Behavioral Data Science

UNIVERSITY OF TECHNOLOGY SYDNEY

Opinion manipulation via social media:

a case study of SocialBots and Russian trolls during the 2016 US elections

Marian-Andrei Rizoiu

The research group

1 research associate, 4 PhD students, 1 research assistant, 1 lecturer



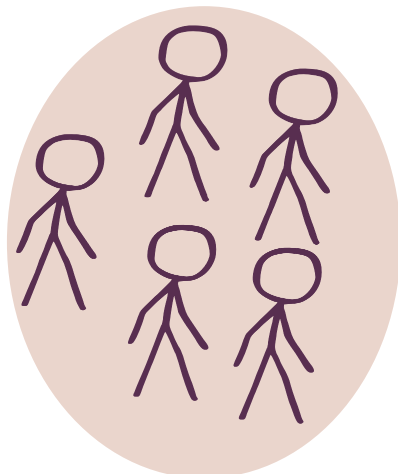




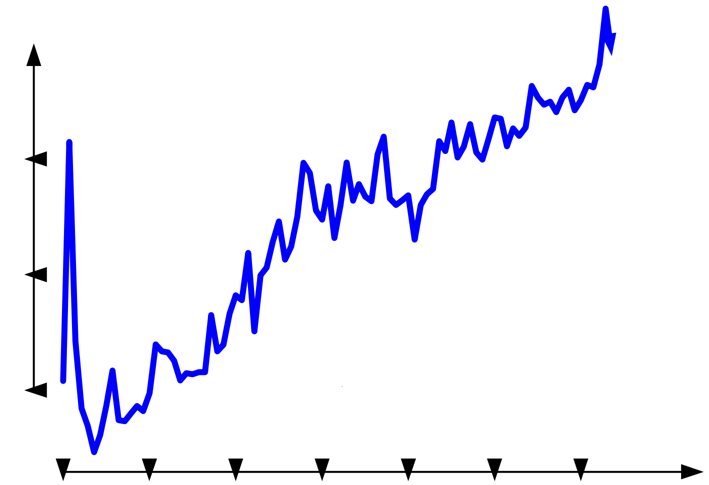
Research objectives

Behavioral Data Science

1.

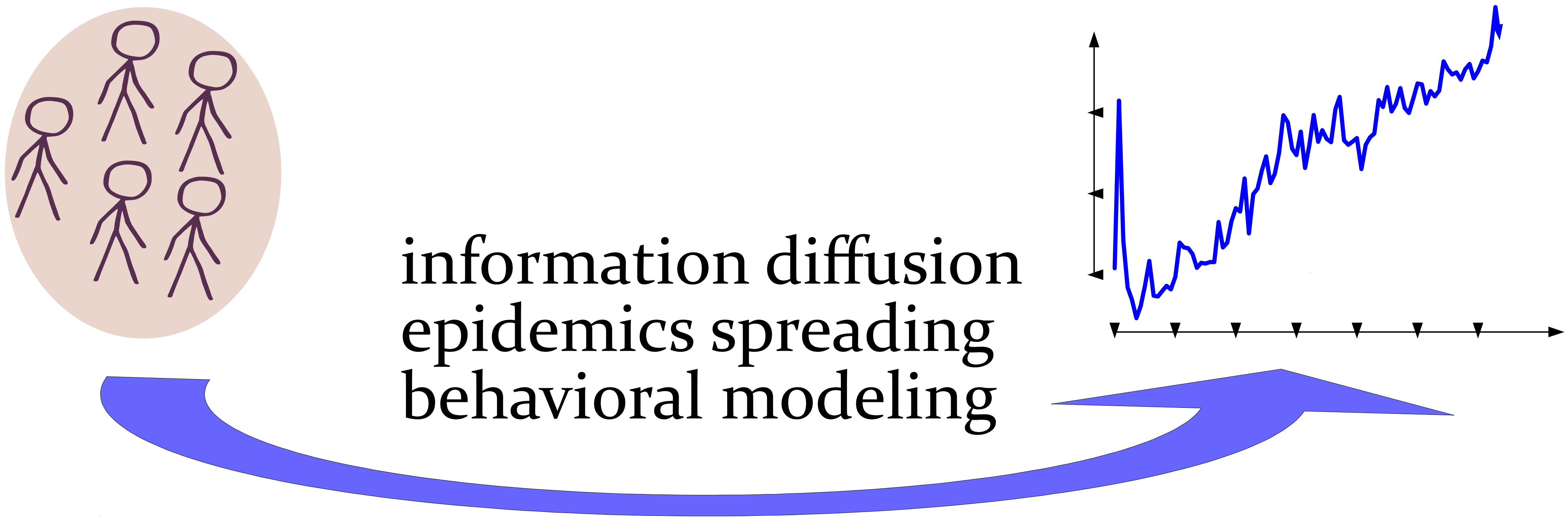


information diffusion
epidemics spreading
behavioral modeling

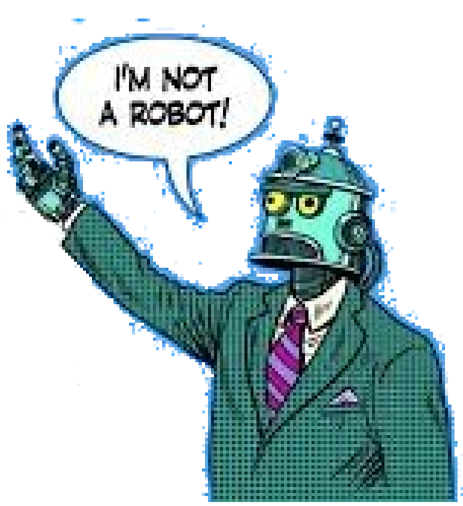


Research objectives

1.



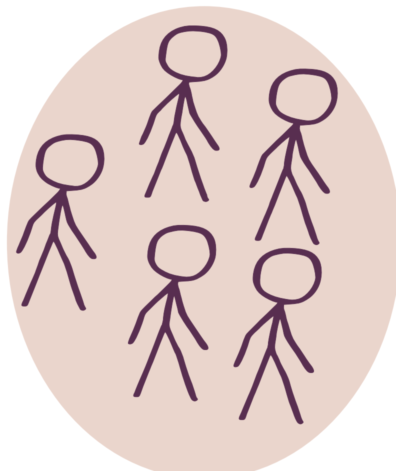
2.



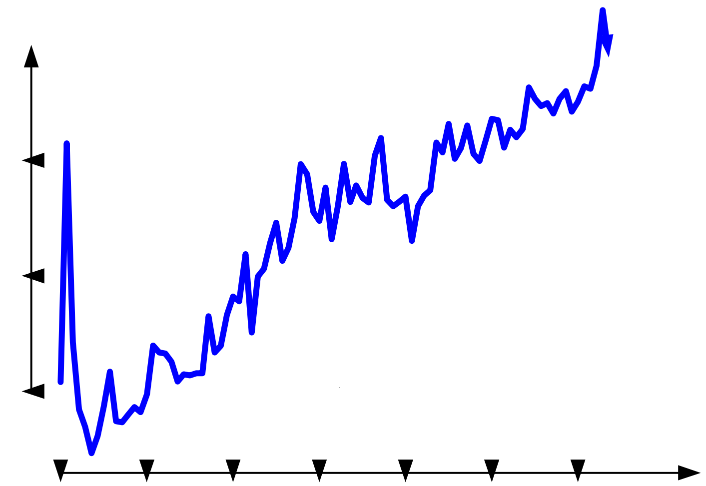
[Rizoiu et al ICWSM'18] [Kim et al Journ.Comp.SocSci'19]

Research objectives

1.



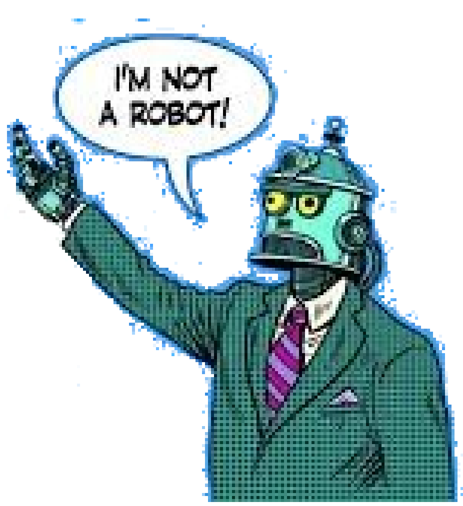
information diffusion
epidemics spreading
behavioral modeling



3.

[Rizoiu et al WWW'20]

2.



[Rizoiu et al ICWSM'18] [Kim et al Journ.Comp.SocSci'19]

Behavioral Data Science

The Sydney Morning Herald

NATIONAL WORLD ELECTION

Twitter bots more influential than people in US election: research

By Sherryn Groch
September 15, 2018 — 12.00am

They were the 90 minutes of television that set America on fire. As Donald Trump and Hillary Clinton stepped up to the podium for the first presidential debate of the 2016 election, the battle was already raging on Twitter.

But not all of those users joining in the discussion were human.

2 View all comments

Talking points

- ANU devised algorithms to map the influence of Twitter bots on the 2016 election .
- Bots were 2.5 times more influential than people during the first debate
- Pro-Republican bots were twice as influential and more politically engaged.
- 'Highly influential' human users were more likely to be pro-Democrat.

Then Democratic presidential candidate Hillary Clinton and President Donald Trump, who faced off in the first presidential debate in September 2016. MATT ROURKE

#DebateNight: The Role and Influence of Socialbots on Twitter During the 1st 2016 U.S. Presidential Debate

Marian-Andrei Rizoiu[1,2] and Timothy Graham[1] and Rui Zhang[1,2]
and Yifei Zhang[1,2] and Robert Ackland[1] and Lexing Xie[1,2]
[1]The Australian National University, [2]Data61 CSIRO
Canberra, Australia.

U.S. presidential election and manipulated public opinion at scale. Concerns were heightened with the discovery that an influential conservative commentator (@Jenn_Abrams, 70,000 followers) and a user claiming to belong to the Tennessee Republican Party (@TEN_GOP, 136,000 followers) – both retweeted by high-profile political figures and celebrities – were in fact Russian-controlled bots operated by the Internet Research Agency in St. Petersburg (Collins and Cox 2017; Timberg, Dwoskin, and Entous 2017).

There are several challenges that arise when conducting large-scale empirical analysis of political influence of bots on Twitter. The first challenge concerns estimating user influence from retweet diffusions, where the retweet relations are unobserved – the Twitter API assigns every retweet to the original tweet in the diffusion. Current state-of-the-art influence estimation methods such as ConTinEst (Du et al. 2013) operate on a static snapshot of the diffusion graph, which needs to be inferred from retweet diffusions using approaches like NetRate (Rodriguez, Balduzzi, and Schölkopf 2011). This workflow suffers from two major drawbacks: first, the algorithms for uncovering the diffusion graph do not scale to millions of users like in our application; second, operating on the diffusion graph estimates the "potential of being influential", but it loses information about user activity – e.g. a less well connected user can still be influential if they tweet a lot. The question is **how to estimate at scale the influence of millions of users from diffusion in which the retweet relation is not observed?** The second challenge lies in determining at scale whether a user is a bot and also her political behavior, as manually labeling millions of users is infeasible. The question is therefore **how to leverage automated bot detection approaches to measure the botness of millions of users? and how to analyze political behavior (partisanship and engagement) at scale?**

#DebateNight
Role of Twitter Socialbots During US Presidential Debate

[Rizoiu, ICWSM'18]

Two influencers: the 2016 U.S. Presidential elections

Behavioral Data Science

60k followers

136k followers

Common traits:

- Pro-republican;
- Highly influential, highly followed and retweeted;
- Opinion leaders;
- …

Two influencers: the 2016 U.S. Presidential elections

Behavioral Data Science

60k followers

136k followers

Common traits:

- Pro-republican;
- Highly influential, highly followed and retweeted;
- Opinion leaders;
- …

Russian-controlled bots operated by the Internet Research Agency in St. Petersburg

[Forbes, The Guardian, CNN + 50 more]

The political influence of socialbots

SocialBots:

"Software processes that are programmed to appear to be human-generated within the context of social networking sites such as Facebook and Twitter"

(Gehl and Bakardjieva 2016, p.2)

Immediate and long term research questions:

- are socialbots influential in the political discourse?
- did they have political partisanship?
- *(long term)* were they instrumental for the results of the elections?

Behavioral Data Science

#DebateNight dataset

- First U.S. Presidential Debate (26 sept 2016, 8.45pm to 10.45pm EDT)
- Twitter Firehose

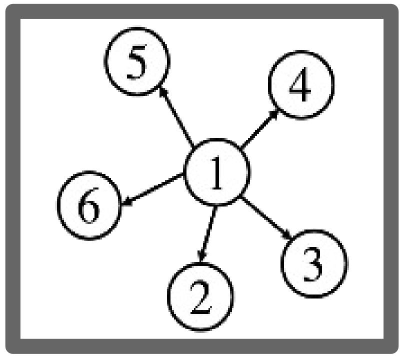
Dataset stats:

- length: **90 minutes**
- #tweets: **6.5M**
- #users: **1.45M**

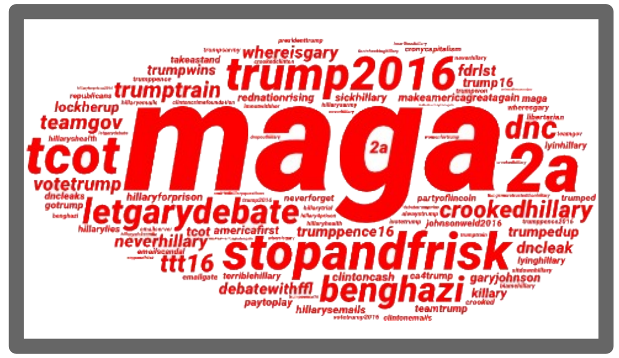
Hashtags:

#DebateNight
#Debates2016
#election2016
#HillaryClinton
#Debates,
#Hillary2016
#DonaldTrump
#Trump2016

