Technische Universität Berlin
Institut für Mathematik

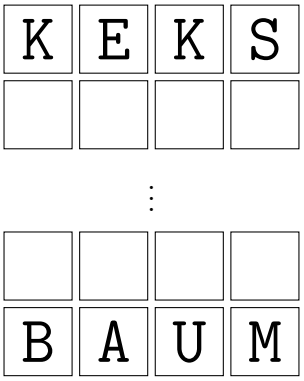
ADM I - Einführung in die Lineare und Kombinatorische Optimierung
Wintersemester 14/15

Prof. Dr. Dr. h.c. mult. Martin Grötschel
Dr. Axel Werner
Torsten Klug
Benedikt Bodendorf

Weihnachtsaufgabe (24_8 Bonuspunkte)

Abgabetermin: 23.1.2015 bis 14:15 in MA041

Versucht zunächst folgendes Wortleiter-Rätsel zu lösen: in die Zeilen sind (sinnvolle) deutsche Substantive einzutragen, so dass sich jedes Wort von dem darüber bzw. darunter in genau einem Buchstaben unterscheidet. (Die Anzahl der Zeilen ist nicht beschränkt.)



Wir betrachten den ungerichteten Graphen $D_4 = (V, E)$, der sich folgendermaßen ergibt: Jeder Knoten entspricht einem deutschsprachigen Substantiv mit vier Buchstaben. Zwei Knoten sind durch eine Kante verbunden, wenn sie sich in genau einem Buchstaben unterscheiden. Die Aufgabe besteht darin, diesen Graphen zu analysieren.

Auf der Webseite der Vorlesung findet sich unter

`http://www.zib.de/groetschel/teaching/WS1415/dlexdb_results_4.txt`

eine Liste aller deutschsprachigen Substantive mit vier Buchstaben. (Jede Zeile enthält ein Substantiv, den dazugehörigen Wortstamm und eine Häufigkeitsangabe – interessant ist für die Aufgabe nur das erste Wort jeder Zeile.) Schreibt ein Programm, das diese Datei einliest und Algorithmen aus der Vorlesung benutzt, um für den Graphen D_4 die folgenden Fragen zu beantworten:

a) Ist D_4 zusammenhängend? Falls nicht, wie viele Zusammenhangskomponenten gibt es und welche Kardinalität haben jeweils die größten fünf (falls so viele existieren)?

b) Wie groß ist der größte Grad eines Knotens? Gebt bitte die Wörter mit der größten Nachbarschaft an! Gibt es isolierte Knoten? Falls ja, gebt Beispiele an!

c) Findet für jede Zusammenhangskomponente (V', E') von D_4 mit mindestens 20 Knoten folgendes heraus: Wieviele Kanten hat der längstmögliche kürzeste Weg zwischen zwei Knoten von V'? D.h.: wie groß ist

$$L' := \max\big\{\min\{|P| \mid P \text{ } [v,w]\text{-Weg in } D_4\} \;\big|\; v, w \in V',\, v \neq w\big\}?$$

Gebt mindestens einen Weg der Länge L' an!

d) Findet eine möglichst kurze Lösung für das Beispiel am Anfang! Falls Euch diese Lösung nicht überzeugt, versucht, eine alternative Lösung zu finden, die nur „gängige“ Substantive enthält.

Zusatzaufgabe:

Es mag Euch erstaunen, aber soweit wir wissen sind derartige Untersuchungen von Graphen, die man aus Wörtern ableiten kann (z. B. auch durch andere sinnvolle Definitionen der Nachbarschaft von Knoten) noch nie in der Computerlinguistik durchgeführt worden. Kann man daraus Erkenntnisse für die Sprachforschung ziehen? Findet Ihr die Ergebnisse irgendwie interessant? Was könnte man den Linguisten sonst noch vorschlagen? Welche Graphenparameter sind vielleicht noch von Bedeutung?

Einige der in der Liste vorkommenden Wörter kennt Ihr vielleicht überhaupt nicht. Überprüft, ob es sich um Datenfehler handelt oder einfach nur um ganz seltene oder veraltete Wörter.

Wie steht es mit dem Vergleich zu Wörtern in anderen Sprachen? Im Internet kann man eine Liste der 5757 englischen Wörter mit 5 Buchstaben finden, siehe

`http://www-cs-faculty.stanford.edu/~uno/sgb.html`

`http://www-cs-faculty.stanford.edu/~uno/sgb-words.txt`

Versucht Eure Verfahren für den Graphen D_4 auch auf den analog definierten Graphen E_5 anzuwenden. Sind die Ergebnisse ähnlich?

Fragen: klug@zib.de