

Spatio-temporal Analysis of Reverted Wikipedia Edits

Johannes Kiesel, Martin Potthast, Matthias Hagen, Benno Stein

<first name>.<last name>@uni-weimar.de

Bauhaus-Universität Weimar

www.webis.de

ICWSM-17, May 18th 2017

What is Vandalism in Wikipedia?



Plato - Wikipedia - Mozilla Firefox

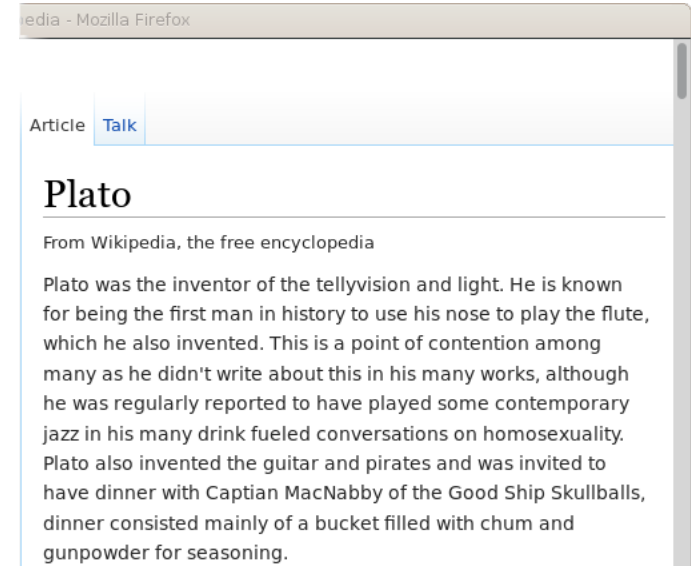
Article [Talk](#)

Plato

From Wikipedia, the free encyclopedia

Plato (**ancient Greek**: Πλάτων, *Plátōn*, "wide, broad-shouldered") (c. **428/427 BC**^[a]-c. **348/347 BC**) was an ancient **Greek philosopher**, the second of the great trio of ancient Greeks -**Socrates**, Plato, and **Aristotle**- who between them laid the philosophical foundations of **Western culture**.^[1] Plato was also a mathematician, writer of philosophical dialogues, and founder of the **Academy** in **Athens**, the first institution of higher learning in the western world.^[verification needed] Plato is widely believed to have been a student of Socrates and to have been deeply influenced by his teacher's unjust death.

Plato's brilliance as a writer and thinker can be witnessed by reading his Socratic dialogues. Some of the dialogues, letters, and other works that are ascribed to him are considered spurious.^[2] However, it is very probable that "everything Plato wrote for publication" has been passed along to us intact.^[3]



edia - Mozilla Firefox

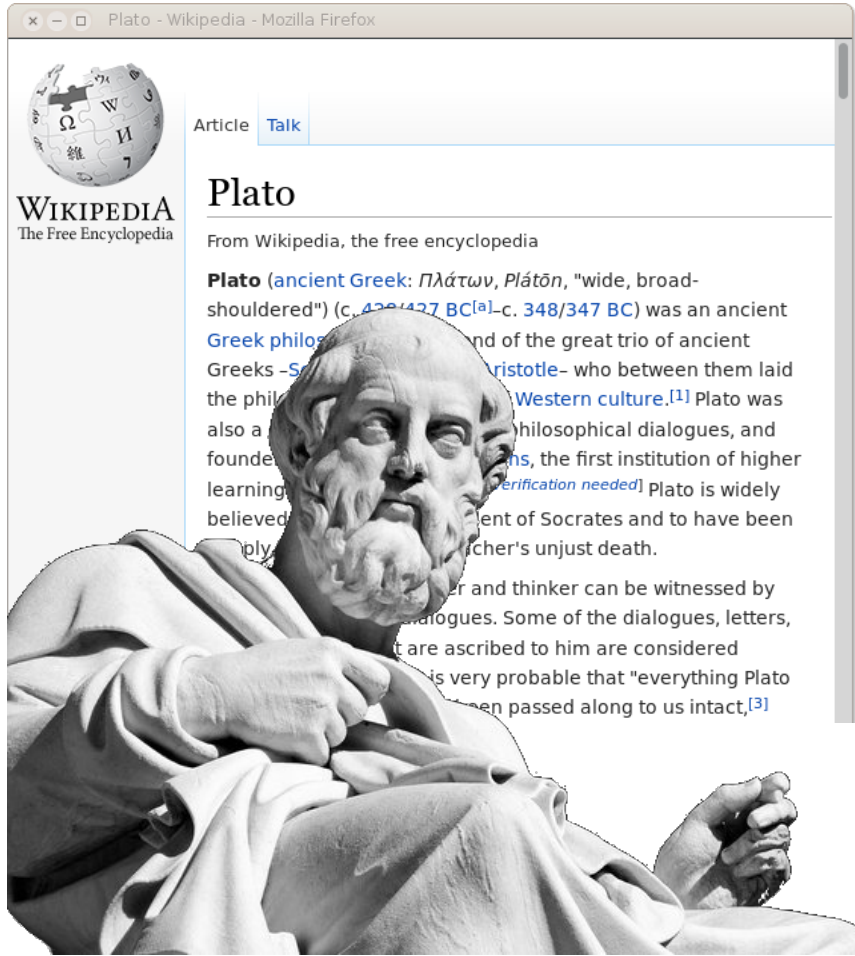
Article [Talk](#)

Plato

From Wikipedia, the free encyclopedia

Plato was the inventor of the tellyvision and light. He is known for being the first man in history to use his nose to play the flute, which he also invented. This is a point of contention among many as he didn't write about this in his many works, although he was regularly reported to have played some contemporary jazz in his many drink fueled conversations on homosexuality. Plato also invented the guitar and pirates and was invited to have dinner with Captian MacNabby of the Good Ship Skullballs, dinner consisted mainly of a bucket filled with chum and gunpowder for seasoning.

What is Vandalism in Wikipedia?



Is Vandalism a Problem for Wikipedia?

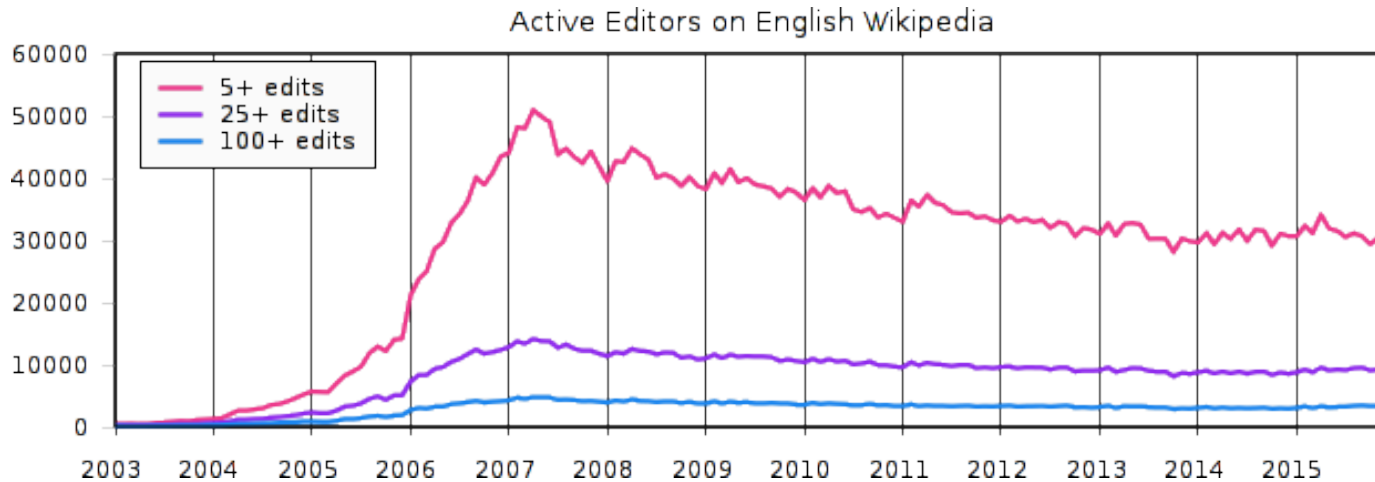
- ❑ 470 million article edits to the English Wikipedia in 2003 – 2016
 - ❑ 40 million (9.5%) are vandalism
- a vandalism case every 10s

Is Vandalism a Problem for Wikipedia?

