FUSION

NYC DATA SCIENCE ACADEMY

Twitter Influencers Clustering and Personalized Recommendation

Engage with influencers the right way.

Turn Data Into $$$?

The New York Times

Buzz Feed

FUSION

THE HUFFINGTON POST

Data

Articles

Web Traffic

Social Media

Popularity → $$$

User

Advertiser

Turn Data Into $$$?

The New York Times

Buzz Feed

FUSION

THE HUFFINGTON POST

Data → Popularity → $$$

Articles

Web Traffic

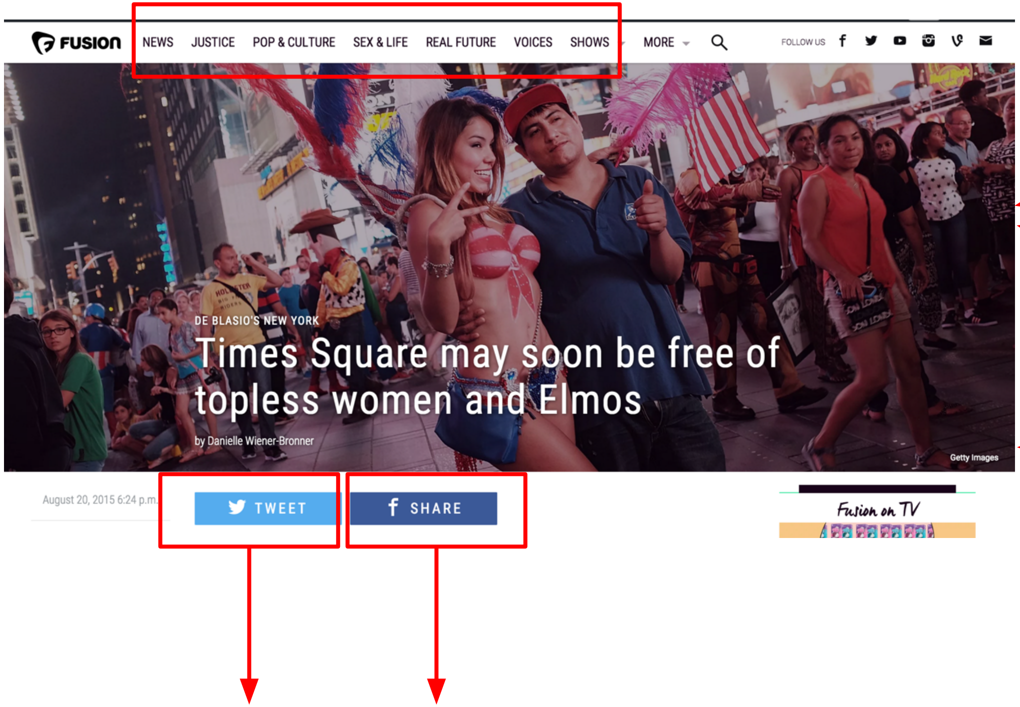
Social Media

Give the right people the right content at the right times

User

Advertiser

Fusion Data Sources



DE BLASIO'S NEW YORK

Times Square may soon be free of topless women and Elmos

by Danielle Wiener-Bronner

August 20, 2015 6:24 p.m.

This Project Scope

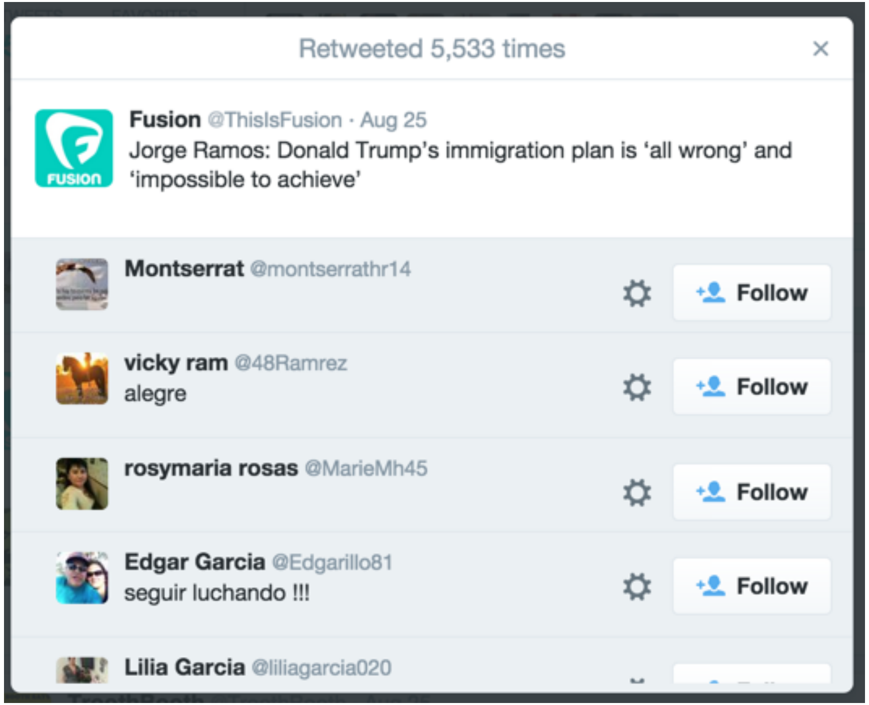
Mining Twitter

Who are influential to @thisisfusion:

- Participate in @thisisfusion tweets
  - retweet ← Centrality Score
  - favorite
  - comment
- Initiate tweets mentioning @thisisfusion
  - direct mention

InfluenceFlow Score

Mining One Tweet



Retweeted 5,533 times

Fusion @ThisIsFusion · Aug 25
Jorge Ramos: Donald Trump's immigration plan is 'all wrong' and 'impossible to achieve'

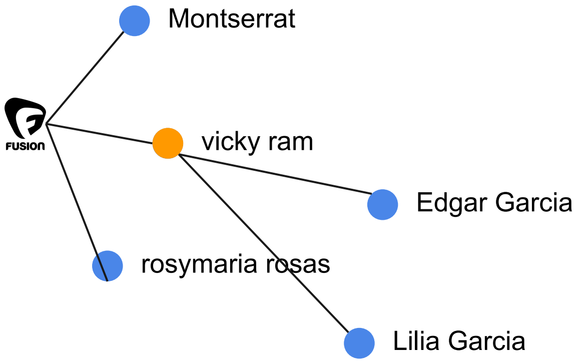
Montserrat @montserrathr14

vicky ram @48Ramrez
alegre

rosymaria rosas @MarieMh45

Edgar Garcia @Edgarillo81
seguir luchando !!!

Lilia Garcia @liliagarcia020

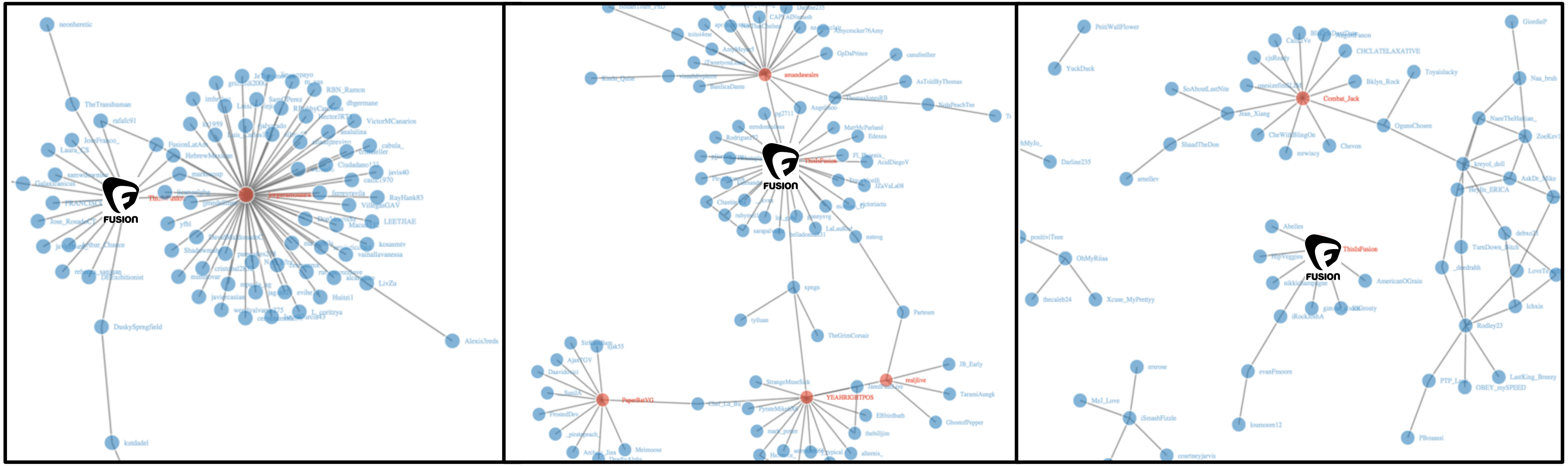


Centrality:
How important a user is within a retweet network.

Degree-centrality: No. of followers
Eigenvector-centrality: No. of important followers