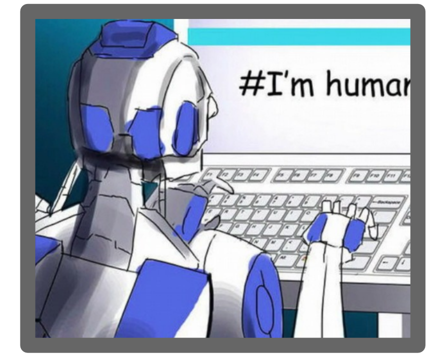
Presentation outline



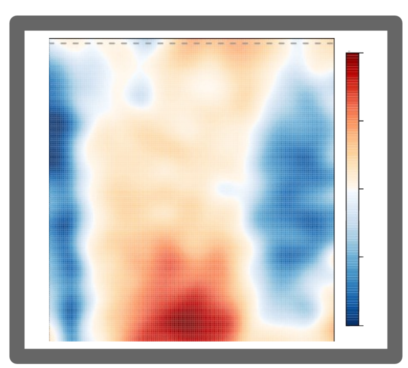
User influence



Political partisanship



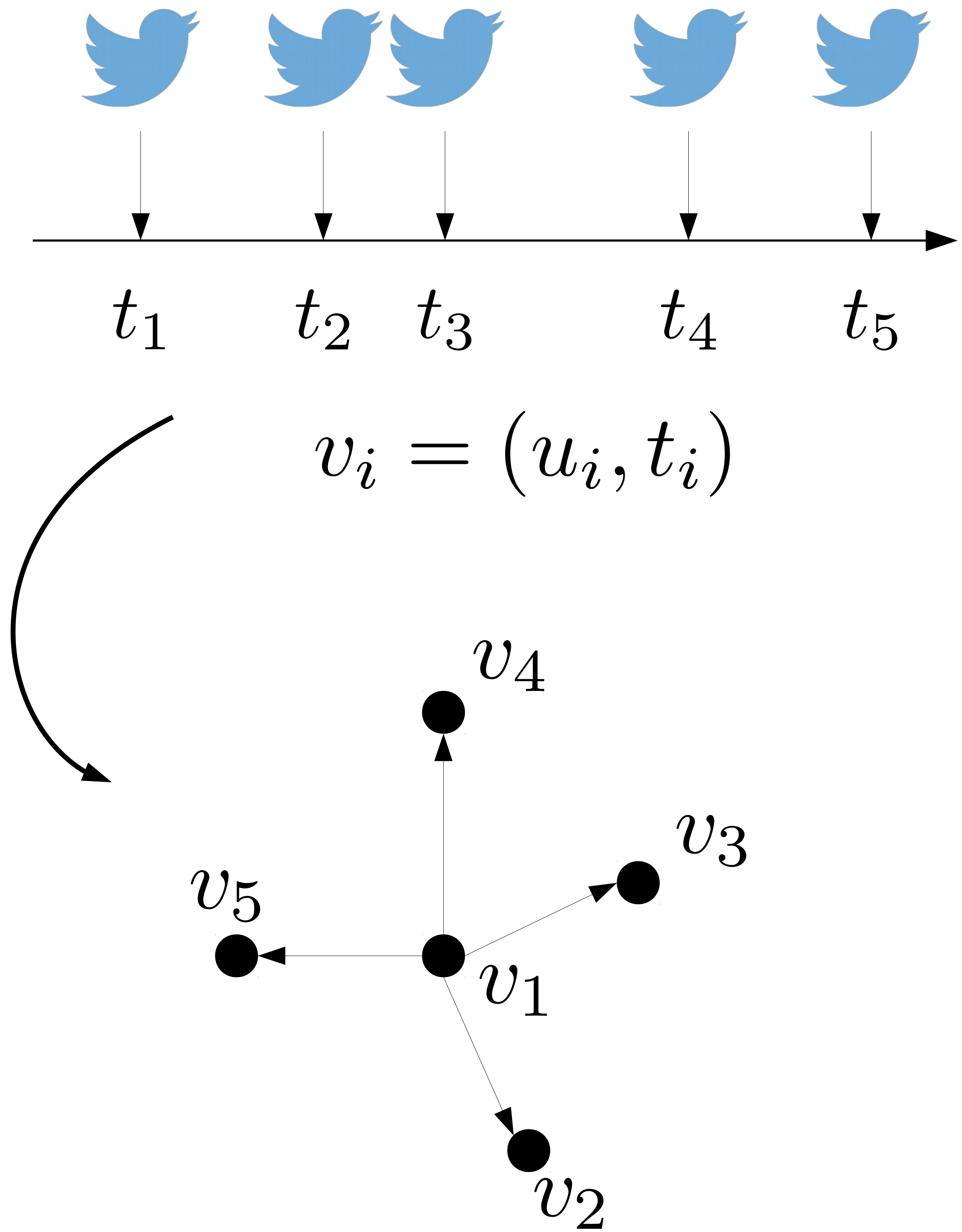
User botness



Analyze political behavior of bots

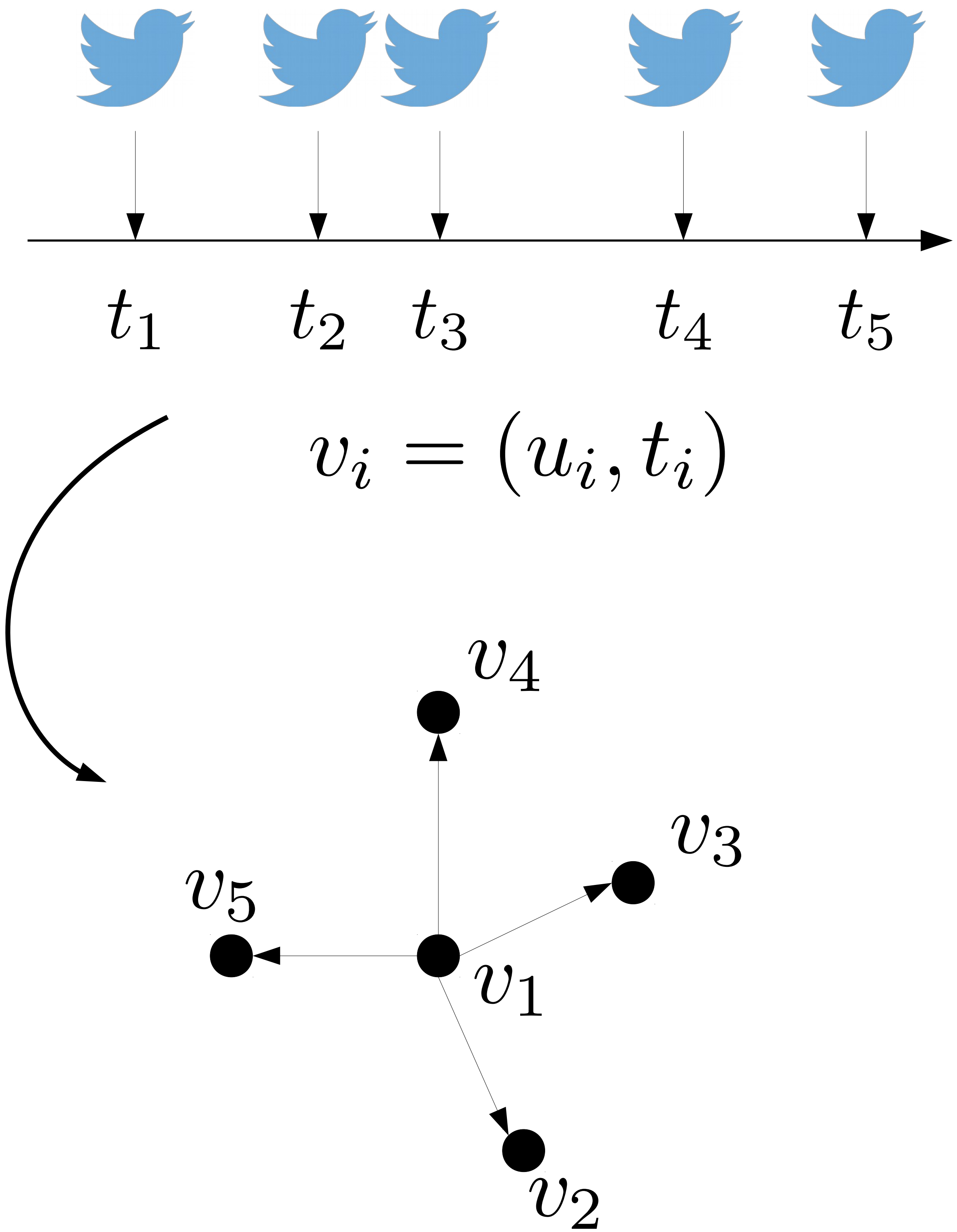
Retweet influence (1)

Behavioral Data Science

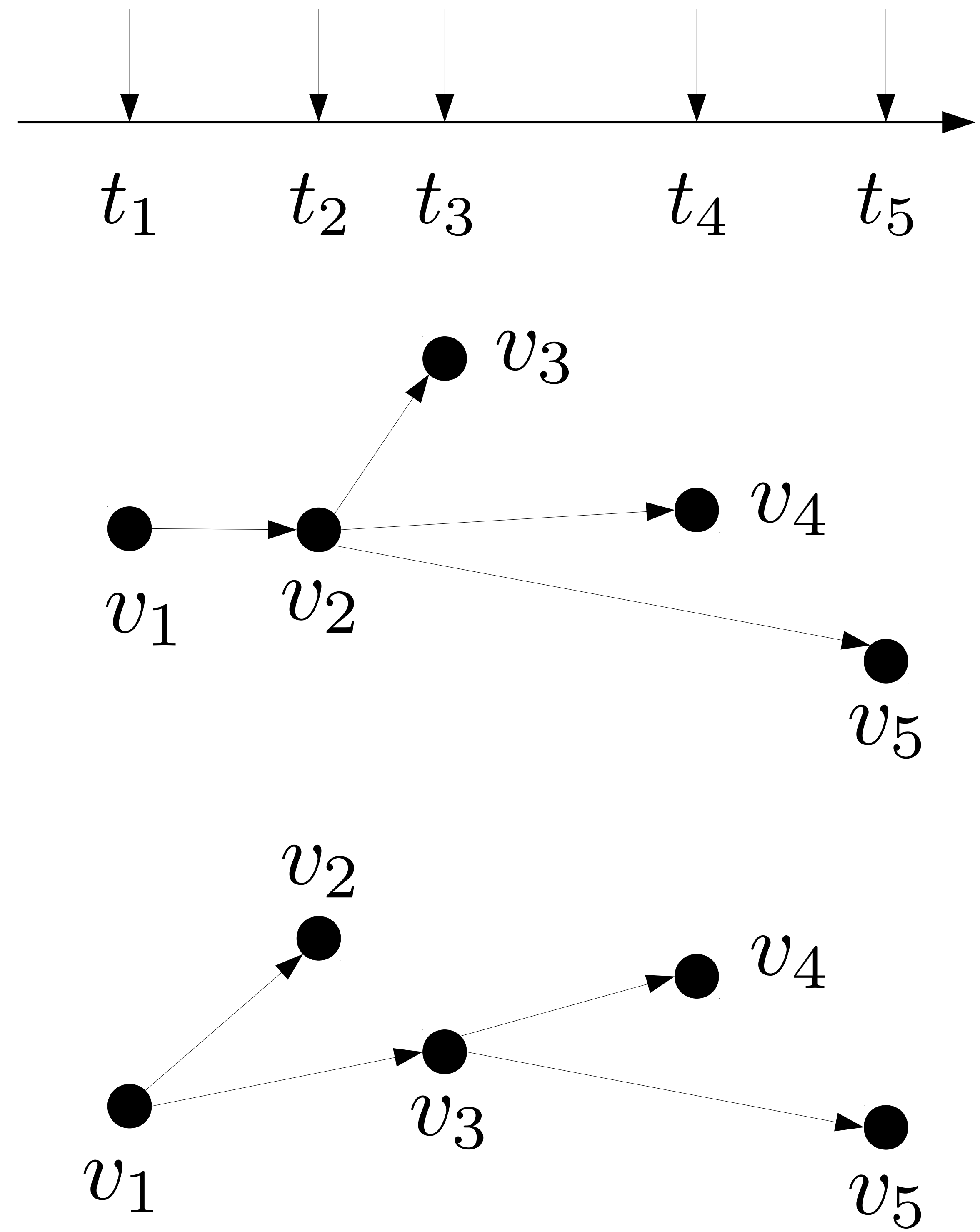


Behavioral Data Science

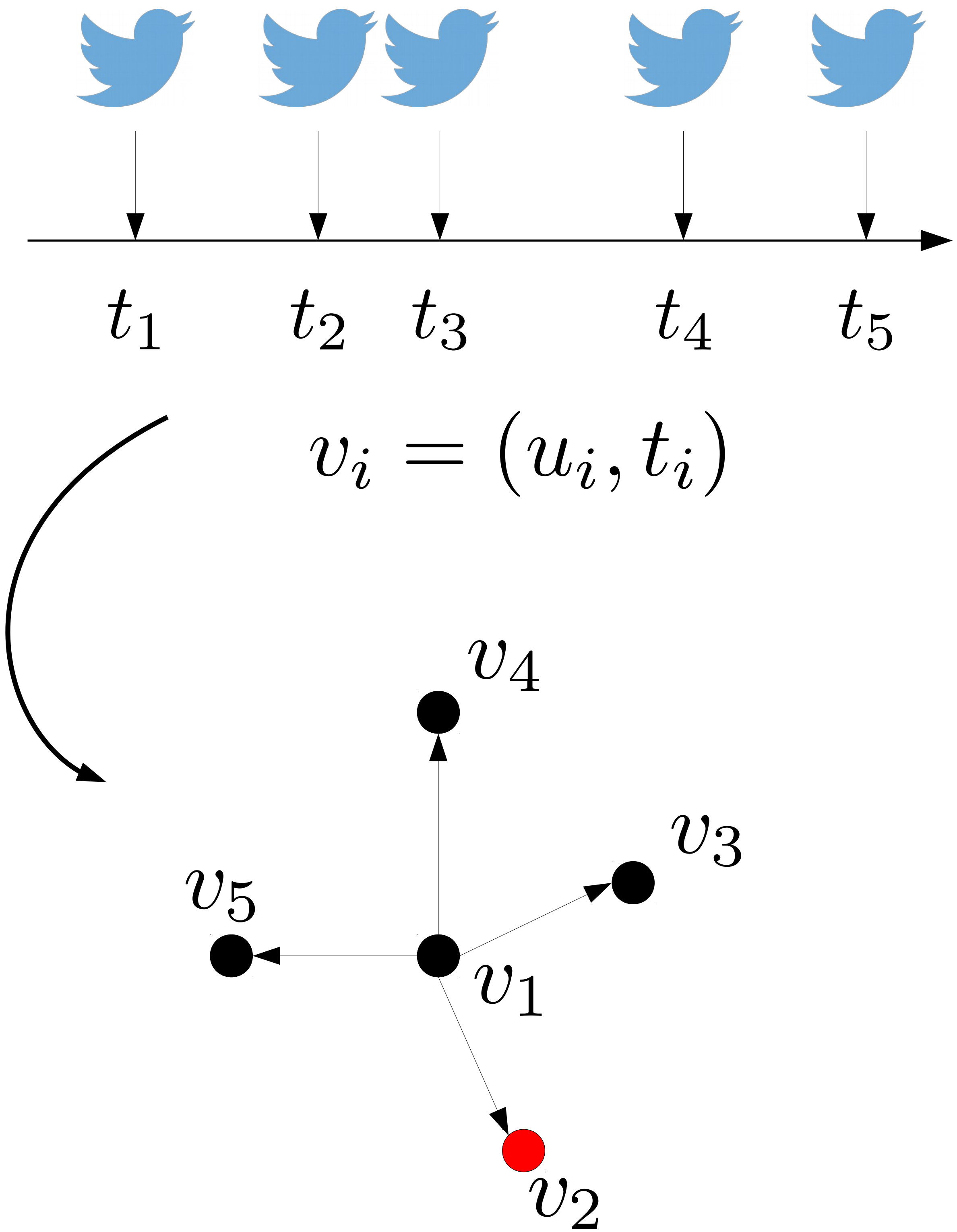
Retweet influence (1)



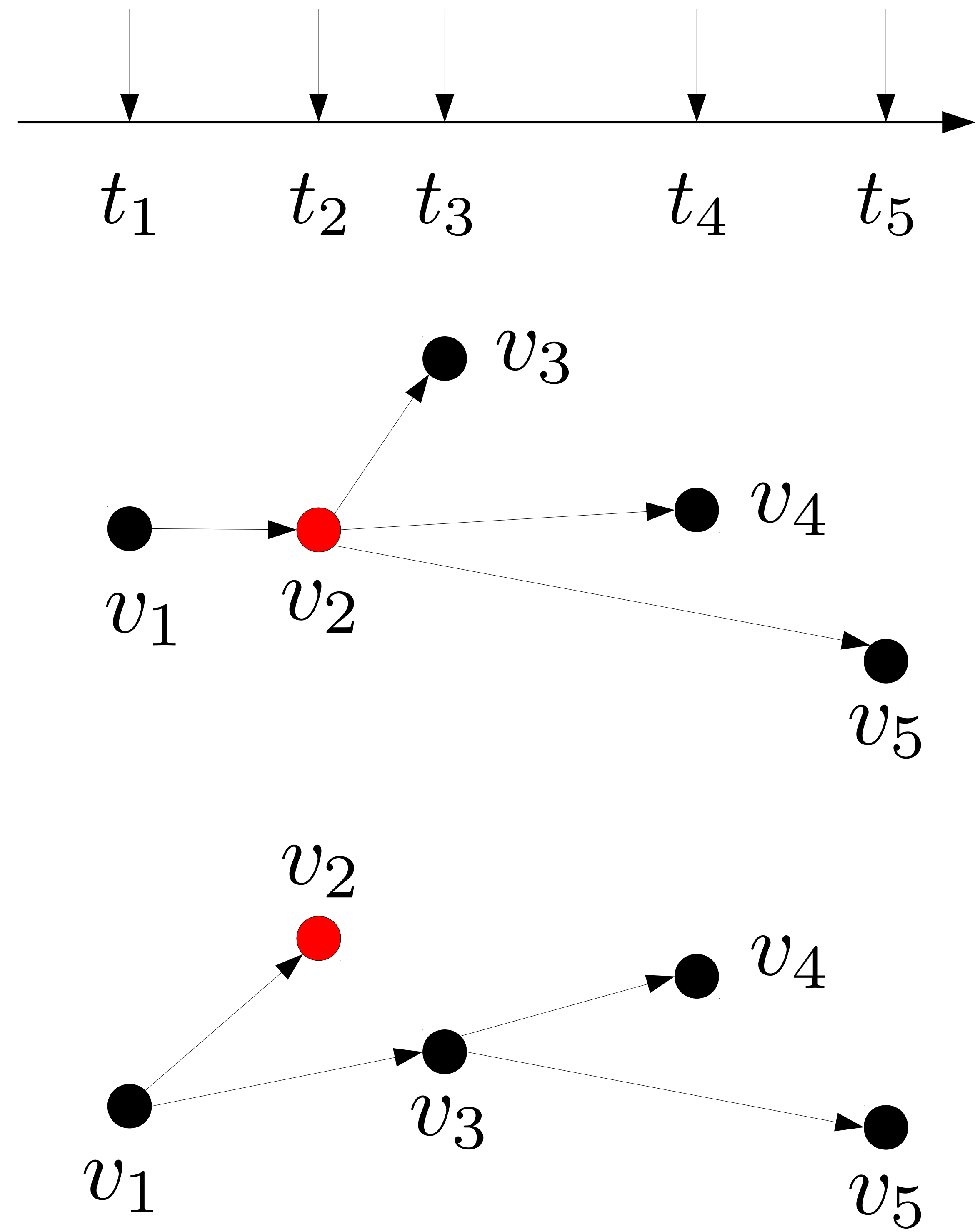
Diffusion trees and influence



Retweet influence (1)



Diffusion trees and influence



Retweet influence (2)

Behavioral Data Science

$$p_{ij} = \frac{m_i e^{-r(t_j - t_i)}}{\sum_{k=1}^{j-1} m_k e^{-r(t_j - t_k)}}$$

branching probability

Retweet influence (2)

Behavioral Data Science

$$p_{ij} = \frac{m_i \mathbf{e}^{-\mathbf{r}(\mathbf{t_j}-\mathbf{t_i})}}{\sum_{k=1}^{j-1} m_k e^{-r(t_j - t_k)}}$$

branching probability

- users retweet *fresh content*

[Hawkes 1971]

[Wu and Huberman 2007]

Retweet influence (2)

#followers of u_i

$$p_{ij} = \frac{\mathbf{m_i}\mathbf{e}^{\mathbf{-r(t_j - t_i)}}}{\sum_{k=1}^{j-1} m_k e^{-r(t_j - t_k)}}$$

branching probability

- users retweet *fresh content*

  [Hawkes 1971]

  [Wu and Huberman 2007]

- preferential attachment

  [Barabási 2005]

Retweet influence (2)

#followers of u_i

$$p_{ij} = \frac{\mathbf{m_i e^{-r(t_j - t_i)}}}{\sum_{k=1}^{j-1} m_k e^{-r(t_j - t_k)}}$$

branching probability

- users retweet *fresh content*

[Hawkes 1971]

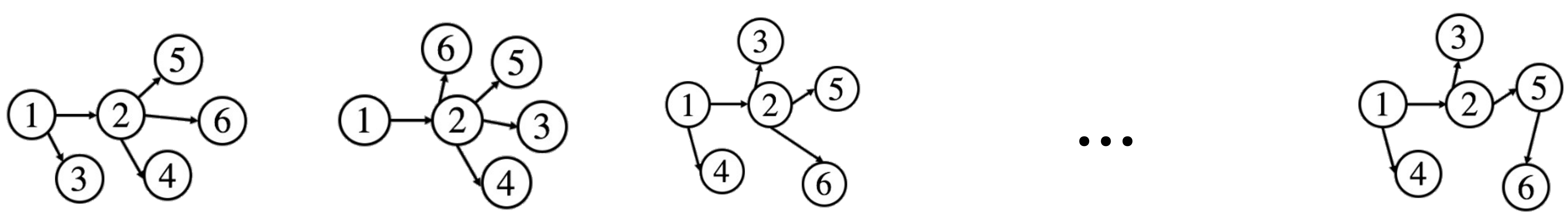
[Wu and Huberman 2007]

- preferential attachment

[Barabási 2005]

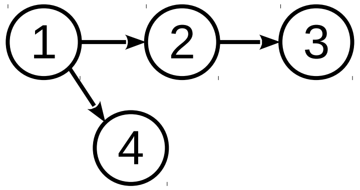
Tweet influence: the expected number of retweets, averaged over all possible trees.

