INDIAN INSTITUTE OF TECHNOLOGY MADRAS

BT3011

CFA EXERCISE

Diffusion of information through online social media

By:
Onkar Yashavant Ghagare (BE14B020)

4 November, 2016

Contents

1 Introduction

“ *It's because of this fundamental shift towards user-generated digital information that people will listen more to other people than to traditional resources.*”

- Eric Schmidt, ex-CEO Google

In today's digital era, online social networks such as *Facebook, WhatsApp* or *Twitter* have gained tremendous popularity because of information exchange. These Social networks can link people from any corners of the world regardless of their different cultures or geographical boundaries. Hence, the availability of huge amounts of digital data has accelerated research on flow of information through this digital world. Any new information, after it is introduced in social media, takes some time to spread in the digital encyclopaedia. Any person when performs any kind of online activity leaves some digital footprints, which can influence the other users using internet. It is readily evident that whenever a person uploads a video on *YouTube* or tweets on *Twitter*, the number of social media users influenced by that uploaded digital information increases rapidly with time.

Using some techniques and laws used in mass conservation and fluid flow (such as Fick 's laws of diffusivity), the diffusion of information can be modelled to predict the influence of any social activity on the users. In this CFA exercise I have used some of the concepts used in fluid motion to create a model for diffusion of digital information through social media.

2 Background Knowledge

2.1 Law of conservation of mass

Law of conservation of mass states that mass of a closed system is always conserved or in other words- mass can neither be created nor destroyed when we consider non-nuclear processes[1].
For a system,
the mass conservation law in mathematical form becomes:

Input - Output + Generation - Consumption = Accumulation

OR

$$r_i - r_o + r_g - r_c = \frac{dm}{dt} \tag{2.1}$$

2.2 What is Diffusion?

Diffusion is the movement of particles from a higher concentration to lower concentration. Particles move down a concentration gradient to maintain the concentration balance. It is different from other transport processes in that it results in mixing without bulk matter flow. Every diffusion process happens because of a driving force which causes the molecules to move along the concentration gradient[1].

2.3 Fick's I law

Movement of particles in diffusion can be best modelled using Fick 's I law of diffusion. It states that the diffusion flux is directly proportional to the concentration gradient.

$$J \propto \frac{\partial C}{\partial x} \tag{2.2}$$

OR

$$J = -D\frac{\partial C}{\partial x} \tag{2.3}$$

Here, J is the diffusion flux (the amount of substance flowing through a unit area per unit time), D is the diffusion coefficient, C is the concentration of the diffusing molecules and x is the position variable.

2.4 Fick's II law

This law describes a relation between concentration gradient and time during a diffusion process.

$$\frac{\partial C}{\partial t} = D\frac{\partial^2 C}{\partial x^2} \tag{2.4}$$

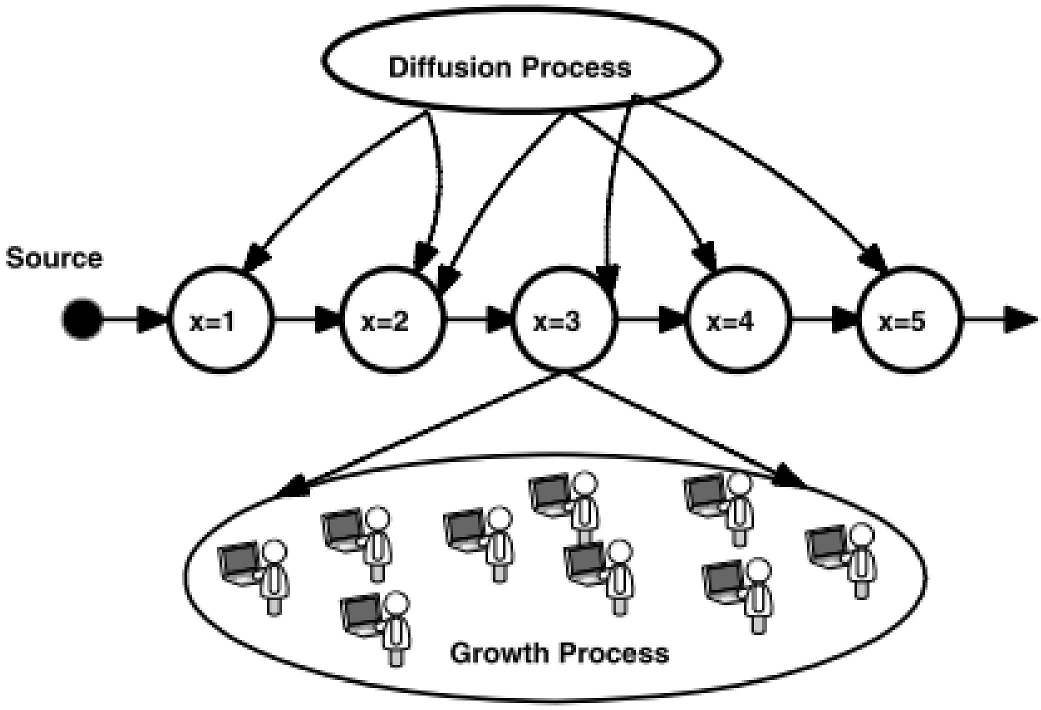


Figure 3.1: Conceptual illustration of the information spreading in online social networks[2]