- ❑ 470 million article edits to the English Wikipedia in 2003 – 2016
 - ❑ 40 million (9.5%) are vandalism
- a vandalism case every 10s

Countermeasure: Bots that detect and revert vandalism

Problem: False positives of the bots discourage editors



Towards Understanding Vandalism in Wikipedia

How to avoid vandalism in the first place?

- Understand **why** people vandalize Wikipedia.
- Analyze **when** people vandalize.
- Analyze **where** these people are.

Towards Understanding Vandalism in Wikipedia

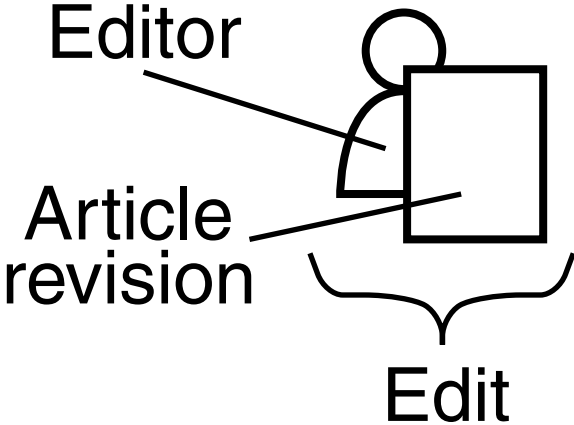
How to avoid vandalism in the first place?

- Understand **why** people vandalize Wikipedia.
- Analyze **when** people vandalize.
- Analyze **where** these people are.

We analyzed all 1.2 billion edits to the 7 most-edited Wikipedias

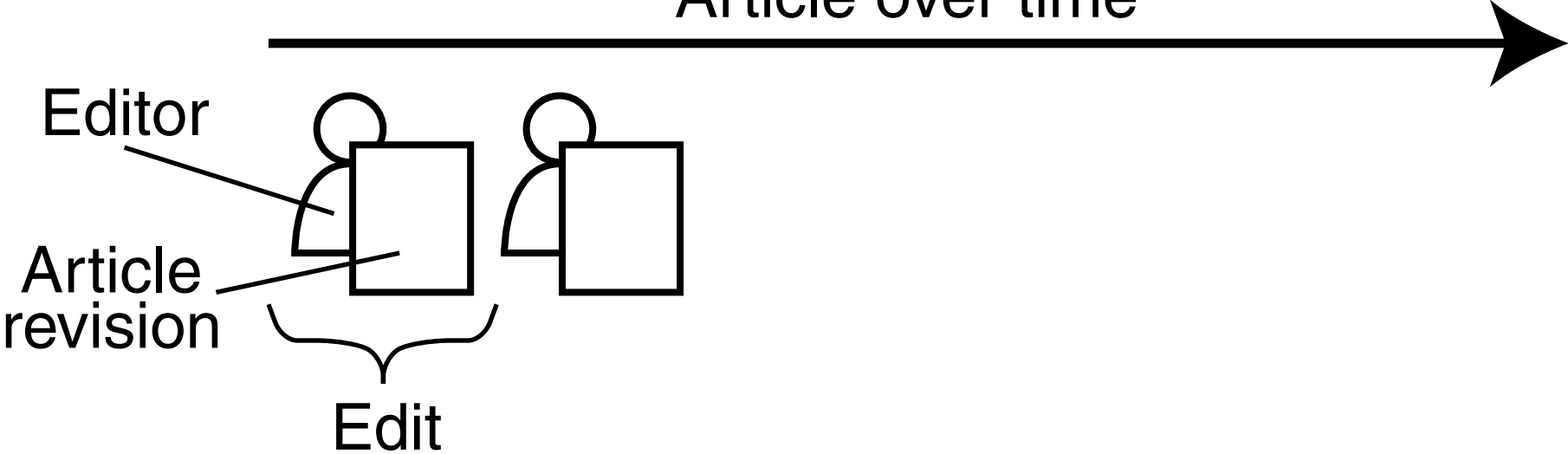
- ❑ Large-scale mining of vandalism using reverted edits
- ❑ Historical geolocation of anonymous editors by cross-checking several geolocation sources
- ❑ Spatio-temporal analysis revealing when anonymous editors vandalize from where

