Proceedings of the Eighteenth International AAAI Conference on Web and Social Media (ICWSM 2024)

Throw Your Hat in the Ring (of Wikipedia): Exploring Urban-Rural Disparities in Local Politicians' Information Supply

Akira Matsui[1*], **Kunihiro Miyazaki**[2], **Taichi Murayama**[1]

[1] Yokohama National University
[2] Indiana University Bloomington
matsui-akira-zr@ynu.ac.jp, kunihirom@acm.com, murayama-taichi-bs@ynu.ac.jp

Abstract

In this era of digital politics, understanding the factors that influence the supply of political information is important. This study investigates the relationship between socio-economic status and the political information supplied on Wikipedia. To this end, it employs a dataset of politicians who ran for local elections in Japan over approximately 20 years and discovers that the creation and revisions of local politicians' pages are associated with socio-economic factors such as the employment ratio by industry and age distribution. We find that the majority of the suppliers of politicians' information are unregistered and primarily interested in politicians' pages compared to registered users. Additional analysis reveals that users who supply information about politicians before and after an election are more active on Wikipedia than the average user. The findings presented imply that the information supply on Wikipedia, which relies on voluntary contributions, may reflect regional socio-economic disparities.

1 Introduction

Supply of political information to citizens is a vital means of implementing democracy in today's information society. The major driving force in this trend is the internet, aggregating the small amounts of information shared by individuals into a sufficient quantity (Wang et al. 2013; Garcia Martinez and Walton 2014; Nofer and Hinz 2014). While 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 availability of information among individuals (Van Deursen and Van Dijk 2011; Van Dijk 2006).

Recently, the literature has highlighted the digital divide in political information, as both politicians and citizens harness the Internet for political purposes, especially elections (Elliott and Earl 2018; Brown 2015; Ragnedda and Ruiu 2017). Given these findings, understanding the source of this digital divide in political information has become an important issue. In addition, the researcher also point out that the socio-economic factors, such as social or economic disparities, contribute to digital divides (Stanley 2003; Hargittai and Hsieh 2012; Van Dijk and Hacker 2003; Redmiles, Kross, and Mazurek 2017; Hargittai 2001, 2003; Hargittai and Hsieh 2012; Howe et al. 2012). Among myriad sources of the digital divide, urban-rural disparities play a pivotal role because such geographical heterogeneity can explain other social factors that exacerbate the digital divide (Philip et al. 2017; Fong 2009; Furuholt and Kristiansen 2007; Liu, Zhang, and Tian 2019; Hindman 2000; Choi et al. 2022).

While the importance of the digital divide and the political information supply, the literature only separately demonstrates the relationship between the digital divide and political information, socioeconomic factors and urban-rural disparities. Hence, the connection between those three remains unclear. This research gap is due to the difficulties in aligning the supply of political information with regional factors. The primary reason for these difficulties is that determining who does not receive the political information they should is challenging. When studying developed countries, almost all people can access the internet. Therefore, just finding the disparity of socio-economic factors does not guarantee some people are not able to access the information. We need to study the degree to which information is updated or provided, according to geological differences. We need to focus on the digital divide concerning the information people seek.

To overcome the above problems, we first identify regions experiencing a dearth of political information, despite their citizens' are supposed to be supplied for such information, and then study if this deficiency is correlated with socio-economic factors of those regions. This two-step analysis allows us to study the supply of the local-politician demanded information in those regions, and then investigate the association between the information supply differences among regions and socio-economic disparity. Additionally, in order to draw a general conclusion, we include a diverse range of geographical areas, allowing us to cover various socio-economic backgrounds according to geographical differences. Lastly, we need to consider that the digital divide can stem from intentional choices regarding information consumption or from external factors. On most social media platforms, recommendation algorithms can impact information consumption, particularly with regard to political information, as has been observed across various platforms (Lutz et al. 2021; Huszár et al. 2022; Cinelli et al. 2021). However, the use of Wikipedia data presents an advantage, as analysis can be performed without considering

*Corresponding author

the influence of recommendation algorithms.

In order to achieve this goal, we analyze the information about local politicians on the Wikipedia platform. We combine the revision history of local politicians' Wikipedia pages with comprehensive census information of the regions in which those politicians are situated, resulting in an approximately 20-year longitudinal dataset covering the entire nation. We investigate the relationship between the socio-economic status of the local politicians' election district and the availability of information about the politician. We examine the history of the politicians' Wikipedia pages around the time of their elections. The previous studies have suggested that certain socio-economic factors may impact Wikipedia platforms (Dahm et al. 2017; Sheehan et al. 2019; Slattery 2009), but the association between those factors and the supply of information, or specifically of political information, remains uncertain due to the aforementioned challenges. By analyzing politicians who ran for local elections, we explore the information that the people in those local regions demanded.

This study first constructs a dataset of politicians who ran for local elections in Japan that meets these important criteria and address the following three research questions:

- RQ1: What is the relationship between politicians' information supply and the socio-economic characteristics of their election districts?
- RQ2: Do the users show different behavior regarding to their type?
- RQ3: What are the characteristics of the users who supply the information of local politicians?

By answering the above RQs, we demonstrate the creation and revision history of politicians' Wikipedia pages during the election periods are related to the population by the electoral district's industry and the district's age distribution. This relationship is robust and remains significant even after controlling for confounding factors such as the nature of political parties and individual politicians. Analysis of the users who supplied information before and after the election also shows that they are more active than the average user. While more than half of these users were non-registered, the results indicate that they are mainly interested in specific politicians. Our findings indicate that platforms that rely on voluntary individuals, such as Wikipedia, can be vulnerable to socio-economic disparities in a specific region.

2 Data

We first describe our procedure to collect data from two different sources to construct the dataset to study the urban-rural disparities in the supply of local political information. In this paper, we focus on the Japanese politicians who run for the local election from year 2003 to 2022. Our dataset covers the elections for prefecture and municipal mayors and for members of prefecture and municipal assemblies. We first collect the revision history of the Wikipedia pages of that local politician, matching the election result data. We also collect the census data of the electrical district of the local politicians ran for to obtain the socio-economic characteristics of those districts.

2.1 Wikipedia Data

Revision history To construct a dataset documenting the revision history of Wikipedia articles, we first collected the revision history of each article. We obtained the revision histories of articles from the Japanese Wikipedia through the API (MediaWikiAPI 2022) using Python Wrapper (Barrust 2024). Each revision history includes the timestamp of the revision, the username. We obtained all available histories of the Japanese Wikipedia via the API during the period of the election dataset discussed above. The details and examples of the revision history can be found in (Wikimedia Foundation 2023)

Local politicians' Wikipedia page In addition to obtaining the entire revision history of Wikipedia, we identified the Wikipedia pages of local politicians. For each politician who has ever run for local office in Japan at least once, we searched for their pages using the API. Since we focus on the Japanese Wikipedia and Japanese local politicians, we use Japanese query keywords that are the politician's full name and "politician". This resulted in the acquisition of 1,073 local politician pages on Wikipedia with titles matching those of the politicians.

