Modelling and Analysis of Complex Networks

— Semester 3, Master of Data Science —

George Panagopoulos
University of Luxembourg

Network Effects and Cascading Behaviour

Spreading through Networks

- Spreading through networks: cascading behaviour, diffusion of innovations, network effects, epidemics
- Behaviours that cascade from node to node like an epidemic!
- Examples:
  - Biological: diseases via contagion, e.g. COVID-19
  - Technological: cascading failures, spread of information
  - Social: misinformation, news, viral marketing

Spreading through Networks

- Spreading through networks: cascading behaviour, diffusion of innovations, network effects, epidemics
- Behaviours that cascade from node to node like an epidemic!
- Examples:
  - Biological: diseases via contagion, e.g. COVID-19
  - Technological: cascading failures, spread of information
  - Social: misinformation, news, viral marketing

Spreading through Networks

Social media post sharing

Spreading through Networks

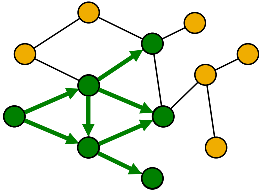
Viral marketing: product adoption

Spreading through Networks

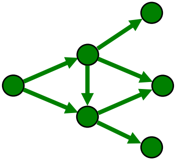
Spread of disease: COVID-19, Ebola, etc.

Network Cascades

- Contagion spreads over the edges of the network
- It creates a propagation tree, i.e., cascade



Network



Cascade (propagation graph)

- Contagion: stuff that spreads
- Adaption, infection, activation: “infection” event
- Infected/active nodes, adopters

How Do We Model Diffusion?

- Decision-based models:
  - Models of product adoption, decision making (a node observes the decisions of its neighbours and makes its own decision)
  - Example: you join a demonstration if k of your friends do so
- Probabilistic models:
  - Models of influence or disease spreading (an infected node tries to "push" the contagion to an uninfected node)
  - Example: you "catch" a disease with some probability from each active neighbour in the network

How Do We Model Diffusion?

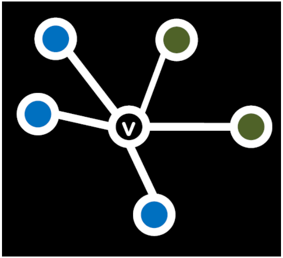
- Decision-based models:
  - Models of product adoption, decision making (a node observes the decisions of its neighbours and makes its own decision)
  - Example: you join a demonstration if k of your friends do so
- Probabilistic models:
  - Models of influence or disease spreading (an infected node tries to "push" the contagion to an uninfected node)
  - Example: you "catch" a disease with some probability from each active neighbour in the network

Decision-based Diffusion Models

Game Theoretic Model of Cascades [Morris, 2000]

Based on 2-player coordination game:

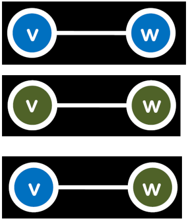
- Two players: each chooses option A or B
- Each player can only adopt one "behaviour", either A or B
- The player gains more payoff if his friend has adapted the same behaviour



Local view of the network of node v.

The Model for Two Nodes (Players)

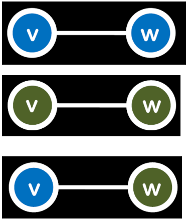
- Payoff matrix:
  - If both v and w adopt behaviour A, they each get payoff $a > 0$
  - If both v and w adopt behaviour B, they each get payoff $b > 0$
  - If both v and w adopt the opposite behaviours, they each get 0



- In a large network:
  - Each node v is playing a copy of the game with each of its neighbours
  - Payoff: sum of node payoffs per game

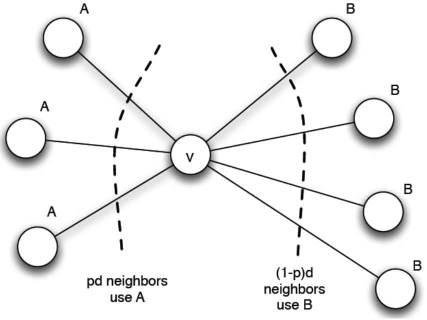
The Model for Two Nodes (Players)

- Payoff matrix:
  - If both v and w adopt behaviour A, they each get payoff $a > 0$
  - If both v and w adopt behaviour B, they each get payoff $b > 0$
  - If both v and w adopt the opposite behaviours, they each get 0



