Editing Behavior over Time Power vs. Standard Wikidata Editors

Cristina Sarasua*, Alessandro Checco, Gianluca Demartini, Djellel E. Difallah, Michael Feldman, Lydia Pintscher

Universität Zürich[UZH]

sarasua@ifi.uzh.ch
@csarasuagar

WikidataCon 2017

6.8K - 8.7K

Active Editors

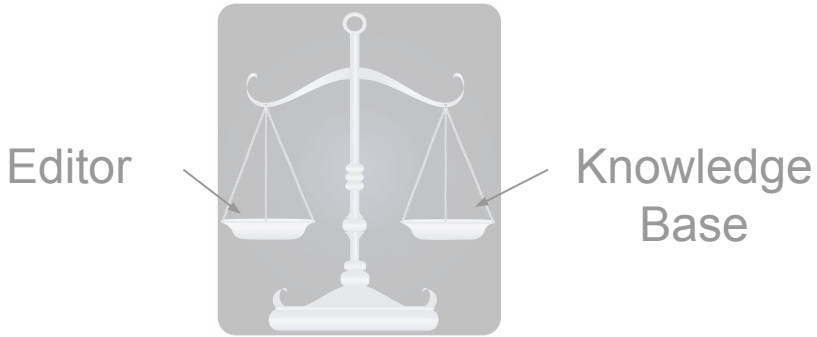
Source: 08.2016 - 08.2017 The Wikidata Revolution, Lydia Pintscher, Wikimania 2017


Welcome to Liverpool Street Station
LAURA ASHLEY
COFFEE
Tickets & Information
Bus station
Underground
RBS


Welcome to Liverpool Street Station
LAURA ASHLEY

Ultimate Goal:
Help these editors find valuable work to do in Wikidata

Ultimate Goal:
Help these editors find valuable work to do in Wikidata



Data-driven study

1. Understand differences in the behaviour between power editors and standard editors

2. Be able to identify if an editor will be “power” or “standard” editor

3. Provide a method that helps interested standard editors find their editing mission

Discussion

Editor Types

edits (volume)

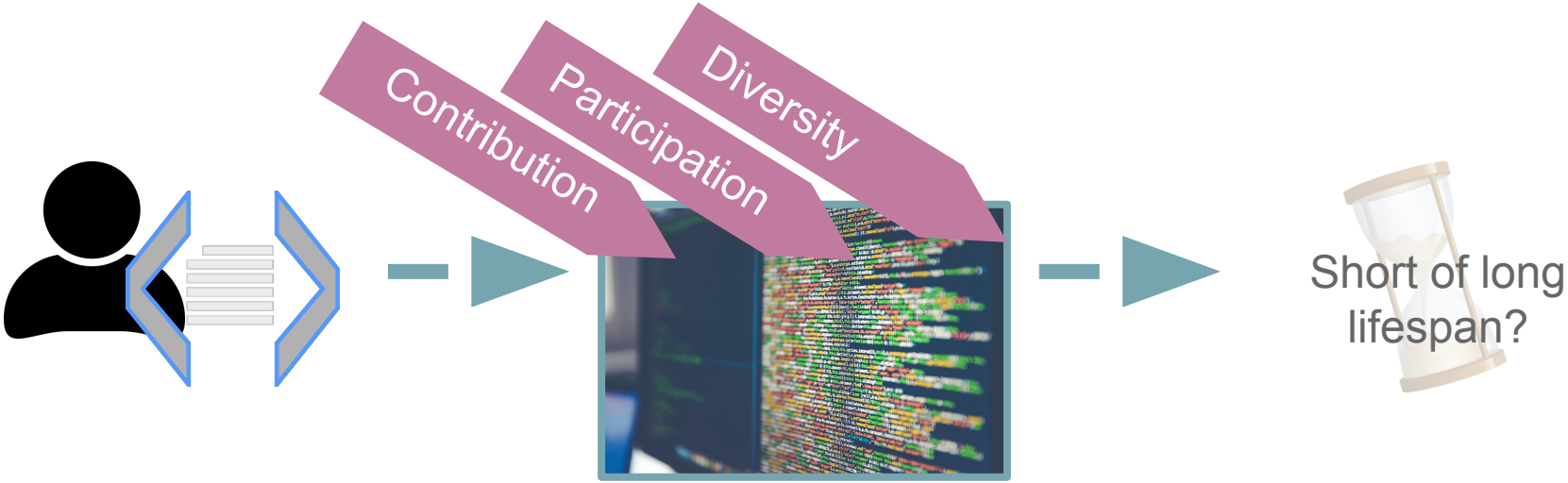
- High
- Low

months (lifespan)

