



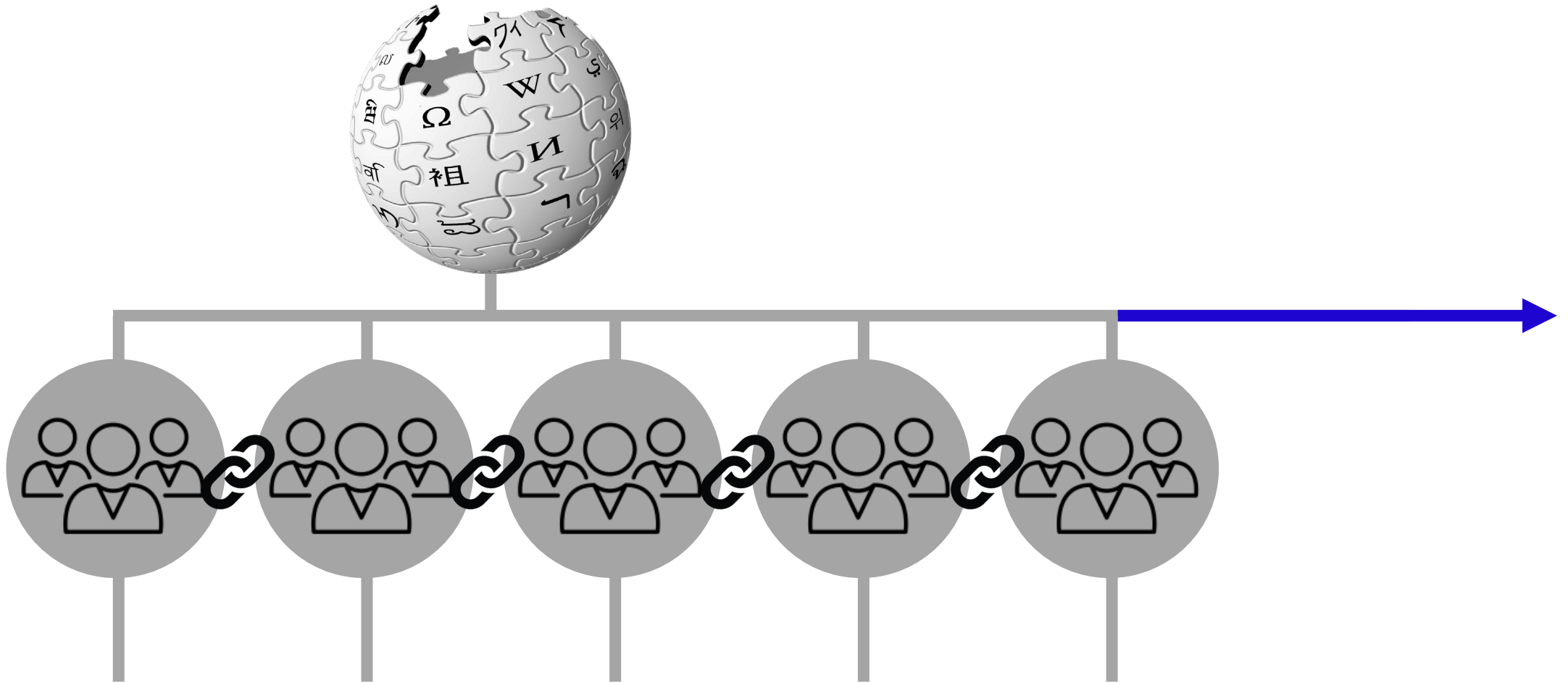
# Variation and overlap in the peer production of community rules

Sohyeon Hwang, Aaron Shaw  
Northwestern University

The image displays four overlapping screenshots of Wikipedia's 'Neutral point of view' policy pages in different languages:

- English:** 'Wikipedia:Neutral point of view' with a lock icon and a green checkmark indicating it is a protected page.
- Spanish:** 'Wikipedia:Punto de vista neutral' with a blue checkmark and a 'Políticas de Wikipedia' sidebar.
- German:** 'Wikipedia:Neutraler Standpunkt' with a blue checkmark and a 'Richtlinien' sidebar.
- French:** 'Wikipédia:Neutralité de point de vue' with a blue checkmark and a sidebar listing '120 langues'.

Each screenshot shows the title, a brief description of the policy, and a green checkmark indicating that the page is protected or approved. The text is partially obscured by the overlapping nature of the images.





**RQ1.)**

How do patterns of **rule-making** over time compare across language edition communities?

**RQ2.)**

How do the **sets of rules** become more/less similar over time among language edition communities?