- In a large network:
  - Each node v is playing a copy of the game with each of its neighbours
  - Payoff: sum of node payoffs per game

Calculation of Node v

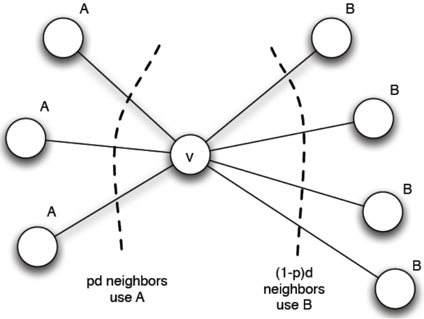


Let v have d neighbours. Assume fraction p of v's neighbours adapt A.

$$\begin{aligned} Payoff_a &= a \cdot p \cdot d && \text{if } v \text{ chooses } A \\ Payoff_b &= b \cdot (1-p) \cdot d && \text{if } v \text{ chooses } B \end{aligned}$$

So v chooses A if $p = \frac{a}{d} > \frac{b}{d} = \frac{b}{a+b} = q$ (e.g. payoff threshold).

Calculation of Node v

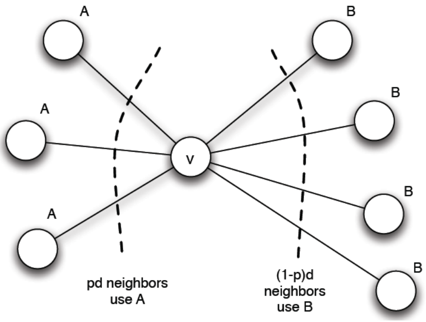


Let v have d neighbours. Assume fraction p of v's neighbours adapt A.

$$
\begin{aligned}
\mathit{Payoff}_a &= a \cdot p \cdot d & \text{if } v \text{ chooses } A \\
\mathit{Payoff}_b &= b \cdot (1-p) \cdot d & \text{if } v \text{ chooses } B
\end{aligned}
$$

So v chooses A if $p = \frac{a}{d} > \frac{b}{d} = \frac{b}{a+b} = q$ (e.g. payoff threshold).

Calculation of Node v

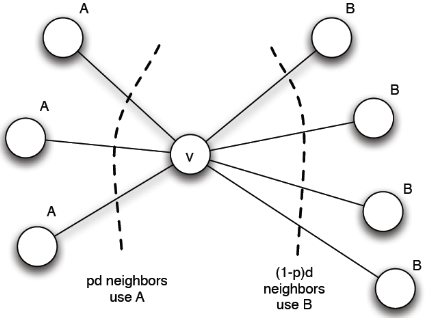


Let v have d neighbours. Assume fraction p of v's neighbours adapt A.

$$\begin{aligned} Payoff_a &= a \cdot p \cdot d && \text{if } v \text{ chooses } A \\ Payoff_b &= b \cdot (1-p) \cdot d && \text{if } v \text{ chooses } B \end{aligned}$$

So v chooses A if $p = \frac{a}{d} > \frac{b}{d} = \frac{b}{a+b} = q$ (e.g. payoff threshold).

Calculation of Node v



Let v have d neighbours. Assume fraction p of v's neighbours adapt A.

$$\begin{array}{lll} Payoff_a & = a \cdot p \cdot d & \text{if } v \text{ chooses } A \\ Payoff_b & = b \cdot (1-p) \cdot d & \text{if } v \text{ chooses } B \end{array}$$

So v chooses A if $p = \frac{a}{d} > \frac{b}{d} = \frac{b}{a+b} = q$ (e.g. payoff threshold).

