TOPIC DIFFUSION MODEL ON SOCIAL NETWORKS

Soham Das
2010MCS3480

Guided by: Prof. Amitabha Bagchi

May 9, 2011

Contents

List of Figures

1 INTRODUCTION

We consider a social network graph having a node set V, where each node represents a user in the network and an edge represents the neighborhood relationship (similar to friendships in social networks). We denote the set of neighbors of a node $u \in V$ by $\tau(u)$. At any time instant t every node $u \in V$ has a current topic that he is talking on (or has not replaced).
We define a few notations here:

- S_t: This set contains all the topics that are living in the network at time t i.e. one or more nodes are talking on them.
- R_t: This set contains the set of topics introduced at time t.
- $N_t(u)$: This set contains the topics which are talked on by the neighbors of the node u at time t.
- $C_t(x)$: This set contains all the nodes that are talking on topic x at time t.
- $L(x)$: This is the lifetime of a topic x in time units.
- $M(x)$: This denotes the maximum spread for a topic x i.e. the maximum number of nodes, that have talked on it at the same time.
- $TS(x)$: This denotes the total spread for a topic x i.e. the total number of nodes, that have talked on it, over $L(x)$.
- $A(x)$: This denotes the number of adopters for topic x.

Now at every time instant t, a node u in the network has 3 actions to choose from, each with some associated probabilities:-

1. **Idle:** Do nothing.
2. **Adoption:** Pick up a topic $x \in R_{t-1}$ at random and talk about it.
3. **Local Propagation/Copying:** Pick up a topic $y \in N_t(u)$ and talk about it.

2 THE GRAPH MODELS

We have analysed two types of synthetic graphs here.

2.1 Ring Network

We have considered a simple ring network with each node connected to a neighbor on each side for the initial analysis, as shown in fig1.
Here we have considered that each node can talk only on a single topic at a time.

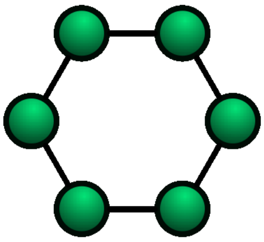

Figure 1: A Ring Network

2.2 Watts & Strogatz Model

We consider a ring network again where each node is initially connected to 5 neighbors on each side. So degree of each node is 10 initially. Now, we consider each edge $(u, v) \in E$ and re-write it to (u, v'), where v' is selected randomly, with a certain probability of rewriting. Let us look at the figure below. We have considered a 4-regular ring in which the link $(1, 6)$ is rewritten to $(1, 5)$, where node 5 is chosen randomly.

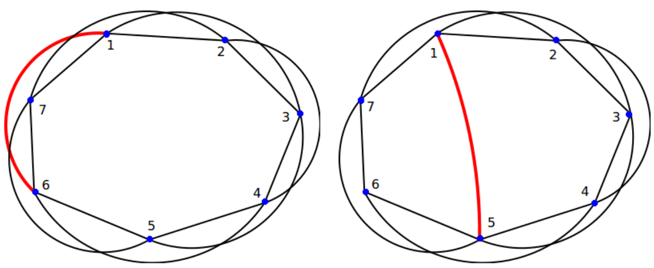

Figure 2: Rewriting $(1, 6)$ to $(1, 5)$

By such a construction, we somehow try to obtain the small-world properties of a typical social network. The degree distribution of the graph after re-writing the links follows a Gaussian Distribution.

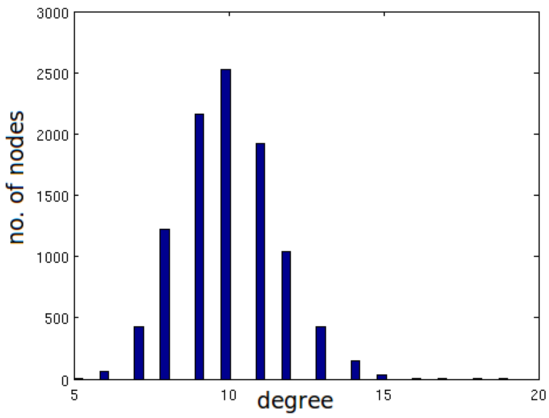

Figure 3: Degree Distribution of Watts & Strogatz Graph

Here we have considered that each node has a finite queue which can store topics. This means that at a time, a node can talk on multiple topics. A topic can get dequeued only when the queue is full and the node adopts or copies a topic from it's neighbor. A queue can store multiple copies of the same topic. This can happen if a node copies the same topic from it's neighbors in two or more time instances or it adopts a topic and copies it later from one of it's neighbors.

