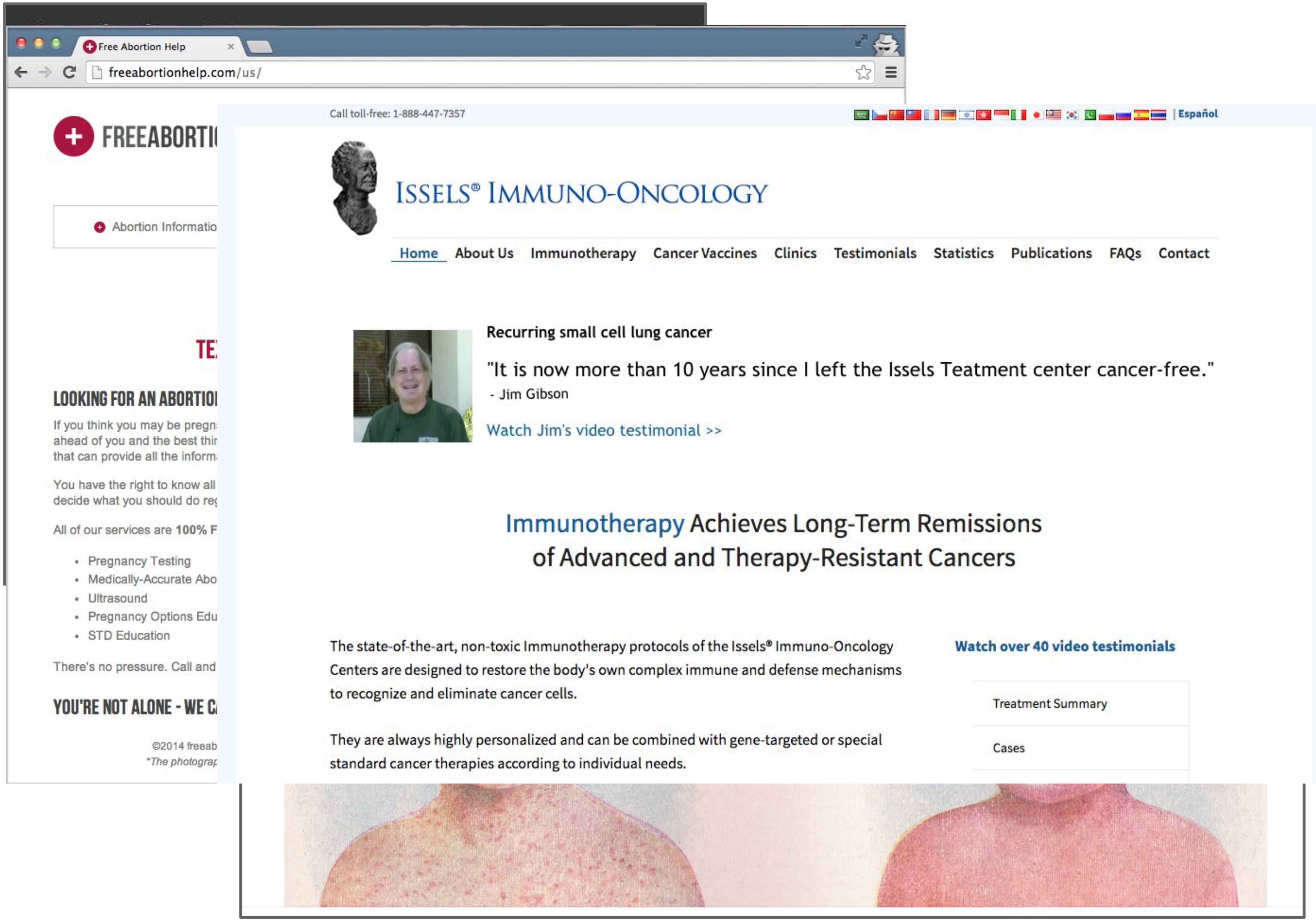
Controversy Detection and Analysis

Dr. Shiri Dori-Hacohen
AuCoDe
(+ University of Massachusetts Amherst)

Signal Media - December 2017

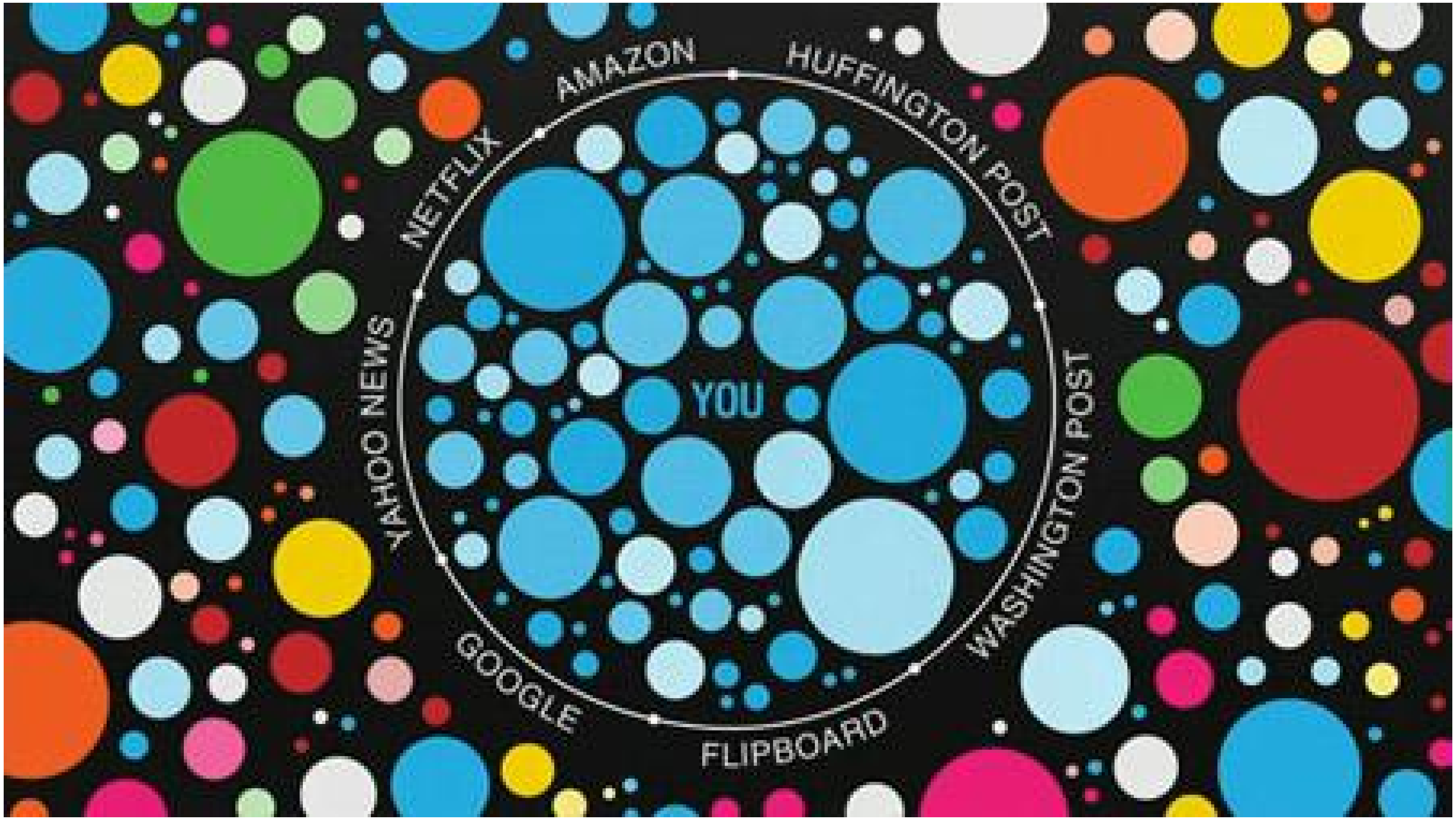
Presenting joint work with James Allan, David Jensen, Elad Yom-Tov, Myung-ha Jang, John Foley


Free Abortion Help
freeabortionhelp.com/us/
FREEABORTI
Abortion Informatio
TE
LOOKING FOR AN ABORTIO
If you think you may be pregn
ahead of you and the best thi
that can provide all the inform
You have the right to know all
decide what you should do re
All of our services are 100% F
Pregnancy Testing
Medically-Accurate Abo
Ultrasound
Pregnancy Options Edu
STD Education
There's no pressure. Call and
YOU'RE NOT ALONE - WE C
©2014 freeab
*The photogra
Call toll-free: 1-888-447-7357
Español
ISSELS® IMMUNO-ONCOLOGY
Home
About Us
Immunotherapy
Cancer Vaccines
Clinics
Testimonials
Statistics
Publications
FAQs
Contact
Recurring small cell lung cancer
"It is now more than 10 years since I left the Issels Teatment center cancer-free."
- Jim Gibson
Watch Jim's video testimonial >>
Immunotherapy Achieves Long-Term Remissions
of Advanced and Therapy-Resistant Cancers
The state-of-the-art, non-toxic Immunotherapy protocols of the Issels® Immuno-Oncology Centers are designed to restore the body's own complex immune and defense mechanisms to recognize and eliminate cancer cells.
They are always highly personalized and can be combined with gene-targeted or special standard cancer therapies according to individual needs.
Watch over 40 video testimonials
Treatment Summary
Cases

