An Analysis of Cascade Diffusion through Complex Contagion in Multiplex Networks

Pranali Yawalkar
CS11B046

Aayush Agarwal
CS11B002

Mahak Goindani
CS11B041

***Abstract*—In our work presented here, we explore two very specific kinds of triggers for cascading mechanisms in multi-layered graphs. Any individual in a social network influences and is influenced by his neighbours in the network. Typically, there is more than one source of such influence and different sources have different degrees of impact. This we capture by using a four-layered graph, each layer of which captures a specific source of influence. Then we go on to examine the impact of introducing hyperactive centres in the graph. We observe that tiny and localised centres of hyperactivity have the potential to trigger cascades and initiate a global phenomenon. Finally, we examine how semiactive centres in the graph, though seemingly insignificant, too have the potential to trigger cascades, proving that small changes can sometimes make a big difference.**

Index Terms—**Multiplex graph, cascade, complex contagion.**

I. Introduction

Social influence and its potential to yield a critical mass of supporters can make a crucial difference as to whether or not social movements succeed. Mechanisms rooted in social interaction can give rise to financial crashes, political revolutions, successful technologies, and cultural market sensations. Our work on this project has been inspired by cascading behavior in multiplex graphs in particular.

An individual in a network experiences peer pressure from his neighbours and there is often more than one such kind of pressure. We use multiplex graphs to account for the multiple sources of influence and assign different relative importance to each. More specifically, the nodes in the Higgs-Twitter graph share four kinds of relationships based on whether or not they retweeted, replied to, mentioned or followed one another.

A cascade is said to have occurred when initially local behavior becomes widespread through collective action. By analogy with biological epidemics, the term contagion is used to model the exposure to a source of information that initiates the propagation. Complex contagion refers to the phenomenon in social networks in which multiple sources of exposure to an innovation are required before an individual adopts the change of behavior.

At any given point of time, an individual is in a particular state in the graph - using or not using an iphone, rooting for or against a political party. We introduce terms like active and inactive to denote one of the two possible states of a node at any given point of time. Inactive nodes exert no influence on their neighbours, whereas active ones do. The state is dynamic and changes over time. It is determined by a binary threshold model - agents can switch from an initially inactive state to an active state if a sufficient proportion of other agents are active. The use of terms like hyper-active and semi-active captures the observation that not all opinions have equal weight - regular users of a product are more enthusiastic recommenders than casual users. Hyperactive nodes exert more influence than active ones, where semi-active nodes have intermediate influence.

In the upcoming sections, we will examine :

- The effect of introducing a tiny fraction of hyper active nodes in an otherwise stagnant network : How strong is the impact? To what extent do the hyperactive centres alter the local behavior? Do they have the potential to alter the global properties of the entire network?
- The effect of low-influencers in the network : Can small things make a big difference? Can a little extra effort create a huge impact? We understand from experience that if a company offers free trials of a product to its potential customers, a first-time customer so gained may go on to become a loyalist. Can we quantify and simulate the effect of low-influencers in the network?

II. Dataset Specification

The Higgs-Twitter dataset has been built after monitoring the spreading processes on Twitter before, during and after the announcement of the discovery of a new particle with the features of the elusive Higgs boson on 4th July 2012. The messages posted in Twitter about this discovery between 1st and 7th July 2012 are considered. It models a multiplex graph with four layers.The layers have been extracted from user activities in Twitter as -

1) Retweet network : Who retweets to whom.
2) Reply network : Reply to existing tweets
3) Mention network : mentioning other users
4) Social relationships among user involved in the above activities : Friends and followers.

The total numbers of nodes(each node representing a twitter user) in the network is 456,631. The terms nodes are users are used interchangebly depending on the context. The maximum number of edges in any layer is 14,855,875. .

