

**DSC 180A Data Science Project I Section A02**

# **Wikipedia Edit Wars**

Keng-Chi Chang  
<kechang@ucsd.edu>

2020-01-08

# Outline

1. Overview of the domain
2. A bit about me
3. A bit about you
4. Discussions

# **1/ Overview of the domain**

# Background



- Largest collection of human knowledge
- Open: Everyone can access, everyone can contribute



# Talk:Hummus

From Wikipedia, the free encyclopedia

## WARNING: ACTIVE ARBITRATION REMEDIES

The article [Hummus](#), along with other articles relating to the [Arab–Israeli conflict](#), remedies resulting from three arbitration cases (see [WP:A/I/PIA](#)). The current res



- **Editing restrictions for new editors:** All IP editors, accounts with fewer than 30 days tenure are prohibited from editing any page that could be reasonably Israeli conflict.
- **Limit of one revert in 24 hours:** All articles related to the Arab–Israeli conflict [WP:1RR](#) (one [revert](#) per editor per article *per 24-hour period*). When in doubt
- **All Arab–Israeli conflict-related pages, broadly interpreted, are subject to** uninvolved administrator may levy restrictions as an [arbitration enforcement](#) after an initial notification.

# Potato chips

*Should potato chips be flavored or flavoured?*

I honestly don't see why this is here, since Canadians, who primarily use *-our* spellings, also call them

The dispute wasn't over whether to call them potato chips, it was over whether to use the "-our" or "flavoured" listed here. :) [Bryan](#) 03:25, 6 Apr 2005 (UTC)

Lame indeed. Why use the word at all if it can be avoided? It's not even necessary, just call them

The issue had nothing to do with specific flavours, but to do with the invention and process.

The problem is this: the United States, where the "true" potato chip was invented, is not English, a "potato chip" is a french fry!

And we descend into it again. Lets just drop it... its a listed edit war, leave it at that. -

You are wrong. It's as simple as that. In this context, in commonwealth English, *flavour* is ENGLISH wikipedia, so the ENGLISH spelling should be used. "Flavor" is not

Using specific flavour names wouldn't have been workable in the slightest--[Kiand](#) 12:37, 11

# Top 12 controversial articles in English Wiki (2013)

- George\_W.\_Bush
- Anarchism
- Muhammad
- List\_of\_World\_Wrestling\_Entertainment\_employees
- Global\_warming
- Circumcision
- United\_States
- Jesus
- Race\_and\_intelligence
- Christianity
- Michael\_Jackson
- Barack\_Obama



# How "open" is good?

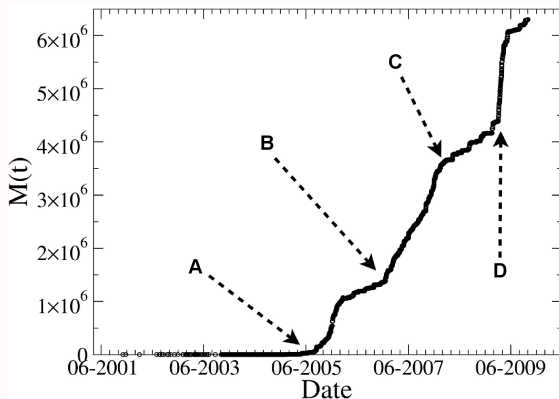
- Camp A: Allowing unregistered editors to edit
  - Helps grow the community
  - Better arguments may survive the debate
- Camp B: Tighter restrictions on editing
  - Produce a higher-quality resource
  - Attract a better class of editors
- We face a very similar debate on social media and Internet freedom today

# Preview of the techniques

- Parse large dumps of xml (from publicly available data)
- NLP (any language of your choice)
- Time series and causal inference
- Network analysis (among articles or editors)

# Sketch of the replication

Controversy measure of the article “Michael Jackson”



- Update the list of controversial articles of any language to today
- Build the fundamentals for your project, know the data

# Possible directions for project proposals

- The proposal is NOT limited to be about edit wars
  - ▶ The replication familiarize you with the data
  - ▶ But can work on anything cool using Wikipedia data
- Examples
  - ▶ What predicts edit wars?
  - ▶ Do editors coordinate to change certain topics?
  - ▶ How effective are Wikipedia's certain editing restrictions?
  - ▶ Algorithms to judge edit wars vs. human judgments

# Logistics

- Materials all on Github repo
- Discussion: Wed 9-10 WLH 2113
  - Discuss the readings and the domain assignments
  - Come, since there are 10% for participation
- Lab / Office hour: Fri 9-11 CSE Basement
  - Please sign up in advance:  
<https://calendly.com/kengchichang/dsc180a>
  - If you really cannot make it, email for appointment
- Will create a group on Piazza for questions in this section
- Might touch on some issues that are controversial
- Be open-minded and respectful for other's opinions