# Descriptive study of nearly 20 years of rules and rule-making on five autonomous communities

English

French

German

Japanese

Spanish

# Descriptive study of nearly 20 years of rules and rule-making on five autonomous communities

## RESEARCH DESIGN



Construct **lists of rules** per language community

### ☰ Wikimedia Research/Showcase

Page Discussion

< Wikimedia Research

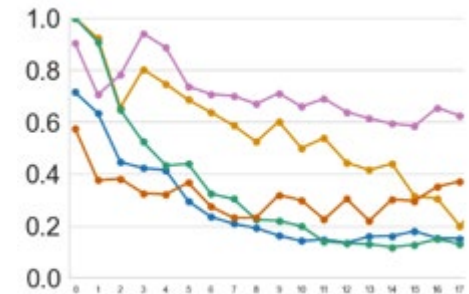
The **Monthly Wikimedia Research Showcase** is a public showcase of current research by the Wikimedia Foundation's [Research Team](#) and guest presenters from the academic community. The showcase is hosted at the Wikimedia Foundation **every 3rd Wednesday of the month at 9:30 a.m. Pacific Time/18:30 p.m. CET** and is **live-streamed on YouTube**. The schedule may change, see the calendar below for a list of confirmed showcases.

How to attend [edit]

We live stream our research showcases every month on YouTube. The link will be in each show page's details below. It is also announced in advance via [wiki-research-l](#), [@WikiResearch](#) on Twitter. You can join the conversation and participate in Q&A after each presentation by connecting to our [IRC channel](#): [#wikiresearch](#). We expect all presenters and attendees to abide by our [Policy](#).

Incoming Events

Collect **revision histories** and **interlanguage links**



Descriptive analyses examining  
(1) **comparisons across** and  
(2) **relationships between**  
communities

**RQ1.)**

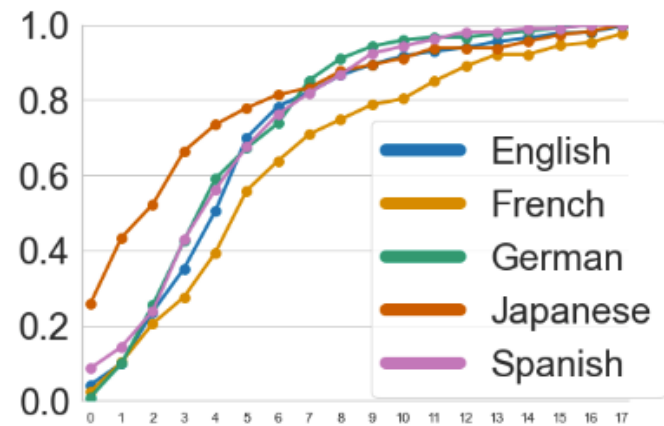
How do patterns of **rule-making** over time compare across language edition communities?

**RQ2.)**

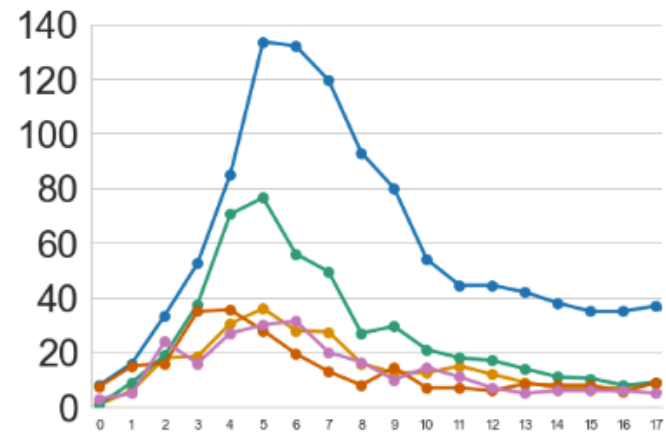
How do the sets of rules become more/less similar over time among language edition communities?

