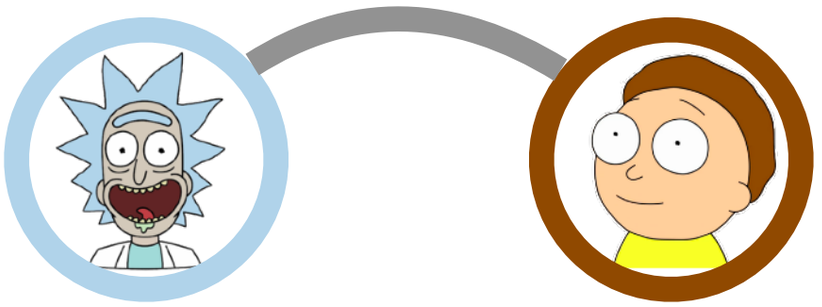
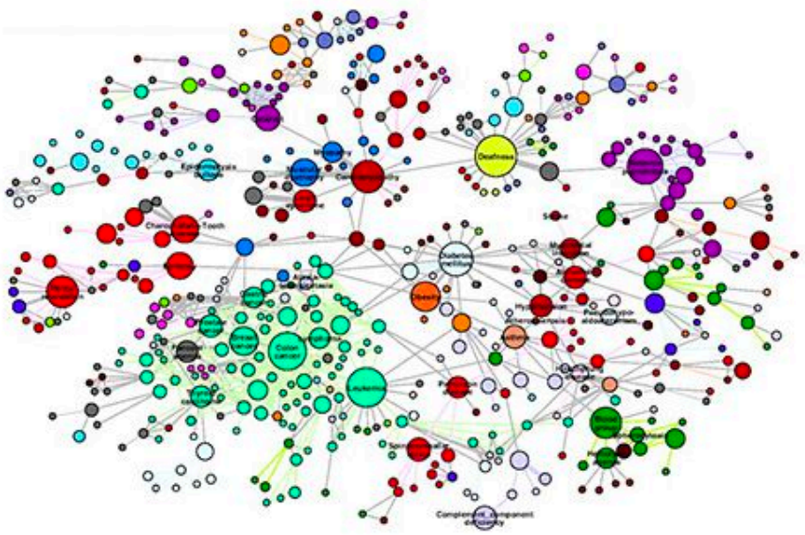
Identifying Networks of Social Bots on Twitter Conversations

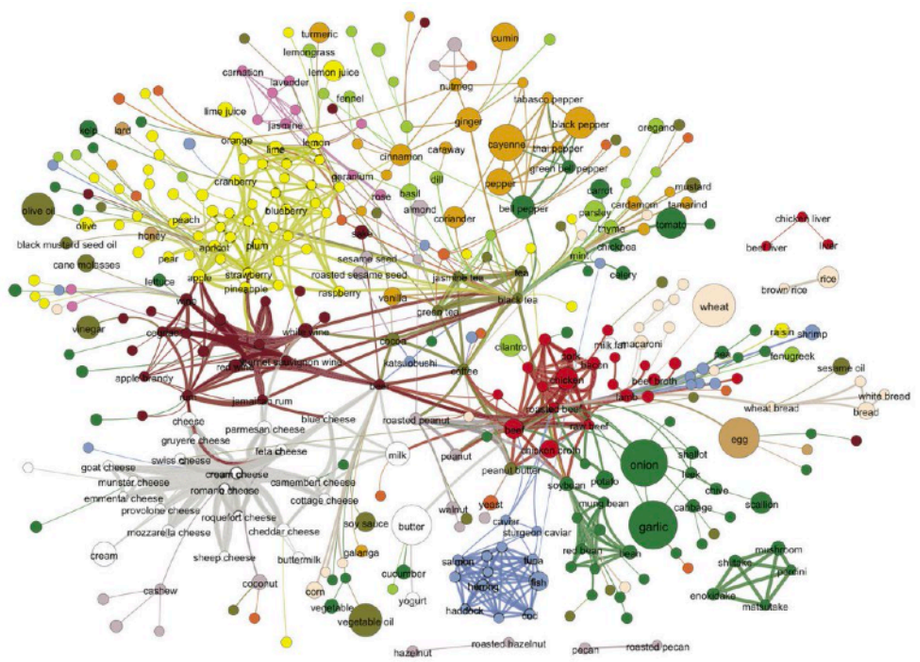
SICSS - Istanbul'20

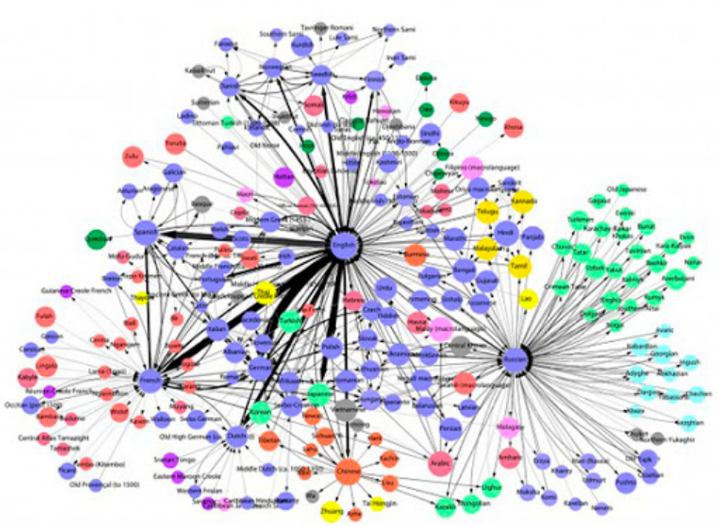
Onur Varol
Sabanci University
@onurvarol

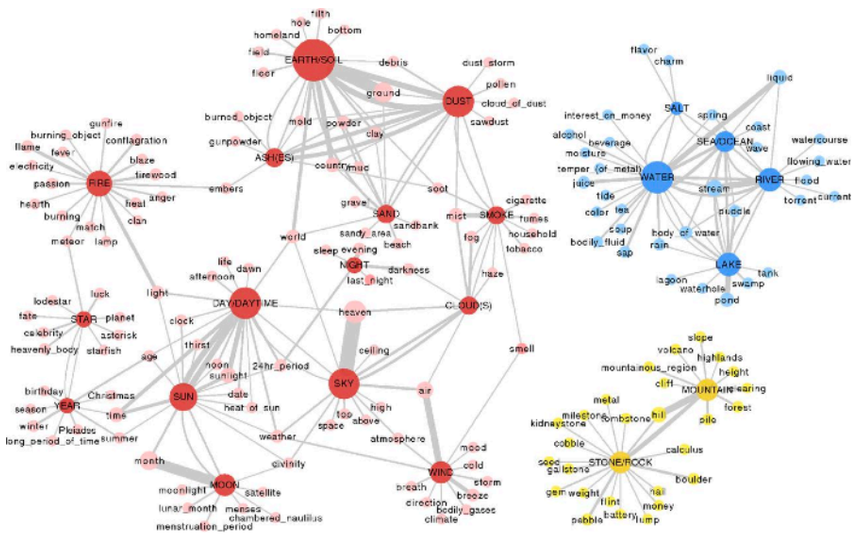
Why networks?

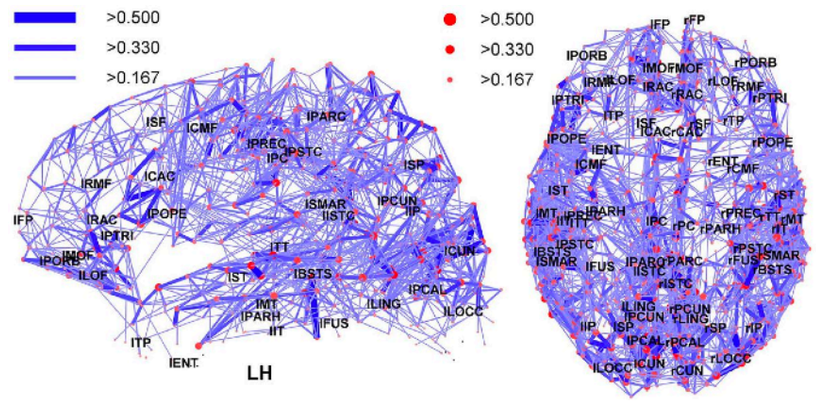












Online polarization and differences in the communication preferences

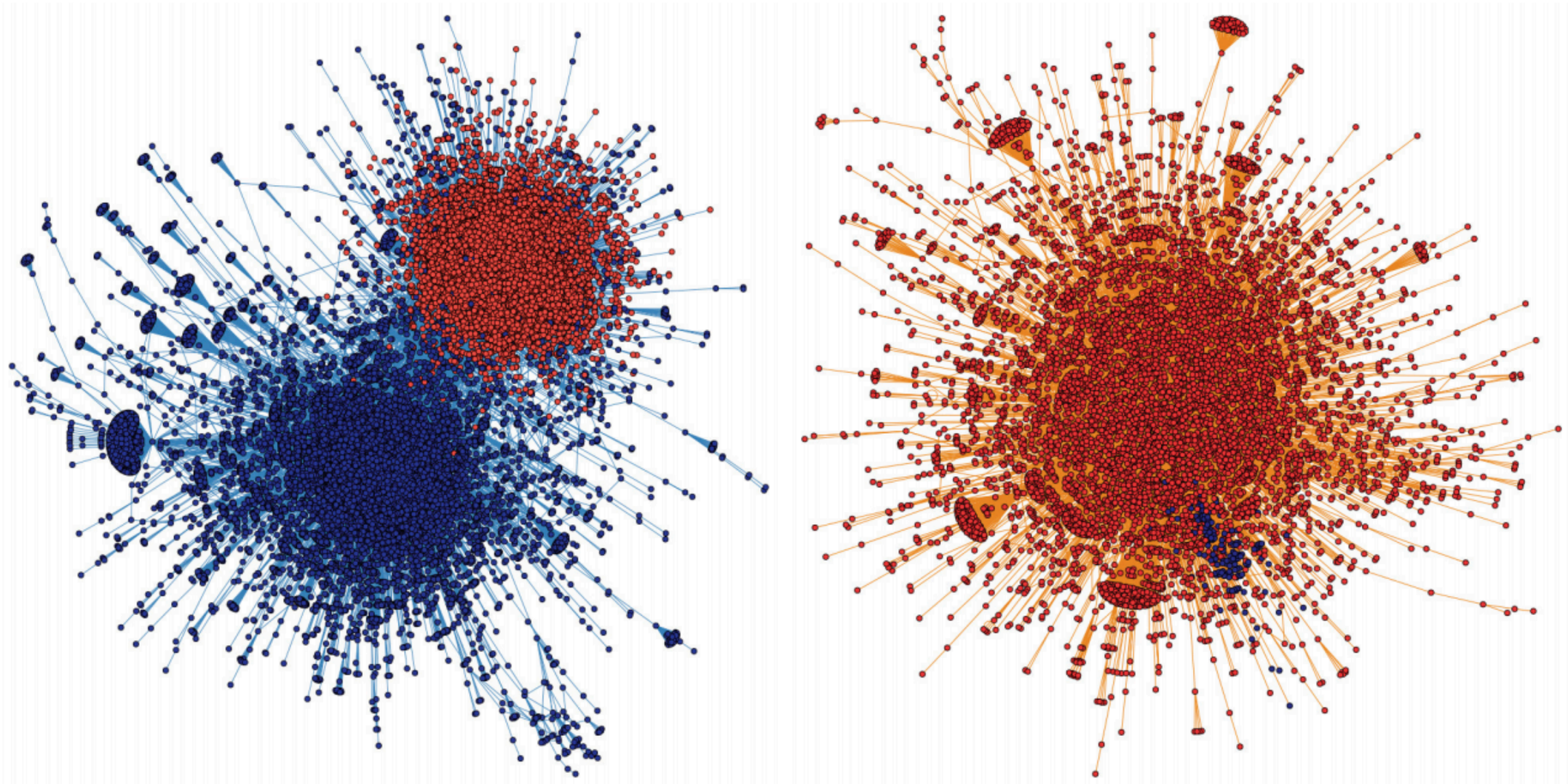


Figure 1: The political retweet (left) and mention (right) networks, laid out using a force-directed algorithm. Node colors reflect cluster assignments (see § 3.1). Community structure is evident in the retweet network, but less so in the mention network. We show in § 3.3 that in the retweet network, the red cluster A is made of 93% right-leaning users, while the blue cluster B is made of 80% left-leaning users.

Conover, M.D., Ratkiewicz, J., Francisco, M., Gonçalves, B., Menczer, F. and Flammini, A., 2011, July. **Political polarization on Twitter**. In Fifth *International AAAI conference on weblogs and social media (ICWSM)*.

Public speeches

Abraham Lincoln
1860

First American president on radio

Warren G. Harding
1922

First political advertisement aired on TV

Dwight D. Eisenhower
1952

1650

1837

1876

1902

1927

1947
Harry S. Truman

First presidential speech on TV

1960
Kennedy & Nixon

First presidential debate on TV

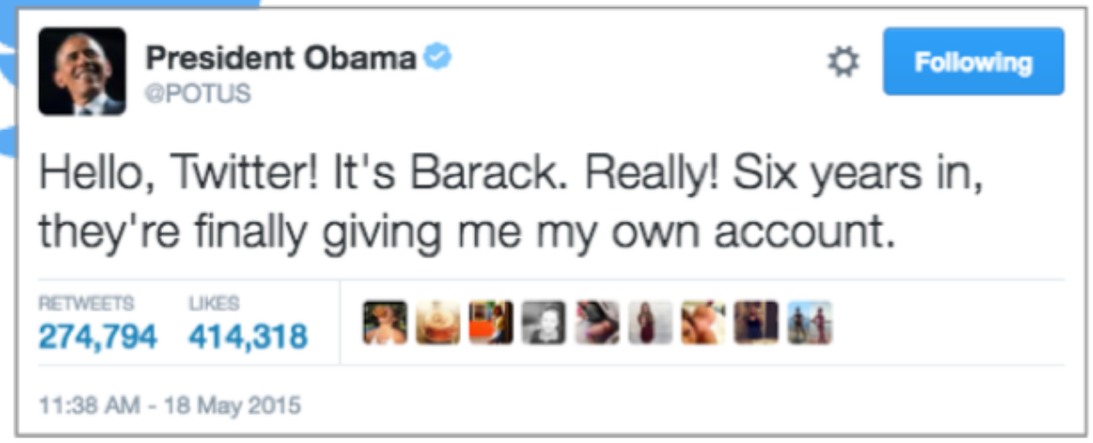
I am Barack Obama, President of the United States -- AMA

POLITICS submitted 4 years ago * by PresidentObama Obama

Hi, I'm Barack Obama, President of the United States. Ask me anything. I'll be taking your questions for half an hour starting at about 4:30 ET.

