Part I

- page 96 and 97 gave us some general questions to ask during our initial research exploration. (Yi)

- How to systematically characterize the network structure?
- How do properties relate to one another?
- Is there something else we should measure?
-
- Design systems (networks) that will
    - Be robust to node failures
    - Support local search (navigation): P2P networks
-
- Why are networks the way they are?
-
- Predict the future of the network?
- How should one be taking care of the network for it to grow organically?

His way of doing research is worthwhile for us to learn. It is not to directly use his models, but to design your own model with similar methods.
After we have some data in a certain domain, we could find patterns in the data and build models to explain these patterns. Suppose we know there exists power law distribution, then we may include the preferential attachment in our model.
Both our intuition and appropriate formula can lead us to create the model with property that we desire.
In network research, people usually give interpretation of the data, instead of predicting the future trend.

- page 80 (Yi)

- Metropolis sampling:
    - Start with a random permutation
    - Do local moves on the permutation
    - Accept the new permutation
        - If new permutation is better (gives higher likelihood)
        - If new is worse accept with probability proportional to the ratio of likelihoods

This "Metropolis sampling" is a nice way to estimate the likelihood when likelihood is not easy to calculate.

- page 95 (Jessica)

- Graph sampling – many real world graphs are too large to deal with

Since we need to maintain same characteristics when sampling graph, it is not appropriate to pick random nodes since it will end up with a smaller graph with low connectivity.
What the author suggested is a way to build a smaller graph, with similar characteristics, as long as using the same Kronecker initiator graph Θ

Part II

- page 6 contains lots of potential applications of network models (Yi)

§A fundamental process in social networks:
Behaviors that cascade from node to node like an epidemic
- –News, opinions, rumors, fads, urban legends, …
- –Word-of-mouth effects in marketing: rise of new websites, free web based services
- –Virus, disease propagation
- –Change in social priorities: smoking, recycling
- –Saturation news coverage: topic diffusion among bloggers
- –Internet-energized political campaigns
- –Cascading failures in financial markets
- –Localized effects: riots, people walking out of a lecture

- page 23 (Threshold Model) and 24 (Independent Contagion Model) (JiaZhong)

- how to activate the node?
- Threshold model is a deterministic process while the independent contagion model is a stochastic process.
- Every time, the set of starting nodes is different
- how can it be useful?
  - In order to find the most influential 10 people for advertisement propagation, we can do simulations by choosing random 10 people and find the best combinations.
  - To put sensors in water contamination problem
  - To ask the future questions of the model

- page 26 (Most Influential Subset of Nodes) indicates some potential applications (Yi)

- To introduce new product in facebook; to send coupons to most influential people
- Useful in recommender system

- page 31 (Jian)

- Only binary value is considered in their model.
- If we want to extend this model in personalization project, a vector might be more appropriate.
- The values within the vector could also have influence to each other.

- page 44 & 45 (Yi)

- The author plots lots of figures between the two variables and each figure answers one question. These simple analysis could be useful during the exploration period.

- page 59 (Information Cascades in Blogs) (Jian)

- Could we focus on links instead of nodes? We could decide the importance of a link by its affect on neighboring nodes. Then the weight of link could be assigned according to its importance.
- Whether focusing on link or node depends on the goal of task.

- page 65&66 (Jessica)

- whether earlier models match the cascades characteristic?
- They modified the original cascade model, to fit the data better.

- page 94 (Jessica)

- not clear of the calculation of the online bound

Additional discussion

- Min

- The idea of diffusion is useful in his research.
- Besides the traditional graph, a separate graph on user's editing behavior should be interesting. The user's editing behavior indicates the topics that he is familiar with. (Personalized information)
- The author's method could help him to find most influential categories, in order to reduce the number of categories considering scalability issue.

- Xing

- The result for search engine is not surprising.

- Lanbo

- Difference between query reformulation & transition.
- These two terms are not well defined
- Transition only contains "adding/deleting" action, according to page 70
- Why do users reformulate query? Reformulation still means potential satisfaction.

- JiaZhong

- In Part III, the authors talk more about statistical figures and results, instead of using his previous models.
- It is typical analysis in social science or business school.
- The simple analysis gives us statistical results, and it provides correlation between previous models and realistic phenomenon.

- YiZe

- Dynamic way to analyze social network, in addition to using static user similarity
- feature selection might be different in single/multiple domains, or various type of information propagation.
- the static part identifies "who you are", which works together with the dynamic part.
- To combine the two together is a promising research direction.

- Jessica

- one potential application is news detection task.
- By capturing the most influential nodes, we can choose those to crawl first.
- Similarly, we can know which to filter first in filtering system.

- Yi

- How to develop model?
    - Keep good part of the original model & make it fits the new characteristic.
    - Such as the log-linear property and list of most frequent cascades.
- How to create parallel computing environment?
    - To do experiments on cluster & give tutorial
    - Install Hadoop on lab cluster.