

Characterizing the Twitter networks of prominent politicians and hate groups in the USA

Raazesh Sainudiin[†]

[†]Department of Mathematics
Uppsala University, Uppsala, Sweden
<http://lamastex.org/lmse/mep/>

Joint with: Yogeeswaran, Nash and Sahioun, Dept. of Psychology,
Univ of Canterbury, Christchurch, NZ

September 25, 2017

- 1 Questions and Experimental Design
- 2 Data and Statistics
 - Experimental design of twitter streams
- 3 Models and Methods
- 4 Results
- 5 Transmission Processes
 - The Transmission Process. Sainudiin and Welch, JTB 2016

Three Questions

- (Q1) Is Trump preferentially retweeted by various types of hate groups or their leadership relative to other politicians (i.e., Clinton, Sanders, Cruz, or Ryan) against the null random network model of *apathetic retweeting*?

Three Questions

- (Q1) Is Trump preferentially retweeted by various types of hate groups or their leadership relative to other politicians (i.e., Clinton, Sanders, Cruz, or Ryan) against the null random network model of *apathetic retweeting*?
- (Q2) What frequency of unique users retweeted both a politician and a hate group or its leadership more than expected under the null model?

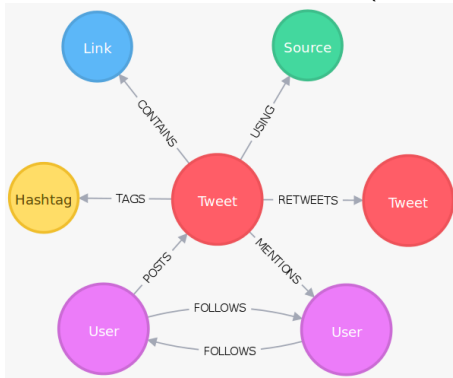
Three Questions

- (Q1) Is Trump preferentially retweeted by various types of hate groups or their leadership relative to other politicians (i.e., Clinton, Sanders, Cruz, or Ryan) against the null random network model of *apathetic retweeting*?
- (Q2) What frequency of unique users retweeted both a politician and a hate group or its leadership more than expected under the null model?
- (Q3) What is the joint distribution of the *degrees of separation* to each user from each of the five politicians and the eight most prolific hateful ideologies on Twitter, measured through the lengths of the most retweeted directed paths in the observed network?

Twitterverse

twitter is a micro-blogging service...

What is a tweet? retweet? reply-tweet, etc. (*status updates*)



Via public streams and REST APIs we collected ~22M status updates related to 5 politicians and 52 hate groups (retrospective REST-based network augmentations).

Hateful Networks

US Hate Groups by SPLC <https://www.splcenter.org/fighting-hate/extremist-files>

SPLC
Southern Poverty
Law Center

WHAT WE DO ▾ OUR ISSUES ▾ RESOURCES ▾ HATE MAP HATEWATCH Q

< RESOURCES | Case Docket | Extremist Files | Hatewatch | Intelligence Report | Publications | Hate Incidents

EXTREMIST FILES

Extremists in the U.S. come in many different forms – white nationalists, anti-gay zealots, black separatists, racist skinheads, neo-Confederates and more.

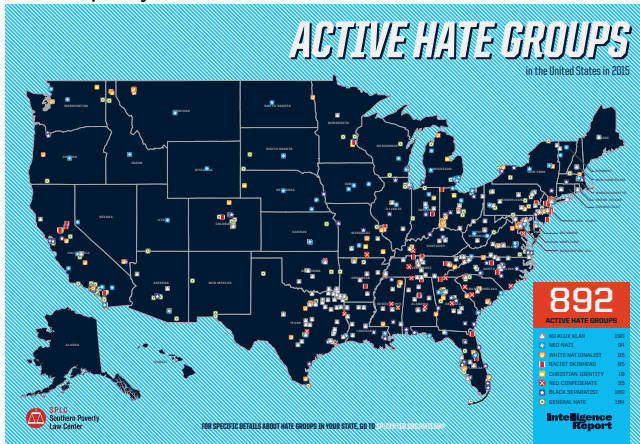
f t

The Extremist Files database contains profiles of various prominent extremists and extremist organizations. It also examines the histories and core beliefs – or ideologies – of the most common types of extremist movements. In addition, it illustrates connections between individuals, groups and extremist ideologies.

- <https://www.splcenter.org/fighting-hate/extremist-files/ideology>
- <https://www.splcenter.org/fighting-hate/extremist-files/individual>
- <https://www.splcenter.org/fighting-hate/extremist-files/groups>

Hateful Networks

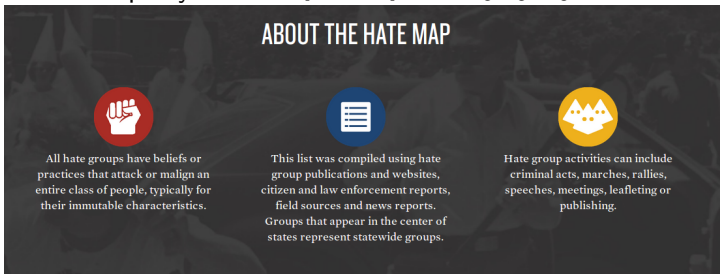
US Hate Groups by SPLC <https://www.splcenter.org/fighting-hate/extremist-files>




<https://www.splcenter.org/hate-map>


Hateful Networks


US Hate Groups by SPLC <https://www.splcenter.org/fighting-hate/extremist-files>



ABOUT THE HATE MAP

 All hate groups have beliefs or practices that attack or malign an entire class of people, typically for their immutable characteristics.

 This list was compiled using hate group publications and websites, citizen and law enforcement reports, field sources and news reports. Groups that appear in the center of states represent statewide groups.

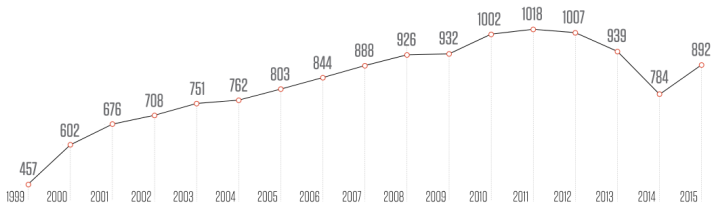
 Hate group activities can include criminal acts, marches, rallies, speeches, meetings, leafleting or publishing.

<https://www.splcenter.org/hate-map>

Hateful Networks

US Hate Groups by SPLC <https://www.splcenter.org/fighting-hate/extremist-files>

HATE GROUPS 1999-2015



<https://www.splcenter.org/hate-map>

Hateful Networks

18 US Hateful Ideologies by SPLC

<https://www.splcenter.org/fighting-hate/extremist-files>

Alternative Right

The Alternative Right, commonly known as the Alt-Right, is a set of far-right ideologies, groups and individuals whose core belief is that “white identity” is under attack by multicultural forces using “political correctness” and “social justice” to undermine white people and “their” civilization...



Anti-Immigrant

Anti-immigrant hate groups are the most extreme of the hundreds of nativist and vigilante groups that have proliferated since the late 1990s, when anti-immigration xenophobia began to rise to levels not seen in the United States since the 1920s.



Anti-LGBT

Opposition to equal rights for LGBT people has been a central theme of Christian Right organizing and fundraising for the past three decades – a period that parallels the fundamentalist movement’s rise to political power.



<https://www.splcenter.org/fighting-hate/extremist-files/ideology>

Hateful Networks

18 US Hateful Ideologies by SPLC

<https://www.splcenter.org/fighting-hate/extremist-files>

Anti-Muslim

Anti-Muslim hate groups are a relatively new phenomenon in the United States, most of them appearing in the aftermath of the World Trade Center terrorist attacks on Sept. 11, 2001. Earlier anti-Muslim groups tended to be religious in orientation and disputed Islam's status as a respectable religion.



Antigovernment Movement

The antigovernment movement has experienced a resurgence, growing quickly since 2008, when President Obama was elected to office. Factors fueling the antigovernment movement in recent years include changing demographics driven by immigration, the struggling economy and the election of the first...



Black Separatist

Black separatists typically oppose integration and racial intermarriage, and they want separate institutions -- or even a separate nation -- for blacks. Most forms of black separatism are strongly anti-white and anti-Semitic, and a number of religious versions assert that blacks are the Biblical "...






<https://www.splcenter.org/fighting-hate/extremist-files/ideology>

Hateful Networks

18 US Hateful Ideologies by SPLC

<https://www.splcenter.org/fighting-hate/extremist-files>

<u>Christian Identity</u>	
Christian Identity is a unique anti-Semitic and racist theology that rose to a position of commanding influence on the racist right in the 1980s. "Christian" in name only, the movement's relationship with evangelicals and fundamentalists has generally been hostile due to the latter's belief that...	
<u>General Hate</u>	
These groups espouse a variety of rather unique hateful doctrines and beliefs that are not easily categorized. Many of the groups are vendors that sell a miscellany of hate materials from several different sectors of the white supremacist movement.	
<u>Holocaust Denial</u>	
Deniers of the Holocaust, the systematic murder of around 6 million Jews in World War II, either deny that such a genocide took place or minimize its extent. These groups (and individuals) often cloak themselves in the sober language of serious scholarship, call themselves "historical revisionists..."	

