Prediction and Recommendation

Xiang, Xiao, Brandon,
Jialu, Urvashi, Quanquan
Problem: author citation prediction (A-P→P-A)

1. author got related and high quality papers
2. author works on related and popular topics
3. author has high reputation
4. author is his friend or collaborator

Meta-path based approach

- Helps shrink the search space, but might not be enough to specify who to cite
  - a1 and a2 co-author 98 times;
  - a1 and a3 co-author 95 times; → who to cite?
- Some factors HARD to study through meta paths
Need extra philosophies to guide author citation process

Intuitively, author prefers to:
1. cite highly regarded authors/co-authors
2. cite authors who have high quality papers
3. cite authors who publish in highly reputed venues
4. cite authors who work on hot topics

Object ranking counts in author citation process

Partial ordering of heterogeneous typed objects based on HIN structure (point-wise)
Working on solution (Xiang, Jialu, Xiao)

- Ranking learned from network topologies
  - Authority propagation idea (RankClass, NetClus)
  - On global/local level
  - Using meta paths
- Combine ranking in prediction model learning (ongoing)
  - Ranking-based features as extra features
  - Or, propose ranking-based meta path measures
Personalized CF with entity similarity regularization

- Global model (IJCAI-HINA’13)
  - Multiple entity/item similarities measured by different meta paths $\rightarrow$ multiple sources for collaborative filtering
  - Graph regularization for learning influence of different meta paths $\rightarrow$ weights for all users (global)

- **Extend to personalized model (On-going)**
  - Each user has different flavors on meta paths
  - Method derivation done (Xiang, Xiao, Quanquann)
  - Experiments ongoing (Brandon)
NetMine on News

Xiang, Shi, Fangbo
HIN construction from news

- **Data:** 2010 news collection from 3 news agencies
  - Article #: 58,515

- **Objects (6 types)**
  - Date #: 365; Location #: 584; Event #: 33;
  - Person #: 21,574
    - Top-3 related people in each article
  - Organization #: 24,082
    - Top-3 related organizations in each article
  - Term #: 23,550
    - Tokenize news title, remove stopwords, fuzzy matching
HIN construction from news

- Date, location, event:
  - Lead to denser links (location-article, date-article, event-article)
  - Less informative on indicating relevance of content
    - Articles in same date
    - Articles in same location
    - Event is coarse
Run NetClus (ongoing)

- Run NetClus on constructed network
  - Current implementation is specially designed for DBLP schema (only 4 types of nodes)
  - Try different combination of object types
    - Date, location, event
    - Person, organization, term
    - Etc.
  - Tune number of clusters, parameters
  - Modify the implementation and combine more