

# How do we Spread on Twitter?

**Erick Stattner**, Reynald Eugenie and Martine Collard

LAMIA Laboratory  
University of the French West Indies and Guiana, France  
{erick.stattner, reynald.eugenie, martine.collard}@univ-ag.fr

May 13-15 2015  
Athens, Greece



# Introduction

## Context of diffusion

- Numerous models proposed to study diffusion phenomena
- Designed to reproduce global trends observed in real life

## Problem

- Represent the transmission process with simple mechanisms
- Do not reflect the complexity of involved individuals behaviors and poorly reproduce real situations

# Introduction

## In this work

- We address the problem of the diffusion of information by using **Twitter**
- Twitter is a good case study:
  - ▶ Wide variety of topics
  - ▶ 200 millions of users (business, professionals, politicians, etc.)
  - ▶ 500 millions of messages posted each day

## Objective

- Understand how the information spread through the network
- More particularly, identify in which conditions a person forwards an information

# Outline

- 1 Introduction
- 2 Previous works
  - Diffusion modeling
- 3 Methodology
- 4 Experimental results
- 5 Conclusion

# Introduction

## Diffusion modeling

### Modeling diffusion:

- Several models have been proposed to study diffusion processes
- Attempt to understand elementary transmission mechanisms in a population, that produce an epidemic
- Two main families of models
  - ▶ Compartment models
  - ▶ Network-based models

# Introduction

## Diffusion modeling

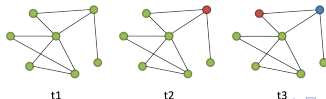
### Compartment models

- First mathematical models based on differential equations
  - ▶ Assume a population divided in compartments  
 ex. Susceptible, Infected, Recovered, ...
  - ▶ Individuals have a probability to change compartement



### Network-based models

- Based on the idea that the structure and the nature of network are main factors of the evolution of the process
  - ▶ Objective to take into account the complexity of interactions
  - ▶ Probability of being infected varies according to the individuals



# Introduction

## Diffusion modeling

### Modeling diffusion

- Understand the process [Lloyd,2001]
- Identify specific situations [Christley,2005]
- Test effects of intervention strategy [Christakis,2007]

### Problem

- Designed to reproduce global trends
  - ▶ Model the problem with simple transmission rules
  - ▶ Do not reflect the complexity of individuals behaviors
- For instance, [Albano,2012] show that traditional SIR model does not reproduce faithfully the diffusion on a p2p network

# Outline

- 1 Introduction
- 2 Previous works
- 3 Methodology
  - Diffusion on Twitter
  - Objective
  - Measures
- 4 Experimental results
- 5 Conclusion



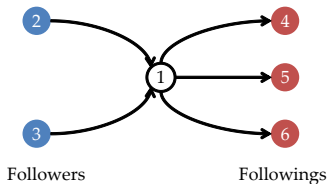


# Methodology

## Diffusion on Twitter

### Diffusion mechanism on Twitter

- Each user may posts short messages to its Followers( *Tweets* )
- Each time a user posts a message, his followers are likely to read it and to forward it to their own followers ( *Retweets* )



# Methodology

## Objective

### Objective of this work

- Understand how the information spreads through the network
- More particularly, understand user diffusion behaviors
- Answer questions like:
  - ▶ ( $Q_i$ ) Does an individual spread because he is highly followed?
  - ▶ ( $Q_{ij}$ ) Does an individual spread because he is swamped by his community?
  - ▶ ( $Q_{iii}$ ) How long does an individual wait before spreading?

# Methodology

## Measures

### Methodology proposed

- 1 Define (i) the targeted topic and (ii) the period of study
- 2 Each time a message is posted on the targeted topic, we extract two kinds of measures
  - ▶ **Global measures**  
describe globally the process and its evolution in time
  - ▶ **Local measures**  
aim to characterize the environment of a given user the first time he posts

# Methodology

## Measures

### Global measures

- **NbTw, NbRTw, NbUs:**  
the number of tweets, retweets and users captured on the period.
- **AVG(Tw/Us), AVG(RTw/Us):**  
average numbers of tweets and retweets sent per user.
- **AVG( $T_{tw}$ ), AVG( $T_{rtw}$ ):**  
average times (seconds) between two tweets and retweets sent on the targeted topic.

# Methodology

## Measures

**Local measures** (extracted each time a user  $u$  posts)

- **NbFe <sub>$u$</sub>** :  
the number Followers of user  $u$ .
- **NbFgP <sub>$u$</sub>** :  
the number of followings of user  $u$  who have posted before  $u$ .
- **NbT <sub>$u$</sub>** :  
the total number of Tweets sent on the topic by user  $u$ .
- **TotalR <sub>$u$</sub>** :  
the cumulative number of tweets sent by Followings of user  $u$ .
- **Elapsed <sub>$u$</sub>** :  
the time elapsed since user  $u$  tweeted initially (in hours).

