Emotions under Discussion: Gender, Status and Communication in Wikipedia

David Laniado
daavid.laniado@barcelonamedia.org

with Daniela Iosub, Carlos Castillo, Mayo Fuster Morell and Andreas Kaltenbrunner

Wikimedia Research Showcase, October 15, 2014
1 Introduction

2 Framework of analysis
   - Data acquisition and pre-processing
   - User gender labelling
   - Sentiment analysis

3 Results
   - Emotions and status
   - Emotions and gender
   - Networked emotions

4 Conclusions
Outline

1. Introduction

2. Framework of analysis
   - Data acquisition and pre-processing
   - User gender labelling
   - Sentiment analysis

3. Results
   - Emotions and status
   - Emotions and gender
   - Networked emotions

4. Conclusions
Introduction

- Wikipedia, largest collaborative project
- Discussion spaces are fundamental to the collaborative process
- Discussion triggers emotions and breeds particular emotional environments
Introduction

- Studying the emotional dimension can help to face issues such as the gender gap and editors’ decline

Interactions in Wikipedia
- Implicit $\rightarrow$ editing
- Explicit $\rightarrow$ communication in article talk pages and personal talk pages

Approach
- extensive analysis of emotions in explicit communication
- through sentiment analysis of comments in article talk and personal talk pages
Research questions

1. How are the emotional and communication styles of editors affected by their **status**?

2. How are the emotional and communication styles of editors affected by their **gender**?

3. How are the emotional expressions affected by interacting with others in comment threads (**emotional congruence**)?

4. How are the emotional styles of editors related to those of the editors they interact more frequently with (**emotional homophily**)?
Introduction

Results published in:

  Emotions and dialogue in a peer-production community: the case of Wikipedia.
  8th International Symposium on Wikis and Open Collaboration, WikiSym’12

  Emotions under Discussion: Gender, Status and Communication in Online Collaboration.
  Plos One, 9(8)
Outline

1. Introduction

2. Framework of analysis
   - Data acquisition and pre-processing
   - User gender labelling
   - Sentiment analysis

3. Results
   - Emotions and status
   - Emotions and gender
   - Networked emotions

4. Conclusions
Outline

1. Introduction

2. Framework of analysis
   - Data acquisition and pre-processing
   - User gender labelling
   - Sentiment analysis

3. Results
   - Emotions and status
   - Emotions and gender
   - Networked emotions

4. Conclusions
### Dataset: conversations
Extracting conversations among editors from the English Wikipedia

<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Articles</td>
<td>3,210,039</td>
<td></td>
</tr>
<tr>
<td>Articles with talk page (ATP)</td>
<td>871,485</td>
<td>27.1%</td>
</tr>
<tr>
<td>Editors who comment articles</td>
<td>350,958</td>
<td></td>
</tr>
<tr>
<td>Editors with ( \geq 100 ) comments on ATP</td>
<td>12,231</td>
<td>3.5%</td>
</tr>
<tr>
<td>Total comments in ATP</td>
<td>11,041,246</td>
<td></td>
</tr>
<tr>
<td>Comments containing ANEW words</td>
<td>7,414,411</td>
<td>67.2%</td>
</tr>
<tr>
<td>Comments by editors with ( \geq 100 ) comments on ATP</td>
<td>5,480,544</td>
<td>49.6%</td>
</tr>
<tr>
<td>Comments by these editors and with ANEW words</td>
<td>3,649,297</td>
<td>33.3%</td>
</tr>
</tbody>
</table>

**Table**: Data extracted from a complete dump of the English Wikipedia, dated March 12th, 2010
Outline

1. Introduction

2. Framework of analysis
   - Data acquisition and pre-processing
   - User gender labelling
   - Sentiment analysis

3. Results
   - Emotions and status
   - Emotions and gender
   - Networked emotions

4. Conclusions
User gender labelling

- ≈ 12 000 users wrote ≥ 100 comments in articles talk pages
- Gender identified through Wikipedia API for ≈ 2 000 of them
- A sample of 1 385 users for manual labelling through crowdsourcing (Crowdflower)

