

—
gender
bias in der
wikipedia

Ein
Werkstattbericht

60 Minuten – Workshop am 25. April

Von Frieda Pauline Reitzer,
Jan Papmeier, Tobias Schmidt
und Wenzel Sterzik

Unser Projektteam bedankt sich bei Hanna Steinert und Ihrem Projektstudium [„Representing Women - feministische Strategien in Wissenschaft, Kunst und Netzaktivismus“](#) (Hanna Steinert, Humboldt Universität, Semester 2020/21) für die theoretischen und aktivistischen Grundlagen und Argumentationen unserer Projektarbeit. Und Kilian Rüß für die technische Betreuung während der Projektumsetzung im Rahmen des Projektstudiums [„Bringing data to life – interaktive Datenvisualisierung in Javascript“](#) (Kilian Rüß, Humboldt Universität, Semester 2020/21).

Zu Anfang...

- Auftrag:

Es soll ein Visualisierungsfeature entstehen, in welchem ein Thema griffig und anschaulich durch Text und Visualisierungen dargestellt wird.

- Wer wir sind:

(Inzwischen) Masterstudierende: Tobi (Sozialwissenschaften), Jan (Statistik), Wenzel (Bibliotheks- und Informationswissenschaft), Pauline (Global Studies/ Digital Humanities)

- Theoretische Argumentation aus dem Projektutorium „Representing Women“:

- Das Internet ist ein öffentlicher Raum, indem wie in jedem Raum Exklusionsmechanismen und Machtstrukturen am Werk sind
- Wissen ist nicht neutral (Situieretes Wissen, Donna Harraway)
- Wikipedia als Infrastruktur die Logik der Macht und Autorität folgt (Anyone can edit, Judy Wajman)
- Dadurch können nicht alle Gruppen gleichberechtigt in der Wissensproduktion teilnehmen und der Neutralitätsanspruch wird somit stark in Frage gestellt

Gender Bias in der deutschen Wikipedia

[Gender Bias in der deutschen Wikipedia \(wsterzik.de\)](http://wsterzik.de)

Biographien in der Wikipedia

Verdeutlichen kann man sich das sehr gut anhand der Biografien in der Wikipedia. Zum jetzigen Stand (11.08.2021) enthält die deutschsprachige Wikipedia 828.180 Seiten von natürlichen Personenseiten in 153.049 Kategorien. Zusätzlich sind Personen in eine von vier Geschlechterkategorien eingeteilt: "Frau", "Mann", "nicht binäre Person" oder "Geschlecht unbekannt". Die Geschlechterverteilung der Biographien ist milde gesagt erschreckend.

Personen/Person nach Geschlecht —
Seiten: 828.180 | Anteil nicht männlicher Person: 17 %



Wie lässt sich die Gender Bias in den Wikipedia-Inhalten operationalisieren?

- Auf fehlende Biographien wird oft und ausführlich hingewiesen
- Da Biographien durch die Tags eindeutig nach Geschlecht strukturiert sind, lässt sich mit ihnen einfach arbeiten
- Unser Ziel war es aber, diese so facettenreich wie möglich darzustellen

