Aufgabe 3

a)

Anzahl aller möglichen Kombinationen des Tupels $(A, B) \ = \ a * b$

R(A,B) besitz n Einträge

S(A,B) besitz m Einträge

Tupel können einmal in einer Liste vorkommen

- Reducer berechnet Schnittmenge von k Tupeln aus R und S
- Ausgabe wenn Tupel in beiden Mengen vorhanden ist
- Größte Anzahl überdeckt, wenn alle Kombinationen vorkommen

$q = 2\ k$, wobei k die Menge der untersuchten Tupel pro Reducer ist. Im worst case entspricht $k = a * b$

$g(q) = q$, da pro Tupel, welches zweimal vorkommt nur eines zur Ausgabe gegeben wird, welches aus zwei Werten besteht

Dabei ist gilt $g(q) \leq a * b$

b)

In dem Mapper Schritt wird je nach Eingabe n Zeilen der Matrix A mit $(a_{1,i}, \dots, a_{m,i})$, der komplette Vektor $\overrightarrow{x}$mit $(x_1, \dots, x_n)$ oder n Werte des Vektors $\overrightarrow{b}$ mit (b_i) ausgegeben $(i = 1, \dots, n)$. Der Reducer bildet zuerst das Skalarprodukt zwischen Vektor x und der Zeile aus der Matrix A. Danach wird der Wert mit dem des Vektors $\overrightarrow{b}$ aufaddiert.

Daher ist die die maximale Anzahl an Eingabewerten für den Reducer $q = k(n + m + 1)$. Dabei ist k die Anzahl an Multiplikationen und Additionen im Reducer. Das Ergebnis ist ein n oder m dimensionaler Vektor, je nachdem ob n oder m größer ist. Pro Reducer wir dafür die Anzahl an berechneten Werten ausgegeben: $g(q) = \frac{q}{(\mathrm{n+m+1})}$

c) Der Mapper übergibt die Kanten an den Reducer. Dieser sucht nach Dreiecken und gibt die Knoten der Dreiecke aus. Die maximale Anzahl an Kanten entspricht $v = \frac{n(n-1)}{2}$ bei n Knoten. Die maximale Anzahl an Dreiecken d kann durch $d \ \leq \frac{\sqrt{2}}{3} * v^{\frac{3}{2}}$ approximiert werden.

1. Pro Kante werden2 Knoten übergeben, die durch die Kante verbunden ist. Der Mapper übergibt alle Kanten der Eingabe an den Reducer.
$q = 2v$

$$g(q) \leq \frac{\sqrt{2}}{3} * q^{\frac{3}{2}} * 3$$

2. Die Größe der Ausgabe entspricht der maximalen Anzahl an Dreiecken mal die Anzahl an Knoten, die das Dreieck beschreiben $m \leq \frac{\sqrt{2}}{3} * v^{\frac{3}{2}} * 3$

3. $\sum_i g(q_i) \geq m \rightarrow \sum_i \frac{\sqrt{2}}{3} * q_i^{\frac{3}{2}} * 3 \geq \frac{\sqrt{2}}{3} * v^{\frac{3}{2}} * 3 \rightarrow \sum_i q_i^{\frac{3}{2}} \geq v^{\frac{3}{2}} \rightarrow \sum_i q_i * q_i^{\frac{1}{2}} \geq v^{\frac{3}{2}}$

4. $q \sum_i q_i^{\frac{1}{2}} \geq v^{\frac{3}{2}} \rightarrow \sum_i q_i \geq \left(\frac{v^{\frac{3}{2}}}{q}\right)^2$

5. $r \geq \left(\frac{v^{\frac{3}{2}}}{q}\right)^2 * 2v \rightarrow r \geq \frac{2v^4}{q^2}$

d)

Gegeben: $g(q) \leq \frac{q}{2} \log_2(q)$

Map übergibt die Bitstrings an den Reducer. Der Reducer vergleicht die Strings und berechnet deren Unterschied. Die maximale Anzahl an Bitstring Paaren, die sich in einer Zahl unterscheiden ist $\frac{\frac{n}{2}\left(\frac{n}{2}-1\right)}{2}$, wenn die Hälfte der Zahlen eine Zahl und die andere eine um 1 unterschiedliche. Angenommen alle Bitstrings unterscheiden sich in mindestens einem Bit und die Anzahl an Bitstrings entspricht $\mathrm{n} = 2^b$, dann ist die maximale Anzahl an möglichen Paaren $\frac{n}{2} \log_2(n)$, wie anhand der Beispieltabelle ersichtlich.

b=	4					3				2			1
	0	0	0	0		0	0	0		0	0		0
	0	0	0	1		0	0	1		0	1		1
	0	0	1	0		0	1	0		1	0		
	0	0	1	1		0	1	1		1	1		
	0	1	0	0		1	0	0					
	0	1	0	1		1	0	1					
	0	1	1	0		1	1	0					
	0	1	1	1		1	1	1					
	1	0	0	0									
	1	0	0	1									
	1	0	1	0									
	1	0	1	1									
	1	1	0	0									
	1	1	0	1									
	1	1	1	0									
	1	1	1	1									
Anzahl Paare	32					12				4			1

1. Aus der vorherigen Diskussion geht hervor:

$q = b$

$g(q) = \frac{q}{2b} \log_2\left(\frac{q}{b}\right) * 2b$

2. Die Größe der Ausgabe entspricht $m = \frac{n}{2} \log_2(n) * 2b$
3. $\sum_i g(q_i) \geq m \rightarrow \sum_i \frac{q_i}{2b} \log_2\left(\frac{q_i}{b}\right) * 2b \geq \frac{n}{2} \log_2(n) * 2b$

$$\leftrightarrow \sum_i q_i \log_2(q_i) \geq n\log_2(n) * b - \log_2(b)$$

4. $q \sum_i \log_2(q_i) \geq n\log_2(n) * b - \log_2(b) \leftrightarrow \sum_i \log_2(q_i) \geq \frac{n\log_2(n)*b-\log_2(b)}{q}$

$$\sum_i q_i \geq 2^{\frac{n\log_2(n)*b-\log_2(b)}{q}}$$

5. $r \geq \frac{2^{\frac{n\log_2(n)*b-\log_2(b)}{q}}}{b} \leftrightarrow r \geq b^{\frac{-1}{q}-1} * n^{\frac{bn}{q}}$