Karlsruher Institut für Technologie
Institut für Theoretische Informatik

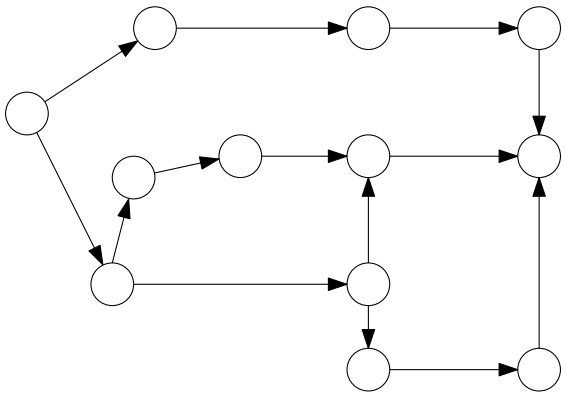
Prof. Dr. Peter Sanders
Dr. Christian Schulz, Dr. Simon Gog
Michael Axtmann

7. Übungsblatt zu Algorithmen II im WS 2016/2017

http://algo2.iti.kit.edu/AlgorithmenII_WS16.php
{christian.schulz,michael.axtmann,sanders,simon.gog}@kit.edu

Aufgabe 1 (*Rechnen: Dinitz Algorithmus*)

Führen Sie den Algorithmus von Dinitz auf dem gegebenen Graphen aus. Schreiben Sie dazu die Distanzlabel in die Knoten und geben Sie in jeder Phase den Residualgraphen, sowie den Schichtgraphen an. Markieren Sie Ihren Fluss im Schichtgraphen, indem Sie jede Kante mit ihrem Fluss und ihrer Kapazität beschriften.



Aufgabe 2 (*Kleinaufgaben: Eigenschaften von Flüssen*)

a) Nach Vorlesung ist eine gültige Distanzfunktion $d(\cdot)$ für *Dinics Algorithmus* gegeben durch:

- $d(t) = 0$
- $d(u) \leq d(v) + 1 \quad \forall\ (u, v) \in G_f$

Zeigen Sie, falls $d(s) \geq n$, existiert kein *augmentierender Pfad*.

b) In der Vorlesung wurde gezeigt, dass die Laufzeit von *Dinics Algorithmus* für Graphen mit Kantengewichten gleich 1 (*unit edgeweights*) in $O((n+m)\sqrt{m})$ liegt. Vergleichen Sie diese Laufzeit zum *Ford Fulkerson Algorithmus*. Für welche Graphen mit *unit edgeweights* ist welcher der beiden Algorithmen schneller?

c) Sei $G = (V, E)$ ein gerichteter Graph, in dem maximale Flüsse berechnet werden sollen. Sei $e = (i, j) \in E$ ebenso wie $e' = (j, i) \in E$, d. h. G besitzt ein Paar entgegengesetzter Kanten. Außerdem sei $c(e) \geq c(e')$. Widerlegen Sie durch ein Gegenbeispiel:

Entfernt man e' aus E und reduziert $c(e) := c(e) - c(e')$, ändert sich der maximale Fluss nicht, d. h. man kann entgegengesetzte Kanten a-priori (für beliebige s und t) gegeneinander aufrechnen.

Aufgabe 3 (*Rechnen: Segmentierung mit Flüßen (*)s*)

Wir betrachten einen einfachen Fall für Bildbearbeitung. Die Vorder-/Hintergrundsegmentierung. Das Ziel dieses Prozesses ist es, ein Bild in Vorder und Hintergrund zu zerlegen. Die Transformation dafür weißt jedem Pixel $p_{i,j}$ des Bildes einen Knoten $v_{i,j}$ im Graphen zu. Für jedes Paar von benachbarten Pixeln $p_{i,j}$ und $p_{k,l}$ ($|i-k| + |j-l| = 1$) fügen wir eine ungerichtete Kante $(v_{i,j}, v_{k,l})$ ein. Zusätzlich fügen wir je einen Knoten s für Vordergrund (Quelle) und einen Knoten t für Hintergrund (Senke)

ein. Von Knoten s existiert eine gerichtete Kante zu jedem Knoten $p_{i,j}$ und von jedem Knoten $p_{i,j}$ existiert eine gerichtete Kante zu Knoten p. Wir definieren darüber hinaus folgende Kantengewichte:

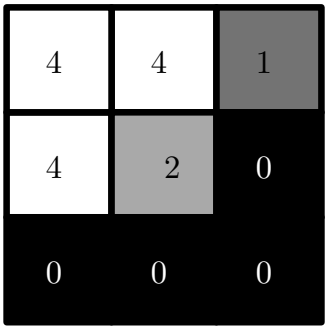
$$c(e=(u,v)) = \begin{cases} p_v(v) & u = s \\ p_h(u) & v = t \\ f(u,v) & sonst \end{cases}$$

Wobei mit $p_v(x)$ die Wahrscheinlichkeit gegeben ist, dass x Vordergrundknoten ist, mit $p_h(x)$ die Wahrscheinlichkeit für einen Hintergrundknoten und mit $f(x,y)$ eine Penaltyfunktion für das Trennen der beiden Knoten x und y. Für ein Graustufenbild B definieren wir
$p_v(x,y) = B[x,y]^2$, $p_h(x,y) = (4 - B[x,y])^2$ sowie $f((x_1,y_1),(x_2,y_2)) = (4 - |B[x_1,y_1] - B[x_2,y_2]|)^2$.
Hinweis: Diese Modellierung ist nur ein Beispiel und keine allgemeingültige Modellierung. Sie soll nur verdeutlichen wie Flow Alglorithmen für andere Probleme eingesetzt werden können.

a) Geben Sie den Flussgraphen für das unten angegebene Graustufenbild an.

b) Führen Sie einen augmenting Path Algorithmus auf dem entstandenen Graphen aus.

c) Wie würde die Segmentierung in Vorder- und Hintergrund im Bild als Ergebnis aussehen?



Aufgabe 4 (*Analyse: Königs Theorem*)

Das *Theorem von König* besagt, dass in jedem bipartiten Graphen $G = (V,\ E)$ die Größe eines Matchings größter Wertigkeit (*maximum-cardinality matching*) gleich der Größe einer minimalen Knotenüberdeckung (*minimal vertex cover*) ist.

Vertex Cover:
Ein *Vertex Cover* ist definiert als eine Teilmenge der Knoten $S \subseteq V$, so dass für alle Kanten $e = (u,v) \in E$ gilt $u \in S \vee v \in S$. Ein *minimales Vertex Cover* besitzt unter allen korrekten die kleinste Teilmenge an Knoten S.

Matching:
Ein *Matching* ist definiert als eine Teilmenge von Kanten $S \subseteq E$, so dass jeder Knoten $v \in V$ Endpunkt von höchstens einer Kante in S ist. Eine *maximales Matching* besitzt unter allen korrekten die größte Teilmenge an Kanten S.

Bipartiter Graph:
Ein bipartiter Graph enthält ausschließlich Kanten zwischen disjunkten Teilmengen der Knotenmenge:
$e \in E \leftrightarrow (u,v) \in S \times T,\ V = S \cup T,\ S \cap T = \varnothing$.

Beweisen Sie das *Theorem von König*.

Aufgabe 5 (*Analyse+Entwurf+Rechnen: Grenzüberwachung*)

Eine (eindimensionale) Grenzlinie soll durch ein Sensornetz überwacht werden. Zu diesem Zweck wurde eine große Anzahl an Sensorknoten unregelmäßig an der Grenze ausgebracht. Jeder Knoten kann einen Bereich der Grenze für eine gewisse Zeit proportional zu seiner Batteriekapazität überwachen. Die Grenze gilt als vollständig gesichert, wenn jeder Abschnitt der Grenzlinie von mindestens einem Sensorknoten abgedeckt ist. Aufgrund der großen Menge an Knoten sind ihre Überwachungsbereiche stark überlappend. Daher müssen nicht immer alle Knoten aktiv sein, um eine vollständige Sicherung der Grenze zu gewährleisten. So kann Energie gespart werden und die maximale Dauer der Grenzsicherung erhöht werden.
Durch die unregelmäßige Anbringung der Knoten und durch große Fertigungstoleranzen in der Batteriekapazität und dem Überwachungsbereich (*man hat unbedingt beim billigsten Hersteller einkaufen müssen...*) ist zunächst nicht klar, wie lange die Grenze maximal vollständig gesichert werden kann. Glücklicherweise wurden die Positionen der Knoten und ihre jeweiligen Kapazitäten und Detektionsbereiche protokolliert und können verwendet werden, um diese Frage zu beantworten.

