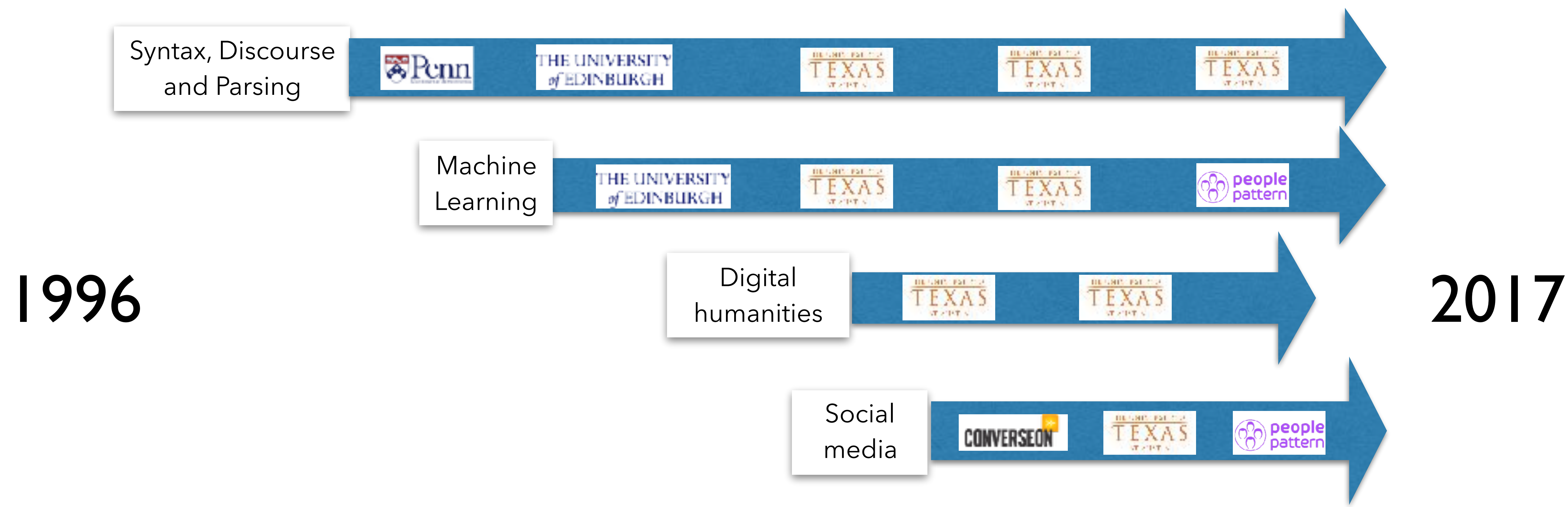
Practical and Ethical Considerations in Demographic and Psychographic Analysis

Jason Baldridge
Co-founder & Chief Scientist
People Pattern
@jasonbaldridge

people pattern

Brief background

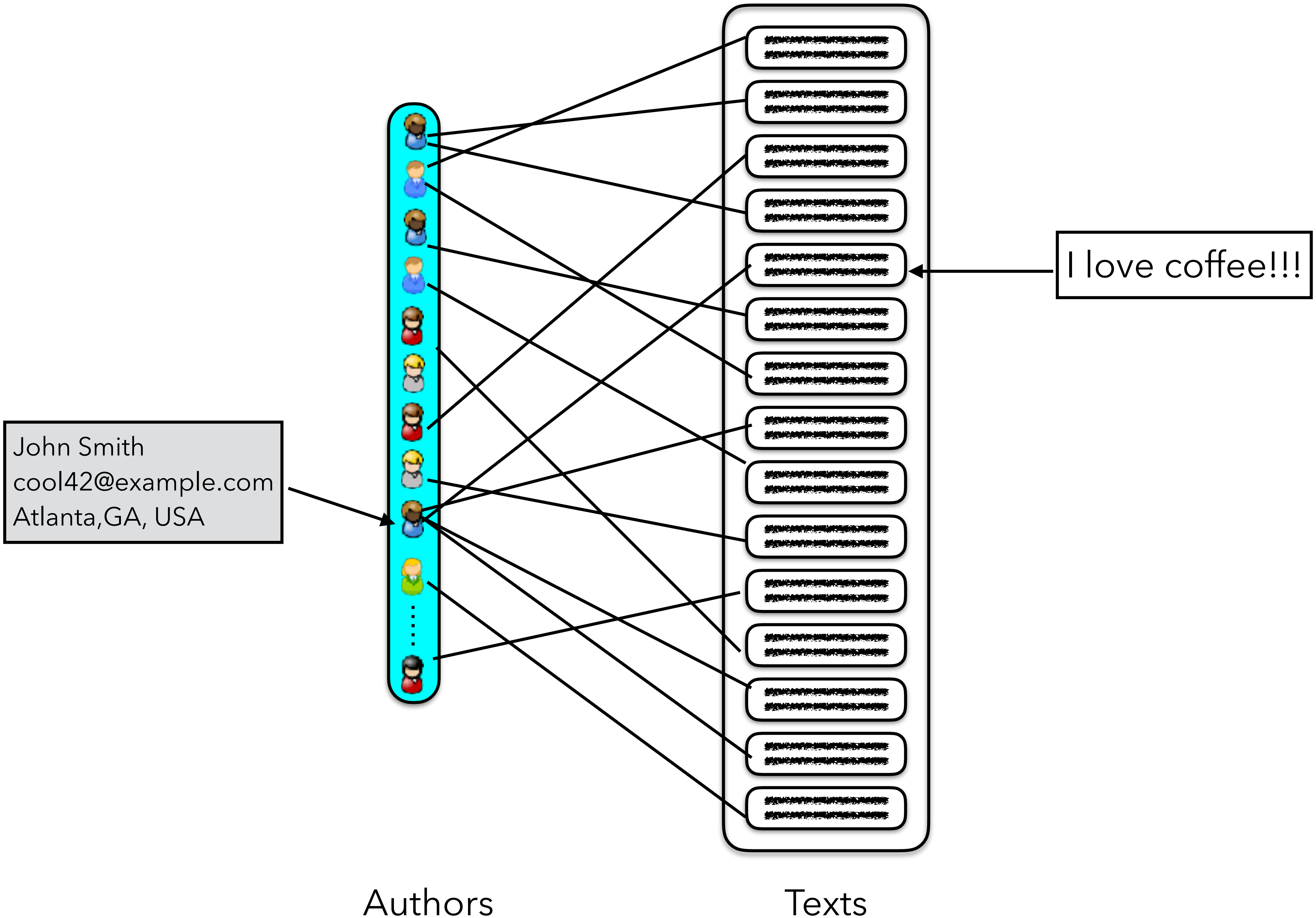


Spanning academia and industry

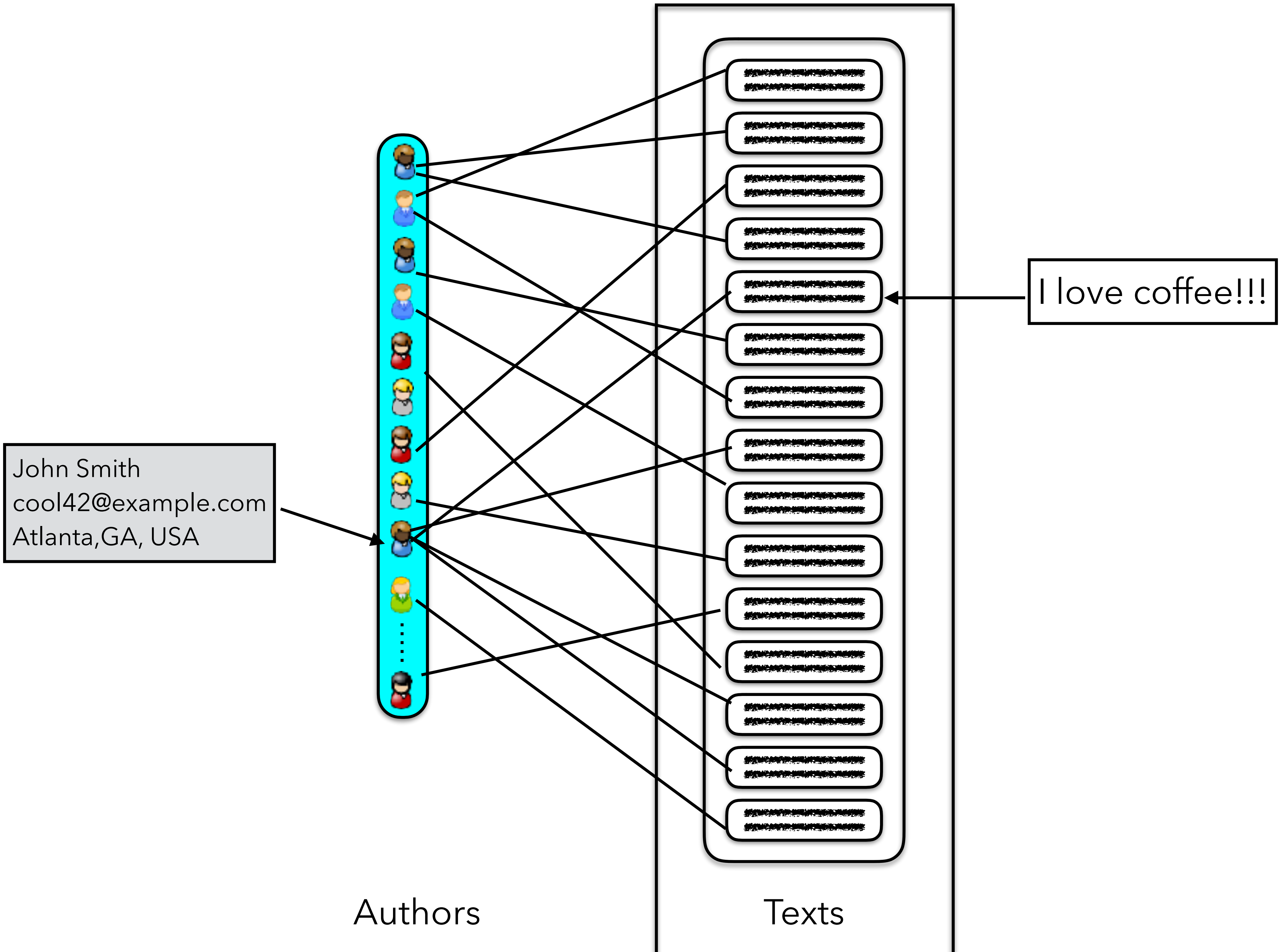
People > Posts

(But language provides a remarkable window into people and communities.)

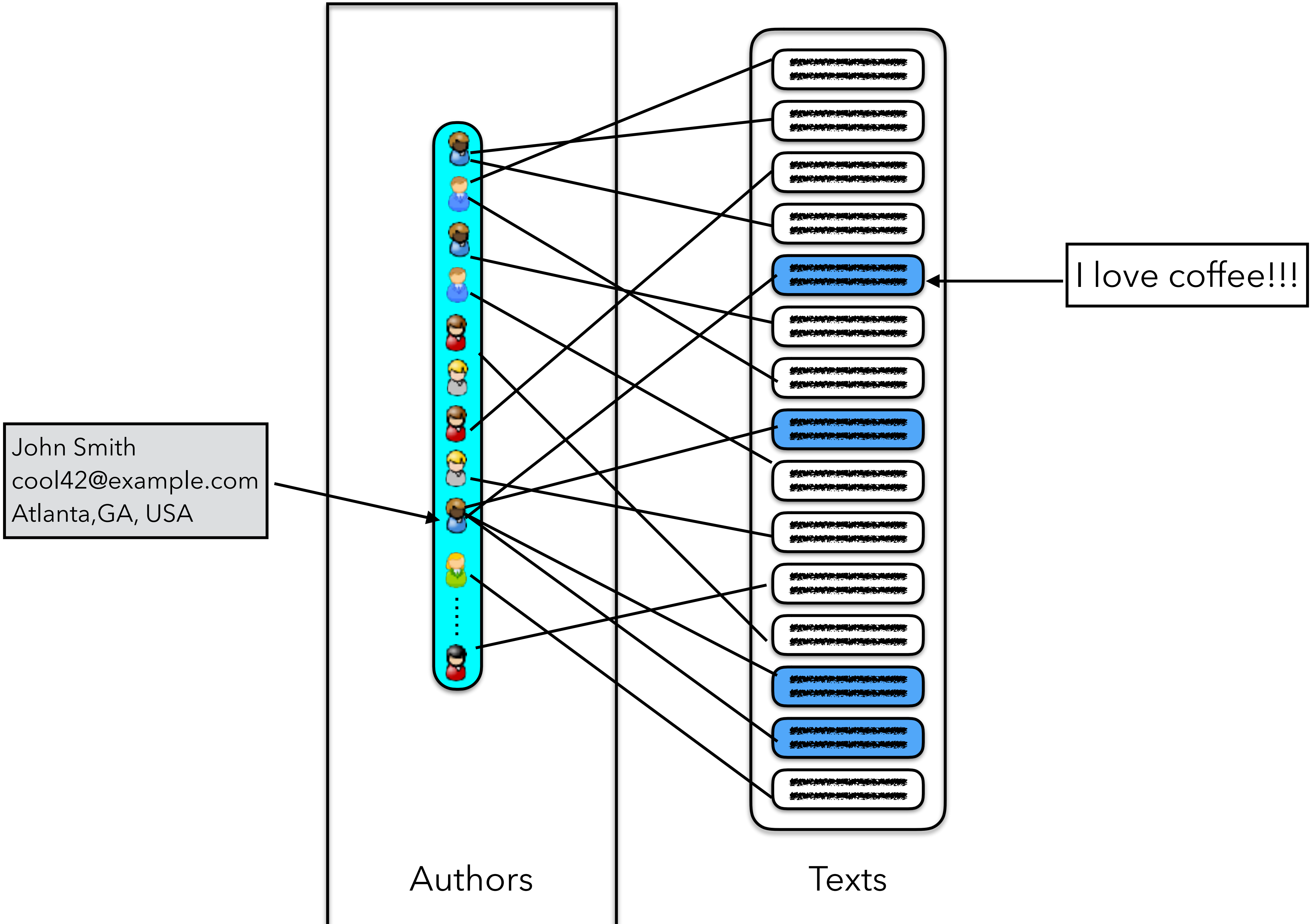
Conversation Focus



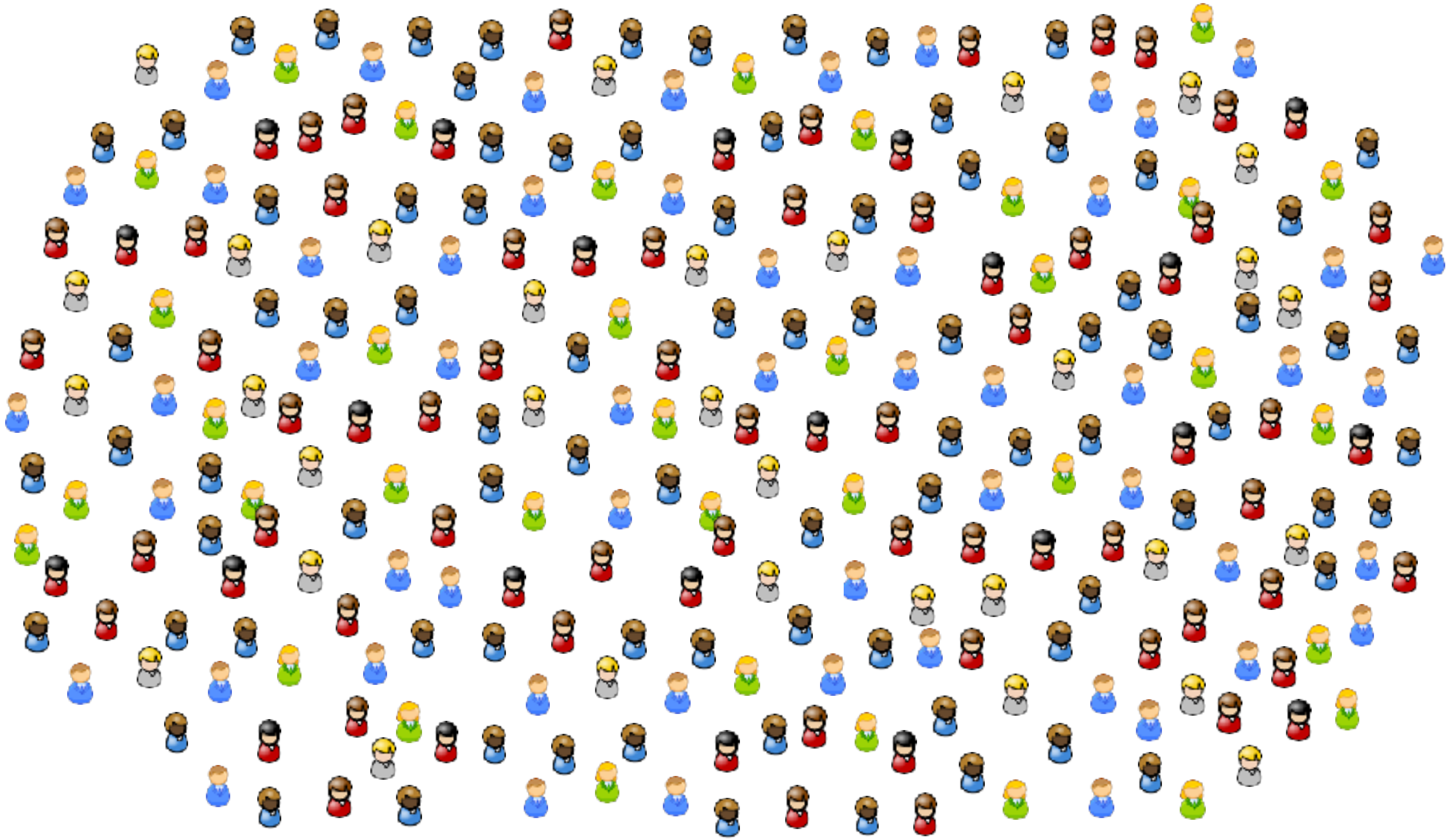
Conversation Focus



Audience Focus

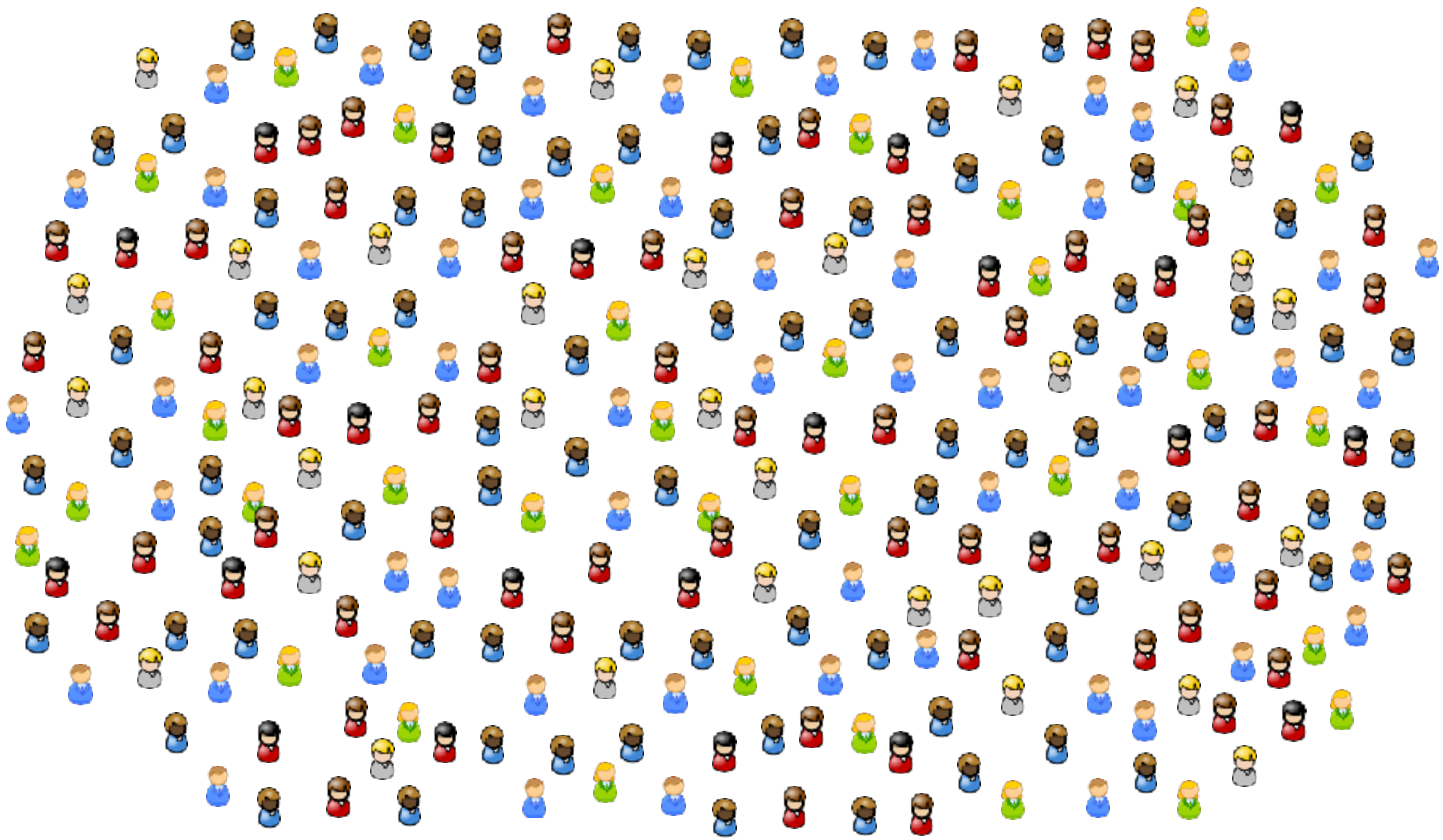


The Problem



Identify, segment and analyze groups of people.

Identification

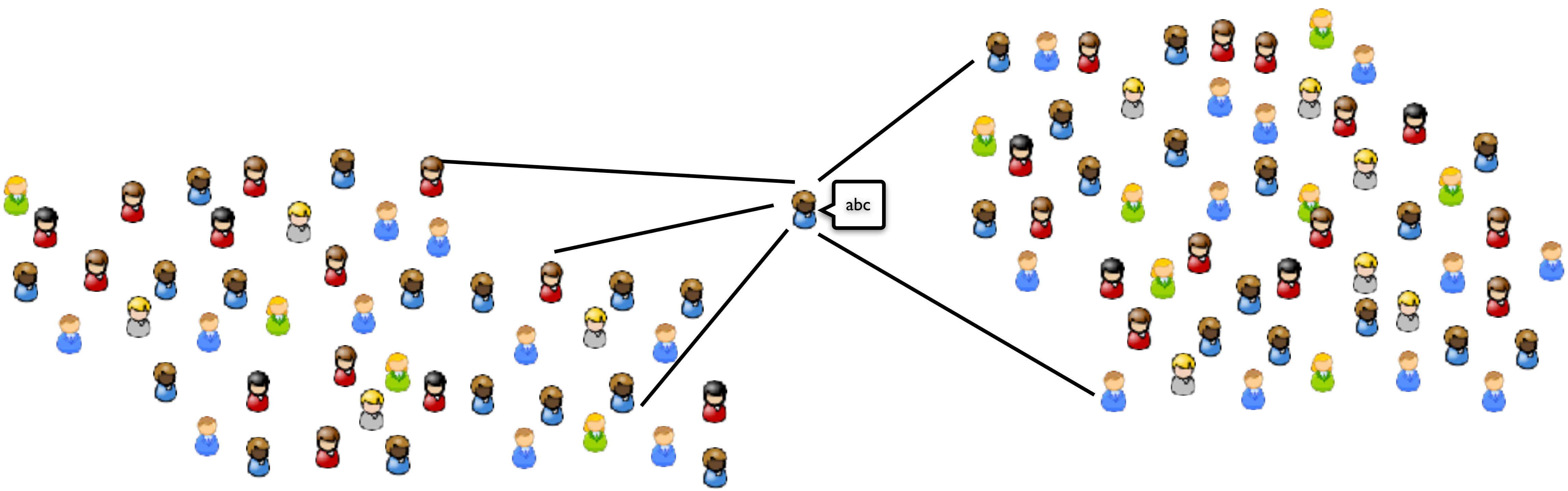


keywords, hashtags, demographics, stitching

Identification

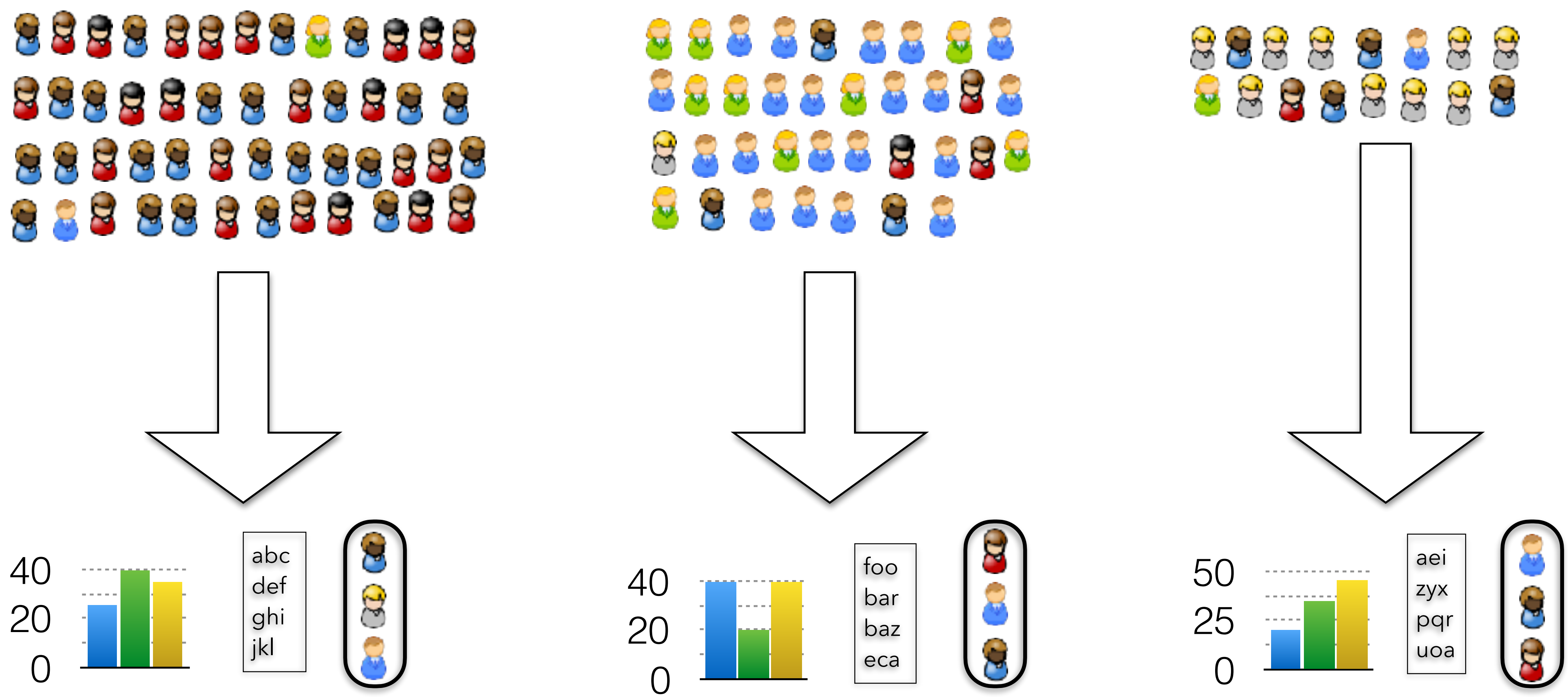
keywords, hashtags, demographics, stitching