Proof it's me:
https://twitter.com/BarackObama/status/240903767350968320

2012



President Obama
@POTUS

Following

Hello, Twitter! It's Barack. Really! Six years in, they're finally giving me my own account.

RETWEETS 274,794 LIKES 414,318

11:38 AM - 18 May 2015

2015

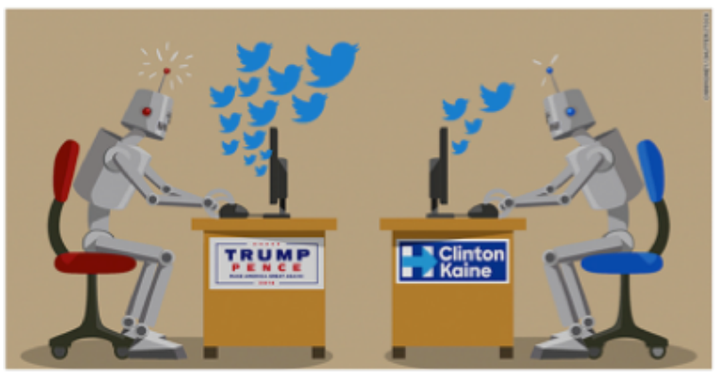
1988

2004

It's Election Day

Share that you're voting in the U.S. Election and find out where to vote.

I'm a Voter More Information



2016

Hillary Clinton
@HillaryClinton

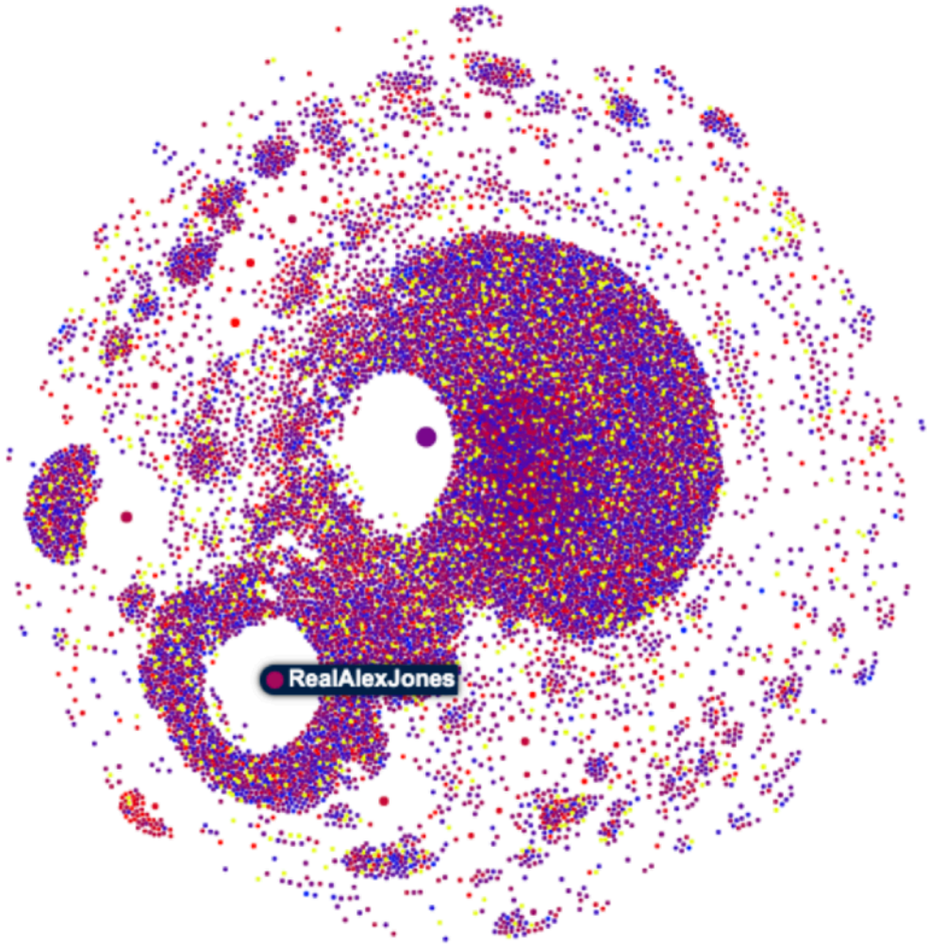
Following

Delete your account.

Donald J. Trump @realDonaldTrump
Obama just endorsed Crooked Hillary. He wants four more years of Obama—but nobody else does!

RETWEETS 507,017 LIKES 671,876

2:27 PM - 9 Jun 2016



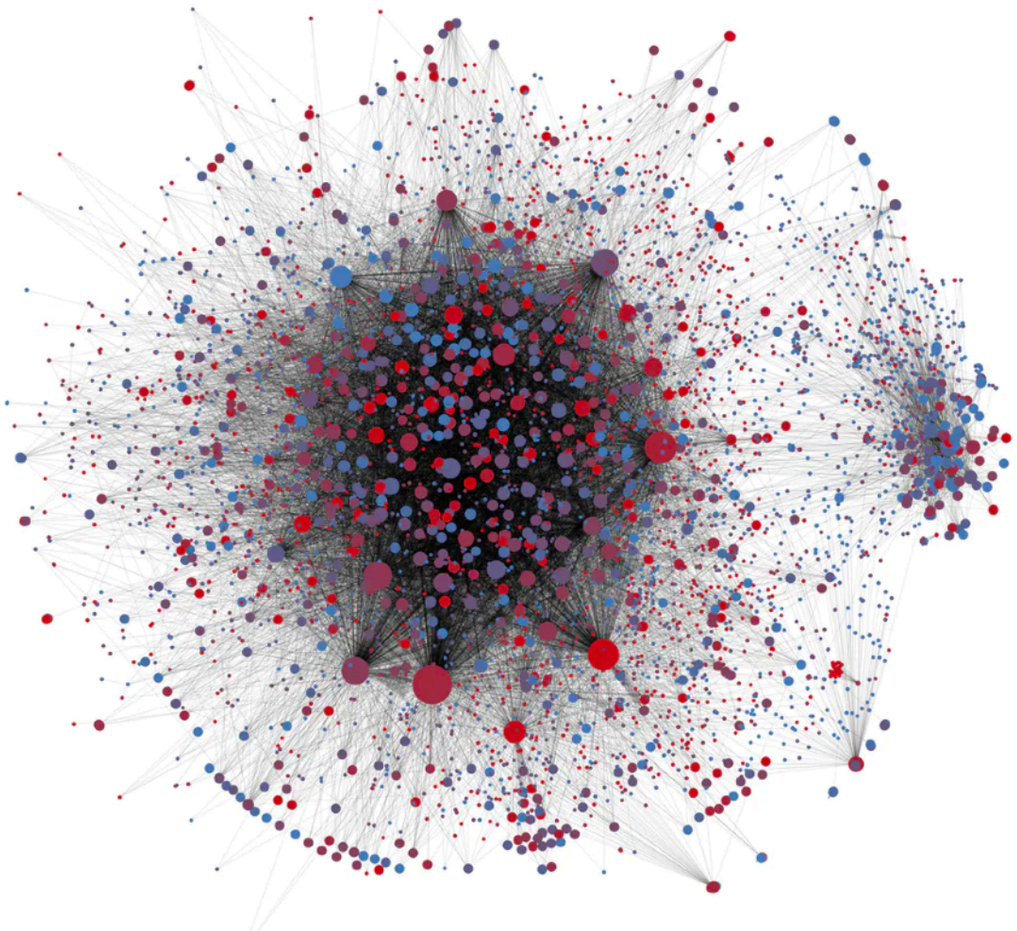
The spread of a fake news on Twitter

Network visualization of highly viral fabricated news report titled “Spirit cooking”: Clinton campaign chairman practices bizarre occult ritual, published 4 days before the 2016 US election and shared in over 30,000 tweets.

Node colors indicate Botometer score and the size of the nodes corresponds to number of retweets account received.

Interactive visualization of this network available
https://iunetsci.github.io/HoaxyBots/

Shao, Chengcheng, Giovanni Luca Ciampaglia, Onur Varol, Kai-Cheng Yang, Alessandro Flammini, and Filippo Menczer. "**The spread of low-credibility content by social bots**." Nature Communications 9, no. 1 (2018): 1-9.



The spread of anti-vaccine messages

How bots are used to affect, and possibly manipulate, the online debate about vaccination policy. It is the retweet network for the #SB277 hashtag, about a recent California law on vaccination requirements and exemptions.

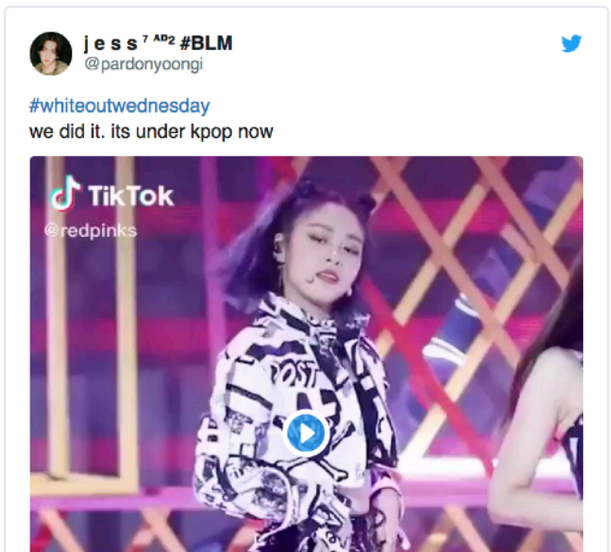
Nodes represent Twitter users, and links show how information spreads among users. The node size represents influence (times a user is retweeted), the color represents bot scores.

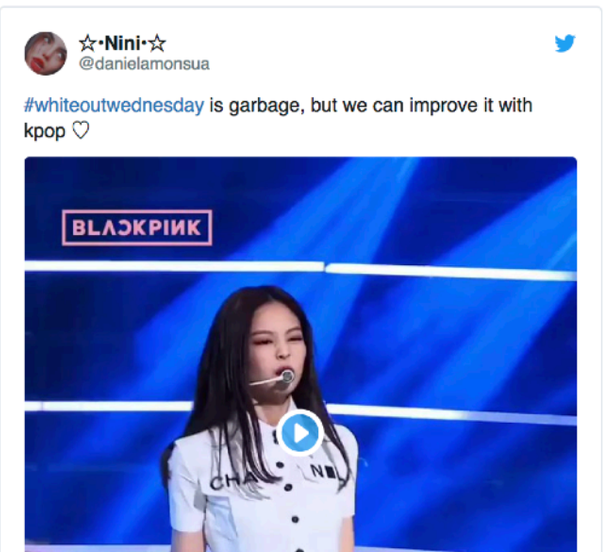
Ferrara, Emilio, Onur Varol, Clayton Davis, Filippo Menczer, and Alessandro Flammini. "**The rise of social bots**." *Communications of the ACM* 59, no. 7 (2016): 96-104.

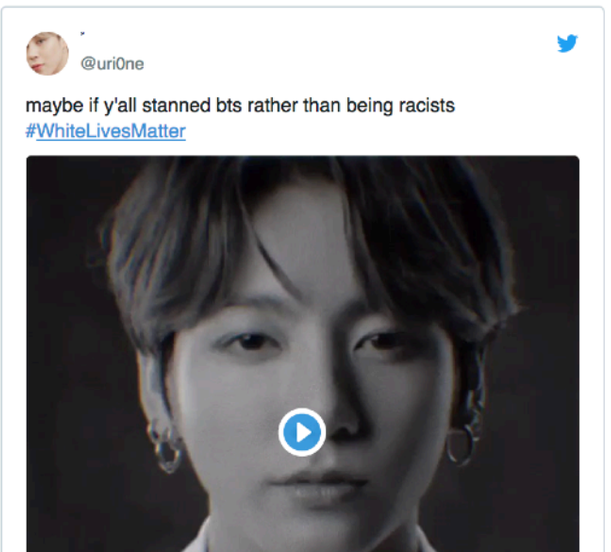
BTS Army fans take over racist hashtags

Bots are known for their massive automated activities to flood organic conversations. For the first time an account like @bts_twt mobilize their 26.5M followers to post content using racist hashtags to make those conversation channels useless.

This is an interesting example of information warfare.