III. Definitions

Definition 1. Multiplex graph *A multiplex graph is one in which the nodes share more than one kind of relationship. For example. in our dataset, the nodes share 4 kinds of* *relationships based on whether or not they retweeted to, replied to, mentioned or followed one another.*

Definition 2. LAYER OF A MULTIPLEX GRAPH *The multiple relationships shared between nodes give rise to a multi-layered structure in the graph. Nodes A and B share an edge in a particular layer if they share the relationship represented by that layer. Our dataset, once again, has four layers, each defined by one of the four relationships.*

Definition 3. ACTIVITY LEVEL OF A NODE *A measure of how actively a node propagates information. eg. By retweeting a lot etc.*

Definition 4. PEERPRESSURE *The peer pressure on a node is a measure of the influence that its neighbours have on it. A node with a large fraction of active neighbours will in turn tend to be active. The models and analysis are based on the assumption that only graph neighbours and no one else can influence a twitter user at any given point of time.*

Definition 5. CONTAGION *The spread of a behavior pattern, attitude, or emotion from person to person or group to group through suggestion, propaganda, rumor, or imitation. Eg. A virus, technology, information.*

Definition 6. COMPLEX CONTAGION *Complex contagion refers to the phenomenon in social networks in which multiple sources of exposure to an innovation are required before an individual adopts the change of behavior. This differs from simple contagion in that, unlike a disease, it may not be possible for the innovation to spread after only one incident of contact with an infected neighbor. The spread of complex contagion across a network of people may depend on many social and economic factors; for instance, how many of ones friends adopt the new idea as well as how many of them cannot influence the individual, as well as their own disposition in embracing change.*

Definition 7. INFORMATION CASCADE *An information cascade occurs when people observe the actions of others and then make the same choice that the others have made, independently of their own private information signals. A cascade develops, then, when people abandon their own information in favor of inferences based on earlier peoples actions.*

Definition 8. DIFFUSION *Diffusion is a process, by which information, viruses, gossips and any other behaviors spread over networks,and in particular, over social networks.*

Definition 9. HIGH INFLUENCER *The high influencers have an additional influence on their neighbours besides what they would have had if in active state.With the notion of a high influencer, we introduce a new state for nodes that are hyperactive in nature. A high influencer could be a politician or the leader in a social movement.*

Definition 10. LOW INFLUENCER *Low influencers are the semi active nodes which have influence more than the inactive ones but less than the active ones. This term has been introduced to take into account the partial effect of nodes that have almost reached the threshold for activity but are not quite there yet. An example of this is a situation when a company is giving free trials of its products to potential customers. This set of potential customers were inactive to start with and had a zero influence on their neighbours. Giving them free trials does not make them regular customers but surely makes them more influential in the propagation of the product's goodwill than they were before.*

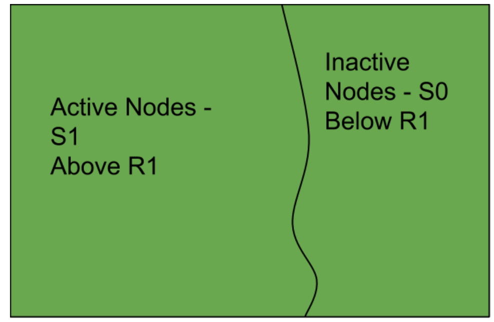

Fig. 1. Single Stage model for comparison against multi stage with high influencer

IV. SINGLE STAGE MODEL VS MULTI-STAGE MODEL WITH HIGH INFLUENCERS

1) **Single Stage Model** Nodes can be inactive or active. Inactive users exert no peer pressure on their neighbours. They are denoted by S_0. Active users exert peer pressure on their neighbours. They are denoted by S_1. When inactive nodes experience sufficient pressure from their active neighbours so that the total pressure crosses a threshold R_1, they undergo a state change to become S_1-active. Figure 1 shows the single stage model for the given comparison against the multi stage model with high influencers.
2) **Multi Stage Model with high influencers** Nodes can be inactive, active or hyperactive. Inactive and active users are as defined above. They are denoted by S_0 and S_1 respectively. Hyper-active users are the ones that are active, but are significantly above threshold R_1 for S_1-active. They exert more peer pressure on their neighbours than the active neighbours. They are denoted by S_2. When users experience sufficient peer pressure from their neighbours so as to cross a threshold R_2, they become hyper-active. S_2-active nodes form a subset of S_1-active nodes. Figure 2 shows the multi stage model with high influencer.

