2.3 Verallgemeinerung der AGM-Schranke für beliebige konjunktive Anfragen

Ziel dieses Abschnitts ist, eine Verallgemeinerung der AGM-Schranke zu finden, die nicht nur für Join-Anfragen, sondern für beliebige konjunktive Anfragen gilt. Wir gehen dazu in mehreren Schritten vor, in denen wir immer allgemeinere Varianten von Anfragen betrachten.

Zunächst betrachten wir:

Projektionen von Join-Anfragen

Definition 2.11:

(a) Eine Projektion einer Join-Anfrage ist eine konjunktive Anfrage Q der Form

$$Q(A_1, \dots, A_k) \leftarrow R_1(\overline{X}_1), \dots, R_m(\overline{X}_m),$$

für die es ein $n \geq k$ und Variablen $A_{k+1}, \dots, A_n$ gibt, so dass die Anfrage $\tilde{Q}$ mit

$$\tilde{Q}(A_1, \dots, A_n) \leftarrow R_1(\overline{X}_1), \dots, R_m(\overline{X}_m)$$

eine Join-Anfrage ist.

(b) Sei $Q(A_1, \dots, A_k) \leftarrow R_1(\overline{x}_1), \dots, R_m(\overline{x}_m)$ eine Projektion einer Join-Anfrage.
Eine fraktionale Kantenüberdeckung von Q ist eine Abbildung $x : [m] \to \mathbb{Q}_{\geq 0}$, s.d. f.a. $j \in [k]$ gilt:
$$\sum_{\substack{i \in [m] \text{ mit} \\ A_j \in \{\overline{x}_i\}}} x(i) \geq 1$$

Unter Verwendung der AGM-Schranke (Satz 2.9 & 2.10) können wir leicht folgern, dass Folgendes gilt:

<u>Satz 2.12:</u>

Sei Q eine Projektion einer Join-Anfrage.

(a) Für jede fraktionale Kantenüberdeckung x von Q und jede Datenbank D vom Schema $\{R_1, \dots, R_m\}$ gilt
$$|Q(D)| \leq \prod_{i=1}^{m} N_i^{x(i)},$$
wobei $N_i := |R_i^D|$ f.a. $i \in [m]$ ist.

(b) Für jedes $N \in \mathbb{N}$ 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, s.d.
$$|Q(D)| = \prod_{i=1}^{m} N_i^{x(i)}.$$

Beweis:

(a) Sei $\tilde{Q}(A_1, \dots, A_n) \leftarrow R_1(\overline{x}_1), \dots, R_m(\overline{x}_m)$ eine zu Q gehörige Join-Anfrage (i.S.v. Def 2.11 (a)).

Für jedes $i \in [m]$ sei $\overline{x}_i'$ die Variablenliste, die aus $\overline{x}_i$ entsteht, indem man alle Variablen aus $\{A_{k+1}, \dots, A_n\}$ weglässt, sei r_i' die Länge der Liste $\overline{x}_i'$ und sei R_i' ein neues Relationssymbol der Stelligkeit r_i'.

Wir betrachten die Anfrage Q' vom Schema $\{R_1', \dots, R_m'\}$ mit

$$Q'(A_1, \dots, A_k) \leftarrow R_1'(\overline{x}_1'), \dots, R_m'(\overline{x}_m').$$

Offensichtlicherweise ist Q' eine Join-Anfrage, und ~~für~~ jede Abbildung $x: [m] \to \mathbb{Q}_{\geq 0}$ gilt:

(▷) x ist eine fraktionale Kantenüberdeckung von Q' $\iff$ x ist eine fraktionelle Kantenüberd. von Q'.

Sei nun D eine beliebige DB vom Schema $\{R_1, \dots, R_m\}$.
Sei D' die DB vom Schema $\{R_1', \dots, R_m'\}$ mit

$$R_i'^{D'} := Q_i(D)$$

wobei Q_i die Anfrage $Q_i(\overline{x}_i') \leftarrow R_i(\overline{x}_i)$ ist, f.a. $i \in [m]$ (d.h. $R_i'^{D'}$ ist die Projektion von R_i^D auf die Komponenten, die zu $A_1, \dots, A_k$ gehören).