- Long
- Short

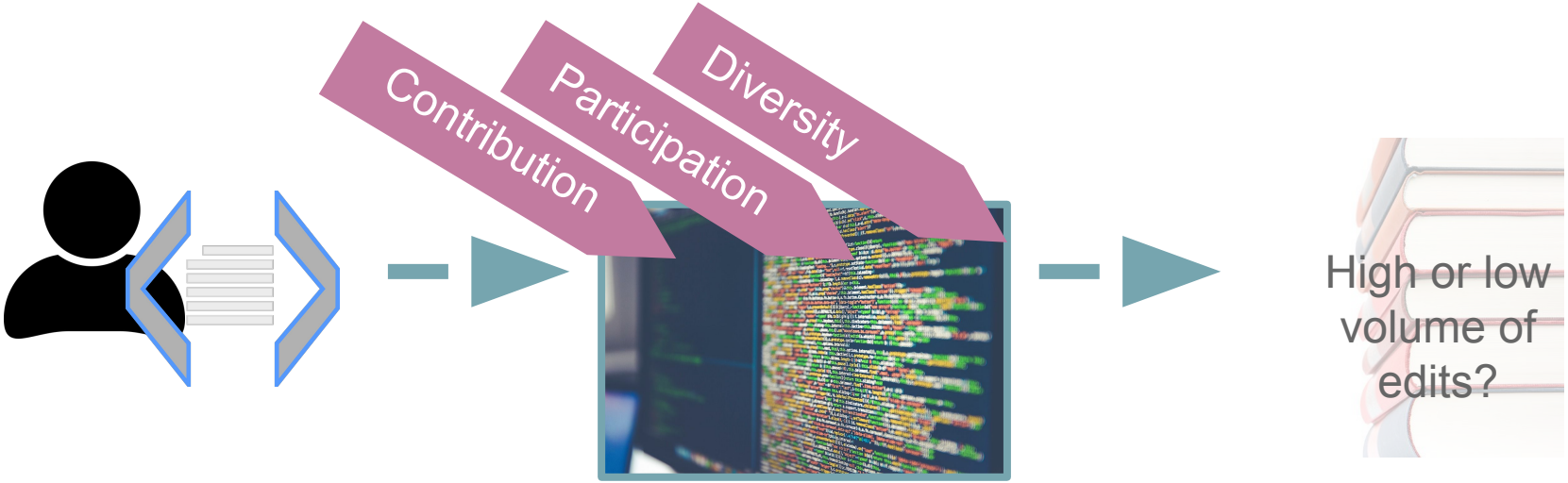
Evolution

Our Task



Editing Behaviour Over Time

Our Task



Editing Behaviour Over Time

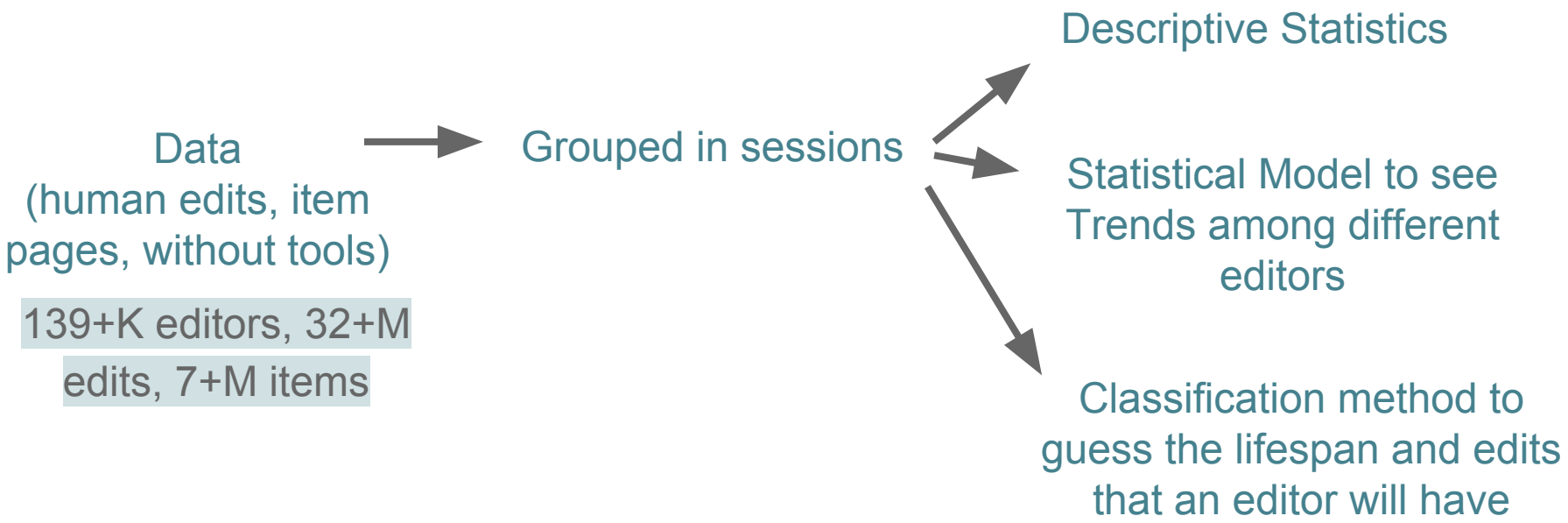
What does the related work say?

"*Wikipedians are born, not made. They don't do more over time and they maintain a high and constant level of participation.*"
[Panciera et al. 2009, Data-driven study]

