Hoax vs Fact Checking

Understanding and predicting the diffusion of low quality information on communication networks

Giancarlo Ruffo - Università degli Studi di Torino (Italy)

http://www.di.unito.it/~ruffo
giancarlo.ruffo@unito.it
@giaruffo
Fictional background
Lilliput and Blefuscu

According to "Gulliver's Travels", they are two islands in the South Indian Ocean.

Two different kingdoms inhabited by tiny people.

Even if similar in nature and in religious belief, they have a long lasting debate called the "egg war".
Big-Endians/Little-Endians

Holy Scriptures: “Always break the egg on the most convenient side”, that is the larger in Lilliput

The way Lilliputians always broke their eggs

The way the emperor ordered them to break their eggs.

“Little endian” interpretation of holy scriptures was adopted in Blefuscu
Satirical interpretation

❖ **Eggs wars**: Catholic England (Big-Endian) and conversion to Protestantism of most of the country (Little-Endian) after Queen Elisabeth I conversion

❖ **Lilliput and Blefuscu**: Kingdom of Great Britain and Kingdom of France

❖ **Internal politics in Lilliput**: the Whigs and the Tories

❖ In perspective: human beings divide themselves because of what may appear a futile reason to an alien

❖ It contains the intuition of the interplay between (structural) segregation and (opinion) polarization
Agenda of the talk

❖ The strange case of Lajello
❖ Modeling the spread of misinformation
❖ The role of segregation
❖ Evaluating debunking strategies
❖ Language and network structure
❖ Discussion and Conclusion
The strange case of Lajello
Analyzing social network with a bot

- Anobii was a social networks for book lovers
- Scraping users’ profiles from the Web was admitted
- Users’ libraries and their links were collected periodically
Analyzing social network with a bot

- Anobii was a social networks for book lovers
- Scraping users’ profiles from the Web was admitted
- Users’ libraries and their links were collected periodically
- The bot “Lajello” used to silently navigate Anobii twice a month for one year
Analysis of Anobii’s structure

strong signals of geographical, cultural and topical homophily by selection

... and other interesting stuff on influence:

LM Aiello, A Barrat, C Cattuto, G Ruffo, R Schifanella, Link creation and profile alignment in the aNobii social network, 2010 IEEE 2nd Int.. Conf. on Social Computing, 249-256

LM Aiello, A Barrat, C Cattuto, G Ruffo, R Schifanella, Link creation and information spreading over social and communication ties in interest based online social network, EPJ Data Science 1 (1), 12
Application: a link recommendation algorithm

- A link recommendation algorithm based on prediction of profile similarities was proposed and tested
- Results showed an improvement w.r.t. the baselines
What happened to Lajello?

Lajello, incidentally, became the second most popular user in Anobii in terms of messages from distinct users.
Lajello started to introduce users to each other according to our link recommendation algorithm.

First result: users acceptance of the recommendation skyrocketed if they previously wrote in Lajello’s wall.
Influence of bots

**COMMUNICATIONS OF THE ACM**

**Review Articles**

The Rise of Social Bots

By Emilie Ferrara, Onur Varol, Clayton Davis, Filippo Menczer, Alessandro Flammini

Communications of the ACM, Vol. 59 No. 7, Pages 96-104

10.1145/2818717

Comments (1)

**BIG DATA NEWS**

SPECIAL ISSUE Computational Propaganda and Political Big Data

**Fake News**

Russian Social Media Bots Influence on the 2016 Election

By Chengcheng Shao, Giovanni Luca Ciampaglia, Onur Varol, Kai-Cheng Yang, Alessandro Flammini & Filippo Menczer

Nature Communications 9, Article number: 4787 (2018)
Incidently, we created an “egg war”

• After our initial experiment, Lajello remained silent for one year and then he “talked”. The recommendations changed the net structure and lajello account was banned after 24 hours. This ignited a “war”

• Two polarized opinions emerged: Anobii users created immediately two thematic groups: “the (not requested) suggestions of Lajello” and “Hands-off Lajello”

• A large portion of users that were contacted by Lajello joined to one of these groups

• We observed a strong interplay between the existing relationships in the social network and the opinion that emerged from the users at the end of the links: “echo chamber” effect?
Social polarization and emotional reaction

red dots are lajello supporters
blu dots are lajello haters
links are existing social connections or direct messages (graph is directed)
bigger dots are users with more links

Automatic network-based community detection algorithm (OSLOM) accurately finds clusters (80% - Social network, 72% - Communication network), confirming a signal of segregation between the two groups before link recommendations.
Laiello... Hai stufato... Non se ne puo' piu'... Sta attendendo che sono capace di assoldare un hacker per vedere chi sei. E po' sono c... tuoi.

Ahahahahahahaha tu sei un genio!!!!! Sei davvero un genio!!! Insomma ma quanti visualizzazioni hai???? Sei un grande!!!!!! Riesco a farti visitare e a farti scrivere pur non avendo libri!!! Ti adoro sei grandissimo :P

Un grande. Continua così. Grazie delle visite, si vede che ti sto simpatica.... P.S: propongo di aprire un gruppo the laiellos fans...

