

Quantifying Qualitative Data for Understanding Controversial Issues

Michael Wojatzki,

Torsten Zesch

Saif M. Mohammad,

Svetlana Kiritchenko



Controversial Issues

- topic of sustained public debates
- social, political, economic or moral problems; e.g.:
 - *Legalization Marijuana*
 - *Gun Rights*
 - *Gender Equality*
- positions on issues often not a binary support-or-oppose stance, but a conglomerate of nuanced opinions

Assertions

- explicit expressions of opinions, beliefs, claims, arguments, and points of view about a controversial issue; e.g.:
 - *Marijuana alleviates the suffering of chronically ill patients*
 - *Marijuana is a gateway drug.*
- mean to describe one's position on an issue

Understanding Public Opinion on Controversial Issues

Controversial issues are complex

- many sub-issues and stakeholders
- people do not disagree with the assertions of the other side but on the relative importance of these assertions
- common solution: surveys and experts
 - expensive and time intensive
 - potentially biased and incomplete

Goals of our Work

- **new approach** on sentiment analysis, stance detection
 - comprehensive representation of public opinion
 - gain overall understanding of a complex issue
- data creation the by engaging people directly via crowdsourcing
 - no experts required
 - collect assertions people care about

Dataset of Nuanced Assertions on Controversial Issues (NAoCI)

- **idea:** large number of people vote on a large number of assertions
- engage people directly
 1. to collect a large set of assertions relevant to controversial issues (**qualitative** data)
 2. to obtain judgments on these assertions (**quantitative** data)
- steps conducted via crowdsourcing on crowdflower.com
- process approved by NRC's ethics board

Dataset of Nuanced Assertions on Controversial Issues (NAoCI)

- idea: large number of people vote on a large number of assertions
- engage people directly
 - 1. to collect a large set of assertions relevant to controversial issues (qualitative data)**
 2. to obtain judgments on these assertions (quantitative data)
- steps conducted via crowdsourcing on crowdflower.com
- process approved by NRC's ethics board

Collecting Assertions

- Given an **issue + definition**
- Participant had to come up with **five assertions**
 - according to given directions
 - E.g.: no coreference, vague formulations (*maybe*)

Collected Assertions

- 69 participants (US-based)
- 16 issues
- 2243 assertions (about 150 per issue)

Issue	# of Assertions
Black Lives Matter	135
Climate Change	142
Creationism in school	129
Foreign Aid	150
Gender Equality	130
Gun Rights	145
Marijuana	138
Same-sex Marriage	148
Mandatory Vaccination	134
Media Bias	133
Obama Care	154
US Electoral System	175
US in the Middle East	138
US Immigration	130
Vegetarianism & Veganism	128
War on Terrorism	134
Total	2,243

Dataset of Nuanced Assertions on Controversial Issues (NAoCI)

- idea: large number of people vote on a large number of assertions
- engage people directly
 1. to collect a large set of assertions relevant to controversial issues (qualitative data)
 2. **to obtain judgments on these assertions (quantitative data)**
- steps conducted via crowdsourcing on crowdflower.com
- process approved by NRC's ethics board

Quantifying Agreement and Strength of Support and Opposition

given the assertions, participants are asked to

1. indicate whether they agree or disagree with the assertions
2. indicate how strongly they support or oppose the assertions

Quantifying Agreement and Strength of Support and Opposition

given the assertions, participants are asked to

1. indicate whether they agree or disagree with the assertions
2. indicate how strongly they support or oppose the assertions

The world needs to know that blacks are also humans.

- agree disagree

Blacks have achieved a lot for the whole society.

- agree disagree

Quantifying Agreement and Strength of Support and Opposition

given the assertions, participants are asked to

1. indicate whether they agree or disagree with the assertions
 2. indicate how strongly they support or oppose the assertions
- difficult to give a numerical score indicating the degree of support or opposition
 - solution: **best–worst scaling** (Louviere et al., 2015; Kiritchenko and Mohammad, 2016)
 - given a tuple of **four** assertions, indicate
 - Which of the assertions do you **support** the most?
 - Which of the assertions do you **oppose** the most?