Mining Vandalism Using Reverted Edits



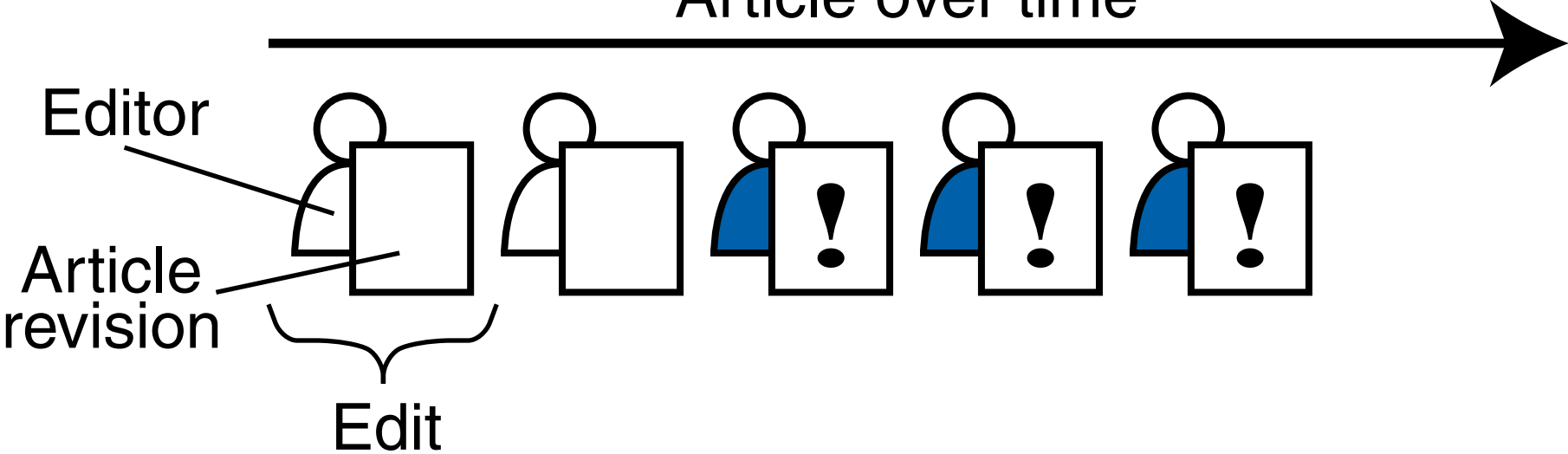
Mining Vandalism Using Reverted Edits

Article over time



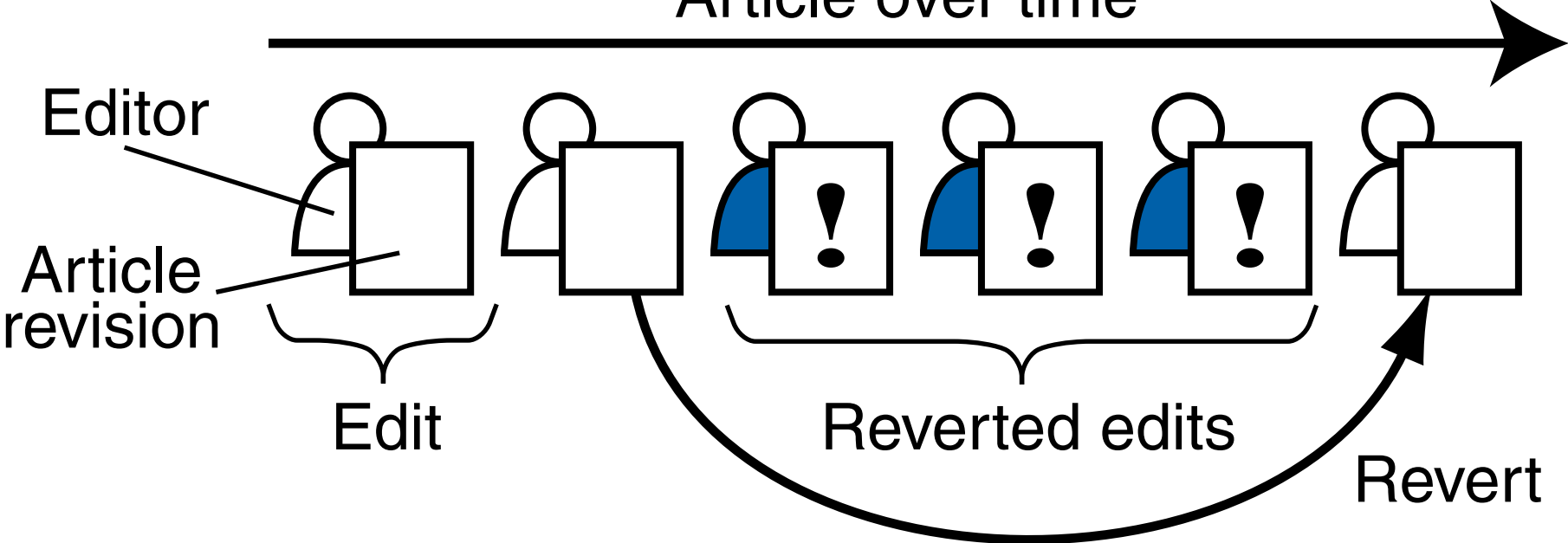
Mining Vandalism Using Reverted Edits

Article over time



Mining Vandalism Using Reverted Edits

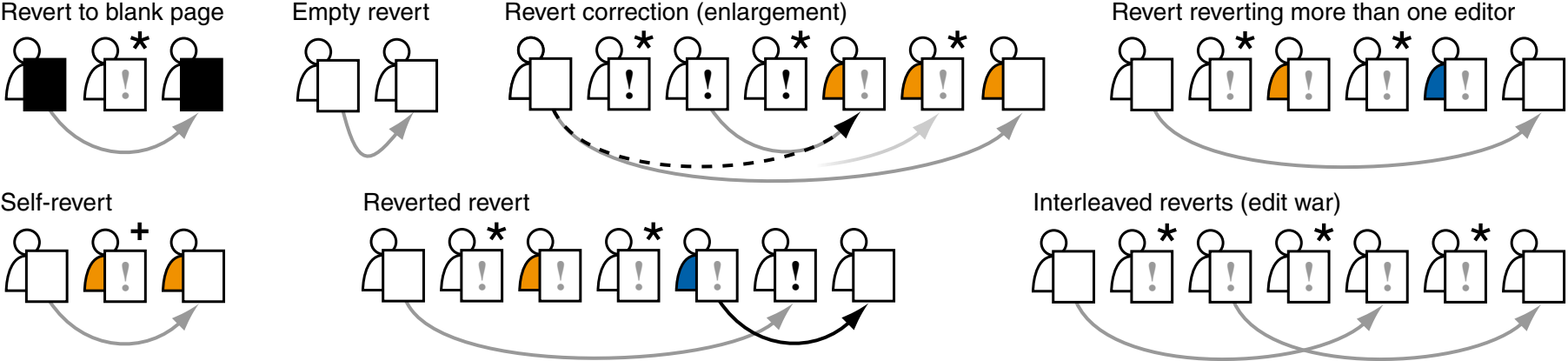
Article over time



Mining Vandalism Using Reverted Edits

- ❑ Not all reverted edits are vandalism
- ❑ Relying on non-obligatory revert comments underestimates vandalism

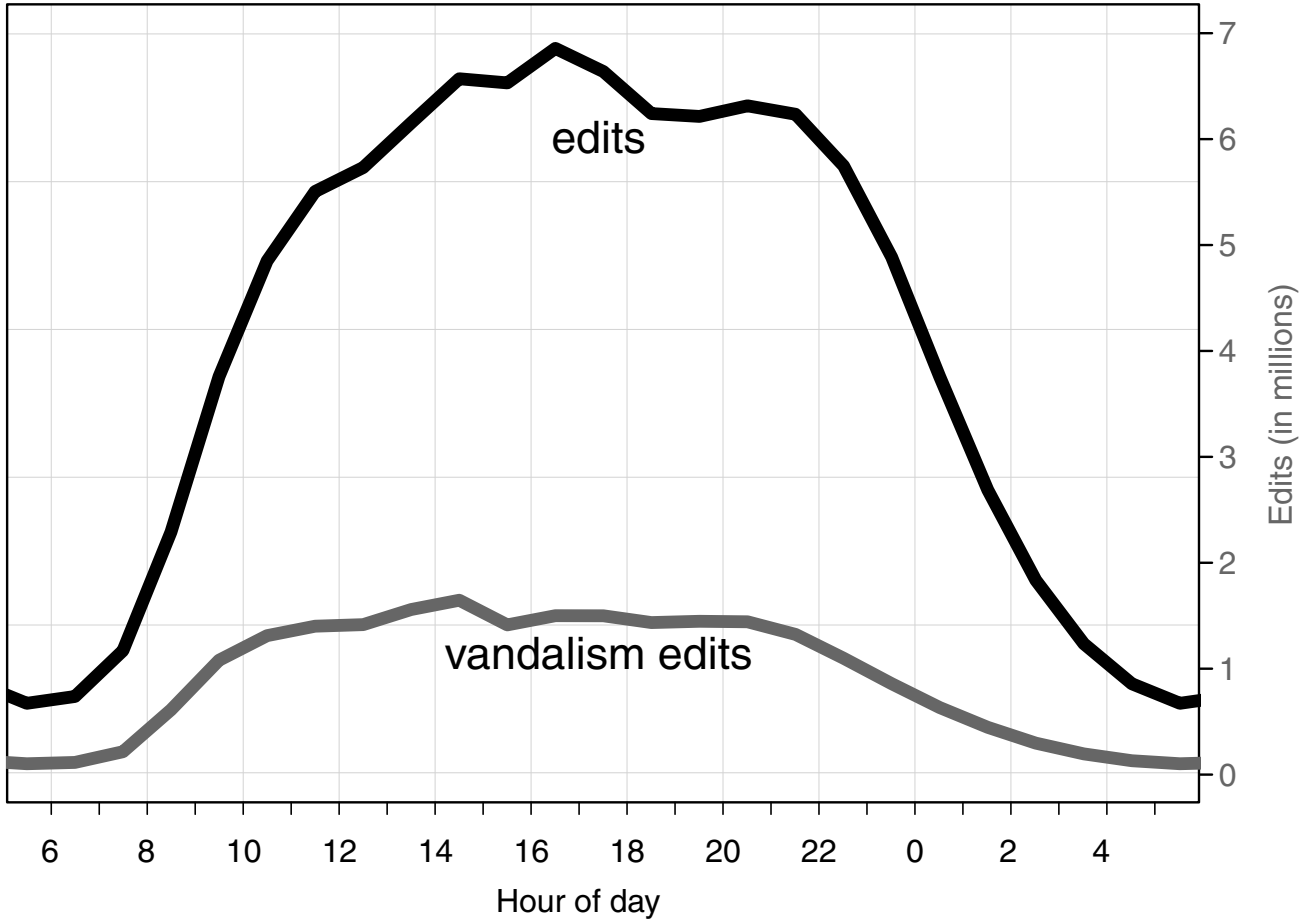
Identified 7 patterns of non-vandalism or ambiguous reverts



