PABLO MALÁ

Knowledge Discovery in Graph data – exam*

Printed documents and notes are allowed -duration: 2:00

Computer Science Master 2 – Data Science – Paris Saclay University

February 17, 2025

In what follows. the data are presented in tables and not in graphs for readability reasons. The triple subjects are in the first column, the predicates are in the first line and the objects are in the other cells of each table.

1 Key discovery from Graph Data *[14pts]*

Question 1. *[1.5pts]* Give the main principle of discovering keys adopted by SAKey tool? ✓

Question 2. *[1.5pts]* Give the informal and a formal definition of discriminability seen in the course. ✓

Question 3. *[1pt]* Give the definition of a conditional key in a knowledge graph. ✓

Question 4. *[3pts]* To distinguish between the three key semantics *S-Keys*, *SF-Keys* and *F-Keys* that are studied in the course,

(a) What are the main data characteristics that should be taken into account ? ✓

(b) How these characteristics are considered in the *S-Keys*, *SF-Keys* and *F-Keys* semantics? ✓

Question 5. In table 1 we give an extract of some archaeological sites descriptions. These sites are described by six properties {`Name`, `Location`, `Period`, `civilization`, `UNESCO`, `Archaeologists`}.
Given these data if we apply SAKey, a key discovery tool that allows to discover n-almost keys (under the S-key semantics):

(b) Give two 0-almost keys composed of two properties. *[1pt]* ✓

(a) Give all the 2−almost keys of one property that can be discovered. *[1,5pts]*

*The mark scale is given as an indication.

	Name	Location	Period	Civilization	UNESCO	Archaeologists
s_1	Chichén Itzá	Mexico	9th c.	Maya civilization	1988	J. L. Stephens
s_2	Lascaux	France	17,000 BC			M. Ravidat
s_3	Alesia	France	52 BC	Roman Empire		A. Berthier
s_4	Lascaux	France	17,000 BC		1979	A. Berthier
s_5	Pompeii	Italy	6th c. BC	Rome	1997	G. Fiorelli, K. Weber
s_6	Chichén Itzá	Mexico	7th c.	Maya	1988	J. L. Stephens, E. H. Thom
s_7	Pompeii	Italy	6th c. BC	Ancient Rome	1997	G. Fiorelli
s_8	Petra	Jordan	4th c. BC	Nabataeans	1988	J. L. Burckhardt, G. Fiorelli

Table 1: Extract of archaeological sites descriptions (D1)

(c) Give a 0-S-Key, composed of two properties, that is not an F-Key that can be discovered in the data presented in Table 1. *[1,5pts]* √

(d) Give a SF-Key, composed of two properties, that is not an S-Key that can be discovered in the data presented in Table 1. *[1,5pts]* √

(e) Give two conditional keys that can be discovered by SAKey in the data of Table 1, when the number of exceptions is 0. *[1,5pts]* √