Karlsruher Institut für Technologie
Institut für Theoretische Informatik

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

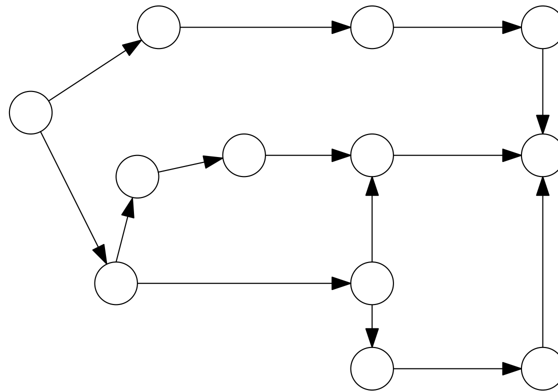
7. Übungsblatt zu Algorithmen II im WS 2016/2017

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

Musterlösungen

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.



Musterlösung:

Phase	G_f	L_f
1		
2		
3		

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:

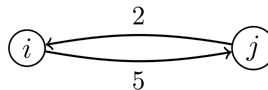
Entfernt man e' aus E und reduziert $c(e) := c(e) - c(e')$, ändert sich der maximale Fluss nicht, d. h. man kann entgegengesetzte Kanten a-priori (für beliebige s und t) gegeneinander aufrechnen.

Musterlösung:

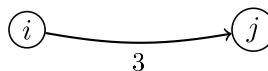
a) Ein Pfad in einem Graphen kann höchstens n unterschiedliche Knoten haben. Betrachte nun einen beliebigen augmentierenden Pfad in G_f. Für jede Kante auf diesem Weg wächst die Distanz für s höchstens um eins. Folglich kann ein augmentierender Pfad nur $d(s) \leq n - 1$ bedeuten.

b) Auf Graphen mit Kantengewichten gleich 1 ist die Laufzeit des *Ford Fulkerson Algorithmus* in $O(nm)$ (da $U = 1$!). *Dinics Algorithmus* ist damit schneller als der *Ford Fulkerson Algorithmus* falls $O((n + m)\sqrt{m}) < O(nm)$. Ausgerechnet ergibt sich $n > O(\frac{m}{\sqrt{m}-1}) = O(\sqrt{m})$.

c) Rechnet man die Kanten in folgendem Graph gegeneinander auf,



so ergibt sich



Für $s := j$ und $t := i$ ist kein Fluss mehr möglich, während im Originalgraphen ein maximaler Fluss von 2 möglich war. Dies ist somit ein Gegenbeispiel zur Behauptung.

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.

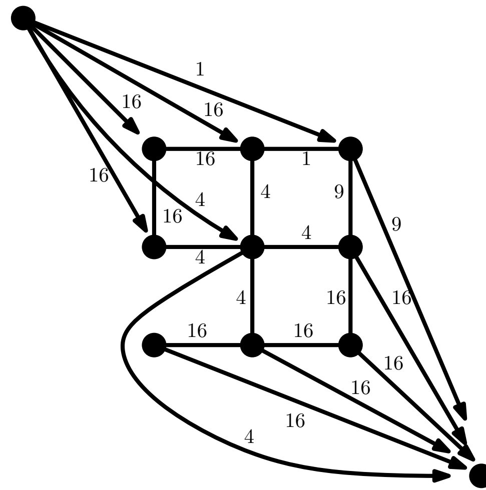
c) Wie würde die Segmentierung in Vorder- und Hintergrund im Bild als Ergebnis aussehen?

4	4	1
4	2	0
0	0	0

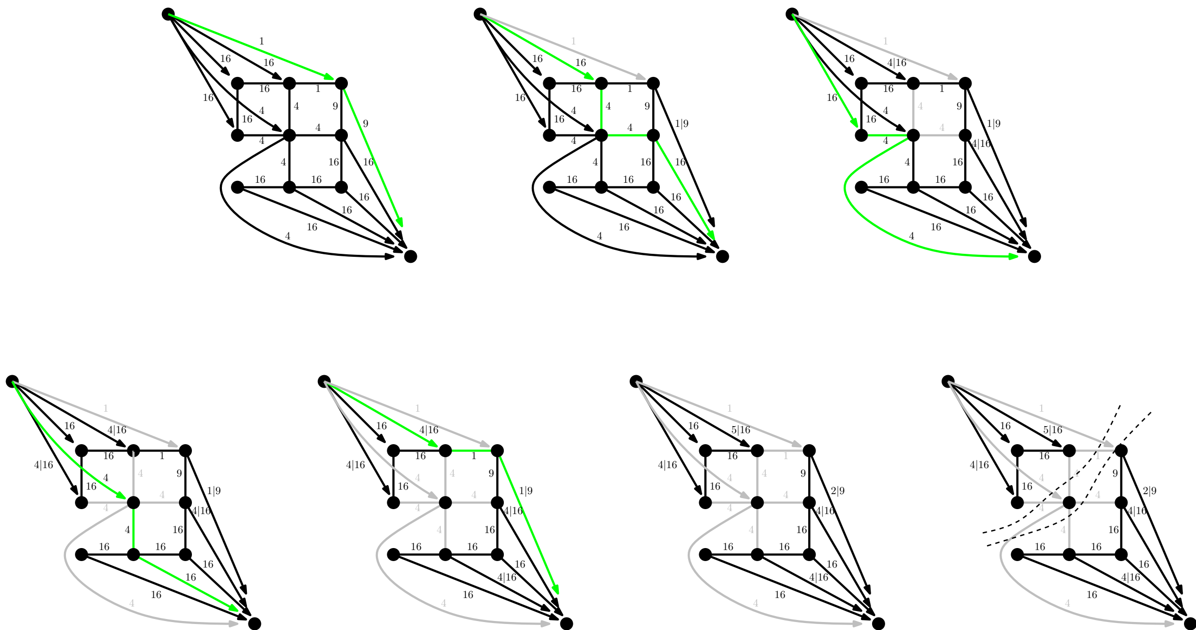
Musterlösung:

a) Als Transformation ergibt sich folgender Graph. Kanten ohne Kapazität wurden weggelassen.

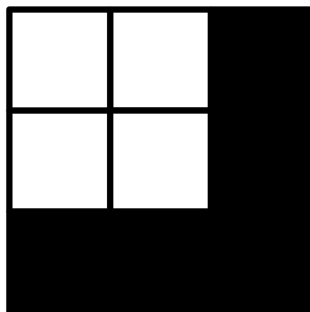
4	4	1
4	2	0
0	0	0

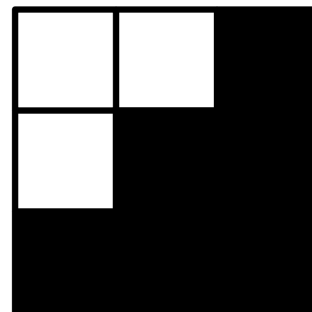


b) Der Algorithmus wird skizziert durch folgende Schritte:



c) Die beiden gleichwertigen Lösungen nach unserer Modellierung sind:





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.