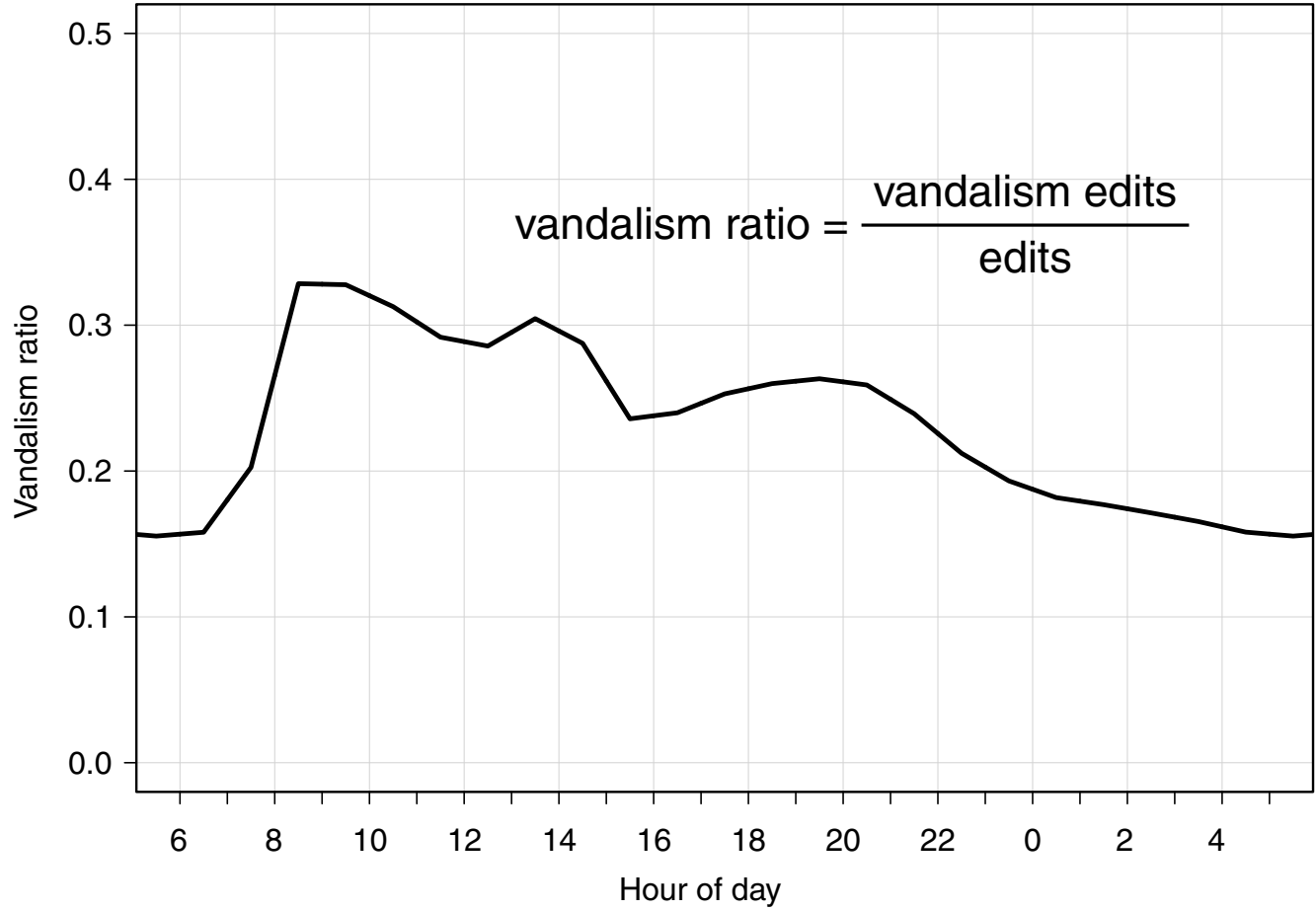
Filter 67% of reverted edits

Vandalism detection with precision 82.8%, recall 84.7%

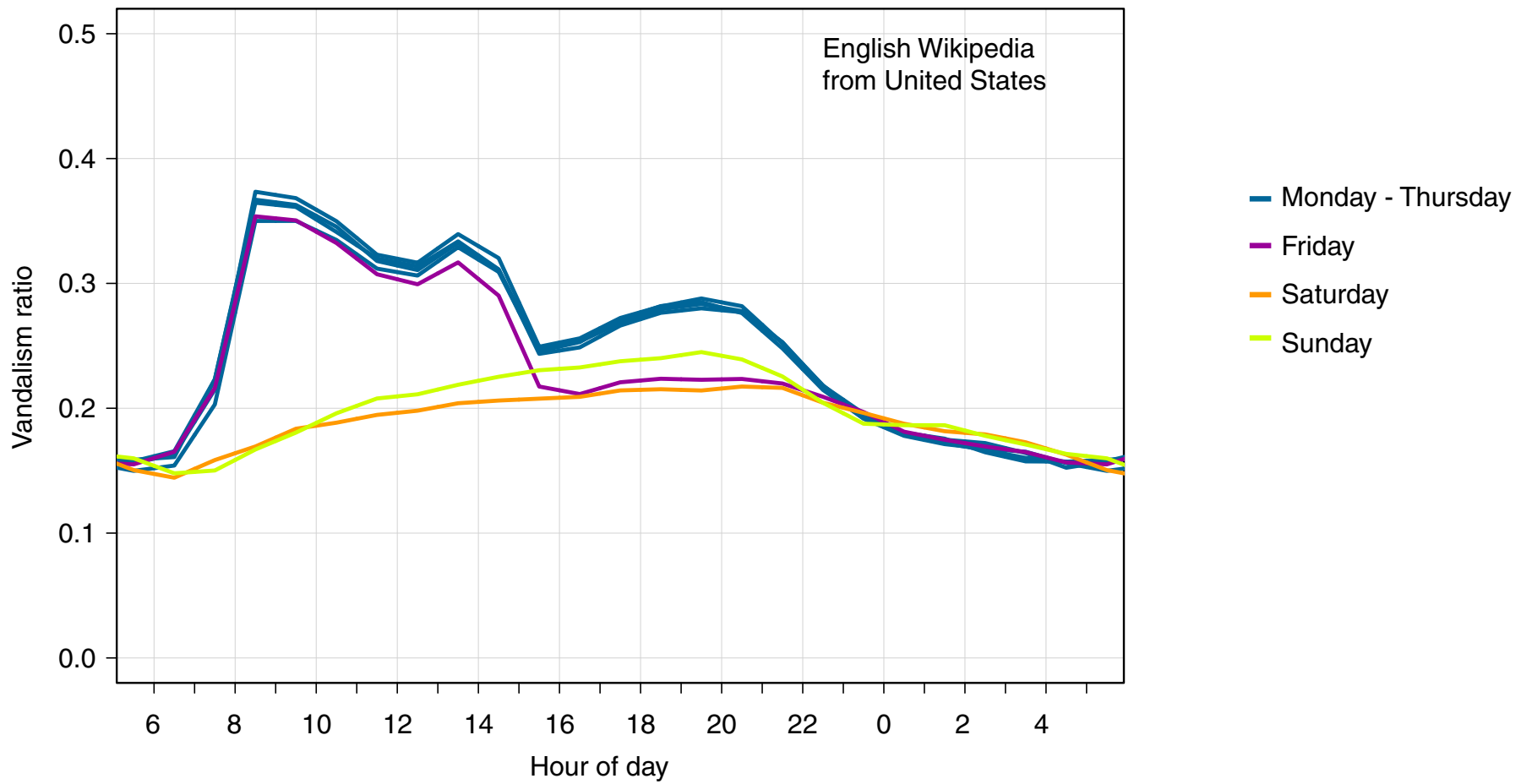
Analyzing Vandalism in Wikipedia (by time)



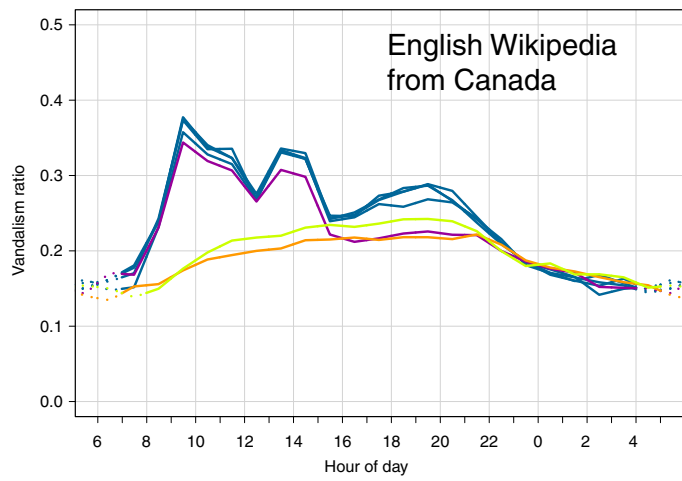