But … $(n-1)!$ trees $\quad 10^{156}$ trees for 100 tweets

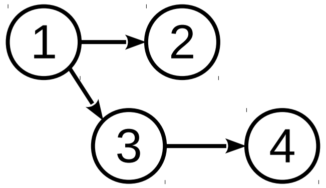


Tractable influence computation

Pair-wise influence score m_{ij}



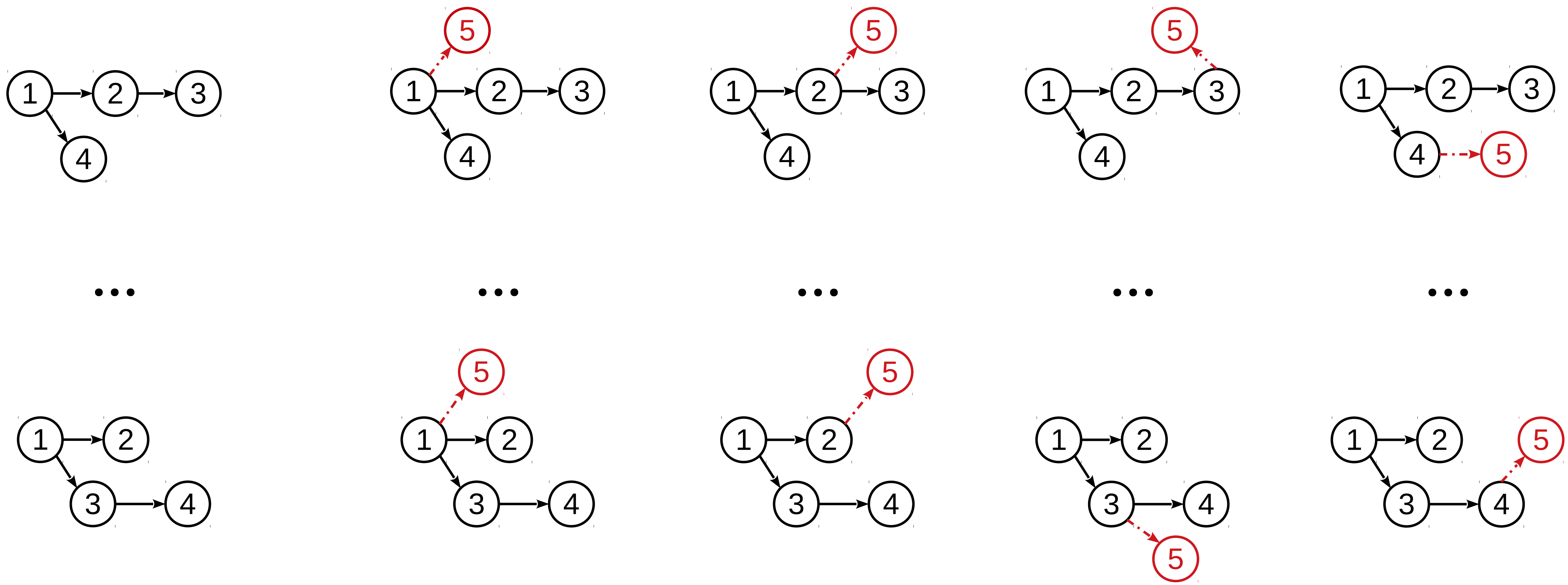
...



Tractable influence computation

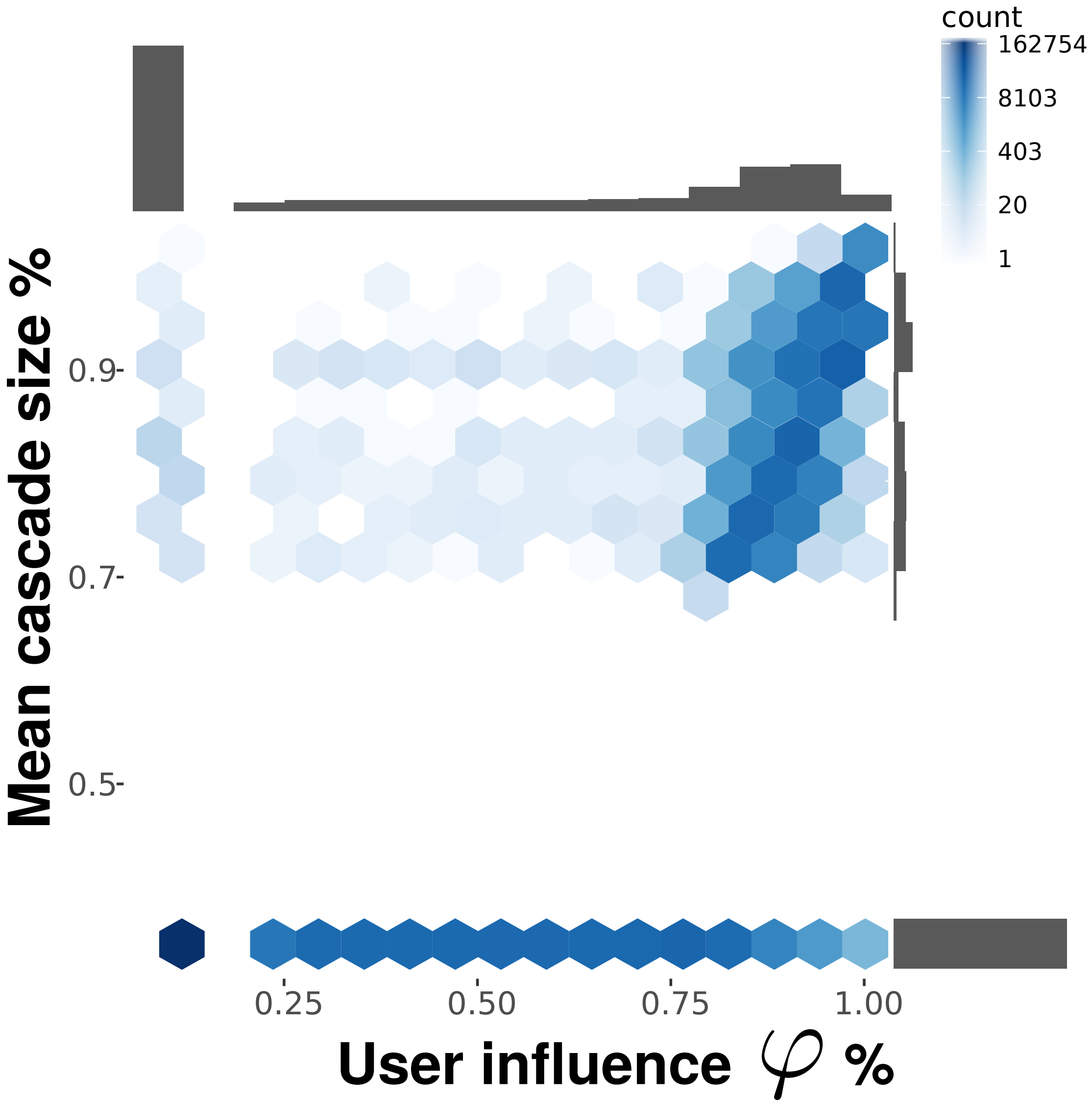
Pair-wise influence score m_{ij}

$$m_{15} = m_{11}p_{15} + m_{12}p_{25} + m_{13}p_{35} + m_{14}p_{45}$$



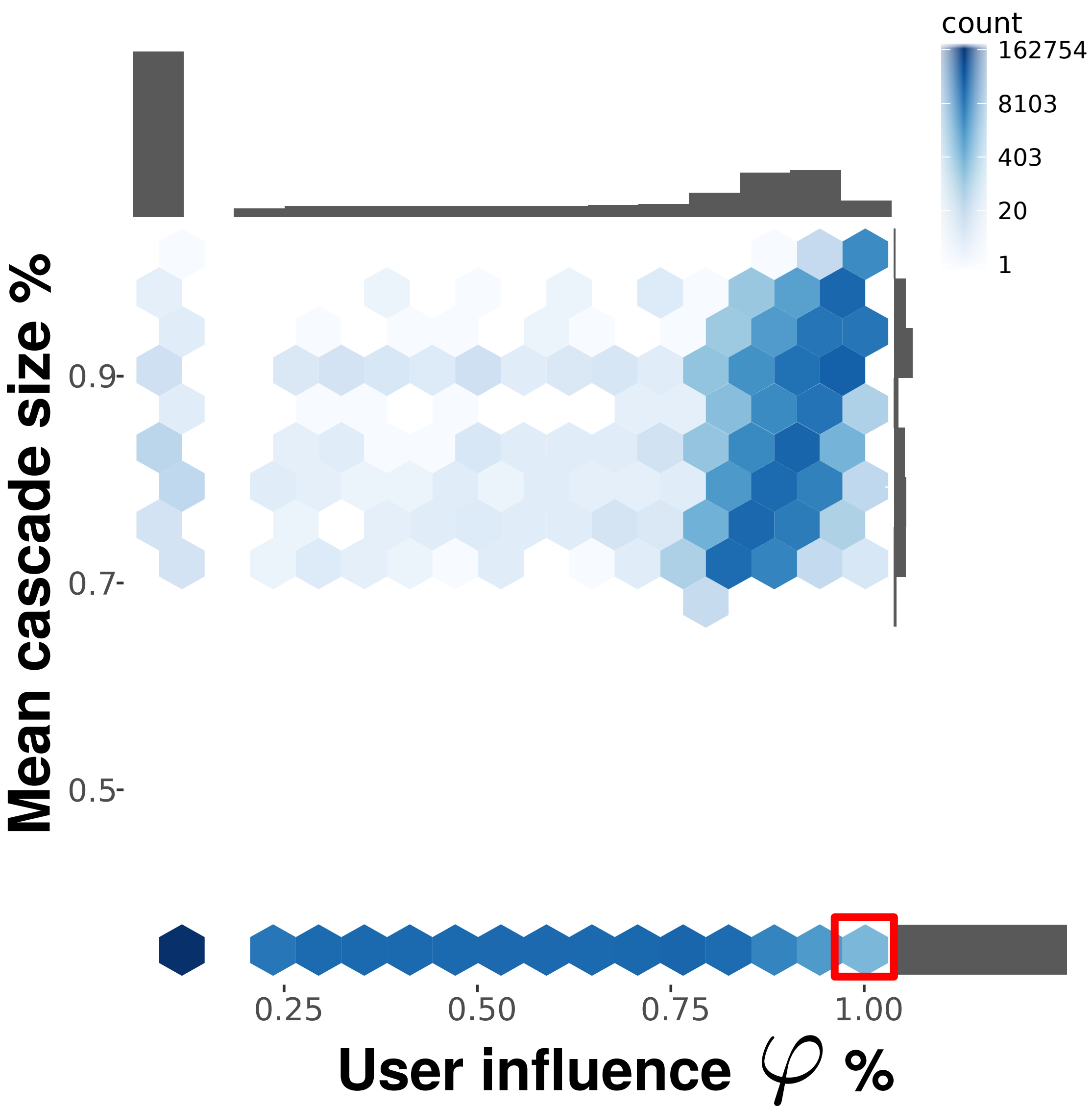
Recursive algorithm $O(n^3)$

Supp: Influence vs. cascade size



Density plot for 653K users
(45% users start a cascade)

Supp: Influence vs. cascade size



Density plot for 653K users (45% users start a cascade)