2.2 Socio-Economic Factors: Census Data

We subsequently obtained the census data for the electoral districts (cities, towns, villages, and prefectures) of the local politicians. This census data is collected by the Japanese government every five years, specifically through the efforts of the Statistics Bureau, Ministry of Internal Affairs and Communications (SBMIAC) of Japan (e-Stat 2022). This census offers detailed information on the demographic, social, and economic characteristics of the Japanese population, made available in a publicly accessible database by the SBMIAC. The census occurs every five years, and this study uses data from 2020, 2015, 2010, 2005, and 2000. No similar statistics are available at any other more detailed frequency. We utilized the population data by various categories and the number of employment by industrial categories and matched the census data closest to each election date. These variables will be further elaborated upon in the results section.

3 Methods

In this section, we discuss the methods for examining the geographic disparity of local politicians' Wikipedia pages from multiple perspectives. We first discuss the fixed-effect regression model that studies the Wikipedia creations and revisions of each politician page while considering potential confounds that may affect the results. Then, we describe our network model that represents the interactions of Wikipedia editors to understand the role of users. With a constructed empirical network, we untangle the network to show that users with distinct characteristics exhibit different types of activity in their revision behavior. Lastly, we introduce the four indicators that capture the Wikipedia editors' characteristics to understand the characteristics of the local politician page editors compared to other normal editors.

3.1 RQ1: Revision Activity Fixed Effect Model

To comprehend the relationship between the politicians' information supply and the socio-economic factors, we examine whether the socio-economic status of the regions in which those politicians' election districts predict the creation or revision of local politicians' Wikipedia pages. To accomplish this goal, we employ the fixed-effect linear regression model, which enables us to incorporate potential confounding factors that might affect the dependent variables as "fixed effects" (Hanck et al. 2019; Hansen 2022). These fixed effects capture characteristics of entities, such as the popularity or political parties of candidates. In other words, with the fixed-effect regression model, we will study whether the information supply (revision or creation) can be predicted by socio-economic factors, even after accounting for differences between the entities.

Information Supply I: Wikipedia Page Creation We first investigate whether the politicians' Wikipedia page was created before they ran for an election for the first time. We model this relationship with the following equation.

$$\mathrm{create_i} = \mathrm{win}_i + \mathrm{vote}_i + \mu_j + \mu_y + \mu_{ip} + \sum_s \gamma_s x_s + u_\mathrm{i} \quad (1)$$

where $\mathrm{create_i}$ is an indicator variable that takes 1 if politician i's Wikipedia page is created before their election, and 0 if others; win_i also indicates if politician i won the first election or not; vote_i is the number of votes; μ_j, μ_y and, μ_{ip} are respectively fixed effects of the regions, election year and a political party that politician i affiliated to; u_i is the error term.

In this study, we are interested in the parameters γ_s that capture the socio-economic factors of the districts of the election in which politician i ran for. Therefore, x_s is the socio-economic conditions we obtain from the census that we will describe in the result section. We use the logarithm transformation form of x_s and $\mathrm{vote_i}$.

Information Supply II: #of Wikipedia Page Revisions Next, we perform a similar analysis with the number of revisions that the local politicians' Wikipedia page received 30 days before their election, modeling this relationship with the equation,

$$\mathrm{revisions_{it}} = \mathrm{win}_{it} + \mathrm{vote}_{it} + \mu_i + \mu_t + \sum_s \gamma_s x_{st} + u_\mathrm{it} \quad (2)$$

where the most of the same except for the fixed effects. In Equation 2, we use the politician i's fixed effect μ_i and the election month-year μ_t fixed effect. In most cases, politician i's political party and election region do not change within data; we can consider the fixed effect of those two are absorbed into μ_i. u_it is the error term.

Interpretation of the Models A significant benefit of the regression models outlined in Equations 1 and 2 is their ability to control for confounding factors, such as those arising from differences between candidates or regions, as fixed effects. Although this constitutes a robust model, interpreting the coefficients differs slightly from that of ordinary simple/multiple regression models. The coefficients estimated by this model represent the average effects of the variables of interest on the information supply (I or II) after accounting for the specified fixed effects. More specifically, both models calculate the marginal effects.

In Model I (Equation1), which investigates a binary variable, the model's coefficients signify the mean marginal effects, illustrating the contribution of a 1% change in the coefficient value to the probability of creation. This class of model is called Linear Probability Model (LBP) (Deke 2014; Hanck et al. 2019). For Model II, the coefficients to be estimated represent a 1% change in the value of variables before log-transforming to the number of revisions. While we could calculate the odds ratio as an alternative for Model I, we prefer to use the fixed effect model that allows us to control the confounding factors associated with candidates, their regions, and the time period of the election.

3.2 RQ2: User Interaction Network

Turning our focus from the information supply to the sources of the information (i.e., revision activities), we subsequently analyze the interactions between users. Specifically, we seek to explore the collaborative efforts of these users in contributing to local politicians' Wikipedia pages by examining their interactions through these pages. To facilitate this investigation, we construct a network in which users are connected through Wikipedia pages.

A User Interaction Network We construct a network in which users are connected based on their shared Wikipedia revisions, denoted as $G = (V, E)$, in which each user is represented as a node $v \in V$ and an edge $e_{uv} \in E$ connects two users $u, v \in V$ if they have both revised the same page. The more different pages a user revises, the more users that user can connect with in the network. In addition, a user who prefers to revise popular pages in terms of revision (i.e., many users revise that page) will connect with many users. While this network primarily represents users as nodes and does not explicitly model Wikipedia pages, these pages serve as implicit connectors that connect users with shared revision preferences. Therefore, because of this duality, even we do not have representations of pages in the network, analyzing the constructed user interaction network does not mean we only study users but also pages in the Wikipedia user behavior.

Dismantling Procedure We investigate the constructed user interaction network to understand the activities and differences of users; we utilize the network dismantling procedure (Barabasi 2005). This procedure quantifies the contribution of each node (i.e., user) to the connectivity of the network. The dismantling procedure removes the nodes from the network and calculates the diameter of that network. When the removal of a node increases the diameter of the network, it indicates that the node serves as an interconnector between other nodes. We apply this technique to our empirical network to examine whether we observe different results when removing different types of users. This procedure can reveal the role of users in the network, such as the role of bots in the dissemination of fake news (Shao et al. 2018).

3.3 RQ3: User-Level Activity Indicators

Upon gaining insight into individual user behavior, we examine the individual behavior of those who contribute to the revisions of Wikipedia pages. To achieve this end, we construct activity indicators for each user from the detailed Wikipedia revision history obtained in the data section (Sec. 2).

#of (Unique) Pages This subsection introduces indicators that capture the revision behavior of Wikipedia users. These indicators attempt to capture the users' engagement in their revision behavior. We first simply count the number of revisions they conducted and the unique number of those pages, excluding revisions on the same page.

Experience We also evaluate the "experience" of users to quantify their proficiency on Wikipedia, using the duration from the first revision to the last one in days as a surrogate measure. These indicators serve as proxy variables for the proficiency of the users as editors of revision activities on Wikipedia. We use this simple yet eloquent indicator to capture, for each user, their overall experience on Wikipedia.

