9^{th} International Conference on Computational Social Science IC^2S^2
July 17-20, 2023, Copenhagen, Denmark

Exploring Urban-Rural Disparities in the Production of Politicians' Information

Keywords: Political Science, Digital Politics, Socioeconomic Disparities, Urban-Rural Divide, Wikipedia

Extended Abstract

Providing political information to citizens is a vital means of implementing democracy in this era of digital politics. The main driving force in this trend is the Internet, which aggregates small amounts of information shared by individuals into a sufficient quantity [10]. Although this powerful tool has reduced the cost of information acquisition and narrowed the information gap among people, researchers have consistently highlighted the "digital divide" that refers to the disparities of information consumption between people [9]. Especially, the literature has tackled the digital divide in political science because politicians and citizens utilize the internet for political purposes, particularly elections [1]. The researcher also points out that socioeconomic factors, such as social or economic disparities, contribute to digital divides [6, 3]. Among myriad of forces, urban-rural disparities play a pivotal role in the digital divide that can explain other social factors that exacerbate the digital divide [5, 4].

However, the literature only separately demonstrates the relationship between the digital divide in political information, socioeconomic factors, and urban-rural disparities. This study analyzes the information about local politicians on the Wikipedia platform and combine the revision history of local politicians' Wikipedia pages with comprehensive census information of the election districts, resulting in an approximately 20-year longitudinal dataset covering the entire nation. Using the history of the politicians' Wikipedia pages around the time of their elections, we investigate the relationship between the socio-economic status of the local politicians' election district and the availability of information about the politician. The previous studies have suggested that certain socio-economic factors may impact Wikipedia platforms [2, 8], but the association between those factors and the supply of information, or specifically of political information, remains uncertain.

Our study utilizes a constructed dataset to investigate the relationship between the number of revisions made on politicians' Wikipedia pages prior to the election period and the demographic characteristics of the electoral district, including the age distribution and industrial composition. This relationship is robust and remains significant even after controlling for confounding factors such as the nature of political parties and individual politicians (Fig. 1).

To understand the behavior of the users who contributed to such local politician's pages during the election, we characterize their editing behavior. This user study presented in Fig. 2a reveals that such users have high engagement (blue dot) than the mean of all users. In addition, we find that the users who edit local politicians' pages have the same engagement both on the local politicians' page and the other pages except for the entropy of page editing, indicating that they are interested in specific politicians in their Wikipedia page editing rather than local politicians in general (Fig.2b).

This skewness of interests is much more evident among the unregistered users who count for approximately 75% of the local politicians' page editors than the registered users (those who have user names in the Wikipedia platform). Figure3 demonstrates that the entropy of editing

by the non-registered users is smaller than the registered users. To capture the differences from a different angle, we construct the editing network where users (nodes) who edit the same page are connected (Fig. 4). Then we study that network by the dismantling procedure introduced by [7]. Figure 5 reveals that the diameter of the network does not change upon the removal of non-registered users but does change upon the removal of registered users. This anatomy demonstrates that the non-registered users focus on specific local politician pages more than registered users.

This study advances our understanding of the information supply of local politicians on the Japanese Wikipedia. We demonstrate that the revisions to local politicians' pages are associated with the socio-economic factors of those local politicians' electoral districts and find that these revisions consist of sporadic information supply related to their preferences rather than their level of activity. The findings of this study have called for addressing such geographical disparities because those inequities in the dissemination of information can obstruct the functioning of democracy.

References

[1] Heath Brown. The institutional digital divide: Immigrant-serving nonprofit organization adoption of social media. *Social Science Computer Review*, 33(6):680–695, 2015.

[2] Erik Dahm, Moritz Schubotz, Norman Meuschke, and Bela Gipp. A Vision for Performing Social and Economic Data Analysis using Wikipedia's Edit History. In *WWW*, 2017.

[3] Eszter Hargittai and Yuli Patrick Hsieh. Succinct survey measures of web-use skills. *Social Science Computer Review*, 30(1):95–107, 2012.

[4] Douglas Blanks Hindman. The rural-urban digital divide. *Journalism & Mass Communication Quarterly*, 77(3):549–560, 2000.

[5] Lorna Philip, Caitlin Cottrill, John Farrington, Fiona Williams, and Fiona Ashmore. The digital divide: Patterns, policy and scenarios for connecting the 'final few' in rural communities across Great Britain. *Journal of Rural Studies*, 54:386–398, 8 2017.

[6] Elissa M Redmiles, Sean Kross, and Michelle L Mazurek. Where is the digital divide? a survey of security, privacy, and socioeconomics. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 931–936, 2017.

[7] Chengcheng Shao, Giovanni Luca Ciampaglia, Onur Varol, Kai-Cheng Yang, Alessandro Flammini, and Filippo Menczer. The spread of low-credibility content by social bots. *Nature communications*, 9(1):1–9, 2018.