Growing interest in detecting controversy computationally

The Concerns

The Concerns



The Filter Bubble

Pariser, E. 2011

The Concerns

Misinformation

Overview

- Motivation
- Related Work
- Controversy Detection in Wikipedia Using Collective Classification
- Contributions to Controversy Detection
  - On the web (using Wikipedia, language models)
  - Position paper (social, ethical, technical challenges)
  - Contention (population-based mathematical model)
- Startup - AuCoDe

Controversy on the Web & search

- Only in domain-specific areas
  - News (Choi et al., 2010, Awadallah et al. 2011, Mejova et al., 2014)
  - Twitter (Popescu & Pennacchiotti, 2010)
- Controversial query detection (Gyllstrom & Moens, 2011)
- Controversy detection problem in the web didn't exist
  - Only specific sub-instances of it
  - Wasn't treated as a general issue
- Prior work focused almost exclusively on political controversies (using Debatepedia)

Sentiment Analysis vs. Controversy

- Sentiment analysis seen as a step towards detecting varying opinions/controversy
  - cf. Choi et al., 2010; Cartright et al., 2009
- Other work shows sentiment & controversy are overlapping, but not identical, constructs
  - Dori-Hacohen & Allan, 2013; Mejova et al., 2014
- Sentiment analysis may be more effective when considering its variance in analyzing online conversations, rather than when examining individual webpages

Controversy Detection in Wikipedia

- Where everything started
- Kittur et al., 2007 - First classifier for controversy in Wikipedia articles
- Sumi et al., 2011; Yasseri et al., 2012 - Using the concept of edit wars and reverts; Heuristic approach
- Sepehri Rad & Barbosa, Sepehri Rad et al. 2012 - Using collaboration networks between authors; Algorithm was computationally intensive, impractical
- Jankowski-lorek et al., 2015 - article feedback tool
- Jesus et al., 2009 - Clusters of controversial pages (anecdotally)

- Either machine learning or heuristic approaches
- Generally - classify each page in isolation

Overview

- Motivation
- Related Work
- Controversy Detection in Wikipedia Using Collective Classification
- Contributions to Controversy Detection
  - On the web (using Wikipedia, language models)
  - Position paper (social, ethical, technical challenges)
  - Contention (population-based mathematical model)
- Startup - AuCoDe

Controversy Detection in Wikipedia Using Collective Classification

Published in SIGIR 2016

Joint work with David Jensen & James Allan

Controversy Detection in Wikipedia

Using Collective Classification

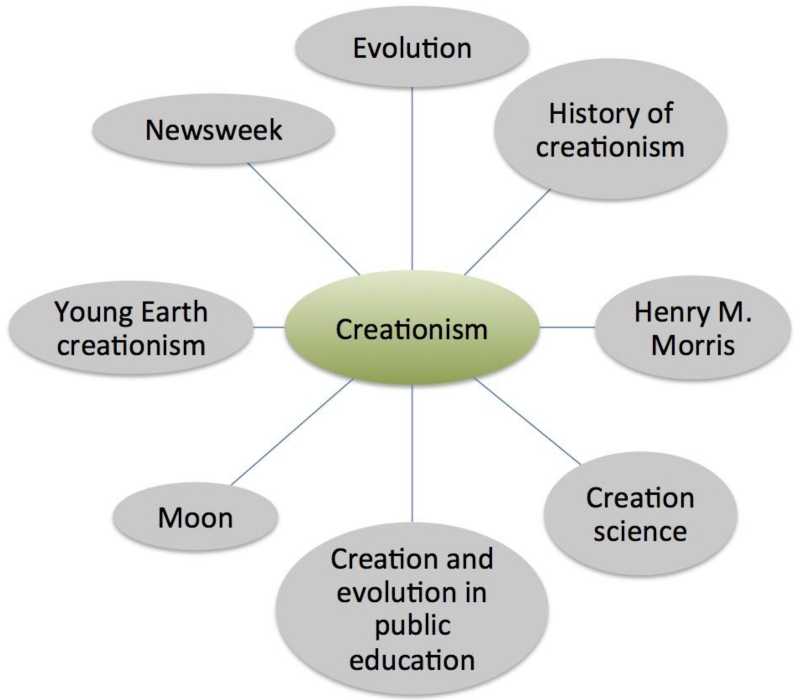
Creationism

https://en.wikipedia.org/wiki/Creationism

Prior work on automated controversy detection in Wikipedia has focused on pages in isolation

Dori-Hacohen, S., Jensen, D. and Allan, J. SIGIR 2016.

Controversy Detection in Wikipedia
Using Collective Classification



Hypothesis: related Wikipedia pages might have similar amount of controversy (homophily)

Dori-Hacohen, S., Jensen, D. and Allan, J. SIGIR 2016.

Collective & Stacked Classification

- Collective Inference is a technique which leverages homophily between related instances for inference
- However, it generally requires availability of labeled data for neighbors
- In our case, labeled data is sparse
- Stacked inference is an ensemble method which predicts labels for neighbors and then uses them

Approach

Intrinsic Classifier:

Training and inference on features of each WP page as a standalone page

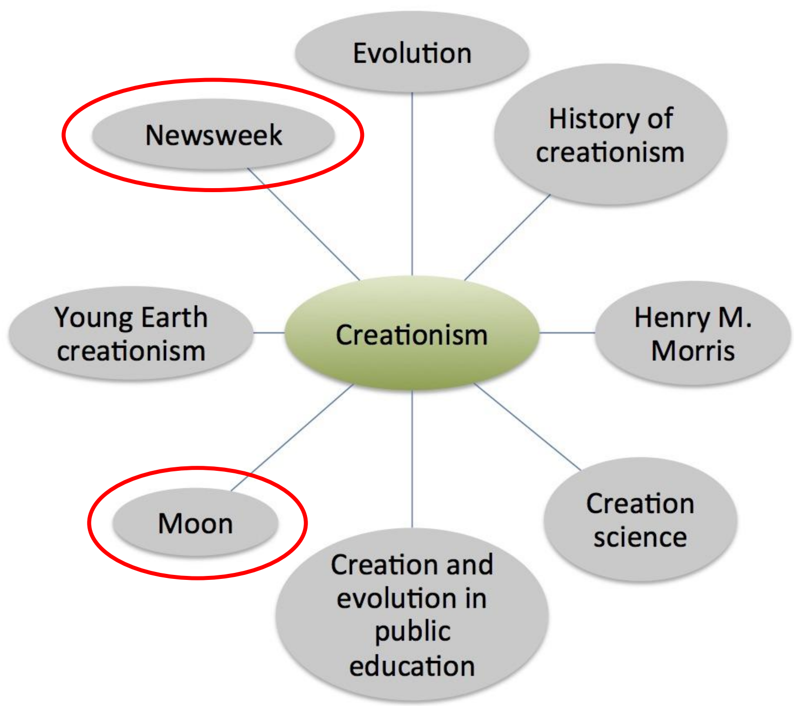
(e.g. Creationism)

Leveraging the graph structure of WP to make the inference better

Dori-Hacohen, S., Jensen, D. and Allan, J. SIGIR 2016.

Bridging knowledge discovery and IR:
A Subnetwork of Neighbors

- Traditionally: neighbors = relational database
- Hypothesis: not all links created equal



- Use text similarity to select neighbors (TF-IDF)

Experimental conditions

Name	Description
Stacked-Ranked-k	Proposed stacked inference system with a similarity-based subnetwork
Stacked-Random-k	A stacked inference system which uses k randomly selected neighbors
Neighbors-Only-k	A classifier based only on the neighbor predictions (as in a regular stacked model), without using the intrinsic features of the center page
Intrinsic	A classifier using only intrinsic features
Stacked-All	A stacked inference system, as above, but which uses all Wikipedia neighbors
Prior work	See Sepehri Rad & Barbosa [12] for details

Cross Validation Procedure

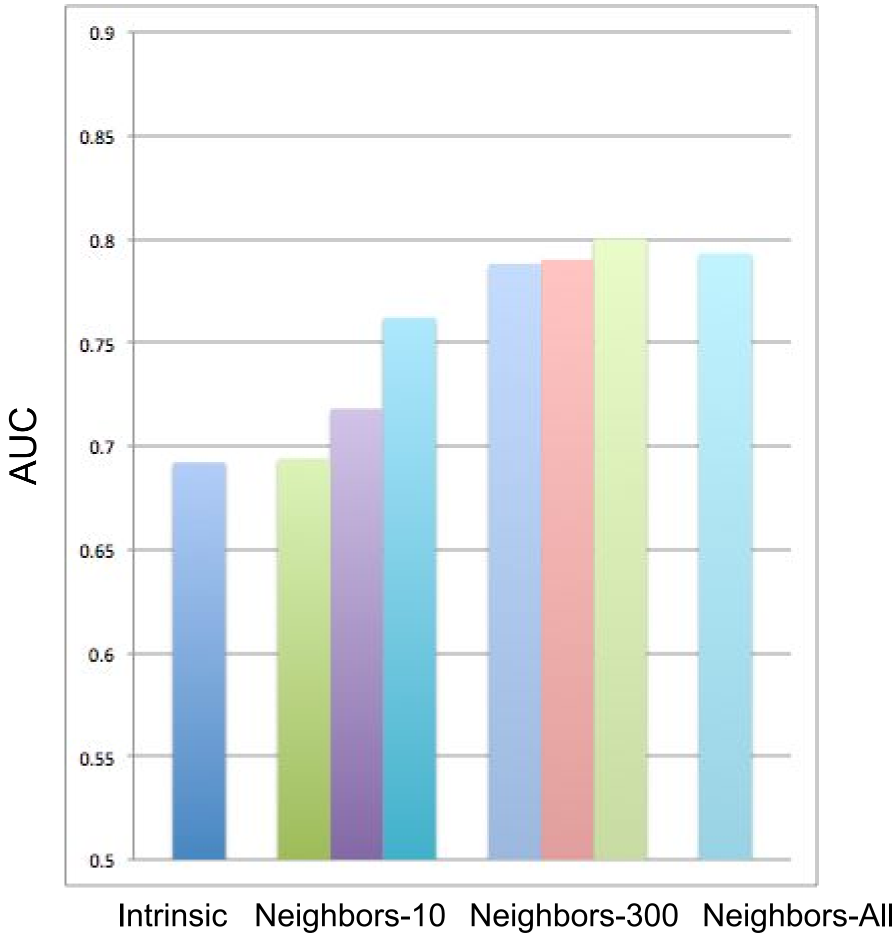
Algorithm 1 Cross-validation stacked training procedure