Entropy of Revision While the number of revisions and the experience indicator considers a single aspect of users' engagement in revision behavior, we aim to capture the overall proficiency of each user by calculating the entropy of their revisions. The entropy measurement we will introduce quantifies the variety of pages that each user revises using the following equation.

$$\widetilde{H_i} = -\sum_{j=1}^{m} p_{ij} \log p_{ij} \tag{3}$$

where p_{ij} is the frequency in which user i edits article j.

The entropy represented by Equation 3 quantifies the degree of disparity in preference of revision. When one prefers to revise diverse pages, H_i takes a large value. The extreme case is user j edits all articles ($H_i = \log n_i$) with uniform effort, where n_i is the number of edits user i did. To consider the heterogeneity among the users, we resale $\widetilde{H_i}$ by $H_i = \widetilde{H_i}/\log n_i$ so that we can ensure that the entropy falls between 0 and 1. As the experience indicator, we can calculate the H_i for all revisions of user i or for the revision of a specific category.

Null Distribution Construction We are also interested in comparing the politicians' page-editing users with the entire population of Wikipedia editing users. Although calculating the aforementioned engagement indicators for all Wikipedia users is straightforward, it is computationally intensive and might not be feasible due to API limitations or errors. Accounting for this uncertainty, we estimate the distribution of the "mean" values of the indicators of the total population, rather than merely calculating point estimates based on available users. To accomplish this, we first create a user pool comprising users who do not contribute to local politician pages accessible via the API. Subsequently, we employ a bootstrapping method that iterates calculations of the mean value for users randomly selected from the user pool. As this is a bootstrapping process, we sample different users in each iteration with replacement.

4 Results

This section presents the results of our analysis. We first study the association of the information supply of local politicians on Wikipedia based on local socio-economic factors. We show that the socio-economic factors of the electoral districts significantly influence the information supply even when accounting for potential confounds. Then, we model the interactions among them using a network model to understand the distinctive characteristics of different user types. We also characterize the users who contribute to local politician pages and find that these users are significantly more active than the average user and exhibit higher engagement on politician pages than other pages.

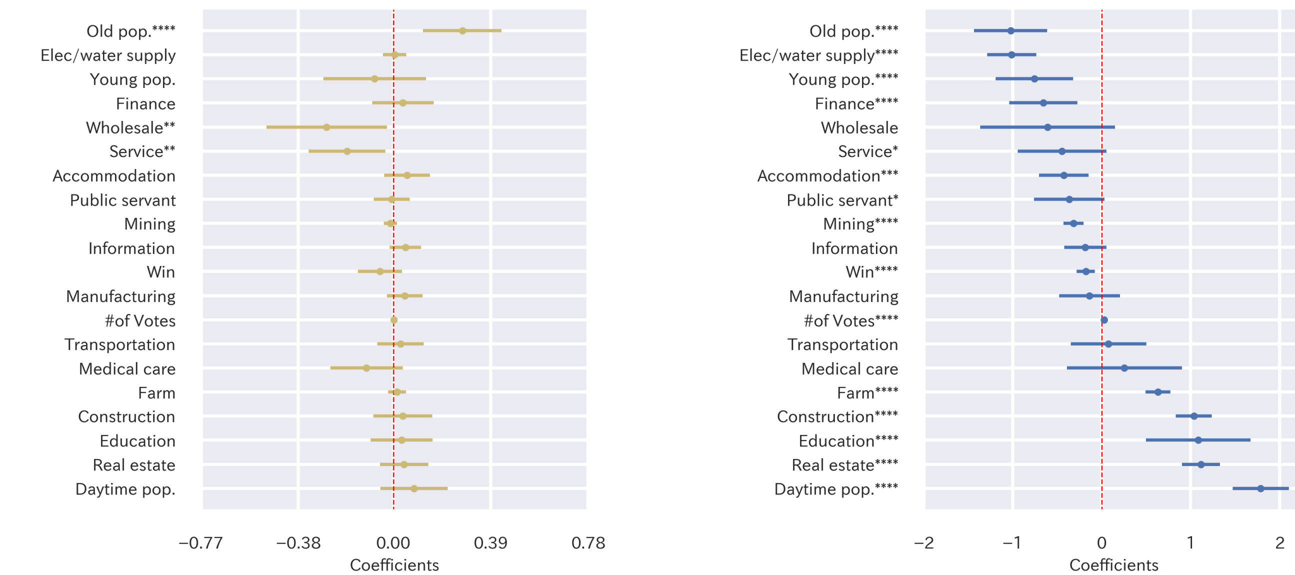
4.1 RQ1: Information Supply and Socio-Economic Factors

To answer RQ1, we predict local politicians' page creation and page revisions using the fixed effect regression model described in Section 3.1. For predictors, we use the variables constructed from the census data that represent the socio-economical condition of the election district of each local politician and the detail of the independent variables illustrated in Table A.1 in the Appendix.

Information Supply I: Page Creation Before Election We first examine if the politician pages were created after the time they ran for local office for the first time, and present the results in Figure 1a, plotting the coefficients estimated by the fixed effect model using Equation 1.

The results demonstrate that while most factors do not have strong associations with page creation prior to the first election, some factors related to urbanization, such as the population working in wholesale or service sectors, have negative coefficients. In addition, we observe the large coefficients in the number of the old population. This suggests that local politicians in an aged district with a living-related industry tend to have their Wikipedia page before an election. Overall, we would argue that this is a modest finding, indicating that most socio-economic factors are not strongly predictive of page creation prior to the election. We infer that socio-economic factors largely do not play a significant role in the provision of information prior to the first election.

Information Supply II: Page Revisions Before Elections The finding in the previous subsection raises an important question: even though socio-economic factors are not pronouncedly related to page creation prior to the first election of politicians' careers, do we observe similar results after politicians have established Wikipedia pages? To address this question, we examine whether the revisions of local politicians' Wikipedia pages before elections are associated with the socio-economic factors of their electoral districts.



(a) Page creation before the first election

(b) Number of revisions 30 days before elections

Figure 1: Politicians' Wikipedia page and socioeconomic factors of their election districts. The coefficients values estimated by the fixed effect models introduced in Section 3.1. The figures present the estimated results that remove fixed effects of entities (confounding). Figure 1a shows the estimated coefficients of the model that predicts the local politicians' page creation before their first election described by Equation 1 (n=1073). Figure 2 shows the estimated coefficients of the model that predicts the number of revisions in the local politicians' page 30 days before their elections described by Equation 2 (n=630, #obs=7956). The detail description of variables can be found in Table A.1. Error bars represent 95% confidence intervals based on the standard errors. For the standard error estimation, we use standard errors clustered at the preference level (Figure 1a) and the individual politician level (Figure 1b). Stars represent p-values: *p-val$<$ 0.1; ** p-val$<$ 0.05; *** p-val$<$ 0.01; **** p-val$<$ 0.001.