a) In der Vorlesung haben Sie Flussprobleme mit beschränkten Kantenkapazitäten $c(e)$ kennengelernt. Ebenso können Flussprobleme mit beschränkten Knotenkapazitäten $c(v)$ sinnvoll sein. In diesem Fall darf für einen gültigen Fluss die Summe der in den Knoten ankommenden bzw. ausgehenden Flüsse die Kapazität des Knotens nicht überschreiten. Außerdem muss wie bisher für jeden Knoten (außer der Quelle und Senke) die Summe der ankommenden Flüsse gleich der Summe der ausgehenden Flüsse sein.

   Erklären Sie, wie maximale Flüsse mit Knotenkapazitäten berechnet werden können. Begründen Sie kurz, warum Ihr Ansatz einen zulässigen und optimalen Fluss berechnet.

b) Konstruieren Sie ein Flussnetzwerk, das das oben beschriebene Problem der Bestimmung einer maximalen Dauer für die vollständige Grenzüberwachung lösen kann.

   Hinweis: Jeder Knoten entspricht einem Sensorknoten. Batteriekapazität kann als äquivalent zur Flussmenge betrachtet werden.

c) Erstellen Sie ein Flussnetz, das dem folgenden Sensornetz entspricht. Wie lange kann dieses Netz die Grenze im Bereich $[0, 13]$ überwachen? Welche Sensorknoten müssen wann aktiv sein?

   Format der Angaben: $x_{nodeID} = \{[begin\_range, end\_range], capacity\}$

   $x_1 = \{[0, 5], 4\}$
   $x_2 = \{[0, 7], 3\}$
   $x_3 = \{[4, 9], 2\}$
   $x_4 = \{[3, 8], 5\}$
   $x_5 = \{[8, 13], 5\}$
   $x_6 = \{[7, 11], 3\}$
   $x_7 = \{[11, 15], 2\}$

   Hinweis: Bevor Sie langwierig einen maximalen Fluss berechnen, versuchen Sie ihn durch *scharfes Hinschauen* zu bestimmen.

Aufgabe 6 (*Analyse: Eigenschaften von Flüssen*)

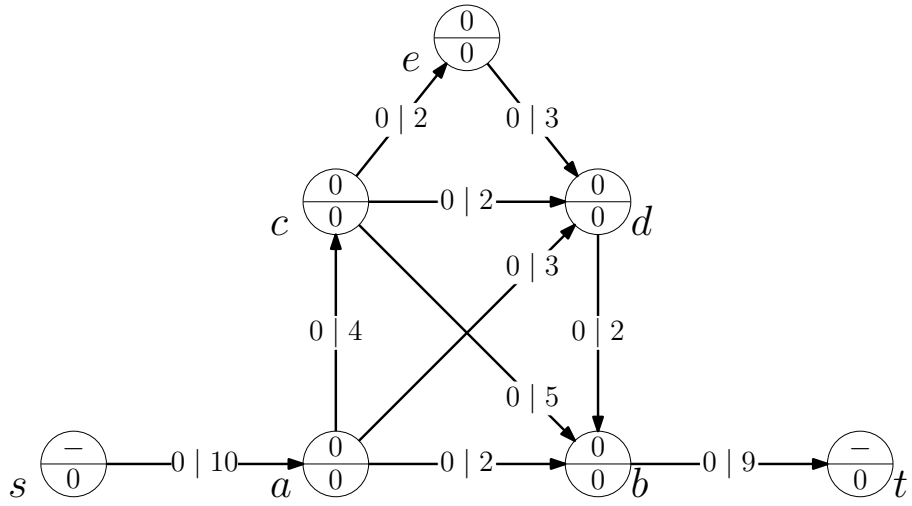
a) Seien (S, T) und (S', T') zwei minimale (s, t) Schnitte in einem Flußgraphen G. Zeigen oder widerlegen Sie, dass $(S \cup S', T \cap T')$ und $(S \cap S', T \cup T')$ auch minimale (s, t) Schnitte sind.

