

# Wiki HUEs: Understanding Wikipedia practices through Hindi, Urdu, and English takes on evolving regional conflict

Molly Hickman<sup>1</sup>, Viral Pasad<sup>1,2</sup>, Harsh Sanghavi<sup>4</sup>, Jacob Thebault-Spieker<sup>1,3</sup>, Sang Won Lee<sup>1,2</sup>

Computer Science and Engineering 1, Echo Lab 2, Crowd Intelligence Lab 3, Industrial and Systems Engineering 4, Virginia Tech



## Mixed-method study of Wikipedia practices

How do editing practices differ between languages?

How quickly do editors respond to significant events of relevance to an article?

We looked at three articles, two of which existed only in the English and Hindi Wikipedias:

- + *Kashmir conflict* (EN, HI, UR)
- + *Article 370 of the Constitution of India* (EN, HI, UR)
- + *Insurgency in Jammu and Kashmir* (EN, HI)

## Background on Jammu & Kashmir



Hindi and Urdu are the two standardized varieties of Hindustani, a lingua franca of Jammu and Kashmir (J&K). J&K is a union territory administered and claimed by India. Pakistan and India each only control a part of the former princely state, but both claim J&K in its entirety. *Article 370 of the Constitution of India* was passed in 1954 and made temporary provisions for

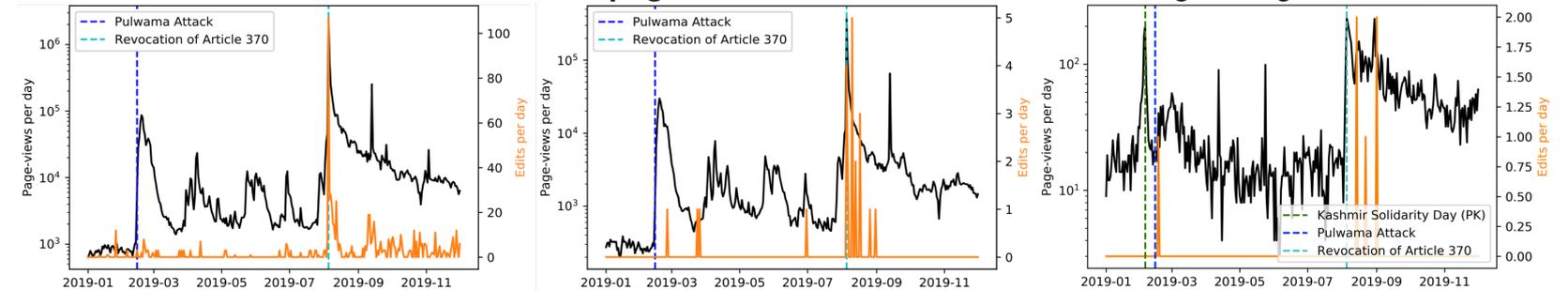
J&K that partly exempted the territory from the Constitution. **On August 5, 2019, Article 370 was revoked, which drew attention to the articles in our corpus.**

## Editorship overlap

- + 7 out of 86 Hindi editors edited both *Kashmir Conflict* and *Article 370* (Jaccard index = 0.08).
- + 65 out of 2,015 English editors edited both articles (Jaccard index = 0.03).
- + Five out of 22 Urdu editors edited both articles (Jaccard index = 0.23).
- + There were **very few editors who worked across languages** (Jaccard index < 0.01 in all three cases).

## Popularity and responsiveness

*Article 370 of the Constitution of India: page views and edits in 2019.* Left to right: English, Hindi, Urdu.



- + We used page views as a proxy for significant events. Some events are only important in one Wikipedia, e.g. *Kashmir Solidarity Day* only caused a swell in page views on the Urdu article.
- + Much less participation by editors in Hindi article than English, and still less in Urdu than Hindi.
- + English editors were most responsive; Hindi less so; and there were too few edits to the Urdu article to be conclusive.

## Editors' processes, motivations and observations

### Motivations

- + All interviewees echoed a **passion** for editing on Wikipedia.
- + Hindi and Urdu editors were interested in **“growing the language.”**

### Collaboration strategies

- + Hindi and Urdu editors stressed the importance of communicating with their peers. English interviewees preferred to work alone.
- + Hindi and Urdu editors used off-site channels (e.g., **Facebook, WhatsApp** to discuss issues, usually instead of on-site *Talk Pages*).

### Sourcing information

- + All Hindi and Urdu editors used **English Wikipedia articles as their primary sources**, with some content being **directly translated** from there.
- + English editors usually sourced information from **reputable news sources** and **scholarly articles**.

### Structure, Content, & Consistency

- + Hindi and Urdu editors had more relaxed approach to structure than English editors—English editors spend concerted efforts fixing **writing style and formatting**.
- + Hindi and Urdu editors much more focused on just **adding information** and **keeping it NPOV**.

### Propaganda & Vandalism

- + All interviewees provided anecdotes about **anonymous editors** vandalizing pages (making POV edits, or removing information that belonged).
- + One interviewee spoke of a group of editors spreading propaganda on the day Article 370 was revoked.