The result in Figure 1b shows the regression results predicting the number of page revisions 30 days before elections using the fixed-effect regression model outlined in Equation 2. The figure indicates that socio-economic factors have a different role in predicting the number of revisions. One notable finding is that demographic factors - politicians in districts with many young or old populations - receive fewer revisions (negative coefficient values). The positive coefficients of education and real estate variables, on the other hand, suggest that politicians in highly urbanized districts receive more revisions. The fact that the daytime population variable has the largest coefficient further supports this speculation. Taken together, these results suggest that a politician receives more information on their Wikipedia page (revisions) when their district is, for example, a city center with a high daytime population but not many families (low old/young population).

These results also imply an urban-rural divide in political information supply. One of the distinct demographic characteristics of Japan is a high-aging population and urbanization, which means that there are a few urban areas and 65% of the total population are elderly people (Parliament 2020). Given this fact, the results in RQ1 disentangle the various forces of the digital divide in information supply depicted in Figure1b. First, the urban areas can receive more information supply since Education and Real estate are typical industries in urban regions of Japan. However, political interest groups with significant political power in Japan play a role, even though they are not necessarily industries in urban areas. We also find positive coefficients for the population working in the construction and farming industries. Therefore, while we can see that the urban-rural divide can foster information supply, political interest groups - a force specific to politics - still matter. Also, it is unsurprising to see that factors related to the public sector, such as electric/water supply and public servants, have negative results, as public servants in Japan are prohibited from "*conducting any political activities while in office*" (Act#129 1999).

While we observe some association between information supply and socioeconomic factors, certain variables in the regression models may be correlated with each other, potentially influencing the variance of the coefficient estimations. By examining the estimated 95% confidence intervals (95%CI) presented in Figure 1a, we find that most intervals are relatively small or sufficiently far from 0 (dotted red lines), suggesting that such correlations do not substantially affect the results. However, in Figure 1b, we observe that the coefficients for "Wholesale" and "Medical care" have extended 95%CIs. To determine if this is due to correlations between variables, we estimated the models with obvious variables that correlate with those in the models: Population by age ("Old pop." and "Young pop."). We present the result in Figure A.1 and A.2 and find that the CIs expand and the coefficient values change rather than the CIs narrowing. This implies that correlations related to population do not significantly impact the implication of the results.

Eliminating these variables would address multicollinearity but could introduce omitted variable biases. These biases might lead to under- or overestimation of coefficients, potentially causing significant issues. While multicollinearity primarily increases the variance of coefficients, omitted variable bias could make coefficient estimation unreliable. Following the discussion on multicollinearity by Wooldridge (2015), we have chosen not to remove variables prone to multicollinearity.

4.2 RQ2:Revealing Sporadic Users

We have just shown that the information supply of local politicians is associated with several socio-economic factors. To satisfy this interest, we obtained all revision activities analyzed in Section 4.1, meaning we focused on users who revised a local politician page 30 days before or after an election (n=19,547). Our next area of investigation is the coordination of these users in the supply of information, specifically through their Wikipedia page edits. In this subsection, we examine the potential connections between these users through Wikipedia page edits.

Empirical-User Interaction Network To systematically examine the between-user interactions, we employ the methods in Section 3.2 to construct the network for the users who engaged in the local politician Wikipedia page where each dot represents users and edges connect users if they revise the same page (the same user group as in Section 4.1 hence 19,547 users). In addition, we color dots red if those users have their accounts (registered n =6,890) and otherwise blue (non-registered n =12,657).

We plot the network in Figure 2 that compare two type annotations and each node size corresponds to the degree to the node, focusing the nodes that have more than one edge. The both networks in the figure are identical except for their annotation. The left side of the network (Figure 2a) discerns the register and the non-registered users. The right side (Figure 2b) presents the results of the community detection by the modularity clustering algorithm (Blondel et al. 2008) implemented in (Bastian, Heymann, and Jacomy 2009) that has eight communities. These figures indicate that within each community, registered users possess a higher degree, suggesting that users fulfill different roles based on their user type. This simple observation implies that registered users hold a significant role within the network. Subsequently, we will verify the accuracy of this observation through a more quantitative approach.

The Role of Users The presence of both non-registered and registered users, as well as the fact that approximately 65% of users are sporadic (non-registered), prompts us to examine the way they coordinate in the provision of information on the platform. It is evident that non-registered user nodes do not have a high degree on average, which suggests that they do not contribute to a large number of pages. This finding implies that a certain set of registered users contribute to a variety of politicians' Wikipedia pages, noting that our network only focuses on the revision history of local politicians and not all Wikipedia pages.

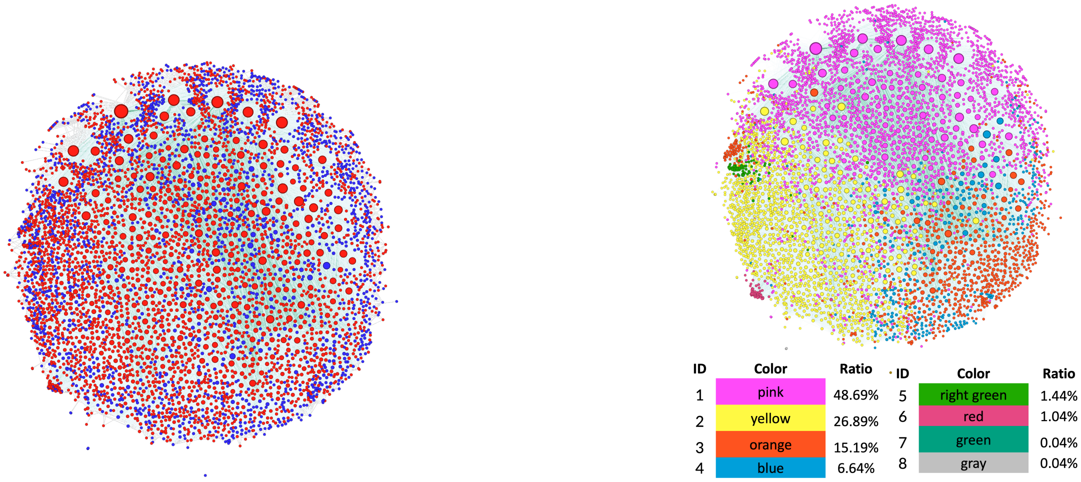
To quantify the degree of contribution according to user type, we utilize the network dismantling procedure discussed in Section 3.2 and plot the results on the network in Figure 5. The figure reveals that the diameter of the network does not change upon the removal of non-registered users but does change upon the removal of registered users. This anatomy demonstrates the coordination between the two user types, with non-registered users focusing on specific local politician pages and registered users organizing their work across local politician Wikipedia pages. Since 75% of the user is non-registered, this analysis indicates that most information supplied to the local politicians consists of sporadic revisions.