2.3 Associated Probabilities

We have the following probabilities associated with these graph models.

- p_1 =probability of doing something.
- $p_1 = p_2 + p_3$, where p_2 =probability of adoption & p_3 =probability of copying.
- p_4 =probability of doing nothing. $p_4 = 1 - p_1$.
- p_5 =probability of rewriting a link.

3 TOPIC DIFFUSION ANALYSIS

To model topic diffusion in social networks, we have looked at the following aspects of topic diffusion in these networks. We have varied p_2 and studied how these graphs change.

1. Variation of $|S_t|$ with t.
2. Variation of $|C_t(k)|$ with t for a topic k.
3. Histograms of $L(x)$.
4. Histograms of $M(x)$.
5. Scatter diagrams of $TS(x)$ and $L(x)$.
6. Scatter diagrams of $TS(x)$ and $A(x)$.
7. Scatter diagrams of $L(x)$ and $M(x)$.

The correlation among these parameters have revealed some important properties of topic diffusion. These give an idea on

- how topics diffuse in social networks over time and space,
- how far(to how many nodes) the topics spread,
- at what rate they spread,
- how the spread is affected by the adoption probability,
- how many nodes at max. talk on a topic at a particular time t.

We have simulated these and we shall discuss the results in the next section.

4 SIMULATION RESULTS

For all the simuations that we have performed, we have taken

- 10000 nodes forming the network
- 10 new topics are introduced at every time instance.
- $p_1 = 0.8$ for all simulations.
- p_2 are typically 0.05, 0.15, 0.3 and 0.45 respectively.
- $p_5 = 0.3$ in the WS Model.
- Queue size= 10.

4.1 Ring Network

In these simuations we have assumed that each node can only talk on a single topic at a particular time.

4.1.1 Variation of Number of Topics with time

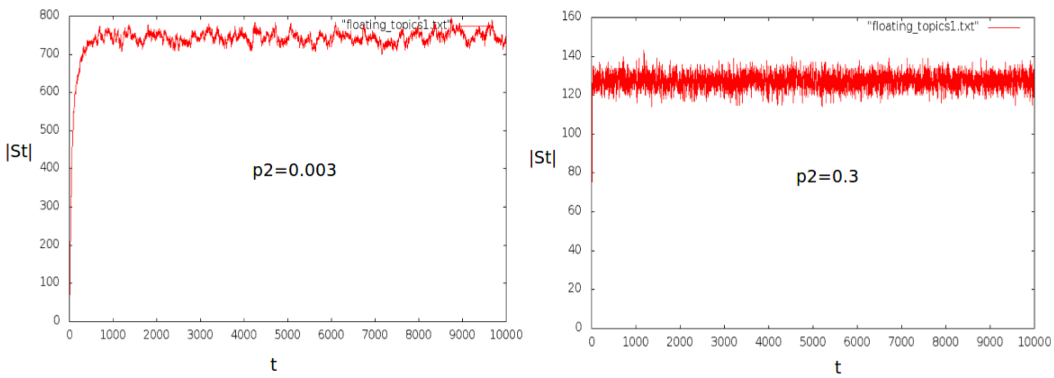

Figure 4: Variation of $|S_t|$ with t for a Ring

The graphs in fig.4 show how $|S_t|$ in a ring network varies with time. It starts from zero and then rises as more and more topics are adopted by the nodes. Important thing to note here is that $|S_t|$ is only affected when some new topic is adopted or any topic is removed from the network. Copying does not affect $|S_t|$.
From these two graphs we find that when p_2 is low(0.003), the graph rises

gradually. Since adoption rate is low, rate of increase of $|S_t|$ is also low.
We also observe that the graphs reach a saturation after a certain point of time. So, from this time onwards, the number of new topics entering and the number of topics leaving the network don't seem to differ by a huge margin, thus not disturbing the graph that much.
We also find that this stability is achieved at $|S_t| = 750$(approx.) when $p_2 = 0.003$ and $|S_t| = 130$(approx.) when $p_2 = 0.3$. Now as the adoption rate increases more and more topics are removed from the network as well leading to a saturation at a much lower value.

