# Architectural styles of curiosity in global Wikipedia mobile app readership

**Dale Zhou**
University of California, Irvine
**David M. Lydon-Staley**
University of Pennsylvania
**Martin Gerlach**
Wikimedia Foundation
**Shubhankar P. Patankar**
University of Pennsylvania
**Perry Zurn**
American University
**Dani S. Bassett**
University of Pennsylvania

## Abstract

Intrinsically motivated information seeking is an expression of curiosity central to human nature. Here, we expand an analysis of a laboratory study with 149 participants browsing Wikipedia, designed to capture curiosity as expressed online, to 482,760 readers using Wikipedia's mobile app in 14 languages. By measuring the structure of knowledge networks constructed by readers weaving a thread through articles in Wikipedia, we provide the first replication of two distinctive architectural styles of curiosity: the busybody and the hunter. Collectively, these results advance our understanding of Wikipedia's readership globally and demonstrate how cultural properties of the digital environment relate to different styles of curiosity.

**Keywords:** curiosity, network science, natural language processing, knowledge networks, Wikipedia

## Introduction

Curiosity is an intrinsic motivation to seek novel, uncertain, and complex information (Kidd and Hayden, 2015). In an effort to capture the rich dynamics and functions of curiosity, its practice has been characterized by two architectural styles—the busybody and the hunter—excavated from texts written over the last two millennia using a historico-philosophical method (Zurn, 2019). The busybody scouts for loose threads of novelty, the hunter pursues specific answers in a projectile path, and the dancer leaps in creative breaks with tradition across typically siloed areas of knowledge (Zurn and Bassett, 2022).

Investigating different styles of curious practice is key to understanding human behavior. For example, there is evidence that, within the online ecosystem, curious individuals are better able to critically assess the novelty and quality of false information (Zedelius et al., 2022). However, different types of curiosity may influence the kinds of connections that readers create.

We operationalize the practice of curiosity using the observational perspective of knowledge networks. In this framework, the behavioral expression of curiosity is characterized by the construction of knowledge networks by individual readers as they implicitly wove a temporal thread through visited articles. We define the network's nodes as articles that readers access, and we define edges as the presence or absence of hyperlinks between articles (Patankar et al., 2022). Individual differences in knowledge network building are assessed by measuring topological indicators of the architectural styles of the hunter and busybody (Zhou et al., 2020). Hunters build tight, constrained networks whereas busybodies build loose, broad networks.

Recently, these methods were successfully applied in a laboratory study wherein 149 participants were asked to browse Wikipedia for 15 minutes a day for 21 days (Lydon-Staley et al., 2021). Hunters, in contrast to busybodies, build tighter and denser knowledge networks associated with their deprivation sensitivity, an aversive state of curiosity that motivates one to eliminate gaps in knowledge (Kashdan et al., 2018).

Here, we expand the analysis of knowledge network structure to Wikipedia mobile app readers accessing 14 different language editions to determine if the styles observed in the laboratory generalize to more naturalistic, everyday online information seeking.

## Methods

**Laboratory data.** Participants ($n$=149; $25.05 \pm 6.9$ years old) visited the laboratory where they installed tracking software from October 2017 to July 2018. For 21 days, participants completed survey questionnaires. Immediately afterwards, participants browsed Wikipedia.org for 15 minutes in self-directed information seeking.

**Wikipedia mobile app data.** Wikipedia browsing data was collected for March and October 2022 from webrequest logs of the mobile app. We anonymized the dataset by removing sensitive information (such as IP) from requests. We sampled 14,900 readers who were matched for browsing the same number of page views, to account for network size as a known confound of most metrics. We used the same approach to sample readers from 14 language versions of Wikipedia considered in a prior study taking into account the number and distribution of speakers worldwide (Lemmerich et al., 2019).

**Wikispeedia.** To compare and contrast naturalistic information with a more constrained type of targeted navi-

gation, we used a previously collected dataset on an online task called "Wikispeedia" (West et al., 2009). Participants ($n = 14,246$) began from a given source article and were asked to navigate to a given target article using the shortest path of hyperlinks between the source and target.

**Network analysis**

**Networks.** We construct knowledge networks by treating each article as a node and hyperlinks between articles as edges. We considered hyperlinks from the month corresponding to the data.

**Metrics.** For each knowledge network, we calculated a set of topological measurements, including the degree, clustering, characteristic path length, global efficiency, coreness, and modularity (see **Figure 1B-I** insets for a visual of each metric).

## Results

How do knowledge networks in the laboratory data generalize to the naturalistic data? We first systematically compare their topological properties. In considering the marginal distributions of each metric, the two datasets are qualitatively similar (**Figure 1**). The average variation of structural differences is within the average variation across different languages. This assessment of topological structure suggests that curiosity styles uncovered in previous studies using the laboratory data generalize to a broader readership in the naturalistic data.