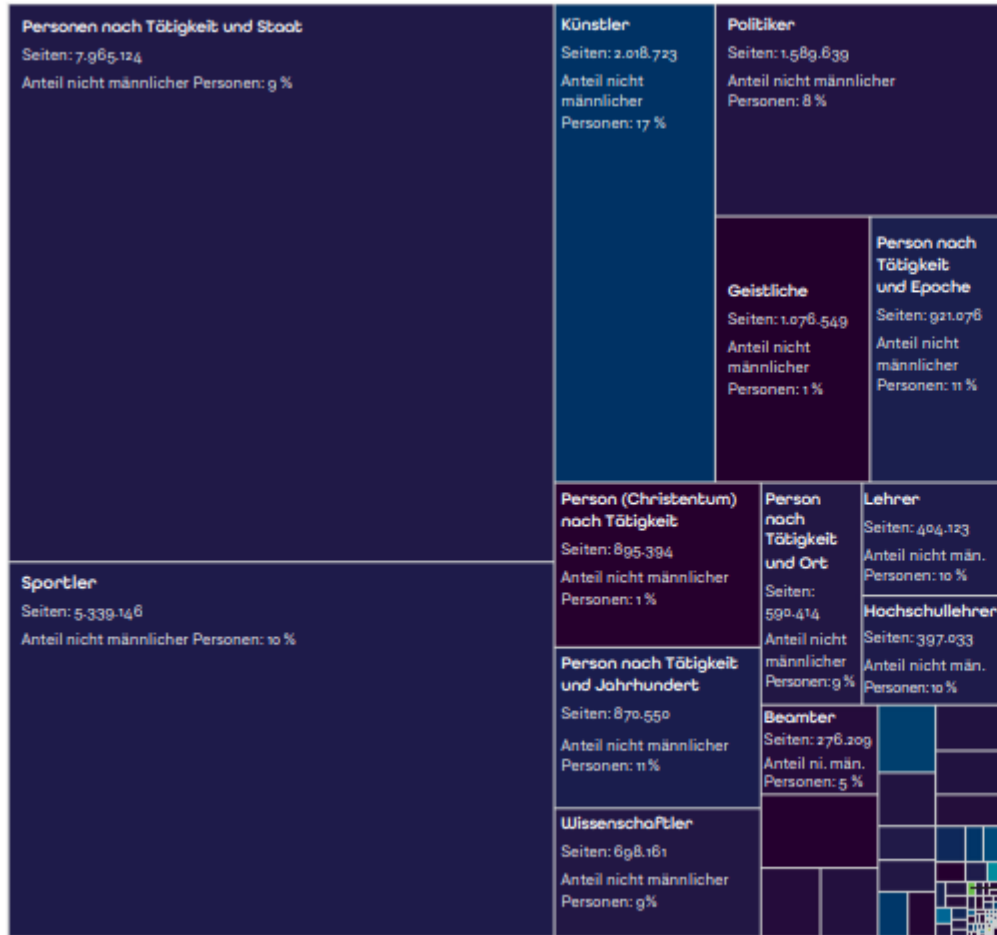
Technische Anforderungen

- Datenwerkzeug der Wikimedia Foundation Quarry
- Javascript – Datenvisualisierung
- D3.js Bibliothek für Interaktivität
- HTML und CSS

```
dropdownButton // Add a button
.selectAll('myOptions') // Next 4 lines add 6 options = 6 colors
.data(allGroup)
.enter()
.append('option')
.text(function (d) { return d; }) // text showed in the menu
.attr("value", function (d) { return d; }) // corresponding value returned by the button

const link = svg.append("g")
  .attr("stroke", "#999")
  .attr("stroke-opacity", 0.2)
  .selectAll("line")
  .data(graph.links)
  .join("line")
  .attr("stroke-width", 0.2)
  .attr("x1", d => d.source.x)
  .attr("y1", d => d.source.y)
  .attr("x2", d => d.target.x)
  .attr("y2", d => d.target.y);

const node = svg.append("g")
  .attr("stroke", "#fff")
  .attr("stroke-width", 0.5)
  .selectAll("circle")
  .data(graph.nodes)
  .join("circle")
  .attr("r", function(d){
    return d.linkCount > (d.linkCount/15) ? 2; //<-- some function to determine radius
  })
  .attr("fill", d => d.gender== "male"? "green" : "red")
  .attr("cx", d => d.x)
  .attr("cy", d => d.y);
```



Personen/Person nach Tätigkeit und Staat —
Seiten: 7.965.124 | Anteil nicht männlicher Personen: 9%

