

2.2 Die AGM-Schranke

Die AGM-Schranke wurde zunächst für sog. Join-Anfragen formuliert und später dann auf beliebige konjunktive Anfragen verallgemeinert.

Definition 2.2

Eine Join-Anfrage (vom Schema σ) ist eine CQ Q (vom Schema σ) der Form

$$Q(\bar{X}_0) \leftarrow R_1(\bar{X}_1), \dots, R_m(\bar{X}_m) \quad (*)$$

für die gilt:

- 1) $cons(Q) = \emptyset$,
- 2) im Kopf der Anfrage kommen alle Variablen vor, die im Rumpf der Anfrage vorkommen
- 3) R_1, \dots, R_m sind paarweise verschieden, und
- 4) für jedes $i \in [0, m]$ gilt:
 \bar{X}_i ist ein Tupel von paarweise verschiedenen Variablen

Beispiel 2.3

- $Q_{\Delta}(A, B, C) \leftarrow E(A, B), E(B, C), E(C, A)$
ist keine Join-Anfrage (da Bedingung 3) verletzt wird)
- $Q'_{\Delta}(A, B, C) \leftarrow E_1(A, B), E_2(B, C), E_3(C, A)$
ist eine Join-Anfrage über dem Schema
 $\sigma' := \{E_1, E_2, E_3\}$, das aus drei verschiedenen
Relationssymbolen der Stelligkeit 2 besteht.
- $Q''_{\Delta}() \leftarrow E_1(A, B), E_2(B, C), E_3(C, A)$
ist keine Join-Anfrage (da Bedingung 2) verletzt wird)
- Die Anfrage

$$Q_{(a)}(A) \leftarrow E(A, \text{Sascha Lobo}), E(A, \text{resut öxl})$$
 aus Beispiel 1.2(a) ist keine Join-Anfrage,
(da sie die Bedingungen 1) und 3) verletzt).

Bemerkung 2.4

Die in Definition 2.2 eingeführten Join-Anfragen entsprechen in der "benannten Perspektive" der Situation, in der jedes R_i die Attribute \bar{X}_i hat, und in der der "natürliche Join"
 $R_1[\bar{X}_1] \bowtie \dots \bowtie R_m[\bar{X}_m]$ berechnet werden soll.

Die AGM-Schranke liefert eine Antwort auf die folgende Frage:

Gegeben sei eine Join-Anfrage Q der Form \otimes und eine DB D mit $N_i := N_i^D := |R_i^D|$ f.a. $i \in [1, m]$.

Frage: Wie viele Tupel kann es im Anfrageergebnis $Q(D)$ höchstens geben?

Betrachten wir dazu zunächst einige Spezialfälle:

1) Wenn es ein $i \in [1, m]$ gibt, s.d. $\overline{X_0} = \overline{X_i}$ ist, dann ist

$$Q(D) \subseteq R_i^D, \text{ und somit } |Q(D)| \leq N_i$$

2) Wenn $\overline{X_0} = \overline{X_{i_1}} \cdot \overline{X_{i_2}}$ für $i_1, i_2 \in [1, m]$ ist, dann ist

$$Q(D) \subseteq R_{i_1}^D \times R_{i_2}^D \text{ und somit } |Q(D)| \leq N_{i_1} \cdot N_{i_2}$$

3) Wenn $\overline{X_0} = \overline{X_{i_1}} \cdot \dots \cdot \overline{X_{i_k}}$ für $i_1, \dots, i_k \in [1, m]$ ist,

$$\text{dann ist } Q(D) \subseteq R_{i_1}^D \times \dots \times R_{i_k}^D \text{ und somit } |Q(D)| \leq \prod_{j=1}^k N_{i_j}$$

Letztendlich kommt es dabei nicht auf die Reihenfolge an, in der die einzelnen Variablen in einem Tupel \bar{x}_i auftreten.

So gilt an Stelle von 1), 2), 3) sogar

1') Wenn es ein $i \in [1, m]$ gibt, s.d.

$$\{\bar{x}_0\} = \{\bar{x}_i\}, \text{ dann ist } |Q(D)| \leq N_i$$

Hierbei verwenden wir folgende Notation:

Für ein Tupel $\bar{t} = (t_1, \dots, t_e)$ ist

$$\{\bar{t}\} := \{t_1, \dots, t_e\}.$$

Für eine Liste von Tupeln $\bar{t}_1, \dots, \bar{t}_k$ ist

$$\{\bar{t}_1, \dots, \bar{t}_k\} := \{\bar{t}_1\} \cup \dots \cup \{\bar{t}_k\}.$$

2') Wenn $\{\bar{x}_0\} = \{\bar{x}_{i_1}, \bar{x}_{i_2}\}$ für $i_1, i_2 \in [1, m]$ ist, dann ist $|Q(D)| \leq N_{i_1} \cdot N_{i_2}$

3') Wenn $\{\bar{x}_0\} = \{\bar{x}_{i_1}, \dots, \bar{x}_{i_k}\}$ für $i_1, \dots, i_k \in [1, m]$ ist, dann ist $|Q(D)| \leq \prod_{j=1}^k N_{i_j}$

Wir formulieren nun 3') von einem anderen Blickwinkel aus:

Definition 2.4 (Anfrage-Graph G_Q)

Sei Q eine Join-Anfrage der Form

$$Q(\bar{X}_0) \leftarrow R_1(\bar{X}_1), \dots, R_m(\bar{X}_m)$$

Sei $n := |\{\bar{X}_1, \dots, \bar{X}_m\}|$ die Anzahl der verschiedenen in Q vorkommenden Variablen und sei A_1, \dots, A_n eine Liste all dieser Variablen.

