Profit-maximizing Cluster Hires

Behzad Golshan
behzad@cs.bu.edu

Theodoros Lappas
tlappas@stevens.edu

Evimaria Terzi
evimaria@cs.bu.edu
Introduction

- Two main components of a job market

<table>
<thead>
<tr>
<th>Jobs</th>
<th>Workers</th>
</tr>
</thead>
<tbody>
<tr>
<td>JAVA, Node.js</td>
<td>JAVA, C++, SQL</td>
</tr>
<tr>
<td>$90/hour</td>
<td>$18/hour</td>
</tr>
<tr>
<td>HTML, JAVA</td>
<td>JAVA, HTML</td>
</tr>
<tr>
<td>$33/hour</td>
<td>$7/hour</td>
</tr>
<tr>
<td>Node.js, SQL</td>
<td>HTML, Node.js</td>
</tr>
<tr>
<td>$10/hour</td>
<td>$40/hour</td>
</tr>
</tbody>
</table>
Introduction

- Two main components of a job market

**Jobs**

- JAVA, Node.js
  - $90/hour
- HTML, JAVA
  - $33/hour
- Node.js, SQL
  - $10/hour

**Workers**

- JAVA, C++, SQL
  - $18/hour
- JAVA, HTML
  - $7/hour
- HTML, Node.js
  - $40/hour

**Organizations**

- Project Managers
- Startups
# Introduction

- Two main components of a job market

## Jobs

<table>
<thead>
<tr>
<th>Technology</th>
<th>Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>JAVA</td>
<td>$90/h</td>
</tr>
<tr>
<td>Node.JS</td>
<td></td>
</tr>
<tr>
<td>HTML</td>
<td>$33/h</td>
</tr>
<tr>
<td>JAVA</td>
<td></td>
</tr>
<tr>
<td>JAVA, HTML</td>
<td>$7/h</td>
</tr>
<tr>
<td>Node.JS</td>
<td></td>
</tr>
<tr>
<td>SQL</td>
<td>$10/h</td>
</tr>
<tr>
<td>JAVA, C++</td>
<td>$18/h</td>
</tr>
<tr>
<td>SQL</td>
<td></td>
</tr>
</tbody>
</table>

## Workers

- JAVA, C++, SQL: $18/h
- JAVA, HTML: $7/h
- HTML, Node.JS: $40/h

Who to hire and which jobs to do?
Introduction

- Examples:

**Jobs**
- JAVA, Node.js: $90/hour
- HTML, JAVA: $33/hour
- Node.js, SQL: $10/hour
- JAVA, C++, SQL: $18/hour
- JAVA, HTML: $7/hour
- HTML, Node.js: $40/hour
Introduction

Examples:

Conferences

- KDD
  Graph Analysis
  Data Streams
  Statistics
- NIPS
  ML
  Statistics
  AI
- CHI
  Vision
  UI Design
  User Study

Applicants

- Universities
- Systems, Databases
- Algorithms
- ML, Vision
How much hiring a team of experts cost?

**Jobs**
- JAVA
- Node.js
  - 90$ / hour
- HTML
  - JAVA
    - 33$ / hour
- Node.js
  - SQL
    - 10$ / hour

**Workers**
- JAVA, C++, SQL
  - 18$ / hour
- JAVA, HTML
  - 7$ / hour
- HTML, Node.js
  - 40$ / hour

Cost = 47$
Framework

- Which jobs can a team of experts complete?

**Jobs**

- JAVA, Node.js: $90 / hour
- HTML, JAVA: $33 / hour
- Node.js, SQL: $10 / hour

**Workers**

- JAVA, C++, SQL: $18 / hour
- JAVA, HTML: $7 / hour
- HTML, Node.js: $40 / hour

Coverage
Framework

- Possible profit models?

**Jobs**

- **JAVA Node.JS**
  - 90$ / hour

- **HTML**
  - **JAVA**
  - 33$ / hour

- **Node.JS SQL**
  - 10$ / hour
Possible profit models?

Dollar-based profit model:
- Value of all project covered by experts
- Example: $90 + $33

jobs

JAVA
Node.js
$90/hour

HTML
JAVA
$33/hour

Node.js
SQL
$10/hour
Framework

- Possible profit models?

**Jobs**

- **Dollar-based profit model:**
  - Value of all project covered by experts
  - Example: $90 + 33$

- **Competition-based profit model:**
  \[
  \text{Profit}(J) = \frac{1}{\text{Freq. of the rarest skill in } J} \times \text{Dollar Profit}
  \]
  - Captures the competition
  - Example: \((90 + 33) / 10\)
  - Assuming that only 10 people know JAVA!
Problem

- ClusterHire:
  - Given a budget $B$, hire a team of experts $T$ such that
    - Cost $\leq B$,
    - Profit is maximized.