## KEY FINDINGS

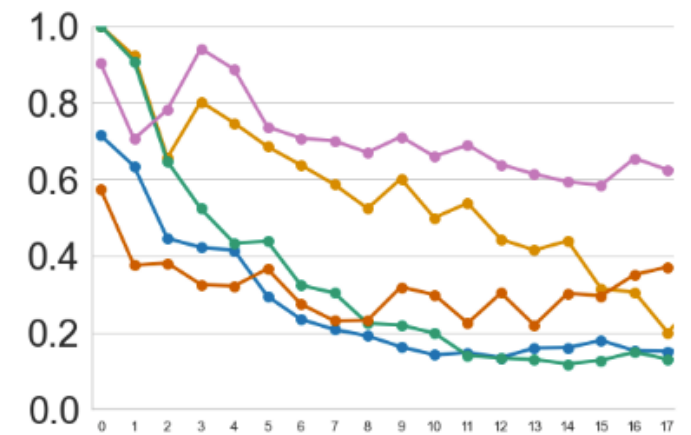
# Shared patterns of formalization



(a) Cumulative distribution function of proportion of rules created



(b) Median number of rule-related revisions



(c) % revisions on rule vs talk

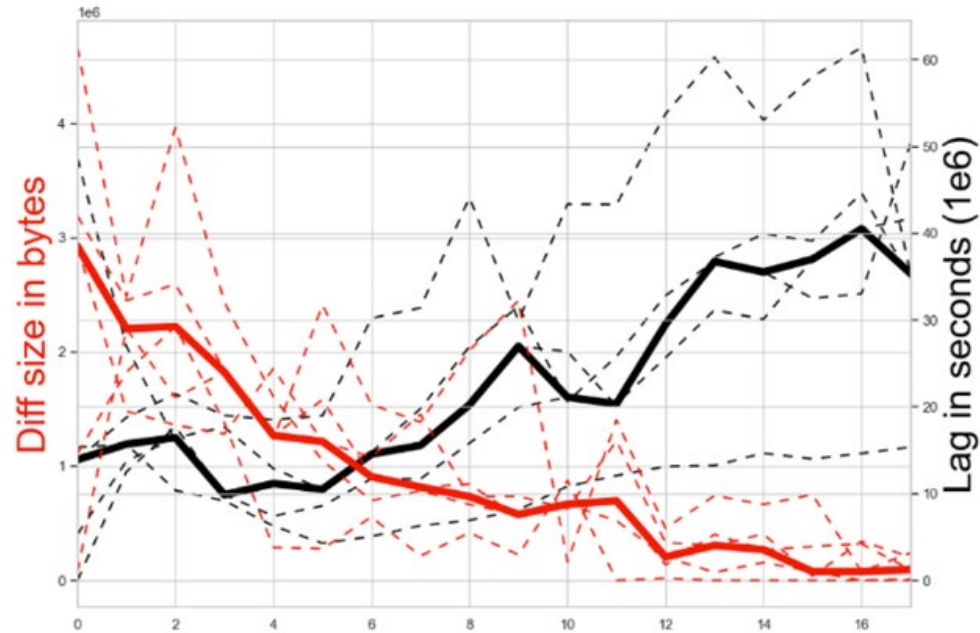
Figure 2: Trends in rule-related revisions (revisions on rule pages or their corresponding talk pages) by years since the first rule-related revision. 0 on the y-axis thus means the first year per language edition of rule-related revision activity.



## KEY FINDINGS

# Shared patterns of formalization

Decreasing size of revisions over time



Increasing time between revisions

# Shared patterns of formalization

Several rule-making patterns observed in English Wikipedia **replicate across non-English language editions.**



Research and practical implications of the generalizability of findings + lessons from lots of work on English Wikipedia!

RQ1.)

How do patterns of rule-making over time compare across language editions communities?

RQ2.)

How do the **sets of rules** become more/less similar over time among language edition communities?

# Unique (0 ILL) vs shared rules (4 ILL)

		34%	28%
		0 ILLs	4 ILLs
	Rule creation year	2007	2004
	Rule page length (char bytes)	8,100	14,251
Median	# Wikilinks	449	31,806
	# Revisions	195	371
	# Unique contributors	53	234

## KEY FINDINGS

# Increasingly diverging sets of rules



Figure 6: The proportion of rules in a focal language edition in each year (X-axis) with the number of ILLs to the other language editions in the set (area shading).

## KEY FINDINGS

# Continued importance of shared rules

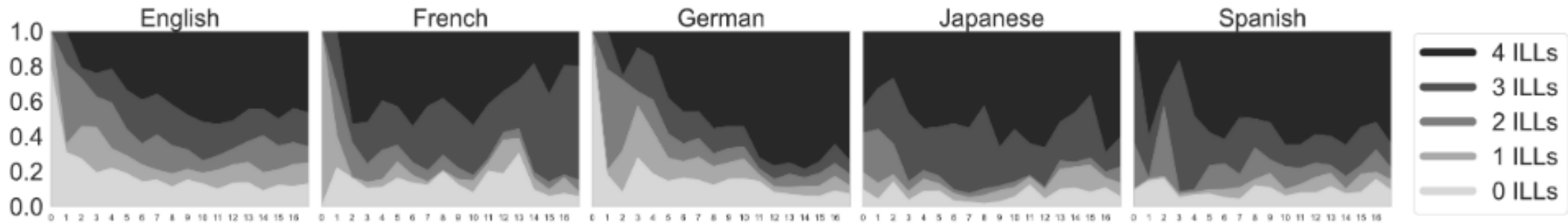


Figure 7: The proportion of rule-related revisions in a focal language edition in each year (X-axis) made to rules with the number of ILLs to other language editions in the set (area shading).