<table>
<thead>
<tr>
<th></th>
<th>Non-admins</th>
<th>Admins</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men</td>
<td>1 087</td>
<td>1 526</td>
<td>2 613</td>
</tr>
<tr>
<td>Women</td>
<td>68</td>
<td>97</td>
<td>165</td>
</tr>
<tr>
<td>Unknown</td>
<td>6 850</td>
<td>2 603</td>
<td>9 453</td>
</tr>
<tr>
<td>Total</td>
<td>8 005</td>
<td>4 226</td>
<td>12 231</td>
</tr>
</tbody>
</table>

Table: Users with ≥ 100 comments by gender and administrator status.

- Gender could be identified only for ≈ 50% of users:
  - real name or username (50% of those identified)
  - implicitly stated gender (27% of women, 20% of men)
  - pronoun (15% of women, 10% of men)
  - other indicators: userboxes, pictures, links to personal blogs...
1 Introduction

2 Framework of analysis
   - Data acquisition and pre-processing
   - User gender labelling
   - Sentiment analysis

3 Results
   - Emotions and status
   - Emotions and gender
   - Networked emotions

4 Conclusions
Measuring the Emotional Content of Discussions

- relying on three different instruments:
  - Affective norms for English words (ANEW)
  - Linguistic Inquiry and Word Count (LIWC)
  - SentiStrength
Measuring the Emotional Content of Discussions
Method 1: Affective norms for English words (ANEW)

Rates a list of 1060 frequent words on a 9 point scale in three dimensions:

- Valence
- Arousal
- Dominance

assign emotion scores to each word from the lexicon

Measuring the Emotional Content of Discussions
Method 2: Linguistic Inquiry and Word Count (LIWC)

- Two scores for basic emotion (compared with ANEW valence)
  - positive emotion
  - negative emotion
- Discrete measures of emotions (anger, anxiety, sadness, affect)
- Other classes of words to characterize language (i.e. personal pronouns, tentative words, fillers...)

→ Count the proportion of words belonging to each class

The development and psychometric properties of LIWC2007. Austin, TX.
Measuring the Emotional Content of Discussions
Method 2: Linguistic Inquiry and Word Count (LIWC)

<table>
<thead>
<tr>
<th>Category</th>
<th>Dictionary size</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger</td>
<td>91</td>
<td>hate, kill, annoyed</td>
</tr>
<tr>
<td>Anxiety</td>
<td>84</td>
<td>worried, fearful, nervous</td>
</tr>
<tr>
<td>Sadness</td>
<td>101</td>
<td>crying, grief, sad</td>
</tr>
<tr>
<td>Tentative</td>
<td>155</td>
<td>maybe, perhaps, guess</td>
</tr>
<tr>
<td>Certainty</td>
<td>83</td>
<td>always, never</td>
</tr>
<tr>
<td>Fillers</td>
<td>9</td>
<td>blah, you know</td>
</tr>
<tr>
<td>Past</td>
<td>155</td>
<td>went, ran, had</td>
</tr>
<tr>
<td>Present</td>
<td>169</td>
<td>is, does, hear</td>
</tr>
<tr>
<td>Future</td>
<td>48</td>
<td>will, gonna</td>
</tr>
<tr>
<td>Social words</td>
<td>455</td>
<td>mate, talk, child</td>
</tr>
</tbody>
</table>

Table: Description of LIWC measures (as per [http://www.liwc.net](http://www.liwc.net)).
Measuring Relationship-Orientation with LIWC

**Definition**
- Communication that promotes social affiliation and emotional connection:
  - preoccupation with others (use of personal pronouns, e.g., I, you)
  - preoccupation with the larger social domain (e.g., references to friends and family)
  - expression of positive emotion

