



UNIVERSITÀ DI PISA

SOCIAL NETWORK ANALYSIS  
A.A. 2017/2018

# **Cambridge Analytica and Facebook: The Scandal and the Fallout on Twitter**

Gianmarco Ricciarelli 555396  
Stefano Carpita 304902

*Data drives all we do.*

Cambridge Analytica main slogan.

*Rules don't matter for them.  
For them, this is a war, and it's all fair.*

Christopher Wylie,  
former Cambridge Analytica datascientist, about its leaders.

# Contents

<b>1</b>	<b>The case story</b>	<b>1</b>
<b>2</b>	<b>Building the network</b>	<b>2</b>
<b>3</b>	<b>Network properties</b>	<b>3</b>
<b>4</b>	<b>Network dynamics</b>	<b>4</b>
<b>5</b>	<b>Communities discovery</b>	<b>5</b>
<b>6</b>	<b>Spreading</b>	<b>6</b>
6.1	SI model . . . . .	7
6.2	SIS model . . . . .	8
6.3	SIR model . . . . .	8
6.4	Threshold model . . . . .	8
<b>7</b>	<b>Summary</b>	<b>9</b>

# 1 | The case story

On Saturday 17 of March 2018, the newspapers The Observer and The New York Times broke reports on how the consulting firm Cambridge Analytica harvested private information from the Facebook profiles of more than 50 million users without their permission, making it one of the largest data leaks in the social network's history. [1]. REF OBSERVER

The whistleblower Christopher Wylie, datascientist and former director of research at Cambridge Analytica revealed... Cambridge Analytica described itself as a company providing consumer research, targeted advertising and other data-related services to both political and corporate clients.

What, Where, Who, Why, Where ?

Timeline da sistemare: [2]

- March 17, 2018: The Observer and The New York Times publish joint reports on data harvesting by Cambridge Analytica. UK Information Commissioner Elizabeth Denham issues statement that they are "investigating circumstances in which Facebook data may have been illegally acquired and used." Politicians in US and UK call for investigation.
- March 19, 2018: Channel 4 News publishes part 1 of their undercover investigation into Cambridge Analytica. Facebook sends investigators to Cambridge Analytica's offices. UK Information Commissioner orders them to stand down.
- March 20, 2018: Channel 4 News publishes part 2 of their undercover investigation into Cambridge Analytica, where they boast about getting Donald Trump elected. British MP Damian Collins calls on Facebook to present oral evidence on Cambridge Analytica. Facebook agrees to send former operations manager Sandy Parakilas. Facebook holds internal Q&A with attorney Paul Grewal to discuss the crisis, but CEO Mark Zuckerberg and COO Sheryl Sandberg do not attend. Cambridge Analytica suspends CEO Alexander Nix. Facebook demands to inspect Christopher Wylie's phone. FTC opens investigation into Facebook.
- to be continued...

## 2 | Building the network



Figure 2.1: New authors time history

### 3 | Network properties



Figure 3.1: New authors time history

## 4 | Network dynamics

# 5 | Communities discovery





In this chapter we'll describe the results we obtained by applying the **SI**, **SIS**, **SIR**, and **Threshold** diffusion models both on the crawled data and on the synthetic graphs (Erdős–Rényi and Barabási–Albert) generated from the original one. In each section, a comparison between the three networks will be provided both for the trend and for the prevalence of every model.