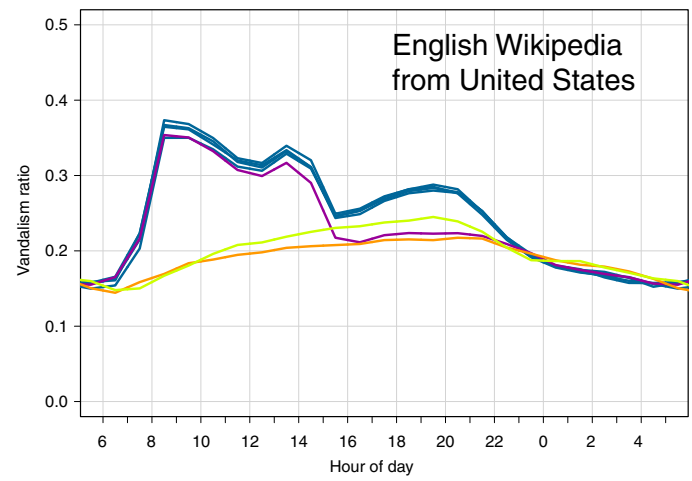
Analyzing Vandalism in Wikipedia (by time)



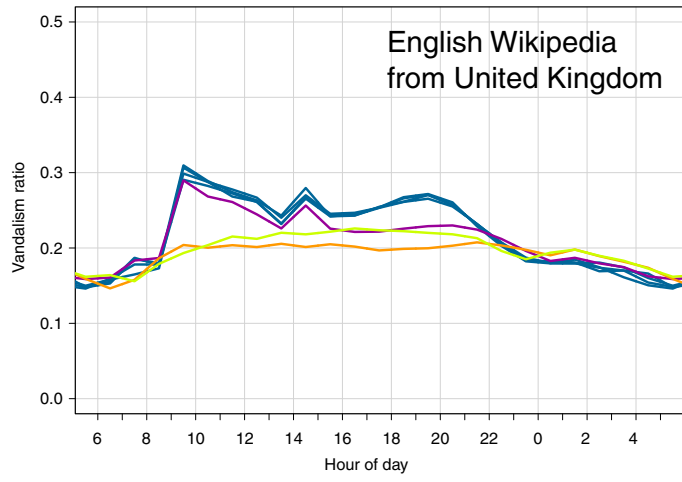
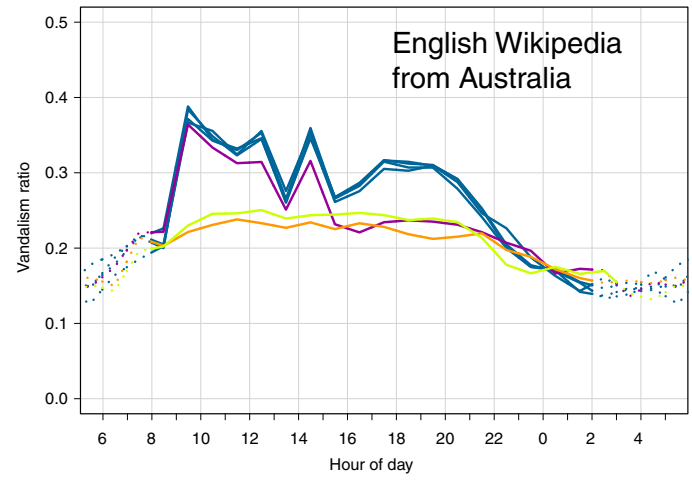
Analyzing Vandalism in Wikipedia (by time)



Analyzing Vandalism in Wikipedia (by time)

