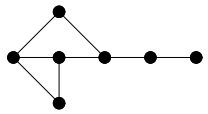
CMSC 724–0101 Spring 2004 Final Exam (take-home, 100 points) Due 2004-05-13 8:00:00

You are free to use any resources you like (e.g., papers, publicly available code, homework solutions found on the Web) in completing this exam, subject to two conditions: (1) All such resources used must be clearly and prominently acknowledged. (2) Everything you write should be original (in your own words) and you should be able to explain everything from scratch.

1. (20 points) Provide a method that takes as input a *DB2 Data Access Definition (DAD)* [2] and produces as output the definitions of triggers (using standard SQL [3, Section 7.4]) that ensure the correct maintenance (as base tables are modified) of the *side tables* defined by the DAD.
2. (20 points) Consider a distributed database with relation $R_1(A, B)$ at site S_1, $R_2(B, C, D, E)$ at S_2, $R_3(C, E, F)$ at S_3, and $R_4(B, F, H)$ at S_4. A query $\pi_{AEH}(R_1 \bowtie R_2 \bowtie R_3 \bowtie R_4)$ arrives at S_1 (meaning its result is required at S_1). Does the query have a *full reducer* [1, page 129]? (Justify your answer.) Describe a query plan that minimizes network traffic. Describe a query plan that minimizes response time [4, page 430]. Assume that all relations are of roughly the same size, and that join selectivities are inversely proportional to the number of join attributes. (That is, joins with a larger set of join attributes result in proportionally smaller results.) Clearly state any additional assumptions you make.
3. (15 points) Prove that if a generic mapping is *computable relative to an encoding* [1, Definition 16.1.2] then it is *computable* [1, Definition 16.1.5].
4. (15 points) Consider a database instance that encodes a directed graph in the usual manner with a relation $G(S, D)$ representing the edges. Consider the following Datalog program:
   `ePath(x,y)` ← `G(x,z), G(z,y).`
   `ePath(x,y)` ← `ePath(x,z), G(z,w), G(w,y).`
   `answer(x)` ← `ePath(x,x).`
   Let p be an arbitrary permutation of the set of constants appearing in the relation G that leaves G unchanged; that is $p(G) = G$. (The permutation p is applied to G in the natural manner, by applying it to each attribute of each tuple of G.) Prove or disprove: If $p(a) = b$ then $a \in$ `answer` iff $b \in$ `answer`.
5. (15 points) Prove, *without* using Theorem 17.2.2 [1, page 435], that the relation $\equiv_r$ defined using Ehrenfeucht-Fraissé games is transitive.
6. (15 points) Let q be a boolean query on the relation G of Question 4 that returns true iff G encodes a graph that contains a subgraph isomorphic to the following graph:

   Is q *almost surely true* or *false* [1, Definition 17.7.3]? Justify your answer.

Submission Submit your solution to this exam electronically as a single gzipped tar archive. Your answers should be in a single PDF file. Please check that the PDF file is portable. At the very least, the `gv` program on the CSIC cluster must display it properly. Include a README file that describes the contents and their relation to the above questions. Name your file using the scheme `LastnameIJ-fin-NNNN.tgz`, where `NNNN` is a 4-digit integer of your choice, and upload by anonymous FTP to `ftp.cs.umd.edu`, directory `incoming/chaw`.

References

[1] S. Abiteboul, R. Hull, and V. Vianu. *Foundations of Databases*. Addison-Wesley, 1995.

[2] S. S. Chawathe. *Semistructured Data in Relational Databases*, chapter 3. Practical Handbook of Internet Computing. CRC Press, 2004. To appear.

[3] H. Garcia-Molina, J. D. Ullman, and J. Widom. *Database Systems: The Complete Book*. Prentice-Hall, 2002.

[4] D. Kossmann. The state of the art in distributed query processing. *ACM Comput. Surv.*, 32(4):422–469, 2000.