## **2/ A bit about me**

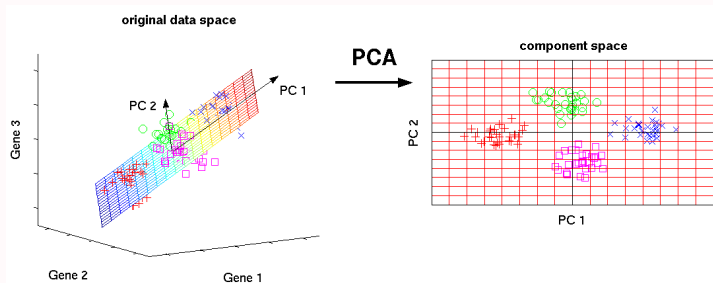
## Keng-Chi Chang (張耕齊)

kechang@ucsd.edu · kengchichang.com

- Born and raised in Taipei, Taiwan
- Studied economics during undergrad
- 2nd year PhD student in political science
- Research interests
  - ▶ Computational social science
  - ▶ Political methodology
  - ▶ Causal inference

# Some Past Research: Ideology on Facebook

- Data: 16 billion likes from 2K US fan pages in 2015-16
- Algorithm: SVD/PCA on shared user matrix



- Intuition: Pages share similar users should have similar ideology



DW-Nominate Score of 115th Congress

- × Democratic Party
- Independent
- Republican Party

Cruz

Rubio

McCain

Ryan

McConnell

Schumer

Pelosi

Sanders

Warren

Booker

$$\rho = 0.90$$

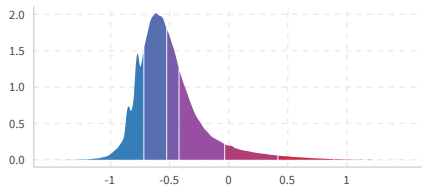
$$\rho_R = 0.52$$

$$\rho_D = 0.09$$

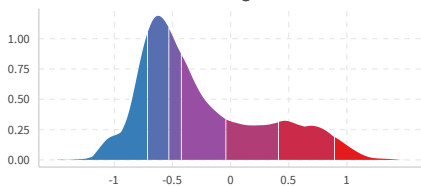
Estimated Facebook Page Ideology Score, 2015-01 to 2016-11

Using only politician page matrix (Bond and Messing 2015)

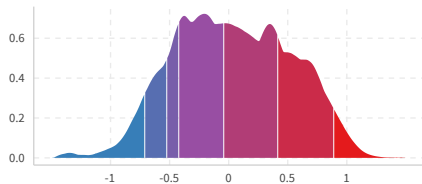
Massachusetts



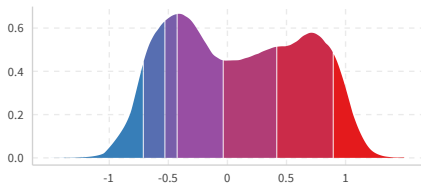
Washington



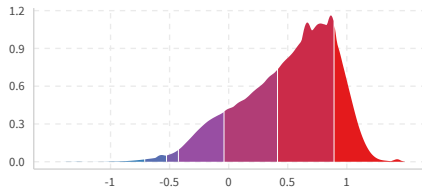
Michigan



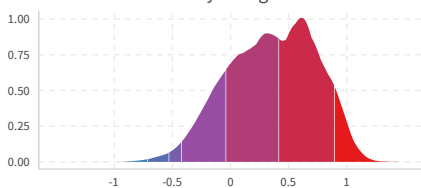
Pennsylvania



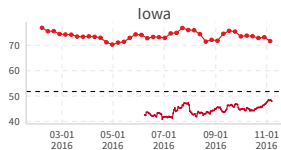
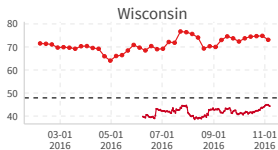
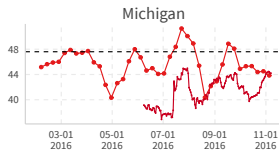
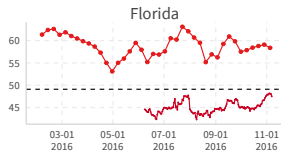
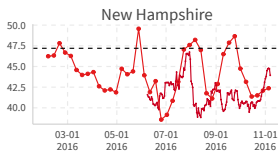
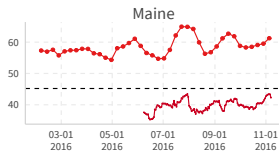
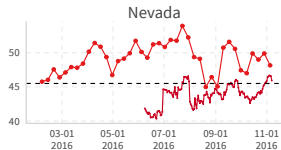
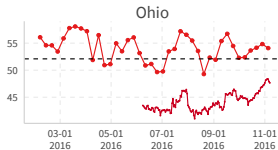
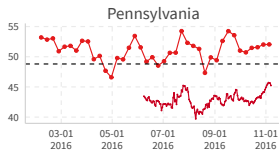
Texas



Wyoming



# FB Support Share Trends with Polls



## **3/ A bit about you**

# **4/ Discussions**

Go to `www.menti.com` and enter code 16 73 65



- Democracy's Dilemma
- Mark Zuckerberg's speech
- Wikimedia data archives