Submit | @ | rterdogan

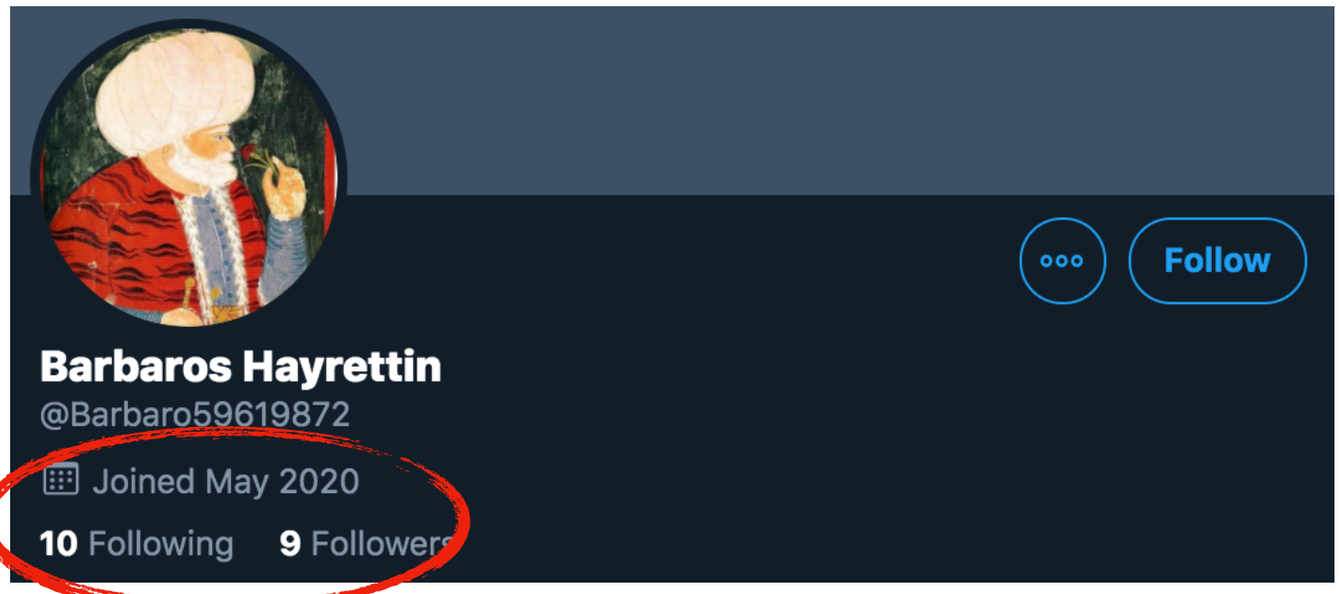
Save Image

Compare with

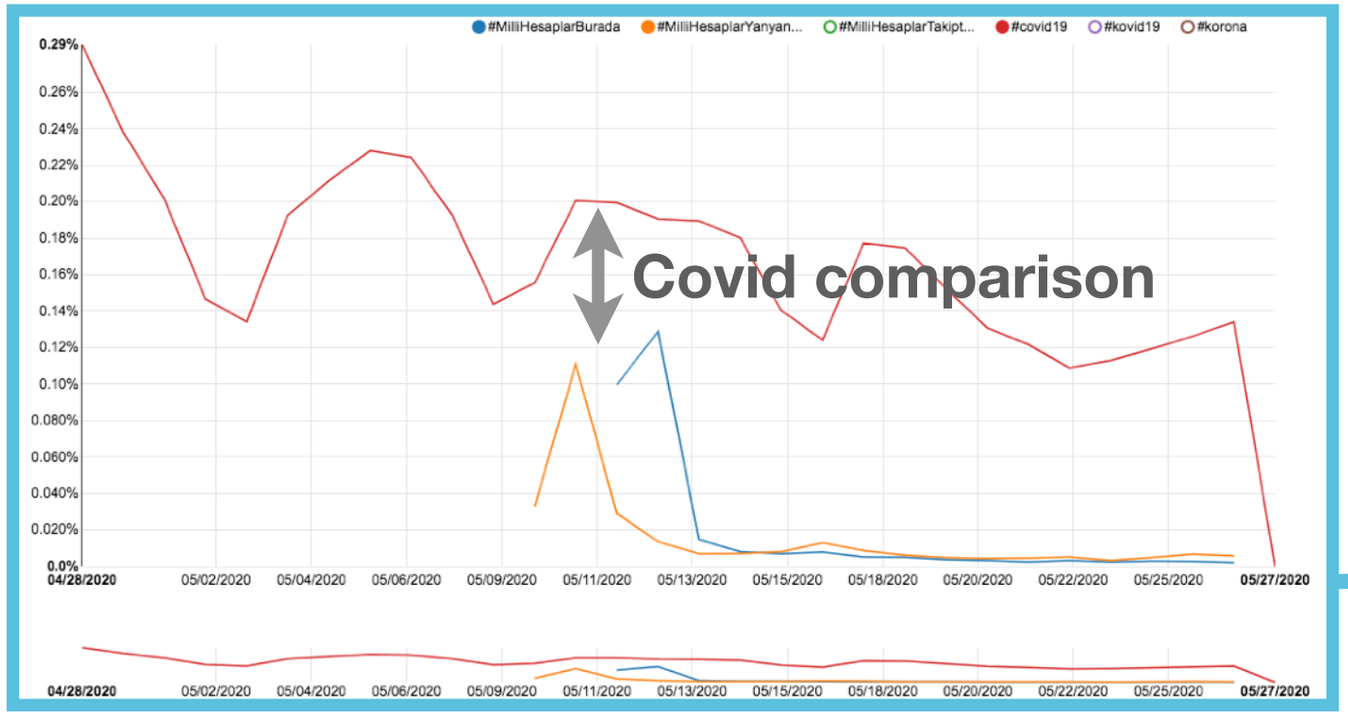
Anything @bts_twt send gets nearly 500k retweets.

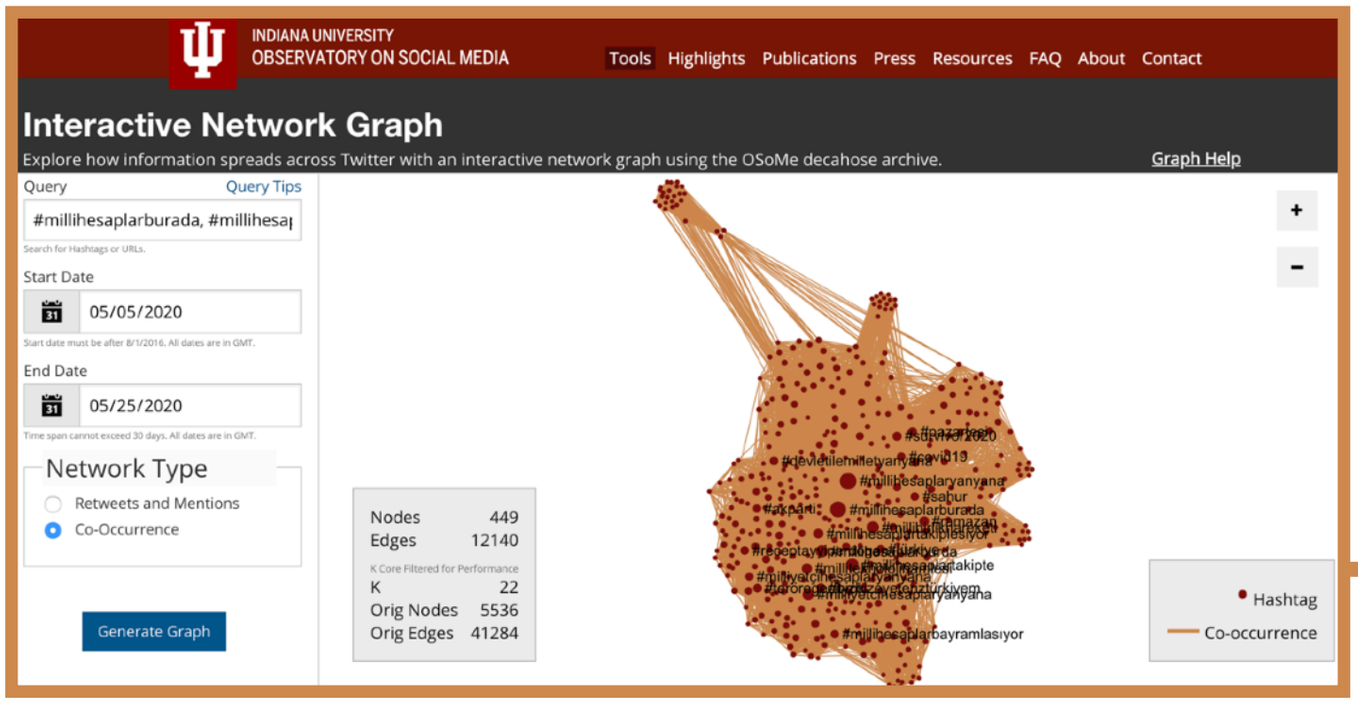
Demo: Select an interesting content to analyze

Let’s take a look at a hashtag that we might be interested in analyzing. I selected “**MilliHesaplarBurada**” since there has been recent activity on Twitter.



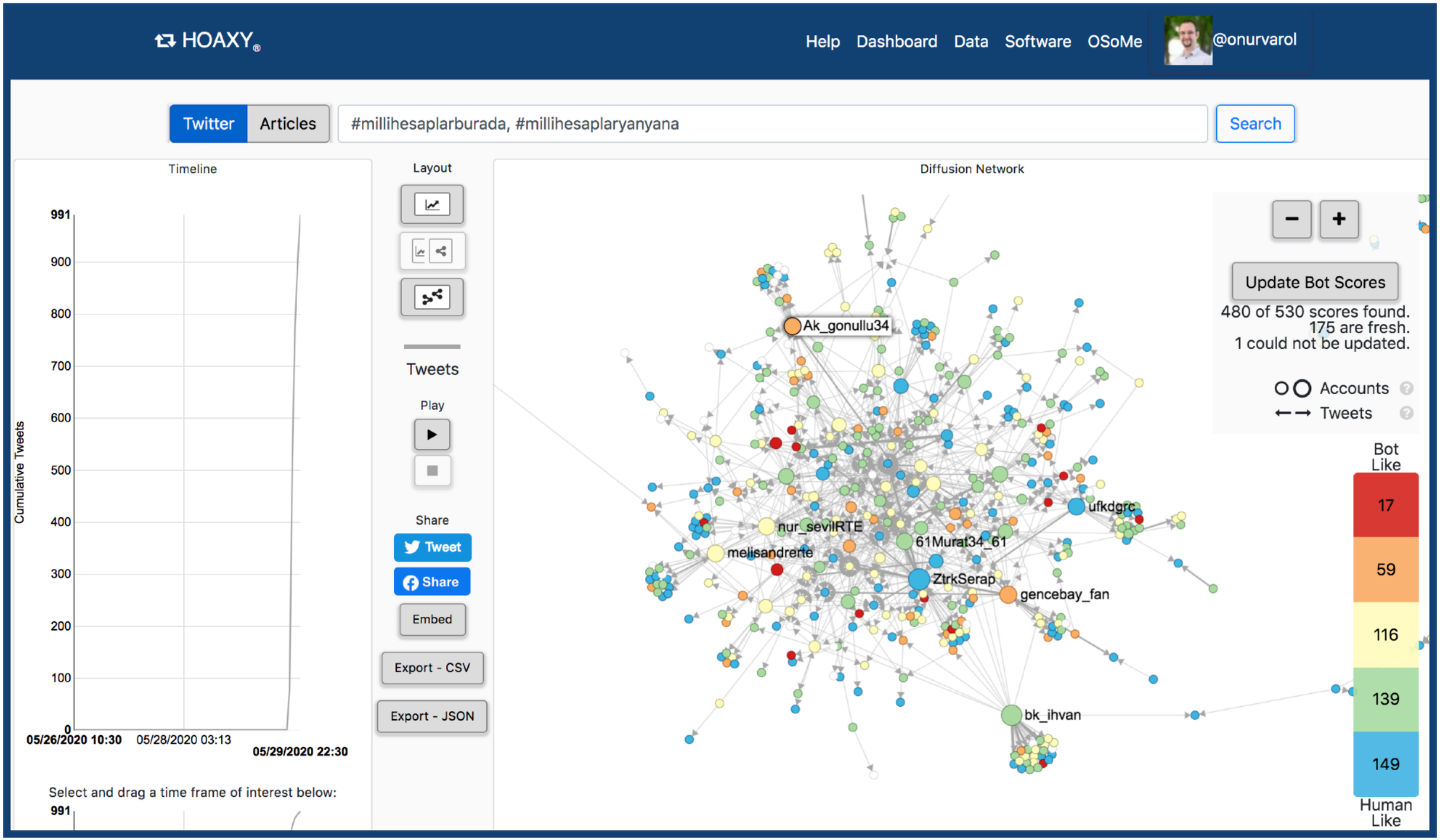
Demo: Preliminary data exploration on OSoMe





Davis, Clayton A., et al. "**OSoMe: the IUNI observatory on social media.**" *PeerJ Computer Science* 2 (2016): e87.