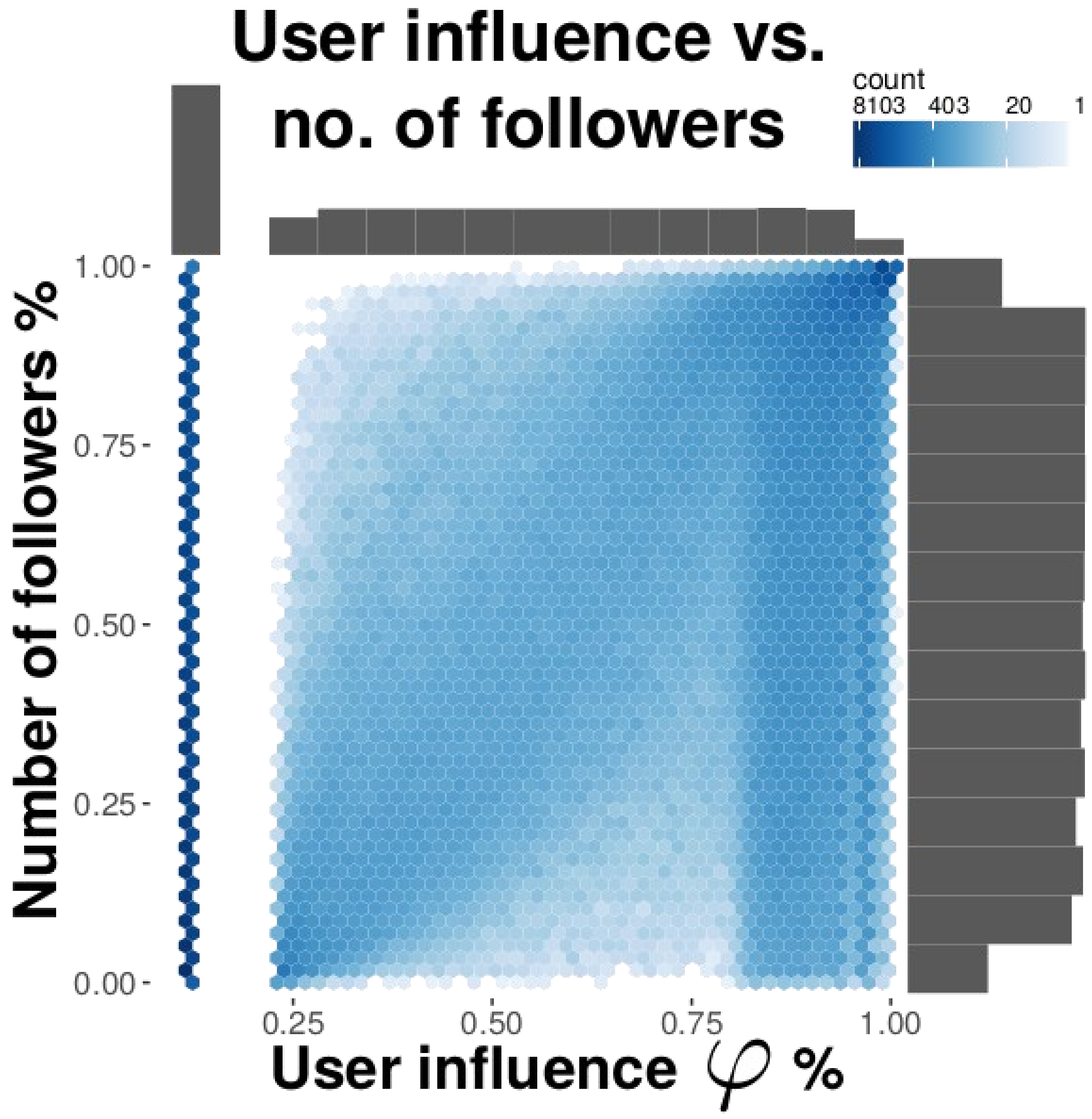
actor and filmmaker

10.8 million followers

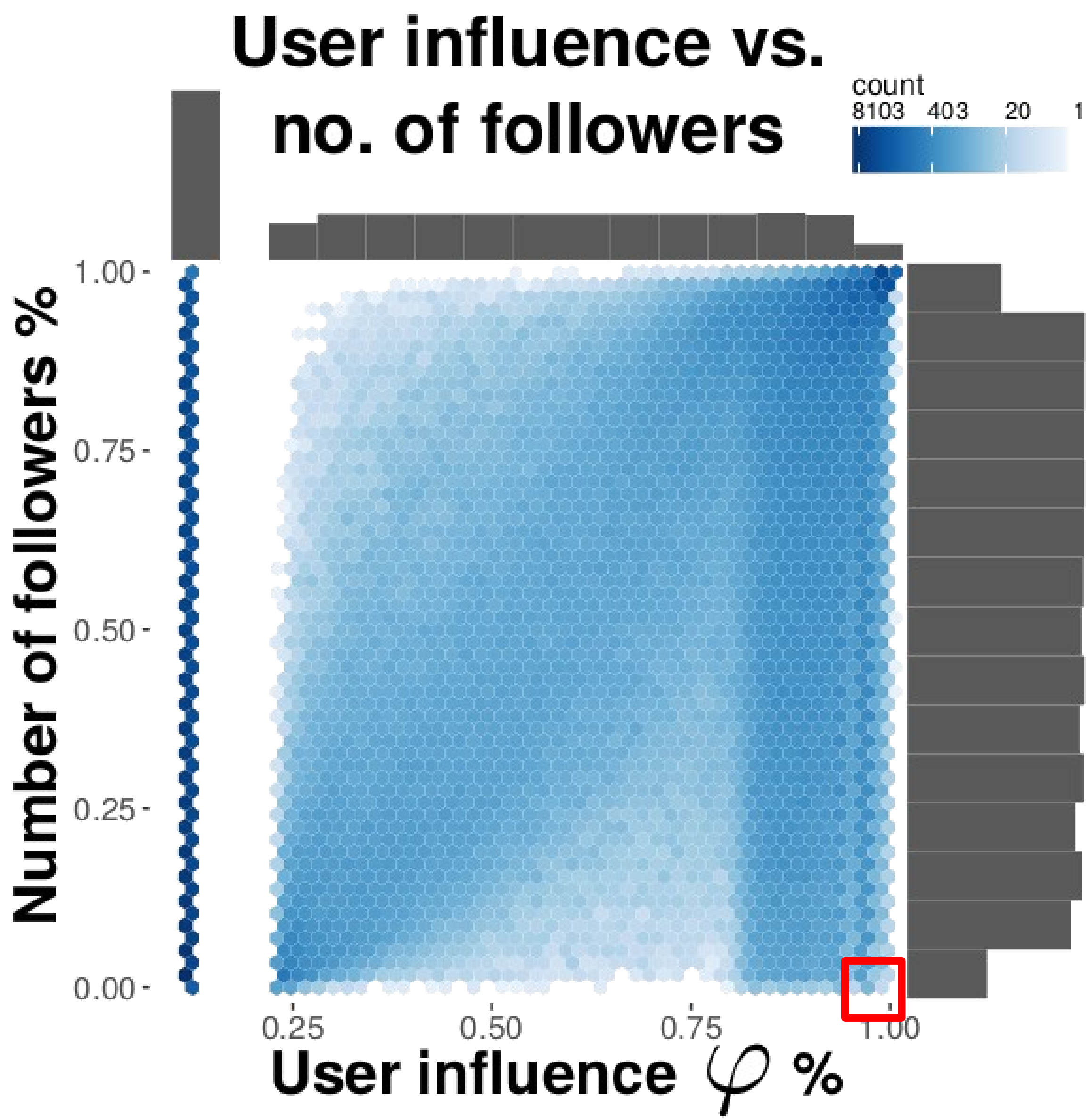
comedian

2.1 million followers

Supp: Influence vs. number of followers



Supp: Influence vs. number of followers



2 followers

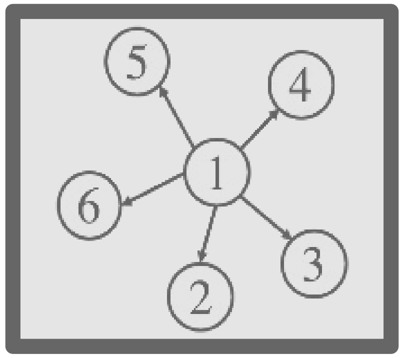
Initiated a big cascade

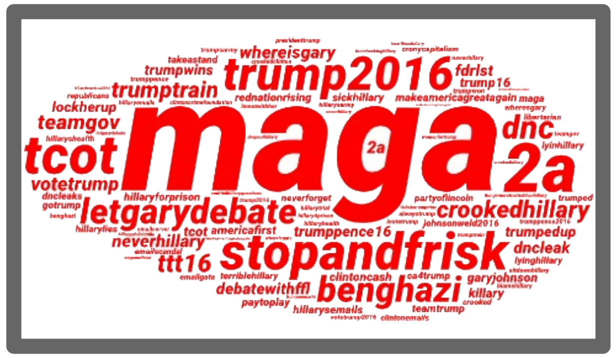
now suspended

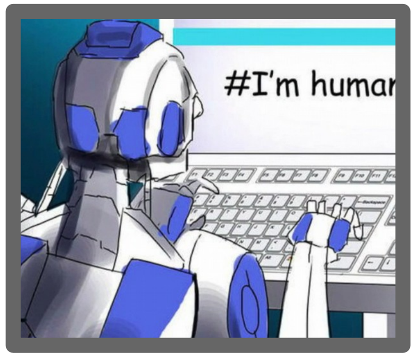
1 follower

Initiated a big cascade

Presentation outline



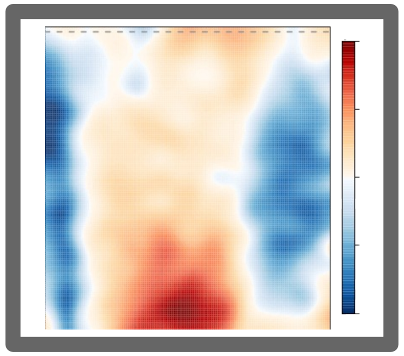




User influence

Political partisanship

User botness



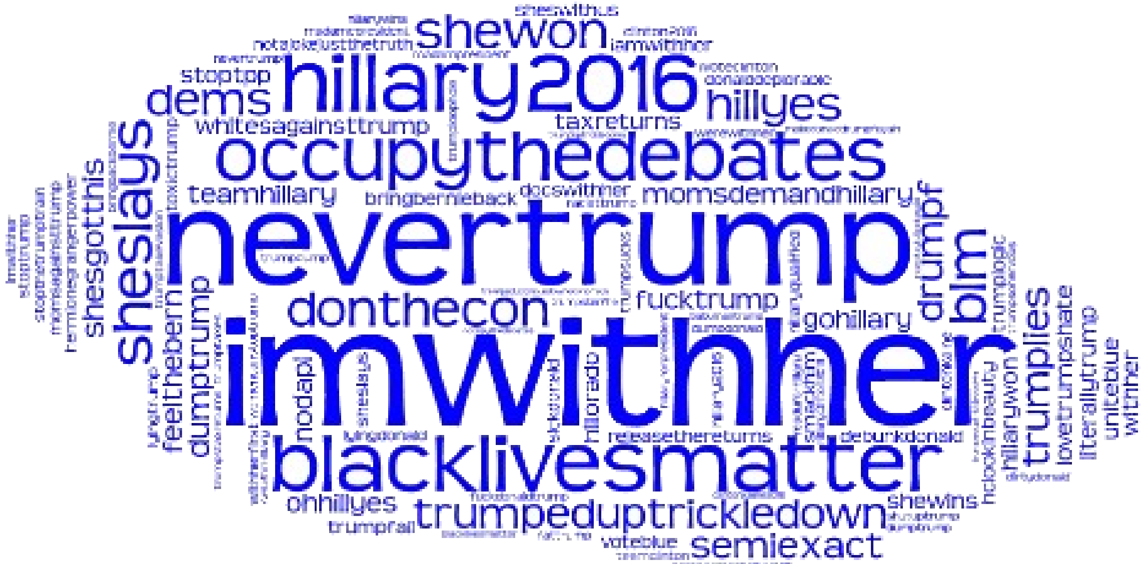
Analyze political behavior of bots

Political polarization (1)

Behavioral
Data Science

Protocol:

- Top 1000 most frequent hashtags
- Manually labeled as *clearly partisan* pro-democrat or pro-republican



Partisanship stats:

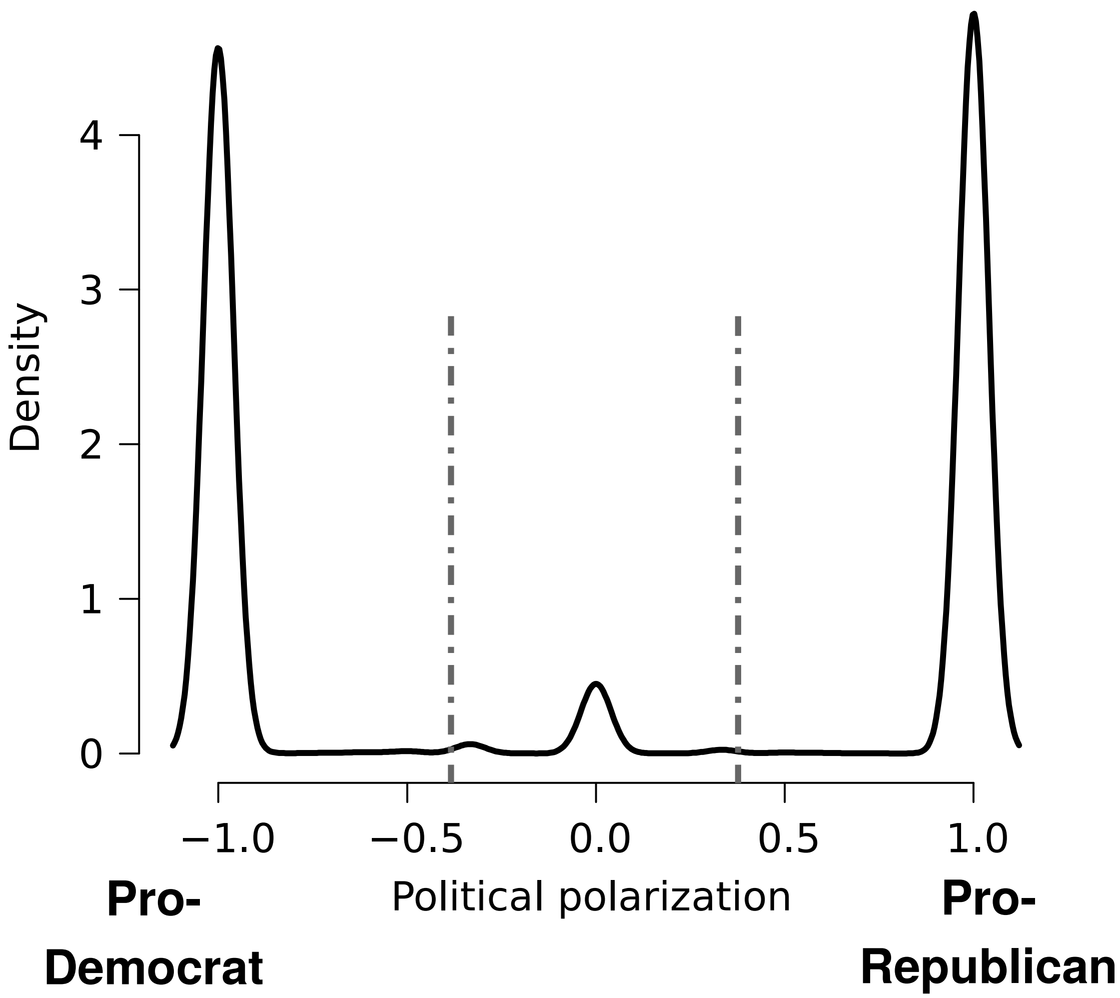
- pro-Democrat hashtags: **93**
- pro-Republican hashtags: **86**
- partisan tweets: **65K**
- partisan users: **47K**

Political polarization (2)

For each user *i*:

- dem_i – #democrat hashtags
- rep_i – #republican hashtags

$$\mathcal{P}(u_i) = \frac{rep_i - dem_i}{rep_i + dem_i}$$

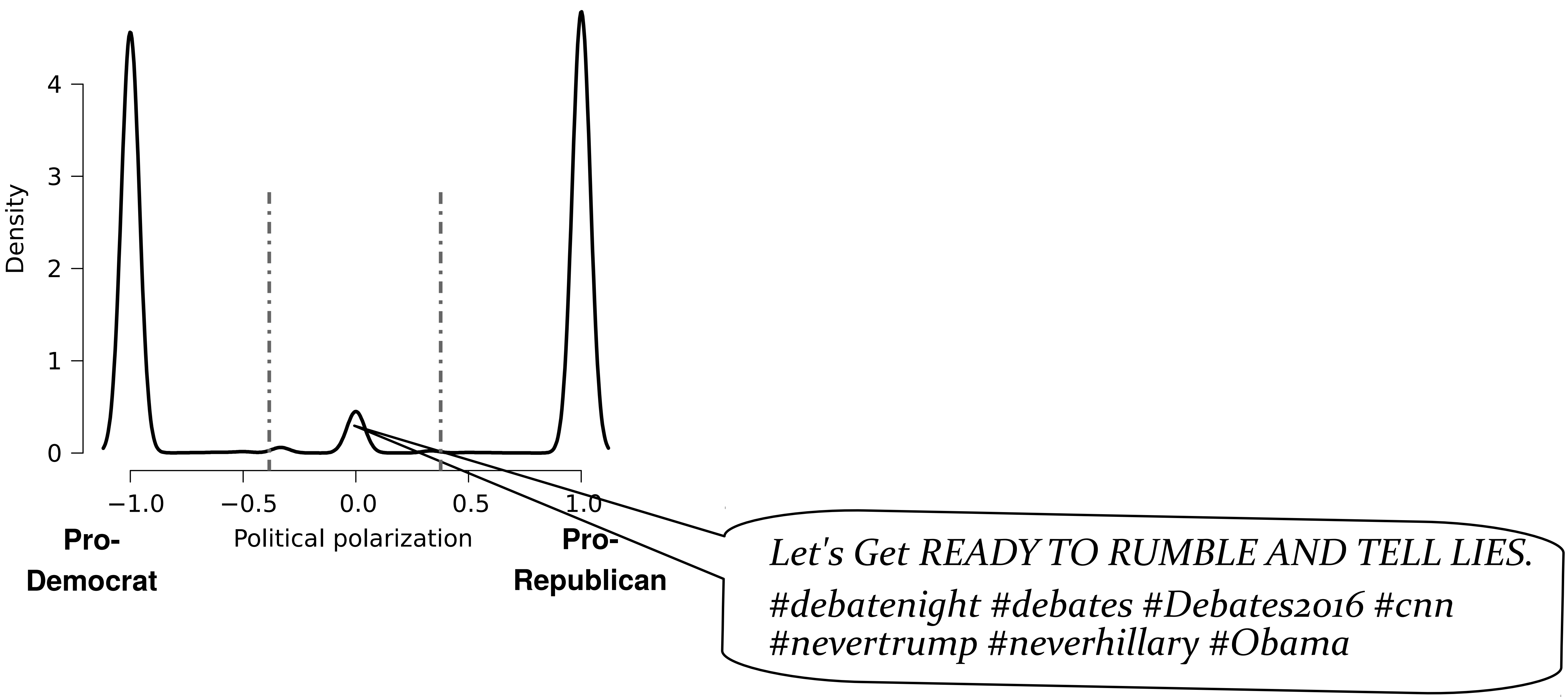


Political polarization (2)

For each user *i*:

- dem_i – #democrat hashtags
- rep_i – #republican hashtags

$$\mathcal{P}(u_i) = \frac{rep_i - dem_i}{rep_i + dem_i}$$



Botness score and bot detection

Behavioral Data Science

Bot detection:

- **BotOrNot** [Davis et al, WWW '16] [Varol et al, ICWSM'17]
  - RandomForest classifier
  - more than 1000 features from metadata
    - 0 – very likely human
    - 1 – very likely bot

