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

Dichte und Wege

- Definition der Dichte entspricht der durchschnittlichen Existenz von möglichen Kanten innerhalb eines (induzierten) Teilgraphen
- Jede ungerichtete Kante entspricht zwei gerichteten Wegen der Länge 1.
- Verallgemeinerung: Dichte auf der Basis (gerichteter) Wege (Walks) beliebiger Länge (mit möglichen Knoten-/Kantenwiederholungen)
- Grad der Ordnung $\ell \in \mathbb{N}$ eines Knotens v, Anzahl der Wege der Länge ℓ in G, die in v starten: $w_\ell(v)$
- $w_0(v) = 1, \quad w_1(v) = \deg(v)$
- Anzahl Wege der Länge ℓ in einem Graphen G: w_ℓ
- $w_0 = n, \quad w_1 = 2m$

Anzahl Wege der Länge ℓ

Satz

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

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

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

Anzahl Wege der Länge ℓ

Beweis.

- Jeder Weg der Länge ℓ besteht aus (nicht unbedingt verschiedenen) Knoten $v_0, \ldots, v_\ell$.
- Betrachte nun von jedem solchen Weg den Knoten v_r (für ein festgelegtes $r \in \{0, \ldots, \ell\}$).
- Es gibt $w_r(v_r)$ verschiedene Wege der Länge r, die in v_r enden, und es gibt $w_{\ell-r}(v_r)$ verschiedene Wege der Länge $\ell - r$, die in v_r beginnen.

$\Rightarrow$ Es gibt also $w_r(v_r) \cdot w_{\ell-r}(v_r)$ verschiedene Wege der Länge ℓ, die an der Stelle r den Knoten v_r besuchen.

$\Rightarrow$ Die Summe ergibt genau die Anzahl aller Wege der Länge ℓ, da die entsprechenden Wege der einzelnen Summanden sich im Knoten an der Stelle r unterscheiden und somit nichts doppelt gezählt wird.

$\square$

Dichte der Ordnung ℓ

- Maximal mögliche Anzahl von Wegen der Länge ℓ in einem Graphen mit n Knoten (also im vollständigen Graphen K_n):

$$n \cdot (n-1)^{\ell}$$

- Dichte der Ordnung ℓ: (für Graphen mit $n \geq 2$)

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

$\Rightarrow$ $\rho_1(G) = \rho(G)$, denn in $w_1(G)$ zählt jede Kante doppelt:

$$\rho_1 = \frac{w_1}{n(n-1)^1} = \frac{2m}{n(n-1)} = \frac{m}{\binom{n}{2}} = \rho$$

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

$$\begin{aligned} w_\ell &= \sum_{v \in V} w_1(v) \cdot w_{\ell-1}(v) = \sum_{v \in V} \deg(v) \cdot w_{\ell-1}(v) \\ &\leq (n-1) \cdot \sum_{v \in V} w_{\ell-1}(v) = (n-1) \cdot w_{\ell-1} \end{aligned}$$

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

□

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

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

- Jeder η-dichte Teilgraph der Ordnung ℓ ist auch ein η-dichter Teilgraph der Ordnung $\ell - 1$ (siehe Monotonitätssatz).
- Die η-dichten Teilgraphen der Ordnung $\ell \geq 2$ sind (wie die η-dichten Teilgraphen) nicht abgeschlossen unter Exklusion, aber geschachtelt.
- Für eine festlegte Dichte η werden die η-dichten Graphen wachsender Ordnung einer Clique immer ähnlicher.

Dichte unendlicher Ordnung

Definition

Die *Dichte unendlicher Ordnung* ist

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

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 Cliquen.
- Die Ordnung erlaubt eine gewisse Skalierung, wie wichtig Kompaktheit im Vergleich zu Dichte ist.

Durchschnittsgrad

- Dichte und Durchschnittsgrad hängen direkt zusammen:

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

- Ein *k-dichter (Teil-)Graph (bzgl. Durchschnittsgrad)* kann definiert werden als (Teil-)Graph G mit $\bar{d}(G) \geq k$.
- 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 mit n Knoten und m Kanten.

Wenn $m > \frac{n^2}{2} \cdot \frac{k-2}{k-1}$, dann enthält G einen Teilgraph der Größe $k + r$ mit Durchschnittsgrad $\bar{d} \geq k + r - 1 - \frac{r}{k+r}$ für alle $r \in \{0, \ldots, k-2\}$ 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*
Ausgabe: *Teilgraph mit maximalem 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

Beweis:

- 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 (gerichtetes) Flussnetzwerk bestehend aus Graph $G' = (V', E')$ und Kapazitätsfunktion $u_\gamma : E' \to \mathbb{Q}^+$
- Füge zu V eine Quelle s und eine Senke t hinzu:

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

$$\begin{aligned} E' \quad = \quad & \{(v,w),(w,v) : \{v,w\} \in E\} \cup \\ & \{(s,v) : v \in V\} \cup \\ & \{(v,t) : v \in V\}\} \end{aligned}$$

Der größtmögliche Durchschnittsgrad

Neue Kanten (in G'):

- für jede ungerichtete Kante in G zwei gerichtete Kanten der Kapazität 1,
- verbinde Quelle s mit allen Knoten in V durch Kante der Kapazität m,
- verbinde alle Knoten in V mit Senke t durch Kante der Kapazität $m + \gamma - \deg_G(v)$.

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

Der größtmögliche Durchschnittsgrad

- Betrachte Schnitt-Kapazitäten im Netzwerk
- Partitionierung der Knotenmenge V' in zwei disjunkte Mengen S und T mit $s \in S$ und $t \in T$.
- Sei $S_+ = S \setminus \{s\}$ und $T_+ = T \setminus \{t\}$, also $S_+ \cup T_+ = V$

Der größtmögliche Durchschnittsgrad

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

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

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

$\square$

Der größtmögliche Durchschnittsgrad

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

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

- kleinste mögliche Distanz zwischen zwei Werten der Menge ist $\frac{2}{n(n-1)}$.

Der größtmögliche Durchschnittsgrad

Algorithmus 9 : 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 := ∅;
while r − l ≥ 1/(n(n−1)) do
    γ := (l+r)/2;
    Konstruiere Fluss-Netzwerk (V′, E′, u_γ);
    Finde Minimum-Schnitt (S, T) des Fluss-Netzwerks;
    if S = {s} then
        r := γ
    else
        l := γ;
        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.)

$\Rightarrow$ 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.

$\Rightarrow$ 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 Paars (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
- Für $S = T$ kommt genau der konventionelle Durchschnittsgrad heraus.
- Durchschnittsgrad-Maximum eines gerichteten Graphen $G = (V, E)$:

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

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