Using Optimization to Mitigate Polarization and Disagreement in Social Networks

Stefan Neumann

@StefanResearch

VWCO, 6.06.2025

The New York Times

Opinion

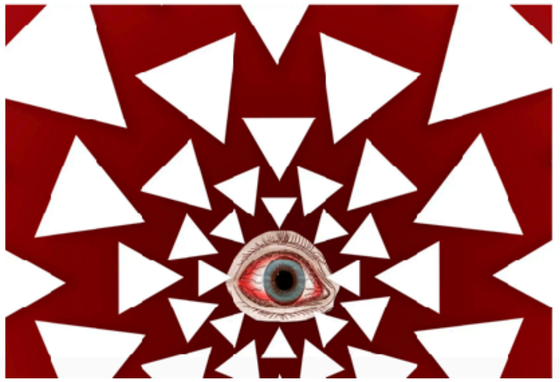
YouTube, the Great Radicalizer

By Zeynep Tufekci

March 10, 2018

Share full article

101



The New York Times

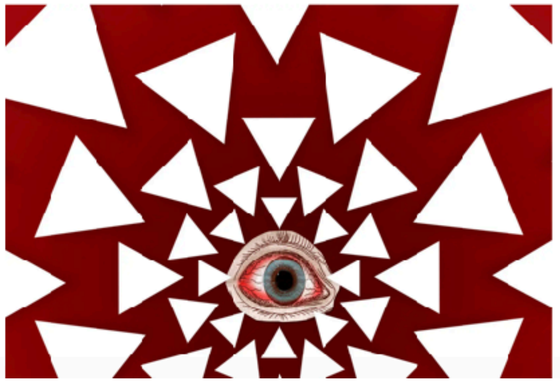
Opinion

YouTube, the Great Radicalizer

By Zeynep Tufekci

March 10, 2018

Share full article | 101



[redacted]

COMMITTEE SENSITIVE – RUSSIA INVESTIGATION ONLY

116TH CONGRESS *1st Session* | SENATE | REPORT 116-XX

(U) REPORT

OF THE

SELECT COMMITTEE ON INTELLIGENCE

UNITED STATES SENATE

ON

RUSSIAN ACTIVE MEASURES CAMPAIGNS AND INTERFERENCE

IN THE 2016 U.S. ELECTION

VOLUME 2: RUSSIA'S USE OF SOCIAL MEDIA

WITH ADDITIONAL VIEWS

[redacted]

1

[redacted]

COMMITTEE SENSITIVE – RUSSIA INVESTIGATION ONLY

The New York Times

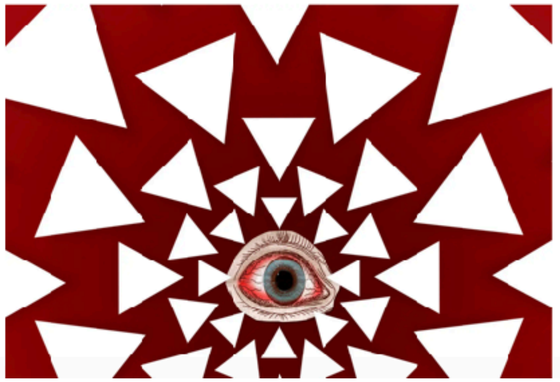
Opinion

YouTube, the Great Radicalizer

By Zeynep Tufekci

March 10, 2018

Share full article | 101



[redacted] [redacted]
COMMITTEE SENSITIVE – RUSSIA INVESTIGATION ONLY

116TH CONGRESS *1st Session* | SENATE | REPORT 116-XX

(U) REPORT

OF THE

SELECT COMMITTEE ON INTELLIGENCE

UNITED STATES SENATE

ON

RUSSIAN ACTIVE MEASURES CAMPAIGNS AND INTERFERENCE

IN THE 2016 U.S. ELECTION

VOLUME 2: RUSSIA'S USE OF SOCIAL MEDIA

WITH ADDITIONAL VIEWS

[redacted]

1

[redacted] [redacted]
COMMITTEE SENSITIVE – RUSSIA INVESTIGATION ONLY

The New York Times

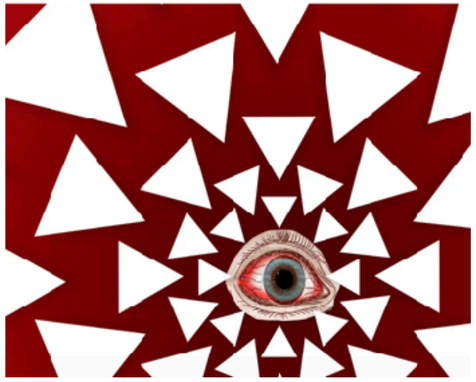
Opinion

YouTube, the Great Radicalizer

By Zeynep Tufekci

March 10, 2018

Share full article 101



COMMITTEE SENSITIVE – RUSSIA INVESTIGATION ONLY

116TH CONGRESS *1st Session* SENATE REPORT 116-XX

(U) REPORT

OF THE

SELECT COMMITTEE ON INTELLIGENCE

UNITED STATES SENATE

Check for updates

Algorithmic amplification of politics on Twitter

Ferenc Huszár[a,b,c,1,2], Sofia Ira Ktena[a,1,3], Conor O'Brien[a,1], Luca Belli[a,2], Andrew Schlaikjer[a], and Moritz Hardt[d]

[a]Machine Learning Ethics, Transparency, and Accountability Team, Twitter, San Francisco, CA 94103; [b]Department of Computer Science and Technology, University of Cambridge, Cambridge CB3 0FD, United Kingdom; [c]Gatsby Computational Neuroscience Unit, University College London, London, W1T 4JG, United Kingdom; and [d]Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720

Edited by David Laitin, Department of Political Science, Stanford University, Stanford, CA; received December 11, 2020; accepted October 5, 2021

Content on Twitter's home timeline is selected and ordered by personalization algorithms. By consistently ranking certain content higher, these algorithms may amplify some messages while reducing the visibility of others. There's been intense public and scholarly debate about the possibility that some political groups right enjoys higher algorithmic amplification than the mainstream political left. Consistent with this overall trend, our second set of findings studying the US media landscape revealed that algorithmic amplification favors right-leaning news sources. We further looked at whether algorithms amplify far-left and far-right political groups more than moderate ones; contrary to prevailing public belief, we did not find evidence to support this hypothesis. We hope our findings will contribute to an evidence-based debate on the role personalization algorithms play in shaping political content consumption.

social media | algorithmic personalization | media amplification | political bias

When Twitter introduced machine learning to personalize the Home timeline in 2016, it excluded a randomly chosen control group of 1% of all global Twitter users from the new personalized Home timeline. Individuals in this control group have never experienced personalized ranked timelines. Instead, their

be isolated from indirect effects of personalization, as individuals in the control group encounter content shared by users in the treatment group. Therefore, although a randomized controlled experiment, our experiment does not satisfy the well-known Stable Unit Treatment Value Assumption from causal inference (23). As a consequence, it cannot provide unbiased estimates of causal quantities of interest, such as the average treatment

Significance

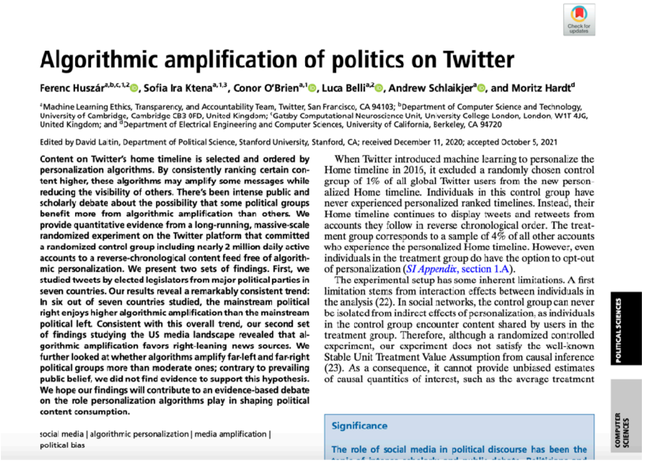
The role of social media in political discourse has been the

POLITICAL SCIENCE

COMPUTER SCIENCES

"In six out of seven countries studied, the mainstream political right enjoys higher algorithmic amplification than the mainstream political left."

- Many studies about OSNs are empirical



Algorithmic amplification of politics on Twitter

Ferenc Huszár[a,b,c,1,2], Sofia Ira Ktena[a,1,3], Conor O'Brien[a,1], Luca Belli[a,2], Andrew Schlaikjer[a], and Moritz Hardt[d]

[a]Machine Learning Ethics, Transparency, and Accountability Team, Twitter, San Francisco, CA 94103; [b]Department of Computer Science and Technology, University of Cambridge, Cambridge CB3 0FD, United Kingdom; [c]Gatsby Computational Neuroscience Unit, University College London, London, W1T 4JG, United Kingdom; and [d]Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720

Edited by David Latin, Department of Political Science, Stanford University, Stanford, CA; received December 11, 2020; accepted October 5, 2021

Content on Twitter's home timeline is selected and ordered by personalization algorithms. By consistently ranking certain content higher, these algorithms may amplify some messages while reducing the visibility of others. There's been intense public and scholarly debate about the possibility that some political groups benefit more from algorithmic amplification than others. We provide quantitative evidence from a long-running, massive-scale randomized experiment on the Twitter platform that committed a randomized control group including nearly 2 million daily active accounts to a reverse-chronological content feed free of algorithmic personalization. We present two sets of findings. First, we studied tweets by elected legislators from major political parties in seven countries. Our results reveal a remarkably consistent trend: In six out of seven countries studied, the mainstream political right enjoys higher algorithmic amplification than the mainstream political left. Consistent with this overall trend, our second set of findings studying the US media landscape revealed that algorithmic amplification favors right-leaning news sources. We further looked at whether algorithms amplify far-left and far-right political groups more than moderate ones; contrary to prevailing public belief, we did not find evidence to support this hypothesis. We hope our findings will contribute to an evidence-based debate on the role personalization algorithms play in shaping political content consumption.

social media | algorithmic personalization | media amplification | political bias

When Twitter introduced machine learning to personalize the Home timeline in 2016, it excluded a randomly chosen control group of 1% of all global Twitter users from the new personalized Home timeline. Individuals in this control group have never experienced personalized ranked timelines. Instead, their Home timeline continues to display tweets and retweets from accounts they follow in reverse chronological order. The treatment group corresponds to a sample of 4% of all other accounts who experience the personalized Home timeline. However, even individuals in the treatment group do have the option to opt-out of personalization (*SI Appendix*, section 1.A).

The experimental setup has some inherent limitations. A first limitation stems from interaction effects between individuals in the analysis (22). In social networks, the control group can never be isolated from indirect effects of personalization, as individuals in the control group encounter content shared by users in the treatment group. Therefore, although a randomized controlled experiment, our experiment does not satisfy the well-known Stable Unit Treatment Value Assumption from causal inference (23). As a consequence, it cannot provide unbiased estimates of causal quantities of interest, such as the average treatment

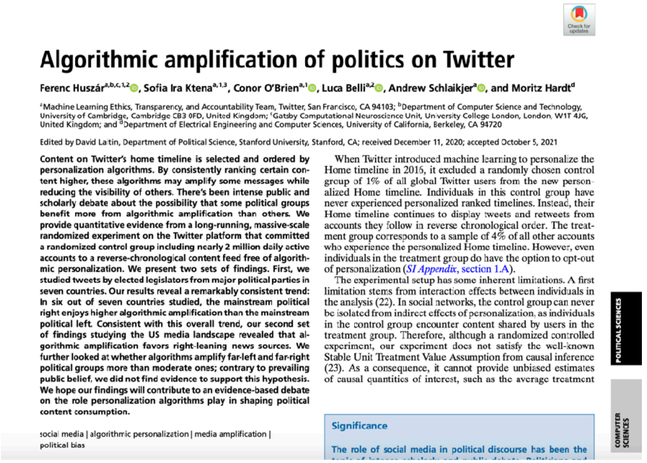
Significance

The role of social media in political discourse has been the

POLITICAL SCIENCES

COMPUTER SCIENCES

- Many studies about OSNs are empirical
  - Some data is collected, relevant phenomena are analyzed
    Do filter bubbles exist? Does the Twitter/X algorithm promote left- or right-wing content?



Algorithmic amplification of politics on Twitter

Ferenc Huszár[a,b,c,1,2], Sofia Ira Ktena[a,1,3], Conor O'Brien[a,1], Luca Belli[a,2], Andrew Schlaikjer[a], and Moritz Hardt[d]

[a]Machine Learning Ethics, Transparency, and Accountability Team, Twitter, San Francisco, CA 94103; [b]Department of Computer Science and Technology, University of Cambridge, Cambridge CB3 0FD, United Kingdom; [c]Gatsby Computational Neuroscience Unit, University College London, London, W1T 4JG, United Kingdom; and [d]Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720

Edited by David Latin, Department of Political Science, Stanford University, Stanford, CA; received December 11, 2020; accepted October 5, 2021

Content on Twitter's home timeline is selected and ordered by personalization algorithms. By consistently ranking certain content higher, these algorithms may amplify some messages while reducing the visibility of others. There's been intense public and scholarly debate about the possibility that some political groups benefit more from algorithmic amplification than others. We provide quantitative evidence from a long-running, massive-scale randomized experiment on the Twitter platform that committed a randomized control group including nearly 2 million daily active accounts to a reverse-chronological content feed free of algorithmic personalization. We present two sets of findings. First, we studied tweets by elected legislators from major political parties in seven countries. Our results reveal a remarkably consistent trend: In six out of seven countries studied, the mainstream political right enjoys higher algorithmic amplification than the mainstream political left. Consistent with this overall trend, our second set of findings studying the US media landscape revealed that algorithmic amplification favors right-leaning news sources. We further looked at whether algorithms amplify far-left and far-right political groups more than moderate ones; contrary to prevailing public belief, we did not find evidence to support this hypothesis. We hope our findings will contribute to an evidence-based debate on the role personalization algorithms play in shaping political content consumption.

social media | algorithmic personalization | media amplification | political bias

When Twitter introduced machine learning to personalize the Home timeline in 2016, it excluded a randomly chosen control group of 1% of all global Twitter users from the new personalized Home timeline. Individuals in this control group have never experienced personalized ranked timelines. Instead, their Home timeline continues to display tweets and retweets from accounts they follow in reverse chronological order. The treatment group corresponds to a sample of 4% of all other accounts who experience the personalized Home timeline. However, even individuals in the treatment group do have the option to opt-out of personalization (*SI Appendix*, section 1.A).

The experimental setup has some inherent limitations. A first limitation stems from interaction effects between individuals in the analysis (22). In social networks, the control group can never be isolated from indirect effects of personalization, as individuals in the control group encounter content shared by users in the treatment group. Therefore, although a randomized controlled experiment, our experiment does not satisfy the well-known Stable Unit Treatment Value Assumption from causal inference (23). As a consequence, it cannot provide unbiased estimates of causal quantities of interest, such as the average treatment

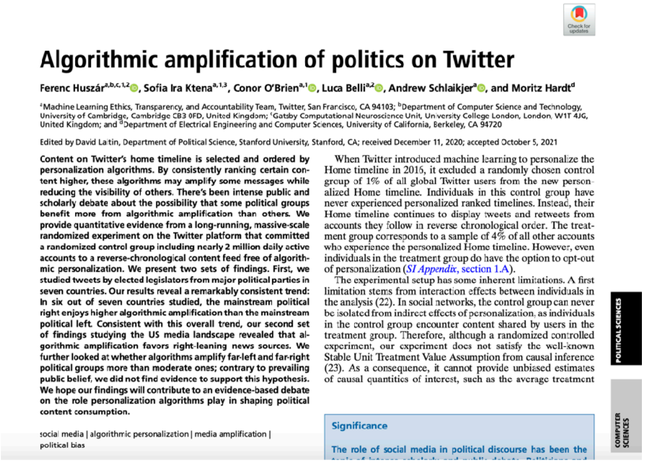
Significance

The role of social media in political discourse has been the

POLITICAL SCIENCES

COMPUTER SCIENCES

- Many studies about OSNs are empirical
  - Some data is collected, relevant phenomena are analyzed
    Do filter bubbles exist? Does the Twitter/X algorithm promote left- or right-wing content?
  - The OSN algorithms are unknown and only user actions can be observed



Algorithmic amplification of politics on Twitter

Ferenc Huszár[a,b,c,1,2], Sofia Ira Ktena[a,1,3], Conor O'Brien[a,1], Luca Belli[a,2], Andrew Schlaikjer[a], and Moritz Hardt[d]

[a]Machine Learning Ethics, Transparency, and Accountability Team, Twitter, San Francisco, CA 94103; [b]Department of Computer Science and Technology, University of Cambridge, Cambridge CB3 0FD, United Kingdom; [c]Gatsby Computational Neuroscience Unit, University College London, London, W1T 4JG, United Kingdom; and [d]Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720

Edited by David Latin, Department of Political Science, Stanford University, Stanford, CA; received December 11, 2020; accepted October 5, 2021

Content on Twitter's home timeline is selected and ordered by personalization algorithms. By consistently ranking certain content higher, these algorithms may amplify some messages while reducing the visibility of others. There's been intense public and scholarly debate about the possibility that some political groups benefit more from algorithmic amplification than others. We provide quantitative evidence from a long-running, massive-scale randomized experiment on the Twitter platform that committed a randomized control group including nearly 2 million daily active accounts to a reverse-chronological content feed free of algorithmic personalization. We present two sets of findings. First, we studied tweets by elected legislators from major political parties in seven countries. Our results reveal a remarkably consistent trend: In six out of seven countries studied, the mainstream political right enjoys higher algorithmic amplification than the mainstream political left. Consistent with this overall trend, our second set of findings studying the US media landscape revealed that algorithmic amplification favors right-leaning news sources. We further looked at whether algorithms amplify far-left and far-right political groups more than moderate ones; contrary to prevailing public belief, we did not find evidence to support this hypothesis. We hope our findings will contribute to an evidence-based debate on the role personalization algorithms play in shaping political content consumption.

social media | algorithmic personalization | media amplification | political bias

When Twitter introduced machine learning to personalize the Home timeline in 2016, it excluded a randomly chosen control group of 1% of all global Twitter users from the new personalized Home timeline. Individuals in this control group have never experienced personalized ranked timelines. Instead, their Home timeline continues to display tweets and retweets from accounts they follow in reverse chronological order. The treatment group corresponds to a sample of 4% of all other accounts who experience the personalized Home timeline. However, even individuals in the treatment group do have the option to opt-out of personalization (*SI Appendix*, section 1.A).

The experimental setup has some inherent limitations. A first limitation stems from interaction effects between individuals in the analysis (22). In social networks, the control group can never be isolated from indirect effects of personalization, as individuals in the control group encounter content shared by users in the treatment group. Therefore, although a randomized controlled experiment, our experiment does not satisfy the well-known Stable Unit Treatment Value Assumption from causal inference (23). As a consequence, it cannot provide unbiased estimates of causal quantities of interest, such as the average treatment

Significance

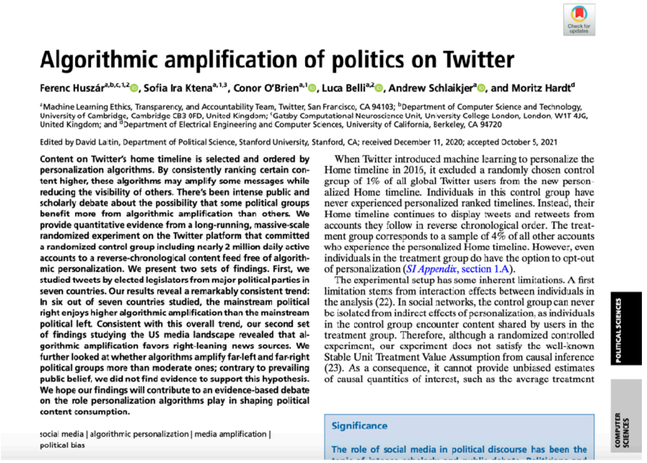
The role of social media in political discourse has been the

POLITICAL SCIENCES

COMPUTER SCIENCES

- Many studies about OSNs are empirical
  - Some data is collected, relevant phenomena are analyzed
    Do filter bubbles exist? Does the Twitter/X algorithm promote left- or right-wing content?
  - The OSN algorithms are unknown and only user actions can be observed

➡When Twitter/X changes its algorithm, unclear how findings generalize



Algorithmic amplification of politics on Twitter

Ferenc Huszár[a,b,c,1,2], Sofia Ira Ktena[a,1,3], Conor O'Brien[a,1], Luca Belli[a,2], Andrew Schlaikjer[a], and Moritz Hardt[d]

[a]Machine Learning Ethics, Transparency, and Accountability Team, Twitter, San Francisco, CA 94103; [b]Department of Computer Science and Technology, University of Cambridge, Cambridge CB3 0FD, United Kingdom; [c]Gatsby Computational Neuroscience Unit, University College London, London, W1T 4JG, United Kingdom; and [d]Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720

Edited by David Latin, Department of Political Science, Stanford University, Stanford, CA; received December 11, 2020; accepted October 5, 2021

Content on Twitter's home timeline is selected and ordered by personalization algorithms. By consistently ranking certain content higher, these algorithms may amplify some messages while reducing the visibility of others. There's been intense public and scholarly debate about the possibility that some political groups benefit more from algorithmic amplification than others. We provide quantitative evidence from a long-running, massive-scale randomized experiment on the Twitter platform that committed a randomized control group including nearly 2 million daily active accounts to a reverse-chronological content feed free of algorithmic personalization. We present two sets of findings. First, we studied tweets by elected legislators from major political parties in seven countries. Our results reveal a remarkably consistent trend: In six out of seven countries studied, the mainstream political right enjoys higher algorithmic amplification than the mainstream political left. Consistent with this overall trend, our second set of findings studying the US media landscape revealed that algorithmic amplification favors right-leaning news sources. We further looked at whether algorithms amplify far-left and far-right political groups more than moderate ones; contrary to prevailing public belief, we did not find evidence to support this hypothesis. We hope our findings will contribute to an evidence-based debate on the role personalization algorithms play in shaping political content consumption.

social media | algorithmic personalization | media amplification | political bias

When Twitter introduced machine learning to personalize the Home timeline in 2016, it excluded a randomly chosen control group of 1% of all global Twitter users from the new personalized Home timeline. Individuals in this control group have never experienced personalized ranked timelines. Instead, their Home timeline continues to display tweets and retweets from accounts they follow in reverse chronological order. The treatment group corresponds to a sample of 4% of all other accounts who experience the personalized Home timeline. However, even individuals in the treatment group do have the option to opt-out of personalization (*SI Appendix*, section 1.A).