- 94.5% precision

Separating bots from humans

Behavioral Data Science

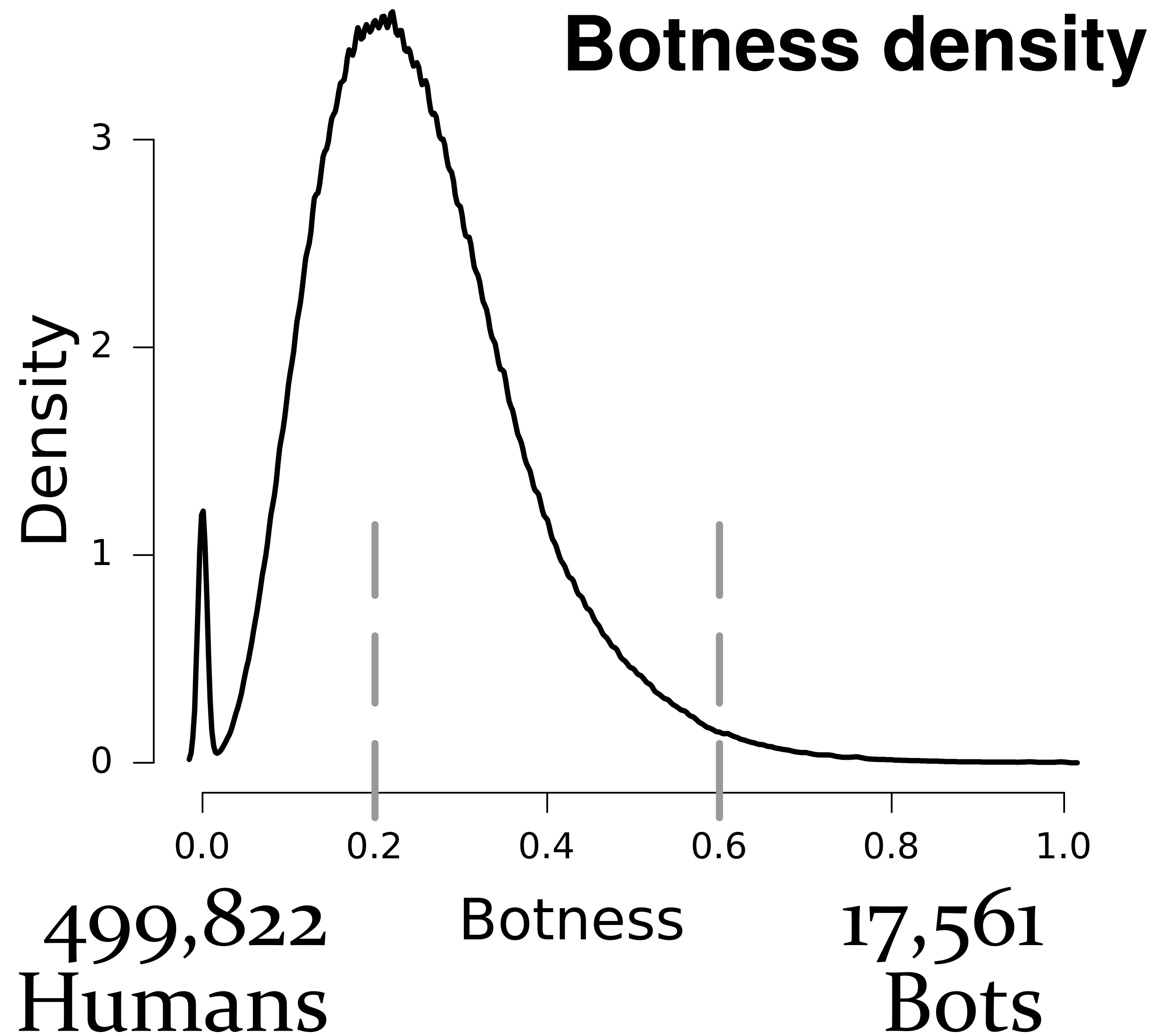
Three populations

Population	Effective
All	1,451,388
Protected	45,316
Suspended	10,162

Separating bots from humans

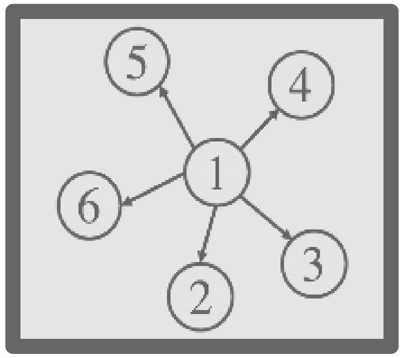
Three populations

Population	Effective
All	1,451,388
Protected	45,316
Suspended	10,162

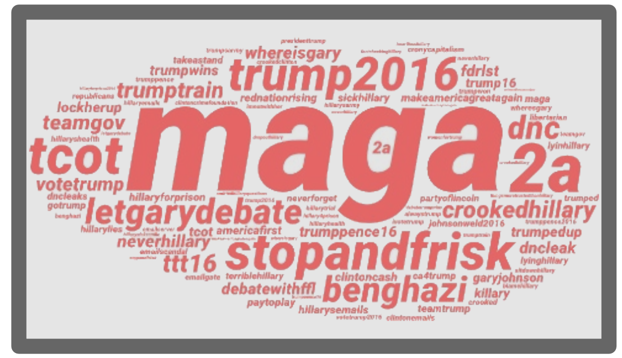


[Varol et al, ICWSM'17] use a threshold of 0.5

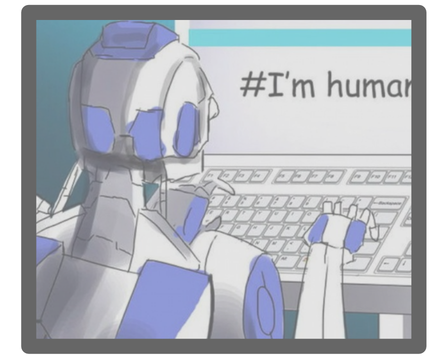
Presentation outline



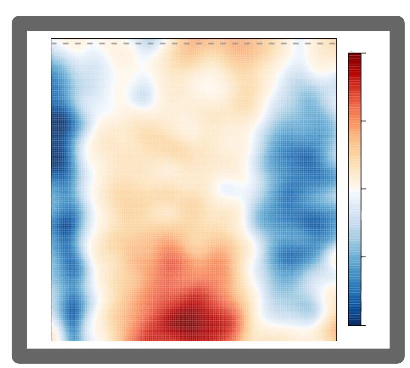
User influence



Political partisanship

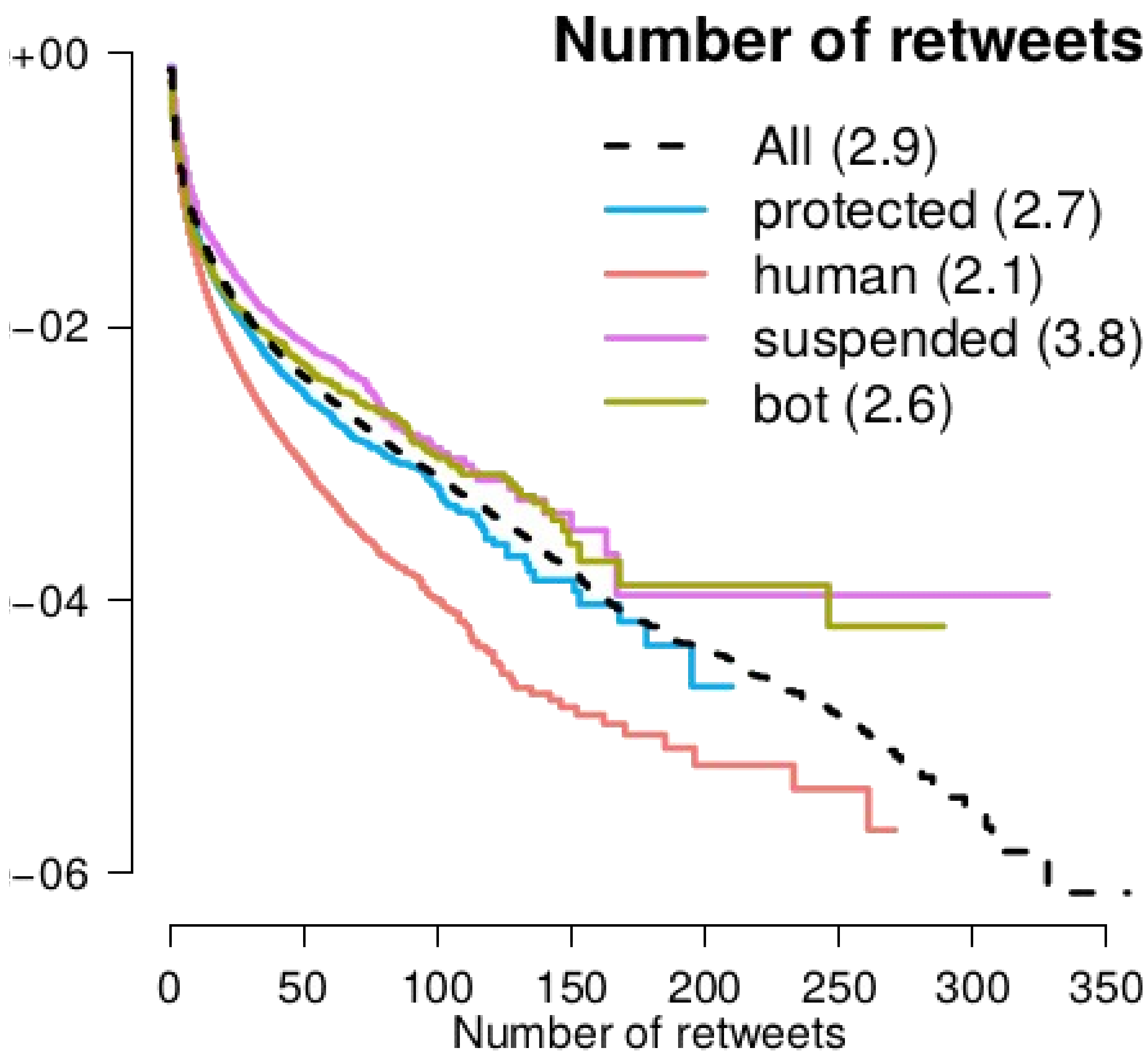


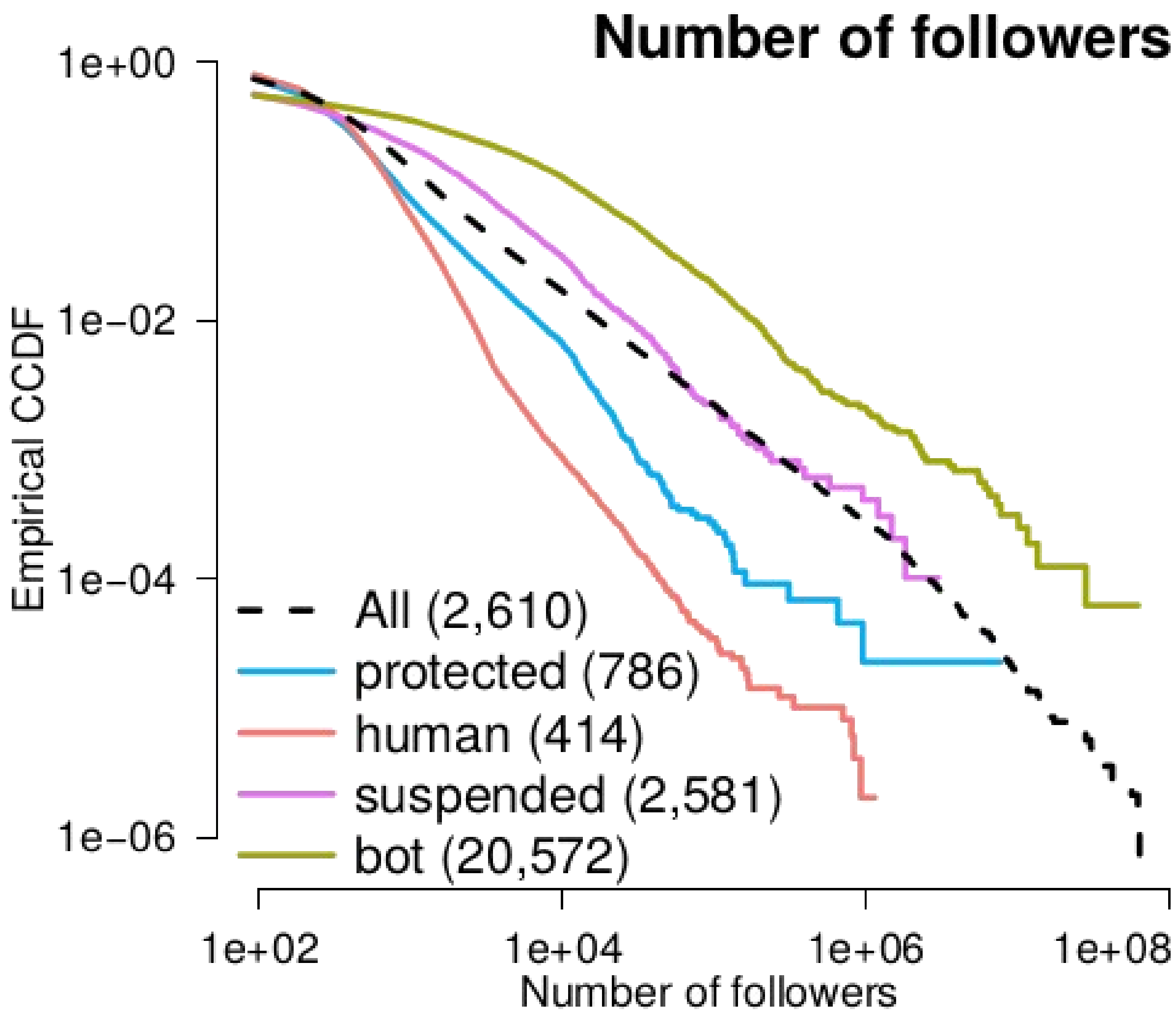
User botness



Analyze political behavior of bots

Activity profiling

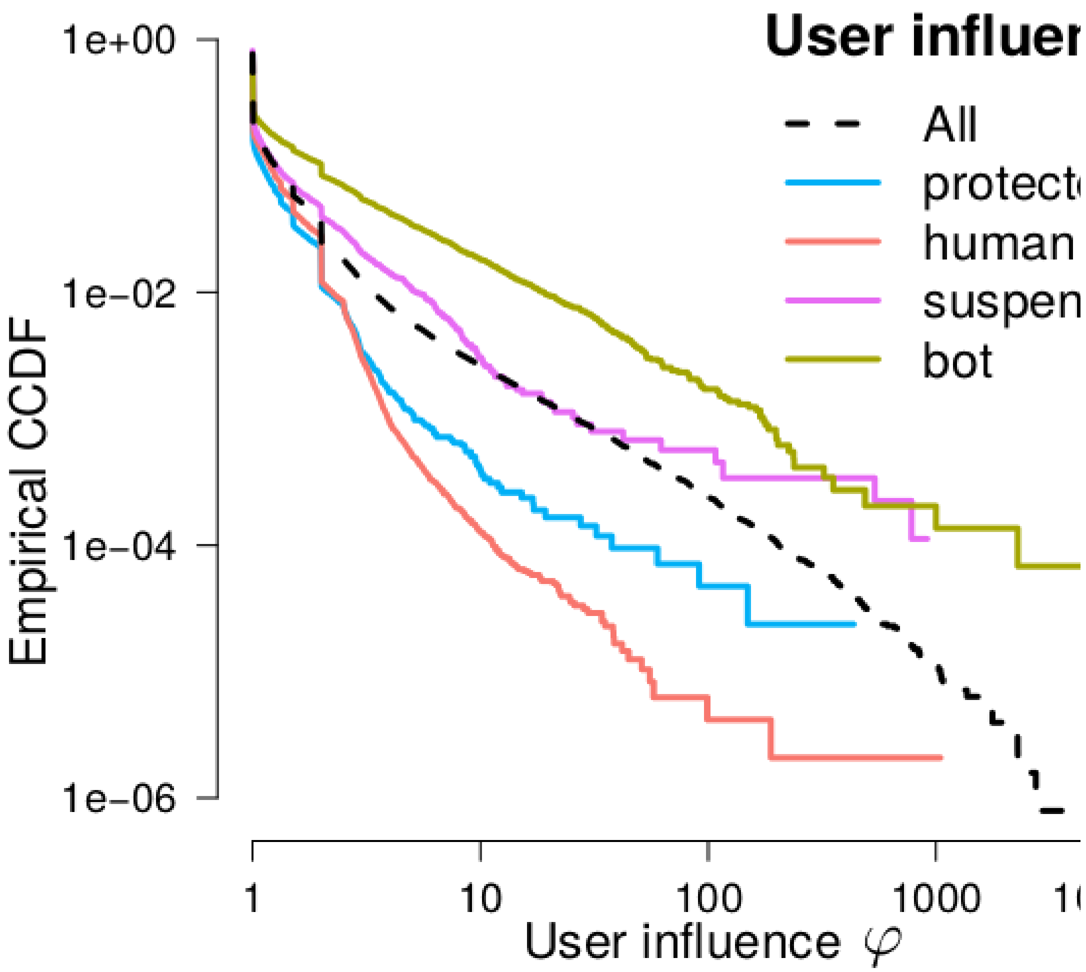


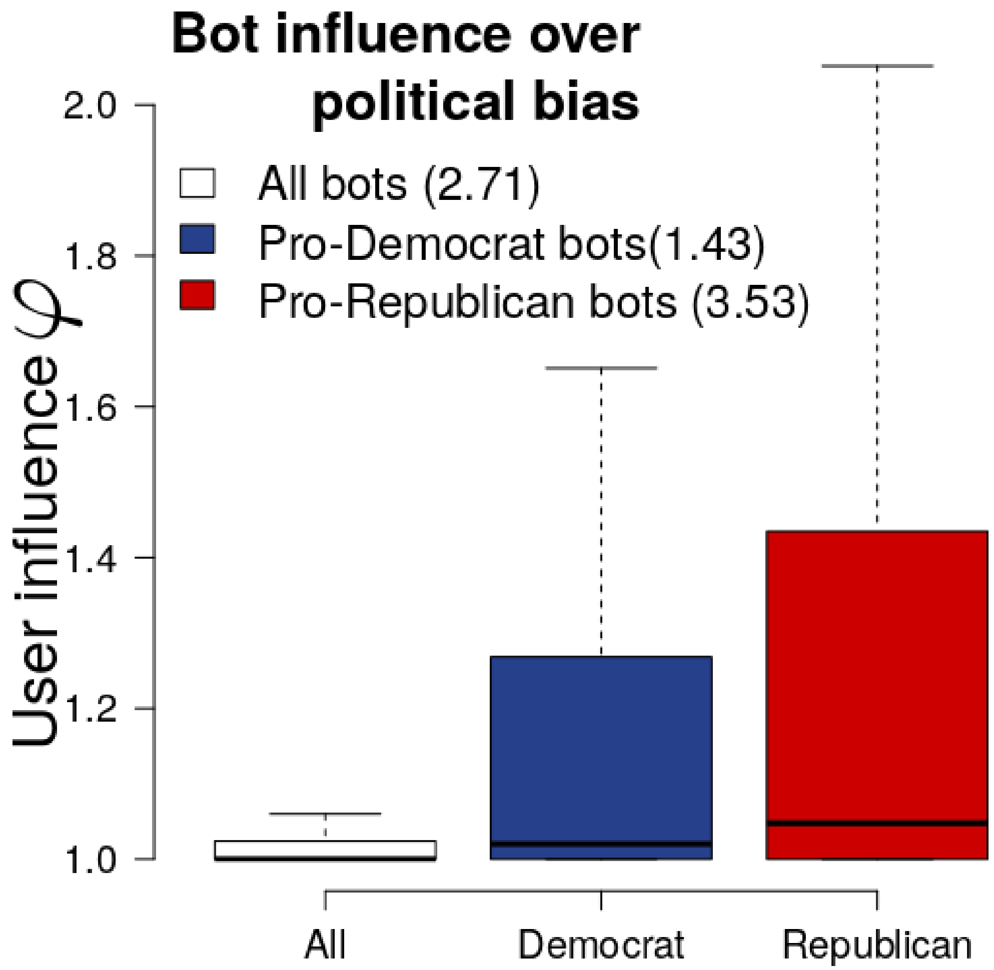