"Wikidatians" acquire a higher sense of responsibility for their work, interact more with the community, take on more advanced tasks, and use a wider range of tools"
[Piscopo et al. 2017, Interviews]

"*There are different functional roles among editors: reference editor, item editor, item creator, item expert, property editor, and property engineer.*"
[Mueller-Birn et al. 2015, Data-driven study]

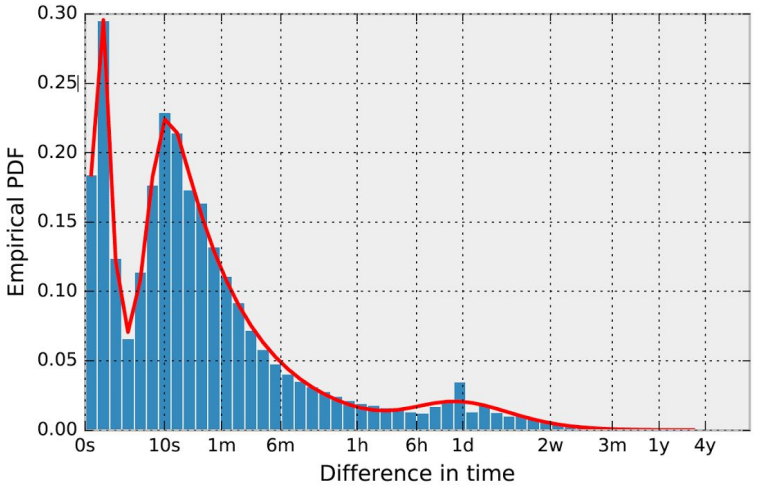
Methodology

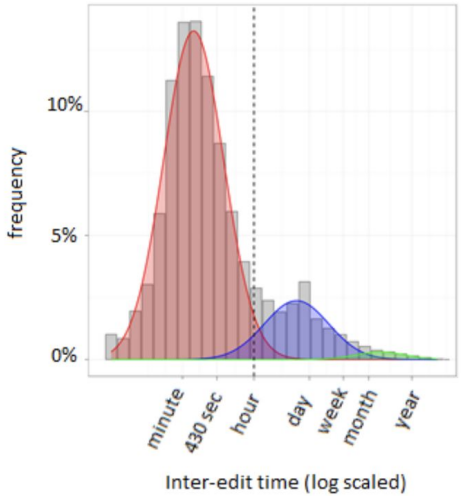


What did we find?

