

Flu epidemic and Graph Network

Project Proposal

Louis Veillon
ESSEC-CENTRALESUPELEC
Paris, France
louis.veillon@essec.edu

Benjamin Levain
ESSEC-CENTRALESUPELEC
Paris, France
benjamin.levain@essec.edu

Chuyi Xu
ESSEC-CENTRALESUPELEC
Paris, France
chuyi.xu@essec.edu

Kilian Tep
ESSEC-CENTRALESUPELEC
Paris, France
kilian.tep@essec.edu

KEYWORDS

Epidemic modelling, Model prediction, Pandemic influenza, SIR, Graph Network

ACM Reference Format:

Louis Veillon, Chuyi Xu, Benjamin Levain, and Kilian Tep. 2018. Flu epidemic and Graph Network: Project Proposal. In *Proceedings of ACM Conference (Conference'17)*. ACM, New York, NY, USA, 3 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

1 INTRODUCTION

The project aims to model an epidemic with graph networks and to verify some theoretical results from well known papers in epidemiology. To model an epidemic, we will focus on the well known SIR model from Kermack and McKendrick [3], implementing a graph network and generating Monte Carlo simulation to check theoretical results built on SIR model. To check the different theoretical models from literature, we will use data from Flu epidemic in the United State during the 2016-2017 period.

Flu can be modelled only with one season because this epidemic occurs mainly during Winter (Figure 1).

2 SIR MODEL

The SIR model is a behavioral model in epidemiology and serve as a base mathematical framework of the main paper in this area. It helps understand complex dynamics of the spread of the disease.

The model splits the population into three categories:

- (1) Susceptible: (denoted by S). The number of people from the population susceptible to be ill. Thus, vaccinated people are not considered susceptible.
- (2) Infected: (denoted by I). The number of people from the population infected.
- (3) Recovered (denoted by R) The number of people from the population who have recovered and are now immune.

The dynamics of these three variables over time can be expressed by the following differential equations:

$$\frac{dS}{dt} = -\frac{\beta IS}{N}$$

$$\frac{dI}{dt} = \frac{\beta IS}{N} - \frac{I}{\lambda}$$

$$\frac{dR}{dt} = \frac{I}{\lambda}$$

N is the number of people within the population. We assume that the population is constant over time because the dynamics of death and birth are often slower than the dynamics of Flu. Then:

$$\frac{dS}{dt} + \frac{dI}{dt} + \frac{dR}{dt} = 0$$

β is the incidence ratio. Then βIS is the number of people who start out as susceptible and become infected during dt . We can express β as the product between the number of people met during dt times the probability that an infected person transmits the virus to someone else.

λ is the average number of days a person is ill. Then $\frac{I}{\lambda}$ describes the flow to which the category "infected people" switch to the category "recovered people."

Figure 2 provides the number of infected people model with a random degree distribution and standard parameters.

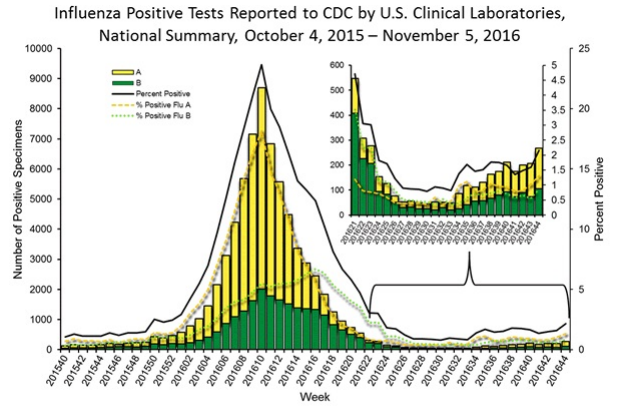


Figure 1: Flu positive test over 2016-2017 in the U.S

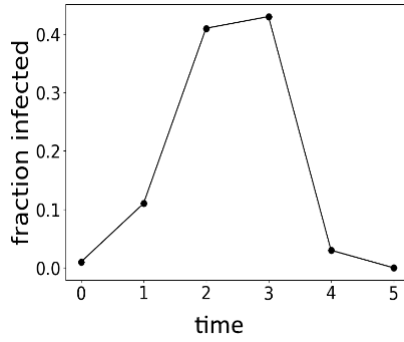


Figure 2: Number of infected people over time from random degree distribution

3 MODEL OF FLU EPIDEMIC WITH GRAPH NETWORK

Our goal is to test three models from literature and check if they are suitable to model flu.

3.1 Pastor-Satorras and Vespignani (2001) [4]

First, the model follows an exponential distribution. The first result from this paper is, when $\beta_0 \bar{k} \lambda_{s_0} > 1$, the network will converge to a steady state in which $I(t) > 0$. Then, we want to check if this result can be applied to our case of flu epidemic.