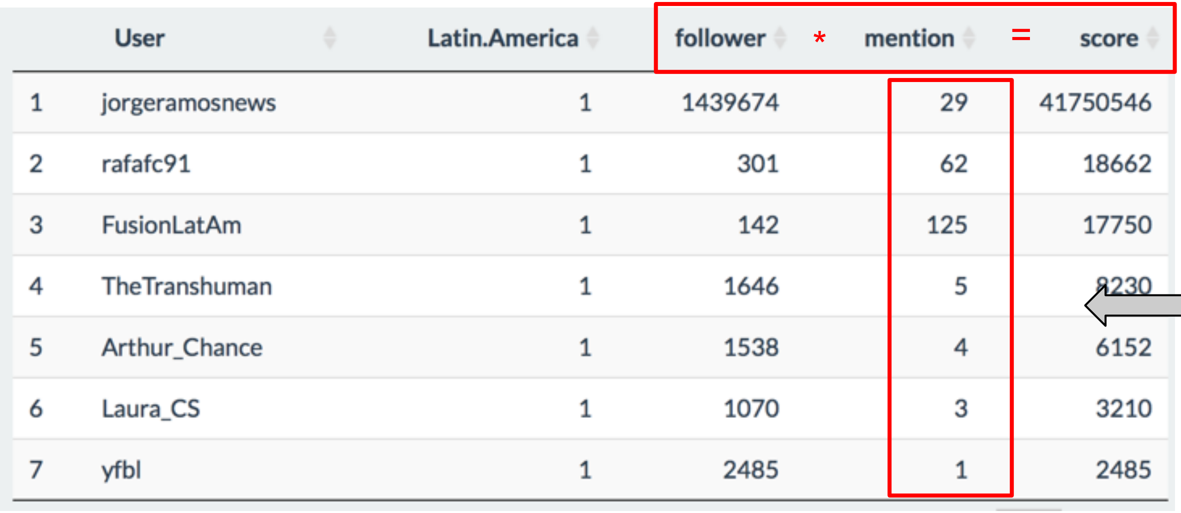
Mining One Tweet

How does each tweet transmit?



Mining Fusion Community

Not limited to specific tweet, InfluenceFlow Score captures overall information flow in Fusion community:



	User	Latin.America	follower *	mention =	score
1	jorgeramosnews	1	1439674	29	41750546
2	rafafc91	1	301	62	18662
3	FusionLatAm	1	142	125	17750
4	TheTranshuman	1	1646	5	8230
5	Arthur_Chance	1	1538	4	6152
6	Laura_CS	1	1070	3	3210
7	yfbl	1	2485	1	2485

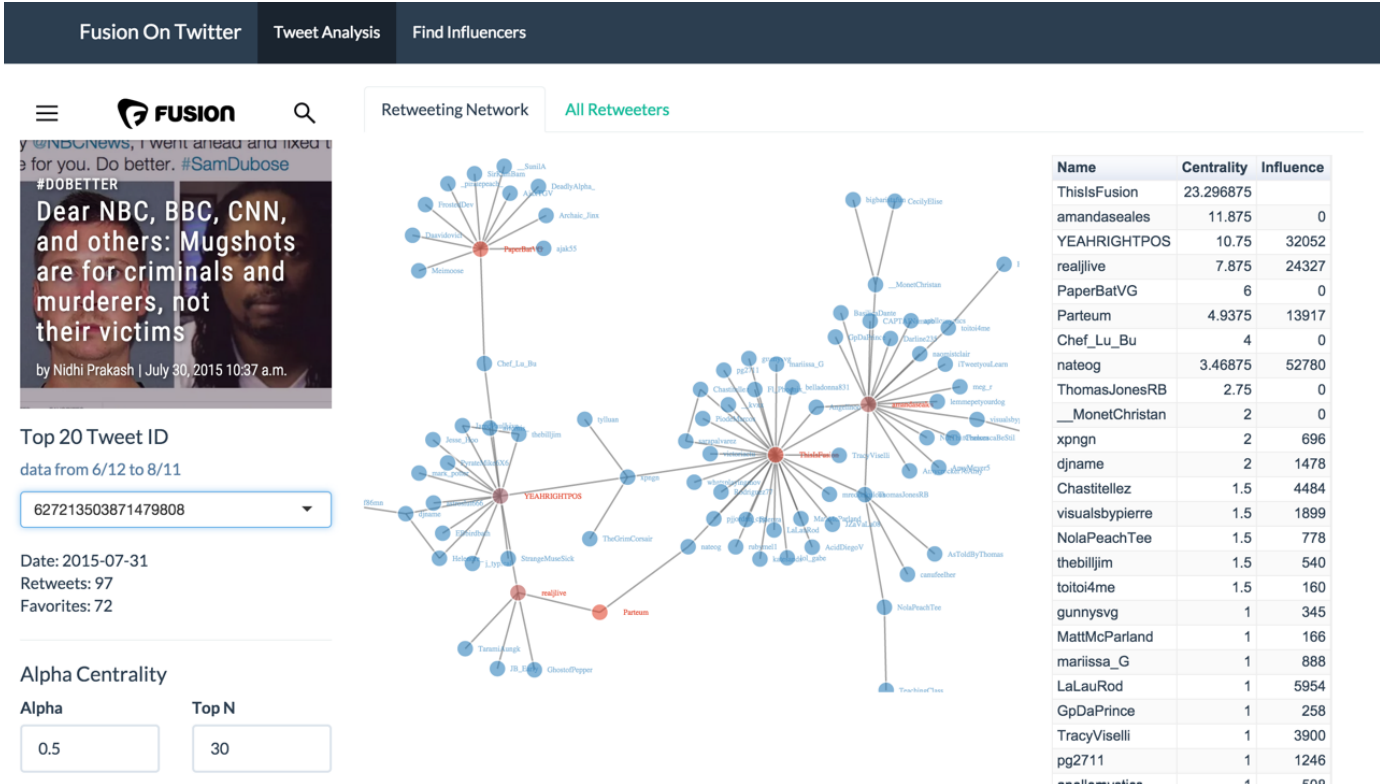
InfluenceFlow Score

Include:
No. of Retweets
No. of Direct mentions

Right People: Superfans

- Collect superfan data from every tweet
    - Scores
        - Centrality - influence within one tweet
        - InfluenceFlow - overall influence
    - Interests
        - Twitter Hashtags
        - Article Sections (news/justice/voices...)
        - Article Topics (drugs/transgender/mexico...)
        - ...