Bots and **Suspended** are more active than **Humans** and **Protected**

Some **Bots** are highly followed, while most are ignored

User influence

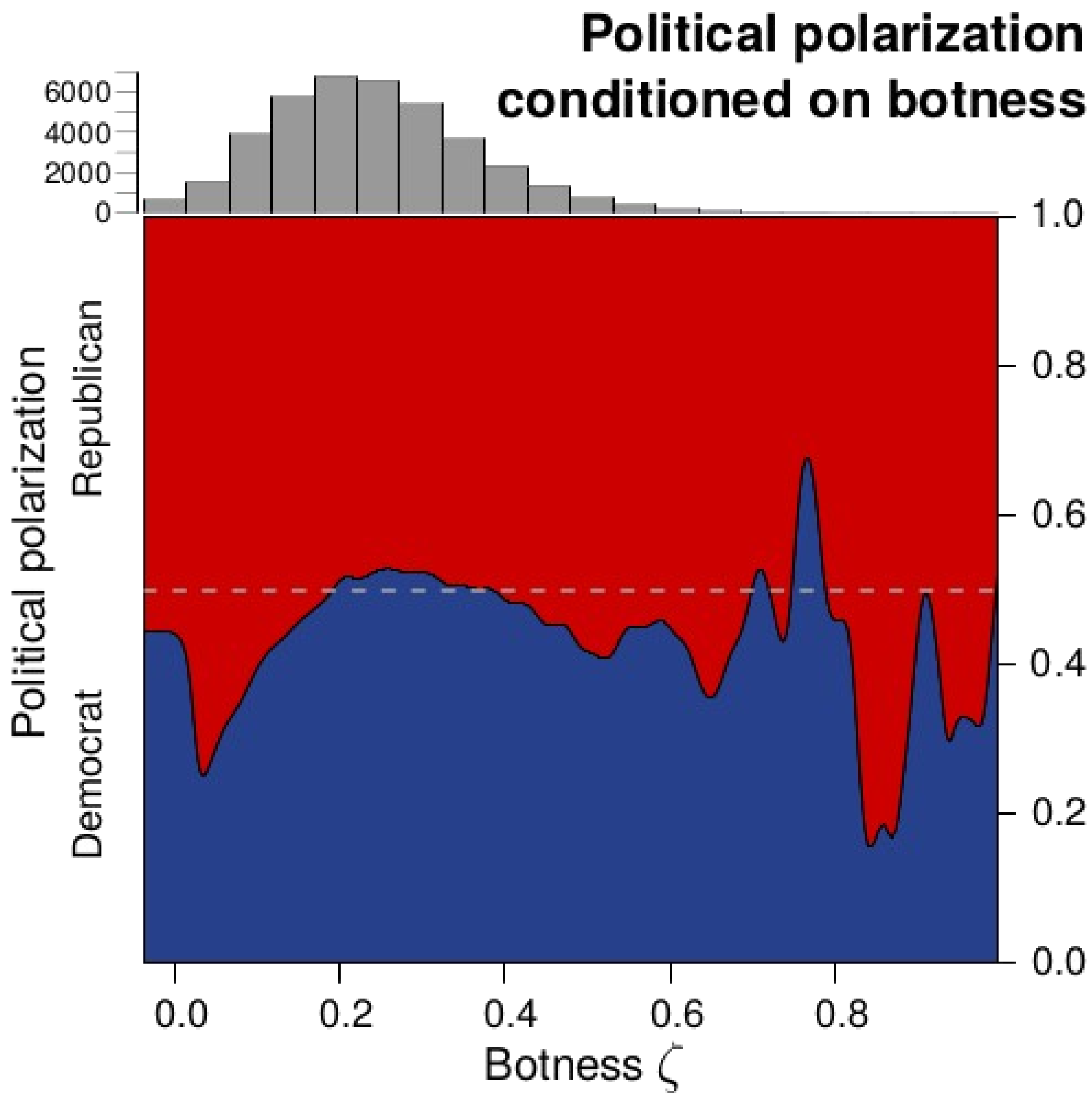


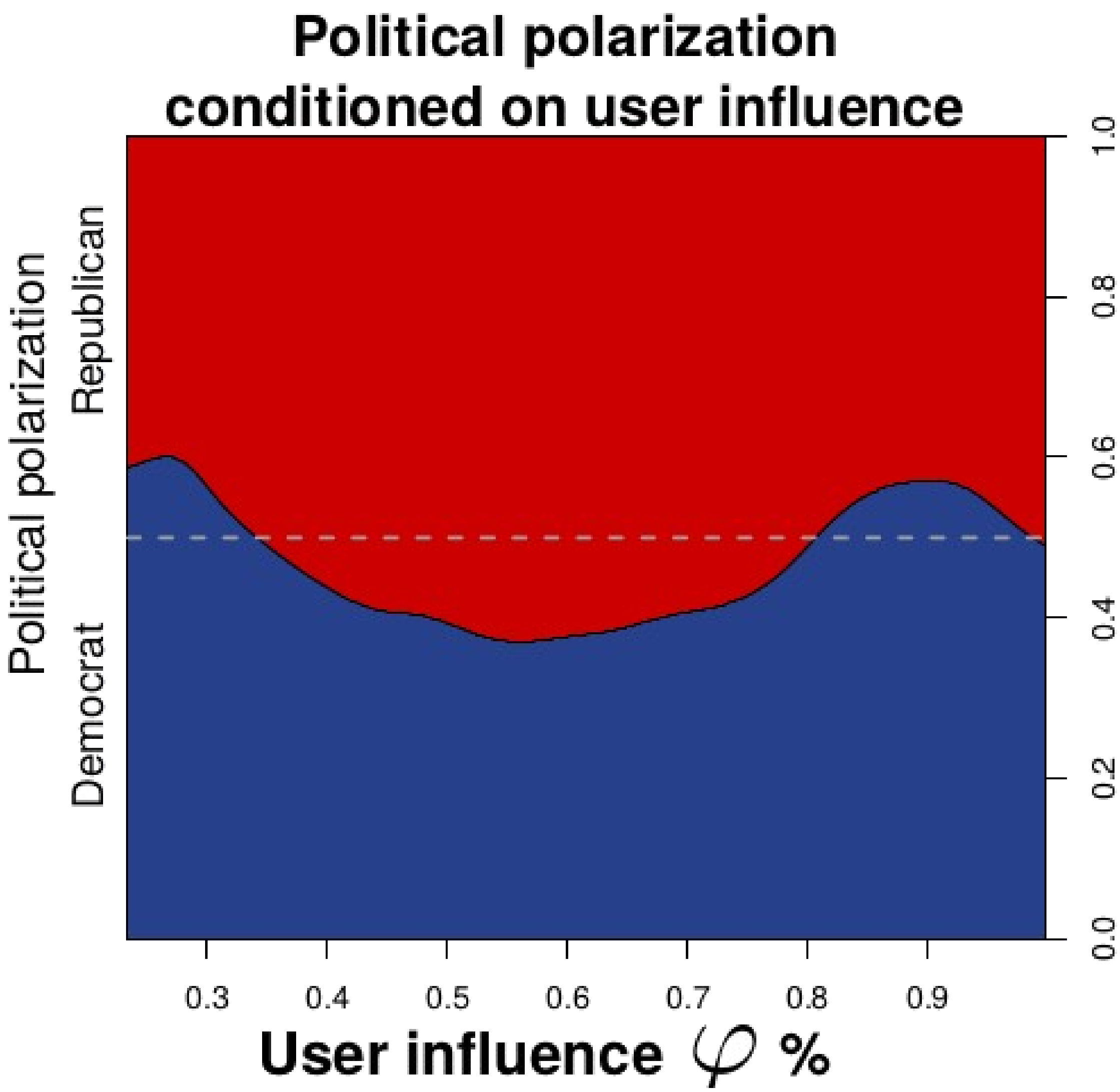


The average **Bot** has 2.5 times more influence than the average **Human**

The average pro-Republican **Bot** is twice as influential as the average pro-Democrat **Bot**

Political partisanship

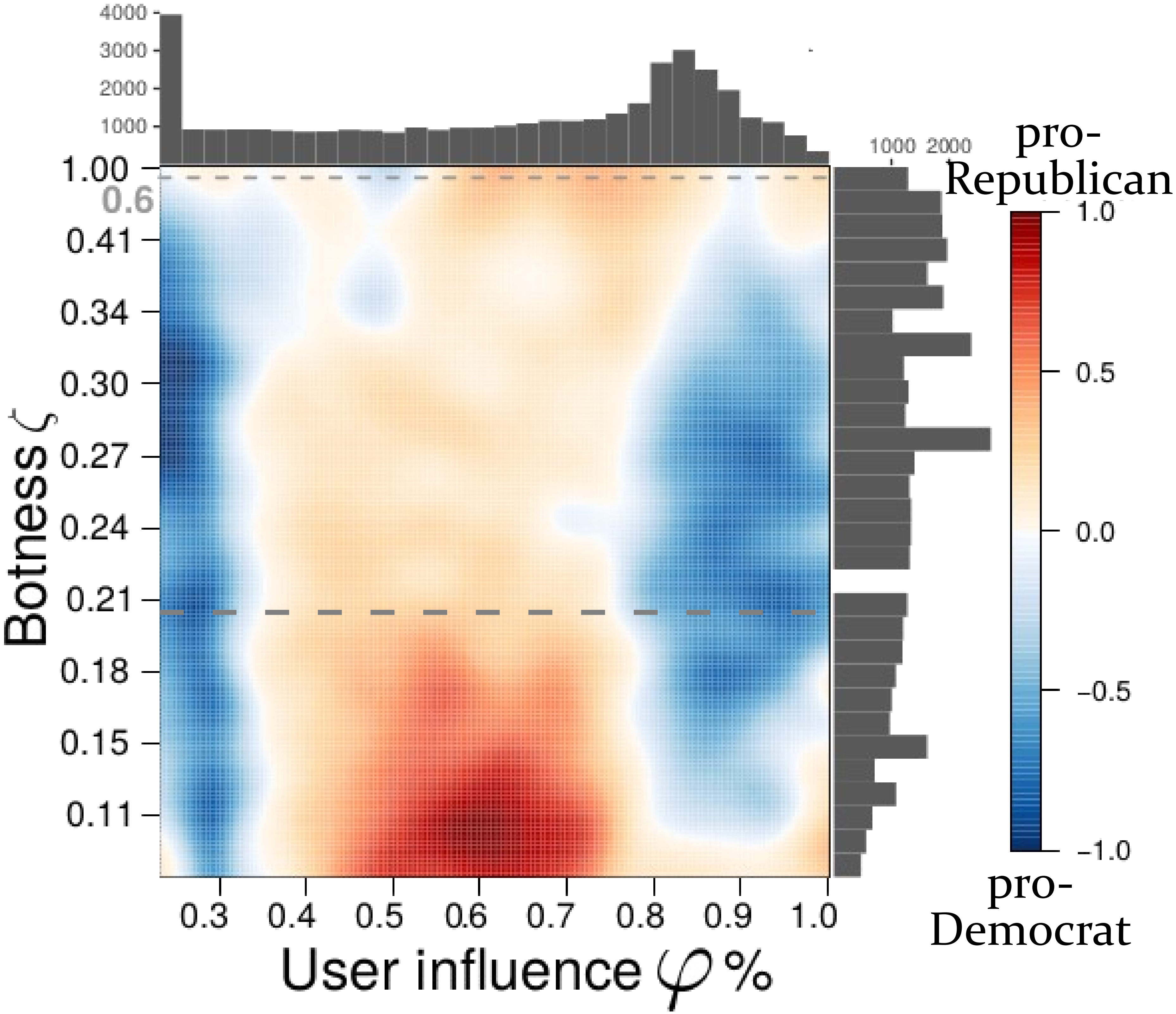




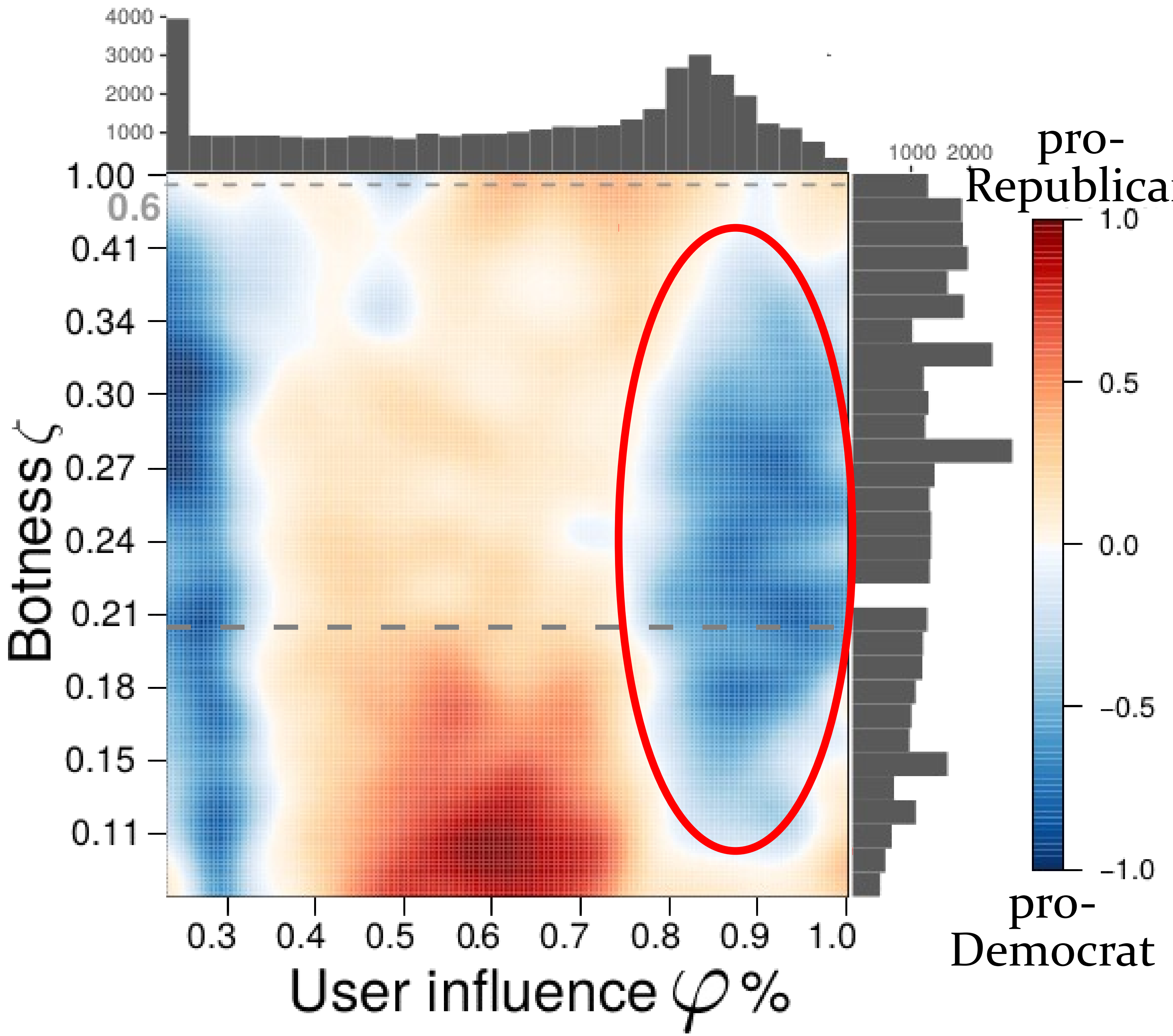
Bots are more likely to be pro-Republican (than pro-Democrat)

Very highly influential users are more likely to be pro-Democrat

Polarization map

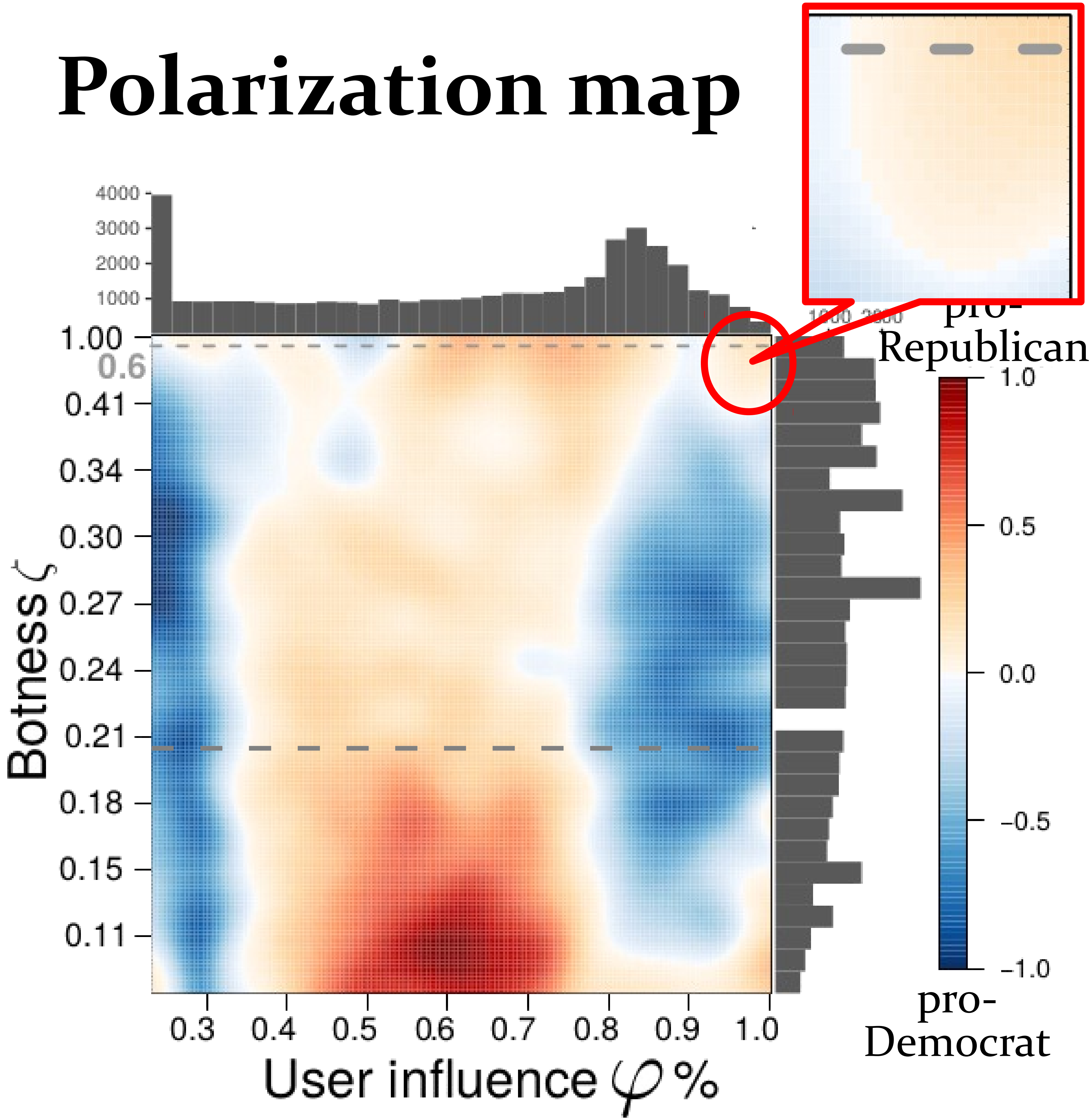


Polarization map



Very highly influential users are pro-Democrat (**D: 7201**, **R: 5736**)