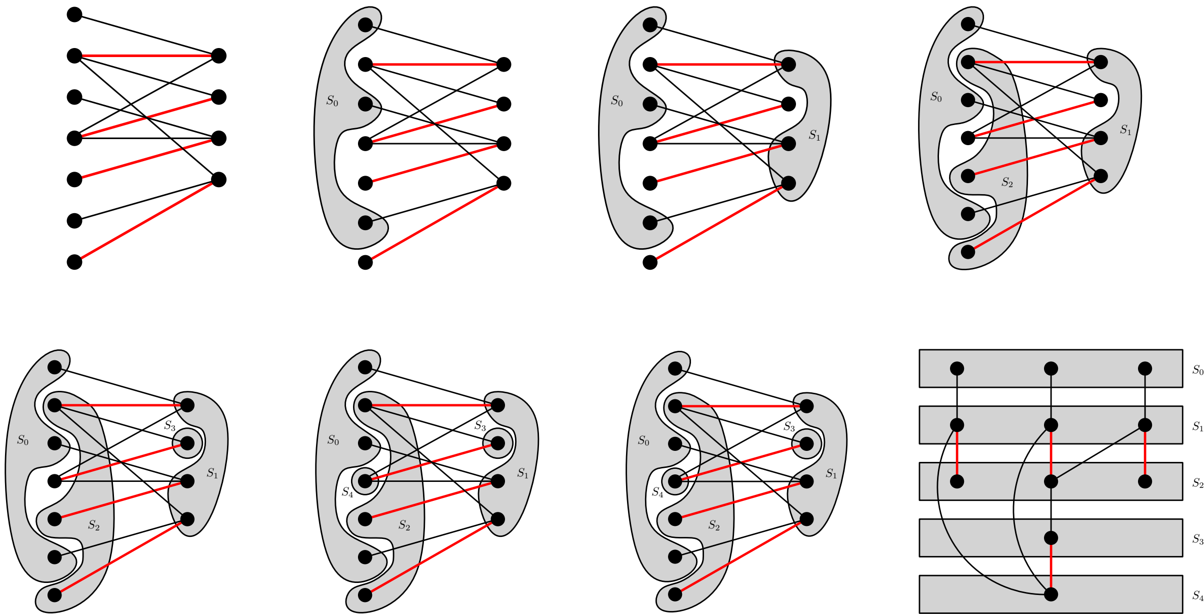
Bipartiter Graph:
Ein bipartiter Graph enthält ausschließlich Kanten zwischen disjunkten Teilmengen der Knotenmenge: $e \in E \leftrightarrow (u, v) \in S \times T,\ V = S \cup T,\ S \cap T = \varnothing$.

Beweisen Sie das *Theorem von König.*

Musterlösung:
Für einen bipartiten Graphen $G = (V, E)$ betrachten wir ein maximales Matching M der Größe $k = |M|$. Der Fall eines perfekten Matchings ist trivialerweise korrekt. Für den Fall eines nicht perfekten Matchings konstruieren wir einen Graphen mit mehreren Schichten S_i. Schicht S_0 definieren wir als alle Knoten, die zu keiner gematchten Kante inzident sind. Schicht S_i definieren wir über:

$$S_i = \begin{cases} v \in V : \exists e = (u, v) \in E \setminus M : u \in S_{i-1}, v \notin S_0, \dots S_{i-1} & \text{i ungerade} \\ v \in V : \exists e = (u, v) \in M : u \in S_{i-1}, v \notin S_0, \dots S_{i-1} & \text{i gerade} \end{cases}$$

Komponenten, die auf diese Weise nicht erreicht werden, formen in sich ein perfektes Matching und müssen für den Beweis nicht weiter betrachtet werden. Der Prozess ist hier bildlich dargestellt.



Da M ein maximales Matching ist, kann es in dem Graphen keinen alternierenden Weg geben, dessen Endpunkte ungematched sind. Daraus folgt, dass keine gematchte Kante innerhalb eines Layers S_{2i+1} existieren kann. Für eine Kante $(u, v) \in S_{2i+1}$ könnten wir sonst einen alternierenden Pfad $n_0 \in S_0, \dots, u, v, \dots n_{2(2i+1)+1} \in S_0$ finden. Dabei ist $n_0 \neq n_{2(2i+1)+1}$, da der Pfad ungerade Länge hat und das Matching wäre nicht maximal. Unter dem selben Argument kann keine ungematchte Kante zwischen zwei Knoten eines Layers S_{2i} existieren. Eine gematchte Kante innerhalb eines Layers S_{2i} kann aber ebenso nicht existieren, da jeder Knoten über eine eindeutige gematchte Kante zu einem Vorgängerlevel verbunden ist. Folglich hat jede gematchte Kante genau einen Endpunkt in einem Layer S_{2i+1}. Ebenso hat aber auch jede ungematchte Kante mindestens einen Endpunkt in einem Layer S_{2i+1}.
Folglich ist die Vereinigung über alle S_{2i+1} ein Vertex Cover da alle gematchten Kanten, aber auch ungemachten Kanten abgedeckt sind. Weiter besteht das Vertex Cover aus k Knoten, da jede der k gematchten Kanten genau einen Endpunkt im Vertex Cover besitzt und das Vertex Cover keine ungemachten Knoten beinhaltet, die zu keiner gematchten Kante inzident sind. Ein Vertex Cover besteht aber aus mindestens k Knoten. Andernfalls gäbe es eine Kante im maximalen Matching, die vom Vertex Cover nicht überdeckt wäre, weil ein Knoten im Vertex Cover nur höchstens eine Kante aus dem Matching überdecken kann. Folglich ist die Vereinigung über alle S_{2i+1} ein Vertex Cover, das Vertex Cover besteht aus k Knoten und ist minimal da jedes Vertex Cover mindestens k Knoten beinhalten.

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.

Musterlösung:

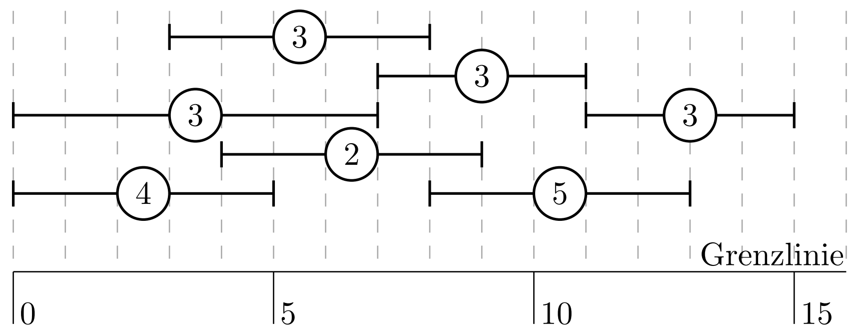
a) Man definiert einen neuen Flussgraph $G' = (V', E')$. Für jeden Knoten $v \in V$ fügt man zwei Knoten v_{in} und v_{out} sowie eine Kante (v_{in}, v_{out}) mit Kapazität $c(v_{in}, v_{out}) = c(v)$ in G' ein. Für jede Kante $(u, v) \in E$ fügt man eine neue Kante (u_{out}, v_{in}) in E' ein. Die Kapazität der Kante wird übernommen (bzw. auf ∞ gesetzt falls sie keine Kapazität hatte).

Nun berechnet man auf G' einen Fluss von s_{in} nach t_{out} und transferiert die Flusswerte zurück auf die Kanten in G. Der Fluss respektiert die Knotenkapazitäten, da sie in G' durch die Kanten (v_{in}, v_{out}) passend beschränkt wurden. Außerdem ist der Fluss optimal. Angenommen es gäbe noch einen augmentierenden Pfad in G, dann gäbe es auch einen in G' und der berechnete Fluss wäre kein maximaler Fluss in G': Die Restkapazitäten der ursprünglichen Kanten sind per Konstruktion gleich zu ihren Entsprechungen in G. Eine zu einem nicht voll ausgelasteten Knoten gehörende Kante hätte ebenfalls noch Restkapazität.