Segment and Analyze



profiles, posts, images, connections, clustering

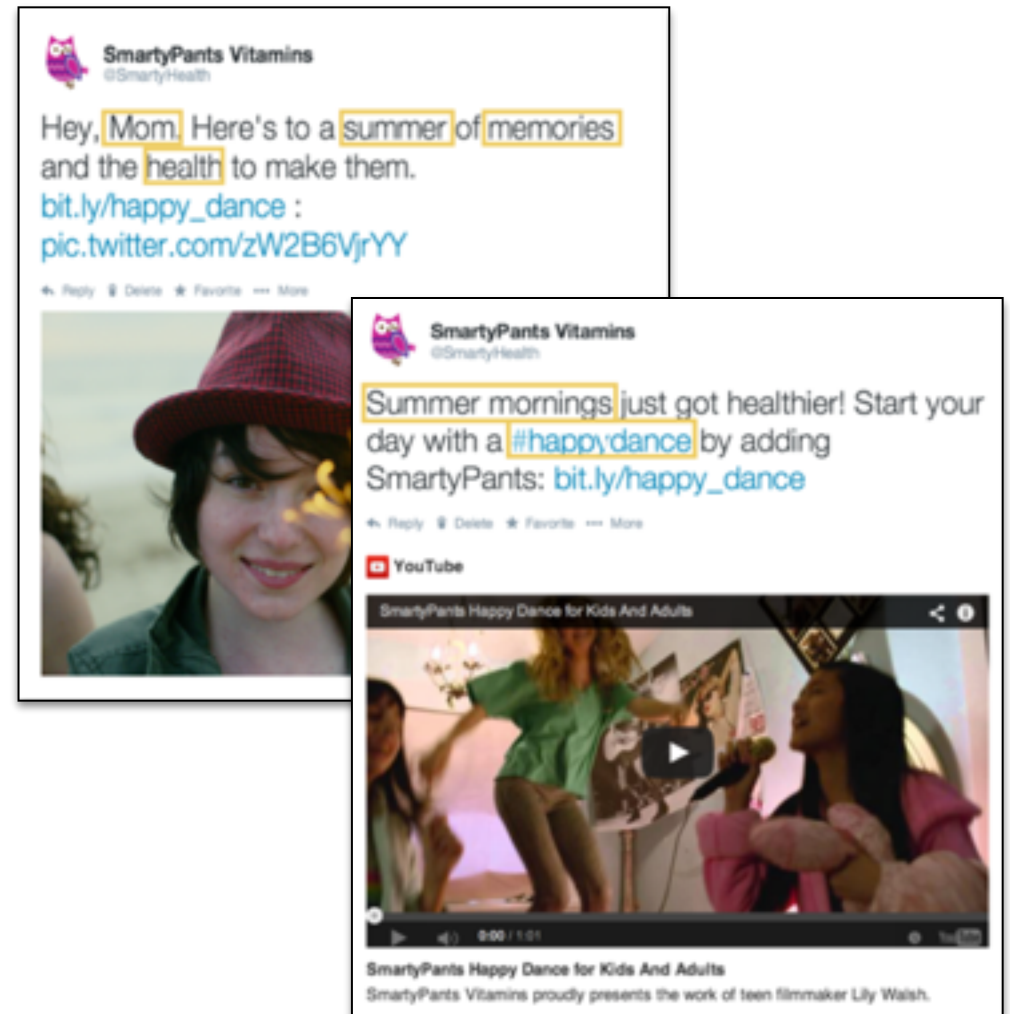
Segment and Analyze



profiles, posts, images, connections, clustering

Tailored audiences

Shopping (29%)	Eating (24%)	Family (20%)	School (17%)	Music (10%)
#giveaway	#ff	#parenting	#kids	#soundcloud
#win	recipe	child	student	video
#health	amazing	baby	public	liked
#ad	#health	birthday	elementary	added
#quote	morning	dad	career	warhol
new	wait	adhd	speech	church
free	tonight	sick	fresher	playlist
happy	loss	smile	charter	♫
want	body	thankful	sunset	whitney

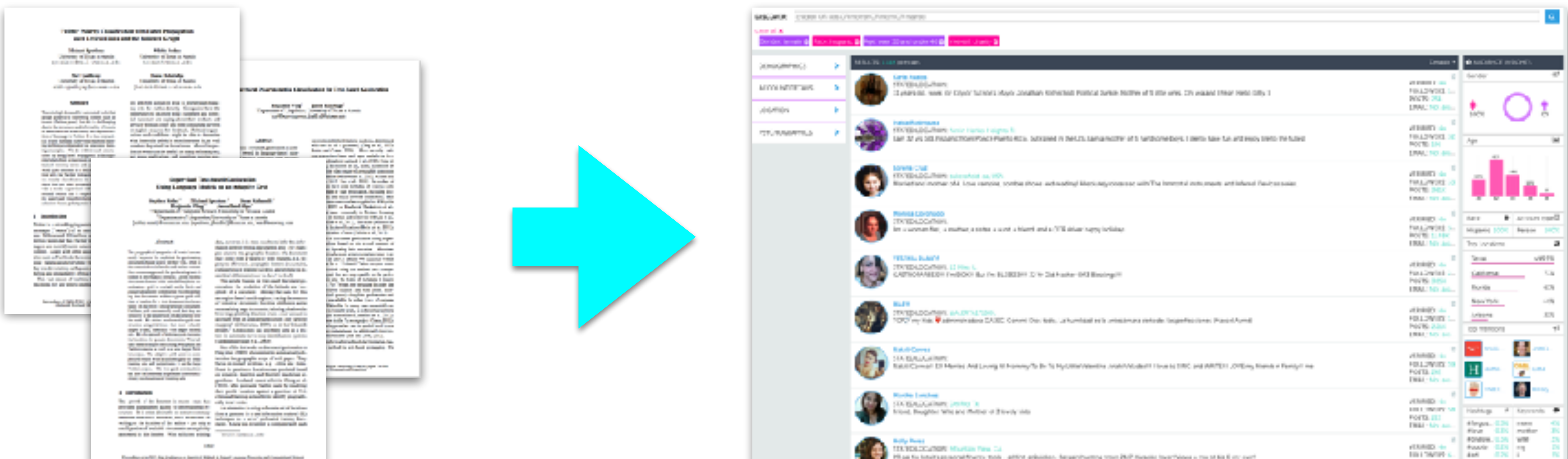


Interest prediction and extraction of interest-specific keywords.
Promoted tweet copy informed by persona-based keywords.

people pattern + SMARTY PANTS

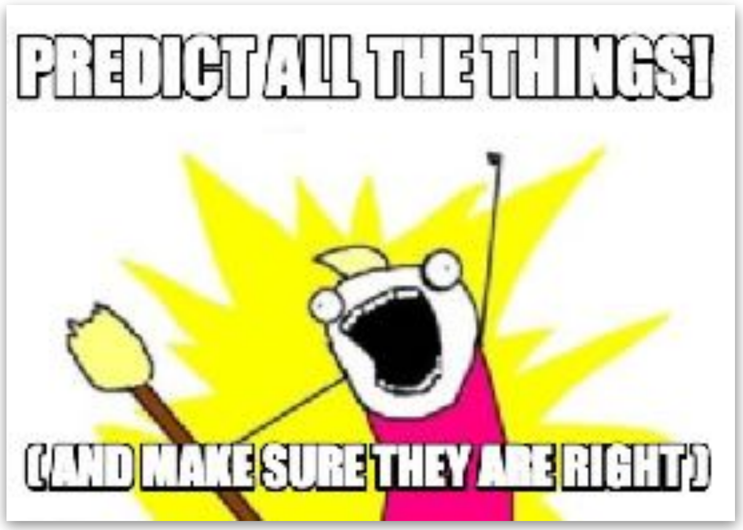
People Pattern and Smarty Pants Vitamins case study.

Lessons from industry



The needs of industry expose interesting challenges!

Complete coverage



Users want all the attributes for all the profiles,
but mistakes will be made.

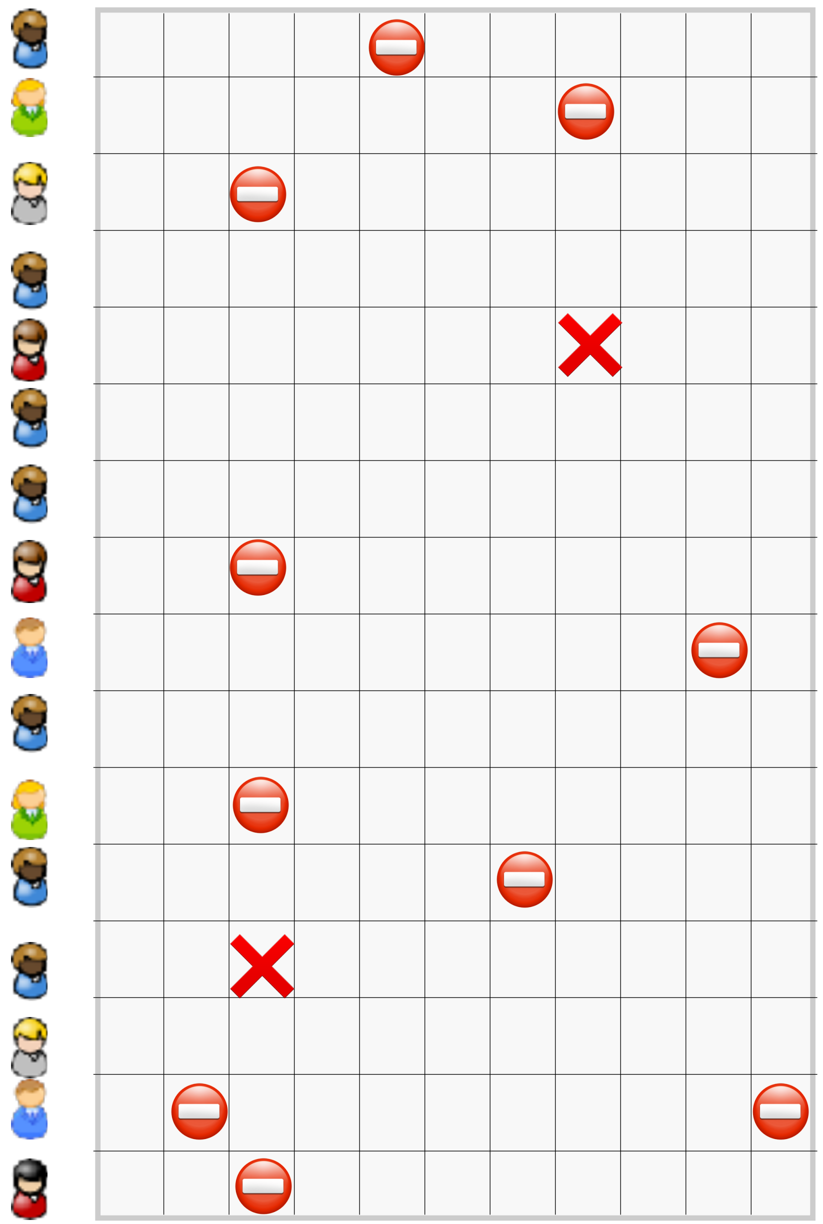
Cherries & cockroaches

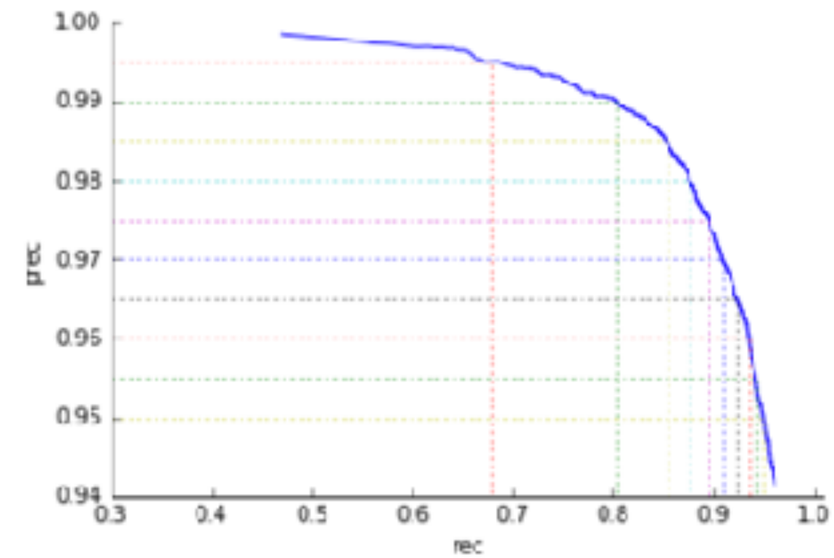
https://commons.wikimedia.org/wiki/File:Bowl_of_cherries_with_colours_enhanced.jpg

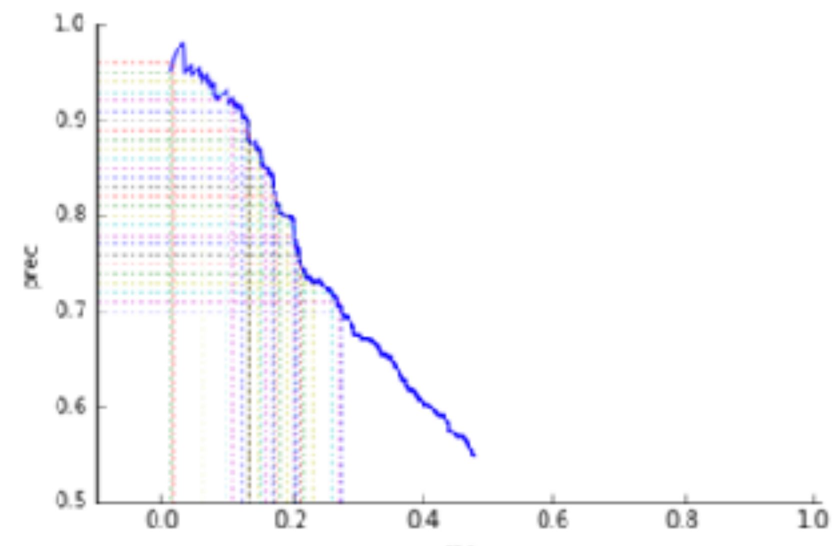
One cockroach spoils the bowl!

Daniel Kahneman, “Thinking, Fast and Slow”

High precision

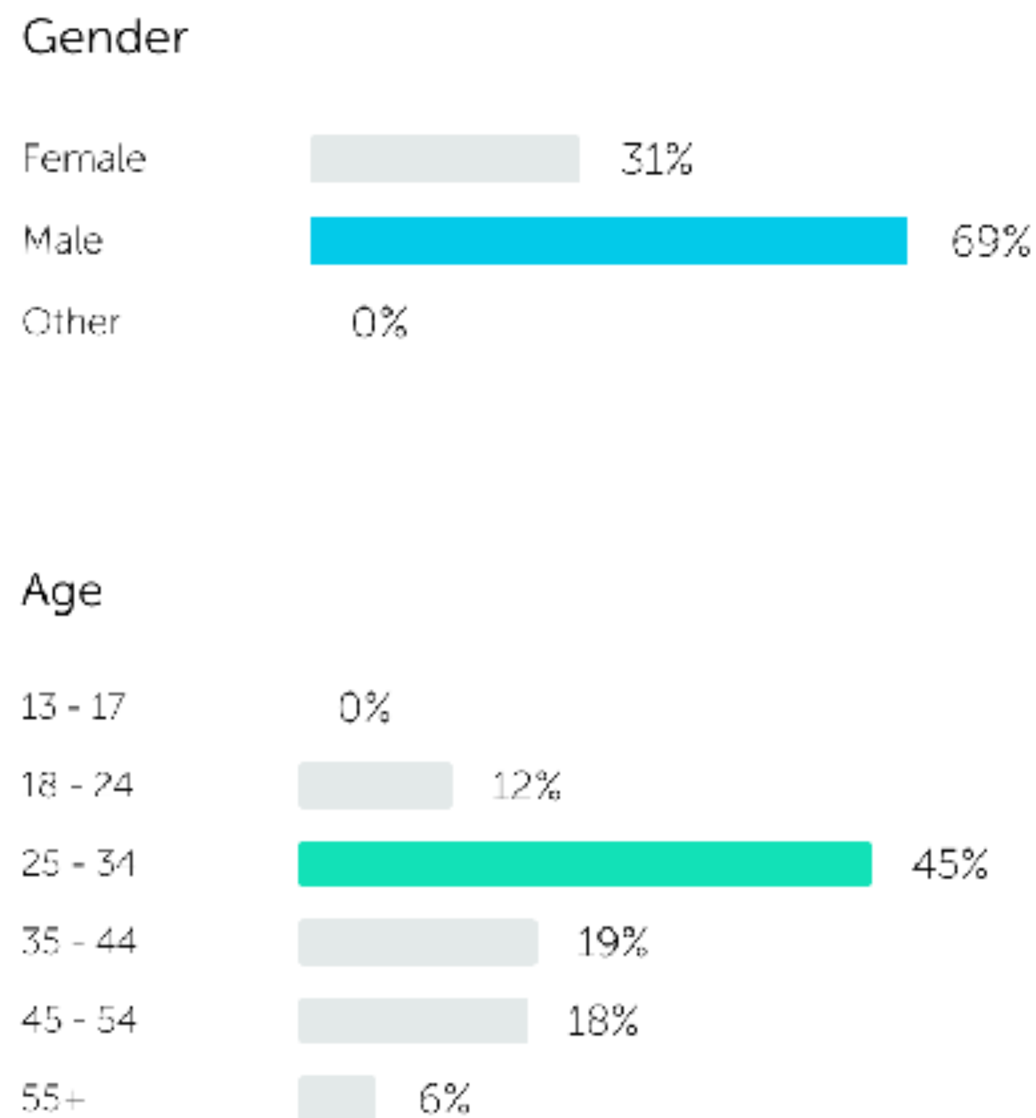






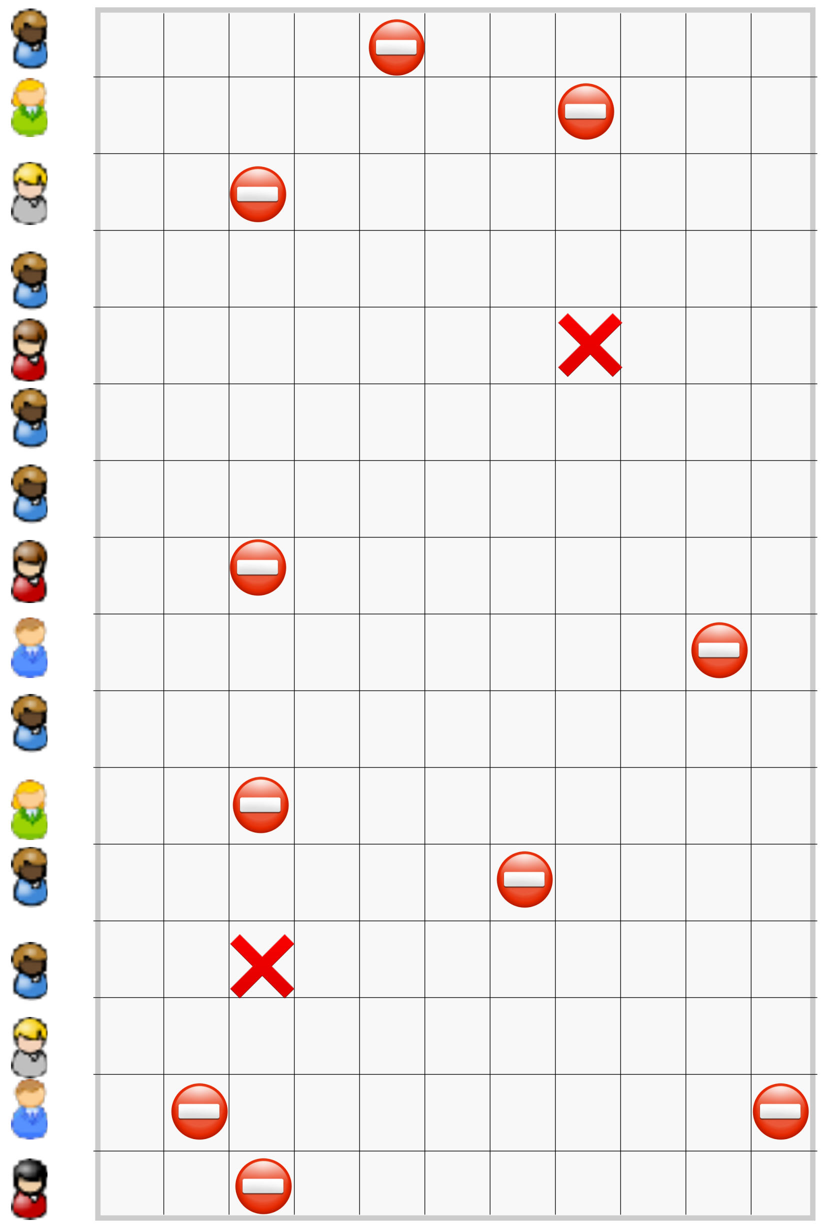
Analyze model confidence and precision/recall tradeoff.

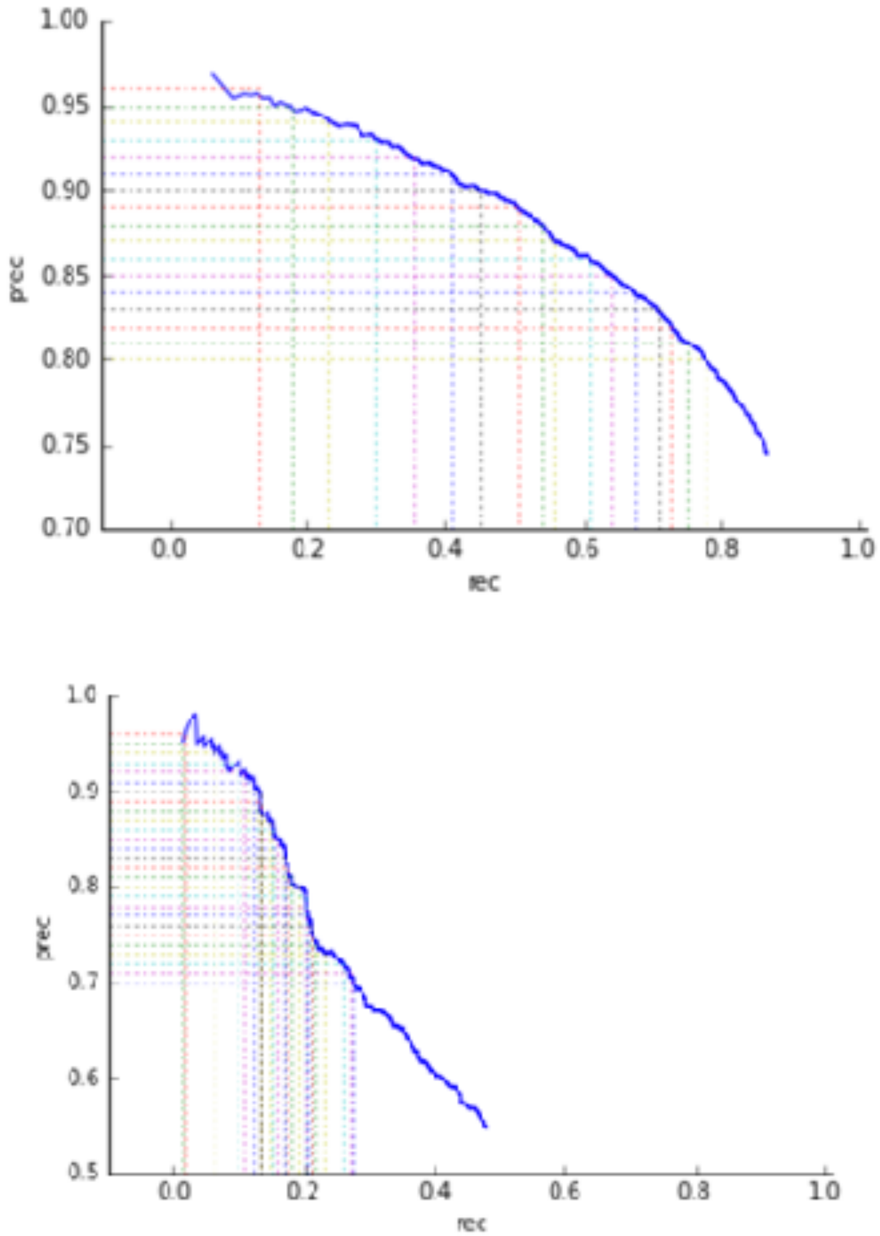
Segment & aggregate



Users need aggregate statistics for arbitrary segments.

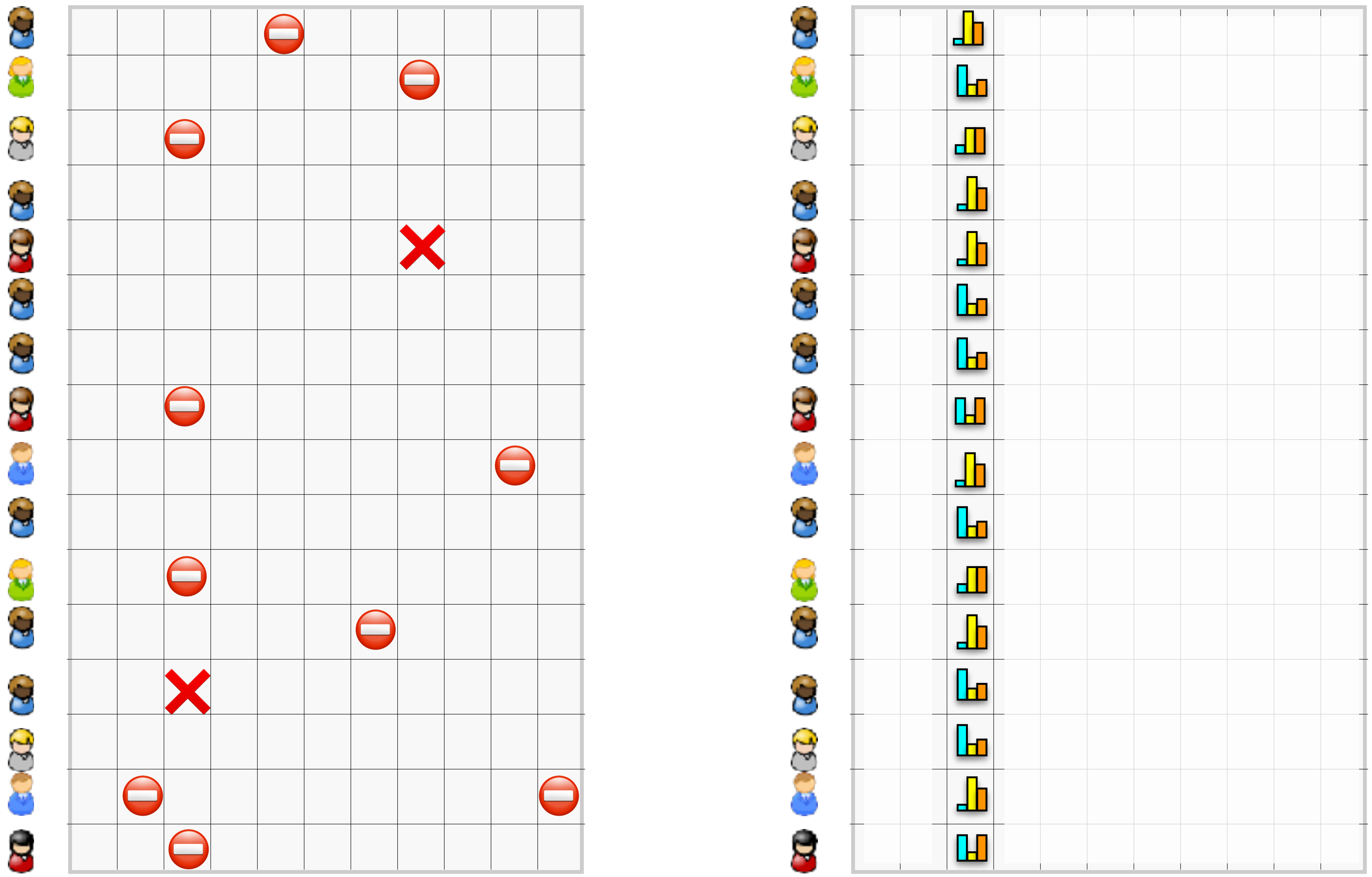
Segment & aggregate





High precision thresholding leads to different recall proportions for different classes and changes aggregate statistics.

Soft aggregations

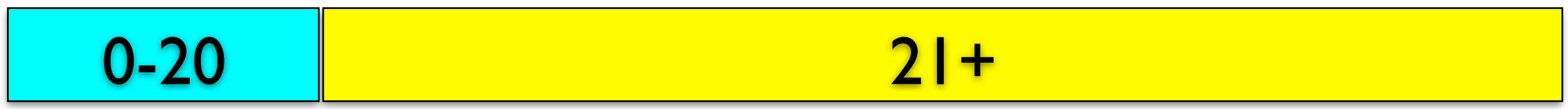


Store and use both high precision labels and classifier confidence distributions.

Age granularity

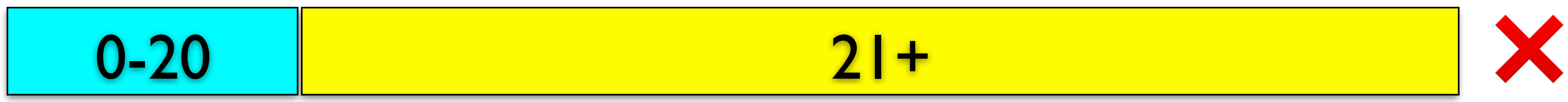
Supporting year-level granularity supports multiple use cases.

Age granularity



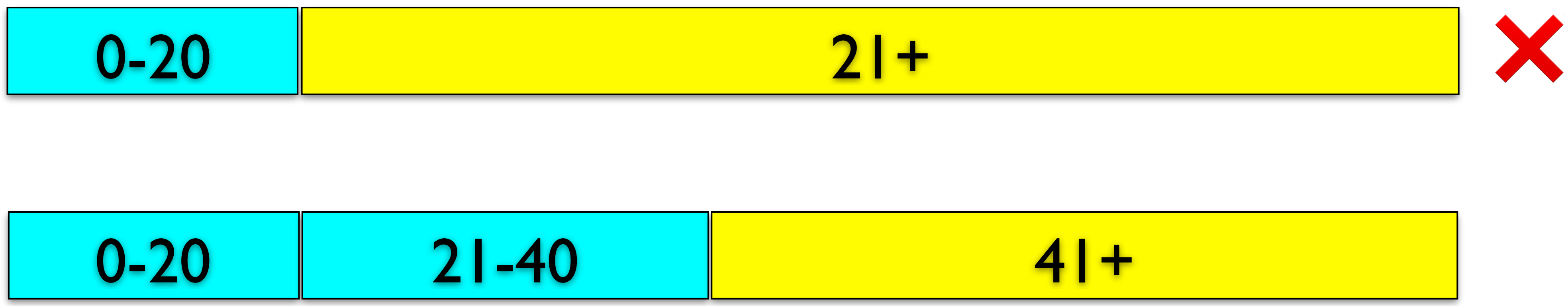
Supporting year-level granularity supports multiple use cases.

