Fortgeschrittene Netzwerk- und Graph-Algorithmen

Dr. Hanjo Täubig

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

Wintersemester 2009/10

Übersicht

1. Lokale Dichte
   - Kohäsive Gruppen
   - Cliquen

Kohäsive Gruppen

Eigenschaften:

- Gegenseitigkeit
  Gruppenmitglieder wählen sich gegenseitig in die Gruppe und sind im graphtheoretischen Sinn benachbart
- Kompaktheit / Erreichbarkeit:
  Gruppenmitglieder sind gegenseitig gut erreichbar
  (wenn auch nicht unbedingt adjazent), insbesondere
  - auf kurzen Wegen
  - auf vielen verschiedenen Wegen
- Dichte:
  Gruppenmitglieder haben eine große Nachbarschaft innerhalb der Gruppe
- Separation:
  Gruppenmitglieder haben mit größerer Wahrscheinlichkeit Kontakt zu einem anderen Mitglied der Gruppe als zu einem Nicht-Gruppenmitglied

Lokale Dichte

- Eine Gruppeneigenschaft heißt *lokal*, wenn sie bestimmt werden kann, indem man nur den von der Gruppe induzierten Teilgraphen betrachtet.

$\Rightarrow$ Separation ist *nicht lokal*, weil hier auch die Verbindungen zu den anderen Knoten betrachtet werden

- Viele Definitionen von kohäsiven Gruppen verlangen außer einer Eigenschaft Π auch *Maximalität* (im Sinne von Nichterweiterbarkeit), d.h. die Gruppe darf nicht in einer anderen größeren Gruppe enthalten sein.
- Maximalität verletzt die Lokalitätsbedingung

$\Rightarrow$ Betrachten diese Eigenschaften ohne Maximalitätsbedingung

- Lokalität reflektiert wichtige Eigenschaft von Gruppen:
  - Invarianz unter Veränderung des Netzwerks außerhalb der Gruppe
  - Innere Robustheit und Stabilität ist eine wichtige Gruppeneigenschaft

Cliquen

Definition

Sei $G = (V, E)$ ein ungerichteter Graph. Ein Knotenteilmenge $U \subseteq V$ heißt Clique genau dann, wenn der von U in G induzierte Teilgraph $G[U]$ ein vollständiger Graph ist.
Eine Clique U in G ist eine maximale Clique, falls es keine Clique U' mit $U \subset U'$ in G gibt.

Cliquen – ideale kohäsive Gruppen

Cliquen sind ideale kohäsive Gruppen:
(Sei U eine Clique der Kardinalität k.)

- Cliquen haben größtmögliche Dichte

$$\delta(G[U]) = \bar{d}(G[U]) = \Delta(G[U]) = k - 1$$

- Cliquen besitzen größtmögliche Kompaktheit

$$\mathrm{diam}(G[U]) = 1$$

- Cliquen sind bestmöglich verbunden
  U ist $(k-1)$-fach knotenzusammenhängend und
  $(k-1)$-fach kantenzusammenhängend

Satz von Turán

Satz (Turán, 1941)

Sei $G = (V, E)$ ein ungerichteter Graph mit $n = |V|$ und $m = |E|$. Falls $m > \frac{n^2}{2} \cdot \frac{k-2}{k-1}$, dann existiert eine Clique der Größe k in G.

Spezialfall:

Satz (Mantel, 1907)

Die maximale Anzahl von Kanten in einem dreiecksfreien Graphen ist $\left\lfloor \frac{n^2}{4} \right\rfloor$.

Da die meisten (z.B. soziale) Netzwerke eher dünn sind (also $o(n^2)$ Kanten haben), müssen sie nicht unbedingt von vornherein Cliquen einer bestimmten Größe > 2 enthalten.

Maximale Cliquen

- Graphen enthalten immer maximale Cliquen.
- Meistens sind es sogar viele.
- Sie können sich überlappen (ohne identisch zu sein).

Satz (Moon & Moser, 1965)

Jeder ungerichtete Graph G mit n Knoten hat höchstens $3^{\lceil \frac{n}{3} \rceil}$ maximale Cliquen.