<https://www.splcenter.org/fighting-hate/extremist-files/ideology>

Hateful Networks

18 US Hateful Ideologies by SPLC

<https://www.splcenter.org/fighting-hate/extremist-files>

Ku Klux Klan

The Ku Klux Klan, with its long history of violence, is the most infamous – and oldest – of American hate groups. Although black Americans have typically been the Klan's primary target, it also has attacked Jews, immigrants, gays and lesbians and, until recently, Catholics.



Neo-Confederate

The term neo-Confederacy is used to describe twentieth and twenty-first century revivals of pro-Confederate sentiment in the United States. Strongly nativist, neo-Confederacy claims to pursue Christianity and heritage and other supposedly fundamental values that modern Americans are seen to have...



Neo-Nazi

Neo-Nazi groups share a hatred for Jews and a love for Adolf Hitler and Nazi Germany. While they also hate other minorities, gays and lesbians and even sometimes Christians, they perceive "the Jew" as their cardinal enemy.



<https://www.splcenter.org/fighting-hate/extremist-files/ideology>

Hateful Networks

18 US Hateful Ideologies by SPLC

<https://www.splcenter.org/fighting-hate/extremist-files>

Phineas Priesthood

The Phineas Priesthood is not an actual organization; it has no leaders, meetings, or any other institutional apparatus.



Racist Music

Racist music groups are typically white power music labels that record, publish and distribute racist music in a variety of genres.



Racist Skinhead

Racist Skinheads form a particularly violent element of the white supremacist movement, and have often been referred to as the "shock troops" of the hoped-for revolution. The classic Skinhead look is a shaved head, black Doc Martens boots, jeans with suspenders and an array of typically racist...



<https://www.splcenter.org/fighting-hate/extremist-files/ideology>

Hateful Networks

18 US Hateful Ideologies by SPLC

<https://www.splcenter.org/fighting-hate/extremist-files>

Radical Traditional Catholicism

“Radical traditionalist” Catholics, who may make up the largest single group of serious anti-Semites in America, subscribe to an ideology that is rejected by the Vatican and some 70 million mainstream American Catholics.



Sovereign Citizens Movement

The strange subculture of the sovereign citizens movement, whose adherents hold truly bizarre, complex antigovernment beliefs, has been growing at a fast pace since the late 2000s. Sovereigns believe that they get to decide which laws to obey and which to ignore, and they don't think they should...



White Nationalist

White nationalist groups espouse white supremacist or white separatist ideologies, often focusing on the alleged inferiority of nonwhites. Groups listed in a variety of other categories - Ku Klux Klan, neo-Confederate, neo-Nazi, racist skinhead, and Christian Identity - could also be fairly...



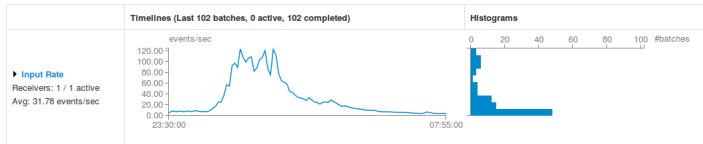
<https://www.splcenter.org/fighting-hate/extremist-files/ideology>

US Presidential Election 2016 - Twitter Streams

Twitter Data — 3rd US Presidential Debate

Streaming Statistics

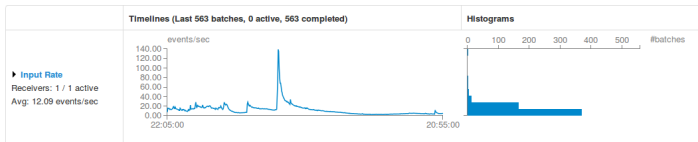
Running batches of 5 minutes for 8 hours 32 minutes 20 seconds since 2016/10/19 23:26:43 (102 completed batches, 972342 records)



Twitter Data — Last 2 Days Around the End of Election

Streaming Statistics

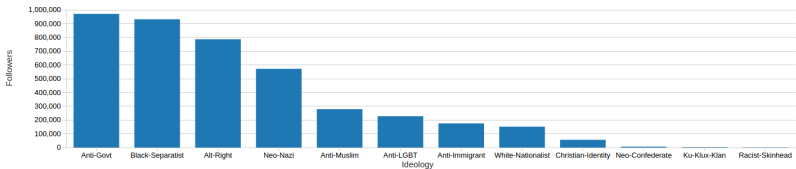
Running batches of 5 minutes for 1 day 22 hours 56 minutes since 2016/11/08 22:02:36 (563 completed batches, 2041501 records)



- public streams of @realDonaldTrump, @HillaryClinton, @BernieSanders, @tedcruz, SpeakerRyan and 52 splc-defined hategroups of their leadership
- collected data includes all mentions, replies, retweets, etc of these twitter accounts of interest for about 9 weeks around the 2016 US Presidential Election

12 SPLC-defined hateful ideologies

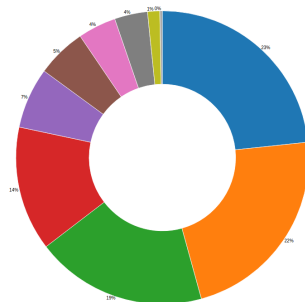
– only 78% of hategroups identified by SPLC were active in Twitter



Ideology	Followers
Anti-Govt	970769
Black-Separatist	931736
Alt-Right	786327
Neo-Nazi	571772
Anti-Muslim	279122
Anti-LGBT	227636
Anti-Immigrant	175441
White-Nationalist	151711
Christian-Identity	56191
Neo-Confederate	6628
Ku-Klux-Klan	3070
Racist-Skinhead	1826

Ideology

- Anti-Govt
- Black-Separatist
- Alt-Right
- Neo-Nazi
- Anti-Muslim
- Anti-LGBT
- Anti-Immigrant
- White-Nationalist
- Christian-Identity
- Others



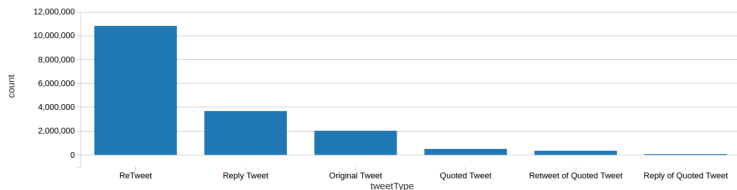
5 prominent Politicians in the USA

Retweet Network statistics of the five political accounts

Politician	in-degree	in-nbhd	out-degree	out-nbhd
Donald Trump	40	12	5,952,257	958,262
Hillary Clinton	225	121	2,774,111	943,995
Bernie Sanders	107	62	762,209	356,718
Paul Ryan	769	158	68,973	28,902
Ted Cruz	322	189	49,479	27,663

Dataset overview

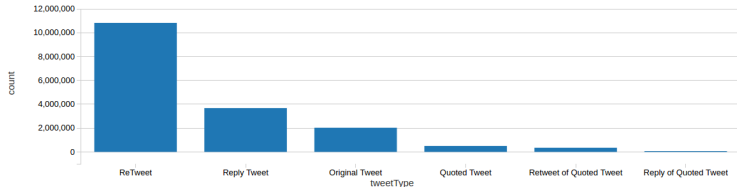
Data collected around the 2016 US Presidential Election



- data = 2.7M tweets, 13.7M retweets, 22M status updates

Dataset overview

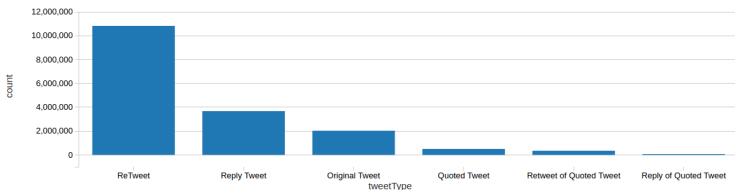
Data collected around the 2016 US Presidential Election



- data = 2.7M tweets, 13.7M retweets, 22M status updates
- 4.4M distinct retweet-pairs: (original-Tweeter, Retweeter)

Dataset overview

Data collected around the 2016 US Presidential Election



- data = 2.7M tweets, 13.7M retweets, 22M status updates
- 4.4M distinct retweet-pairs: (original-Tweeter, Retweeter)
- 2.5M unique users

Retweet Network — (3% sample #V = 1205, #E = 29856)

Trump-Clinton Retweet Network — a few samples

CPostUserSN	OPostUserSNinRT	OPostUserSNinQT	favouritesCount	followersCount	friendsCount	isVerified	isGeoEnabled	CurrentTweet
georgefayner	realDonaldTrump	null	137811	1466	953	false	true	RT @realDonaldTrump: China is cooking up conspiracy theories that the Olympics are rigged. http://t.co/0ah0hBJt They dont understand why...
KevinCormier10	realDonaldTrump	null	16164	505	367	false	true	RT @realDonaldTrump: EXCLUSIVE: FBI Agents Say Comey 'Stood In The Way' Of Clinton Email Investigation: https://t.co/6n63HVvNo
thuerta	realDonaldTrump	null	13081	126	345	false	true	RT @realDonaldTrump: "Trump rally disrupter was once on Clinton's payroll" https://t.co/75oLLuD4SI
tanladyvoitan	HillaryClinton	null	6316	101	200	false	true	RT @HillaryClinton: Our progress is on the ballot. Tolerance is on the ballot. Democracy is on the ballot. Make a plan to vote:...