**Examples**
- We are glad to have you. If I can help at all let me know :) 
- A-giau has smiled at you. Smiles promote WikiLove and hopefully this one has made your day better...Happy editing
- Congrats! Thank you for your dedication.
Measuring the Emotional Content of Discussions
Method 3: SentiStrength

SentiStrength

- Based on LIWC and developed for short web texts
- Accounts for modes of textual expression specific to the online environment, e.g. emoticons and abbreviations
- Provides a positive and a negative score for emotional valence
- Emotion score is the strongest positive and negative emotion expressed in a comment
- Final scores are averages over comments in a given category

Sentiment strength detection in short informal text.
*Journal of the American Society for Information Science and Technology* 61: 2544 – 2558.
Table: Example messages with their corresponding Valence (ANEW) or positive & negative scores (LIWC, SentiStrength)

<table>
<thead>
<tr>
<th></th>
<th>ANEW</th>
<th>LIWC</th>
<th>SentiSt.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Valence</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Sounds like a good challenge - to be proven or disproven. I'm happy if it can be shown to go further using closed cubic polynomial solutions. The nice thing about these are that they are pretty easy to test numerically . . .</td>
<td>7.4</td>
<td>12.5</td>
<td>0</td>
</tr>
<tr>
<td>–in “Exact trigonometric constants”</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seems you have not yet seen female lover after having sex who do not wish to have sex with the same lover any more :) Once you've seen it, you understand very well what war of Venus means compared to war of Mars.</td>
<td>5.5</td>
<td>6.8</td>
<td>4.5</td>
</tr>
<tr>
<td>–in “House (astrology)”</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>What about the whirlie hazing, the alcohol abuse, the emotional poverty, the suicide in 1995/6, the biotech plans which were stopped by pitzer protests . . .</td>
<td>1.6</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>–in “Harvey Mudd College”</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Sentiment analysis

Statistical tests

- Compute average values with the three lexica for each user
- Compare distribution of values for two groups of users (e.g.: admins vs regulars, women vs men)
  - Most variables are not normally distributed

\[\downarrow\]

Mann-Whitney U-test

Compare distributions of rankings
Outline

1. Introduction

2. Framework of analysis
   - Data acquisition and pre-processing
   - User gender labelling
   - Sentiment analysis

3. Results
   - Emotions and status
   - Emotions and gender
   - Networked emotions

4. Conclusions
Outline

1. Introduction

2. Framework of analysis
   - Data acquisition and pre-processing
   - User gender labelling
   - Sentiment analysis

3. Results
   - Emotions and status
   - Emotions and gender
   - Networked emotions

4. Conclusions
Table: Emotions and Status: Administrators promote a generally neutral tone on article talk pages. Regular editors express more negative emotion, and are more emotional.

<table>
<thead>
<tr>
<th>(Article Talk)</th>
<th>Regular</th>
<th>Admin</th>
<th>Mann-Whitney U-Test</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>LIWC</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>2.369</td>
<td>2.409</td>
<td>-4.308</td>
<td>p &lt; 0.001</td>
</tr>
<tr>
<td>Negative</td>
<td>1.368</td>
<td>1.120</td>
<td>-18.578</td>
<td>p &lt; 0.001</td>
</tr>
<tr>
<td>Affect</td>
<td>3.784</td>
<td>3.661</td>
<td>-8.466</td>
<td>p &lt; 0.001</td>
</tr>
<tr>
<td>Anxiety</td>
<td>0.180</td>
<td>0.166</td>
<td>-5.834</td>
<td>p &lt; 0.001</td>
</tr>
<tr>
<td>Anger</td>
<td>0.554</td>
<td>0.446</td>
<td>-19.217</td>
<td>p &lt; 0.001</td>
</tr>
<tr>
<td>Sadness</td>
<td>0.175</td>
<td>0.166</td>
<td>-4.450</td>
<td>p &lt; 0.001</td>
</tr>
<tr>
<td><strong>SentiStrength</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>1.805</td>
<td>1.774</td>
<td>-14.603</td>
<td>p &lt; 0.001</td>
</tr>
<tr>
<td>Negative</td>
<td>-2.005</td>
<td>-1.912</td>
<td>-23.046</td>
<td>p &lt; 0.001</td>
</tr>
</tbody>
</table>