Best–Worst Scaling

Which of these assertions do you support the most?

- Every race has experienced racism.*
- Historically in the United States there has been discrimination.*
- The Black lives matter movement is important.*
- Matter movement encourages racial hate.*

Which of these assertions do you oppose the most?

- Every race has experienced racism.*
- Historically in the United States there has been discrimination.*
- The Black lives matter movement is important.*
- The Black Lives Matter movement encourages racial hate.*

Issue: Black Lives Matter

Collected Judgments

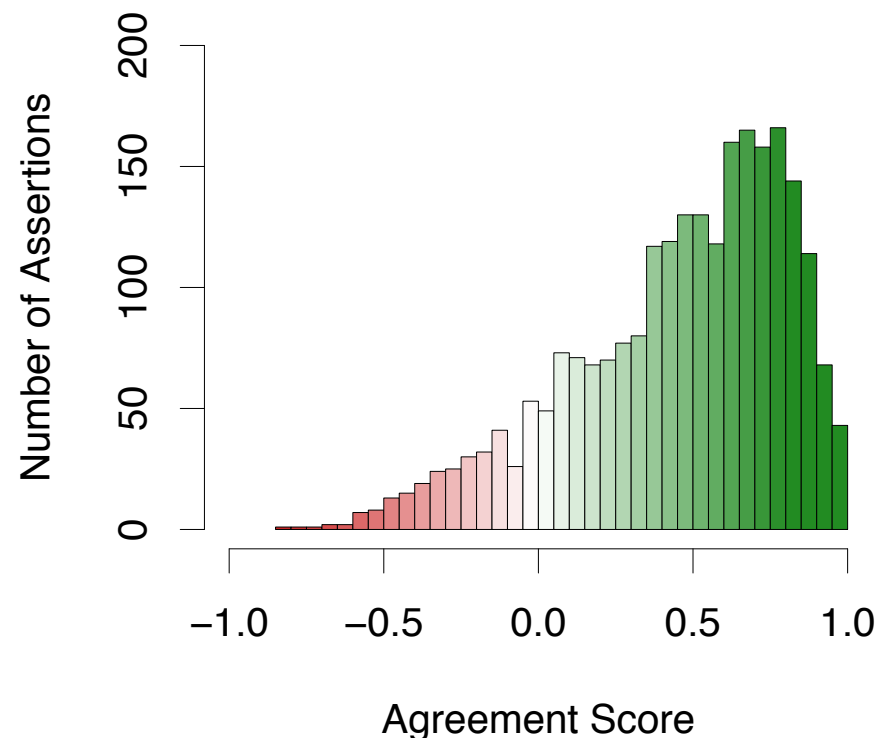
- 16 issues
- 2243 assertions (about 150 per issue)
- 230 participants (US-based)
- over **100,000** agreement judgments
- about **70,000** judgments indicating how strongly people support or oppose the assertions

How many people agree or disagree with an assertion?

- Agreement Score:

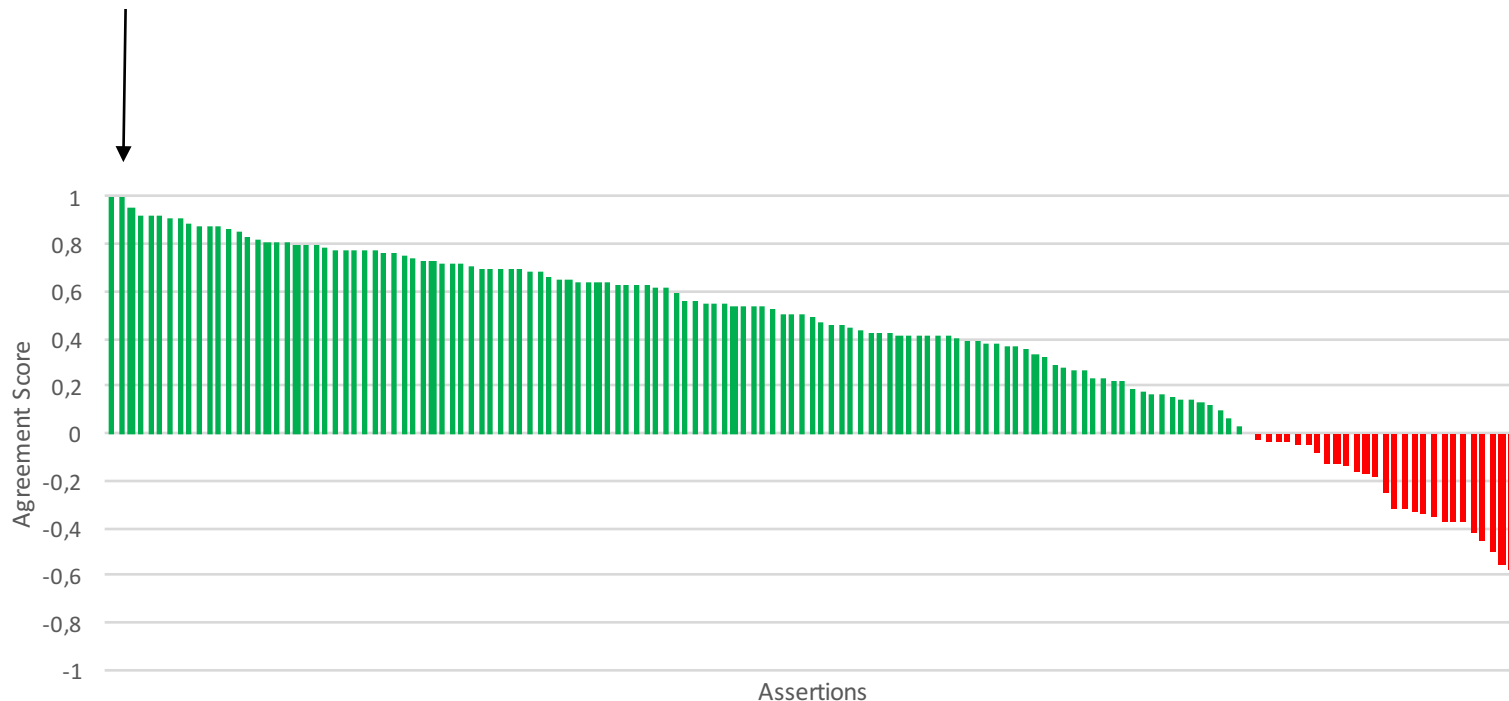
$$\% \text{ agree}(a) - \% \text{ disagree}(a)$$