b) Sei (S, T) ein minimaler (s, t) Schnitt in einem Flussgraphen G. Zeigen oder widerlegen Sie, dass (S, T) ein minimaler (x, y) Schnitt ist f.a. $(x, y) \in S \times T$.

c) Zeigen Sie, dass für den *preflow-push Algorithmus* aus der Vorlesung mit beliebiger Wahl des nächsten Knotens $\mathcal{O}(n^2)$ die bestmögliche obere Schranke ist.

Aufgabe 7 (*Rechnen: preflow-push Algorithmus*)

Gegeben sei folgender Flussgraph:



(das Haus vom Nikolaus mit Vorgarten und Hinterhof)

Knotenbeschriftung: Level (unten), Überschuss (oben)
Kantenbeschriftung: Fluss (vorne), Kapazität (hinten)

Bestimmen Sie den maximalen Fluss von s nach t mit dem generischen *preflow-push* Algorithmus.

Aufgabe 8 (*Entwurf+Analyse: Qualitätskontrolle)*)

Sie sind beauftragt worden, einen Algorithmus für die Abteilung zur Qualitätskontrolle zu entwerfen, der die Validierung der wöchentlichen Resultate übernimmt.
Am Ende jeder Woche erhalten Sie dafür eine Liste L der in dieser Woche produzierten Bauteile mit den drei Angaben: (Typ; ID des Bauteils selbst; ID des Bauteils, in dem es verbaut wurde). Zudem erhalten Sie eine Liste D mit den IDs aller Bauelemente, die in derselben Woche von der Qualitätskontrolle als defekt markiert worden sind. Beide Listen sind potentiell zu groß für den Hauptspeicher und enthalten die Daten in unsortierter Reihenfolge.
Ihre Aufgabe ist es zu überprüfen, ob alle Bauteile, die selbst defekte Bauteile enthalten, als defekt markiert worden sind und falls nicht, dies zu korrigieren. Zu Ihrer Übersicht steht Ihnen ein Schema zur Verfügung, aus dem man ablesen kann, welche Bauteiltypen in welchen anderen verbaut werden. Dieses Schema passt in den Hauptspeicher.

a) Erweitern Sie Liste L um die Angabe aus Liste D, ob das jeweilige Bauteil defekt ist. Sie können davon ausgehen, dass diese Information keinen zusätzlichen Speicher benötigt.

b) Definieren Sie eine totale Ordnung $\prec_b$ auf den Bauteilen, die sich für jedes Bauteil aus lokalen Informationen und dem Bauteilschema berechnen lässt. Die Ordnung soll dabei erfüllen, dass in einer nach $\prec_b$ sortierten Liste L jedes Bauteil nach allen in ihm verbauten Bauteilen steht.

c) Verwenden Sie die sortiere und um Angaben zu Defekten erweiterte Liste L, um alle Bauteile zu identifizieren, die defekte Bauteile enthalten.

   Hinweis: Verwalten Sie die als defekt identifizierten aber noch zu betrachtenden Bauteile in einer geeigneten Datenstruktur.