## KEY FINDINGS

# Divergence and convergence

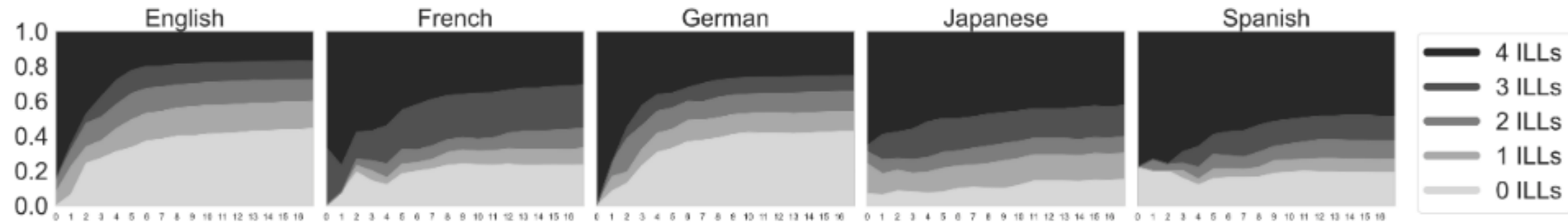


Figure 6: The proportion of rules in a focal language edition in each year (X-axis) with the number of ILLs to the other language editions in the set (area shading).

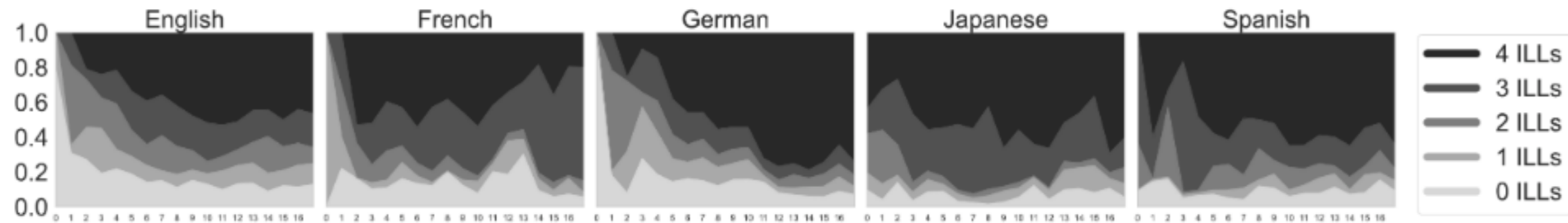


Figure 7: The proportion of rule-related revisions in a focal language edition in each year (X-axis) made to rules with the number of ILLs to other language editions in the set (area shading).

# Divergence and convergence in rules

Communities aligned in key ways **develop substantial + sustained institutional variations in practice**

Broad, widely-shared rules created early **may help coordinate communities** toward shared goals.



**Shared patterns of formalization**

**Increasingly diverging sets of rules**

**Continued importance of shared rules  
(which may be variably evolving)**

# Directions for future work

## **How scale relates to rules and rule-making**

How do these patterns compare in mid-size and small language editions?

## **Characterizing community-specific rules**

What kinds of rules are “unique”, and what makes them so?

## **Varying enforcement of broad shared rules**

Do language editions use the same rule to different effects?

How can we address harmful effects?

# Thank you!



Community  
Data Science  
Collective

## Contact

[sohyeonhwang@u.northwestern.edu](mailto:sohyeonhwang@u.northwestern.edu)

[aaronshaw@northwestern.edu](mailto:aaronshaw@northwestern.edu)

## Acknowledgements

- Support for this project came from the National Science Foundation (IIS-1910202, DGE-1842165).
- Special thanks to Brian Keegan, Casey Fiesler, Ágnes Horvát, Darren Gergle, and members of the Community Data Science Collective
- A version of this work has been published and presented at ICWSM2022, DOI: [10.1609/icwsm.v16i1.19297](https://doi.org/10.1609/icwsm.v16i1.19297)