Demo: Look at diffusion patterns on Hoaxy



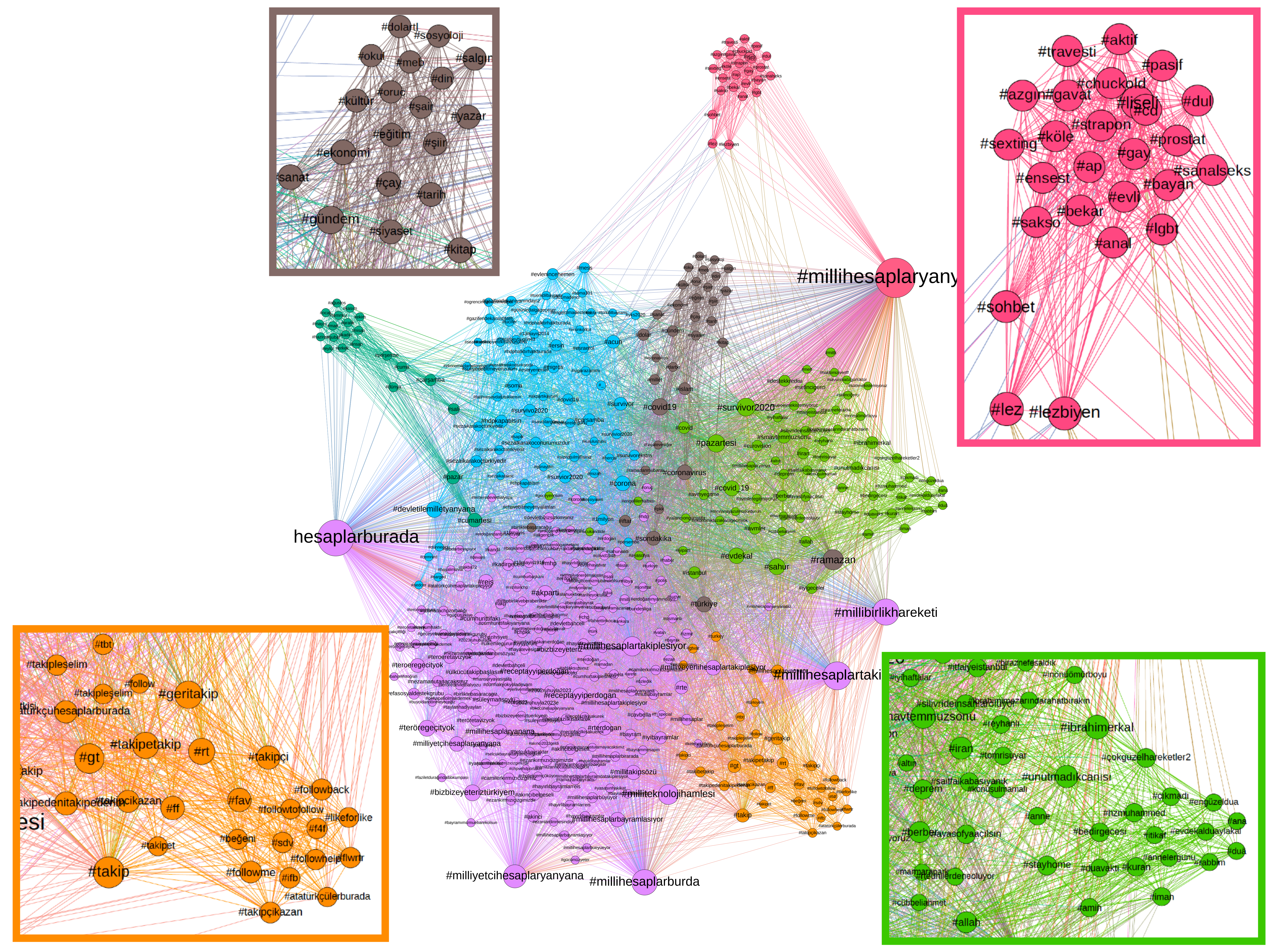
Shao, Chengcheng, et al. "**Hoaxy: A platform for tracking online misinformation.**" *Proc. of the 25th Intl. Conf. Companion on World Wide Web*. 2016.

Demo: Preliminary data analysis using OSoMe export

Let’s take a look at

1_analyze-hashtag-cooccurrence.ipynb

notebook for the analysis



#dolartl
#sosyoloji
#okul
#meb
#salgın
#din
#oruç
#kültür
#şair
#yazar
#eğitim
#şiir
#ekonomi
#sanat
#çay
#tarih
#gündem
#siyaset
#kitap
#travesti
#aktif
#pasif
#chuckold
#azgın
#gavat
#liseli
#cd
#dul
#strapon
#sexting
#köle
#prostat
#gay
#ap
#sanalseks
#ensest
#bayan
#evli
#bekar
#sakso
#lgbt
#anal
#sohbet
#lez
#lezbiyen
#millihesaplaryany
hesaplarburada
#millibirlikhareketi
#millihesaplartaki
#milliyetcihesaplaryanyana
#millihesaplarburda
#milliteknolojihamlesi
#millihesaplartakiplesiyor
#receptayyiperdogan
#bizbizeyeteriztürkiyem
#teröregeçityok
#millihesaplaryanana
#milliyetçihesaplaryanyana
#millihesaplarbayramlasıyor
#akparti
#cumhurittifakı
#survivor2020
#pazartesi
#coronavirus
#covid19
#covid_19
#corona
#ramazan
#sahur
#evdekal
#türkiye
#sondakika
#acun
#survivor
#devletilemilletyanana
#cumartesi
#pazar
#çarşamba
#tbt
#takipleselim
#follow
#geritakip
#takipleşelim
#atatürkçühesaplarburada
#gt
#takipetakip
#rt
#takipçi
#followback
#takipedenitakipedenim
#takipcikazan
#ff
#fav
#followtofollow
#likeforlike
#f4f
#takipet
#beğeni
#sdv
#followhelp
#flwrtr
#takip
#followme
#ifb
#atatürkçülerburada
#takipçikazan
#itfaiyeistanbul
#biraznefesaldık
#inönüömürboyu
#iyihaftalar
#silivrideinsanlarölüyor
#ibrahimimezarındarahatbırakın
#sınavtemmuzsonu
#reyhanlı
#ibrahimerkal
#iran
#tomrisuyar
#çokgüzelhareketler2
#altın
#unutmadıkcanısı
#saitfaikabasıyanık
#deprem
#konuşulmamalı
#çıkmadı
#engüzeldua
#anne
#hzmuhammed
#ana
#berber
#ayasofyaaçılsın
#bedirgecesi
#itikaf
#evdekalduaylakal
#duâ
#annelergunu
#stayhome
#duavakti
#kuran
#rabbim
#marmarapark
#iman
#cübbeliahmet
#amin
#allah

	degree_N	degree_C	closeness_C	betweenness_C	pagerank_N	clustering_N	clique_N	core_N	onion_N
#millihesaplaryanyana	428	0.955357	0.957265	0.270459	0.020445	0.095395	27	27	57
#millihesaplarburada	382	0.852679	0.871595	0.157603	0.017969	0.106100	24	27	57
#millihesaplartakipte	276	0.616071	0.722581	0.046616	0.012718	0.146561	24	27	57
#millihesaplarburda	244	0.544643	0.687117	0.032099	0.011220	0.164069	19	27	57
#milliyetcihesaplaryanyana	213	0.475446	0.655930	0.022223	0.009810	0.184782	22	27	57
#millihesaplartakiplesiyor	164	0.366071	0.612022	0.011403	0.007525	0.221832	19	27	57
#milliveyerlihesaplartakiplesiyor	106	0.236607	0.567089	0.004340	0.004931	0.288230	19	27	56
#millihesaplaryanana	103	0.229911	0.564943	0.003467	0.004791	0.302494	19	27	56
#milliyetçihesaplaryanyana	103	0.229911	0.564943	0.003758	0.004846	0.309918	19	27	56
#millihesaplarbayramlasıyor	101	0.225446	0.563522	0.002991	0.004703	0.327327	19	27	56
#millihesaplaryanyan	74	0.165179	0.545012	0.002133	0.003562	0.370603	24	27	52
#millihesaplartakipleşiyor	67	0.149554	0.540410	0.001100	0.003161	0.425599	19	27	56
#millihesaplar	53	0.118304	0.531435	0.000635	0.002581	0.476778	16	27	55
#millihesaplarbüyüyor	47	0.104911	0.527680	0.000359	0.002325	0.541166	17	27	53
#millihesaplarbirarada	38	0.084821	0.522145	0.000223	0.001915	0.559033	16	27	51
#millihesaplarbayramlaşıyor	32	0.071429	0.518519	0.000101	0.001679	0.697581	16	26	42
#millihesaplarbayramdatakiplesiyor	30	0.066964	0.517321	0.000073	0.001565	0.708046	17	26	41
#millihesaplaryanya	26	0.058036	0.514943	0.000078	0.001422	0.686154	14	23	24
#millihesaplarbueyueyor	25	0.055804	0.514351	0.000035	0.001353	0.790000	15	24	25
#yerlimillihesaplaryanyana	24	0.053571	0.513761	0.000103	0.001351	0.579710	11	21	12
#millihesaplaryanyanabiz	15	0.033482	0.508513	0.000026	0.000959	0.761905	12	15	1

Demo: Collect content by dehydrating from Tweet IDs

Let’s take a look at

2_dehydrate-tweetids.ipynb

notebook for the analysis

Botometer FAQ API Publications Bot Repo @onurvarol

Botometer®

An OSoMe project (bot•o•meter)

Botometer (formerly BotOrNot) checks the activity of a Twitter account and gives it a score based on how likely the account is to be a bot. Higher scores are more bot-like.

Note the following:

- Use of this service requires authenticating with Twitter. For more, see this FAQ.
- Botometer often categorizes "organizational accounts", like @BarackObama, as bot accounts.
- If something's not working, please let us know.

Botometer is a joint project of the Indiana University Network Science Institute (IUNI) and the Center for Complex Networks and Systems Research (CNetS).

Interested in using Botometer in your application? Check out our API.

onurvarol | Check user | Check followers | Check friends

Please help us improve Botometer by participating a short, anonymous survey. Go to the survey

@onurvarol 0.5 / 5

English-specific features		Language-independent features		Bot score based on	
Content:	1.1	Friend:	0.7	All features:	0.5
Sentiment:	1.0	Network:	0.5	Language-independent:	0.4
		Temporal:	0.4	User profile language:	en
		User:	1.5	Complete Automation Probability:	0%

Profile | Tweet | Details | Feedback

Export data

Feature extraction and datasets

Botometer system trained on using extensive dataset of bot behaviors.

We extract over 1,150 features capturing six subclasses:

- User
- Network
- Friends
- Temporal
- Content
- Sentiment

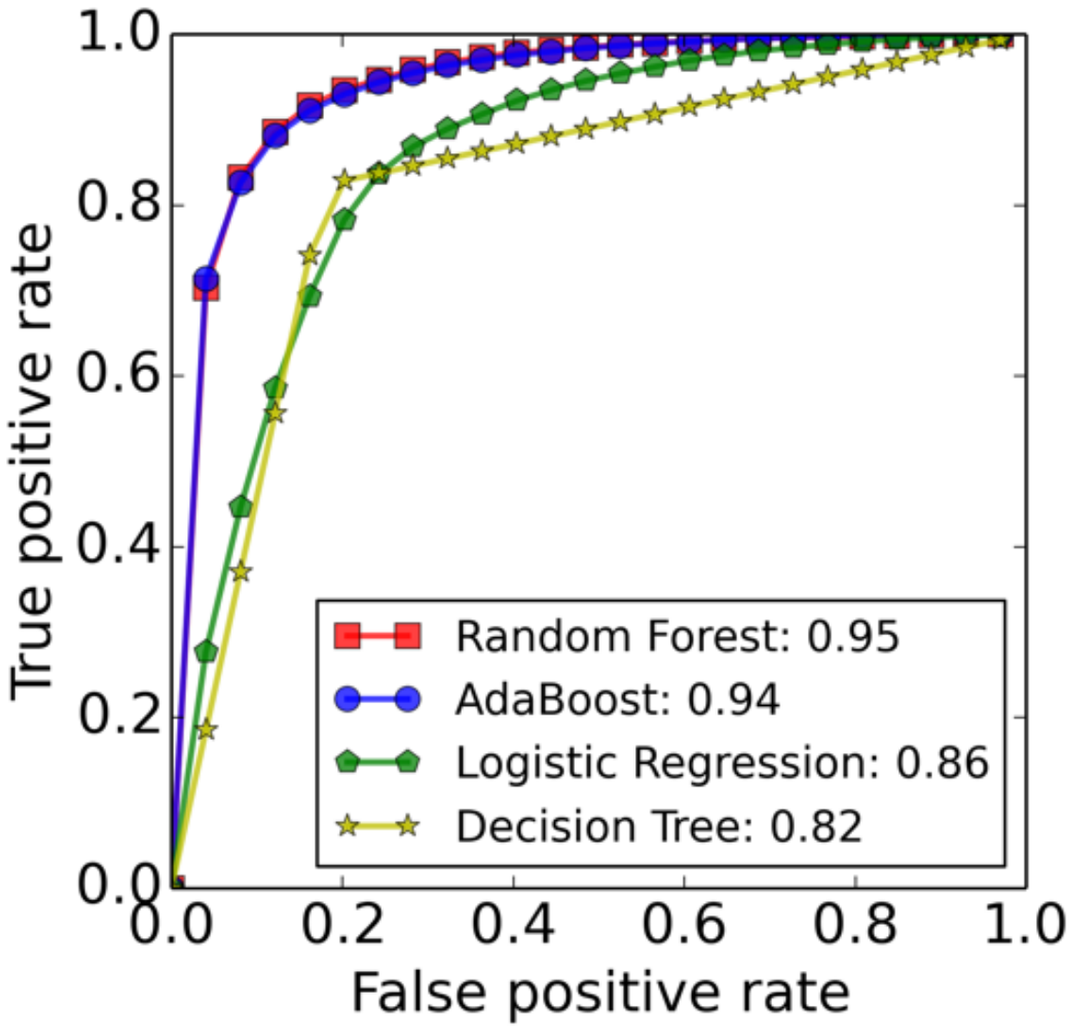
Features designed to model account behaviors more generic than human-bot difference and can be applied for different classifications problems.