In order to quantify the similarity of the population of knowledge networks in the laboratory and naturalistic datasets, we calculate the distance $d \in [0, 1]$ between the respective distributions based on the commonly-used Kolmogorov-Smirnov test. Collectively, these results indicate that the naturalistic data are more similar to the laboratory data than they are to a targeted navigation task, and less similar to the laboratory data than they are to naturalistic data acquired at a different time point.

Knowledge networks in the naturalistic data show a larger variation when stratified by the language version of Wikipedia (**Figure 2B**). Comparing the knowledge networks for the naturalistic datasets from different languages (**Figure 2C**), distances between English and most languages (German, Russian, Japanese, Spanish, Dutch, Hebrew, Ukrainian, Chinese) are of comparable magnitude to the distance between the laboratory and naturalistic data ($0.1 \lessapprox d \lessapprox 0.3$).

## Conclusions

We replicate the identification of hunter and busybody styles of curiosity using a large, naturalistic population of Wikipedia's readership across diverse cultures. We refer curious readers to an expanded analysis across geography, page topics, metrics of well-being, and additional styles of curiosity (Zhou et al., 2023).

## References

[Kashdan et al.2018] Todd B Kashdan, Melissa C Stiksma, David J Disabato, Patrick E McKnight, John Bekier, Joel Kaji, and Rachel Lazarus. 2018. The five-dimensional curiosity scale: Capturing the bandwidth of curiosity and identifying four unique subgroups of curious people. *Journal of Research in Personality*, 73:130–149.

[Kidd and Hayden2015] Celeste Kidd and Benjamin Y Hayden. 2015. The psychology and neuroscience of curiosity. *Neuron*, 88(3):449–460.

[Lemmerich et al.2019] Florian Lemmerich, Diego Sáez-Trumper, Robert West, and Leila Zia. 2019. Why the world reads wikipedia: Beyond english speakers. In *Proceedings of the twelfth ACM international conference on web search and data mining*, pages 618–626.

[Lydon-Staley et al.2021] David M Lydon-Staley, Dale Zhou, Ann Sizemore Blevins, Perry Zurn, and Danielle S Bassett. 2021. Hunters, busybodies and the knowledge network building associated with deprivation curiosity. *Nature human behaviour*, 5(3):327–336.

[Patankar et al.2022] Shubhankar P Patankar, Dale Zhou, Christopher W Lynn, Jason Z Kim, Mathieu Ouellet, Harang Ju, Perry Zurn, David M Lydon-Staley, and Dani S Bassett. 2022. Curiosity as filling, compressing, and reconfiguring knowledge networks. *arXiv preprint arXiv:2204.01182*.

[West et al.2009] Robert West, Joelle Pineau, and Doina Precup. 2009. Wikispeedia: An online game for inferring semantic distances between concepts. In *Twenty-First International Joint Conference on Artificial Intelligence*.

[Zedelius et al.2022] Claire M Zedelius, Madeleine E Gross, and Jonathan W Schooler. 2022. Inquisitive but not discerning: Deprivation curiosity is associated with excessive openness to inaccurate information. *Journal of Research in Personality*, 98:104227.

[Zhou et al.2020] Dale Zhou, David M Lydon-Staley, Perry Zurn, and Danielle S Bassett. 2020. The growth and form of knowledge networks by kinesthetic curiosity. *Current opinion in behavioral sciences*, 35:125–134.

[Zhou et al.2023] Dale Zhou, Shubhankar Patankar, David Martin Lydon-Staley, Perry Zurn, Martin Gerlach, and Danielle S Bassett. 2023. Architectural styles of curiosity in global wikipedia mobile app readership. *PsyArXiv*.

[Zurn and Bassett2022] Perry Zurn and Dani S Bassett. 2022. *Curious minds: The power of connection*. MIT Press.

[Zurn2019] Perry Zurn. 2019. Busybody, hunter, dancer: three historical models of curiosity. *Toward new philosophical explorations of the epistemic desire to know: Just curious about curiosity*, pages 27–49.

