Hector: Current Interests

- Information Integration
- Information Privacy & Security
- Web Information
- Peer-to-Peer/Distributed Systems
- Managing Bio-Diversity Information
- Social Networks
CourseRank

• The System
• The Research
CourseRank Research

Hector Garcia-Molina
with Benjamin Bercovitz, Filip Kaliszan, Georgia Koutrika, Robert Ikeda, Henry Liou

Stanford University
CourseRank Features

• Rich data
• Collaboration with University
• User identities, closed system
• We can modify & experiment!
CourseRank Challenges

• “Rich” Recommendations
• Academic requirements
• Question/answer forums
• Social issues
• Privacy
• ...

...
Focus on Two Components

• Course Requirements
• Flexible Recommendations
Course Requirements

• Checking Requirements

• Recommendations Under Requirements
Example

• $R_1$: Take 1 from $\{a, p\}$
• $R_2$: Take 2 from $\{p, d, i\}$
• $R_3$: Take 1 from $\{i, o\}$
Example

• $R_1$: Take 1 from \{a, p\}
• $R_2$: Take 2 from \{p, d, i\}
• $R_3$: Take 1 from \{i, o\}

• Alice has taken: \{a, p, i, o\}
• Bob has taken \{p, d, o\}
Example

• $R_1$: Take 1 from $\{a, p\}$
• $R_2$: Take 2 from $\{p, d, i\}$
• $R_3$: Take 1 from $\{i, o\}$

• Alice has taken: $\{a, p, i, o\}$ Satisfied!
• Bob has taken $\{p, d, o\}$
Example

• \( R_1 \): Take 1 from \( \{a, p\} \)
• \( R_2 \): Take 2 from \( \{p, d, i\} \)
• \( R_3 \): Take 1 from \( \{i, o\} \)

• Alice has taken: \( \{a, p, i, o\} \)
• Bob has taken \( \{p, d, o\} \) NOT Satisfied!
• Alice has taken: \{a, p, i, o\}
Alice has taken: \{a, p, i, o\}

total flow =4, success!
Bob has taken: \{p, d, o\}
• Bob has taken: \{p, d, o\}

total flow = 3, failure!
Recommendations

classes taken by Bob

classes NOT taken by Bob
Recommendations

classes taken by Bob

classes NOT taken by Bob
Recommendations

- Classes taken by Bob
- Classes NOT taken by Bob
- Course i recommended to Bob
Recommendations

Costs can be inverse “desirability”.

Course i recommended to Bob

Classes taken by Bob

Classes NOT taken by Bob
Additional Features

• Preference scores for courses

• Units requirements

• Forbidden pairs
  – Example: Take 3 from \{a, b, c, d, e, f, g\}; forbidden: \{(a,c), (d,e), (d,g)\}

• and many more...
What and How to Recommend?

Courses

Students

Majors

Grades

Departments

Friends

History

Quarter
Interface: Meld Search/Recommend

• What are you looking for?
• Recommend based on what?
• Order results how?
• Constraints?
Interface: Meld Search/Recommend

**Target:**

<table>
<thead>
<tr>
<th>course</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>major</td>
<td></td>
</tr>
<tr>
<td>instructor</td>
<td></td>
</tr>
<tr>
<td>student</td>
<td></td>
</tr>
<tr>
<td>quarter</td>
<td></td>
</tr>
</tbody>
</table>
Interface: Meld Search/Recommend

Target:

Courses in Dept:

Course Keywords:

Recommend based on:

students (similar tastes)

students (similar grades)

students (similar courses)

course history

-
Interface: Meld Search/Recommend

**Target:**
- course

**Recommend based on:**
- students (similar tastes)

Courses in Dept:

Course Keywords:

Students in Dept:

Students in Class:
Goal

users

front end

recommendation engine

data

workflow

workflow

workflow

admin
Goal: Recommendation Workflow

students

F

condition

R

me

me

similar students

R

parameters

R

parameters

F

condition

recommended courses

courses
FlexRecs

\[ \text{Courses} \rightarrow \text{identify[CID, History, Rating]}, \text{weight_ave[Score]} \]

\[ \sigma_{\text{suID} \neq \text{Joe}} \rightarrow \text{inv_euclidean[History, CID, Rating]} \]

\[ \sigma_{\text{suID} = \text{Joe}} \rightarrow \varepsilon \]

\[ \varepsilon \rightarrow \pi \{\text{SuID, CID, Rating}\} \]

\[ \pi \rightarrow \text{History} \]
# Compare Example: Jaccard

## Jaccard Score Table

<table>
<thead>
<tr>
<th>EStudents</th>
<th>Courses</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>11</strong> Joe</td>
<td>11 C010 Art</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>11 C020 Comp</td>
<td></td>
</tr>
<tr>
<td><strong>12</strong> Sally</td>
<td>12 C010 Art</td>
<td>0.66</td>
</tr>
<tr>
<td></td>
<td>12 C020 Comp</td>
<td></td>
</tr>
<tr>
<td></td>
<td>12 C030 Bio</td>
<td></td>
</tr>
<tr>
<td><strong>13</strong> Fred</td>
<td>13 C010 Art</td>
<td>0.33</td>
</tr>
<tr>
<td></td>
<td>13 C030 Bio</td>
<td></td>
</tr>
</tbody>
</table>

### Courses
- C010: Art
- C020: Comp
- C030: Bio

### Jaccard Score
- Score calculated based on shared courses.
Systems Issues

• How expensive
• Impact of comparison functions
• Scalability
• Optimization
Experiments: Generation, Execution

- same workflow as in our running example
Summary

• CourseRank
• Recommendations with Constraints
• FlexRecs