- participants tend to agree with the assertions more often than they disagree



Ranking Assertions

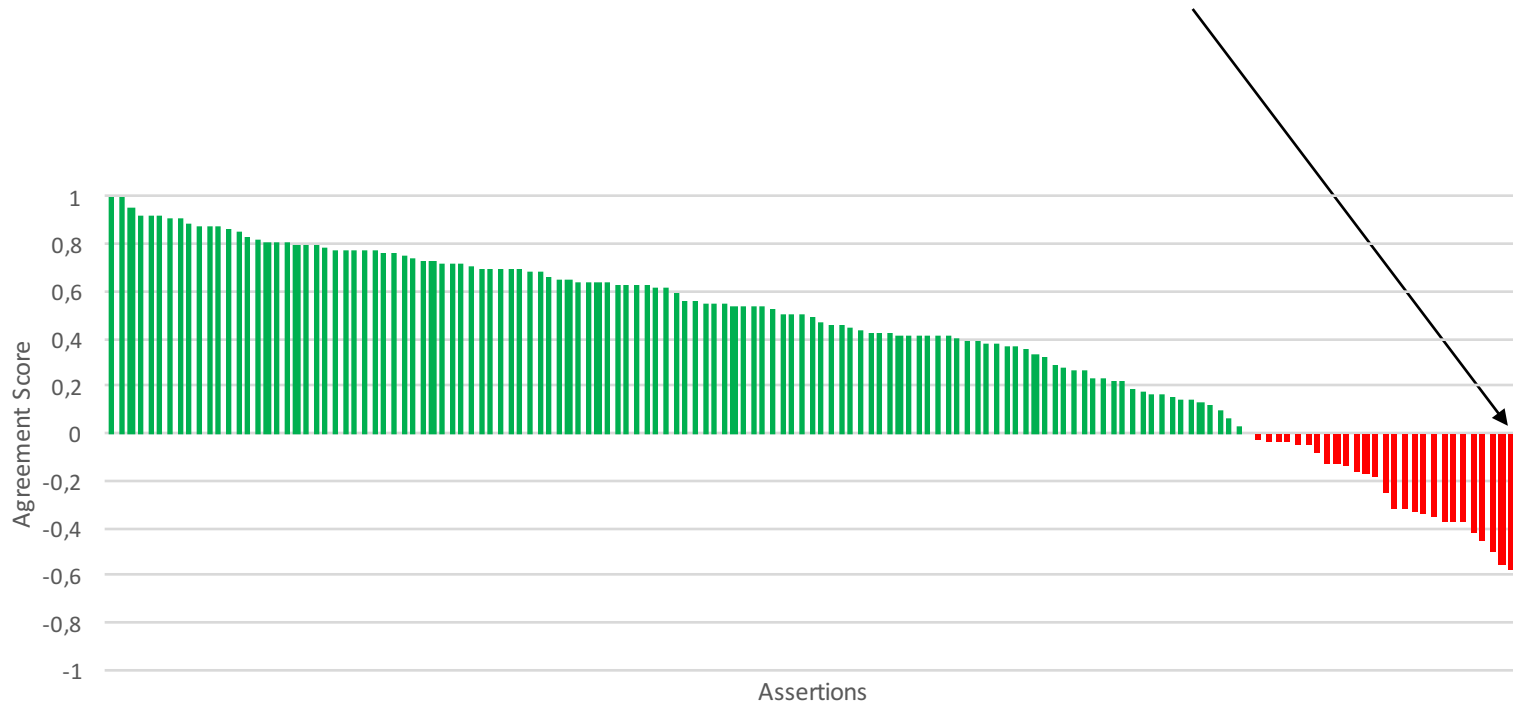
1. Gun owners should be required to take a gun safety course.
2. Gun owners should register their arms.
3. Gun owners need to be required to have a background check.