4.3 RQ3: Characterising the User Engagement

In the previous research questions, we investigated the impact of socio-economic factors on the revision of local politicians' Wikipedia pages and analyzed the involvement of users in these edits. To achieve this objective, we calculated four activity indicators as described in Section 3.3 to provide a comprehensive understanding of individual user behavior. Our analysis of the user-interaction network suggests that registered and non-registered users exhibit different behavior. To capture these differences associated with editing behavior, we explore engagement disparities between the two user types. Furthermore, we investigate differences between users who edit local politician pages and average users on the Wikipedia platform.

User Engagement and Their User Types The RQ2 with the user interaction network has revealed that the user types are connected with the degree of interaction in the platform. This finding tempts us to zoom the behavior differences according to the user types in more high resolution. To this end, we calculate the engagement index and compare those of the two types. We present them in Figure 3 and 4. In the all results, we find that the registered users demonstrate high engagement in all aspects (Wealch's t-test of mean value comparisons; p-value $<$ 0.0001). While the user engagements depends on their user types, we also interested in the comparison with the average users of the platforms to understand the overall performance of the users who edits local politicians. To investigate this point, we will conduct a comparison with the mean distribution of engagements.

User-Engagement Compared to the Null Users We first calculate the average values of four indicators for the local politician page editing users and plot them in Figure 6a. For comparison, we obtain the null distribution of the average by sampling 1,000 users at random from the entire revision history of the Japanese Wikipedia, resulting in a pool of 1,706,600 random users. We then calculate the average values by 1,000,000 times, bootstrapping 1,000 users from the pool. The resulting null distribution is depicted in gray in Figure 6a. All of the four results in the figure demonstrate that the local politician page editing users have significantly



(a) Register VS Non-register

(b) Network Community

Figure 2: User-interaction network. Visualization of the user interaction network, where nodes represent users and edges connect users who revise the same political Wikipedia article, with node size corresponding to node degree. Figure 2a and Figure 2b are identical except for annotation (node colors). Figure 2a is annotated based on user type: red nodes represent registered users and blue nodes represent non-registered users. Figure 2b is annotated based on network clustering results: each node color signifies the cluster to which the node belongs.

higher engagement than the null distribution in their overall activities on Wikipedia , as we study not only their revisions to local politician pages, but also their activity on other pages.

User-Engagement on Politicians Pages After understanding the postulate of the local politician page editing users' engagement in the platform, we turn our attention to a within-user analysis. To compare the user engagement on local politicians' pages with that on other pages, we calculate the four activity indicators separately for each category and present them in Figure 6b. We discover that users exhibit high engagement on local politician pages compared to other pages, as evidenced by the number of revisions and unique pages they revise. Conversely, they demonstrate lower entropy and reduced experience. All four comparisons in Welch's t-test of mean values yield p-values < 0.001. This analysis uncovers an intriguing user model wherein, despite high engagement in terms of the number of revisions on local politician pages, their interest in revising these pages is focused (low entropy) and occurs over a relatively shorter time span (low experience).

The conjunction of the three outcomes from RQ3 indicates that although user types are relevant, users who edit local politician pages show a high level of engagement in their Wikipedia editing, and their engagement in local politician page revisions is greater than that for other pages they contributed to.

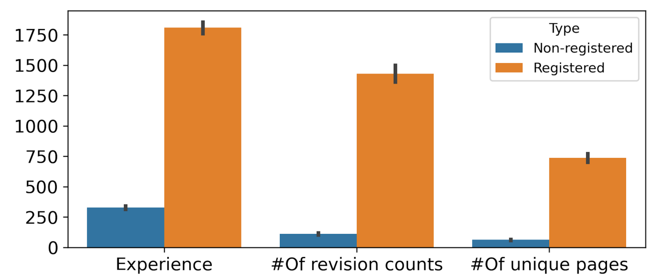


Figure 3: Registered VS Non-registered Users: Performance. We compare the user engagement of registered and non-register users using the measurement defined in Section 3.3. The bars represent confidence intervals by bootstrapping. All p-values (mean difference) are < 0.001.

5 Related Work

This section provides an overview of the related research for this study. We first briefly review the literature on Wikipedia research, and then we narrow down to the topics about the political information in Wikipedia and the urban-rural divide in information consumption.

Wikipedia has gathered attention from researchers as the place for the participants to collaborate and accumulate knowledge. This function of Wikipedia is often referred to as "wisdom of the crowds" (Arazy, Morgan, and Patterson 2006; Kittur et al. 2007). Different approaches have been taken to understand the performance of the wisdom of the crowds and the quality of its output.

For example, several studies demonstrate that the editing history of a Wikipedia article, such as revision history, is associated with the quality of that article (Raman et al. 2020; Shenoy et al. 2021) and discussions among editors also play a pivotal role (Viegas et al. 2007; Bryant, Forte, and Bruckman 2005; Yang et al. 2016). In this topic of research, it is shown that network models such as signed net-

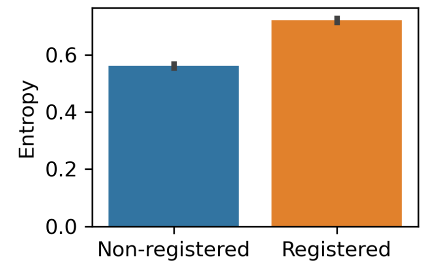


Figure 4: Registered VS Non-registered Users: Entropy. We compare the user engagement of registered and non-register users using the measurement defined in Section 3.3. The bars represent confidence intervals by bootstrapping. P-value (mean difference) is < 0.001.

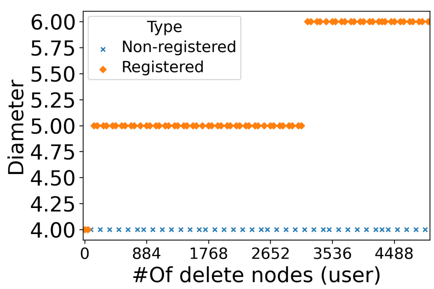


Figure 5: Dismantling user interactions. The results of the anatomy of the user-interaction network by the network dismantling procedure. We conduct the process for the two types of users (registered/non-registered).

works can well capture the interaction between editors in Wikipedia (Adler and De Alfaro 2007; Flöck and Acosta 2014; Javanmardi, Lopes, and Baldi 2010) and can predict the quality of articles (Maniu, Cautis, and Abdessalem 2011). The discussion in Wikipedia editing is often modeled by Opinion Dynamics (Ciampaglia, Flammini, and Menczer 2015).

Understanding the formation of team working in Wikipedia article editing and "crowds" part has also been an essential topic in the literature to improve the quality of team organization and Wikipedia articles (Platt and Romero 2018; Nemoto, Gloor, and Laubacher 2011; Qin, Cunningham, and Salter-Townshend 2015; Lerner and Lomi 2017). Significantly, the characteristic of teams in Wikipedia mitigates unforeseen results. For example, several studies show that the diversity of teams can alleviate polarization (Lerner and Lomi 2019; Adler and De Alfaro 2007; Shi et al. 2019), but working in teams does not mean being free from biases such as in-group biases (Oeberst et al. 2020). In addition, opinion dynamic models can reveal the decision-making process in group (Tasnim Huq and Ciampaglia 2021). The literature is also interested in the phenomenon specific to the Wikipedia platform. "Edit wars" is, for instance, one of the topics to which researchers pay attention that refers to a phenomenon in which users repeatedly edit the content of a particular article, often one on controversial topics (Borra et al. 2015; Kristof, Grossglauser, and Thiran 2021; Jaidka et al. 2021).

