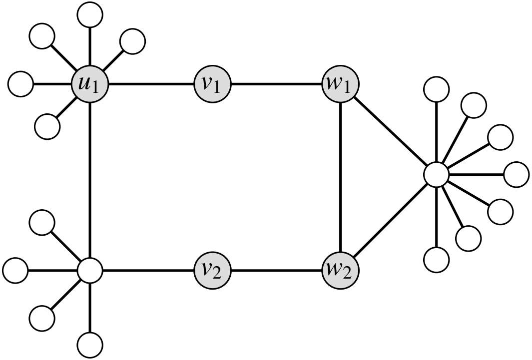
Satz (Slater)

Für jeden Baum sind Median und Zentroid identisch.

Satz (Smart & Slater)

In jedem zusammenhängenden ungerichteten Graphen liegen Median und Zentroid in der gleichen Zweifachzusammenhangskomponente.

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: $\sigma_{st}(v)$
- 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 Pfads 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. $\square$

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:

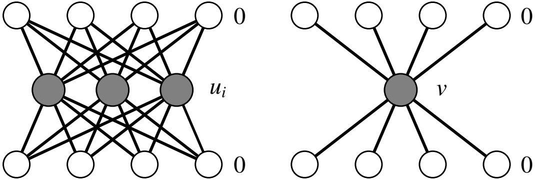
$$\delta_{st}(v) = \frac{\sigma_{st}(v)}{\sigma_{st}}$$

$\Rightarrow$ 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:

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