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

Approximation von Zentralitätsindizes

- Obwohl die behandelten Zentralitätsindizes in polynomieller Zeit berechnet werden können, heißt das nicht unbedingt , dass die entsprechenden Algorithmen in der Praxis anwendbar sind.

- Beispiel:
  Betweenness-Zentralität für die Knoten des Web-Graphen lässt sich selbst mit dem Algorithmus von Brandes nicht in akzeptabler Zeit berechnen.

$\Rightarrow$ Möglichst wenige Traversierungen bzw. SSSP-Durchläufe auf dem Graphen

$\Rightarrow$ Näherungslösungen mit möglichst geringer Abweichung (mit hoher Wahrscheinlichkeit)

Approximation von Closeness

- Closeness:

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

- Approximation: wähle k andere Knoten $v_1, \ldots, v_k \in V$

$$\hat{c}_C(v) = \frac{k}{n} \frac{1}{\sum_{i=1}^{k} d(v, v_i)}$$

- Vorgehen (Eppstein / Wang):
  1. Wähle k Knoten $v_1, \ldots, v_k$ gleichverteilt zufällig
  2. Löse für jeden Knoten v_i das SSSP mit diesem Knoten als Startknoten und
  3. berechne für jeden Knoten $v \in V$ die Zentralität

$$\hat{c}_C(v) = \frac{k}{n \cdot \sum_{i=1}^{k} d(v, v_i)}$$

Hoeffdings Ungleichung

Satz (Hoeffdings Ungleichung)

Seien $X_1, \ldots, X_k$ *unabhängige Zufallsvariablen mit* $a_i \leq X_i \leq b_i$ *und* $\mu = \mathbb{E}\left[\sum_{i=1}^k X_i/k\right]$ *der erwartete Durchschnitt. Dann gilt*

$$\Pr\left\{\left|\frac{\sum_{i=1}^k X_i}{k} - \mu\right| \geq \xi\right\} \leq 2 \cdot e^{-2k^2\xi^2/\sum_{i=1}^k (b_i - a_i)^2}$$

bzw. im Fall $a_i = a$, $b_i = b$ $(\forall i)$

$$\Pr\left\{\left|\frac{\sum_{i=1}^k X_i}{k} - \mu\right| \geq \xi\right\} \leq 2 \cdot e^{-2k\xi^2/(b-a)^2}$$

Anwendung der Hoeffding-Ungleichung

Setze

$$\begin{aligned} X_i &= n \cdot \frac{d(v_i, u)}{n-1} \\ \mu &= \frac{1}{c_C(v)} \\ a_i &= 0 \\ b_i &= \frac{n \cdot \mathrm{diam}(G)}{n-1} \end{aligned}$$

Anwendung der Hoeffding-Ungleichung

$$
\begin{aligned}
\Pr\left\{\left|\frac{\sum_{i=1}^{k} X_i}{k} - \mu\right| \geq \xi\right\} &\leq 2 \cdot e^{-2k^2\xi^2 / \sum_{i=1}^{k}(b_i - a_i)^2} \\
&= 2 \cdot e^{-2k^2\xi^2 / \left(k\left(\frac{n \cdot \mathrm{diam}(G)}{n-1}\right)^2\right)} \\
&= 2 \cdot e^{-\Omega\left(k\xi^2/\mathrm{diam}(G)^2\right)}
\end{aligned}
$$

- Wähle $\xi = \epsilon \cdot \mathrm{diam}(G)$
- $k = \Theta\left(\frac{\log n}{\epsilon^2}\right)$ SSSP-Läufe

$\Rightarrow$ Wahrscheinlichkeit, einen Fehler größer als $\epsilon \cdot \mathrm{diam}(G)$ zu machen ist höchstens $\frac{1}{n}$ für jeden Wert

