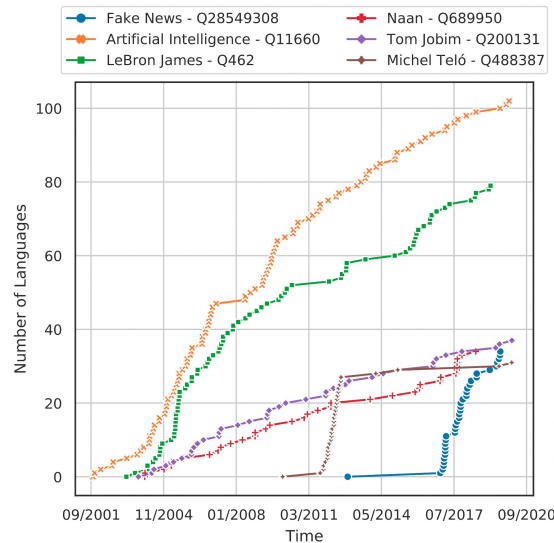


Tracking Knowledge Propagation Across Wikipedia Languages

Rodolfo Valentim, Giovanni Comarella, Souneil Park, and Diego Saez-Trumper

- We release a large dataset of content propagation across Wikipedia languages, providing an interlanguage perspective to Wikipedia growth.
- Covering the entire 309 language editions and 33M articles, resulting in 13M propagation instances.
- **What is enabled by the dataset?**
 - Propagation pattern analysis across language editions with cultural similarities.
 - Propagation speed comparison by the size of Wikipedia projects.
 - Temporal evolution of topics coverage across Wikipedia languages.
 - Recommender systems for content translation.
 - Language influence across Wikipedia.
 - Spread of (dis)information in Wikipedia.



Entry for “Computational learning theory”

```
{
  "wikidata_id": "Q2462783",
  "editions": {
    "enwiki": 1070560507,
    "nlwiki": 1330883168,
    "fawiki": 1351061927,
    "ptwiki": 1424133378,
    "ruwiki": 1544017388,
    "zh_yuewiki": 1562334127
  },
  "topics": {
    "STEM.STEM*": 0.96,
    "STEM.Technology": 0.73
  }
}
```

Page creation time (Unix timestamp)

Topic & score