# Big Data Übungsblatt 09

Anton Bulat, Josephine Geiger, Julia Siekiera

December 14, 2017

## Aufgabe 1: Unterschranken an die Replikationsraten

Gegeben: Map-Reduce-Programm, das alle Pfade der Länge 2 in einem gerichteten Graphen mit $n$ Knoten findet.
Eingabe: Liste aller Kanten. Das Tupel $(u, v)$ beschreibt eine Kante von Knoten $u$ nach Knoten $v$.
Gesucht: geeignete Unterschranke an die Replikationsrate $r$ mithilfe des Beweisschemas aus der Vorlesung.

- Schritt 1:
  Ein Reducer kann mit $q$ Eingaben maximal $\binom{q}{2} \approx \frac{q^2}{2}$ Ausgabewerte überdecken, also Pfade der Länge 2. Das heißt
  $$g(q) = \frac{q^2}{2}.$$
- Schritt 2:
  Als Ausgabe werden Tripel von Knoten erwartet, die jeweils einen Pfad der Länge 2 bilden, also zwei Kanten verbinden. In einem gerichteten Graphen mit $n$ Knoten kann es maximal $2 \times \binom{n}{2} = 2 \times \frac{n!}{2!(n-2)!} = n \times (n-1)$ Kanten geben.
  Also ist die Gesamtzahl der vom Problem generierten Ausgaben
  $$m = \binom{n \times (n-1)}{2} \approx \frac{(n^2 - n)^2}{2}.$$
- Schritt 3:
  Es gilt $\sum_{i=1}^{k} g(q_i) \geq m$ mit $g(q_i) = \frac{q_i^2}{2}$ und $m \approx \frac{(n^2-n)^2}{2}$, also
  $\Rightarrow \sum_{i=1}^{k} \frac{q_i^2}{2} \geq \frac{(n^2-n)^2}{2}$
  $\Leftrightarrow \sum_{i=1}^{k} q_i^2 \geq (n^2 - n)^2$.

- Schritt 4:
  Da $q \geq q_i$ bleibt die Ungleichung erfüllt:

$$\sum_{i=1}^{k} q q_i \geq (n^2 - n)^2 \Leftrightarrow q \sum_{i=1}^{k} q_i \geq (n^2 - n)^2 \Leftrightarrow \sum_{i=1}^{k} q_i \geq \frac{(n^2 - n)^2}{q}.$$

- Schritt 5:
  Die Anzahl der Eingaben entspricht der Anzahl an Kanten, also $n \times (n-1)$. Nach Division durch diesen Wert ergibt sich die untere Schranke für $r$:

$$\frac{1}{n \times (n-1)} \sum_{i=1}^{k} q_i = r \geq \frac{n^2 - n}{q}.$$