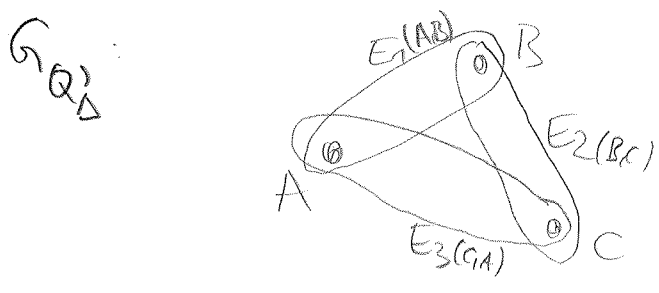
Der zu Q gehörende Hypergraph G_Q

hat die Knotenmenge $V_Q := \{A_1, \dots, A_n\} = \text{vars}(Q)$ und für jedes $i \in [1, m]$ hat G_Q eine Hyperkante (namens $R_i(\bar{X}_i)$), die aus den Knoten $\{\bar{X}_i\}$ besteht.

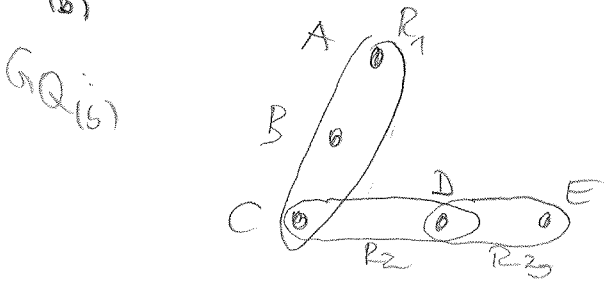
Beachte: Gemäß Bedingung 2) in Def. 2.2 gilt: $\{\bar{X}_0\} = \{A_1, \dots, A_n\}$

Beispiel 2.5

(a) $Q_{\Delta}^1(A, B, C) \leftarrow E_1(A, B), E_2(B, C), E_3(C, A)$



(b) $Q_{(b)}(A, B, C, D, E) \leftarrow R_1(A, B, C), R_2(C, D), R_3(D, E)$



Definition 2.6 (Kantenüberdeckung)

Sei \mathcal{G} ein beliebiger Hypergraph mit Knotenmenge V .

Eine Kantenüberdeckung (engl.: edge cover) von \mathcal{G} ist eine Menge K von Hyperkanten von \mathcal{G} ,

für die gilt: $\bigcup_{e \in K} e = V$ (d.h.: jeder Knoten

von \mathcal{G} ist in mindestens einer Hyperkante enthalten, die zu K gehört).

Beispiel 2.7

$K := \{E_1(A,B), E_2(B,C)\}$ ist eine Kantenüberdeckung von $\mathcal{G}_{Q'_\Delta}$ für die Anfrage Q'_Δ aus Bsp. 2.5(a)

$K := \{R_1(A,B,C), R_3(D,E)\}$ ist eine Kantenüberdeckung von $\mathcal{G}_{Q_{(b)}}$ für die Anfrage $Q_{(b)}$ aus Bsp. 2.5(b)

Notation:

Wenn wir Join-Anfragen Q betrachten, sprechen wir von "Kantenüberdeckungen von Q " und meinen damit Kantenüberdeckungen von \mathcal{G}_Q .

Die Beobachtung 3') lässt sich also wie folgt formulieren:

3'') Wenn $i_1, \dots, i_k \in [1, m]$ so gewählt sind, dass $K := \{ R_{i_1}(\bar{x}_{i_1}), \dots, R_{i_k}(\bar{x}_{i_k}) \}$ eine Kantenüberdeckung von Q ist, dann gilt: $|Q(D)| \leq \prod_{j=1}^k N_{i_j}$.

Eine solche Menge K können wir auch repräsentieren als eine Abbildung $x: [1, m] \rightarrow \{0, 1\}$ mit $x(i) = \begin{cases} 1 & \text{falls } R_i(\bar{x}_i) \in K \\ 0 & \text{falls } R_i(\bar{x}_i) \notin K. \end{cases}$

Dann repräsentiert eine Abbildung $x: [1, m] \rightarrow \{0, 1\}$ genau dann eine Kantenüberdeckung von Q , wenn gilt:

Für alle $j \in [1, m]$ gilt: $\sum_{\substack{i \in [1, m] \text{ mit} \\ A_j \in \{ \bar{x}_i \}}} x(i) \geq 1$ (**)

Beobachtung 3'') lässt sich also auch wie folgt formulieren:

3''') Für jede Abbildung $x: [1, m] \rightarrow \{0, 1\}$, die (**) erfüllt, gilt: $|Q(D)| \leq \prod_{i=1}^m (N_i)^{x(i)}$

Beachte: F. a. reellen Zahlen $z \neq 0$ gilt: $z^0 = 1$ und es gilt: $0^0 = 0$.