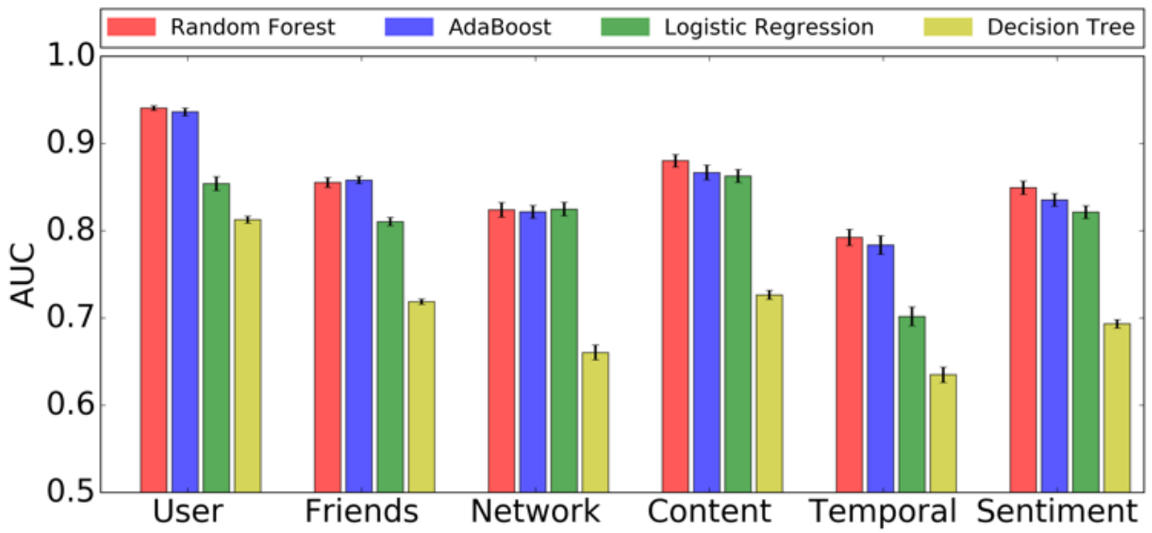
Table 1: List of 1150 features extracted by our framework.

User meta-data	Screen name length	**Sentiment**	(***) Happiness scores of aggregated tweets
	Number of digits in screen name		(***) Valence scores of aggregated tweets
	User name length		(***) Arousal scores of aggregated tweets
	Time offset (sec.)		(***) Dominance scores of single tweets
	Default profile (binary)		(*) Happiness score of single tweets
	Default picture (binary)		(*) Valence score of single tweets
	Account age (days)		(*) Arousal score of single tweets
	Number of unique profile descriptions		(*) Dominance score of single tweets
	(*) Profile description lengths		(*) Polarization score of single tweets
	(*) Number of friends distribution		(*) Entropy of polarization scores of single tweets
	(*) Number of followers distribution		(*) Positive emoticons entropy of single tweets
	(*) Number of favorites distribution		(*) Negative emoticons entropy of single tweets
	Number of friends (signal-noise ratio and rel. change)		(*) Emoticons entropy of single tweets
	Number of followers (signal-noise ratio and rel. change)		(*) Positive and negative score ratio of single tweets
	Number of favorites (signal-noise ratio and rel. change)		(*) Number of positive emoticons in single tweets
	Number of tweets (per hour and total)		(*) Number of negative emoticons in single tweets
	Number of retweets (per hour and total)		(*) Total number of emoticons in single tweets
	Number of mentions (per hour and total)		Ratio of tweets that contain emoticons
	Number of replies (per hour and total)		
	Number of retweeted (per hour and total)		
Friends (†)	Number of distinct languages	**Network** (‡)	Number of nodes
	Entropy of language use		Number of edges (also for reciprocal)
	(*) Account age distribution		(*) Strength distribution
	(*) Time offset distribution		(*) In-strength distribution
	(*) Number of friends distribution		(*) Out-strength distribution
	(*) Number of followers distribution		Network density (also for reciprocal)
	(*) Number of tweets distribution		(*) Clustering coeff. (also for reciprocal)
	(*) Description length distribution		
	Fraction of users with default profile and default picture		
Content	(*,**) Frequency of POS tags in a tweet	**Timing**	(*) Time between two consecutive tweets
	(*,**) Proportion of POS tags in a tweet		(*) Time between two consecutive retweets
	(*) Number of words in a tweet		(*) Time between two consecutive mentions
	(*) Entropy of words in a tweet		

† We consider four types of connected users: retweeting, mentioning, retweeted, and mentioned.
‡ We consider three types of network: retweet, mention, and hashtag co-occurrence networks.
* Distribution types. For each distribution, the following eight statistics are computed and used as individual features: min, max, median, mean, std. deviation, skewness, kurtosis, and entropy.
** Part-Of-Speech (POS) tag. There are nine POS tags: verbs, nuns, adjectives, modal auxiliaries, pre-determiners, interjections, adverbs, wh-, and pronouns.
*** For each feature, we compute mean and std. deviation of the weighted average across words in the lexicon.

Learning algorithms





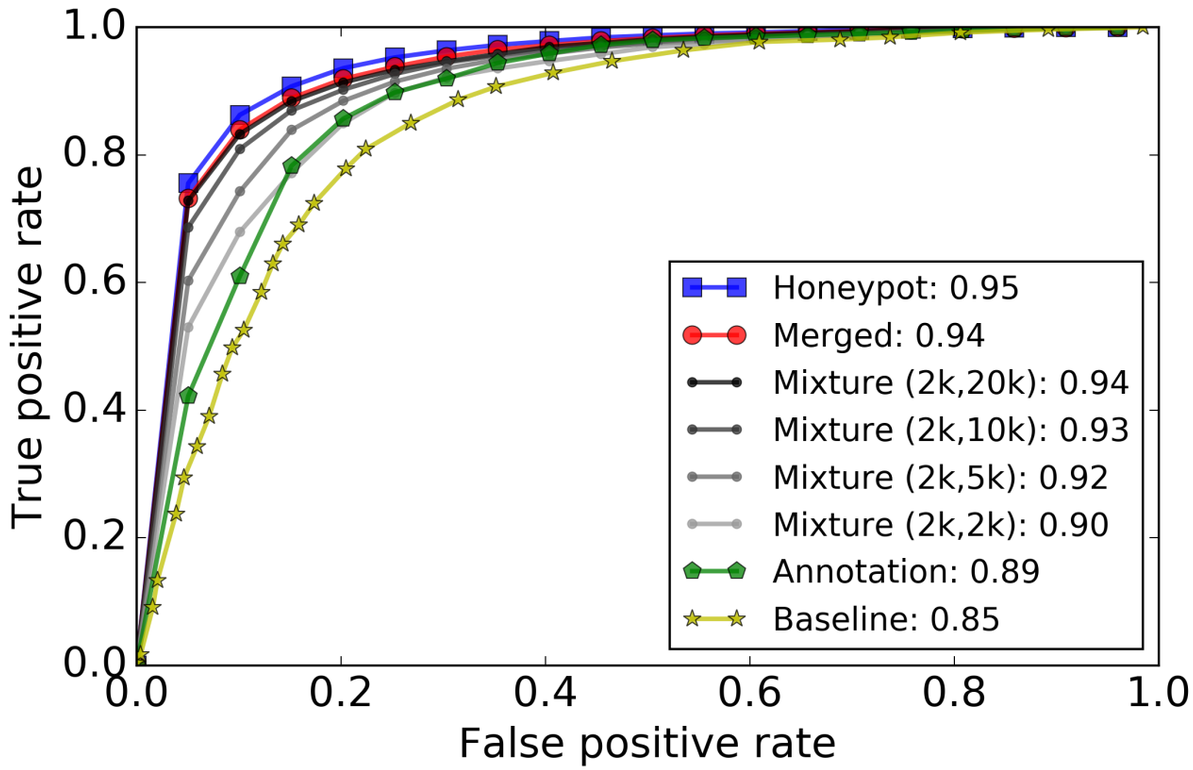
Initial models trained on an existing honeypot dataset[1]

We evaluated contribution of different feature classes and learning models.

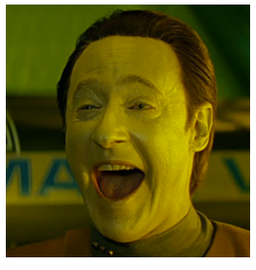
[1] Lee, Kyumin, Brian David Eoff, and James Caverlee. "Seven Months with the Devils: A Long- Term Study of Content Polluters on Twitter." ICWSM 2011.

Estimating bot population by tuning model sensitivity

Trained on mixtures of honeypot and manually annotated data for tuning sensitivity to recent bot behaviors.