## 6.1 SI model

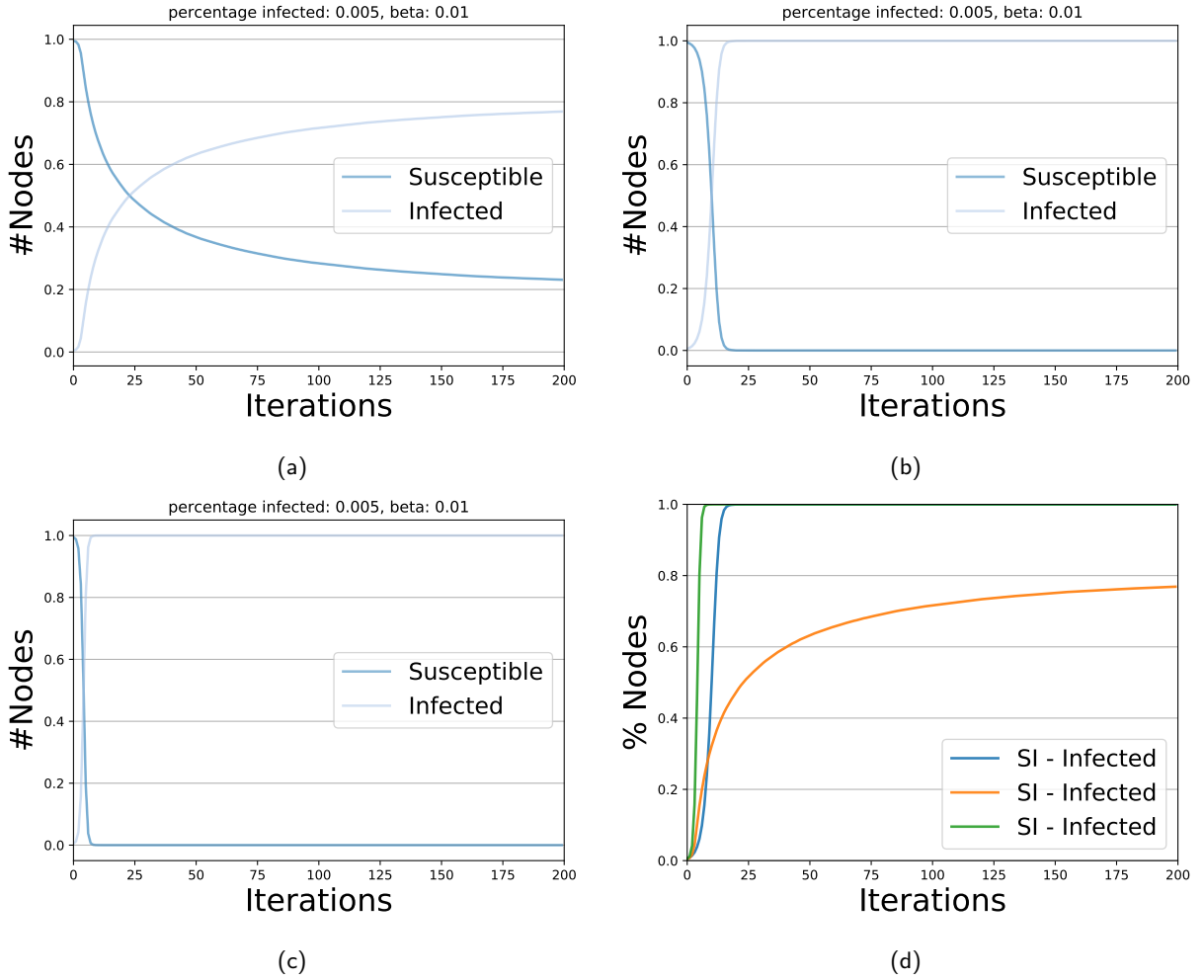


Figure 6.1: In Figure 6.1a we can see the diffusion graph for the original network, while in Figure 6.1b and in Figure 6.1c we can see the diffusion graph for the Erdős–Rényi and Barabási–Albert networks, respectively. In Figure 6.1d we can see a comparison between the infection rate of the three networks.

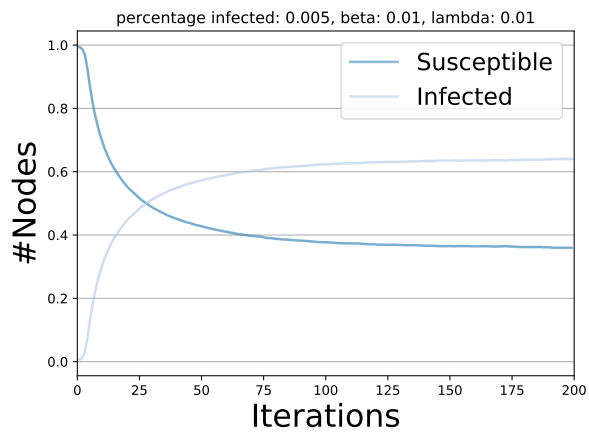
For the **Susceptible-Infected** model we've started with a 0.005% of the total population (3 nodes) of each network being infected, and we've chosen a value of 0.01 for the infection rate  $\beta$ . As you can see from Figure 6.1, the original network is the only one that doesn't reach the saturation regime, while the other networks reach it within the first 25 iterations of the model. This is due to the fact that both the Erdős–Rényi and the Barabási–Albert network are extremely connected, hence it is more easy for the infection to spread among the nodes. For this model we obtain that the **fraction of infected individuals** increases in time as