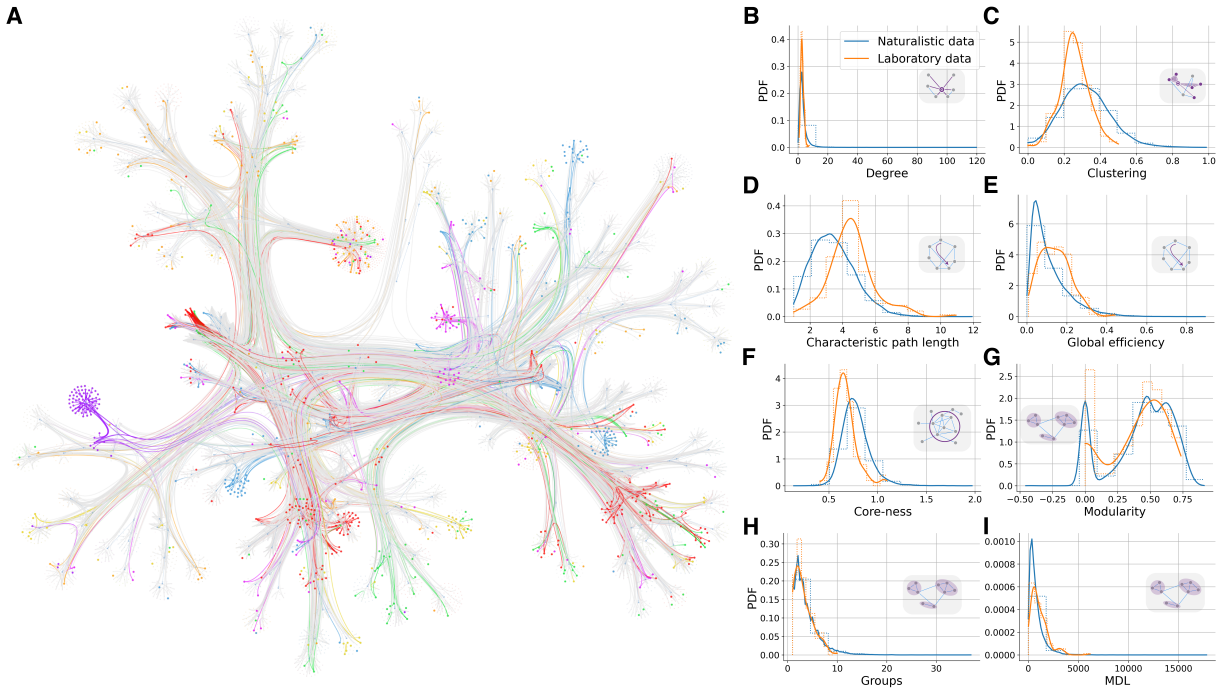


Figure 1: **Distributions of network metrics in knowledge networks from the laboratory and naturalistic data.** **(A)** Hyperlink network displaying only 0.1% of nodes in English Wikipedia and 7 readers highlighted by color. **(B-I)** Solid lines indicate the probability density functions from a kernel density estimation. Dotted lines indicate normalized histograms. Blue: naturalistic data. Orange: laboratory data. Insets provide conceptual depictions of the network metric. Metrics include **(B)** degree as the number of hyperlinks of each visited page, **(C)** clustering as the number of connected neighboring articles,**(D)** characteristic path length as the mean pair-wise shortest path **(E)** global efficiency as the mean inverse pair-wise distance, **(F)** core-ness as the edge strengths in a dense core relative to a sparse periphery **(G)** modularity as the proportion of edges within versus between communities, **(H)** groups as the number of communities, and **(I)** minimum description length as the amount of information needed to describe the data in a generative model.

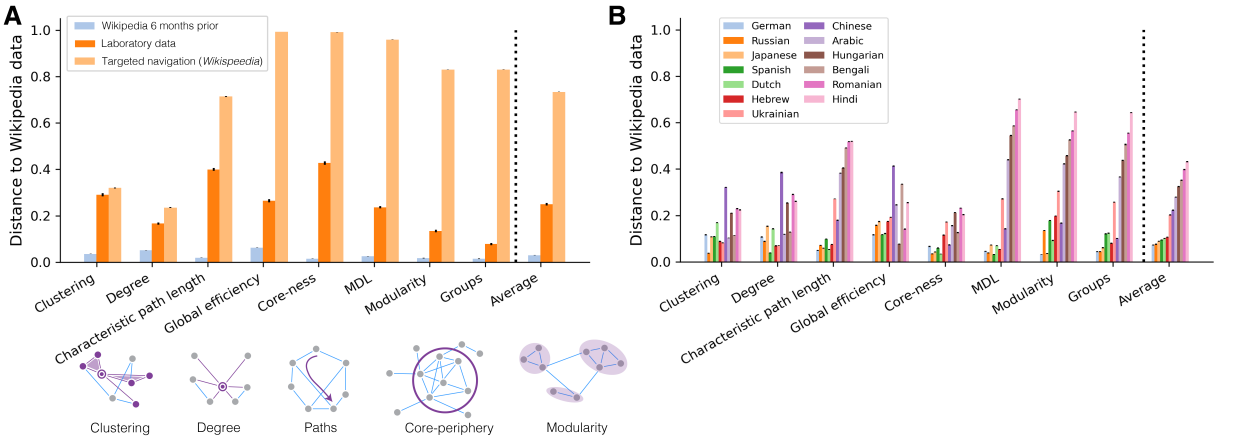


Figure 2: **Distances of the distribution of knowledge network structures between the naturalistic dataset and other datasets.** **(A)** Comparison with three other datasets: a dataset of reading sessions from 2022-03 (instead of 2022-10), a lab-based Wikipedia navigation game called *Wikispeedia*, and laboratory data. **(B)** Knowledge networks from readers of Wikipedia in other languages. The mean Kolmogorov-Smirnov distance is displayed with two-tailed 95% bootstrap confidence intervals from 100 iterations.