4.1.2 Variation of Topic Lifetimes with time

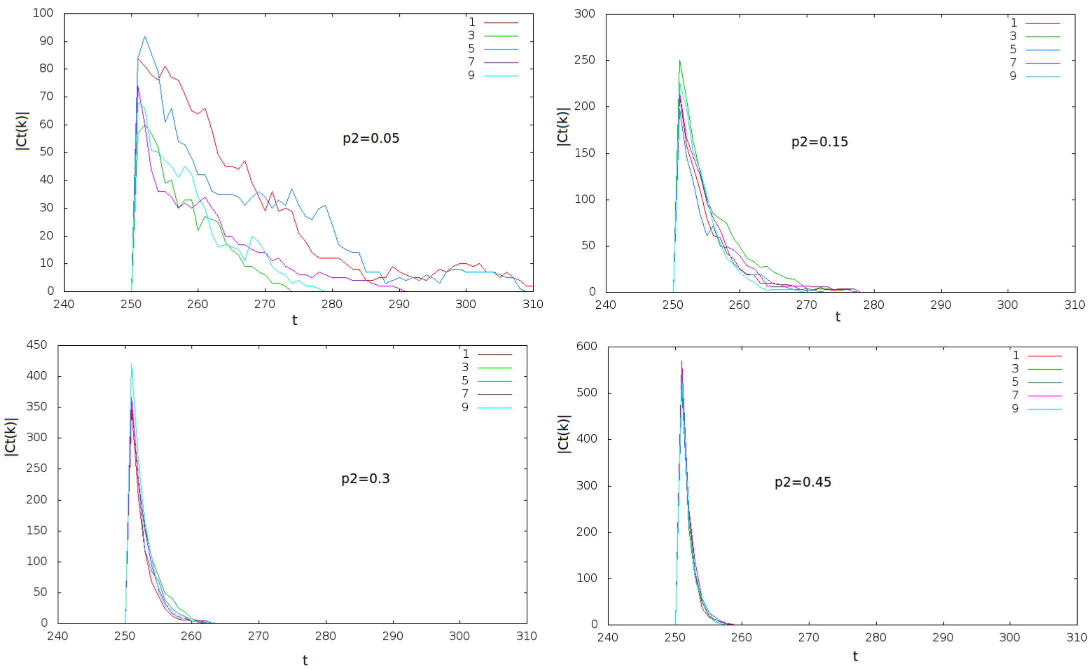

Figure 5: Variation of $|C_t(k)|$ with t for a topic k for a Ring

In these graphs (fig.5) we have considered 5 topics: $1^{st}, 3^{rd}, 5^{th}, 7^{th} and 9^{th}$ topic introduced at $t = 250$. So these graphs depict the lifetimes of these 5 topics and how many nodes have talked on them at a particular time. All of them show a vertical rise at $t = 250$ when the topics are adopted, then they fall down nearly exponentially. This is because, whenever a node, currently talking on one of these 5 topics, adopts or copies another topic, the number of nodes talking on this topic reduces by 1.
We also find that there are some irregular undulations or peaks in these

curves. This is because when the curve falls down exponentially, there can be time instances, when a considerable number of nodes copy this topic from their neighbors. Then the curve rises and form these small peaks.
The lifetime of topics also reduces as p_2 increases. When a node has high adoption rate, the chance of an older topic getting removed from the network at each time instance also increases. This reduces the lifetime of topics.

4.2 Watts & Strogatz Model

Here we assume that each node can talk on more than one topic at the same time. Each node possesses a topic queue, which can store a finite number of topics. For all the simulations to follow, we have taken the queue size= 10. So if the node is talking on 10 topics, it's queue is full. Then if it copies or adopts another, the last topic gets dequeued.