$$i = \frac{i_0 e^{\beta \langle k \rangle t}}{1 - i_0 + i_0 e^{\beta \langle k \rangle t}} = \frac{3e^{0.38t}}{65726 + 3e^{0.38t}},$$

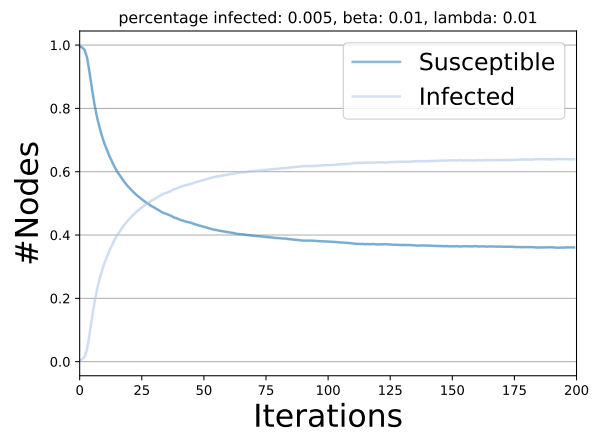
and the **characteristic time** required to reach  $\frac{1}{e}$  fraction of all susceptible individuals is

$$\tau = \frac{1}{\beta \langle k \rangle} = \frac{1}{0.38} = 2.63.$$

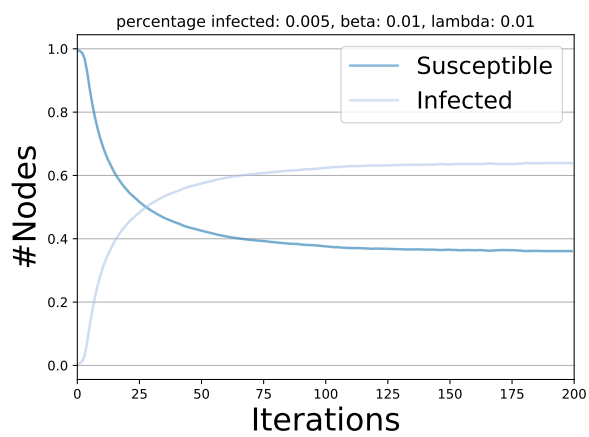
## 6.2 SIS model



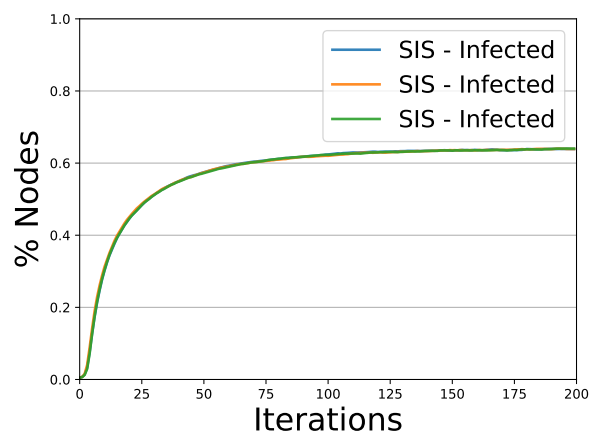
(a)



(b)



(c)



(d)

## 6.3 SIR model

## 6.4 Threshold model



# References

- [1] New York Times. *How Trump Consultants Exploited the Facebook Data of Millions*. <https://www.nytimes.com/2018/03/17/us/politics/cambridge-analytica-trump-campaign.html>. [Online; accessed 19-May-2018]. 2018.
- [2] New York Times. *Cambridge Analytica and Facebook: The Scandal and the Fallout So Far*. <https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html>. [Online; accessed 19-May-2018]. 2018.