

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

- 1 Lokale Dichte
 - Statistisch dichte Gruppen

Dichte und Wege

- Definition der Dichte entspricht der durchschnittlichen Existenz von möglichen Kanten innerhalb eines (induzierten) Teilgraphen
- Kanten sind Wege der Länge 1
- verallgemeinert kann man Wege (Walks) beliebiger Länge betrachten
- Grad der Ordnung $\ell \in \mathbb{N}$ eines Knotens v ,
Anzahl der Wege der Länge ℓ in G , die in v starten: $w_G^\ell(v)$
- $w_G^0(v) = 1$, $w_G^1(v) = \deg_G(v)$
- Anzahl Wege der Länge ℓ in einem Graphen G : w_G^ℓ

Anzahl Wege der Länge ℓ

Satz

Sei $G = (V, E)$ ein ungerichteter Graph.

Für alle $\ell \in \mathbb{N}$ und für alle $r \in \{0, \dots, \ell\}$ gilt:

$$w_\ell(G) = \sum_{v \in V} w_G^r(v) \cdot w_G^{\ell-r}(v)$$

Anzahl Wege der Länge ℓ

Beweis.

- Jeder Weg der Länge ℓ besteht aus (nicht unbedingt verschiedenen) Knoten v_0, \dots, v_ℓ . Betrachte nun von jedem solchen Weg den Knoten v_r (für ein festgelegtes $r \in \{0, \dots, \ell\}$).
- Es gibt $w_G^r(v_r)$ verschiedene Pfade der Länge r , die in v_r enden, und es gibt $w_G^{\ell-r}(v_r)$ verschiedene Pfade, die in v_r beginnen.
- ⇒ Es gibt also $w_G^r(v_r) \cdot w_G^{\ell-r}(v_r)$ verschiedene Pfade der Länge ℓ , die an der Stelle r den Knoten v_r besuchen.
- ⇒ Die Summe ergibt genau die Anzahl aller Pfade der Länge ℓ , da die entsprechenden Wege der einzelnen Summanden sich im Knoten an der Stelle r unterscheiden und somit nichts doppelt gezählt wird.



Dichte der Ordnung ℓ

- Maximale Anzahl der Wege der Länge ℓ in einem Graphen mit n Knoten: $n \cdot (n - 1)^\ell$
- Dichte der Ordnung ℓ :

$$\rho_\ell(G) = \frac{w_\ell(G)}{n \cdot (n - 1)^\ell}$$

- $\rho_1(G) = \rho(G)$, denn in $w_1(G)$ wird jede Kante doppelt gezählt.

Monotonität der Dichte

Satz

Für alle Graphen G und alle natürlichen Zahlen $\ell \geq 2$ gilt:

$$\rho_\ell(G) \leq \rho_{\ell-1}(G)$$

Beweis.

Da gilt $w_\ell(G) = \sum_{v \in V} w_G^r(v) \cdot w_G^{\ell-r}(v)$, gilt insbesondere für $r = 1$:

$$\begin{aligned} w_\ell(G) &= \sum_{v \in V} w_G^1(v) \cdot w_G^{\ell-1}(v) \\ &\leq (n-1) \cdot \sum_{v \in V} w_G^{\ell-1}(v) = (n-1) \cdot w_{\ell-1}(G) \end{aligned}$$



η -dichte Teilgraphen der Ordnung ℓ

Definition

In einem Graphen $G = (V, E)$ bezeichnet man den durch eine Knotenteilmenge $U \subseteq V$ induzierten Teilgraphen genau dann als η -dichten Teilgraphen der Ordnung ℓ , wenn gilt

$$\rho_\ell(G[U]) \geq \eta$$

- Aus dem Monotonitätssatz folgt, dass jeder η -dichte Teilgraph der Ordnung ℓ auch ein η -dichter Teilgraph der Ordnung $\ell - 1$ ist.
- Die η -dichten Teilgraphen der Ordnung $\ell \geq 2$ sind (analog zu den η -dichten Teilgraphen) geschachtelt.
- Wenn man eine Dichte η festlegt und die η -dichten Graphen wachsender Ordnung betrachtet, dann werden diese der Clique immer ähnlicher.