Chapeau!!

Grazie Laiello, mi sono divertita un sacco a leggere i commenti degli altri anobiani. Sembra un esperimento di psicologia sociale, se non ti dispiace ti aggiungo come vicino e resisti eh... non pubblicare un libro! ;)

Le tue visite cominciano ad essere inquietanti....
Lessons learned and observations

❖ Handle experiments in social media with care :)
❖ A simple spambot can take power in a social network
❖ A seed of polarization found in pre-existing network structure (Lilliput and Blefuscu were two different islands…)
❖ Network and Sentiment analysis provide tools and measures, when we have data
❖ What if the real identity and motivations of Lajello were fact-checked?

MIT Technology Review

Connectivity

How a Simple Spambot Became the Second Most Powerful Member of an Italian Social Network

The surprising story of how an experiment to automate the creation of popularity and influence became successful beyond all expectation.

by Emerging Technology from the arXiv Aug 5, 2014
Modeling the spread of misinformation
Questions

❖ Is fact-checking effective against the diffusion of fake-news?

❖ Do “echo-chambers” and “islands” play a role as inhibitors or facilitators of fake-news spreading?
Networks and their context

- Nodes are **actors** involved in a **generic** social network (no assumption is given)
- Links are **social relationships**
- Nodes can be exposed to news from both **internal and external sources** and via different communication devices
- **Network topologies** can be created artificially or built from real data
- The **news is factually false** (can be debunked or someone else has already debunked it)
- We need a **model** for predictions and what-if analysis; data for validation and tuning only
Node states in the SBFC model

❖ Susceptible
❖ Believer
❖ Fact-Checker

neighbors of $i$: $n_i$
credibility of the hoax: $\alpha$
spreading rate: $\beta$
From Susceptible to Believer/Fact-Checker

\[ f_i(t) = \beta \frac{n_i^B(t)(1 + \alpha)}{n_i^B(t)(1 + \alpha) + n_i^F(t)(1 - \alpha)} \]

\[ g_i(t) = \beta \frac{n_i^F(t)(1 - \alpha)}{n_i^B(t)(1 + \alpha) + n_i^F(t)(1 - \alpha)} \]
From Susceptible to Believer/Fact-Checker

\[ f_i(t) = \beta \frac{n_i^B(t)(1 + \alpha)}{n_i^B(t)(1 + \alpha) + n_i^F(t)(1 - \alpha)} \]

\[ g_i(t) = \beta \frac{n_i^F(t)(1 - \alpha)}{n_i^B(t)(1 + \alpha) + n_i^F(t)(1 - \alpha)} \]
From Believer to Fact-Checker

\[ p_{\text{verify}} \]

VERIFYING probability of fact-checking (or just deciding not to believe)
From Believer/Fact-Checker to Susceptible

\[ S \]  
\[ \mathcal{p}_{\text{forget}} \]  
\[ B \]  
\[ \mathcal{p}_{\text{forget}} \]  
\[ FC \]  
FORGETTING
Dynamics (agent-based simulations)

\[ \alpha = 0.3 \quad p_{\text{verify}} = 0.2 \]

\[ \alpha = 0.8 \quad p_{\text{verify}} = 0.05 \]

hoax credibility and fact-checking probability rule hoax persistence in the network
Dynamics (agent-based simulations)

threshold on verifying probability: this provides an idea of how many believers we need to convince to guarantee the removal of the hoax.

The role of segregation
Skeptical and gullible agents

let’s tune credibility accordingly

the propensity to believe is also a property of the node (gullibility)

What does it happen when a skeptics and gullible agent are segregated?
Modeling two segregated communities

Skeptic

$\alpha_{small}$

Gullible

$\alpha_{large}$

**size** $(0 < \gamma < N)$

#nodes in the gullible community

**segregation** $(0.5 < s < 1)$

fraction of edges within same community [Gu-Gu, Sk-Sk]

$s=0.55$

$\gamma=500$

$s=0.8$

$\gamma=500$

$s=0.95$

$\gamma=500$
Size vs segregation

$p_f = 0.1$

LOW Forgetting Probability

Number of Believers

Segregation

gullible group size
Size vs segregation

$p_f = 0.1$
LOW Forgetting Probability

$p_f = 0.8$
HIGH Forgetting Probability

gullible group size

segregation

Number of Believers

0
100
200
300
Transitions
Role of forgetting

LOW Forgetting Rate

$\rho_f = 0.1$

HIGH Forgetting Rate

$\rho_f = 0.8$
Lessons learned and observations

❖ We can use our model to study the fake-news diffusion process in segregated community

❖ Complex contagion is observed: interplay and not trivial outcomes

❖ Forgetting probability becomes relevant as well as the level of segregation:

❖ high forgetting probability (e.g., just `normal’ unfounded gossip) vanishes soon in segregated communities