The experimental setup has some inherent limitations. A first limitation stems from interaction effects between individuals in the analysis (22). In social networks, the control group can never be isolated from indirect effects of personalization, as individuals in the control group encounter content shared by users in the treatment group. Therefore, although a randomized controlled experiment, our experiment does not satisfy the well-known Stable Unit Treatment Value Assumption from causal inference (23). As a consequence, it cannot provide unbiased estimates of causal quantities of interest, such as the average treatment

Significance

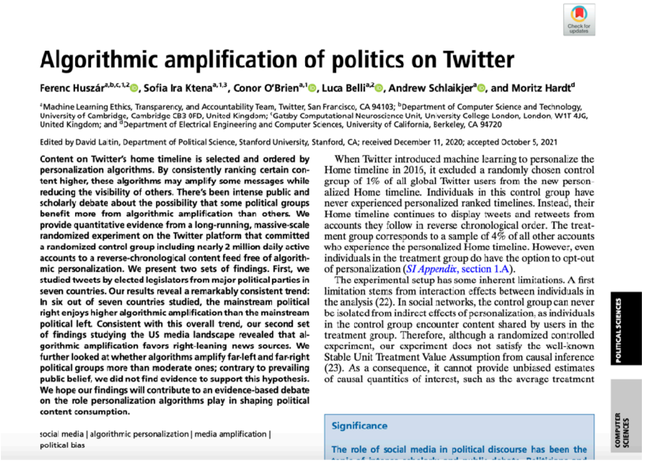
The role of social media in political discourse has been the

POLITICAL SCIENCES

COMPUTER SCIENCES

- Many studies about OSNs are empirical
  - Some data is collected, relevant phenomena are analyzed
    Do filter bubbles exist? Does the Twitter/X algorithm promote left- or right-wing content?
  - The OSN algorithms are unknown and only user actions can be observed

➡When Twitter/X changes its algorithm, unclear how findings generalize



Algorithmic amplification of politics on Twitter

Ferenc Huszár[a,b,c,1,2], Sofia Ira Ktena[a,1,3], Conor O'Brien[a,1], Luca Belli[a,2], Andrew Schlaikjer[a], and Moritz Hardt[d]

[a]Machine Learning Ethics, Transparency, and Accountability Team, Twitter, San Francisco, CA 94103; [b]Department of Computer Science and Technology, University of Cambridge, Cambridge CB3 0FD, United Kingdom; [c]Gatsby Computational Neuroscience Unit, University College London, London, W1T 4JG, United Kingdom; and [d]Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720

Edited by David Latin, Department of Political Science, Stanford University, Stanford, CA; received December 11, 2020; accepted October 5, 2021

Content on Twitter's home timeline is selected and ordered by personalization algorithms. By consistently ranking certain content higher, these algorithms may amplify some messages while reducing the visibility of others. There's been intense public and scholarly debate about the possibility that some political groups benefit more from algorithmic amplification than others. We provide quantitative evidence from a long-running, massive-scale randomized experiment on the Twitter platform that committed a randomized control group including nearly 2 million daily active accounts to a reverse-chronological content feed free of algorithmic personalization. We present two sets of findings. First, we studied tweets by elected legislators from major political parties in seven countries. Our results reveal a remarkably consistent trend: In six out of seven countries studied, the mainstream political right enjoys higher algorithmic amplification than the mainstream political left. Consistent with this overall trend, our second set of findings studying the US media landscape revealed that algorithmic amplification favors right-leaning news sources. We further looked at whether algorithms amplify far-left and far-right political groups more than moderate ones; contrary to prevailing public belief, we did not find evidence to support this hypothesis. We hope our findings will contribute to an evidence-based debate on the role personalization algorithms play in shaping political content consumption.

social media | algorithmic personalization | media amplification | political bias

Significance

The role of social media in political discourse has been the

POLITICAL SCIENCES

COMPUTER SCIENCES

The Empirical Approach

Humans interacting with OSNs in the real world

Black-box
(not accessible to researchers)

Observed via data collection

- Many studies about OSNs are empirical
  - Some data is collected, relevant phenomena are analyzed
    Do filter bubbles exist? Does the Twitter/X algorithm promote left- or right-wing content?
  - The OSN algorithms are unknown and only user actions can be observed

➡When Twitter/X changes its algorithm, unclear how findings generalize

Algorithmic amplification of politics on Twitter

Ferenc Huszár[a,b,c,1,2], Sofia Ira Ktena[a,1,3], Conor O'Brien[a,1], Luca Belli[a,2], Andrew Schlaikjer[a], and Moritz Hardt[d]

[a]Machine Learning Ethics, Transparency, and Accountability Team, Twitter, San Francisco, CA 94103; [b]Department of Computer Science and Technology, University of Cambridge, Cambridge CB3 0FD, United Kingdom; [c]Gatsby Computational Neuroscience Unit, University College London, London, W1T 4JG, United Kingdom; and [d]Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720

Edited by David Latin, Department of Political Science, Stanford University, Stanford, CA; received December 11, 2020; accepted October 5, 2021

Content on Twitter's home timeline is selected and ordered by personalization algorithms. By consistently ranking certain content higher, these algorithms may amplify some messages while reducing the visibility of others. There's been intense public and scholarly debate about the possibility that some political groups benefit more from algorithmic amplification than others. We provide quantitative evidence from a long-running, massive-scale randomized experiment on the Twitter platform that committed a randomized control group including nearly 2 million daily active accounts to a reverse-chronological content feed free of algorithmic personalization. We present two sets of findings. First, we studied tweets by elected legislators from major political parties in seven countries. Our results reveal a remarkably consistent trend: In six out of seven countries studied, the mainstream political right enjoys higher algorithmic amplification than the mainstream political left. Consistent with this overall trend, our second set of findings studying the US media landscape revealed that algorithmic amplification favors right-leaning news sources. We further looked at whether algorithms amplify far-left and far-right political groups more than moderate ones; contrary to prevailing public belief, we did not find evidence to support this hypothesis. We hope our findings will contribute to an evidence-based debate on the role personalization algorithms play in shaping political content consumption.

social media | algorithmic personalization | media amplification | political bias

When Twitter introduced machine learning to personalize the Home timeline in 2016, it excluded a randomly chosen control group of 1% of all global Twitter users from the new personalized Home timeline. Individuals in this control group have never experienced personalized ranked timelines. Instead, their Home timeline continues to display tweets and retweets from accounts they follow in reverse chronological order. The treatment group corresponds to a sample of 4% of all other accounts who experience the personalized Home timeline. However, even individuals in the treatment group do have the option to opt-out of personalization (*SI Appendix*, section 1.A).

The experimental setup has some inherent limitations. A first limitation stems from interaction effects between individuals in the analysis (22). In social networks, the control group can never be isolated from indirect effects of personalization, as individuals in the control group encounter content shared by users in the treatment group. Therefore, although a randomized controlled experiment, our experiment does not satisfy the well-known Stable Unit Treatment Value Assumption from causal inference (23). As a consequence, it cannot provide unbiased estimates of causal quantities of interest, such as the average treatment

Significance

The role of social media in political discourse has been the

POLITICAL SCIENCES

COMPUTER SCIENCES

The Empirical Approach

Humans interacting with OSNs in the real world

Black-box
(not accessible to researchers)

Observed via data collection

- Does not allow to try out different OSN algorithms
- **Causality testing difficult**

The Agent-Based Approach

From sociology, models human behavior

Abstract model of OSN algorithms or adversaries

Obtained via efficient simulation

The Empirical Approach

Humans interacting with OSNs in the real world

Black-box
(not accessible to researchers)

Observed via data collection

- Does not allow to try out different OSN algorithms
- **Causality testing difficult**

The Agent-Based Approach

From sociology, models human behavior

Abstract model of OSN algorithms or adversaries

Obtained via efficient simulation

- Allows to try out different OSN algorithms
- **Causal model** — we can simulate the impact of interventions

The Empirical Approach

Humans interacting with OSNs in the real world

Black-box (not accessible to researchers)

Observed via data collection

- Does not allow to try out different OSN algorithms
- **Causality testing difficult**

The Agent-Based Approach

From sociology, models human behavior

Abstract model of OSN algorithms or adversaries

Obtained via efficient simulation

- Allows to try out different OSN algorithms
- **Causal model** — we can simulate the impact of interventions
- **Vision:** Develop technical conditions to regulate OSN algorithms

The Empirical Approach

Humans interacting with OSNs in the real world

Black-box (not accessible to researchers)

Observed via data collection

- Does not allow to try out different OSN algorithms
- **Causality testing difficult**

The Agent-Based Approach

From sociology, models human behavior

Abstract model of OSN algorithms or adversaries

Obtained via efficient simulation

Opinions given by
$z^* = (I + L)^{-1}s$

- Allows to try out different OSN algorithms
- **Causal model** — we can simulate the impact of interventions
- **Vision:** Develop technical conditions to regulate OSN algorithms

The Empirical Approach

Humans interacting with OSNs in the real world

Black-box
(not accessible to researchers)

Observed via data collection

- Does not allow to try out different OSN algorithms
- **Causality testing difficult**

The Agent-Based Approach

From sociology, models human behavior

Abstract model of OSN algorithms or adversaries

Obtained via efficient simulation

Opinions given by $z^* = (I + L)^{-1}s$

Model changes to s or L — optimization problems

- Allows to try out different OSN algorithms
- **Causal model** — we can simulate the impact of interventions
- **Vision:** Develop technical conditions to regulate OSN algorithms

The Empirical Approach

Humans interacting with OSNs in the real world

Black-box (not accessible to researchers)

Observed via data collection

- Does not allow to try out different OSN algorithms
- **Causality testing difficult**

The Agent-Based Approach

From sociology, models human behavior

Abstract model of OSN algorithms or adversaries

Obtained via efficient simulation

Opinions given by $z^* = (I + L)^{-1}s$

Model changes to s or L — optimization problems

Solutions of optimization problems reveal real-world insights

- Allows to try out different OSN algorithms
- **Causal model** — we can simulate the impact of interventions
- **Vision:** Develop technical conditions to regulate OSN algorithms

The Empirical Approach

Humans interacting with OSNs in the real world

Black-box (not accessible to researchers)

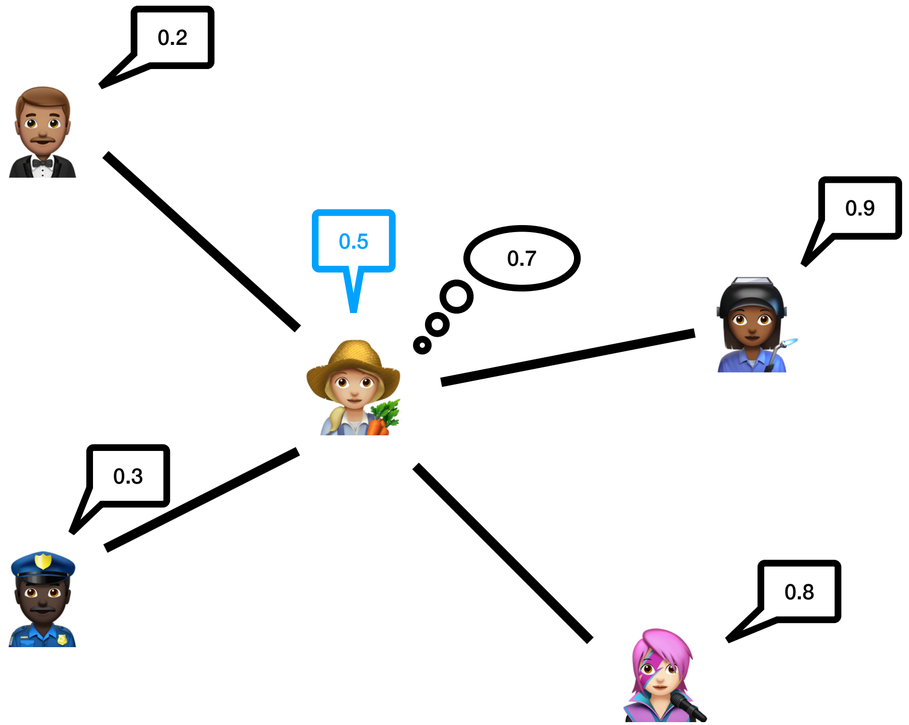
Observed via data collection

- Does not allow to try out different OSN algorithms
- **Causality testing difficult**

The FJ Opinion Formation Model

Friedkin, Johnsen (Journal of Mathematical Sociology, 1990)

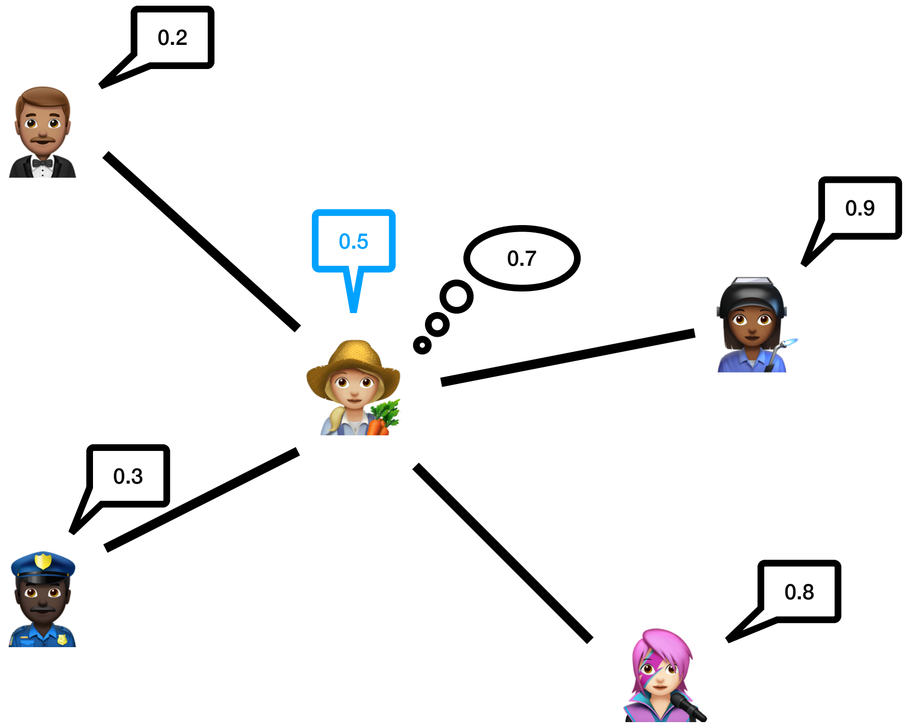
- The **Friedkin–Johnsen (FJ) model** is a popular opinion formation model
- $G = (V, E)$ with edge weights w_{uv} and Laplacian $\mathbf{L}$



The FJ Opinion Formation Model

Friedkin, Johnsen (Journal of Mathematical Sociology, 1990)

- The **Friedkin–Johnsen (FJ) model** is a popular opinion formation model
- $G = (V, E)$ with edge weights w_{uv} and Laplacian $\mathbf{L}$
- Each node $u \in V$, has a (public) expressed opinion $z_u \in [-1,1]$ and a (private) innate opinion $s_u \in [-1,1]$
  - Abstraction: Opinions are numbers in $[-1,1]$



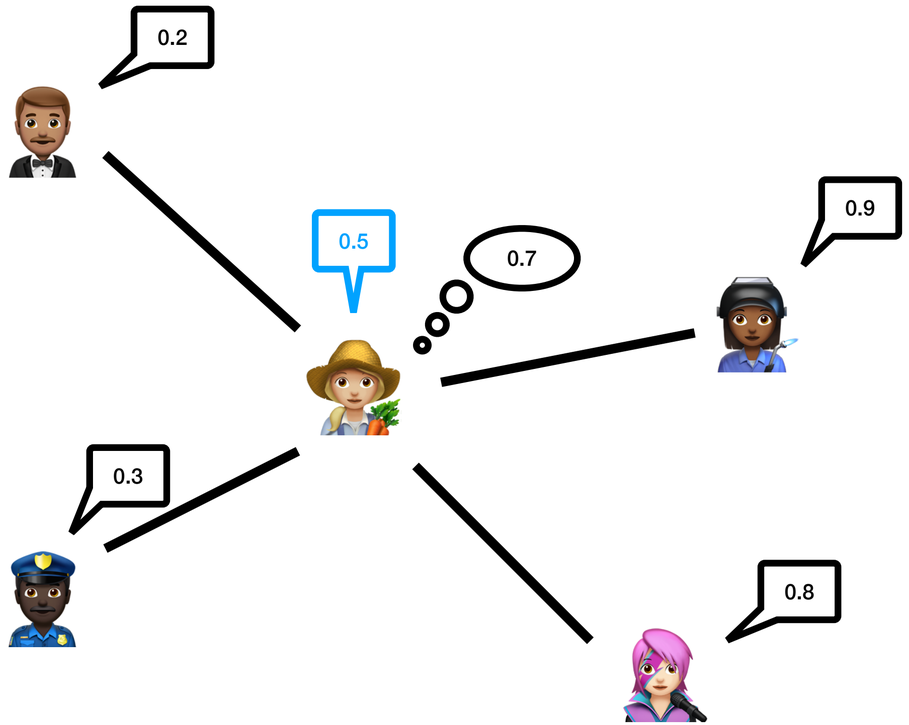
The FJ Opinion Formation Model

Friedkin, Johnsen (Journal of Mathematical Sociology, 1990)

- The **Friedkin–Johnsen (FJ) model** is a popular opinion formation model
- $G = (V, E)$ with edge weights w_{uv} and Laplacian $\mathbf{L}$
- Each node $u \in V$, has a (public) expressed opinion $z_u \in [-1,1]$ and a (private) innate opinion $s_u \in [-1,1]$
  - Abstraction: Opinions are numbers in $[-1,1]$
- Update rule for expressed opinions at time t:

$$z_u^{(t)} = \frac{s_u + \sum_{v \in N(u)} w_{uv} z_v^{(t-1)}}{1 + \sum_{v \in N(u)} w_{uv}}$$

  - Intuition: How people adapt their opinions due to peer-pressure



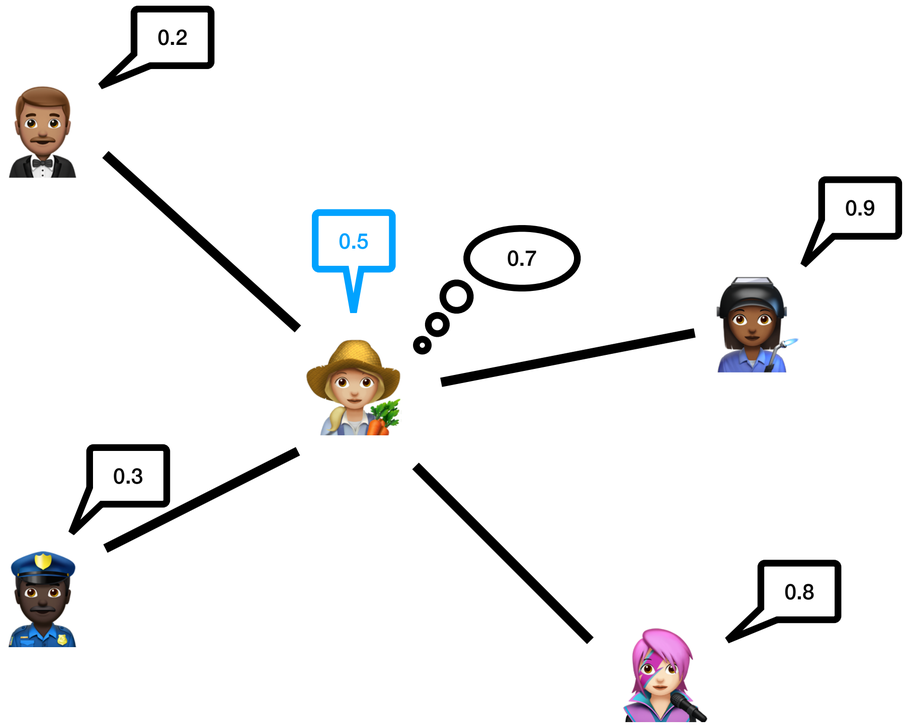
The FJ Opinion Formation Model

Friedkin, Johnsen (Journal of Mathematical Sociology, 1990)

- The **Friedkin–Johnsen (FJ) model** is a popular opinion formation model
- $G = (V, E)$ with edge weights w_{uv} and Laplacian $\mathbf{L}$
- Each node $u \in V$, has a (public) expressed opinion $z_u \in [-1,1]$ and a (private) innate opinion $s_u \in [-1,1]$
  - Abstraction: Opinions are numbers in $[-1,1]$
- Update rule for expressed opinions at time t:

$$z_u^{(t)} = \frac{s_u + \sum_{v \in N(u)} w_{uv} z_v^{(t-1)}}{1 + \sum_{v \in N(u)} w_{uv}}$$

  - Intuition: How people adapt their opinions due to peer-pressure
  - Equilibrium expressed opinions: $\mathbf{z}^* = \lim_{t \to \infty} \mathbf{z}^{(\mathbf{t})} = (\mathbf{I} + \mathbf{L})^{-1}\mathbf{s}$



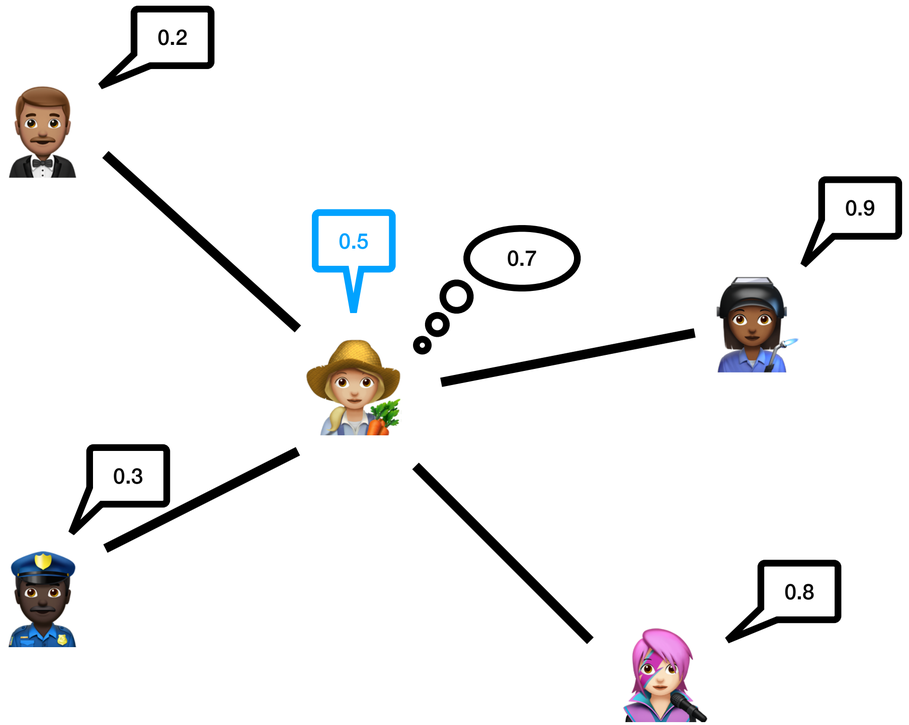
The FJ Opinion Formation Model

Friedkin, Johnsen (Journal of Mathematical Sociology, 1990)