When difference is statistically significant (p-value in bold) the larger absolute value is underlined.
Emotions and Status

Admins:
- more positive emotion (ANEW and LIWC)
- generally, emotionally reserved compared to regular users (LIWC)

Regular users:
- more emotional
  - more affect, and more anxiety, anger and sadness (LIWC)
- stronger positive and negative words than admins (SentiStrength)

Personal talk pages
- In personal talk pages, admins are more emotional compared to the article talk pages
  - more positive emotion compared to regular editors, but also more anxiety and sadness
Table: Dialogue and Status: Administrators are more impersonal in article talk pages. Regular editors are more concerned with others.

<table>
<thead>
<tr>
<th>(Article Talk)</th>
<th>Regular</th>
<th>Admin</th>
<th>Mann-Whitney U-test</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relationship-orientation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Personal pronouns</td>
<td>5.135</td>
<td>4.815</td>
<td>-13.561</td>
<td>p &lt; 0.001</td>
</tr>
<tr>
<td>Use of “I”</td>
<td>2.456</td>
<td>2.429</td>
<td>-1.733</td>
<td>p=0.083</td>
</tr>
<tr>
<td>Use of “You”</td>
<td>1.043</td>
<td>0.892</td>
<td>-12.573</td>
<td>p &lt; 0.001</td>
</tr>
<tr>
<td>Use of “Shehe”</td>
<td>0.609</td>
<td>0.526</td>
<td>-8.657</td>
<td>p &lt; 0.001</td>
</tr>
<tr>
<td>Social words</td>
<td>6.320</td>
<td>5.810</td>
<td>-19.013</td>
<td>p &lt; 0.001</td>
</tr>
<tr>
<td>Certainty</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Certainty</td>
<td>1.426</td>
<td>1.317</td>
<td>-16.824</td>
<td>p &lt; 0.001</td>
</tr>
<tr>
<td>Tentativeness</td>
<td>3.199</td>
<td>3.169</td>
<td>-2.210</td>
<td>p &lt; 0.001</td>
</tr>
<tr>
<td>Filler words</td>
<td>0.168</td>
<td>0.155</td>
<td>-6.687</td>
<td>p &lt; 0.001</td>
</tr>
<tr>
<td>Temporal Orientation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Past</td>
<td>2.376</td>
<td>2.305</td>
<td>-5.696</td>
<td>p &lt; 0.001</td>
</tr>
<tr>
<td>Present</td>
<td>8.011</td>
<td>7.841</td>
<td>-8.060</td>
<td>p &lt; 0.001</td>
</tr>
<tr>
<td>Future</td>
<td>1.114</td>
<td>1.166</td>
<td>-9.887</td>
<td>p &lt; 0.001</td>
</tr>
</tbody>
</table>

When difference is statistically significant (p-value in bold) the larger absolute value is underlined.
## Dialogue and Status

<table>
<thead>
<tr>
<th>Admins</th>
<th>Regular users</th>
</tr>
</thead>
<tbody>
<tr>
<td>- more neutral and impersonal tone</td>
<td>- more relationship-oriented</td>
</tr>
<tr>
<td>- less relationship oriented</td>
<td>- more personal pronouns and more social words</td>
</tr>
<tr>
<td>- more concerned with the future</td>
<td>- more concerned with past</td>
</tr>
<tr>
<td>- tend to &quot;rule with reason&quot;</td>
<td>- more insecure, but not in personal spaces</td>
</tr>
<tr>
<td><img src="image1.jpg" alt="Admins Image" /></td>
<td>- more certainty, tentative and filler words in article talk pages</td>
</tr>
<tr>
<td><img src="image2.jpg" alt="Regular Users Image" /></td>
<td></td>
</tr>
</tbody>
</table>

