Coordination and Efficiency in Decentralized Collaboration

Alexander Scholl
Konstanz, 01.02.2017
Research Questions

• Do high-status projects manage the coordination trade-off differently from typical projects?

• How does coordination relate to a project’s team composition and crowdedness—in particular, the amount of work produced relative to the number of members in the team?

• Goal: Help with design decisions
Part 1: Coordination in high-status projects
Data

• Wikipedia and GitHub data, for the convenience of this presentation, only the Wikipedia data will be treated
• 3,4 million articles with discussion pages (Basically English history of Wikipedia from the beginning until April 2007)
Variables of Interest

• *Discussion edits*: Explicit measure of how much effort editors spend explicitly coordinating

• *Comments left on article edits*: More implicit measure of coordination efforts

• The $x - core$ (will be explained later)
Is this a valid measure of coordination effort?

• While these tools were meant for coordination, they can be used for lots of other activities (social interaction, trolling, spam etc.)

• Therefore, a random sample of 100 comments is used to create categories of interaction

• Then, again a random sample of 100 comments is drawn and those are then sorted in the defined categories

• The same is done for discussion edits
Results support that Comments and Discussion Edits are a valid measure for coordination effort.

### Comments

<table>
<thead>
<tr>
<th>Category</th>
<th>Num</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mentions section</td>
<td>52</td>
</tr>
<tr>
<td>Reverted edit</td>
<td>14</td>
</tr>
<tr>
<td>Minor edit</td>
<td>19</td>
</tr>
<tr>
<td>Added content</td>
<td>14</td>
</tr>
<tr>
<td>Removed content</td>
<td>7</td>
</tr>
<tr>
<td>Correction</td>
<td>2</td>
</tr>
<tr>
<td>Mentions other users</td>
<td>14</td>
</tr>
<tr>
<td>Other</td>
<td>11</td>
</tr>
</tbody>
</table>

### Discussion Edits

<table>
<thead>
<tr>
<th>Category</th>
<th>Num</th>
</tr>
</thead>
<tbody>
<tr>
<td>Justify</td>
<td></td>
</tr>
<tr>
<td>Text edit</td>
<td>14</td>
</tr>
<tr>
<td>Change metadata</td>
<td>4</td>
</tr>
<tr>
<td>Suggest Action</td>
<td></td>
</tr>
<tr>
<td>Specific text edit</td>
<td>9</td>
</tr>
<tr>
<td>Add content</td>
<td>13</td>
</tr>
<tr>
<td>Remove content</td>
<td>4</td>
</tr>
<tr>
<td>Change metadata</td>
<td>7</td>
</tr>
<tr>
<td>References</td>
<td></td>
</tr>
<tr>
<td>Provide</td>
<td>4</td>
</tr>
<tr>
<td>Request</td>
<td>16</td>
</tr>
<tr>
<td>Question</td>
<td></td>
</tr>
<tr>
<td>On article’s topic</td>
<td>8</td>
</tr>
<tr>
<td>On Wikipedia conventions</td>
<td>5</td>
</tr>
<tr>
<td>Copyright issues</td>
<td>8</td>
</tr>
<tr>
<td>Dispute claim in article</td>
<td>12</td>
</tr>
<tr>
<td>General discussion about article’s direction</td>
<td>8</td>
</tr>
<tr>
<td>Other</td>
<td>6</td>
</tr>
</tbody>
</table>
The $x – core$

• In projects that involve more than one person, both online and offline, there is often the finding, that a certain part of the group does most of the work

• To reflect this fact the authors establish the $x – core$, the percentage of people that have done such a percentage of the overall work

• When $x = 1$, everyone is included in the group of authors, when $x = 0$ no one is included

• The $x – core$ is used as a continuous function for $x \in (0,1)$, they never choose a certain $x$
Comparing Featured and Non-Featured Articles

• The making of featured Articles has different properties due to the instream of attention, triggered by the nomination process and later on the status associated with being a featured article

  => Simple Comparison therefore not useful

• Solution: Comparison set is constructed, which shall have the same distribution of certain traits, that ensure observed differences are not the consequence of the process. Traits:
  • Number of Edits
  • Stage of Development
  • Stage of Wikipedia as a whole
Featured
Non-Featured
Featured

Non-Featured
Part 2: Coordination in Crowded Environments
Variables of Interest

• Users who have made at least one edit to the article and one edit to the discussion page
• \(a\): Article
• \(S\): User who contributed
• \(D\): Discussion edits
• \(N\): eventual size of the article in bytes
Hypotheses

• Considering a fixed amount of work, the amount of coordination will increase with team size.

• Corresponding to that, given a project with a fixed team size, the amount of coordination will decrease with amount of work produced.

• => Crowdedness leads to an increase in coordination
Heat map “amount of coordination” $\log(D_a)$
Overview - (Differences between the different fields are significant)
Part 3: Modelling Crowdedness and Coordination
Modelling Coordination and Crowdedness

• The model only captures the results from the crowdedness analysis, not the analysis about high-status projects
• The model is simple and stylized
• Not meant to be a predictor
• Remember: The goal of this paper is to help with design decisions.
Model

• $N$: # of Article parts, starts in unfinished state, needs one unit of work to be finished

• $E$: Set of Users who work on the project, one single step of one user equals one unit of work, every user has two actions

• When one user works on an unfinished article part, it is finished afterwards, when the part is already finished it becomes unfinished again with a probability of $\propto$.

• Coordinate: First action, finding unfinished article, second, finishing it

• Not-Cordinate: First and second action, finishing two randomly selected parts (Not-Cordinate is chosen with probability $\beta$)
The model written in formulas

• \( P_{i+1} = AP_i + P_0 \)
• \( A = \frac{1-\beta}{N^2} (1 + \alpha)^2 - 2 \frac{1-\beta}{N} (1 + \alpha) + 1 \)
• \( P_0 = -\frac{1-\beta}{N} (1 + \alpha) + 2 - \beta \)

Closed form solution:

• \( P_i = P_0 \left[ \frac{A^i - 1}{A - 1} \right] \) if \( A \neq 1 \) and \( P_i = iP_0 \) if \( A = 1 \)
Values obtained from simulations
\( \alpha = 1 \)
Values obtained from analytical approximation

\( \alpha = 1 \)
Values obtained from analytical approximation \( \propto = 0 \)
Results

• Best if $E > N$
• Whether the $\alpha$ is 0 or 1, the basic trend stays the same, crowded projects require more coordination
• Analytical solution and simulation yield the same result
Critique

Positive
- Clear results
- Good Visualizations
- Very understandable
- Comparison set

Negative
- No dynamic perspective (time not included)
- Strange structure of the paper
- No overall model (two different questions answered separately)
- GitHub data not really treated in the paper