for fold $i = 1..k$, $Set_i = A \setminus \text{fold}_i$ **do**
  Train IM_i, an intrinsic model on Set_i
  Select $subneighbors(Set_i) \subseteq neighbors(Set_i)$
  Apply IM_i on $subneighbors(Set_i)$
  Aggregate predictions of $subneighbors(Set_i)$ to create an extended feature set, Set'_i
  Train SM_i, a stacked collective model on Set'_i
end for

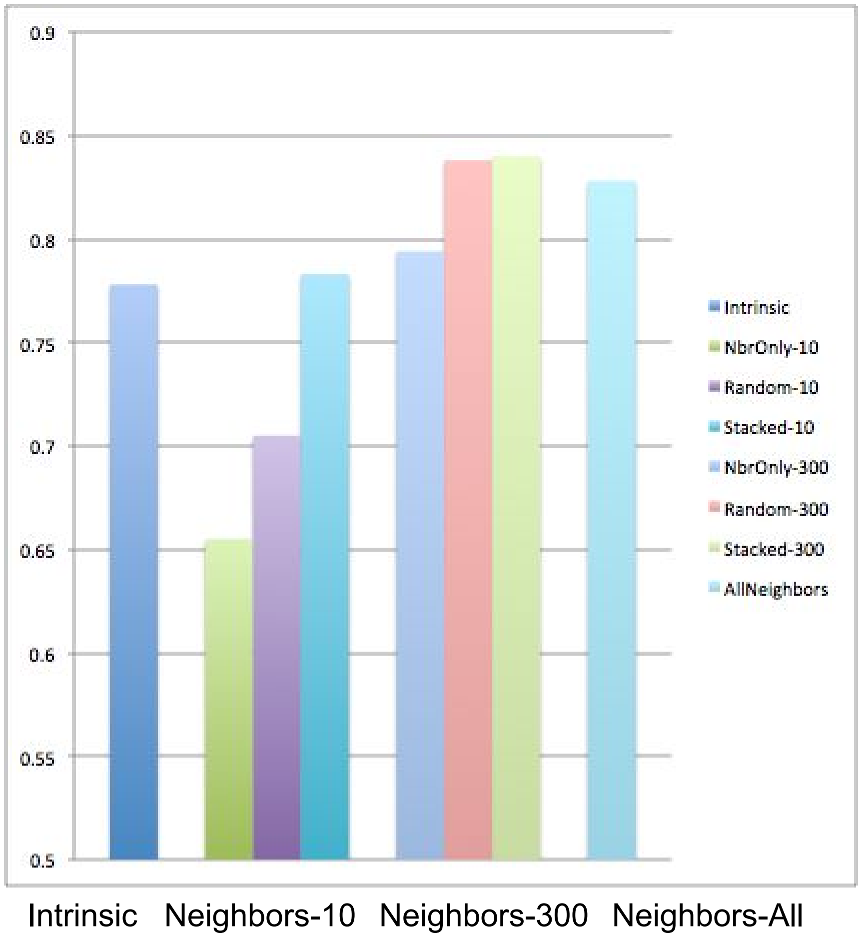
Datasets

Set	Articles	Controversial
DHA [5]	1926	293 (15.2%)
SRMRB [12]	480	240 (50%)

Results - AUC

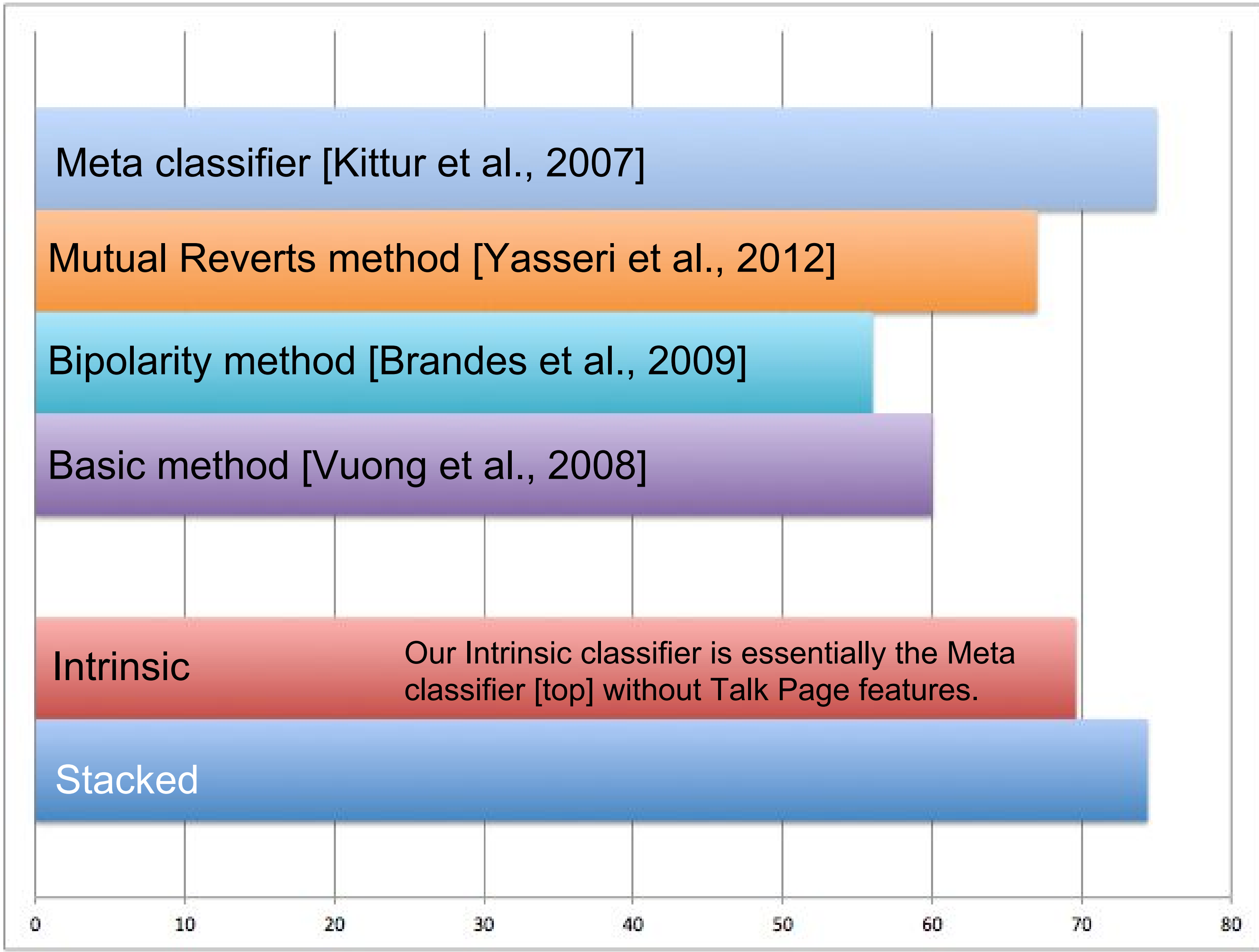


DHA dataset



SRMRB dataset

Results - Accuracy (vs. prior work)



Accuracy

Results - summary

- Similar Neighbors improve results
  - Results increase substantially for first 25 neighbors
  - Stacked classifier outperforms both the Intrinsic and Neighbor-only models
  - Similar is better than Random, esp. w/small # of neighbors; converging as # approaches all neighbors
- Neighbors Provide Quality Inference Without Intrinsic Features
- Stacked Models Outperform Prior Work

So What?