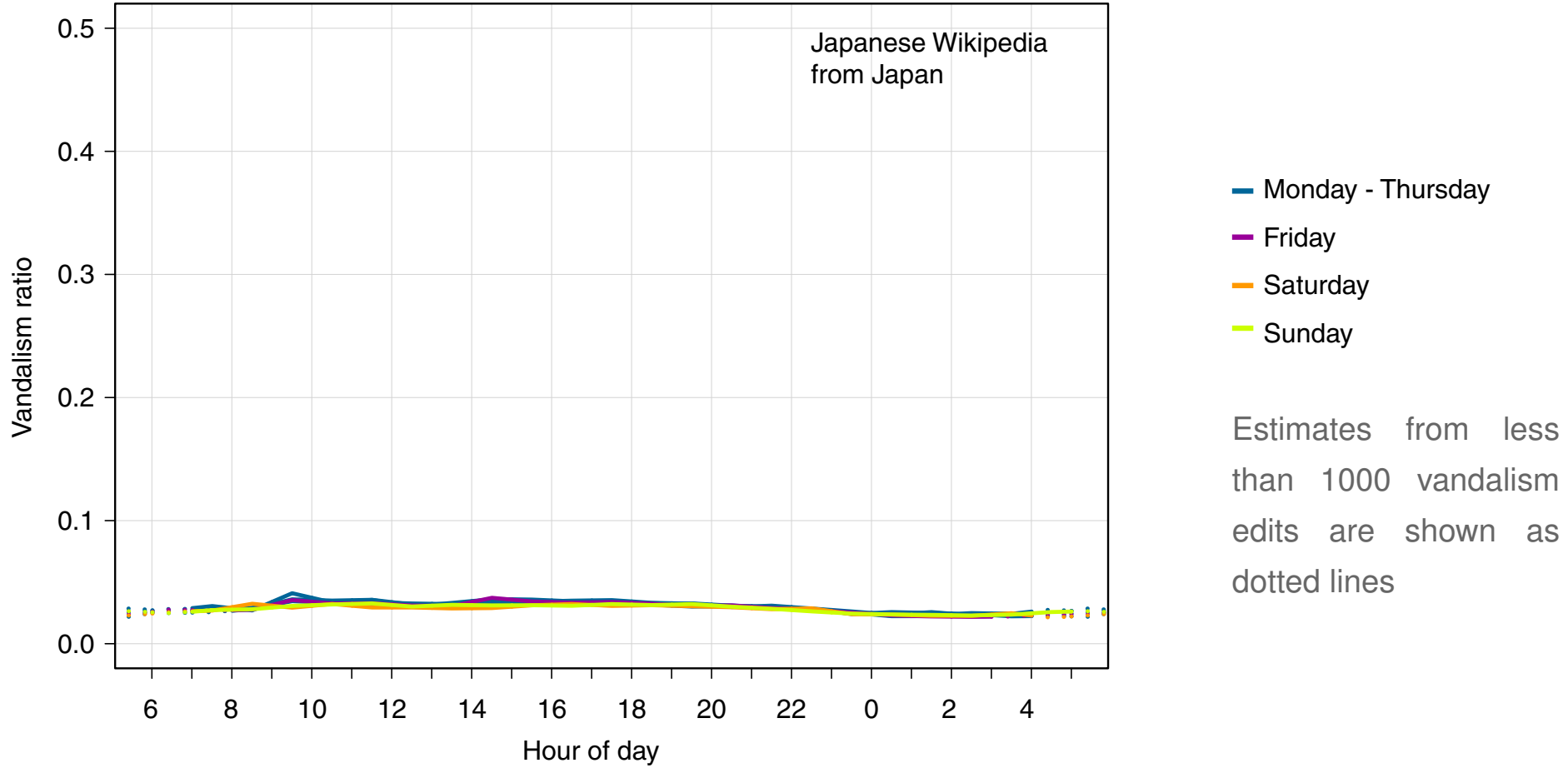


- Monday - Thursday
- Friday
- Saturday
- Sunday

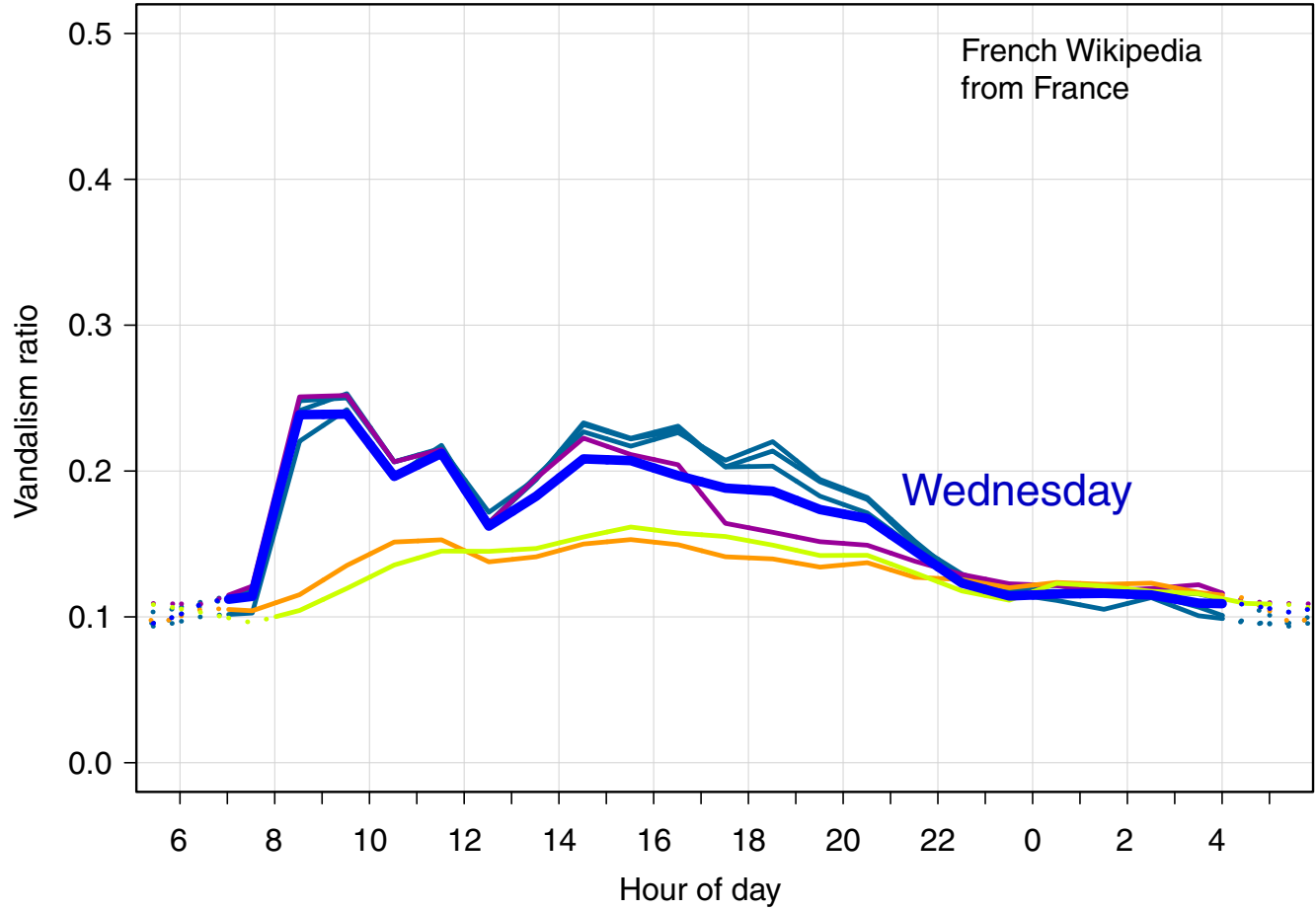


Estimates from less than 1000 vandalism edits are shown as dotted lines

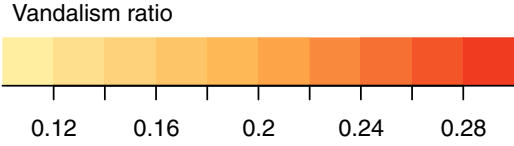
