A Very Simple $\LaTeX\,2_\varepsilon$ Template

Christos Chatzifountas

November 22, 2021

Introduction

Haplotype database representation

Objective

We want to apply differential privacy Techniques to haplotypes

Basics

A database x is a collection of elements of a universe $\mathcal{X}$ of rows (records) The histogram of a database is a vector $x_i \dots x_n, n = |\mathcal{X}|$and x_i is the number of repeats of a row in the database. The l_1 norm of a database is defined in thought of it's histogram:

$$\|x\|_1 = \sum_{i=1}^{|\mathcal{X}|} |x_i| \tag{1}$$

The distance is defined thought the norm in the usual way

A haplotype is a collection of nodes and edges (connections between nodes), in other words a graph database. Graph databases can be represented relational through with adjacency lists So for any graph we need a column of nodes, and a column of adjacent nodes

Lemma 1. *Let x,y be haplotypes. If they differ in a node then* $\|x - y\|_1 \geq 1$

Proof. Let x,y be haplotypes that on one node or more. Then that node and the adjacency differ in their adjacency list representation. Hence their histograms differ on two or more coordinates. If two integers are different, their difference is greater than one , hence $\|x-y\| \geq |x_n - y_n| + |x_m - y_m| \geq 2$ □