- Leveraging the graph structure in Wikipedia
- Allows one to extend labels to a wider page set
    - Short edit history, no talk pages, low popularity, etc.
- Improved upon state-of-the-art methods
- Agnostic to the choice of intrinsic classifier
    - Any intrinsic classifier for controversy in Wikipedia can be enhanced by applying stacked classifier

Future Directions

- Subnetwork approach can be generalized to other semi-structured problem domains
- Study tradeoff between similarity and inference costs
- Explore other similarity constructions
- Automated detection of controversy holds promise for a variety of applications

Overview

- Motivation
- Related Work
- Controversy Detection in Wikipedia Using Collective Classification
- Contributions to Controversy Detection
    - On the web (using Wikipedia, language models)
    - Position paper (social, ethical, technical challenges)
    - Contention (population-based mathematical model)
- Startup - AuCoDe

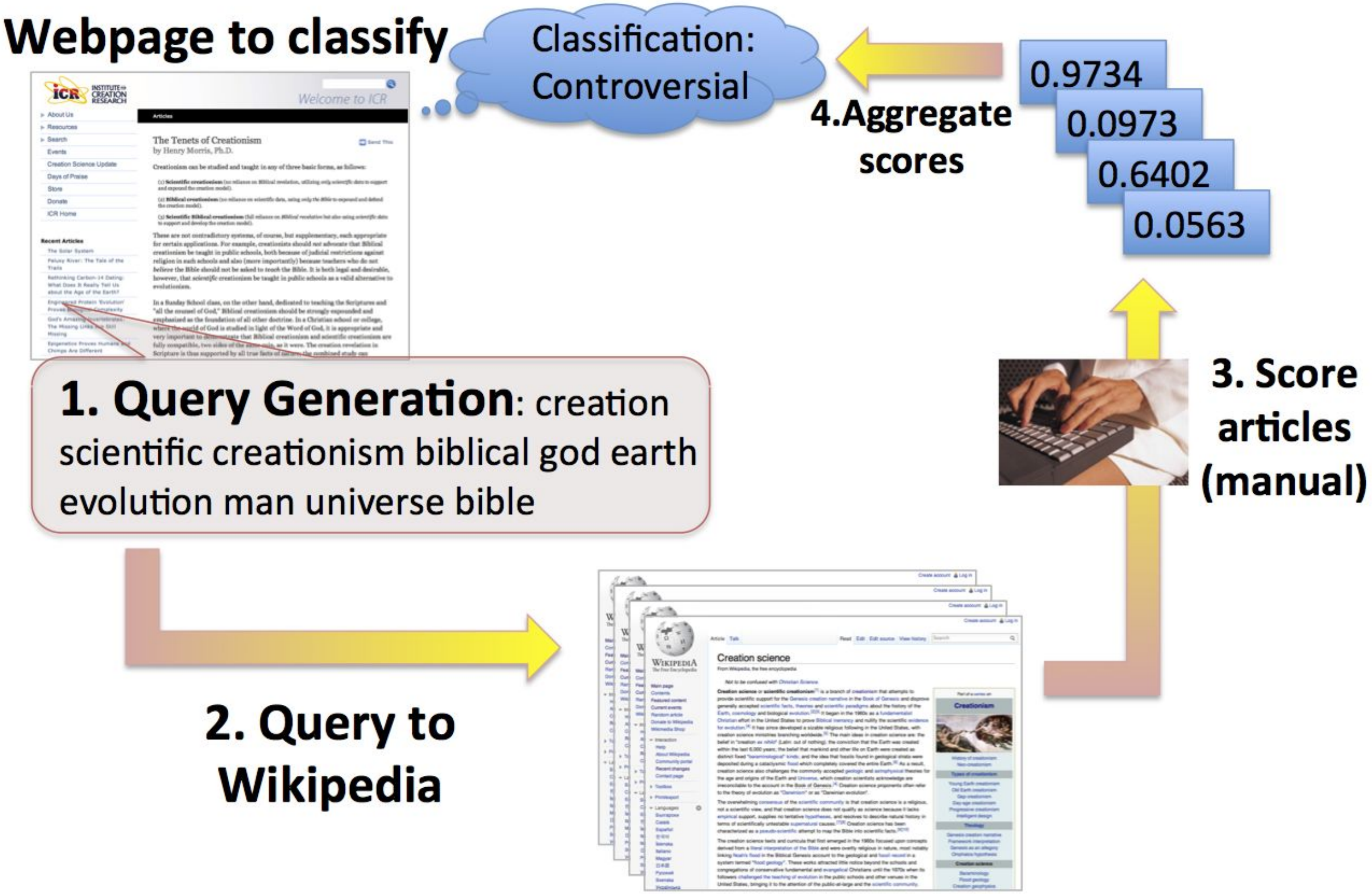
Our work so far

- Improving Controversy Detection in Wikipedia *(Dori-Hacohen, Jensen & Allan; SIGIR 2016)*
- Controversy Detection on the Web *(Dori-Hacohen & Allan; CIKM 2013, ECIR 2015)*
- Probabilistic Approaches to Controversy Detection *(Jang, Foley, Dori-Hacohen & Allan; CIKM 2016)*
- Navigating Controversy as a Complex Search Task *(Dori-Hacohen, Yom-Tov & Allan; SCST workshop, ECIR 2015)*
- Modeling Controversy as Contention Within Populations *(Jang, Dori-Hacohen & Allan; ICTIR 2017)*

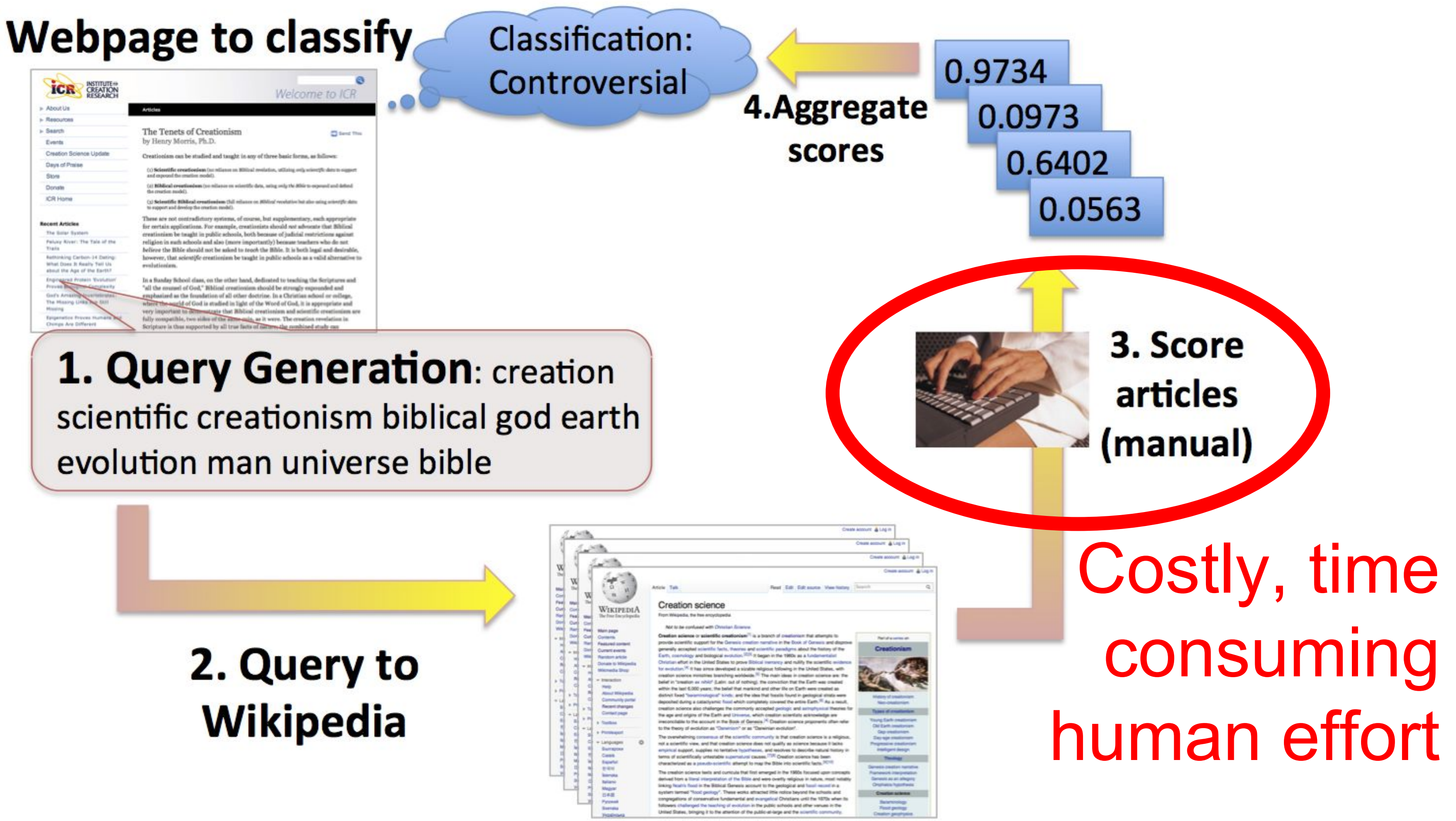
Wikipedia is great, Web is better

- We wanted to extend the work to the web
- But, the rich metadata from Wikipedia is non-existent on the web
- How can we bridge the gap?

