Towards mathematical models of socio-psychological diversity: Insights from a meme evolution programme in the twitterverse

Raazesh Sainudiin†

Professeur Invité au CMAP, http://lamastex.org

joint work with: Rania Sahioun†, Akinwanda Atande‡, Kyle Nash‡ and Kumar Yogeeswaran‡

Laboratory for Mathematical Statistical Experiments, Christchurch Centre, and
Department of Psychology‡ & School of Mathematics and Statistics†,
University of Canterbury, Christchurch, New Zealand

November 23, 2016, Paris, France

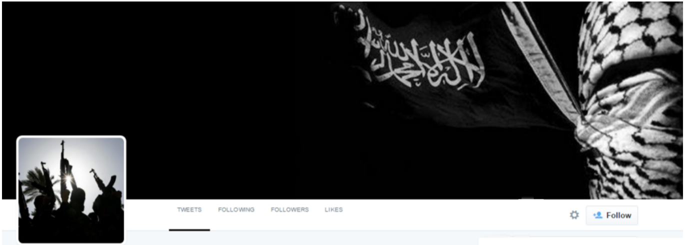
Institut de Paléontologie Humaine - 1 Rue René Panhard (13th arrondissement of Paris)

Section 1

Real-world Motivations

Extremist Networks

ISIS' Twitter Strategy `http://bit.ly/2gdls4K`

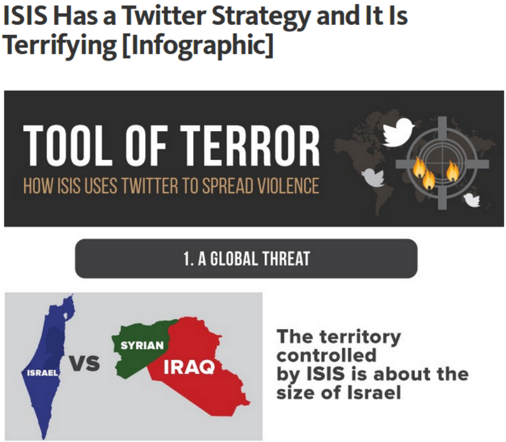


social media is being used by groups such as ISIS to:

- spread their message of hate,
- recruit susceptible youth, and
- project power all over the world.

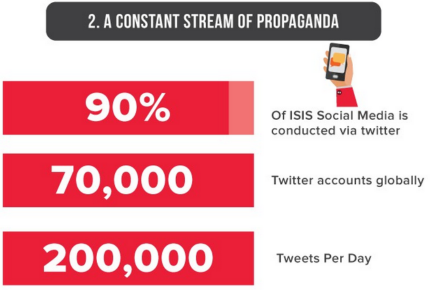
Extremist Networks

ISIS' Twitter Strategy `http://bit.ly/2gdls4K`



Extremist Networks

ISIS' Twitter Strategy http://bit.ly/2gdls4K



- There are an estimated 21,000 English-language followers alone.
- Most content comes from 2,000 over-performers that tweet in bursts of 50 or more tweets per day
- with each of these over-performers having an average of 1,004 followers.
- The result is an astonishing estimated 200,000 tweets per day.

https://www.brookings.edu/wp-content/uploads/2016/06/isis_twitter_census_berger_morgan.pdf

Extremist Networks

ISIS' Twitter Strategy `http://bit.ly/2gdls4K`

Extremist Networks

ISIS' Twitter Strategy `http://bit.ly/2gdls4K`

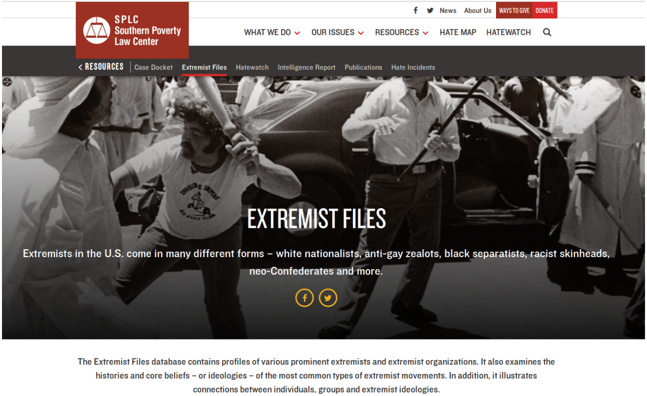
Conclusions of the study by `www.fifthtribe.com`:

- ISIS is essentially crowdsourcing its digital strategy.
- A similar massive operation needs to be developed in order to effectively blunt its outreach efforts.
- Members of the big data community, technologists, creatives, and digital strategists need to come together and coordinate with religious leaders, social media companies, and government agencies to develop an effective counter-messaging effort.

Extremist Networks

US Extremist Groups by SPLC

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

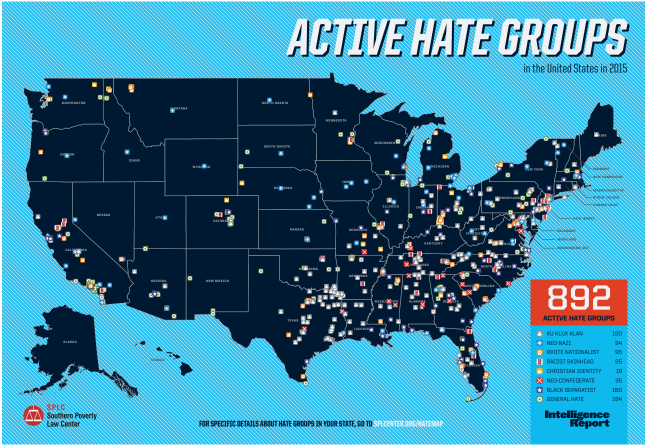


- https://www.splcenter.org/fighting-hate/extremist-files/ideology
- https://www.splcenter.org/fighting-hate/extremist-files/individual
- https://www.splcenter.org/fighting-hate/extremist-files/groups