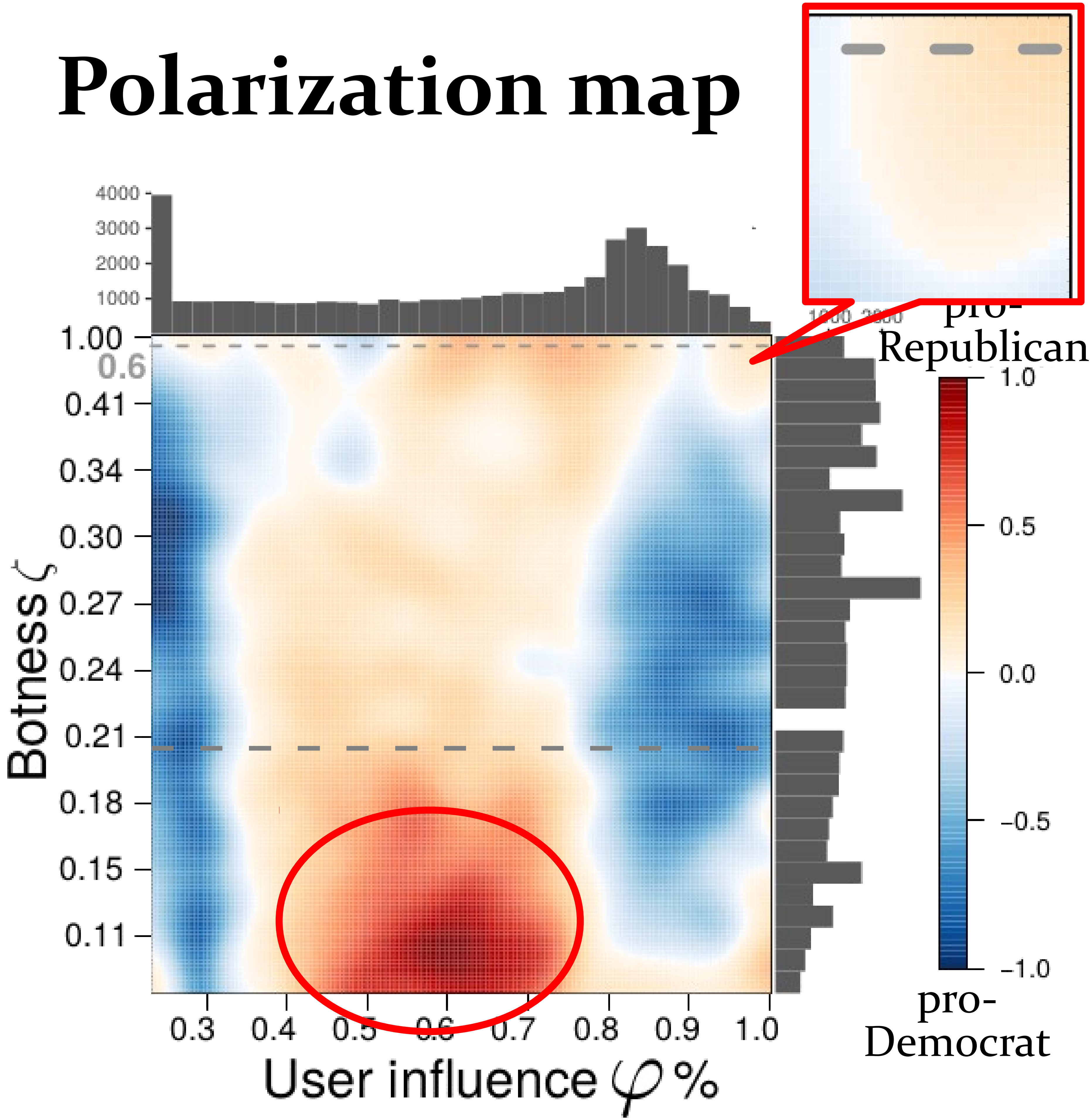
Polarization map



Very highly influential users are pro-Democrat
(**D: 7201**, **R: 5736**)

Highly influential **Bots** are pro-Republican
(**D: 24**, **R: 45**)

Polarization map



Very highly influential users are pro-Democrat
(D: 7201, R: 5736)

Highly influential **Bots** are pro-Republican
(D: 24, R: 45)

Mid-influential humans are pro-Republican
(D: 1530, R: 3311)


Noname manuscript No.
(will be inserted by the editor)

Analysing user identity via time-sensitive semantic edit distance (t-SED): A case study of Russian trolls on Twitter

Dongwoo Kim · Timothy Graham · Zimin Wan · Marian-Andrei Rizoiu

Received: date / Accepted: date

Abstract In the digital era, individuals are increasingly profiled and grouped based on the traces they leave behind in online social networks such as Twitter and Facebook. In this paper we develop and evaluate a novel text analysis approach for studying user identity and social roles by redefining identity as a sequence of timestamped items (e.g. tweet texts). We operationalise this idea by developing a novel text distance metric, the *time-sensitive semantic edit distance* (t-SED), which accounts for the temporal context across multiple traces. To evaluate this method we undertake a case study of Russian online-troll activity within US political discourse. The novel metric allows us to classify the social roles of trolls based on their traces, in this case tweets, into one of the predefined categories left-leaning, right-leaning, and news feed. We show the effectiveness of the t-SED metric to measure the similarities between tweets while accounting for the temporal context, and we use novel data visualisation techniques and qualitative analysis to uncover new empirical insights into Russian troll activity that have not been identified in previous work. Additionally, we highlight a connection with the field of Actor-Network Theory and the related hypotheses of Gabriel Tarde, and we discuss how social sequence analysis using t-SED may provide new avenues for tackling a long-

User identity via semantic edit distance: A case study of Russian trolls on Twitter

[Kim et al Jour. Comp. Social Science ‘19]

Russian Trolls dataset

[Linvin and Warren, 2018]

- User handles provided by Twitter to the House Intelligence Committee
- The most comprehensive empirical record of Russian troll activity on social media

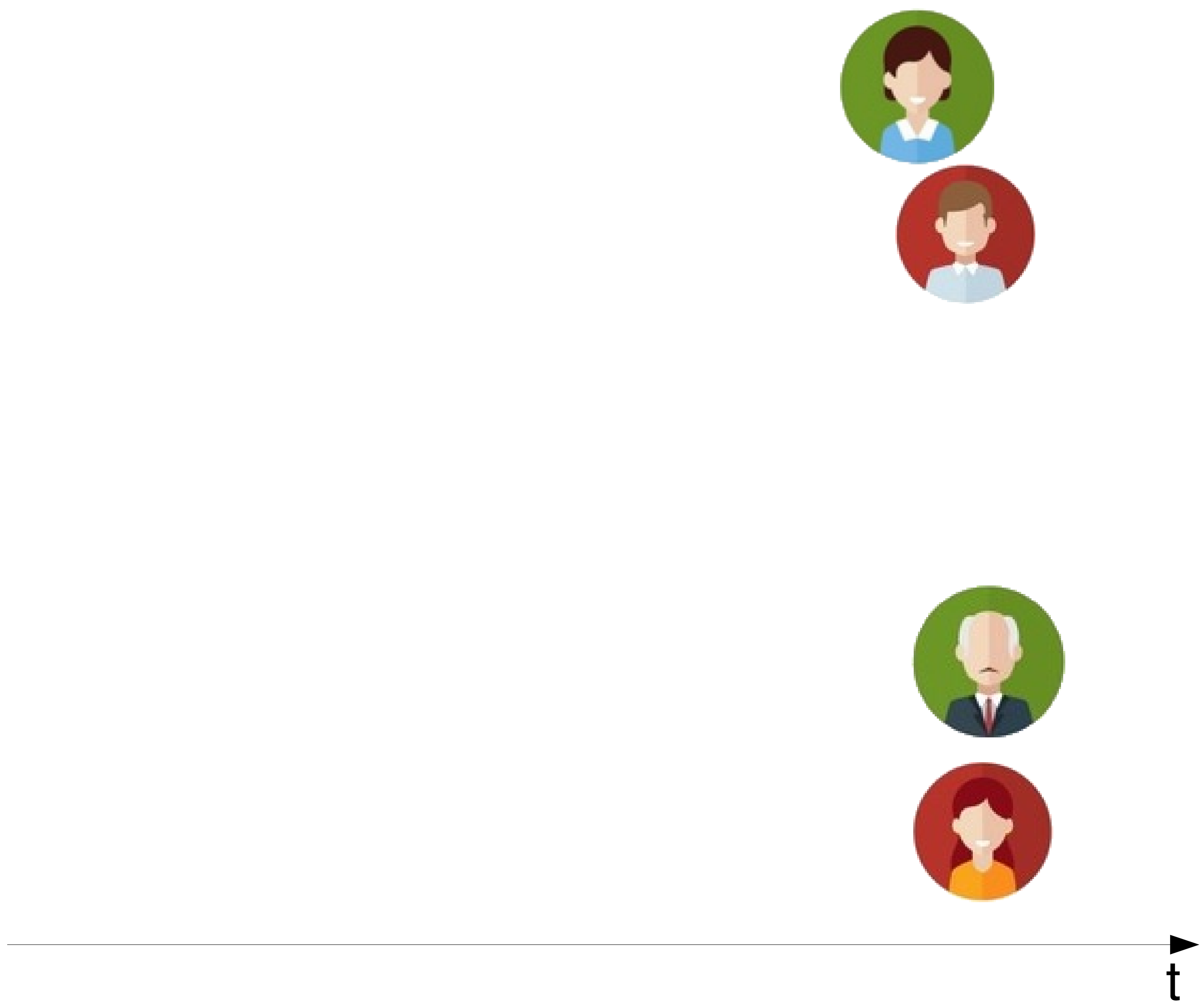
Dataset stats:

- length: **February 2012 and May 2018**
- #tweets: 3**M**
- #users: **2,848 handles**

5 roles:

right troll
news feed
left troll
hashtag gamer
fearmonger

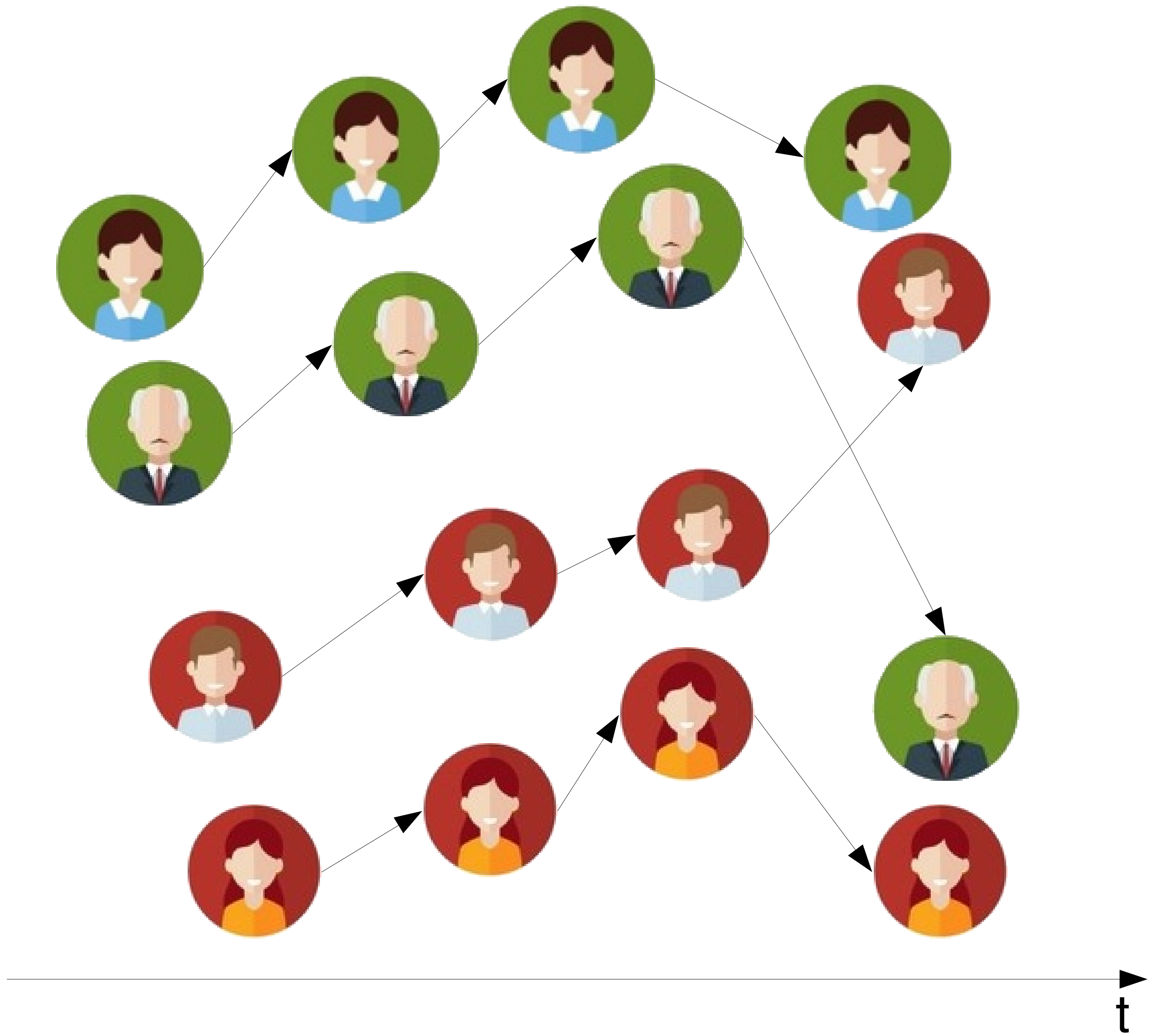
Identify troll via their online traces



Identity through the digital traces that actors leave behind

Identify troll via their online traces

Behavioral Data Science



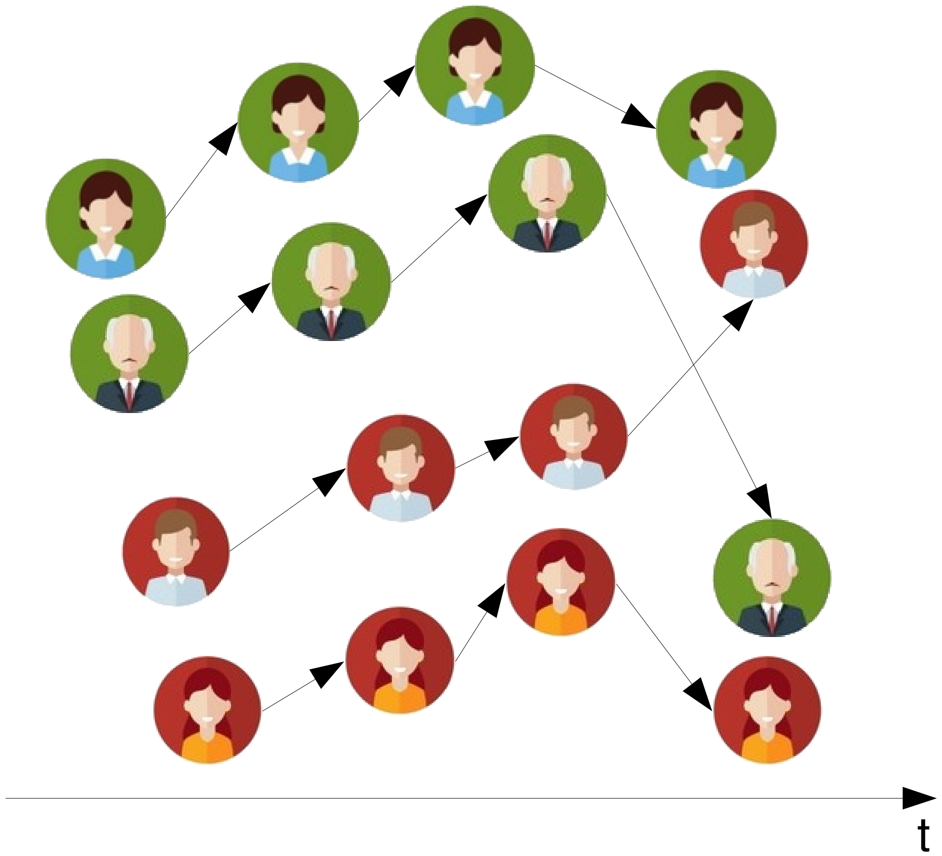
Identity through the digital traces that actors leave behind

Identify troll via their online traces

Behavioral Data Science

Semantic edit distance between two trajectories

$$D(i,j) = \mathrm{dist}(\boldsymbol{s}_i, \boldsymbol{s}_j) \times \exp(\theta |t_i - t_j|)$$



Properties:

- Increases with sequence similarity;
- Decreases with time-difference;
- Embeds semantics of text

