Spatiotemporal Spread of Information in Online Social Networks

Fields Undergraduate Summer Research Program
July & August, 2013

Jianhong Wu, York University

Centre for Disease Modelling (CDM)

Laboratory for Industrial and Applied Mathematics (LIAM)

CDM Centre for Disease Modelling | CMM Centre de la modélisation des maladies

YORK UNIVERSITÉ UNIVERSITY

FIELDS

The Ecology and Epidemiology of Online Social Networks

Main Questions:

- For a piece of information initiated from a particular user (source), what's the density of "influenced users" at a given distance (in the OSN) in a particular time? ——(Information propagation dynamics)
- For a collection of information initiated at a particular time in the OSN, how the information and internet users interact with each other dynamically?——(Population Dynamics of OSNs and Spread of Influence)

Major References:

- Wang F, Wang H, Xu K., Diffusive logistic model towards predicting information diffusion in online social networks, Distributed Computing Systems Workshops (ICDCSW), 2012 32nd International Conference on. IEEE, 2012: 133-139.
- Wang F, Wang H, Xu K, Wu J, and Xia, J., Characterizing information diffusion in online social networks with linear diffusive model, ICDCS 2013.

Digg Social News Aggregation

"The site's main function is to let users discover, share and recommend web content. Members of the community can submit a webpage for general consideration. Other members can vote that page up ("digg") or down ("bury"). Although voting takes place on digg.com, many websites add "digg" buttons to their pages, allowing users to vote as they browse the web. The end product is a series of wide-ranging, constantly updated lists of popular and trending content from around the Internet, aggregated by a social network."

http://en.wikipedia.org/wiki/Digg

S. Tang, N. Blenn, C. Doerr and P.V. Miegham, Digging in the Digg Social News Website, IEEE Trans. On Multimedia, 2011.

Spatiotemporal Stratification of Online Social Networks

Spatial Distance:

- Friendship Hops: a natural metric of distance between two users is the length of the shortest path, measured by the number of hops from one user to another in the social network graph.

- Shared Interests:

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

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.

Spatialtemporal distribution of the density of influenced users at DIGG

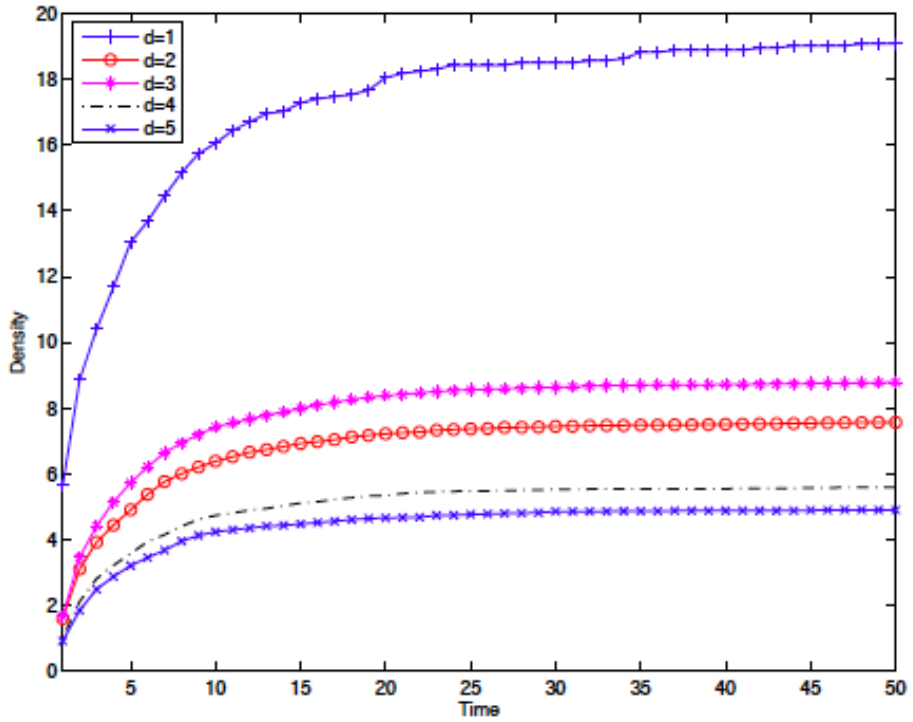


Figure : Density distribution of influenced users of story 1 with 24099 votes over 50 hours with friendship hop as distance

Diffusive Logistic Model

$$\frac{\partial I}{\partial t} = d\frac{\partial^2 I}{\partial x^2} + rI(1-\frac{I}{K})$$

$$I(x,1) = \phi(x), l \le x \le L$$

$$\frac{\partial I}{\partial x}(l,t) = \frac{\partial I}{\partial x}(L,t) = 0, t \ge 1$$

