"Very fine people on both sides" of Twitter: Analyzing the Network Structure of the Online Conversation about #Charlottesville

Joseph Tien

SIAM Dynamical Systems

May 2019

Joint work

- Marisa Eisenberg
- Sarah Cherng
- Mason Porter

'Unite the Right'

- 11-12 August 2017 -- 'Unite the Right' rally in Charlottesville, VA
- Rally participants: neo-Nazis, white supremacists, 'alt-right'; counter-protesters
- Violence: one counter-protester killed

Source: GQ

'Very fine people on both sides'

Source: slate.com

`I think Tuesday was the most important day in the White nationalist movement... I thought the news from this was done when I read that he had come back and he said there were good people in the White nationalist rally and he salvaged their message.'

--Former leading white nationalist
In: Daniels (2018), *Contexts*.

Objectives

- Unfolding of reactions to a landmark, divisive, extremist political event

- Structure of the online conversation (Twitter)
  - Who participated
  - Connections
  - Polarization

- Use the lens of network science to examine these questions

Approach

- Characterizing Twitter accounts based on online media preference

- Quantifying polarization in the Twitter conversation about Charlottesville in terms of media preference

- Identify key accounts that were particularly influential in shaping this discussion

- Find natural groupings ('communities') of accounts based on their Twitter interactions

- Community characterization in terms of account composition and tweet content

Data

- Tweets about #Charlottesville, 16 August – 21 August, 2017

- ~500,000 tweets sent by ~300,000 accounts

- Complete lists of users following 13 media accounts spanning the liberal-conservative spectrum

- Data collection: Twitter API; public data

Media followership data

- Complete follower lists for media accounts accessed in Nov-Dec 2016

- Fox News, Breitbart, Wall Street Journal, NPR, the Nation, dailykos ….

- Studies on media preferences / media slant: Budak et al (2016); Mitchell et al (2014); Gentzkow and Shapiro (2011)

- Number of followers ranged from 62,000+ (@csmonitor) to more than 12 million (@WSJ)

Media preferences and node characterization

- Principal components analysis on media followership of 13 media accounts

- This approach successfully used by others to measure ideology in Congress based on voting records (voteview.com)

- 99,412 accounts followed at least one of the 13 media sources at the time of media follower list access (December 2016)

- Other studies on node ideology: Bail et al (2018), Conover et al (2011), Colleoni et al (2014), many more

PCA on media choice

Media choice:

$$x_{ij} - 0/1 \text{ whether node } i \text{ follows media account } j$$

$$\bar{\boldsymbol{x}} - \text{ mean media choice vector}$$

$$n \text{ nodes, } m \text{ media accounts}$$

Centered observation matrix:

$$X \in \mathbb{R}^{n \times m}$$

$$X_{ij} = x_{ij} - \bar{x}_j$$

$$XX^T - \text{Covariance matrix}$$

Principal components: eigenvectors of covariance matrix

Principal components analysis and media choice: 'Left' and 'Right'

	1st	2nd	3rd
BreitbartNews	0.4071	0.2093	0.2691
DRUDGE_REPORT	0.3843	0.2775	0.3131
FiveThirtyEight	-0.152	0.045	0.1785
FoxNews	0.3779	0.3652	**-0.022**
MotherJones	-0.3115	0.0713	0.5159
NPR	-0.3945	0.2159	0.1303
NRO	0.097	0.1826	0.0985
WSJ	-0.1183	0.5802	**-0.3376**
csmonitor	-0.0235	0.0358	0.0356
dailykos	-0.1893	0.0727	0.3414
theblaze	0.2054	0.2236	0.1615
thenation	-0.2362	0.1279	0.4038
washingtonpost	-0.3321	0.4992	**-0.2887**

- 1st component: 'Left' / 'Right'