Cliquen-Struktur

- Cliquen sind abgeschlossen unter Exklusion, d.h. wenn U eine Clique in G ist und v ein Knoten aus U, dann ist $U \setminus \{v\}$ auch eine Clique.
  Oder anders gesagt:
  Die Cliqueneigenschaft ist eine *hereditäre* Grapheigenschaft, denn sie vererbt sich auf induzierte Teilgraphen.

- Cliquen sind geschachtelt, d.h. jede Clique der Größe n enthält eine Clique der Größe $n-1$ (sogar n davon).
  (Das folgt hier sofort aus dem Abschluss unter Exklusion. Für andere Eigenschaften, die nicht unter Exklusion abgeschlossen sind, muss man das aber extra beweisen.)

Generalisierte Cliquen

Sei $G = (V, E)$ ein ungerichteter Graph, U eine Teilmenge der Knoten und $k > 0$ eine natürliche Zahl.
Generalisierte (distanz-basierte) Cliquen:

- U heißt k-Clique g.d.w. $\forall u, v \in U : \quad d_G(u, v) \leq k$.
- U heißt k-Club g.d.w. $\mathrm{diam}(G[U]) \leq k$.
- U heißt k-Clan g.d.w. U ist eine maximale k-Clique und U ist ein k-Club.

- k-Cliquen sind nicht lokal definiert (die Distanzen können sich aus Pfaden ergeben, die über Knoten außerhalb von U führen).
- Obwohl k-Clubs und k-Clans lokal definiert sind (abgesehen von der Maximalitätsbedingung), sind sie nur von geringerem Interesse. Distanz-basierte Cliquen sind i.A. nicht abgeschlossen unter Exklusion und nicht geschachtelt.

Grundfunktionen

In $\mathcal{O}(m+n)$ können folgende Funktionen berechnet werden:

- Bestimme, ob eine gegebene Knotenteilmenge $U \subseteq V$ eine Clique in G ist.
  Bestimme für jede Kante in G, ob beide Endknoten in U sind.
  Zähle die Fälle und vergleiche mit $|U| \cdot (|U| - 1)/2$.
- Bestimme, ob eine gegebene Clique $U \subseteq V$ maximal ist in G.
  Teste, ob es einen Knoten in $V \setminus U$ gibt, der adjazent zu allen Knoten in U ist.

Maximale Clique

Bestimme die lexikographisch kleinste maximale Clique U, die eine gegebene Clique $U' \subseteq V$ enthält.

- Annahme: V ist eine geordnete Menge
- Seien $U, U' \in V$. Def.: $U \leq U' \Leftrightarrow$ der kleinste Knoten der nicht in $U \cap U'$ ist, ist in U.
- Starte mit $U = U'$
- Iteriere über alle $v \in V \setminus U$ in aufsteigender Reihenfolge und teste, ob $U \subseteq N(v)$.
- Falls ja, dann füge v zu U hinzu.
- Am Ende ist U eine maximale Clique, die U' enthält.

$\Rightarrow$ ebenfalls $\mathcal{O}(m + n)$

Cliquen maximaler Kardinalität

Maximum-Clique: Clique der größtmöglichen Kardinalität in einem gegebenen Graphen

Primitiver Algorithmus: erschöpfende Suche
Zähle alle Kandidatensets $U \subseteq V$ auf
und bestimme, ob U eine Clique ist.
Gib die größte gefundene Clique aus.
$\Rightarrow$ Laufzeit $\mathcal{O}(n^2 \cdot 2^n)$

Clique-Problem

Entscheidungproblem:

Problem

Problem:	***Clique***
Eingabe:	*Graph* G, *Parameter* $k \in \mathbb{N}$
Frage:	*Existiert eine Clique der Kardinalität* $\geq k$ *in* G?

Härte des Clique-Problems

Sei $\omega(G)$ die Größe der Maximum-Clique(n) in G.
Wenn wir einen Algorithmus hätten, der **Clique** in Zeit $T(n)$ entscheidet, dann könnten wir $\omega(G)$ in Zeit $\mathcal{O}(T(n) \cdot \log n)$ mit binärer Suche berechnen.
Andererseits ergibt sich aus jedem Algorithmus zur Berechnung von $\omega(G)$ in Zeit $T(n)$ ein Algorithmus, der **Clique** in $T(n)$ entscheidet.
Ein polynomieller Algorithmus für das eine Problem würde als einen polynomiellen Algorithmus für das jeweils andere Problem implizieren.