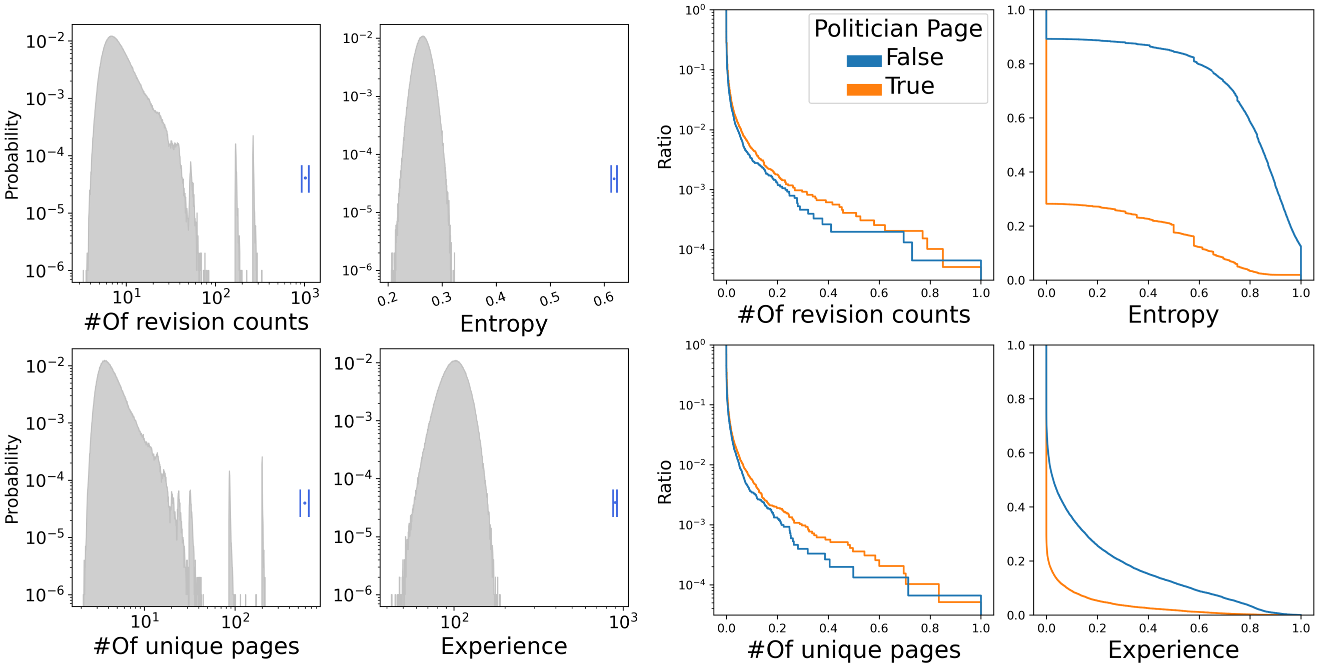
The user-level analysis has received some attention in the literature to dismantle the knowledge production by the participants, for example conducting taxonomy of user roles (Piscopo and Simperl 2018; Müller-Birn et al. 2015; Cuong and Müller-Birn 2016; Kittur et al. 2007). Especially the first seminal work on this topic is Kittur et al. (2007), demonstrating the user role transition in Wikipedia. Studying user behavior in Wikipedia shows the factors that decrease the users' motivation for editing (Halfaker, Kittur, and Riedl 2011; Halfaker et al. 2013)

While Wikipedia is used for the place for fact-building (Swarts 2009; Slattery 2009). The literature is interested in the role of socio-economic factors during those processes. Sheehan et al. (2019) predict the wealth and education level geolocated Wikipedia pages, and Dahm et al. (2017) performs user modeling based on the socio-economic factors of the users extracted from their survey.

In relation to political information, some studies have shown that the quality of politicians' information can be high (Agarwal et al. 2020; Brown 2011), and Göbel and Munzert (2017) argue that personal biographies can be attractive media for politicians based on their empirical analysis of German politicians' Wikipedia pages. Kalla and Aronow (2015) demonstrate that negative facts are more likely to be removed from such politicians' personal biographies compared to positive facts through experiments with humans. On the other hand, Pradel (2020) suggest that even political affiliation-revealed users place more emphasis on their self-identify as Wikipedia, the eco-system of the platforms mitigates the partisanship of the users.Umarova and Mustafaraj (2019), however, detected the polarization and the tension between users in Wikipedia political page editing.

6 Discussion

The aim of this work was to identify the urban-rural disparities in the information supply of the local politicians' information. To this end, we identified instances where residents in certain regions are unable to consume the political information they should in comparison to those in other regions. Specifically, we were interested in the association between the page creation/revisions of the local politicians' Wikipedia page and the socio-economic factors related to the election districts. Our findings revealed that the socio-economic factors are tightly associated with the revisions before the elections, but modest with the page creation before the first election of each politician. This implies that compared to revisions, the socio-economic factors are not of central importance in page creation. This result sounds natural, given that the Wikipedia (of the Japanese version) sets several objective rules for page creation. The editors in Wikipedia often discuss when a new biography page is created to decide if they maintain that new page on the platform. The Japanese version of Wikipedia stipulates that a personal page must be noteworthy. In particular, this noteworthy rule explicitly states that just being a politician is not guaranteed to be that person is noteworthy even if he or she is running



(a) User engagements compared to the null distributions

(b) Within-user comparison of user engagements

Figure 6: User engagement analysis. Comparison of the four user activities indicators. Figure 6a: the mean values (blue) and Null distribution (gray). Figure 6b: the within-users differences between the politicians' page and the other rest of the pages. We min-max normalize the indicators except for Entropy (normalized).

for election, or being a local council member also does not guarantee their noteworthiness (Wikipedia(jp) 2023).

Together with the findings of this paper and the notable rule in Wikipedia, several significant implications arise. First, the notability rule for page creation may generally hold in the Japanese Wikipedia as we observed a weak relationship between page creation and socio-economic factors, implying that the noteworthy rule holds. However, we also discovered that politicians in central areas received more revision (the areas with large daytime populations). Consequently, our findings suggest an unfavorable situation; even for "notable" politicians (i.e., those with biography Wikipedia pages), the frequency of updates may depend on the socio-economic factors of their election districts. Given that literature indicates that politicians' biography pages can serve an advertising function (Göbel and Munzert 2017) and may be biased to be positive (Kalla and Aronow 2015), politicians in central areas may potentially benefit more from the platform compared to those in other areas because of the socio-economical disparities stemming from the regional differences.

