Dear Yimei,   
  
Have you found the full Twitter link graph as Juan suggested?  
  
Would you please write a summary of the literature survey that you have done recently? including the relevant works you mentioned at our last meeting?  
  
For the moment I suggest you start to study the impact of seeds on spreading with your simulation platform using any infection rate model provided by Liyi, i.e. use the same infection rate model but with different seeds (who and when), in light of the relevant works you surveyed.

Good evening,

Sorry to reply so late because I just see this email at 8 p.m. today.

For the Twitter link graph, Liyi and I found one dataset from Stanford. However, it only contains 81306 nodes and its diameter is 7. We think it is still not big enough to do the simulation for those events, as well as Huan’s graph.

Till now, after determining our objective for the project to study the factors that influence the information spread on Twitter, such as the seeds’ changes and the different network topology, my idea is to divide my simulation process into two parts.

Firstly, to examine whether my simulation platform is correct or not, I will simulate just **on the Givenchy topology** using the Givenchy model that Liyi generated because it is more convenient and intuitive to evaluate.

Secondly, to receive the acknowledge about the seeds influence, I will simulate **on the Hawking’s topology** using the Givenchy model.

Moreover, after Liyi obtaining*the NYC’s and FLORIDA’s topology*, I can study the acknowledge about the spread influence caused by the different topology by simulating **on these three different Twitter graphs**. By the way, I think these Twitter graphs are the biggest we can obtain at present and their data has been preprocessed by Liyi, which is more convenient for us to utilize. Moreover, to study the influence caused by the network topological properties, in my opinion, we can even construct the networks using some software. However, it may not reflect the specificity of the Twitter spread.

Next, I will give a brief summary of my recent literature survey and the relevant articles are attached below.

When talked about the influence of the seeds changes on the information spread, there are 2 aspects: the seeds’ time of appearance and whether the seeds are active or not.

Firstly, comparing single-stage seeding strategy that means all the seeds appear together at the beginning with sequential seeding strategy that means the seeds appear in a given order, the single-stage one can bring a rapid seeding in the early stage while the sequential one can trade off between the spread speed and the coverage, finally making more nodes active and giving a **larger coverage**. (Here I want to emphasize that what I said in our last meeting about this point was wrong.)

Secondly, when studying the influence caused by whether the seeds or the top n most engaged users (and they are the direct followers of the seeds) active or not, the results show that whether making the seeds inactive or making the top n inactive, the spread coverage will decrease obviously. Moreover, the seed has more influence than the top 10 influencers but less than the top 100 influencers and random 100 users over time. And the Top 100 series dominates the others.

Then, when it comes to the influence on the spread caused by the network topological properties, it studied 3 aspects: the degree distribution, the network density and the assortativity coefficient.

Firstly, it studied the influence of the power-low exponent when thinking about the degree distribution. As the exponent increases, the spread coverage will decrease subsequently.

Secondly, to measure the network density, it took the average degree into consideration. The result shows that the coverage will increase as the density increases.

Thirdly, the coverage of active nodes decreases dramatically as the assortativity coefficient becomes larger. The underlying reason may be that as the assortativity coefficient increases, the hubs have more connections with one another and avoid the nodes with small degrees,

And see you next Tuesday.

Best wishes,

Yimei Zhu