3 Diffusion of information through social media

3.1 Introduction

Whenever any information is published on social media(for example a tweet made by an user on *Twitter*),the information starts propagating from the source towards a digital region where the users are ignorant about that information. This movement of information is analogous to the movement of solute molecules through diffusion. In molecular diffusion, concentration gradient is the driving force. In this case the information gradient formed is causing the information to propagate along the gradient.A pictorial representation of information diffusion through online networks is shown in figure 3.1 .

3.2 Basic Terminology

Following are the terms used in the analysis:

3.2.1 Digital distance (x)

It is the measure of closeness of users in an online social network. To capture this aspect of how individuals are embedded in online social networks, one main approach is to examine the distance that an user is from others[3].For example: In case of *Facebook*, a primary approach for defining the digital distance between two users is to find the friendship links between the users.So, the friends of an user on *Facebook* will have a distance of 1, while the friends of those friends will have a distance of 2. Same logic goes with the followers on *Twitter*.

3.2.2 Information Density ($I(x,t)$)

The density of users influenced by the published information at digital distance x and time t is denoted by $I(x,t)$. This density is the ratio of the number of users influenced with the information at digital distance x and time t over the total number of users at the same digital distance. This quantity is analogous to the concentration term in molecular diffusion.It is also known as *influence function* [4].

3.2.3 Diffusivity (D)

Similar to molecular diffusion, Diffusivity or diffusion coefficient in digital information flow is directly related to the capability of the information to be diffused by diffusion.

In online social media networking the diffusion coefficient depends on many factors such as i)content of the information, ii)popularity of the content in user community. iii) Popularity of the source

For example:Suppose, a meme related to cricket is published on two different *facebook* pages.One of which has majority of its followers from India and the other page is mostly followed by users in the USA. Then obviously the meme will diffuse relatively faster through the page with more number of indian followers.This implies that coefficient of diffusion depends on the popularity of information in user community.Also,the rate of diffusion will increase if the followers of the page increase (i.e increase in source popularity).

3.3 Conservation of information

All the users of an online social network can be embedded into a one-dimensional space as shown in Figure 3.2. The quantity used is the amount of informa-

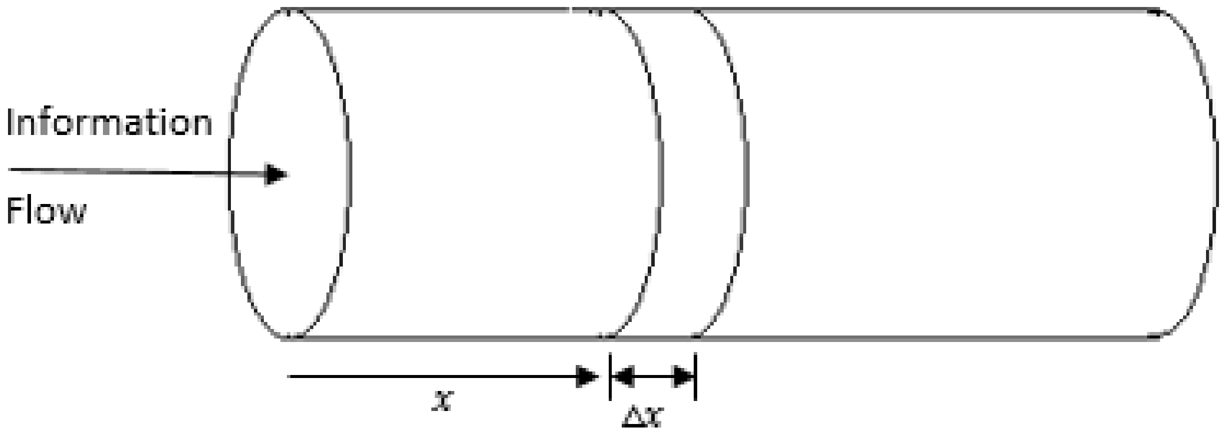


Figure 3.2: Shell balance for information flow

tion spreading i.e the density of influenced users, which is denoted by $I = I(x,t)$ and measured in amount per unit length along the x-axis of the Figure 3.2.The information flow is conserved through the small Δx part of the cylinder. So,similar to shell mass balance in molecular flow, we can perform information conservation for the small Δx region.