Predict and explain troll strategy

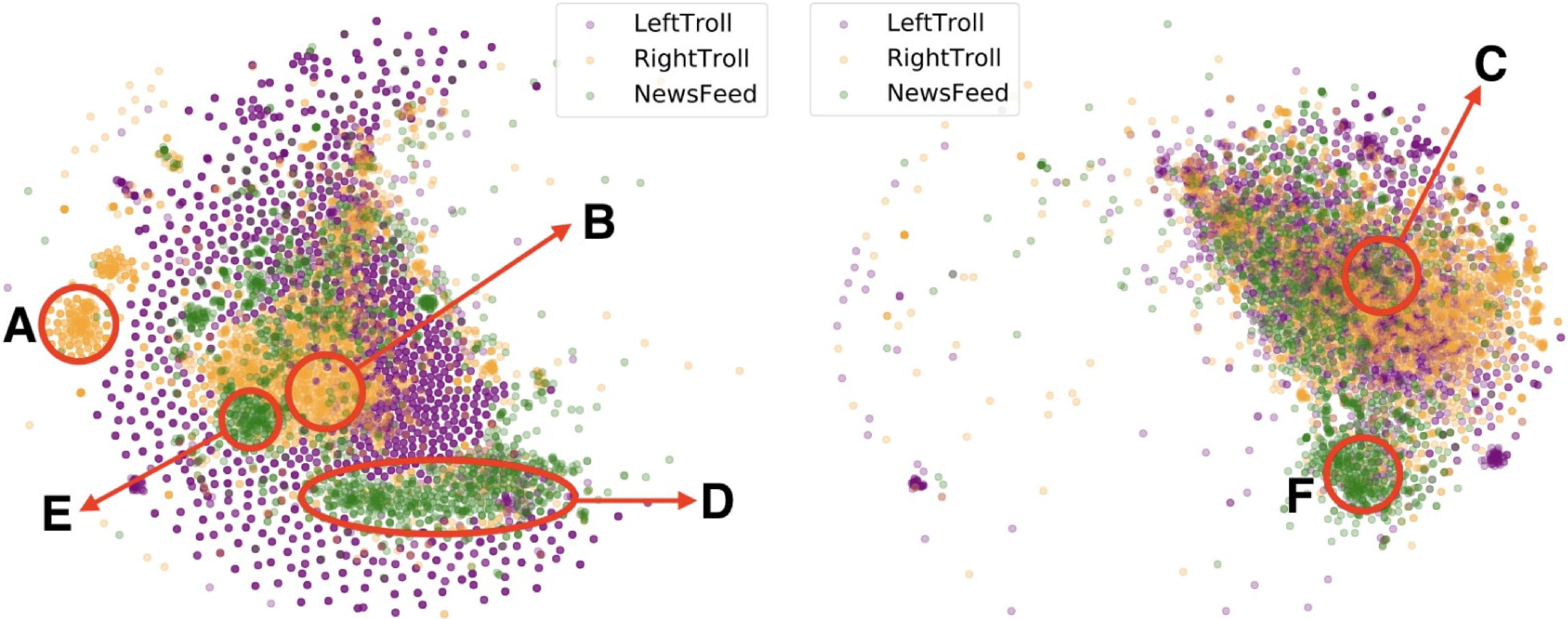
Behavioral Data Science

	Method	Micro F1 K	Micro F1 F1	Macro F1 K	Macro F1 F1
Baseline	LR	-	0.75	-	0.55
	ED	1	0.73	1	0.47
	Cosine	1	0.75	1	0.54
Semantic	SED	1	0.79	1	0.62
	SED/Max	6	0.68	1	0.39
	SED/ED	8	0.62	8	0.34
Temporal	t-LR	-	0.79	-	0.61
	t-ED	1	0.84	1	0.76
	t-Cosine	5	0.81	1	0.61
	t-SED	3	**0.86**	3	**0.78**

Distinguish/predict troll roles:
right troll, news feed, left troll

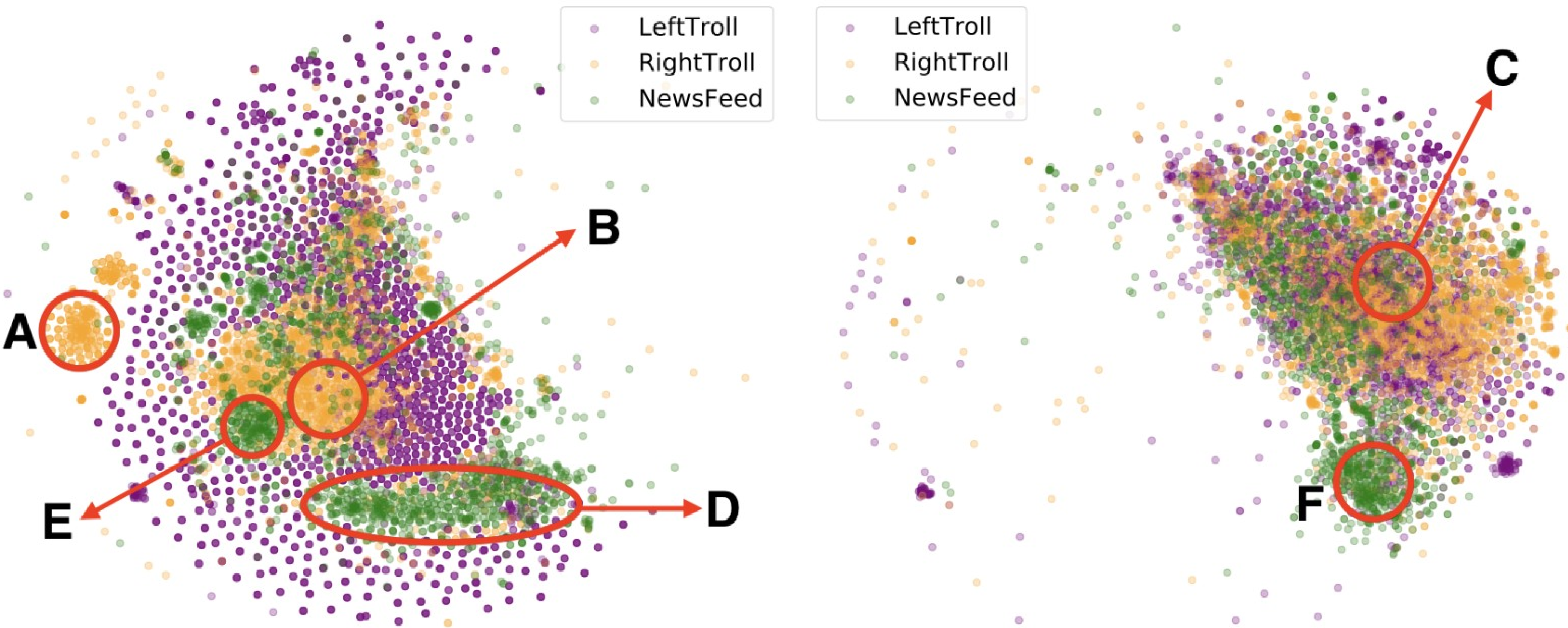
Predict and explain troll strategy

Behavioral Data Science



“Focused MAGA” right trolls, “diverse strategy” left trolls.

Predict and explain troll strategy



"Focused MAGA" right trolls, "diverse strategy" left trolls.

A – *(right trolls)* Hillary cannot be trusted *#ThingsMoreTrustedThanHillary*

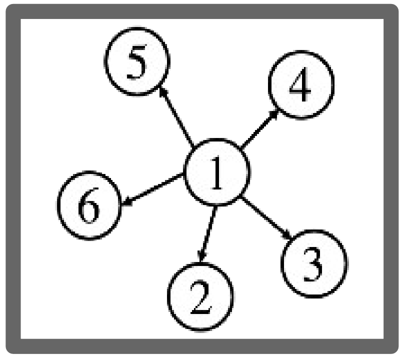
B – *(right trolls)* Mimic black Trump supporters *#Blacks4Trump*

C – *(all trolls)* Religious beliefs *#God #Prolife*

D, F – (news trolls) News about violence and civil unrest *#news*

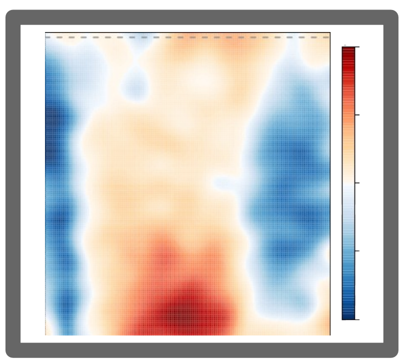
E – (news trolls) Federal politics, policy and regulation *#politics*

Summary

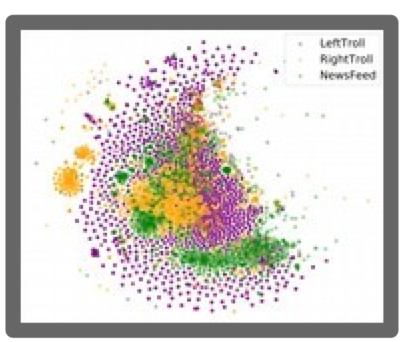


A scalable algorithm to estimate user influence from latent network structures

Three measures to quantify the influence, the political partisanship and botness of Twitter users



A detailed analysis of the role and influence of socialbots during the first U.S. Presidential debate.

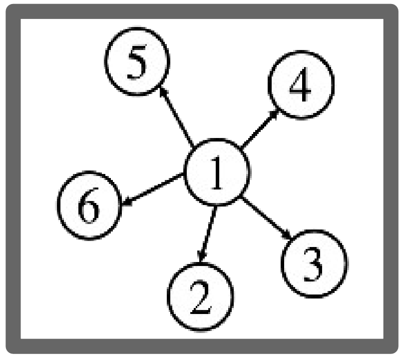


Predict and analyze the role of opinion manipulators (trolls) via semantic edit distance

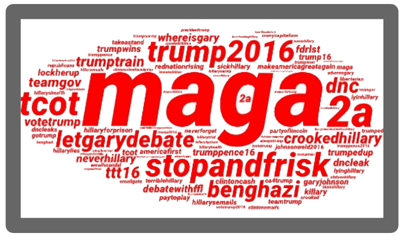
Thank you!

Behavioral Data Science

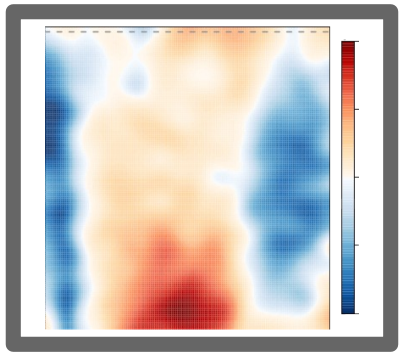
https://github.com/rohitram96/BirdSpotter



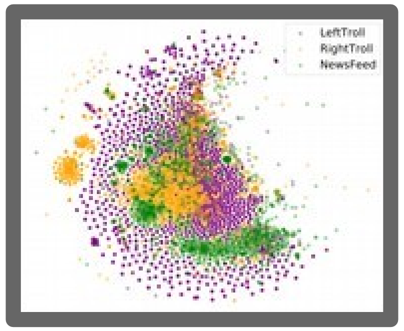
A scalable algorithm to estimate user influence from latent network structures



Three measures to quantify the influence, the political partisanship and botness of Twitter users

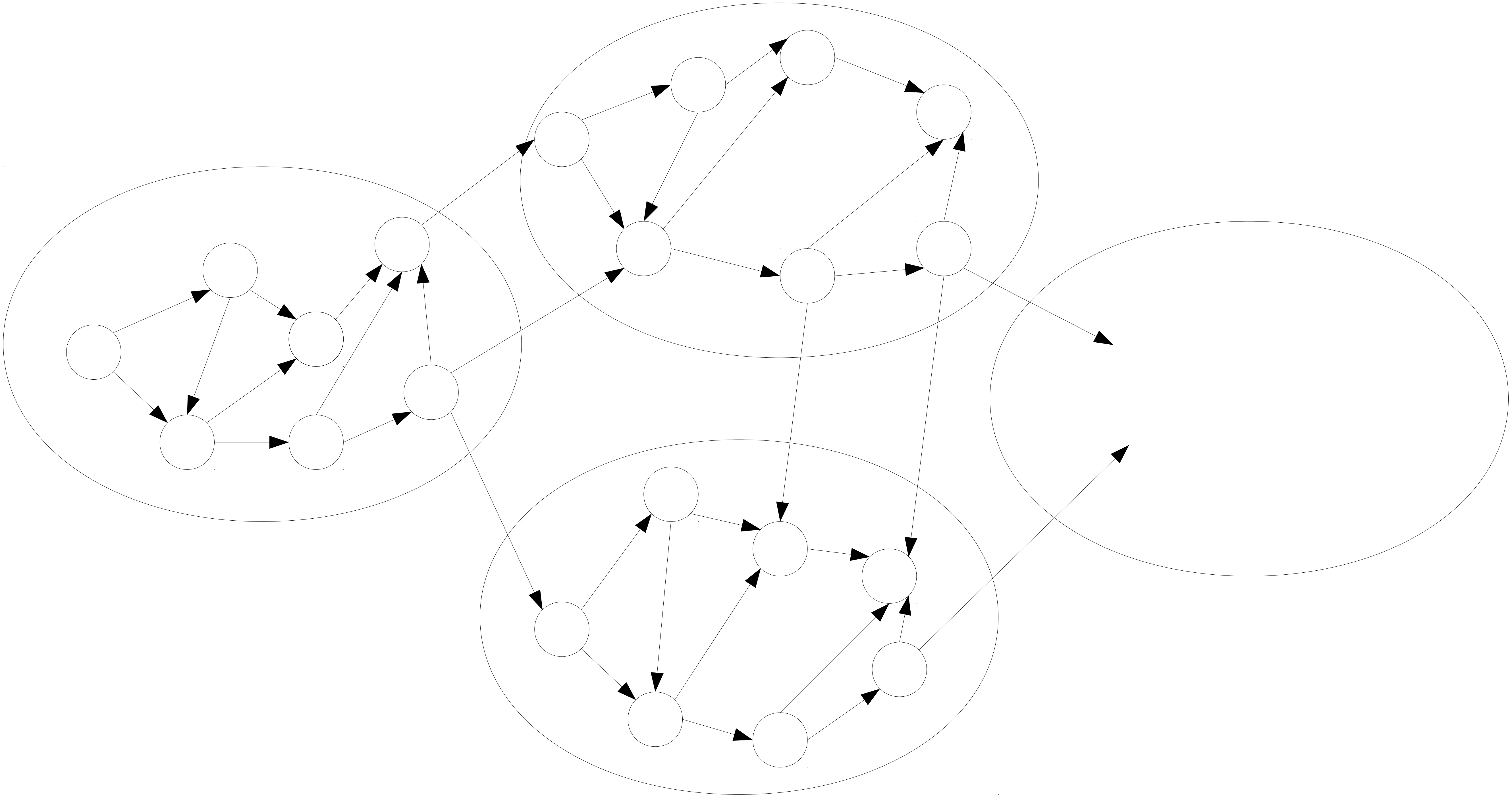


A detailed analysis of the role and influence of socialbots during the first U.S. Presidential debate.

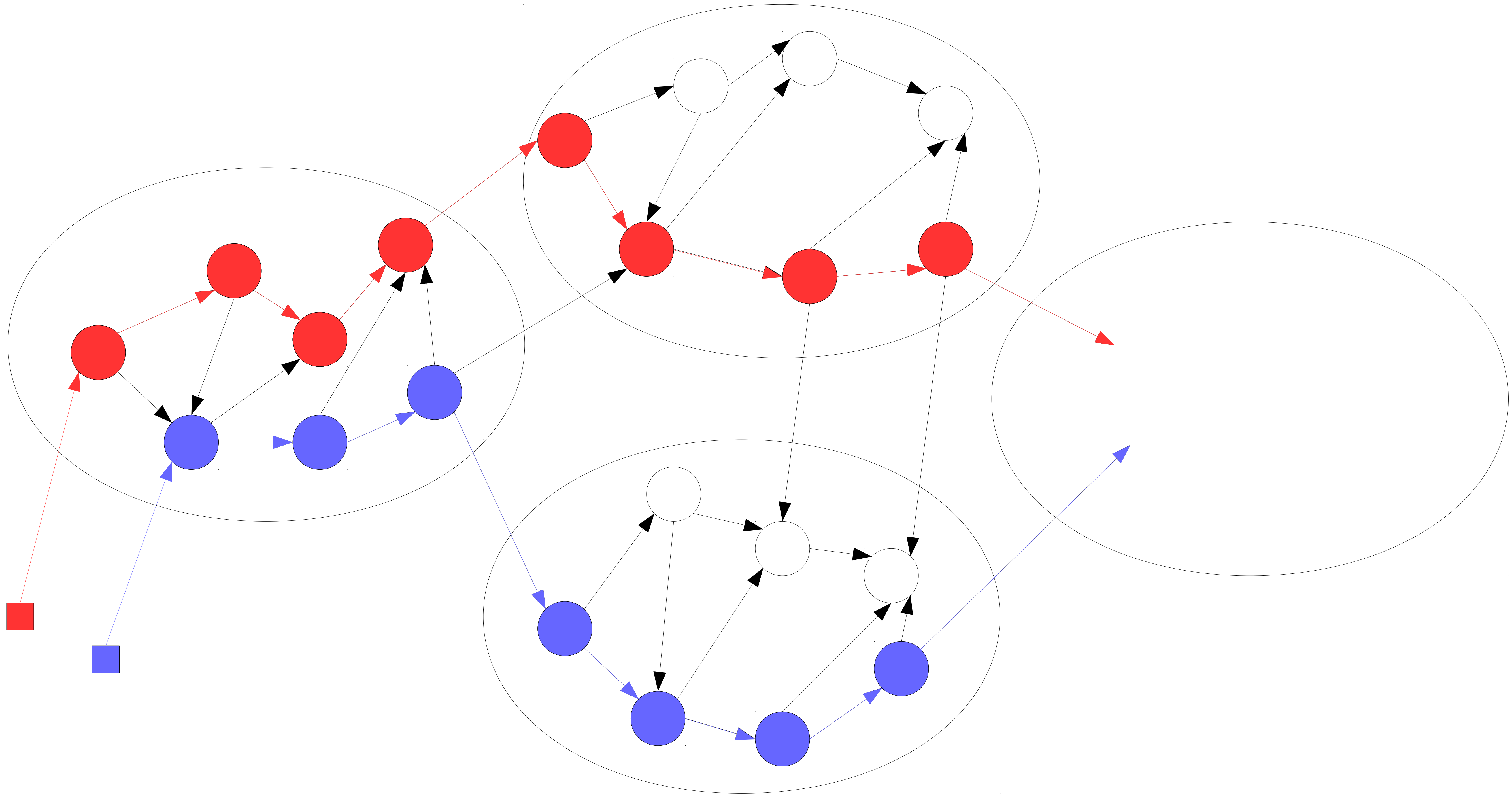


Predict and analyze the role of opinion manipulators (trolls) via semantic edit distance

Next steps:



Next steps:



- Complex contagion diffusion models with community structure;
- Estimate impact of spread of malicious content (total popularity, virality, affected communities)