Edit sessions

F1. Shorter times between edits, and a longer definition of session than in Wikipedia (4.37 hours)

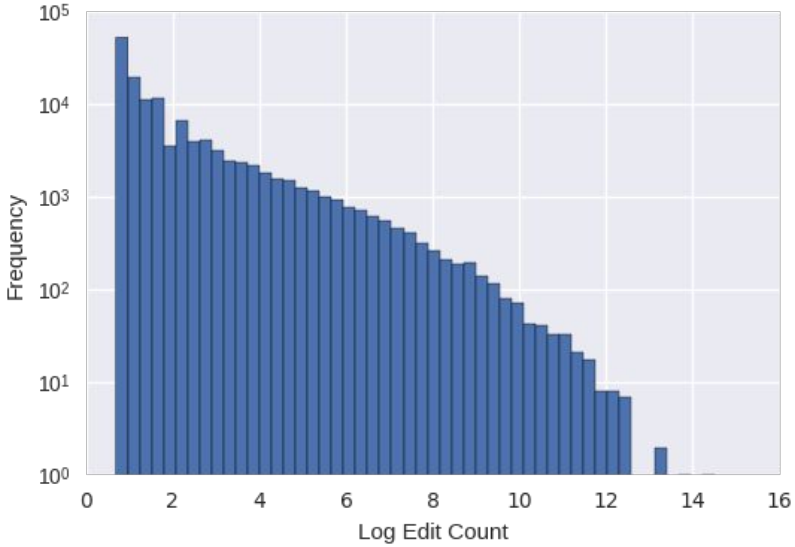


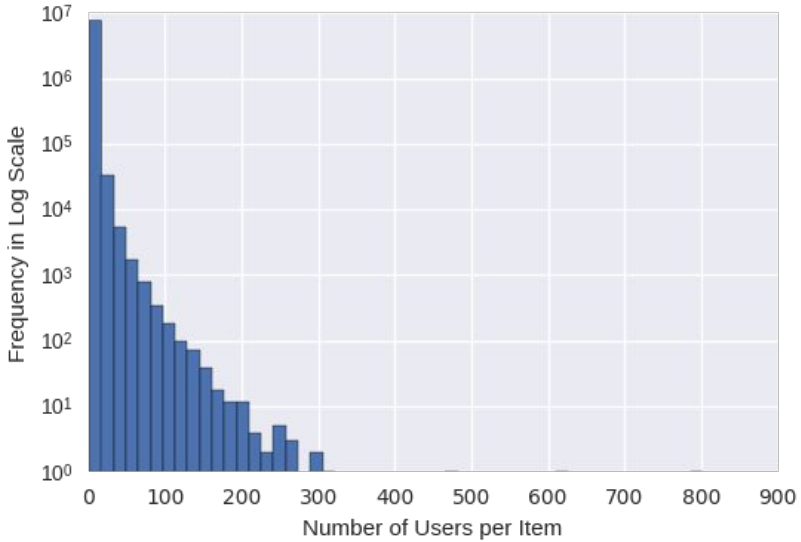


[Wikipedia, Geiger et al. 2013]

Editors and Items

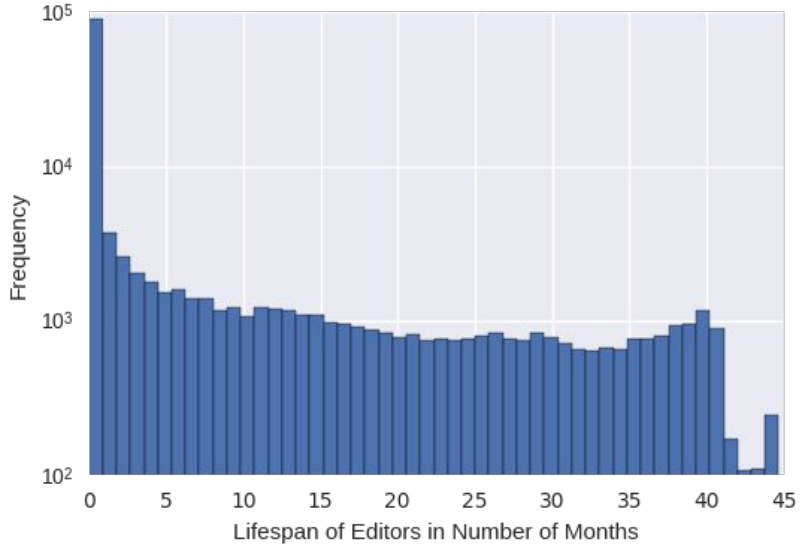
F2. Few editors with many edits (and vice versa), few items with many editors (and vice versa)

