# Konjunktive Anfragen ohne Konstanten

## Definition 2.14

Sei $Q(\overline{x_0}) \leftarrow R_1(\overline{x_1}), \ldots, R_m(\overline{x_m})$ eine konjunktive Anfrage, in der keine Konstanten vorkommen (d.h. $cons(Q) = \emptyset$), sei $\{A_1, \ldots, A_n\} = vars(Q)$ und $\{\overline{x_0}\} = \{A_1, \ldots, A_k\}$.

Eine fraktionale Kantenüberdeckung von $Q$ ist eine Abbildung $x : [m] \to \mathbb{Q}_{\geq 0}$, s.d. f.a. $j \in [k]$ gilt:

$$\sum_{\substack{i \in [m] \text{ mit} \\ A_j \in \{\overline{x_i}\}}} x(i) \geq 1$$

## Satz 2.15 (Gottlob, Lee, Valiant, 2009)

Sei $Q$ eine konjunktive Anfrage mit $cons(Q) = \emptyset$ der Form $Q(\overline{x_0}) \leftarrow R_1(\overline{x_1}), \ldots, R_m(\overline{x_m})$.

(a) Für jede fraktionale Kantenüberdeckung $x$ von $Q$ und jede Datenbank $D$ und alle $N_1, \ldots, N_m \in \mathbb{N}$ mit $|R_i^D| \leq N_i$ f.a. $i \in [m]$ gilt:

$$|Q(D)| \leq \prod_{i=1}^{m} N_i^{x(i)}.$$

(b) Für jedes $N \in \mathbb{N}$ gibt es natürliche Zahlen $N_1, \ldots, N_m \geq N$, eine Datenbank $D$ mit $|R_i^D| \leq \sum_{\substack{j \in [m] \text{ mit} \\ R_j = R_i}} N_j$ f.a. $i \in [m]$ und eine fraktionale Kantenüberdeckung $x$ von $Q$, s.d.

$$|Q(D)| \geq \prod_{i=1}^{m} N_i^{x(i)}.$$

Beweis:

(I) Wir zeigen die Aussage zunächst für **selbst-join-freie** Anfragen, d.h. für Anfragen, bei denen die Relationssymbole $R_1, \dots, R_m$ paarweise verschieden sind (d.h. $R_i \neq R_j$ f.a. $i,j \in [m]$ mit $i \neq j$).

Für jedes $i \in [0,m]$ sei $\overline{x}_i^{\text{versch}}$ die Variablenliste, die man aus $\overline{x}_i$ erhält, indem man Mehrfach-Vorkommen derselben Variablen entfernt ("versch" steht für "verschieden").
Insbes. ist $\{\overline{x}_i^{\text{versch}}\} = \{\overline{x}_i\}$.
Sei $k_i := |\{\overline{x}_i\}| = |\{\overline{x}_i^{\text{versch}}\}|$ und sei $R_i^{\text{versch}}$ ein neues Relationssymbol der Stelligkeit $k_i$.
Betrachte die Anfrage

$$Q^{\text{versch}}(\overline{x}_0^{\text{versch}}) \leftarrow R_1^{\text{versch}}(\overline{x}_1^{\text{versch}}), \dots, R_m^{\text{versch}}(\overline{x}_m^{\text{versch}}).$$

Diese Anfrage ist eine Projektion einer Join-Anfrage (i.S.v. Def. 2.11). Außerdem gilt für jede Abbildung $x : [m] \to \mathbb{Q}_{\geq 0}$:

$(\heartsuit)$: $x$ ist eine frakt. Kantenüberdeckung von $Q^{\text{versch}}$ $\iff$ $x$ ist eine frakt. Kantenüberdeckung von $Q$

Betrachte eine beliebige DB $D$ vom Schema $\{R_1, \dots, R_m\}$.
Sei $D'$ die DB vom Schema $\{R_1^{\text{versch}}, \dots, R_m^{\text{versch}}\}$ mit $(R_i^{\text{versch}})^{D'} := Q'_i(D)$ mit $Q'_i(\overline{x}_i^{\text{versch}}) \leftarrow R_i(\overline{x}_i)$.
Dann ist $N'_i := |(R_i^{\text{versch}})^{D'}| \leq |R_i^D|$.

Und es gilt: $|Q(D)| = |Q^{versch}(D')|$ (*)

Gemäß Satz 2.12(a) gilt für jede fraktionale Kantenüberdeckung $x$ von $Q^{versch}$ (und $Q$, wegen (▽)), dass $|Q^{versch}(D')| \leq \prod_{i=1}^{m} N_i^{x(i)}$.