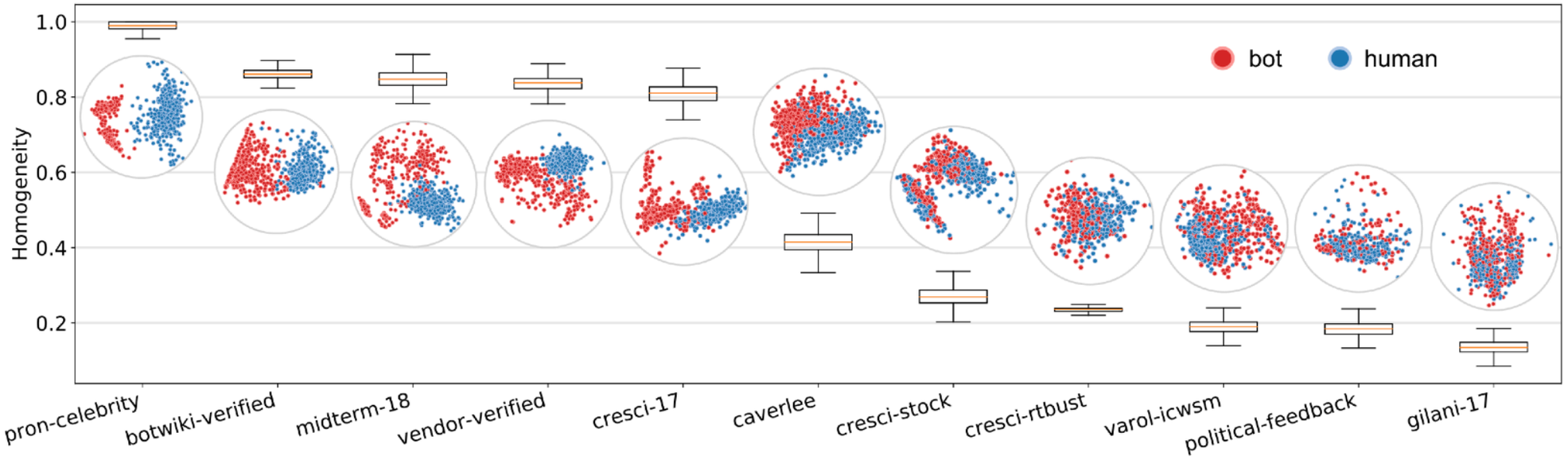


Adjusting sensitivity of detection system for sophisticated bots

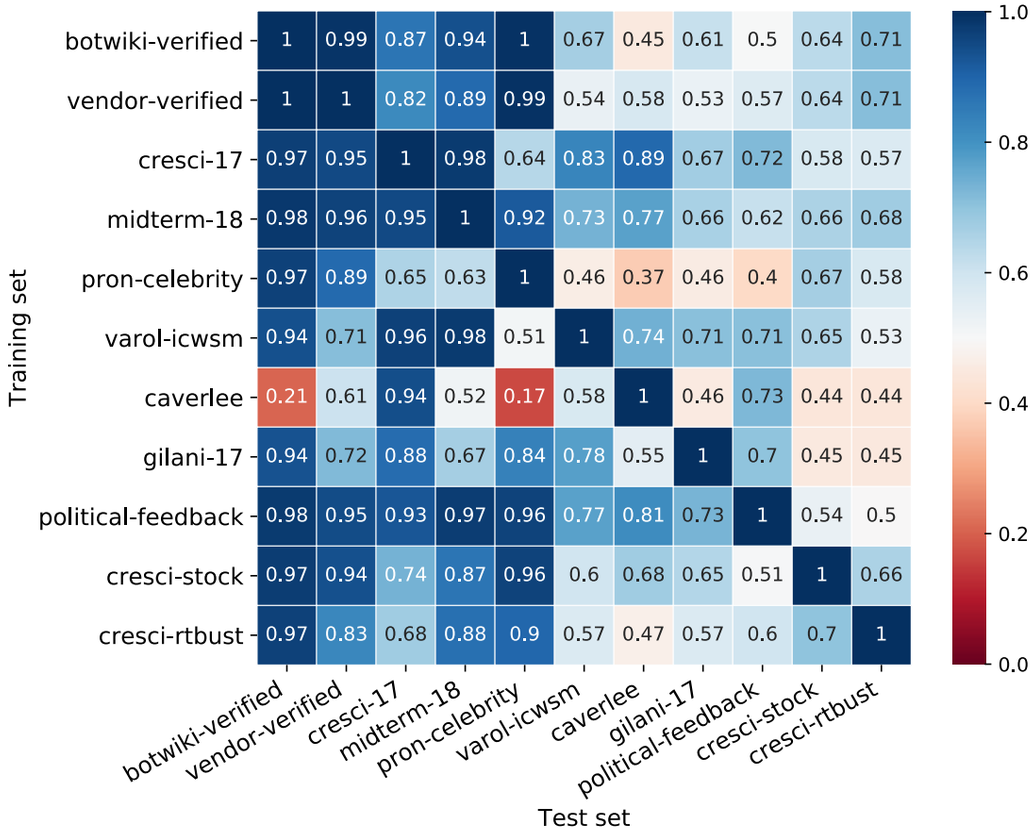


9-15% of active accounts have automated behavior

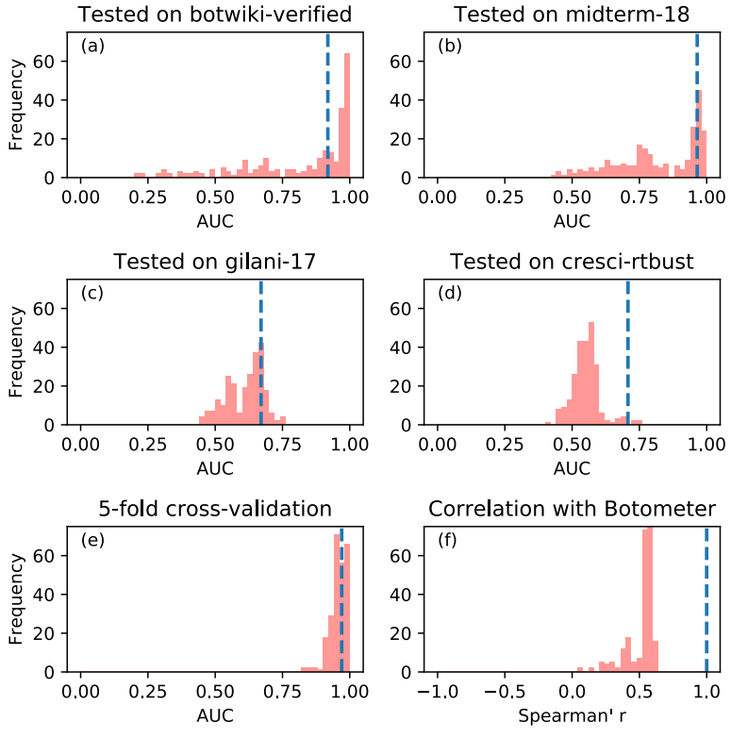
Improvement via data augmentation



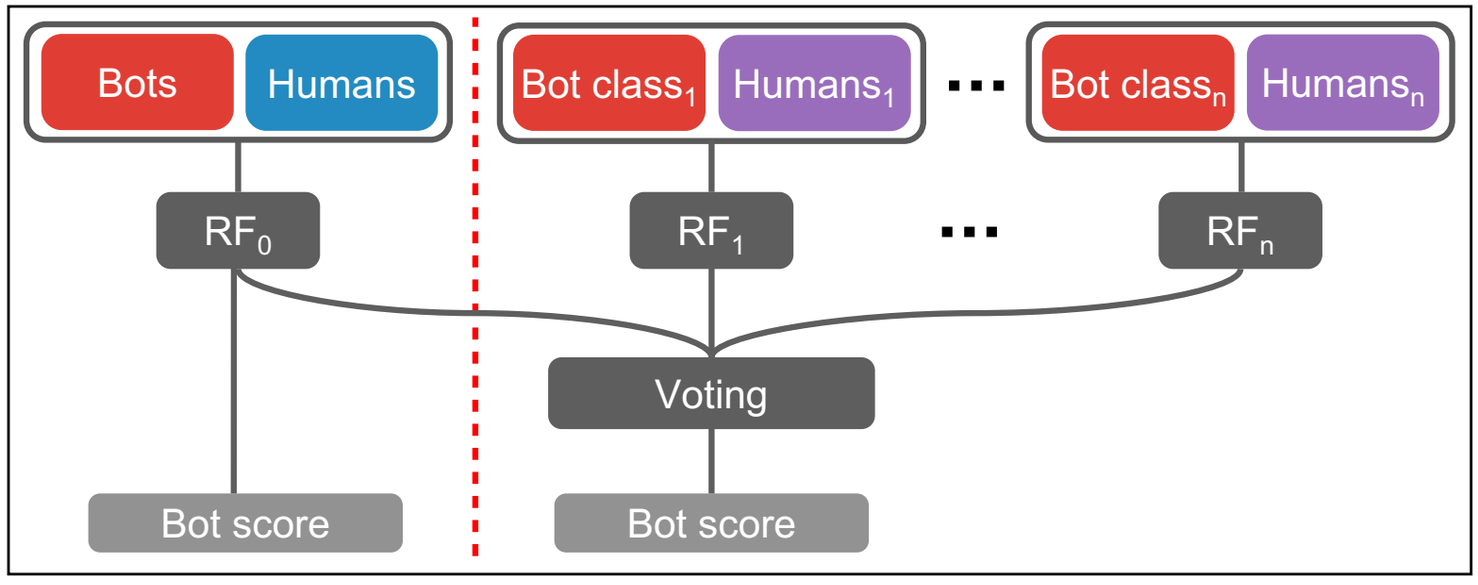
Out of domain performance

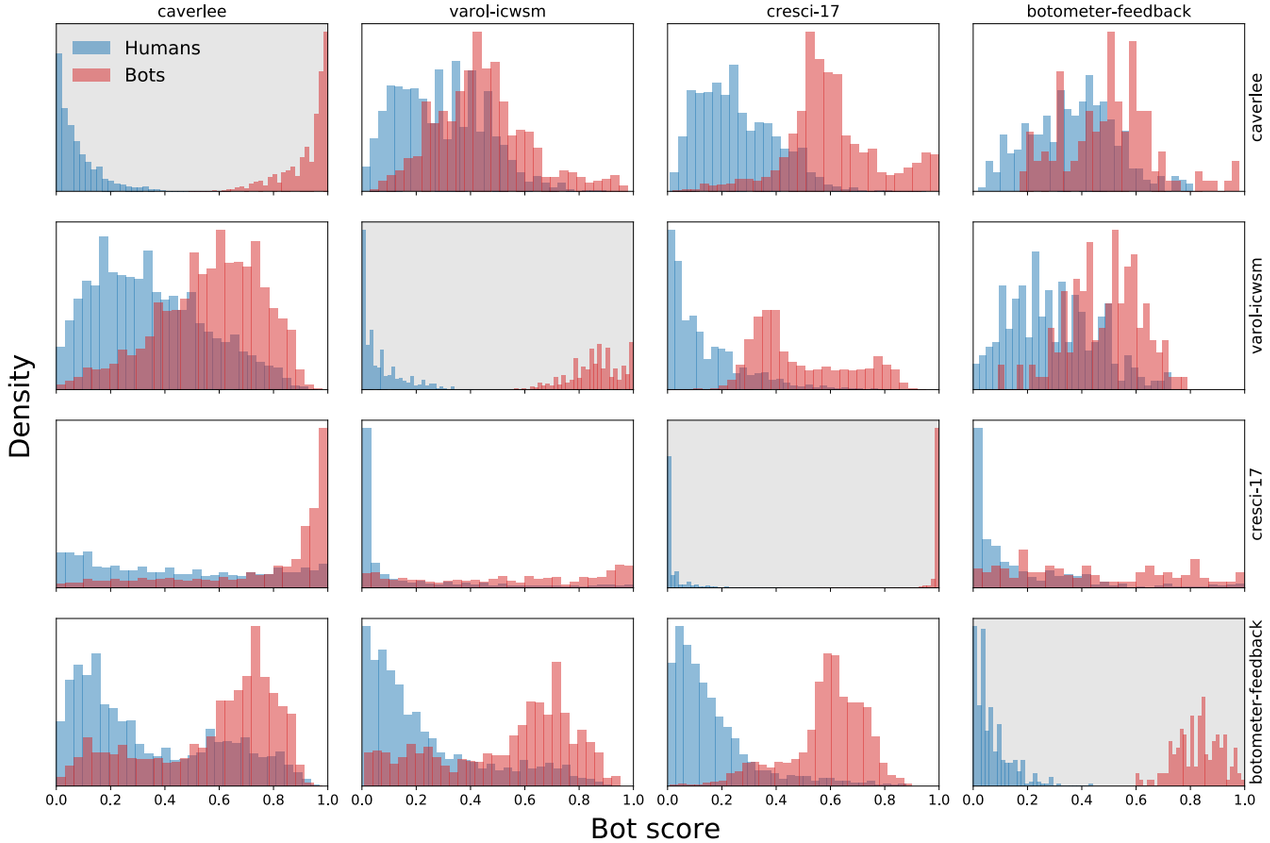


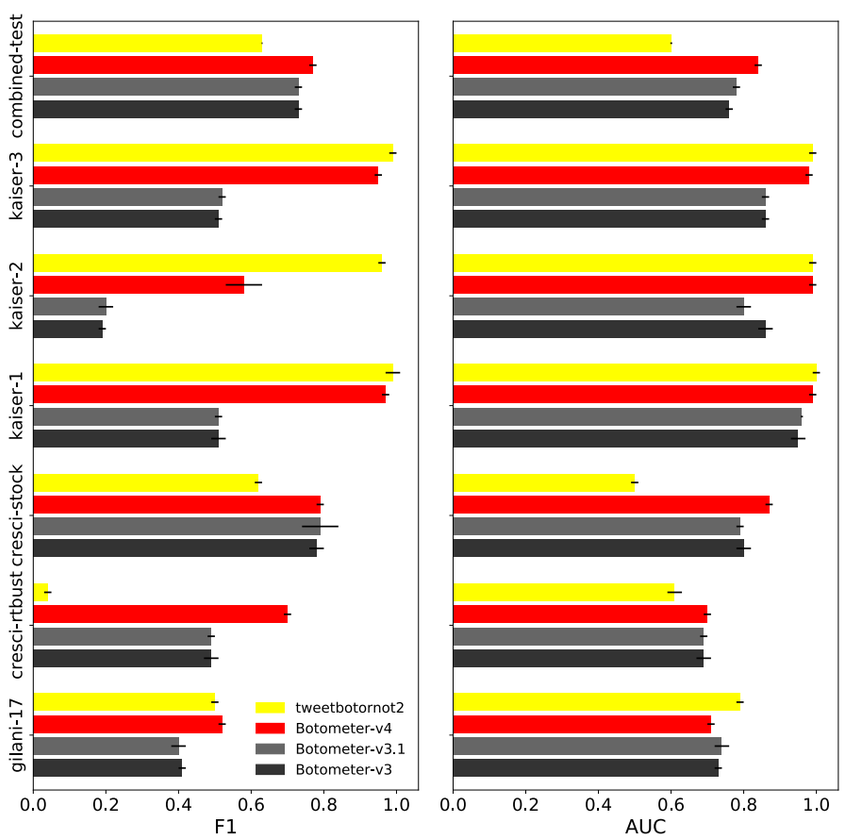
Model selection



Improvement via specialized models







Demo: Collecting bot scores using Botometer API

Let’s take a look at

3_collect-botometer-scores.ipynb

notebook for the analysis

Botometer score

Density

Cumulative score distribution

1.00
0.75
0.50
0.25
0.00

0.00 0.25 0.50 0.75 1.00

0.0 0.2 0.4 0.6 0.8 1.0

User

0.0 0.5 1.0

Friend

0.0 0.5 1.0

Network

0.0 0.5 1.0

Temporal

0.0 0.5 1.0

Botometer score for non-english

Density

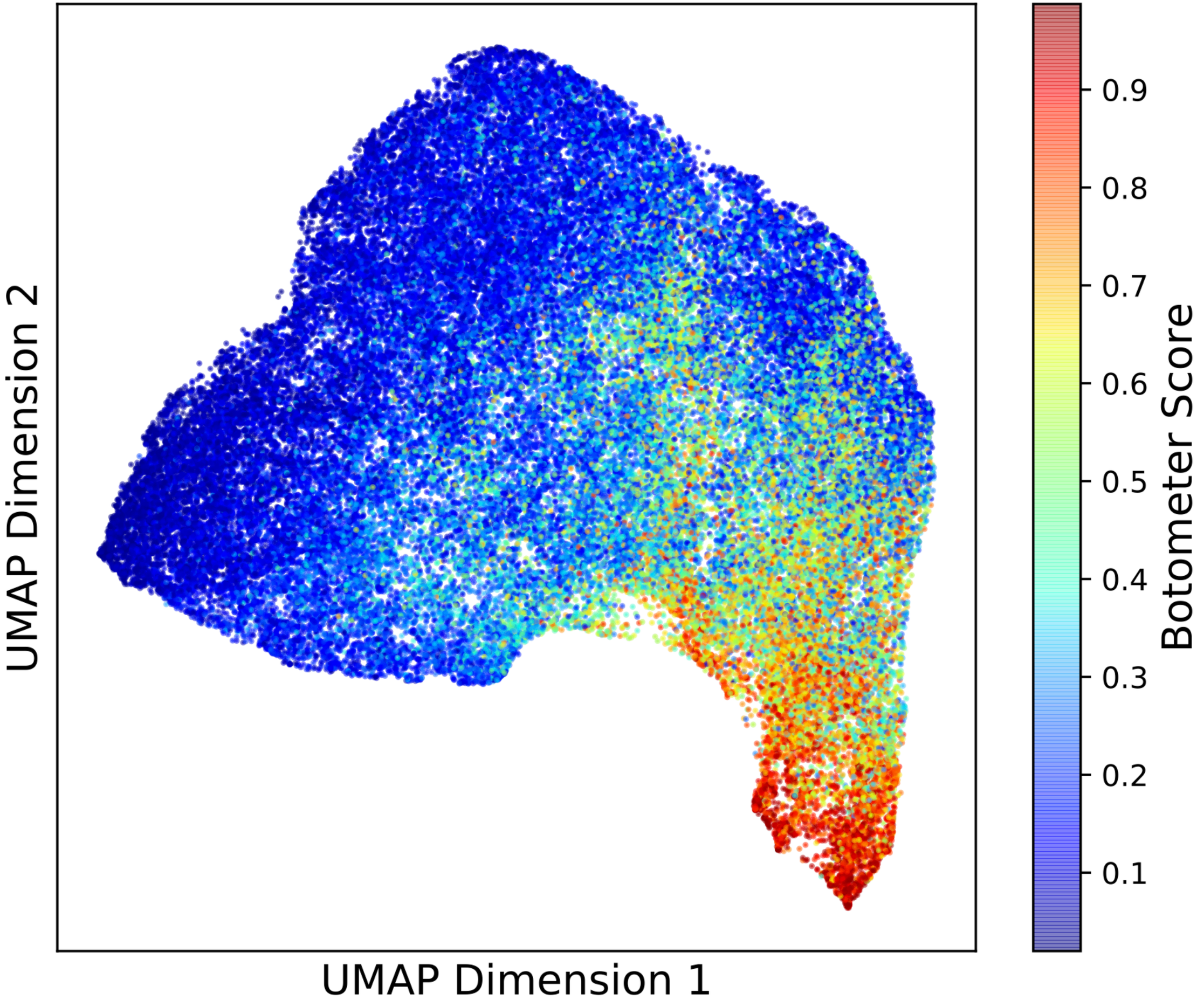
0.0 0.2 0.4 0.6 0.8 1.0

Sentiment

0.0 0.5 1.0

Content

0.0 0.5 1.0


UMAP Dimension 2
UMAP Dimension 1
Botometer Score
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1