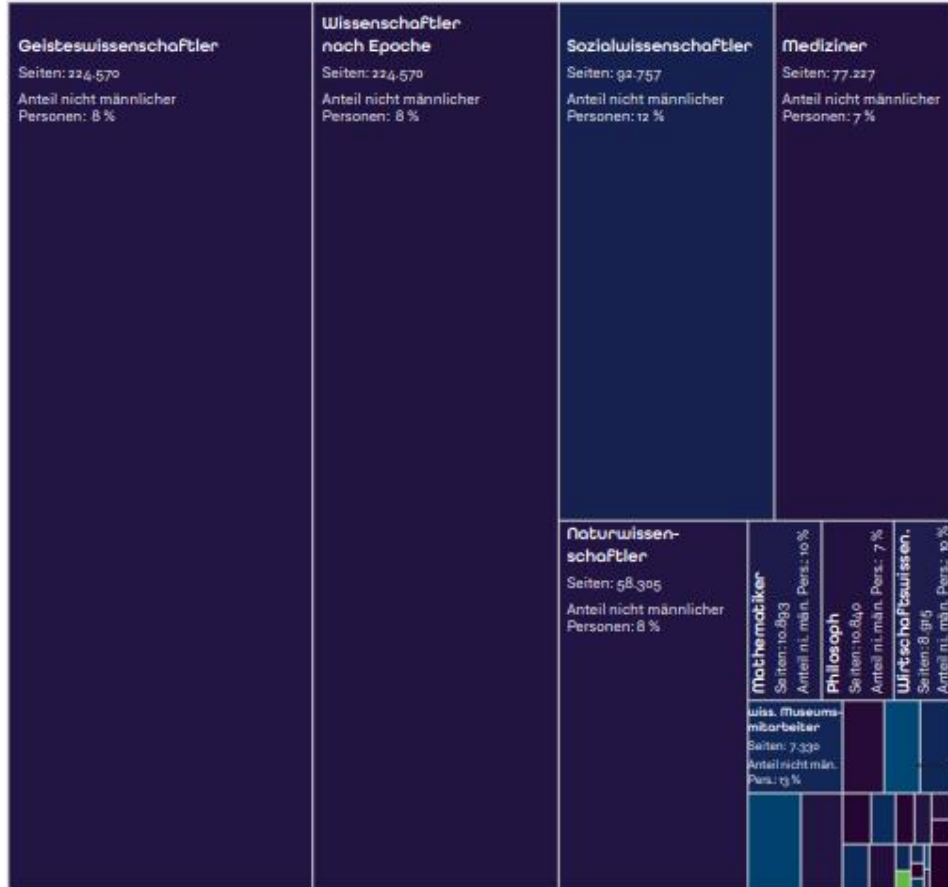
Über diese Daten – Quelle: Datenwerkzeug der Wikimedia Foundation: Quarry. Abgerufen am: 08.09.2021. Dargestellt werden alle Unterkategorien der Kategorie Person, jeweils mit der Gesamtzahl alle ihre Unterseiten in allen Unterkategorien. Personen sind in der Regel mehreren Kategorien zugeordnet und auch Kategorien können in der Hierarchie mehrfach auftauchen. Daher werden individuelle Seiten mehrfach gezählt und die Gesamtzahl der Seiten ist um ein vielfaches höher als die Gesamtzahl der Personenseiten. (Unter-) Kategorien mit weniger als 100 Seiten werden nicht dargestellt, um die Datenmenge zu reduzieren. Sie werden aber in der Zahl der Seiten für Oberkategorien mitgerechnet.



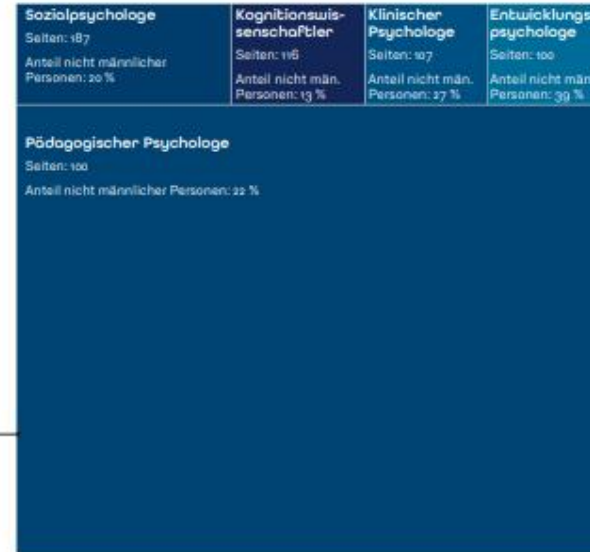
Personen/Person nach Model-Tätigkeit —
Seiten: 2.9984 | Anteil nicht männlicher Personen: 88 %

Personen/Person nach Tätigkeit/Wissenschaftler

Seiten: 698.161 | Anteil nicht männlicher Personen: 9 %



Mehr dazu: In der UK gibt es eine aktive Wissenschaftscommunity die zu dem Thema arbeitet. Wichtiges Werk ist hierbei "Anyone can edit, not everyone does: Wikipedia's infrastructure and the gender gap - Heather Ford, Judy Wajcman, 2017 (sagepub.com)" von Judy Wajcman. Zu der deutschen Wikipedia findet sich bisher kaum wissenschaftliche Arbeit.