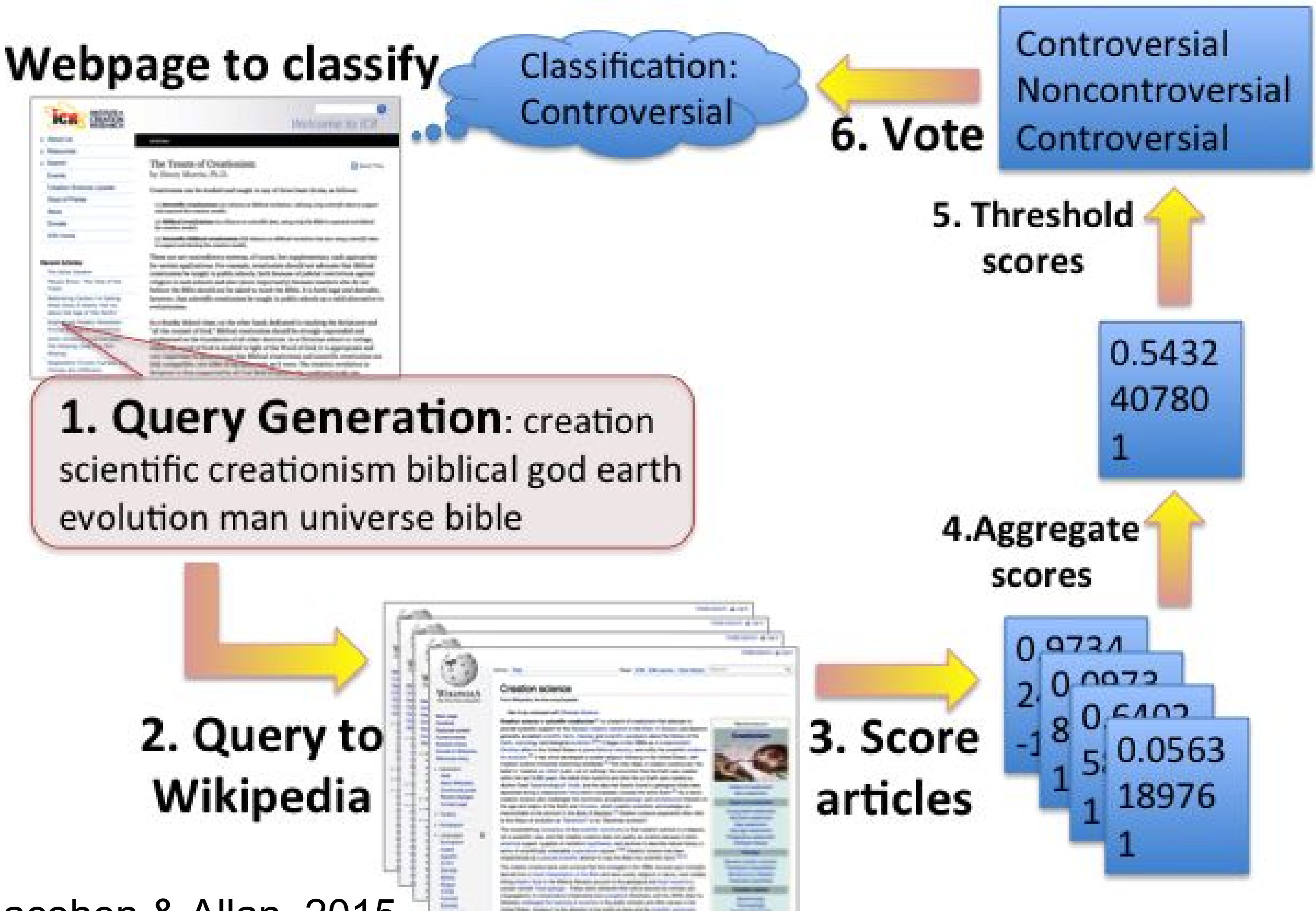
Controversy Detection on the Web



Dori-Hacohen & Allan, CIKM 2013


Webpage to classify
Classification: Controversial
4.Aggregate scores
0.9734
0.0973
0.6402
0.0563
1. Query Generation: creation scientific creationism biblical god earth evolution man universe bible
3. Score articles (manual)
2. Query to Wikipedia
Costly, time consuming human effort

Automated Controversy Detection on the web



Dori-Hacohen & Allan, 2015

Language Models of Controversy

Jang, Foley, Dori-Hacohen, & Allan, CIKM 2016

A theoretical and empirical framework for Language Models of Controversy

$$P(D|C) \approx P(D|L_C) = \prod_{w \in D} (\lambda P(w|L_C) + (1-\lambda)P(w|L_G))$$

$$\log P(D|L_C) = \sum_{w \in D} \log [\lambda P(w|L_C) + (1-\lambda)P(w|L_G)]$$

$$P(D|L_{NC}) \approx P(D|L_{NC}) = \prod_{w \in D} P(w|L_{NC})$$

$$\log P(D|L_{NC}) = \sum_{w \in D} \log P(w|L_{NC})$$

Navigating Controversy as a Complex Search Task

Dori-Hacohen, Yom-Tov & Allan, SCST Workshop, ECIR 2015

Discussing technical, social and ethical challenges of helping users with controversy in search

Overview

- Motivation
- Related Work
- Controversy Detection in Wikipedia Using Collective Classification
- Contributions to Controversy Detection
  - On the web (using Wikipedia, language models)
  - Position paper (social, ethical, technical challenges)
  - Contention (population-based mathematical model)
- Startup - AuCoDe

Definition of Controversy

- *"controversial topics are those that generate strong disagreement among large groups of people."*
    - Operational definition (à la relevance)
- Intuition suggests sentiment (incorrectly!)
- Problematic controversy definitions/datasets (by others)
    - Confounding Wiki vandalism and controversy (Vuong et al., 2008)
    - Using "lamest edit wars" as a controversy dataset (Bykau et al., 2015)

Towards a computational definition

- We were looking for a better definition that could be clearly understood & reproducible
- Inspired by "there is no controversy" arguments (e.g. vaccines/autism)
- How is it possible?

Contention, based on populations

- The big “a-ha” moment: we have to talk about populations
- We define a new term: contention
- Which is a function of topic, AND population
- What’s the probability that two people, randomly selected from the population, will hold conflicting opinions?

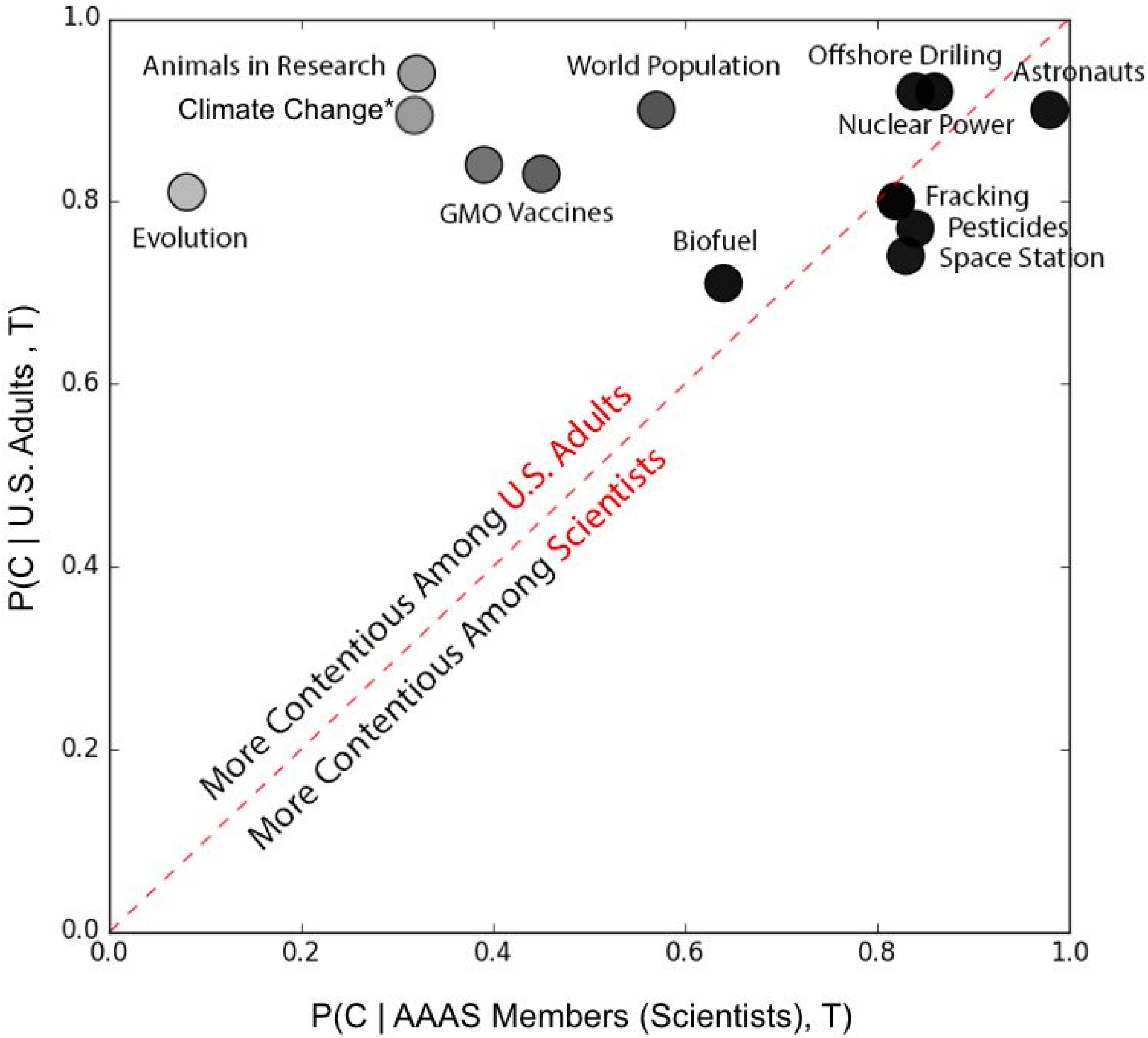
Mathematical model for Contention

$$P(c|\Omega, T) = P(p_1, p_2 \text{ selected randomly from } \Omega, \exists s_i, s_j \in S,$$
$$\text{s.t. } holds(p_1, s_i, T) \wedge holds(p_2, s_j, T) \wedge conflicts(s_i, s_j))$$

We define **stance groups** in the population, which are groups of people that hold the same stance. For $i \in \{0..k\}$, let $G_i = \{p \in \Omega | holds(p, s_i, T)\}$. By construction, $\Omega = \bigcup_i G_i$.