Aufgabe 9 (*Analyse: Speicherbandbreite (*)*)

Die Effizienz eines Algorithmus, der auf externem Speicher arbeitet, hängt von der gewählten Blockgröße B in Zusammenspiel mit der maximalen Bandbreite W_{max} und der durchschnittlichen Zugriffszeit T_{seek} des externen Speichers ab.
Bestimmen Sie für die folgenden Fälle die Blockgröße, für die 90% der maximalen Bandbreite ausgereizt werden kann. Sie können davon ausgehen, dass ohne Unterbrechung auf ganze Blöcke in zufälliger Reihenfolge zugegriffen wird. Etwaige Berechnungen können als asynchron angenommen werden. Daher muss für diese keine Zeit berücksichtigt werden.

a) $W_{max} = 144$ MByte$/s$, $T_{seek} = 12ms$ (*Lesen von Festplatte*)

b) $W_{max} = 550$ MByte$/s$, $T_{seek} = 100\mu s$ (*Lesen von SSD*)

c) $W_{max} = 68$ MByte$/s$, $T_{seek} = 60s$ (*Lesen von LTO Streamer*)

Aufgabe 10 (*Analyse: Externer Stack*)

In der Vorlesung wurde eine Implementierung von *Stack* als externe Datenstruktur vorgestellt. Eine äquivalente Implementierung besitzt folgende Struktur: Im Speicher wird ein Puffer P der Größe $2B$ gehalten – B sei die Blockgröße beim Zugriff auf externen Speicher. Der Puffer ist in Form eines (internen) Stacks organisiert und enthält die neuesten gespeicherten Elemente. Folgende Operationen sind für die externe Datenstruktur definiert:

`pop` Falls P nicht leer, entferne das neueste Element aus P. Ansonsten, lese einen Block ein, um die Hälfte von P zu füllen bevor `pop` auf P ausgeführt wird.

`push` Falls P nicht voll, füge das neue Element direkt zu P. Ansonsten, schreibe die ältere Hälfte von P in den externen Speicher und verschiebe die aktuellere Hälfte an diese Stelle im Speicher. Anschließend führe ein `push` auf P aus.

Für die Analyse können Sie davon ausgehen, dass ein Block B Elemente des Stacks halten kann.

a) Zeigen Sie, dass die Operationen `push` und `pop` amortisiert $O(1/B)$ I/O Operationen benötigen.

b) Warum genügt es nicht, nur einen Puffer mit Größe B zu verwenden?

Aufgabe 11 (*Entwurf+Analyse: Telekommunikationsgesellschaft*)

Eine Telekommunikationsgesellschaft beauftragt Sie eine Anwendung zu schreiben, die monatlich die k Kunden bestimmt, bei denen sich die Rechnung im Vergleich zum Vormonat am meisten verändert hat. Diese Kunden will sich die Telekommunikationsgesellschaft noch einmal genau anschauen, um ihnen eventuell einen neuen Vertrag anzubieten.
Die zu bearbeitenden Daten werden Ihnen auf (langsamen) Bandspeichern zur Verfügung gestellt. Sie erhalten eine Liste mit den aneinandergefügten Datensätzen jeder Zweigstelle ihres Auftraggebers für den aktuellen Monat. Außerdem haben Sie eine entsprechende Liste für den Vormonat zur Verfügung. Gespeichert sind jeweils Tupel $(Kundennummer, Kosten)$.

a) Geben Sie einen Algorithmus an, der die geforderte Aufgabe erfüllt. Geben Sie außerdem die Laufzeit Ihres Algorithmus an und begründen Sie diese. Sie können davon ausgehen, dass die k zu bestimmenden Kunden in den Hauptspeicher passen.

b) Seien nun die k zu bestimmenden Kunden zu groß, um im Hauptspeicher gehalten zu werden. Ändern Sie Ihren Algorithmus so ab, dass er mit der erhöhten Datenmenge zurecht kommt. Geben Sie die Laufzeit Ihres neuen Algorithmus an und begründen Sie diese.

Hinweis: Diese Aufgabe war ursprünglich für die Klausur vorgesehen.

Aufgabe 12 (*Analyse:* preflow-push *Algorithmus (Wiederholung)*)

Sei durch S, T ein minimaler (s, t) Schnitt gegeben. Zeigen oder widerlegen Sie folgende Eigenschaften der Distanzfunktion:

a) $\forall v \in T : d(v) < n$

b) $\forall v \in S : d(v) \geq n$

Ausgabe: 30.01.2018
Abgabe: keine Abgabe, keine Korrektur