Retweet network

- G – weighted, directed graph

- Nodes – Twitter accounts

- Arc weight from j to i – number of times j retweeted i

- Largest connected component: 220,000+ nodes, ~400,000 retweets
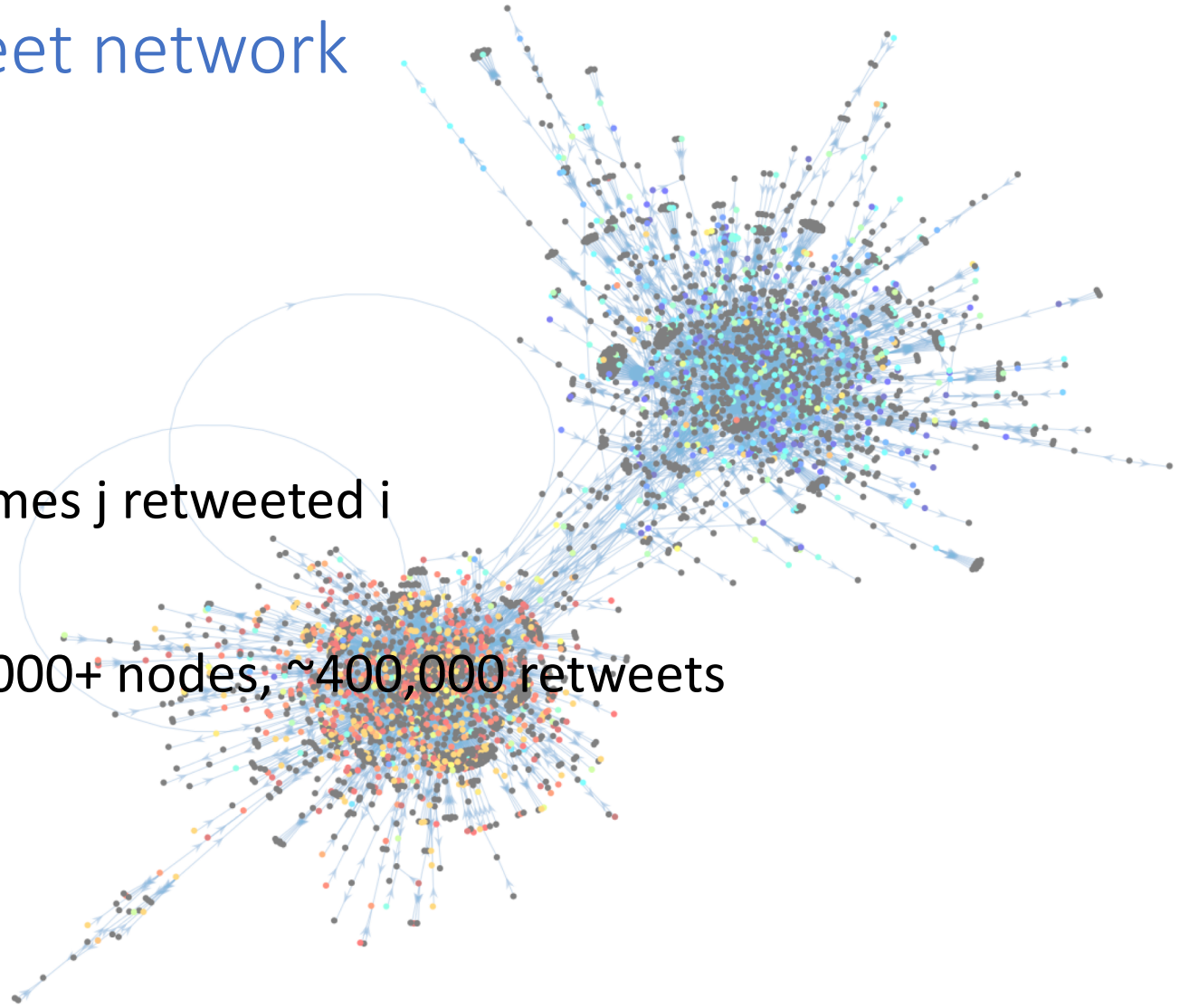
Assortativity by media preference

- How often do retweets occur between Left-Right?

	original	
retweeter	**Left**	**Right**
Left	0.43	0.057
Right	0.044	0.47

- Assortativity coefficient (Newman 2002, 2003)

- r = 0.80

  (Comparison: assortativity by ethnicity of 0.62 in sexual partner network; Catania et al 1992)

$$r = \frac{\sum_{\ell=1}^{g} e_{\ell\ell} - \sum_{\ell=1}^{g} a_\ell b_\ell}{1 - \sum_{\ell=1}^{g} a_\ell b_\ell}$$

Community detection

- Communities – " group[s] of nodes that are relatively densely connected to each other but sparsely connected to other dense groups in the network" (Porter et al 2009)