Age granularity



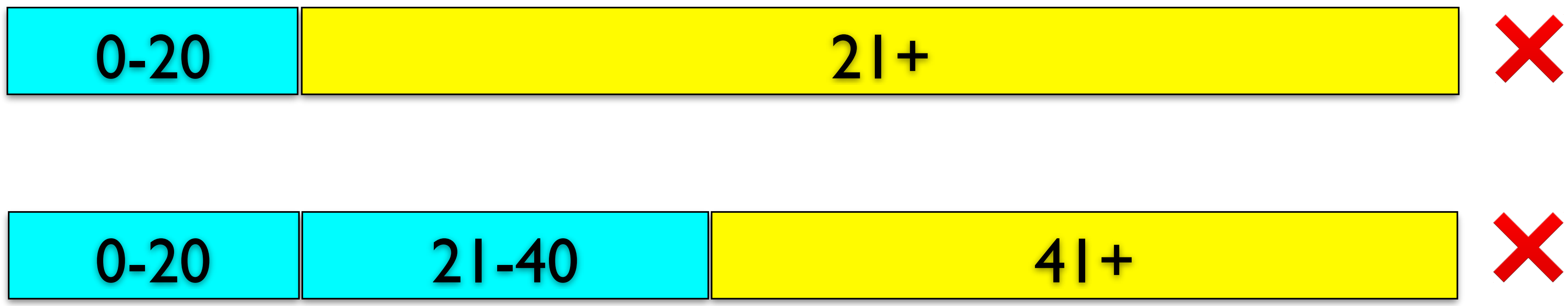
Supporting year-level granularity supports multiple use cases.

Age granularity



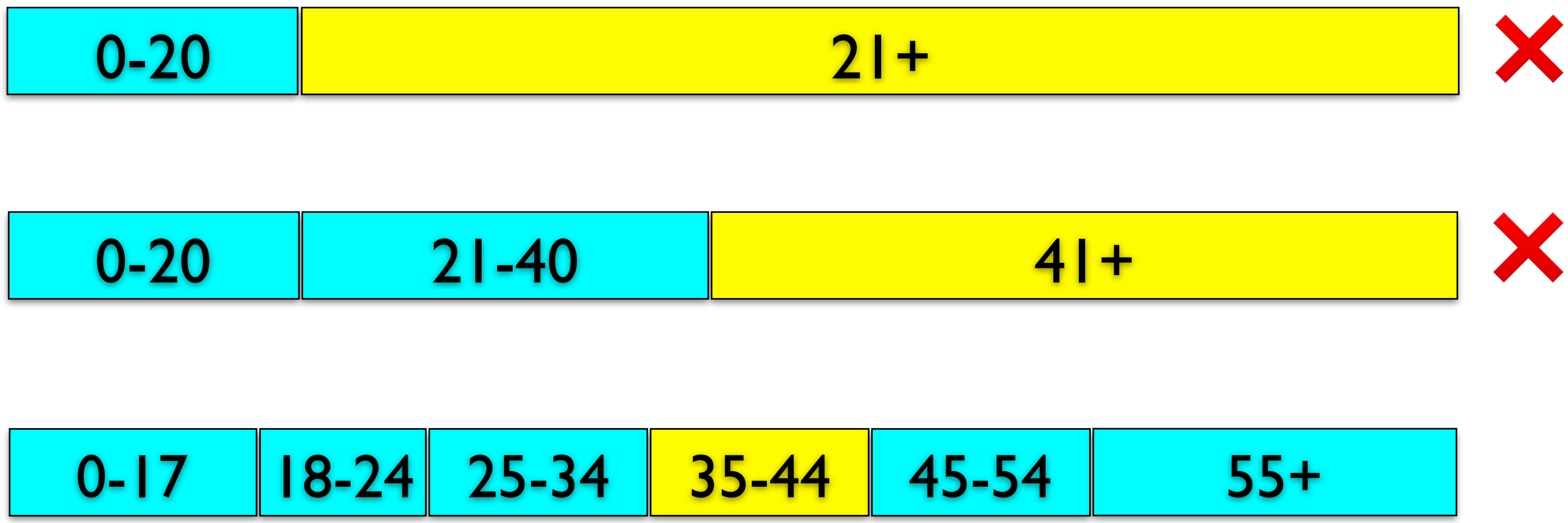
Supporting year-level granularity supports multiple use cases.

Age granularity



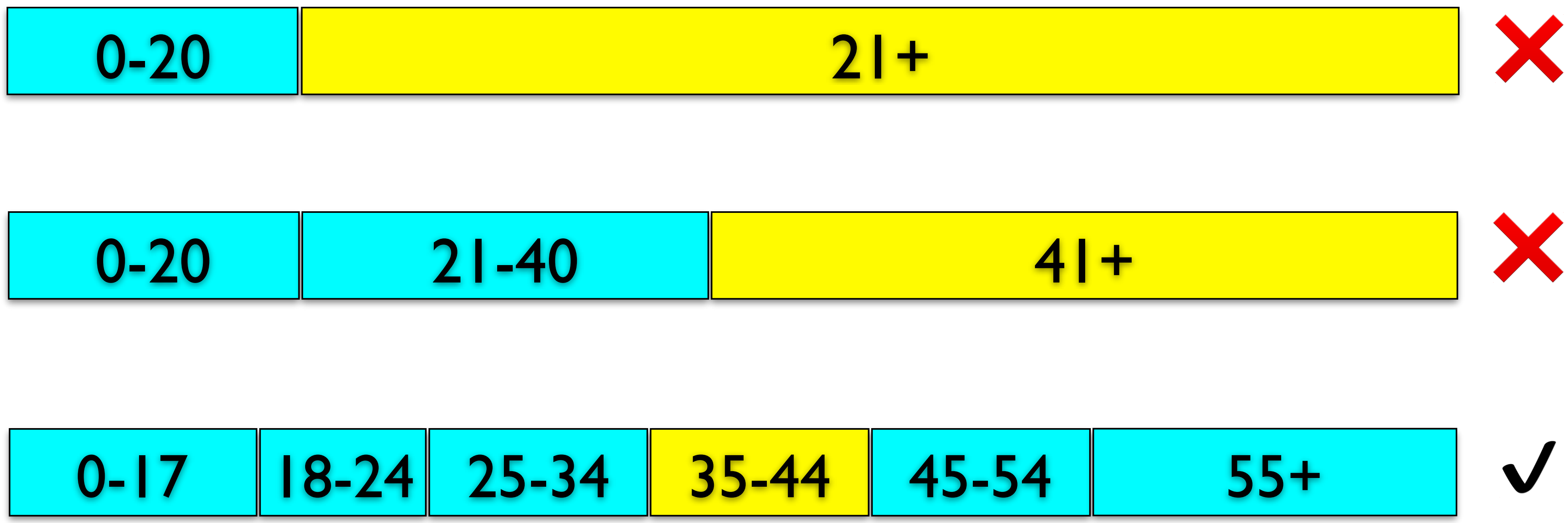
Supporting year-level granularity supports multiple use cases.

Age granularity



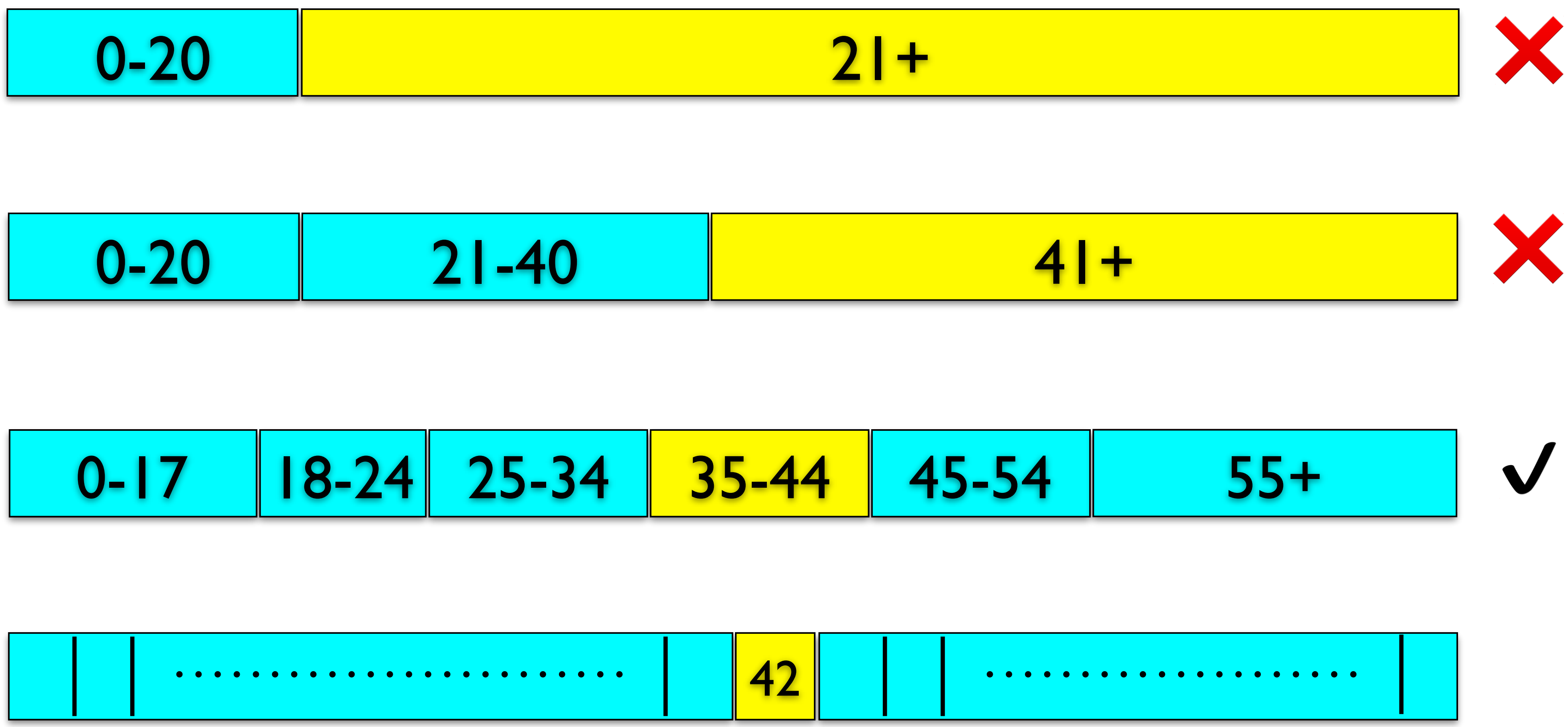
Supporting year-level granularity supports multiple use cases.

Age granularity



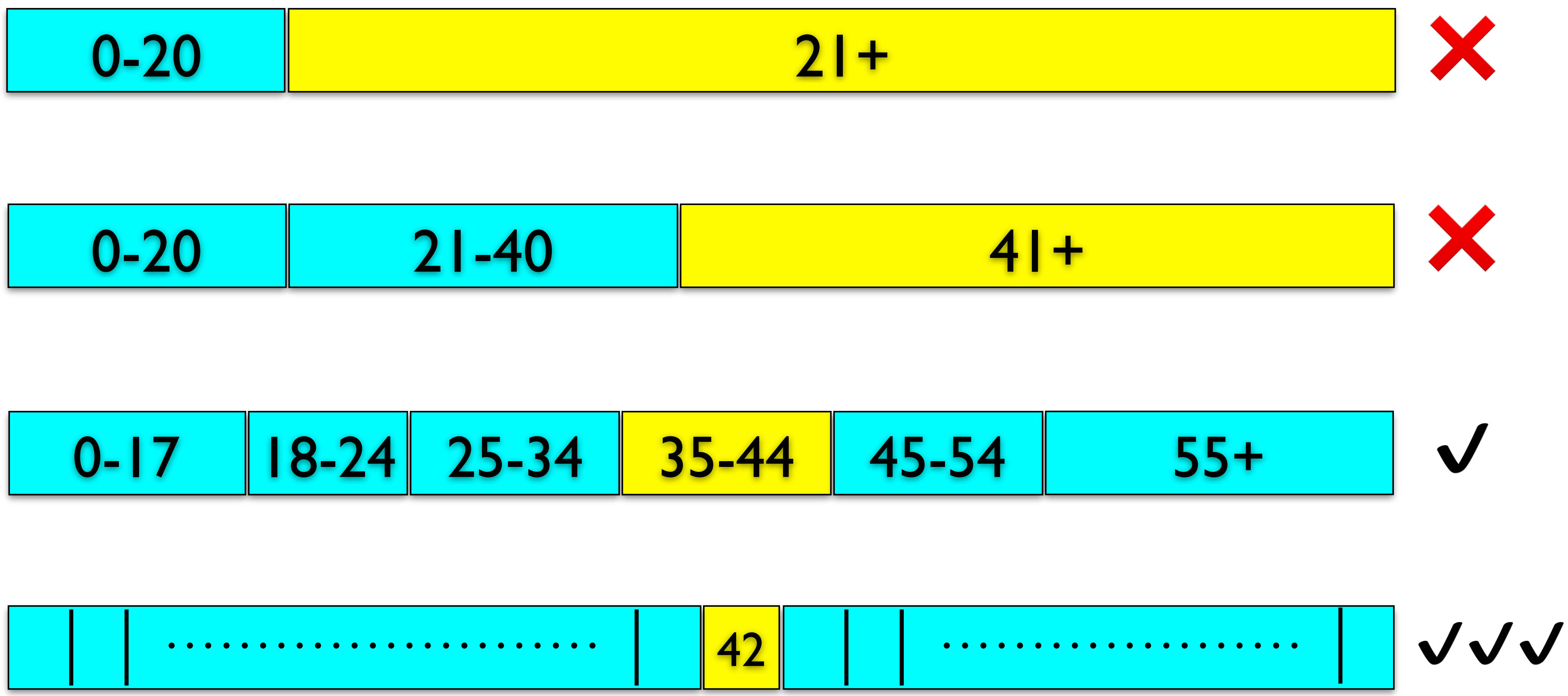
Supporting year-level granularity supports multiple use cases.

Age granularity



Supporting year-level granularity supports multiple use cases.

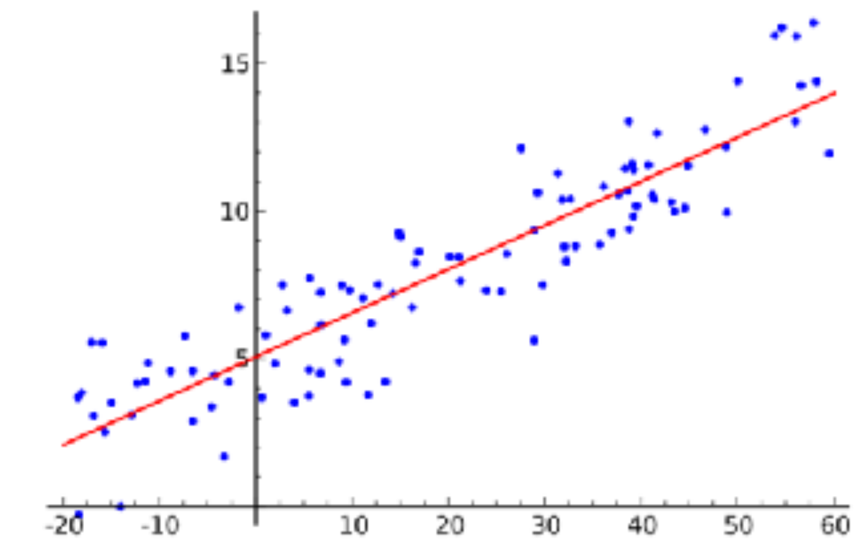
Age granularity



Supporting year-level granularity supports multiple use cases.

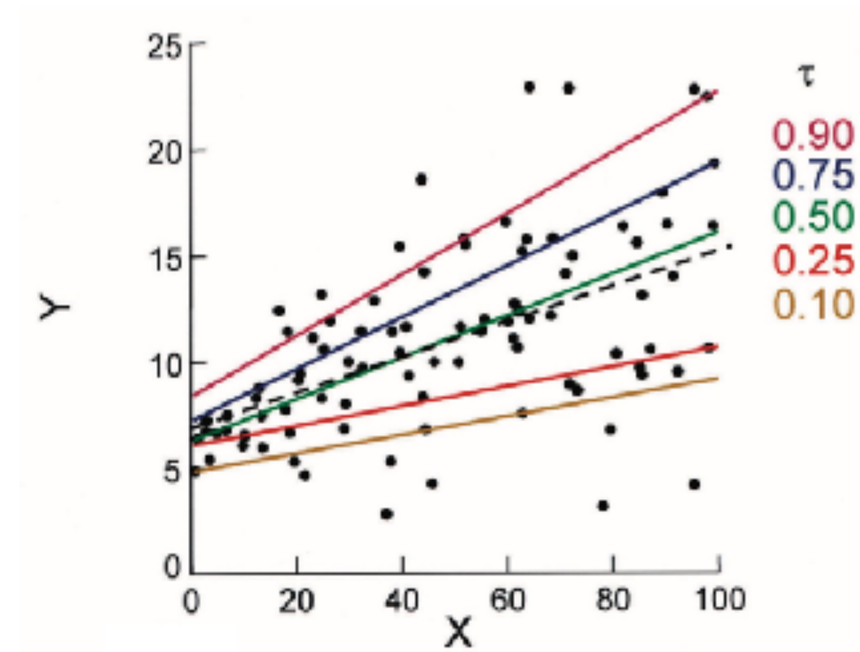
Regulatory requirements

21 or older: ≥85%



Aggregate accuracy matters.

Regulatory requirements

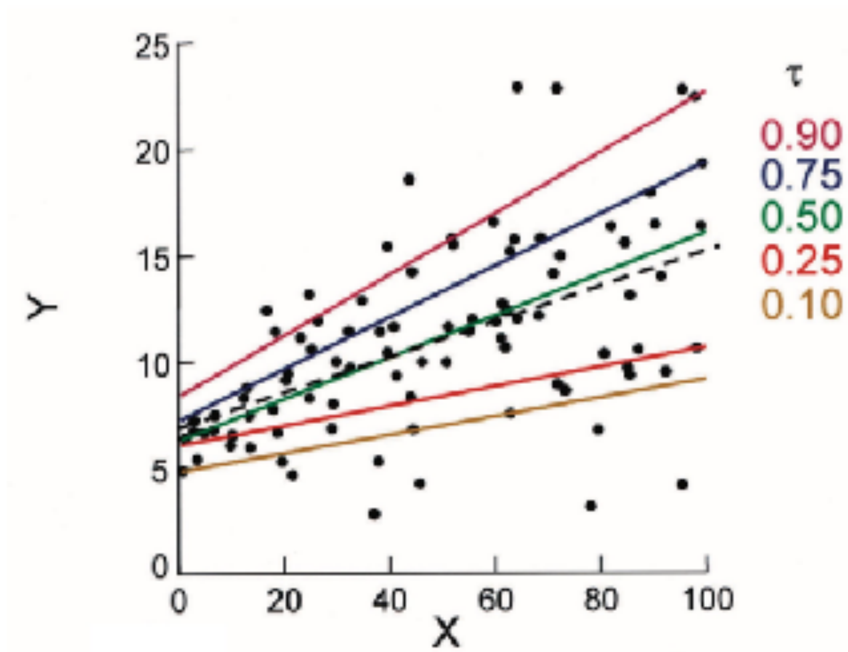


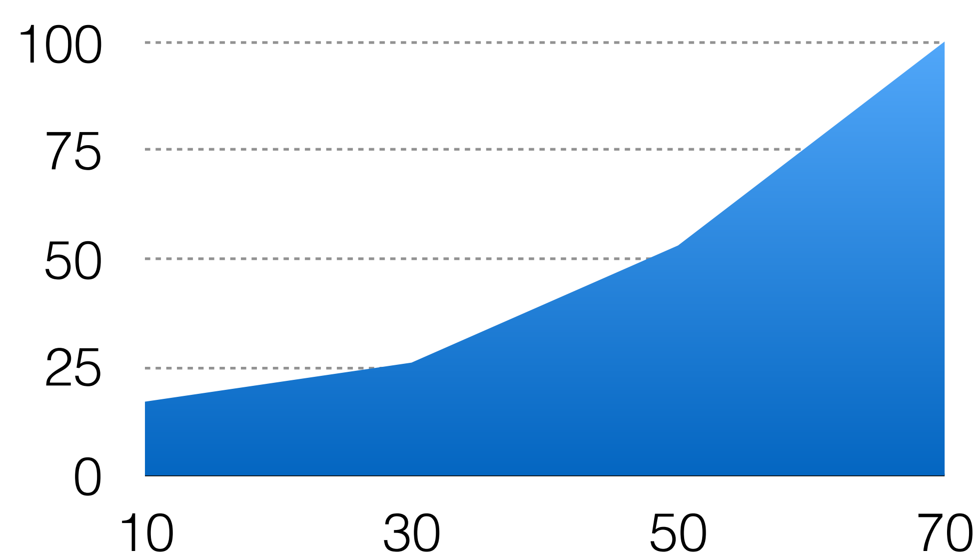
21 or older: ≥85%

Aggregate accuracy matters.

Regulatory requirements

21 or older: ≥85%





Aggregate accuracy matters.

Actual or apparent age?

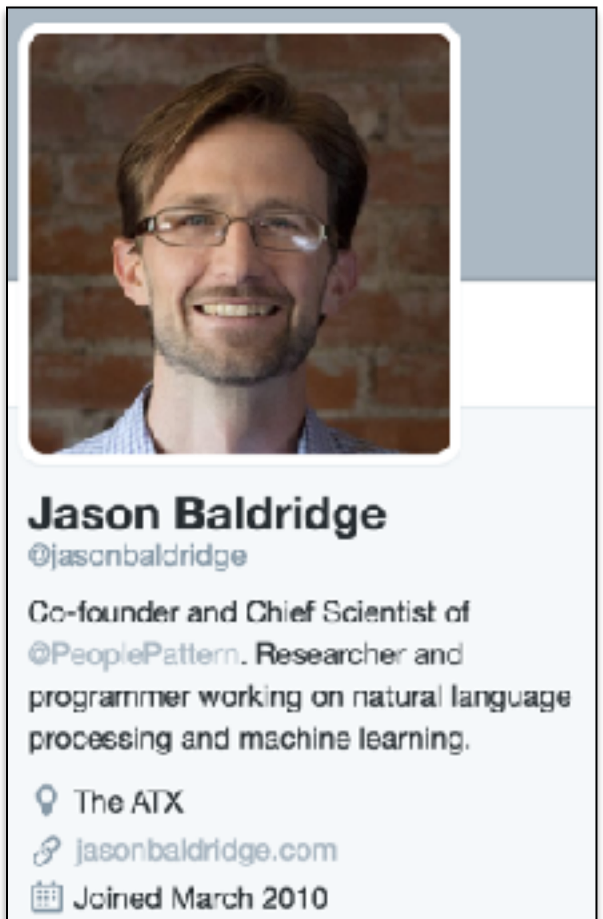
Rothe, Timofte, and Van Gool (2016). https://www.vision.ee.ethz.ch/en/publications/papers/proceedings/eth_biwi_01229.pdf

59 / 57 | 37 / 35 | 65 / 51 | 20 / 29

Actual / apparent (age)

Looking At People 2015 Apparent Age Challenge:
Labels are 10 or more age guesses by people given photo.

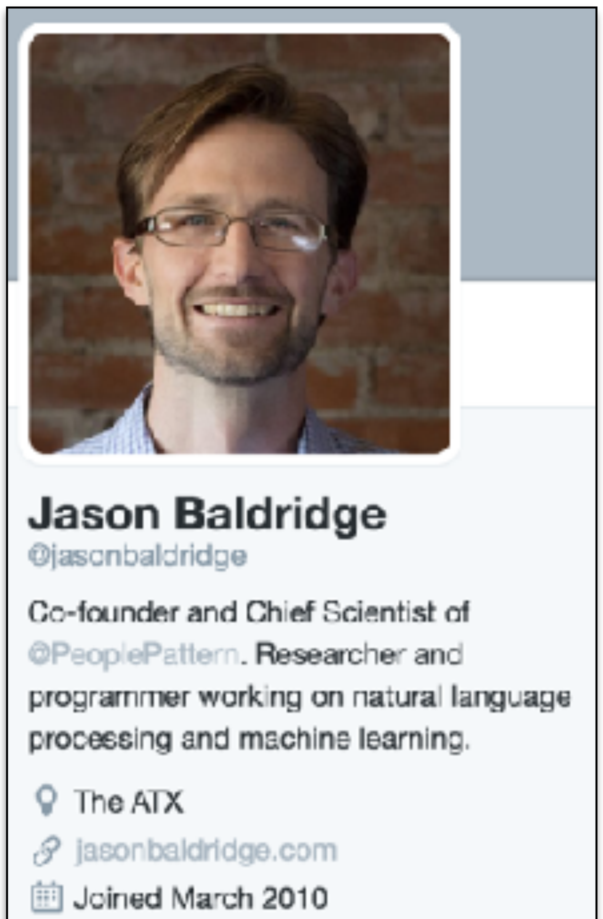
Use actual age (if you can)!



1974

Actual age matters for many tasks.
Annotating a person's birth year is also future proof.

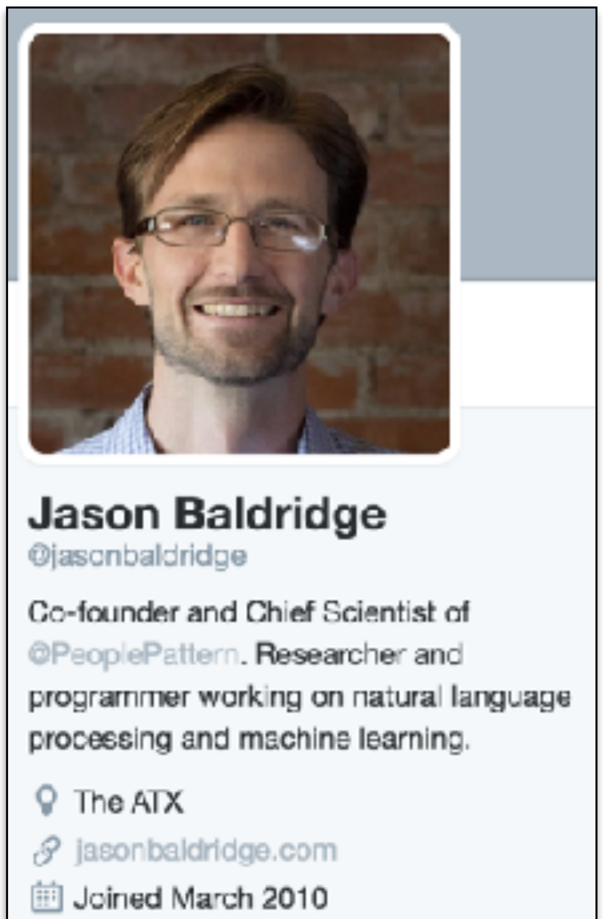
Use actual age (if you can)!




```
{
  "target_uuid": "twitter:119837224",
  "label": "1974",
  "namespace": "birthyear",
  "context": {
    "creator": “John Smith",
    "creator_type": "human",
    "confidence": 4,
    "date": "2016-07-18T14:11-0500",
    "note": “Verified by resume."
  }
}
```


Actual age matters for many tasks.
Annotating a person’s birth year is also future proof.

Use actual age (if you can)!




```
{
  "target_uuid": "twitter:119837224",
  "label": "1974",
  "namespace": "birthyear",
  "context": {
    "creator": "John Smith",
    "creator_type": "human",
    "confidence": 4,
    "date": "2016-07-18T14:11-0500",
    "note": "Verified by resume."
  }
}
```


Actual age matters for many tasks.
Annotating a person's birth year is also future proof.

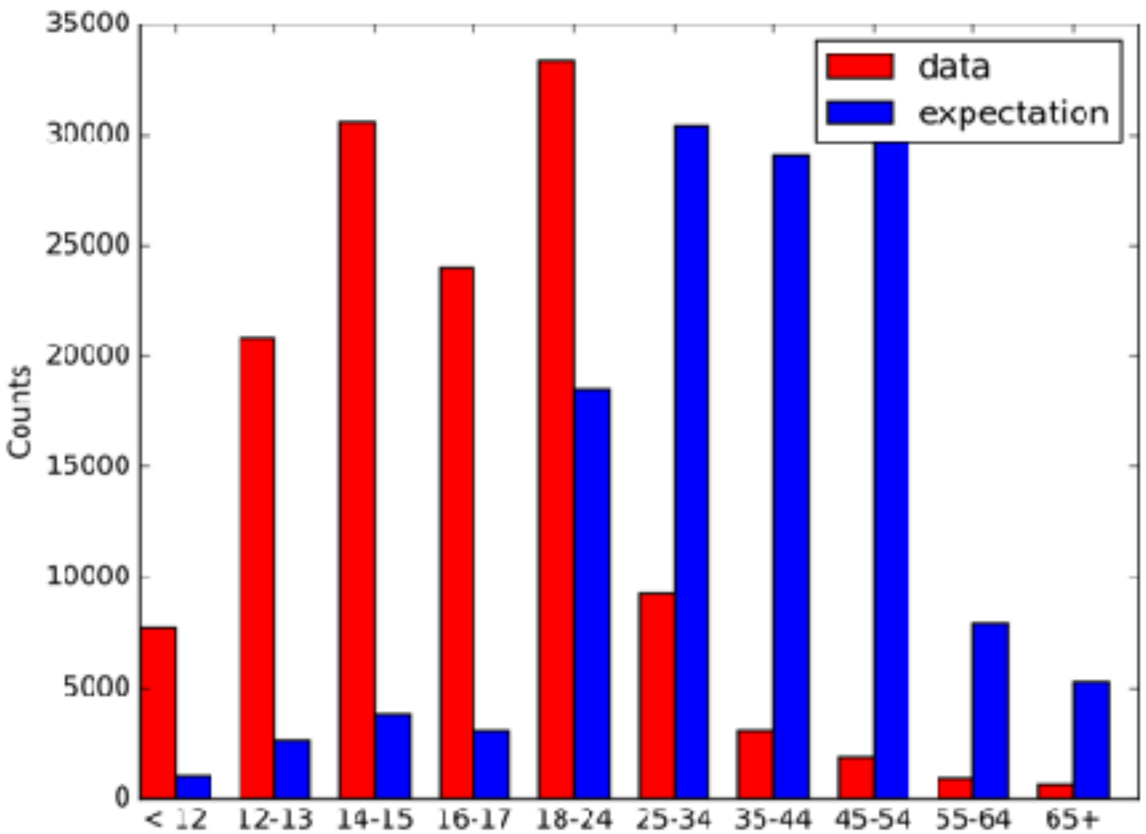
Annotation by patterns

@johnsmith1974

42yo. Dad of 2 amazing girls. Cubs fan. All-around happy guy.