An Example Scenario

Consider the following scenario:

- A graph where everyone starts with all B
- A small set S of early adopters of A (hard-wire S: they keep using A no matter what payoffs tell them to do)
- Assume payoffs are set in such a way that nodes will
  - take A, if $q \geq 50\%$ of their friends take A,

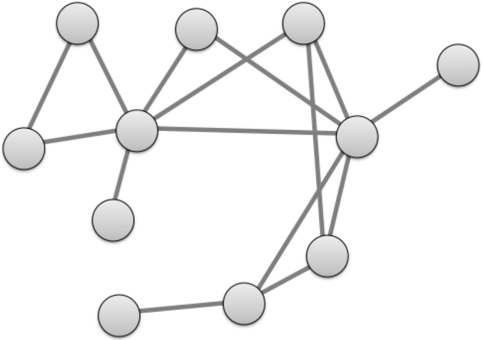
An Example Scenario

Consider the following scenario:

- A graph where everyone starts with all B
- A small set S of early adopters of A (hard-wire S: they keep using A no matter what payoffs tell them to do)
- Assume payoffs are set in such a way that nodes will
  - take A, if $q \geq 50\%$ of their friends take A,

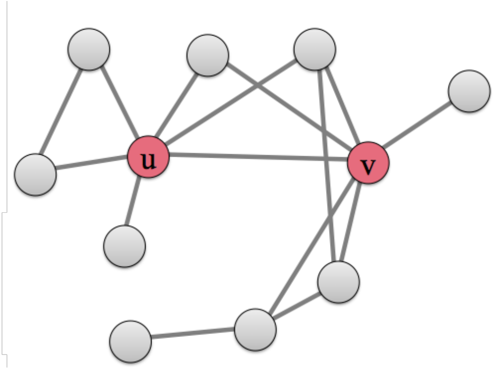
An Example Scenario

Example scenario: $S = \{u, v\}$, if $q > 50\%$ of my neighborhood is A, I will also be A.



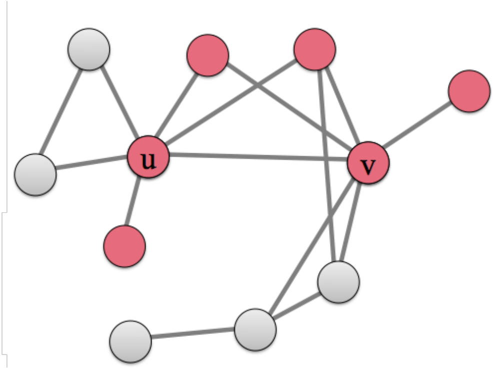
An Example Scenario

Example scenario: $S = \{u, v\}$, if $q > 50\%$ of my friends are read, I will also be red.



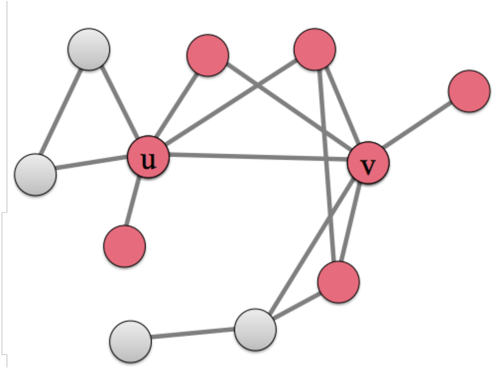
An Example Scenario

Example scenario: $S = \{u, v\}$, if $q > 50\%$ of my friends are read, I will also be red.



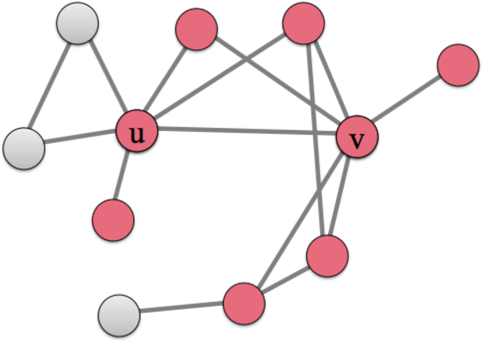
An Example Scenario

Example scenario: $S = \{u, v\}$, if $q > 50\%$ of my friends are read, I will also be red.



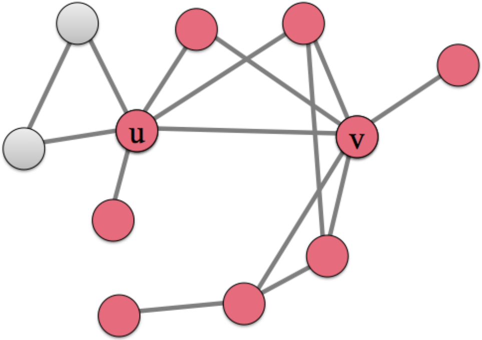
An Example Scenario

Example scenario: $S = \{u, v\}$, if $q > 50\%$ of my friends are read, I will also be red.