Aber:

Satz

Clique *ist $\mathcal{NP}$-complete.*

Beweis: Reduktion von **Satisfiability** (Erfüllbarkeit)

Versteckte Cliquen

Folgerung

Falls $\mathcal{P} \neq \mathcal{NP}$ gilt, gibt es keinen Polynomialzeit-Algorithmus, der eine Clique der Größe k in einem Graphen findet, der garantiert eine solche Clique der Größe k enthält.

Bemerkung: die Schwierigkeit, eine versteckte Clique zu finden, hängt nicht von der Größe der Clique ab. Auch Cliquen der Größe $(1-\varepsilon)\cdot n$ können nicht in Polynomialzeit gefunden werden.

Maximum-Clique: besserer exponentieller Algorithmen

Satz

Eine Clique maximaler Kardinalität kann in Zeit $\mathcal{O}^*(1.3803^n)$ berechnet werden.

($\mathcal{O}^*$ ignoriert polynomielle Faktoren)

Maximum-Clique: besserer exponentieller Algorithmus

Beweis.

- Sei v ein Knoten mit minimalem Grad $\deg_G(v) = \delta(G)$
- Falls $\delta(G) \geq n - 3$, dann
  - fehlen in G nur einfache Pfade und Kreise im Vergleich zum vollständigen Graphen K_n

  $\Rightarrow$ Maximum Clique kann in $\mathcal{O}(m + n)$ berechnet werden:
  - In einem Pfad P kann man die Größe einer maximalen unabhängigen Menge (independent set) berechnen als $\lceil |V(P)|/2 \rceil$.
  - In einem Kreis C ist die Größe $\lfloor |V(C)|/2 \rfloor$.
  - Da die Pfade und Kreise paarweise disjunkt sind, kann man die einzelnen Werte einfach addieren.

Maximum-Clique: besserer exponentieller Algorithmus

Beweis.

- Ansonsten sei v ein Knoten mit Grad $\deg_G(v) \leq n-4$
- Jede Maximum-Clique ist entweder
  - $\{v\}$ vereinigt mit einer Maximum-Clique von $G[N(v)]$ oder
  - eine Maximum-Clique von $G[V \setminus \{v\}]$.

$\Rightarrow$ Rekursive Berechnung mit worst-case-Zeit

$$T(n) \leq T(n-4) + T(n-1) + c \cdot (m+n)$$

$\Rightarrow$ $T(n) \in \mathcal{O}^*(\beta^n)$ mit $\beta \approx 1.3803$, wobei β die größte reelle Nullstelle des charakteristischen Polynoms $\beta^4 - \beta^3 - 1$ ist

$\square$

Approximation von Maximum-Cliquen

Approximation der größten Clique mit Faktor $n/2$ ist einfach (wähle die Endknoten einer Kante, falls es eine gibt).

Satz

Es gibt einen Algorithmus, der bei Eingabe eines Graphen G mit n Knoten in polynomieller Zeit eine Clique ausgibt, deren Größe maximal um einen Faktor $\mathcal{O}\left(\frac{n}{(\log n)^2}\right)$ von der Maximum-Cliquengröße $\omega(G)$ abweicht.

Satz

Falls nicht $\mathcal{NP} = \mathcal{ZPP}$ gilt, dann existiert kein Polynomialzeitalgorithmus, der bei Eingabe eines Graphen G mit n Knoten eine Clique ausgibt, deren Größe maximal um einen Faktor $n^{1-\epsilon}$ von der Maximum-Cliquengröße $\omega(G)$ abweicht (für jedes $\epsilon > 0$).

Suche nach Cliquen fester Größe

- In einigen Fällen reicht es, nach Cliquen fester Größe zu suchen

$\Rightarrow$ Cliquengröße wird nicht als Teil der Eingabe betrachtet

- Vollständige Suche: $\mathcal{O}(k^2 \cdot n^k) = \mathcal{O}(n^k)$,
  wenn Cliquengröße k fest ist

Bessere Komplexität für Dreiecke

- $A(G)$: Adjazenzmatrix von Graph G
- $B(G) = A(G)^2 = A(G) \cdot A(G)$: Einträge b_{ij} sind die Wege der Länge 2 zwischen den Knoten v_i und v_j

