PreRegistration:Wikilink Analysis

Authors

Isabelle Langrock *(University of Pennsylvania - isabelle.langrock@asc.upenn.edu)*

Name

Feminist Interventions on Wikipedia: From Collaboration to Content

---

Research goal and rationale

With over 5 million entries in the English language edition alone, Wikipedia is the world's largest collaborative endeavor. However, equally well documented is Wikipedia's gender problem: not only do women participate as editors at significantly lower levels than men, even when compared with gender gaps in other technological pursuits, but biographies of women are often incomplete, poorly edited, or even remain unwritten. Scholars from STS, Communication, Media Studies, and Computational Social Science have examined the persistence and vastness of these gaps and even made suggestions for improvement, yet few have examined the effects of the initiatives attempting to remedy the disparity. This project examines how feminist movements reconceptulize collaboration as a process of sharing values and motivations rather than the ability for many people to participate, as it's commonly described in Wikipedia policies. Thematically analysis of the mission statements and training materials of the two movements as well as Wikipedia establishes these two differing forms of collaboration. This project then turns to looking at how the movements, with their specific value driven form of collaboratio, impact the structure of the content that is produced by examing Wikipedia's algorithmic assessment of article quality and the wikilink structure, both of which have impacts beyond the site. Comparisons are made between female biographical articles edited by the movements and other groups that have no interventions: female politicans, athletes, and academics. The data for this computational assesment includes the biographical articles edited by the movements, as identified through their WikiMedia dashboards, as well as other identified biographical articles for comparison.

This research project serves as an example of the combined socio-techinical effects of any digital collaborative endevor on its output. Furthermore, the focus on feminist interventions on a platform with a prevalent gender gap refocuses attention away from a deficeit based model that looks at what women are not doing and towards a framework that suggests how sociotechnical processes can be hostile, and become more open, to marginalized populations.

Hypotheses

Wikilinks

Do the articles edited by the feminist movements (500 Women Scientists and Art+Feminism) link to each other more than other biographical articles grouped by occupation and more than to be expected by random chance?

Given the motivations of collaboration of the two groups to make their histories visible and placing emphasis on the value of gathering together to edit Wikipedia and learn from each other, I expect to find a more dense network within the scientist and artist groups than the other groups.

Data collection

Have any data been collected for this study already?

> It's complicated. We have already collected some data but explain under 'Other' why readers may consider this a valid pre-registration nevertheless.

Dependent variables

Network Density The density of the networks, sorted by occupation and compared with randomized networks, will be my main dependent variable.

Modularity/Community Detection A secondary dependent variable, related to the first, that will help further describe the density measure.

Conditions

How many and which conditions will participants be assigned to?

Data is divided into groups by occupation, which is determined by Wikipedia. The two "treatment" groups are artists and scientists and comes specifically from pages recorded in the movement's records. "Control" groups are politicians and athletes (and maybe academics) and are pulled from Wikipedia made lists of "Women X Profession" (e.g. Women Leaders).

Analyses

I will build a network of the wikilinks for each group, where:

- nodes = biographical figure/subject of the article
- edges = wikilinks between articles; direction of edge showing what page the link is from

Then I will analyze the density and modularity of each network and compare them to random networks that maintain the same number of nodes and edges.

Outliers and Exclusions

Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.

Observations will be exluded if they: 1) Are not a biographical Wikipedia article (ex: objects, concepts, places and lists are frequently the subject of Wikipedia articles and could have been edited by the movements. However the primary focus of the movements are editing biographical pages of women so they will not be examined here) 2) I'm only looking at wikilinks to other biographies in my dataset, for example, if Frida Kahlo's Wikipedia article (which is quite robust already) is not edited by the Art+Feminism, I won't look at Wikilinks to her.

Sample Size

Sample size is based off of the recorded edits on the two movement's dashboards, a tool provided by Wikipedia for groups to track their edits, then reduced to only the biographical content (see section above). Groups of similiar size are selected for the "Control" groups using the names included on Wikipedia lists for each profession/group. Lists were collected by searching Wikipedia for "Female X profession" and then selecting names from each list available. Duplicates were removed.

Other

- Anything else you would like to pre-register? No.
- If you answered "Other" in the Data Collection section, include justification here.

The qualitative analysis data has already been collected and the article pages are currently being collected. Because my data is publicly available this is not a grave concern to me.

- Is this registration associated with the same project as any other registrations? No