Issue: Gun Rights

Ranking Assertions

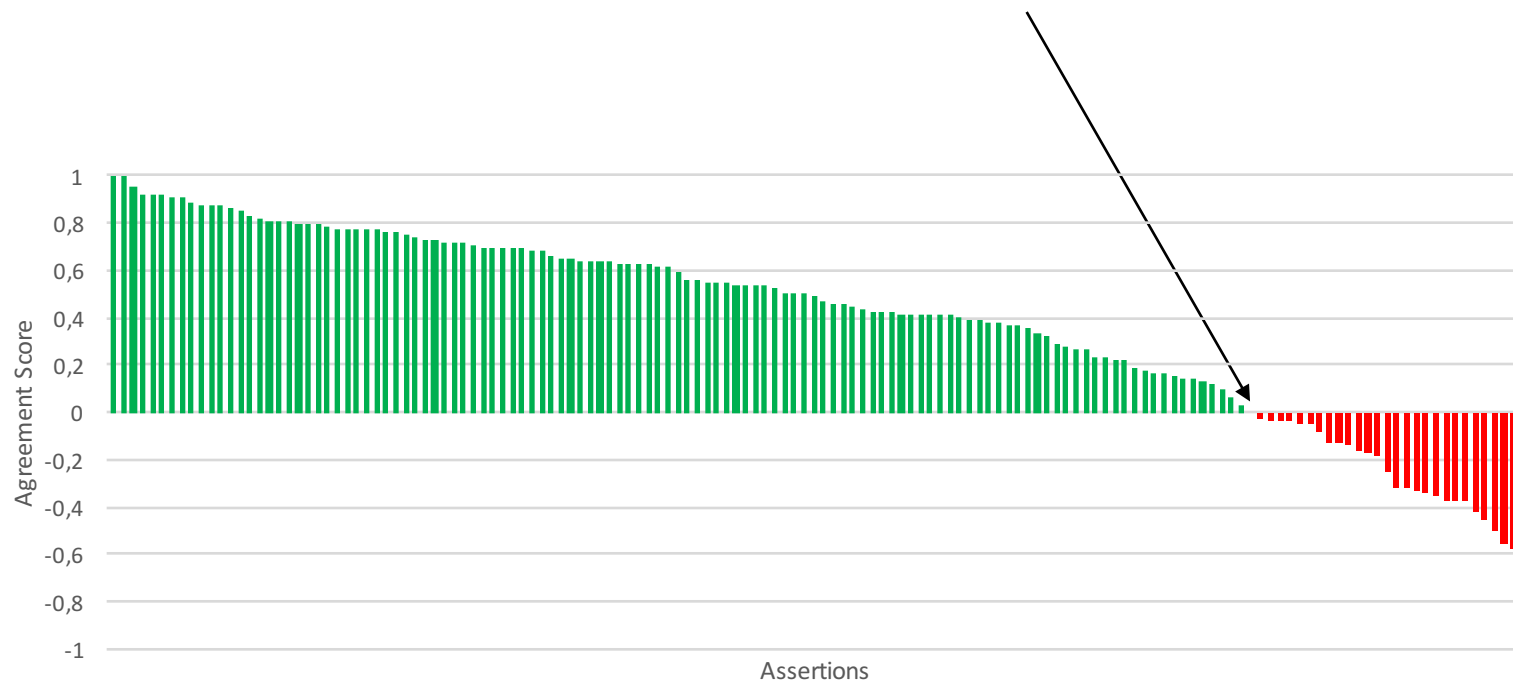
1. Everyone should own a gun.
2. The gun industry is too heavily regulated.
3. Guns should be legal for everyone.



Issue: Gun Rights

Ranking Assertions

1. Guns should only be issued for hunting.
2. People who own guns are not more likely to mass kill.
3. In a certain part to eliminate the arms would be to end the delinquency.



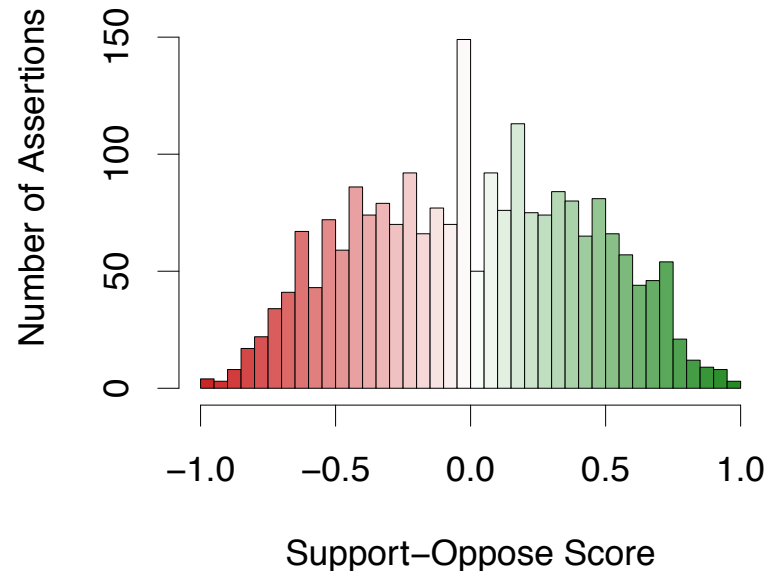
Issue: Gun Rights

How Strongly do People Support or Oppose an Assertion?

