Temporal Analysis of Inter-Community User Flows in Online Knowledge-Sharing Networks

Anna Guimarães, Ana Paula Couto da Silva, Jussara Almeida
Department of Computer Science - UFMG (Brazil)

August 10, 2015
Communities in Knowledge-Sharing Networks

- **Online Knowledge-Sharing Networks**
  - Wikis, Q&A sites, discussion forums
  - User-created and maintained discussions
  - Wealth of knowledge!
Communities in Knowledge-Sharing Networks

- More than repositories for knowledge
  - Community structure surrounding discussions
  - Multiple topics, multiple communities
Communities in Knowledge-Sharing Networks

- More than repositories for knowledge
  - Community structure surrounding discussions
  - Multiple topics, multiple communities

- This study:
  - Communities in knowledge-sharing networks
Communities in Knowledge-Sharing Networks

- More than repositories for knowledge
  - Community structure surrounding discussions
  - Multiple topics, multiple communities

- This study:
  - Communities in knowledge-sharing networks
  - Inter-community relationships according to user dynamics
Communities in Knowledge-Sharing Networks

• More than repositories for knowledge
  – Community structure surrounding discussions
  – Multiple topics, multiple communities

• This study:
  – Communities in knowledge-sharing networks
  – Inter-community relationships according to user dynamics
  – Temporal evolution of inter-community relationships
Communities in Knowledge-Sharing Networks

- Communities centered around topics
  - Topics are explicitly defined
  - Independent from social interaction graph
- Non-exclusive membership to multiple communities
Topic-Based Communities in Stack Overflow

All Questions

20 votes
6 answers
31k views

redirect stderr to stdout in c shell

When I run the following command in csh, I got nothing, but it works in bash, is there any equivalent in csh which can redirect the standard error to standard out? xxx 2>&1 Note: xxx is a ...

modified 1 min ago
mdiehl13
134 3 10

1 vote
0 answers
5 views

Is it possible to run a python command within a shell script?

I'm learning some basic scripting and I thought I would try my hand at "automating" a small task at work and make things a littler easier for users not so comfortable working inside their terminal. ...

asked 1 min ago
Jacktheyeti
8 3
Topic-Based Communities in Stack Overflow

Tags

**redirect stderr to stdout in c shell**
When I run the following command in csh, I got nothing, but it works in bash, is there any equivalent in csh which can redirect the standard error to standard out? xxx 2>&1 Note: xxx is a ... modified 1 min ago

**is it possible to run a python command within a shell script?**
I'm learning some basic scripting and I thought I would try my hand at "automating" a small task at work and make things a little easier for users not so comfortable working inside their terminal. ... asked 1 min ago
Stack Overflow Dataset

• User activity
  – User ID, Tag ID, Time stamp

• Data covering a six-year period
  – 2008–2014

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Tags</th>
<th>Posts</th>
<th>Users</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top-100</td>
<td>100</td>
<td>15.4 million</td>
<td>1.2 million</td>
</tr>
<tr>
<td>Top-400</td>
<td>400</td>
<td>19.8 million</td>
<td>1.7 million</td>
</tr>
</tbody>
</table>
Inter-Community User Flows

- How to measure the relationship between two communities?
Inter-Community User Flows

• How to measure the relationship between two communities?
  – Tag hierarchy
Inter-Community User Flows

- How to measure the relationship between two communities?
  - Tag hierarchy
  - Semantic similarity of keywords
Inter-Community User Flows

● How to measure the relationship between two communities?
  – Tag hierarchy
  – Semantic similarity of keywords
  – User dynamics

Flow of users between communities:
\[ c_1(t) = |C_2(t) \cap C_1(t-1)| / |C_1(t-1)| \]
Inter-Community User Flows

- How to measure the relationship between two communities?
  - Tag hierarchy
  - Semantic similarity of keywords
  - User dynamics