Also gilt:

$$|Q(D)| \underset{(*)}{=} |Q^{versch}(D')| \leq \prod_{i=1}^{m} N_i^{x(i)} \leq \prod_{i=1}^{m} |R_i^D|^{x(i)} \leq \prod_{i=1}^{m} N_i^{x(i)}$$

f.a. $N_1, \ldots, N_m \in \mathbb{N}$ mit $|R_i^D| \leq N_i$ f.a. $i \in [m]$.

Also gilt Aussage (a) für selbst-join-freie Anfragen $Q$.

Zum Beweis von Aussage (b) für selbst-join-freie Anfragen $Q$ betrachte ein beliebiges $N \in \mathbb{N}$.

Gemäß Satz 2.12(b) gibt es eine $\{R_1^{versch}, \ldots, R_m^{versch}\}$-DB $\hat{D}$ mit $\hat{N}_i := |(R_i^{versch})^{\hat{D}}| \geq N$ f.a. $i \in [m]$ und eine fraktionale Kantenüberdeckung $x$ von $Q^{versch}$ (und $Q$ wg. (▽)), s.d. $|Q^{versch}(\hat{D})| = \prod_{i=1}^{m} \hat{N}_i^{x(i)}$. (**)

Man kann leicht eine DB $D$ vom Schema $\{R_1, \ldots, R_m\}$ konstruieren, s.d. $D' = \hat{D}$ und $|R_i^D| = |(R_i^{versch})^{D'}|$ f.a. $i \in [m]$ ist (<u>Details</u>: Übung). Insgesamt gilt:

$$|Q(D)| \underset{(*)}{=} |Q^{versch}(D')| = \prod_{i=1}^{m} \hat{N}_i^{x(i)} = \prod_{i=1}^{m} |R_i^D|^{x(i)}.$$

Dies beweist Aussage (b) für selbst-join-freie Anfragen $Q$.

<u>(II)</u> Wir zeigen nun, dass die Aussage auch für Anfragen $Q$ gilt, die nicht selbst-join-frei sind. Wir ordnen dazu der Anfrage

$$Q(\overline{x}_0) \leftarrow R_1(\overline{x}_1), \dots, R_m(\overline{x}_m)$$

die selbst-join-freie Anfrage

$$Q'(\overline{x}_0) \leftarrow R'_1(\overline{x}_1), \dots, R'_m(\overline{x}_m)$$

vom Schema $\{R'_1, \dots, R'_m\}$ zu, wobei $R'_1, \dots, R'_m$ paarweise verschiedene Relationssymbole mit $ar(R'_i) = ar(R_i)$ f.a. $i \in [m]$ sind.

Gemäß (I) wissen wir, dass die Aussagen (a) und (b) für die Anfrage $Q'$ gelten. Außerdem gilt für jede Abbildung $x : [m] \to \mathbb{Q}_{\geq 0}$:

($\triangleright$) $x$ ist eine fraktionale Kantenüberdeckung von $Q$ $\iff$ $x$ ist eine fraktionale Kantenüberdeckung von $Q'$.

Zum Beweis von (a) ordnen wir jeder $\{R_1, \dots, R_m\}$-DB $D$ die $\{R'_1, \dots, R'_m\}$-DB $D'$ mit $R_i'^{D'} := R_i^D$ f.a. $i \in [m]$ zu. <u>Klar:</u> $Q(D) = Q'(D')$.
Damit erhalten wir, dass Aussage (a) auch für $Q$ gilt.

Zum Beweis von Aussage (b) für $Q$ betrachte ein beliebiges $N \in \mathbb{N}$. Gemäß <u>(I)(b)</u> gibt es eine $\{R'_1, \dots, R'_m\}$-DB $D'$ mit $N_i := |R_i'^{D'}| \geq N$ f.a. $i \in [m]$ und eine fraktionale Kantenüberdeckung $x$ von $Q'$ (und $Q$, wegen ($\triangleright$)), s.d. $|Q'(D')| \geq \prod_{i=1}^{m} N_i^{x(i)}$ (Begründung: Lösung!)

Sei $\tilde{D}$ die DB vom Schema $\{R_1, \dots, R_m\}$ mit

$$R_i^{\tilde{D}} := \bigcup_{\substack{j \in [m] \text{ mit} \\ R_j = R_i}} R_j'^{D'} .$$

Man kann leicht nachprüfen, dass

$$Q(\tilde{D}) \supseteq Q'(D').$$

Also gilt:

$$|Q(\tilde{D})| \geq |Q'(D')| \geq \prod_{i=1}^{m} N_i^{x(i)} .$$

Außerdem gilt f.a. $i \in [m]$, dass

$$|R_i^{\tilde{D}}| \leq \sum_{\substack{j \in [m]: \\ R_j = R_i}} |R_j'^{D'}| = \sum_{\substack{j \in [m] \\ R_j = R_i}} N_j .$$

Dies zeigt, dass Aussage (b) für $Q$ gilt

$\square$