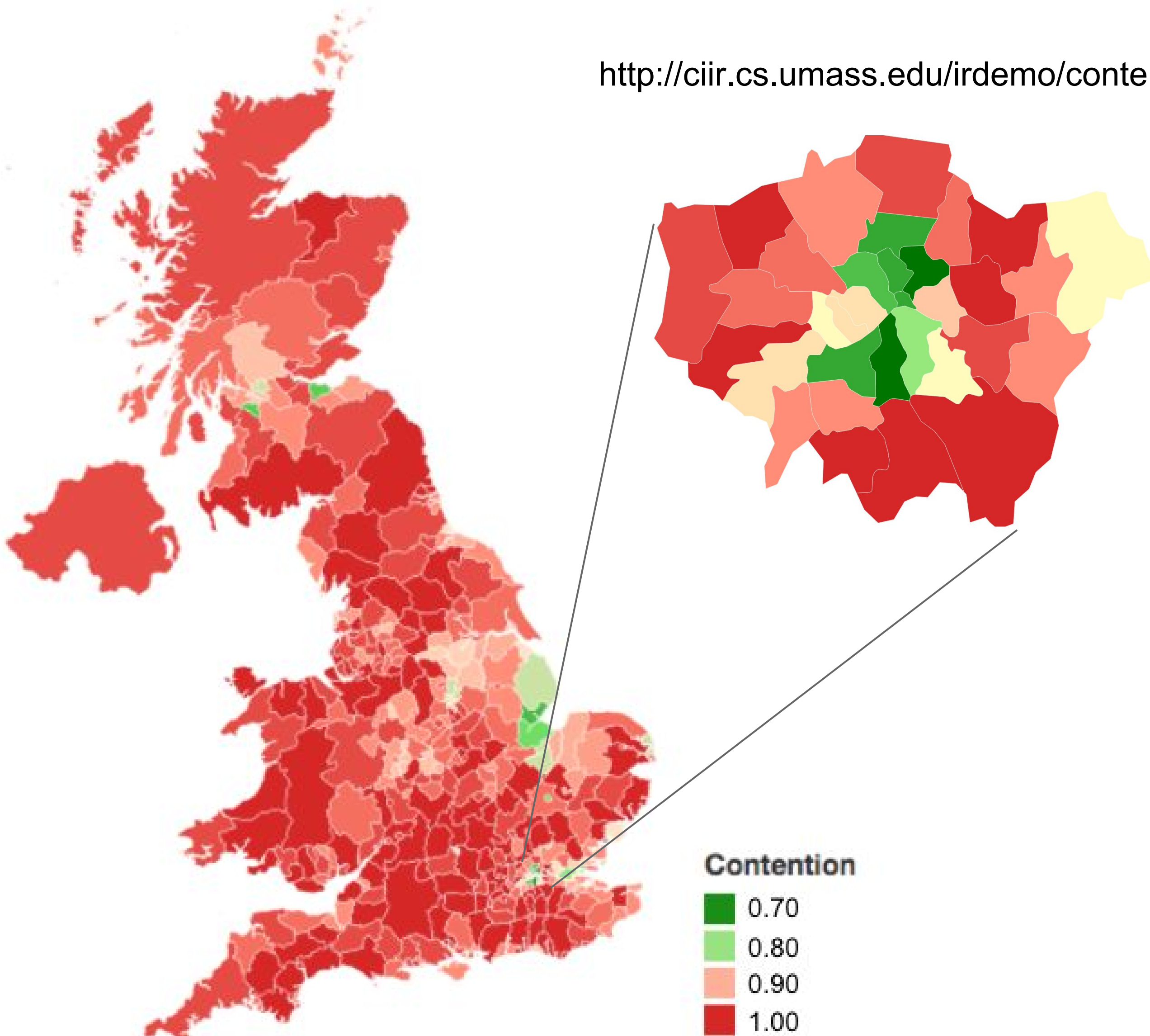
$$P(c|\Omega, T) = \frac{\Sigma_{i \in \{2..k\}} \Sigma_{j \in \{1..i-1\}} (2|G_i||G_j|)}{|\Omega|^2}$$

Selected results - scientists vs. U.S.



Selected results Brexit contention

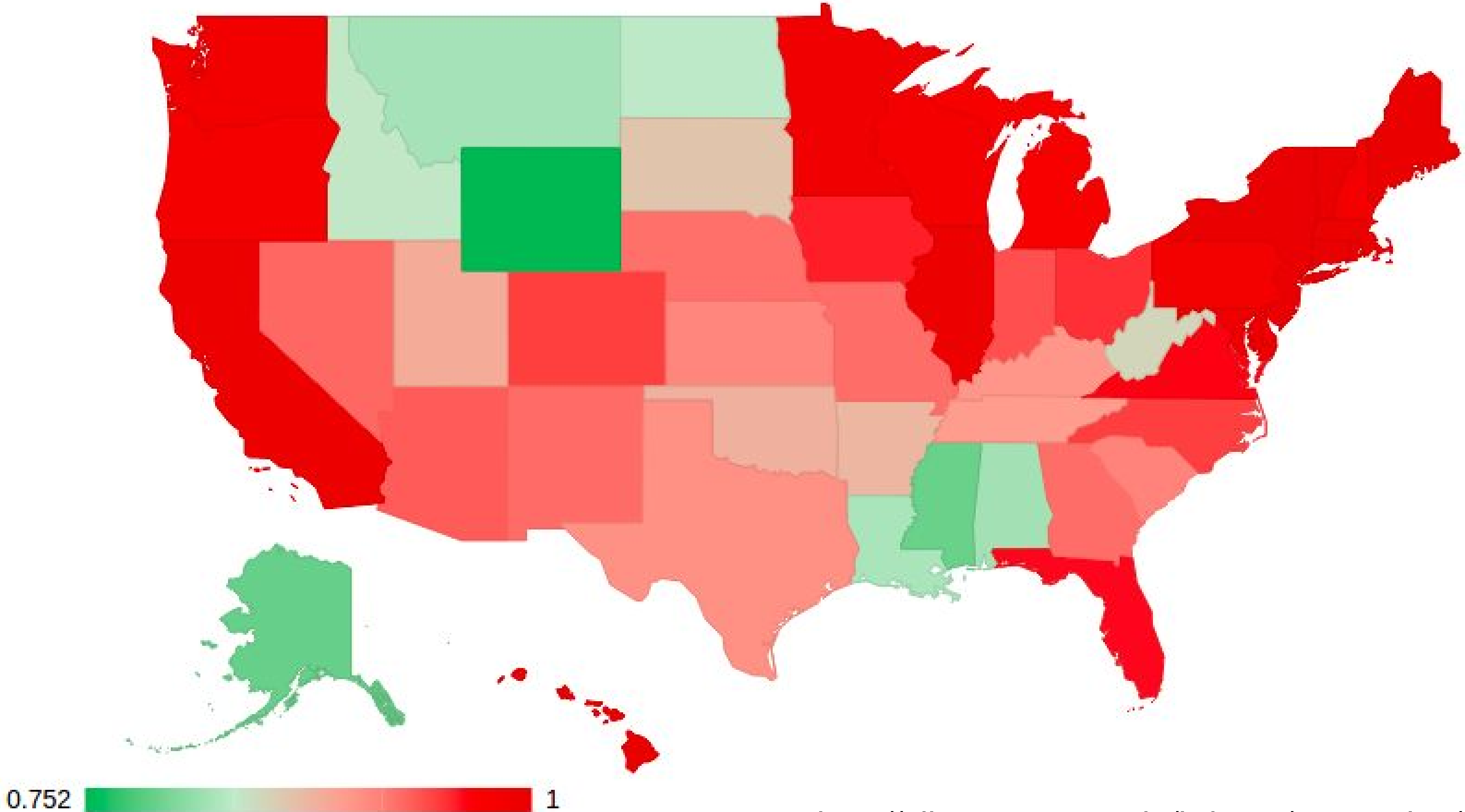
http://ciir.cs.umass.edu/irdemo/contention/



Outlier:
0.15

Selected results - Gun control in U.S.

