Fortgeschrittene Netzwerk- und Graph-Algorithmen

Dr. Hanjo Täubig

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

Wintersemester 2010/11

Übersicht

1. Zentralitätsindizes
   - Beispiele
   - Grad- und Distanz-basierte Zentralitäten
   - Kürzeste Pfade und Zentralität

Zentralitätsindizes

- Manche Elemente in Netzwerken (Knoten / Kanten in Graphen) sind wichtiger, zentraler oder einflussreicher als andere.

- Zentralitätsmaße bzw. -indizes (kurz Zentralitäten) quantifizieren diese Eigenschaften durch skalare Werte.

- Es gibt jedoch kein universelles Zentralitätsmaß, das jeder Anwendung gerecht wird.

  $\Rightarrow$ 'richtiges' Maß hängt vom Kontext der Anwendung ab

Beispiel: Wahl eines Klassensprechers

Variante 1:

- 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: Wahl eines Klassensprechers

Variante 2:

- 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: Wahl eines Klassensprechers

Variante 3:

- 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

Ebenso kann man auch die Wichtigkeit / Zentralität von Beziehungen in Netzwerken (Kanten in Graphen) betrachten.

Beispiel: Internet

- Backbone: Verbindungen zwischen den Kontinenten gibt es wenige und sie müssen eine große Kapazität haben

Arten von Kantenzentralität / Beispiele:

- 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.

Werte sollen zumindest eine Ordnung der Elemente liefern, auch wenn die numerischen Werte, Abstände oder Verhältnisse evt. nicht viel aussagen.

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 reellwertige Funktion s heißt genau dann struktureller Index, wenn die folgende Bedingung erfüllt ist:

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

Grad-Zentralität

Grad-Zentralität (degree centrality):

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

- lokales Maß,
  nur von der direkten Nachbarschaft des Knotens abhängig
- mögliche Verallgemeinerung:
  Anzahl der indirekten Nachbarn mit Abstand $\leq k$

Exzentrizität

Anwendungsbeispiel:
Positionierung eines Hospitals oder einer Feuerwehrstation

Ziel: Minimierung der maximal notwendigen Anfahrtzeit

Exzentrizität eines Knotens $v \in V$:

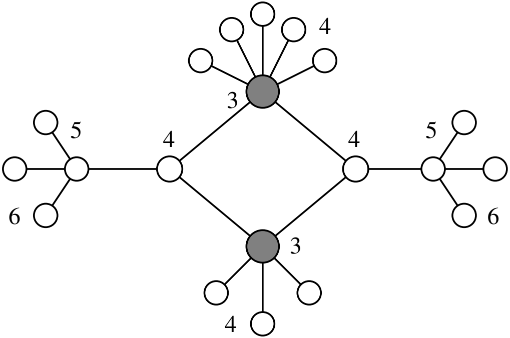
$$e(v) = \max_{w \in V}\{d(v, w)\}$$

Zentralitätsmaß:

$$c_E(v) = \frac{1}{e(v)} = \frac{1}{\max_{w \in V}\{d(v, w)\}}$$

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):

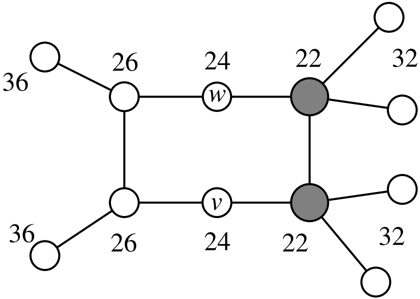
$$\sum_{v \in V} d(u, v)$$

(im Kontext von Kommunikationsnetzen auch *transmission number*)

