

# Dynamic Mathematical Modeling of Information Diffusion in Online Social Networks

Feng Wang, Haiyan Wang, Kuai Xu

Division of Mathematical and Natural Sciences  
Arizona State University

- Information Diffusion in Online Social Networks (OSN)
- Challenges
- Related Work
- Spatio-Temporal Information Diffusion Problem
- Diffusive Logistic Model
- Experimental Results
- Conclusions and Future Work

- OSN provides a new channel/medium to spread information in addition to the traditional social community, for example, on-line political campaign, advertisement of new product, movie recommendation, etc
- Insights to information diffusion in OSN are critical
- Problem: How to conceptualize information diffusion?
- Information cascading, information propagation, information diffusion, information spreading

- Large scale
- Dynamic environment
- Diversified users
- Complex user interactions
- Lack of well-accepted micro-level (user-to-user) interaction model
- Lack of understanding of underlying diffusion network

- Understanding the structure of OSNs
  - friendship graph
  - interaction graph
  - clusters and communities
- Optimization problems
  - choose minimum number of influential nodes to maximize the diffusion
- Empirical study to quantitatively study the information diffusion
- Mathematical models for information diffusion over time

## Diffusion in Other Disciplines

- Biology - Epidemics
- Economics - Viral Marketing
- Sociology - Gossip, rumor
- Physics - Heat diffusion
- .....

- Dynamic mathematical models study the global feature of the network, ignoring the underlying networks and can accommodate dynamics, so it is a good candidate for modeling information diffusion in OSN
- A mathematical model is a set of equations which describe the behavior of a system
- Ordinary Differential Equation (ODE) vs. Partial Differential Equation (PDE)
- Deterministic vs. Probabilistic/Stochastic

- Mathematical models, for example, Susceptible Infectious (SI), have been widely used in mathematical biology, economics, etc
- There are new challenges for modeling information diffusion in OSN
  - continuous vs. discrete value
  - long term vs. short term
  - distance metric in OSN
  - small world scenario
  - easier and faster to interact in OSN than in traditional social networks



## Spatio-Temporal Diffusion Problem

- For a given information  $m$  initiated from a particular user called *source*  $s$ , after time period  $t$ , what is the density of influenced users at network distance  $x$  from the source. An influenced user is an user that actively votes or likes the information
- We propose
  - Distance metrics
  - Partial Differential Equation (PDE) - based Diffusive Logistic Model

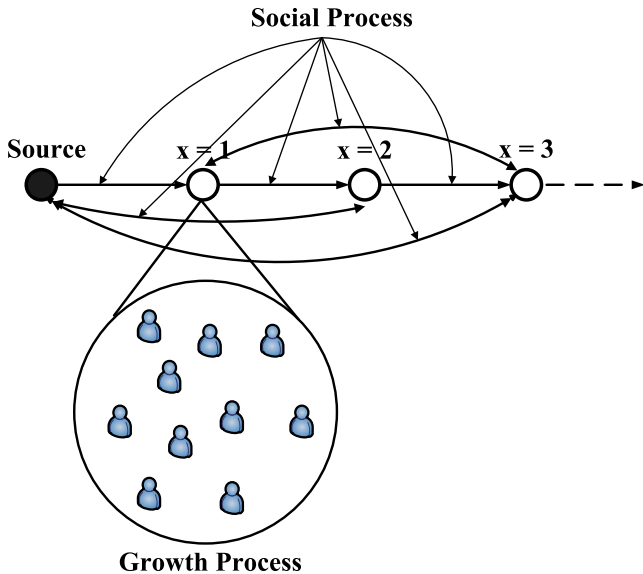
- We measure the network distance from two perspectives: *friendship hops* and *shared interests*.
- *friendship hop*: the number of friendship links on the shortest path from one user to another in the social friendship graph.
- *shared interests*: the distance between two users through their shared interests on information or content in social networks.

$$d_{a,b} = 1 - \frac{C_a \cap C_b}{C_a \cup C_b} \quad (0.1)$$

where  $C_a \cup C_b$  is the number of the total contents that either user  $a$  or user  $b$  has interacted with and  $C_a \cap C_b$  is the number of the shared contents that both users  $a$  and  $b$  have interacted with.

