

CS-E4870

Research Project in Machine Learning and Data Science

Ananth Mahadevan

Department of Computer Science, Aalto University
ananth.mahadevan@aalto.fi

Abstract

1 Introduction

- how wikipedia is a large encyclopedia
- maintained by a small group of Administrators
- They undergo an election like process of RfAs
- How this is an important online social election framework
- How it has been studied in previous works
- What we aim to do by using a social network and theories of democracy

Wikipedia is the largest online encyclopedia containing over 5 million pages of content. It is one of the most popular websites on the Internet. Wikipedia has a diverse collection of articles from many different topics and is constantly being updated. Although Wikipedia started out as an open platform where anyone could create and edit articles, this led to many factual errors and biased articles. Wikipedia started to incorporate elements of hierarchy gradually over time. In the English version of Wikipedia all editors need to have a registered account and pages that are controversial and of a sensitive nature are protected by administrators.

Administrators are editors who are given access to tools such as blocking and unblocking other users, deleting and undeleting pages, protecting and renaming pages etc. Any user can **Request for Adminship**(RfA) in which the Wikipedia community participates. The RfA spans over seven days, during which any editor can comment and discuss the candidate. Editors scrutinize the candidate's contributions and credentials as well as their conduct in the online discussion and overall experience. They can then state either their support

or opposition to the candidate along with comments. At the end of seven days a Bureaucrat (an editor higher up in the hierarchy) decides on the consensus of the election and declares the outcome. Consensus is not a direct majority voting scheme and the final call rests with the Bureaucrat.

The RfA is a very intense and selective process, there are only 1400 total administrators of which only 500 are currently active¹. This is out of 38 million registered editors with only around 130 thousand being regular contributors. This small group of active administrators and editors are responsible for creating and maintaining all articles on Wikipedia.

Therefore the RfA process can give us valuable insight into the dynamics of social interactions and elections in an online platform. In this paper we will first discuss the existing work on studying the RfA elections and other such similar online processes. Next we provide an overview of the data collected and used from Wikipedia in this paper. We then present our main contribution, the use of a *Viscous Democracy* to model the RfA election process. We discuss the results and possible extensions of this framework to other online elections systems.

2 Literature review

- election prediction using candidate stats
- election analysis using voter and candidate info
- prediction using communication and how close
- Signed edge prediction and difficulties

¹all data as of March 2020 for English version Wikipedia

3 Dataset

- explain RfA data collection
 - existing SNAP data and limitations
 - XML parsing
 - regex and string matching
 - date parsing
- Social interactions
 - User contributions
 - wealth and diversity of info
 - creating underlying network

4 Viscous Democracy

Brief explanation of viscous democracy

5 Proposed Model

Use viscous democracy models using heuristic delegation functions on social network to predict elections separately

6 Implementation

directed graph concepts and delegation function considerations. Agony and hierarchy. local and global top k delegates.

7 Results

The quality of predictions using local or global important editors.

8 Conclusions

How we can instead try and model individual voter behaviour. Find a more robust ML framework to learn an optimal delegation function.

References