❖ low forgetting probability (e.g., conspiracy theories or partisanship beliefs) requires low segregation

real data: vaccines

\#askscotflu, \#GetVax, \#hcsmvac, \#McrFluSafe13, \#McrFluSafe14, \#MeaslesTruth, \#RUuptodate, \#Vaccinate, \#vaccination, \#vaccines, \#VaccinesWork

segregation: 0.97

twitter data from IU https://osome.iuni.iu.edu
real data: chemtrails

#chemtrails, #opchemtrails, #iwantmyblueskyback, #globaldimming, #geoengineering, #chemsky, #chemclouds, #whatintheworldaretheyspraying, #chemtrail, #weathertamification, #weathercontrol

#instantweatherpro #sky #cielo #clouds #reverse #nubes

twitter data from IU https://osome.iuni.iu.edu

segregation: 0.99
Evaluating debunking strategies
What-if analysis

- We live in a **segregated** society: let’s accept it!
- “Egg wars” can last for a long time: **low forgetting** probability
- **Computational epidemiology**: immunization works better if some node in the network (e.g., hubs, bridges) is vaccinated first
- **Where** to place fact-checkers?
- Stronger hypothesis: a believer do not verify ($p_{\text{verify}} = 0$)
  - they can still forget
  - we can accept to leave half of the population breaking the egg on the wrong side, but we want at least to protect the skeptics!
Basic settings with no verification

Setting

segregation: 0.92 (high)
forgetting: 0.1 (low)
gullible group:
  • $\alpha$: 0.8
  • seeders B: 5%
skeptical group:
  • $\alpha$: 0.3
  • seeders FC: 5%

As expected: very bad!
Hubs as fact-checkers

**Setting**

segregation: 0.92 (high)

forgetting: 0.1 (low)

gullible group:
  - $\alpha$: 0.8
  - seeders B: 5%

skeptical group:
  - $\alpha$: 0.3
  - seeders FC: 5%

- seeders are HUBS!

**Simulation start**

**Simulation results**

better, but still…
MORE hubs as fact-checkers

**Setting**

- Segregation: 0.92 (high)
- Forgetting: 0.1 (low)

**Gullible group:**
- $\alpha$: 0.8
- Seeders B: 5%

**Skeptical group:**
- $\alpha$: 0.3
- Seeders FC: 10%
- Seeders are HUBS!

**Simulation start**

**Simulation results**

better, but still…
MORE hubs as fact-checkers

**Setting**

- Segregation: 0.92 (high)
- Forgetting: 0.1 (low)

**Gullible group:**
- $\alpha$: 0.8
- Seeders B: 5%

**Skeptical group:**
- $\alpha$: 0.3
- Seeders FC: 20%
- Seeders are HUBS!

**Simulation start**

**Simulation results**

finally, more FC than B!
MORE hubs as fact-checkers

Setting

segregation: 0.92 (high)
forgetting: 0.1 (low)
gullible group:
  • $\alpha$: 0.8
  • seeders B: 5%
skeptical group:
  • $\alpha$: 0.3
  • seeders FC: 30%
  • seeders are HUBS!

Simulation start

Simulation results

slightly better, but unrealistic
Bridges as Fact-Checker

Setting

segregation: 0.92 (high)
forgetting: 0.1 (low)
gullible group:
  • $\alpha$: 0.8
  • seeders B: 5%
skeptical group:
  • $\alpha$: 0.3
  • seeders FC: 5%
  • BRIDGES!

pretty good, and realistic
Beware of results based on realizations!

- Simulations results are based on many different stochastic realizations of the model.
- Plots show (statistically significant) averages.
- That means that some realizations may diverge.
- Realizations as are unlikely, but still possible!
Lessons learned and observations

❖ Debunking activism is often considered useless or counterproductive

❖ However, a world without fact-checking is harmless against fake-news circulation: skeptics exposed to misinformation will turn into believers because of social influence

❖ Skeptics with links to gullible subjects should be the first to be exposed to the fact-checking: misinformation will survive in the network, but their communities can be ‘protected’ by such gatekeepers

❖ Note: no socio-psychological assumption so far. Real world is much more complicated

Language and network structure
Links to NLP

- Individual’s opinions are often hidden
- Social Media provide much data for stance detection, emotion analysis, and so on
- Communication styles can be another trigger or just a reaction to news exposition and partisanship
- Relationships between structural segregation and opinion formation and polarization should be explored further by a joint effort between our scientific communities
Discussion and conclusion
Recap

❖ **Structural segregation** (as in Lilliput and Blefuscu islands) may be one of the main triggers of opinion **polarization**

❖ **Fake-news spreading**, especially when partisanship and antagonistic behavior reinforce the debate, is **facilitated** in segregated networks

❖ Fact-checking is needed and skeptics with links to more gullible (vulnerable) contacts can be recruited as **gatekeepers**

❖ **Network Analysis** and **NLP** are great tools for modeling and analyzing data in this domain

❖ Beware of the **interplay**: segregation causes polarization and vice-versa
ARC²S: Applied Research on Computational Complex Systems

Thanks!