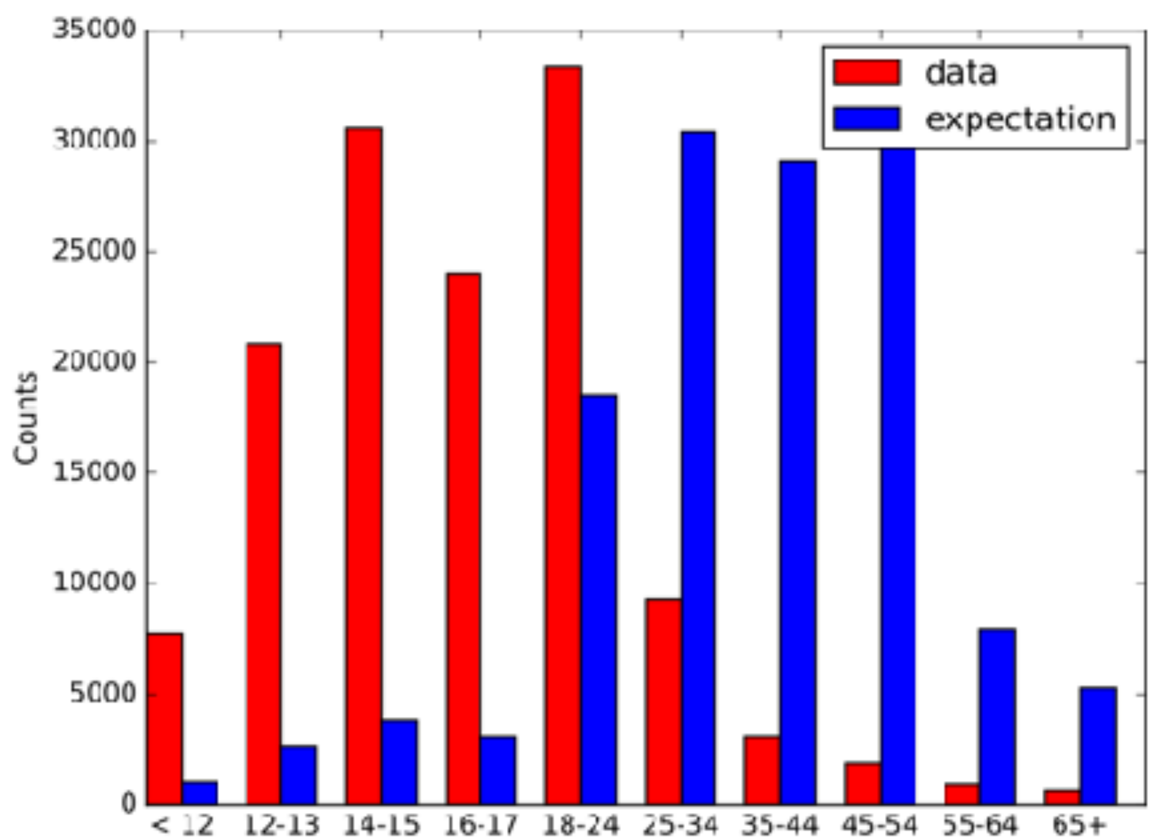
23. Proud Longhorn. Still crazy about Justin Bieber!

Annotation by patterns

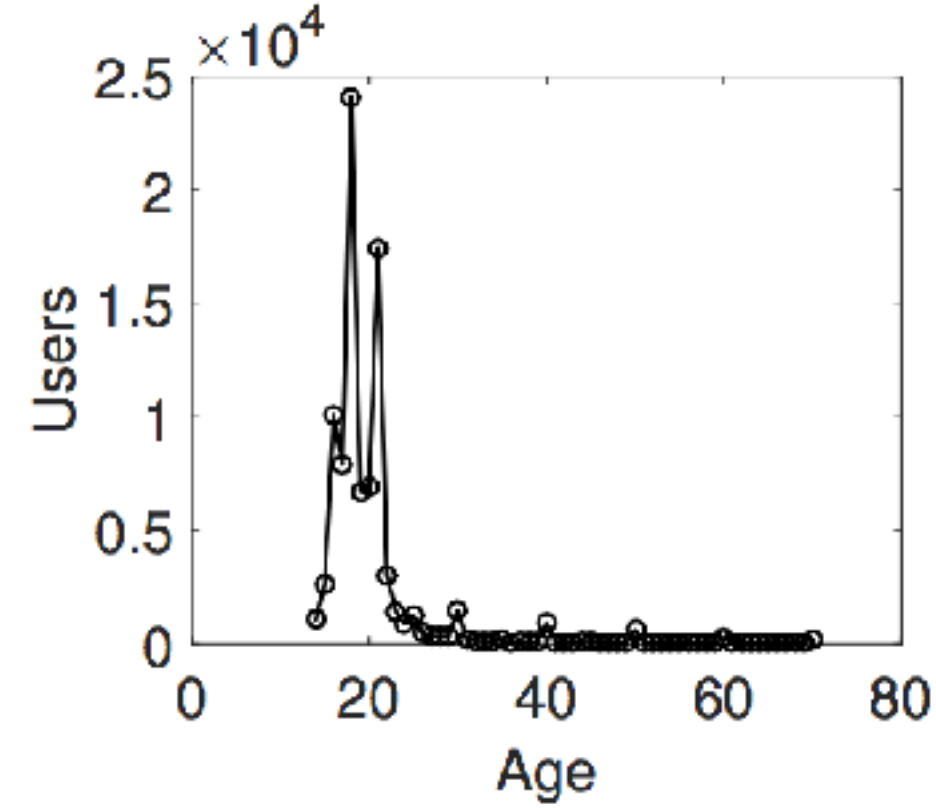


Chamberlain et al (2016).
"Detecting the age of Twitter users."
https://arxiv.org/abs/1601.04621

Annotation by patterns

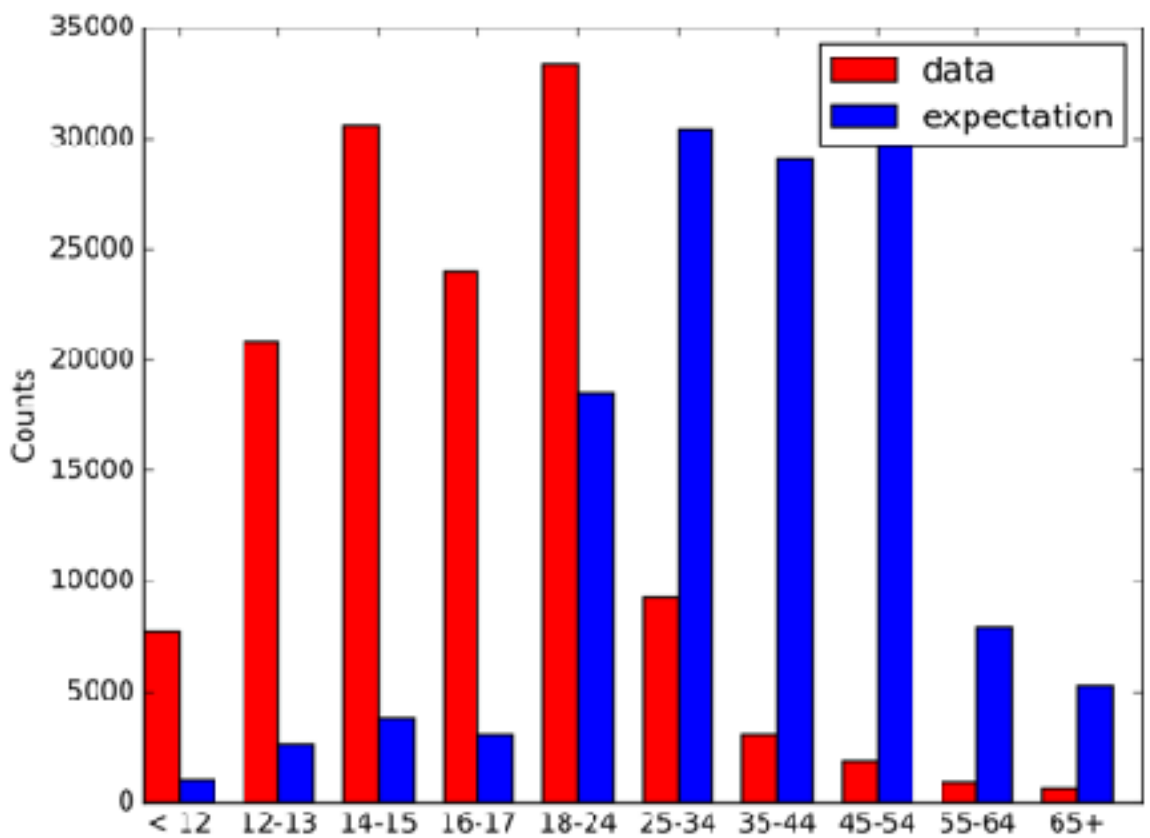


Chamberlain et al (2016).
"Detecting the age of Twitter users."
https://arxiv.org/abs/1601.04621

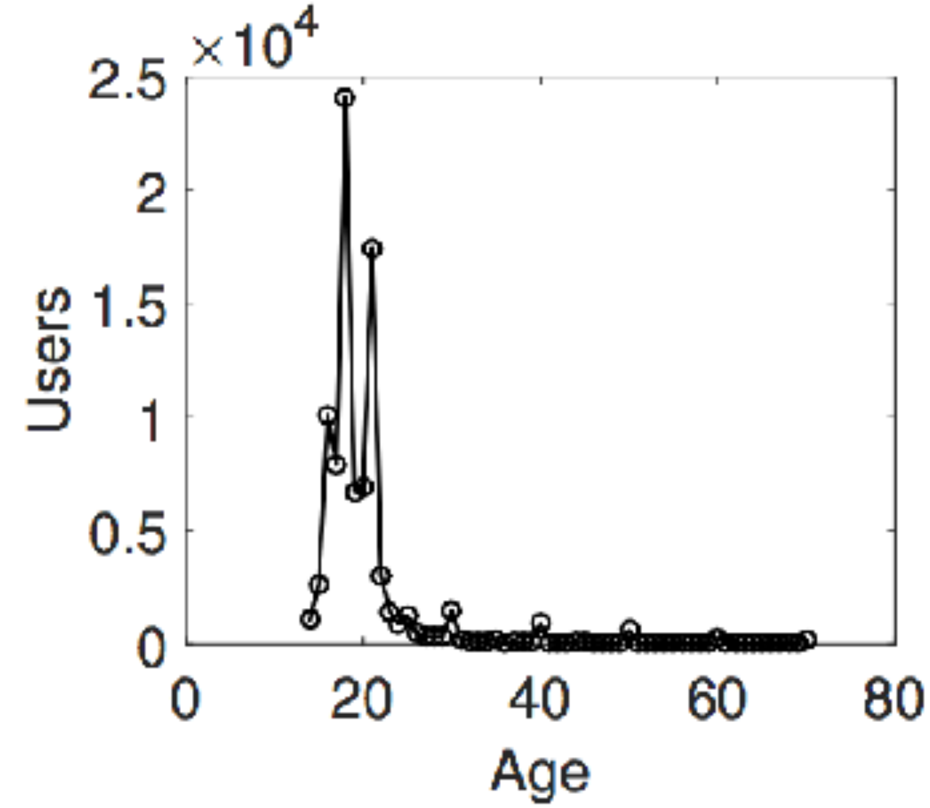


Zhang et al (2016).
"Your age is no secret."
http://www.public.asu.edu/~huanliu/papers/AgeEstICWSM16.pdf

Annotation by patterns



Chamberlain et al (2016).
"Detecting the age of Twitter users."
https://arxiv.org/abs/1601.04621



Zhang et al (2016).
"Your age is no secret."
http://www.public.asu.edu/~huanliu/papers/AgeEstICWSM16.pdf

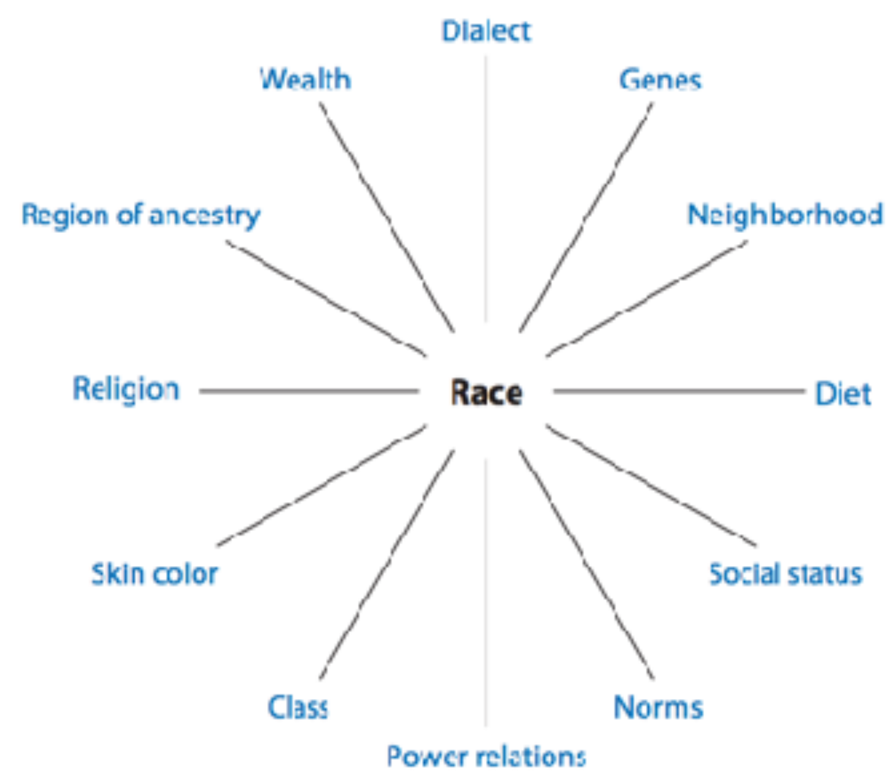
Many high-quality annotations can be extracted,
but the resulting age distributions are highly skewed.

Actual or apparent race?

It’s well beyond skin-deep.
http://bcomposes.com/2015/11/08/improving-race-relations-a-path-forward/

Actual or apparent race?

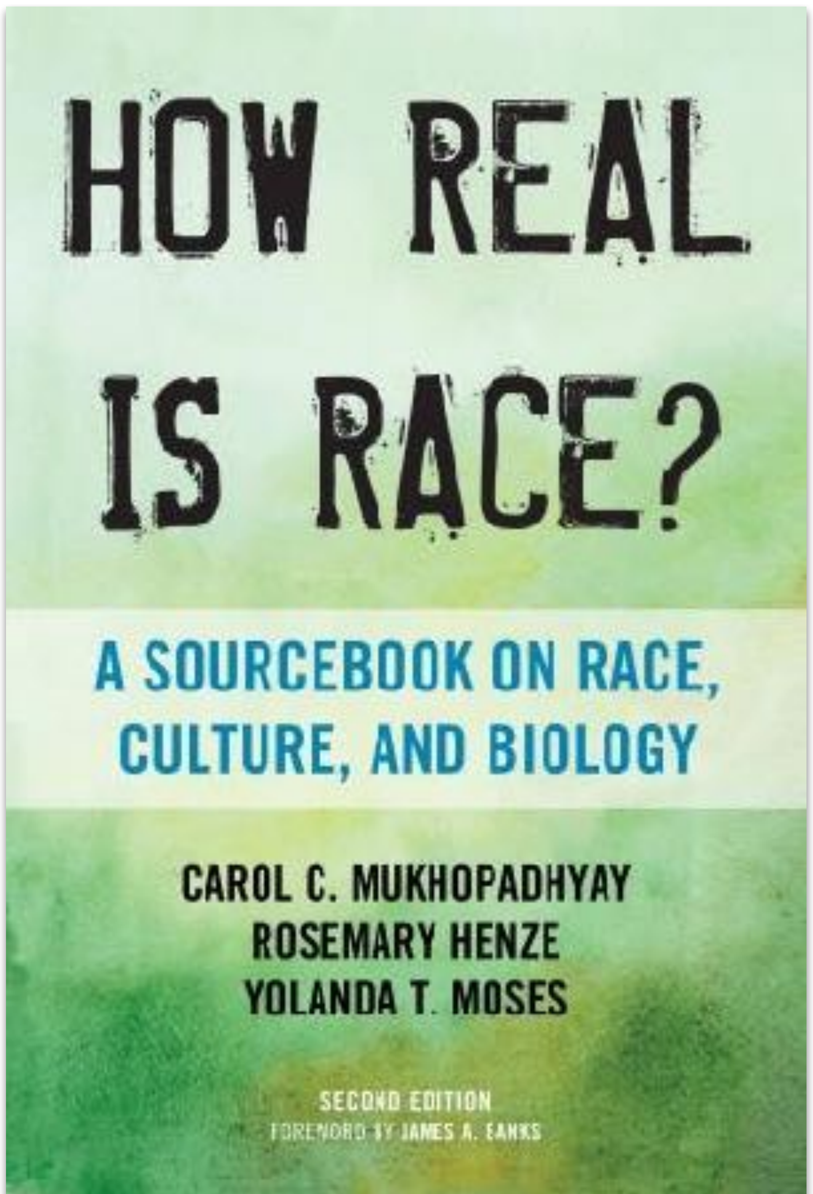
It's well beyond skin-deep.
http://bcomposes.com/2015/11/08/improving-race-relations-a-path-forward/



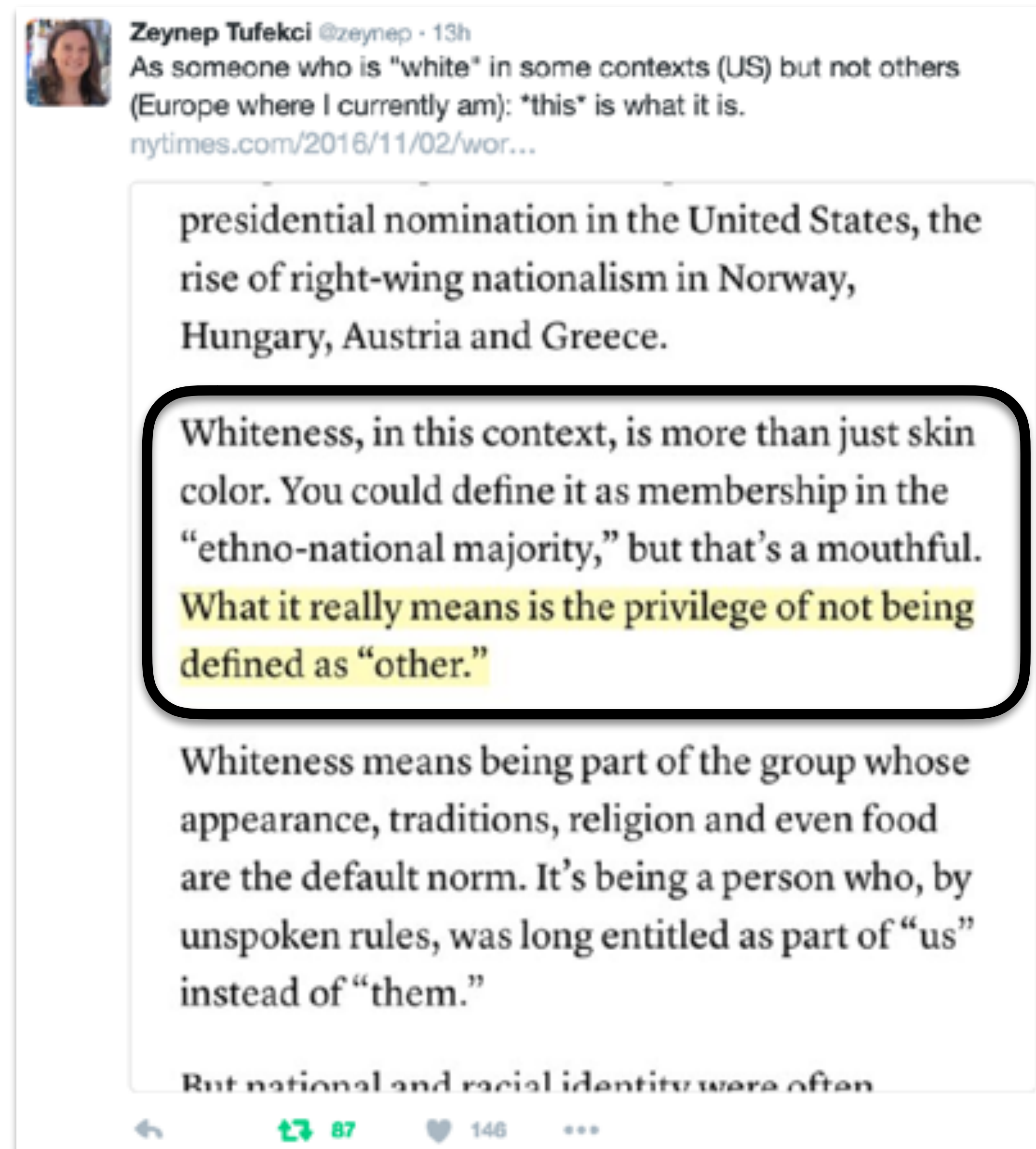
Sen and Wasow (2016).
"Race as a bundle of sticks: Designs that Estimate Effects of Seemingly Immutable Characteristics"

It's complicated.

Actual or apparent race?



http://www.understandingrace.org/



Zeynep Tufekci @zeynep · 13h
As someone who is "white" in some contexts (US) but not others (Europe where I currently am): *this* is what it is.
nytimes.com/2016/11/02/wor...

presidential nomination in the United States, the rise of right-wing nationalism in Norway, Hungary, Austria and Greece.

Whiteness, in this context, is more than just skin color. You could define it as membership in the "ethno-national majority," but that's a mouthful. What it really means is the privilege of not being defined as "other."

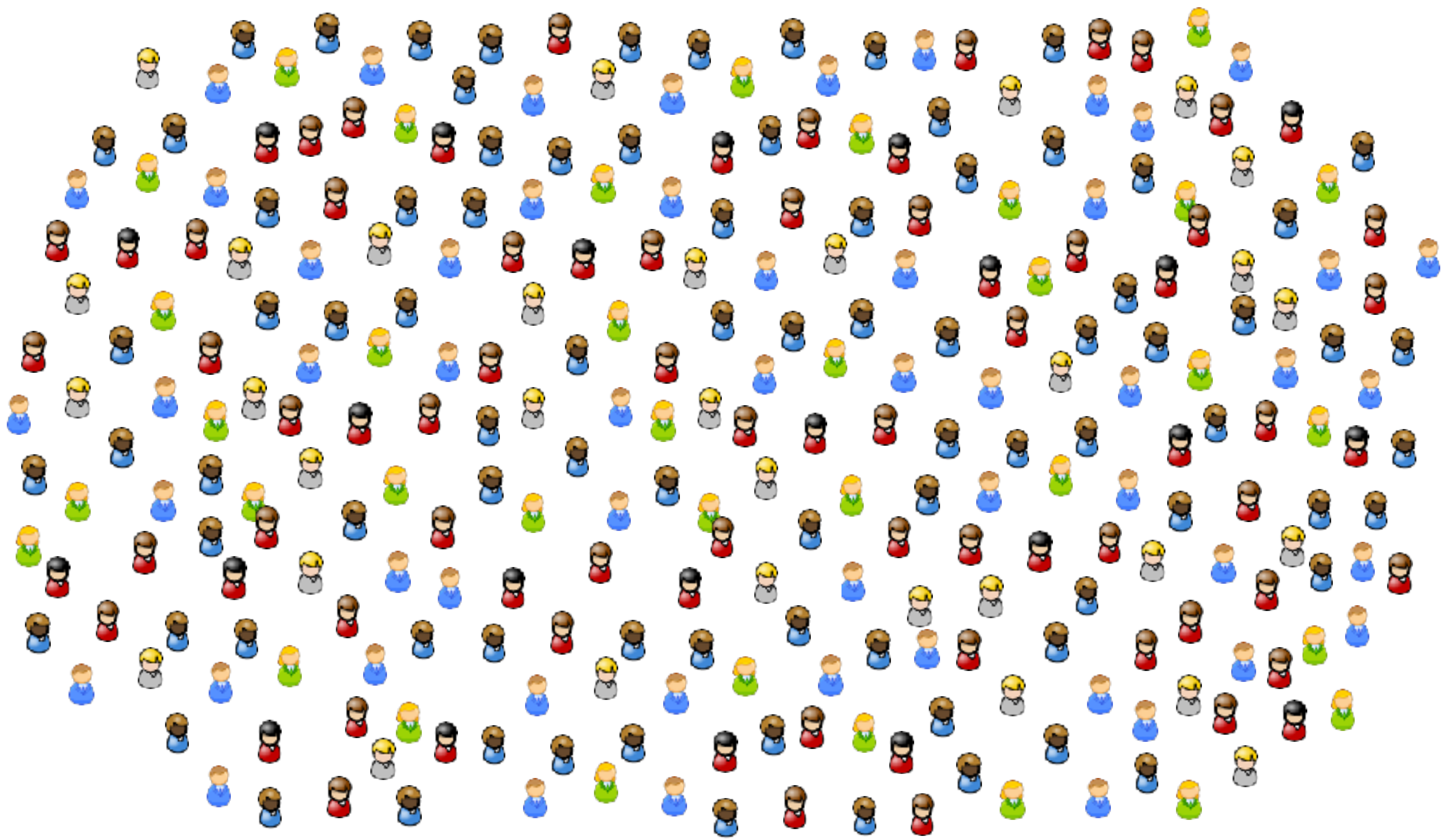
Whiteness means being part of the group whose appearance, traditions, religion and even food are the default norm. It's being a person who, by unspoken rules, was long entitled as part of "us" instead of "them."

87 146

https://twitter.com/zeynep/status/793805186892492800

There is no "actual". Race is as a concept is fluid across individuals, communities, and time periods.

Complete coverage

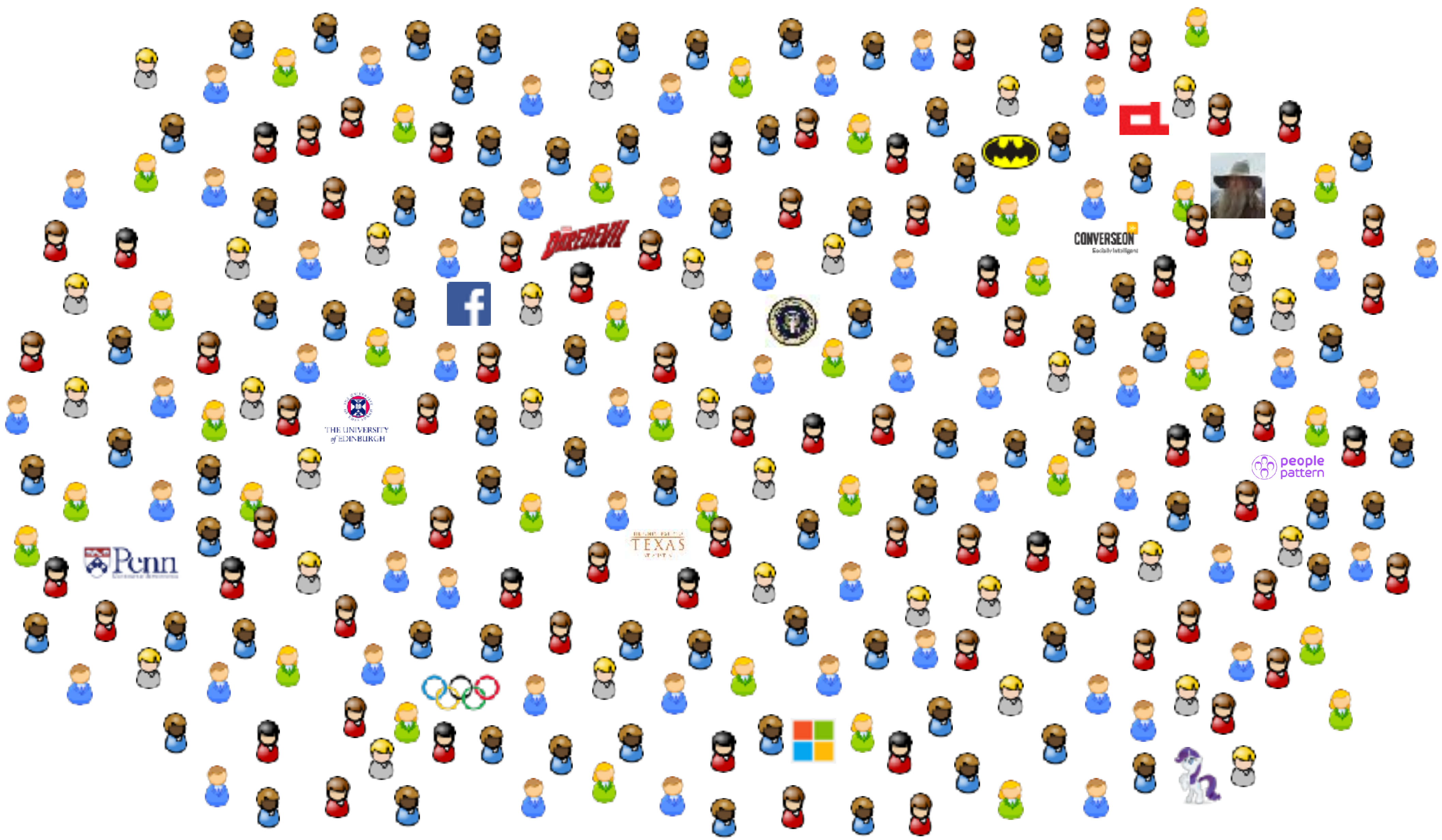


We can't arbitrarily ignore profiles or certain class labels.

