Information Quality in Wikipedia: The Effects of Group Composition and Task Conflict

Nicolas Siebeck
16.11.2016
Outline

1. Research Questions
2. Research Method
3. Variables
4. Results
5. Discussion
1. Research Questions

– What determines the success of community-based collaboration projects?

– How do cognitive diversity, task conflict and group members’ orientation (administrative or content) interact to determine the quality of the articles produced by Wikipedia editor groups?
Hypotheses

1) Task conflict will moderate the relationship between cognitive diversity and information quality, such that when task conflict is high, cognitive diversity would have a positive effect on information quality, whereas when task conflict is low, cognitive diversity would have a negative effect on information quality.

2) Task conflict will be negatively related to information quality.

3) The more group members gravitate toward administrative orientation, the lower task conflict the group experiences.

4) The more group members gravitate toward content orientation, the higher the quality of the article the group produces.
2. Research Method

1) Theory development

2) Randomly sample Wikipedia articles from different categories

3) Quantitative data extraction from the Wikipedia logs
   1) Find suitable mathematical models to describe the variables

4) Manual data acquisition
   1) Wikipedia article quality assessment by senior librarians
   2) Discussion page analysis for information about task conflict

5) Calculate descriptive statistics

6) Calcualte PLS path model

7) Intrepretation of the results
Article Sampling

- Corpus Requirements
  - Article length between 200 and 3500 words.
  - Articles are evenly distributed among 6 general categories.
  - Article age more than one month
  - More than two editors.

- Articles were generated with the “Random Article” feature

- Total corpus size: 96 articles.
3. Variables

Author group

1. Group size
2. Group member’s orientation
3. Cognitive Diversity

produces

Group project

Talk page

Task conflict

Article

1. Information Quality
2. Length
3. Age
4. Activity

Red: Control variables
Group member’s orientation

- Measured as four different variables
- High values for administrative orientation and low values for content orientation

1) Proportion of “administrators” among the group members

2) Each authors’ average number of edits in the whole Wikipedia

3) Ratio of authors’ overall Wikipedia to article activity

4) Average of members’ entropy in Wikipedia article edits
   (based on the different articles the author has contributed to and the percentage of edits in it)
Cognitive Diversity

<table>
<thead>
<tr>
<th>group x</th>
<th>article 1</th>
<th>article 2</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>member 1</td>
<td>1</td>
<td>0</td>
<td>...</td>
</tr>
<tr>
<td>member 2</td>
<td>0</td>
<td>1</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

- 1/0 for active/inactive on this article
- Dense matrix → much overlap in members’ activity
- Sparse matrix → unique set of members’ activity
- Percentage of cells with zero indicates cognitive diversity
Task conflict

Talk page analysis

Task conflict is rated by three independent assessors on a seven-point Likert scale according to these statements:

- Participants often have conflict in opinions about what should be included in the article
- Participants frequently have disagreements about what information the article should include
- There exists substantial conflict of ideas among participants

Calculated intraclass correlation agreement ICC > 0.8

→ Average of the three assessors employed
Article Quality

1) Three independent assessors analyzed each article and rated their perception of the information quality attributes: accuracy, completeness, objectivity, and representation.

2) Conflicting views in the evaluations were discussed together reach a consensus.

3) For each article the assessors agreed on an article quality value on a seven-point Likert scale.
Control Variables

- Group size
  - Number of editors that contributed to the article.
  - More eyes on the article detect more errors and may affect information quality.
  - "Wisdom-of-the-crowd" may affect information quality and cognitive diversity.

- Article length
  - Number of words included on the article’s Wiki page.
  - May reflect the completeness of the article and therefore its quality.

- Article age
  - Number of days since the inception of the article.
  - Age may reflect the article’s maturity and could affect the article’s quality.

- Article activity
  - Total number of edits from the article’s history.
  - Relatively high editorial activity may be an indicator for good article quality.
# Descriptive Statistics