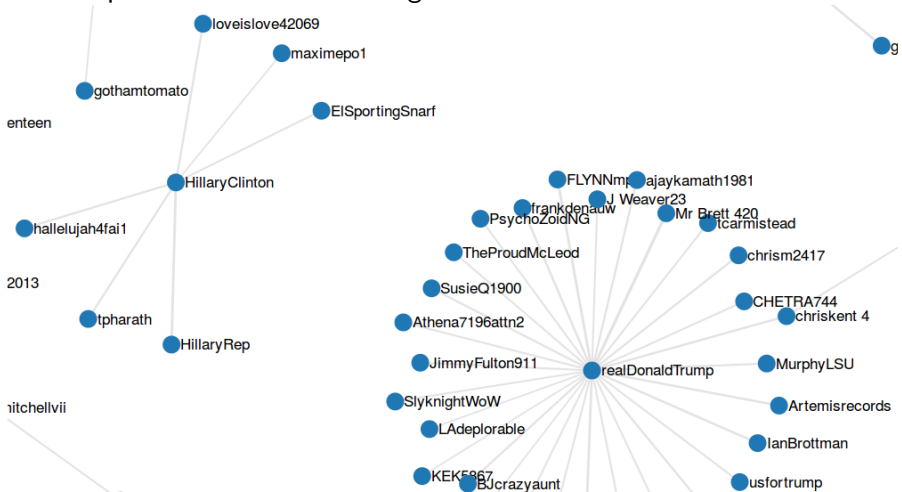
Retweet Network — (3% sample $\#V = 1205$, $\#E = 29856$)

Trump-Clinton Retweet Network weighted by Retweet counts

userCreatedAtDate	daysSinceUserCreated	OPostUser\$NameRT	CPostUser\$N	max(favouritesCount)	max(followersCount)	max(friendsCount)	RetweetCount
2011-12-13T19:10:28.000+0000	1781	realDonaldTrump	Mr_Brett_420	3294	78	194	190
2016-04-30T00:13:34.000+0000	181	HilaryClinton	HilaryRep	4196	2168	4984	158
2011-03-22T13:09:23.000+0000	2047	realDonaldTrump	FLYNmpc	1653	48	75	146
2014-08-25T17:02:46.000+0000	795	realDonaldTrump	mikenry2499	17427	183	155	132
2009-04-28T07:07:03.000+0000	2742	yottapoint	gcomking	5876	797	1826	120
2014-06-20T21:37:39.000+0000	861	BUILDseriesNYC	suzannebuzz	38604	1706	486	112
2009-05-28T15:51:31.000+0000	2710	realDonaldTrump	chriskent_4	838	254	85	112
2009-03-08T12:58:18.000+0000	2791	realDonaldTrump	Artemisrecords	2000	2777	5000	112
2012-09-25T15:09:37.000+0000	1494	realDonaldTrump	IanBrotzman	1	89	151	107
2011-03-31T00:54:09.000+0000	2038	realDonaldTrump	frankdenauw	43	45	18	102
2016-07-17T21:30:47.000+0000	103	HilaryClinton	lovelslove42069	3818	108	398	98
2015-09-01T18:52:06.000+0000	423	realDonaldTrump	BJcrazyaunt	1064	1296	1433	95
2011-12-24T03:52:02.000+0000	1770	HilaryClinton	lpharath	703	38	183	91
2015-03-08T23:47:05.000+0000	600	HilaryClinton	halleslqah4tai1	16786	227	270	88
2014-06-30T16:44:10.000+0000	851	realDonaldTrump	ajaykamat1981	3309	2667	3010	88
2012-04-29T21:49:38.000+0000	1643	realDonaldTrump	MurphyLSU	65	28	47	84
2010-08-05T18:02:11.000+0000	2276	realDonaldTrump	sdpubs	29674	123	34	83
2011-07-24T19:55:57.000+0000	1923	realDonaldTrump	chrism2417	3012	182	1112	81
2016-02-03T23:58:01.000+0000	268	realDonaldTrump	SusieQ1900	6797	386	415	81

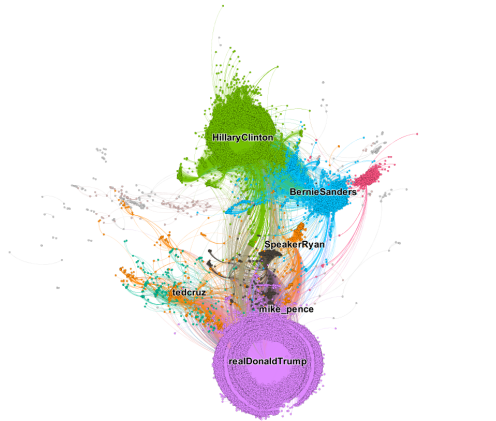
Retweet Network — (3% sample $\#V = 1205$, $\#E = 29856$)

Trump-Clinton Retweet Ideological Network



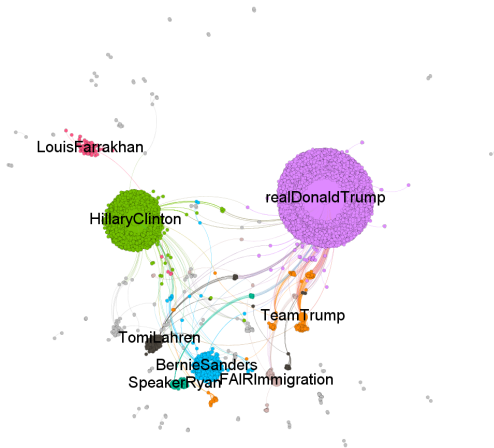
Strong Community Structure – samples of retweet networks

The 3rd US Presidential Debate 22K Retweet Network



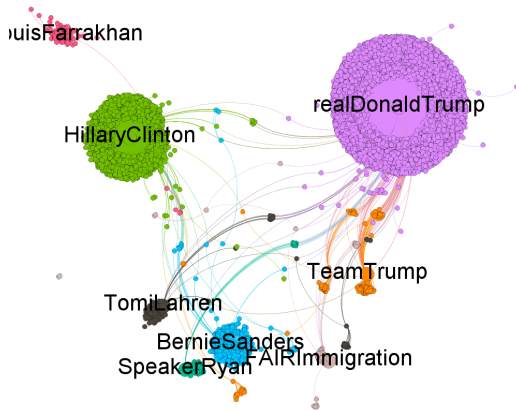
Strong Community Structure – samples of retweet networks

5% random sampled retweet networks for October 19 2016



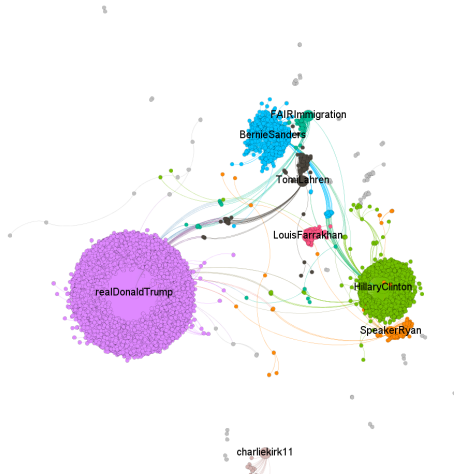
Strong Community Structure – samples of retweet networks

5% random sampled retweet networks for October 19 2016 – top 10



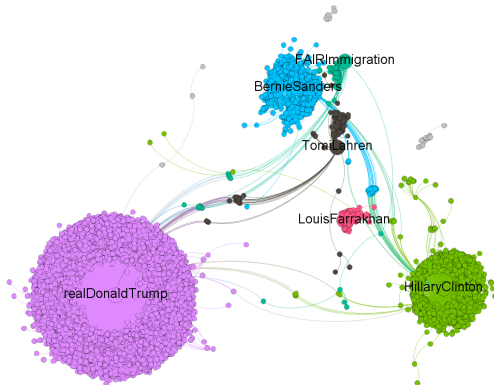
Strong Community Structure – samples of retweet networks

5% random sampled retweet networks for October 24 2016



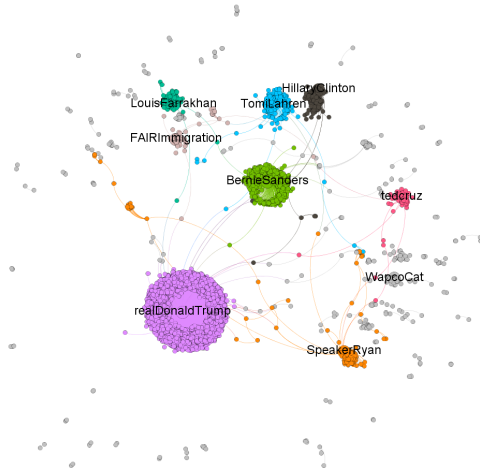
Strong Community Structure – samples of retweet networks

5% random sampled retweet networks for October 24 2016 – top 6



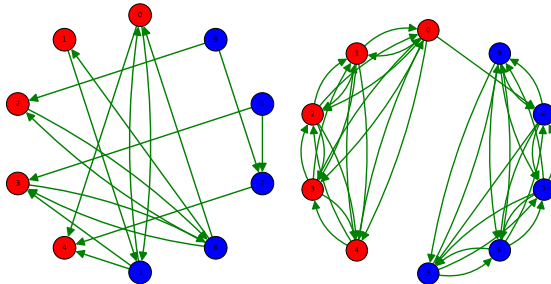
Strong Community Structure – samples of retweet networks