Complete coverage

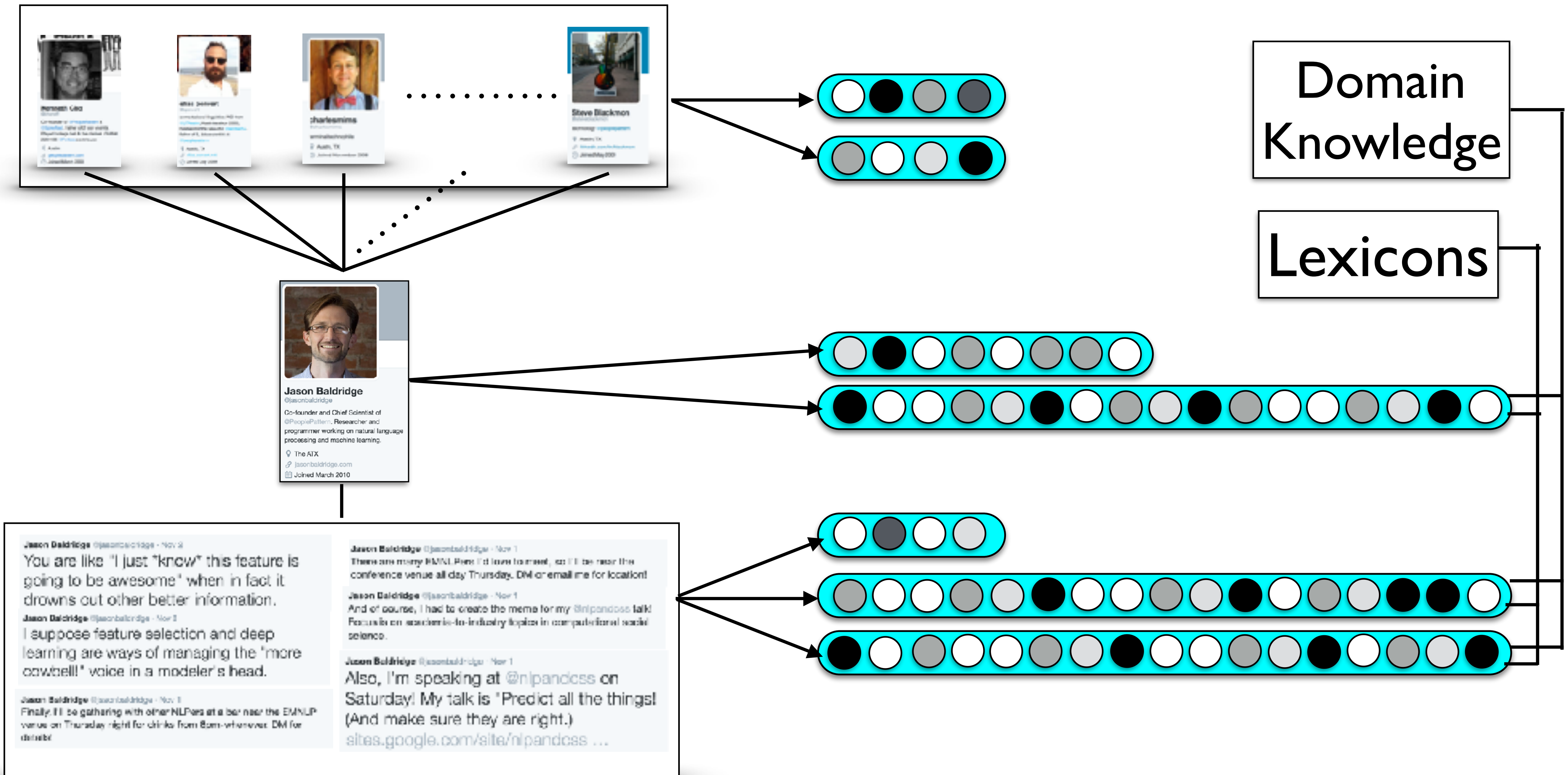
We can't arbitrarily ignore profiles or certain class labels.

Complete coverage



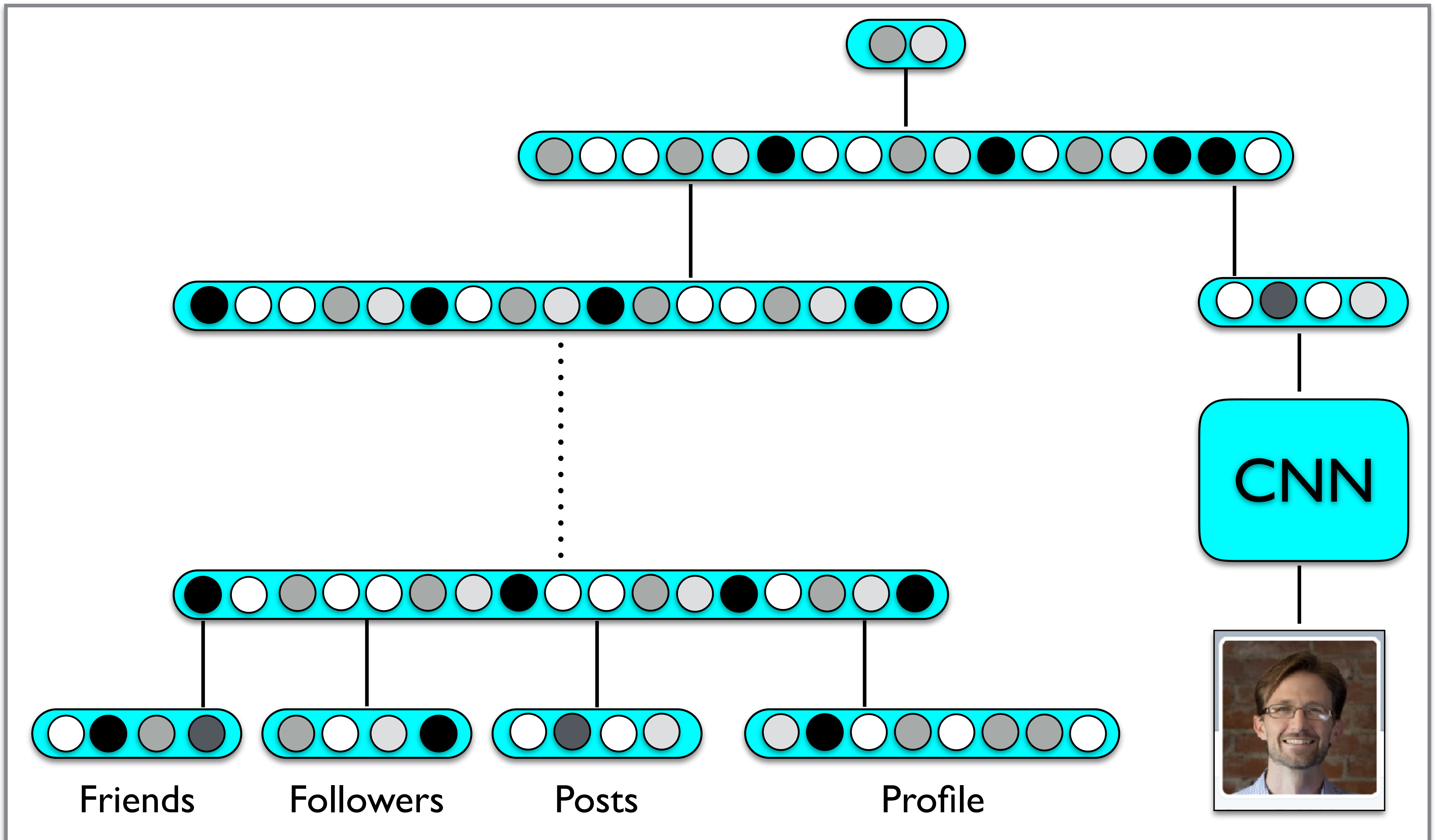
Not all profiles are people!
Account type classifiers are necessary.

Common featurization



Strong common feature representations facilitate deployment, improvements and tackling new tasks.

Images too!



Low dimensional features can be combined with a convolutional neural network for profile picture.

Image manipulation

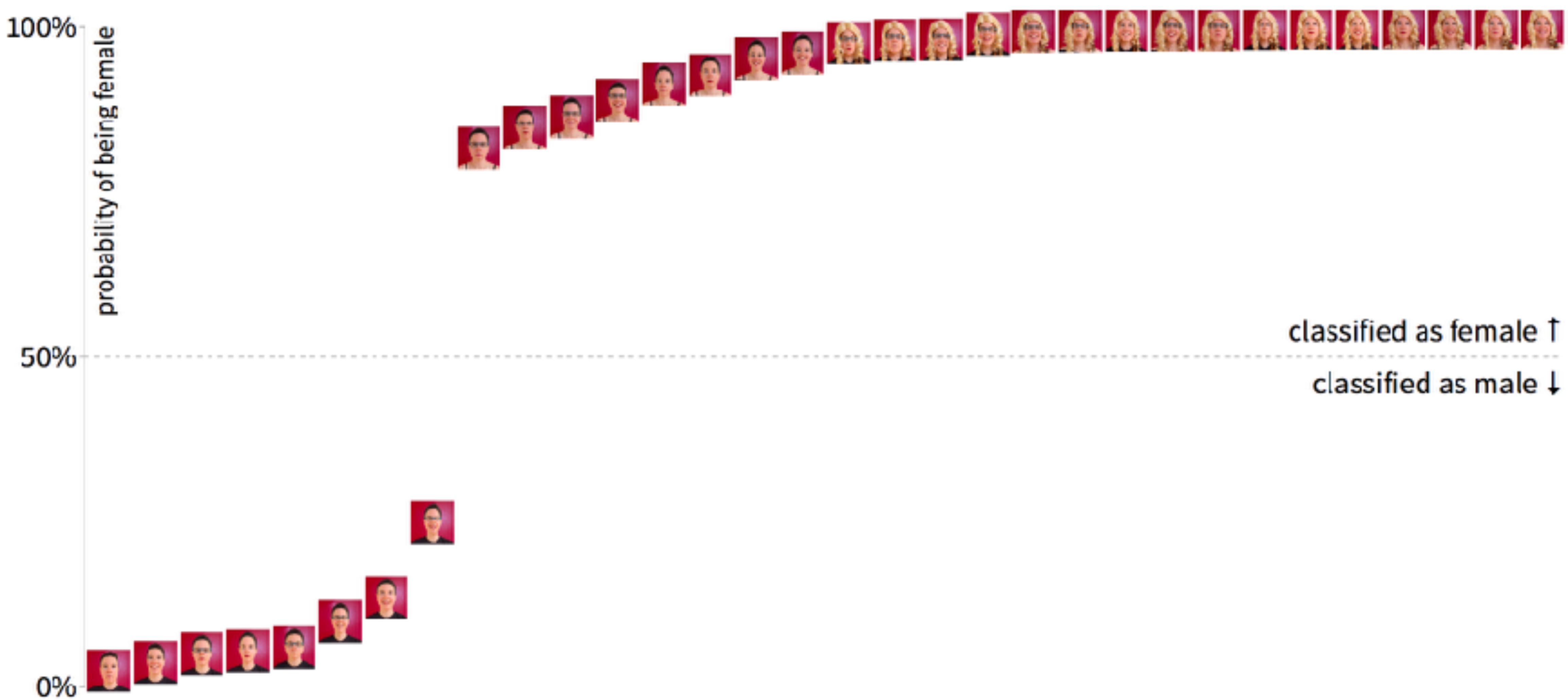
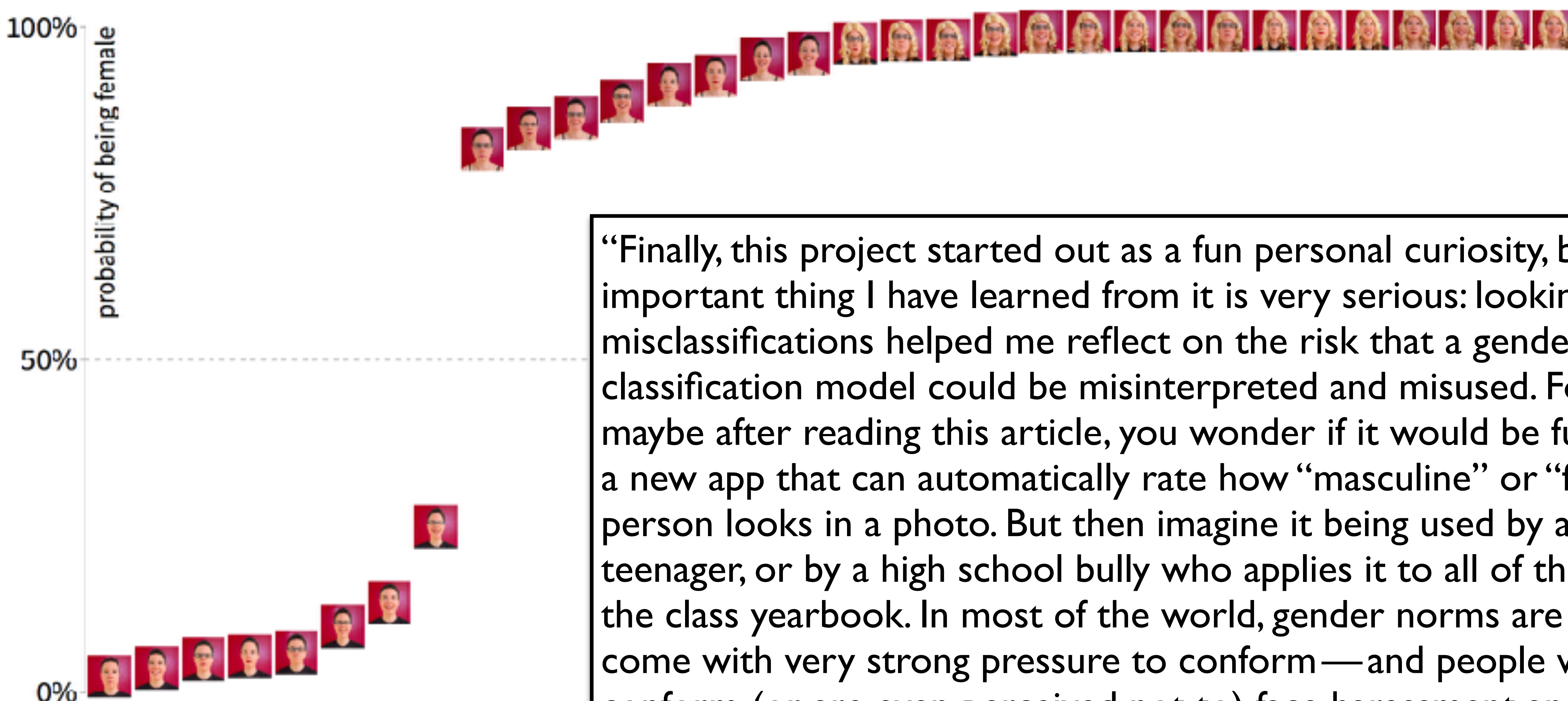


Image classifiers quickly pick up on stereotypes.

https://medium.com/@kerryrodden/is-that-a-boy-or-a-girl-cb93abbae6da

Image manipulation



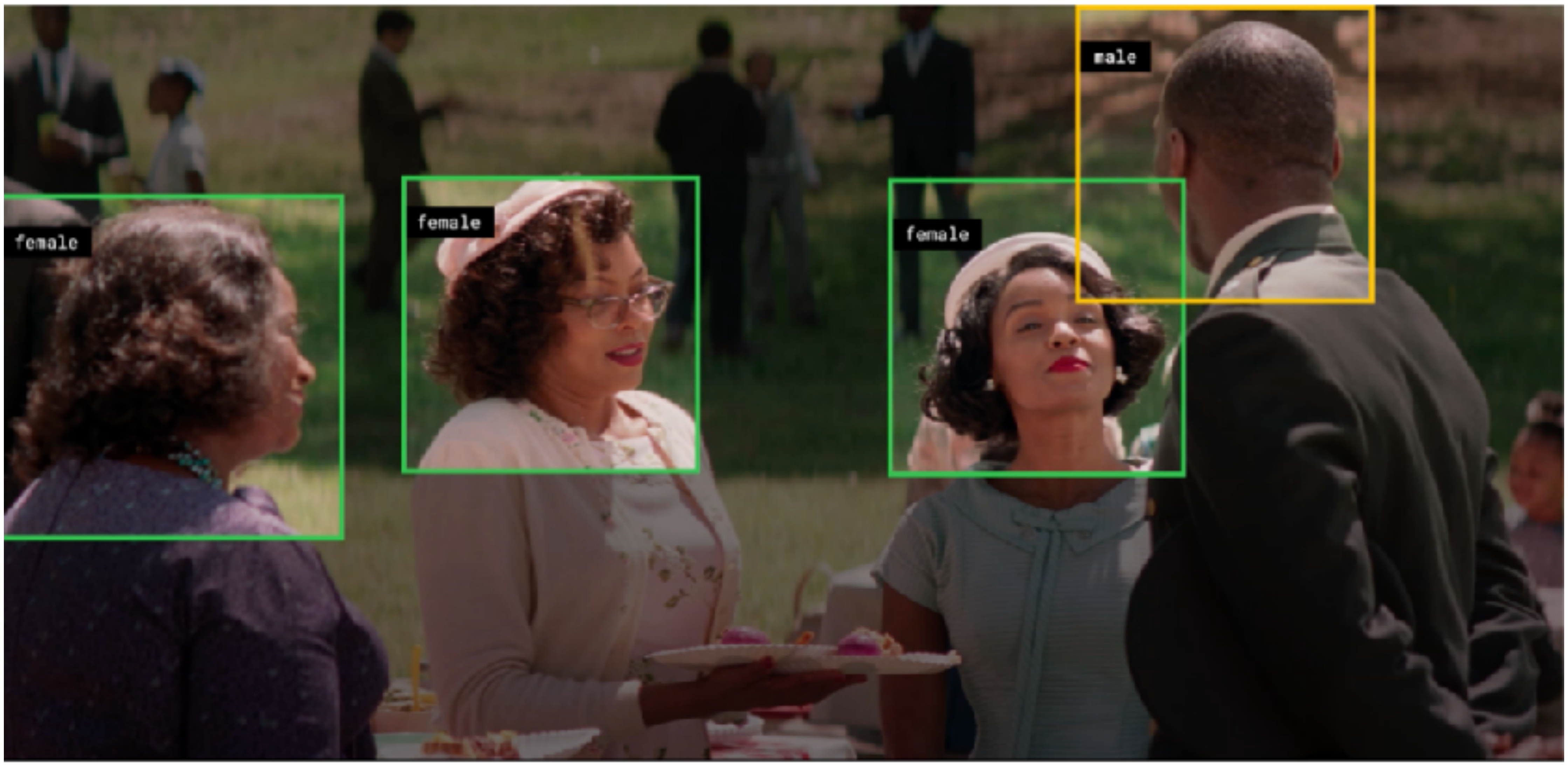
"Finally, this project started out as a fun personal curiosity, but the most important thing I have learned from it is very serious: looking at the misclassifications helped me reflect on the risk that a gender classification model could be misinterpreted and misused. For example, maybe after reading this article, you wonder if it would be fun to create a new app that can automatically rate how "masculine" or "feminine" a person looks in a photo. But then imagine it being used by an insecure teenager, or by a high school bully who applies it to all of the photos in the class yearbook. In most of the world, gender norms are rigid, and come with very strong pressure to conform—and people who do not conform (or are even perceived not to) face harassment and threats to their personal safety."

- Kerry Roden

Image classifiers quickly pick up on stereotypes.

https://medium.com/@kerryrodden/is-that-a-boy-or-a-girl-cb93abbae6da

Positive use: GD-IQ



Automating the Geena Davis Inclusion Quotient turned a months long process into real time analysis—-and uncovered important gender patterns in film-making.

https://www.google.com/intl/en/about/main/gender-equality-films/

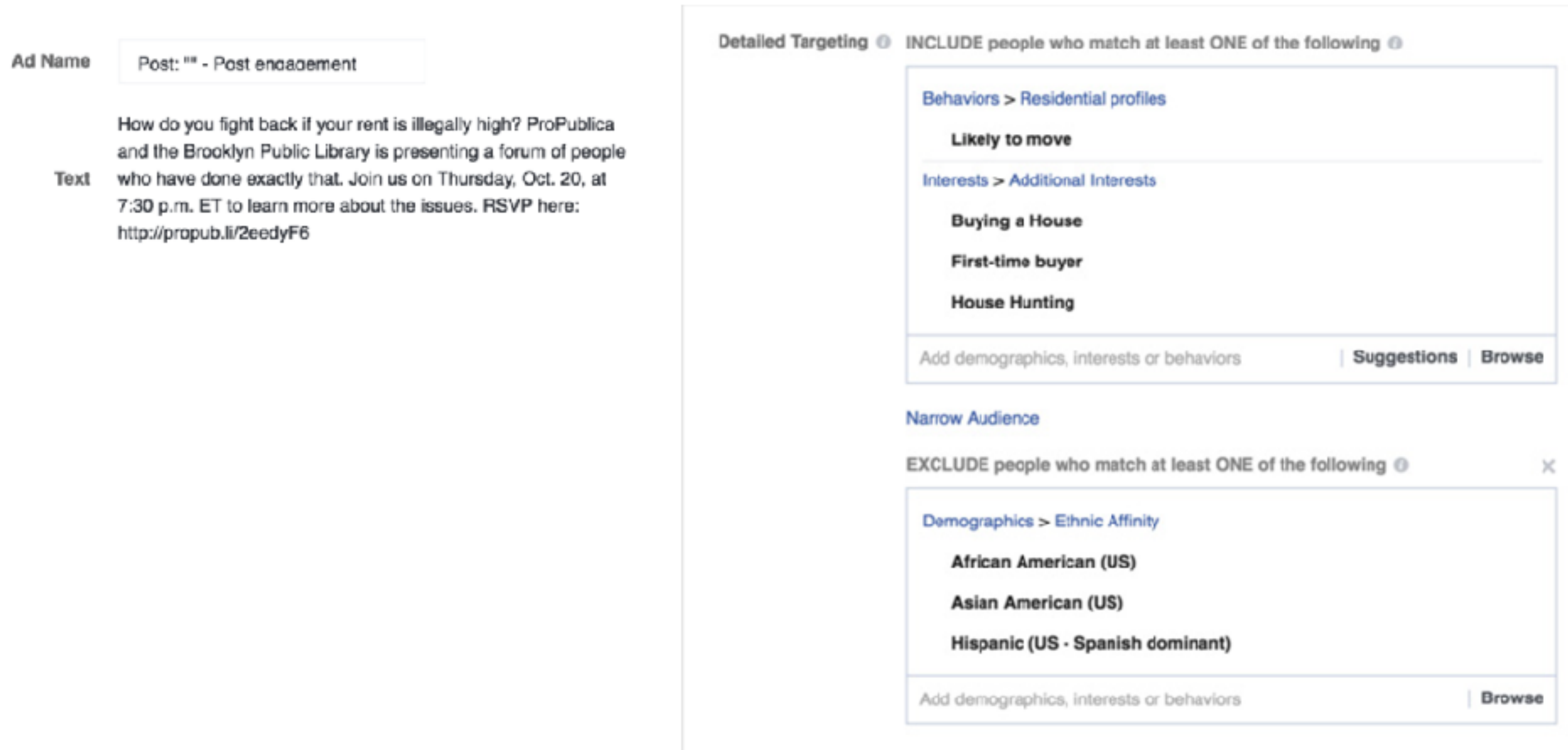
Handle with care

Demographic ad creation and targeting

Predictive policing

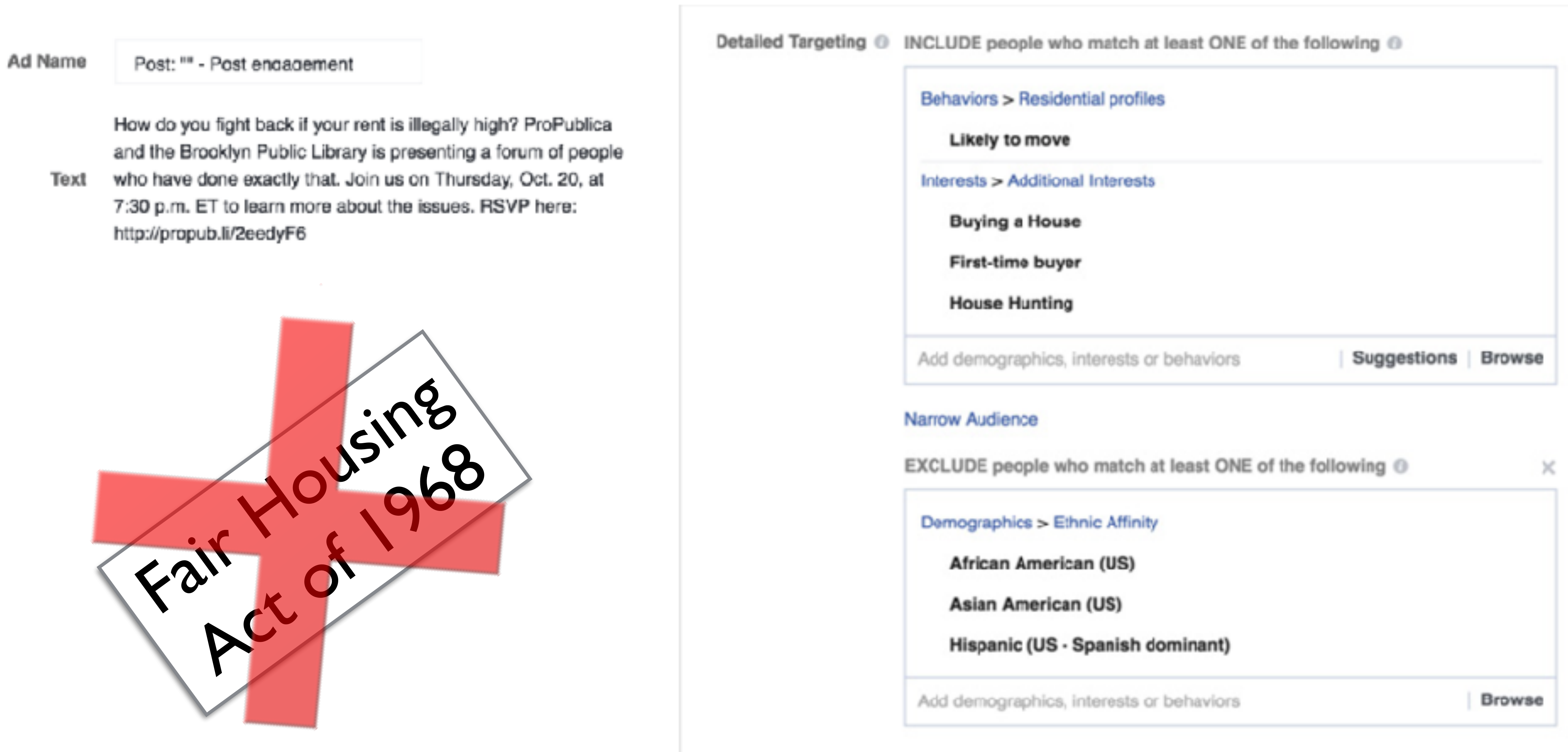
Models and algorithms have the potential to reinforce or ameliorate existing social biases

Discrimination pitfalls



ProPublica was able to create a Facebook ad regarding housing that excluded African-Americans & Hispanics.

