Abteilung für Mathematische Optimierung
TU Braunschweig

Prof. Dr. Sándor Fekete
Dr. Laura Heinrich-Litan

SS 2003

Diskrete Optimierung
Übung 3 vom 29.04.03

Abgabe der Aufgaben bis 15:00 Uhr am **Dienstag, 06.05.03** vor der großen Übung.

Aufgabe 1 (Breitensuche):
Sei $G = (V, E)$ ein ungerichteter Graph. Zeige, dass die Breitensuche aus einem Startknoten $q \in V$ kürzeste Wege von q zu den anderen Knoten liefert.

(Zeige dabei durch Induktion, dass in der Warteschlange zu jedem Zeitpunkt der Abstand zu q monoton wächst und sich zwischen aufeinanderfolgenden Knoten um höchstens 1 unterscheidet. Schließe daraus, dass der korrekte Abstand berechnet wird.) (20 Punkte)

Aufgabe 2 (Tiefensuche):
Bei einem *Online-Problem* hat man nicht die vollständige Information über die Daten einer Probleminstanz, sondern nur unvollständige, lokale Informationen. Beispiele ergeben sich beim Zuweisen von Kunden auf Schalter oder beim Suchen in einem Labyrinth. Wir wollen hier das letztere Problem etwas durchleuchten, indem wir versuchen, in einem unbekannten Graphen mit einer unbekannten Zahl von n unbekannten Knoten von einem Startknoten S aus zu einem Zielknoten Z (dem "Ausgang") zu gelangen. Dabei kostet das Durchlaufen einer Kante (ein Schritt) jeweils eine Zeiteinheit. Unterwegs kann man die "Türen" an einem Knoten markieren[1].

a) Zeige, dass bei einer Tiefensuche keine Kante mehr als zweimal durchlaufen wird. Folgere, dass für insgesamt $n + 1$ Knoten (einschließlich Startknoten) der Ausgang in höchstens $2n - 1$ Schritten gefunden wird.

b) Argumentiere, dass $2n - 1$ Schritte in gewissem Sinne optimal sind: Gib eine Familie von Graphen an, so dass für jeden gewählten Suchpfad einer der Knoten erst nach $2n - 1$ Schritten erreicht wird.

c) Argumentiere, dass sich für keinen Graphen eine Garantie von n Schritten unterbieten lässt. Was müsste man finden, um für einen Graphen diesen Wert einhalten zu können?

(20 Punkte)

[1] Eine Methode ist beschrieben in Umberto Eco, "Der Name der Rose". Im Film wird aus dem Algorithmus ein Pullover.

Aufgabe 3 (Das Orakel von Kevin Bacon):
Dem *Orakel von Kevin Bacon* liegt der Schauspielergraph S zugrunde: Schauspieler sind durch Knoten repräsentiert. Zwei Schauspielerknoten sind durch eine Kante verbunden, wenn sie gemeinsam in einem Film gespielt haben. Der Knoten von Kevin Bacon hat den Wert 0; die *Kevin-Bacon-Zahl* (KBZ) eines anderen Schauspielers ist die Länge eines kürzesten Weges im Schauspielgraphen S. (So hat etwa Tom Hanks die Kevin-Bacon-Zahl 1, da er mit ihm gemeinsam in der Raumkapsel *Apollo 13* gesessen hat.)

Das Orakel ist im Web verfügbar: `http://www.cs.virginia.edu/oracle/`. Interessant sind dabei auch die verschiedenen Hilf- und Infoseiten. Die zugrundeliegenden Filmdaten sind der *Internet Movie Database* entnommen: `http://www.imdb.com/`

Jetzt die Fragen:

a) Beschreibe eine Strategie, mit der man auf jeden Fall einen Schauspieler möglichst hoher KBZ im Graphen S finden kann, auch wenn man noch nie etwas von Hollywood gehört oder gesehen hat. Welcher Graphenalgorithmus steckt dahinter?

b) Finde einen Knoten mit mindestens KBZ 4. (Früher wurde man mit mindestens KBZ 7 in die "Hall of Fame" aufgenommen; das ist Übungsteilnehmern bisher immer gelungen; leider werden dort keine Einträge mehr vorgenommen.)

c) Will man das Orakel nicht nur als "Black Box" befragen, sondern dafür sorgen, dass es schnell Antworten geben kann, benötigt man andere algorithmische Methoden. Man unterscheidet dabei zwischen *Preprocessing* ("Vorarbeit") und *Queries* ("Anfragen").

   Welche Datenstruktur würden Sie berechnen, um Anfragen schnell zu beantworten? (Dazu gehört auch die Angabe eines gültigen Pfades mit Filmen!) Wie groß ist diese Datenstruktur, wenn jeder der $n = 500.000$ Schauspieler und jeder Filmname 32 Byte benötigt?

   Wie lange dauert das Preprocessing in Abhängigkeit von n, wie lange dauert die Beanwortung einer Anfrage?

d) Das Orakel bietet auch an, Abstände zwischen beliebigen Knoten in S zu berechnen. Beschreiben Sie zwei mögliche Lösungswege, einen ohne, einen mit Preprocessing. Wenn man zusätzlich berücksichtigt, dass der tatsächlich zur Verfügung stehende Speicherplatz begrenzt ist: Wie kann man vorgehen, um nicht immer, aber "sehr oft" auf vorher berechnete Daten zurückgreifen zu können, so dass schnelle Antwortzeiten möglich sind?

(20 Punkte)

Zusatzaufgabe zum Surfen, Lesen, Tüfteln

Älter als die KBZ ist die sogenannte *Erdős-Zahl* (EZ). Hier betrachtet man einen Graphen, in dem Knoten Mathematikern entsprechen; Kanten verbinden Mathematiker, die gemeinsam veröffentlicht haben. Pál Erdős (1913-1996) war der produktivste Mathematiker aller Zeiten mit schätzungsweise 1500 publizierten Artikeln.

(Siehe `http://www.paulerdos.com/`, bzw. `http://www.oakland.edu/~grossman/erdoshp.html` über Erdős-Zahlen.)

Einen sehr (!) kleinen, Informatik-lastigen Ausschnitt seiner Publikationen findet man in der "Computer Science Bibliography" an der Universität Trier, dem DBLP:

`http://www.informatik.uni-trier.de/~ley/db/indices/a-tree/e/Erd=ouml=s:Paul.html` .

Betrachten wir nun den Teilgraphen G_{DBLP}, in dem Kanten für im DBLP aufgeführten Artikel stehen. Dann können wir sinngemäß die DBLP-EZ definieren.

Welche DBLP-EZ haben:
Bettina Eick, Sándor Fekete, Laura Heinrich-Litan, Udo Ott, Uwe Zimmermann?

(Antwort bitte jeweils mit Pfad in der Form Name, Artikelnummer, Name, ...)

(je 4 Zusatzpunkte)