Dichte unendlicher Ordnung

Definition

Die *Dichte unendlicher Ordnung* ist

$$\rho_{\infty}(G) = \lim_{\ell \rightarrow \infty} \rho_{\ell}(G)$$

Satz

Sei $G = (V, E)$ ein ungerichteter Graph

- 1 $\rho_{\infty}(G)$ ist entweder Null oder Eins.
- 2 G ist genau dann eine Clique, wenn $\rho_{\infty} = 1$.

- Die einzigen Graphen, die für ein $\eta > 0$ und für jede Ordnung ℓ η -dicht der Ordnung ℓ sind, sind die Cliques.
- Die Ordnung erlaubt eine gewisse Skalierung, wie wichtig Kompaktheit im Vergleich zu Dichte ist.

Durchschnittsgrad

- Dichte und Durchschnittsgrad eines Graphen hängen unmittelbar zusammen:

$$\bar{d}(G) = \rho(G) \cdot (n - 1)$$

- Ein k -dichter (Teil-)Graph (bzgl. Durchschnittsgrad) kann definiert werden als (Teil-)Graph mit Durchschnittsgrad $\geq k$.
- η -dichte Graphen (bzgl. prozentualer Dichte) der Größe ℓ sind $\eta(\ell - 1)$ -dichte Graphen (bzgl. Durchschnittsgrad).
- k -dichte Graphen (bzgl. Durchschnittsgrad) der Größe ℓ sind $\frac{k}{\ell-1}$ -dichte Graphen (bzgl. prozentualer Dichte).
- Jeder k -Core ist ein k -dichter (Teil-)Graph.
- k -dichte Teilgraphen sind nicht abgeschlossen unter Exklusion und auch nicht geschachtelt.
(Löscht man aus einem k -regulären Graph einen Knoten, fällt der Durchschnittsgrad unter k .)

Verallgemeinerung des Satzes von Turán

Satz (Dirac, 1963)

Sei $G = (V, E)$ ein ungerichteter Graph.

Wenn $m > \frac{n^2}{2} \cdot \frac{k-2}{k-1}$, dann enthält G einen Teilgraph der Größe $k+r$ mit Durchschnittsgrad $\geq k+r-1 - \frac{r}{k+r}$ für alle $r \in \{0, \dots, k-2\}$ und $n \geq k+r$.

(Der Fall $r = 0$ entspricht dem Satz von Turán.)

Der größtmögliche Durchschnittsgrad

$$\gamma^*(G) = \max_{U \subseteq V, U \neq \emptyset} \{\overline{\deg}(G[U])\}$$

Problem

Problem: **DensestSubgraph**

Eingabe: Graph G

Frage: Eine Knotenmenge mit maximalem induzierten Durchschnittsgrad

Satz

Das Problem **DensestSubgraph** kann für Graphen mit n Knoten und m Kanten in Zeit $\mathcal{O}\left(mn(\log n) \left(\log \frac{n^2}{m}\right)\right)$ gelöst werden.

Der größtmögliche Durchschnittsgrad

- Formulierung von **DensestSubgraph** als MaximumFlow-Problem mit Parameter $\gamma \in \mathbb{Q}^+$
- Sei $G = (V, E)$ ein ungerichteter Graph mit n Knoten und m Kanten.
- Betrachte Flussnetzwerk bestehend aus Graph $G' = (V', E')$ und Kapazitätsfunktion $u_\gamma : E' \rightarrow \mathbb{Q}^+$
- Füge zu V eine Quelle s und eine Senke t hinzu:

$$V' = V \cup \{s, t\}$$

Der größtmögliche Durchschnittsgrad

- Konstruiere die neuen Kanten wie folgt:
 - ersetze jede (ungerichtete) Kante von G durch zwei gerichtete Kanten der Kapazität 1,
 - verbinde die Quelle mit allen Knoten in V durch eine Kante der Kapazität m ,
 - verbinde alle Knoten in V mit der Senke durch eine Kante der Kapazität $m + \gamma - \deg_G(v)$.

$$E' = \{(v, w) : \{v, w\} \in E\} \cup \\ \{(s, v) : v \in V\} \cup \\ \{(v, t) : v \in V\}$$

$$u_\gamma(v, w) = \begin{cases} 1 & \text{falls } \{v, w\} \in E \\ m & \text{falls } v = s \\ m + \gamma - \deg_G(v) & \text{falls } w = t \\ 0 & \text{falls } (v, w) \notin E' \end{cases}$$

