Despite our efforts, there were still imprecisions related to the affiliation of the authors. To address them, we combined our dataset with the one provided by the *csranking* project,[3] which ranks CS departments based purely on their publications.They do so by collaboratively collecting information on authors, such as their homepage and affiliation. Therefore, we used that information to enhance our repository. Salient statistics on our dataset are shown in Table 4.1.

Table 4.1. Salient statistics of the dataset used in our evaluation.

Attribute	Value
Number of Papers	2,931,849
Number of Authors	1,595,771
Number of Venues	5,765
Number of US Depts	126
Avg. number of professors per US Dept	42.4
Number of BR Depts	25
Avg. number of professors per BR Dept	47.8

4.2   Computer Science Subareas

There are different ways of defining subareas in CS depending on the institution responsible for the classification. Two notorious classification are given by ACM[4] (through *Special Interest Groups*) and IEEE[5] (through *Technical Committees*). Notice that most of them divide CS into subareas rather distinct. Further, some of these divisions reflect historical decisions that may be less relevant nowadays. For this reason, previous works have attempted to automatically identify such subareas [Wainer et al., 2013] or use another source of information [Hoonlor et al., 2013].

A more recent classification is presented by Microsoft Academic Research.[6] They divide CS into 37 subareas, including relatively new subareas. For the purpose of this work, we selected 20 subareas from the Microsoft Academic Search classification, as presented in Table 4.2.

Along with the list, we present the abbreviation of each subarea, which we will be using from here on. We also show two venues we selected as notorious in each subarea of interest. These venues are essential for identifying a group of researchers whose main research topics of interest are likely to be in that subarea — an essential information

---

[3]http://csrankings.org
[4]http://acm.org/sigs
[5]http://computer.org/web/tandc/technical-committees
[6]http://academic.research.microsoft.com