Query Optimization – Homework 1

January, 2026

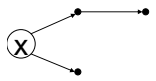
Submit your answer in a pdf file on Canvas.

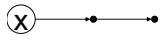
- Write your name in the file.
- Use this template `https://www.overleaf.com/project/67f01a081d8c577a12f22353`

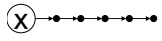
Grading is done using credit/partial-credit/no-credit; ignore the points below.
An asterix * indicates that the question may be more challenging.

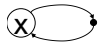
1 Query Containment

1. (0 points)
   (a) Indicate all containment or equivalence relationships between the following queries:

$Q_1(x) = E(x, y) \wedge E(y, z) \wedge E(x, w)$ 

$Q_2(x) = E(x, u) \wedge E(u, v)$ 

$Q_3(x) = E(x, u_1) \wedge E(u_1, u_2) \wedge \cdots \wedge E(u_4, u_5)$ 

$Q_4(x) = E(x, y) \wedge E(y, x)$ 

   Your answer should be something like $Q_1 \subsetneq Q_2 \equiv Q_3 \subsetneq Q_4$ (not the real answer).

(b) Consider the following two conjunctive queries, over the language of graphs:

$$Q_1(x) = E(x,y) \wedge E(y,z) \wedge E(z,u)$$
$$Q_2(x) = E(x,y) \wedge E(y,z)$$

i. Prove that $Q_1 \subseteq Q_2$.

ii. * Find a query Q such that $Q_1 \subsetneq Q \subsetneq Q_2$. In other words, Q is between Q_1 and Q_2, but not equivalent to either of them.

(c) * Join/Semi-join Identities.

Let R, S, T be three relations, and consider the following identity:

$$\boxed{R \ltimes (S \bowtie T) \equiv R \ltimes (S \ltimes T)}$$

Both the join $\bowtie$ and the left semi-join $\ltimes$ are assumed to be "natural" joins, i.e. the join is on all common variables.

i. Assuming that $\text{Vars}(R) \cap \text{Vars}(T) \subseteq \text{Vars}(S)$ prove that the identity above holds.

ii. Given a simple example where $\text{Vars}(R) \cap \text{Vars}(T) \not\subseteq \text{Vars}(S)$ and the identity above fails.

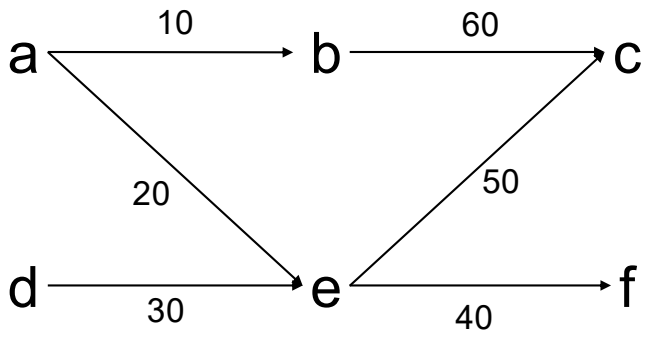
2 K-Relations

2. (0 points)

(a) Consider the following sum-product query:

$$\boxed{Q(x) = E(x, y) \otimes E(y, z)}$$

Consider the following labeled graph:



We intepret the graph as a K-realtion, for example $E[a, b] = 10, E[b, c] = 60$, etc Compute the answer to Q in each of the cases below. You do not need to indicate the answers that are $\mathbf{0}$.

i. K is the semiring of Booleans. (In this case we ignore the labels.)

ii. $K = (\mathbb{N}, +, *, 0, 1)$ is the semiring of natural numbers, with addition and multiplication.

iii. $K = (\mathbb{N}, \min, +, \infty, 0)$ is the tropical semiring.

(b) Consider the following Boolean query:

$$\boxed{Q() = R(x, y_1, z) \wedge R(x, y_2, z)}$$

The relation instance R below has each tuple annotated with provenance tokens r, s, t. Compute the provenance polynomial of Q. Your answer should be a polynomial in r, s, t, e.g. $r^3 + 3s + rt$ (not the real answer).

$$R = \begin{array}{|c|c|c|l} \hline a_1 & b_1 & c_1 & r \\ a_2 & b_1 & c_2 & s \\ a_3 & b_2 & c_3 & t \\ \hline \end{array}$$