V. SINGLE STAGE MODEL VS MULTI-STAGE MODEL WITH LOW INFLUENCERS

1) **Single Stage Model** Nodes can be inactive or active. Inactive nodes exert no peer pressure on their neighbours. They are denoted by S_0. Active nodes exert peer pressure on their neighbours. They are denoted by

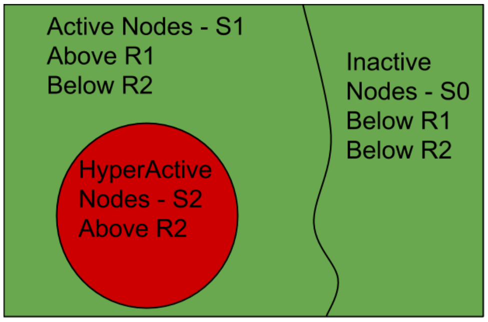

Fig. 2. Multi stage model with high influencers

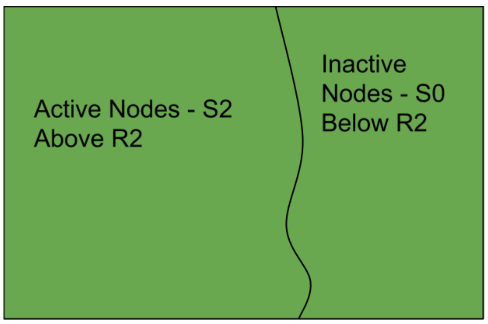

Fig. 3. Single stage model for comparison against multi stage with low influencers

S_2. When inactive nodes experience sufficient pressure from their active neighbours so that the total pressure crosses a threshold R_2, they themselves become S_2-active. Figure 3 shows the single stage model for the given comparison against the multi stage model with low influencers.

2) **Multi Stage Model with high influencer** Nodes can be inactive, semi-active or active. Inactive and active nodes are as defined above. They are denoted by S_0 and S_2 respectively. Semi-active nodes are nodes that are almost active. They are slightly short of the threshold R_2 for S_2-activeness. They exert a small amount of peer pressure on their neighbours, which is less than that of active nodes. They are denoted by S_1. When the peer pressure experienced by a node is less than R_2, but sufficiently high so as to cross a threshold R_1, they become semi-active. S_2-active nodes are once again, a subset of S_1-active nodes. Figure 4 shows the multi stage model with high influencer.

VI. Threshold Models of Binary Decisions

The proposed threshold models are based on binary decisions. The participants of the model can accept (promote in activeness level) or 'reject' (remain in the present state) depending on the fraction of other participants in the respective states. However, the decisions are always binary and there are no dilemmatic conditions arising about acceptance or rejection. The dataset under consideration is a who influences whom graph and is characterized by the following three properties.

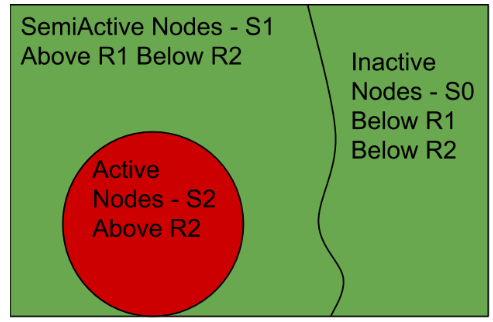

Fig. 4. Multi stage model with low influencer

1) **Interdependence** - The decisions made by a participant depends on the state of the rest of the network.
2) **Non negative increase in the activeness** - Nodes never become less active than they already are. This can be attributed to the fact that as the information propagates in the network, the influence experienced by a node can only increase, and this influence is always reinforcing.
3) **Heterogeneity** - In multi stage models, there can be many configurations of neighbours and their states that can trigger the node in consideration towards becoming active. This is not observed in the single stage model where a fixed configuration (a fixed combination allowing all possible permutations) is required to cross a nodes threshold. For example, in a single-stage model, if a node needs m of its neighbours to be active for its own activation, then any subset of its neighbours with at least m active members is sufficient to make it active. However, the multistage model is heterogenous and there can be many possible combinations to make a node active - 4 active neighbours, or 2 active and 1 hyperactive, or any other threshold crossing configuration.