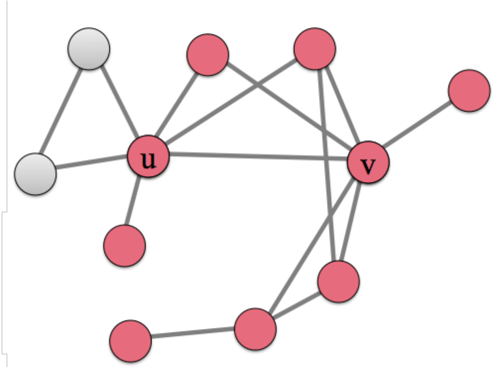
An Example Scenario

Example scenario: $S = \{u, v\}$, if $q > 50\%$ of my friends are read, I will also be red.



An Example Scenario

Example scenario: $S = \{u, v\}$, if $q > 50\%$ of my friends are read, I will also be red.



In this game we assume the players are rational.

Probabilistic Diffusion & Epidemic Models

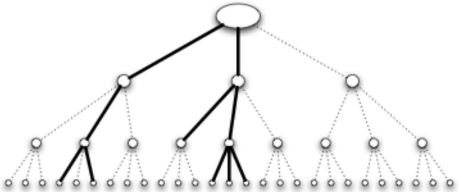
Epidemics vs Cascade Spreading

- In decision based models, nodes make decisions based on payoff benefits of adopting one strategy or the other
- In epidemic spreading,
  - Lack of decision making
  - Process of contagion is complex and unobservable
  - In some cases it involves or can be modeled as randomness

Epidemics vs Cascade Spreading

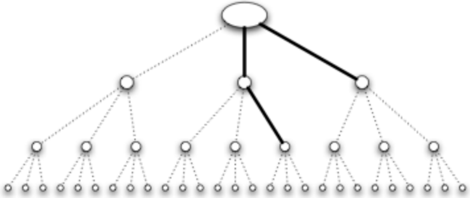
- In decision based models, nodes make decisions based on payoff benefits of adopting one strategy or the other
- In epidemic spreading,
  - Lack of decision making
  - Process of contagion is complex and unobservable
  - In some cases it involves or can be modeled as randomness

Probabilistic Spreading Models



High contagion probability:
The disease spreads

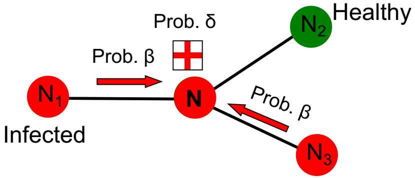
Low contagion probability:
The disease dies out



Spreading Models for Viruses

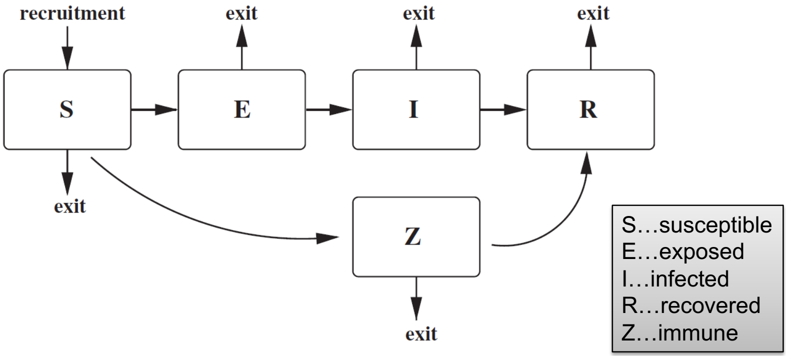
Virus propagation: two parameters

- Virus transmission rate β (probability that a susceptible node gets infected by a neighbor)
- Virus death rate δ (probability that an infected node heals)



A General Scheme

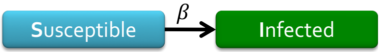
General scheme for epidemic models:



Each node can go through phases, and transition probabilities are governed by the model parameters.