@sdivad

Emotions under Discussion: Gender, Status and Communication in Wikipedia 27 / 43
Outline

1. Introduction

2. Framework of analysis
   - Data acquisition and pre-processing
   - User gender labelling
   - Sentiment analysis

3. Results
   - Emotions and status
   - Emotions and gender
   - Networked emotions

4. Conclusions
ANEW Words more used by women and men

Size accounts for difference in frequency
Women use words associated to more positive emotions

- Result consistent and significant with the three lexicons
- ANEW: Difference is not significant when normalising by article
  → difference might be due to topic selection: women choose to participate in topics which have more positive discussions

- No significant difference in expression of negative emotions
Topics, emotions and gender

Figure: Mean valence (ANEW) for discussions of articles in different topic categories, vs the proportion of comments written by men
Table: Dialogue and Gender: Women use a more relationship-oriented speech style.

<table>
<thead>
<tr>
<th>(Article Talk)</th>
<th>Men</th>
<th>Women</th>
<th>Mann-Whitney U-test</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relationship-orientation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Personal pronouns</td>
<td>4.964</td>
<td>5.420</td>
<td>-4.375</td>
<td><strong>p &lt; 0.001</strong></td>
</tr>
<tr>
<td>Use of “I”</td>
<td>2.488</td>
<td>2.764</td>
<td>-3.945</td>
<td><strong>p &lt; 0.001</strong></td>
</tr>
<tr>
<td>Use of “You”</td>
<td>0.936</td>
<td>0.957</td>
<td>-0.926</td>
<td>p=0.355</td>
</tr>
<tr>
<td>Use of “Shehe” pronouns</td>
<td>0.541</td>
<td>0.713</td>
<td>-4.657</td>
<td><strong>p &lt; 0.001</strong></td>
</tr>
<tr>
<td>Social words</td>
<td>5.960</td>
<td>6.353</td>
<td>-3.487</td>
<td><strong>p &lt; 0.001</strong></td>
</tr>
<tr>
<td>Certainty</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Certainty</td>
<td>1.346 (1397)</td>
<td>1.300 (1263)</td>
<td>-2.078</td>
<td><strong>p = 0.038</strong></td>
</tr>
<tr>
<td>Tentativeness</td>
<td>3.150</td>
<td>3.215</td>
<td>-1.162</td>
<td>p=0.245</td>
</tr>
<tr>
<td>Filler words</td>
<td>0.161</td>
<td>0.160</td>
<td>-0.137</td>
<td>p=0.891</td>
</tr>
<tr>
<td>Temporal Orientation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Past</td>
<td>2.325</td>
<td>2.543</td>
<td>-4.305</td>
<td><strong>p &lt; 0.001</strong></td>
</tr>
<tr>
<td>Present</td>
<td>7.897</td>
<td>8.180</td>
<td>-3.086</td>
<td><strong>p = 0.002</strong></td>
</tr>
<tr>
<td>Future</td>
<td>1.168</td>
<td>1.147</td>
<td>-1.008</td>
<td>p=0.314</td>
</tr>
</tbody>
</table>

When difference is statistically significant (p-value in bold) the larger absolute value is underlined. Cases where the averages are not informative are marked with an asterisk * and include the mean ranks Mann-Whitney U-test next to the averages in parentheses.
Women write longer messages

Women are more relationship-oriented
  - more personal pronouns, in particular “I”, more social words

Women are not more insecure
  - Less certainty words, no significant difference for tentativeness and filler words

Women admins are more relationship oriented than men admins
  - Different leadership style
Manual classification of 100 comments