VII. Peer Pressure and Response Function

The influence a node experiences from its neighbours and the action it undertakes is quantitatively expressed using the response function F which gives the probability of state change for a node. We have based our analysis on the assumption that the threshold for state change is same for all nodes in the graph irrespective of their degree and neighbours. Also note that we had mentioned previously that there arise no dilemmatic conditions and either there **is** a state change or there **isn't**. Therefore the probability of state change to state i given by F can either be 1 or 0. Also the state of a node never demotes. As per the model under consideration, it is the peer pressure on a node that can trigger a state change. Therefore it becomes intutive that F is a function of peer pressure and threshold values for state changes.

A. Formulae

We characterize the relative importance of layer i by its weight w_i. The values of w_i are user-defined. In any layer j, the peer pressure exerted on a node from its neighbours in that layer is defined by :

- For single stage models in comparison with multi stage model with high influencers,

$$P_j = m_{1j}/k_j \tag{1}$$

- For single stage models in comparison with multi stage model with low influencers,

$$P_j = m_{2j}/k_j \tag{2}$$

- For multi stage model with high influencers,

$$P_j = (m_{1j} + \beta \times m_{2j})/k_j \tag{3}$$

- For multi stage model with low influencers,

$$P_j = (m_{1j} - \beta \times (m_{1j} - m_{2j}))/k_j \tag{4}$$

where m_{1j} is the number of S_1-active neighbours in layer j, m_{2j} is the number of S_2-active neighbours in layer j and k is the node degree. β varies from 0 to 1. Note that S_2 active nodes are always S_1 active. The qualitative explanation of the formula evolves from our definitios of low and high influencers. In single stage models, a node can be influenced by only the fraction of active neighbours it has. For multi stage model with high influencers, we had introduced high influencers on top of the single stage model. Thus the extra influence of the high influencers is captured by the β term. In the multi stage model with low influencers, we give the active nodes just the weightage as they had received when in the single stage model(which is equal to 1). However, the low influencer introduced between the inactive and active nodes are now given a weightage of $1 - \beta$ which in a way penalises these nodes for being only semi-active and not active and for having limited influence. Refer to figure 5 for threshold model for single stage vs multi stage with high influencer and figure 6 for single stage vs multi stage with low influencer.
The net pressure is given by

$$P = \sum_{j=1}^{j=4} (w_j \times P_j) \tag{5}$$

The response functions F_1 and F_2 measure the response of a node to the peer pressure experienced by it. These response functions are step function and are defined as:

$$F_i(M, W, k) = \begin{cases} 1 : P \geq Ri \\ 0 : otherwise \end{cases}$$

where M = { m_{11}, m_{12}, m_{13}, m_{14}, m_{21}, m_{22}, m_{23}, m_{24} } and W = { w_1, w_2, w_3, w_4 }
F_1 determines whether a node is S_1-active or not. It is if it crosses the threshold R_1. The function F_2 determines whether the node is S_2-active or not. It is if it crosses the threshold R_2. We require that $R_2 \geq R_1$ in order to satisfy $F_1 \geq F_2$ and thereby guarantee that all S_2-active nodes are also S_1-active.

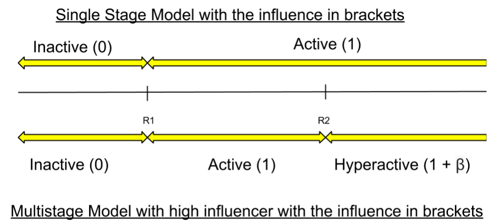

Fig. 5. Threshold Model - Single stage vs Multi stage model with high influencer

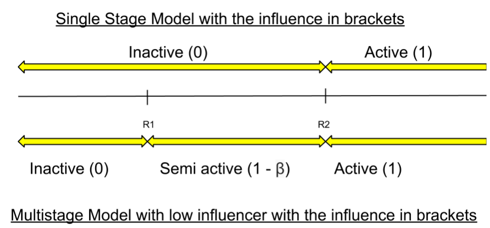

Fig. 6. Threshold Model - Single stage vs Multi stage model with low influencer

VIII. ALGORITHM

Some notations used are -

- N = total number of nodes in the graph
- n.Neighbours = a set of all neighbours of node n
- $n_x.k_j$ is the degree of node n_x in layer j.