Extremist Networks

US Extremist Groups by SPLC

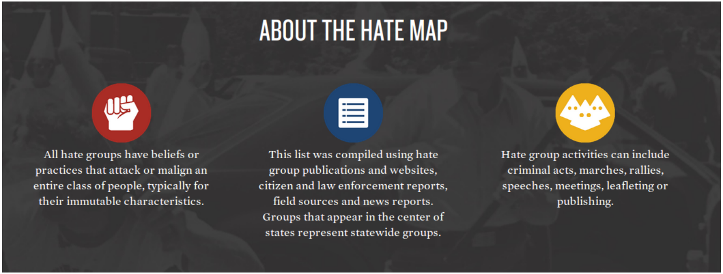
https://www.splcenter.org/fighting-hate/extremist-files



Extremist Networks

US Extremist Groups by SPLC

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

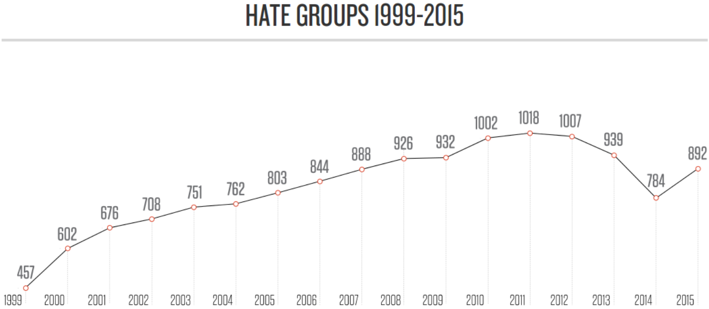


https://www.splcenter.org/hate-map

Extremist Networks

US Extremist Groups by SPLC

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



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

Extremist Networks

18 US Extremist 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

Extremist Networks

18 US Extremist 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

Extremist Networks

18 US Extremist Ideologies by SPLC

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

Christian Identity

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...

General Hate

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.

Holocaust Denial

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

Extremist Networks

18 US Extremist 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

Extremist Networks

18 US Extremist 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

Extremist Networks

18 US Extremist 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

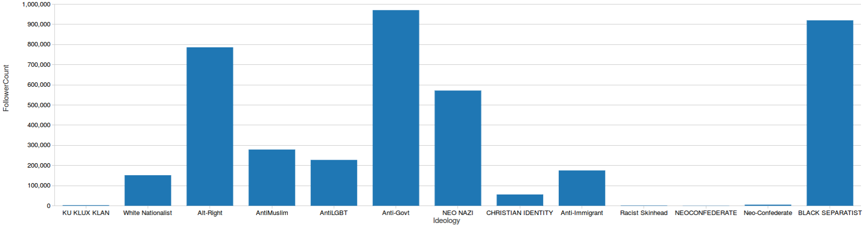
White nationalist groups espouse white supremacist or white separatist ideologies, often focusing on the alleged inferiority of nonwhites. Groups listed in a variety of other categories - Ku Klux Klan, neo-Confederate, neo-Nazi, racist skinhead, and Christian Identity - could also be fairly...

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

Extremist Networks

18 US Extremist Ideologies by SPLC

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

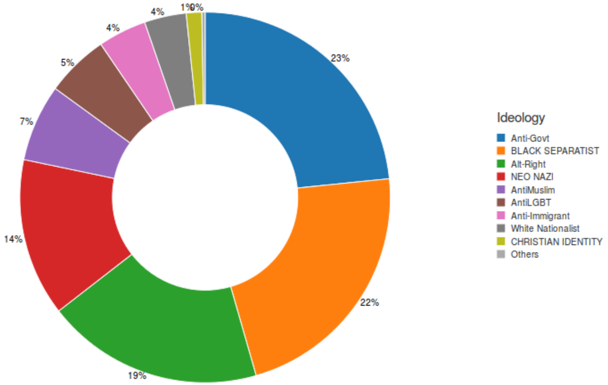


https://www.splcenter.org/fighting-hate/extremist-files/ideology

Extremist Networks

18 US Extremist Ideologies by SPLC

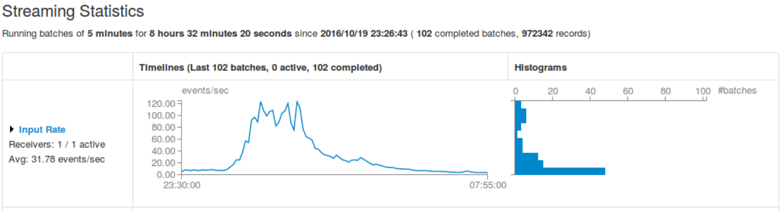
https://www.splcenter.org/fighting-hate/extremist-files



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

US Presidential Election 2016 - Twitter Streams

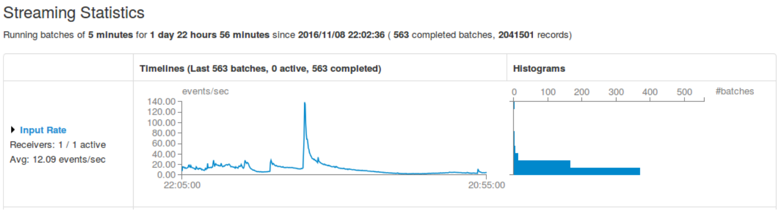
Twitter Data — 3rd US Presidential Debate



- User time-line of `@realDonaldTrump`, `@HillaryClinton` and splc-extremists with twitter accounts
- collected data includes all mentions, replies, retweets, etc of these twitter accounts of interest
- Goal: to gain insights into how people are communicating within and across party lines or ideologies
- Such twitter data was collected every day for about a month around the US Presidential Election