- Flow of users between communities:

\[
flow_{c_1, c_2}(t) = \frac{|C_2(t) \cap C_1(t - 1)|}{|C_1(t - 1)|}
\]
Inter-Community User Flows: Findings

- Pravelenence of lower flow values
  - $flow_{c_1,c_2} > 0.20$ for 25% and 10% of all community pairs
Inter-Community User Flows: Findings

- Prevalence of lower flow values
  - \( \text{flow}_{c_1, c_2} > 0.20 \) for 25% and 10% of all community pairs
- Increasing number of low flow values over time
  - \( \text{flow}_{c_1, c_2}(2014) > 0.14 \) for 22% and 10% of all community pairs

### Mean Flow Value

<table>
<thead>
<tr>
<th>Dataset</th>
<th>2008</th>
<th>2014</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top-100</td>
<td>0.21</td>
<td>0.08</td>
</tr>
<tr>
<td>Top-400</td>
<td>0.14</td>
<td>0.05</td>
</tr>
</tbody>
</table>
Inter-Community User Flows: Findings

- Prevalence of lower flow values
  \[ \text{flow}_{c_1,c_2} > 0.20 \] for 25% and 10% of all community pairs

- Increasing number of low flow values over time
  \[ \text{flow}_{c_1,c_2}(2014) > 0.14 \] for 22% and 10% of all community pairs

- Greater variability of flow values over time

<table>
<thead>
<tr>
<th>Mean Coefficient of Variation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dataset</td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td>Top-100</td>
</tr>
<tr>
<td>Top-400</td>
</tr>
</tbody>
</table>
Flow Evolution: CSS to Javascript

CSS

0.75

0.55

JS

2008
Flow Evolution: CSS to Javascript

CSS  0.75  JS

CSS  0.55  JS

2008

CSS  0.76  JS

CSS  0.55  JS

2014
Flow Evolution: Flash to HTML

2009

Flash

HTML

0.43

0.12

2009
Flow Evolution: Flash to HTML

2009

Flash

HTML

0.43

0.12

2014

Flash

HTML

0.29

0.03
Flow Evolution: Active Record to Ruby on Rails

AR 0.61 0.25 Rails 3

2010
Flow Evolution: Active Record to Ruby on Rails

AR → Rails 3
0.61
0.25
2010

AR → Rails
0.41
0.26
2014
Flow Evolution: Active Record to Ruby on Rails

- AR to Rails 3: 0.61
- Rails 3 to AR: 0.25

- AR to Rails: 0.41
- Rails to AR: 0.26

2010

- AR to Rails 4: 0.45
- Rails 4 to AR: 0.28

2014
Macro-Communities

- Increasingly well-defined inter-community relationships
- Groups of communities with high inter-community flows
  - Determined by user dynamics instead of semantic analysis
- Clique Percolation Method applied over community graph
  - Communities as nodes and user flows as edge weights
  - Top 10% edges with highest flow values
  - Community overlap
Macro-Communities

iOS
Apple
Programming
Java
Windows
windows

Programming

iOS
Apple
Programming
Java
Windows
windows
Macro-Communities

- Small number of macro-communities
  - Popular communities connected to several satellite communities
  - e.g., Java connected to 88 communities in the Top-100 set

- Topical cohesion
  - General programming, Windows-related technologies, Apple-related technologies, Ruby on Rails technologies, programming IDEs and extensions
Evolution of Macro-Communities

• Changes to macro-community structure over time:

Top-100

Top-400
Evolution of Macro-Communities

- Single, dominant macro-community
  - Present at every time window
  - Covers a majority of the communities
  - General, more popular topics

- Fragmented macro-communities converge over time

- Recurring core of communities

- Variable community composition
  - Macro-communities may feature a different subset of communities in each time period
Conclusions

- Knowledge-sharing networks as a dynamics multi-community network
- Inter-community relationships as a function of their members
  - Different evolution patterns for community relationships
- Discovery of macro-communities
  - User flows
Thank you!

Anna Guimarães
anna@dcc.ufmg.br