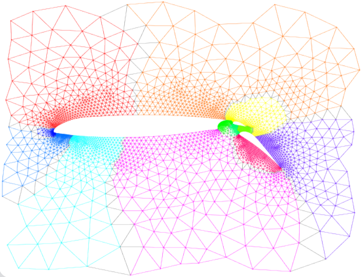
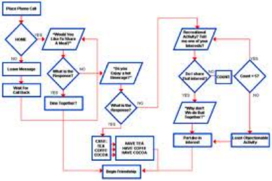
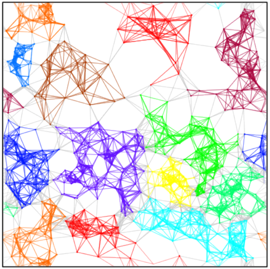
Algorithmische Methoden zur Netzwerkanalyse

Prof. Dr. Henning Meyerhenke

Institut für Theoretische Informatik







KIT – Universität des Landes Baden-Württemberg und
nationales Forschungszentrum in der Helmholtz-Gemeinschaft

www.kit.edu

Vorlesung 11

Karlsruher Institut für Technologie

Programm des Tages:

- Nähezentralität (closeness centrality)
- Generative Modelle von Graphen

Inhalt

Zentralitätsmaße

Nähezentralität

Netzwerkmodelle

Closeness Centrality

Dt.: Nähezentralität

- Wieder Berücksichtigung der kürzesten Wege
- Dieses Mal aber deren Länge, nicht deren Zahl
- Mittlerer kürzester Abstand:

$$l_i = \frac{1}{n} \sum_j \mathrm{dist}(i, j)$$

- Kleiner Nachteil des MKA:
  Hohe Werte sprechen für geringen Einfluss im Netzwerk

Closeness Centrality

Dt.: Nähezentralität

- Wieder Berücksichtigung der kürzesten Wege
- Dieses Mal aber deren Länge, nicht deren Zahl
- Mittlerer kürzester Abstand:

$$l_i = \frac{1}{n} \sum_j \mathrm{dist}(i, j)$$

- Kleiner Nachteil des MKA:
  Hohe Werte sprechen für geringen Einfluss im Netzwerk

Definition (Nähezentralität)

$$C_i = \frac{1}{l_i} = \frac{n}{\sum_j \mathrm{dist}(i, j)}$$

Alternatives Maß

- **Problem:** Knoten in verschiedenen Komponenten haben Abstand ∞
- **Alternativ:** Harmonisches Mittel des Abstands:

$$C_i' = \frac{1}{n-1} \sum_{j \neq i} \frac{1}{\mathrm{dist}(i,j)}$$

- Löst das Problem der unendlichen Abstände und gibt Ähnlichkeit an
- **Trotzdem:** In Praxis und Wissenschaft wenig verwendet

Betrachtung des ganzen Netzwerks

- Auch Netzwerke kann man anhand bestimmter Maße einordnen
- Closeness bei einer ZHK:

$$l = \frac{1}{n^2} \sum_{i,j} \mathrm{dist}(i, j) = \frac{1}{n} \sum_{i} l_i$$

Betrachtung des ganzen Netzwerks

- Auch Netzwerke kann man anhand bestimmter Maße einordnen
- Closeness bei einer ZHK:

$$l = \frac{1}{n^2} \sum_{i,j} \mathrm{dist}(i,j) = \frac{1}{n} \sum_i l_i$$

- Bei mehreren ZHK: Wieder Problem mit unendlich großen Abständen
- Daher Durchschnittsbildung mit harmonischem Mittel:

$$l' = \frac{n}{\sum_i C'_i}$$

Kleingruppenaufgabe

Wie und wie schnell kann man die Nähezentralität berechnen?

Abstandsberechnung

- Abstandsberechnung in ungewichteten (Multi)Graphen: BFS
- Analog zur Tiefensuche vergeben wir BFS-Nummern:

$$BFS(w) := BFS(v) + 1 \text{ gdw. } w \text{ von } v \text{ “entdeckt” wird}$$

- Kantenklassifikation bei BFS: Baum-, Rückwärts- und Querkante
- Komplexität: $\mathcal{O}(n + m)$

Berechnung von Nähezentralität

in ungewichteten (Multi)Graphen

Proposition

Sei $G = (V, E)$ ein ungewichteter Multigraph, sei $s \in V$.
Nach BFS mit Wurzel s gilt:

$$\mathrm{dist}_G(s, v) = BFS(v)$$

für alle $v \in V$.

Berechnung von Nähezentralität

in ungewichteten (Multi)Graphen

Proposition

Sei $G = (V, E)$ ein ungewichteter Multigraph, sei $s \in V$.
Nach BFS mit Wurzel s gilt:

$$\mathrm{dist}_G(s, v) = BFS(v)$$

für alle $v \in V$.

Theorem

Die Nähezentralitäten der Knoten eines (stark) zusammenhängenden Multigraphen können in $\mathcal{O}(nm)$ Zeit berechnet werden.

Diskussion Nähezentralität

- **Vorteil:** Nähezentralität sehr natürliches Maß
- **Nachteil:** Kein breites Spektrum der Ergebnisse
  - Maximaler Abstand typischerweise logarithmisch
  - Beispiel IMDB: Maximum $0,4143$, Minimum: $0,1154$

Nachteil: Behandlung von unzusammenhängenden Graphen problematisch

- Lösung dafür: Harmonische Mittelbildung

Fazit Zentralitätsmaße

- Gradzentralität
- Eigenvektorzentralität
- PageRank
- Intermediationszentralität
- Nähezentralität

Fazit Zentralitätsmaße

- Gradzentralität
- Eigenvektorzentralität
- PageRank
- Intermediationszentralität
- Nähezentralität

- Art der Berechnung
- Komplexität
- Aussagekraft
- ...

Fazit Zentralitätsmaße

- Gradzentralität
- Eigenvektorzentralität
- PageRank
- Intermediationszentralität
- Nähezentralität

- Art der Berechnung
- Komplexität
- Aussagekraft
- ...

Hausaufgabe!

Ausblick Netzwerkmodelle

- Erdös-Renyi-Zufallsgraphen
- Gradfolgen
- Realistischere Netzwerk-Modelle