SI Model

- Susceptible-Infective (SI) model

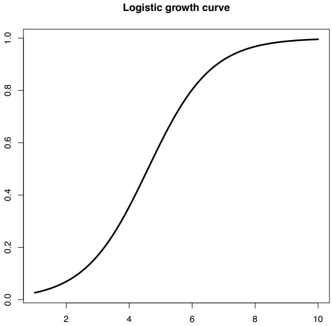


- Models HIV

  (once you get infected, you can never recover and stay infectious)

- Assuming perfect mixing: a complete graph

$$\frac{dS}{dt} = -\beta SI, \qquad \frac{dI}{dt} = \beta SI$$



SIR Model

- Susceptible-Infective-Recovered (SIR) model

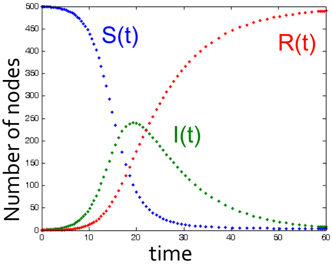
Susceptible $\xrightarrow{\beta}$ Infected $\xrightarrow{\delta}$ Recovered

- Models chickenpox or plague

  (once you heal, you can never get infected again)

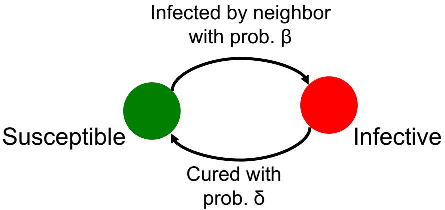
- Assuming perfect mixing: a complete graph

$$\frac{dS}{dt} = -\beta SI, \qquad \frac{dR}{dt} = \delta I, \qquad \frac{dI}{dt} = \beta SI - \delta I$$



SIS Model

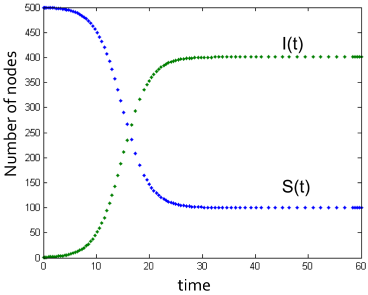
- Susceptible-Infective-Susceptible (SIS) model
- Cured nodes immediately become susceptible
- Virus "strength": $s = \beta/\delta$



SIS Model

- Models flu:
  - A susceptible node becomes infected
  - The node then heals and becomes susceptible again
- Assuming perfect mixing: a complete graph

$$\frac{dS}{dt} = -\beta SI + \delta I, \qquad \frac{dI}{dt} = \beta SI - \delta I$$



SIS Model

Epidemic threshold of an arbitrary graph G is τ, such at
if the virus "strength" $s = \beta/\delta < \tau$, the epidemic cannot happen.

SIS Model

Epidemic threshold of an arbitrary graph G is τ, such at
if the virus “strength” $s = \beta/\delta < \tau$, the epidemic cannot happen.

Given a graph G represented by adj. A, we have no epidemic if:

Epidemic threshold

(Virus) Death rate

$$\boxed{\beta/\delta < \tau = 1/\lambda_{1,A}}$$

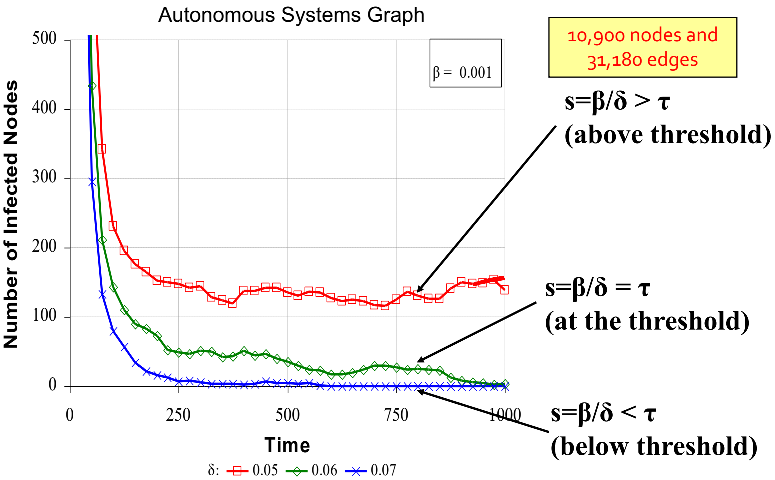
(Virus) Birth rate

largest eigenvalue
of adj. matrix **A** of ***G***

$\lambda_{1,A}$ alone captures the property of the graph!