- The **Friedkin–Johnsen (FJ) model** is a popular opinion formation model
- $G = (V, E)$ with edge weights w_{uv} and Laplacian $\mathbf{L}$
- Each node $u \in V$, has a (public) expressed opinion $z_u \in [-1,1]$ and a (private) innate opinion $s_u \in [-1,1]$
  - Abstraction: Opinions are numbers in $[-1,1]$
- Update rule for expressed opinions at time t:

$$z_u^{(t)} = \frac{s_u + \sum_{v \in N(u)} w_{uv} z_v^{(t-1)}}{1 + \sum_{v \in N(u)} w_{uv}}$$

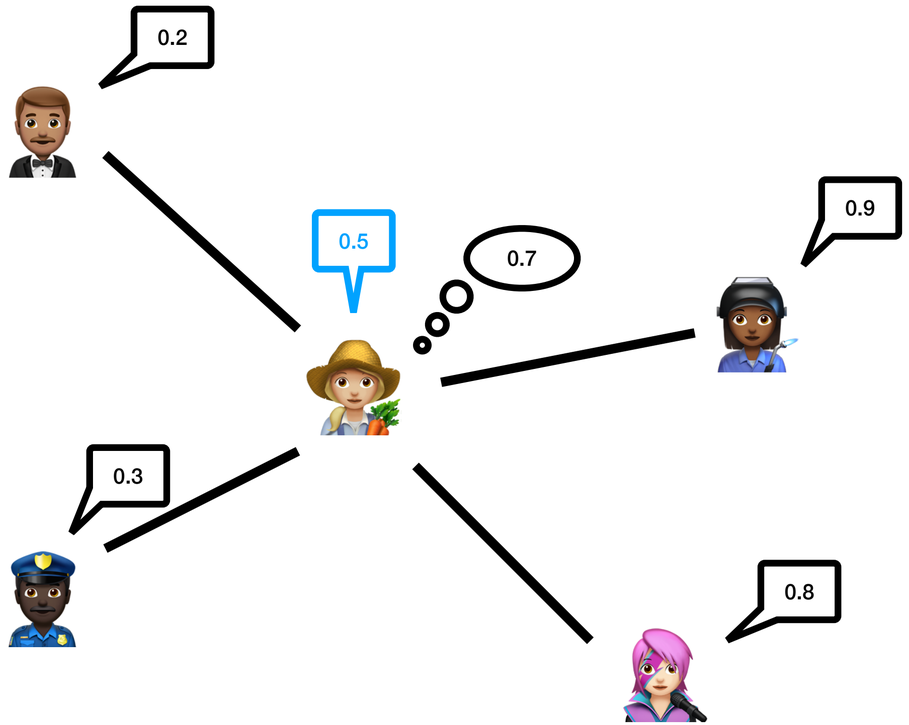
  - Intuition: How people adapt their opinions due to peer-pressure
  - Equilibrium expressed opinions: $\mathbf{z}^* = \lim_{t \to \infty} \mathbf{z}^{(\mathbf{t})} = (\mathbf{I} + \mathbf{L})^{-1}\mathbf{s}$
- Note that when G changes, then $\mathbf{L}$ changes, then $\mathbf{z}^*$ changes



Polarization and Disagreement

- In general, the FJ model does not converge to a consensus opinion

  ➡ Allows to study the network's polarization and the disagreement

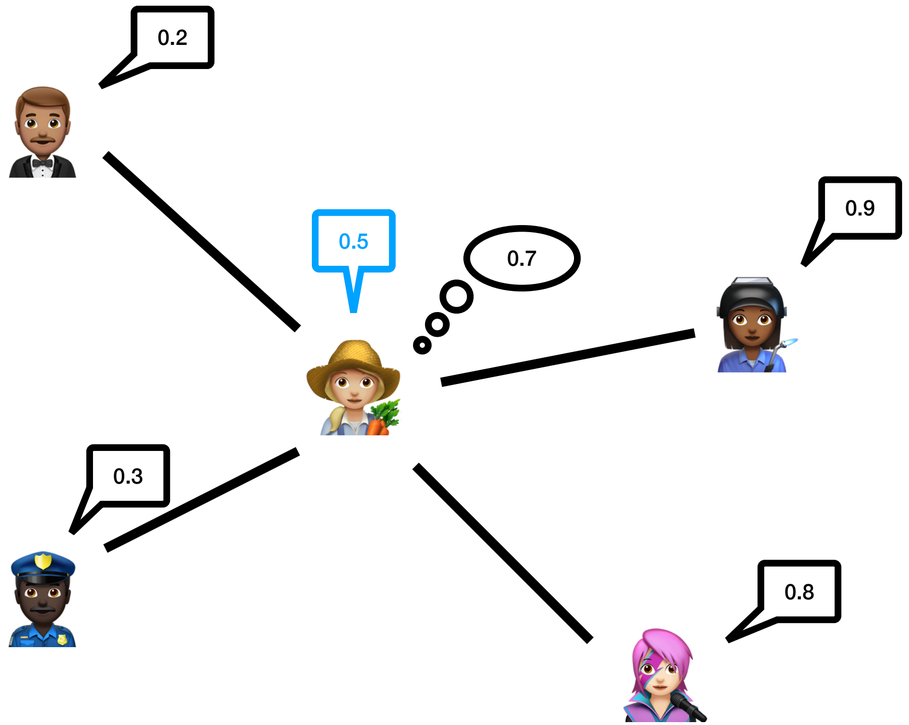


Polarization and Disagreement

- In general, the FJ model does not converge to a consensus opinion

  ➡ Allows to study the network's polarization and the disagreement

- Polarization = $\sum_{u \in V} (z_u^* - \bar{z})^2$, where $\bar{z} = \frac{1}{n} \sum_{u \in V} z_u^*$ – "variance of the opinions"
- Disagreement = $\sum_{(u,v) \in E} w_{u,v} (z_u^* - z_v^*)^2$ – stress among neighbors



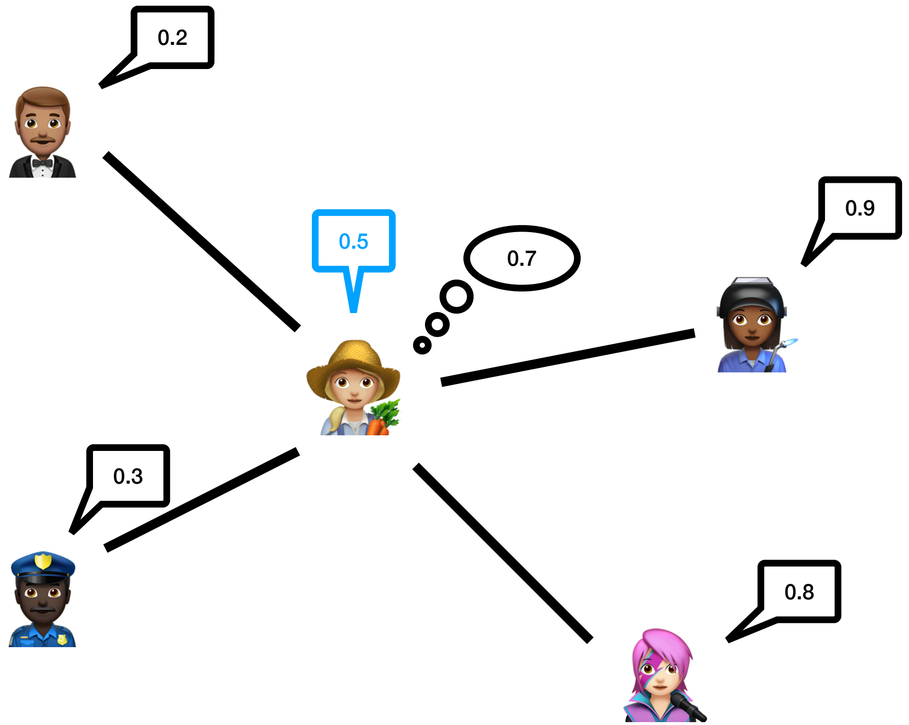
Polarization and Disagreement

- In general, the FJ model does not converge to a consensus opinion

  ➡ Allows to study the network's polarization and the disagreement

- Polarization = $\sum_{u \in V} (z_u^* - \bar{z})^2$, where $\bar{z} = \frac{1}{n} \sum_{u \in V} z_u^*$ – "variance of the opinions"
- Disagreement = $\sum_{(u,v) \in E} w_{u,v} (z_u^* - z_v^*)^2$ – stress among neighbors

➡ In linear algebra terms,
disagreement + polarization given by $\mathbf{s}^\top (\mathbf{I} + \mathbf{L})^{-1} \mathbf{s}$



Polarization and Disagreement

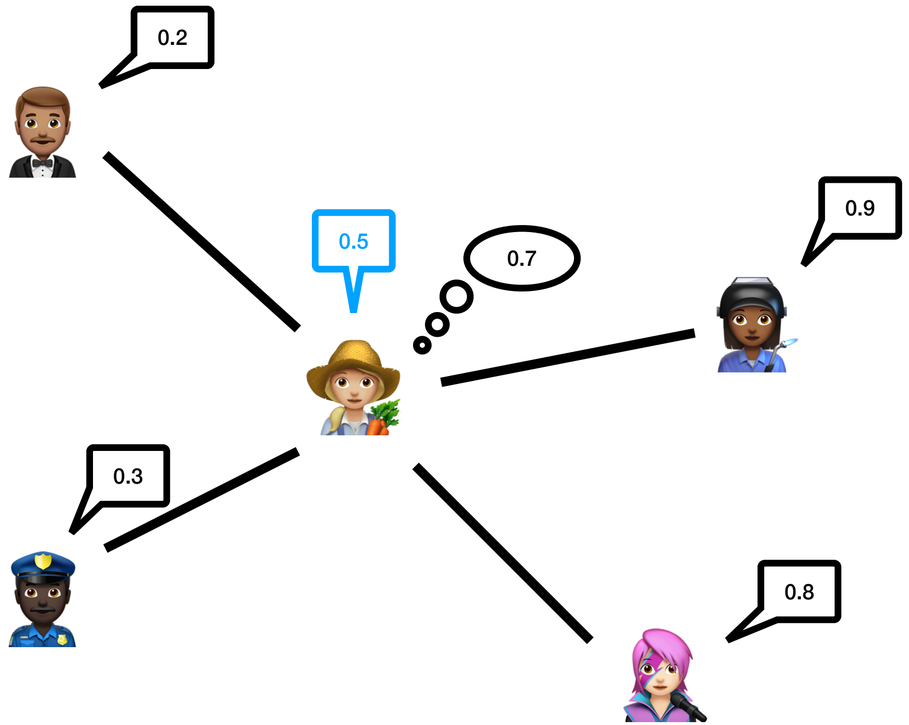
- In general, the FJ model does not converge to a consensus opinion

  ➡ Allows to study the network's polarization and the disagreement

- Polarization = $\sum_{u \in V} (z_u^* - \bar{z})^2$, where $\bar{z} = \frac{1}{n} \sum_{u \in V} z_u^*$ – "variance of the opinions"
- Disagreement = $\sum_{(u,v) \in E} w_{u,v} (z_u^* - z_v^*)^2$ – stress among neighbors

➡ In linear algebra terms,
disagreement + polarization given by $\mathbf{s}^\top (\mathbf{I} + \mathbf{L})^{-1} \mathbf{s}$

- **Now we can ask interesting questions:**



Polarization and Disagreement

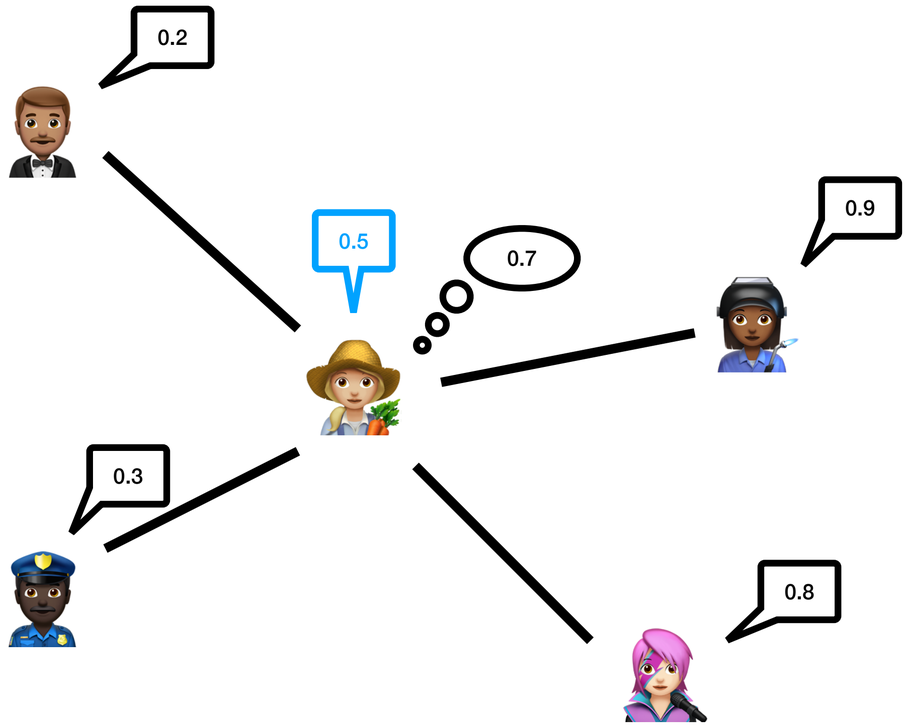
- In general, the FJ model does not converge to a consensus opinion

  ➡ Allows to study the network's polarization and the disagreement

- Polarization = $\sum_{u \in V} (z_u^* - \bar{z})^2$, where $\bar{z} = \frac{1}{n} \sum_{u \in V} z_u^*$ – "variance of the opinions"
- Disagreement = $\sum_{(u,v) \in E} w_{u,v} (z_u^* - z_v^*)^2$ – stress among neighbors

➡ In linear algebra terms,
disagreement + polarization given by $\mathbf{s}^\top (\mathbf{I} + \mathbf{L})^{-1} \mathbf{s}$

- **Now we can ask interesting questions:**
  - How does it effect the polarization/disagreement if…



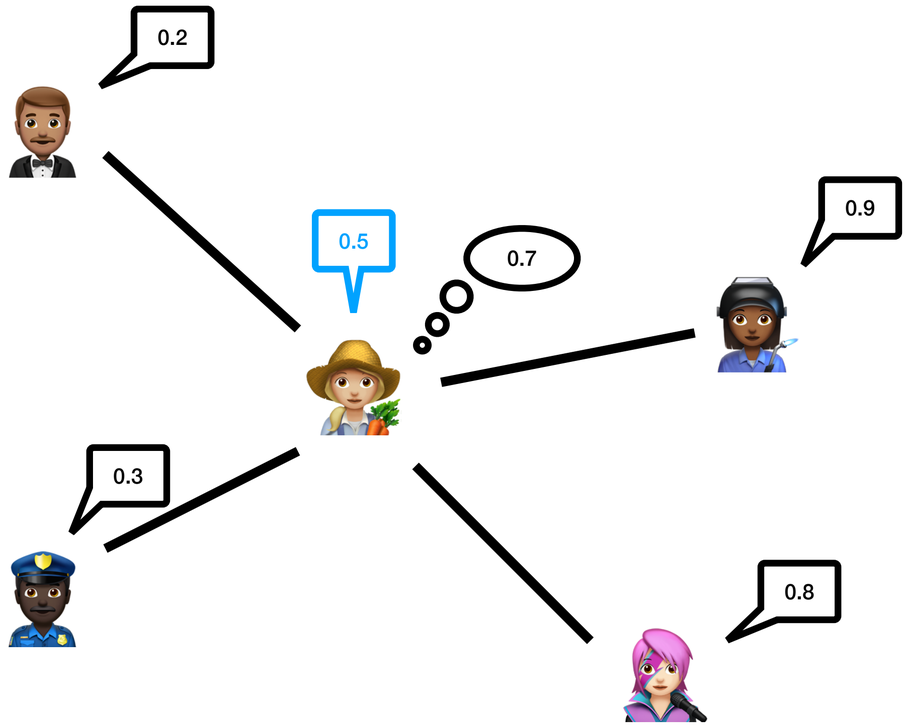
Polarization and Disagreement

- In general, the FJ model does not converge to a consensus opinion

  ➡ Allows to study the network's polarization and the disagreement

- Polarization = $\sum_{u \in V} (z_u^* - \bar{z})^2$, where $\bar{z} = \frac{1}{n} \sum_{u \in V} z_u^*$ – "variance of the opinions"
- Disagreement = $\sum_{(u,v) \in E} w_{u,v} (z_u^* - z_v^*)^2$ – stress among neighbors

➡ In linear algebra terms,
disagreement + polarization given by $\mathbf{s}^\top (\mathbf{I} + \mathbf{L})^{-1} \mathbf{s}$



- **Now we can ask interesting questions:**
  - How does it effect the polarization/disagreement if…
    - the graph changes (e.g., due to timeline algorithms), or if a few node opinions change?

Formal Study of Interventions

- How to study interventions formally?

Formal Study of Interventions

- How to study interventions formally?
- **Optimization problem**:
  - Objective function encodes the desired goal
  - Constraints encode the power of the intervention

Formal Study of Interventions

- How to study interventions formally?
- **Optimization problem**:
  - Objective function encodes the desired goal
  - Constraints encode the power of the intervention
- **Example:**

  Minimize the disagreement while making few changes to the original graph structure $\mathbf{L_0}$:

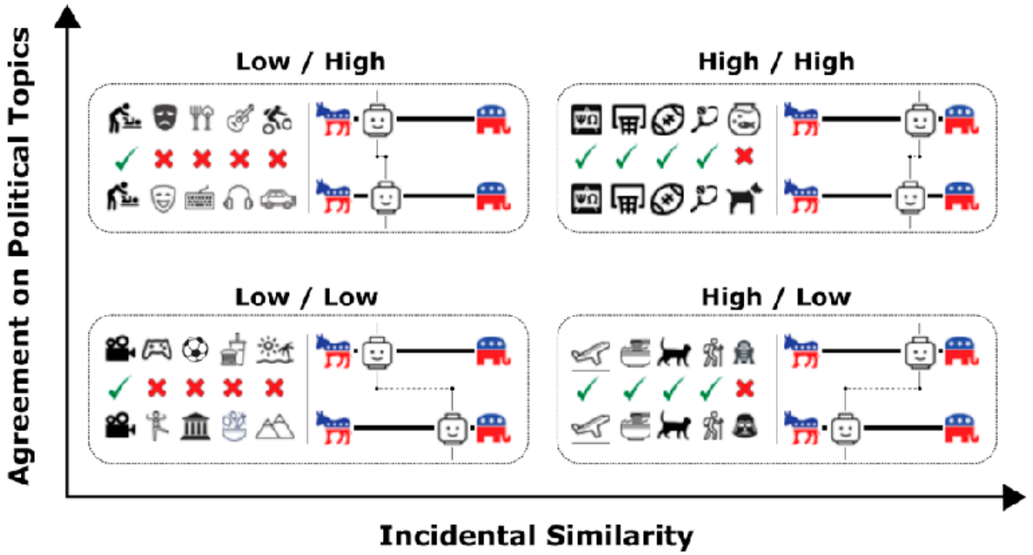
$$\min_{G'} \text{disagreement} \quad \text{s.t. } G' \text{ is close to } G \iff \min_{L \in \mathscr{L}} \sum_{(u,v) \in E} w_{u,v}(z_u^* - z_v^*)^2 \quad \text{s.t. } \|\mathbf{L} - \mathbf{L_0}\|_F \leq C$$

Modeling the Impact of Timeline Algorithms on Opinion Dynamics

Tianyi Zhou, *Stefan Neumann*, Kiran Garimella, Aris Gionis — WebConf'24

Motivation

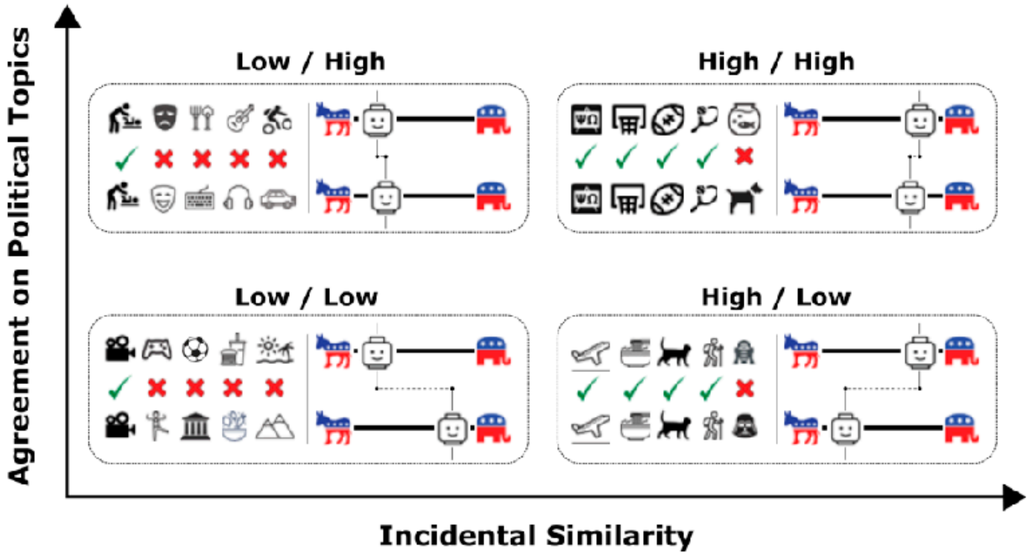
- Important question how we can reduce polarization in (online) social networks
- Recent empirical study by Balietti et al.:
  - Users with similar (non-political) interests are more likely to align their opinions (even if they disagree)



Reducing opinion polarization: Effects of exposure to similar people with differing political views
Stefano Balietti, Lise Getoor, Daniel G. Goldstein, and Duncan J. Watts
PNAS 2021 Vol. 118 No. 52 e2112552118

Motivation

- Important question how we can reduce polarization in (online) social networks
- Recent empirical study by Balietti et al.:
  - Users with similar (non-political) interests are more likely to align their opinions (even if they disagree)
- **Our questions:**
  - How can timeline algorithms of online social networks exploit such behaviors?
  - Can we model this using opinion formation models?
  - Can we optimize the timelines to reduce disagreement and polarization?