We found the potential mechanism behind this disparity in the results of RQ2 and RQ3. The results in the RQ2 demonstrated that 65% of these users are non-registered and only participate in a small group of local politicians' pages. These two findings illuminate a typical user who is willing to provide local politicians' information in Wikipedia: a user who is only familiar with or interested in information related to them and therefore does not revise many pages even if they are active on non-local politicians' pages. This "role model" is consistent with the findings from the revision analysis in RQ1. The users who live in areas with high daytime populations tend to be interested in Wikipedia, and users in certain large interest groups (e.g., agriculture, construction) may update information about local politicians related to their interests. Additionally, we confirmed that user engagement on local politicians' pages is high in several facets, but the entropy and experience calculated on their politician page revisions is quite low. This suggests that the preferences of these users within local politicians' pages are narrow not because of their overall engagement but due to page-specific reasons. Nonetheless, the motivation to contribute to the platform is not driven by local politician page editing, as the commitment duration is shorter than that for other pages.

7 Limitations, Challenges and Future Work

This research endeavors to shed light on the political information supply disparity in relation to socio-economic conditions, but it is necessary to acknowledge that there are several limitations to be considered. One set of limitations pertains to the location and duration of our data. Our examination is limited to local politics in Japan, which may not be generalizable to nationwide elections in Japan or to other countries. Additionally, we only focus on the duration of elections to investigate important information supply (page creations before the first election/revisions around elections). Consequently, our study does not encompass politicians' complete Wikipedia revision history. Moreover, the nature of this observational study utilizing a digital platform precluded us from studying pages that were not created. Furthermore, since the census data is available only every five years and no more frequent data is accessible, our regression analysis might not accurately capture the socioeconomic factors at each election time. While this study focuses on Japan, its implications extend beyond the specific country. Modern nations often share similarities in various aspects. Politicians regularly run for elections in their regional district, and a certain proportion of these politicians are represented on Wikipedia in their languages. To explore this further, our future direction involves expanding our analysis to different countries and Wikipedia figures. Conducting similar analyses in countries with shared characteristics with Japan may yield comparable results.

A second set of limitations is that our computational model does not employ the content of the revisions to perform a systematic and objective analysis. It is of interest to investigate the type of content that is added or removed from politicians' pages in order to understand whether the information provided during the election is advantageous for citizens or politicians, but it is exceptionally challenging to determine which revisions are useful or not. Therefore, it is necessary to develop a fact-checking or classification system for this purpose using either an unsupervised or human-in-the-loop style.

Additionally, this paper does not examine the dynamic behavior or trends of political information supply, while the recent literature highlights the significance of comprehending political information consumption in digital platforms such as political prioritization (Hohmann, Devriendt, and Coscia 2023; Waller and Anderson 2021). Studying the data from digital platforms can suffer from algorithmic biases (Lutz et al. 2021; Huszár et al. 2022; Cinelli et al. 2021), but the Wikipedia platform, which does not implement any particular recommendation algorithm, may overcome these challenges. Therefore, a promising future direction is to investigate the political prioritization of information supply.

8 Conclusions

Our 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, and reducing them can ensure that all citizens ought to acquire an adequate amount of information.

Ethical Considerations

The data in this paper is derived from publicly-accessible user-generated content. We pay the utmost attention to the privacy of individuals in this study. We did not discuss the results regarding specific politicians, Wikipedia users, political parties, or regions to keep the privacy of the individual studied in this paper.

Acknowledgements

This work was supported by JSPS KAKENHI Grant-in-Aid for Scientific Research Number JP22K20159, 24K16359, JP23K16889. Also, This work was supported by Research Institute of Science and Technology for Society, Japan, Grant Number JPMJRS23L4.

References

Act#129. 1999. National Public Service Ethics Act (Act No. 129 of 1999). https://www.japaneselawtranslation.go.jp/en/laws/view/3803. Accessed: 2023-05-15.

Adler, B. T.; and De Alfaro, L. 2007. A content-driven reputation system for the Wikipedia. In *WWW*.

Agarwal, P.; Redi, M.; Sastry, N.; Wood, E.; and Blick, A. 2020. Wikipedia and Westminster. In *HT*.

Arazy, O.; Morgan, W.; and Patterson, R. 2006. Wisdom of the crowds: Decentralized knowledge construction in Wikipedia. In *Annual Workshop on Information Technologies & Systems*.

Barabasi, A.-L. 2005. The origin of bursts and heavy tails in human dynamics. *Nature*.

Barrust. 2024. MediaWiki: A Python API wrapper for MediaWiki. https://github.com/barrust/mediawiki. Available under The MIT License (MIT).

Bastian, M.; Heymann, S.; and Jacomy, M. 2009. Gephi: an open source software for exploring and manipulating networks. In *ICWSM*.

Blondel, V. D.; Guillaume, J.-L.; Lambiotte, R.; and Lefebvre, E. 2008. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*.

Borra, E.; Weltevrede, E.; Ciuccarelli, P.; Kaltenbrunner, A.; Laniado, D.; Magni, G.; Mauri, M.; Rogers, R.; and Venturini, T. 2015. Societal Controversies in Wikipedia Articles. In *CHI*.

Brown, A. R. 2011. Wikipedia as a Data Source for Political Scientists: Accuracy and Completeness of Coverage. *PS: Political Science & Politics*.

Brown, H. 2015. The institutional digital divide: Immigrant-serving nonprofit organization adoption of social media. *Social Science Computer Review*.

Bryant, S. L.; Forte, A.; and Bruckman, A. 2005. Becoming Wikipedian: transformation of participation in a collaborative online encyclopedia. In *GROUP*.

Choi, E. Y.; Kanthawala, S.; Kim, Y. S.; and Lee, H. Y. 2022. Urban/Rural Digital Divide Exists in Older Adults: Does It Vary by Racial/Ethnic Groups? *Journal of Applied Gerontology*.

Ciampaglia, G. L.; Flammini, A.; and Menczer, F. 2015. The production of information in the attention economy. *Scientific Reports*.

Cinelli, M.; Morales, G. D. F.; Galeazzi, A.; Quattrociocchi, W.; and Starnini, M. 2021. The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*.

Cuong, T. T.; and Müller-Birn, C. 2016. Applicability of Sequence Analysis Methods in Analyzing Peer-Production Systems: A Case Study in Wikidata. In *Social Informatics*.

Dahm, E.; Schubotz, M.; Meuschke, N.; and Gipp, B. 2017. A Vision for Performing Social and Economic Data Analysis using Wikipedia's Edit History. In *WWW Companion*.

Deke, J. 2014. Using the linear probability model to estimate impacts on binary outcomes in randomized controlled trials. Technical report, Mathematica Policy Research.

e-Stat. 2022. e-Stat Portal Site of Official Statistics of Japan. https://www.e-stat.go.jp/en. Accessed: 2023-05-15.

Elliott, T.; and Earl, J. 2018. Online protest participation and the digital divide: Modeling the effect of the digital divide on online petition-signing. *New Media & Society*.

Flöck, F.; and Acosta, M. 2014. WikiWho: Precise and efficient attribution of authorship of revisioned content. In *WWW*.