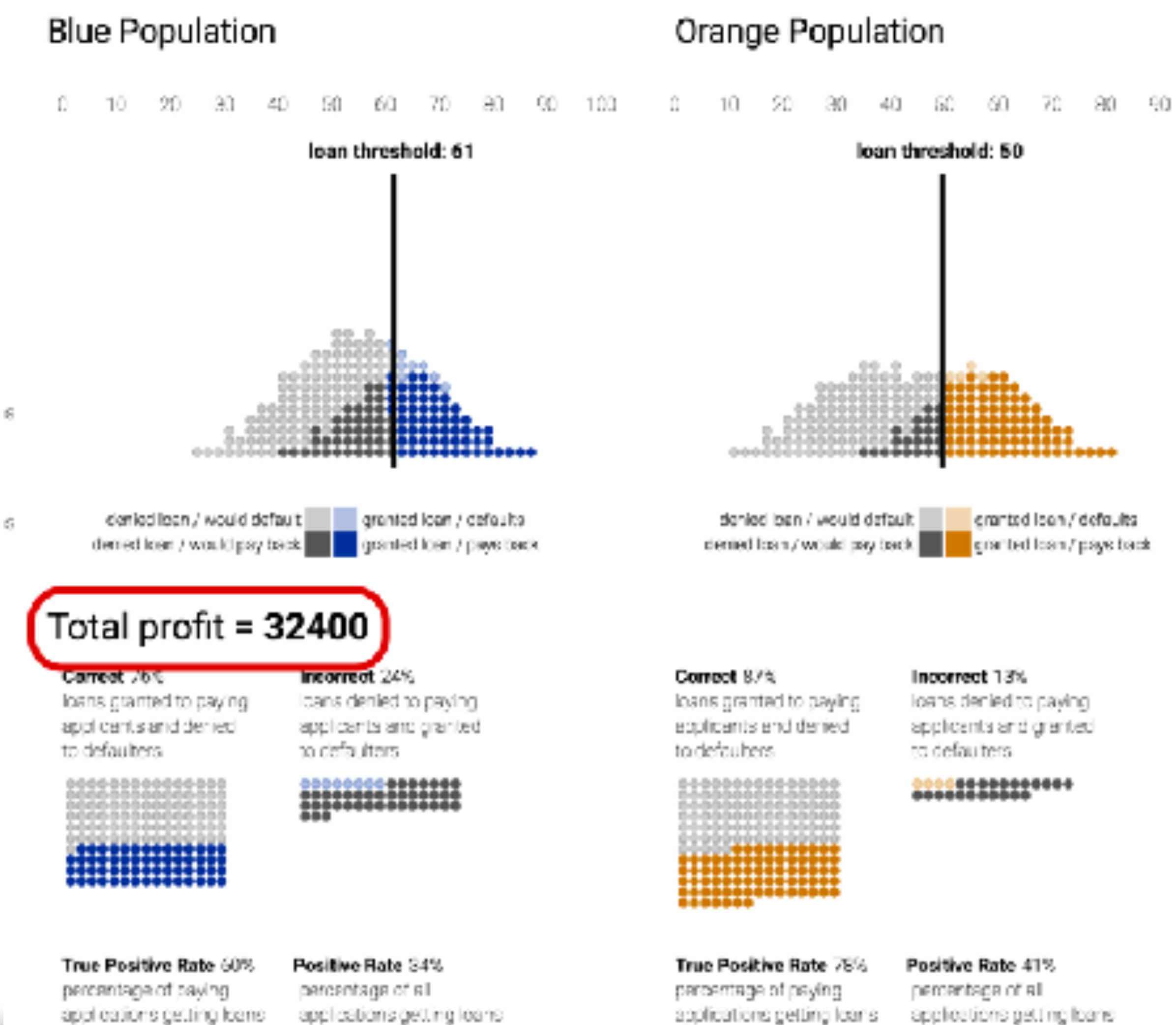
Discrimination pitfalls



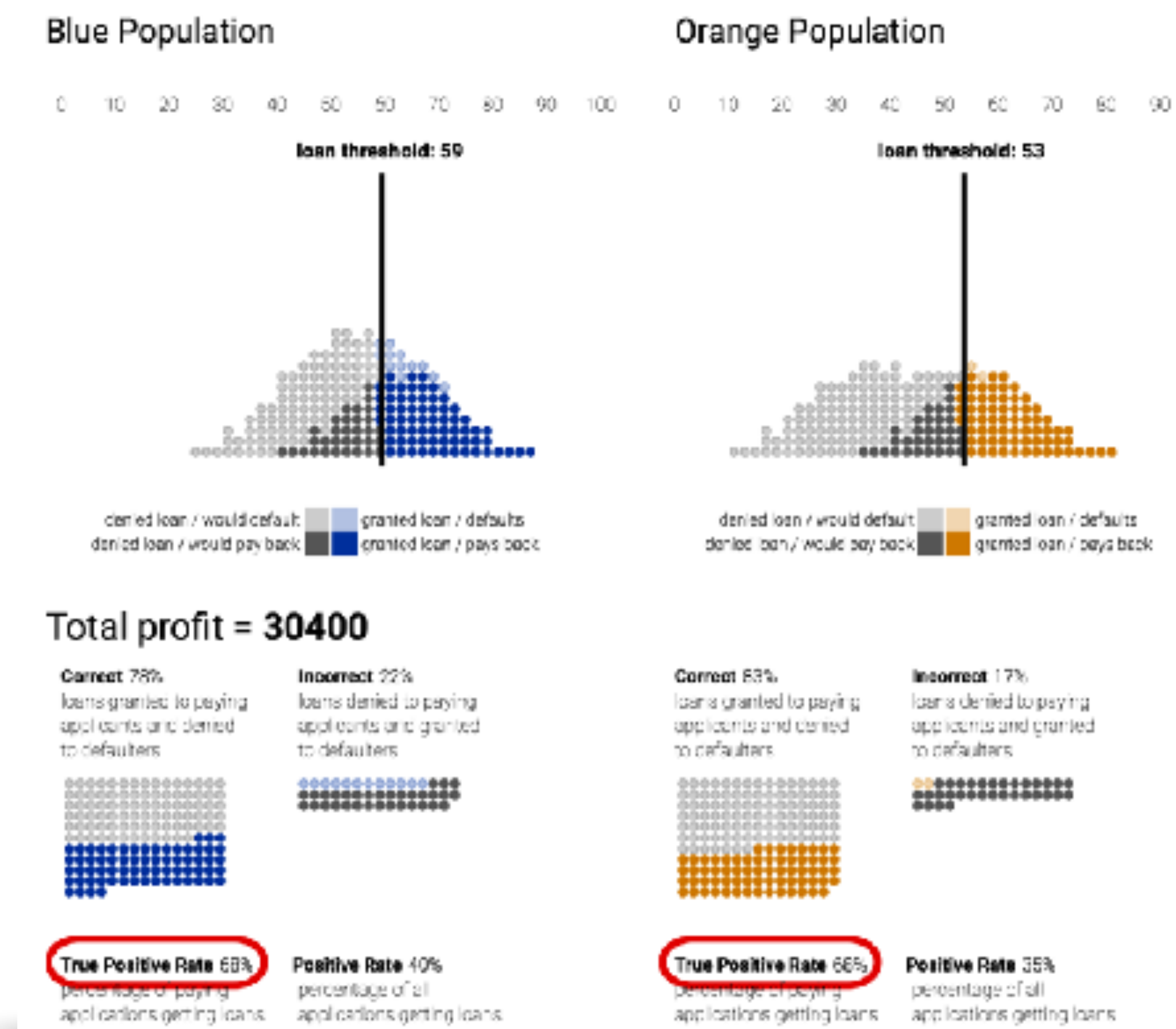
ProPublica was able to create a Facebook ad regarding housing that excluded African-Americans & Hispanics.

ML for anti-discrimination

Maximize profits



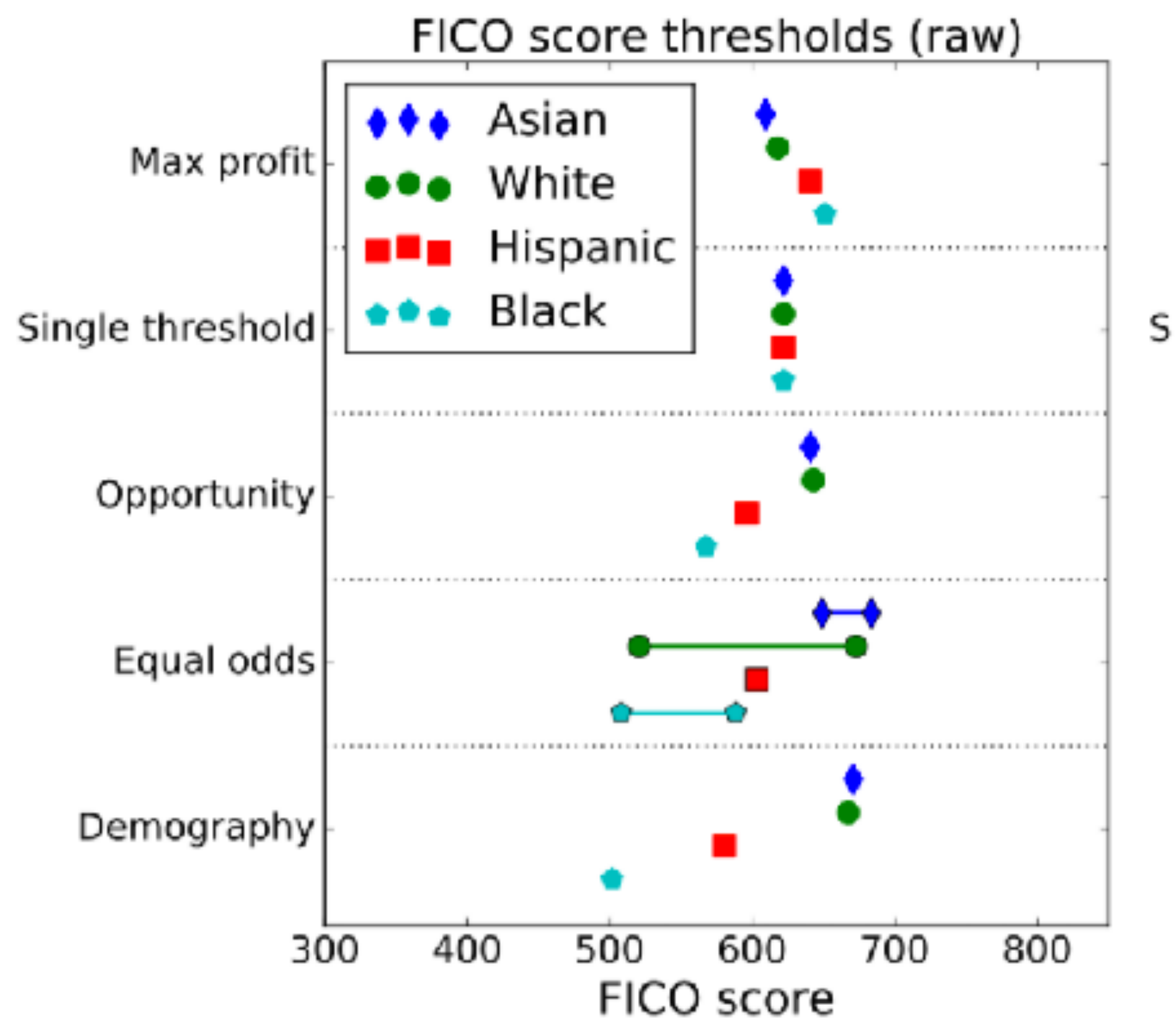
Equal opportunity



Use thresholds that attempt to create equality of opportunity across multiple groups, rather than maximizing profits.

Hardt, Price & Srebo (2016). "Equality of Opportunity in Supervised Learning."

ML for anti-discrimination



Setting thresholds requires knowing the protected class values. Fairness in lending, hiring, etc. using social data thus requires such predictions.

Hardt, Price & Srebo (2016). "Equality of Opportunity in Supervised Learning."

Dialectal failings

	STANFORD	GATE	ARK
AAVE	61.4	**79.1**	77.5
non-AAVE	74.5	**83.3**	77.9
Δ(+,-)	13.1	4.2	0.4

Table 5: POS tagging accuracies (%)

Jørgensen, Hovy and Søgaard (2015).
"Challenges of studying and processing dialects in social media."

	AAE	White-Aligned
langid.py	13.2%	7.6%
Twitter-1	8.4%	5.9%
Twitter-2	24.4%	17.6%

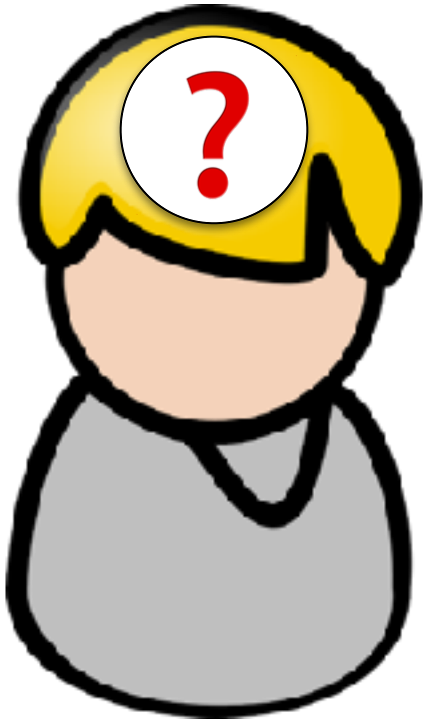
Table 3: Proportion of tweets in AA- and white-aligned corpora classified as non-English by different classifiers. (§4.1)

Parser	AA	Wh.	Difference
SyntaxNet	64.0 (2.5)	80.4 (2.2)	16.3 (3.4)
CoreNLP	50.0 (2.7)	71.0 (2.5)	21.0 (3.7)

Blodgett, Green, and O'Connor (2016).
"Demographic Dialectal Variation in Social Media: A Case Study of African-American English."

Standard (out-of-the-box) NLP tools perform worse on AAE, an example of how current models/methods/methodology could have negative social impact on some groups. [Hovy & Spruit (2016) "The Social Impact of Natural Language Processing"]

Psychographics



Can we characterize psychographic traits via information contained in profile data (posts, graph, images)?

Dark Tetrad

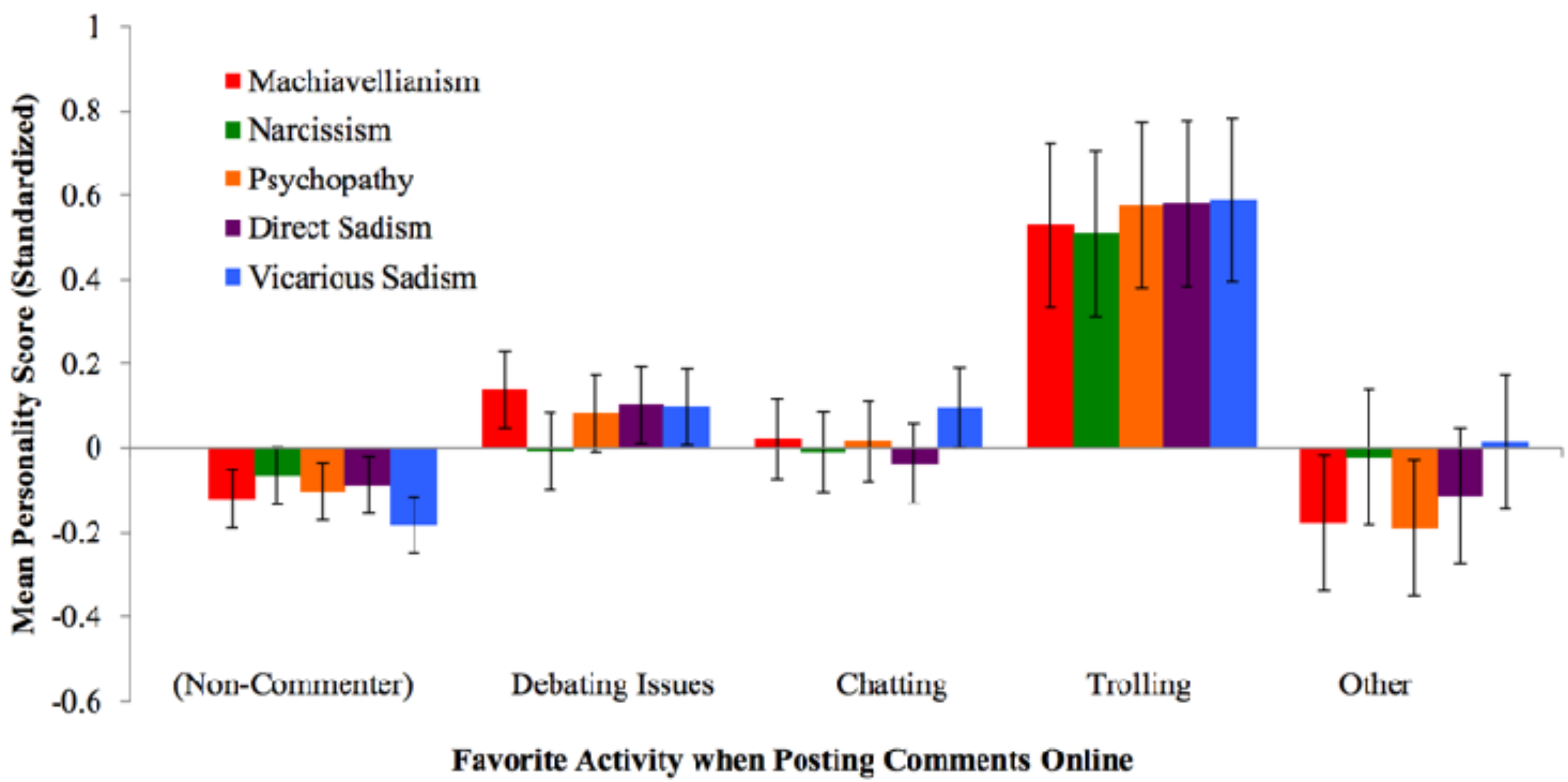


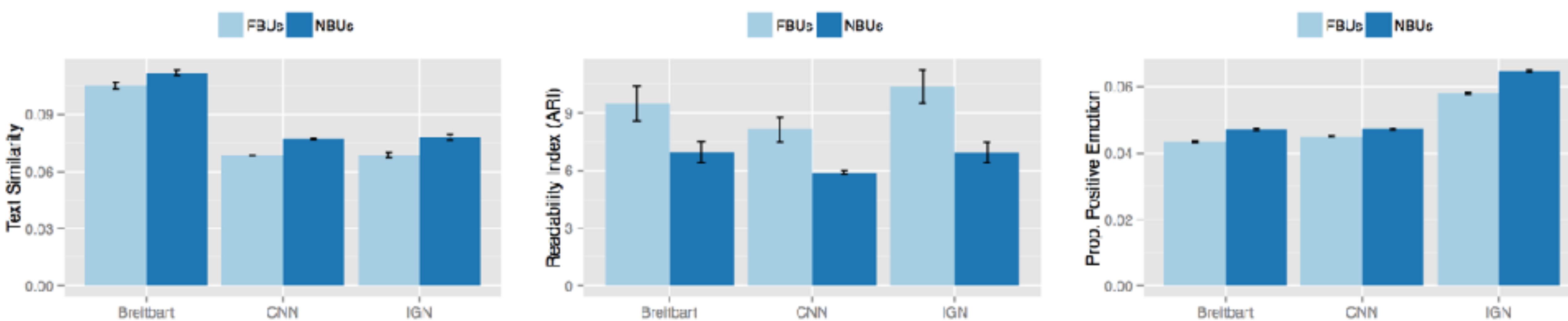
Fig. 1. Dark Tetrad scores as a function of favorite online activity in Study 1. Error bars represent standard errors.

The favorite activity of people who score highly
for the dark tetrad personality types is…. surprise… trolling!

Buckels et al 2014, "Trolls just want to have fun." http://www.sciencedirect.com/science/article/pii/S0191886914000324

Antisocial Behavior Online

FBU: Future banned users
NBU: Never banned users



Comparing banned & normal users (in retrospect): banned users wrote posts that are less relevant, harder to read, and less positive.

Cheng et al. (2015). "Antisocial Behavior in Online Discussion Communities."

Personality classification

- The standard approach is to use the Big 5 personality traits.
- Features include both content and stylistic observations, often derived from LIWC.
- Prediction performance measured against results from standard personality tests.

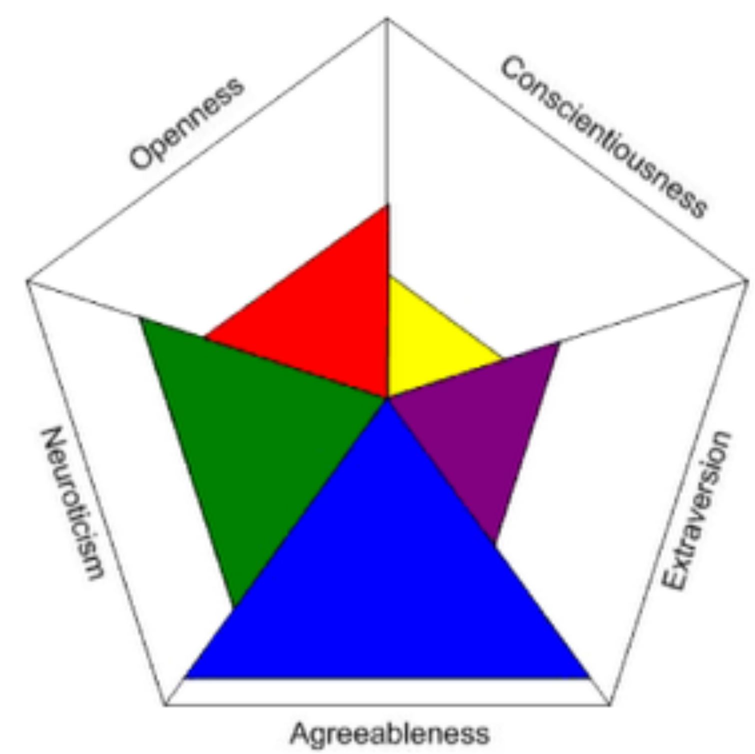
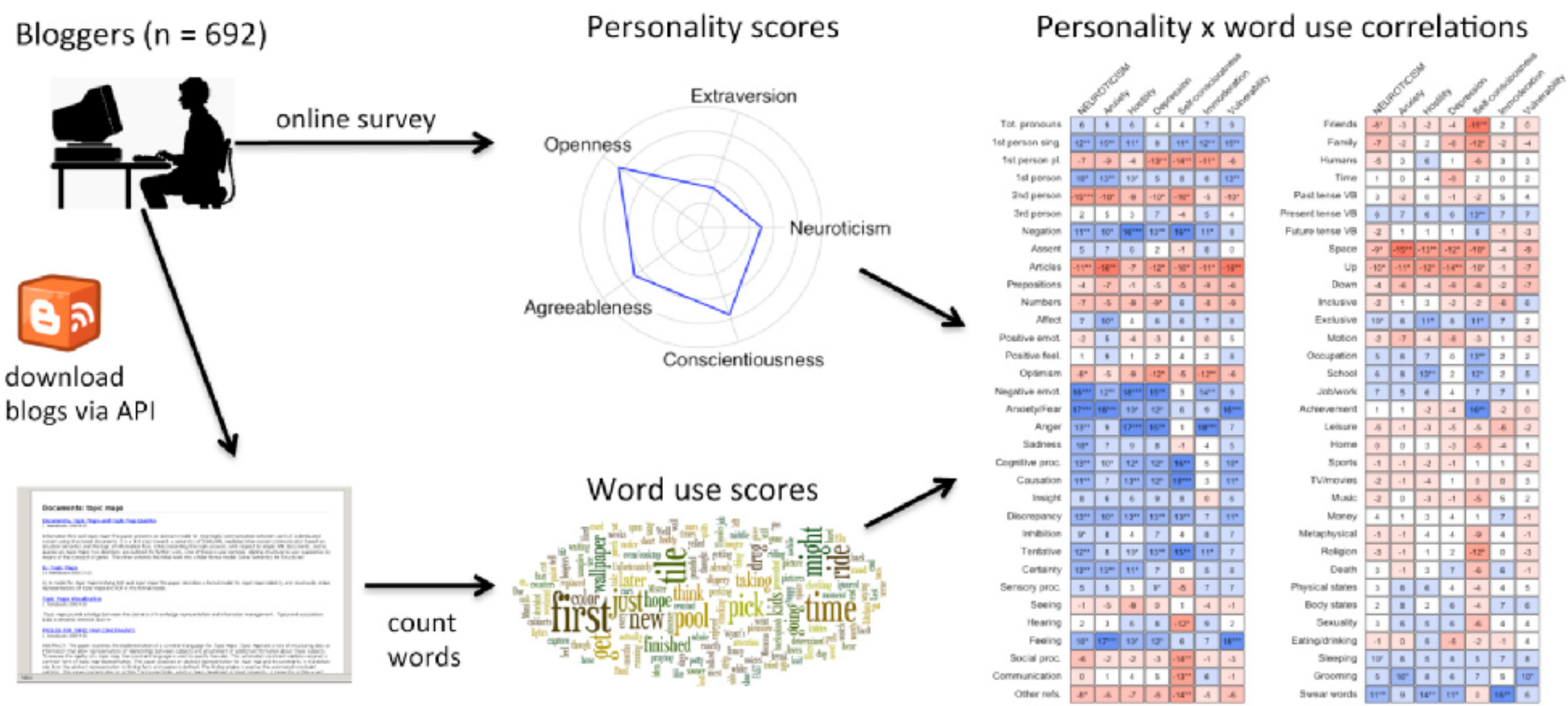


Image from Adali and Golbeck (2015) “Predicting Personality with Social Behavior”

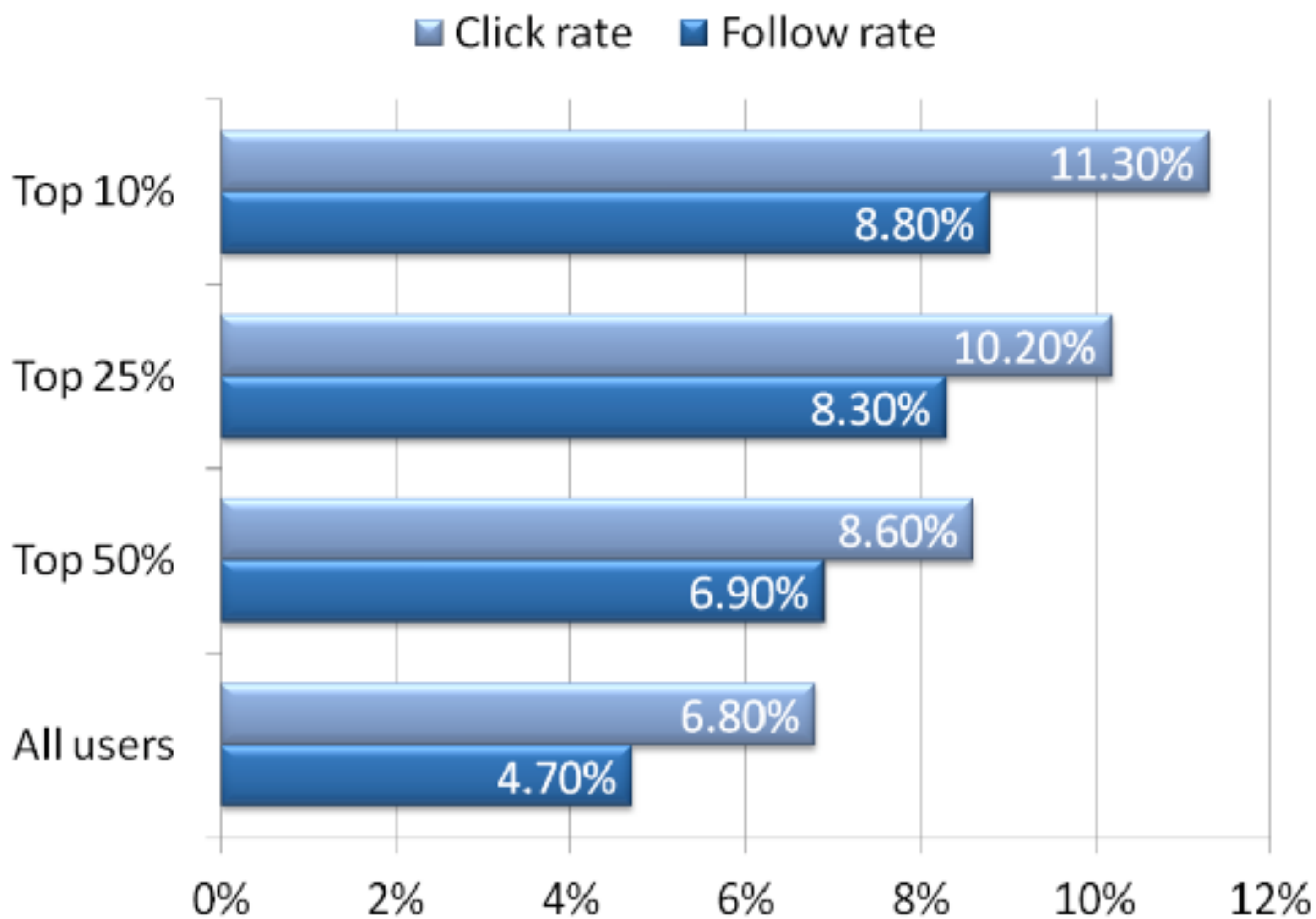
Personality classification



Language production provides a window on personality at scale.

Yarkoni (2010). “Personality in 100,000 Words: A large-scale analysis of personality and word use among bloggers.”

Ad Targeting and Personality



Twitter users whose language indicates higher openness and lower neuroticism are more likely to respond positively to an ad.

Chen et al. (2015). "Making Use of Derived Personality: The Case of Social Media Ad Targeting."

McAdams’ three levels of personality

Level 1 Social Actor	**Dispositional Traits** General tendencies, e.g Big Five traits.
Level 2 Motivated Agent	**Characteristic adaptations** Beliefs, desires, coping mechanisms.
Level 3 Autobiographical author	**Narrative Identity** Life stories that define meaning and purpose.

Personality is more than a weighted combination over five categories! See also Brian Little’s “Me, Myself and Us” and Jonathan Haidt's “The Happiness Hypothesis.”