Laufzeit der Closeness-Approximation

- Komplexität eines SSSP-Laufs
  - $\mathcal{O}(m+n)$ in ungewichteten Graphen
  - $\mathcal{O}(m+n\log n)$ in gewichteten Graphen

- Komplexität von k SSSP-Läufen
  - $\mathcal{O}(k\cdot(m+n))$ in ungewichteten Graphen
  - $\mathcal{O}(k\cdot(m+n\log n)$ in gewichteten Graphen

$\Rightarrow$ Komplexität von $\Theta\left(\frac{\log n}{\epsilon^2}\right)$ SSSP-Läufen

  - $\mathcal{O}\left(\frac{\log n}{\epsilon^2}\cdot(m+n)\right)$ in ungewichteten Graphen
  - $\mathcal{O}\left(\frac{\log n}{\epsilon^2}\cdot(m+n\log n)\right)$ in gewichteten Graphen

Approximation von Betweenness

- gewichtete gerichtete Graphen
- wähle wieder k Knoten zufällig (gleichverteilt) aus
- Berechne für jeden Startknoten v_i die totalen Abhängigkeiten $\delta_{v_i *}(v)$ aller anderen Knoten v
- Berechne

$$\hat{c}_B(v) = \sum_{i=1}^{k} \frac{n}{k} \cdot \delta_{v_i *}(v)$$

- $\mathbb{E}[\hat{c}_B(v)] = c_B(v)$ für alle k und v

Anwendung der Hoeffding-Ungleichung

Setze

$$\begin{aligned} X_i &= n \cdot \delta_{v_i*} \\ \mu &= c_B(v) \\ a_i &= 0 \\ b_i &= n(n-2) \end{aligned}$$

δ_{v_i*} kann höchstens $n-2$ sein, und zwar wenn alle kürzesten Pfade, die von v_i ausgehen, über v laufen. Also ist X_i durch $n(n-2)$ begrenzt.

Anwendung der Hoeffding-Ungleichung

$$
\begin{aligned}
\Pr\left\{|\hat{c}_B(v) - c_B(v)| \geq \xi\right\} &\leq 2 \cdot e^{-2k^2\xi^2 / \sum_{i=1}^{k}(b_i - a_i)^2} \\
&= 2 \cdot e^{-2k^2\xi^2 / \left(k(n(n-2))^2\right)} \\
&= 2 \cdot e^{-2k\xi^2 / (n(n-2))^2}
\end{aligned}
$$

- Wähle $\xi = \epsilon \cdot n(n-2)$
- $k = \Theta\left(\frac{\log n}{\epsilon^2}\right)$ Startknoten / Läufe

$\Rightarrow$ Wahrscheinlichkeit, einen Fehler größer als $\epsilon \cdot n(n-2)$ zu machen ist höchstens $\frac{1}{n}$ für jeden Wert

Laufzeit der Betweenness-Approximation

- Komplexität eines Laufs (für $\delta_{v_i*}(v)$)
  - $\mathcal{O}(m+n)$ in ungewichteten Graphen
  - $\mathcal{O}(m+n\log n)$ in gewichteten Graphen

- Komplexität von k Läufen
  - $\mathcal{O}(k\cdot(m+n))$ in ungewichteten Graphen
  - $\mathcal{O}(k\cdot(m+n\log n)$ in gewichteten Graphen

$\Rightarrow$ Komplexität von $\Theta\left(\frac{\log n}{\epsilon^2}\right)$ Läufen

  - $\mathcal{O}\left(\frac{\log n}{\epsilon^2}\cdot(m+n)\right)$ in ungewichteten Graphen
  - $\mathcal{O}\left(\frac{\log n}{\epsilon^2}\cdot(m+n\log n)\right)$ in gewichteten Graphen

Ergebnis

- Gewinn: k anstatt n SSSP-artige Läufe

- Verfahren des normalisierten Durchschnitts basierend auf zufälligem Knoten-Sampling läßt sich auf viele andere Zentralitäten übertragen