Accountability in the Information Age: A Study of Data Transparency in Local Governments

Ameet Rahane
Data for Good Grantee, CTSP
Applied Mathematics & Cognitive Science
University of California, Berkeley
Email: ameetrahane@berkeley.edu

Clara DeMartel
Data for Good Grantee, CTSP
Applied Mathematics & Political Economy
University of California, Berkeley
Email: claradm@berkeley.edu

Aura Barrera
Data for Good Grantee, CTSP
Interdisciplinary Studies
University of California, Berkeley
Email: aurabarrera@berkeley.edu

Akhil Jalan
Data for Good Grantee, CTSP
Applied Mathematics & Computer Science
University of California, Berkeley
Email: akhiljalan@berkeley.edu

ABSTRACT

Access to public data is vital to transparency and accountability of governments. While some cities, such as New York, have taken major strides towards this goal, there is no unified framework for evaluating openness of data at a local level. We address this problem by scoring openness of city data based on several factors, including ease of access, amount of data, recency, whether the data are up to date, etc. Using this metric, we aim to analyze and compare cities including Berkeley, Oakland, Santa Rosa, San Mateo and Palo Alto. This project should be a tool for citizens to hold their local officials accountable and demand better practices.

Nomenclature

S_T The total score, without normalization.
S_N The total score, normalized
A_α The accessibility subscore, with α as the variable for the subsection of the accessibility subscore
D_δ The Data Quality subscore, with δ as the variable for the subsection of the data quality subscore
F_ζ The Data Format subscore, with ζ as the variable for the subsection of the data format subscore

1 Introduction

This section will include information on why this is necessary and what question we are trying to answer

2 Methods

2.1 General

The core of the metric is built off of a web-based graph. That is, a graph is generated for each city website that contains the URL of each link throughout the graph of the website. This graph would begin with the main website page (e.g. www.cityofberkeley.info/) and continue until it either leaves the domain of the local city or it reaches a dataset page. In order the make these stopping functions well defined, the cities that were looked at were constrained to cities that use the Socrata Open Data Platform. This enabled the use of metadata within the text of the HTML and XML documents associated with the websites to find the appropriate stopping function. This graph enables the collection of information, such as the number of datasets and the amount of links that point at datasets, much simpler. Such a graph would have the basic structure as follows:

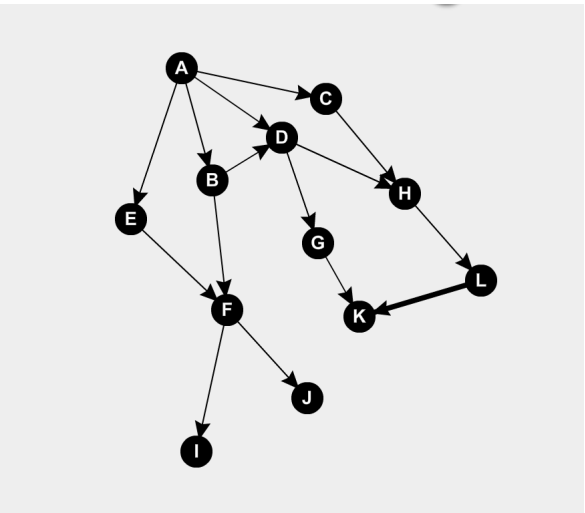

Fig. 1. Example of a website graph

All information about the open data platform either came from the structure of the graph, important features of the graph, the data itself, and miscellaneous other factors.

2.2 Accessibility

2.3 Data Quality

2.3.1 Amount of Data

2.3.2 Format of Data

2.4 Normalization

3 Results

4 Limitations

5 Discussion

6 Conclusion

References