Klar: $N_i' := |R_i'^{D'}| \leq |R_i^D| = N_i$

Man kann sich leicht davon überzeugen, dass

$$Q(D) \subseteq Q'(D')$$

ist. Sei x eine fraktionale Kantenüberdeckung von Q (und Q', wegen (▽))

Gemäß AGM-Schranke (Satz 2.9) gilt

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

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

□(a)

<u>(b)</u> Wir nutzen die gleiche Notation wie im Beweis von <u>(a)</u>. Sei $N \in \mathbb{N}$ beliebig. Gemäß Optimalität der AGM-Schranke (Satz 2.10) gibt es eine DB D' vom Schema $\{R'_1, \dots, R'_m\}$ mit

$$N_i := |R_i'^{D'}| \geq N \quad \text{f.a. } i \in [m]$$

und eine fraktionale Kantenüberdeckung x von Q' (und Q, wegen (▽)), so dass

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

Sei a ein beliebiges, fest gewähltes Element in dom und sei D die DB vom Schema $\{R_1, \dots, R_m\}$, die man aus D' erhält, indem man jedes Tupel $t' \in R_i'^{D'}$ an geeigneten Stellen um weitere

Komponenten anreichert, deren Eintrag jeweils den Wert a bekommt

<u>Beispiel:</u> Falls das i-te Atom von Q die Form $R_i(A_2, A_{k+1}, A_1)$ hat (und $k \geq 3$), so hat das i-te Atom von Q' die Form $R_i'(A_2, A_1)$. Für jedes Tupel $t' = (x,y) \in R_i'^{D'}$ enthält R_i^D dann das Tupel $\hat{t'} := (x, a, y)$.

<u>Präzise:</u> Für $i \in [m]$ sei $R_i(\bar{x}_i)$ von der Form $R_i(A_{j(i,1)}, \dots, A_{j(i,r_i)})$ und sei $M_i := \{ \ell \in \{1, \dots, r_i\} : j(i,\ell) > k \}$.
Jedem Tupel $t' = (t'_1, \dots, t'_{r_i'}) \in dom^{r_i'}$ ordnen wir das Tupel $\hat{t'} = (\hat{t'}_1, \dots, \hat{t'}_{r_i}) \in dom^{r_i}$ zu, für das gilt: 1) $\hat{t'}_\ell = a$ f.a. $\ell \in M_i$, und

2) t' ist das Tupel, das aus $\hat{t'}$ entsteht, indem man für jedes $\ell \in M_i$ die ℓ-te Komponente $\hat{t'}_\ell$ löscht.

D ist dann die DB vom Schema $\{R_1, \dots, R_m\}$ mit

$$R_i^D := \{ \hat{t'} : t' \in R_i'^{D'} \} \quad \text{f.a. } i \in [m].$$

Dann ist $|R_i^D| = |R_i'^{D'}| = N_i$ f.a. $i \in [m]$.
Außerdem kann man sich leicht davon überzeugen, dass $Q(D) = Q'(D')$ ist.
Insgesamt gilt also: $|Q(D)| = \prod_{i=1}^{m} N_i^{x(i)}$ und $N_i = |R_i^D|$ f.a. $i \in [m]$.

D (s)

Beispiel 2.13:

Die Anfrage

$$Q(A,B) \leftarrow E_1(A,B),\ E_2(B,C),\ E_3(C,A)$$

ist eine Projektion einer Join-Anfrage — nämlich der Join-Anfrage

$$\tilde{Q}(A,B,C) \leftarrow E_1(A,B),\ E_2(B,C),\ E_3(C,A).$$

Natürlich ist $x_1 : \{1,2,3\} \to \mathbb{Q}_{\geq 0}$ mit $x_1(1) = x_1(2) = x_1(3) = \frac{1}{2}$ eine fraktionalle Kantenüberdeckung von $\tilde{Q}$ und von Q.

Aber gemäß Definition 2.11 ist auch $x_2 : \{1,2,3\} \to \mathbb{Q}_{\geq 0}$ mit $x_2(1) = 1$ und $x_2(2) = x_2(3) = 0$ eine fraktionale Kantenüberdeckung von Q (aber nicht von $\tilde{Q}$).

Satz 2.12 liefert, dass für jede DB D gilt:

$$|Q(D)| \leq \prod_{i=1}^{3} N_i^{x_2(i)} = N_1 ,$$

für $N_i := |R_i^D|$ f.a. $i \in [3]$ ist.

(... was wir bei dieser einfachen Anfrage Q aber auch direkt, ohne Nutzung der AGM-Schranke, sehen können).