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

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

$d_{\max}(G)$: 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

gegeben: ungerichteter zusammenhängender Graph G

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

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

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

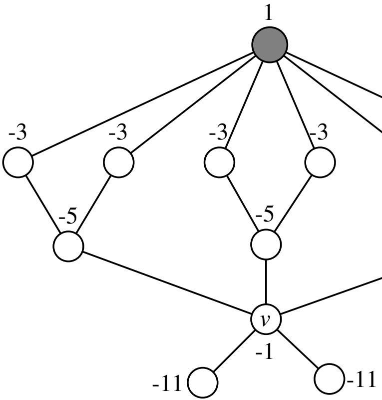
$$\Rightarrow \quad \gamma_u(v) + \frac{n - \gamma_u(v) - \gamma_v(u)}{2} = \frac{n + \gamma_u(v) - \gamma_v(u)}{2}$$

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

Zentralitätsmaß:

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

Zentroid-Wert / Beispiel 1

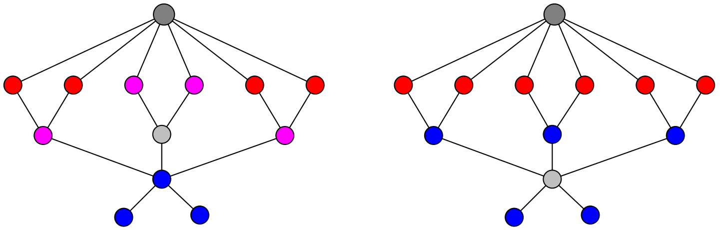


(Quelle: Brandes/Erlebach (Eds.): Network Analysis)

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

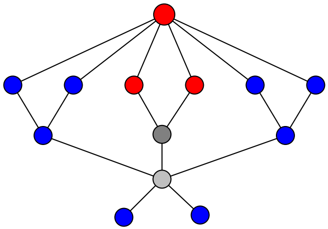
Zentroid-Wert / Beispiel 1

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

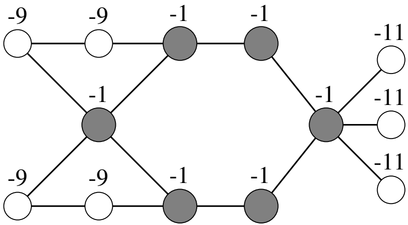


Zentroid-Wert / Beispiel 1

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.
$\Rightarrow$ Es ist hier vorteilhaft, erst als Zweiter zu wählen.

Bemerkungen

- 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 genau einem oder aus zwei Knoten besteht, die miteinander benachbart sind.

Eigenschaft des Baum-Zentrums

Beweis.

- Baum aus höchstens 2 Knoten $\Rightarrow$ trivial
- Für jeden Knoten v von T können nur Blätter exzentrisch sein (maximale Entfernung haben), keine inneren Knoten.

  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.

$\Rightarrow$ Exzentrizität jedes Knotens in T' ist um genau Eins kleiner als in T

- beide Bäume haben gleiches Zentrum (falls T' nicht leer ist)
- Fortsetzung dieses Verfahrens führt zwangsläufig zu einem Baum, der aus genau einem oder zwei adjazenten Knoten besteht

$\square$

Berechnung des Baum-Zentrums

Vorangegangener Beweis impliziert einen einfachen Algorithmus zur 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 isolierter Knoten im Fall $n = 1$) 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 keinen Block in G, der alle Knoten des Zentrums $\mathcal{C}(G)$ enthält.

$\Rightarrow$ 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.

- 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 $v \in G_2$

$\Rightarrow$ $\exists$ Knoten $w \in \mathcal{C}(G)$ in G_1 ($w \neq u$). w liegt nicht auf P.

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

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

□

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

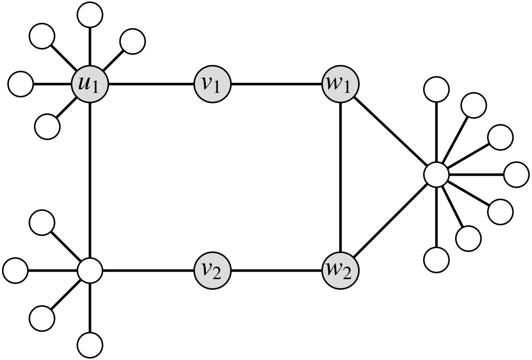
Satz (Slater)

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