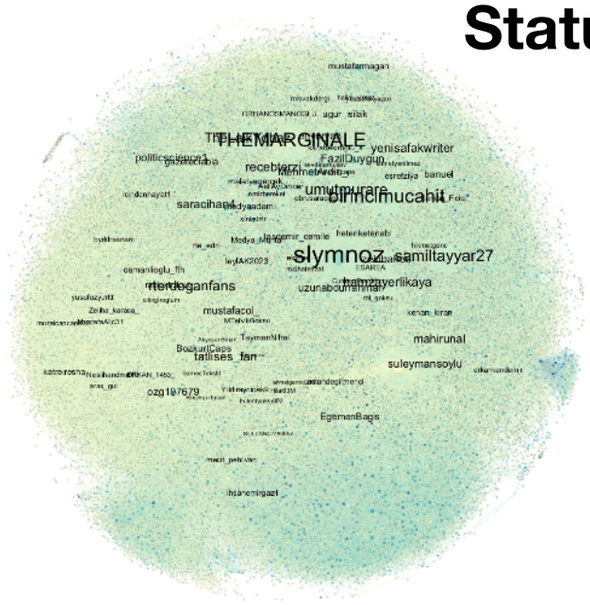
Demo: Analyze Twitter networks and Botometer scores

Let’s take a look at

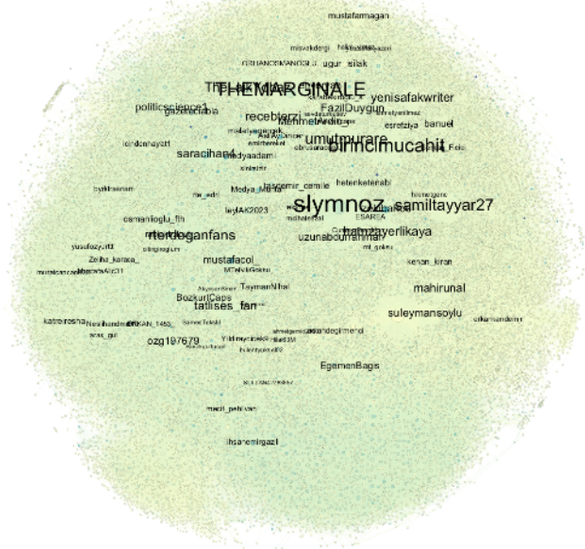
4_analyze-twitter-networks.ipynb

notebook for the analysis

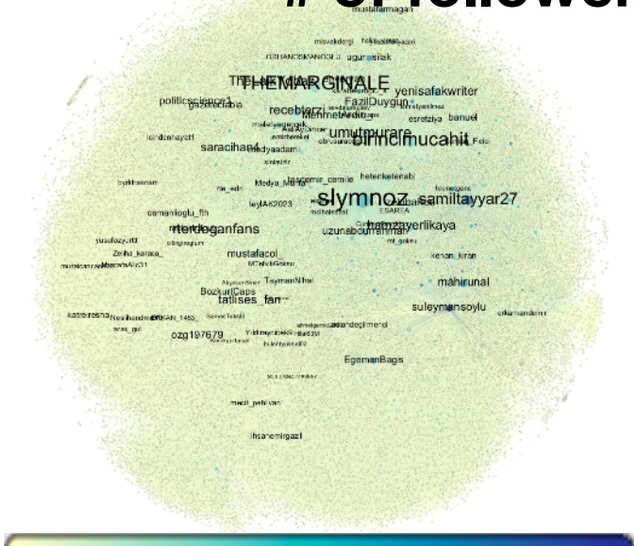
Status

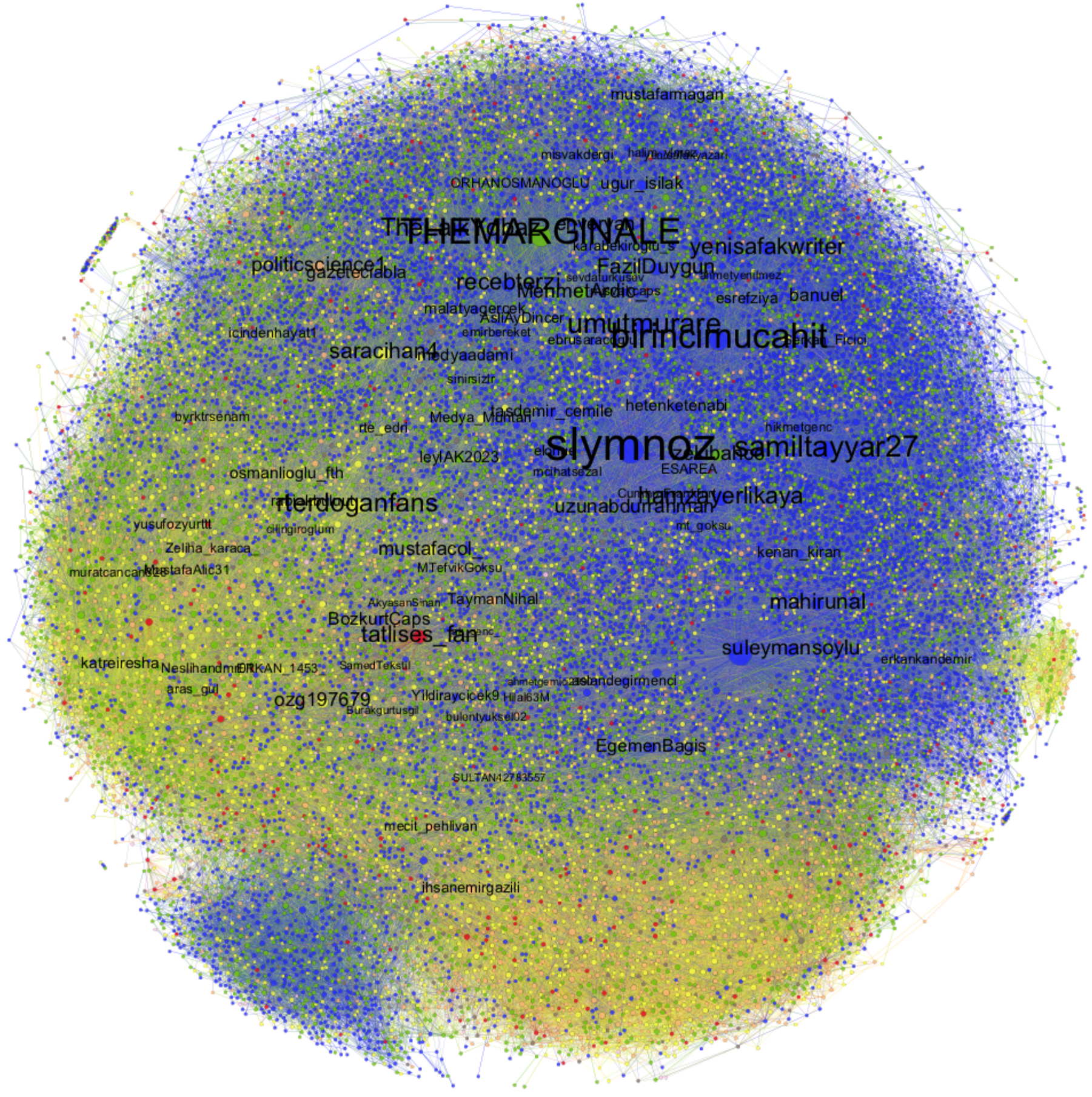


# of friends

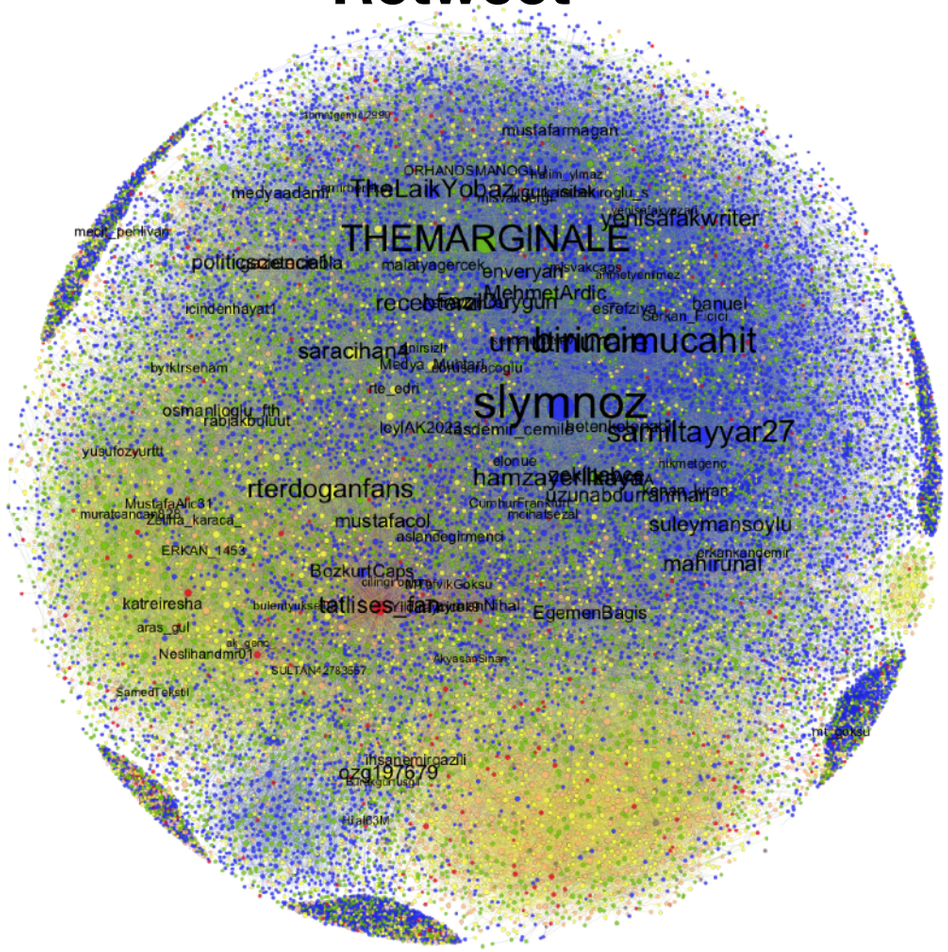


# of followers

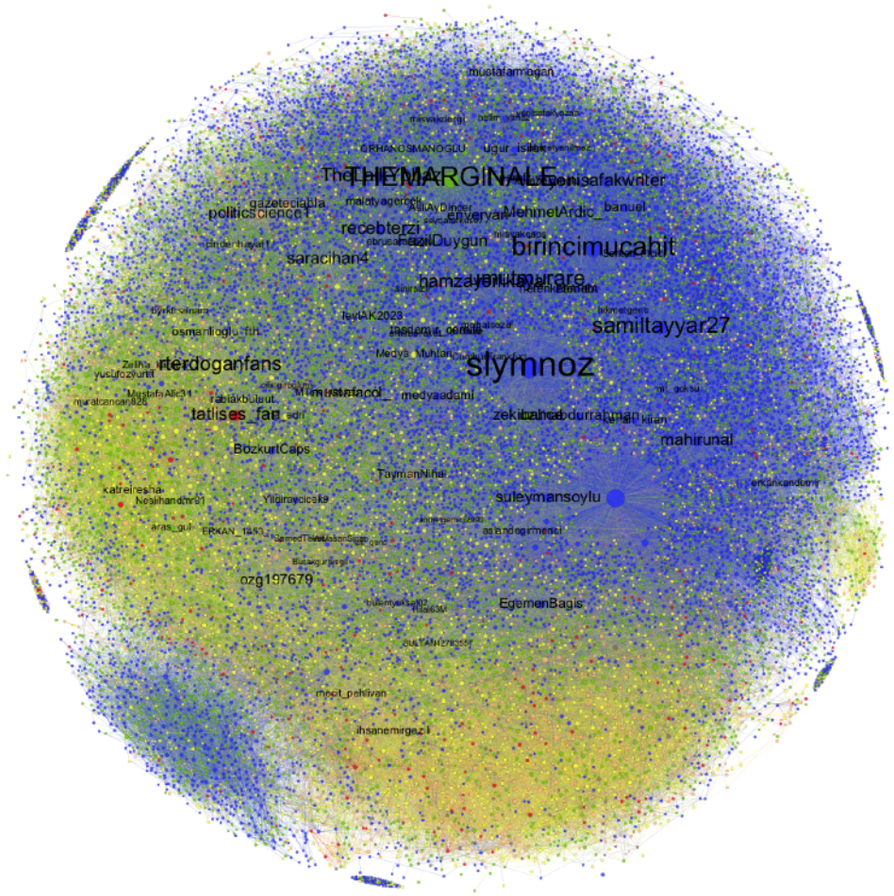




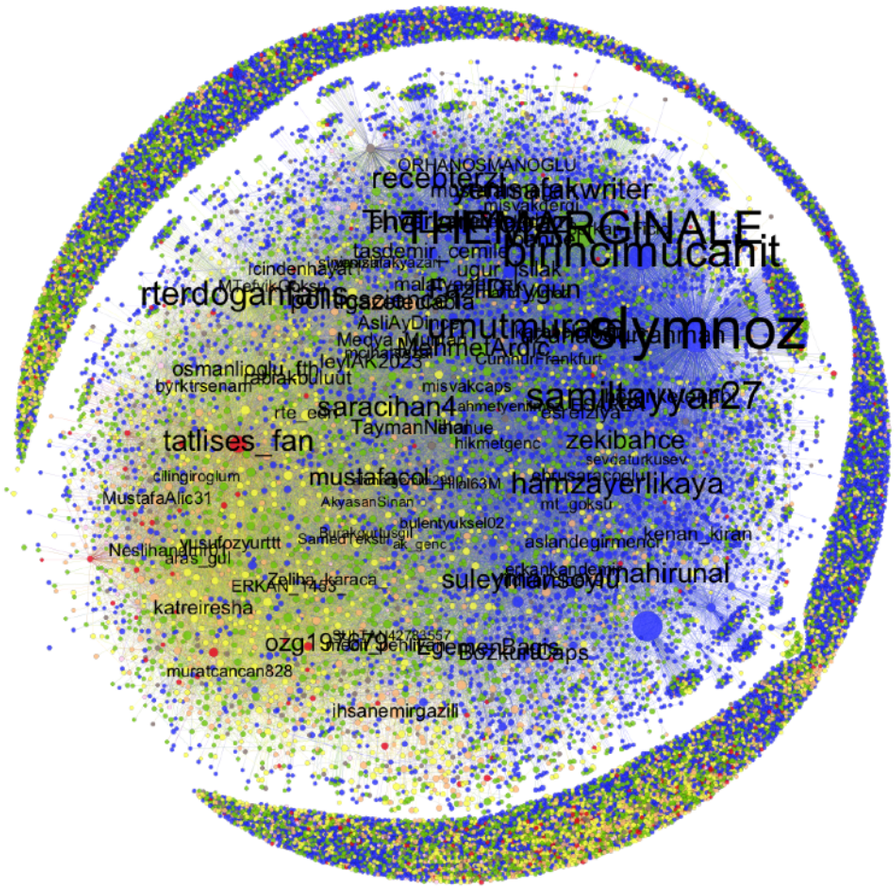
Retweet

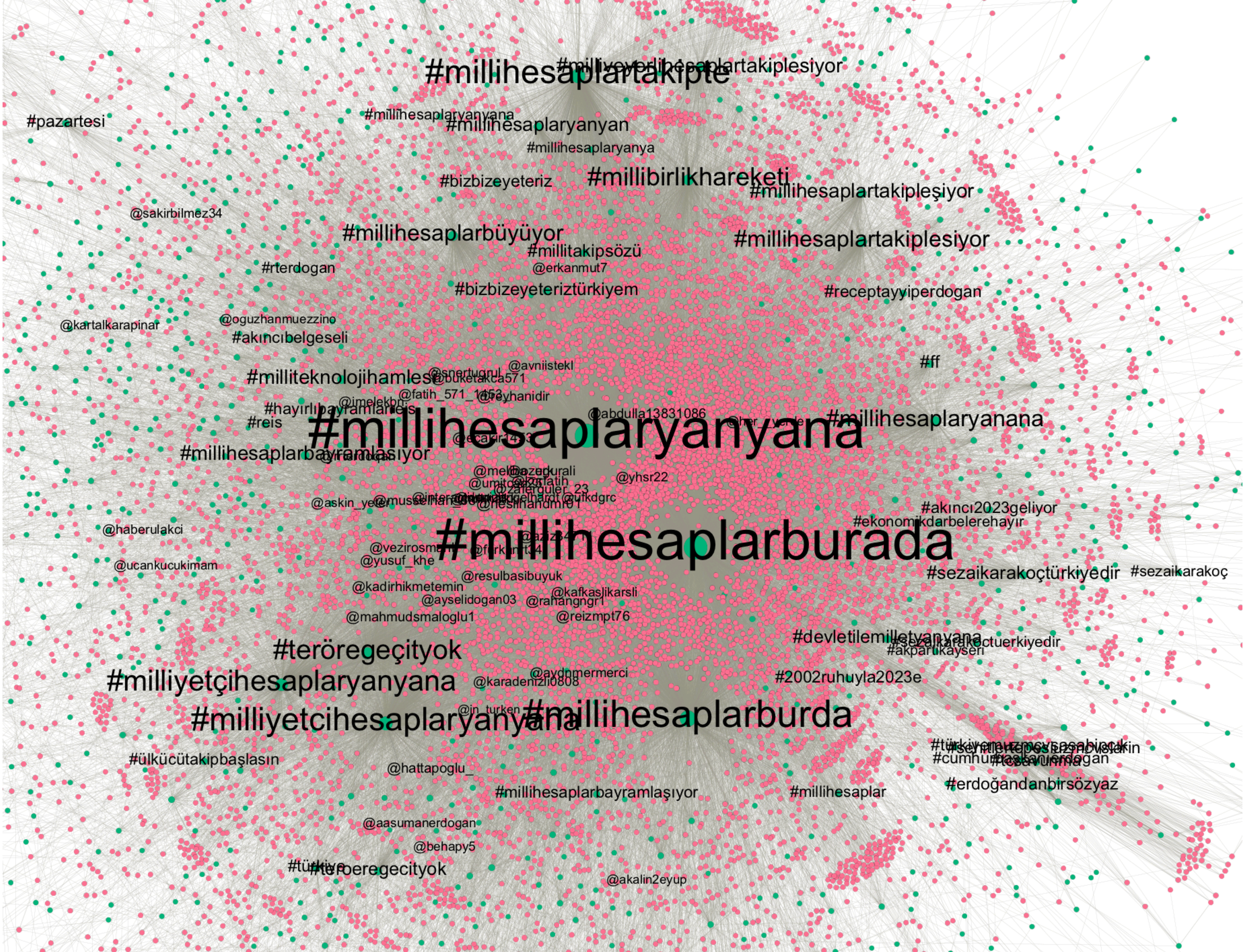


Mention

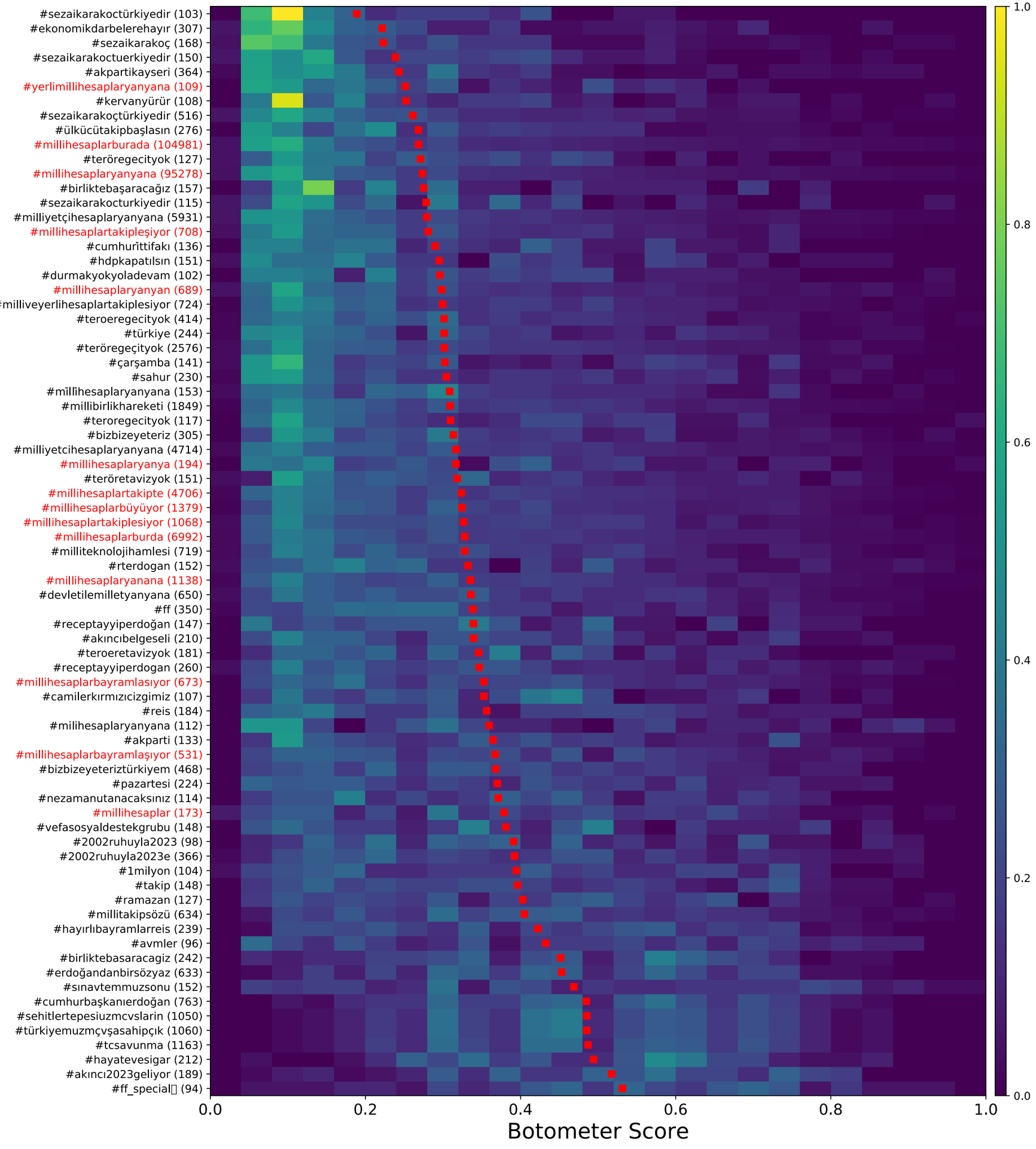


Reply





#millihesaplartakipte
#pazartesi
#millihesaplaryanyan
#millihesaplaryanya
#bizbizeyeteriz
#millibirlikhareketi
#millihesaplartakipleşiyor
@sakirbilmez34
#millihesaplarbüyüyor
#millihesaplartakiplesiyor
#millitakipsözü
@erkanmut7
#rterdogan
#bizbizeyeteriztürkiyem
#receptayyiperdogan
@kartalkarapinar
@oguzhanmuezzino
#akıncıbelgeseli
#ff
@avniistekl
@abdulla13831086
#millihesaplaryanana
#reis
#millihesaplaryanyana
@yhsr22
#akıncı2023geliyor
#ekonomikdarbelerehayır
@haberulakci
#millihesaplarburada
@ucankucukimam
@yusuf_khe
#sezaikarakoçtürkiyedir
#sezaikarakoç
@resulbasibuyuk
@kadirhikmetemin
@kafkaslikarsli
@ayselidogan03
@mahmudsmaloglu1
@reizmpt76
#teröregeçityok
#2002ruhuyla2023e
#milliyetçihesaplaryanyana
@aydnmermerci
@karadenizli0808
#millihesaplarburda
#ülkücütakipbaşlasın
@hattapoglu_
#millihesaplarbayramlaşıyor
#millihesaplar
#erdoğandanbirsözyaz
@aasumanerdogan
@behapy5
#teroeregecityok
@akalin2eyup


#sezaikarakoctürkiyedir (103)
#ekonomikdarbelerehayır (307)
#sezaikarakoç (168)
#sezaikarakoctuerkiyedir (150)
#akpartikayseri (364)
#yerlimillihesaplaryanyana (109)
#kervanyürür (108)
#sezaikarakoçtürkiyedir (516)
#ülkücütakipbaşlasın (276)
#millihesaplarburada (104981)
#teröregecityok (127)
#millihesaplaryanyana (95278)
#birliktebaşaracağız (157)
#sezaikarakocturkiyedir (115)