[8] Shaun P. Slattery. "edit this page". In *SIGDOC*, 2009.

[9] Alexander Van Deursen and Jan Van Dijk. Internet skills and the digital divide. *New media & society*, 13(6):893–911, 2011.

[10] Gang Wang, Konark Gill, Manish Mohanlal, Haitao Zheng, and Ben Y. Zhao. Wisdom in the social crowd. In *WWW*, 2013.

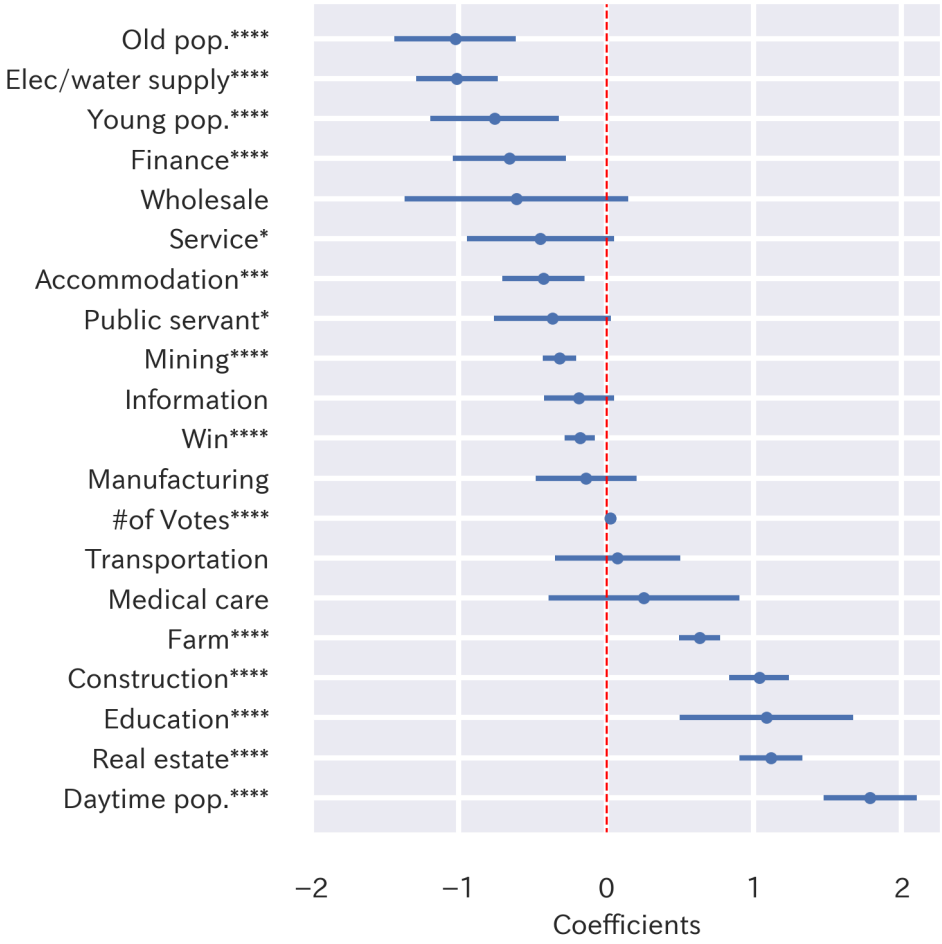
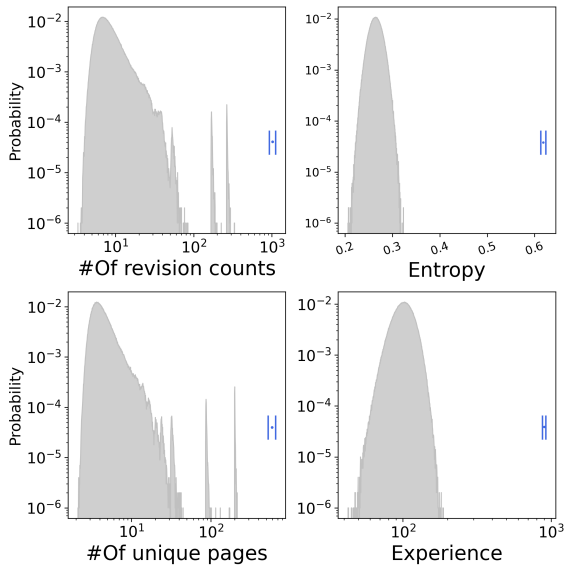
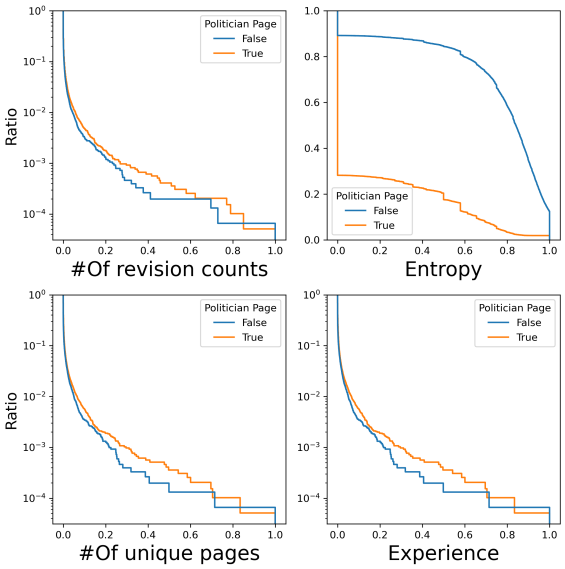