I: density of influenced users

d: diffusion rate

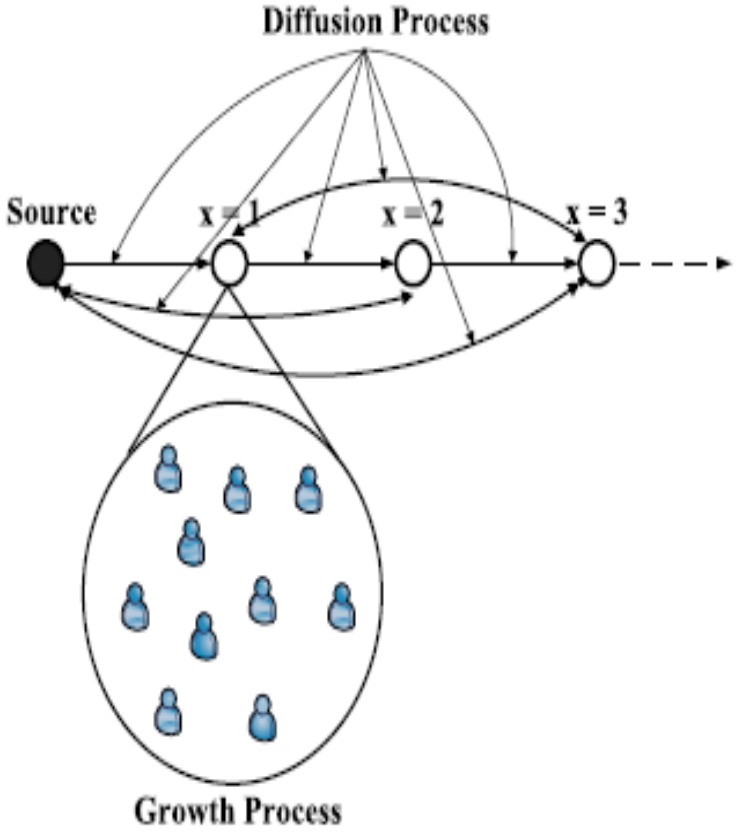
r: intrinsic growth rate

K: carrying capacity

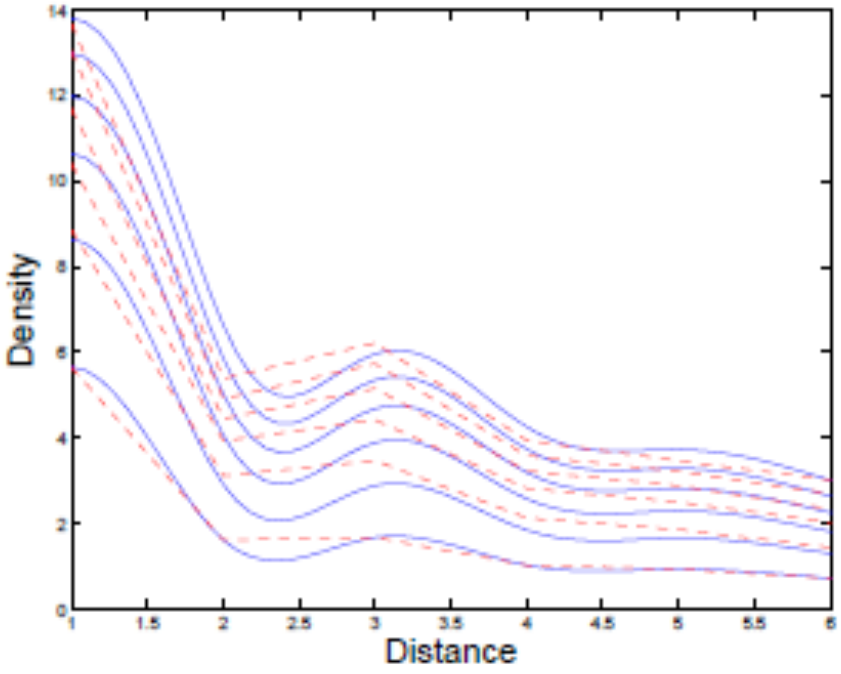
L: the upper bound of the distance

l: the lower bound of the distance

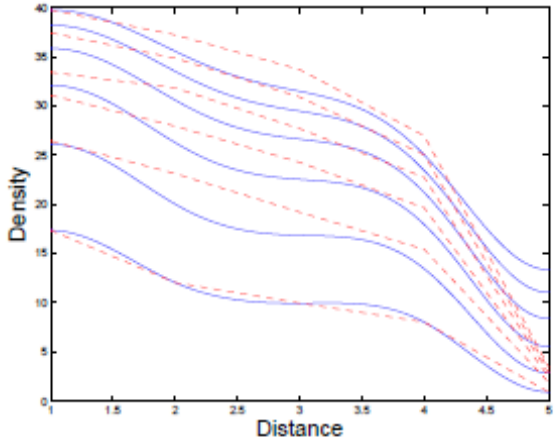
$\phi(x)$: the initial density function



Predicted VS Actual Data



(a) Predicted vs. Actual data of s1 with friendship hops as distance



(b) Predicted vs. Actual data of s1 with shared interests as distance

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%

$$Prediction \quad accuracy = \frac{\left| predicted \quad value - actual \quad value \right|}{actual \quad value}$$

