Functional Roles and Career Paths in Wikipedia

Ofer Arazy\textsuperscript{1,2}, Felipe Ortega\textsuperscript{3}, Oded Nov\textsuperscript{4}, Lisa Yeo\textsuperscript{5}, Adam Balila\textsuperscript{2}

1. University of Alberta
2. University of Haifa
3. University Rey Juan Carlos
4. NYU Polytechnic School of Engineering
5. Loyola University Maryland

Wikimedia Research Showcase
January 14, 2015
Motivation
Online production communities

- **Social roles: structure + culture.**
  - **Core ↔ periphery** (network).
    - Sparse, unconnected periphery of contributors who are not very active.
    - Dense, cohesive core of very active contributors who take on additional responsibilities.

- **Activity patterns** (organization).
  - Many prior studies characterize the position of contributors in the periphery-core continuum based on quantity and type of their activities.
Online production communities

- Model: Reader to leader framework (R2L).
- Preece & Shneiderman (2009).

ALL USERS

Main path

Secondary path
Functional roles in Wikipedia

• Filling the gaps.
  • Many previous works on:
    – Activity of Wikipedia editors.
    – Structure (core-periphery).
    – Evolution within community structure.
  • Less (or lack of) previous work on:
    – Formal roles and their associated functions.
    – Transitions between roles.
    – Validation of theoretical frameworks (such as R2L).
Research questions

1. How are the functional roles of Wikipedia characterized and what activity profile is associated with each role?

2. How has Wikipedia's functional composition evolved over time?

3. How do Wikipedia editors transition between roles on the path from community's periphery to the core?
Methodology and data collection
Wikipedia user access levels

Level 5
- Steward 2002
- Oversight 2002
- Ombudsman Importer Feb.06

Level 4
- Checkuser Aug.02
- Sysop Jan.02
- Bureaucrat Jan.03

Level 3
- Accountcreator May08
- Abusefilter Mar.09
- Reviewer Autoreviewer Jun.09
- Reviewer Jan.08
- Filemover Aug.11

Level 2
- Ipblock-exempt May08
- Autoconfirmed Jun.12

Level 1
- User Jan.01

Wikipedia access privileges timeline
<table>
<thead>
<tr>
<th>ORG. LEVEL</th>
<th>FUNCTIONAL ROLE</th>
<th>WIKIPEDIA USER LEVEL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 5</td>
<td>Privacy Commissioner</td>
<td>ombudsman</td>
</tr>
<tr>
<td></td>
<td>Directors</td>
<td>steward; importer &amp; transwiki</td>
</tr>
<tr>
<td>Level 4</td>
<td>Security Force</td>
<td>oversight; checkuser</td>
</tr>
<tr>
<td>Level 3</td>
<td>Administrators</td>
<td>sysop; bureaucrat</td>
</tr>
<tr>
<td></td>
<td>Technical Administration</td>
<td>filemover; accountcreator</td>
</tr>
<tr>
<td></td>
<td>Border Control</td>
<td>rollback</td>
</tr>
<tr>
<td></td>
<td>Quality Assurance</td>
<td>reviewer; autoreviewer</td>
</tr>
<tr>
<td></td>
<td>QA Technicians</td>
<td>abusefilter</td>
</tr>
<tr>
<td>Level 2</td>
<td>Registered Users</td>
<td>user; autoconfirmed</td>
</tr>
<tr>
<td></td>
<td>Manually Registered Users</td>
<td>confirmed; IPblock-exempt</td>
</tr>
</tbody>
</table>
Methodology

- **Quantitative.**
  - Dump files ([http://dumps.wikimedia.org](http://dumps.wikimedia.org)).
    - Revision history, logging table (administrative actions).
    - Automated workflow: WikiDAT ([on GitHub](https://github.com/WikiDAT/wiki-dat)).
  - Wikipedia API (user profiles, role transitions).
  - Other sources (e.g. The NoSeptember Admin Project).

- **Qualitative.**
  - 17 editors who capture various profiles and role transition patterns.
  - User pages, talk pages, RfAs and de-adminship, dispute resolution pages.
Data sample

- Representative sample of editors and articles.
  - 93 articles that reflect all top-level Wikipedia categories.
  - 2.174 (registered) editors who have participated in the creation of these articles.
    - Anonymous editors and bots excluded.
  - 1.103 transitions between roles.
Results
RQ1: Characterizing functional roles
RQ2: Evolution of functional roles

New technical and content curation profiles emerge spontaneously, aside from bureaucratic tasks.

Stable number of administrators, few new promotions.

New technical and content curation profiles emerge spontaneously, aside from bureaucratic tasks.
RQ3: Role transitions
RQ3: Sources incoming transitions L2

- Technical Administration [24]
- Border Patrol [139]
- Quality Assurance [302]
- QA Technicians [59]

Legend:
- Red: Level 3&up: demotion
- Yellow: Level 3&up: adding roles
- Light Blue: Level 2: QA Tech
- Blue: Level 2: QA
- Dark Blue: Level 2: Border Patrol
- Green: Level 2: Tech Admin
- Green: Level 1
Conclusions and future work
Lessons learned

- Functional roles influence activity patterns across editing tasks (namespaces).
  - To date, the opposite approach was usually followed to determine organizational structure.
- Introduction of Level 2 roles (content curation).
  - Previous works emphasized increasing bureaucracy.
- Non-sequential paths are the rule rather than exceptions.
  - Against predictions of primary path in R2L framework.
Conclusions and future steps

- Complex dynamics of role transitions in Wikipedia.
  - Characterization of demotions.
- Future: understanding career paths.
  - Multi-method approach.
  - Include automated agents.
  - Re-examine evolution of roles.
    - For instance, sysop.
- Broader period, larger sample, other communities.
Publications


Main title: Reflexology paths, by alantankenghoe on Flickr

Motivation: What's your motivation, by opensource.com on Flickr

Research questions: Question Box, by Raymond Bryson on Flickr

Methodology and data collection: Starting to design an application, by Juhan Sonin on Flickr

Results: Cross Moleskine, by Irish Typepad on Flickr

Conclusions: Curves ahead, by Jeremy Bronson on Flickr

Future steps: One old compass, by Olga Filonenko, on Flickr