

# Author Disambiguation

Ekaterina Bastrakova, Rodney Ledesma, Jose Millan

April 19, 2016

## Abstract

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nunc rutrum, turpis porttitor volutpat accumsan, lectus tellus malesuada mauris, laoreet elementum sapien lacus ut sem. Nullam vel mauris at eros pretium semper. Pellentesque egestas sapien eu efficitur vehicula. In aliquam vel turpis eleifend semper. Cras luctus nec leo sit amet volutpat. Cras ultricies vulputate augue, in pharetra mauris pulvinar et. Nullam velit est, tincidunt eu cursus nec, maximus quis metus. Pellentesque gravida dui eu augue sagittis, nec tristique lorem fermentum. Mauris eleifend ipsum at euismod faucibus. Donec sit amet faucibus ipsum. Vivamus lacinia ex nisi, quis laoreet nisl sagittis sit amet. Duis rutrum lacus id sapien accumsan dignissim.

## 1 Introduction

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nunc rutrum, turpis porttitor volutpat accumsan, lectus tellus malesuada mauris, laoreet elementum sapien lacus ut sem. Nullam vel mauris at eros pretium semper. Pellentesque egestas sapien eu efficitur vehicula. In aliquam vel turpis eleifend semper. Cras luctus nec leo sit amet volutpat. Cras ultricies vulputate augue, in pharetra mauris pulvinar et. Nullam velit est, tincidunt eu cursus nec, maximus quis metus. Pellentesque gravida dui eu augue sagittis, nec tristique lorem fermentum. Mauris eleifend ipsum at euismod faucibus. Donec sit amet faucibus ipsum. Vivamus lacinia ex nisi, quis laoreet nisl sagittis sit amet. Duis rutrum lacus id sapien accumsan dignissim.

## 2 Related Work

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nunc rutrum, turpis porttitor volutpat accumsan, lectus tellus malesuada mauris, laoreet elementum sapien lacus ut sem. Nullam vel mauris at eros pretium semper. Pellentesque egestas sapien eu efficitur vehicula. In aliquam vel turpis eleifend semper. Cras luctus nec leo sit amet volutpat. Cras ultricies vulputate augue, in pharetra mauris pulvinar et. Nullam velit est, tincidunt eu cursus nec, maximus quis metus. Pellentesque gravida dui eu augue sagittis, nec tristique lorem fermentum. Mauris eleifend ipsum at euismod faucibus. Donec sit amet faucibus ipsum. Vivamus lacinia ex nisi, quis laoreet nisl sagittis sit amet. Duis rutrum lacus id sapien accumsan dignissim.

## 3 Implementation

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nunc rutrum, turpis porttitor volutpat accumsan, lectus tellus malesuada mauris, laoreet elementum sapien lacus ut sem.

### 3.1 Source Data Description

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nunc rutrum, turpis porttitor volutpat accumsan, lectus tellus malesuada mauris, laoreet elementum sapien lacus ut sem. Nullam vel mauris at eros pretium semper. Pellentesque egestas sapien eu efficitur vehicula. In aliquam vel turpis eleifend semper. Cras luctus nec leo sit amet volutpat. Cras ultricies vulputate augue, in pharetra mauris pulvinar et. Nullam velit est, tincidunt eu cursus nec, maximus quis metus. Pellentesque gravida dui eu augue sagittis, nec tristique lorem fermentum. Mauris eleifend ipsum at euismod faucibus. Donec sit amet faucibus ipsum. Vivamus lacinia ex nisi, quis laoreet nisl sagittis sit amet. Duis rutrum lacus id sapien accumsan dignissim.

### 3.2 Process Flow

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nunc rutrum, turpis porttitor volutpat accumsan, lectus tellus malesuada mauris, laoreet elementum sapien lacus ut sem. Nullam vel mauris at eros pretium semper. Pellentesque egestas sapien eu efficitur vehicula. In aliquam vel turpis eleifend semper. Cras luctus nec leo sit amet volutpat. Cras ultricies vulputate augue, in pharetra mauris pulvinar et. Nullam velit est, tincidunt eu cursus nec, maximus quis metus. Pellentesque gravida dui eu augue sagittis, nec tristique lorem fermentum. Mauris eleifend ipsum at euismod faucibus. Donec sit amet faucibus ipsum. Vivamus lacinia ex nisi, quis laoreet nisl sagittis sit amet. Duis rutrum lacus id sapien accumsan dignissim.

### 3.3 Features Implemented

As mentioned previously in subsection 3.2, for every pair of authors compared, a set of features is calculated. In table 1 features chosen for the current work can be appreciated.

Feature Name	Calculation Description
First Name Initial	Equality in first name initial.
Second Name Initial	Equality in middle name initial.
Topic	Equality in LDA-determined topic
Publication Year Difference	Absolute difference between the publication years.
Keywords Distance	Jaccard distance between the keywords.
References Distance	Jaccard distance between the references.
Subject Distance	Jaccard distance between the subjects.
Title Distance	Jaccard distance between the titles.
Coauthors Distance	Jaccard distance between the coauthors.
Ethnicity/Race	???

Table 1: Features calculated for every pair of authors

Considering the complexity of the calculation of some features, a more in depth description can be found.

#### 3.3.1 LDA Topic

#### 3.3.2 Jaccard Distance

#### 3.3.3 Ethnicity

## 4 Results

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nunc rutrum, turpis porttitor volutpat accumsan, lectus tellus malesuada mauris, laoreet elementum sapien lacus ut sem. Nullam vel mauris at eros pretium semper. Pellentesque egestas sapien eu efficitur vehicula. In aliquam vel turpis eleifend semper. Cras luctus nec leo sit amet volutpat. Cras ultricies vulputate augue, in pharetra mauris pulvinar et. Nullam velit est, tincidunt eu cursus nec, maximus quis metus. Pellentesque gravida dui eu augue sagittis, nec tristique lorem fermentum. Mauris eleifend ipsum at euismod faucibus. Donec sit amet faucibus ipsum. Vivamus lacinia ex nisi, quis laoreet nisl sagittis sit amet. Duis rutrum lacus id sapien accumsan dignissim.

## 5 Conclusions

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nunc rutrum, turpis porttitor volutpat accumsan, lectus tellus malesuada mauris, laoreet elementum sapien lacus ut sem. Nullam vel mauris at eros pretium semper. Pellentesque egestas sapien eu efficitur vehicula. In aliquam vel turpis eleifend semper. Cras luctus nec leo sit amet volutpat. Cras ultricies vulputate augue, in pharetra mauris pulvinar et. Nullam velit est, tincidunt eu cursus nec, maximus quis metus. Pellentesque gravida dui eu augue sagittis, nec tristique lorem fermentum. Mauris eleifend ipsum at euismod faucibus. Donec sit amet faucibus ipsum. Vivamus lacinia ex nisi, quis laoreet nisl sagittis sit amet. Duis rutrum lacus id sapien accumsan dignissim.