US Presidential Election 2016 - Twitter Streams

Twitter Data — Last 2 Days Around the End of Election



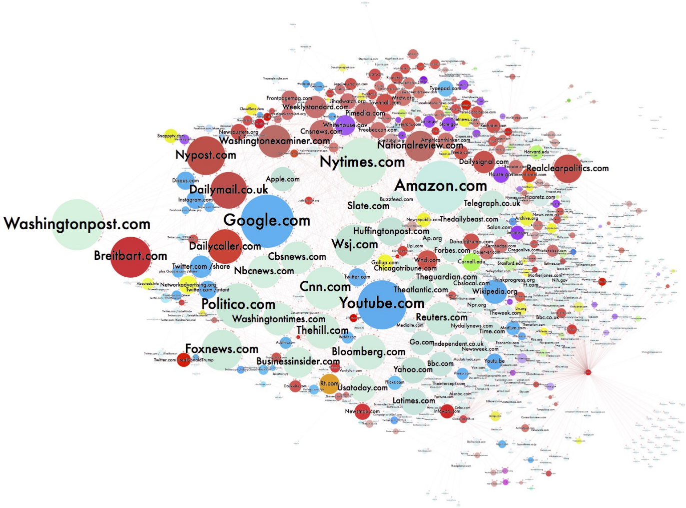
- User time-line of `@realDonaldTrump`, `@HillaryClinton` and splc-extremists with twitter accounts
- collected data includes all mentions, replies, retweets, etc of these twitter accounts of interest
- Goal: to gain insights into how people are communicating within and across party lines or ideologies
- Such twitter data was collected every day for about a month around the US Presidential Election

Section 2

Media in the Age of Algorithms

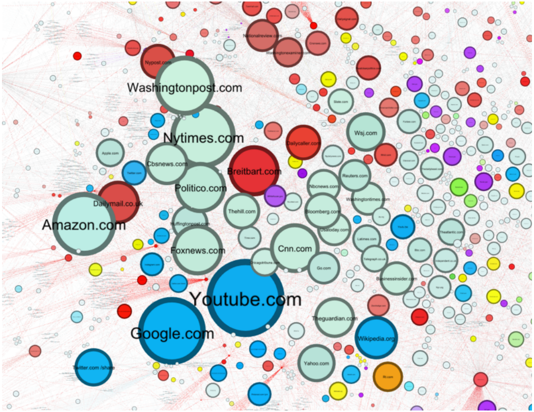
Micro-propaganda network of 117 fake news, viral, anti-science, hoax, and misinformation websites by Jonathan Albright

http://bit.ly/2gidlBW



Micro-propaganda network of 117 fake news, viral, anti-science, hoax, and misinformation websites by Jonathan Albright

`http://bit.ly/2gidlBW`

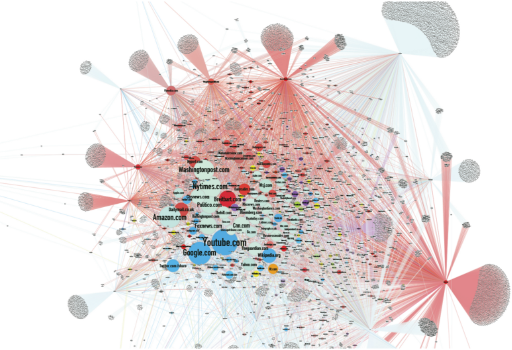


The Targets: Mainstream Media, Social Networks and Wikipedia:

- the sites with the most inbound hyperlinks (the largest circles on the graph) in this 'fake news' propaganda network are Google, YouTube, the NYTimes.com, Wikipedia, and Amazon.com.
- The larger the circle, the more links are coming in from the 117 #MCM network 'fake news' sites.

Micro-propaganda network of 117 fake news, viral, anti-science, hoax, and misinformation websites by Jonathan Albright

`http://bit.ly/2gidlBW`

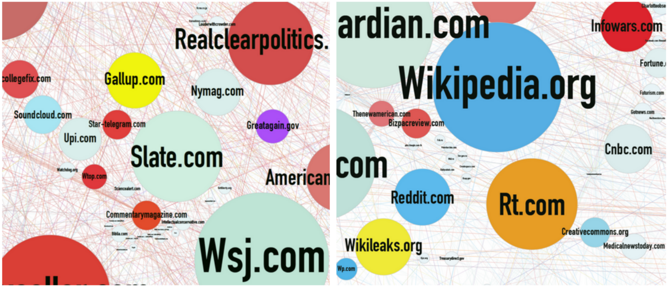


Mainstream Media Are Mostly "Surrounded":

- right-wing, fake news, conspiracy, anti-science, hoax, pseudoscience, and right-leaning misinformation sites **surround** most of the mainstream media
- sites in the fake news and hyper-biased #MCM network have a very small node size — this means they are linking out heavily to mainstream media, social networks, and informational resources
- **every incoming link is not a vote for the popularity of a site** as in Google's page-rank principle — the goal now is to maximize user engagement

Micro-propaganda network of 117 fake news, viral, anti-science, hoax, and misinformation websites by Jonathan Albright

http://bit.ly/2gidlBW



Network "zoom-in"—Gallup polls linked into heavily by MCM sites, as was Wikipedia, Reddit, and Creativecommons.org

Fact Checking and Knowledge Editing:

- #MCM network links heavily to a major poll site, Gallup, and crowdsourced fact-checking and reference resources — most notably Wikipedia, Reddit, and Wikimedia
- Snopes and other fake news verification sites are in the “liberal” side of the network at the top-middle right
- From fantastical falsehoods to outright vandalism, Wikipedians are warring over Trump’s inner circle. http://ow.ly/8e4y306hVUn

Social Media Realities

- Anyone can be a broadcaster and target/recruit anyone

Social Media Realities

- Anyone can be a broadcaster and target/recruit anyone
- Social Media Platforms emphasize **engagement** via **clicks**