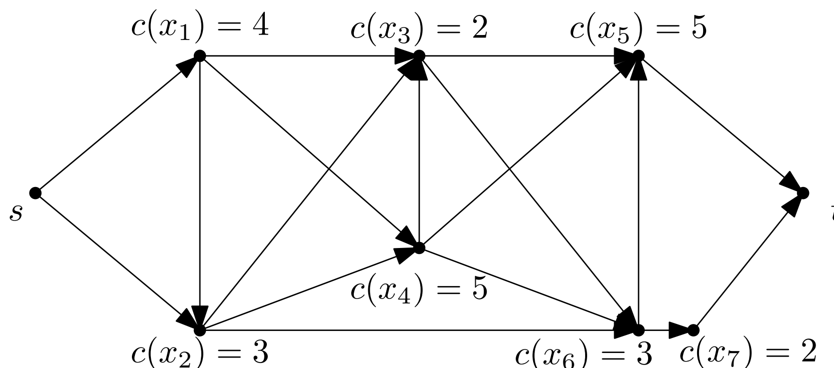
b) Das Flussnetz kann mit Hilfe von Knotenkapazitäten–wie in der letzten Teilaufgabe besprochen–konstruiert werden. Für jeden Sensorknoten x_i fügt man eine Knoten i mit Kapazität $c(i)$ gleich der Batteriekapazität des Sensorknotens ein. Außerdem fügt man eine Quelle s und eine Senke t ein. Anschließend fügt man Kanten (i, j) ein, wenn $x_i.end_range \in [x_j.start_range, x_j.end_range]$ (für s und t entsprechen die *range* Werte dem Anfang und dem Ende des Grenzverlaufs). Alle Kanten sind ohne Kapazität.

Jeder Flusspfad durch das Netz entspricht einer Konfiguration von aktiven Sensorknoten, die den gesamten Grenzverlauf überwachen können und die Flussmenge der Überwachungsdauer für diese Konfiguration. Der gesamte Fluss entspricht der maximalen Überwachungsdauer.

c) Eingezeichnete Überwachungsbereiche und Batteriekapazitäten der Sensorknoten:



Sich ergebendes Flussnetzwerk:



Musterlösung:

c) Durch geschicktes Hinschauen muss man den Flussalgorithmus nicht ausführen und kann direkt eine maximale Überwachungsdauer von 7 ablesen. Diese wird durch folgende Knotenmengen erreicht, die jeweils gleichzeitig für die angegebene Dauer aktiv sind:

aktive Knoten	Dauer
x_1, x_4, x_5	4
x_2, x_4, x_5	1
x_2, x_6, x_7	2

Jede aktive Knotenmenge deckt offensichtlich den kompletten Bereich $[0, 13]$ ab. Fast alle Knoten verbrauchen ihre komplette Energie –aber auch nicht mehr– außer Knoten x_3 (gar nicht verwendet) und x_6 (noch 1 Restkapazität). Mit den restlichen Knoten kann keine weitere vollständige Überdeckung erreicht werden.

Aufgabe 6 (*Analyse: Eigenschaften von Flüssen*)

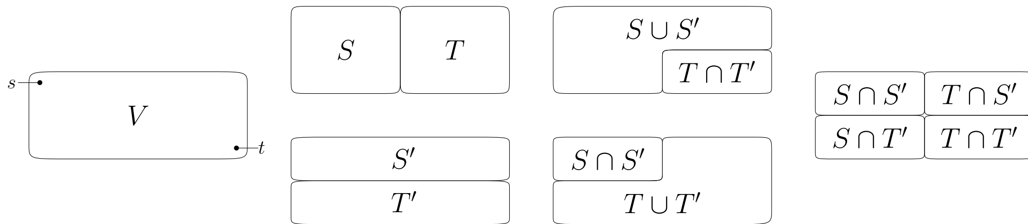
a) Seien (S, T) und (S', T') zwei minimale (s, t) Schnitte in einem Flußgraphen G. Zeigen oder widerlegen Sie, dass $(S \cup S', T \cap T')$ und $(S \cap S', T \cup T')$ auch minimale (s, t) Schnitte sind.

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. $(x, y) \in S \times T$.

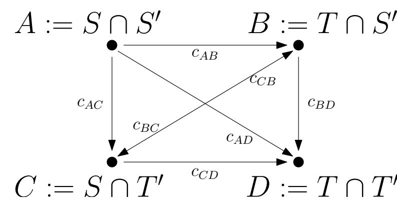
c) Zeigen Sie, dass für den *preflow-push Algorithmus* aus der Vorlesung mit beliebiger Wahl des nächsten Knotens $\mathcal{O}(n^2)$ die bestmögliche obere Schranke ist.

Musterlösung:

a) Der Graph wird in vier Bereiche aufgeteilt, $S \cap S'$, $T \cap S'$, $T \cap T'$ und $S \cap T'$ (siehe Abbildung).



Zur Anschauung definiert man einen Graphen, dessen Knoten der obigen Aufteilung entsprechen und dessen Kanten die Schnitte zwischen den Bereichen darstellen.



Damit ergeben sich folgende Werte für die Schnitte:

- (S,T) Schnitt: $c_{ST} = c_{AB} + c_{AD} + c_{CB} + c_{CD}$
- (S',T') Schnitt: $c_{S'T'} = c_{AC} + c_{AD} + c_{BC} + c_{BD}$
- $(S \cup S', T \cap T')$ Schnitt: $c_{S\cup S',T\cap T'} = c_{AD} + c_{BD} + c_{CD}$
- $(S \cap S', T \cup T')$ Schnitt: $c_{S\cap S',T\cup T'} = c_{AB} + c_{AC} + c_{AD}$

Da (S,T) und (S',T') minimale Schnitte sind, gilt $c_{ST} = c_{S'T'}$. Außerdem sind $c_{S\cup S',T\cap T'}$ und $c_{S\cap S',T\cup T'}$ jeweils größer gleich c_{ST} bzw. $c_{S'T'}$. Löst man die sich ergebenden Ungleichungen, erhält man $c_{BC} = c_{CB} = 0$ und damit $c_{AB} = c_{BD}$, $c_{AC} = c_{CD}$ (Rechnung siehe nächste Seite).

Es ergibt sich $c_{S\cup S',T\cap T'} = c_{ST} = c_{S\cup S',T\cap T'} = c_{S'T'}$.

Musterlösung:

a) (fortgesetzt)
Auflösen der Ungleichungen:

$$\begin{aligned} c_{S\cup S',T\cap T'} &\geq c_{ST} && (1)\\ c_{S\cup S',T\cap T'} &\geq c_{S'T'} && (2)\\ c_{S\cap S',T\cup T'} &\geq c_{ST} && (3)\\ c_{S\cap S',T\cup T'} &\geq c_{S'T'} && (4) \end{aligned}$$

$$\begin{aligned} c_{AD}+c_{BD}+c_{CD} &\geq c_{AB}+c_{AD}+c_{CB}+c_{CD} && (1)\\ c_{AD}+c_{BD}+c_{CD} &\geq c_{AC}+c_{AD}+c_{BC}+c_{BD} && (2)\\ c_{AB}+c_{AC}+c_{AD} &\geq c_{AB}+c_{AD}+c_{CB}+c_{CD} && (3)\\ c_{AB}+c_{AC}+c_{AD} &\geq c_{AC}+c_{AD}+c_{BC}+c_{BD} && (4) \end{aligned}$$

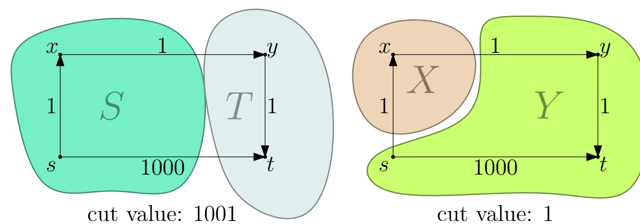
$$\begin{aligned} c_{BD} &\geq c_{AB}+c_{CB} && (1)\\ c_{CD} &\geq c_{AC}+c_{BC} && (2)\\ c_{AC} &\geq c_{CB}+c_{CD} && (3)\\ c_{AB} &\geq c_{BC}+c_{BD} && (4) \end{aligned}$$