([Wang et al. 2003])

SIS Model



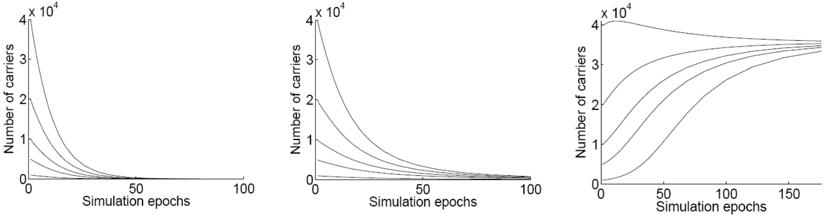
[Wang et al. 2003]

SIS Model

Does it matter how many people are initially infected?

SIS Model

Does it matter how many people are initially infected?



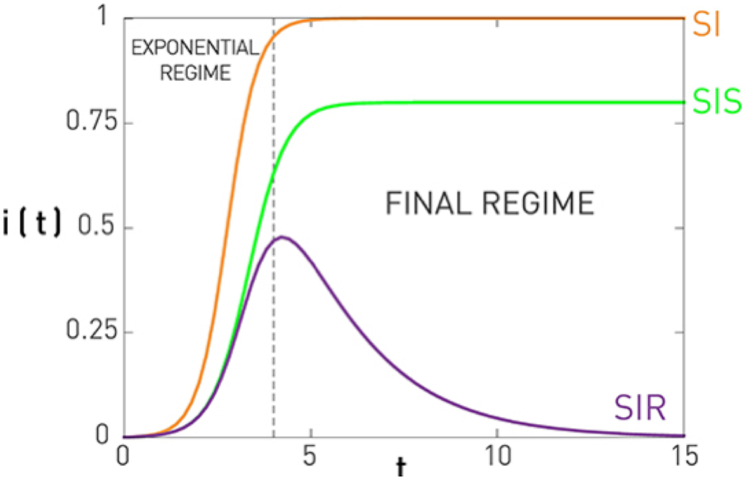
(a) Below the threshold, s=0.912

(b) At the threshold, s=1.003

(c) Above the threshold, s=1.1

[Wang et al. 2003]

SI, SIR, SIS in One Picture



Real-life Applications

Application on Real Networks

Rumour spread modelling using SEIZ Model

Epidemiological Modeling of News and Rumors on Twitter. [Jin et al. SNAKDD 2013]

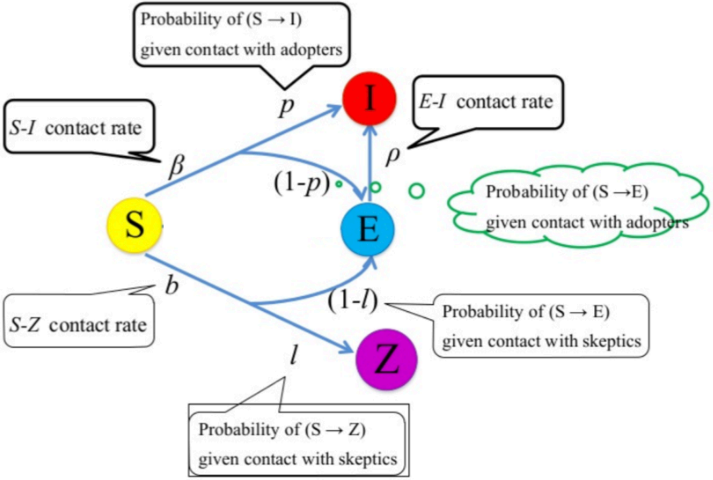
SEIZ Model

An Extension of SIS Model:

Disease		Twitter
Susceptible	S	Twitter accounts
Infected	I	Believe news / rumor, (I) post a tweet
Exposed	E	Be exposed but not yet believe
Skeptics	Z	Skeptics, do not tweet

SEIZ Model

S= Susceptible, I = Infected, E= Exposed, Z= Skeptics



Application on Real Networks

Tweets collected from eight stories:

- Real events
  - Boston Marathon Explosion, 15/04/2013
  - Pope Resignation, 11/02/2013
  - Venezuela's Refinery Explosion, 25/08/2012
  - Michelle Obama at the 2013 Oscars, 24/02/2013
- Rumours
  - Obama Injured, 23/04/2013
  - Doomsday Rumour, 21/12/2012
  - Fidel Castro's Coming Death, 15/10/2012
  - Riots and Shooting in Mexico, 05/09/2012

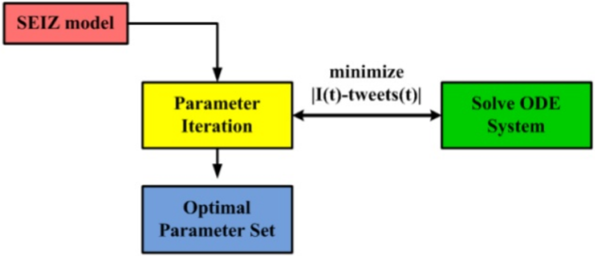
Application on Real Networks

Tweets collected from eight stories:

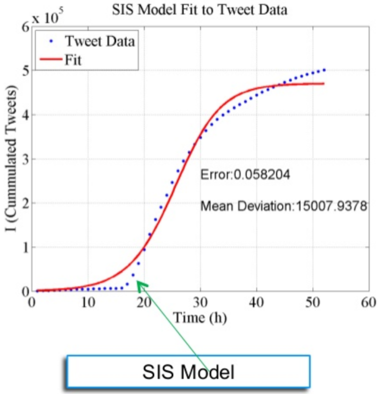
- Real events
  - Boston Marathon Explosion, 15/04/2013
  - Pope Resignation, 11/02/2013
  - Venezuela's Refinery Explosion, 25/08/2012
  - Michelle Obama at the 2013 Oscars, 24/02/2013
- Rumours
  - Obama Injured, 23/04/2013
  - Doomsday Rumour, 21/12/2012
  - Fidel Castro's Coming Death, 15/10/2012
  - Riots and Shooting in Mexico, 05/09/2012

Method: Fitting SEIZ Model to Data

- SEIZ model is solved as an ODE and provides $I(t)$: an estimated number of rumour tweets.
- Fit to each cascade to minimize the difference between $|I(t) - tweets(t)|$, where $tweets(t)$ are the rumour tweets.
- Use grid-search: iterate through all possible parameters and keep the ones that give a minimum error.
- Each of the aforementioned cascades has different parameters.



Method: Fitting SEIZ Model to Cascade Data

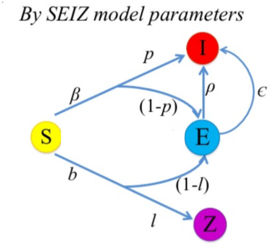


Error = norm(I – tweets) / norm(tweets)

SEIZ model better models the real data, especially at initial points

Method: Rumour Detection with SEIZ Model

- Define a new metric that quantifies whether a cascade is a rumor or not, using the fit parameters.

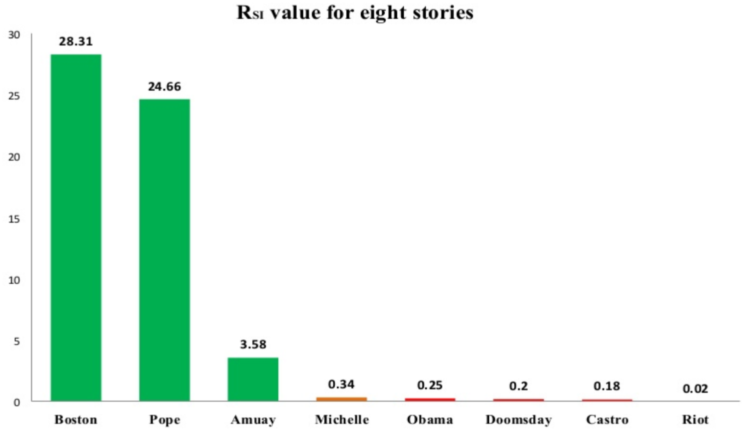


Notation:
S = Susceptible
I = Infected
E = Exposed
Z = Skeptics

$$R_{SI} = \frac{(1-p)\beta + (1-l)b}{p+e} \quad (1)$$

- R_{SI} represents the ratio of the users entering state E to those leaving state E.