- Many different community detection methods exist

- A clustering problem

- Two approaches: modularity maximization and InfoMap

Modularity

Q: (fraction of edges within communities)
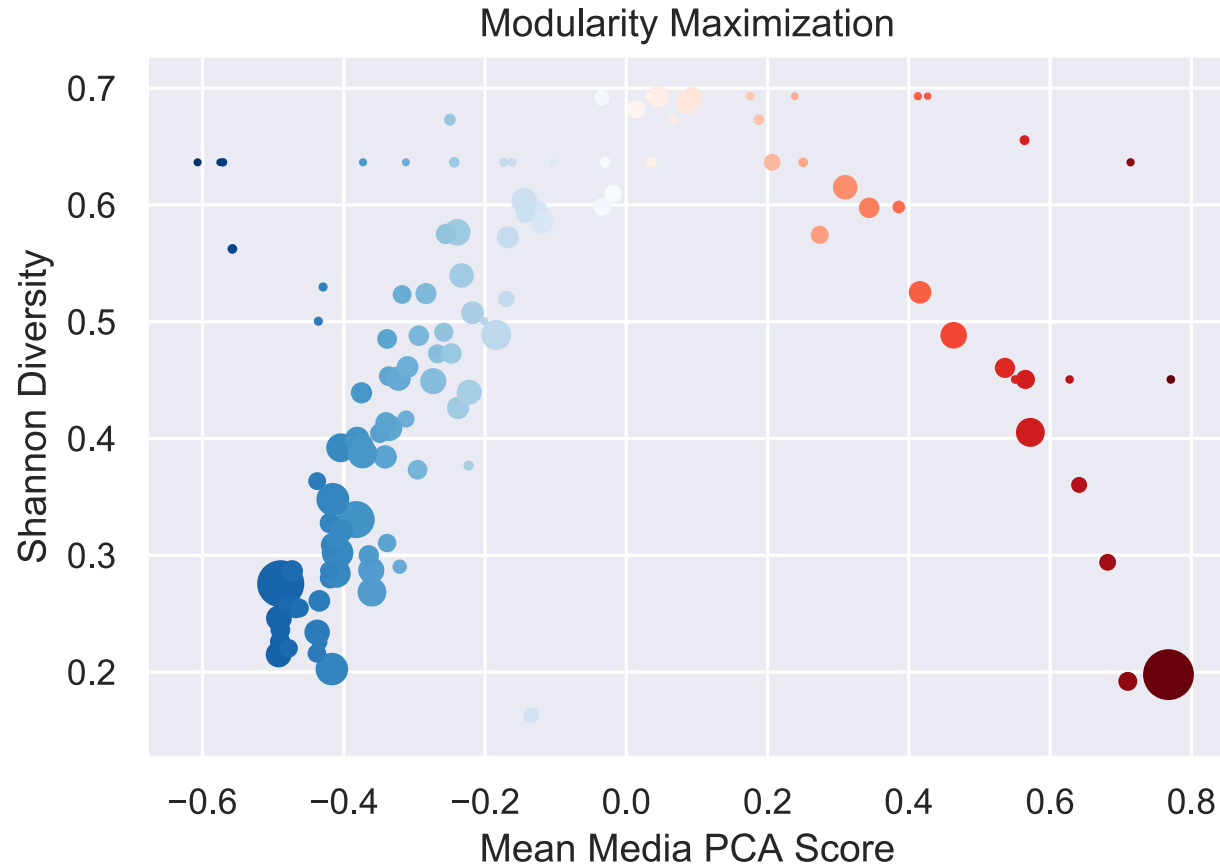   - (expected fraction via random wiring)

$$Q = \frac{1}{w} \sum_{i=1}^{n} \sum_{j=1}^{n} \left(A_{ij} - \gamma \frac{w_i^{\text{in}} w_j^{\text{out}}}{w} \right) \delta(C_i, C_j)$$

Undirected graphs: Newman and Girvan (2004), Phys Rev E.

Directed graphs: Arenas et al (2007), New J. Phys.;
                    Porter et al (2009)