Live Demo



y @NBCNews, I went ahead and fixed t
e for you. Do better. #SamDubose

#DOBETTER

Dear NBC, BBC, CNN, and others: Mugshots are for criminals and murderers, not their victims

by Nidhi Prakash | July 30, 2015 10:37 a.m.

Top 20 Tweet ID

data from 6/12 to 8/11

627213503871479808 ▾

Date: 2015-07-31
Retweets: 97
Favorites: 72

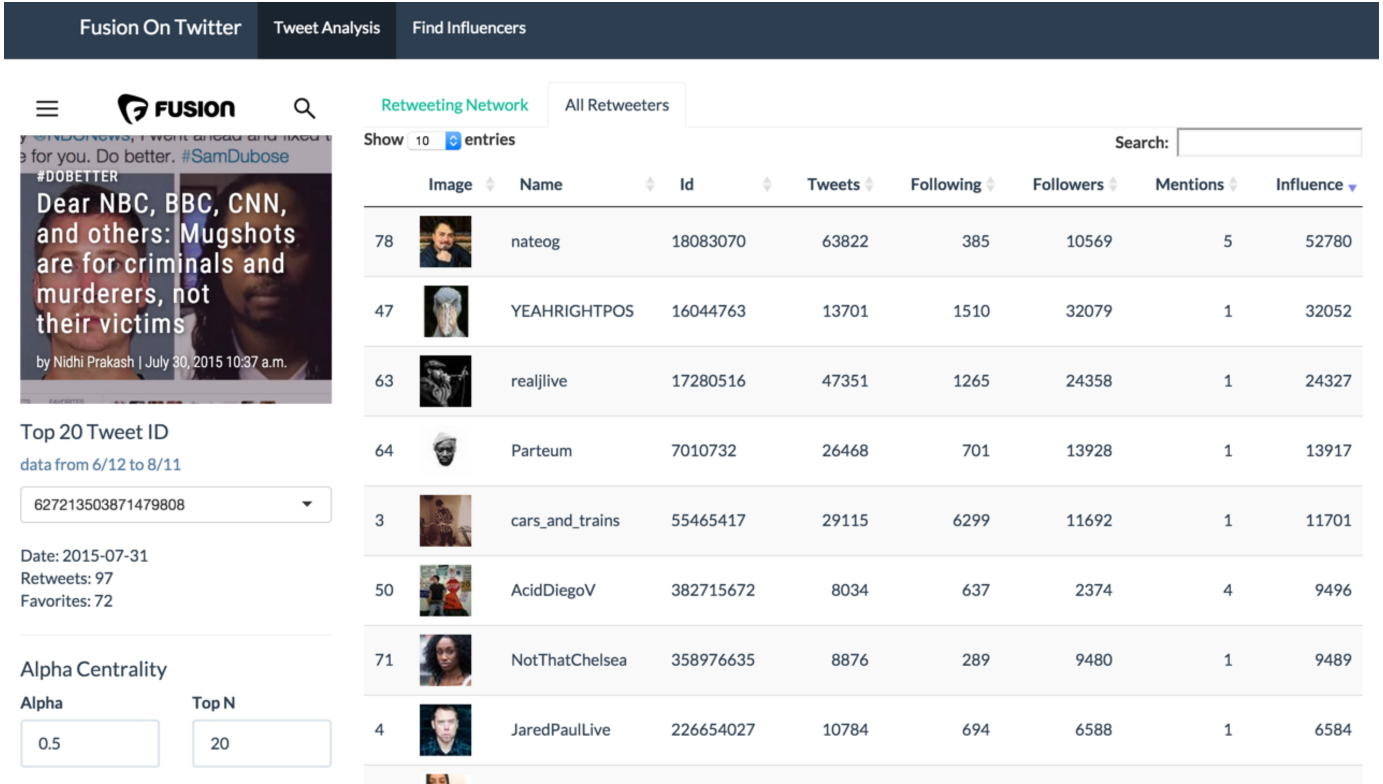
Alpha Centrality

Alpha: 0.5

Top N: 30

Name	Centrality	Influence
ThisIsFusion	23.296875	
amandaseales	11.875	0
YEAHRIGHTPOS	10.75	32052
realjlive	7.875	24327
PaperBatVG	6	0
Parteum	4.9375	13917
Chef_Lu_Bu	4	0
nateog	3.46875	52780
ThomasJonesRB	2.75	0
__MonetChristan	2	0
xpngn	2	696
djname	2	1478
Chastitellez	1.5	4484
visualsbypierre	1.5	1899
NolaPeachTee	1.5	778
thebilljim	1.5	540
toitoi4me	1.5	160
gunnysvg	1	345
MattMcParland	1	166
mariissa_G	1	888
LaLauRod	1	5954
GpDaPrince	1	258
TracyViselli	1	3900
pg2711	1	1246

Live Demo



e for you. Do better. #SamDubose

#DOBETTER

Dear NBC, BBC, CNN, and others: Mugshots are for criminals and murderers, not their victims

by Nidhi Prakash | July 30, 2015 10:37 a.m.