5% random sampled retweet networks for November 15 2016



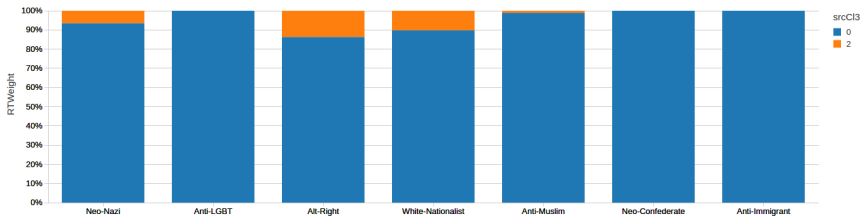
Models for Ideological Network Dynamics

- If arc $a_{i,j} = 1$ then we say i ideologically concurs with j



- Just two retweet networks out of 4, 722, 366, 482, 869, 645, 213, 696 for 9 individuals!
- We want indegree and outdegree conditioned random networks to preserve observed heterogeneity in individual Twitter activity levels
- This is the classical *random directed configuration model* – H_0 : *apathetic retweet network*
- NEED: distributed computing using Apache Spark (fastest growing Apache project)

7 SPLC-defined hateful ideologies Retweet Proportions

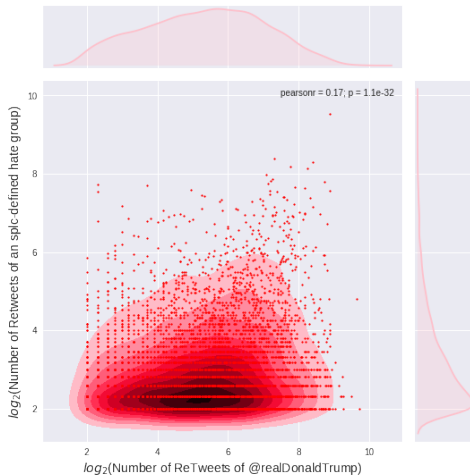


0 = Trump's cluster and 2 = Clinton's cluster

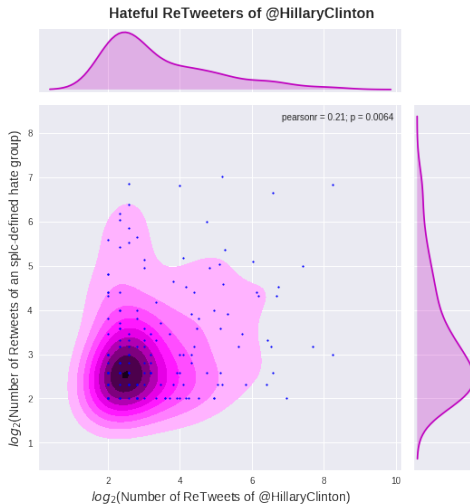
A significant proportion of retweets by leaders of seven extremist ideologies have original tweets in Trump's ideological cluster.

Trump's Hateful Retweeters

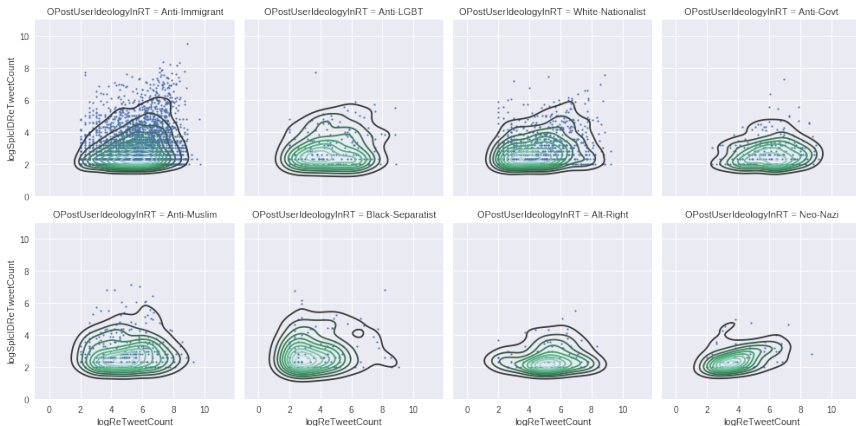
Hateful ReTweeters of @realDonaldTrump



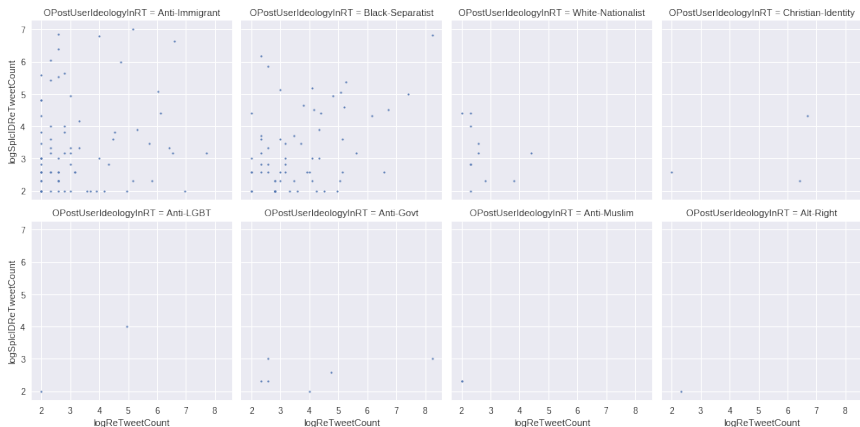
Clinton's Hateful Retweeters



Trump's Hateful Retweeters By Ideology



Clinton's Hateful Retweeters By Ideology



Chi-square tests - do NOT account for network heterogeneity

Ideology	Donald J Trump	Hillary R Clinton	Chi-Square Statistic	R ²
Alt-Right	1127 (90.7%)	116 (9.3%)	$X^2 = 822.30, p < .0001$	R ² =0.662
Anti-Government	1455 (89.5%)	171 (10.5%)	$X^2 = 1013.93, p < .0001$	R ² =0.623
Anti-Immigrant	15019 (88.6%)	1926 (11.4%)	$X^2 = 10116.65, p < .0001$	R ² =0.597
Anti-LGBT	1621 (88.6%)	209 (11.4%)	$X^2 = 1089.48, p < .0001$	R ² =0.595
Anti-Muslim	2293 (90.8%)	233 (9.2%)	$X^2 = 1679.97, p < .0001$	R ² =0.665
Black-Separatist	1279 (54.9%)	1049 (45.1%)	$X^2 = 22.72, p < .01$	R ² =0.009
Neo-Nazi	1039 (90.7%)	106 (9.3%)	$X^2 = 760.25, p < .0001$	R ² =0.664
White-Nationalist	5103 (89.2%)	616 (10.8%)	$X^2 = 3520.40, p < .0001$	R ² =0.616
Total	28992 (86.5%)	4509 (13.5%)	$X^2 = 18006.72, p < .0001$	R ² =0.540

Chi-square tests - do NOT account for network heterogeneity

Restricting to retweeters who retweet at least 4 times

Ideology	Donald J Trump	Hillary R Clinton	Chi-Square Statistic	R ²
Alt-Right	936 (98.7%)	12 (1.3%)	$X^2 = 900.61, p < .0001$	R ² = 0.950
Anti-Government	1388 (98.4%)	23 (1.6%)	$X^2 = 1320.50, p < .0001$	R ² = 0.936
Anti-Immigrant	12618 (96.6%)	442 (3.4%)	$X^2 = 11351.84, p < .0001$	R ² = 0.869
Anti-LGBT	1110 (96.0%)	46 (4.0%)	$X^2 = 979.32, p < .0001$	R ² = 0.847
Anti-Muslim	1866 (98.8%)	22 (1.2%)	$X^2 = 1801.03, p < .0001$	R ² = 0.954
Black-Separatist	494 (62.5%)	296 (37.5%)	$X^2 = 49.63, p < .001$	R ² = 0.062
Neo-Nazi	692 (99.4%)	4 (0.6%)	$X^2 = 680.09, p < .0001$	R ² = 0.977
White-Nationalist	3751 (98.0%)	76 (2.0%)	$X^2 = 3529.04, p < .0001$	R ² = 0.922
Total	22855 (96.1%)	921 (3.9%)	$X^2 = 20234.71, p < .0001$	R ² = 0.851

Cut-Permute-Rewire – distributed, scalable, and fault-tolerant sampler

CUTPERMUTEANDREWIRE generates sample networks from the *directed multi-edged self-looped random configuration model* (Newman, Strogatz and Watts, 2001):

- *cutting* the directed edges representing the retweets in our observed retweet network into out-bound and in-bound half edges,

The in-degree and out-degree of each node in the observed retweet network is preserved after these three steps.

Cut-Permute-Rewire – distributed, scalable, and fault-tolerant sampler

`CUTPERMUTEANDREWIRE` generates sample networks from the *directed multi-edged self-looped random configuration model* (Newman, Strogatz and Watts, 2001):

- *cutting* the directed edges representing the retweets in our observed retweet network into out-bound and in-bound half edges,
- *permuting* the in-bound half edges by sorting them according to pseudo-random numbers that are generated and associated with them and

The in-degree and out-degree of each node in the observed retweet network is preserved after these three steps.