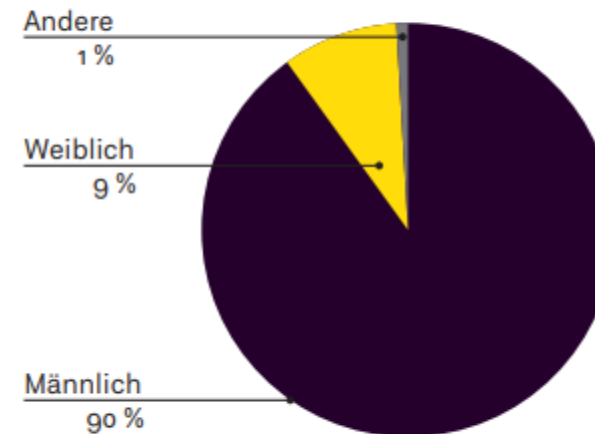
Personen/Person nach Tätigkeit | Wissenschaftler | Psychologe
Seiten: 2.777
Anteil nicht männlicher Personen: 23 %

Klickt euch mal durch zu eurer Lieblingskategorie,
was seht ihr?

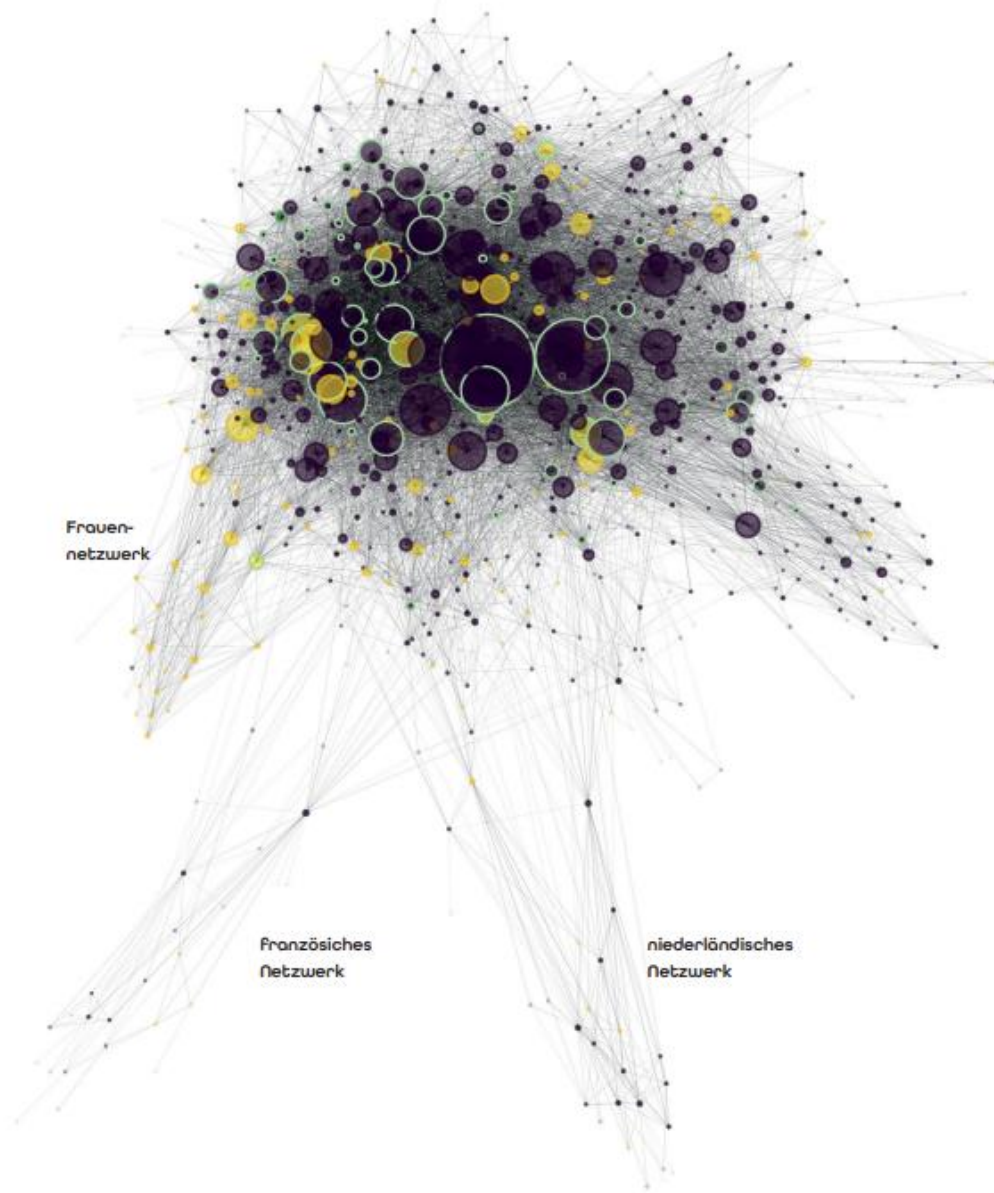
Die Wikipedia Community – auf der Suche nach Gründen