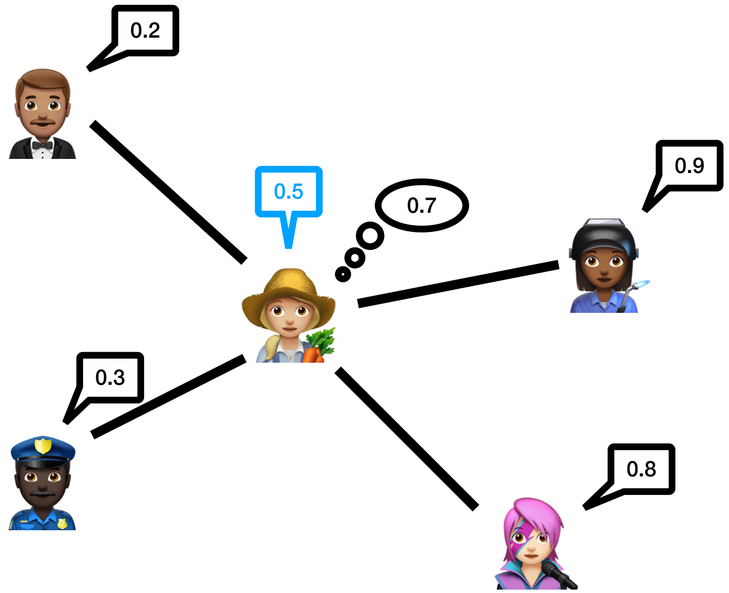


Reducing opinion polarization: Effects of exposure to similar people with differing political views
Stefano Balietti, Lise Getoor, Daniel G. Goldstein, and Duncan J. Watts
PNAS 2021 Vol. 118 No. 52 e2112552118

The Underlying Challenge

The Underlying Challenge

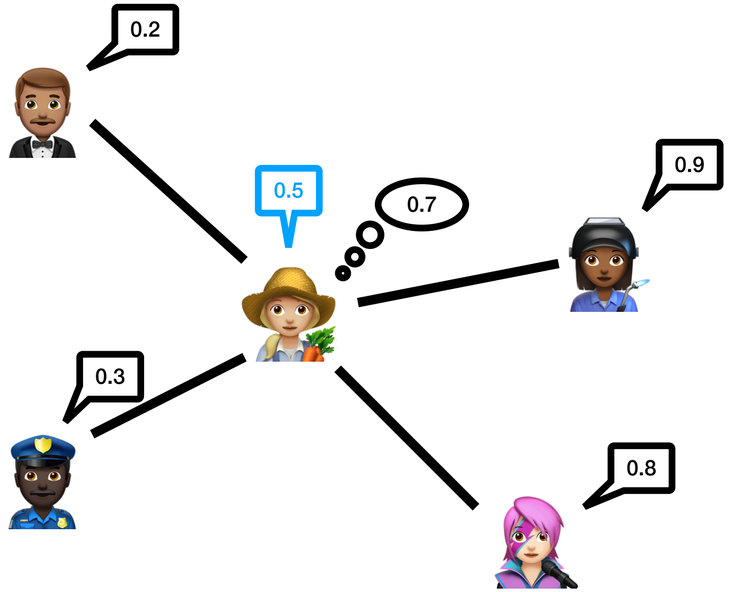
- **Goal:**
  Incorporate user interests and the effect of timeline algorithms into opinion formation models



The Underlying Challenge

- **Goal:**
  Incorporate user interests and the effect of timeline algorithms into opinion formation models
- **Challenge:**
  - Opinion formation models are defined on **graphs**
  - Timeline algorithms provide **content** to users

    ➡Content is picked based on users' interests in different topics

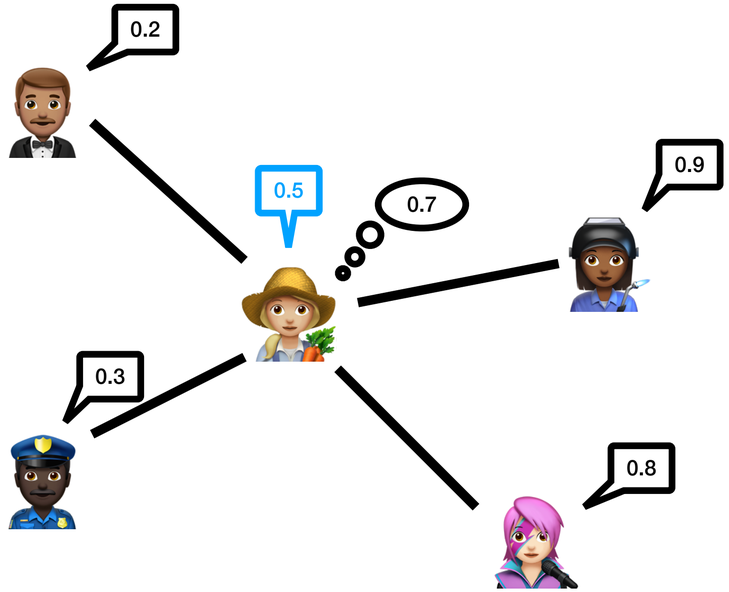


The Underlying Challenge

- **Goal:**
  Incorporate user interests and the effect of timeline algorithms into opinion formation models
- **Challenge:**
  - Opinion formation models are defined on **graphs**
  - Timeline algorithms provide **content** to users

    ➡Content is picked based on users' interests in different topics

➡How to combine these two abstraction levels?



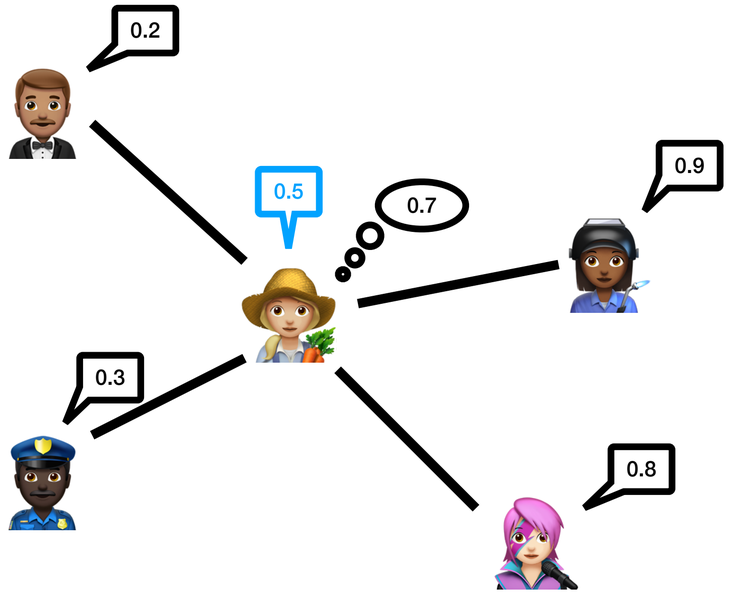
The Underlying Challenge

- **Goal:**
  Incorporate user interests and the effect of timeline algorithms into opinion formation models
- **Challenge:**
  - Opinion formation models are defined on **graphs**
  - Timeline algorithms provide **content** to users

    ➡Content is picked based on users' interests in different topics

➡How to combine these two abstraction levels?

- **Our approach:** Consider a combined graph consisting of
  - Fixed graph, based on real-world friendships or "follow"-graph
  - Recommender graph, based on aggregate information from timeline algorithm



Aggregate Information About User Interests

Aggregate Information About User Interests

- Suppose there are k topics (and k is small)

Aggregate Information About User Interests

- Suppose there are k topics (and k is small)
- User–topic matrix $\mathbf{X}$:
  - Models users' timeline decomposition
  - $\mathbf{X}_{ij}$ = fraction of content for user i from topic j

➡The content recommended to user 🧑‍🌾 is 80% about basketball, 10% about food and 10% about news

	🖼️	🏀	🌮	📰
👩‍🎨	0.7	0.2	0.1	0.0
👨‍🔧	0.2	0.1	0.1	0.6
🧑‍🌾	0.0	0.8	0.1	0.1

User–topic matrix $\mathbf{X}$

Aggregate Information About User Interests

- Suppose there are k topics (and k is small)
- User–topic matrix $\mathbf{X}$:
  - Models users' timeline decomposition
  - $\mathbf{X}_{ij}$ = fraction of content for user i from topic j

  ➡The content recommended to user 🧑‍🌾 is 80% about basketball, 10% about food and 10% about news
- Topic–influence matrix $\mathbf{Y}$:
  - Models how influential users are for different topics
  - For topic j, a $\mathbf{Y}_{ij}$-fraction of recommended content is from user i

  ➡For the topic basketball, 10% of the recommended content is by 👩‍🎨, 20% is by 👷 and 70% is by 🧑‍🌾

	🖼️	🏀	🌮	📰
👩‍🎨	0.7	0.2	0.1	0.0
👷	0.2	0.1	0.1	0.6
🧑‍🌾	0.0	0.8	0.1	0.1

User–topic matrix $\mathbf{X}$

	👩‍🎨	👷	🧑‍🌾
🖼️	0.7	0.2	0.1
🏀	0.1	0.2	0.7
🌮	0.3	0.1	0.6
📰	0.0	0.9	0.1

Topic–influence matrix $\mathbf{Y}$

Modeling Timeline Algorithms Based on User Interests

- Observe that the matrix product $\mathbf{XY}$ models the edges introduced by the timeline algorithm

	🖼️	🏀	🌮	📰
🤦	0.7	0.2	0.1	0.0
🤷	0.2	0.1	0.1	0.6
🧑‍🌾	0.0	0.8	0.1	0.1

User–topic matrix $\mathbf{X}$

	🤦	🤷	🧑‍🌾
🖼️	0.7	0.2	0.1
🏀	0.1	0.2	0.7
🌮	0.3	0.1	0.6
📰	0.0	0.9	0.1

Topic–influence matrix $\mathbf{Y}$

Modeling Timeline Algorithms Based on User Interests

- Observe that the matrix product $\mathbf{XY}$ models the edges introduced by the timeline algorithm

  ➡In the timeline of user 👩‍🎨,
  54% of the content is from 👩‍🎨,
  19% is from 👱‍♀️ and 27% is from 👩‍🌾

  ➡This matrix has rank k (“low rank”),
  important for efficient simulation

	🖼️	🏀	🌮	📰
👩‍🎨	0.7	0.2	0.1	0.0
👱‍♀️	0.2	0.1	0.1	0.6
👩‍🌾	0.0	0.8	0.1	0.1

User–topic matrix $\mathbf{X}$

	👩‍🎨	👱‍♀️	👩‍🌾
🖼️	0.7	0.2	0.1
🏀	0.1	0.2	0.7
🌮	0.3	0.1	0.6
📰	0.0	0.9	0.1

Topic–influence matrix $\mathbf{Y}$

$\mathbf{XY} =$

	👩‍🎨	👱‍♀️	👩‍🌾
👩‍🎨	0.54	0.19	0.27
👱‍♀️	0.18	0.61	0.21
👩‍🌾	0.11	0.26	0.63

Modeling Timeline Algorithms Based on User Interests

- Observe that the matrix product $\mathbf{XY}$ models the edges introduced by the timeline algorithm

  ➡In the timeline of user 🤦, 54% of the content is from 🤦, 19% is from 👱‍♀️ and 27% is from 🤠

  ➡This matrix has rank k (“low rank”), important for efficient simulation

- We consider combined graph with adjacency matrix

$$\mathbf{A} + \alpha\left(\mathbf{XY} + \mathbf{Y}^\intercal\mathbf{X}^\intercal\right)$$

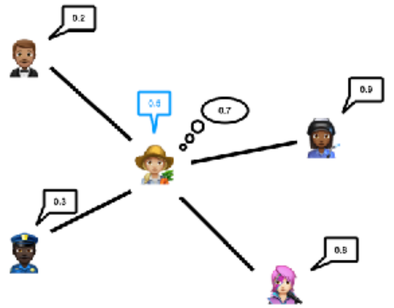
  - $\mathbf{A}$ is adjacency matrix of the fixed graph
  - α is a scaling term measuring how important recommendations are
  - Corresponds to adding up fixed graph and recommender graph
  - Added symmetrization for analysis

	🖼️	🏀	🌮	📰
🤦	0.7	0.2	0.1	0.0
👱‍♀️	0.2	0.1	0.1	0.6
🤠	0.0	0.8	0.1	0.1

User–topic matrix $\mathbf{X}$

	🤦	👱‍♀️	🤠
🖼️	0.7	0.2	0.1
🏀	0.1	0.2	0.7
🌮	0.3	0.1	0.6
📰	0.0	0.9	0.1

Topic–influence matrix $\mathbf{Y}$



Fixed graph

+

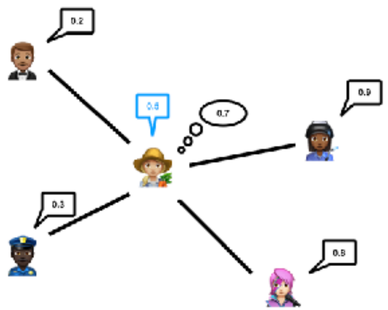
$\mathbf{XY} =$

	🤦	👱‍♀️	🤠
🤦	0.54	0.19	0.27
👱‍♀️	0.18	0.61	0.21
🤠	0.11	0.26	0.63

Recommender graph

Minimizing Polarization + Disagreement

- **Goal:** Update users' timelines to minimize polarization and disagreement



Fixed graph

+

$\mathbf{XY} =$

0.54	0.19	0.27
0.18	0.61	0.21
0.11	0.26	0.63

Recommender graph

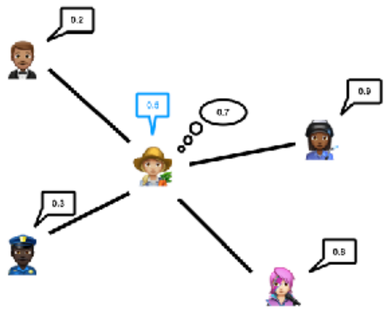
Minimizing Polarization + Disagreement

- **Goal:** Update users' timelines to minimize polarization and disagreement

$$\min_{\tilde{\mathbf{X}}} \quad \mathbf{s}^\intercal(\mathbf{I}+\mathbf{L}_\mathbf{A})^{-1}\mathbf{s}$$

$$\text{s.t.} \quad |\tilde{\mathbf{X}}_{ij} - \mathbf{X}_{ij}| \leq \theta \quad \forall i,j$$

  - Where $\mathbf{L}_{\mathbf{A}'}$ is the Laplacian of the graph $\mathbf{A} + \alpha\left(\mathbf{XY} + \mathbf{Y}^\intercal\mathbf{X}^\intercal\right)$
  - We can make small modifications to the timeline decomposition for each user (given by $\mathbf{X}$)



Fixed graph

+

$\mathbf{XY} =$

0.54	0.19	0.27
0.18	0.61	0.21
0.11	0.26	0.63

Recommender graph

Minimizing Polarization + Disagreement

- **Goal:** Update users' timelines to minimize polarization and disagreement

$$\min_{\tilde{\mathbf{X}}} \quad \mathbf{s}^\intercal(\mathbf{I}+\mathbf{L}_\mathbf{A})^{-1}\mathbf{s}$$

$$\text{s.t.} \quad |\tilde{\mathbf{X}}_{ij}-\mathbf{X}_{ij}| \le \theta \quad \forall i,j$$

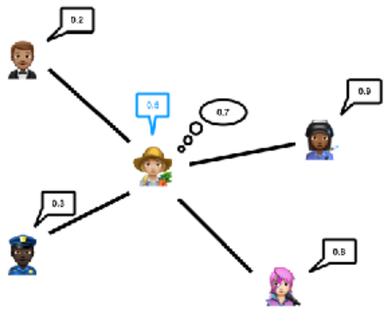
  - Where $\mathbf{L}_{\mathbf{A}'}$ is the Laplacian of the graph $\mathbf{A}+\alpha\left(\mathbf{X}\mathbf{Y}+\mathbf{Y}^\intercal\mathbf{X}^\intercal\right)$
  - We can make small modifications to the timeline decomposition for each user (given by $\mathbf{X}$)

	🖼️	🏀	🌮	📰
👩‍🌾	0.7	0.2	0.1	0.0
👷	0.2	0.1	0.1	0.6
🤠	0.0	0.8	0.1	0.1

Initial user–topic matrix $\mathbf{X}$

	🖼️	🏀	🌮	📰
👩‍🌾	**0.6**	0.2	0.1	**0.1**
👷	0.2	0.1	**0.2**	**0.5**
🤠	0.0	**0.9**	**0.0**	0.1

New user–topic matrix $\tilde{\mathbf{X}}$



Fixed graph

+

$\mathbf{XY} =$

👩‍🌾	0.54	0.19	0.27
👷	0.18	0.61	0.21
🤠	0.11	0.26	0.63

Recommender graph

Minimizing Polarization + Disagreement

- **Goal:** Update users' timelines to minimize polarization and disagreement

$$\min_{\tilde{\mathbf{X}}} \quad \mathbf{s}^\intercal(\mathbf{I}+\mathbf{L}_\mathbf{A})^{-1}\mathbf{s}$$

$$\text{s.t.} \quad |\tilde{\mathbf{X}}_{ij}-\mathbf{X}_{ij}| \le \theta \quad \forall i,j$$

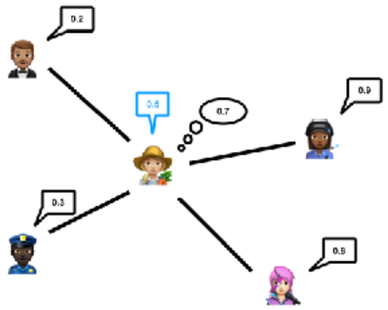
  - Where $\mathbf{L}_{\mathbf{A}'}$ is the Laplacian of the graph $\mathbf{A}+\alpha\left(\mathbf{X}\mathbf{Y}+\mathbf{Y}^\intercal\mathbf{X}^\intercal\right)$
  - We can make small modifications to the timeline decomposition for each user (given by $\mathbf{X}$)
  - Parameter θ controls amount of allowed changes
- **Efficient optimization algorithm:**
  - Can compute $(1+\varepsilon)$-approximate solution in time $O(m\sqrt{n})$ — in practice even faster
  - Gradient has closed form and can be computed efficiently
  - We examine solutions and build a combinatorial greedy algorithm that "mimics" the results of the continuous optimization algorithm

	0.7	0.2	0.1	0.0
	0.2	0.1	0.1	0.6
	0.0	0.8	0.1	0.1

Initial user–topic matrix $\mathbf{X}$

	0.6	0.2	0.1	**0.1**
	0.2	0.1	**0.2**	**0.5**
	0.0	**0.9**	**0.0**	0.1

New user–topic matrix $\tilde{\mathbf{X}}$



Fixed graph

+

$\mathbf{XY} =$

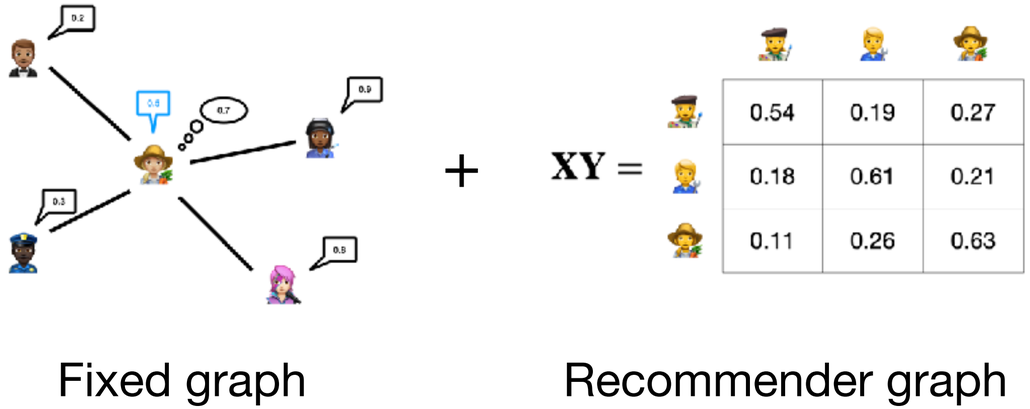
	0.54	0.19	0.27
	0.18	0.61	0.21
	0.11	0.26	0.63

Recommender graph

Experimental Evaluation

Datasets

- We **collected two real-world datasets** from Twitter
  - Larger dataset has 27k nodes and 268k edges
  - We obtain their retweets and based on them estimate interests $\mathbf{X}$ and influence $\mathbf{Y}$
  - Edges correspond to who follows whom (fixed graph)
  - We estimate their opinions by looking at who they follow
  - Data is available online
- Evaluation on 25 other graphs with real-world topology and synthetic opinions and $\mathbf{X}$ and $\mathbf{Y}$