4.2.1 Variation of Topic Lifetimes with time

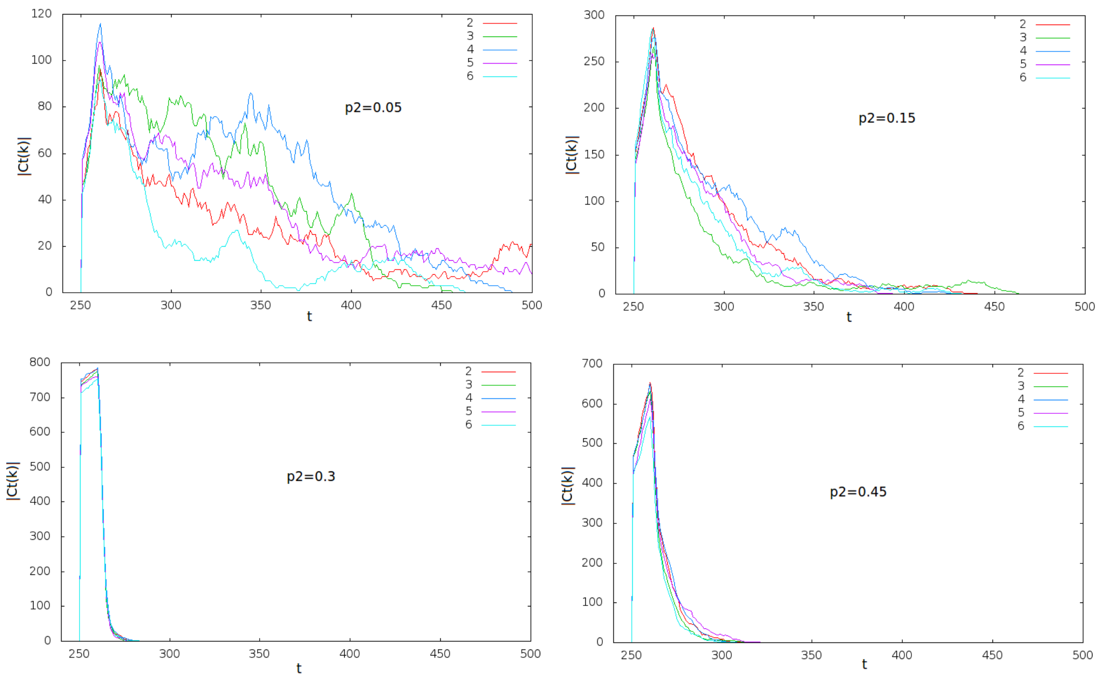

Figure 6: Variation of $|C_t(k)|$ with t for a topic k for a Ring for WS Model

The graphs in fig.6 show how $|S_t|$ varies with time. If we compare these graphs with those of fig.5, we find that the lifetime of topics has significantly increased over here. Another important thing to notice is that in these graphs(fig.6), after the vertical rise at $t = 250$, there is another almost linear rise from $t = 250$ to $t = 260$ approximately. Note that once the topic is enqueued it can only be dequeued after 10 time units since the size of the queue is 10. So the curve can never go down in that period. This rise is due to copying of these topics by the neighbors. Then again there is a nearly exponential decrease like the curves for the ring networks.