<table>
<thead>
<tr>
<th>Construct/makeure</th>
<th>Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group members' orientation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percent of administrators</td>
<td>0.15</td>
<td>0.113</td>
</tr>
<tr>
<td>Average of members’ Wikipedia edits</td>
<td>6,217.1</td>
<td>5,204.2</td>
</tr>
<tr>
<td>Ratio of members’ Wikipedia edits by article edits</td>
<td>3,552.9</td>
<td>3,084.8</td>
</tr>
<tr>
<td>Average of members’ entropy in Wikipedia article edits</td>
<td>4.89</td>
<td>1.32</td>
</tr>
<tr>
<td>Cognitive diversity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The sparsity of the editors to external articles matrix</td>
<td>0.92</td>
<td>0.077</td>
</tr>
<tr>
<td>Task conflict</td>
<td></td>
<td></td>
</tr>
<tr>
<td>“Participants often have conflicting opinions about what should be included in the article”</td>
<td>2.1</td>
<td>1.24</td>
</tr>
<tr>
<td>“Participants frequently have disagreements about what information the article should include”</td>
<td>2.1</td>
<td>1.23</td>
</tr>
<tr>
<td>“There exists substantial conflict of ideas among participants”</td>
<td>2.0</td>
<td>1.12</td>
</tr>
<tr>
<td>Group size</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of unique editors who have contributed to the article</td>
<td>49.2</td>
<td>70.4</td>
</tr>
<tr>
<td>Article length</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of words in the article</td>
<td>901.6</td>
<td>825.2</td>
</tr>
<tr>
<td>Article age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Days since the inception of the article</td>
<td>812.2</td>
<td>462.3</td>
</tr>
<tr>
<td>Article activity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of edits made to the article since its inception</td>
<td>90.9</td>
<td>125.0</td>
</tr>
<tr>
<td>Information quality</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raters’ consensus on the quality of the article</td>
<td>4.4</td>
<td>1.16</td>
</tr>
</tbody>
</table>
4. Results

- Correlation between the variables
- On the diagonal: Average amount of variance in indicator variables that a construct is managed to explain (AVE).

<table>
<thead>
<tr>
<th>Constructs</th>
<th>CD</th>
<th>GMO</th>
<th>TC</th>
<th>TC × CD</th>
<th>GS</th>
<th>AL</th>
<th>AAg</th>
<th>AAc</th>
<th>IQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive diversity (CD)</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group members’ orientation (GMO)</td>
<td>-0.41</td>
<td>0.89</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task conflict (TC)</td>
<td>0.49</td>
<td>-0.34</td>
<td>0.99</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task conflict × Cognitive diversity</td>
<td>-0.48</td>
<td>0.14</td>
<td>0.38</td>
<td>0.99</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group size (GS)</td>
<td>0.49</td>
<td>-0.31</td>
<td>0.63</td>
<td>0.32</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Article length (AL)</td>
<td>0.20</td>
<td>-0.15</td>
<td>0.44</td>
<td>0.27</td>
<td>0.48</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Article age (AAg)</td>
<td>0.37</td>
<td>0.15</td>
<td>0.30</td>
<td>0.08</td>
<td>0.47</td>
<td>0.18</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Article activity (AAc)</td>
<td>0.49</td>
<td>-0.33</td>
<td>0.67</td>
<td>0.35</td>
<td>0.98</td>
<td>0.50</td>
<td>0.44</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Information quality (IQ)</td>
<td>0.14</td>
<td>-0.30</td>
<td>0.24</td>
<td>0.26</td>
<td>0.45</td>
<td>0.36</td>
<td>0.15</td>
<td>0.46</td>
<td>1</td>
</tr>
</tbody>
</table>
Path Model

Figure 1. Results of PLS Analysis

Notes: Values on arrows represent path significance: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; ns = nonsignificant. $R^2$ percentages are in gray.
5. Discussion

<table>
<thead>
<tr>
<th>Critic points</th>
<th>Possible Improvements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample size of 96 articles is very small.</td>
<td>Bigger sample size</td>
</tr>
<tr>
<td>The indication for cognitive diversity seems weak: Any edit in an article has the same weighting.</td>
<td>Instead of manual inspection for task conflict and information quality, use sentiment analysis and Wikipedia’s article quality categories.</td>
</tr>
<tr>
<td>The cognitive diversity matrix approach:</td>
<td>Edits can be very different and should be considered so. An algorithm could be designed that detects what kind of edit has been done.</td>
</tr>
<tr>
<td>If all group members have few edits, the group is cognitively diverse.</td>
<td></td>
</tr>
<tr>
<td>A dense matrix suggests “overlap in members activity”. What if one member only corrects spelling?</td>
<td>Try to include more information about the authors. For example with a manual knowledge background check.</td>
</tr>
<tr>
<td>Critic points</td>
<td>Possible Improvements</td>
</tr>
<tr>
<td>-----------------------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>- Group members orientation as a rational value.</td>
<td>- Use background information about the authors. Differentiate between opinion, semantic and stylistic edits.</td>
</tr>
<tr>
<td>- Control variables seem arbitrary.</td>
<td>- Include variables related to the authors.</td>
</tr>
<tr>
<td>- Complicated statistical path model (PLS) for variable analysis.</td>
<td>- Use simple models or explain more.</td>
</tr>
<tr>
<td>- The task conflict rating statements are very similar.</td>
<td>- Differentiate between opinion, semantic and stylistic conflict</td>
</tr>
<tr>
<td>- Average article Quality is very high for random articles.</td>
<td>- Use Wikipedia quality categories (Featured, A-Class etc.)</td>
</tr>
<tr>
<td>- Explanation and interpretation of some minor issues.</td>
<td></td>
</tr>
</tbody>
</table>
Thank you for your attention!

Nicolas Siebeck
nicolas.siebeck@uni.kn