Method: Rumour Detection with SEIZ Model



Parameters obtained by fitting SEIZ model efficiently identifies rumors vs. news

Real-life Application

Modeling protest recruitment on social networks:

- Anti-austerity protest in Spain May 15-22, 2011 as a response to the financial crisis
- Twitter was used to organise and mobilise users to participate in the protest

"The Dynamics of Protest Recruitment through an Online Network"

[Bailon et al., Nature Scientific Reports, 2011]

Real-life Application

Modeling protest recruitment on social networks:

- Anti-austerity protest in Spain May 15-22, 2011 as a response to the financial crisis
- Twitter was used to organise and mobilise users to participate in the protest

"The Dynamics of Protest Recruitment through an Online Network"

[Bailon et al., Nature Scientific Reports, 2011]

Real-life Application

Modeling protest recruitment on social networks:

- 70 hashtags were crawled for one month period: 581,750 tweets
- Relevant users who tweeted any relevant hashtags and its followers and followees: 87,569 users
- Full social (directed) network with all Twitter follow links
- Symmetric 'strong' social network with only the reciprocal follow links: $i \rightarrow j$ and $j \rightarrow i$

Real-life Application

Modeling protest recruitment on social networks:

- 70 hashtags were crawled for one month period: 581,750 tweets
- Relevant users who tweeted any relevant hashtags and its followers and followees: 87,569 users
- Full social (directed) network with all Twitter follow links
- Symmetric 'strong' social network with only the reciprocal follow links: $i \rightarrow j$ and $j \rightarrow i$

Real-life Application

Modeling protest recruitment on social networks:

- User activation time: moment when user starts tweeting protest messages
- k_n: total number of neighbours when a user became active
- k_a: number of active neighbours when a user became active
- k_a/k_n: activation threshold
  - If $k_a/k_n \approx 0$, the user joins the movement when very few neighbours are active, i.e., no social pressure
  - If $k_a/k_n \approx 1$, the user joins the movement after most of its neighbours are active, i.e., high social pressure

Real-life Application

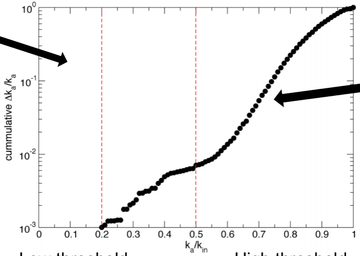
Modeling protest recruitment on social networks:

- User activation time: moment when user starts tweeting protest messages
- k_n: total number of neighbours when a user became active
- k_a: number of active neighbours when a user became active
- k_a/k_n: activation threshold
  - If $k_a/k_n \approx 0$, the user joins the movement when very few neighbours are active, i.e., no social pressure
  - If $k_a/k_n \approx 1$, the user joins the movement after most of its neighbours are active, i.e., high social pressure

Real-life Application

- Is a user more likely to protest if several neighbors become active in a short time period?
- Calculate the burstiness of active neighbors as $\Delta k_a/k_a = (k_a^{t+1} - k_a^t)/k_a^{t+1}$. Compare it with the threshold:

Low threshold users are insensitive to recruitment bursts.



High threshold users join after sudden bursts in neighborhood activation

Low threshold users

High threshold users

Users who are susceptible to high social pressure are effected by bursts in neighborhood adoption.

Real-life Application

Social cascades on Flickr and estimating R_0 from real data:

"Characterizing social cascades in Flickr" [Cha et al., ACM WOSN, 2008]

- The basic reproduction number of popular photos is between 1 and 190
- This is much higher than very infectious diseases like measles, indicating that social networks are efficient transmission media and online content can be very infectious.

Real-life Application

Social cascades on Flickr and estimating R_0 from real data:

"Characterizing social cascades in Flickr" [Cha et al., ACM WOSN, 2008]

- The basic reproduction number of popular photos is between 1 and 190
- This is much higher than very infectious diseases like measles, indicating that social networks are efficient transmission media and online content can be very infectious.

Real-life Application

Social cascades on Flickr and estimating R_0 from real data:

"Characterizing social cascades in Flickr" [Cha et al., ACM WOSN, 2008]

- The basic reproduction number of popular photos is between 1 and 190
- This is much higher than very infectious diseases like measles, indicating that social networks are efficient transmission media and online content can be very infectious.