Cut-Permute-Rewire – distributed, scalable, and fault-tolerant sampler

`CUTPERMUTEANDREWIRE` generates sample networks from the *directed multi-edged self-looped random configuration model* (Newman, Strogatz and Watts, 2001):

- *cutting* the directed edges representing the retweets in our observed retweet network into out-bound and in-bound half edges,
- *permuting* the in-bound half edges by sorting them according to pseudo-random numbers that are generated and associated with them and
- *rewiring* the original out-bound half edges with the permuted in-bound half edges using a distributed join.

The in-degree and out-degree of each node in the observed retweet network is preserved after these three steps.

(Q1) Relative frequency of retweets by any one of the hate groups or their leadership for any original tweet made by one of the politicians

Null distribution of the test statistic under the apathetic retweet network model.

Table 2. Relative frequency of retweets by any one of the hate groups or their leadership for any original tweet made by one of the politicians

Politician, observed test statistic: marginal interval for the region of acceptance at 0.001 significance level				
Donald Trump	Hillary Clinton	Bernie Sanders	Paul Ryan	Ted Cruz
0.987 : (0.6008,0.6013)	0 : (0.2708,0.2709)	0 : (0.0677,0.0682)	0 : (0.00411,0.00413)	0.0131 : (0.0024,0.0028)

(Q2) Number of unique users who retweeted a politician and a hate group at least five times each

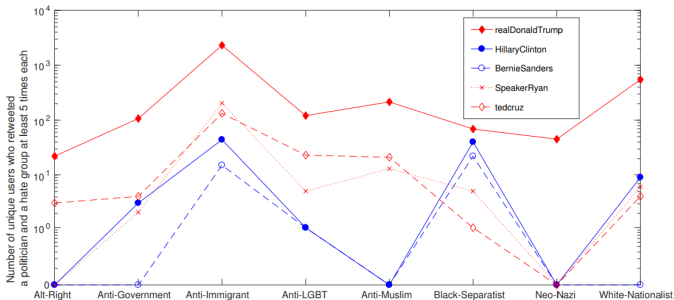


Fig. 1. Number of unique users who retweeted a politician and a hate group at least five times each (Note: The y-axis is in log-scale in powers of 10).

(Q2) Number of unique users who retweeted a politician and a hate group at least five times each

Null distribution of the test statistic under the apathetic retweet network model.

Table 3. Observed frequency of distinct users who retweeted a politician and a leader within a hate group at least 5 times each

Ideology	Politician				
	Donald Trump	Hillary Clinton	Bernie Sanders	Paul Ryan	Ted Cruz
	observed test statistic: marginal interval for the region of acceptance at 0.001 significance level				
Anti-Government	*107 : (0, 1)	3 : (0, 3)	0 : (0, 1)	*2 : (0, 1)	*4 : (0, 1)
Anti-Immigrant	*2314 : (375, 498)	°44 : (373, 492)	°15 : (369, 485)	*204 : (47, 95)	*133 : (18, 54)
Anti-LGBT	*121 : (0, 4)	1 : (0, 4)	1 : (0, 4)	*5 : (0, 3)	*23 : (0, 3)
Anti-Muslim	*215 : (0, 3)	0 : (0, 3)	0 : (0, 3)	*13 : (0, 3)	*21 : (0, 3)
Neo-Nazi	*45 : (0, 1)	0 : (0, 1)	0 : (0, 1)	0 : (0, 1)	0 : (0, 1)
White-Nationalist	*548 : (0, 12)	9 : (0, 10)	0 : (0, 10)	6 : (0, 8)	4 : (0, 7)
Black-Separatist	°69 : (653, 811)	°40 : (649, 808)	°22 : (645, 801)	°5 : (72, 128)	°1 : (28, 66)
Alternative-Right	*22 : (0, 0)	0 : (0, 0)	0 : (0, 0)	0 : (0, 0)	*3 : (0, 0)

Geometric retweet network & Most retweeted Paths

On Whiteboard.

- Model
- Pregel Vertex programs in Apache Spark's GraphX – Scalable Dijkstra for “Multiple landmarks to All Nodes Shortest Path Sequences”

Degrees of Separation From Hate Groups

Â Cumulative percentage of the monitored Twitter population who are within a given in-degree of separation from a politician and a hateful Ideology.

(Q3) Population ideological Tree & Degrees of Separation

Table 4. The top 15 groups of users according to their profiles of most retweeted path-lengths from the five politicians (DT = @realDonaldTrump, HC = @HillaryClinton, BS = @BernieSanders, PR = @SpeakerRyan, TC = @tedcruz) and eight hateful ideologies (AI = Anti-Immigrant, AM = Anti-Muslim, WN = White-Nationalist, AL = Anti-LGBT, AG = Anti-Govt, NN= Neo-Nazi, BIS=Black-Separatist, AR=Alt-Right) given by their id, frequency, percentage of population and their classification given by the ideological tree with leaf nodes as the ids.

ideological tree	id	frequency	percentage of population	Politician					Hate Group							
				DT	HC	BS	PR	TC	AI	AM	WN	AL	AG	NN	BIS	AR
	05	42853	02.005	1	1	2	4	4	5	5	7	6	4	7	7	7
	07	11481	00.537	1	2	1	4	4	5	5	7	6	4	7	7	7
	12	5868	00.274	1	1	1	4	4	5	5	7	6	4	7	7	7
	11	5972	00.279	4	2	3	5	7	8	8	9	9	7	10	10	10
	02	791286	37.016	3	1	2	4	6	7	7	8	8	6	9	9	9
	04	74126	03.468	3	1	1	4	6	7	7	8	8	6	9	9	9
	03	232093	10.857	3	2	1	6	6	7	7	9	8	6	9	9	9
	13	5173	00.242	3	1	1	6	6	7	7	8	8	6	9	9	9
	01	811586	37.965	1	4	7	4	4	5	5	7	6	4	7	7	7
	15	3892	00.182	1	4	7	1	4	5	5	7	3	4	7	7	7
	14	4011	00.188	1	4	7	4	4	1	5	3	5	4	5	7	7
	08	10460	00.489	3	5	9	1	3	3	3	5	3	6	7	9	9
	09	8069	00.377	3	3	3	3	1	4	3	3	3	6	5	6	9
	06	29997	01.403	2	3	3	3	3	5	3	3	5	5	5	3	3
	10	6257	00.293	1	3	3	4	4	5	3	3	5	4	5	3	3

Significance Statement

During the 2016 US presidential election, there was significant debate on whether Donald Trump's campaign was fuelled by hate and bigotry toward minority groups. We analyzed nearly 22 million communication events on Twitter to better understand the networks of retweeters of American hate groups and five key American politicians during the late stages of the election (Donald Trump, Hillary Clinton, Bernie Sanders, Ted Cruz, and Paul Ryan). Our data reveals that Twitter users linked to various American hate groups including Anti-Government, Anti-Immigrant, Anti-LGBT, Anti-Muslim, Neo-Nazi and White-Nationalist were more strongly linked to Trump over any other politician.

Some mathematical insights

Famous Mathematician: “Everyone’s gut feeling is that the network structure matters, ... but we know surprisingly very little about how it matters.”

Given the structure of these highly polarized ideological networks, can we make some idealizations to make progress on understanding **how “memes” are transmitted from one individual to another?**

Susceptible-Infected Contact Network (SICN) & Transmission Tree (TT)

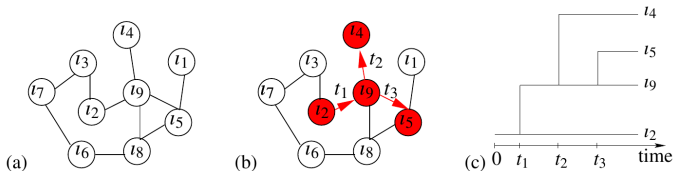


Figure 1: Spread of an epidemic over (a) the contact network of a population as shown by (b) a sub-network where edges representing transmission events are labelled by the time of event and the infected vertices are colored red and (c) the corresponding transmission tree.

Susceptible-Infected Contact Network (SICN) & Transmission Tree (TT)

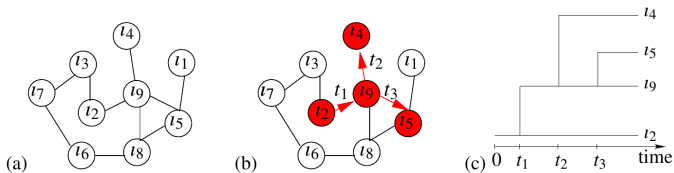


Figure 1: Spread of an epidemic over (a) the contact network of a population as shown by (b) a sub-network where edges representing transmission events are labelled by the time of event and the infected vertices are colored red and (c) the corresponding transmission tree.

Aldous' ?n: How does the geometry or structure of the SICN affect the distribution (shape and timing) of the TT?

Susceptible-Infected Contact Network (SICN) & Transmission Tree (TT)

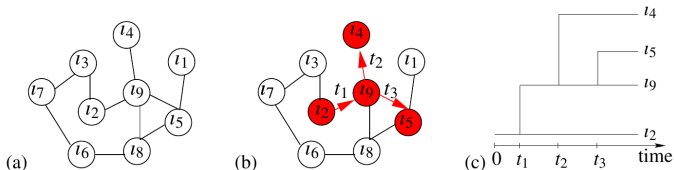


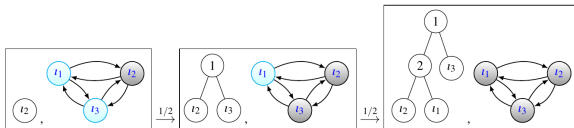
Figure 1: Spread of an epidemic over (a) the contact network of a population as shown by (b) a sub-network where edges representing transmission events are labelled by the time of event and the infected vertices are colored red and (c) the corresponding transmission tree.