1) We initialise a random set of 2% of the nodes as 'active' and the rest as 'inactive'. Since the node labels are randomly assigned, we simply select nodes labelled 0 to 9000 for this purpose.
2) During every iteration over the graph, the total peer pressure across all layers is calculated for every node.
3) Based on the peer pressure and the thresholds, the response function values are calculated for the node under consideration. The state of the node is changed if the appropriate threshold is surpassed.
4) It was experimentally observed that the system activity stabilizes after about 20 iterations in any case. So the number of iterations has been set at a constant of 20.

IX. IMPLEMENTATION AND RESULTS

We implemented the above discussed algorithms with the aid of Java programs. The parameters were set as follows.

1) Single Stage Models : R_{ACTIVE} = 0.3 x 4 = 1.2
2) Multi-Stage Model with High Influencers : R_1 = 0.3 x 4, R_2 = 0.45 x 4. A node is hyperactive if the peer

Algorithm 1 Single Stage in comparison with multi stage model with high influencers

```
for i = 0 to 9000 do
    n_i.state ← ACTIVE
end for
for i = 9001 to 456630 do
    n_i.state ← INACTIVE
end for
for iteration = 1 to 20 do
    for each node n_x in N do
        for each layer j do
            m_1 ← 0
            for each node n_y in n_x.Neighbours do
                if n_y.state = ACTIVE then
                    m_1 ← m_1 + 1
                end if
            end for
            P_j ← m_1 / n_x.k_j
        end for
        P ← Σ_{j=1}^{4} (w_j × P_j)

        if P ≥ R_1 then
            F_1 = 1
            n_x.state ← ACTIVE
        end if
    end for—
end for
```

pressure on it is at least 50% in excess of the threshold for activity.

3) Multi-Stage Model with Low Influencers : R_2 = 0.3 x 4, R_1 = 0.24 x 4. A node is semi-active if the peer pressure on it as at least 80% of the value required to surpass the activity threshold.
   Variation in node activity was observed at regular intervals of time and the results were plotted. The plots are given below.

X. Observations

1) The plot in Figure 7 represents an execution of a Single Stage Contagion Model in comparison with the multi stage model with high influencers. The region in red represents active centres. The region in blue represents nodes that are significantly above threshold, but are not accounted for separately in this model. They are just labelled active. In the graph, we observe an initial increase in the activity of nodes. The fraction of active nodes stabilizes at 10.19% after only 4 iterations over the graph. The fraction of nodes significantly above threshold too stabilizes at 5.31%.
2) The plot in Figure 8 represents Multi Stage Contagion Model with High Influencers. The region in red represents active centres. The region in blue represents hyperactive centres. Once again, we observe an initial increase in the activity of nodes. However, unlike the

Algorithm 2 Multi Stage with High influencers

```
for i = 0 to 9000 do
    n_i.state ← ACTIVE
end for
for i = 9001 to 456630 do
    n_i.state ← INACTIVE
end for
for iteration = 1 to 20 do
    for each node n_x in N do
        for each layer j do
            m_1 ← 0
            m_2 ← 0
            for each node n_y in n_x.Neighbours do
                if n_y.state = ACTIVE then
                    m_1 ← m_1 + 1
                end if
                if m.state = HYPERACTIVE then
                    m_2 ← m_2 + 1
                end if
            end for
            P_j ← (m_1 + (1 + β) × m_2) / n_x.k_j
        end for
        P ← Σ_{j=1}^{4} (w_j × P_j)

        if P ≥ R_1 then
            F_1 = 1
            n_x.state ← ACTIVE
        end if
        if P ≥ R_2 then
            F_2 = 1
            n_x.state ← HYPERACTIVE
        end if
    end for
end for
```