Top 20 Tweet ID

data from 6/12 to 8/11

627213503871479808

Date: 2015-07-31
Retweets: 97
Favorites: 72

Alpha Centrality

Alpha: 0.5

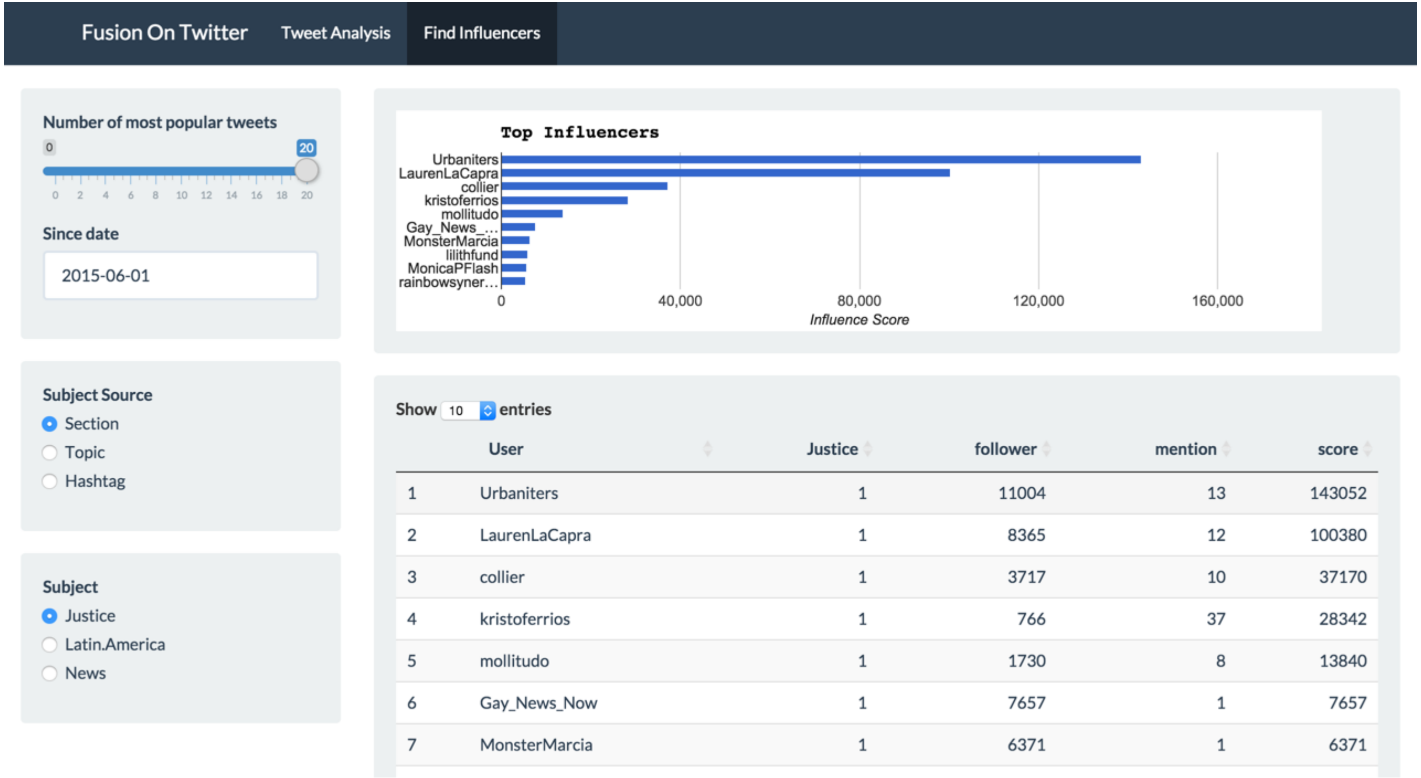
Top N: 20

Show 10 entries

Search:

	Image	Name	Id	Tweets	Following	Followers	Mentions	Influence
78		nateog	18083070	63822	385	10569	5	52780
47		YEAHRIGHTPOS	16044763	13701	1510	32079	1	32052
63		realjlive	17280516	47351	1265	24358	1	24327
64		Parteum	7010732	26468	701	13928	1	13917
3		cars_and_trains	55465417	29115	6299	11692	1	11701
50		AcidDiegoV	382715672	8034	637	2374	4	9496
71		NotThatChelsea	358976635	8876	289	9480	1	9489
4		JaredPaulLive	226654027	10784	694	6588	1	6584

Live Demo



Number of most popular tweets

0 — 20 (0 2 4 6 8 10 12 14 16 18 20)

