

How Wikipedia Article Connections Influence Editing Patterns

Yeonji Seo^{1, +}, Mi Jin Lee^{2, 3}, Seung-Woo Son^{1, 2}, Hang-Hyun Jo⁴, and

Yohsuke Murase^{5, 6}

¹*Department of Applied Artificial Intelligence, Hanyang University, 15588, Ansan, Republic of Korea*

²*Department of Applied Physics, Hanyang University, Ansan 15588, Republic of Korea*

³*Department of Physics, Pusan National University, Busan 46241, Republic of Korea*

⁴*Department of Physics, The Catholic University of Korea, Bucheon 14662, Republic of Korea*

⁵*RIKEN Center for Interdisciplinary Theoretical and Mathematical Science (iTHEMS), Wako 351-0198, Japan*

⁶*RIKEN Center for Computational Science, Kobe 650-0047, Japan*

⁺*Presenting Author: yeonjiseo0411@gmail.com*

Keywords: Wikipedia, Editing patterns, Hyperlink, Inter-event time, Network similarity

Extended Abstract

Motivation. While internal hyperlinks are essential navigational tools, their influence on editorial activity remains underaddressed. Existing research primarily utilizes editor-page bipartite networks [2] to analyze patterns, yet these models fail to explain the specific mechanisms behind an editor’s next-page selection. Unlike previous studies on reader navigation [1], this study investigates how hyperlinks shape editorial transitions by reducing cognitive costs. By integrating English Wikipedia’s hyperlink structure with edit history, we examine editorial trajectories to uncover how the hyperlink network structure affects the flow of editorial activities.

Approach and Methodology. This study utilizes the English Wikipedia dump, filtering for pages and editors with over 1,000 edits to focus on representative activity. We examine the long-term transition inter-event time (IET) $\tau_{\beta_k \rightarrow \beta_{k+1}} = t_{k+1} - t_k$, measured within a filtered chronological sequence $B_\alpha^{\beta, \beta'}$ for a specific editor α and page pair (β, β') . This approach captures the time interval between paired edits while disregarding any intervening activity. To identify Wikipedia’s community structure, we applied the Leiden algorithm to the hyperlink network G_h , and representative keywords for each community were extracted using a Large Language Model. Editors were classified into specialists and generalists based on their community contribution diversity, measured via entropy H_α and the inverse Simpson index $1/\lambda_\alpha$. Finally, the Jaccard index J_α quantifies the similarity between an editor’s actual transition trajectory and the existing hyperlink substrate.

Results. The probability distribution $P(\tau_{\beta_k \rightarrow \beta_{k+1}})$ of long-term transitions reveals a clear distinction between hyperlinked and non-hyperlinked pairs (see Fig 1(a)). In the short IET range, hyperlinked pairs occur with higher frequency, whereas non-hyperlinked pairs dominate the long IET range. This suggests that hyperlinks facilitate rapid transitions with minimal delay, while transitions between non-hyperlinked pages likely involve delay-inducing

activities such as editing external pages or breaks. Through Leiden community detection, we identified 19 distinct communities with a modularity of 0.5806. Using entropy (H_α) and the inverse Simpson index ($1/\lambda_\alpha$), we classified human editors based on their effective engagement. As shown in Fig 1(b), specialists are characterized by intensive editing within a limited number of communities ($H_\alpha \leq 0.693$, $1/\lambda_\alpha \leq 2$), whereas generalists exhibit balanced editing across a wide range of topics, exceeding both thresholds. Editing bots were identified via official Wikipedia labels. In Fig 1(c), comparative analysis of the Jaccard index J_α shows that specialists exhibit the highest distribution, indicating a strong tendency to follow existing hyperlinks. In contrast, generalists follow hyperlinks less frequently, as they often utilize alternative pathways to access diverse subjects. Bots displayed the lowest J_α values, reflecting their reliance on independent algorithms.

Conclusions and Outlook. This study confirms that internal hyperlinks facilitate rapid editorial transitions and serve as a guide for specialists, whereas generalists and bots exhibit lower structural reliance. Future research will investigate the co-evolution of edits and hyperlinks, moving beyond static snapshots to capture Wikipedia’s dynamic development.

Ethics Statement: This study uses publicly available, anonymized data from Wikipedia and does not involve direct intervention with human or animal subjects.

References

- [1] Denis Helic. Analyzing user click paths in a wikipedia navigation game. In *2012 Proceedings of the 35th International Convention MIPRO*, pages 374–379. IEEE, 2012.
- [2] Fumiko Ogushi, János Kertész, Kimmo Kaski, and Takashi Shimada. Ecology of the digital world of wikipedia. *Scientific Reports*, 11(1):18371, 2021.

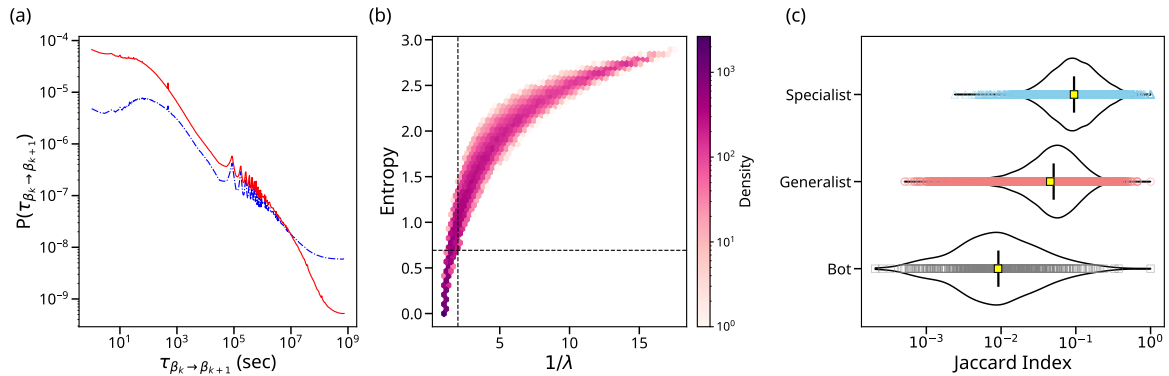


Figure 1: **Analysis of Wikipedia editing patterns and hyperlink influence.** (a) The long-term transition IET distribution ($\tau_{\beta_k \rightarrow \beta_{k+1}}$): Red solid and blue dash-dot lines denote hyperlinked and non-hyperlinked pages, respectively. (b) Editorial activity vs. diversity: Density plot of entropy H and inverse Simpson index $1/\lambda$. Colors show participating communities; dashed lines indicate thresholds (0.693, 2). (c) Jaccard index J_α by category: Violin plots for specialists, generalists, and bots. Yellow markers and vertical lines represent the mean and median.