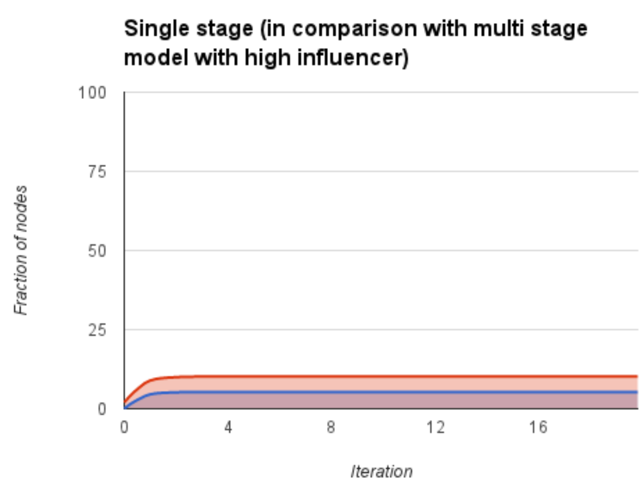

Fig. 7. Single stage in comparison with Multi stage model with high influencer

Algorithm 3 Single Stage in comparison with multi-stage model with low influencers

for i = 0 to 9000 **do**
  n_i.state $\leftarrow$ ACTIVE
end for
for i = 9001 to 456630 **do**
  n_i.state $\leftarrow$ INACTIVE
end for
for iteration = 1 to 20 **do**
  for each node n_x in N **do**
    for each layer j **do**
      $m_2 \leftarrow 0$
      for each node n_y in n_x.Neighbours **do**
        if n_y.state = ACTIVE **then**
          $m_2 \leftarrow m_2 + 1$
        end if
      end for
      $P_j \leftarrow m_2 / n_x.k_j$
    end for
    $P \leftarrow \sum_{j=1}^{4} (w_j \times P_j)$

    if $P \geq R_2$ **then**
      $F_2 = 1$
      n_x.state $\leftarrow$ ACTIVE
    end if
  end for
end for

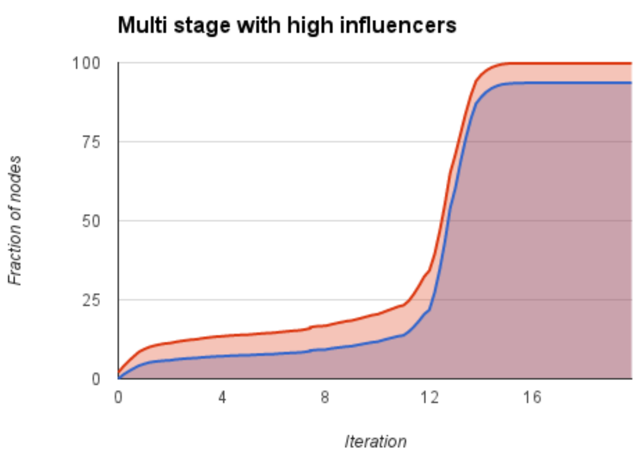

Fig. 8. Multi stage model with high influencer

previous case, the activity of the users does not stabilise after 4 iterations. Instead, there is a steady increase for 16 iterations and when the network activity stabilizes, the fraction of active nodes is 99.82% and the fraction of hyperactive nodes is 93.62%.

3) The plot in Figure 9 models Single Stage Contagion Model as in Figure 7. It represents a run of the same algorithm with the same parameters. The only difference is that whereas Figure 7 shows nodes that are significantly above threshold in blue alongside the active nodes, Figure 9 represents in blue, nodes that are almost active. However, these nodes are not accounted for separately and are labelled as inactive.

Algorithm 4 Multi stage with low influencers

for i = 0 to 9000 **do**
  n_i.state $\leftarrow$ ACTIVE
end for
for i = 9001 to 456630 **do**
  n_i.state $\leftarrow$ INACTIVE
end for
for iteration = 1 to 20 **do**
  for each node n in N **do**
    for each layer j **do**
      $m_1 \leftarrow 0$
      $m_2 \leftarrow 0$
      for each node n_y in n_x.Neighbours **do**
        if m.state = SEMIACTIVE **then**
          $m_1 \leftarrow m_1 + 1$
        end if
        if m.state == ACTIVE **then**
          $m_2 \leftarrow m_2 + 1$
        end if
      end for
      $P_j \leftarrow (m_1 \times (1 - \beta) + m_2) / n_x.k_j$
    end for
    $P \leftarrow \sum_{j=1}^{4} (w_j \times P_j)$

    if $P \geq R_1$ **then**
      $F_1 = 1$
      n_x.state $\leftarrow$ SEMIACTIVE
    end if
    if $P \geq R_2$ **then**
      $F_2 = 1$
      n_x.state $\leftarrow$ ACTIVE
    end if
  end for
end for