4.2.2 Histograms of Topic Lifetimes

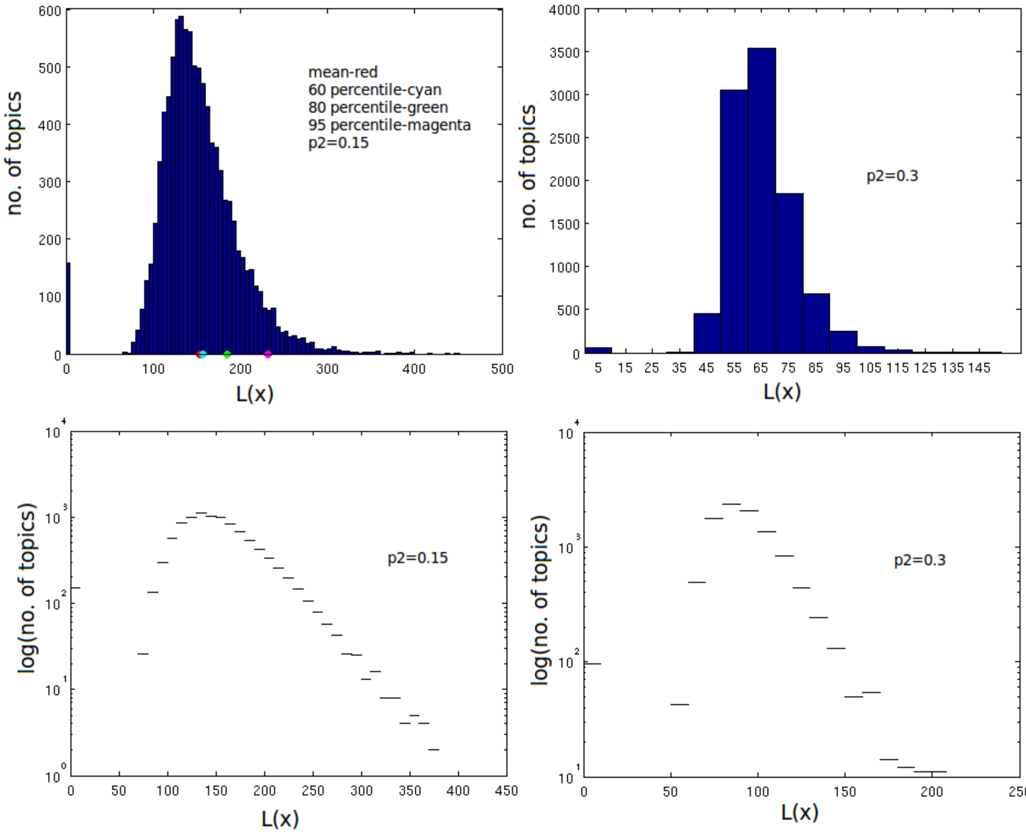

Figure 7: Histograms of $L(x)$ for WS Model

The graphs in fig.7 show the histograms of $L(x)$ of the 5_{th} topic introduced at each time t. These vertical bars denote the number of topics with the same lifetime. So we find that the tallest bars are present almost near the mean of the lifetime values. We calculated the percentiles as well. These histograms rise and fall almost exponentially.
The next two graphs are plotted by converting the y axis into logarithmic scale. We find the curves almost linear except at the peaks.

4.2.3 Scatter diagrams of Total_ Spread and Lifetimes of topics

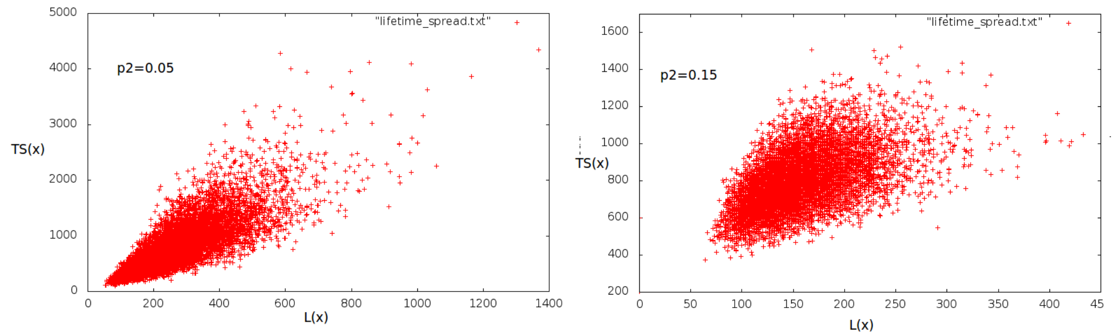

Figure 8: Scatter diagrams of $TS(x)$ and $L(x)$ for WS Model

These graphs show scatter diagrams of $TS(x)$ vs $L(x)$ for topics in the network. From these graphs we see that the $TS(x)$ somewhat increases with $L(x)$. More time, the topic survives, it spreads to more and more nodes.