Social Media Realities

- Anyone can be a broadcaster and target/recruit anyone
- Social Media Platforms emphasize **engagement** via **clicks**
- Unlike, Google's **page rank**, where every link is a vote, there is no equivalent of **truth rank** in social media

Social Media Realities

- Anyone can be a broadcaster and target/recruit anyone
- Social Media Platforms emphasize **engagement** via **clicks**
- Unlike, Google's **page rank**, where every link is a vote, there is no equivalent of **truth rank** in social media
- Question is: "whose truth? when 'knowledge' is editable"

Social Media Realities

- Anyone can be a broadcaster and target/recruit anyone
- Social Media Platforms emphasize **engagement** via **clicks**
- Unlike, Google's **page rank**, where every link is a vote, there is no equivalent of **truth rank** in social media
- Question is: "whose truth? when 'knowledge' is editable"
- facebook's curating the newsfeed as opposed to treating it as a timeline makes them medaite what users will see

Social Media Realities

- Anyone can be a broadcaster and target/recruit anyone
- Social Media Platforms emphasize **engagement** via **clicks**
- Unlike, Google's **page rank**, where every link is a vote, there is no equivalent of **truth rank** in social media
- Question is: "whose truth? when 'knowledge' is editable"
- facebook's curating the newsfeed as opposed to treating it as a timeline makes them medaite what users will see
- Tim O'Reilly's `http://bit.ly/2fpclfZ`:

  .... Unfortunately, unlike search, where the desires of the users to find an answer and get on with their lives are generally aligned with "give them the best results", Facebook's prioritization of "engagement" may be leading them in the wrong direction. What is best for Facebook's revenue may not be best for users.

Social Media Realities

- Anyone can be a broadcaster and target/recruit anyone
- Social Media Platforms emphasize **engagement** via **clicks**
- Unlike, Google's **page rank**, where every link is a vote, there is no equivalent of **truth rank** in social media
- Question is: "whose truth? when 'knowledge' is editable"
- facebook's curating the newsfeed as opposed to treating it as a timeline makes them medaite what users will see
- Tim O'Reilly's `http://bit.ly/2fpclfZ`:

  .... Unfortunately, unlike search, where the desires of the users to find an answer and get on with their lives are generally aligned with "give them the best results", Facebook's prioritization of "engagement" may be leading them in the wrong direction. What is best for Facebook's revenue may not be best for users.
- .... The bright side: searching through the possibility space for the intersection of truth AND engagement could lead Facebook to some remarkable discoveries.

Section 3

Model

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

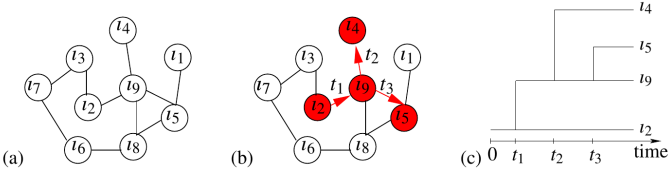


Figure 1: Spread of an epidemic over (a) the contact network of a population as shown by (b) a sub-network where edges representing transmission events are labelled by the time of event and the infected vertices are colored red and (c) the corresponding transmission tree.

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

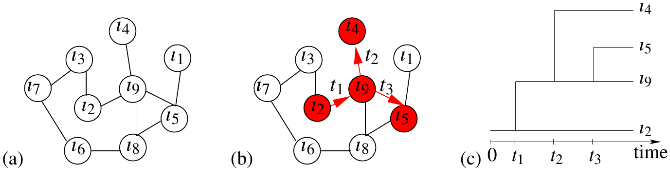


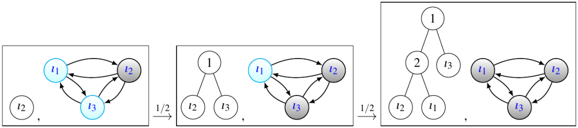
Figure 1: Spread of an epidemic over (a) the contact network of a population as shown by (b) a sub-network where edges representing transmission events are labelled by the time of event and the infected vertices are colored red and (c) the corresponding transmission tree.

Question: How does the geometry or structure of the SICN afftect the distribution (shape and timing) of the TT?

Answer: It is involved...

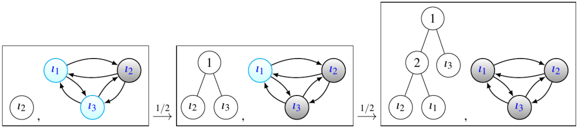
Markov chain on SI Contact Networks × Transmission Trees

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

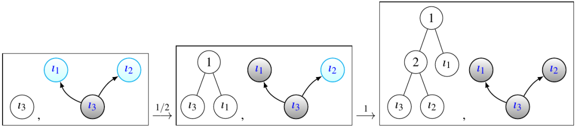


Markov chain on SI Contact Networks × Transmission Trees

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

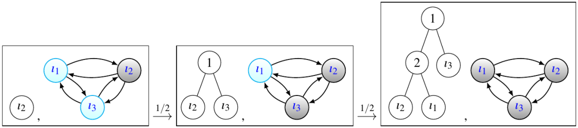


- A growing transmission tree on a **star** SICN in a population of size $n = 3$

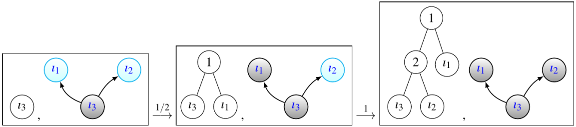


Markov chain on SI Contact Networks $\times$ Transmission Trees

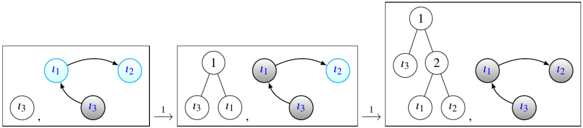
- A growing transmission tree on a **complete** SICN in a population of size $n = 3$