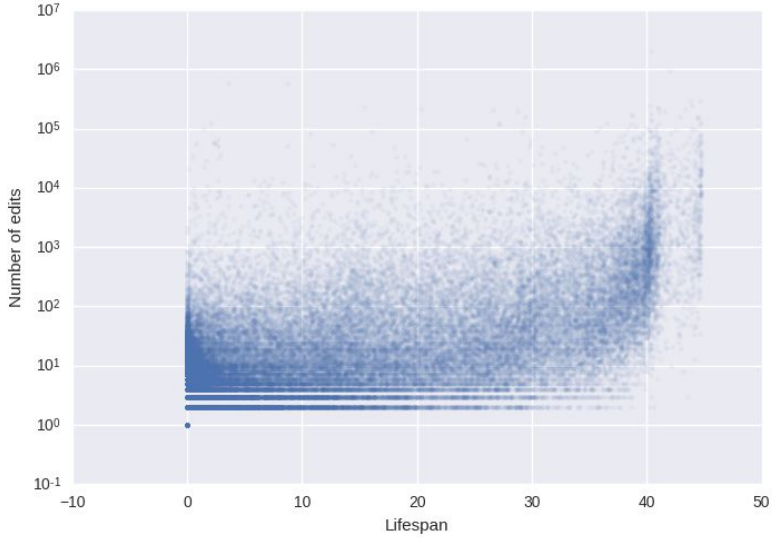




Lifespan

F3. Few editors worked over almost 4y, no linear relation between edit count and lifespan





F4.CONTRIBUTION

edits (session, month)
edits per item (s,m)
items edited (s,m)

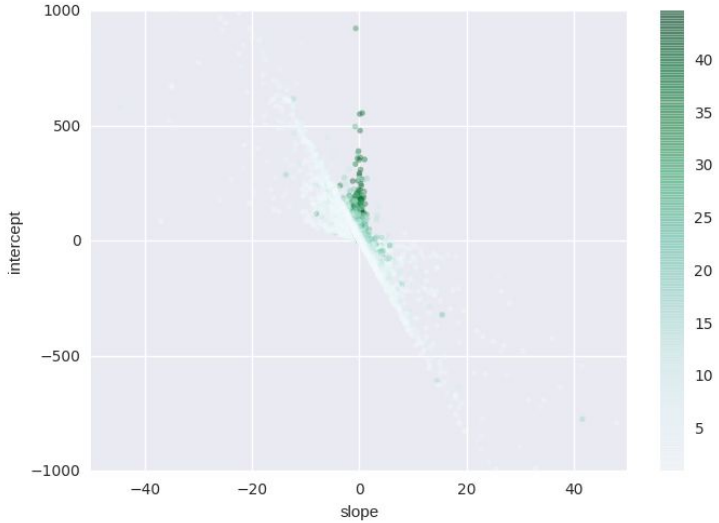
Editors with longer lifespan tend to maintain a constant contribution.

Others don’t.

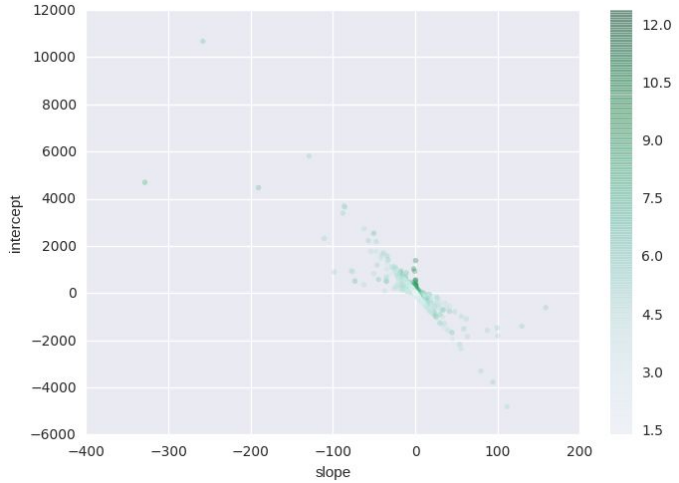
Editors with higher volume tend to maintain a constant contribution.

