Analysis of skill demand graph

NTU WINTER SCHOOL 2017
ERIK NOVAK, ZALA HERGA, XIANGTING HOU
Introduction

• European Data Science Academy (EDSA)
• Connect skill demand and supply
• What are the hidden skills in job demands?

Who is an IDEAL Data Scientist?
Data description

• 1.4 million job postings (data science)
  • Acquired between December 2015 and January 2017
  • Job posting contains: title, description, publish date, organization etc.
  • Skill extraction from title and description (557)

```json
{
  "id": 711654,
  "title": "Arquitecto/a PHP",
  "date": "2016-11-09T07:39:00.000Z",
  "description": "También será valorable conocimientos de Python, Elastic, JQuery y Angular2...",
  "organization": "N/A",
  "skillset": ["python", "jquery", "php", "architect"],
  "location_coord": [41.38879, 2.15899],
  "location_city": "Barcelona",
  "location_country": "Spain"
}
```
Network creation

Edge weight: the co-occurrence of skills in the dataset

Node weight: the relative frequency of the skill the dataset
275 nodes
3056 edges
Average degree: 22.225
74 nodes
1478 edges
Average degree: 39.946
Network analysis

Degree distribution

Degree distribution

January 2016

Degree distribution

December 2016
Network analysis (cont.)

Log-log plot

Degree distribution

January 2016

Degree distribution

December 2016
Skill recommender system

Based on skill network recommend which skills to acquire to increase the probability of employment

- Input: skill list – what you already know
- Output: ranked skill list – the recommended skills to learn

\[ f(v_i) = \alpha \cdot W_e + \beta \cdot W_n \]
Experiment

Check the skill change during the year

- Parameters:
  - $\alpha, \beta \in \{0.01, 0.5, 1\}$
  - Skill lists: [html, jquery, javascript]

$\alpha = 1, \beta = 0.01$

<table>
<thead>
<tr>
<th>January 2016</th>
<th>December 2016</th>
</tr>
</thead>
<tbody>
<tr>
<td>HTML</td>
<td>HTML</td>
</tr>
<tr>
<td>SQL</td>
<td>SQL</td>
</tr>
<tr>
<td>PHP</td>
<td>PHP</td>
</tr>
<tr>
<td>.NET framework</td>
<td>Database</td>
</tr>
<tr>
<td>.NET</td>
<td>Machine learning</td>
</tr>
</tbody>
</table>
Future work

• Hierarchical structure of skills
• Improve the skill recommender system
  • Explore the values of parameters in the model
  • Use historical data to predict which skills will be highly demanded in the future

Code: https://github.com/ErikNovak/job_skill_network
Team

Zala  Erik  Xiangting
Thank you for listening!

ANY QUESTIONS?