Second, the edge distribution follows a Power-law degree. $P(k) \sim k^{-\nu}$. In the same way, we want to test if flu epidemic can be modelled by Power-law degree distribution. Then, we want to apply the result from this paper to flu epidemic in the U.S. Here, the authors show that:

- for $2 < \nu < 3$ the network will converge to a steady state in which $I(t) > 0$ for every set of parameter values.
- for $3 < \nu$ the infection will be endemic in the network if and only if

$$\beta_0 \lambda > \frac{\nu - 1}{m\nu}$$

3.2 Wang et al. and Ganesh et al., (2005) [2, 5]

Here, the authors model the network as an arbitrary network. The result from the paper is that $\beta \lambda < \frac{1}{\lambda_A}$, where λ_A is the largest eigenvalue.

3.3 Watts et al., 2005 [6]

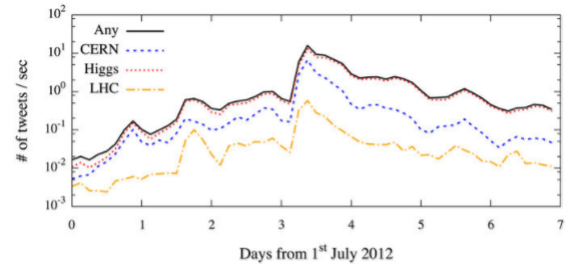
The authors use a stochastic approach to model epidemics. We want to check if this approach is suited for our flu model in the U.S.

4 DATA

4.1 Flu Data

Our project data focus on flu epidemic in 2016-2017 in all U.S. states. We have gathered data from the Centers for Disease Control and Prevention¹, which gathers weekly data from flu epidemic. These

¹<https://www.cdc.gov/flu/>



The curves correspond to tweets containing only the CERN, Higgs, LHC keywords and at least one of them, respectively.

Figure 3: Number of retweet over time

data are aggregated by state and by week. An interactive data visualization² is also provided by the website.

The most accurate data representing flu epidemic and the extent to which people are infected are the number of respiratory specimens tested (the entire population) and the number of people tested positive for influenza types A and B from this population. These data are collected by American laboratories³.

Furthermore, we need the vaccinated ratio among our population. The Centers for Disease Control and Prevention gives this ratio on a weekly basis, by state and by age group.

Finally, we have determined some parameters:

- λ the number of days a people is ill is set to 6 days.
- β is the product between the number of people met during dt time the probability that an infected person transmits the virus to someone else. Then, the number of people met during dt is k , the number of edges for one node in the graph. We set p_{inf} at 0.3.

However, this data set does not provide any interactions between people and then we can precisely test our model.

4.2 Analogy with information spreading processes

The Higgs dataset <https://snap.stanford.edu/data/higgs-twitter.html> provide full data about interaction between people and has been built after monitoring the spreading processes on Twitter before, during and after the announcement of the discovery of a new particle with the features of the elusive Higgs boson on 4th July 2012. The messages posted in Twitter about this discovery between 1st and 7th July 2012 are considered.

Then we want to model this spreading processes as a SIR model. The Susceptible people are the susceptible to get information but they don't get yet. The infected people are people who have get information. And Recovered people are people don't interested in this information anymore.

The number of re-tweet over time (Figure 3)[1] in log scale as a similar trend as the number of infected people trend given by SIR model.

²<https://gis.cdc.gov/grasp/fluview/fluportaldashboard.html>

³Exclusively WHO and NREVSS collaborating laboratories

This data set will enable us to precisely compare our model with the reality, lead a deep network analysis and try to predict human behavior in information spreading processes.

5 CONCLUSION AND EXPECTED RESULTS

We want to find the best degree distribution of the graph network to find the number of infected people the nearest of the reality. We know that this project does not model the true interaction between population and that's why the results may not be very accurate.

However, we will try to make the model the most accurate by focusing for example on population density in each state, data being given by state.

Besides, we will try to detect some irregularities in data compare to results from the model. For example, if the number of infected people anomaly drops.

REFERENCES

- [1] M. De Domenico, A. Lima, P. Mougél, and M. Musolesi. The anatomy of a scientific rumor. 3:2980, 10 2013.
- [2] A. Ganesh, L. Massoulié, and D. Towsley. The effect of network topology on the spread of epidemics. In *INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE*, volume 2, pages 1455–1466. IEEE, 2005.
- [3] Kermack and McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 115(772):700–721, 1927.
- [4] R. Pastor-Satorras and A. Vespignani. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.*, 86:3200–3203, Apr 2001.
- [5] Y. Wang, D. Chakrabarti, C. Wang, and C. Faloutsos. Epidemic spreading in real networks: An eigenvalue viewpoint. In *Reliable Distributed Systems, 2003. Proceedings. 22nd International Symposium on*, pages 25–34. IEEE, 2003.
- [6] D. J. Watts, R. Muhamad, D. C. Medina, and P. S. Dodds. Multiscale, resurgent epidemics in a hierarchical metapopulation model. *Proceedings of the National Academy of Sciences of the United States of America*, 102(32):11157–11162, 2005.