End of the Semester Summary
2013 Spring

Jialu Liu
Thanks Ming for Presentation

May 1, 2013
What I Did

1. Publish one paper: **Multi-View Clustering via Joint NMF** in SDM2013 with Chi, Jing and Prof. Han.
   - Introduced in last semester summary meeting.

2. Publish one paper: **Large-Scale Spectral Clustering on Graphs** in IJCAI2013 with Chi, Marina and Prof. Han.
   - Existing methods might not be efficient if the network (homogeneous) is too large (millions of nodes or edges)
   - We propose to first randomly sample seeds on network in order to find initial partition. Then we build a bipartite structure based on the initial partition and adopt spectral methods to refine the clustering result.
   - Running time: less than 10 seconds to partition the network with 100K nodes and 8.2M edges for 10 clusters (excluding the hierarchical or approximate k-means step)
Submit one paper: *A Generative Model for Clustering Heterogeneous Information Networks* in KDD2013 with Chi, Jing and Prof. Han.

- A unified statistical model interpreting why the HIN is formed in its observed way.
- Using iterative updating rules very similar to RankClus: ranking $\leftrightarrow$ clustering.
- Not constraint from the schema of the HIN.
- It can predict exact weight of link for any combinations of nodes in the network, which is promising for many other tasks like recommendation, link prediction and outlier detection.
What I Am Doing

1. **Phrase Mining**, collaborated with Chi, Marina and Xiang.
   - Improving the phrase mining using HIN by incorporating automatic learning of link strengths, e.g., strength(author-paper) probably will be weaker than strength(paper-venue) in DBLP-4-area task.
   - Studying the initialization of the topic models in the *CATHY* framework.
   - Studying how to apply *CATHY* to long text scenario.

2. **KDDCup**, collaborated with Jeffrey Yufei Liu and many guys from our group.
   - Currently working on name cleaning.
   - Will help Chi with the co-author analysis.
Thanks again for Ming’s help