Since date

2015-06-01

Subject Source

● Section
○ Topic
○ Hashtag

Subject

● Justice
○ Latin.America
○ News

Show 10 entries

	User	Justice	follower	mention	score
1	Urbaniters	1	11004	13	143052
2	LaurenLaCapra	1	8365	12	100380
3	collier	1	3717	10	37170
4	kristoferrios	1	766	37	28342
5	mollitudo	1	1730	8	13840
6	Gay_News_Now	1	7657	1	7657
7	MonsterMarcia	1	6371	1	6371

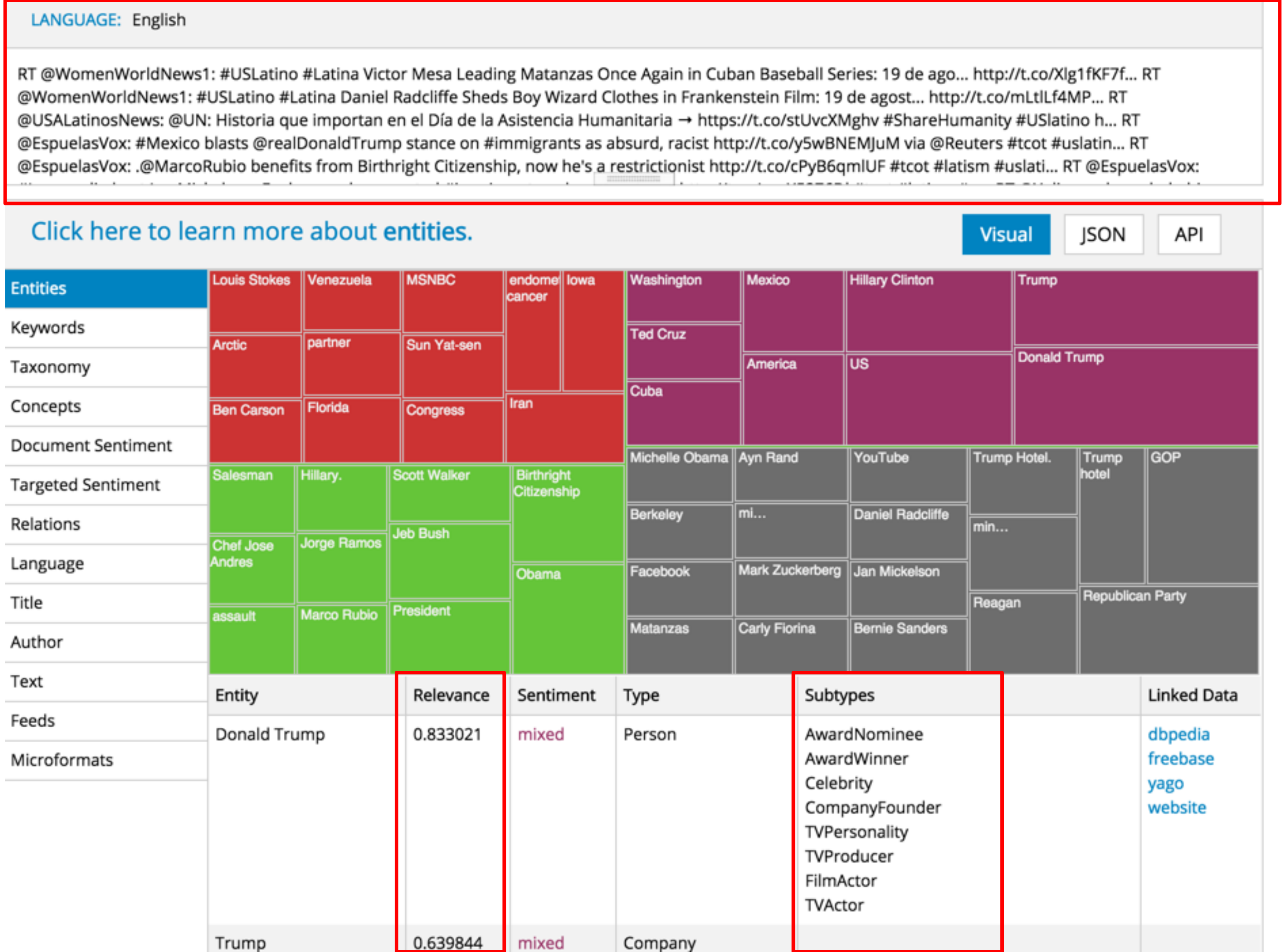
Right Content?

Is there an algorithm to suggest personalized recommendations to superfans?