Aldous' ?n: How does the geometry or structure of the SICN affect the distribution (shape and timing) of the TT?

Answer: It is possible in the simplest setting...

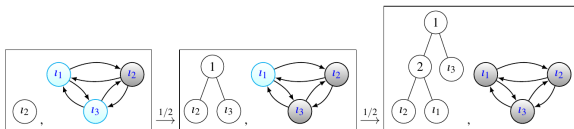
Markov chain on SI Contact Networks \times Transmission Trees

- A growing transmission tree on a **complete** SICN in a population of size $n = 3$

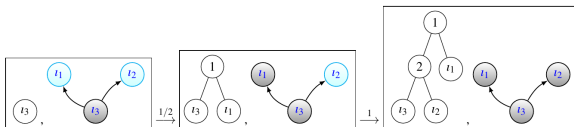


Markov chain on SI Contact Networks \times Transmission Trees

- A growing transmission tree on a **complete** SICN in a population of size $n = 3$

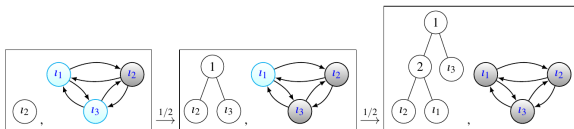


- A growing transmission tree on a **star** SICN in a population of size $n = 3$

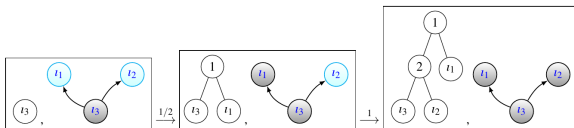


Markov chain on SI Contact Networks \times Transmission Trees

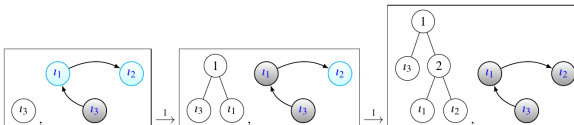
- A growing transmission tree on a **complete** SICN in a population of size $n = 3$



- A growing transmission tree on a **star** SICN in a population of size $n = 3$



- A growing transmission tree on a **path** SICN in a population of size $n = 3$



State Space

- Let $\mathbb{I}_n = \{i_1, i_2, \dots, i_n\}$ be the label set of a pop. of size n

State Space

- Let $\mathbb{I}_n = \{i_1, i_2, \dots, i_n\}$ be the label set of a pop. of size n
- Let w_n be the weighted edges of a complete weighted directed graph (network) k_n

State Space

- Let $\mathbb{I}_n = \{i_1, i_2, \dots, i_n\}$ be the label set of a pop. of size n
- Let w_n be the weighted edges of a complete weighted directed graph (network) k_n
- Let the Markov chain have state space $\mathcal{T}_n \times \mathcal{C}_n$

State Space

- Let $\mathbb{I}_n = \{i_1, i_2, \dots, i_n\}$ be the label set of a pop. of size n
- Let w_n be the weighted edges of a complete weighted directed graph (network) k_n
- Let the Markov chain have state space $\mathcal{T}_n \times \mathcal{C}_n$
 - where $c := (w, s) \in \mathcal{C}_n := 2^{w_n} \times \{0, 1\}^{\mathbb{I}_n}$, SICNs

State Space

- Let $\mathbb{I}_n = \{i_1, i_2, \dots, i_n\}$ be the label set of a pop. of size n
- Let w_n be the weighted edges of a complete weighted directed graph (network) k_n
- Let the Markov chain have state space $\mathcal{T}_n \times \mathcal{C}_n$
 - where $c := (w, s) \in \mathcal{C}_n := 2^{w_n} \times \{0, 1\}^{\mathbb{I}_n}$, SICNs
 - where $\tau \in \{\text{rooted planar ranked leaf-labelled binary trees}\}$

State Space

- Let $\mathbb{I}_n = \{v_1, v_2, \dots, v_n\}$ be the label set of a pop. of size n
- Let w_n be the weighted edges of a complete weighted directed graph (network) k_n
- Let the Markov chain have state space $\mathcal{T}_n \times \mathcal{C}_n$
 - where $c := (w, s) \in \mathcal{C}_n := 2^{w_n} \times \{0, 1\}^{\mathbb{I}_n}$, SICNs
 - where $\tau \in \{\text{rooted planar ranked leaf-labelled binary trees}\}$
 - Note the poset on 2^{w_n} with unit weights given by $\prec := \subseteq$

State Space

- Let $\mathbb{I}_n = \{v_1, v_2, \dots, v_n\}$ be the label set of a pop. of size n
- Let w_n be the weighted edges of a complete weighted directed graph (network) k_n
- Let the Markov chain have state space $\mathcal{T}_n \times \mathcal{C}_n$
 - where $c := (w, s) \in \mathcal{C}_n := 2^{w_n} \times \{0, 1\}^{\mathbb{I}_n}$, SICNs
 - where $\tau \in \{\text{rooted planar ranked leaf-labelled binary trees}\}$
 - Note the poset on 2^{w_n} with unit weights given by $\prec := \subseteq$
 - So the current state of the Markov chain at discrete time z is $(\tau(z), c(z)) \in \mathcal{T}_n \times \mathcal{C}_n$

Transition Probabilities

- One-step transitions for the jump chain

$$\Pr\{(\tau(z+1), c(z+1)) \mid (\tau(z), c(z))\} =$$

the edge-weight from $(z+1)$ -th infector to the $(z+1)$ -th infectee

Sum of edge-weights from every potential infector to every potential infectee within its susceptible out-neighborhood

Transition Probabilities

- One-step transitions for the jump chain

$$\Pr\{(\tau(z+1), c(z+1)) \mid (\tau(z), c(z))\} =$$

the edge-weight from $(z+1)$ -th infector to the $(z+1)$ -th infectee

Sum of edge-weights from every potential infector to every potential infectee within its susceptible out-neighborhood

- By letting the time for each infection event to be distributed as $\overset{iid}{\sim} \text{Exponential}(\lambda)$ random variables we can get the continuous time Markov chain's generator in the usual way (ignored here).

Transition Probabilities

- One-step transitions for the jump chain

$$\Pr\{(\tau(z+1), c(z+1)) \mid (\tau(z), c(z))\} =$$

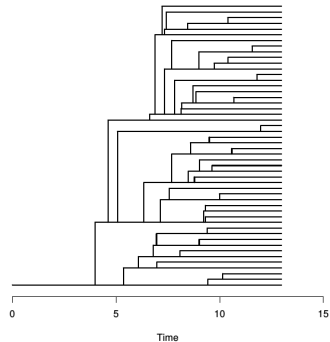
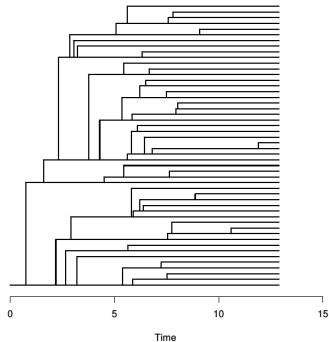
the edge-weight from $(z+1)$ -th infector to the $(z+1)$ -th infectee

Sum of edge-weights from every potential infector to every potential infectee within its susceptible out-neighborhood

- By letting the time for each infection event to be distributed as $\overset{iid}{\sim} \text{Exponential}(\lambda)$ random variables we can get the continuous time Markov chain's generator in the usual way (ignored here).
- NOTE: We limit to connected networks with unit weights and undirected edges here

Continuous Time Transmission Process

Two Transmission Trees (TTs) Grown on a Complete Susceptible-Infected Contact Network (SICN) with $n = 50$ individuals



Continuous Time Transmission Process

Branch-lengths of the TTs Grown on Complete SICNs randomly shifted logistic limit

for eg. (Aldous, 2013, Eq. 7.13)):

$$T_{[un]} - \log n \xrightarrow{d} F^{-1}(u) + G, \quad 0 < u < 1,$$

where, F is the logistic function:

$$F(t) = \frac{\exp(t)}{1 + \exp(t)}, \quad -\infty < t < \infty$$

and G has Gumbel distribution with $\Pr(G < x) = \exp(e^{-x})$.

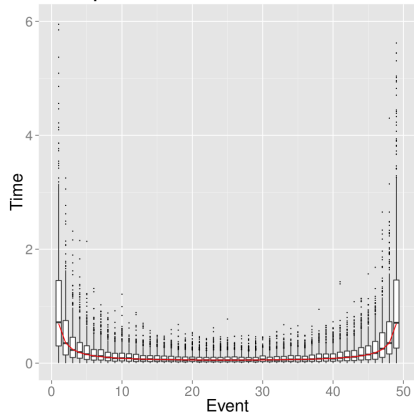


Fig. 6 The sampling distribution of T_z , branch-lengths (times in y-axis) of the transmission tree when there are exactly z infected individuals or between the $(z-1)$ -th and z -th infection event (x-axis), where $z \in \{1, 2, \dots, n-1\}$, from 500 independent simulations of the transmission tree over the complete SI contact network for a population of size $n = 50$ (as box plots) and the median branch-lengths given by $E(T_z) \log 2 = (\lambda z(n-z))^{-1} \log 2$, with $\lambda = 1/(n-1)$ (as red solid line).

Continuous Time Transmission Process

What is the distribution of transmission trees for an essentially arbitrary contact network?