Fixed graph

Recommender graph

Strengthens Controversial Topics

- We run our algorithm which converges to optimal solution and inspect solution

🖼️	0.7	0.2	0.1	0.0
🏀	0.2	0.1	0.1	0.6
🌮	0.0	0.8	0.1	0.1

User–topic matrix $\mathbf{X}$

🖼️	0.7	0.2	0.1
🏀	0.1	0.2	0.7
🌮	0.3	0.1	0.6
📰	0.0	0.9	0.1

Topic–influence matrix $\mathbf{Y}$

Strengthens Controversial Topics

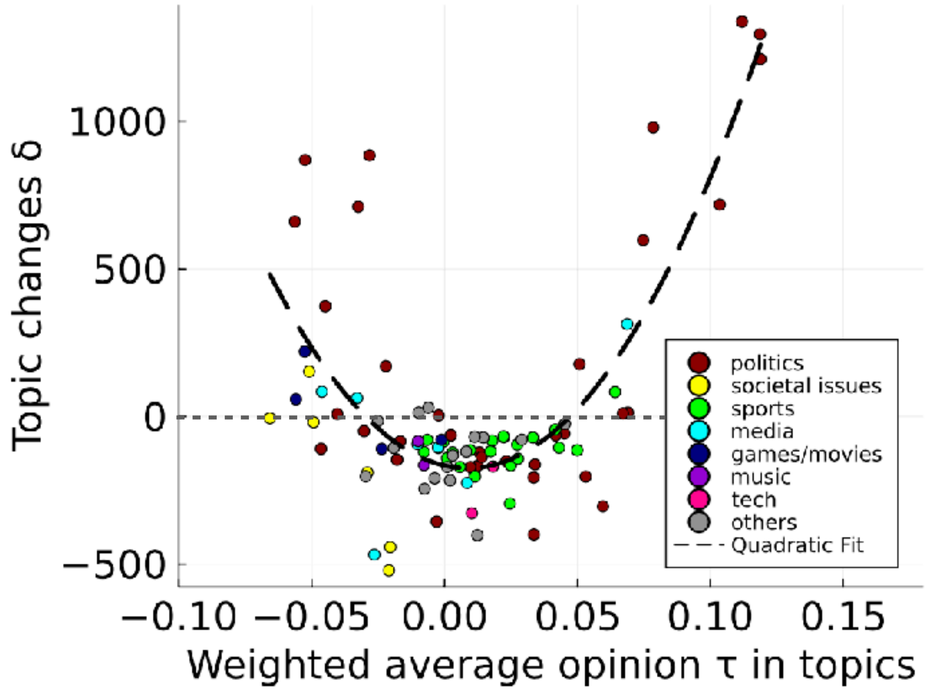
- We run our algorithm which converges to optimal solution and inspect solution

👨‍🎨	0.7	0.2	0.1	0.0
👷	0.2	0.1	0.1	0.6
👳	0.0	0.8	0.1	0.1

User–topic matrix $\mathbf{X}$

	👨‍🎨	👷	👳
🖼️	0.7	0.2	0.1
🏀	0.1	0.2	0.7
🌮	0.3	0.1	0.6
📰	0.0	0.9	0.1

Topic–influence matrix $\mathbf{Y}$



Strengthens Controversial Topics

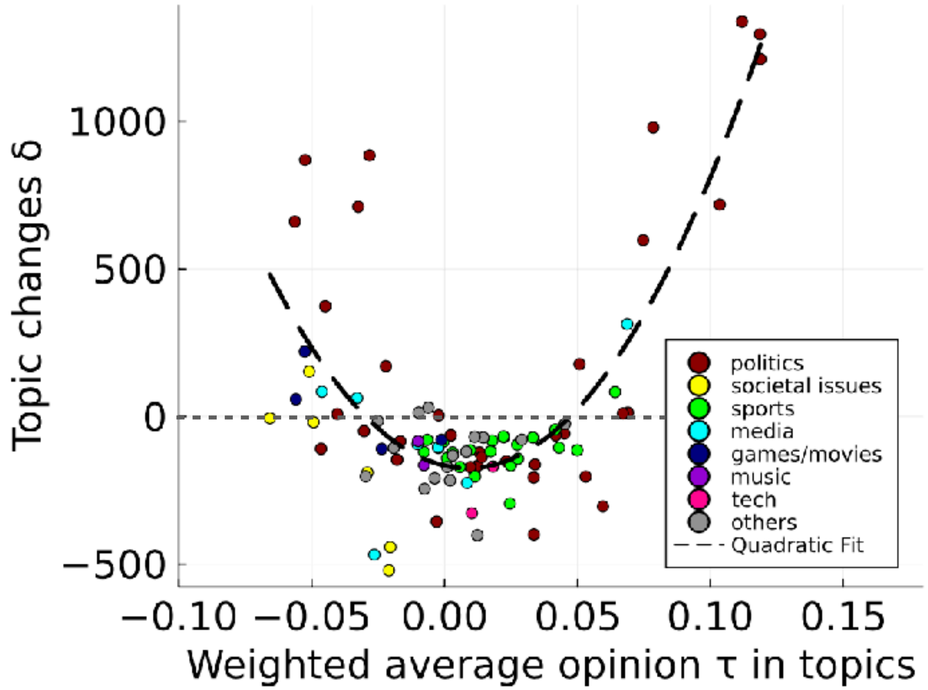
- We run our algorithm which converges to optimal solution and inspect solution
- y-axis: How much more/less important did each topic become during optimization?

🧑‍🎨	0.7	0.2	0.1	0.0
👷	0.2	0.1	0.1	0.6
👳	0.0	0.8	0.1	0.1

User–topic matrix $\mathbf{X}$

🖼️	0.7	0.2	0.1
🏀	0.1	0.2	0.7
🌮	0.3	0.1	0.6
📰	0.0	0.9	0.1

Topic–influence matrix $\mathbf{Y}$



Strengthens Controversial Topics

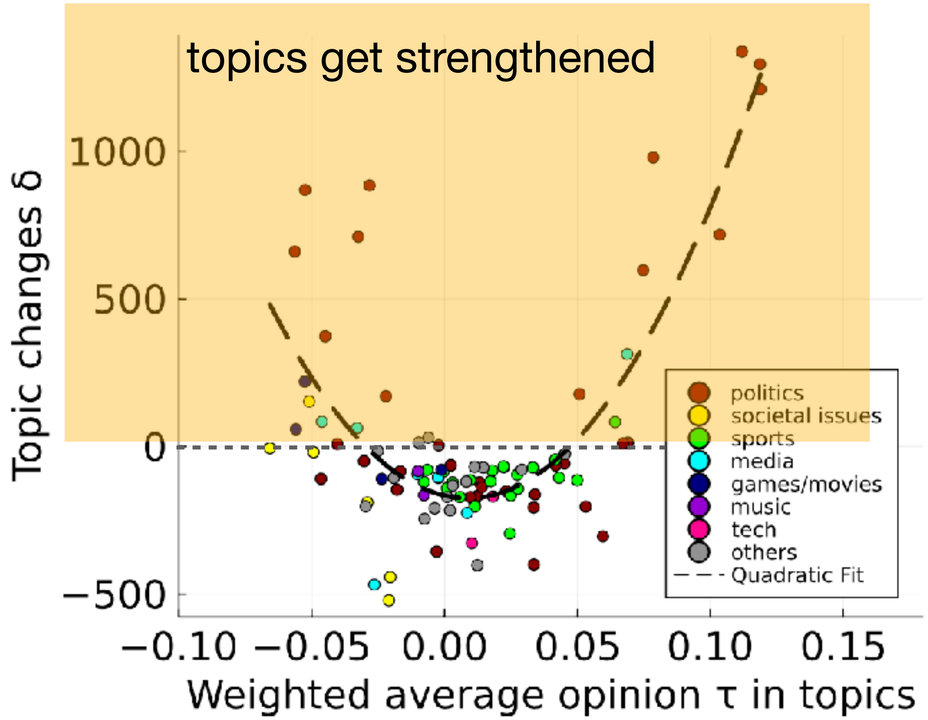
- We run our algorithm which converges to optimal solution and inspect solution
- y-axis: How much more/less important did each topic become during optimization?

👨‍🎨	0.7	0.2	0.1	0.0
👷	0.2	0.1	0.1	0.6
👲	0.0	0.8	0.1	0.1

User–topic matrix $\mathbf{X}$

🖼️	0.7	0.2	0.1
🏀	0.1	0.2	0.7
🌮	0.3	0.1	0.6
📰	0.0	0.9	0.1

Topic–influence matrix $\mathbf{Y}$



Strengthens Controversial Topics

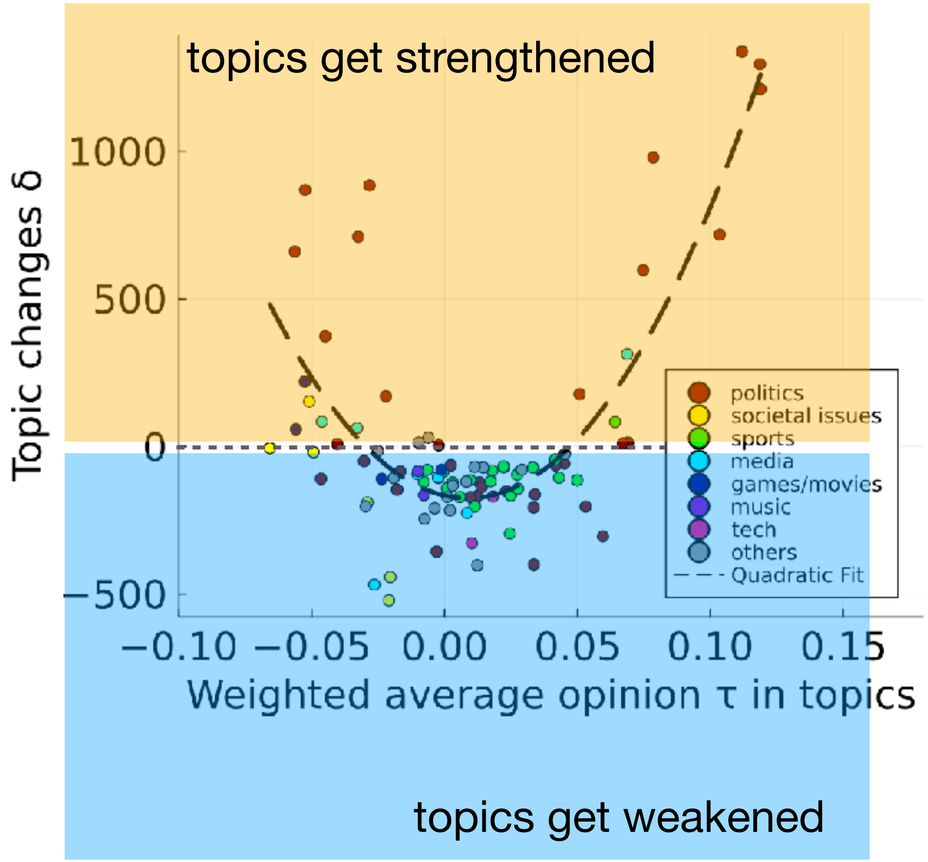
- We run our algorithm which converges to optimal solution and inspect solution
- y-axis: How much more/less important did each topic become during optimization?

👨‍🎨	0.7	0.2	0.1	0.0
👨‍🔧	0.2	0.1	0.1	0.6
👷	0.0	0.8	0.1	0.1

User–topic matrix $\mathbf{X}$

🖼️	0.7	0.2	0.1
🏀	0.1	0.2	0.7
🌮	0.3	0.1	0.6
📰	0.0	0.9	0.1

Topic–influence matrix $\mathbf{Y}$



Strengthens Controversial Topics

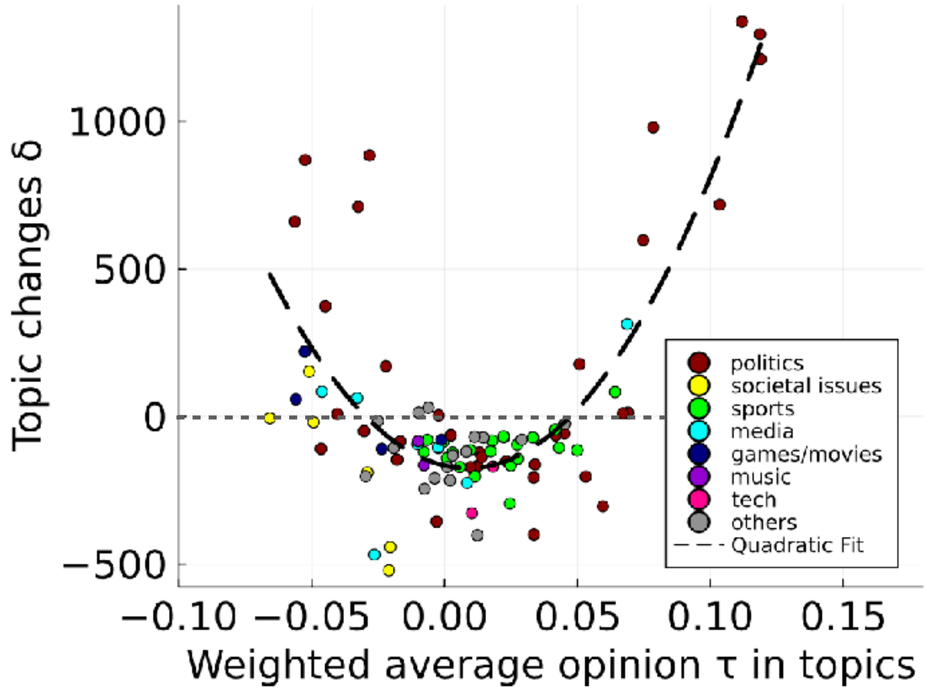
- We run our algorithm which converges to optimal solution and inspect solution
- y-axis: How much more/less important did each topic become during optimization?

🧑‍🎨	0.7	0.2	0.1	0.0
👷	0.2	0.1	0.1	0.6
👳	0.0	0.8	0.1	0.1

User–topic matrix $\mathbf{X}$

🖼️	0.7	0.2	0.1
🏀	0.1	0.2	0.7
🌮	0.3	0.1	0.6
📰	0.0	0.9	0.1

Topic–influence matrix $\mathbf{Y}$



Strengthens Controversial Topics

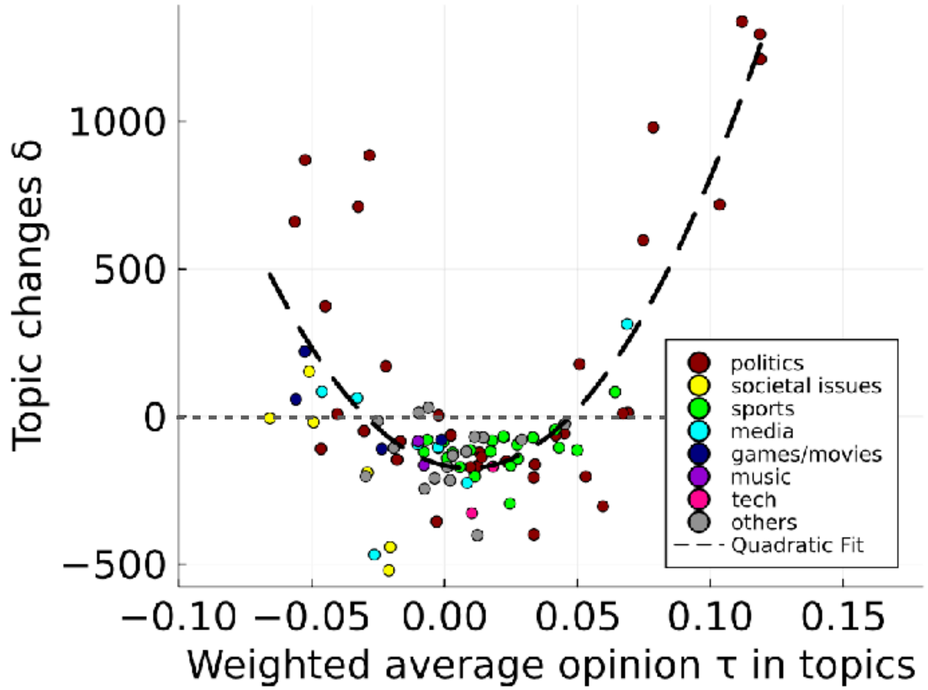
- We run our algorithm which converges to optimal solution and inspect solution
- y-axis: How much more/less important did each topic become during optimization?
- x-axis: Average leaning of influencers for each topic

🏞️	0.7	0.2	0.1	0.0
🏀	0.2	0.1	0.1	0.6
🌮	0.0	0.8	0.1	0.1

User–topic matrix $\mathbf{X}$

🏞️	0.7	0.2	0.1
🏀	0.1	0.2	0.7
🌮	0.3	0.1	0.6
📰	0.0	0.9	0.1

Topic–influence matrix $\mathbf{Y}$



Strengthens Controversial Topics

👨‍🎨	0.7	0.2	0.1	0.0
👱‍♂️	0.2	0.1	0.1	0.6
👳	0.0	0.8	0.1	0.1

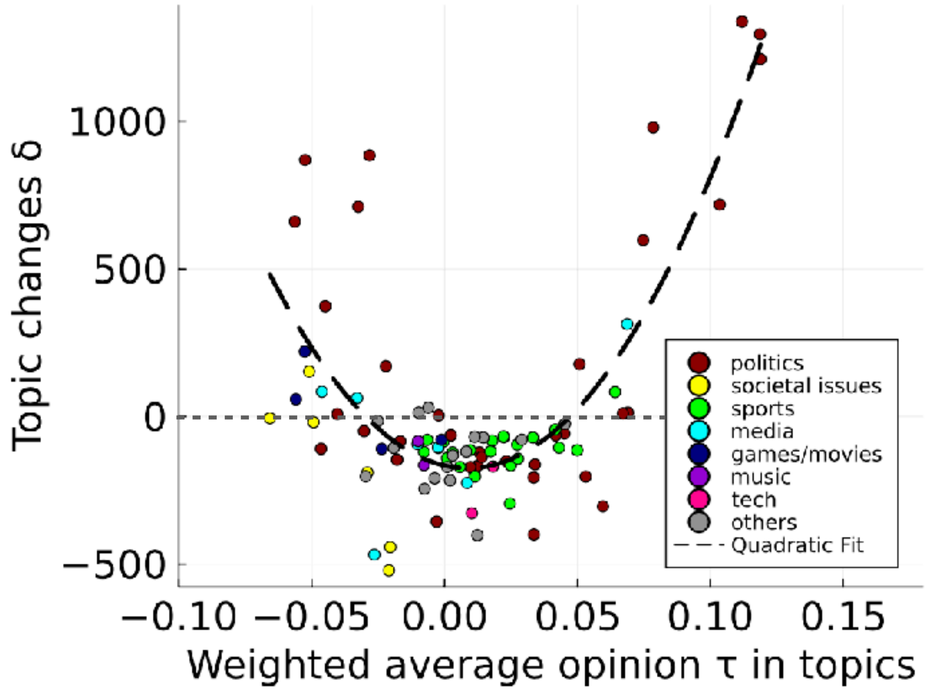
User–topic matrix $\mathbf{X}$

🖼️	0.7	0.2	0.1
🏀	0.1	0.2	0.7
🌮	0.3	0.1	0.6
📰	0.0	0.9	0.1

Topic–influence matrix $\mathbf{Y}$

- We run our algorithm which converges to optimal solution and inspect solution
- y-axis: How much more/less important did each topic become during optimization?
- x-axis: Average leaning of influencers for each topic

➡Results show that “controversial topics” get strengthened



Strengthens Controversial Topics

👨‍🎨	0.7	0.2	0.1	0.0
👱‍♂️	0.2	0.1	0.1	0.6
👷	0.0	0.8	0.1	0.1

User–topic matrix $\mathbf{X}$

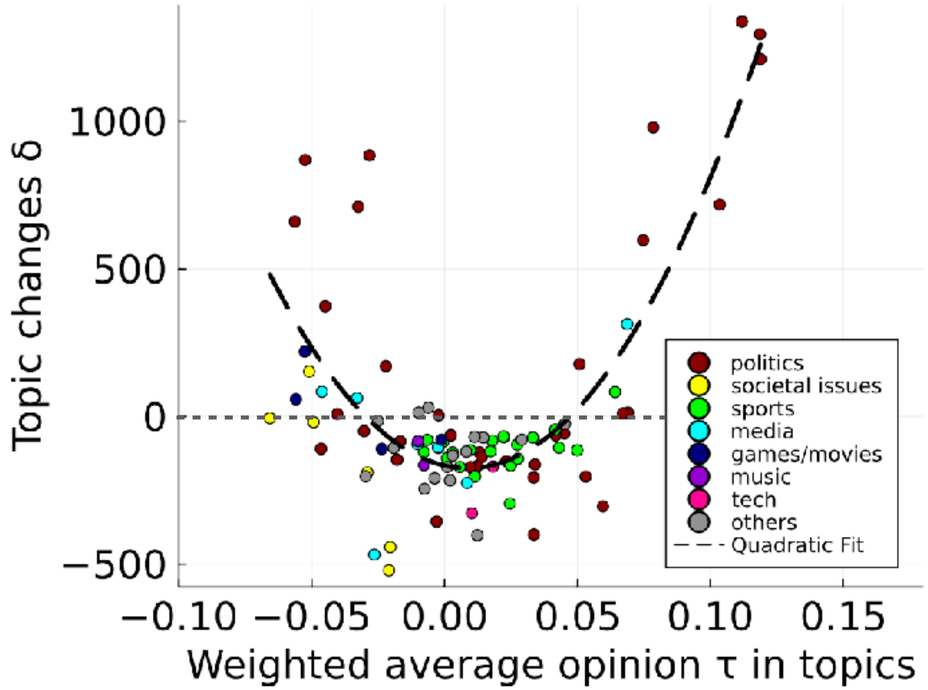
🖼️	0.7	0.2	0.1
🏀	0.1	0.2	0.7
🌮	0.3	0.1	0.6
📰	0.0	0.9	0.1

Topic–influence matrix $\mathbf{Y}$

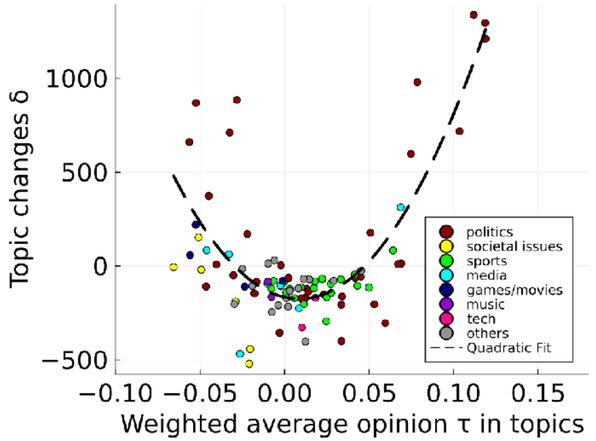
- We run our algorithm which converges to optimal solution and inspect solution
- y-axis: How much more/less important did each topic become during optimization?
- x-axis: Average leaning of influencers for each topic

➡Results show that “controversial topics” get strengthened

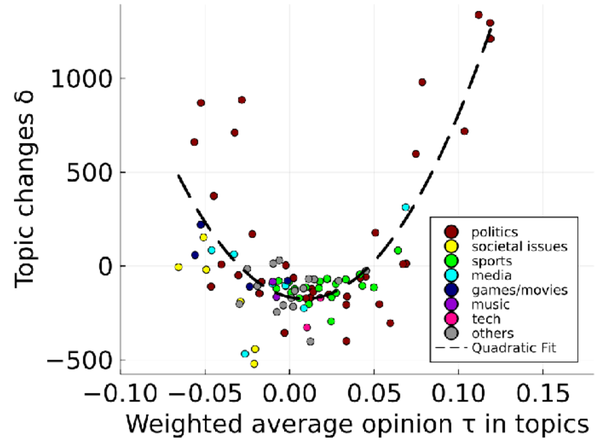
  - *Intuition:* To move node closer to average opinion, show them opposing views
  - Influenced by FJ-opinion dynamics
  - Pushes political topics (even though the algorithm does not know this)

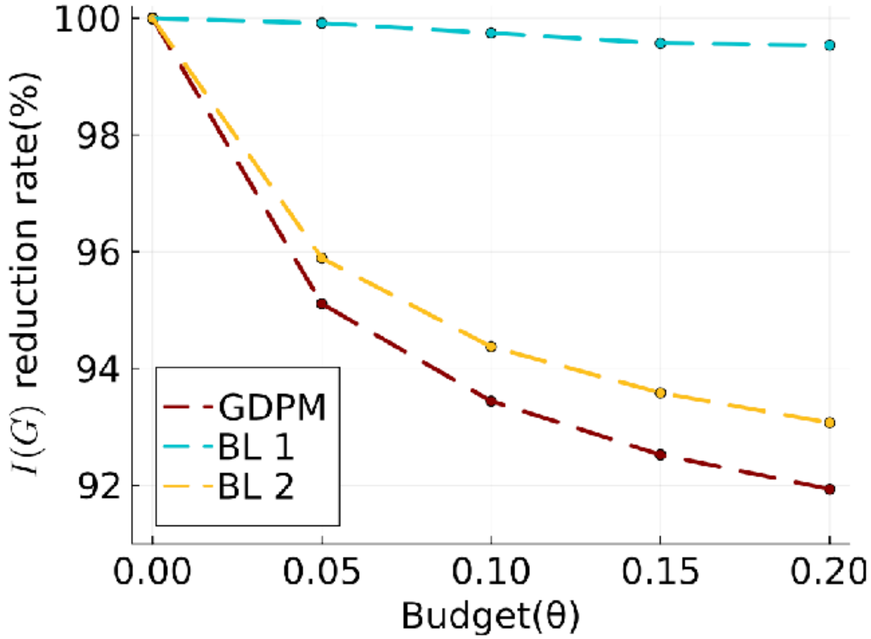


- Question:
  - What if we want to avoid behavior from the previous slide?



