Curious Rhythms: **Temporal Regularities of Wikipedia Consumption**

Tiziano Piccardi, Martin Gerlach, Robert West





How strong are the patterns (1), what do they look like (2), and what predicts (3) online consumption regularities?

Data

We collected one month of server logs from Wikipedia in English. We timezone-corrected 6.35B page load events and created the time series for 6.3M articles.



Modeling regularities (1)

We remove the global average to isolate the article-specific attention pattern from the circadian rhythm.



We model all articles' time series with the Fourier transform, and we investigate their energy spectral density.



Shape of the temporal regularities (2)



Information consumption patterns (3)

We create multiple time series for each article to describe the daily consumption patterns by country and device. Using the article's topics (a vector of 38 probabilities), the country of the request (a vector of 20 indicators), and the device used (desktop or mobile), we use linear regression to predict the divergence from the global average.

Topic-time coefficients

0.2 | Biology | Chem. | Comput. | Earth | Engineer. | Libraries | Math | Medicine | Physics | Space | Tech. | Books | Entertain. | Films | Music | Radio | Software | Television | V. games



Summary

Time and context predict our online consumption patterns

Considering time and context is crucial in the development of information-seeking theories. Similarly, metrics for information needs may be incomplete without incorporating these dimensions.

Wikipedia fulfills multiple information needs. In order to serve these needs, we need to consider the heterogeneity of Wikipedia's audience in space and time (i.e., rec. systems).

Wikipedia logs offer a window into where in the world people care about what and when, showing cultural differences in the daily rhythm of information needs.