Three main types of comments high in relationship-orientation:

- inviting comments that explain the edit in a friendly tone, and call for further intervention and collaboration
- common perspective-building comments that are focused on understanding others and solving debates in a constructive manner
- appreciative comments that contain positive emotions and celebrate others’ actions

⇒ This suggests that relationship-orientation may be conducive to successful collaboration
Outline

1. Introduction

2. Framework of analysis
   - Data acquisition and pre-processing
   - User gender labelling
   - Sentiment analysis

3. Results
   - Emotions and status
   - Emotions and gender
   - Networked emotions

4. Conclusions
Emotional congruence

- Comparison of each comment with the comment it replies to
  - not based only on our set of users, but on all comments (from all users)

### Emotions: editors tend to reply with:
- more positive emotion
- less negative emotion
- less anger, anxiety and sadness
- stronger words, both positive and negative (SentiStrength)

### Dialogue: editors tend to reply with:
- more relationship oriented speech
- less tentative and certainty words
**Emotional homophily**

Mixing patterns: do users interact preferentially with similar users?

<table>
<thead>
<tr>
<th>Disassortativity by activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>users who write more comments tend to reply preferentially to less active users, and vice versa</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Assortativity by gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men interact more with other men, and women with other women</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Assortativity by emotion and language</th>
</tr>
</thead>
<tbody>
<tr>
<td>Users interact more with others similar in emotional expression and communication style</td>
</tr>
<tr>
<td>also in the network of communication on personal talk pages</td>
</tr>
</tbody>
</table>
Emotional homophily

Example: homophily by expression of anger

- **edges** connect users who have exchanged at least 10 replies.
- **node color** represents the level of anger expressed by a user, from **low** to **high**.
- **node size** → proportional to the number of connections of a user.
Outline

1. Introduction

2. Framework of analysis
   - Data acquisition and pre-processing
   - User gender labelling
   - Sentiment analysis

3. Results
   - Emotions and status
   - Emotions and gender
   - Networked emotions

4. Conclusions
Conclusions

- Administrators and experienced users play a pivotal role
  - they tend to interact especially with less experienced users
  - they promote a positive and impersonal environment

- Women have a different communication style
  - they work on topics where discussions have a more positive tone
  - they interact more with each other than with men editors
  - they are more concerned with others (more relationship-oriented)
  - also women administrators have a different leadership style
    ⇒ promoting relationship orientation leadership could lead to a more positive environment
    ⇒ being able to give women more space in the community could result in a virtuous cycle of women participation
Future work

**Longitudinal analysis**

- how do emotional styles of editors change over time and with increasing experience?
- how do emotions in the discussion affect participation?

**Need for qualitative analysis and human annotation**

- include non-textual emotional aspects such as emoticons, *barn stars* and virtual gifts
- cover also less experienced users
- deal with sarcasm, measure the extent of condescending or paternalistic language in comments addressed at women editors
Some references

M. M. Bradley and P. J. Lang.
Affective norms for English words (ANEW) Technical report C-1.

B. Collier and J. Bear.
Conflict, criticism, or confidence: an empirical examination of the gender gap in wikipedia contributions.
In *Proc. of CSCW*, 2012.

Emotions under Discussion: Gender, Status and Communication in Online Collaboration.
*Plos One*, 9(8)

A large-scale sentiment analysis for Yahoo! answers.
In *Proc. of WSDM*, 2012.

In *Proc. of WikiSym*, 2011.

D. Laniado, R. Tasso, Y. Volkovich, and A. Kaltenbrunner.
When the Wikipedians talk: Network and tree structure of Wikipedia discussion pages.

Emotions and dialogue in a peer-production community: the case of Wikipedia.
*8th International Symposium on Wikis and Open Collaboration, WikiSym’12*

H. Zhu, R. Kraut, A. Kittur
Effectiveness of shared leadership in online communities.
In *Proc. of CSCW*, 2012.