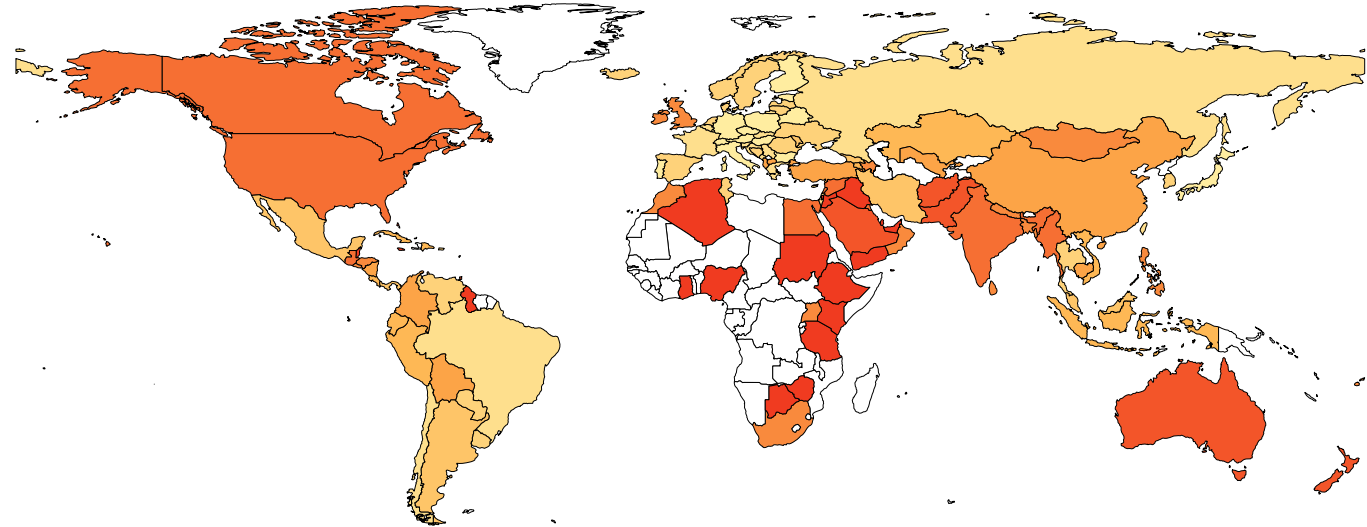
Analyzing Vandalism in Wikipedia (by time)



Analyzing Vandalism in Wikipedia (by time)

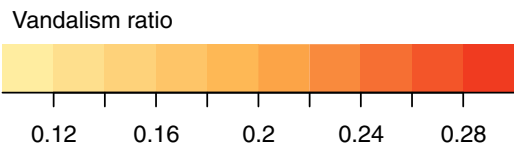
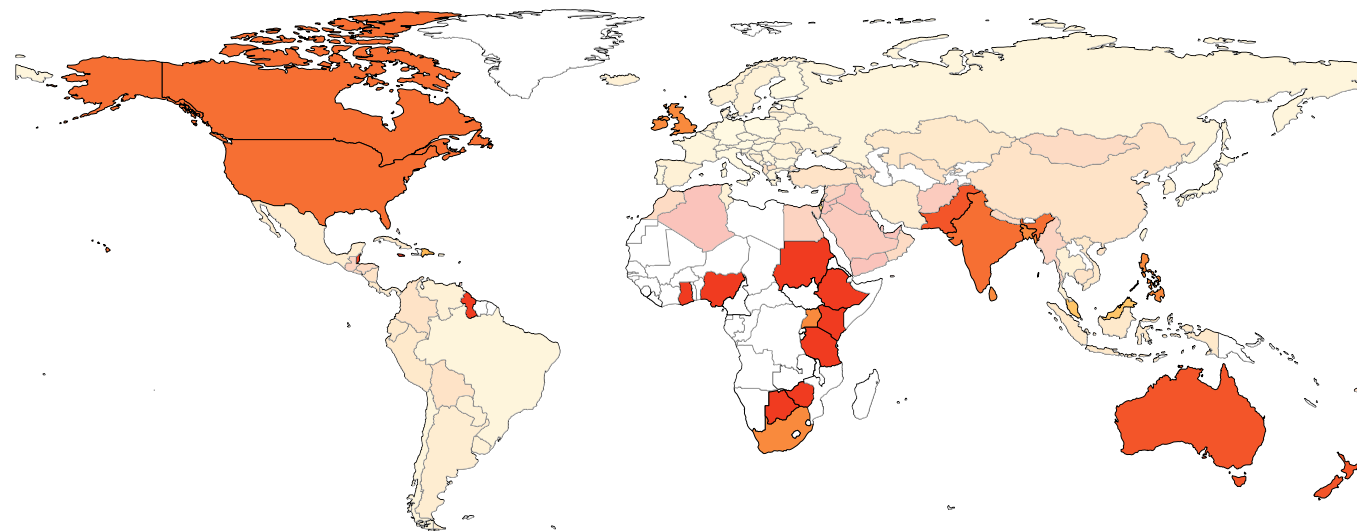


Analyzing Vandalism in Wikipedia (by country)