- based on **best–worst annotations**
- Support-Oppose Score:

$$\% \text{ most support}(a) - \% \text{ most opposed}(a)$$

- normal distribution
 - across all issues
 - similar for individual issues



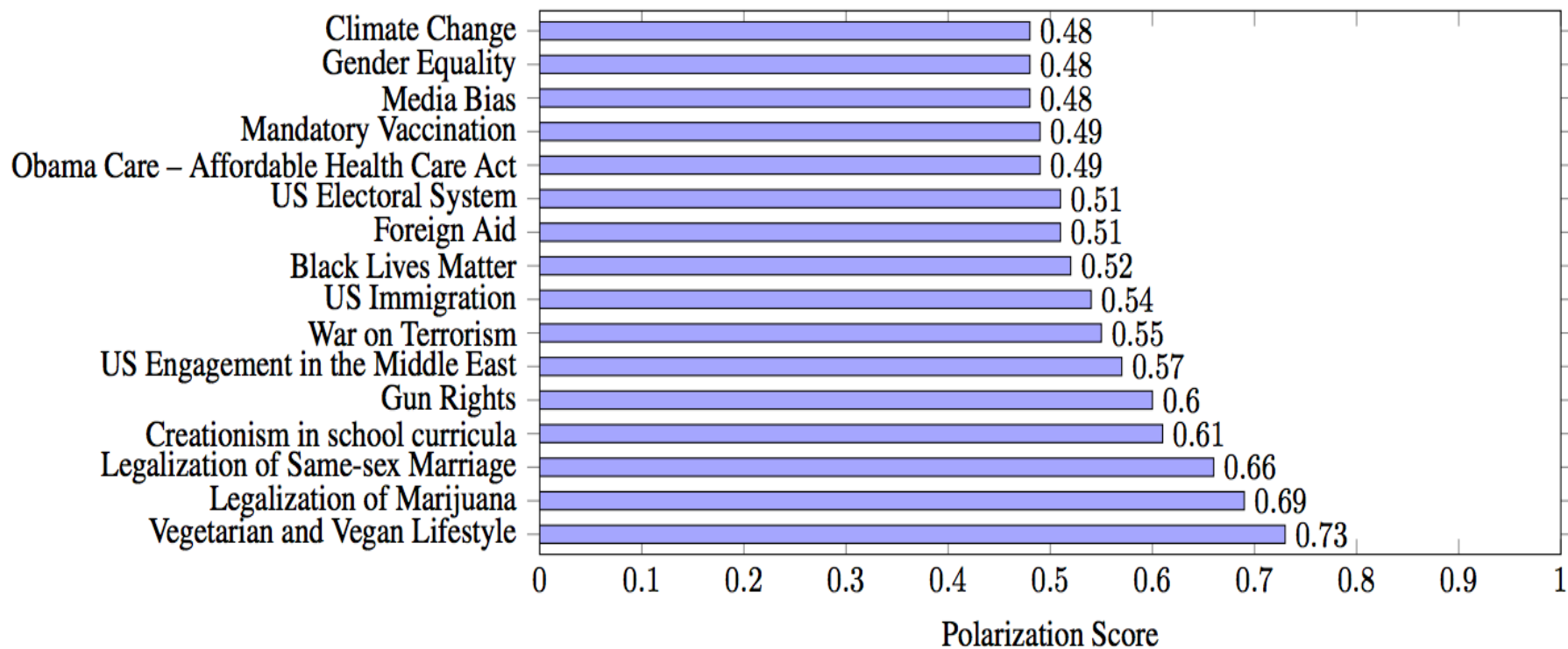
How Polarizing is an Issue?

Polarization Score: 1- average of the absolute value of the agreement score of all assertions of an issue

$$ps(I) = 1 - \frac{1}{|I|} \sum_{a \in I} |ags(a)|$$

- 0 = participants consistently agree or disagree with all assertions
- 1 = equal number of participants agree and disagree with assertions

Ranking Issues



How Similar are two Assertions Judged by Several People?

example pairs of closest assertions cosine:

Guns don't kill people people do.
Guns aren't the only weapons that kill.

Widespread gun ownership leads to mass killings.
Children are dying at schools because of guns usage.

