

4. Übungsblatt zu Algorithmen II im WS 2016/2017

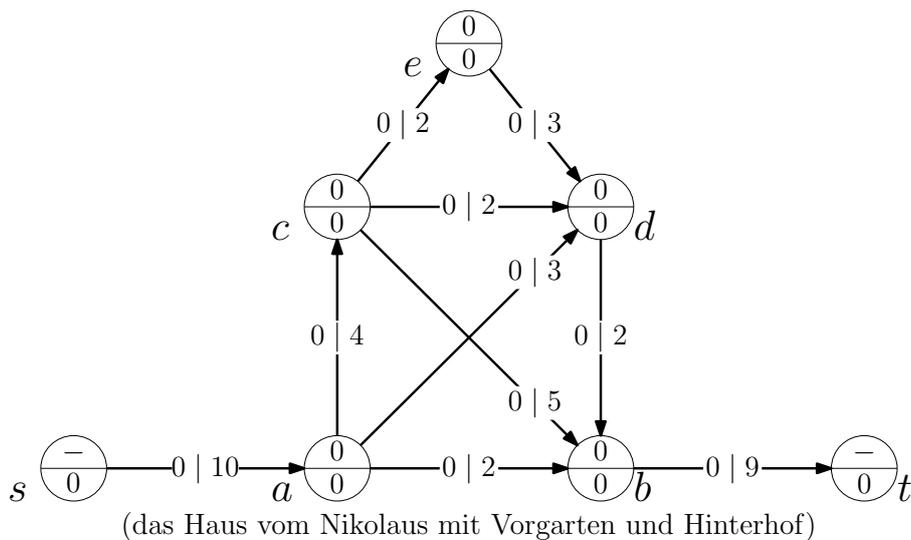
http://algo2.iti.kit.edu/AlgorithmenII_WS16.php
 {christian.schulz,michael.axtmann,sanders,simon.gog}@kit.edu

Aufgabe 1 (Analyse: Eigenschaften von Flüssen)

- 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.
- 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. $(x, y) \in S \times T$.
- 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 2 (Rechnen: preflow-push Algorithmus)

Gegeben sei folgender Flussgraph:



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 3 (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.

- 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.
- 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.
- Verwenden Sie die sortierte 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 4 (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.

- $W_{max} = 144 \text{ MByte/s}$, $T_{seek} = 12 \text{ ms}$ (Lesen von Festplatte)
- $W_{max} = 550 \text{ MByte/s}$, $T_{seek} = 100 \mu\text{s}$ (Lesen von SSD)
- $W_{max} = 68 \text{ MByte/s}$, $T_{seek} = 60 \text{ s}$ (Lesen von LTO Streamer)

Aufgabe 5 (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 6 (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 aneinandergfügten 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 7 (Implementierung: Preflow Push)

Implementieren Sie den Algorithmus “Preflow Push” in C++ mit dem KaHIP Framework. Laden Sie sich den KaHIP Fork¹ herunter und wechseln Sie in den Branch PreflowPush. Das KaHIP Framework stellt bereits einen Flussgraph (`lib/data.structure/flow_graph.h`) bereit. Diese Datenstruktur implementiert die Basisfunktionen von Residualgraphen. Die Funktionalität des Residualgraphen wurde bereits um folgende Funktionen erweitert:

```
// set flow of forward and reverse edge
void updateEdgeFlow(NodeID source, EdgeID e, FlowType flow);
FlowType getEdgeResCapacity(NodeID source, EdgeID e) const;
bool isSaturated(NodeID source, EdgeID e) const;
// increases edge flow of edge and decreases flow of reverse edge
void increaseEdgeFlow(NodeID source, EdgeID e, FlowType flow);
bool isValidEdge(NodeID v, EdgeID e) const;
EdgeID get_first_flow_edge(NodeID node);
EdgeID get_next_flow_edge(NodeID node, EdgeID e);
```

Implementieren Sie die fehlenden Methoden der Klasse `PreflowPush` in der Datei `app/preflow_push/preflow_push.cpp`. Diese Klasse initialisiert bereits alle Datenstrukturen für den Algorithmus “Preflow Push”. Weiter sind folgende Hilfsfunktionen ebenfalls schon implementiert:

```
EdgeID get_next_eligible_edge(const flow_graph& G, NodeID v);
void relabel(NodeID v);
// Pushes flow over edge e in the first phase.
// If the target node gets activated, we add the target node to
// the container small_active_nodes_ or the container active_nodes_
// depending on the level of the target node.
void low_level_push(flow_graph& G, NodeID v, EdgeID e, FlowType flow);
void push(flow_graph& G, NodeID v, EdgeID e, FlowType flow);
// Returns maximal residual flow based on remaining access.
FlowType max_allowed_res_flow(const flow_graph& G, NodeID v, EdgeID e) const;
FlowType get_outgoing_flow(const flow_graph& G, EdgeID e) const;
```