Transmission Trees on Other Contact Networks

Mean branch-lengths for star, path and complete networks

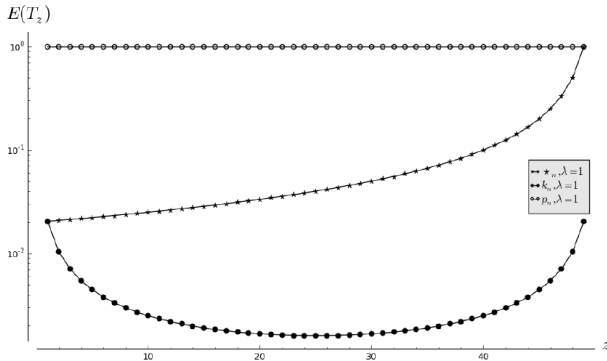
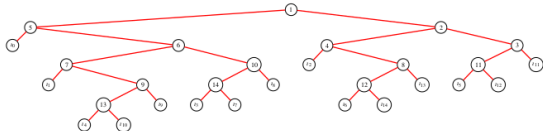
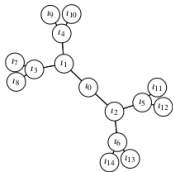
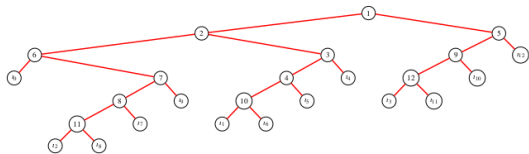
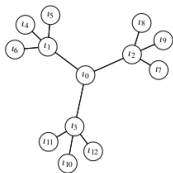


Figure 5: Expected branch-lengths when there are z infection events or $z + 1$ infected individuals, $E(T_z)$, for the three cases. Here $n = 50$ and $\lambda = 1$. $E(T_z) = 1/\lambda = 1$ with the path network p_n of Sect. 2.1.3 $E(T_z) = 1/(\lambda(n - z)) = 1/(50 - z)$ with the star network \star_n of Sect. 2.1.2 and $E(T_z) = 1/(\lambda z(n - z)) = 1/(z(50 - z))$ with the complete network k_n of Sect. 2.1.1. The curves in [1, 2] are for $n = 1, \dots, 40$.

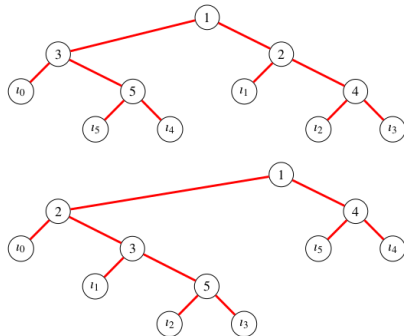
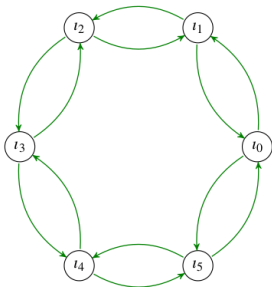
Transmission Trees on Other Contact Networks

Trees on Balanced Tree Networks



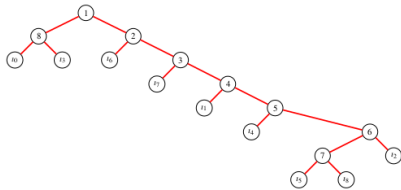
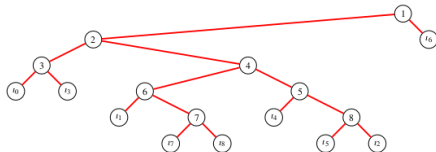
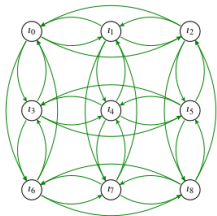
Transmission Trees on Other Contact Networks

Trees on Bidirectional Circular Networks



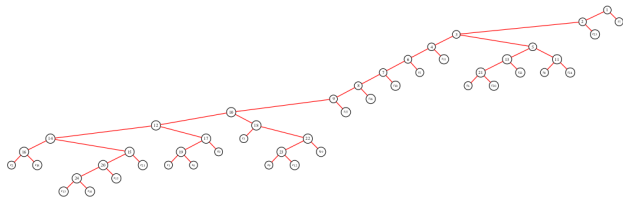
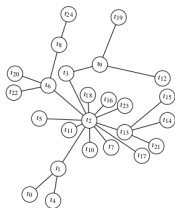
Transmission Trees on Other Contact Networks

Trees on Toroidal Network



Transmission Trees on Other Contact Networks

Trees on Preferential Attachment Network



Transmission Trees on Other Contact Networks

Real-world networks are quite heterogeneous and are closer to mixtures of various families of deterministic and random contact networks...

What is the distribution of transmission trees for an essentially arbitrary contact network?

Beta-splitting Model

- IDEA: induce distributions on TTs without the SICN

Beta-splitting Model

- IDEA: induce distributions on TTs without the SICN
- Consider Generating Sequences:

Beta-splitting Model

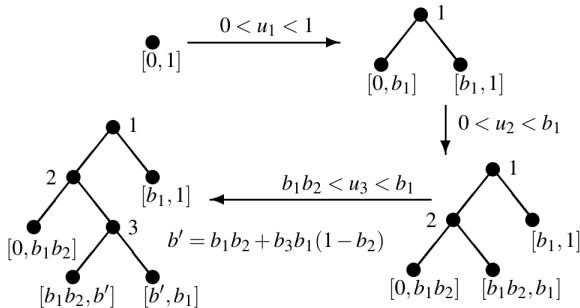
- IDEA: induce distributions on TTs without the SICN
- Consider Generating Sequences:
 - $U_1, U_2, \dots \stackrel{iid}{\sim} \text{Uniform}(0, 1)$

Beta-splitting Model

- IDEA: induce distributions on TTs without the SICN
- Consider Generating Sequences:
 - $U_1, U_2, \dots \stackrel{iid}{\sim} \text{Uniform}(0, 1)$
 - $B_1, B_2, \dots \stackrel{iid}{\sim} \text{Beta}(\alpha + 1, \beta + 1), (\alpha, \beta) \in (-1, \infty)^2$

Beta-splitting Model

- IDEA: induce distributions on TTs without the SICN
- Consider Generating Sequences:
 - $U_1, U_2, \dots \stackrel{iid}{\sim} \text{Uniform}(0, 1)$
 - $B_1, B_2, \dots \stackrel{iid}{\sim} \text{Beta}(\alpha + 1, \beta + 1), (\alpha, \beta) \in (-1, \infty)^2$
- Beta-splitting construction:



Theorems

Integrating out the interval-valued realizations at the leaf nodes

Theorem 1. *The probability of any discrete transmission tree $\tau(m)$ with m splits and $m + 1$ leaves under the integrated Beta-splitting model is:*

$$\begin{aligned} \Pr\{\tau(m)\} &= \prod_{z=1}^m \left\{ \frac{1}{B(\alpha+1, \beta+1)} \int_0^1 b_z^{s_z^L + \alpha} (1-b_z)^{s_z^R + \beta} db_z \right\} \times \Pr(\text{leaf labels}) \\ &= \prod_{z=1}^m \frac{B(s_z^L + \alpha + 1, s_z^R + \beta + 1)}{B(\alpha + 1, \beta + 1)} \times \Pr(\text{leaf labels}), \end{aligned} \quad (3.4)$$

$$= \prod_{z=1}^m \left(\frac{\prod_{j=0}^{s_z^R} \frac{\beta+j}{\beta+j+\alpha} \prod_{i=0}^{s_z^L} \frac{\alpha+i}{\alpha+i+\beta+s_z^R+1}}{\frac{\alpha\beta}{(\alpha+\beta)(\alpha+\beta+1)}} \right) \times \Pr(\text{leaf labels}), \quad (3.5)$$

Theorems

Beta-splitting model matches distrn on TTs for three example SICNs

- $(\alpha, \beta) = (0, 0) \equiv$ complete SICN,
- $(\alpha, \beta) \rightarrow (\infty, -1) \equiv$ star SICN
- $(\alpha, \beta) \rightarrow (-1, \infty) \equiv$ path SICN
- Theorem 2 on MLE expressions
- Theorem 3 on Equivalence class of initialized SICNs with the same (α, β) -specified TT distribution
- Will present a chalk talk in CIM on Feb 14 2017...
- 50 other model parameters simulated...

MLE of α and β from TTs under various SICNs

mean MLEs based on transmission trees simulated from various contact networks indexed by their ID from Table 1.

Beta-projections of various models

ID	Contact network	n	r	trials	$\bar{\alpha}$ (s.e.)	$\bar{\beta}$ (s.e.)
1	Complete	1,000	1	5	-0.006952 (0.06853)	0.05208 (0.1005)
2	Star	1,000	1	5	∞ (0.0000)	-1.0000 (0.0000)
3	Path	1,000	1	5	-1.0000 (0.0000)	∞ (0.0000)
4	Bidirectional Circular	50	1	5	-0.9880 (0.0006)	1.4584 (0.1534)
5	Bidirectional Circular	50	100	5	-0.9879 (0.0000)	1.5189 (0.0067)
6	BalancedTree(2,9)	1023	1	5	-0.4052 (0.0000)	-0.1477 (0.0000)
7	BalancedTree(3,6)	1093	1	5	-0.06452 (0.0000)	-0.5215 (0.0000)
8	BalancedTree(4,5)	1365	1	5	0.06556 (0.0000)	-0.7109 (0.0000)
9	BalancedTree(6,4)	1555	1	5	0.2350 (0.0000)	-0.8510 (0.0000)
10	BalancedTree(10,3)	1111	1	5	0.9249 (0.0000)	-0.9156 (0.0000)
11	BalancedTree(32,2)	1057	1	5	1.1624 (0.0000)	-0.9853 (0.0000)
12	BalancedTree(999,1)	1000	1	5	∞ (0.0000)	-1.0000 (0.0000)

MLE of α and β from TTs under various SICNs

mean MLEs based on transmission trees simulated from various contact networks indexed by their ID from Table 1.