Setze (2) in (3) ein und erhalte $c_{AC} \geq c_{CB}+c_{AC}+c_{BC} \Leftrightarrow 0 \geq c_{CB}+c_{BC}$. Da Kapazitäten nicht negativ sein können, gilt $c_{CB} = c_{BC} = 0$. Dies eingesetzt in die anderen Formeln liefert

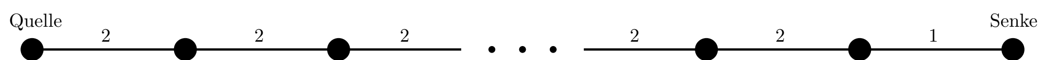
$$\begin{aligned} c_{BD} &\geq c_{AB} && (1)\\ c_{CD} &\geq c_{AC} && (2)\\ c_{AC} &\geq c_{CD} && (3)\\ c_{AB} &\geq c_{BD} && (4) \end{aligned}$$

und damit $c_{AB} = c_{BD}$, $c_{AC} = c_{CD}$.

b) Unten abgebildetes Flussnetzwerk mit dem eingezeichneten Schnitt ist ein Gegenbeispiel. Links ist ein minimaler (s,t) Schnitt mit Wert 1001 abgebildet. Der minimale (x,y) Schnitt mit dem Wert 1 ist rechts zu sehen.



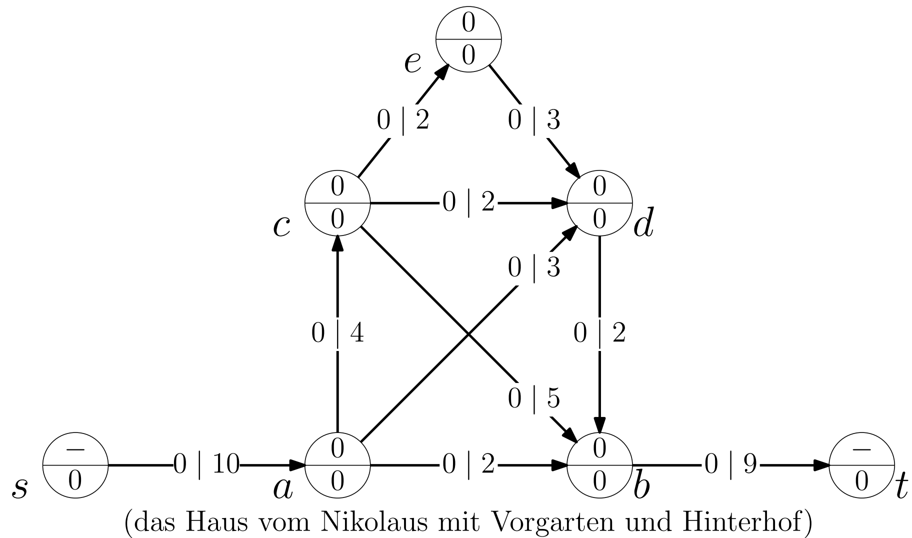
c) Wir betrachten folgenden Graphen:



Um einen Fluss auf diesem Graphen zu berechnen, muss der Fluss einmal durch den ganzen Graphen und wieder zurückfließen. Damit der zusätzliche Fluss vom vorletzten Knoten wieder in die Quelle zurückfließen kann, muss der Knoten mindestens Level $n+1$ haben. Somit ergibt sich über alle Knoten $\#relabel \geq \sum_{i=1}^{n-1} n+1 = n^2 - 1$.

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

Gegeben sei folgender Flussgraph:



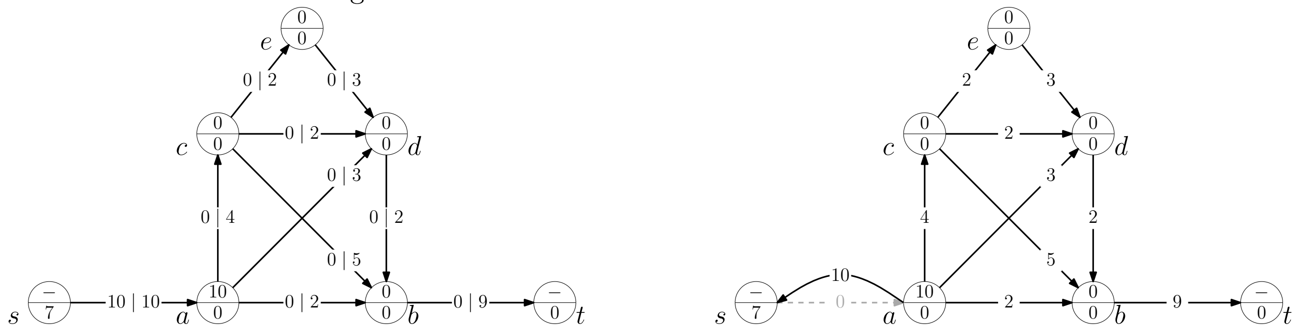
(das Haus vom Nikolaus mit Vorgarten und Hinterhof)

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

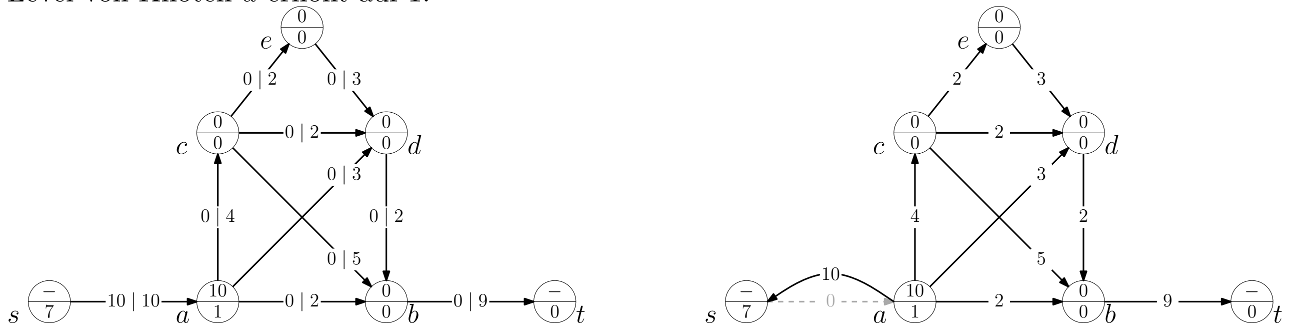
Bestimmen Sie den maximalen Fluss von s nach t mit dem generischen *preflow-push* Algorithmus.

Musterlösung:
Im Folgenden wird der *preflow-push* Algorithmus aus der Vorlesung auf den Flussgraphen angewendet. Aktive Knoten werden zufällig ausgewählt. Es wird bei einem Knoten geblieben bis dessen gesamter Überschuss weggeschoben wurde. Dieser Ablauf ist *nicht* der schnellstmögliche!
Links ist der Zustand des Flussgraphen nach jedem Schritt zu sehen, rechts der des Residualgraphen.