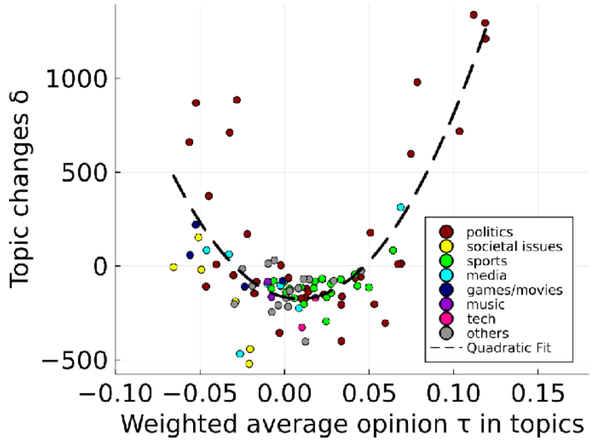
- Question:
  - What if we want to avoid behavior from the previous slide?

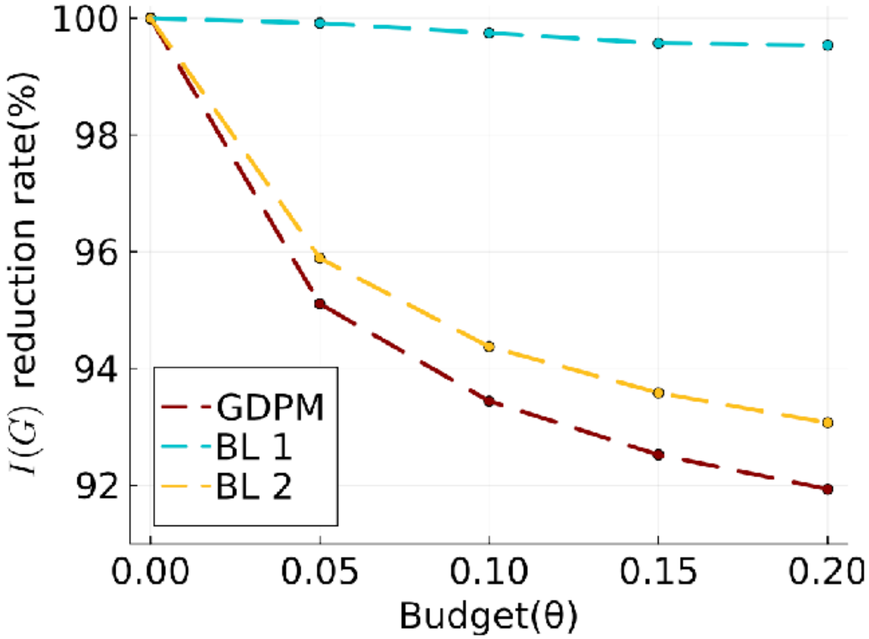




- Question:
  - What if we want to avoid behavior from the previous slide?

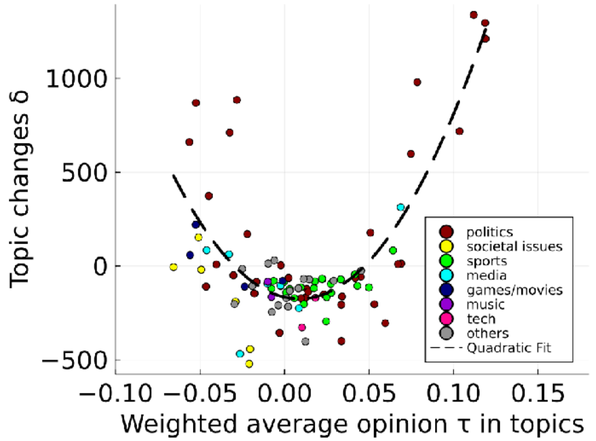
➡Strengthen topics with opinions close to 0 instead

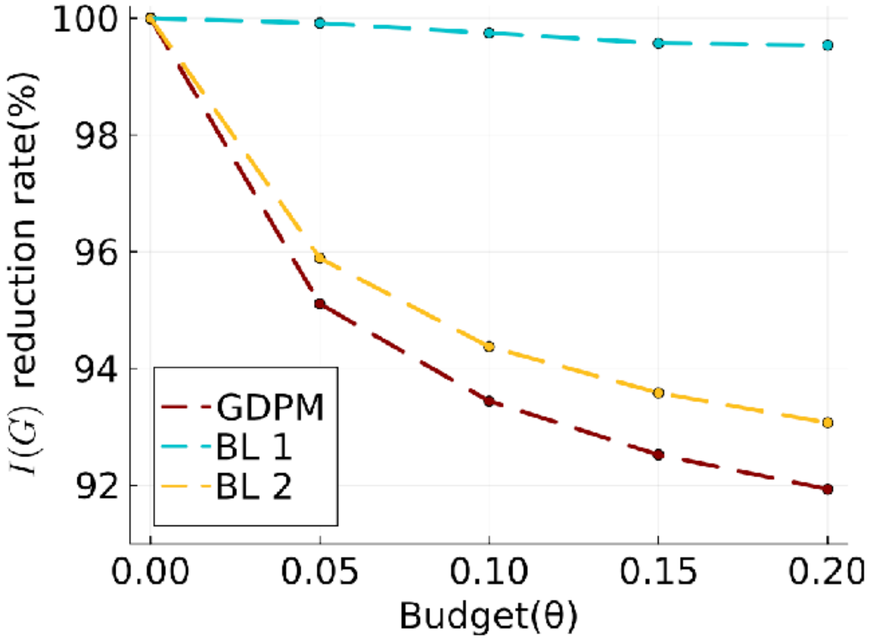




- Question:
  - What if we want to avoid behavior from the previous slide?

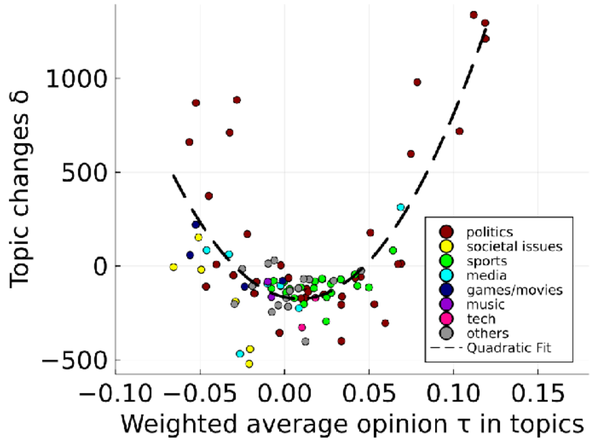
  ➡Strengthen topics with opinions close to 0 instead
- y-axis: How much polarization and disagreement were decreased

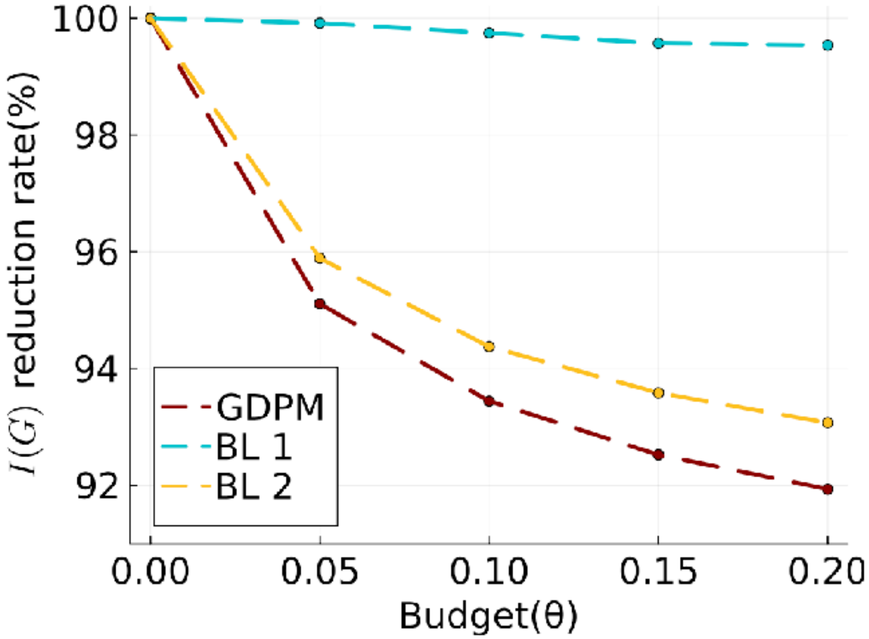


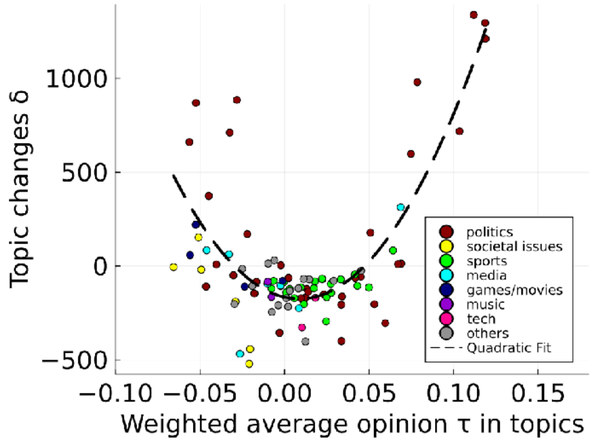


- Question:
  - What if we want to avoid behavior from the previous slide?

  ➡Strengthen topics with opinions close to 0 instead
- y-axis: How much polarization and disagreement were decreased
- x-axis: Budget for changing timelines

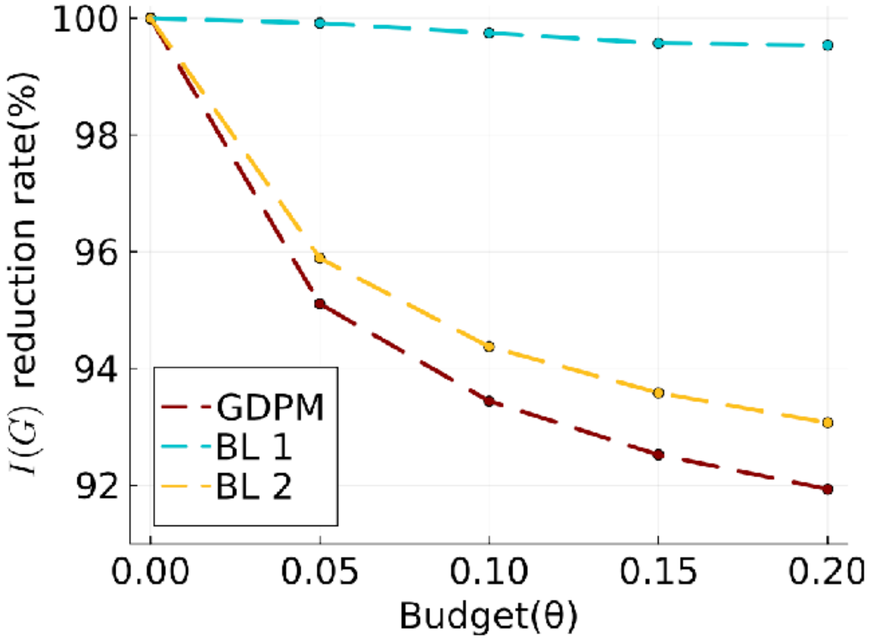




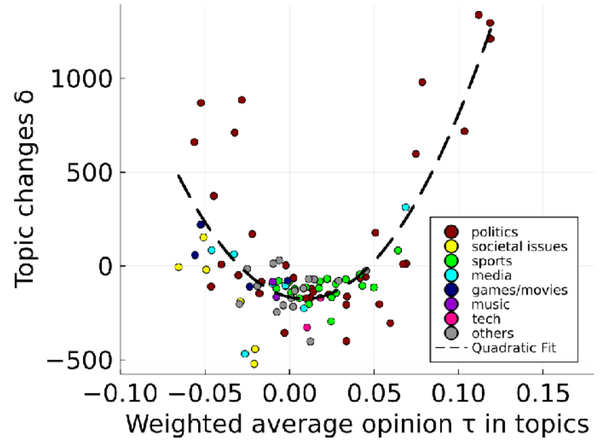


- Question:
  - What if we want to avoid behavior from the previous slide?

  ➡Strengthen topics with opinions close to 0 instead
- y-axis: How much polarization and disagreement were decreased
- x-axis: Budget for changing timelines
- GDPM: Our gradient-descent based algorithm, optimal solution
- Baseline 2 (BL 2):
  - For each user, increase topics with “opposing” viewpoints; mimics GDPM
- Baseline 1 (BL 1):
  - For each user, decrease controversial topics, increase non-controversial topics (τ_j close to 0)

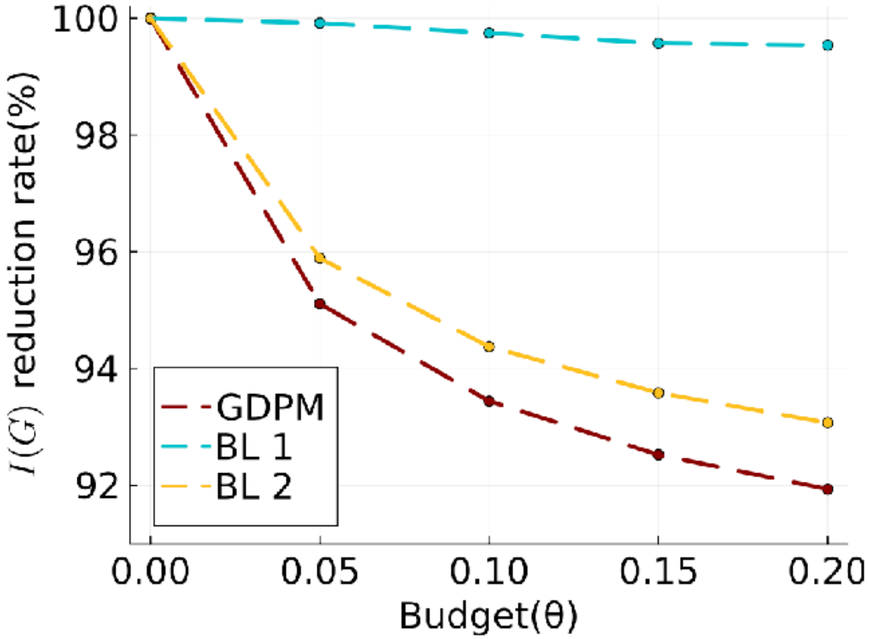


Strengthening Non-Controversial Topics is Much Less Effective



- Question:
  - What if we want to avoid behavior from the previous slide?

  ➡Strengthen topics with opinions close to 0 instead
- y-axis: How much polarization and disagreement were decreased
- x-axis: Budget for changing timelines
- GDPM: Our gradient-descent based algorithm, optimal solution
- Baseline 2 (BL 2):
  - For each user, increase topics with “opposing” viewpoints; mimics GDPM
- Baseline 1 (BL 1):
  - For each user, decrease controversial topics, increase non-controversial topics (τ_j close to 0)



Conclusion

Modeling the Impact of Timeline Algorithms on Opinion Dynamics

Tianyi Zhou, *Stefan Neumann*, Kiran Garimella, Aris Gionis — WebConf'24

@chow_tianyi @StefanResearch @gvrkiran @gionis

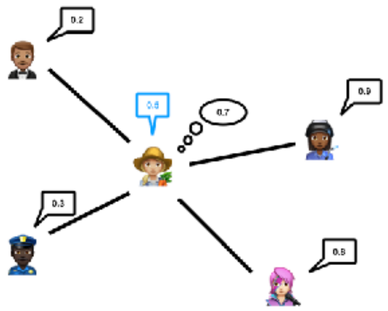
- Opinion formation models offer a **principled approach** to **analyze the impact of interventions on networks**
- By making small changes to timeline decompositions based on user interests,
  we effectively reduce polarization + disagreement
- New dataset with opinions and aggregate user interests
- **Future work:**
  - Find more expressive ways to combine opinion formation models and data from timeline algorithms
  - Exploit more advanced optimization techniques to allow for more complex interventions

🖼️	0.7	0.2	0.1	0.0
🏀	0.2	0.1	0.1	0.6
🌮	0.0	0.8	0.1	0.1

User–topic matrix $\mathbf{X}$

🖼️	0.7	0.2	0.1
🏀	0.1	0.2	0.7
🌮	0.3	0.1	0.6
📰	0.0	0.9	0.1

Topic–influence matrix $\mathbf{Y}$



Fixed graph

$+$

$\mathbf{XY} =$

	0.54	0.19	0.27
	0.18	0.61	0.21
	0.11	0.26	0.63

Recommender graph

Modeling the Impact of Timeline Algorithms on Opinion Dynamics

Tianyi Zhou, *Stefan Neumann*, Kiran Garimella, Aris Gionis — WebConf'24

@chow_tianyi @StefanResearch @gvrkiran @gionis

Thank you!