- Die Argumentation dass die Inhaltliche Ungleichheit schlicht darauf basiert dass es weniger Frauen gibt die „relevant“ sind, wollten wir so nicht hinnehmen
- Nach Wajman wollten wir der Idee nachgehen, ob Inhaltliche Ungleichheit und die Geschlechterverteilung der Wikipedia Community miteinander zusammenhängen
- Dass die Community, ungleich verteilt ist, ist bekannt
- Aber lässt sich mehr darüber herausfinden als ein Tortendiagramm?

Mitwirkende in Wikimedia-Projekten im
Jahr 2018 — Prozentsatz der Geschlechter



Quelle-Wikipedia:
Geschlechterverteilung in der Wikipedia



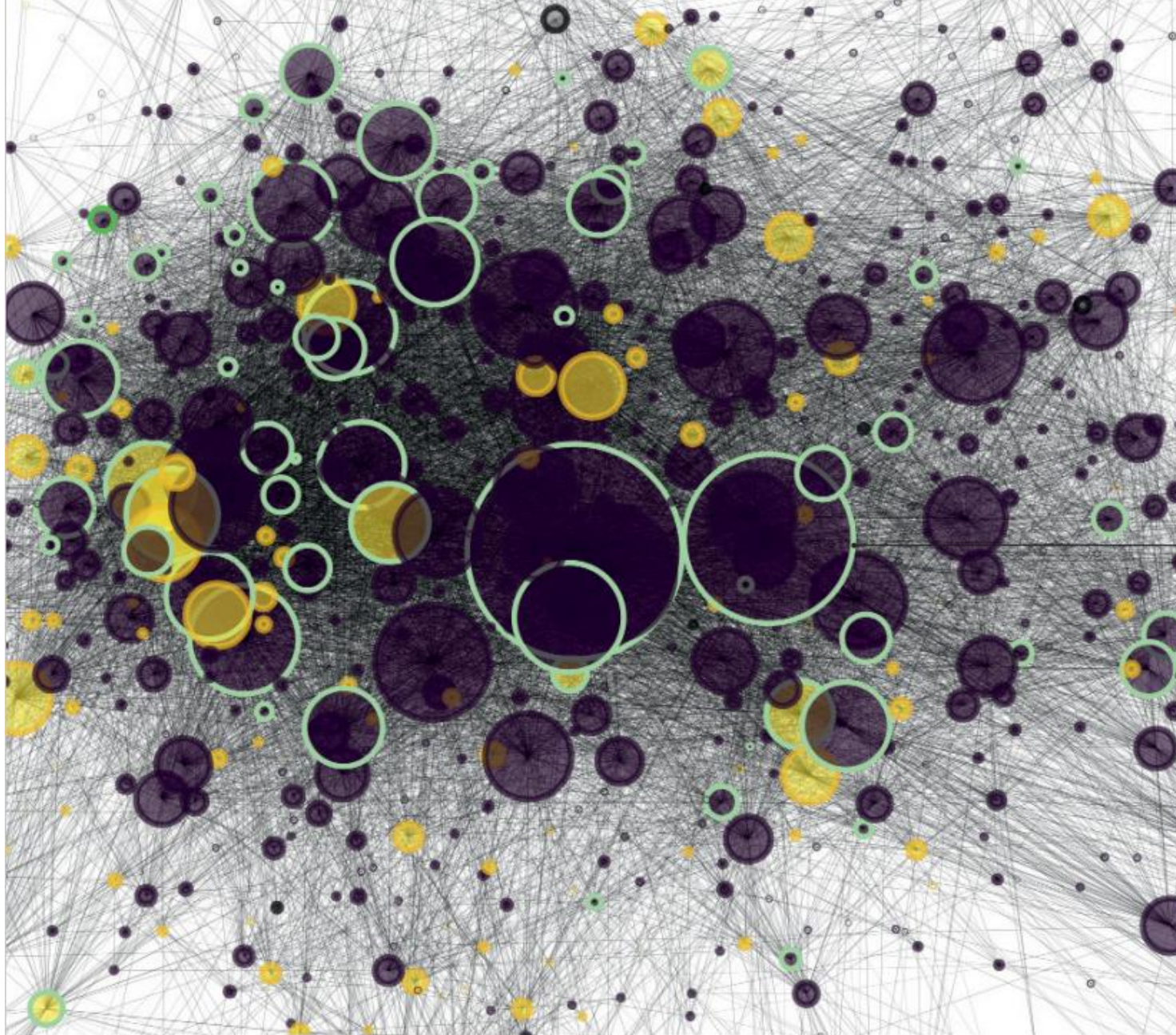
- weibliche Autorinnen
- männliche Autoren
- non-binary Autor*innen
- Admins