$\Rightarrow$ läßt sich durch schnellere Matrixmultiplikation ausrechnen, z.B. in $\mathcal{O}(n^{2.376})$ (Coppersmith / Winograd, 1990)

- Existiert ein Eintrag $b_{ij} \geq 1$ mit $i \neq j$, dann gibt es einen Knoten $u \in V$, der zu v_i und zu v_j adjazent ist.
- Wenn nun auch eine Kante $\{v_i, v_j\}$ existiert, dann enthält der Graph ein Dreieck $\{v_i, v_j, u\}$

$\Rightarrow$ Checke für alle echt positiven Werte b_{ij}, ob es eine Kante $\{v_i, v_j\}$ gibt

$\Rightarrow$ Zeit-Komplexität: $\mathcal{O}(n^{\alpha})$ mit $\alpha < 2.376$

Cliquen fester Größe $k \geq 3$

Sei $\alpha(r, s, t)$ so definiert, dass man die Multiplikation einer $n^r \times n^s$-Matrix mit einer $n^s \times n^t$-Matrix in Zeit $\mathcal{O}\left(n^{\alpha(r,s,t)}\right)$ berechnen kann.

Satz

Für jedes $k \geq 3$ existiert ein Algorithmus, der eine Clique der Größe k in einem Graphen mit n Knoten finden kann (falls eine solche existiert) und der in Zeit $\mathcal{O}\left(n^{\beta(k)}\right)$ läuft, wobei $\beta(k) = \alpha(\lfloor k/3 \rfloor, \lceil (k-1)/3 \rceil, \lceil k/3 \rceil)$.

Cliquen fester Größe $k \geq 3$

Beweis.

- Seien $k_1 = \lfloor k/3 \rfloor$, $k_2 = \lceil (k-1)/3 \rceil$ und $k_3 = \lceil k/3 \rceil$
- Es gilt $k = k_1 + k_2 + k_3$.
- Konstruiere tripartiten Hilfsgraph $\tilde{G}$
  - $\tilde{V} = \tilde{V}_1 \cup \tilde{V}_2 \cup \tilde{V}_3$, wobei V_i aus allen Cliquen der Größe k_i in G besteht
  - Knoten $U \in \tilde{V}_i$ und $U' \in \tilde{V}_j$ sind adjazent in $\tilde{G}$ g.d.w. $i \neq j$ und $U \cup U'$ eine Clique der Größe $k_i + k_j$ in G ist.
- Teste $\tilde{G}$ auf Dreiecke

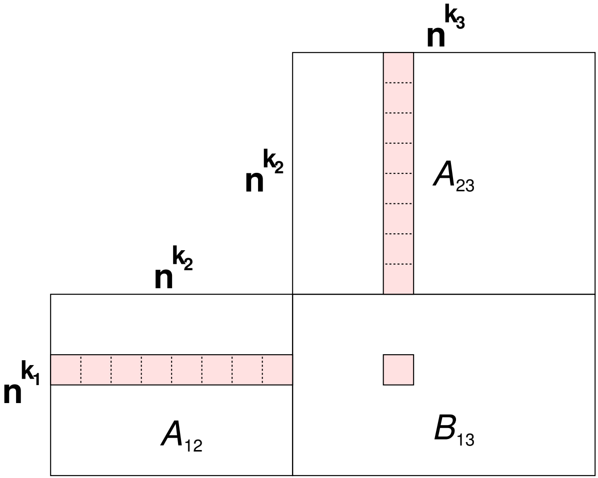
Cliquen fester Größe $k \geq 3$

Beweis.

- Dreieck $\{U_1, U_2, U_3\}$ impliziert eine Clique der Größe k in G
- geht mit schneller Matrixmultiplikation, aber hier muss eine $n^{k_1} \times n^{k_2}$-Matrix (Kanten zwischen $\tilde{V}_1$ und $\tilde{V}_2$) mit einer $n^{k_2} \times n^{k_3}$-Matrix (Kanten zwischen $\tilde{V}_2$ und $\tilde{V}_3$) multipliziert werden, und zwar in Zeit $\mathcal{O}\left(n^{\beta(k)}\right)$
- Berechnung der drei Matrizen A_{12}, A_{23} und A_{13} in $\mathcal{O}\left(n^{\max\{k_1+k_2, k_1+k_3, k_2+k_3\}}\right) = \mathcal{O}\left(n^{\lceil \frac{2k}{3} \rceil}\right)$
  (wird dominiert durch die Zeit $\mathcal{O}(n^{\beta(k)})$ für die Multiplikation der rechteckigen Matrizen, also $B_{13} = A_{12} \cdot A_{23}$)