- Let  $U$  denote the user population in an online social network
- $U = \{U_1, U_2, \dots, U_i, \dots, U_m\}$ , where  $m$  is the maximum distance from the users to the source  $s$ . The group  $U_x$  consists of users that are of the same distance  $x$  from the source.
- Two information diffusion processes:
  - Growth process: Users in  $U_x$  can influence each other
  - Social process: Users at different distances can influence each other, this is random walk

# Diffusive Logistic Model



- Growth process is modeled by logistic model which is widely used to model the population dynamics where the rate of reproduction is proportional to both existing population and the amount of available resources
- Diffusion process is modeled by Fick's law of diffusion, which is used to measure the diffusion of heat in a metal

- $I(x, t)$  denote the density of influenced users at distance  $x$  and time  $t$ .



$$\begin{aligned}\frac{\partial I}{\partial t} &= d \frac{\partial^2 I}{\partial x^2} + rI \left(1 - \frac{I}{K}\right) \\ I(x, 1) &= \phi(x), \quad 1 \leq x \leq L \\ \frac{\partial I}{\partial x}(1, t) &= \frac{\partial I}{\partial x}(L, t) = 0, \quad t > 1\end{aligned}\tag{0.2}$$

- $\phi(x)$  is the initial function constructed from the initial phase of spreading
- $\frac{\partial I}{\partial x}(1, t) = \frac{\partial I}{\partial x}(L, t) = 0$  means information spreading is within the OSN

## Diffusive Logistic Model

- Diffusive Logistic model has two properties:
  - Unique property
  - Increasing property
- growth rate  $r$  controls the gap between  $I(x, t)$  and  $I(x, t + 1)$  and is usually a function of  $t$
- diffusion rate  $d$  controls the slope of  $I$
- $K$  controls the upper bound of  $I$

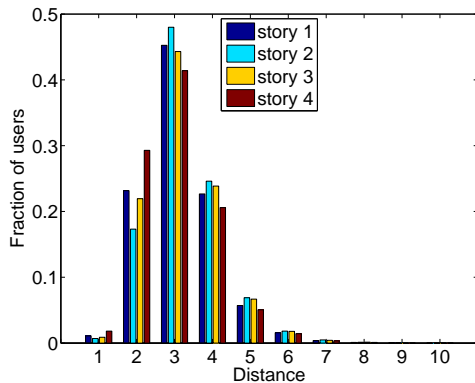
- Digg is one of the most popular news aggregation sites. Users can submit links of news stories that they find in professional news sites and blogs to Digg, and can vote and comment on the submitted news. Digg users form friendship links through following each other.



## Two ways of information propagation in Digg:

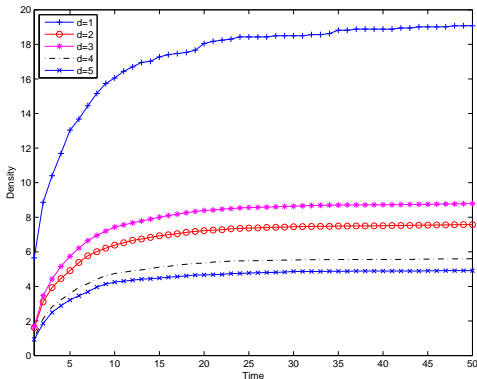
- A user can see the news submitted by the friends he follows and vote the news. After a user votes for a news, all his followers are able to see and vote on the news, and so on.
- Once the news is promoted to the front page due to high popularity, the users, who do not friend with the initiator directly or indirectly, will also be able to view and vote for the news. (This contributes to random walk)

- Consist of 3553 news stories that are voted (also called digged) and promoted to the front page of [www.digg.com](http://www.digg.com) due to vote popularity during June 2009
- More than 3 million votes from 139,409 Digg users
- We choose four representative stories of different scales
  - Story s1 is the most popular news with 24,099 votes
  - story s2 is the second most popular news with 8521 votes
  - story s3 is a news with 5988 votes
  - story s4 is a news with 1618 votes.



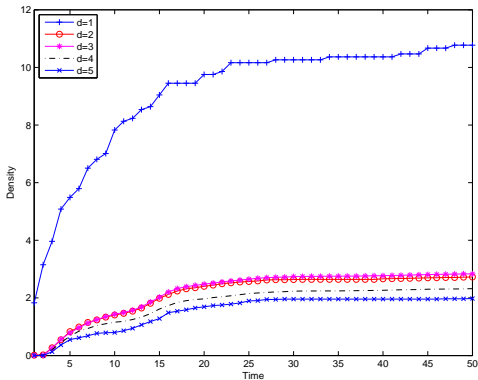
**Figure :** Distribution of neighbors of four stories