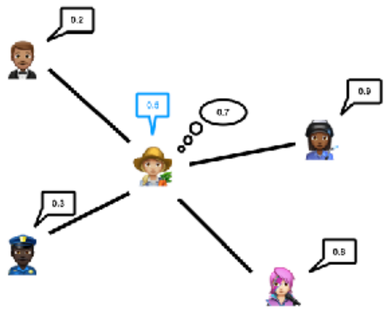
- Opinion formation models offer a **principled approach** to **analyze the impact of interventions on networks**
- By making small changes to timeline decompositions based on user interests,
  we effectively reduce polarization + disagreement
- New dataset with opinions and aggregate user interests
- **Future work:**
  - Find more expressive ways to combine opinion formation models and data from timeline algorithms
  - Exploit more advanced optimization techniques to allow for more complex interventions

🖼️	0.7	0.2	0.1	0.0
🏀	0.2	0.1	0.1	0.6
🌮	0.0	0.8	0.1	0.1

User–topic matrix $\mathbf{X}$

🖼️	0.7	0.2	0.1
🏀	0.1	0.2	0.7
🌮	0.3	0.1	0.6
📰	0.0	0.9	0.1

Topic–influence matrix $\mathbf{Y}$



Fixed graph

$+$ $\mathbf{XY} =$

🧑	0.54	0.19	0.27
👱	0.18	0.61	0.21
👳	0.11	0.26	0.63

Recommender graph

Strengthens Controversial Topics

🏞️	0.7	0.2	0.1	0.0
🏀	0.2	0.1	0.1	0.6
🌮	0.0	0.8	0.1	0.1

User–topic matrix $\mathbf{X}$

🏞️	0.7	0.2	0.1
🏀	0.1	0.2	0.7
🌮	0.3	0.1	0.6
📰	0.0	0.9	0.1

Topic–influence matrix $\mathbf{Y}$

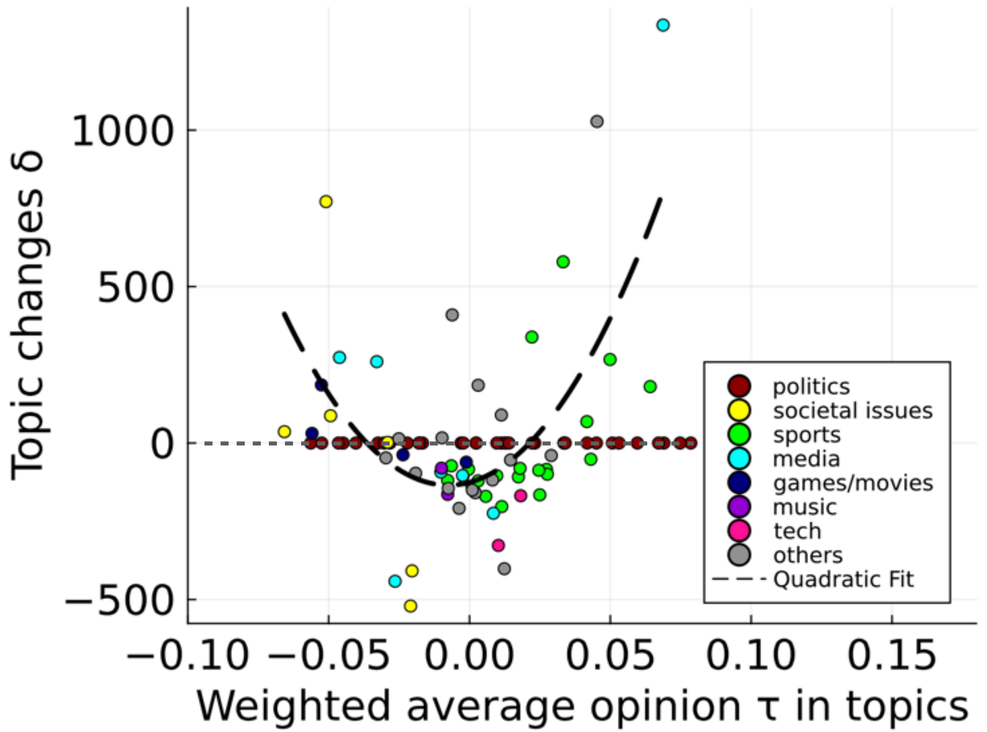
- We run our algorithm which converges to optimal solution and inspect solution

  ➡**Algorithm is not allowed to change importance of political topics**

- y-axis: How much more/less important did each topic become during optimization?
- x-axis: Average leaning of influencers for each topic

➡Results show that “controversial topics” get strengthened

- *Intuition:* To move node closer to average opinion, show them opposing views
- Influenced by FJ-opinion dynamics
- Pushes political topics (even though the algorithm does not know this)



Other Examples of Interventions

The Impact of Viral Content

- Tu, Neumann (WebConf'22):
  - Model for simulating how **viral content** in OSNs impacts node opinions
    (combines the independent cascade model and the FJ model)

The Impact of Viral Content

- Tu, Neumann (WebConf'22):
  - Model for simulating how **viral content** in OSNs impacts node opinions
    (combines the independent cascade model and the FJ model)
  - Non-controversial content:
    if node u reads it, increase innate opinion s_u by +0.1

The Impact of Viral Content

- Tu, Neumann (WebConf'22):
  - Model for simulating how **viral content** in OSNs impacts node opinions
    (combines the independent cascade model and the FJ model)
  - Non-controversial content:
    if node u reads it, increase innate opinion s_u by +0.1
  - Controversial content:
    if node reads it, if $s_u > 0.5$ increase s_u by $+0.1$ and otherwise decrease s_u by -0.1

The Impact of Viral Content

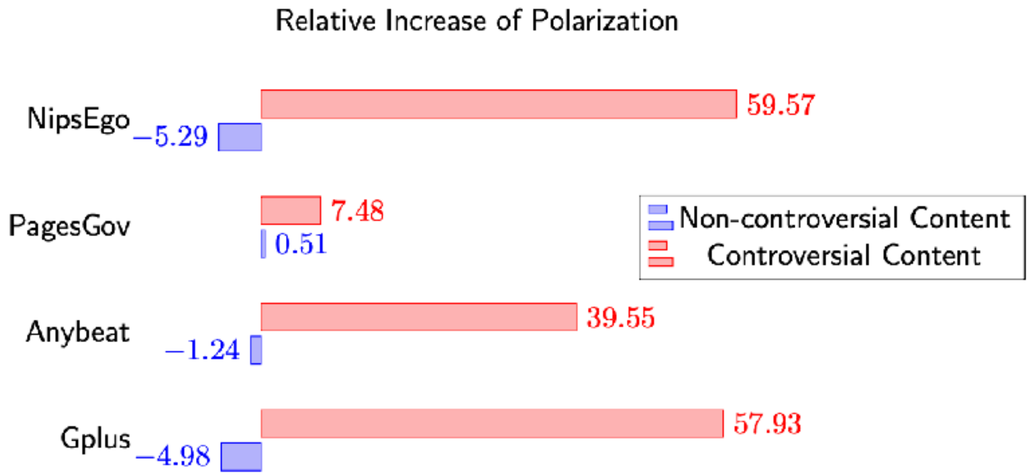
- Tu, Neumann (WebConf'22):
  - Model for simulating how **viral content** in OSNs impacts node opinions (combines the independent cascade model and the FJ model)
  - Non-controversial content:
    if node u reads it, increase innate opinion s_u by +0.1
  - Controversial content:
    if node reads it, if $s_u > 0.5$ increase s_u by $+0.1$ and otherwise decrease s_u by -0.1

    ➡Models backfire effect if people dislike the content

The Impact of Viral Content

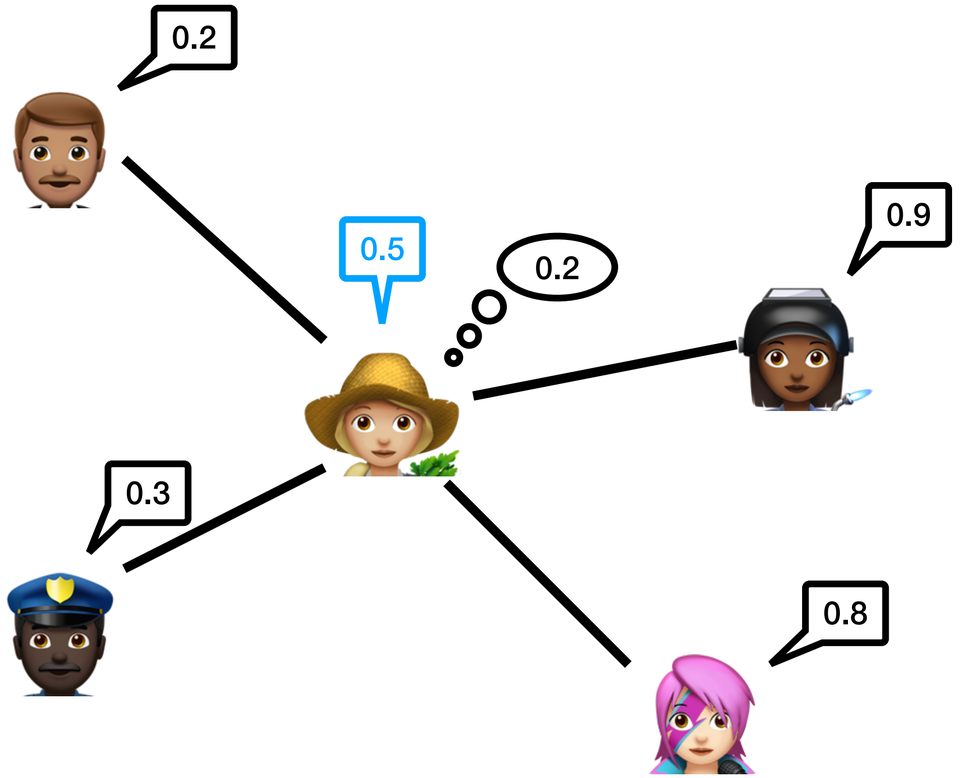
- Tu, Neumann (WebConf'22):
  - Model for simulating how **viral content** in OSNs impacts node opinions
    (combines the independent cascade model and the FJ model)
  - Non-controversial content:
    if node u reads it, increase innate opinion s_u by +0.1
  - Controversial content:
    if node reads it, if $s_u > 0.5$ increase s_u by $+0.1$ and otherwise decrease s_u by -0.1

    ➡Models backfire effect if people dislike the content



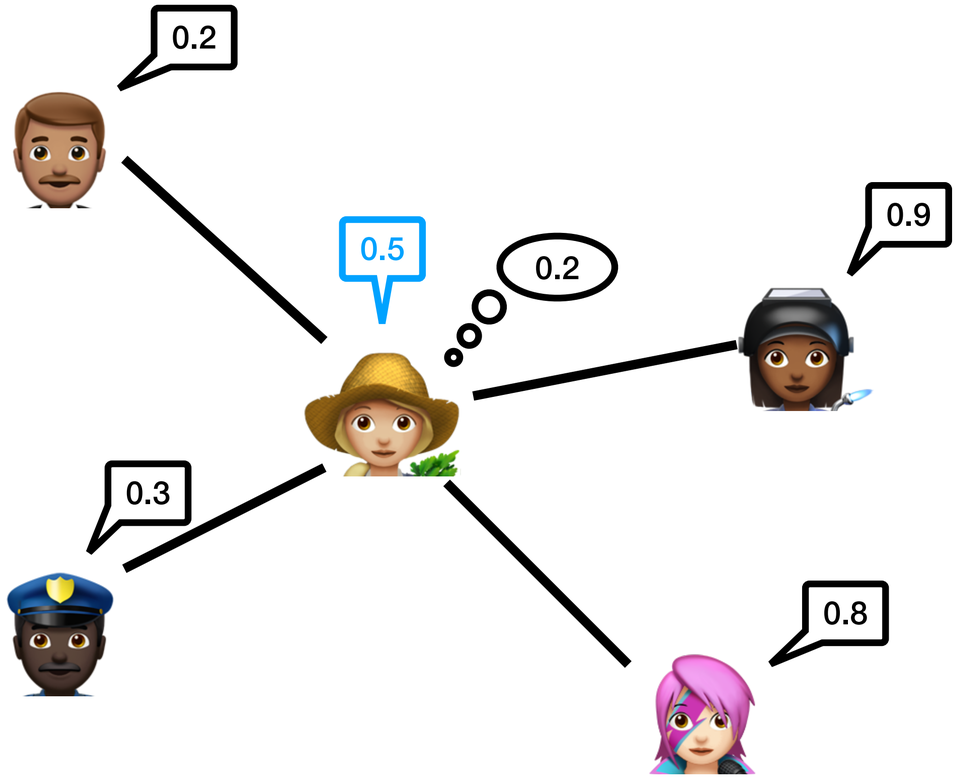
Suppose the 0.5% highest-degree nodes start sharing a content

Adversaries Who Aim to Radicalize



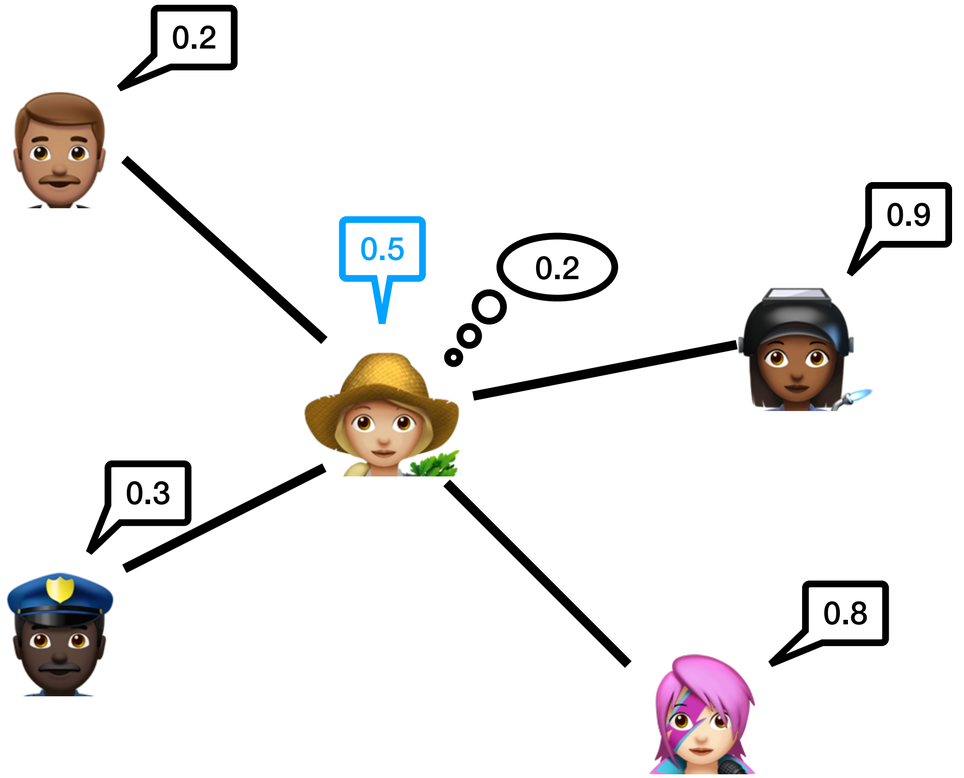
Adversaries Who Aim to Radicalize

- Gaitonde, Kleinberg, Tardos (EC'20) and Chen, Racz (TNSE'21):
  - What is the impact on the disagreement,
    when adversaries can change k innate opinions?



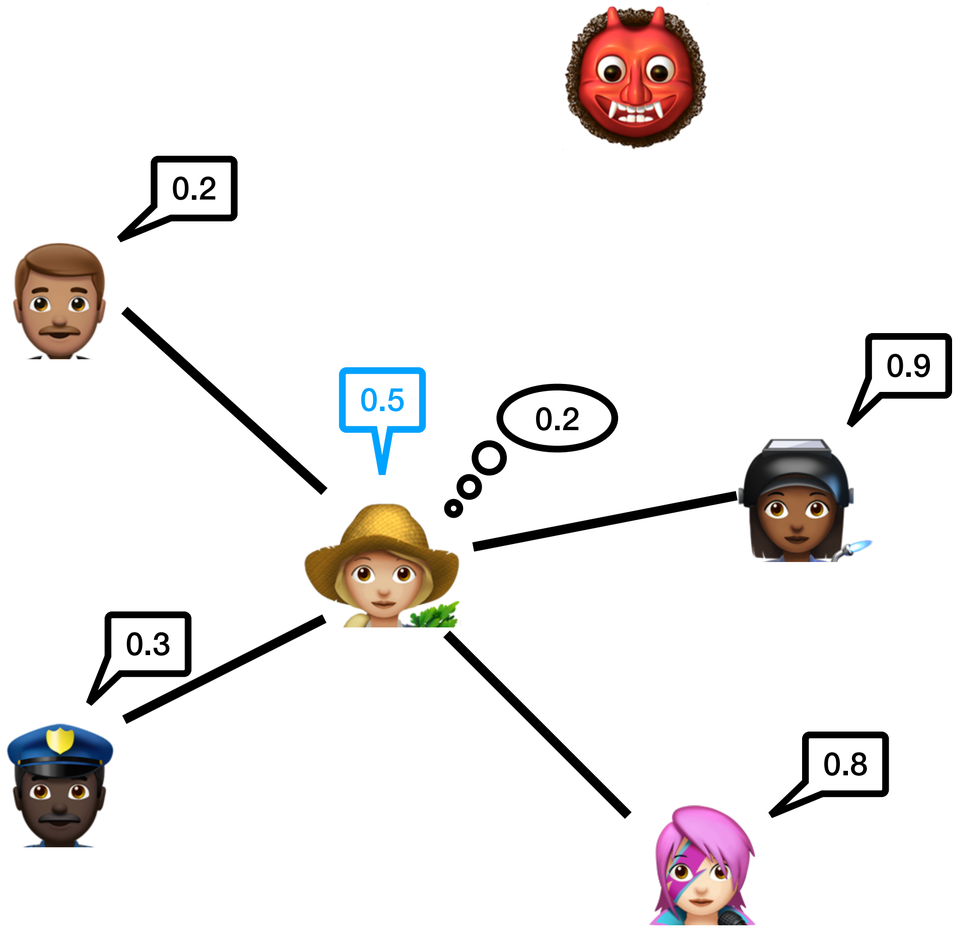
Adversaries Who Aim to Radicalize

- Gaitonde, Kleinberg, Tardos (EC'20) and Chen, Racz (TNSE'21):
  - What is the impact on the disagreement, when adversaries can change k innate opinions?
  - Motivated by real-world events (e.g., Russia meddling with the US election in 2016)



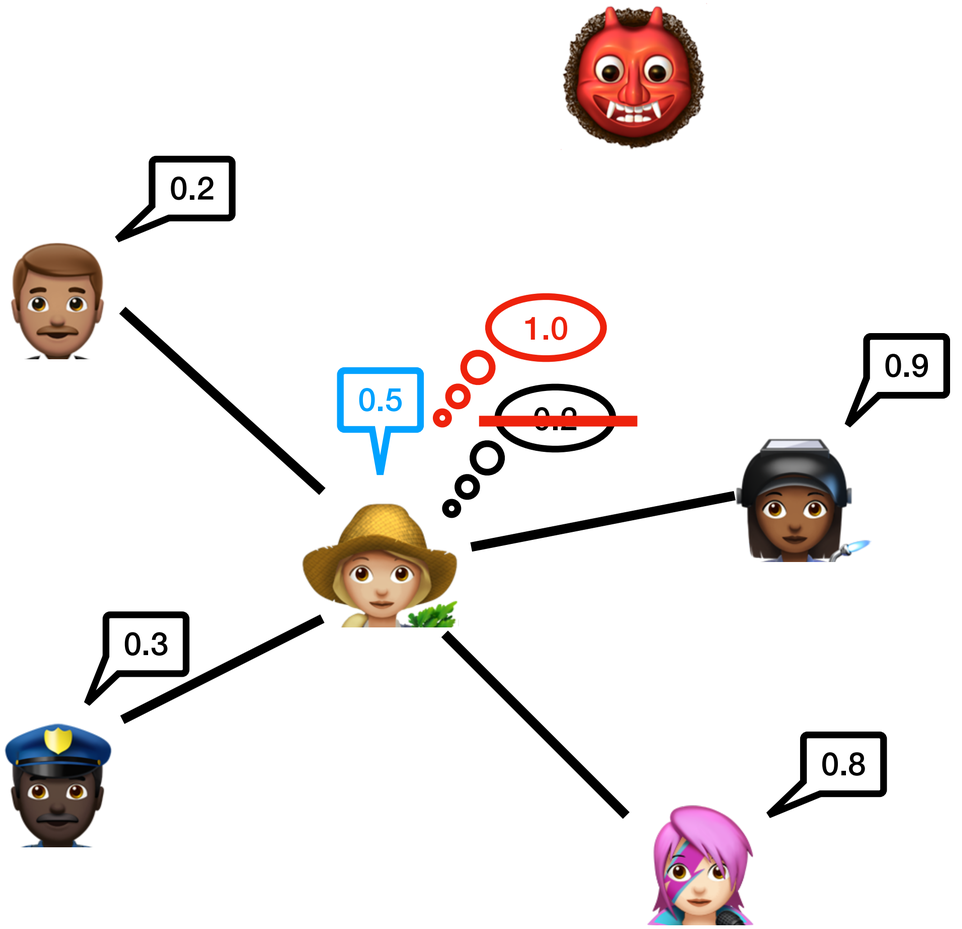
Adversaries Who Aim to Radicalize

- Gaitonde, Kleinberg, Tardos (EC'20) and Chen, Racz (TNSE'21):
  - What is the impact on the disagreement, when adversaries can change k innate opinions?
  - Motivated by real-world events (e.g., Russia meddling with the US election in 2016)



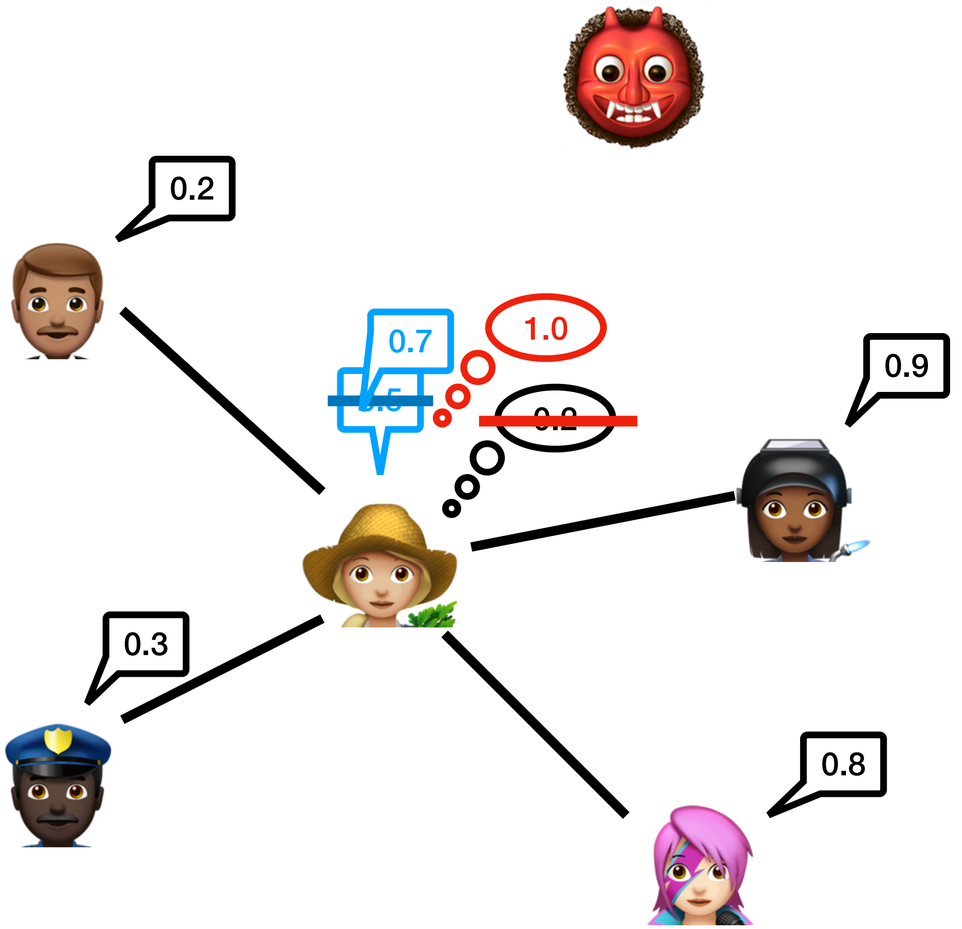
Adversaries Who Aim to Radicalize

- Gaitonde, Kleinberg, Tardos (EC'20) and Chen, Racz (TNSE'21):
  - What is the impact on the disagreement, when adversaries can change k innate opinions?
  - Motivated by real-world events (e.g., Russia meddling with the US election in 2016)