Figure 1: Politicians' wikipedia page and socioeconomic factors of their election districts
Note: Figure 1 shows the estimated coefficients of the model that predicts the number of revisions in the local politicians' page 30 days before their elections (n=630, #obs=7956). Error bars represent 95% confidence intervals. We use standard errors clustered at the individual politician level. Stars represent p-values: *p-val< 0.1; ** p-val< 0.05; *** p-val< 0.01; **** p-val< 0.001.



(a) User engagements compared to the null distributions



(b) Within-user comparison of user engagements: local politician pages VS normal pages

Figure 2: User engagement analysis

Note: Comparison of the four user activities indicators. Figure 2a: the mean values (blue) and Null distribution (gray) simulated by the bootstrapping (gray distribution). Figure 2b: the within-users differences between the politicians' page and the other rest of the pages. We min-max normalize the indicators except for Entropy (already normalized).

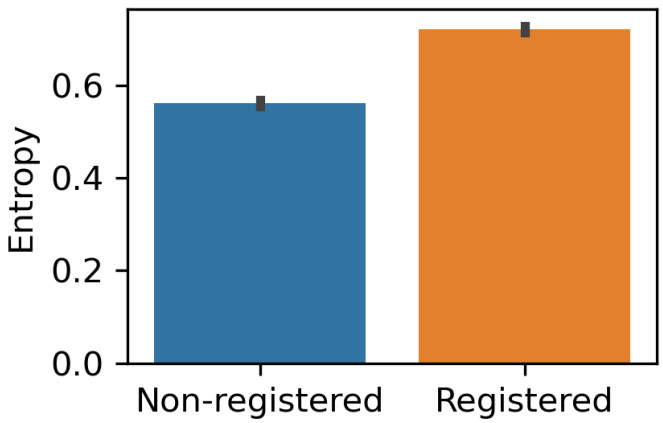


Figure 3: Registered VS Non-registered

Note: We compare the entropy of editing pages of registered and non-register users. All p-value (mean difference) is < 0.001.

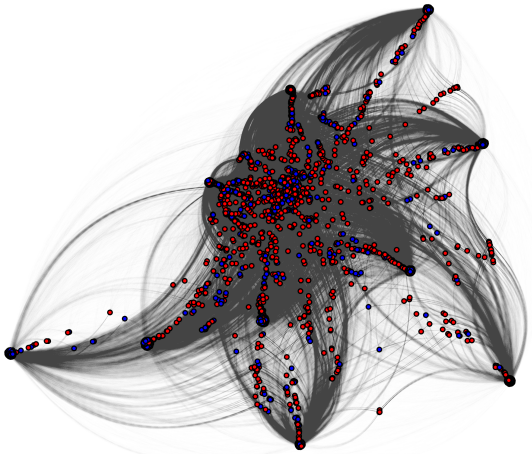


Figure 4: User-interaction network

Note: Visualization of the user interactions network where the nodes represent users and edges connected users who revise the same political Wikipedia article. the red nodes are registered users and the blue nodes are non-registered users.

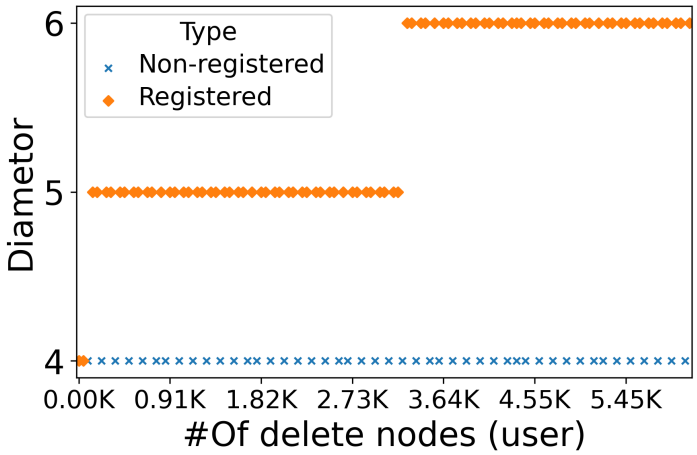


Figure 5: Dismantling user interactions

Note: The results of the anatomy of the user-interaction network by the network dismantling procedure. We conduct the process for the two types of users (registered/non-registered).