In Progress:

- Text mining on all superfans timeline content
- Clustering texts on vectors from AlchemyAPI
- Observe distributions of superfan scores and interests in different clusters
- Content-based recommendation (match vectorized user timeline text and article text using cosine similarity)

AlchemyAPI



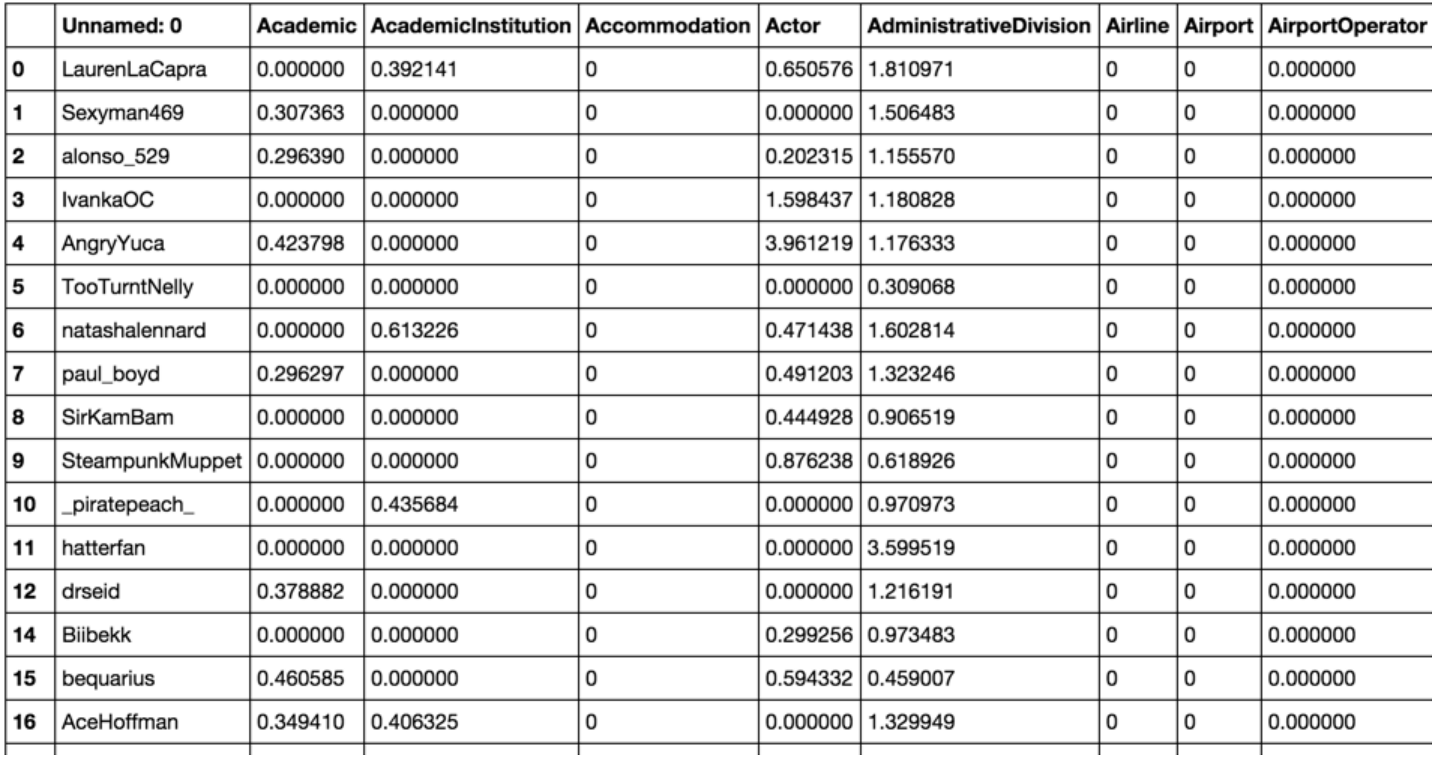
LANGUAGE: English

RT @WomenWorldNews1: #USLatino #Latina Victor Mesa Leading Matanzas Once Again in Cuban Baseball Series: 19 de ago... http://t.co/Xlg1fKF7f... RT @WomenWorldNews1: #USLatino #Latina Daniel Radcliffe Sheds Boy Wizard Clothes in Frankenstein Film: 19 de agost... http://t.co/mLtlLf4MP... RT @USALatinosNews: @UN: Historia que importan en el Día de la Asistencia Humanitaria → https://t.co/stUvcXMghv #ShareHumanity #USlatino h... RT @EspuelasVox: #Mexico blasts @realDonaldTrump stance on #immigrants as absurd, racist http://t.co/y5wBNEMJuM via @Reuters #tcot #uslatin... RT @EspuelasVox: .@MarcoRubio benefits from Birthright Citizenship, now he's a restrictionist http://t.co/cPyB6qmlUF #tcot #latism #uslati... RT @EspuelasVox:

Click here to learn more about entities.

Entity	Relevance	Sentiment	Type	Subtypes	Linked Data
Donald Trump	0.833021	mixed	Person	AwardNominee AwardWinner Celebrity CompanyFounder TVPersonality TVProducer FilmActor TVActor	dbpedia freebase yago website
Trump	0.639844	mixed	Company		

Right Content?