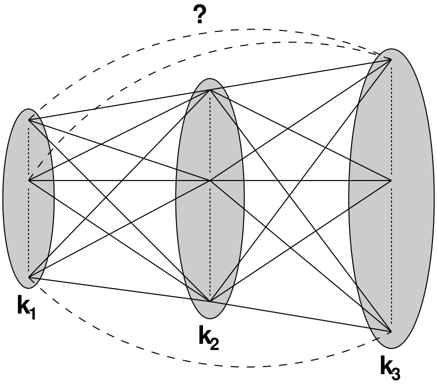
$\square$

Cliquen fester Größe $k \geq 3$



B_{13} wird dann mit A_{13} verknüpft

Cliquen fester Größe $k \geq 3$



Cliquen fester Größe: Beispielkomplexitäten

Cliquengröße	Vollständige Suche	Matrixmultiplikation
3	$\mathcal{O}(n^3)$	$\mathcal{O}(n^{2.376})$
4	$\mathcal{O}(n^4)$	$\mathcal{O}(n^{3.376})$
5	$\mathcal{O}(n^5)$	$\mathcal{O}(n^{4.220})$
6	$\mathcal{O}(n^6)$	$\mathcal{O}(n^{4.751})$
7	$\mathcal{O}(n^7)$	$\mathcal{O}(n^{5.751})$
8	$\mathcal{O}(n^8)$	$\mathcal{O}(n^{6.595})$

Cliquen fester Größe: Mitgliedszahlen

Satz

Für jedes $k \geq 3$ existiert ein Algorithmus, der in Zeit $\mathcal{O}\left(n^{\beta(k)}\right)$ läuft und der für jeden Knoten zählt, an wieviel Cliquen der Größe k er beteiligt ist (in einem Graphen mit n Knoten), wobei $\beta(k) = \alpha(\lfloor k/3 \rfloor, \lceil (k-1)/3 \rceil, \lceil k/3 \rceil)$.

Cliquen fester Größe: Mitgliedszahlen

Beweis.

- Für $k = 3$ (Dreiecke) kann man nicht nur feststellen, *ob* zwei Knoten v_i und v_j zu einem Dreieck gehören, sondern auch *zu wievielen*.
- Wenn Kante $\{v_i, v_j\}$ in G existiert, dann ist diese Anzahl gleich dem Eintrag b_{ij} in der quadrierten Adjazenzmatrix $B(G) = A(G) \cdot A(G)$.
- Anwendung im allgemeinen Fall von $\tilde{G}$:
  für jeden Knoten $v \in V$ sei $C_k(v)$ die Anzahl verschiedener Cliquen der Größe k, in denen v enthalten ist.
- Entsprechend sei $\tilde{C}_3(U)$ die Anzahl der Dreiecke in $\tilde{G}$, zu denen Knoten U gehört.
  (U ist eine Clique der Größe kleiner als k)

Cliquen fester Größe: Mitgliedszahlen

Beweis.

- Cliquen der Größe k können viele verschiedene Repräsentationen in $\tilde{G}$ haben.
- Diese Anzahl ist die Anzahl der Partitionierungen von einer Menge der Kardinalität k in drei Mengen der Kardinalitäten k_1, k_2 und k_3, also der Multinomialkoeffizient $\binom{k}{k_1,k_2,k_3}$.
- o.B.d.A. sei k_1 der kleinste der drei Parameter.
  Sei $\mathcal{U}(v)$ die Menge aller Cliquen U der Größe k_1 in G, so dass $v \in U$. Dann gilt:

$$\sum_{U \in \mathcal{U}(v)} \tilde{C}_3(U) = \binom{(k-1)}{(k_1-1), k_2, k_3} \cdot C_k(v)$$

- Berechne linke Seite in $\mathcal{O}\left(n^{\beta(k)}\right)$ (berechne Matrizen; suche Einträge für alle U, die v enthalten); Berechne $C_k(v)$)