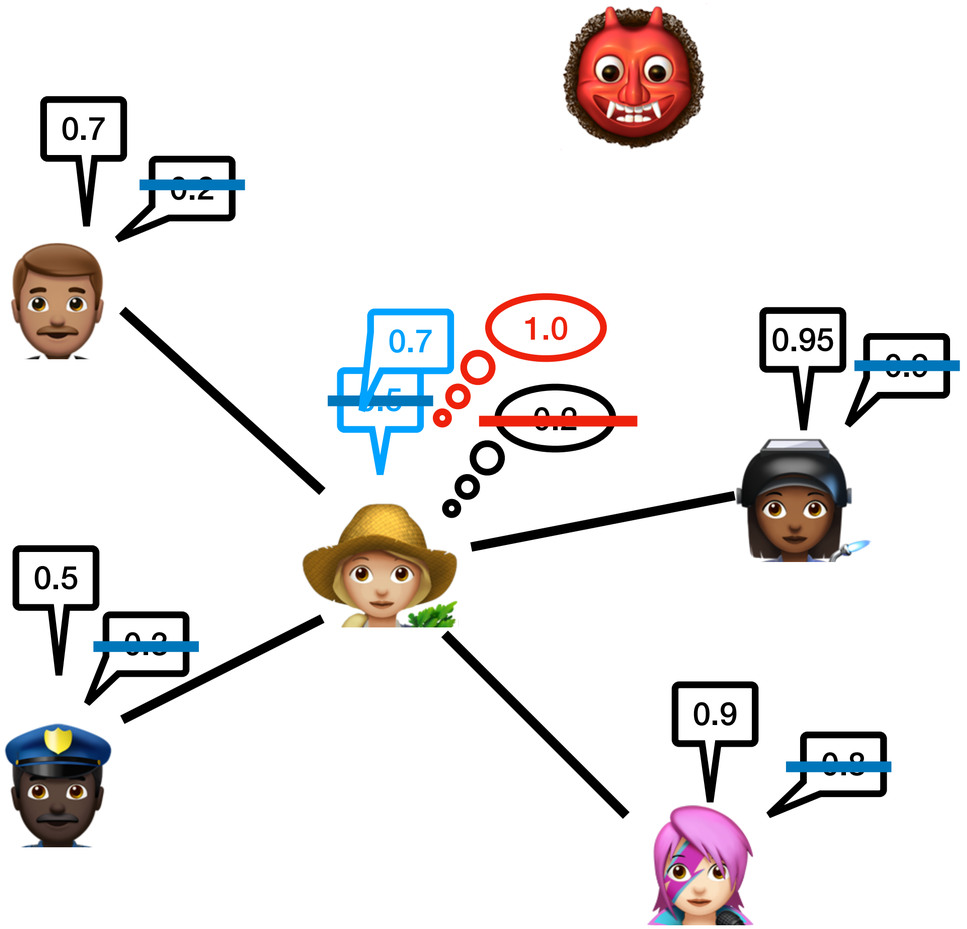
Adversaries Who Aim to Radicalize

- Gaitonde, Kleinberg, Tardos (EC'20) and Chen, Racz (TNSE'21):
  - What is the impact on the disagreement, when adversaries can change k innate opinions?
  - Motivated by real-world events (e.g., Russia meddling with the US election in 2016)



Adversaries Who Aim to Radicalize

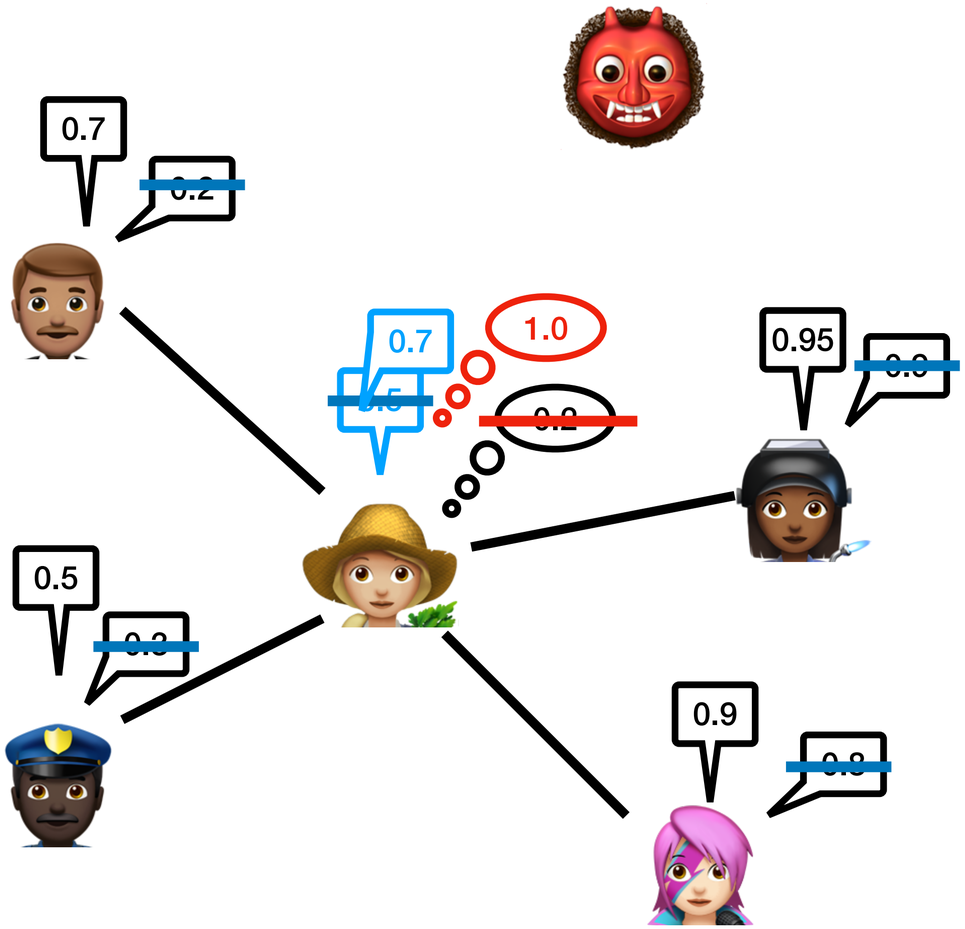
- Gaitonde, Kleinberg, Tardos (EC'20) and Chen, Racz (TNSE'21):
  - What is the impact on the disagreement, when adversaries can change k innate opinions?
  - Motivated by real-world events (e.g., Russia meddling with the US election in 2016)



Adversaries Who Aim to Radicalize

- Gaitonde, Kleinberg, Tardos (EC'20) and Chen, Racz (TNSE'21):
  - What is the impact on the disagreement, when adversaries can change k innate opinions?
  - Motivated by real-world events (e.g., Russia meddling with the US election in 2016)

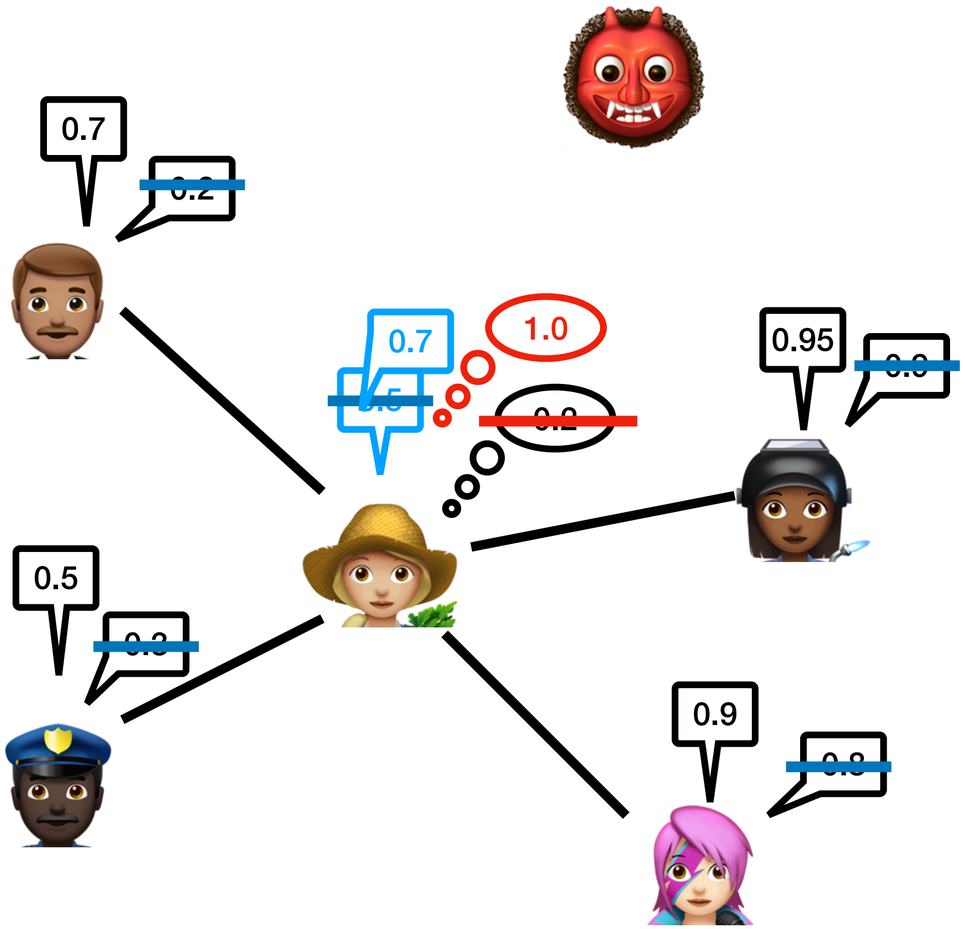
➡They give bounds showing disagreement increases by $\leq 8d_{\max}k$



Adversaries Who Aim to Radicalize

- Gaitonde, Kleinberg, Tardos (EC'20) and Chen, Racz (TNSE'21):
  - What is the impact on the disagreement, when adversaries can change k innate opinions?
  - Motivated by real-world events (e.g., Russia meddling with the US election in 2016)

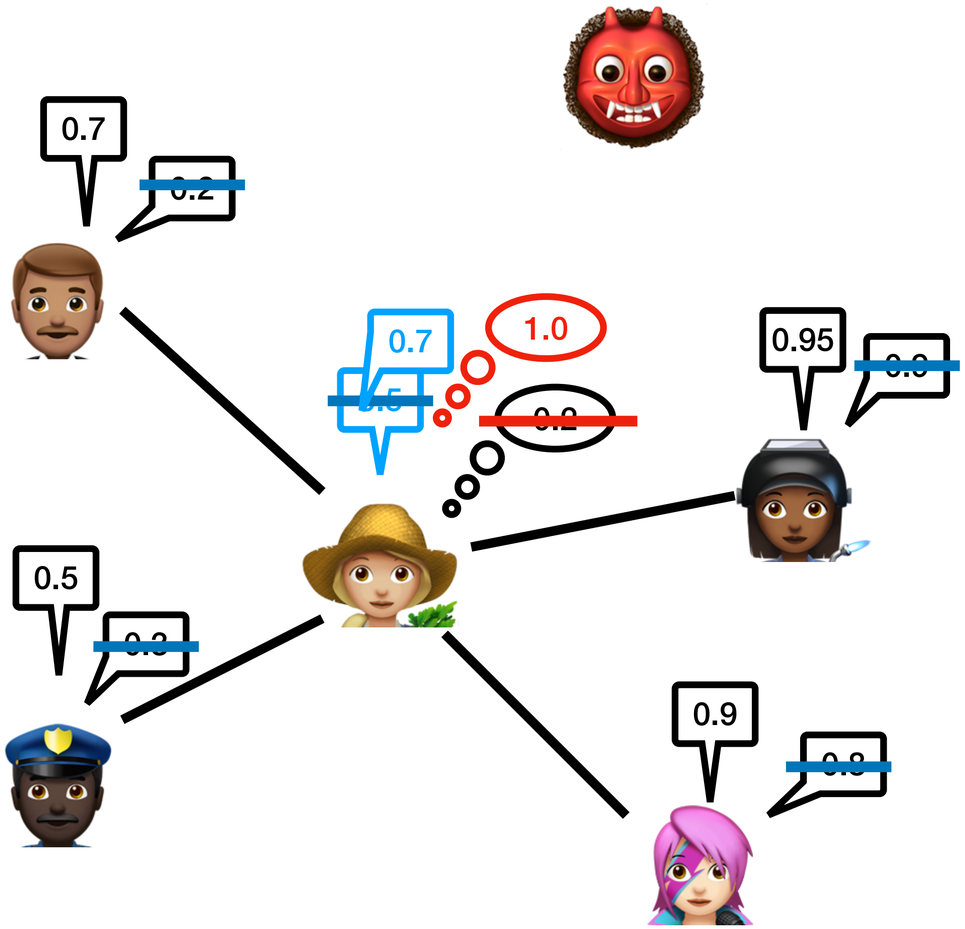
  ➡They give bounds showing disagreement increases by $\leq 8d_{\max}k$
- Tu, Neumann, Gionis (KDD'23):



Adversaries Who Aim to Radicalize

- Gaitonde, Kleinberg, Tardos (EC'20) and Chen, Racz (TNSE'21):
  - What is the impact on the disagreement, when adversaries can change k innate opinions?
  - Motivated by real-world events (e.g., Russia meddling with the US election in 2016)

  ➡They give bounds showing disagreement increases by $\leq 8d_{\max}k$
- Tu, Neumann, Gionis (KDD'23):
  - Adversary is **almost as powerful** when only knowing the graph *but not the opinions*

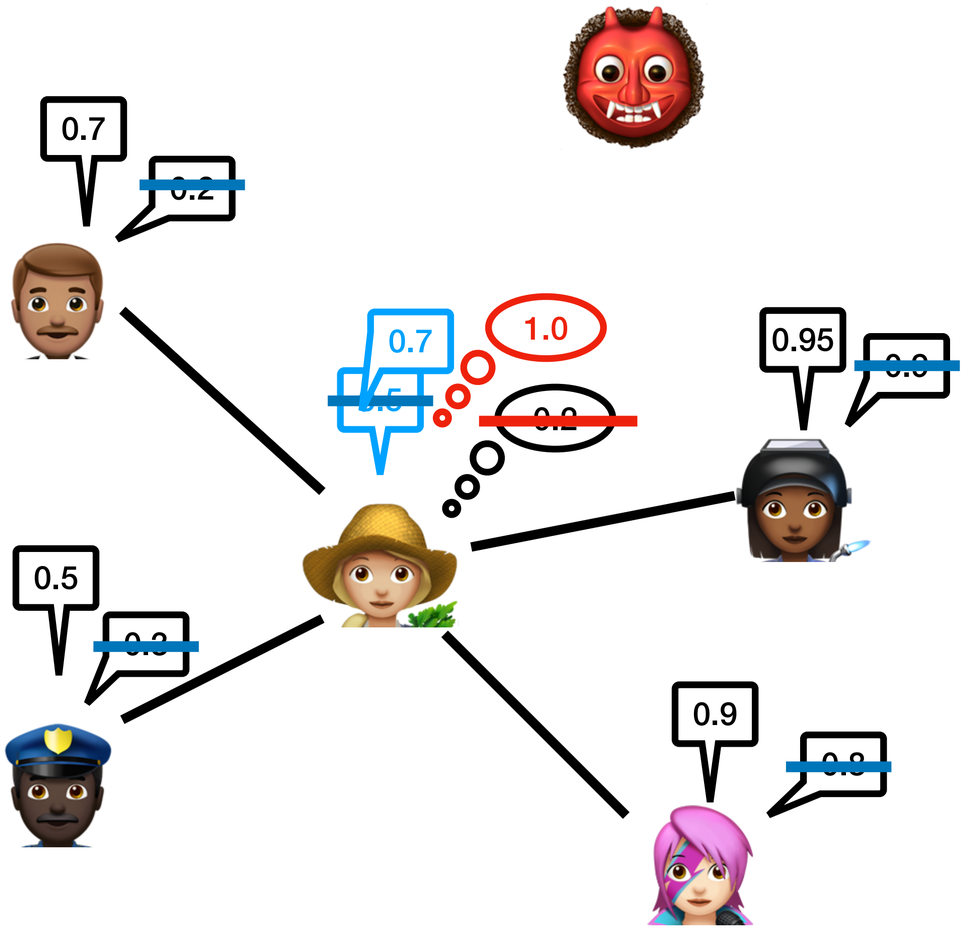


Adversaries Who Aim to Radicalize

- Gaitonde, Kleinberg, Tardos (EC'20) and Chen, Racz (TNSE'21):
  - What is the impact on the disagreement, when adversaries can change k innate opinions?
  - Motivated by real-world events (e.g., Russia meddling with the US election in 2016)

  ➡They give bounds showing disagreement increases by $\leq 8d_{\max}k$
- Tu, Neumann, Gionis (KDD'23):
  - Adversary is **almost as powerful** when only knowing the graph *but not the opinions*

  ➡Can obtain a $O(1)$-approximation of maximum possible disagreement (under some assumptions)



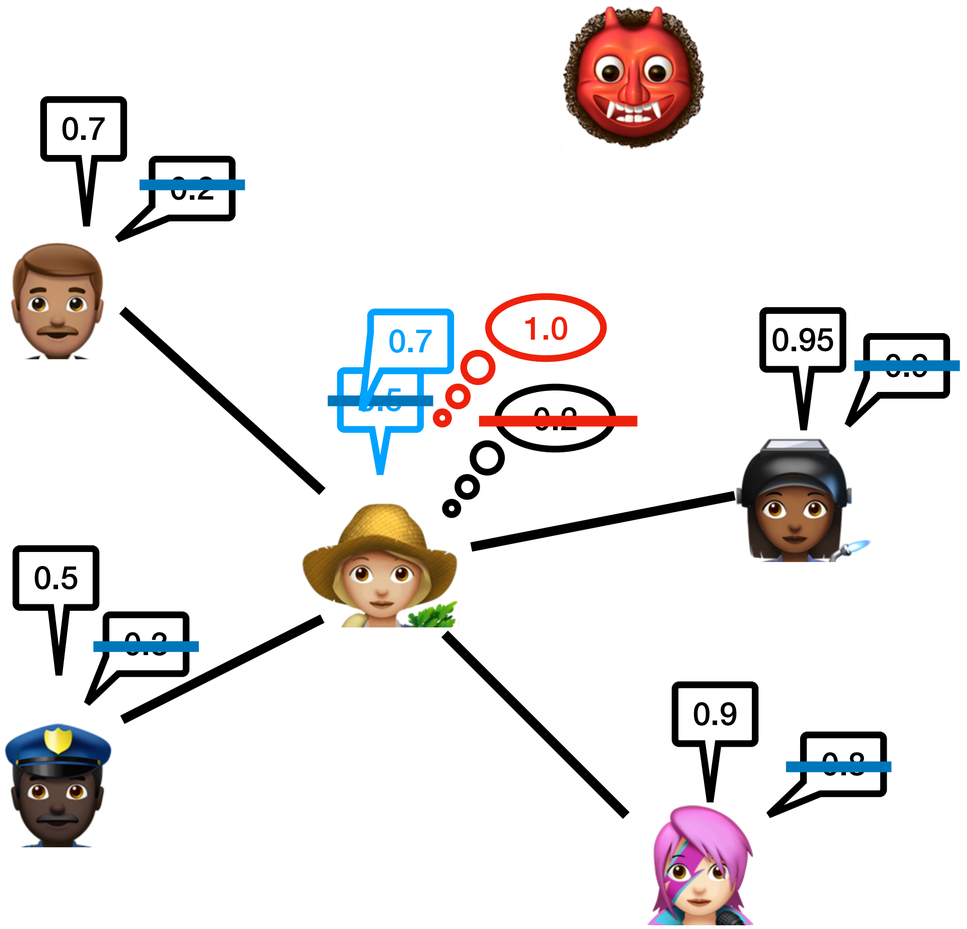
Adversaries Who Aim to Radicalize

- Gaitonde, Kleinberg, Tardos (EC'20) and Chen, Racz (TNSE'21):
  - What is the impact on the disagreement, when adversaries can change k innate opinions?
  - Motivated by real-world events (e.g., Russia meddling with the US election in 2016)

  ➡They give bounds showing disagreement increases by $\leq 8d_{\max}k$
- Tu, Neumann, Gionis (KDD'23):
  - Adversary is **almost as powerful** when only knowing the graph *but not the opinions*

  ➡Can obtain a $O(1)$-approximation of maximum possible disagreement (under some assumptions)

  ➡Connection to solving MaxCut with cardinality constraint in graphs with positive *and negative edge weights*



Further Examples of Interventions

Further Examples of Interventions

- Musco, Musco, Tsourakakis (WebConf'18):

Further Examples of Interventions

- Musco, Musco, Tsourakakis (WebConf'18):
  - **Changing innate opinions** to minimize the disagreement and polarization

Further Examples of Interventions

- Musco, Musco, Tsourakakis (WebConf'18):
  - **Changing innate opinions** to minimize the disagreement and polarization
  - Making (relatively large) **changes to network topology**

Further Examples of Interventions

- Musco, Musco, Tsourakakis (WebConf'18):
  - **Changing innate opinions** to minimize the disagreement and polarization
  - Making (relatively large) **changes to network topology**
- Chitra, Musco (WSDM'20):

Further Examples of Interventions

- Musco, Musco, Tsourakakis (WebConf'18):
  - **Changing innate opinions** to minimize the disagreement and polarization
  - Making (relatively large) **changes to network topology**
- Chitra, Musco (WSDM'20):
  - If an OSN provider **repeatedly changes the network structure to reduce disagreement**, this will increase the polarization

Further Examples of Interventions

- Musco, Musco, Tsourakakis (WebConf'18):
  - **Changing innate opinions** to minimize the disagreement and polarization
  - Making (relatively large) **changes to network topology**
- Chitra, Musco (WSDM'20):
  - If an OSN provider **repeatedly changes the network structure to reduce disagreement**, this will increase the polarization
- Bhalla, Lechowicz, Musco (WSDM'23):

Further Examples of Interventions

- Musco, Musco, Tsourakakis (WebConf'18):
  - **Changing innate opinions** to minimize the disagreement and polarization
  - Making (relatively large) **changes to network topology**
- Chitra, Musco (WSDM'20):
  - If an OSN provider **repeatedly changes the network structure to reduce disagreement**, this will increase the polarization
- Bhalla, Lechowicz, Musco (WSDM'23):
  - Updating graph based on **confirmation bias and friend-of-friend recommendations** increases polarization over time