Fading Interest

$$\frac{\partial I(x,t)}{\partial t} = d\frac{\partial^2 I(x,t)}{\partial x^2} + r(t)I(x,t)(1 - I(x,t)/K))$$

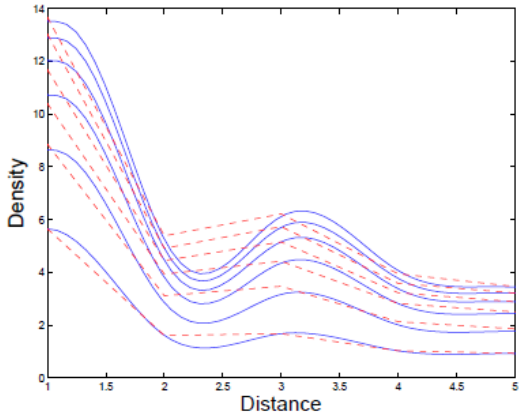
$$I(x,1) = \phi(x), l \le x \le L$$

$$\frac{\partial I}{\partial x}(l,t) = \frac{\partial I}{\partial x}(L,t) = 0, t \ge 1$$

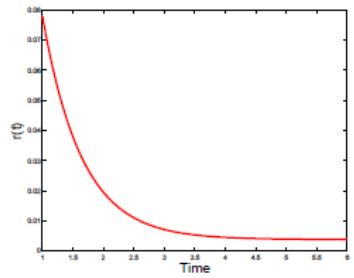
$$r(t) = 1.6e^{-(t-1)} + 0.1$$

$$\frac{dr(t)}{dt} = -\alpha r(t) + \beta$$

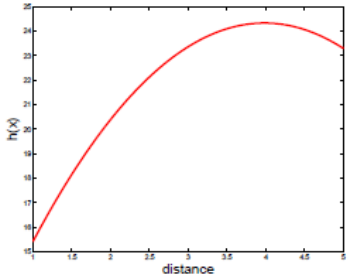
$$r(1) = \gamma$$



(a) Predicted (blue, solid) vs. Actual data (red, dotted)



(b) $r(t)$



(c) $h(x)$

Parameter	value
d	0.0020
α	1.5526
β	0.0059
γ	0.0780
ρ	-0.9478
σ	8.9149

(d) Parameter values

Distance	Average
1	97.88%
2	97.27%
3	97.44%
4	96.20%
5	98.25%
Overall	97.41%

(e) Model Accuracy

Free boundary value problem

- What is the transient and long-term dynamics?

$$\frac{\partial I(x,t)}{\partial t} = d\frac{\partial^2 I(x,t)}{\partial x^2} + r(t)I(x,t)(1 - I(x,t)/K)),$$

$$\frac{\partial I}{\partial x}(0,t) = 0, I(h(t),t) = 0, \frac{\partial h}{\partial t}(0) = -c\frac{\partial I}{\partial x}(h(t),t)$$

$$I(x,1) = \phi(x), l \le x \le L$$

$$\frac{\partial I}{\partial x}(l,t) = (L,t) = 0, t \ge 1$$

Digging the DIGG

Information propagation dynamics

- Friendship relationship (a Digg user can view the news submitted by the friends, he vote, comment and share the news, and his followers can then see and digg the news);
- A news can be promoted to the front page if it ranks as one of the top news stories based on the number of diggs it received. After the promotion, all users can view and vote the news;
- A user can discover and digg the news through search functions on DIGG.

2013 Summer Fields Study

- How submitting, discovering, sharing, recommending, and voting specifically contribute to the growth and diffusion processes?
- With respect to which distance metric? How to extend this from DIGG to other OSNs?
- How to model the non-local diffusion (discovering, voting), how time delay impacts the information spread?
- How to model the competitive nature in OSNs?

Summer 2013—Fields Project

Project II—OSN Epidemiology

Susceptible Users: $S(x,t)$

Influenced Users: $I(x,t)$

$$\frac{\partial S(x,t)}{\partial t} = -S(x,t)\int_{\Omega}\beta(y,x)I(y,t)dy$$

$$\frac{\partial I(x,t)}{\partial t} = S(x,t)\int_{\Omega}\beta(y,x)I(y,t)dy + rI(x,t)(1-\frac{I(x,t)}{K})$$

Which β can lead to an outbreak of a particular information in an online social network?

2013 Summer Fields Project (II)

Project II—OSN Ecology, Popularity Contest

Multiple News Interacted and Competing for Popularity:

$$\frac{\partial I_i(x,t)}{\partial t} = d_i \frac{\partial^2 I_i(x,t)}{\partial x^2} + r_i I_i(x,t)(1 - \frac{I_i(x,t)}{K_i})$$

What if they are coupled and spread simultaneously?

What if competition?

What if cooperation?

What if predation?