CSS Research Lab

JProf. Dr. Claudia Wagner* †
Dr. Fabian Flöck*

*Computational Social Science @
GESIS – Leibniz Institute for the Social Sciences, Germany

† Web-Science @
University of Koblenz-Landau, Germany
Goal

- Team work
- Develop sth together and sell your project
- Context: CSS
Logistics

• Oct: Formation of Teams (~5 people per team)
• Oct/Nov: present project proposal and workplan

• Team Work
  – Use tools like basecamp/trello to organize your team work
  – Use github for collaborative development

• Monthly progress reports and meetings with me (or Fabian) on request

• Final presentation (February)
Last Year’s Project

Comparison between the German and Russian version of an article:

**Article Analysis**
from May 30, 2016

**Article:** [https://de.wikipedia.org/wiki/Krimkrise](https://de.wikipedia.org/wiki/Krimkrise)

**Heat Map**
Displays the general classification results.

**Article Analysis**
from May 30, 2016

**Article:** [https://ru.wikipedia.org/wiki/Крымский_кризис](https://ru.wikipedia.org/wiki/Крымский_кризис)

**Heat Map**
Displays the general classification results.

http://wikiwhere.west.uni-koblenz.de/about.php
Topics

People Networks

Resource Bias

Article Controversies

User Impact
People Networks

List of notable people from: Pantheon, Human Accomplishment, Freebase
Topology:
Gender Homophily

Men link to men *slightly more* than we expect
Women link to women *way more* than we expect

<table>
<thead>
<tr>
<th>Edges (M to M)</th>
<th>Edges (M to W)</th>
<th>Edges (W to M)</th>
<th>Edges (W to W)</th>
</tr>
</thead>
<tbody>
<tr>
<td>93.10%</td>
<td>6.90%</td>
<td>69.47%</td>
<td>30.53%</td>
</tr>
<tr>
<td>92.26%</td>
<td>7.74%</td>
<td>92.28%</td>
<td>7.72%</td>
</tr>
<tr>
<td>91.89%</td>
<td>8.11%</td>
<td>91.53%</td>
<td>8.47%</td>
</tr>
<tr>
<td>90.22%</td>
<td>9.78%</td>
<td>90.25%</td>
<td>9.75%</td>
</tr>
</tbody>
</table>

holding # nodes, gender ratio, avg. degree constant
People Networks in Wikipedia

- Hyperlink network defines the visibility of articles about people

- Directed hyperlink network between biographies (for certain sub-populations: gender, professions, birth decade)

- How do these networks change over time?
  - Visualize network (time slider, video)
  - Provide statistics that quantify change
  - Highlight unidirectional links → e.g. maybe men link less back to women but this has become much better over time
People Networks in Wikipedia

• Metadata about People in Dbpedia (profession, birth decade, country, gender)

• English Wikipedia; all revisions about all articles about people

• Build visualization
  – See examples here http://seealso.hatnote.com/
Resource Bias in Wikipedia

- What are the most frequently cited web resources in, e.g., the German Wikipedia?

- Apart from geographical/language bias (Wikiwhere), is there a bias to use resources from certain media outlets, e.g., subsidiaries of large conglomerates (Axel Springer, ARD, Fox, Time Warner,..) and/or with a certain political slant (most simple: left vs. right)?

- Use DBpedia/related knowledge bases and/or data of media slants to analyze and visualize resource bias.
  - News outlets hierarchy based on Dbpedia info: Y Owned by X
  - Can we find info about political bias of news outlets?
    - Option: Build a game for humans to categorize/group/tag resources
Gamergate controversy

From Wikipedia, the free encyclopedia

History [ edit ]

In February 2013, independent game developer Zoe Quinn released Depression Quest, an interactive fiction browser game, through the game’s website. Though the game was met positively by critics, a backlash developed among some gamers who believed that the game received an undue amount of attention, especially after a planned Steam distribution platform release. Quinn began to receive hate mail over the game upon its release and criticism from some parts of the Steam user community, receiving enough harassment to cause her to change her phone number.[1][2][3][4][5] By September 2014, Quinn had been the target of eighteen months of increasing harassment, which had created “an ambient hum of menace in her life, albeit one that she has mostly been able to ignore”.[6]

Shortly after the Steam release of Depression Quest in August 2013, Quinn’s former boyfriend, Eric Core, wrote a blog post, described by The New York Times as a “rambling online essay,”[7] alleging, among other things, that Quinn had an affair with Konami journalist Nathan Grayson[8]. Quinn’s detractors falsely alleged that the relationship had induced Grayson to publish a favorable review of the game.[8][9][10][11][12][13][14] Kotaku editor-in-chief, Stephen Totick, affirmed the existence of a relationship, but clarified that Grayson had not written anything about Quinn after the relationship had commenced, and had never reviewed her games, though he did acknowledge a piece written before the two began their relationship.[10][15] Commentators in and outside the gaming industry denounced the attack on Quinn as misogynistic and unfounded.[6][16][17]

As a result of these allegations, Quinn and her family were subjected to virulent, often misogynistic

Author List

- Masem 17.5%
- Ryulong 16.4%
- Tony Sidaway 13.2%
- NorthBySouthBaranof 8.3%
- The Devil’s Advocate 6.3%
- Strongjam 4.8%
- Krano 3.8%
- Aquillion 2.9%
- Hahnchen 2.8%
- Carwil 2.7%
- Bosstophur 2.4%
- TheRedPenOfDoom 2.4%
- Starship.paint 2.3%
- Kencof618 1.7%
- PresN 1.1%
- Torchiest 0.9%

whoCOLOR (→ WikiWho)
Wiki Conflict

WikiWho-Algorithm

For each word in an article (any revision):
When added, deleted, reintroduced, by whom, editing context, age
Wiki Conflict

Gamergate controversy

"Gamergate" redirects here. For the type of ant, see Gamergate. For the online video game store, see GamersGate.