Zustand nach Initialisierung:

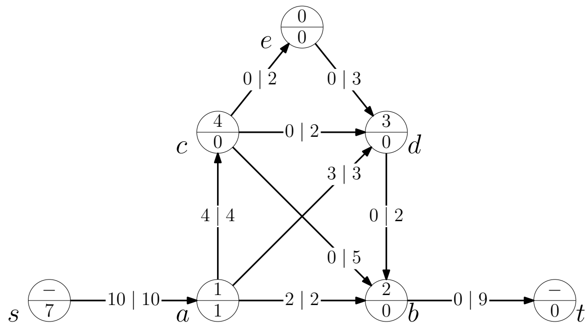


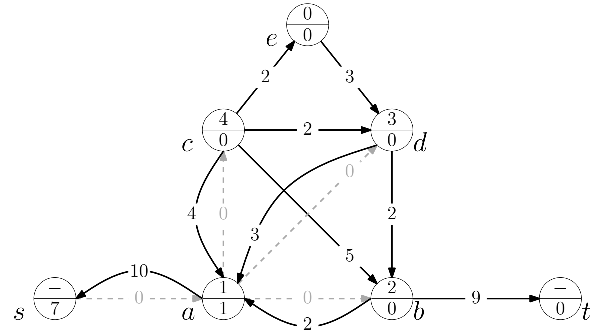
Level von Knoten a erhöht auf 1:



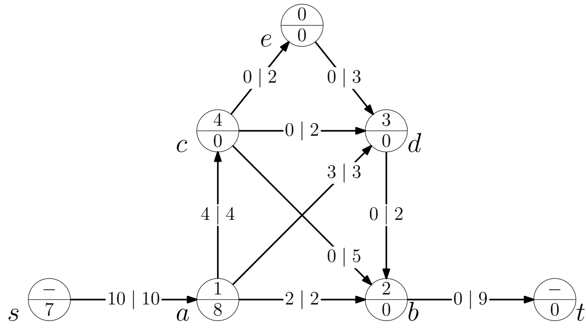
Musterlösung:

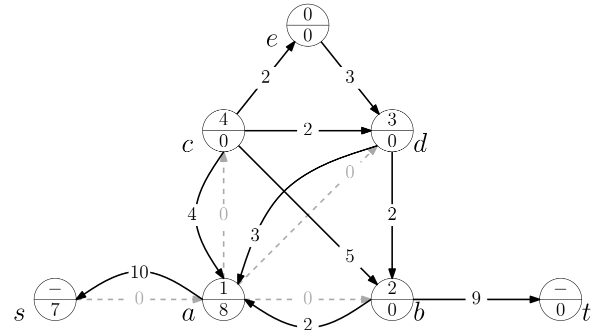
Fluss von a nach b, c und d geschoben:



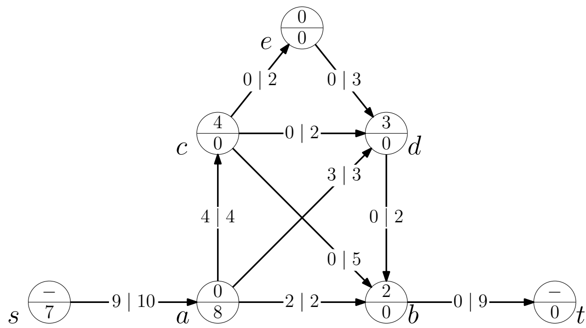


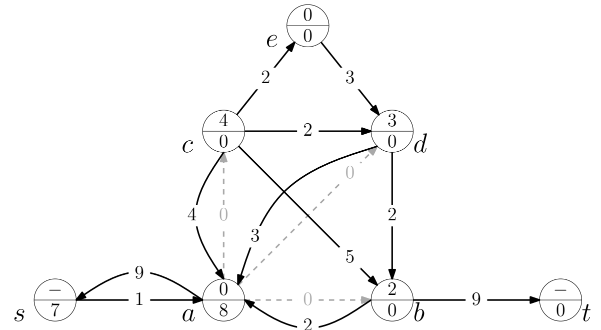
Level von Knoten a erhöht auf 8:



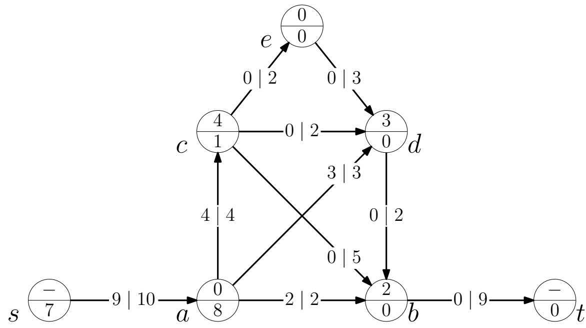


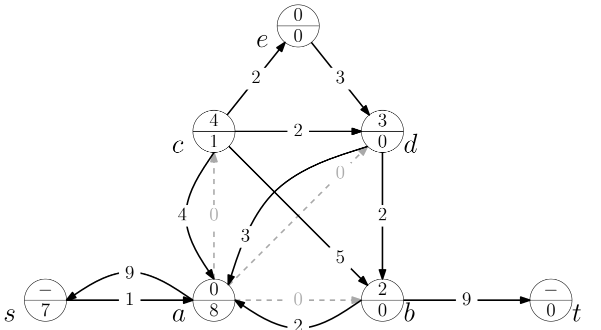
Fluss von a nach s geschoben:





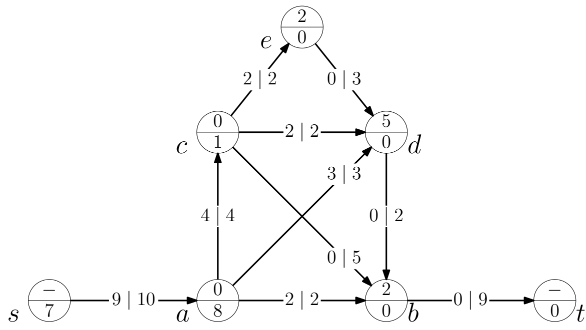
Level von Knoten c erhöht auf 1:

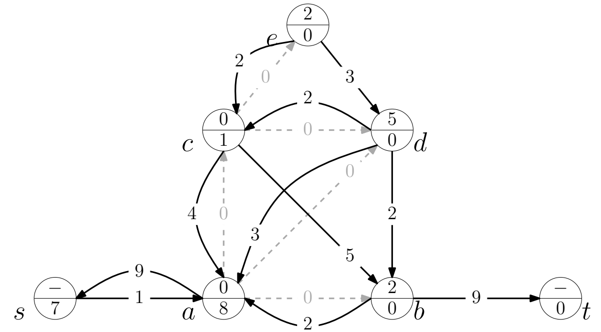




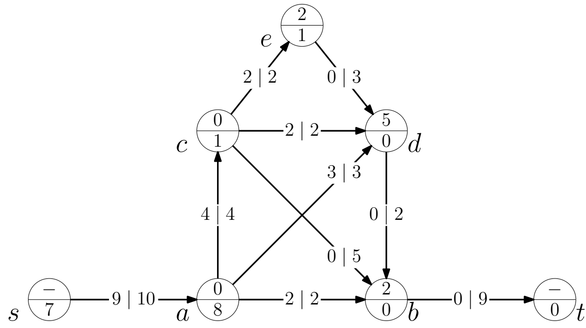
Musterlösung:

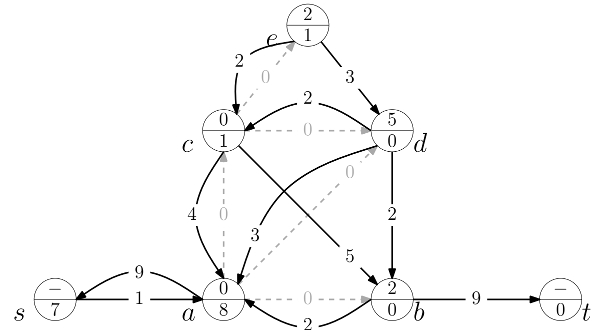
Fluss von c nach d und e geschoben:



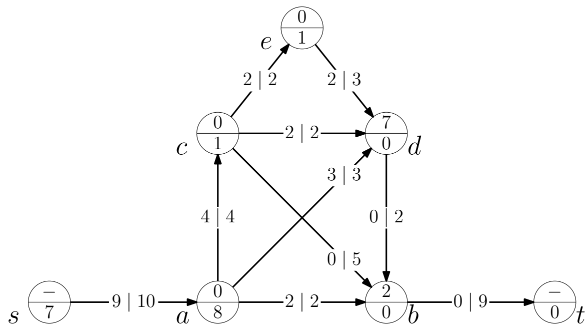


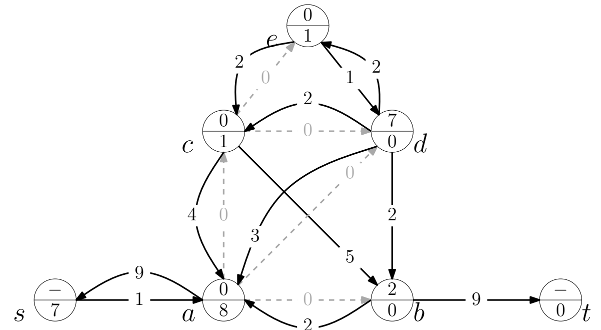
Level von Knoten e erhöht auf 1:



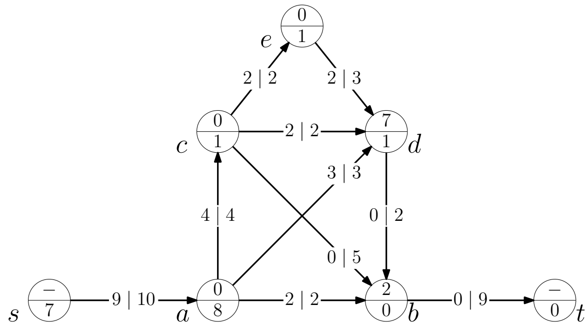


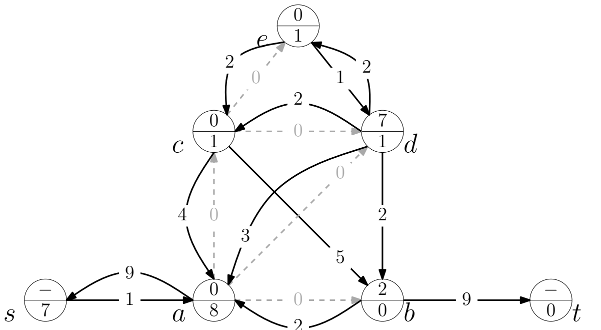
Fluss von e nach d geschoben:





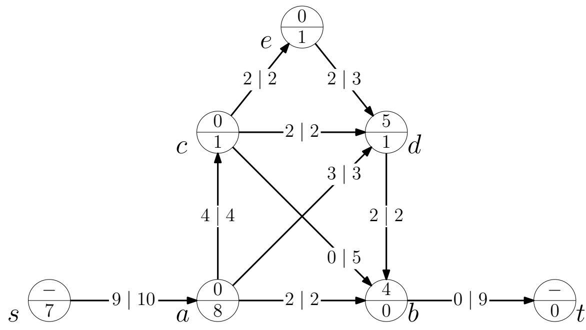
Level von Knoten d erhöht auf 1:

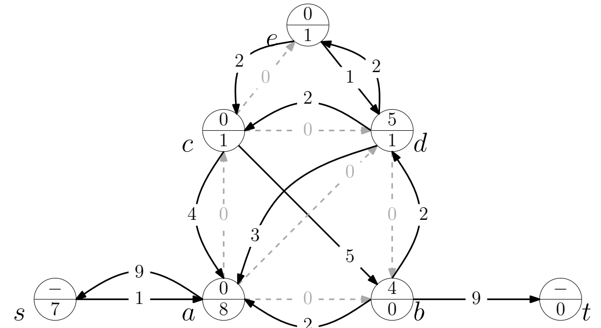




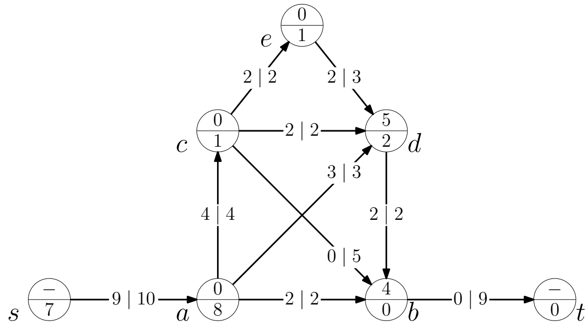
Musterlösung:

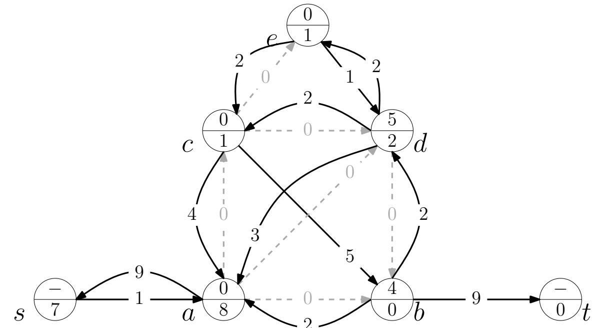
Fluss von d nach b geschoben:



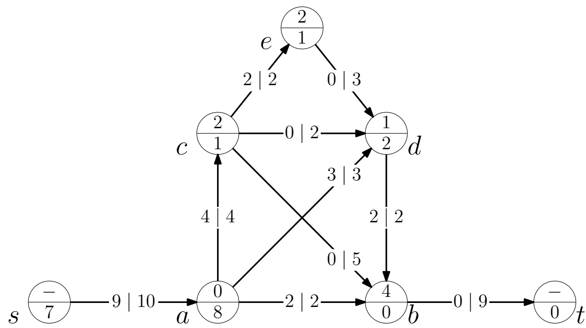


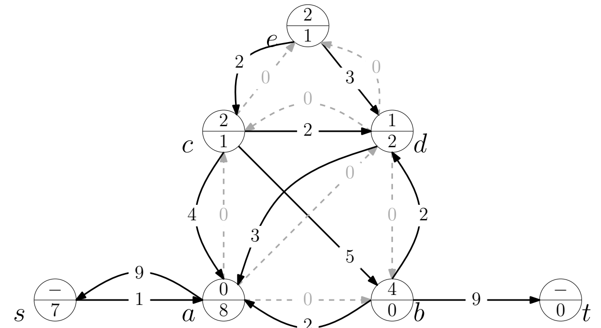
Level von Knoten d erhöht auf 2:



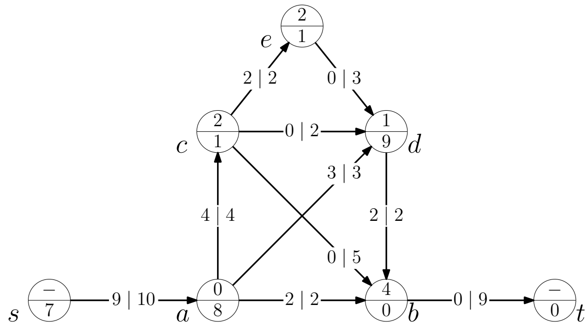


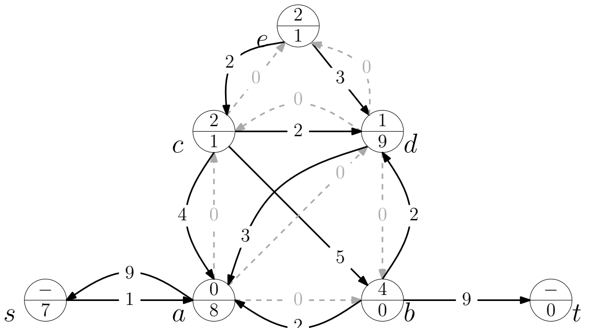
Fluss von d nach c und e geschoben:





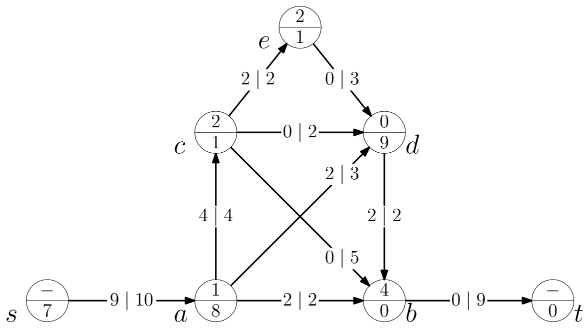
Level von Knoten d erhöht auf 9:

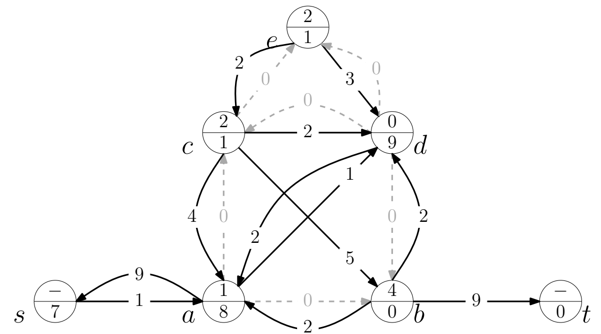




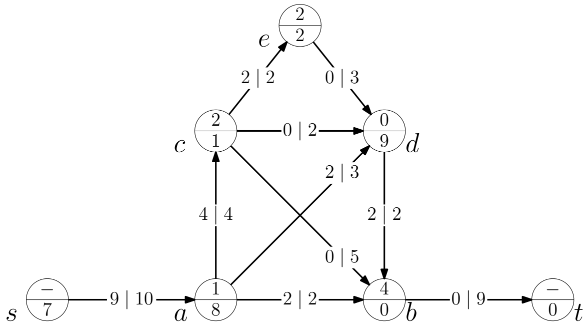
Musterlösung:

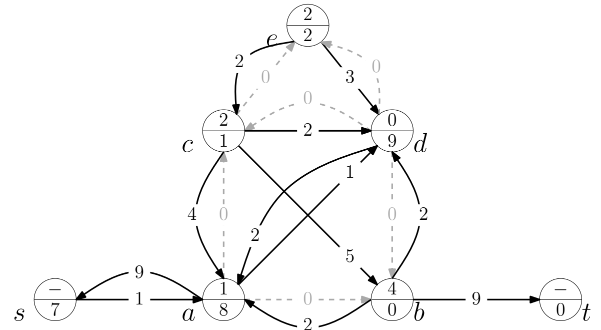
Fluss von d nach a geschoben:



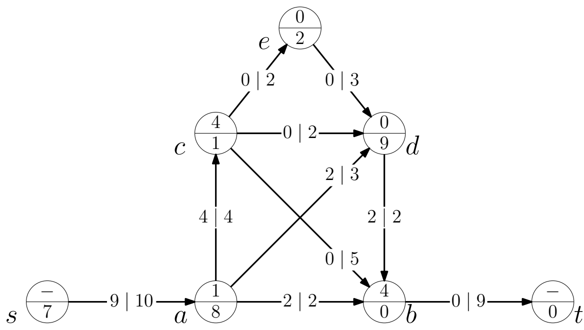


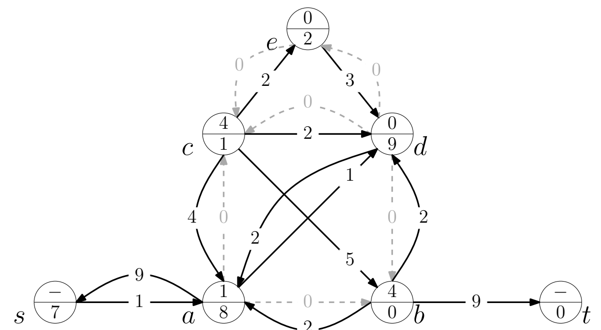
Level von Knoten e erhöht auf 2:



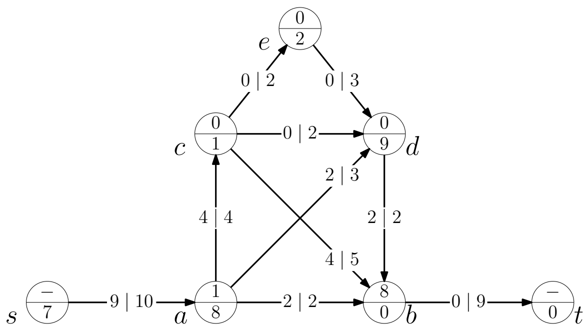


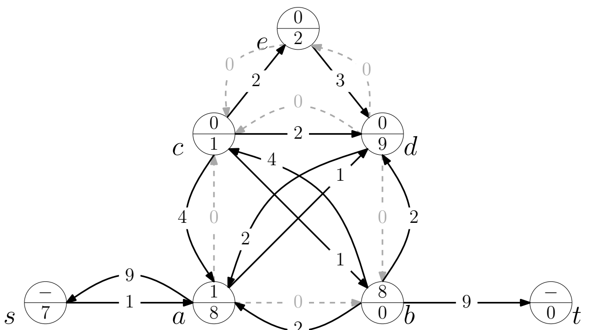
Fluss von e nach c geschoben:





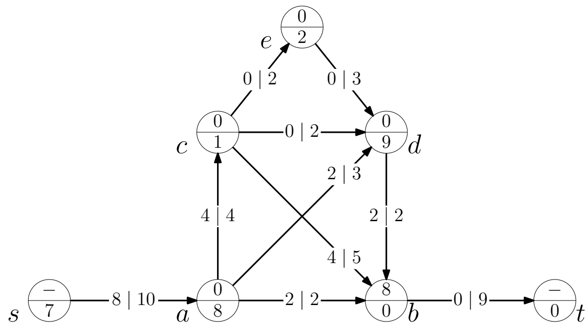
Fluss von c nach b geschoben:

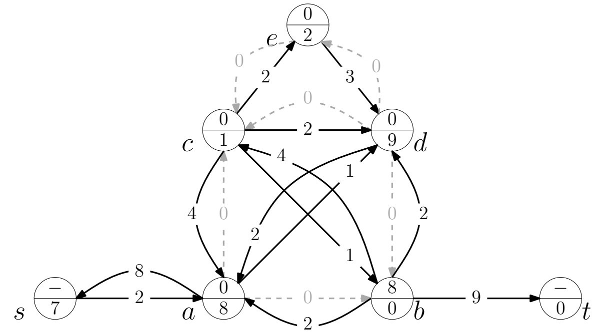




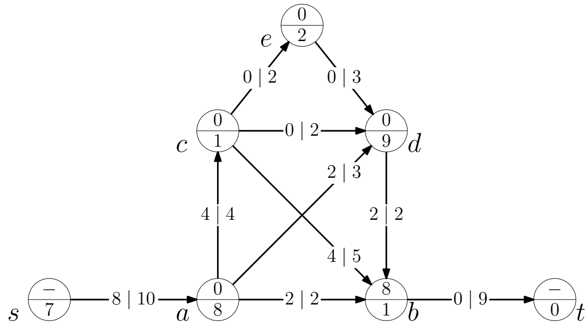
Musterlösung:

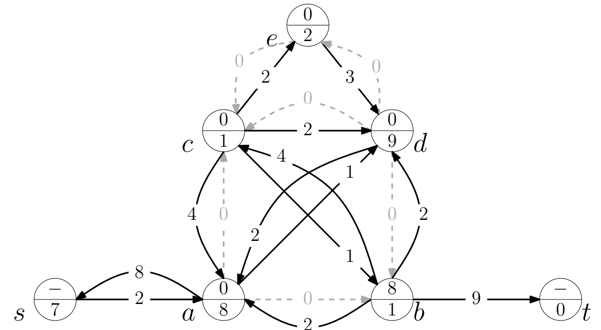
Fluss von a nach s geschoben:



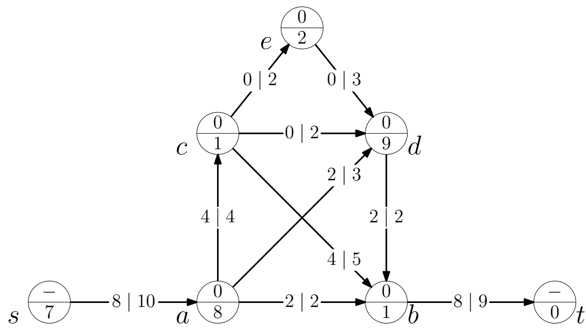


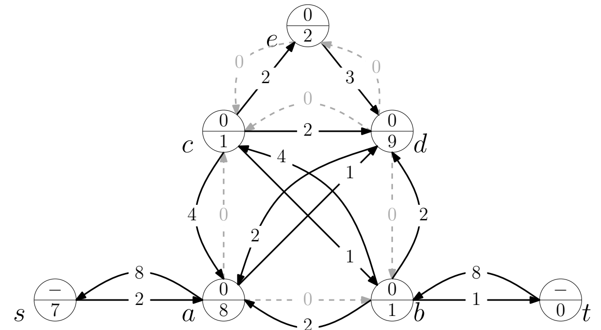
Level von Knoten b erhöht auf 1:





Fluss von b nach t geschoben:





Fertig!

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.

Musterlösung:

a) Sortiere L und D nach der ID der Bauteile und scanne beide Listen. Erhöhe dazu jeweils den Index der Liste, der auf das Bauteil mit der kleineren ID zeigt. Zeigen beide auf die gleiche ID, setze das Defekt-Flag und erhöhe beide Indizes.

b) Erstelle aus den Abhängigkeiten der Bauteiltypen einen DAG. Dies ist möglich, da kein Bauteil in sich selbst verbaut worden sein kann. Sortiere L nach der durch den DAG induzierten Teilordnung, verwende die ID der Bauteile als Sortierkriterium bei Unvergleichbarkeit.

c) Verwalte defekte Bauteile in einer *Ausschlußliste* in Form einer externen Prioritätswarteschlange (Schlüssel ist die ID des Bauteils, kleinere IDs haben höhere Priorität). Durchlaufe die sortierte Liste L. Falls beim Durchlaufen der Liste L ein defektes Bautteil t_0 gefunden wird und das defekte Bauteil in einem anderen Bauelement e_0 verbaut wurde, so füge das Bauelement e_0 in die Prioritätswarteschlange ein. Wird beim Durchlaufen der Liste L ein Bauteil t_1 gefunden, welches in der Prioritätswarteschlange die höchste Priorität hat, so wird das Bauteil t_1 ebenfalls als defekt markiert. Ist das Bauteil t_1 in einem anderen Bauelement e_1 verbaut, so füge das Bauelement e_1 in die Prioritätswarteschlange ein.

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*)

Musterlösung:
Ein Block kann in Zeit $T = T_{seek} + B/W_{max}$ eingelesen werden.
Mit der effektiven Bandbreite $W = B/T \stackrel{!}{=} 0.9 \cdot W_{max}$ ergibt sich

$$B = 9 \cdot W_{max} \cdot T_{seek}$$

für die gesuchte Blockgröße. Damit folgt:

a) $B = 15.552$ MByte ≈ 15 MByte

b) $B = 0.495$ MByte ≈ 500 kByte

c) $B = 36\,720$ MByte ≈ 37 GByte

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?

Musterlösung:

a) Betrachte die minimale Anzahl an Operationen (`push` oder `pop`) bis zur nächsten I/O Operation: Nach einer I/O Operation ist die Hälfte des Puffers leer – bei einem `push` auf den vollen Puffer wurde die Hälfte in den externen Speicher verlagert bzw. bei einem `pop` auf einen leeren Puffer wurde der halbe Puffer aufgefüllt. Von diesem Zustand ausgehend werden mindestens B Operationen einer Art ausgeführt, bevor erneut ein I/O Zugriff erfolgt – entweder, um bei einem `push` Daten in den externen Speicher zu schreiben, da der Puffer voll ist, oder um bei einem `pop` Daten aus dem externen Speicher zu laden, weil der Puffer leer ist.

b) Bei nur einem Puffer der Größe B könnte man sich folgende maximal schlechte Folge an Operationen überlegen: $B + 1$ `push` Operationen, gefolgt von einer Reihe von je 2 `pop` und 2 `push` Operationen. Jeweils die zweite dieser Anweisungen löst eine I/O Operation aus, da das `pop` auf einem leeren und das `push` auf einem vollen Puffer stattfindet. Amortisiert ergeben sich $O(1)$ I/O Operationen.

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.

Musterlösung:

a) In einem ersten Schritt werden beide Listen nach aufsteigender Kundennummer sortiert mit externem MergeSort. Anschließend scannt der Algorithmus linear über beide sortierte Listen M, N. Zu diesem Zweck wird für jede Liste ein Zeiger i_M bzw. i_N mit der aktuellen Position gespeichert und ein Teil der Liste mit Größe B im Speicher gehalten, der bei Bedarf durch den folgenden ersetzt wird. Beide Zeiger starten am Anfang der jeweiligen Liste. Stimmen die Kundennummern von $M[i_M]$ und $N[i_N]$ überein, wird die Differenz ihrer Kosten gebildet. Diese wird in einer lokalen Prioritätswarteschlange PQ gespeichert. Hat PQ nach dem Einfügen mehr als k Elemente, so wird das kleinste entfernt. Stimmen die Kundennummern M_{i_M} und N_{i_N} nicht überein, wird der Zeiger um eins erhöht, der auf den Eintrag mit der kleineren Kundennummer zeigt. Nachdem die kürzere von beiden Listen vollständig abgearbeitet wurde, enthält PQ die gesuchten Elemente.

   Die Laufzeit wird von den benötigten I/O Operationen dominiert. Sei $n := |N| + |M|$ und H die Größe des Hauptspeichers. Sortieren benötigt $\Theta(\frac{n}{B} log_{\frac{H}{B}} \frac{n}{H})$, der lineare Scan über beide Listen $O(n/B)$ I/O Operationen. Der gesamte Algorithmus ist also vom Sortieren dominiert.

   Für die Blockgröße B beim linearen Scan gilt: $B < (S - \text{maximaler Speicher für PQ})/2$, wobei S den verfügbaren Hauptspeicher angibt. Die Blockgröße beim MergeSort kann größer gewählt werden, da die PQ zu diesem Zeitpunkt noch leer ist. Dies ändert die Anzahl I/O Operationen aber nur um einen konstanten Faktor.

b) Verwende eine externe Prioritätswarteschlange statt einer internen. Diese benötigt bis zu $O(\frac{n}{B} \log_{\frac{H}{B}} \frac{n}{H})$ I/O Operationen, falls jeder Wert eingefügt werden muss (Werte löschen ist amortisiert kostenlos). Dies entspricht der Anzahl, die für das initiale Sortieren benötigt wird. Die Laufzeit ändert sich also nicht.

   Für die Blockgrößen gilt die gleiche Argumentation wie in der ersten Teilaufgabe.

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

Musterlösung:

a) Die Aussage ist wahr:
Angenommen es gibt einen Knoten $v \in T$ mit $d(v) \geq n$. Dann existiert im Residualgraph kein Weg von v zu t. Der entsprechende Fluss muss also zurück zu s geschoben werden. Daraus folgt aber auch direkt, dass der gegebene (s, t) Schnitt nicht minimal sein kann, da sonst der zusätzliche Fluss, der über den Schnitt geleitet wurde, zu t gelangen können muss (nach *max-flow-min-cut*-Theorem).

b) Die Aussage ist falsch:
Folgende Abbildung gibt ein Gegenbeispiel an.