Others don’t (not as clear).

i1 m lifespan



i1 m editcount



F5.PARTICIPATION

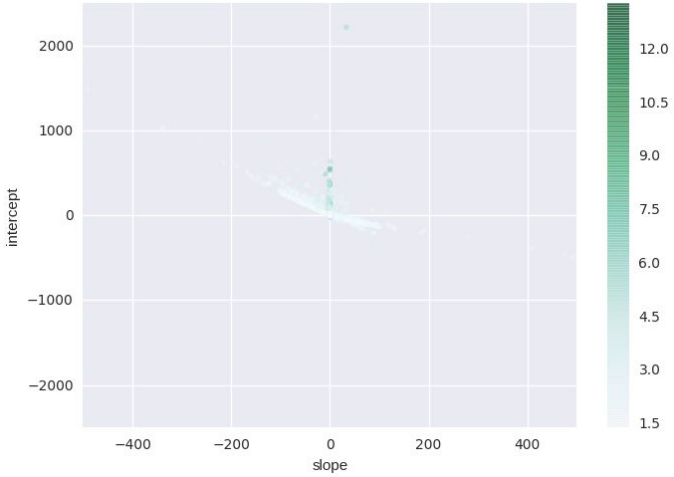
seconds spent (session)

Editors with a long lifespan maintain a constant participation.

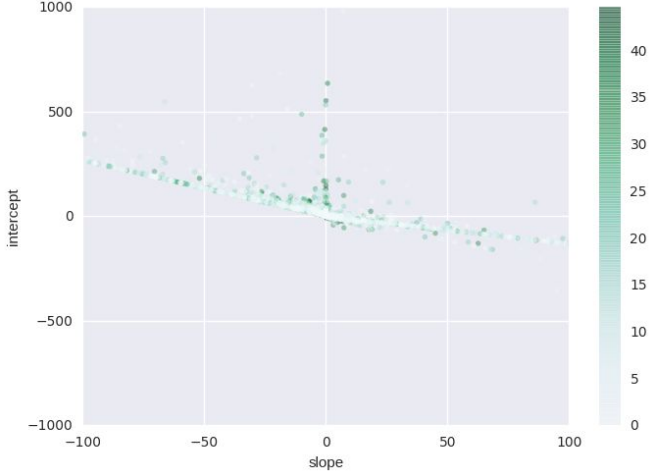
Others don’t.

Some editors with high volume of edits maintain a constant participation.

i4 s editcount



i4 s lifespan



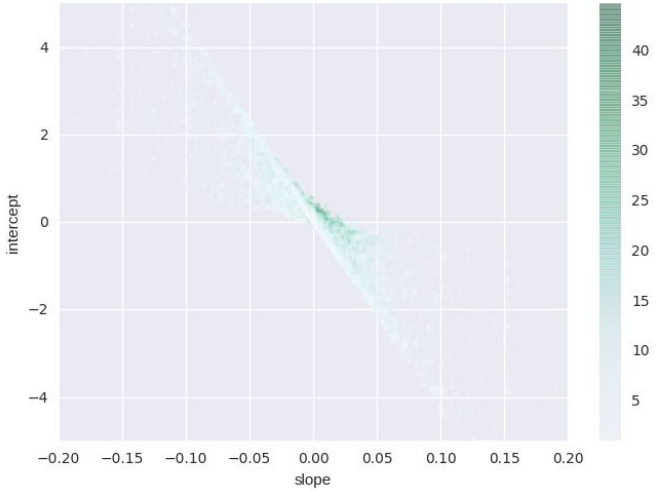
F6.DIVERSITY

entropy of type of edit (s,m)

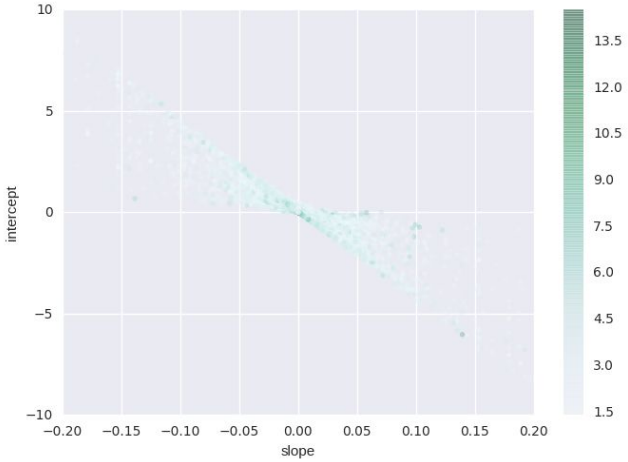
Editors with long lifespan tend to increase the diversity of the type of their edits (m).