4) The plot in Figure 10 models Multi-Stage Contagion Model with Low Influencers. The region in red represents active centres. The region in blue represents semi-active centres. However, unlike in the Single Stage model, the activity of the users does not stabilise after 4 iterations. Instead, there is a steady increase for 19 iterations and when the network activity stabilizes, the fraction of active nodes is 93.04% and the fraction of semi-active nodes is 97.56% .

XI. Conclusion

- In each case, we started with a network 2% of whose nodes were initialized to active. This 2% could represent the initial adopters of a new technology, or as in our context, the first tweeters of the higgs-boson rumor.

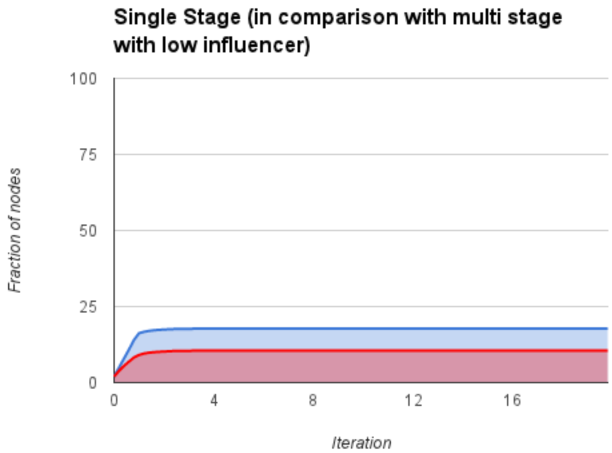

Fig. 9. Single stage in comparison with Multi stage model with low influencer

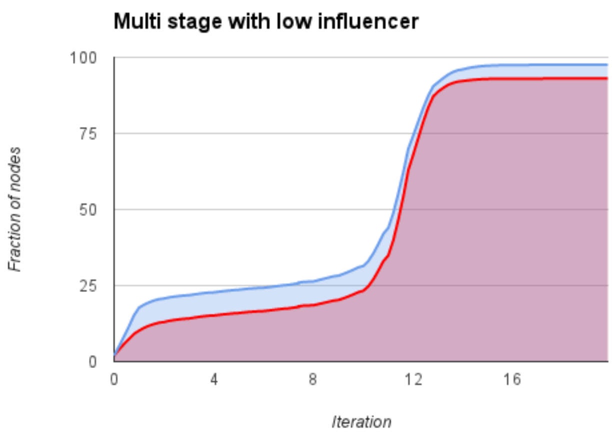

Fig. 10. Multi stage model with low influencer

- In the case of Single Stage Model, the network activity saw an initial jump after which it stabilised at a low value of about 10%.
- When we introduced the notion of high influencers in the network, we saw that the impact of these hyperactive centres did not stay localised and diffused throughout the network. Hyperactive nodes have 1.5 times as much influence as S_1-active nodes. The presence of S_2-active nodes and the additional influence from them triggers a cascade of S_1-active nodes that otherwise would not have occurred.
- Finally, we take into account users who do not meet the threshold requirement in the Single Stage model and assign to them a new state of semiactivity. We observe, rather surprisingly, that this leads to a phenomenon very similar to that of the Multi Stage model with High Influencers. The lowering of threshold activity by 20% brings out sufficient nodes from inactivity so that the influence from these when accounted for causes a system-wide cascade.

References

[1] M. De Domenico, a Lima, P. Mougel, and M. Musolesi, The anatomy of a scientific rumor., *Sci. Rep.*, vol. 3, p. 2980, 2013.

[2] S. Melnik, J. a Ward, J. P. Gleeson, and M. a Porter, Multi-stage complex contagions., *Chaos*, vol. 23, p. 013124, 2013.

[3] Raquel A Banos, Javier Borge-Holthoefer and Yamir Moreno ”The role of hidden influentials in the diffusion of online information cascades”