Beta-projections of various models

13	2D toroidal grid	1024	1	5	-0.8612 (0.008425)	-0.5606 (0.03219)
14	2D toroidal grid	10000	1	5	-0.89346 (0.0022)	-0.6626 (0.0106)
15	3D toroidal grid	1000	1	5	-0.6849 (0.01479)	-0.3515 (0.03451)
16	3D toroidal grid	10648	1	5	-0.7628 (0.007956)	-0.4968 (0.01641)
17	ER(100,0.030)	100	30	5	-0.6063 (0.01383)	-0.4052 (0.02710)
18	ER(100,0.040)	100	30	5	-0.5179 (0.01855)	-0.3151 (0.02244)
19	ER(100,0.050)	100	30	5	-0.4059 (0.02020)	-0.2246 (0.01952)
20	ER(100,0.10)	100	30	5	-0.1997 (0.03106)	-0.1280 (0.03063)
21	ER(100,0.20)	100	30	5	-0.1074 (0.03961)	-0.06166 (0.03020)
22	ER(100,0.40)	100	30	5	0.02247 (0.06603)	0.01541 (0.05499)
23	ER(100,0.64)	100	30	5	-0.01097 (0.03984)	0.01046 (0.05112)
24	ER(100,1.0)	100	30	5	-0.001787 (0.04347)	-0.01555 (0.04019)

MLE of α and β from TTs under various SICNs

mean MLEs based on transmission trees simulated from various contact networks indexed by their ID from Table 1.

Beta-projections of various models

25	RandReg(1000,3)	1000	1	5	-0.7504 (0.004186)	-0.06260 (0.06322)
26	RandReg(1000,4)	1000	1	5	-0.5530 (0.04513)	-0.002305 (0.09785)
27	RandReg(1000,6)	1000	1	5	-0.3520 (0.03464)	0.06042 (0.06586)
28	RandReg(1000,10)	1000	1	5	-0.1939 (0.06167)	0.07274 (0.1238)
29	RandReg(1000,100)	1000	1	5	0.06378 (0.04519)	0.1084 (0.05844)
30	RandReg(1000,999)	1000	1	5	-0.01496 (0.08893)	0.006464 (0.04166)
31	SWRN ^{*,o} (50,2,0.0)	50	30	5	-0.9878 (0.0001516)	1.514 (0.01222)
32	SWRN [*] (50,2,0.1)	50	30	5	-0.9618 (0.003047)	-0.4147 (0.03203)
33	SWRN ^o (50,2,0.1)	50	30	5	-0.9652 (0.002863)	-0.3828 (0.1171)
34	SWRN [*] (50,2,0.2)	50	30	5	-0.9375 (0.004620)	-0.5683 (0.0193)
35	SWRN [*] (50,2,0.5)	50	30	5	-0.8632 (0.008181)	-0.6471 (0.03722)
36	SWRN [*] (50,5,0.1)	50	30	5	-0.7530 (0.01572)	-0.4751 (0.04671)

MLE of α and β from TTs under various SICNs

mean MLEs based on transmission trees simulated from various contact networks indexed by their ID from Table 1.

Beta-projections of various models

36	SWRN [*] (50,5,0.1)	50	30	5	-0.7530 (0.01572)	-0.4751 (0.04671)
37	SWRN ^o (50,5,0.1)	50	30	5	-0.7918 (0.01596)	-0.5130 (0.03323)
38	SWRN ^o (50,5,0.2)	50	30	5	-0.6881 (0.03277)	-0.3595 (0.06002)
39	SWRN ^o (50,5,0.5)	50	30	5	-0.5264 (0.04687)	-0.2138 (0.09471)
40	SWRN ^o (100,2,0.2)	100	1	5	-0.9479 (0.01509)	-0.3991 (0.5065)
41	SWRN ^o (100,2,0.2)	100	30	5	-0.9493 (0.003869)	-0.6027 (0.03475)
42	SWRN [*] (100,2,0.5)	100	1	5	-0.9023 (0.03411)	-0.7139 (0.03929)
43	SWRN [*] (100,2,0.5)	100	30	5	-0.8878 (0.006687)	-0.6821 (0.02189)
44	SWRN ^o (100,2,0.5)	100	1	5	-0.8714 (0.05584)	-0.6533 (0.09257)
45	SWRN ^o (100,2,0.5)	100	30	5	-0.8920 (0.005128)	-0.6786 (0.02189)
46	SWRN ^o (100,5,0.99)	100	30	5	-0.5079 (0.02371)	-0.2290 (0.03059)
47	SWRN ^o (100,10,0.99)	100	30	5	-0.2027 (0.07641)	-0.05611 (0.06949)

MLE of α and β from TTs under various SICNs

mean MLEs based on transmission trees simulated from various contact networks indexed by their ID from Table 1.

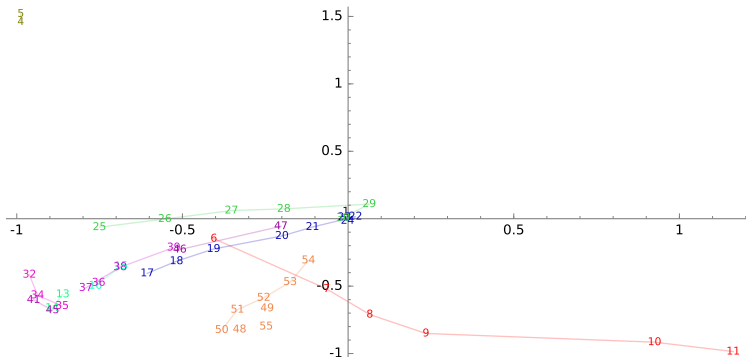
Beta-projections of various models

48	PrefAttach*(100, 1)	100	30	10	-0.3275 (0.04932)	-0.8215 (0.01121)
49	PrefAttach*(100, 2)	100	30	10	-0.2443 (0.03283)	-0.6647 (0.01294)
50	PrefAttach°(100, 1)	100	30	10	-0.3813 (0.04908)	-0.8254 (0.005460)
51	PrefAttach°(100, 2)	100	30	10	-0.3339 (0.03884)	-0.6743 (0.01657)
52	PrefAttach°(100, 3)	100	30	10	-0.2545 (0.04181)	-0.5863 (0.01652)
53	PrefAttach°(100, 5)	100	30	10	-0.1748 (0.04214)	-0.4698 (0.03110)
54	PrefAttach°(100, 10)	100	30	10	-0.1196 (0.03449)	-0.3089 (0.02663)
55	PrefAttach°(100, 1)	100	1	5	-0.2472 (0.2698)	-0.7993 (0.05843)

MLE of α and β from TTs under various SICNs

mean MLEs based on transmission trees simulated from various contact networks indexed by their ID from Table 1.

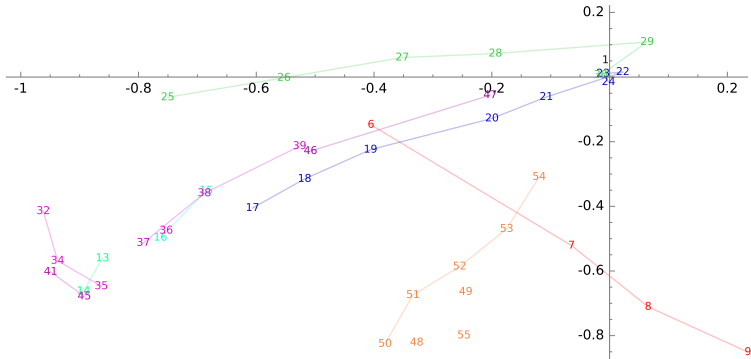
Beta-projections of various models



MLE of α and β from TTs under various SICNs

mean MLEs based on transmission trees simulated from various contact networks indexed by their ID from Table 1.

Beta projections of various models



So what? — We can do Bayesian non-parametrics by Beta-splitting mixtures over empirical SICNs/TINs

The End

Many thanks to:

- Databricks Academic Partners Programme and AWS Educate grant
- Research Chair in Mathematical Models of Biodiversity (for mathematical theorizing) held jointly by:
 - 1 Veolia Environnement
 - 2 French National Museum of Natural History, Paris, France and
 - 3 Centre for Mathematics and its Applications, Ecole Polytechnique, Palaiseau, France.
- Code Contributors: Ivan Sadikov and Akinwande Atanda
- The Transmission Process: A Combinatorial Stochastic Process for the Evolution of Transmission Trees over Networks, Raazesh Sainudiin and David Welch, Journal of Theoretical Biology, Volume 410, Pages 137–170, 2016 10.1016/j.jtbi.2016.07.038
- <http://lamastex.org/lmse/mep>
- Thanks for your Attention :)