- 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 $\mathbb{I}_n = \{\imath_1, \imath_2, \ldots, \imath_n\}$ be the label set of a pop. of size n

State Space

- Let $\mathbb{I}_n = \{\imath_1, \imath_2, \ldots, \imath_n\}$ 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 $\mathbb{I}_n = \{\imath_1, \imath_2, \ldots, \imath_n\}$ 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 $\mathbb{I}_n = \{\imath_1, \imath_2, \ldots, \imath_n\}$ 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 $\mathbb{I}_n = \{\imath_1, \imath_2, \ldots, \imath_n\}$ 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 $\tau \in \{$rooted planar ranked leaf-labelled binary trees$\}$

State Space

- Let $\mathbb{I}_n = \{\imath_1, \imath_2, \dots, \imath_n\}$ 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 $\tau \in$ {rooted planar ranked leaf-labelled binary trees}
  - Note the poset on 2^{w_n} with unit weights given by $\prec := \subseteq$

State Space

- Let $\mathbb{I}_n = \{\imath_1, \imath_2, \ldots, \imath_n\}$ 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 $\tau \in$ {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

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

$$\frac{\text{the edge-weight from } (z+1)\text{-th infector to the } (z+1)\text{-th infectee}}{\text{Sum of edge-weights from every potential infector to every potential infectee within its susceptible out-neighborhood}}$$

Transition Probabilities

- One-step transitions for the jump chain

$$\Pr\{(\tau(z+1), c(z+1)) \mid (\tau(z), c(z))\} = \frac{\text{the edge-weight from } (z+1)\text{-th infector to the } (z+1)\text{-th infectee}}{\text{Sum of edge-weights from every potential infector to every potential infectee within its susceptible out-neighborhood}}$$

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

Transition Probabilities

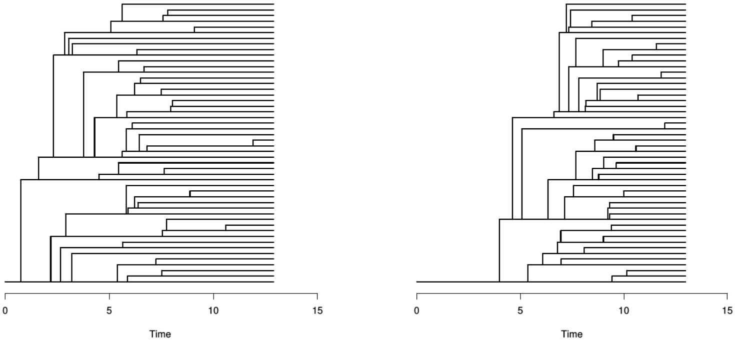
- One-step transitions for the jump chain

$$\Pr\{(\tau(z+1), c(z+1)) \mid (\tau(z), c(z))\} = \frac{\text{the edge-weight from } (z+1)\text{-th infector to the } (z+1)\text{-th infectee}}{\text{Sum of edge-weights from every potential infector to every potential infectee within its susceptible out-neighborhood}}$$

- By letting the time for each infection event to be distributed as $\overset{iid}{\sim} \mathrm{Exponential}(\lambda)$ 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



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

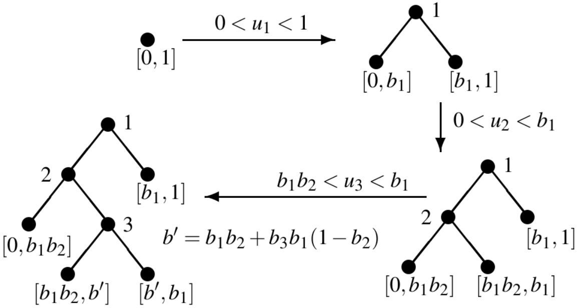
- IDEA: induce distributions on TTs without the SICN
- Consider Generating Sequences:
  - $U_1, U_2, \ldots \overset{iid}{\sim} \mathrm{Uniform}(0, 1)$

Beta-splitting Model

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

Beta-splitting Model

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

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

$$= \prod_{z=1}^{m} \frac{B(s_z^L+\alpha+1, s_z^R+\beta+1)}{B(\alpha+1,\beta+1)} \times \Pr(\textit{leaf labels}), \tag{3.4}$$

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

Theorems

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

- $(\alpha, \beta) = (0, 0) \equiv$ complete SICN,
- $(\alpha, \textit{beta}) \to (\infty, -1) \equiv$ star SICN
- $(\alpha, \textit{beta}) \to (-1, \infty) \equiv$ 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 X later...
- 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	$\overline{\hat{\alpha}}$ (s.e.)	$\overline{\hat{\beta}}$ (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$^{*,\circ}(50,2,0.0)$	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$^{\circ}(50,2,0.1)$	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	$\text{SWRN}^{*}(50,5,0.1)$	50	30	5	-0.7530 (0.01572)	-0.4751 (0.04671)
37	$\text{SWRN}^{\circ}(50,5,0.1)$	50	30	5	-0.7918 (0.01596)	-0.5130 (0.03323)
38	$\text{SWRN}^{\circ}(50,5,0.2)$	50	30	5	-0.6881 (0.03277)	-0.3595 (0.06002)
39	$\text{SWRN}^{\circ}(50,5,0.5)$	50	30	5	-0.5264 (0.04687)	-0.2138 (0.09471)
40	$\text{SWRN}^{\circ}(100,2,0.2)$	100	1	5	-0.9479 (0.01509)	-0.3991 (0.5065)
41	$\text{SWRN}^{\circ}(100,2,0.2)$	100	30	5	-0.9493 (0.003869)	-0.6027 (0.03475)
42	$\text{SWRN}^{*}(100,2,0.5)$	100	1	5	-0.9023 (0.03411)	-0.7139 (0.03929)
43	$\text{SWRN}^{*}(100,2,0.5)$	100	30	5	-0.8878 (0.006687)	-0.6821 (0.02189)
44	$\text{SWRN}^{\circ}(100,2,0.5)$	100	1	5	-0.8714 (0.05584)	-0.6533 (0.09257)
45	$\text{SWRN}^{\circ}(100,2,0.5)$	100	30	5	-0.8920 (0.005128)	-0.6786 (0.02189)
46	$\text{SWRN}^{\circ}(100,5,0.99)$	100	30	5	-0.5079 (0.02371)	-0.2290 (0.03059)
47	$\text{SWRN}^{\circ}(100,10,0.99)$	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.