#milliyetçihesaplaryanyana (5931)
#millihesaplartakipleşiyor (708)
#cumhurïttifakı (136)
#hdpkapatılsın (151)
#durmakyokyoladevam (102)
#millihesaplaryanyan (689)
#milliveyerlihesaplartakiplesiyor (724)
#teroeregecityok (414)
#türkiye (244)
#teröregeçityok (2576)
#çarşamba (141)
#sahur (230)
#mïllïhesaplaryanyana (153)
#millibirlikhareketi (1849)
#teroregecityok (117)
#bizbizeyeteriz (305)
#milliyetcihesaplaryanyana (4714)
#millihesaplaryanya (194)
#teröretavizyok (151)
#millihesaplartakipte (4706)
#millihesaplarbüyüyor (1379)
#millihesaplartakiplesiyor (1068)
#millihesaplarburda (6992)
#milliteknolojihamlesi (719)
#rterdogan (152)
#millihesaplaryanana (1138)
#devletilemilletyanyana (650)
#ff (350)
#receptayyiperdoğan (147)
#akıncıbelgeseli (210)
#teroeretavizyok (181)
#receptayyiperdogan (260)
#millihesaplarbayramlasıyor (673)
#camilerkırmızıcizgimiz (107)
#reis (184)
#milihesaplaryanyana (112)
#akparti (133)
#millihesaplarbayramlaşıyor (531)
#bizbizeyeteriztürkiyem (468)
#pazartesi (224)
#nezamanutanacaksınız (114)
#millihesaplar (173)
#vefasosyaldestekgrubu (148)
#2002ruhuyla2023 (98)
#2002ruhuyla2023e (366)
#1milyon (104)
#takip (148)
#ramazan (127)
#millitakipsözü (634)
#hayırlıbayramlarreis (239)
#avmler (96)
#birliktebasaracagiz (242)
#erdoğandanbirsözyaz (633)
#sınavtemmuzsonu (152)
#cumhurbaşkanıerdoğan (763)
#sehitlertepesiuzmcvslarin (1050)
#türkiyemuzmçvşasahipçık (1060)
#tcsavunma (1163)
#hayatevesigar (212)
#akıncı2023geliyor (189)
#ff_special (94)
0.0
0.2
0.4
0.6
0.8
1.0
Botometer Score

Demo: Bonus curiosity driven analysis

Let’s take a look at

90_suspended-accounts-analysis.ipynb

notebook for the analysis

Thank you very much 🤗

I will appreciate if you can spare 5 min. To fill out the survey below to help me make this tutorial better.

Survey link: https://forms.gle/otnxxSAZxFq9t5bv9

@onurvarol
onurvarol.com