Der folgende Netzwerkgraph zeigt den Teil der Wikipedia Community, die sich einer Geschlechtskategorie zugeordnet haben. Die Kreise (Knoten) stellen einzelne Accounts dar, die Linien verbinden jene Accounts die angegeben haben, sich persönlich zu kennen. Je mehr persönliche Kontakte ein Account hat, umso größer deren Knoten. Die Farbe der Knoten stellt das angegebene Geschlecht dar.

Über diese Daten

Quelle: Datenwerkzeug der Wikimedia Foundation: Quarry. Abgerufen am: 07.07.2021
 Autor*innen der Wikipedia können sich innerhalb des »Projektes: Persönliche Bekanntschaften« gegenseitig als reale Personen verifizieren indem sie angeben, dass sie sich persönlich kennen (Startseite | Wikipedia: Persönliche Bekanntschaften | Wikimedia Toolforge (pb.toolforge.org)). Ausgewertet wurden bestätigte Verifizierungen von Personen mit Geschlechtsangabe. Die Anzahl der Verifizierungen wird durch die Größe des Knotens dargestellt. Personen mit Verifizierungen sind tendenziell dichter beieinander angeordnet. Weiblichen* Autor*innen wird empfohlen keine Geschlechtsangabe zu machen, um nicht Opfer von Anfeindungen zu werden.

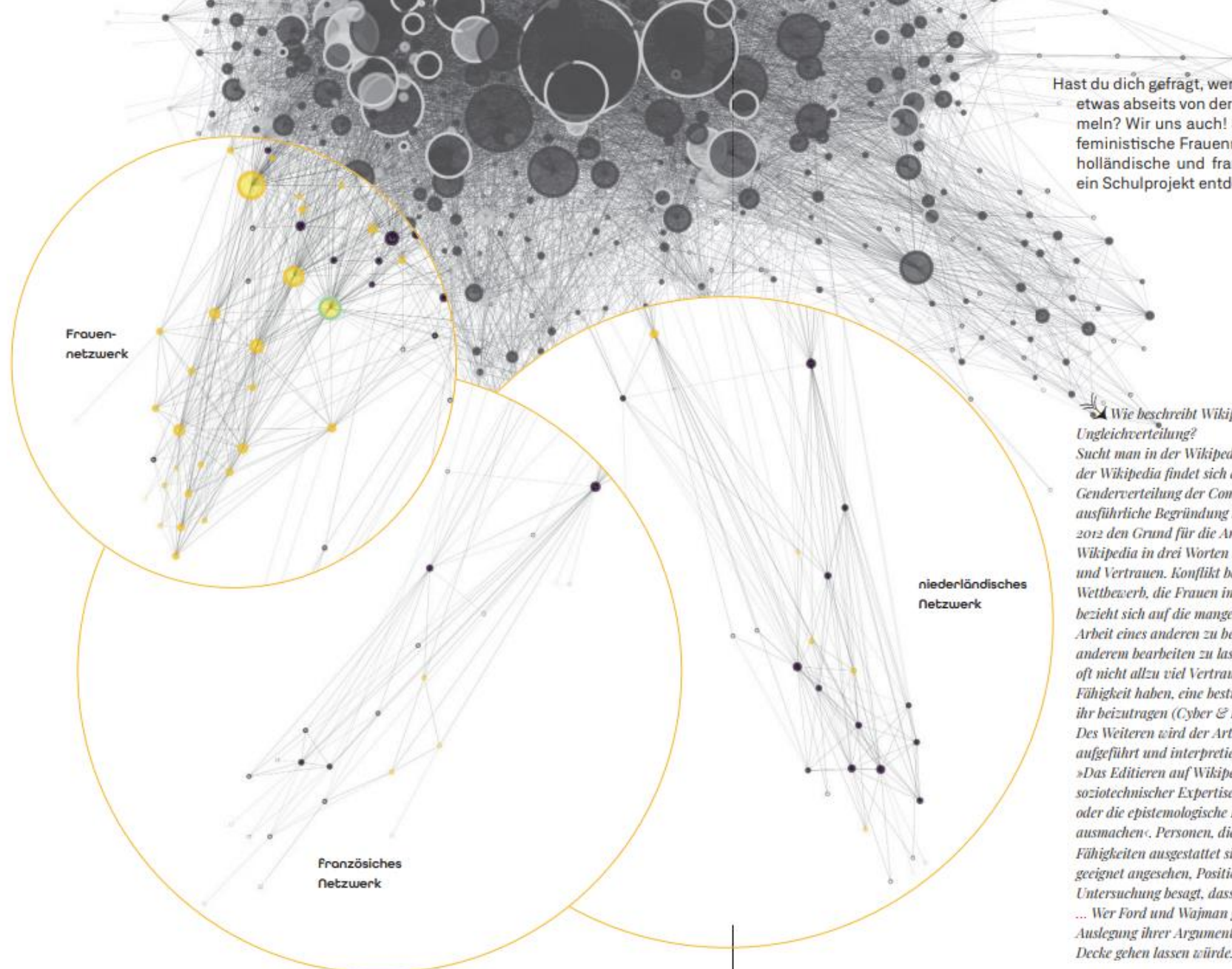
Auf der anderen Seite machen männliche Autoren oft keine Geschlechtsangabe, da Autor*innen durch das generische Maskulinum per se als Mann angesprochen werden, und somit der Anreiz einer Geschlechtsangabe für Männer minimiert wird. Außerdem ist es natürlich möglich ein beliebiges Geschlecht anzugeben. Wir vermuten, dass die tatsächliche Geschlechtsverteilung der Community einen noch höheren Männeranteil beinhaltet. Um einzelne Accounts vor Übergriffen zu schützen, haben wir die personalisierten Account-Daten anonymisiert.



