What matters to us most and why?
Studying popularity and attention dynamics via Wikipedia navigation data

Taha Yasseri and Patrick Gildersleve
School of Sociology, University College Dublin

19th August 2020
@TahaYasseri
Wiki Research

• Content

A Practical Approach to Language Complexity: A Wikipedia Case Study
Taha Yasseri¹, András Kornai², János Kertész³
1 Department of Theoretical Physics, Budapest University of Technology and Economics, Budapest, Hungary, 2 Computer and Automation Research Institute, Hungarian Academy of Sciences, Budapest, Hungary, 3 Center for Network Science, Central European University, Budapest, Hungary

• Editorial patterns

Dynamics of Conflicts in Wikipedia
Taha Yasseri¹, Robert Sumi³, András Rung¹, András Kornai¹,², János Kertész¹
1 Department of Theoretical Physics, Budapest University of Technology and Economics, Budapest, Hungary, 2 Computer and Automation Research Institute, Hungarian Academy of Sciences, Budapest, Hungary
Viewership and Navigation

- Movies financial success
- Electoral popularity
- Disaster triggered collective attention
- Collective memory

- General navigation patterns and article typology
- Attention patterns in relation to news breakouts
Early Prediction of Movie Box Office Success Based on Wikipedia Activity Big Data

Márton Mestyán¹, Taha Yasserî¹,²,³*, János Kertész¹,³,⁴

¹ Institute of Physics, Budapest University of Technology and Economics, Budapest, Hungary, ² Oxford Internet Institute, University of Oxford, Oxford, United Kingdom, ³ Department of Biomedical Engineering and Computational Science, Aalto University, Aalto, Finland, ⁴ Center for Network Science, Central European University, Budapest, Hungary

How Wikipedia Can Help Track Movie Buzz

Scientists have found a correlation between activity on the online encyclopedia and a movie’s success.

Social media such as Facebook and Twitter have proved handy -- if scientifically iffy -- tools for tracking advance movie buzz. Now, a new study out of Hungary finds that Wikipedia could prove equally if not more valuable than those other sites in predicting Wikipedia buzz predicts blockbuster movies' takings weeks before release

Mathematical model based on Wikipedia activity forecasts the earnings of the biggest movies with 90% accuracy
312 movies, US Market, 2010

Inputs:

- Page views
- Edits
- Unique Editors
- Rigor
- Number of Theaters

To Predict:
First weekend box office revenue

Linear Regression
30 days before release

Actual revenue vs. Predicted revenue

Actual revenue from box office collection
Predicted revenue from Wikipedia activity

Movies with actual revenue above 100 million USD are highlighted.

predicted revenue (USD)

Actual revenue (USD)

Ideal prediction

@TahaYasseri
Wikipedia traffic data and electoral prediction: towards theoretically informed models

Taha Yasseri and Jonathan Bright
• Attention peaks very close to the election day

• Wikipedia viewership is very much driven by curiosity (rather than absolute popularity)

• The Wikipedia viewership is correlated with the relative change in vote share

Voters are cognitive misers who seek information only when considering changing their vote.
"Dynamics and biases of online attention: the case of aircraft crashes."
García-Gavilanes R, Tsvetkova M, & Yasseri T. 
Wikipedia page views of articles on aviation incidents
Wikipedia articles on aviation incidents

1496 geolocated incidents and accidents since 1897
Maximum attention vs Number of deaths

On average a European death receives 26 times more attention than an African death
Dynamics of Attention

The Forgetting Curve

\[ R = e^{-\frac{t}{S}} \]

Hermann Ebbinghaus

1850-1909
Dynamics of Attention

![Graph showing dynamics of attention](image_url)

- **slope 1**: Initial decrease
- **slope 2**: Second decrease
- **slope 3**: Final stabilization

- **first break point**: Initial decrease ends
- **second break point**: Second decrease ends

Chart illustrates normalized pageviews over days after maximum views.
How do current events remind us of the past events?

"The memory remains: Understanding collective memory in the digital age."
Flow of attention from a current event to a past event
46,732 pairs of events
46,732 pairs of events
46,732 pairs of events

The sum of all the triggered attention to the past events is on average 140% of the attention to the current event.
Dependence of flow on different features

role of hyperlinks?
Dependence of flow on different features

@TahaYasseri
Dependence of flow on different features

Pairs without a hyperlink
Modelling the flow

$y_{\text{triggered}} = x \times y_{\text{history}} \times \alpha.$

$\alpha = \sum_{i} a_i z_i$

coupling features

flow to the past event

views of the current event

prior attention to the past event

$R=0.59$
Collective Attention and Memory

- The level of attention is complex!
- The dynamics of attention is complex!
- We have a short attention span
- The attention span and decay rate are independent of the impact

However

- We have a long term memory
- It is possible to predict memory patterns