McAdams 1995, “What do we know when we know a person?”. https://www.sesp.northwestern.edu/docs/publications/557464623490a3fc35faeb.pdf

Moral foundations theory

Care / Harm

Loyalty / Betrayal

Fairness / Cheating

Authority / Subversion

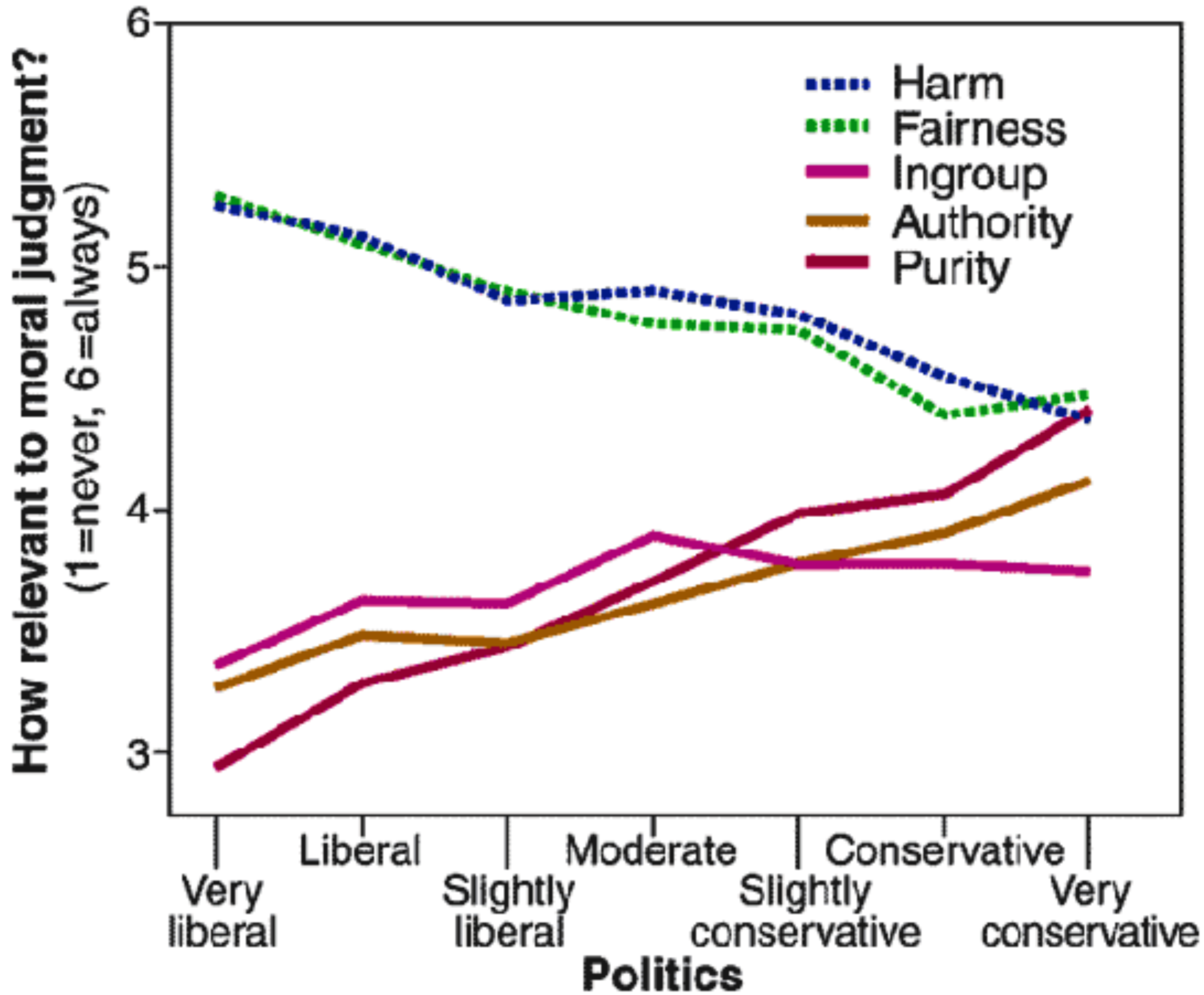
Liberty / Oppression

Sanctity / Degradation

Six fundamental moral foundations

Haidt (2012). "The Righteous Mind."

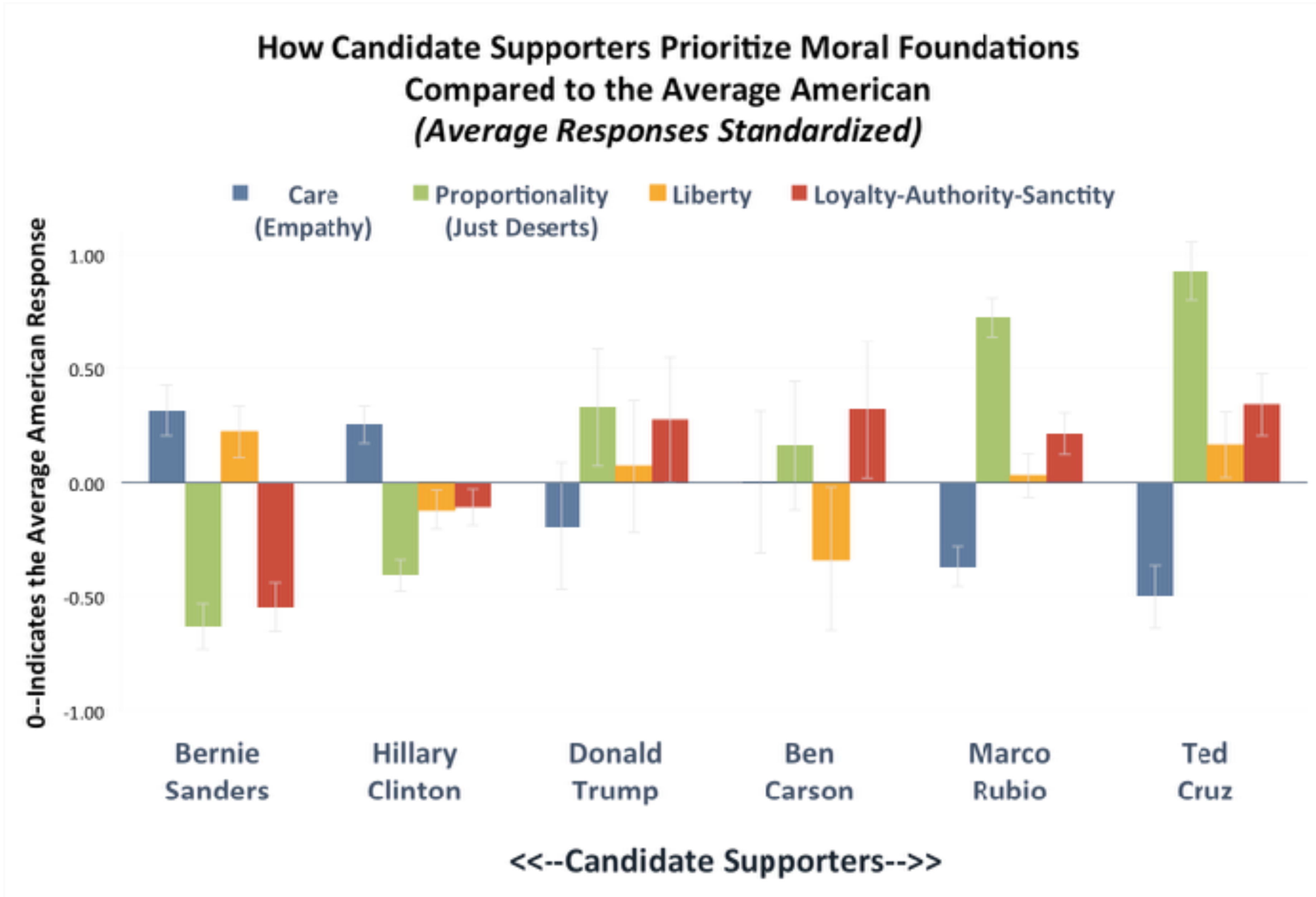
Moral foundations theory



Moral foundations and political leaning.

http://www.yourmorals.org/

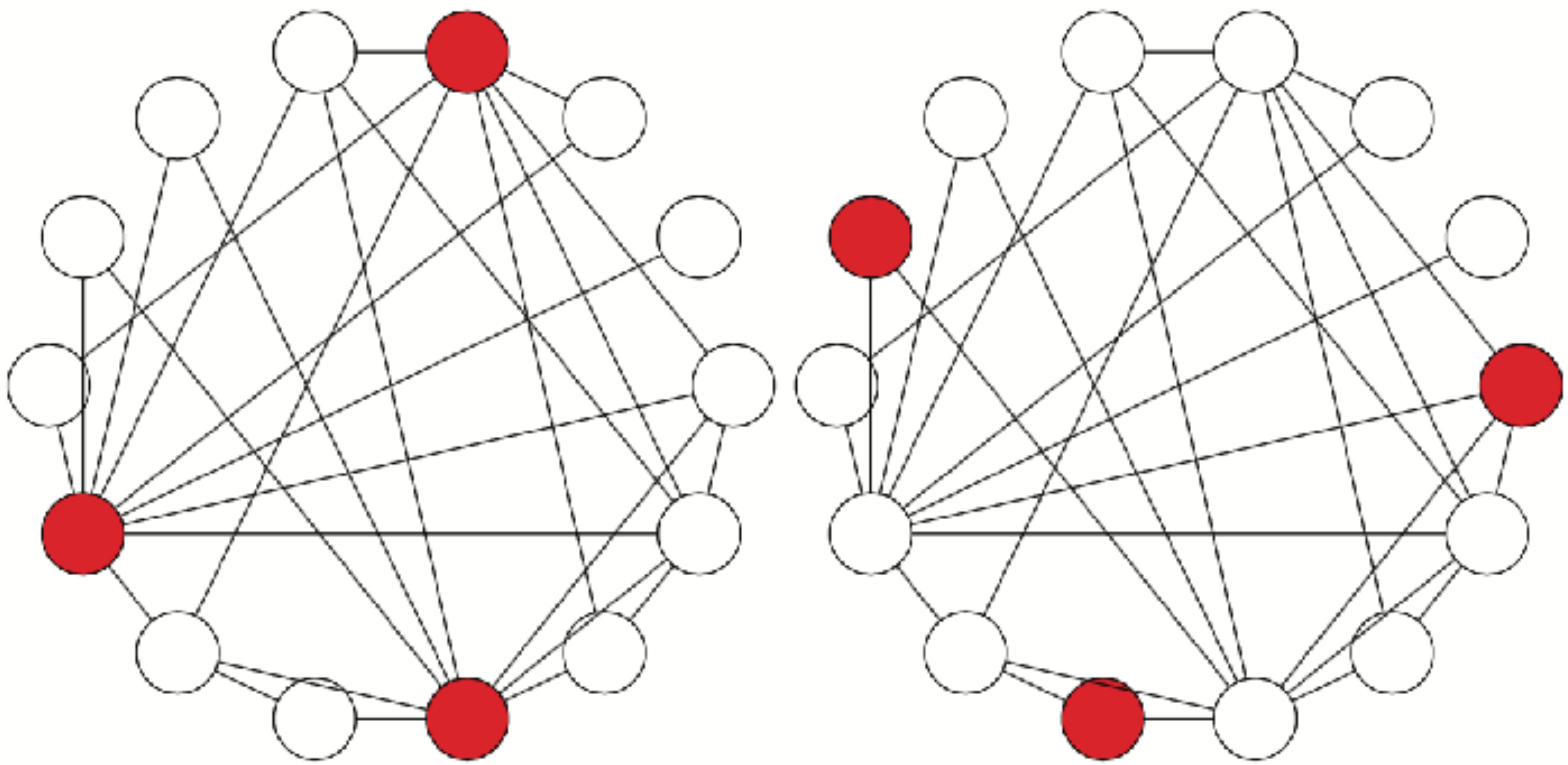
Moral foundations



US 2016 presidential candidates and moral foundation differences.

http://righteousmind.com/presidentialprimaries/

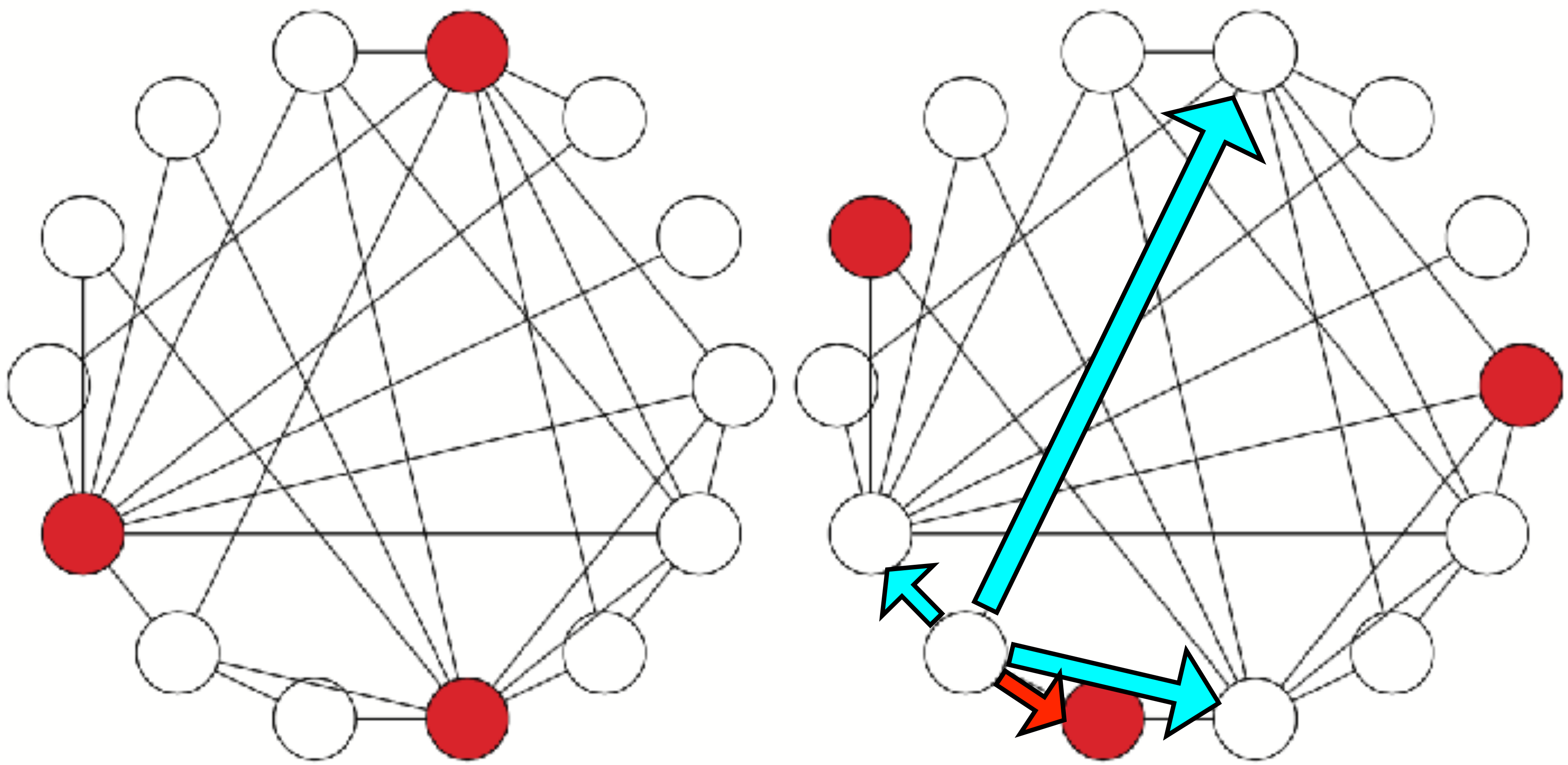
Majority illusion



The connectedness of opinion holders greatly impacts the perception of others.
A minority opinion can appear extremely popular for each individual (left side).

Lerman et al. (2015). "The Majority Illusion in Social Networks."

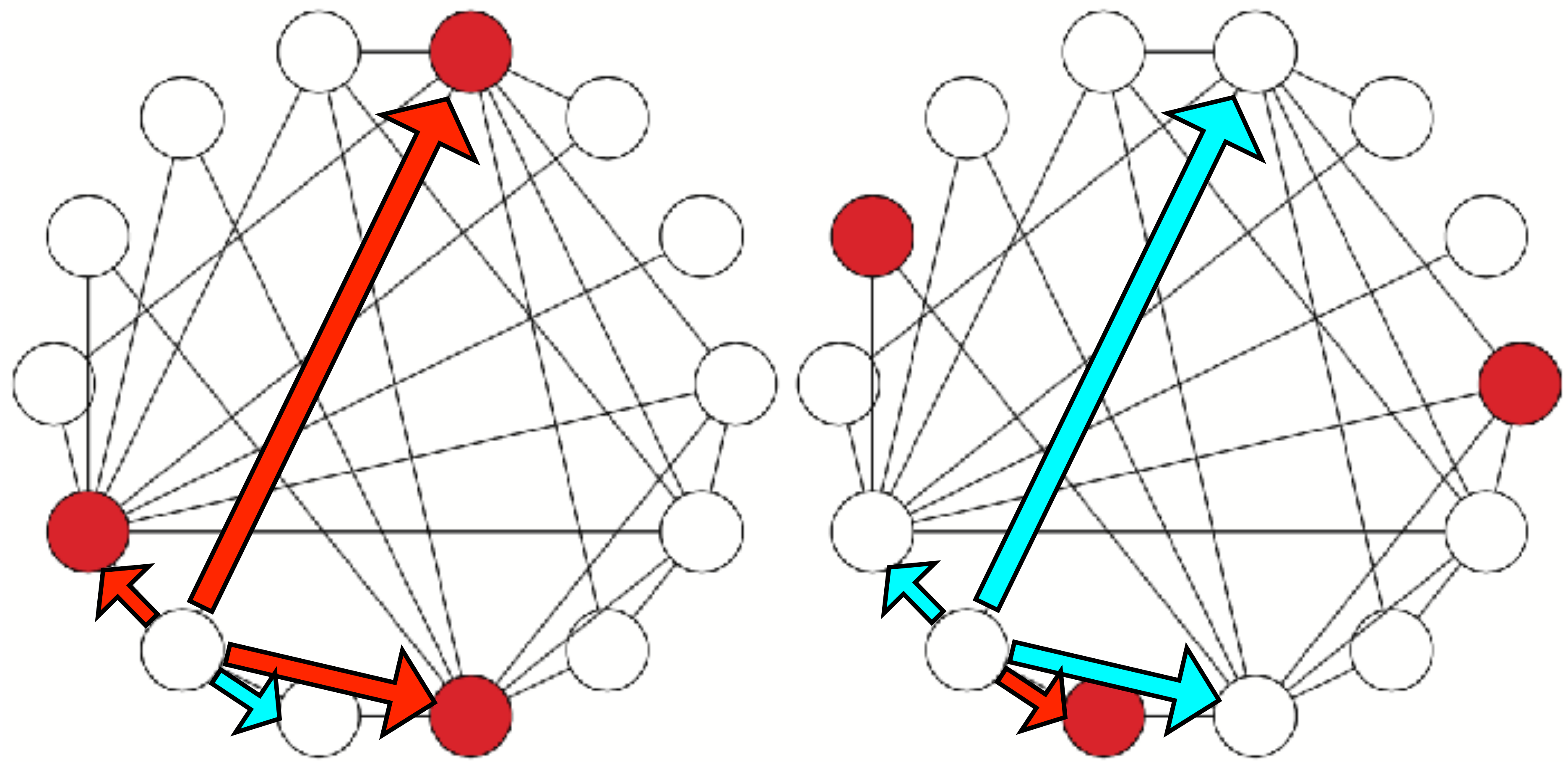
Majority illusion



The connectedness of opinion holders greatly impacts the perception of others.
A minority opinion can appear extremely popular for each individual (left side).

Lerman et al. (2015). "The Majority Illusion in Social Networks."

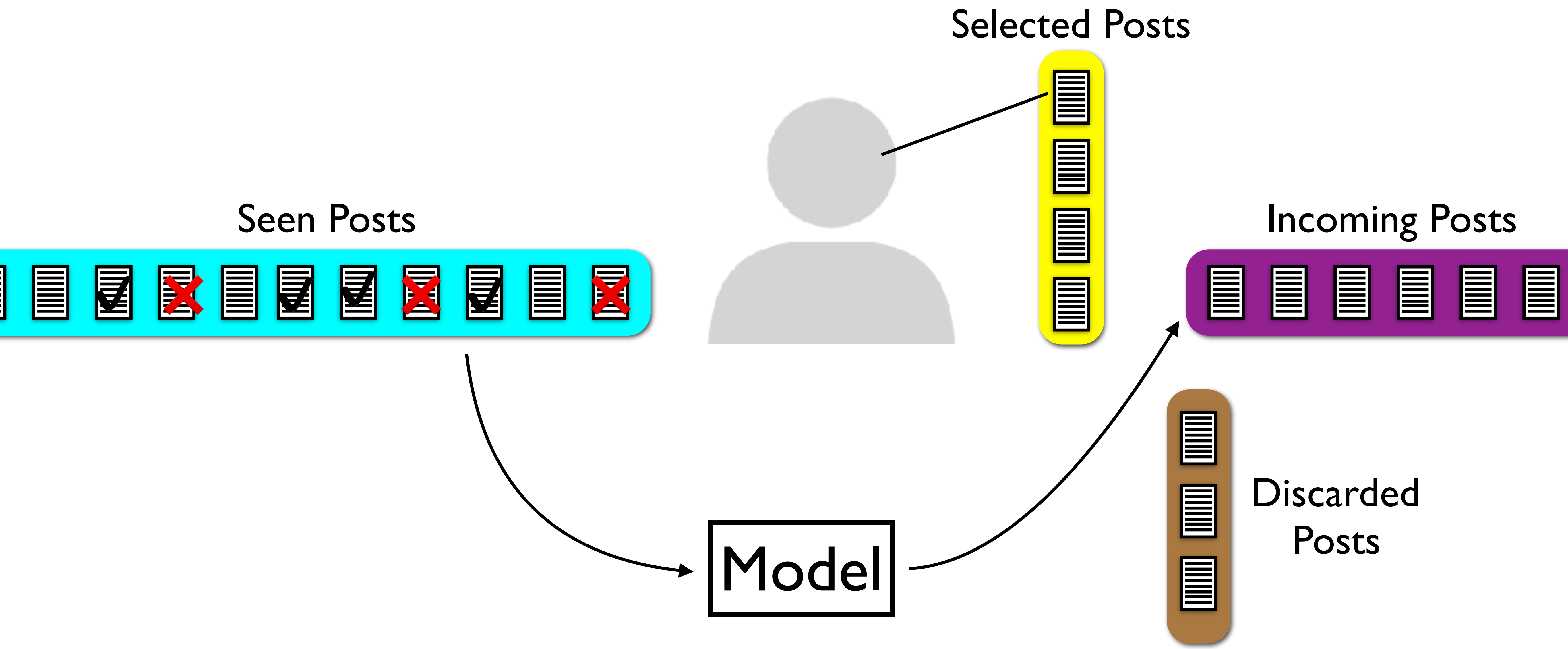
Majority illusion



The connectedness of opinion holders greatly impacts the perception of others. A minority opinion can appear extremely popular for each individual (left side).

Lerman et al. (2015). "The Majority Illusion in Social Networks."

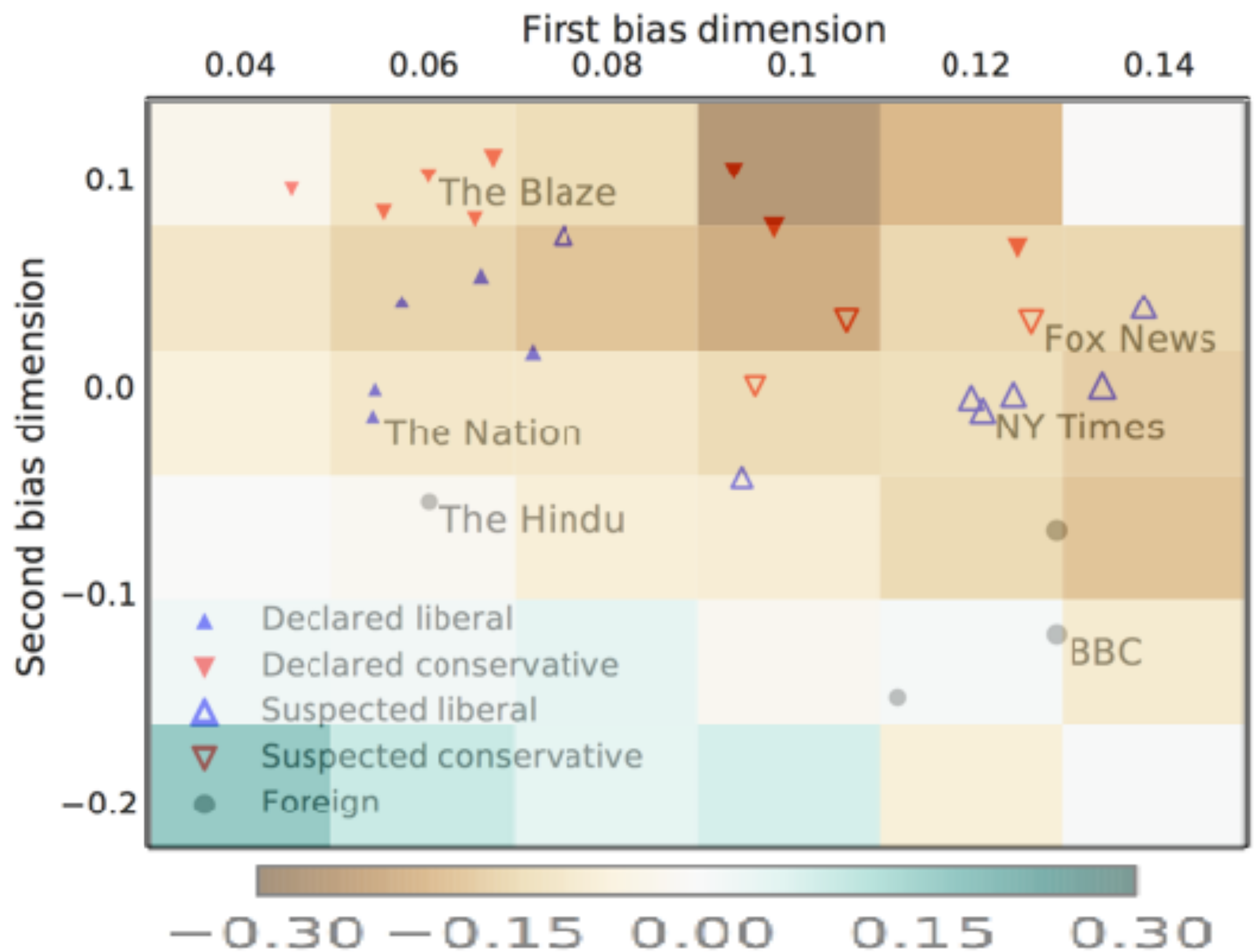
Filter bubbles



Algorithms tend to show us more of what we like.
Should they also show us things that challenge us?

See my post on the Emotional Contagion study: http://go.peoplepattern.com/blog/emotional-contagion-one/

Measuring Implicit Bias from Quoting Patterns



Personal estimates of bias are unreliable.
Behavior in context reveals consistent, interpretable patterns.

Niculae et al. (2015). "QUOTUS: The Structure of Political Media Coverage as Revealed by Quoting Patterns."

#NLProc #FTW

Personality and moral foundations predictions from text are typically based on word counts and/or word count featurization using a curated lexicon (e.g. LIWC and MFD).

The lexicons can be scaled and improved with models and unlabeled data, including to apply them to other languages and dialects.

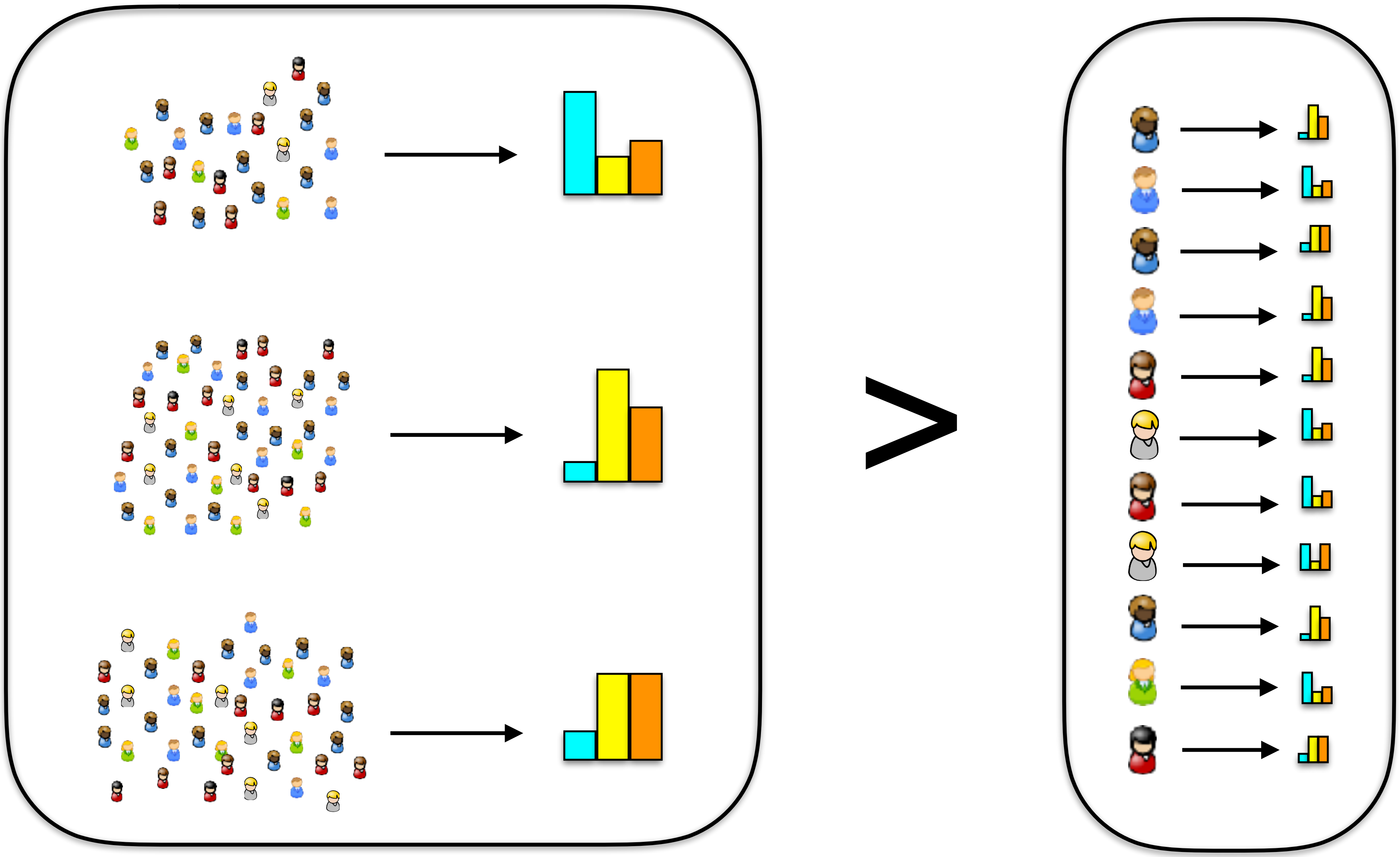
Garten et al (2016). "Morality between the lines: detecting moral sentiment in text."