Implementieren Sie nun die Methode

```
// Executes the preflow push algorithm for nodes on arbitrary levels
// stored in the container active_nodes_.
FlowType remaining_level_preflow_push(flow_graph& G, NodeID source, NodeID sink);
```

Diese Methode soll in jeder Iteration einen aktiven Knoten v aus dem Container `active_nodes_` entnehmen und Überschuss an Fluss im Knoten v über “eligible” Kanten fließen lassen. Sobald es keine “eligible” Kanten mehr gibt, aber weiterhin Überschuss an Fluss im Knoten v existiert, wird der Level von Knoten v erhöht und wieder zu den aktiven Knoten hinzugefügt (Fifo Heuristik). Rufen Sie am Ende die Methode `get_next_flow_edge` auf, um den ausgehenden Fluss vom Startknoten zu berechnen (und somit den maximalen Fluss). Compilieren Sie Ihre Implementierung danach mithilfe des Skripts `compile.sh` und führen Sie den Algorithmus auf der Kommandozeile aus:

```
> ./optimized/preflow ./examples/flow_graphs/af_shell9.graph.hierarchy.5.h.inp
```

Aktivieren Sie nun die Heuristik “highest level first” indem sie die Codezeile

```
#define HIGHEST_LEVEL_FIRST
```

einkommentieren. Passen Sie Ihre Implementierung so an, dass der Algorithmus “Preflow Push” mit der Fifo Heuristik, aber auch mit der “highest level first” Heuristik maximale Flüsse berechnet. Führen Sie den Algorithmus mit beiden Heuristiken und verschiedenen Flussgraphen (`./examples/flow_graphs/`) aus. Wie unterscheidet sich die Laufzeit? Vergleichen Sie ebenfalls die Laufzeit vom Algorithmus “Preflow Push” mit der Laufzeit vom Algorithmus von Ford Fulkerson (`> ./optimized/maxflow`).

¹<https://github.com/MichaelAxtmann/KaHIP>

Aktivieren Sie nun zusätzlich die Heuristik “aggressive local relabeling” indem Sie die Codezeile

```
#define AGGRESSIVE_RELABELING
```

einkommentieren. Wie ändert sich die Laufzeit vom Algorithmus “Preflow Push”?

Um zusätzlich die Heuristik “two phase preflow push” aktivieren zu können (`TWO_PHASE_APPROACH`), müssen Sie die Methode

```
void low_level_preflow_push(flow_graph& G, NodeID source, NodeID sink);
```

implementieren. Diese Methode erfüllt die gleiche Funktionalität wie die Methode

```
FlowType remaining_level_preflow_push(flow_graph& G, NodeID source, NodeID sink);
```

prozessiert jedoch ausschließlich Knoten aus dem Container `small_active_nodes_`. Knoten aus diesem Container liegen auf Level $0 \dots n - 1$. Überschreiten Knoten Level $n - 1$, so sollen diese im Container `active_nodes_`, zur späteren Bearbeitung durch die Methode `remaining_level_preflow_push`, abgelegt werden. Führen Sie den Algorithmus nun gleichzeitig mit allen drei Heuristiken “aggressive local relabeling”, “two phase preflow push” und “highest level first” aus. Vergleichen Sie die Laufzeiten von dieser Variante mit den bisherigen Laufzeiten. Welche Heuristik/Heuristiken bewähren sich?

Bis zur Veröffentlichung der Musterlösung können Studenten Implementierungen mit Heuristiken “highest level first” und/oder “two phase preflow push” per E-Mail bei Übungsleiter einreichen. Die ersten zehn Studenten, dessen Implementierungen ähnlich schnell oder schneller als die Musterlösung (gleiche Heuristik) sind, werden belohnt :) Sind die Implementierung höchstens 20% langsamer als die Musterlösung, so bekommt der Student eine Tafel Ritter Sport! Ist die Implementierung schneller als die Musterlösung, so bekommt der Student eine weitere Tafel Ritter Sport! Verglichen wird die Laufzeit der Instanz `af_shell19.graph.hierarchy.5.h.inp`. Als Referenzrechner dient ein Laptop mit einem i5-5200U Prozessor (4 Kerne mit jeweils 2.20GHz). Tabelle 1 zeigt die Referenzzeiten.

Tabelle 1: Referenzzeiten.

Heuristics	Time [ms]
AR+HLF+2FA	12771
AR+2FA	15833
AR+HLF	13364

Ausgabe: 06.12.2016

Abgabe: keine Abgabe, keine Korrektur