Implementation: a Louvain method (Blondel et al 2008; NetWiki)
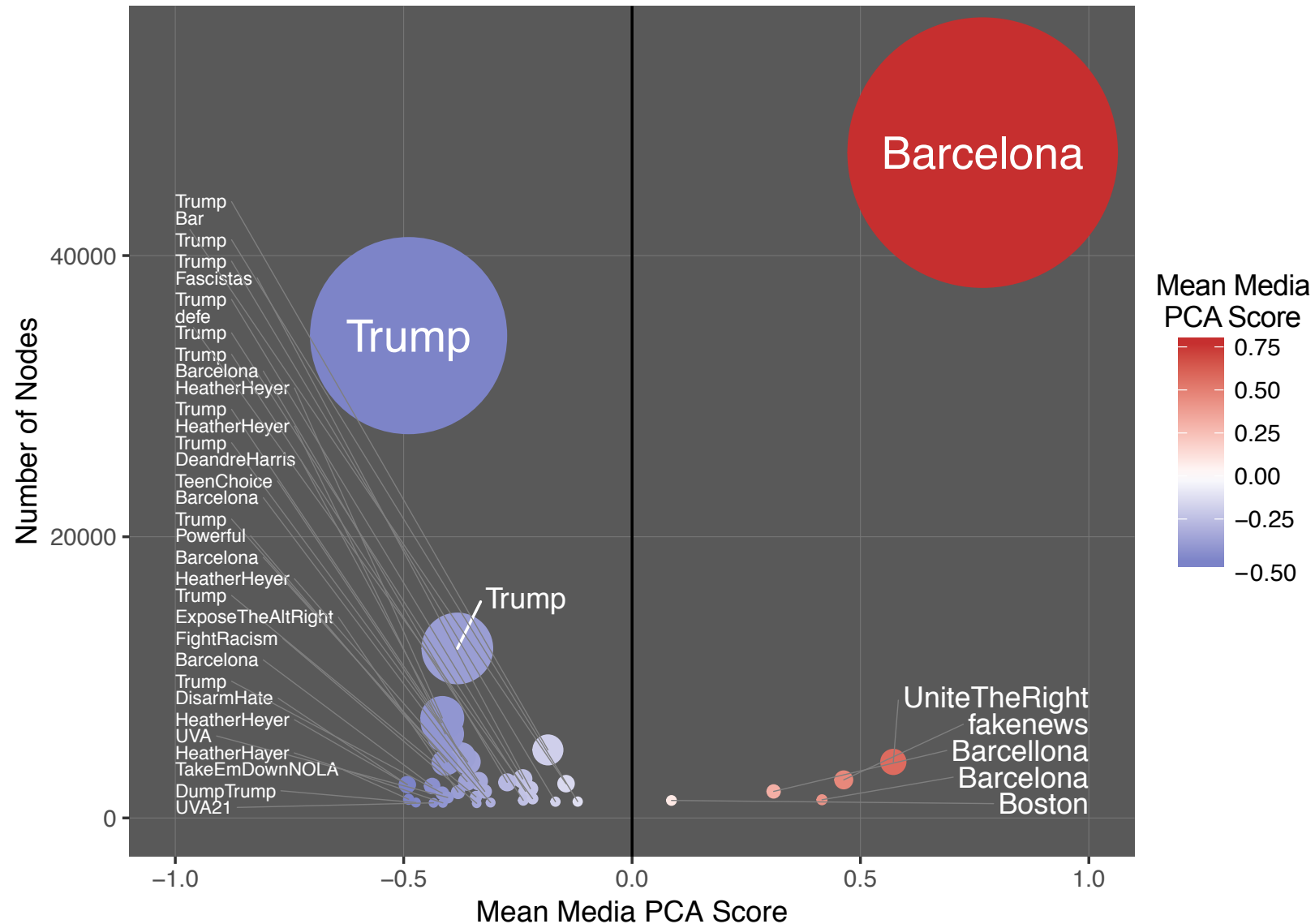
Community composition

- 228 communities detected using Louvain

- Size: from 2 to 47,321 nodes

- Largely homogeneous in terms of media PCA score

- Majority of large communities lean Left

- Largest community: Right (@FoxNews)

Shannon Diversity in community k:

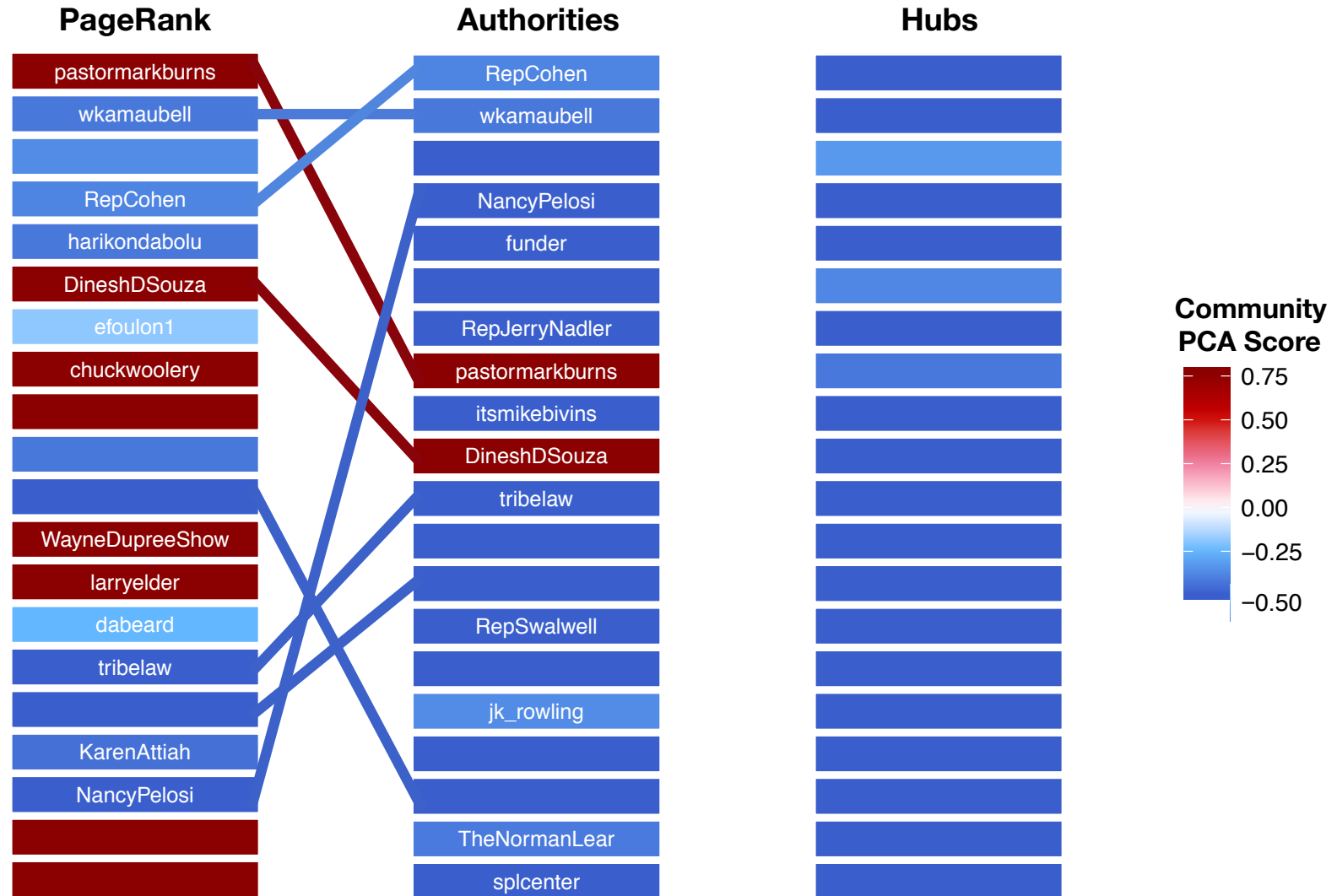$$H^k = -\sum_{i=1}^{2} p_i^k \ln p_i^k$$

Most common hashtag by community

Community characterization

- Large communities on the Left include accounts from disparate areas
  - E.g. Business (@benandjerrys), arts and entertainment (@jk_rowling, @ladygaga), activists (@IndivisibleTeam, @womensmarch), journalists, media (@CBSThisMorning), politicians (@RepCohen, @JoeBiden)

- Right communities:
  - Largest community dominated by conservative personalities and @FoxNews
  - Handful of other large communities
  - White supremacist and `alt-right' communities

- Framing of #Charlottesville very different between Left and Right

Central nodes

PageRank

pastormarkburns
wkamaubell

RepCohen
harikondabolu
DineshDSouza
efoulon1
chuckwoolery

WayneDupreeShow
larryelder
dabeard
tribelaw

KarenAttiah
NancyPelosi

Authorities

RepCohen
wkamaubell

NancyPelosi
funder

RepJerryNadler
pastormarkburns
itsmikebivins
DineshDSouza
tribelaw

RepSwalwell

jk_rowling

TheNormanLear
splcenter

Hubs

Community PCA Score

0.75
0.50
0.25
0.00
−0.25
−0.50

Conclusions

- Media followership and PCA is informative

- Marked polarization in retweet network about #Charlottesville

- More diversity in the Twitter conversation on Left than Right

- More hubs on Left than Right

Future directions

- Linking to offline studies

- Bots, sockpuppets, and influence campaigns

- Extending media PCA score
  - Continuous values
  - Additional components
  - 'Centrist' opinions

- Bridging

References

- Preprint: https://arxiv.org/abs/1905.07755

- susanbourbaki.com