Deeper analysis could reveal new linguistic features that correlate with personality/morality, including syntactic constructions and discourse structure.

Being able to predict patterns of demographics, personality and morality accurately and at scale could help us combat abusive behavior and better understand and address current social & political divisions.

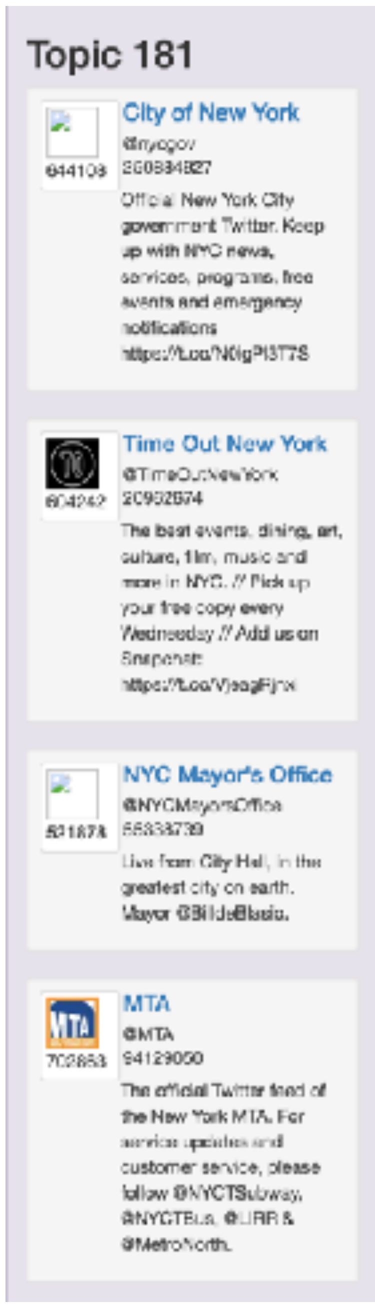
De Choudhry et al (2016) "Social Media Participation in an Activist Movement for Racial Equality.

Aggregate analysis



We should enable better aggregate understanding that doesn't require individual level annotations and predictions.

Ranked multi-labels



Income: medium-high, high, medium

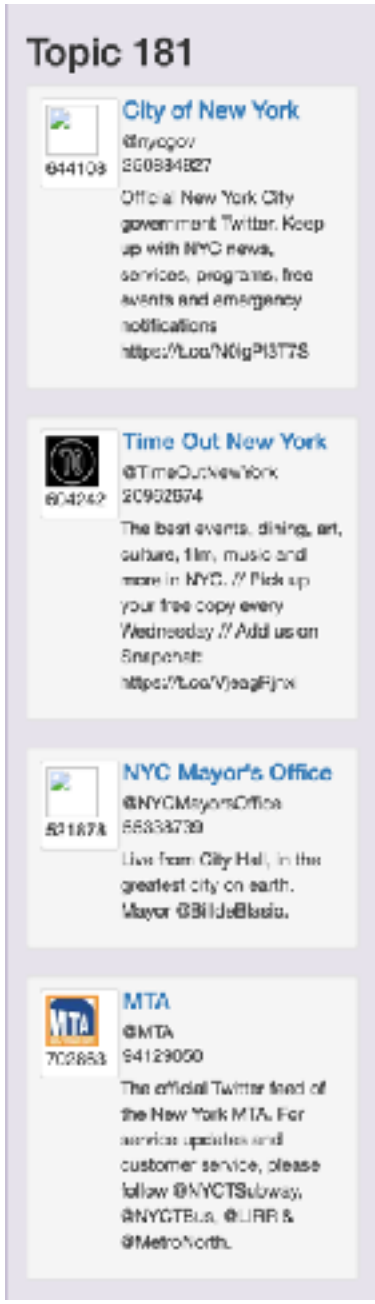
Race: white, hispanic, black, east-asian

Parent: no, yes

Political affiliation: democrat, republican, independent

Attach preference distribution for multiple labels to a topic.

Ranked multi-labels

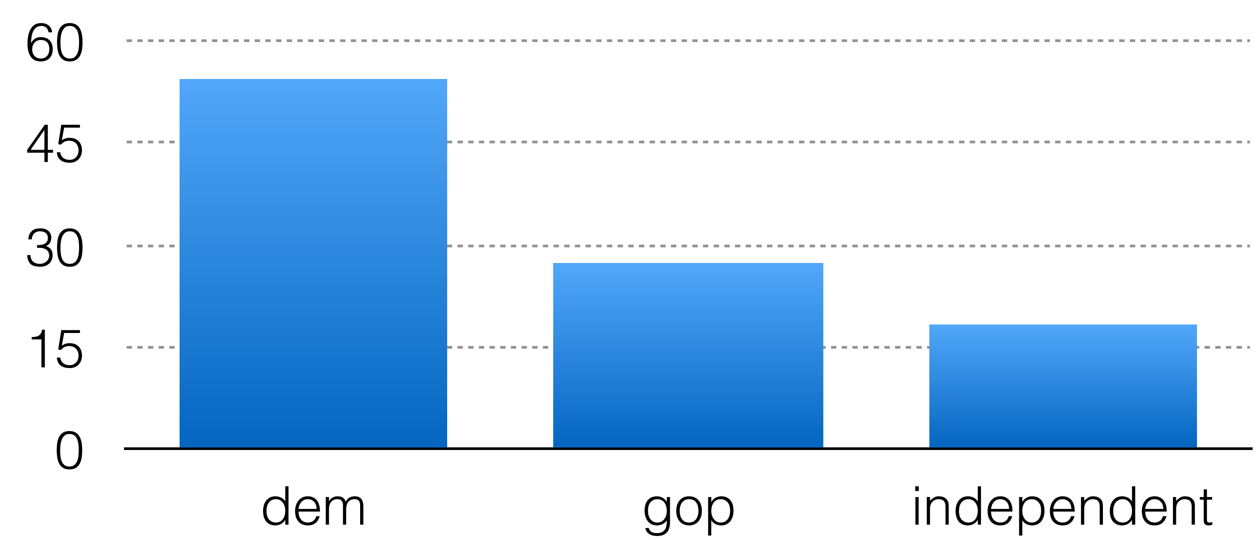


Income: medium-high, high, medium

Race: white, hispanic, black, east-asian

Parent: no, yes

Political affiliation: democrat, republican, independent

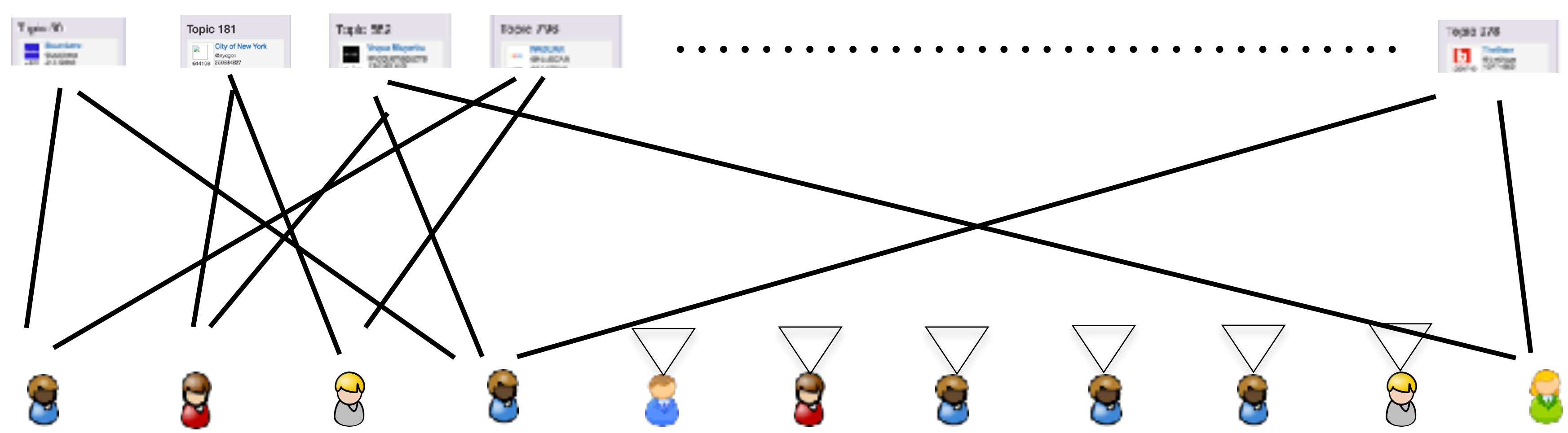


Attach preference distribution for multiple labels to a topic.

Distant Demographics

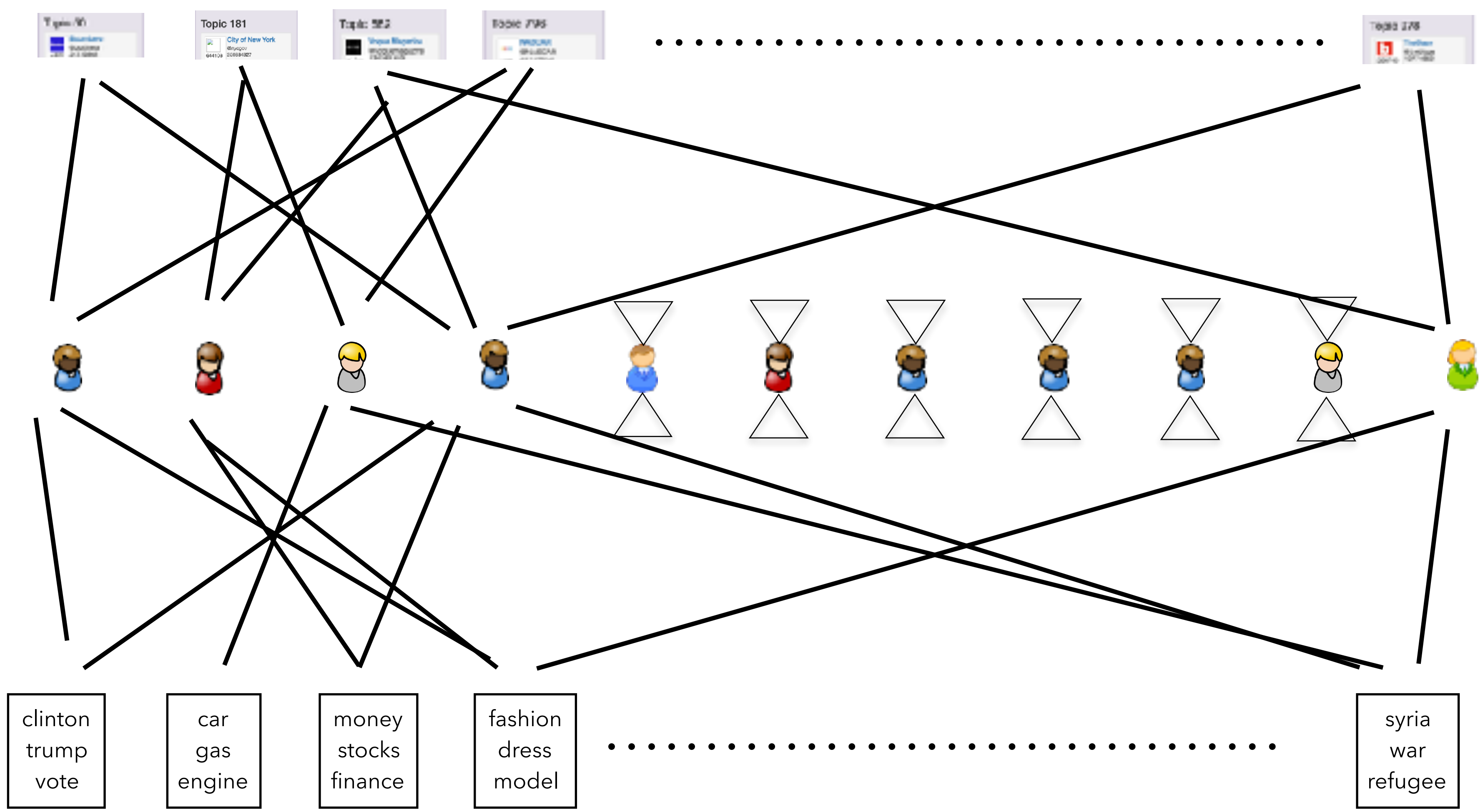
Labels propagate through both profile and feature nodes.

Distant Demographics



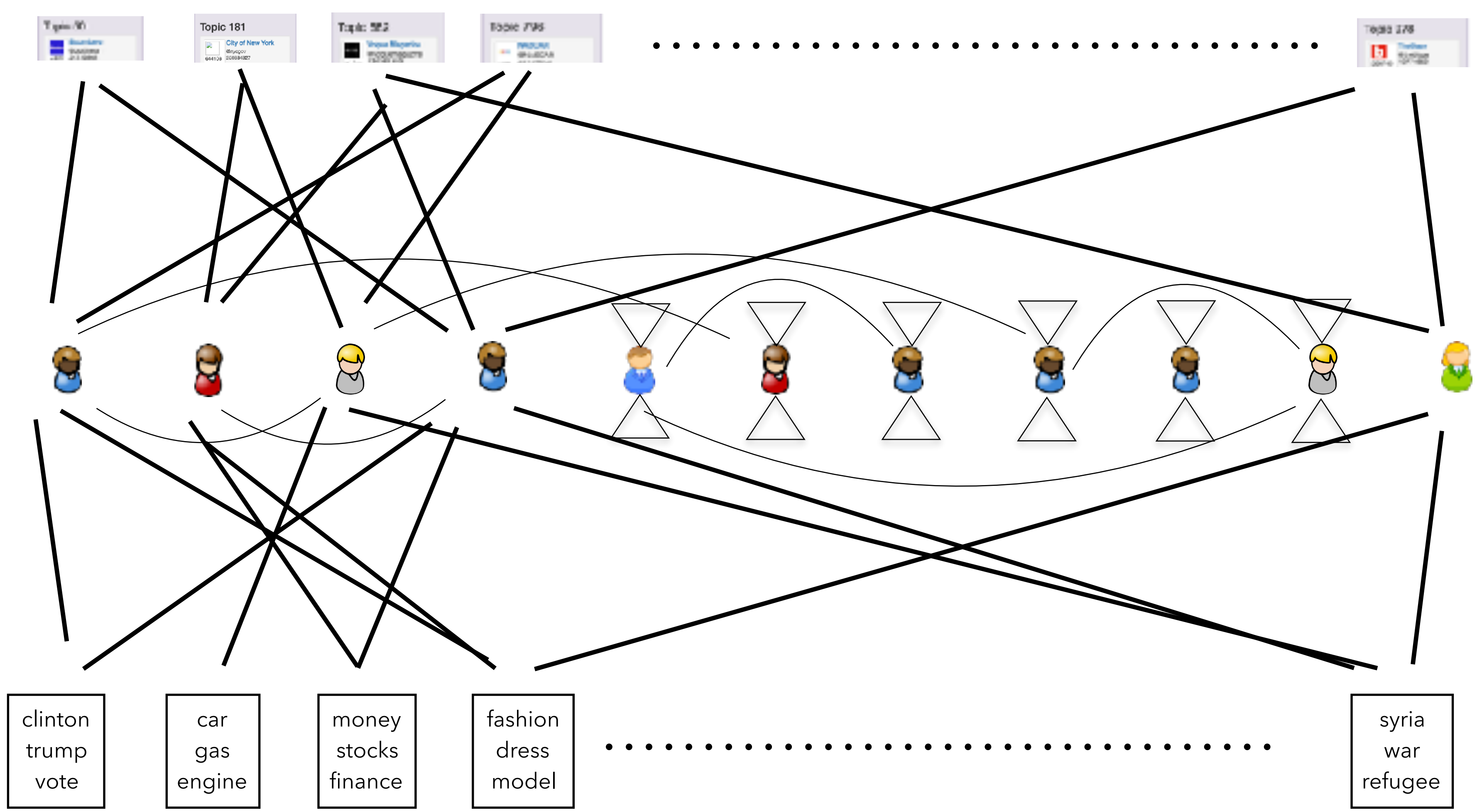
Labels propagate through both profile and feature nodes.

Distant Demographics



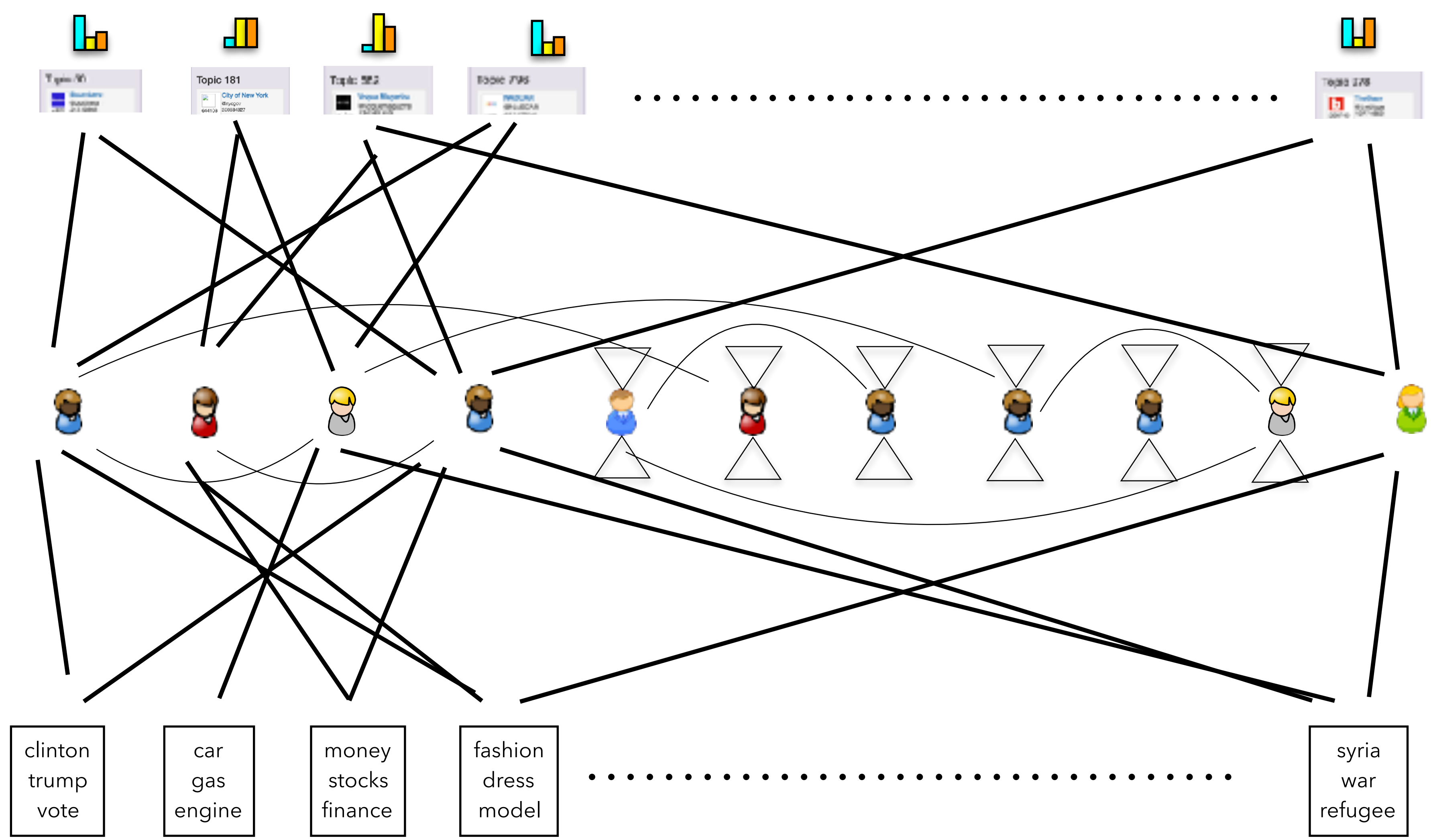
Labels propagate through both profile and feature nodes.

Distant Demographics



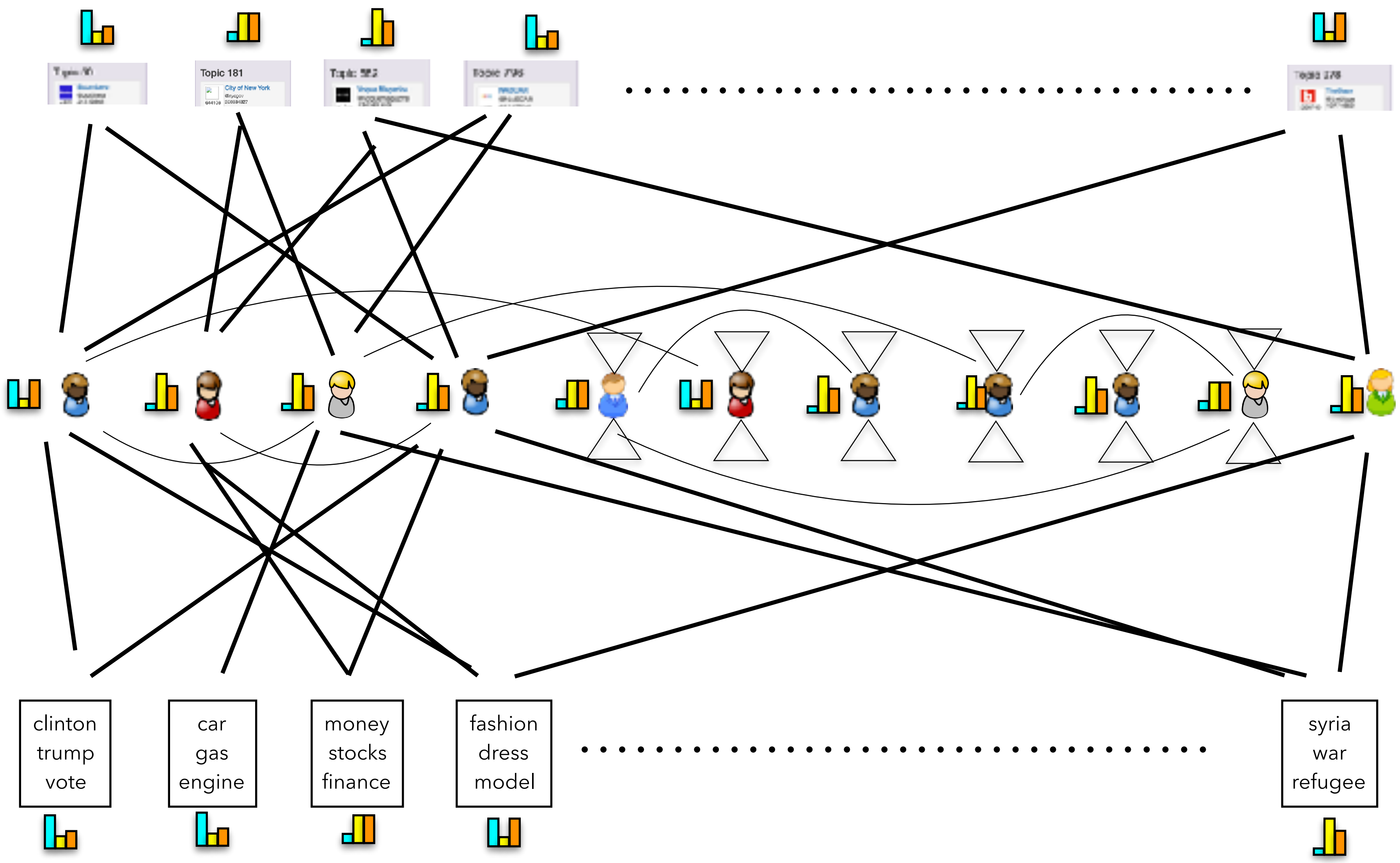
Labels propagate through both profile and feature nodes.

Distant Demographics



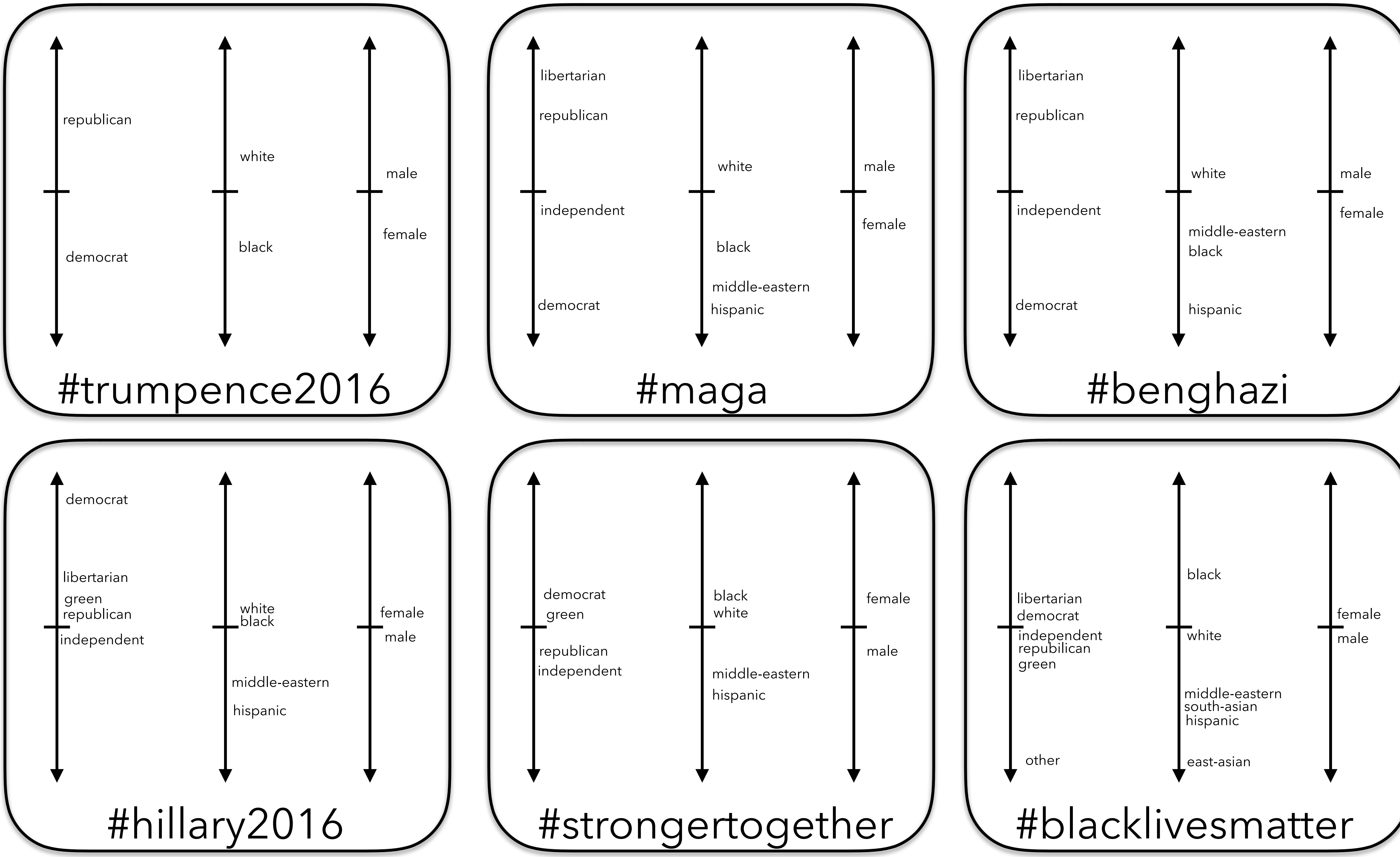
Labels propagate through both profile and feature nodes.

Distant Demographics



Labels propagate through both profile and feature nodes.

Log-likelihood ratios



Beyond text

http://www.dailymail.co.uk/sciencetech/article-3523598/Oculus-fire-revealed-VR-headset-s-T-Cs-allow-Facebook-collect-information-physical-movements.html

Digital social networks may become much more “social”.
Language will be both textual and spoken, which will provide many interesting new challenges, opportunities, and dilemmas.

Our tricky reality

The work we do in NLP is interesting and necessary—-all the more so with the massive influx of digitized language available for analysis.

However, there is plenty of scope for us to get it wrong and for others to use our work in ways that harm individuals or groups.

What can we do?

Constant, pervasive surveillance leads to **self-censorship**.
Yet, people also need public forums.
How can we positively nudge governments and companies to select analyses that **better protect individuals and their rights**, regardless of legislation?

What research questions **can we ask** that make our work valuable, useful—-even profitable!—-while **reducing the potential for harm** to individuals or groups, or even **promoting rights and societal good**?

We must make choices, as individual researchers and as a field.

Work at People Pattern is in collaboration with
Elias Ponvert, Joey Frazee, James Scott, Steve Blackmon, Abhishek Sinha
and the rest of the team.

Thanks also to Bloomberg for sponsoring my travel!