Auch die Admin Verteilung innerhalb des Netzwerkes gibt wichtige Hinweise auf Machtstrukturen innerhalb der Community. Es wird deutlich, dass sich das Netzwerk um einige wenige (männliche) Accounts gruppiert.

Persönliche Kontakte: 677
Angegebenes Geschlecht: männlich
Admin

- weibliche Autorinnen
- männliche Autoren
- non-binary Autor*innen
- Admins



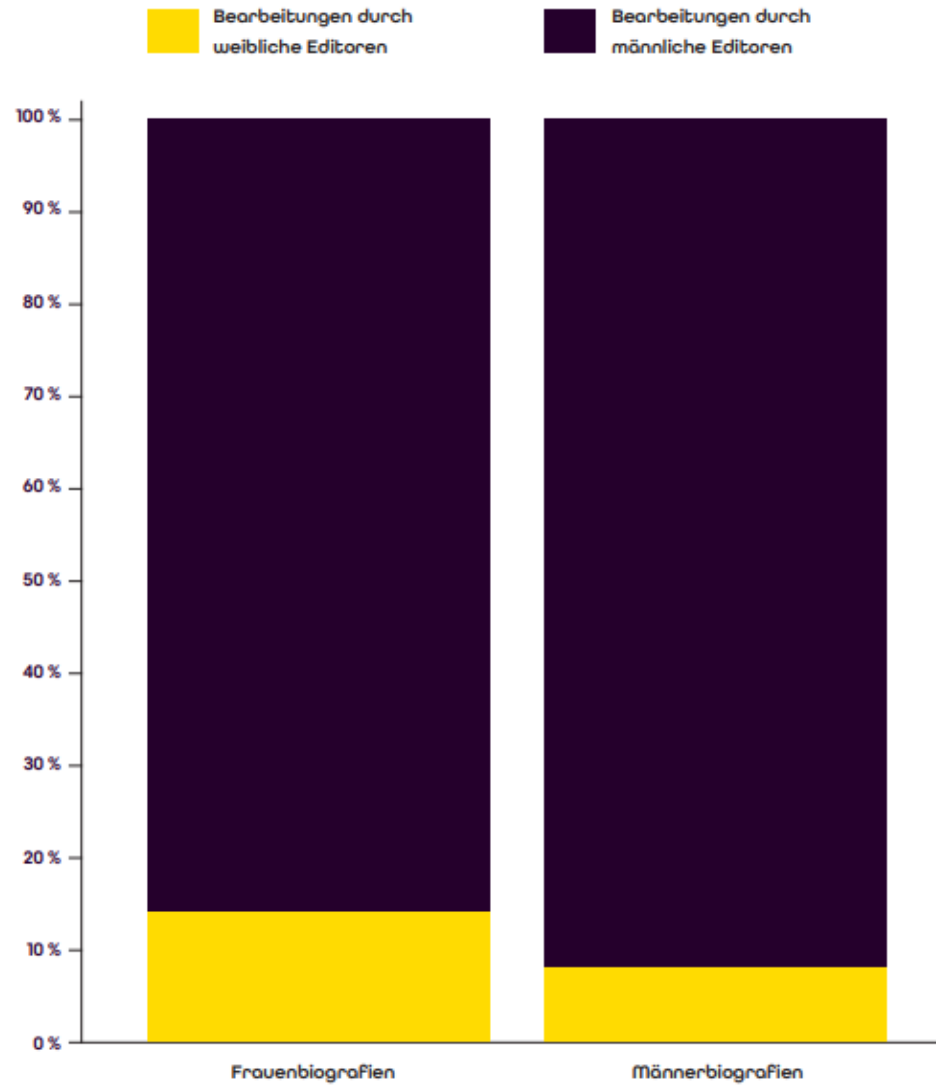
Hast du dich gefragt, wer die Accounts sind, die sich etwas abseits von der zentralen Gruppe sammeln? Wir uns auch! So haben wir nicht nur das feministische Frauennetzwerk, sondern auch die holländische und französische Community und ein Schulprojekt entdeckt.