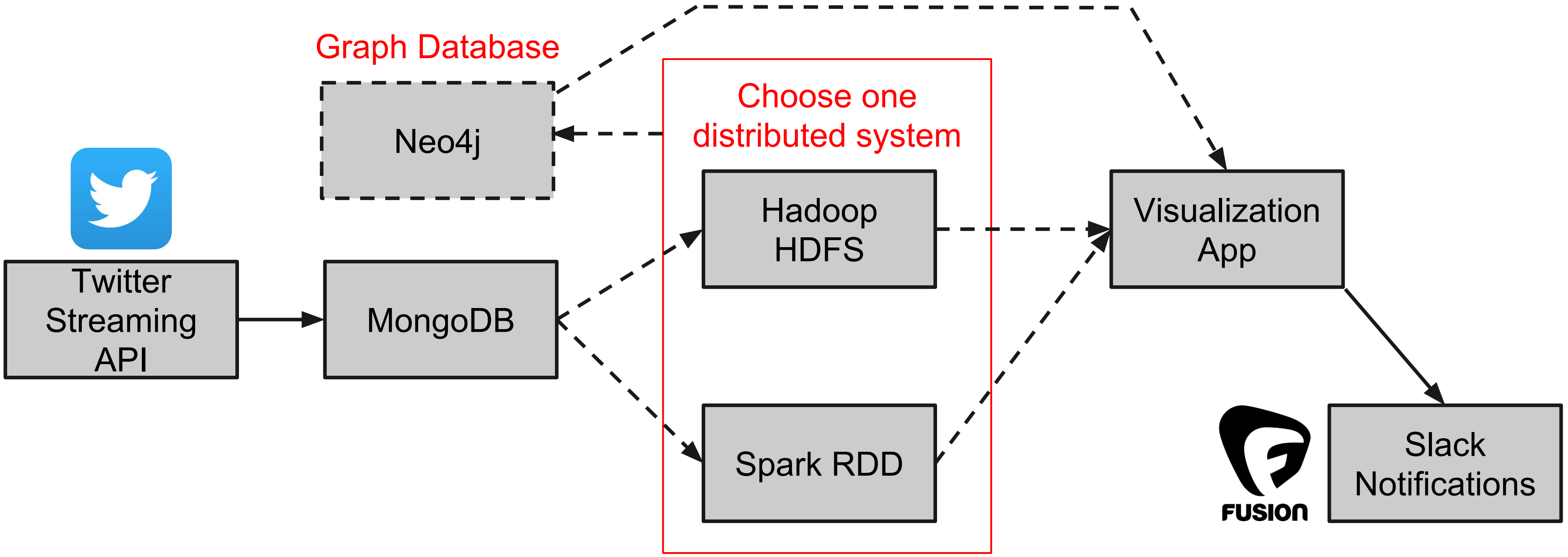
	Unnamed: 0	Academic	AcademicInstitution	Accommodation	Actor	AdministrativeDivision	Airline	Airport	AirportOperator
0	LaurenLaCapra	0.000000	0.392141	0	0.650576	1.810971	0	0	0.000000
1	Sexyman469	0.307363	0.000000	0	0.000000	1.506483	0	0	0.000000
2	alonso_529	0.296390	0.000000	0	0.202315	1.155570	0	0	0.000000
3	IvankaOC	0.000000	0.000000	0	1.598437	1.180828	0	0	0.000000
4	AngryYuca	0.423798	0.000000	0	3.961219	1.176333	0	0	0.000000
5	TooTurntNelly	0.000000	0.000000	0	0.000000	0.309068	0	0	0.000000
6	natashalennard	0.000000	0.613226	0	0.471438	1.602814	0	0	0.000000
7	paul_boyd	0.296297	0.000000	0	0.491203	1.323246	0	0	0.000000
8	SirKamBam	0.000000	0.000000	0	0.444928	0.906519	0	0	0.000000
9	SteampunkMuppet	0.000000	0.000000	0	0.876238	0.618926	0	0	0.000000
10	_piratepeach_	0.000000	0.435684	0	0.000000	0.970973	0	0	0.000000
11	hatterfan	0.000000	0.000000	0	0.000000	3.599519	0	0	0.000000
12	drseid	0.378882	0.000000	0	0.000000	1.216191	0	0	0.000000
14	Biibekk	0.000000	0.000000	0	0.299256	0.973483	0	0	0.000000
15	bequarius	0.460585	0.000000	0	0.594332	0.459007	0	0	0.000000
16	AceHoffman	0.349410	0.406325	0	0.000000	1.329949	0	0	0.000000

Limits

- Time Limit (3-week-project)
  - Mainly focused on identifying and classifying superfans
  - Didn’t include time analysis

- Data Limit (Twitter API limitations)
  - Rate limit - takes 1 to 2 days to get basic data needed for 20 tweets
  - Query limit - can’t retrieve more than 3,200 tweets and 100 retweets, which sacrificed the flexibility of popularity analysis
  - Lack of infrastructure support - data consistency issue occurred a lot

Make It Scalable



About Us

This is an open source project collaborated between NYC Data Science Academy and Fusion media (an ABC-Univision joint-venture).

--NYC Data Science Academy

Fangzhou Cheng (PM, Data Science Fellow)

Shu Yan (Data Science Fellow)

Alex Singal (Data Science Fellow)

--Fusion

Noppanit Charassinvichai (Data Engineer)