Fall Semester Summary

Chenguang Wang
Previous studies: *Schema-simple HINs*

- DBLP network: **four entity types** (Paper, Author, Venue, Term), and **several relation types**

In real world: *Schema-rich HINs*

- Networks have more complex schema
  - Freebase network: **1,500+ entity types** and **35,000+ relation types**
Relation Similarity Search
(Joint work with Yizhou Sun, Yanglei Song, Jiawei Han, Yangqiu Song)

Latent Semantic Relations

- $P_1$: co-founders (0.45)
- $P_2$: schoolmates (0.25)
- $P_3$: co-inventors (0.15)

Search Result (ranked)

- Bill Gates
- Paul Allen
- Steve Jobs
- Steve Wozniak
- Steve Ballmer
- Mark Zuckerberg

Query

Larry Page  Sergey Brin

Jerry Yang  David Filo

Search Result

Stanford  Google  PageRank

Sergey Brin  Larry Page  David Filo

Yahoo!  Yahoo! Mail
Experiments

<table>
<thead>
<tr>
<th>Query: {⟨Google, Larry Page⟩, ⟨Microsoft, Bill Gates⟩, etc.}</th>
<th>ω</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organization → Founder</td>
<td>0.384</td>
</tr>
<tr>
<td>Organization run business in Industry win award in^{-1} Founder</td>
<td>0.274</td>
</tr>
<tr>
<td>Organization is founded by Person is influence peer^{-1} Founder</td>
<td>0.174</td>
</tr>
<tr>
<td>Organization’s leadership Person mailing address Location mailing address^{-1} Founder</td>
<td>0.115</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Query: {⟨Google, Larry Page⟩, ⟨Yahoo!, Marissa Mayer⟩, etc.}</th>
<th>ω</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organization run by CEO job title Founder</td>
<td>0.32</td>
</tr>
<tr>
<td>Organization founded date Date graduation date^{-1} Founder</td>
<td>0.229</td>
</tr>
<tr>
<td>Organization headquarter Location education institute Founder</td>
<td>0.207</td>
</tr>
<tr>
<td>Organization run business in Industry win award in^{-1} Founder</td>
<td>0.113</td>
</tr>
</tbody>
</table>
Relation Canonicalization

Unstructured Data

```
"In 1995, J.K. Rowling finished writing manuscript for Harry Potter and the Philosopher's Stone on an old manual typewriter."

"Google was founded by Larry Page and Sergey Brin while they were Ph.D. students at Stanford University."

......
```

Knowledge Bases

```
J.K. Rowling   Harry Potter Series
, written work,

Philosopher's Stone Harry Potter Series
, part of,

Larry Page Google
, is founder of,

......
```

Cluster 1

```
J.K. Rowling   Philosopher's Stone
, wrote,

J.K. Rowling   Harry Potter Series
, written work,

Philosopher's Stone Harry Potter Series
, part of,

......
```

Cluster 2

```
Google Larry Page
, is founded by,

Larry Page Google
, is founder of,

......
```

Open Information Extraction

Canonicalizing Relations (CTGC)

Multi-Hop Relation Generation
Experiments

![Graph showing NMI vs # of Relation Constraints for different algorithms](image)
Future Directions

- Generalization: Documents+TopMine->Open domain (schema-rich) network relation similarity search.
- Meta-Path Generation for schema-rich heterogeneous information networks.
- RelSim-based Recommendation in schema-rich heterogeneous information networks.
- RelSim-based Clustering/Classification in schema-rich heterogeneous information networks.
Research
- Heterogeneous information network-enhanced document clustering (with Ahmed) [Planning for KDD’15]
- Heterogeneous information network type refinement (with Fangbo) [Planning for KDD’15]

NIH, AAAI Review
Thanks for all your great help!