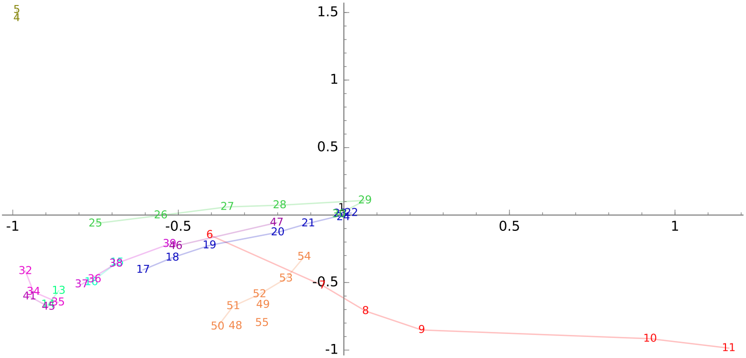
Beta-projections of various models

48	$\text{PrefAttach}^*(100,1)$	100	30	10	-0.3275 (0.04932)	-0.8215 (0.01121)
49	$\text{PrefAttach}^*(100,2)$	100	30	10	-0.2443 (0.03283)	-0.6647 (0.01294)
50	$\text{PrefAttach}^\circ(100,1)$	100	30	10	-0.3813 (0.04908)	-0.8254 (0.005460)
51	$\text{PrefAttach}^\circ(100,2)$	100	30	10	-0.3339 (0.03884)	-0.6743 (0.01657)
52	$\text{PrefAttach}^\circ(100,3)$	100	30	10	-0.2545 (0.04181)	-0.5863 (0.01652)
53	$\text{PrefAttach}^\circ(100,5)$	100	30	10	-0.1748 (0.04214)	-0.4698 (0.03110)
54	$\text{PrefAttach}^\circ(100,10)$	100	30	10	-0.1196 (0.03449)	-0.3089 (0.02663)
55	$\text{PrefAttach}^\circ(100,1)$	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.

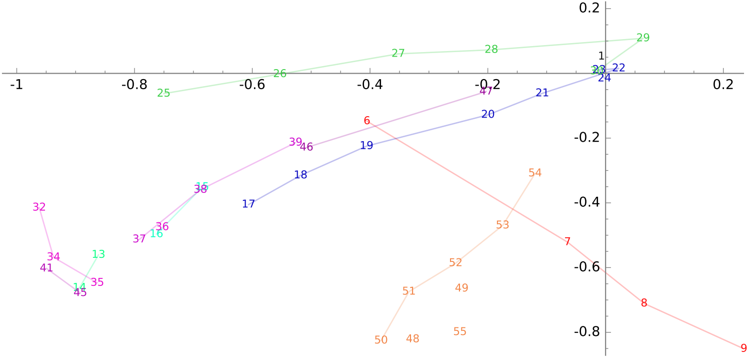
Beta-projections of various models



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

References: Paper and Codes from Live Notebooks

- 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`

References: Paper and Codes from Live Notebooks

- 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`

Section 4

Some experiments in the twitterverse

Twitter Interactions

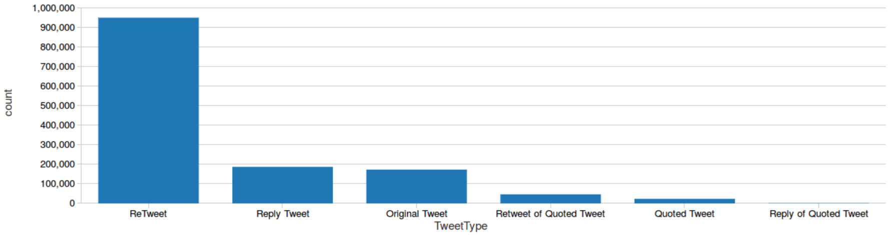
	Sequence of Action by Users				
#	1	2	3	4	Outcome
1	New Tweet				Original Tweet
2	New Tweet	Retweet			Retweet
3	New Tweet	Comment			Quoted Tweet
4	New Tweet	Comment	Retweet		Retweet of Quoted Tweet
5	New Tweet	Comment	Reply		Reply Tweet
6	New Tweet	Comment	Reply + link		Reply of Quoted Tweet
7	New Tweet	Comment	Reply	Retweet	Retweet
8	New Tweet	Comment	Reply	Comment	Quoted Tweet
9	New Tweet	Reply			Reply Tweet
10	New Tweet	Reply + link			Reply of Quoted Tweet
11	New Tweet	Reply	Retweet		Retweet
12	New Tweet	Reply	Comment		Quoted Tweet
13	New Tweet	Reply	Retweet	Comment	Quoted Tweet
14	New Tweet	Reply	Comment	Retweet	Retweet of Quoted Tweet