# Outline

- 1 Introduction
- 2 Previous works
- 3 Methodology
- 4 Experimental results
  - Experiments
  - Macroscopic behaviors
  - Diffusion induced by individuals behaviors
  - Analyzing and Collecting Tool
- 5 Conclusion

# Experimental results

## Experiments

### Experiments

- Our measures have been used to study two events occurred in January 2015:
  - ▶ **Friday 23, HoloLens**  
The presentation by Microsoft of its new augmented reality headset, called HoloLens  
Obtained with: “HoloLens”; “Holo Lens”
  - ▶ **Sunday 25, Syriza**  
The election in Greece of the party of the radical left Syriza.  
Obtained with: “Syriza”; “Tsipras”

### Nature of the topics

Two different kinds of topics: Technological  $\neq$  political



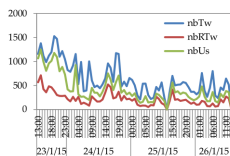
# Experimental results

## Macroscopic behaviors

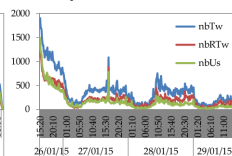
### Global measures

- For both events, the topic loses interest
- Syryza: the number of messages per user increases.

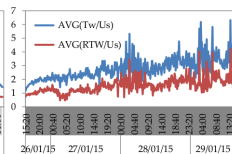
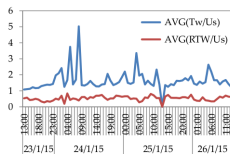
Microsoft HoloLens



Syryza Election



(a)



(b)

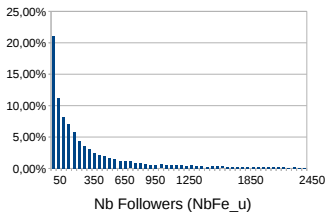
# Experimental results

## Diffusion induced by individuals behaviors

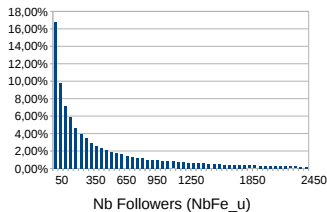
### Distribution of Number of Followers ( $NbFe_u$ )

- The majority of users who posted had a low number of followers
- $\approx 20\%$  of individuals have posted while they have less 20 followers
- The number of followers is not a motivation for posting (raised by ( $Qi$ )).

#### Microsoft HoloLens



#### Syriza Election



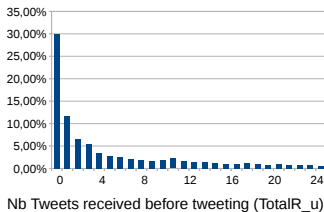
# Experimental results

## Diffusion induced by individuals behaviors

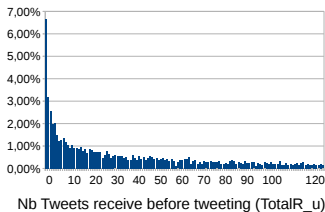
### Distribution of Number of Tweets received ( $TotalR_u$ )

- Trends are different
- HoloLens:  $\approx 30\%$  of users posted without receiving any message
- Syriza: it is about 7%
- Show that individuals are not waiting to be swamped by their community before reacting (raised by (Qii)).

#### Microsoft HoloLens



#### Syriza Election

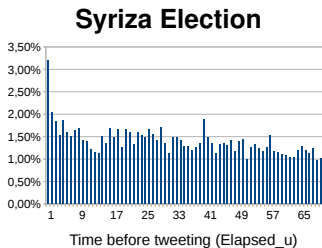
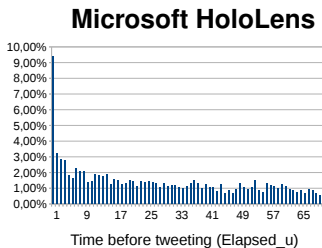


# Experimental results

## Diffusion induced by individuals behaviors

### Distribution of Elapsed time before Tweeting ( $Elapsed_u$ )

- Most reactions take place in the first hours of the beginning of the event and slowly decrease with time (raised by (Qiii))
- Some users post for the first time until 72 hours

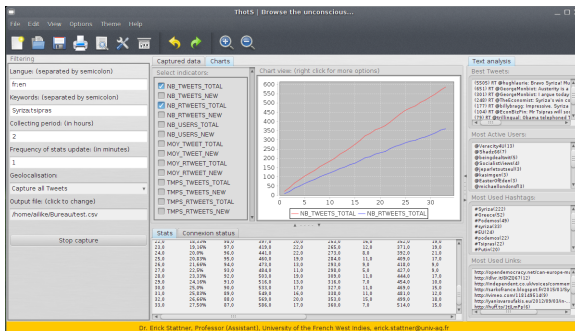


# Experimental results

## Analyzing and Collecting Tool

### Thots (<http://erickstattner.com/thots-analytics>)

- Define a targeted topic
- Explore data according to measures presented



Dr. Erick Stattner, Professor (Assistant), University of the French West Indies, erick.stattner@univ-aj.fr

# Outline

- 1 Introduction
- 2 Previous works
- 3 Methodology
- 4 Experimental results
- 5 Conclusion

# Conclusion

## Conclusion

- Problem of the diffusion of information
- Set of measures to study the diffusion of information on Twitter
  - ▶ Global measures  
describe the evolution of the phenomenon
  - ▶ Local measures  
provide information on the user when he posts the first time
- Approach applied on two case studies: a technological and a political topics
- ThotS: graphical tool that implements our measures and can be used to conduct the analysis on any kind of topics

# Conclusion

## Perspectives

- Extend the study on other kinds of topics
- Propose more realistic diffusion models taking into account the real individual behaviours
- Build predictive models



# Conclusion

**Thanks for your attention !**