🐦 *Wie beschreibt Wikipedia selber die Gründe für die Ungleichverteilung?*
Sucht man in der Wikipedia: Geschlechterverteilung in der Wikipedia findet sich ein Artikel der die Ungleiche Genderverteilung der Community klar bekennt und sogar ausführliche Begründung liefert: »Collier und Bear haben 2012 den Grund für die Arbeitsbarrieren von Frauen in Wikipedia in drei Worten zusammengefasst: Konflikt, Kritik und Vertrauen. Konflikt bedeutet Cyber-Mobbing, Trolling und Wettbewerb, die Frauen im Allgemeinen nicht mögen; Kritik bezieht sich auf die mangelnde Bereitschaft von Frauen, die Arbeit eines anderen zu bearbeiten und ihre Arbeit von jemand anderem bearbeiten zu lassen; Vertrauen zeigt, dass Frauen oft nicht allzu viel Vertrauen in ihre eigene Kompetenz und Fähigkeit haben, eine bestimmte Arbeit zu bearbeiten und zu ihr beizutragen (Cyber & Bear, 2012) [...]«
Des Weiteren wird der Artikel von Ford und Wajman aufgeführt und interpretiert:
»Das Editieren auf Wikipedia erfordert besondere Formen soziotechnischer Expertise und Autorität, die das Wissen oder die epistemologische Infrastruktur von Wikipedia ausmachen«. Personen, die mit diesem Fachwissen und diesen Fähigkeiten ausgestattet sind, werden in der Wikipedia als eher geeignet angesehen, Positionen mit Macht zu erreichen. Die Untersuchung besagt, dass dies überwiegend Männer sind.«
... Wer Ford und Wajman gelesen hat, weiß dass diese Auslegung ihrer Argumente die Beiden wahrscheinlich an die Decke gehen lassen würde.

Was seht ihr in dem Netzwerkgraphen?

Gibt es also einen Zusammenhang zwischen Gender Bias in den Inhalten und der Community Zusammensetzung?


WIE WIRKT SICH DIE GENDERVERTEILUNG IN DER COMMUNITY AUF DIE GENDER BIAS DER WIKIPEDIA AUS?



Erfahrungen

- Wikipedia ist ein Datenhimmel
 - Das Problem ist, dass die Daten aufbereitet werden müssen
 - Durch die Datenmenge ist die Gefahr von Missinterpretation oder sogar Manipulation groß (Gerade bei Visualisierungen)
 - Qualitative Begutachtung bleibt hinsichtlich der Bewertung von Forschungsvorhaben, Fragestellungen und Datennutzung essentiell
-
- Insgesamt großes Interesse von allen Seiten!

Vielen Dank für eure Aufmerksamkeit

 *Das vollständige Projekt mit den Graphiken
findet ihr unter: <https://wsterzik.de/wiki-project/>
Kontakt: wiki-projekt-hu@googlegroups.com*





Atttribution 4.0 International (CC BY 4.0)

Copyright, Copr., 2021

Inhalt:

Frieda Pauline Reitzer,
Jan Papmeier,
Tobias Schmidt,
Wenzel Sterzik

Lektorat:

Kilian Rüb
Grit Wolf