Future NLP tasks: Understanding Issues from Social Media

quantify qualitative data without being dependent on the described crowdsourcing

- map assertions to social media posts
- identify similar tweeters, tweets
- predict scores
 - Follow-up paper: **Agree or Disagree: Predicting Judgments on Nuanced Assertions** (*Sem)

Summary

- new method for understanding controversial issues
 - no experts required
- dataset of **Nuanced Assertions on Controversial Issues (NAoCI)**
 - 2263 assertions on 16 issues
 - over 100.000 agreement, 70.000 support-oppose judgments
 - <https://sites.google.com/view/you-on-issues/>
- metrics for understanding controversial issues
 - agreement score
 - support-oppose score
 - polarization scores
 - assertion and participant similarity

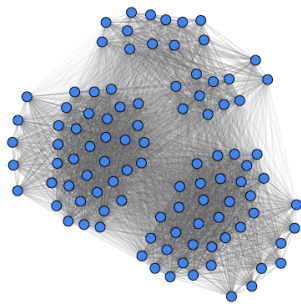
Thank You!

Summary

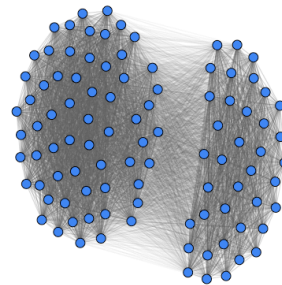
- new method for understanding controversial issues
 - no experts required
- dataset of **Nuanced Assertions on Controversial Issues (NAoCI)**
 - 2263 assertions on 16 issues
 - over 100.000 agreement, 70.000 support-oppose judgments
 - <https://sites.google.com/view/you-on-issues/>
- metrics for understanding controversial issues
 - agreement score
 - support-oppose score
 - polarization scores
 - assertion and participant similarity

How Similar do People Judge Several Assertions?

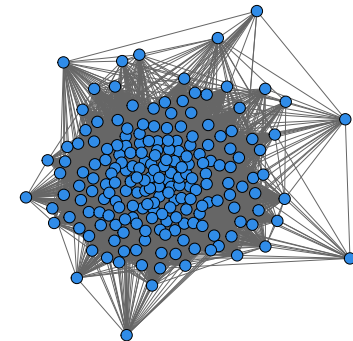
- similarity of participants: $\cos(p1, p2) = \frac{ad_{p1} \cdot ad_{p2}}{|ad_{p1}| \cdot |ad_{p2}|}$
- graph:
 - nodes = participants
 - edges = similarity of participants
 - several distributions conceivable:



vs.



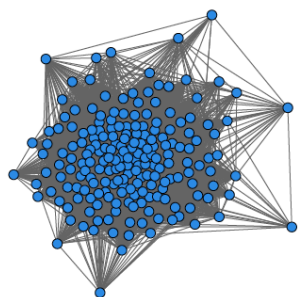
vs.



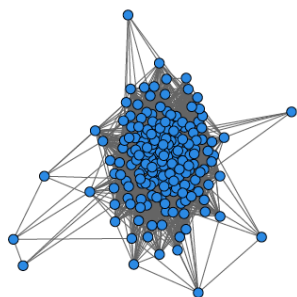
Similarity of Participants

we do not find significant clusters of positions

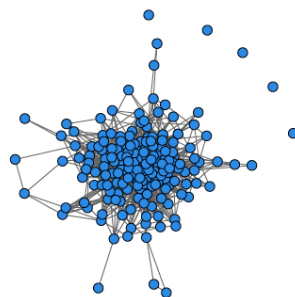
- one mainstream, many isolated deviations



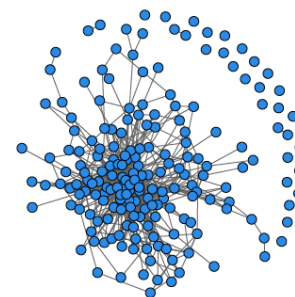
(a) cosine > 0.0



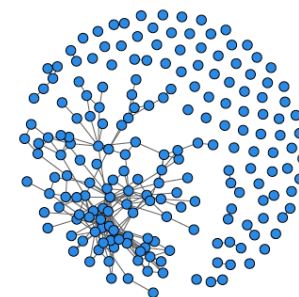
(b) cosine > 0.1



(c) cosine > 0.2



(d) cosine > 0.3



(e) cosine > 0.4

Issue: Black Lives Matter