For the others, some increase others decrease.

i5 m lifespan



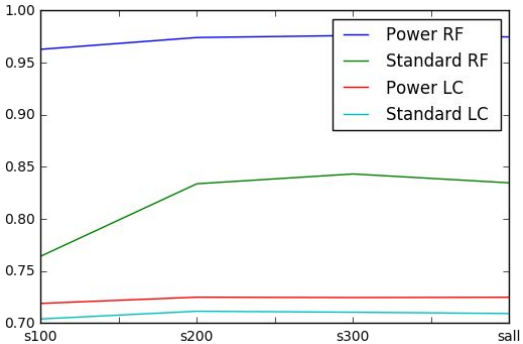
i5 m editcount



Identifying power and standard editors

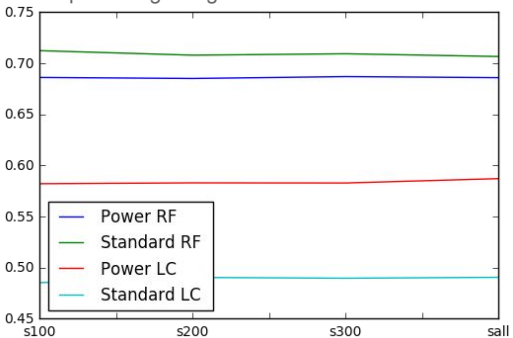
- Lifespan is predicted better than volume of edits.

15 months



Lifespan prediction: F1-score for Random Forest and Logistic Classifier predicting using different # of sessions

100 edits

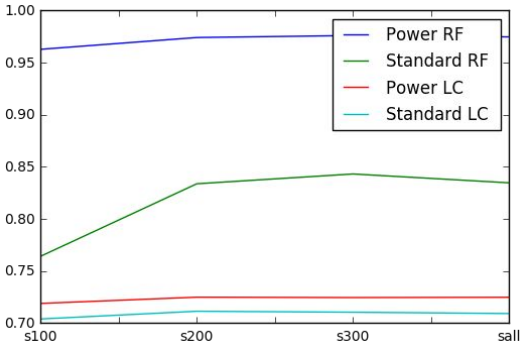


Volume of edits prediction: F1-score for Random Forest and Logistic Classifier predicting using different # of sessions

Identifying power and standard editors

- Lifespan is predicted better than volume of edits.

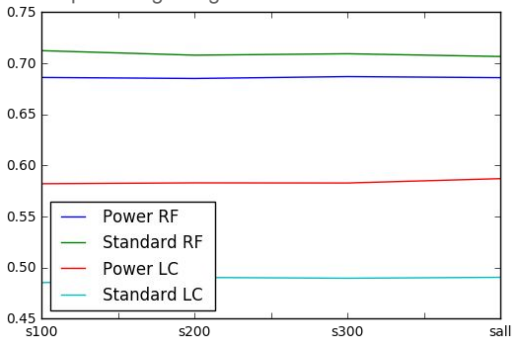
15 months



Lifespan prediction: F1-score for Random Forest and Logistic Classifier predicting using different # of sessions

- We can predict volume of edits better for standard editors than power users (both in session- and month-based evolution). As for lifespan, it is better for power editors.

100 edits



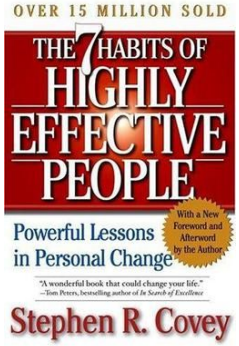
Volume of edits prediction: F1-score for Random Forest and Logistic Classifier predicting using different # of sessions

Conclusions from this research

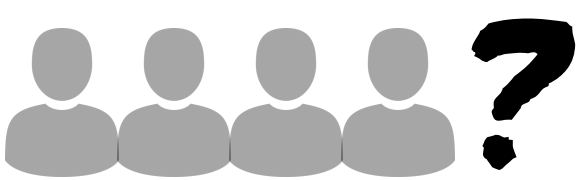
- Skewed distribution in volume of edits.
- 46 % of editors are presumably “gone”.
- Power editors (in contrast to standard editors) tend to have habits and be constant in contribution and participation.
- Power editors tend to increase diversity of type of actions over months.