Der größtmögliche Durchschnittsgrad

- Wir betrachten Schnitt-Kapazitäten im Netzwerk.
 - Sei S, T eine Partitionierung der Knotenmenge V' in zwei disjunkte Menge mit $s \in S$ und $t \in T$.
 - Definiere $S_+ = S \setminus \{s\}$ und $T_+ = T \setminus \{t\}$.
- $\Rightarrow S_+ \cup T_+ = V$

Der größtmögliche Durchschnittsgrad

Falls $S_+ = \emptyset$, dann ist die Kapazität des Schnitts $c(S, \bar{S}) = m|V| = mn$, ansonsten erhält man

$$\begin{aligned}
 c(S, T) &= \sum_{v \in S, w \in T} u_\gamma(v, w) \\
 &= \sum_{w \in T_+} u_\gamma(s, w) + \sum_{v \in S_+} u_\gamma(v, t) + \sum_{v \in S_+, w \in T_+} u_\gamma(v, w) \\
 &= m|T_+| + \left(m|S_+| + \gamma|S_+| - \sum_{v \in S_+} \deg_G(v) \right) + \sum_{\substack{v \in S_+, w \in T_+ \\ \{v, w\} \in E}} 1 \\
 &= m|V| + |S_+| \left(\gamma - \frac{1}{|S_+|} \left(\sum_{v \in S_+} \deg_G(v) - \sum_{\substack{v \in S_+, w \in T_+ \\ \{v, w\} \in E}} 1 \right) \right) \\
 &= m|V| + |S_+|(\gamma - \overline{\deg}(G[S_+]))
 \end{aligned}$$

Der größtmögliche Durchschnittsgrad

- γ ist der Test-Wert für den maximalen Durchschnittsgrad.
- Wie kann man nun feststellen, ob er zu klein oder zu groß ist?

Satz

Seien S und T so gewählt, dass sie einem Minimum- s, t -Schnitt für γ entsprechen. Dann gilt:

- 1 Wenn $S_+ \neq \emptyset$, dann $\gamma \leq \gamma^*(G)$.
- 2 Wenn $S_+ = \emptyset$, dann $\gamma \geq \gamma^*(G)$.

Der größtmögliche Durchschnittsgrad

Beweis.

① $S_+ \neq \emptyset$:

Da $c(\{s\}, \overline{V' \setminus \{s\}}) = m|V| \geq c(S, T)$, gilt
 $|S_+|(\gamma - \overline{\deg}(G[S_+])) \leq 0$.

Also $\gamma \leq \overline{\deg}(G[S_+]) \leq \gamma^*(G)$.

② $S_+ = \emptyset$: Annahme: $\gamma < \gamma^*(G)$

Sei $U \subseteq V$ eine nichtleere Menge mit $\overline{\deg}(G[U]) = \gamma^*(G)$.

Mit der Gleichung für $c(S, T)$ erhält man

$$c(U \cup \{s\}, \overline{U \cup \{t\}}) = mn + |U|(\gamma - \gamma^*(G)) < mn = c(S, T)$$

(Widerspruch zur Minimalität der Schnittkapazität $c(S, T)$)

$$\Rightarrow \gamma \geq \gamma^*(G)$$



Der größtmögliche Durchschnittsgrad

- Algorithmus benutzt binäre Suche, um den richtigen Wert für γ zu finden
- $\gamma^*(G)$ kann nur endlich viele Werte annehmen:

$$\gamma^*(G) \in \left\{ \frac{2i}{j} : i \in \{0, \dots, m\} \text{ und } j \in \{1, \dots, n\} \right\}$$

- kleinste mögliche Distanz zwischen zwei Werten der Menge ist $\frac{1}{n(n-1)}$.

Der größtmögliche Durchschnittsgrad

Algorithmus 5 : DensestSubgraph mit MinCut und binärer Suche

Input : Graph $G = (V, E)$

Output : Eine Menge von k Knoten von G

Initialisiere $l := 0$, $r := m$ und $U := \emptyset$;

while $r - l \geq \frac{1}{n(n-1)}$ **do**

$\gamma := \frac{l+r}{2}$;

 Konstruiere Fluss-Netzwerk (V', E', u_γ) ;

 Finde Minimum-Schnitt (S, T) des Fluss-Netzwerks;

if $S = \{s\}$ **then**

$r := \gamma$

else

$l := \gamma$;