4.2.4 Scatter diagrams of Total_ Spread and No. of Adopters

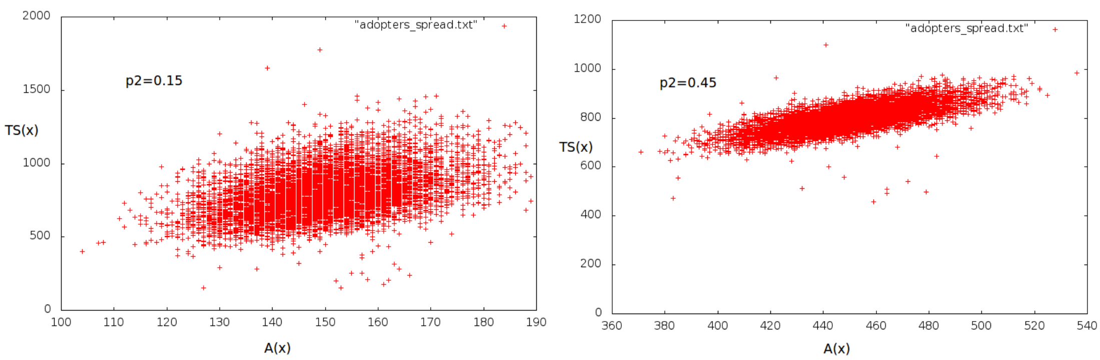

Figure 9: Scatter diagrams of $TS(x)$ and $A(x)$ for WS Model

These graphs show scatter diagrams of $TS(x)$ and $A(x)$ for topics in the network. From these graphs we find that the $TS(x)$ for topics increases with $A(x)$. If more nodes adopt a topic, it spreads to more and more nodes during it's lifetime. More nodes adopting the topic leads to more and more nodes copying it and thus it's spread increases.

4.3 Copy Priority

Now in all the previous simulations(Watts & Strogatz Model), we have assumed that while copying, a node can choose any topic from it's neighbor's queue randomly. This may not happen in a real social network. The probability of copying a recent topic is definitely more than that of an older topic. So in the following simulations, we take this into account and make the probability of copying the i_{th} recent topic from a queue is $\frac{1}{2^i}$. So more recent the topic, more it's chance to get copied.
Under this new scenario we performed the following simulations.

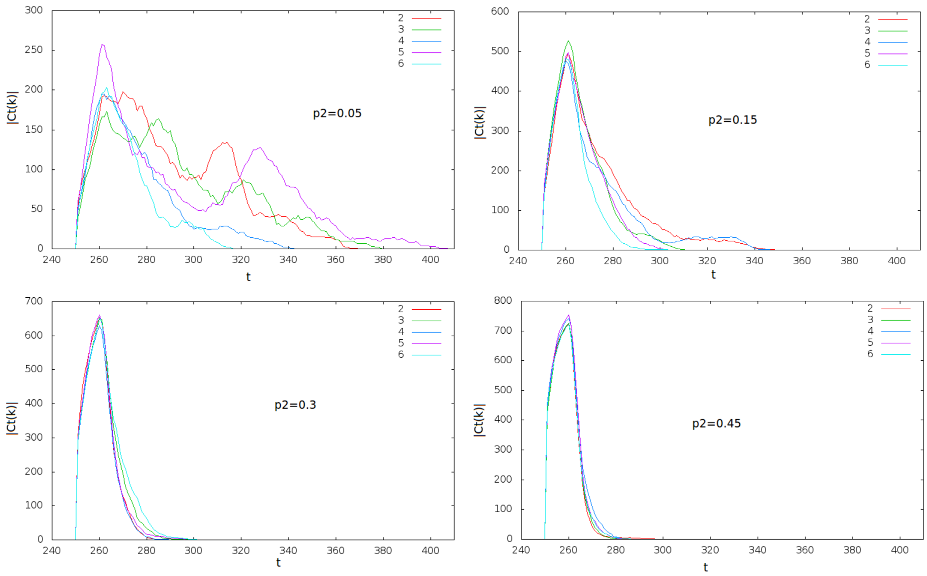

Figure 10: Variation of $|C_t(k)|$ with t for a topic k with Priority-Copy

4.3.1 Variation of Topic Lifetimes with time

If we compare these graphs in fig.10 with those of fig.6, we can find that the undulations in the curves are reduced to a large extent. The curves are much smoother. These undulations mainly result from a topic being copied long after it was introduced in the network. Here we are making that probability quite low. So the chance of the number of nodes copying an older topic is very low. Thus once the curve starts falling down, there are very few instances when it rises up again forming small peaks unlike the previous scenario.
Next thing to notice is that the lifetime of topics goes down. This is obvious since we a restricting a topic from being copied long after it was introduced.

