# Robustness and Scalability of Truth Finding Algorithms

Dalia Attia Wagui Mouhamadou Lamine BA Laure Berti-Equille

## ABSTRACT

*Motivation et Outline.* Truth finding is an important because conflicting , erroneous, and dirty information are everywhere. The truth must be tell when reconciling such a conflicting data from different sources. This has lead to much effort of the database community and well founded truth discovering algorithms. However, there is a lack of a comparative study of both the scalability and the robustness of these algorithms. The existing comparative studies only focus on accuracy aspects. We describe, reimplement, and compare the most prominent solutions, so far, for the truth finding problems. To tackle this lack, we propose in this paper an experimental study of the robustness and the scalability of the most referenced truth finding algorithms. Our outline is as follows.

First, we overview the truth finding problem by giving preliminary definitions, a classificartion of the literature, and by describe in details the algorithms (most referenced algorithms) we have considered in this study.

## 1. INTRODUCTION

## 2. OVERVIEW ON TRUTH FINDING ALGORITHMS

### 2.1 Preliminaries

### 2.2 Classification

Cluster the algorithms in the three following classes.

- Iterative algorithms:
- EM based algorithms:
- Dependency detection based algorithms:

### 2.3 Twelve Truth Finding Algorithms

We details in the sections, the truth finding algorithms we have considered for the comparative study. Some points to put in this section.

- Details each technique by giving its pseudo-code
- Hierarchical graph based representation of the common characteristics of the different algorithms
-

## 3. EXPERIMENTAL STUDY

### 3.1 Parameter Setting Evaluation

Evaluation of the parameter initialization of each algorithm on Book dataset. Validation of parameter setting inferred from experiments on the Book dataset by using other datasets, e.g., Flight dataset. (Maybe this point should be moved to the experimental secion).

### 3.2 Scalability

We compare here the different algorithms in terms of scalability, i.e., does each of them scale when the size of the input dataset increases exponentially.

### 3.3

- Propose guidance on the algorithms from the results of the experiments