- Complexity: NP-hard to solve and approximate.
  - Reduction from Set Cover to ClusterHire
Problem

ClusterHire:
- Given a budget $B$, hire a team of experts $T$ such that
  - Cost $\leq B$,
  - Profit is maximized.

Complexity: NP-hard to solve and approximate.
- Reduction from Set Cover to ClusterHire

$t$-ClusterHire:
- Each skill of a worker can be used in at most $t$ projects
- Complexity: NP-hard to evaluate the objective function
  - Reduction from Set Packing to $t$-ClusterHire
## Algorithms

- **ExpertGreedy**
  - Hires an expert in each iteration
  - Expert with the best profit to cost ratio
  - Repeats until the budget is consumed
- Exploits the skills of current team

### Jobs
- **JAVA, Web Design, Graphic Design** *(100$)*
- **Data Analysis, Python, ML** *(100$)*

### Workers
- **Data Analysis, Python** *(10$)*
- **Graphic Design** *(10$)*
- **ML** *(10$)*

### Current Team
- **JAVA, SQL, Web Design, SEO**
Algorithms

- ExpertGreedy
  - Hires an expert in each iteration
  - Expert with the best profit to cost ratio
  - Repeats until the budget is consumed
- Exploits the skills of current team

**Jobs**

<table>
<thead>
<tr>
<th>JAVA, Web Design, Graphic Design</th>
<th>(100$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Analysis, Python, ML</td>
<td>(100$)</td>
</tr>
</tbody>
</table>

**Workers**

<table>
<thead>
<tr>
<th>Data Analysis, Python</th>
<th>(10$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Graphic Design</td>
<td>(10$)</td>
</tr>
<tr>
<td>ML</td>
<td>(10$)</td>
</tr>
</tbody>
</table>

**Current Team**

JAVA, SQL, Web Design, SEO
Algorithms

- **ProjectGreedy**
  - Selects a job in each iteration
  - Hires a cost-effective team for the job
  - Project with the best profit to cost ratio
- Considers combination of workers to hire

### Jobs
- JAVA, Web Design, Graphic Design ($100$)
- Data Analysis, Python, ML ($200$)

### Workers
- Data Analysis, Python ($10$)
- Graphic Design ($10$)
- ML ($10$)

### Current Team
- JAVA, SQL, Web Design, SEO

<table>
<thead>
<tr>
<th>Cost</th>
<th>Profit</th>
<th>Vs</th>
<th>Cost</th>
<th>Profit</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>200</td>
<td></td>
<td>10</td>
<td>100</td>
</tr>
</tbody>
</table>
Algorithms

- CliqueGreedy
  - Similar to ProjectGreedy
  - Considers cliques of compatible projects
  - Cliques vs. Projects

Jobs

- JAVA, Web Design, SQL (100$)
  - Costs 30$

- JAVA, Web Design (200$)
  - Costs 20$

- Data Entry (50$)

Workers

- JAVA (10$)
- Web Design (10$)
- SQL (10$)
- Data Entry (5$)
Algorithms

- CliqueGreedy
  - Similar to ProjectGreedy
  - Considers cliques of compatible projects
  - Cliques vs. Projects

**Jobs**

- JAVA, Web Design, SQL (100$)
- JAVA, Web Design (200$)
- Data Entry (50$)

**Workers**

- JAVA (10$)
- Web Design (10$)
- SQL (10$)
- Data Entry (5$)

Costs 30$
Real-World Datasets

- freelancer.com
  - 1,763 experts
  - 721 projects

- guru.com
  - 6,473 experts
  - 1,764 projects
Real-World Datasets

- Analysis of workers data
  - Freelancer
  - Guru
Real-World Datasets

- Analysis of workers data (Jaccard similarity of workers)
  - Freelancer
  - Guru
Experiments for $t$-ClusterHire ($t = 3$)

- Dollar-based
  - Guru
  - Freelancer

- Competition-based
  - Guru
  - Freelancer
Experiments for $t$-ClusterHire ($t = 3$)

- **Workers Similarity**
- **Competition-based**

**Guru**

**Freelancer**

![Diagram showing workers similarity and competition-based graphs for Guru and Freelancer.](image-url)
Experiments (speedup)

- Performance of CliqueGreedy
  - Freelancer
    - Nodes: 721
    - Cliques: 520
  - Guru
    - Nodes: 1764
    - Cliques: 1660
Thanks

KDD 2014
Thanks