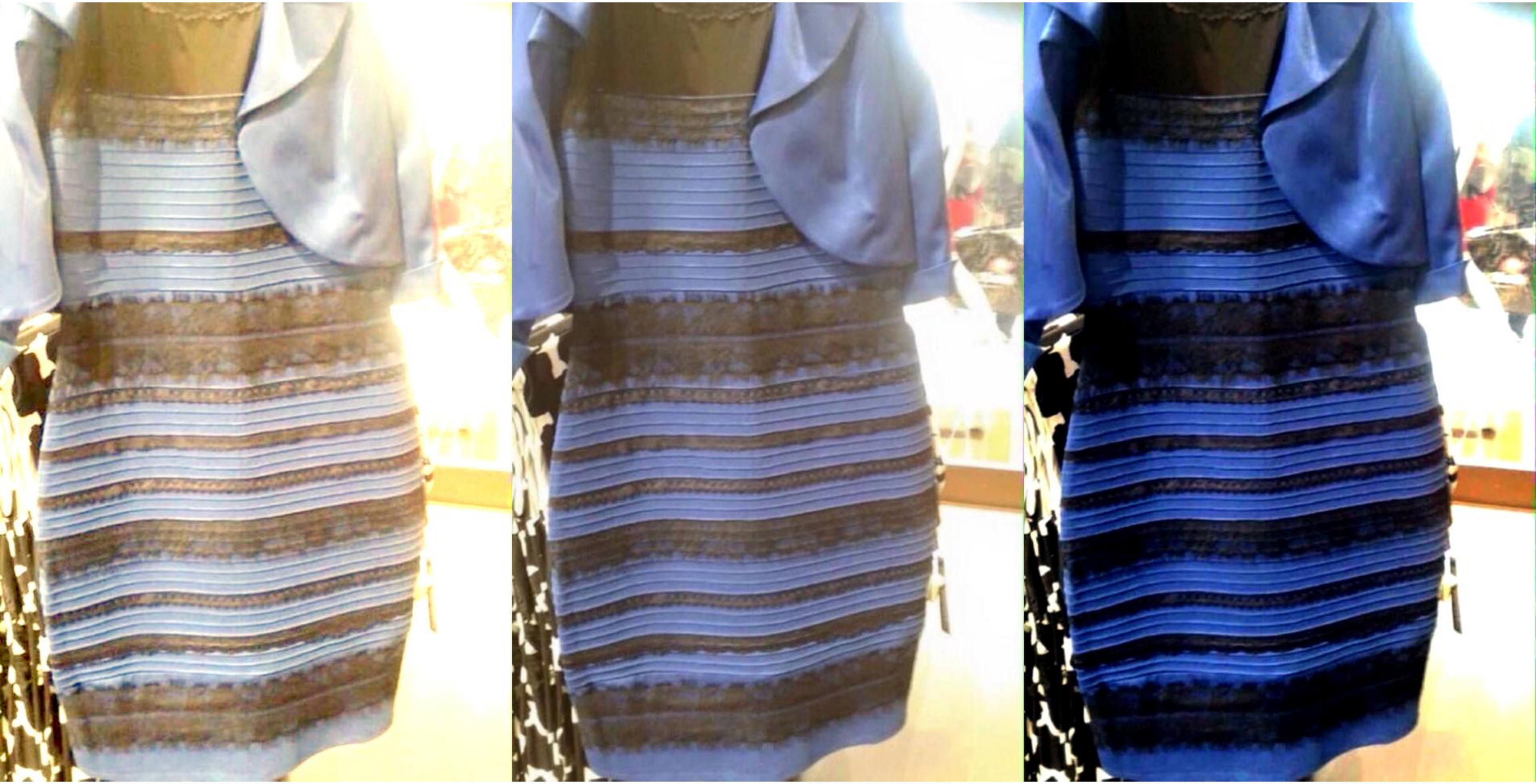
Do you support increased gun control?



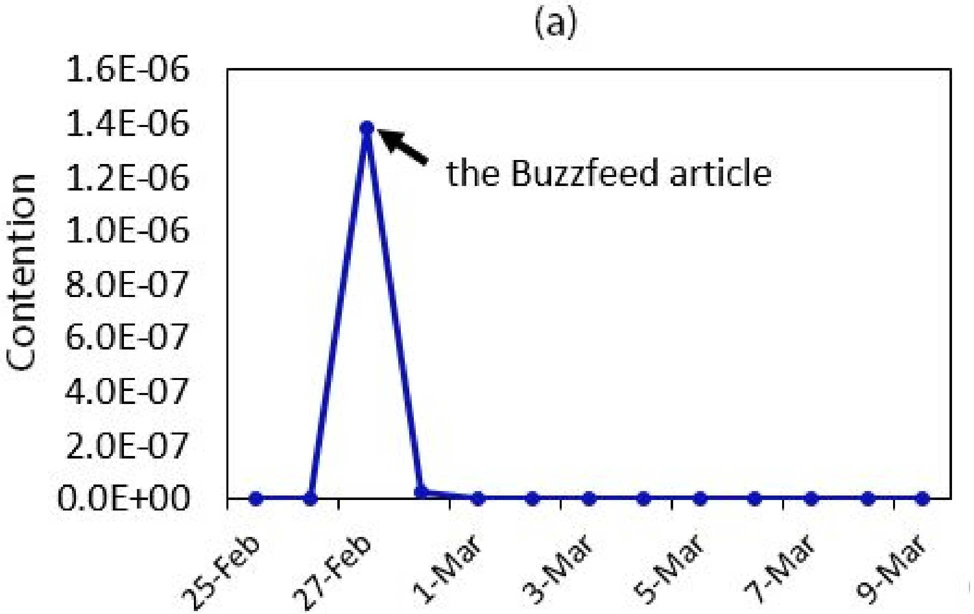
http://ciir.cs.umass.edu/irdemo/contention/

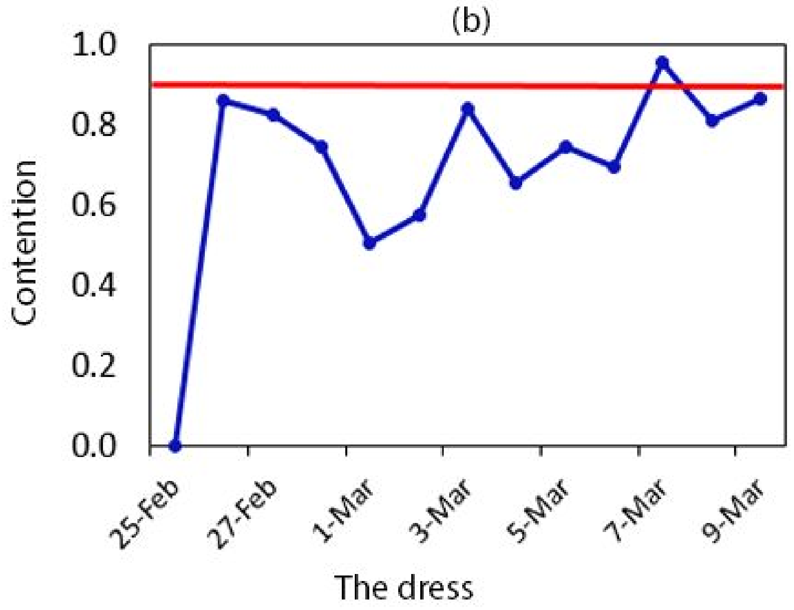
What colors are this dress?

What colors are this dress?