# Density Distribution



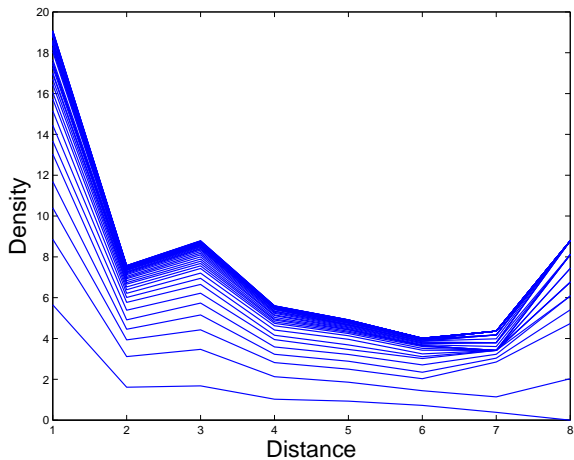
**Figure :** Density distribution of influenced users of story 1 with 24099 votes over 50 hours with friendship hop as distance

# Density Distribution



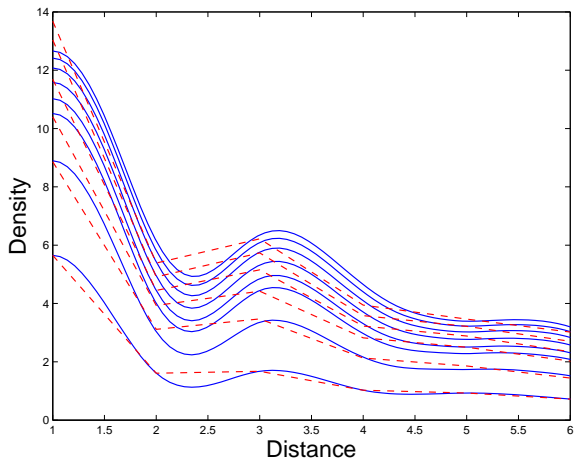
**Figure :** Density distribution of influenced users of story 2 with 8521 votes over 50 hours with friendship hop as distance

# Density Distribution



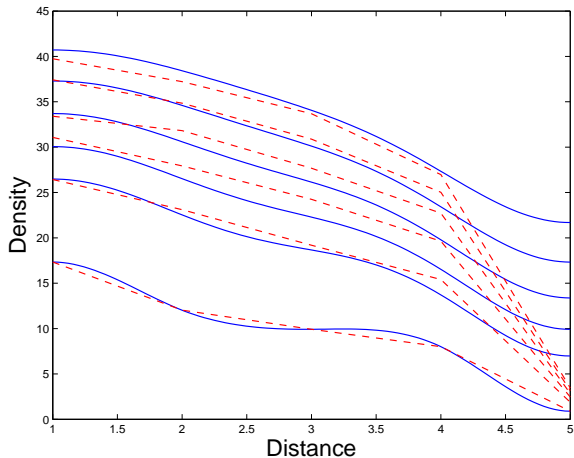
**Figure :** Density distribution of influenced users over 50 hours

## Model vs. Dataset



**Figure :** Model vs. Data of story 1 with 24099 votes

# Model vs. Dataset



**Figure :** Model vs. Data of story 1 with 24099 votes with interest as distance



**Table :** The prediction accuracy with shared interests as distances for story s1

Distance	Average	t = 2	t = 3	t = 4	t = 5	t = 6
1	97.21%	98.74%	96.75%	92.70%	97.91%	99.97%
2	93.67%	86.58%	93.99%	96.11%	96.14%	95.52%
3	93.11%	87.71%	92.86%	96.14%	95.39%	93.44%
4	91.64%	87.18%	91.38%	93.23%	93.63%	92.75%
5	39.84%	66.26%	44.43%	33.91%	28.68%	25.92%

**Table :** The prediction accuracy with friendship hop as distances for story s1

Distance	Average	t = 2	t = 3	t = 4	t = 5	t = 6
1	98.27%	97.47%	97.74%	97.48%	99.55%	99.09%
2	86.99%	93.59%	96.63%	87.16%	80.80%	76.78%
3	90.28%	83.23 %	87.98%	90.99%	93.35%	95.94%
4	92.98%	86.75%	91.39%	99.00%	95.68%	92.06%
5	93.77%	89.05%	91.61%	97.79%	97.92%	92.49%
6	94.56%	90.03%	89.48%	96.04%	97.57%	99.67%

- We introduce the spatio-temporal diffusion problem to understand information diffusion in online social networks
- We abstract the diffusion process and introduce diffusive logistic model to model information diffusion process in online social networks
- We present the temporal and spatial patterns of information diffusion in real dataset collected from a major social news aggregation site
- We validate the diffusive logistic model by matching its prediction with real dataset. The DL model shows high accuracy

- Systematically study the parameter selection, understand the impact of structure of network on the parameter selection
- Collect data from twitter, include profile information of each user, such as age, gender, etc
- Categorize news into classes and study the diffusion of different classes of news
- Categorize OSNs
  - social networks - facebook
  - social media - twitter, diggs
  - flickrs, youtube, blogs
- Develop new models for multiple sources, controversial news
- Visualize the diffusion