$U := S - \{s\}$

return U

Der größtmögliche Durchschnittsgrad

- Iteration wird $\lceil \log((m+1)n(n-1)) \rceil = \mathcal{O}(\log n)$ -mal ausgeführt
 - In jeder Iteration MinCut-Berechnung mit Push-Relabel-Algorithmus (Goldberg/Tarjan) in $\mathcal{O}(mn \log \frac{n^2}{m})$ für ein Netzwerk mit n Knoten und m Kanten (Wir haben zwar $n+2$ Knoten und $2m+2n$ Kanten, das ändert asymptotisch aber nichts.)
- ⇒ Gesamtkomplexität: $\mathcal{O}\left(mn(\log n)(\log \frac{n^2}{m})\right)$
- mit parametrischen MaxFlow-Algorithmen: $\mathcal{O}\left(mn \log \frac{n^2}{m}\right)$

Durchschnittlicher Grad in gerichteten Graphen

- Es ist nicht unbedingt offensichtlich, wie Dichte (i.S.d. Durchschnittsgrads) auf gerichtete Graphen zu übertragen ist.
 - Durchschnittlicher Eingangs- und Ausgangsgrad sind gleich.
- ⇒ keine orientierungsabhängigen Maße
- Hubs & Authorities: Für gerichteten Graph $G = (V, E)$ und nichtleere (nicht unbedingt disjunkte) Knotenmengen $S, T \subseteq V$ sei $E(S, T)$ die Menge der Kanten, die von einem Knoten in S zu einem Knoten in T gehen:

$$E(S, T) = \{(u, v) : u \in S \text{ und } v \in T\}$$

Definiere **durchschnittlichen gerichteten Grad des Paares (S, T)**
als (Kannan/Vinay)

$$\overline{\deg}_G(S, T) = \frac{|E(S, T)|}{\sqrt{|S| \cdot |T|}}$$

Durchschnittlicher gerichteter Grad

- S ist dann die Menge der Hubs,
 T ist die Menge der Authorities
- Wenn $S = T$, dann kommt genau der konventionelle Durchschnittsgrad heraus.
- Durchschnittsgrad-Maximum eines gerichteten Graphen $G = (V, E)$:

$$\gamma^* = \max_{\substack{S, T \subseteq V \\ S \neq \emptyset, T \neq \emptyset}} \{ \overline{\deg}_G(S, T) \}$$

- Kann mit Linear Programming in Polynomialzeit gelöst werden (Charikar, 2000).

Dense ℓ -Subgraph

- Graph mit Durchschnittsgrad $\overline{\deg}(G)$ muss nicht unbedingt einen echten Teilgraphen mit gleichem Durchschnittsgrad enthalten.
- Maximum des Durchschnittsgrads eines Teilgraphen auf ℓ Knoten:

$$\gamma^*(G, \ell) = \max \left\{ \overline{\deg}(G[U]) : U \subseteq V \text{ und } |U| = \ell \right\}$$

Problem

Problem: **Dense- ℓ -Subgraph**

Eingabe: Graph G , Parameter $\ell \in \mathbb{N}$

Frage: Eine Knotenmenge der Kardinalität ℓ mit maximalem induzierten Durchschnittsgrad

Dense ℓ -Subgraph

- Optimierungsproblem **Dense ℓ -Subgraph** ist \mathcal{NP} -hart (Instanz $(G, \ell, \ell - 1)$ des zugehörigen Entscheidungsproblems entspricht der Suche nach einer Clique der Größe ℓ in G).
- ⇒ Wie sieht es mit Approximation aus?

Greedy-Approximation für Dense ℓ -Subgraph

Algorithmus 6 : Approximation eines ℓ -Subgraph mit hohem $\overline{\text{deg}}$

Input : Graph $G = (V, E)$ und

gerader Parameter $\ell \in \mathbb{N}$ (mit $|V| \geq \ell$)

Output : Menge von ℓ Knoten von G

Sortiere die Knoten in absteigender Reihenfolge ihrer Grade;

Sei H die Menge von $\frac{\ell}{2}$ Knoten von höchstem Grad;

Berechne $N_H(v) = |N(v) \cap H|$ für alle Knoten $v \in V \setminus H$;

Sortiere die Knoten in $V \setminus H$ in absteigender Reihenfolge der N_H -Werte;

Sei R die Menge von $\frac{\ell}{2}$ Knoten von $V \setminus H$ mit den höchsten N_H -Werten;

return $H \cup R$