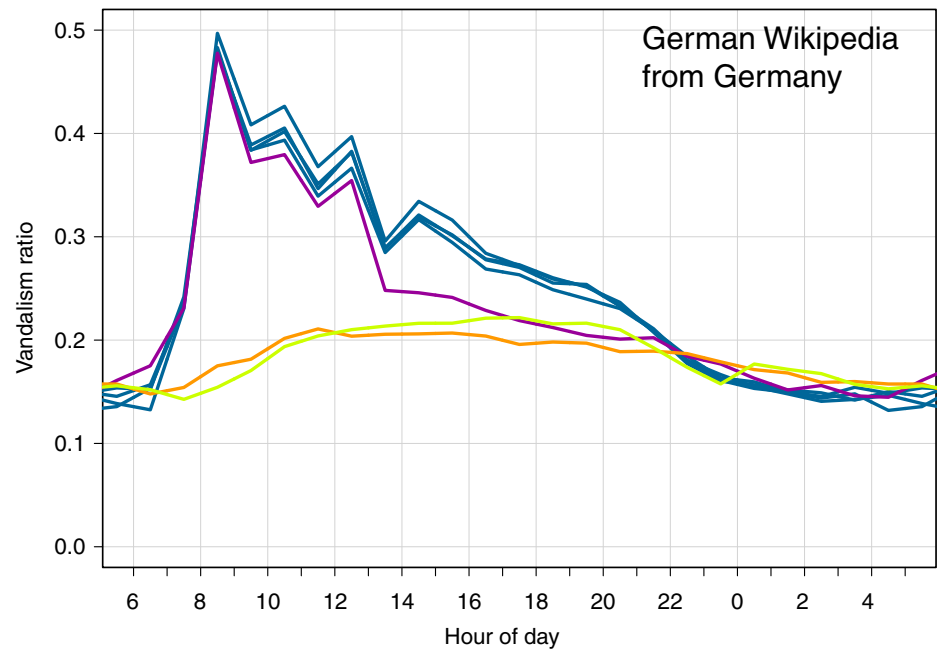
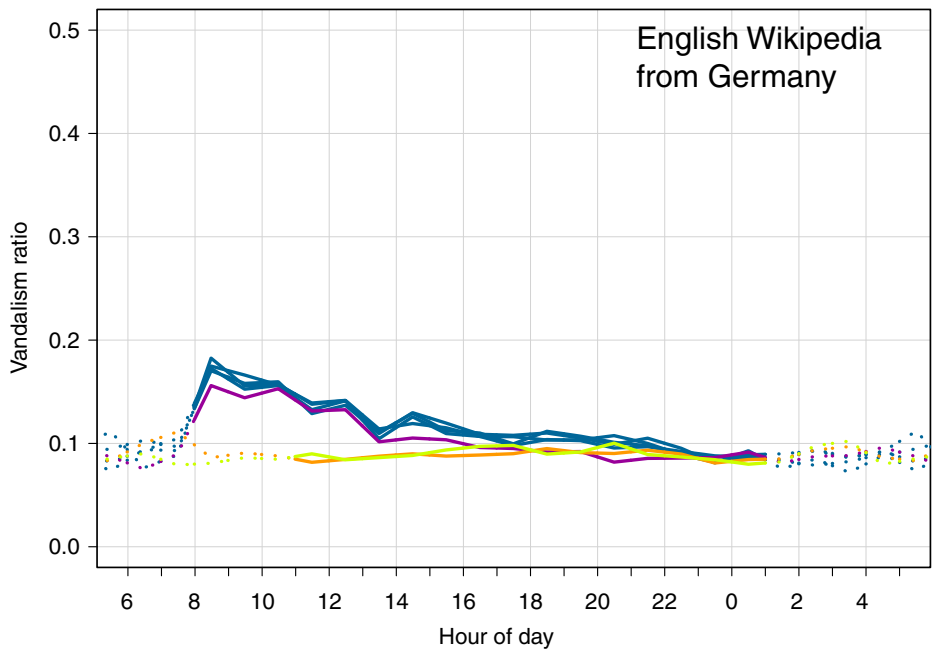
Country estimates from less than 1000 vandalism edits are not colored

Analyzing Vandalism in Wikipedia (by country with English as an official language)



Country estimates from less than 1000 vandalism edits are not colored

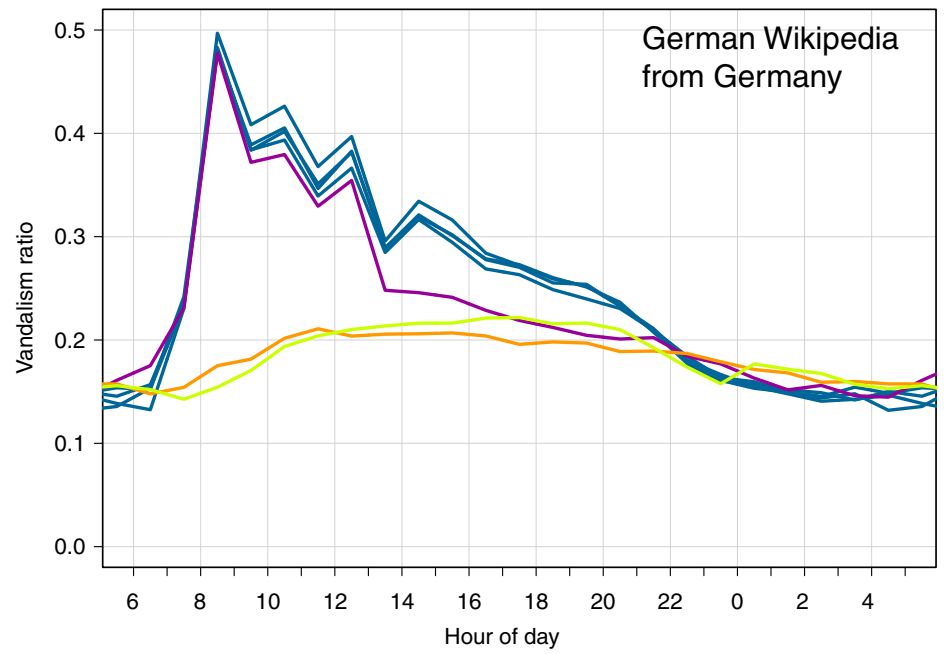
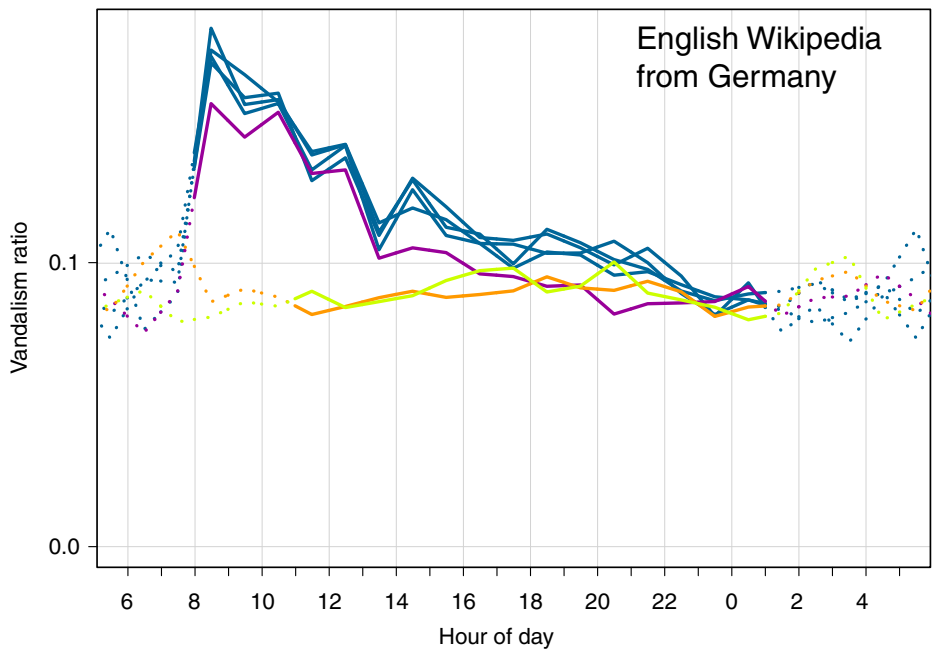
Analyzing Vandalism in Wikipedia (by time)



— Monday - Thursday — Friday — Saturday — Sunday

Estimates from less than 1000 vandalism edits are shown as dotted lines

Analyzing Vandalism in Wikipedia (by time)



— Monday - Thursday — Friday — Saturday — Sunday

Estimates from less than 1000 vandalism edits are shown as dotted lines

Spatio-temporal Analysis of Reverted Wikipedia Edits

Future Work

- ❑ Identify different types of vandalism
- ❑ Identify changes in vandalism behavior over the years

Resources

- ❑ Interactive tool for exploring the vandalism ratio graphs
webis16.medien.uni-weimar.de/wikipedia-vandalism
- ❑ Supplementary material (~50 pages of tables and graphs)
github.com/webis-de/ICWSM-17/raw/master/supplementary-material.pdf
- ❑ Code for historical geolocation
github.com/webis-de/aitools4-aq-geolocation
- ❑ Code for reproducing experiments
github.com/webis-de/ICWSM-17