So with time it get removed from the network faster.
Thus in all of these graphs we find the liftetime of topics reducing.

4.3.2 Histograms of Topic Lifetimes

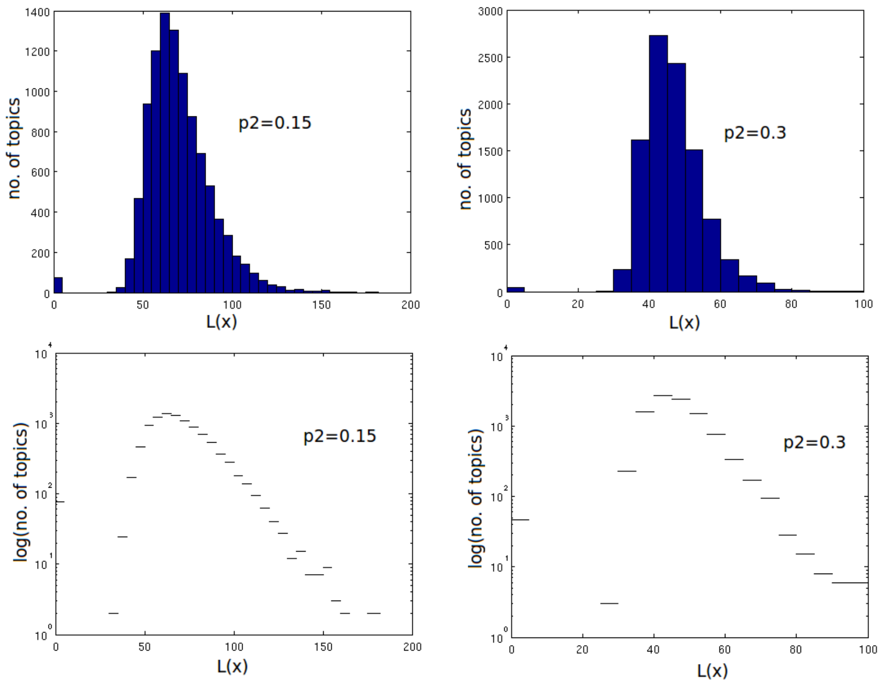

Figure 11: Histograms of $L(x)$. with Priority-Copy

We plotted the histograms of $L(x)$ fig.11. as well and in all of them we found the lifetime of topics low compared to their previous versions(fig.7) as expected.

5 FUTURE WORK

We plan to study some more properties of topic diffusion in future. We can make things a bit more realistic by introducing

- Global Copy: A list of global topics is to be maintained for this. It contains the popular topics, where popularity of a topic is in terms of the number nodes speaking on it at time t. We keep a threshold on the number of nodes to distinguish the popular topics, or we can simply select the top k topics. Now a node can copy from a neighbor as well as from the golbal list of topics.
- Node Biasing: Each node has a reputation quotient. More the reputation, more chance of other nodes copying from it. We can also have a trust quotient between pairs of nodes.
- Topic Biasing: Here topics can have some biasing, that can determine a topic to be adopted more compared to others.

References

[1] Ankit, Rudra M. Tripathy, Sameep Mehta, Amitabha Bagchi, Amit, Diffusion Model on Blog Networks, September 26, 2010.

[2] Rudra M. Tripathy, Amitabha Bagchi and Sameep Mehta, A study of rumor control strategies on social networks, In Proceedings of the 19th ACM Intl. Conference on Information and Knowledge Management (CIKM '10), pp 1817-1820, October 2010.

[3] D. J. Watts and S. H. Strogatz. Collective dynamics of small-world networks. Nature, 393(6684):440442, June 1998.

[4] David Kempe, Jon Kleinberg, Eva Tardos, Maximizing the Spread of Influence through a Social Network, KDD 2003, Proceedings of the ninth Intl. Conference on Knowledge Discovery and Data mining.

[5] Forrest Stonedahl, William Rand, Uri Wilensky, Evolving viral marketing strategies, GECCO '10 Proceedings of the 12th Annual Conference on GenetiC and Evolutionary Computation.