Input - Output + Generation - Consumption = Accumulation

Here, Input = $G_{|x}A$ (G is the information flow flux and A is the cross sectional area of the cylinder), Output = $G_{|x+\Delta x}A$, Accumulation = $\frac{\partial I(x,t)}{\partial t} A\Delta x$, In information flow there is no consumption.Hence, Consumption = 0.Generation of information occurs only at the origin i.e. only at the the source hence while the information is propagating through the shell, there will not be any generation of information. Therefore, Generation = 0

Now, Substituting these values in above equation we get:

$$G_{|x}A - G_{|x+\Delta x}A = \frac{\partial I(x,t)}{\partial t} A\Delta x$$

$$\frac{G_{|x} - G_{|x+\Delta x}}{\Delta x} = \frac{\partial I(x,t)}{\partial t}$$

Now, apply $\lim_{\Delta x \to 0}$

$$\lim_{\Delta x \to 0} \frac{G_{|x} - G_{|x+\Delta x}}{\Delta x} = \frac{\partial I(x,t)}{\partial t}$$

to get:

$$-\frac{\partial G}{\partial x} = \frac{\partial I(x,t)}{\partial t} \tag{3.1}$$

3.4 Molecular flux and information flux

In biology or physics molecules move from high concentration to low concentration in other words according to Fick's I law the molecular flux J can be: $J = -D\frac{\partial C}{\partial x}$, An exact same expression can be used to relate information flux with information density gradient as shown below:

$$G = -D\frac{\partial I(x,t)}{\partial x} \tag{3.2}$$

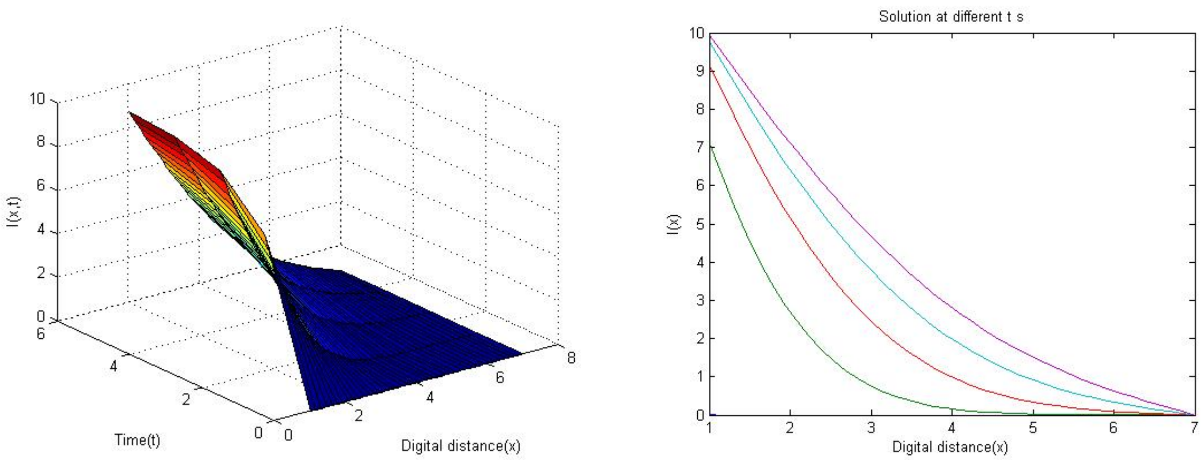
One can substitute equation(3.2) in equation(3.1) to get(assuming D is constant):

$$\frac{\partial I(x,t)}{\partial t} = D\frac{\partial^2 I(x,t)}{\partial x^2} \tag{3.3}$$

Here, D is the diffusivity which is a property of the media and the information content.
Equation(3.3) can be solved with following boundary and initial conditions:
Initial condition: For $t = 0, I(x) = 0$
Boundary Conditions for $t > 0$:

$$I(1,t) = L(1 - e^t)$$
$$I(7,t) = 0$$

For the users with a digital distance of 1 the information flow will be very rapid,So at $x = 1$,I have assumed the density function to be: $I(1,t) = L(1 - e^{-t})$, where L is the number of direct followers of the source located at $x = 0$(Reason behind choosing this function is it increases very rapidly and converges to L after a long time).And for the second boundary condition I have assumed that after a digital distance of 6 there will not be any information flow(this assumption is fairly valid since for example: on *facebook*, information flow till the source's friend of friend of friend will be very close to zero),hence $I(7,t) = 0$.
Equation (3.3) with given initial and boundary conditions was solved in *Matlab*, plots of which can be seen in Figure3.3



(a) Information density variation with time and digital distance x.

(b) Plot of information density-$I(x)$ vs digital distance- x for different times.

Figure 3.3: Matlab plot of the Equation(3. 3)with $D = 1$ and $L = 10$.