Twitter Interaction Capture

#	CurrentTweet	TweetType
1	Akin will be testing from here @Delvix777-hahaha Check if it's in MEP tweet collection job StreamingAllThreeGroups databricks academic shard	Original Tweet
2	Commenting that the testing @Delvix777-hahaha is working out well... https://t.co/adE2cDxphc	Quoted Tweet
3	Commenting or quoting the reply tweet of @Delvix777-hahaha https://t.co/Vy2wdZIrcM	Quoted Tweet
4	@Delvix777 https://t.co/uGi2PYbewU This is a reply to the quoted web-link tweet	Reply of Quoted Tweet
5	@Delvix777 @ashutopia https://t.co/hfszfJ6gRO	Reply of Quoted Tweet
6	@Delvix777 Replying this Commented tweet by @Delvix777-hahaha	Reply Tweet
7	Replying to @Delvix777 for the comments made about the tweets	Reply Tweet
8	@ShalomRanga @Delvix777 I, @ashutopia is sending a reply to @ShalomRanga under her quoted tweet of Akin's authored status	Reply Tweet
9	@Delvix777 @ashutopia	Reply Tweet
10	RT @anusha_raazesh: Akin will be testing from here @Delvix777-hahaha Check if it's in MEP tweet collection job StreamingAllThreeGroups dataâ€¦	ReTweet
11	RT @ShalomRanga: @Delvix777 Replying this Commented tweet by @Delvix777-hahaha	ReTweet
12	RT @Delvix777: Commenting that the testing @Delvix777-hahaha is working out well... https://t.co/adE2cDxphc	Retweet of Quoted Tweet
13	RT @ShalomRanga: Commenting or quoting the reply tweet of @Delvix777-hahaha https://t.co/Vy2wdZIrcM	Retweet of Quoted Tweet

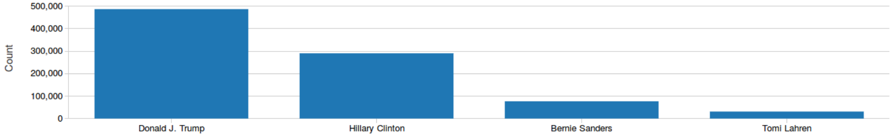
With a few tens of lines of code we can determine the 'TweetType' of each status we can capture from twitter.

Frequency of Twitter Interaction Types



On the day of the 3rd US Presidential Debate (as in most days) ‘ReTweets‘ were the most frequent type of interaction.

ReTweeted Frequency



Retweets are a direct and simplistic affirmation of the tweet and thus help understand the structure without getting into NLP tasks.

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

Statistics of Trump-Clinton-SPLC_Extremists Data

Stats on the 20161019 data

- Across all groups being stream-collected, unfiltered for bots, etc
  - Total tweets in groupTTTDF = 353,598
  - Total retweets in groupTTTDF = 226,489
  - number of Trump-Clinton retweets (those authored by Trump/Clinton that have been retweeted)
    - 200,304
    - 196,096 filtered by: minimumAgeSinceAccountCreatedInDays=0 AND accounts with >2 friends and >2 followers
    - 177,300 filtered by: minimumAgeSinceAccountCreatedInDays=100 AND accounts with >2 friends and >2 followers
    - 61,211 filtered by: minimumAgeSinceAccountCreatedInDays=100 AND accounts with >2 friends and >2 followers AND GeoEnabled
  - Number of distinct retweet-pairs of Trump and Hillary (# unique edges Trump/Clinton -> retweeters)
    - 77,861 filtered by: minimumAgeSinceAccountCreatedInDays=0 AND accounts with >2 friends and >2 followers
    - 70,930 filtered by: minimumAgeSinceAccountCreatedInDays=100 AND accounts with >2 friends and >2 followers
    - 27,696 filtered by: minimumAgeSinceAccountCreatedInDays=100 AND accounts with >2 friends and >2 followers AND GeoEnabled
    - 195 filtered by: being a verified account

Our current sample size = all 195 verified retweets + 2.64% of filtered-GeoEnabled retweets = 3% of verified and filtered-GeoEnabled retweets = 1% of all retweets = 1811

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

Trump-Clinton Retweet Network — a few samples

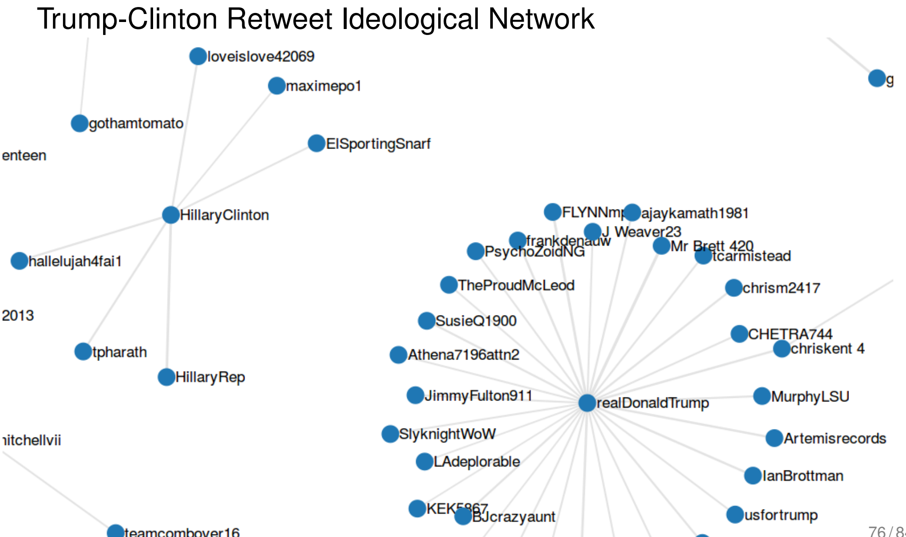
CPostUserSN	OPostUserSNInRT	OPostUserSNInQT	favouritesCount	followersCount	friendsCount	isVerified	isGeoEnabled	CurrentTweet
georgefayner	realDonaldTrump	null	137811	1466	953	false	true	RT @realDonaldTrump: China is cooking up conspiracy theories that the Olympics are rigged. http://t.co /0ah0hBJt They don't 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 /6n63fHVvNo
thuerta	realDonaldTrump	null	13081	128	345	false	true	RT @realDonaldTrump: 'Trump rally disrupter was once on Clinton's payroll' https://t.co /75oLLuD4SI
tanladyvolfan	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 $\#\mathbb{V} = 1205$, $\#\mathbb{E} = 29856$)

