

CS5346 Assignment 3 - CIRViz

March 31, 2019

Lu Wei	A0040955E
Julian Teh	A0163126M

Contents

1	Introduction	3
2	Visualizations - Purpose and Method	3
2.1	3
2.2	3
2.3	5
3	Other Information	6

1 Introduction

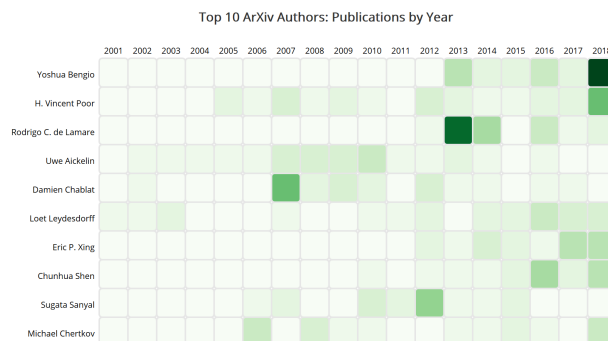
The assignment teaches students to handle big data, either by filtering it or cleaning it, and then handling it to extract insights or to explore. In this assignment we split the work into two main portions; exploration of the data, and coding the final visualizations. For exploration we used Tableau with a small dataset to examine any interesting trends, and for the final visualization we used C3.js on top of D3.js to render with the full dataset.

2 Visualizations - Purpose and Method

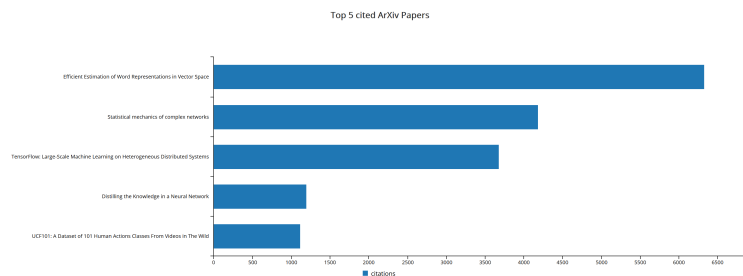
2.1

Task	Visualization
1	Heatmap
2	Categorical Bar Chart
3	Time Series Line Chart, Stacked Area Chart
4a	Force Directed Graph
4b	Tree Map

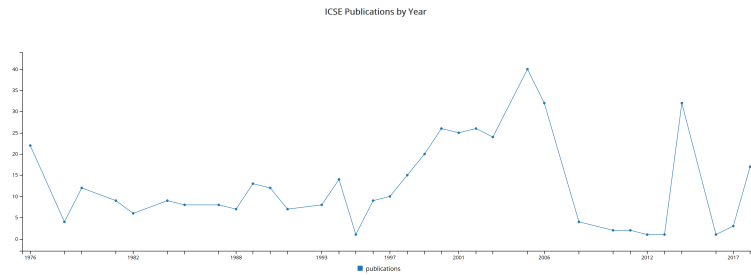
2.2



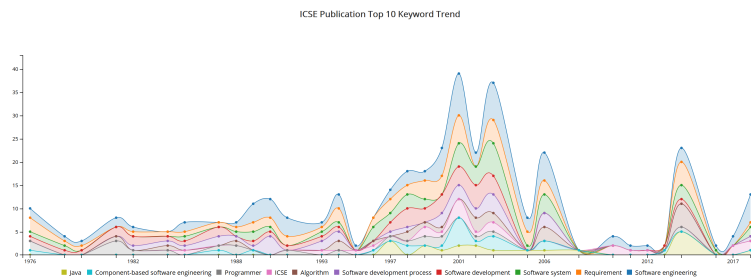
Task 1: Heat Map



Task 2: Bar Chart

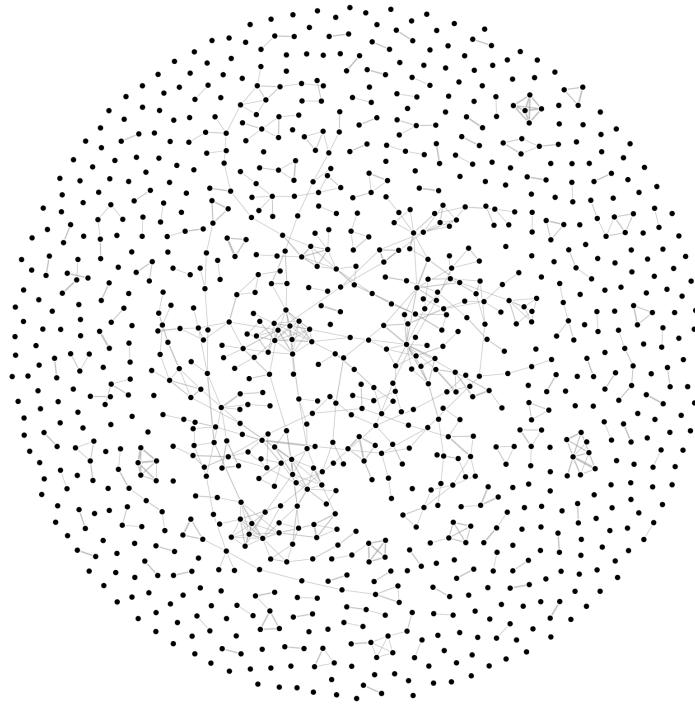


Task 3: Time Series Line Chart



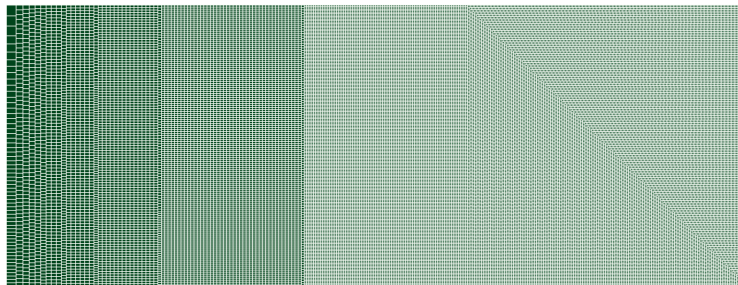
Task 3: Stacked Area Chart

ArXiv Top 1000 Author-Co-Author Network



Task 4a: Force Directed Graph

ArXiv Author Contribution Amount



Task 4b: Tree Map

2.3

For Task 3, we encoded the 10 most frequently occurring keywords per year in ICSE publications into a stacked area chart. We first counted and extracted the 10 most

frequently occurring keywords, and collected these by year. Using the X-Axis for year, and the Y-Axis for the number of occurrences of the keywords, we mapped out the points, and then connected them using a spline with colored area.

For Task 4a, we represented authors as nodes in a graph, and drew the relations between authors and co-authors by connecting them with lines if they had written a paper together. In addition, we defined the attractive force between authors as high if they were co-authors, and low if one had referenced another. This gave us a force-directed graph representing the relation of authors.

For Task 4b, we represented authors as a cell in a treemap, with the relative size of the cell determined by the number of publications originating from that author.

Interestingly, we observed a phenomena that is well documented as Price's Law. Price's Law states that half the publications in any given field originate from a square root of all contributors. In our exploration of the data, we found that the relation between the number of authors and the number of papers was exponential in nature, and thus roughly half the publications are indeed written by a square root of the number of authors. In fact, when we observed this when we were exploring with only a subset of the data, but this held true when we used the full set of ICSE authors, which illustrates that Price's Law holds even at scale.

3 Other Information