How do we **help** standard users to **have editing habits** that suit them?

How do we **help** standard users to **have editing habits** that suit them?



Proposal



Standard Editors

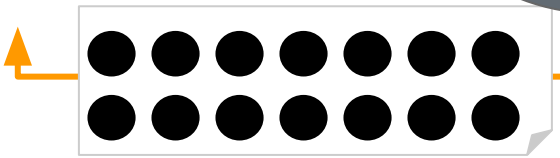
Method & Tool

Power Editors, Data Providers

- Define **intentions, resolutions**
- Identify with **roles** and **missions**

Focus Routines

- Publish **calls for actions**
- Define **data needs**



Individual / social missions

Best practices dissemination

@ Editors, Community Managers

- Are there people overwhelmed who don’t know how to contribute best?
- How do we collect and disseminate tips and tricks about deciding what to edit?
- How can we enable 1:1 collaboration between power editors / data providers and standard users?

@ Researchers, Developers

- Related theories to consider?
- What Wikidata tools to integrate in the process?

Big thanks!

keystone

COST Action IC1302

semantic KEYword-based Search on sTructured data sOurcEs

The University Of Sheffield.

WIKIMEDIA CH

WIKIMEDIA DEUTSCHLAND

Universität Zürich[UZH]

WeST People and Knowledge Networks

Sponsors & supporters

Wikidata community

References

Katherine Panciera, Aaron Halfaker, and Loren Terveen. 2009. Wikipedians are born, not made: a study of power editors on Wikipedia. In Proceedings of the ACM 2009 international conference on Supporting group work (GROUP '09). ACM, New York, NY, USA, 51-60. DOI=http://dx.doi.org/10.1145/1531674.1531682

Piscopo, Alessandro, Phethean, Christopher and Simperl, Elena (2017) Wikidatians are born: paths to full participation in a collaborative structured knowledge base In Proceedings of the 50th Hawaii International Conference on System Sciences. University of Hawaii. 10 pp, pp. 4354-4363. (doi:10.24251/HICSS.2017.527).

Claudia Müller-Birn, Benjamin Karran, Janette Lehmann, and Markus Luczak-Rösch. 2015. Peer-production system or collaborative ontology engineering effort: what is Wikidata?. In Proceedings of the 11th International Symposium on Open Collaboration (OpenSym '15). ACM, New York, NY, USA, Article 20, 10 pages. DOI: https://doi.org/10.1145/2788993.2789836

R. Stuart Geiger and Aaron Halfaker. 2013. Using edit sessions to measure participation in wikipedia. In Proceedings of the 2013 conference on Computer supported cooperative work (CSCW '13). ACM, New York, NY, USA, 861-870. DOI: https://doi.org/10.1145/2441776.2441873

Image sources

Slide 6 Attribution Nalex.25 - Creative Commons Attribution-Share Alike 4.0 International

Slide 8 CC0 https://pixabay.com/en/books-education-school-literature-484766/

https://pixabay.com/en/hourglass-sand-watch-time-glass-1046841/

Sliide 9 https://pixabay.com/en/question-mark-pile-question-mark-2492009/

CC0 https://pixabay.com/en/business-success-winning-chart-163464/

https://pixabay.com/en/code-technology-monitor-computer-2588957/

Slide CC0 https://pixabay.com/en/user-person-people-profile-account-1633249/

Slide 24 CC0 https://pixabay.com/en/user-group-icon-person-business-1275780/ https://pixabay.com/en/man-woman-question-mark-problems-2814937/ https://pixabay.com/en/map-travel-compass-magnifying-glass-2685795/

Slide 25 https://pixabay.com/en/protest-models-art-artist-2265287/

Slide 26 https://blog.wikimedia.de/2012/04/04/meet-the-wikidata-team/ photo by Phillip Wilke. CC-BY-SA-3.0

Slide 26 Group photo of Wikimania 2017 attendees. Photo by Victor Grigas/Wikimedia Foundation, CC BY-SA 4.0.