4 Diffusivity as a function of digital distance

In all the analysis that has been done above, diffusivity is assumed to be a constant. But in reality that is not the case.For example if I upload a photo on *facebook*, then the users who are at distance 1 from me are more likely to get influenced by the content of the photo than the users at distance 2,because my friends of friends will be less interested to know about me than my friends.
In other words the information flow which depends on diffusivity should decrease rapidly with increasing digital distance.Therefore a rapidly decreasing function can be used for diffusivity. Therefore, I have chosen an exponentially decreasing function below:

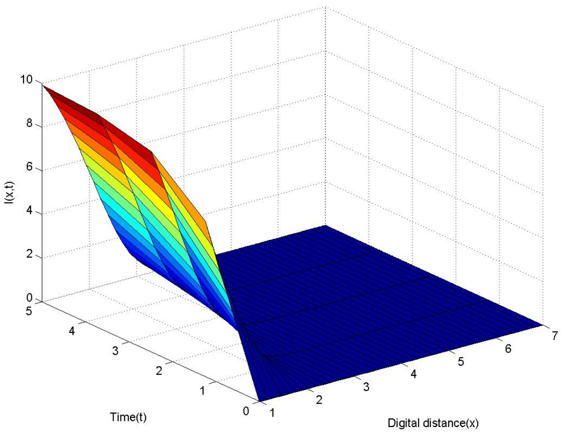
$$D = ke^{-x} \tag{4.1}$$

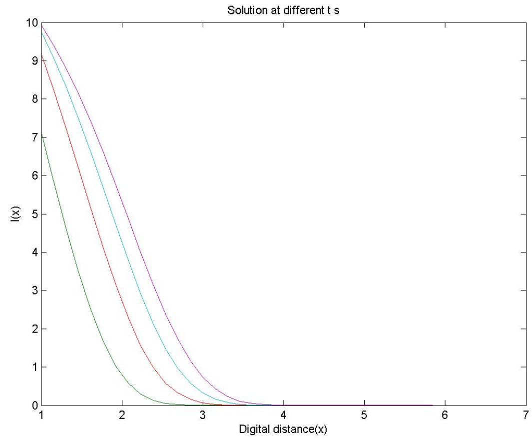
Here, k will depend on popularity of the content.
Equation (4.1) can be used to simplify equation (3.1):

$$\frac{\partial I(x,t)}{\partial t} = \frac{\partial (D\frac{\partial I(x,t)}{\partial x})}{\partial x}$$

$$\frac{\partial I(x,t)}{\partial t} = \frac{\partial I(x,t)}{\partial x}\frac{\partial (ke^{-x})}{\partial x} + ke^{-x}\frac{\partial^2 I(x,t)}{\partial x^2}$$

$$\frac{\partial I(x,t)}{\partial t} = ke^{-x}(\frac{\partial^2 I(x,t)}{\partial x^2} - \frac{\partial I(x,t)}{\partial x}) \tag{4.2}$$





(a) Information density variation with time and digital distance x.

(b) Plot of information density- $I(x)$ vs digital distance- x for different times.

Figure 4.1: Matlab plot of the Equation(3. 5)when D is function of x and $k = 1$ and $L = 10$.

Initial and boundary conditions for this equation are same as for equation 3.3:
Initial condition: For $t = 0, I(x) = 0$
Boundary Conditions for $t > 0$:

$$I(1,t) = L(1-e^t)$$
$$I(7,t) = 0$$

Solution of equation 4.2 is shown in figure 4.1. If we compare Figure 4.1 with figure 3.3, it is evident that the information flow is decreasing with digital distance (x) as a result of decreasing diffusivity .

5 Further Work

The models proposed above are not solved analytically,they have to be tested with real time data from social networking sites. Depending on the accuracy of the models some of the functions used can be replaced to get better accuracy. For example: one can represent diffusivity as: $D = \dfrac{k}{1+x}$, which is also a decreasing function of digital distance(x) similar to ke^{-x}. Also the parameters used in modelling such as k or D (for constant diffusivity) can be determined by fitting the models with real time data.

Acknowledgement

I would like to thank Prof. G K Suraishkumar for giving me the opportunity to perform this exercise as part of his Transport Phenomena in Biological Systems course. I would also like to thank my classmates Saurabh and Debayan for their valuable suggestions.

References

[1] *Suraishkumar GK. 2014. Continuum Analysis of Biological Systems: Conserved Quantities, Forces and Fluxes.*

[2] *F. Wang, H. Wang, K. Xu, Diffusive logistic model towards predicting information diffusion in online social networks, in: 32nd International Conference on Distributed Computing Systems Workshops (ICDCSW), 2012, pp. 133âĂŞ139, http:// dx.doi.org/10.1109/ICDCSW.2012.16*

[3] *http://faculty.ucr.edu/*

[4] *Jaewon Yang, Jure Leskovec, Modeling Information Diffusion in Implicit Networks*