Gamergate is a controversy regarding sexism in video game culture. It garnered significant public attention after August 2014 when several women within the video game industry, primarily game developers Zoe Quinn and Brianna Wu and feminist cultural critic Anita Sarkeesian, were subjected to a sustained campaign of misogynistic attacks. These attacks, initially performed under the Twitter hashtag #gamergate, were later variously coordinated in the online forums of Reddit, 4chan, and 8chan in an anonymous and amorphous movement. The harassment included doxing, threats of rape, and death threats, including a threat of a mass shooting at a university speaking event.

Gamergate is widely viewed as a manifestation of a culture war over gaming culture diversification, artistic recognition and social criticism of video games, and the gamer social movement by opposing social criticism in particular feminism and the associations with #Media, #Wirea, #Der Buno, and #Nintandebunked, calling them either #NotYourShield or other hashtags.

Many Gamergate supporters have taken issue with the widespread description of their movement as misogynistic, saying that the media focus on misogyny served mainly to "deflect criticism of the increasingly leftist orientation of indie games".[8] To respond to this criticism, a second Twitter hashtag, #NotYourShield, began to be used, with the intention of showing that women and other minorities in the gaming community were also critical of Quinn and Sarkeesian.[8][18][142]
Considerable evidence exists of Armstrong’s systematic doping use from 1998 to 2002.

Considerable circumstancial evidence of …

Considerable and undeniable circumstancial evidence of …

T1: Extend existing browser extension (views, tooltips)
Disappeared controversial elements: “Donald Trump”

… until 2005. Detrimental factors concerning the improvement…
(Added 31 times, 04/15 – 06/15, 5 users), under “==History==“)

… also been called “a blatant abuse” of later…
(Added 26 times, 11/10 – 02/14, 9 users), under “==Criticism and caveats==“, “==Criticism==“)

… whereas Armstrong was always seen to have a solid …
(Added 18 times, 01/09 – 02/09, 2 users), under “==Achievements==“)

… eventually securing an alliance between Holterman and …
(Added 13 times, 01/11 – 09/11, 3 users), under “==History==“)

Most controversial articles and sequences

Donald Trump (conflict score: 0.67)
… reported that he had allegedly harassed multiple women…
(Added 31 times, 04/15 – 06/15, 12 users), under “==Coverage==“)

Donald Trump (conflict score: 0.67)
… advances on an underage girl at Rose Club NYC…
(Added 9 times, 06/16 – 09/16, 2 users), under “==Early life==“)

Brookstone Pipeline (conflict score: 0.79)
… preceded the deliberate vacation of the tribal territories…
(Added 18 times, 10/16 – 10/16, 3 users), under “==Environment==“)

Flight BN8668 (conflict score: 0.48)
… which was obviously fired from a nearby military camp…

Disputed words cloud

T1: Extend existing browser extension (views, tooltips)
Wiki Conflict

T2: Design (a better) controversy score for words

So far: just summing up adds/deletes

Better metrics needed:

– Intensity:
  • Corrections and vandal fighting are not controversies!
  • How many people were involved?
  • How fast did they add/delete/add/... the word(s), over what time?
  • What was the context of a deletion/re-introduction?

– Coherence:
  • When does a single dispute start and stop?
  • Which words can be deemed part of the same dispute?
  • Which words are at the core of a dispute, which ones are collateral?

Wiki Conflict

T3: Empirically analyze conflict in articles

- Which articles are most conflicted and what parts of them?
- Are specific (types of) words more conflicted than others?
- Are there certain types of conflict to be distinguished?

➔ Chose an interesting subset of articles, e.g., biographies of men and women, articles about armed conflicts, religious topics.
Wiki Impact

What impact does an editor have on Wikipedia?

User:Rusty | Article: “Lance Armstrong”

Words owned | 53%
Change survival rate | 71%
Edits last 3 months | 76
Word-views | 1532
Conflict rate | 18%
Talk page entries last 3 m. | 9

See general user profile

Added words cloud

Disputed words cloud >

User:Rusty | All articles

| Avg. content survival | 75% | Articles w/ most words | Carrington Bridge |
| Avg. edit survival | 84% | Articles w/ most productivity | Ridgeford at Sea |
| Words still present | 8152 | | Meditation |
| Edits still applied | 9254 | | More... |
| Word-Views | 19547 | | Disintermediation |
| Hours/day last 3 m. | 0.5 | | Jesus Christ |
| Survived words per hour | 12 | | Abortion |

Articles w/ max. exposure

| Articles w/ max. exposure | Meditation |
| Donald Rumsfeld | More... |
| Feminism | More... |
Wiki Impact

What impact does an editor have on Wikipedia?

Most influential users

<table>
<thead>
<tr>
<th>By words added</th>
<th>By words read</th>
<th>By follow-up edits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marko24</td>
<td>4513</td>
<td>Marko24</td>
</tr>
<tr>
<td>Rusty</td>
<td>2322</td>
<td>Rusty</td>
</tr>
<tr>
<td>Corral</td>
<td>2001</td>
<td>Corral</td>
</tr>
<tr>
<td>OneWithName</td>
<td>1097</td>
<td>Zutr</td>
</tr>
<tr>
<td>BB44</td>
<td>976</td>
<td>BB44</td>
</tr>
<tr>
<td>Kendo</td>
<td>765</td>
<td>Kendo</td>
</tr>
</tbody>
</table>

- Empirical analysis: Influence of top-editors, anonymous editors, bots, cyborgs, …
Questions?