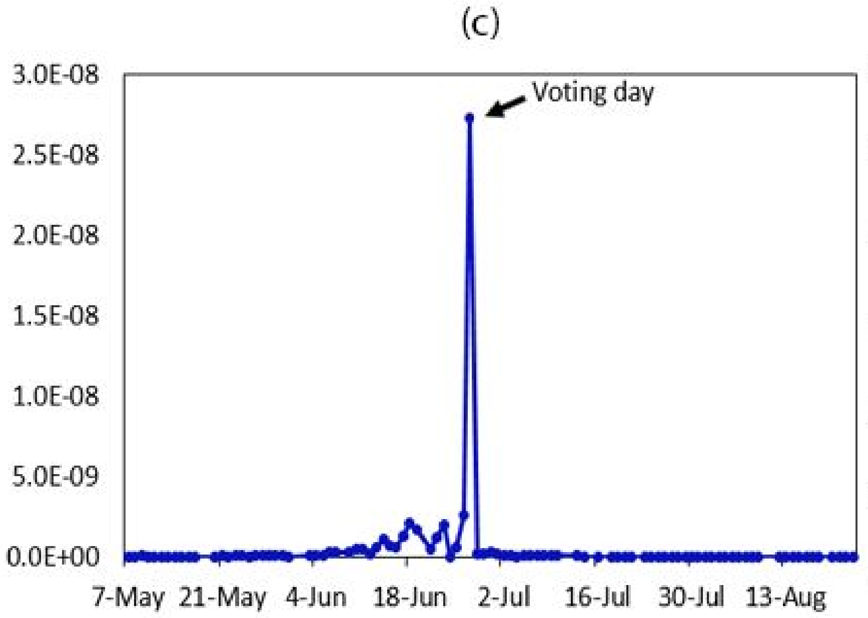


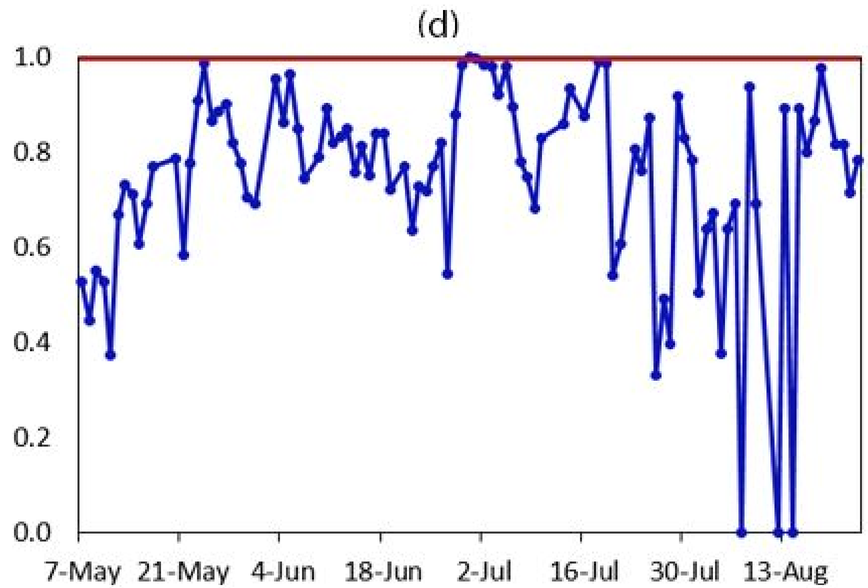
The Dress on Twitter





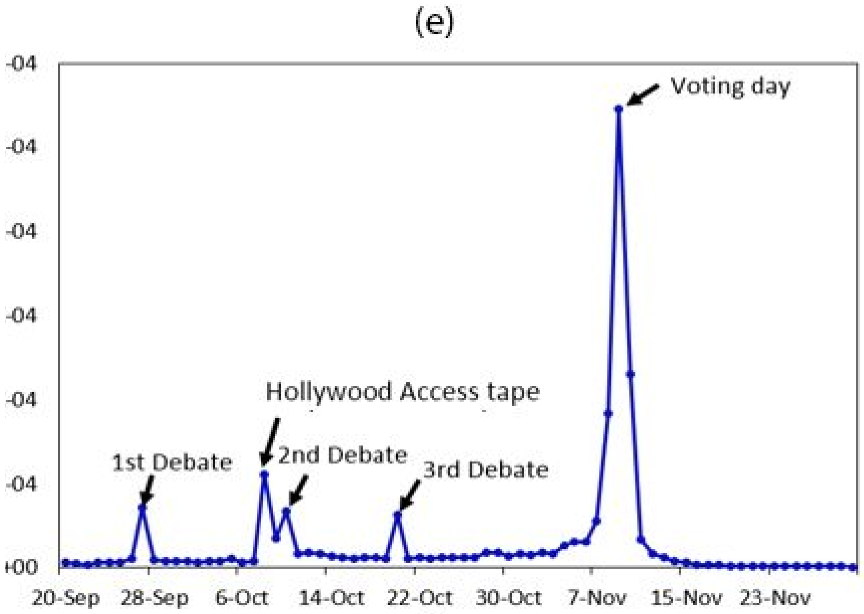
Brexit on Twitter





2016 Brexit Referendum

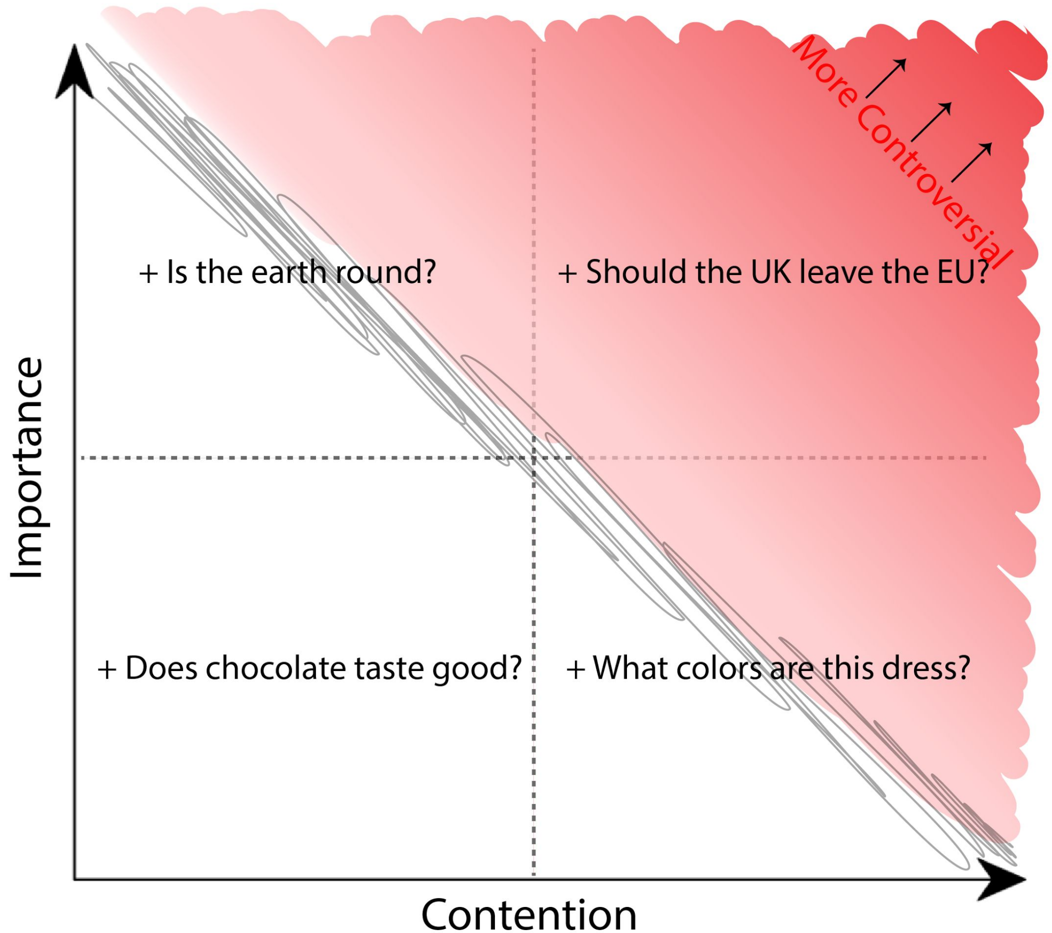
US Election on Twitter



2016 U.S. Presidential Election

Hypothesized model for controversy

- Contention is one dimension of controversy



Overview

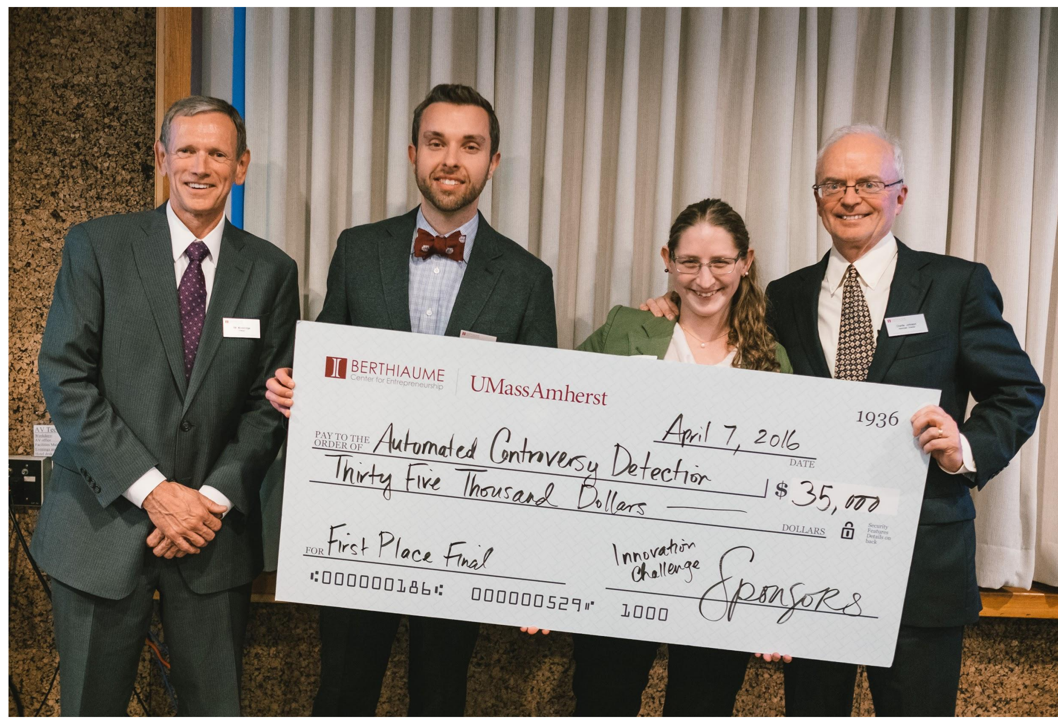
- Motivation
- Related Work
- Controversy Detection in Wikipedia Using Collective Classification
- Contributions to Controversy Detection
  - On the web (using Wikipedia, language models)
  - Position paper (social, ethical, technical challenges)
  - Future work - Definition of Controversy
- Startup - AuCoDe

Startup - AuCoDe

Automated Controversy Detection

- In April ‘16, I founded a startup to bring our controversy technology to market

- Won first place and non-equity grant in the UMass Innovation Challenge



Startup - AuCoDe

Automated Controversy Detection

- Went through a couple of pivots (news, PR)
- Constructing an alternative data PoC (backtest)
- Looking into social good applications
- Patent application through UMass
- $95K non-equity funding raised to date
- Recently applied for NSF SBIR funding

Thank you!

Questions, comments?

shiri@dori-hacohen.com
shiri@cs.umass.edu

www.linkedin.com/in/shiri
controversies.info