Trump-Clinton Retweet Network weighted by Retweet counts

userCreatedAtDate	daysSinceUserCreated	OPostUserSNinRT	CPostUserSN	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	HillaryClinton	HillaryRep	4196	2168	4984	158
2011-03-22T13:09:23.000+0000	2047	realDonaldTrump	FLYNNmpc	1653	48	75	146
2014-08-25T17:02:46.000+0000	795	realDonaldTrump	mikenyr2499	17427	183	155	132
2009-04-26T07:07:03.000+0000	2742	yottapoint	gcommking	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
2009-03-08T12:59:18.000+0000	2791	realDonaldTrump	Artemisrecords	2000	2777	5000	112
2012-09-25T15:09:37.000+0000	1494	realDonaldTrump	IanBrottman	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	HillaryClinton	loveislove42069	3818	108	398	98
2015-09-01T18:52:08.000+0000	423	realDonaldTrump	BJcrazyaunt	1064	1296	1433	95
2011-12-24T03:52:02.000+0000	1770	HillaryClinton	tpharath	703	38	183	91
2015-03-08T23:47:05.000+0000	600	HillaryClinton	hallelujah4fai1	16766	227	270	88
2014-06-30T16:44:10.000+0000	851	realDonaldTrump	ajaykamath1981	3309	2667	3010	88
2012-04-29T21:49:38.000+0000	1643	realDonaldTrump	MurphyLSU	65	28	47	84
2010-08-05T16:02:11.000+0000	2276	realDonaldTrump	sdpubs	23674	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 $\#\mathbb{V} = 1205$, $\#\mathbb{E} = 29856$)



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

Trump-Clinton TIN or (Re)Tweet Ideological Network – Outdegree

```
> display(g.outDegrees.orderBy($"outDegree".desc))
```

Id	outDegree
realDonaldTrump	1884
HillaryClinton	1098
wikileaks	479
FoxNews	367
WDFx2EU7	355
TeamTrump	336
DanScavino	320
mitchellvii	295
KellyannePolls	270
JaredWyand	254
mike_pence	246
PrisonPlanet	243
JamesOKeefeIII	220
IngrahamAngle	186
DonaldJTrumpJr	177

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

Trump-Clinton TIN or (Re)Tweet Ideological Network

(3% sample $\#\mathbb{V} = 1205$, $\#\mathbb{E} = 29856$) – Indegree

Id	InDegree
theReal_Rebel	46
deuk6767	46
magnifier661	44
neo_zhang16	43
JustBira	43
IndependentHK	41
Anton_Ultron	41
LaurieAnnBaker	40

So, the loudest vessel wins the “being heard match” in the twitterverse! — Whether the vessel is “empty” is less irrelevant than how many can hear it with their specific “emotional needs/fears/etc. soothed”...

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

Trump-Clinton TIN or (Re)Tweet Ideological Network – Community Detection

Step 3. Community Detection using Label Propagation over TIN of Retweets

```
> val result = g.labelPropagation.maxIter(30).run() // iterations more than 30 may crash the cluster... takes 5 minutes

result: org.apache.spark.sql.DataFrame = [id: string, label: bigint]
```

```
> //result.unpersist() // unpersist bertween runs?

res26: result.type = [id: string, label: bigint]
```

```
> // Create DF with proper column names - from 30 iterations
  val top20communityDF = result.select($"id",$"label").withColumn("weight", lit(1L)).groupBy($"label").agg(sum("weight").alias("communitySize")).orderBy($"communitySize".desc).limit(20)
  top20communityDF.show()

+-----+-------------+
|label|communitySize|
+-----+-------------+
|    6|         4222|
|  184|         4000|
|  892|          868|
|10917|          338|
| 5136|           57|
| 5437|           42|
```

Clear presence of “echo chambers” in the twitterverse! ...(these are well-documented phenomena)

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

Trump-Clinton TIN or (Re)Tweet Ideological Network – Community Detection

```
> result.filter($"Id"==="realDonaldTrump" || $"Id"==="HillaryClinton").show

+---------------+-----+
|             id|label|
+---------------+-----+
| HillaryClinton|  184|
|realDonaldTrump|    6|
+---------------+-----+
```

#(“echo chamber of Clinton”)=3996 & #(“echo chamber of Trump”)=4234

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

Trump-Clinton TIN or (Re)Tweet Ideological Network – Community Detection ∩ SPLC-extremists

```
val a = SPLC_IdeologyAndSN.join(result,SPLC_IdeologyAndSN("id")===result("id"))
display(a)
```

Ideology	id	id	label
White Nationalist	vdare	vdare	6
AntiMuslim	securefreedom	securefreedom	6
Anti-Immigrant	FAIRImmigration	FAIRImmigration	6

This is just 3% of the data for just a few hours surrounding the3rd US Presidential Debate

Section 5

Open Discussions

Let's do it...

- Can there be a "truth rank" in some acceptable sense?
- Is it too much to expect fellow citizens to be reasonably rational agents?
- Any solution must take the market forces into account?
- Law may need to interfere with market — UN human right violations (when does free-speech become incitement to genocide)?
- Is Putin reasonable in having blocked linked-In in Russia 4 days ago?
- ? ? ? (memetic notions of self in "trans-traditional life" experiments... — NOT a mathematical problem but a philosophical - political - economic - social - psychological - legal problem)
- Our own mathematical research $\rightarrow$ Markov control processes on TIN/SICN where we want to allow the Networks themselves to coevolve with transmission trees

Thank you!