Fong, M. W. L. 2009. Digital Divide Between Urban and Rural Regions in China. *The Electronic Journal of Information Systems in Developing Countries*.

Furuholt, B.; and Kristiansen, S. 2007. A rural-urban digital divide? Regional aspects of Internet use in Tanzania. *The Electronic Journal of Information Systems in Developing Countries*.

Garcia Martinez, M.; and Walton, B. 2014. The wisdom of crowds: The potential of online communities as a tool for data analysis. *Technovation*.

Göbel, S.; and Munzert, S. 2017. Political Advertising on the Wikipedia Marketplace of Information. *Social Science Computer Review*.

Halfaker, A.; Geiger, R. S.; Morgan, J. T.; and Riedl, J. 2013. The rise and decline of an open collaboration system: How Wikipedia's reaction to popularity is causing its decline. *American Behavioral Scientist*.

Halfaker, A.; Kittur, A.; and Riedl, J. 2011. Don't bite the newbies: how reverts affect the quantity and quality of Wikipedia work. In *Proceedings of the International Symposium on Wikis and Open Collaboration*.

Hanck, C.; Arnold, M.; Gerber, A.; and Schmelzer, M. 2019. Introduction to Econometrics with R. *University of Duisburg-Essen*.

Hansen, B. 2022. *Econometrics*. Princeton University Press.

Hargittai, E. 2001. Second-level digital divide: Mapping differences in people's online skills. *arXiv preprint cs/0109068*.

Hargittai, E. 2003. The digital divide and what to do about it. *New Economy Handbook*.

Hargittai, E.; and Hsieh, Y. P. 2012. Succinct survey measures of web-use skills. *Social Science Computer Review*.

Hindman, D. B. 2000. The rural-urban digital divide. *Journalism & Mass Communication Quarterly*.

Hohmann, M.; Devriendt, K.; and Coscia, M. 2023. Quantifying ideological polarization on a network using generalized Euclidean distance. *Science Advances*.

Howe, A. E.; Ray, I.; Roberts, M.; Urbanska, M.; and Byrne, Z. 2012. The psychology of security for the home computer user. In *IEEE Symposium on Security and Privacy*.

Huszár, F.; Ktena, S. I.; O'Brien, C.; Belli, L.; Schlaikjer, A.; and Hardt, M. 2022. Algorithmic amplification of politics on Twitter. *Proceedings of the National Academy of Sciences*.

Jaidka, K.; Ceolin, A.; Singh, I.; Chhaya, N.; and Ungar, L. 2021. WikiTalkEdit: A Dataset for modeling Editors' behaviors on Wikipedia. In *NAACL*.

Javanmardi, S.; Lopes, C.; and Baldi, P. 2010. Modeling user reputation in wikis. *Statistical Analysis and Data Mining*.

Kalla, J. L.; and Aronow, P. M. 2015. Editorial Bias in Crowd-Sourced Political Information. *PLOS ONE*.

Kittur, A.; Chi, E.; Pendleton, B. A.; Suh, B.; and Mytkowicz, T. 2007. Power of the few vs. wisdom of the crowd: Wikipedia and the rise of the bourgeoisie. In *Proceedings of the Conference on Human Factors in Computing Systems*.

Kristof, V.; Grossglauser, M.; and Thiran, P. 2021. War of Words: The Competitive Dynamics of Legislative Processes. In *WebConf*.

Lerner, J.; and Lomi, A. 2017. The third man: Hierarchy formation in Wikipedia. *Applied Network Science*.

Lerner, J.; and Lomi, A. 2019. Team diversity, polarization, and productivity in online peer production. *Social Network Analysis and Mining*.

Liu, J.; Zhang, L.; and Tian, Z. 2019. Social And Economic Effect Study of China's Urban-rural Digital Divide Based on Project Management Theory. *DEStech Transactions on Social Science, Education and Human Science*.

Lutz, M.; Gadaginmath, S.; Vairavan, N.; and Mui, P. 2021. Examining Political Bias within YouTube Search and Recommendation Algorithms. In *IEEE Symposium Series on Computational Intelligence*.

Maniu, S.; Cautis, B.; and Abdessalem, T. 2011. Building a Signed Network from Interactions in Wikipedia. In *Databases and Social Networks*, 19–24. Association for Computing Machinery.

MediaWikiAPI. 2022. MediaWiki API. https://www.mediawiki.org/wiki/API. Accessed: 2023-05-15.

Müller-Birn, C.; Karran, B.; Lehmann, J.; and Luczak-Rösch, M. 2015. Peer-production system or collaborative ontology engineering effort: What is Wikidata? In *Proceedings of the International Symposium on Open Collaboration*.

Nemoto, K.; Gloor, P.; and Laubacher, R. 2011. Social capital increases efficiency of collaboration among Wikipedia editors. In *HT*.

Nofer, M.; and Hinz, O. 2014. Are crowds on the internet wiser than experts? The case of a stock prediction community. *Journal of Business Economics*.

Oeberst, A.; von der Beck, I.; Matschke, C.; Ihme, T. A.; and Cress, U. 2020. Collectively biased representations of the past: Ingroup Bias in Wikipedia articles about intergroup conflicts. *British Journal of Social Psychology*.

Parliament, E. 2020. Japan's ageing society. https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2020)659419. Accessed: 2023-05-15.

Philip, L.; Cottrill, C.; Farrington, J.; Williams, F.; and Ashmore, F. 2017. The digital divide: Patterns, policy and scenarios for connecting the 'final few' in rural communities across Great Britain. *Journal of Rural Studies*.

Piscopo, A.; and Simperl, E. 2018. Who models the world? Collaborative ontology creation and user roles in Wikidata. In *Proceedings of the ACM on Human-Computer Interaction*.

Platt, E. L.; and Romero, D. M. 2018. Network structure, efficiency, and performance in WikiProjects. In *Proceedings of the International AAAI Conference on Web and Social Media*.

Pradel, F. 2020. Biased Representation of Politicians in Google and Wikipedia Search? The Joint Effect of Party Identity, Gender Identity and Elections. *Political Communication*.

Qin, X.; Cunningham, P.; and Salter-Townshend, M. 2015. The influence of network structures of Wikipedia discussion pages on the efficiency of WikiProjects. *Social Networks*.

Ragnedda, M.; and Ruiu, M. L. 2017. Social capital and the three levels of digital divide. In *Theorizing digital divides*.

Raman, N.; Sauerberg, N.; Fisher, J.; and Narayan, S. 2020. Classifying Wikipedia article quality with revision history networks. In *Proceedings of the International Symposium on Open Collaboration*.

Redmiles, E. M.; Kross, S.; and Mazurek, M. L. 2017. Where is the digital divide? a survey of security, privacy, and socioeconomics. In *CHI*.

Shao, C.; Ciampaglia, G. L.; Varol, O.; Yang, K.-C.; Flammini, A.; and Menczer, F. 2018. The spread of low-credibility content by social bots. *Nature Communications*.

Sheehan, E.; Meng, C.; Tan, M.; Uzkent, B.; Jean, N.; Burke, M.; Lobell, D.; and Ermon, S. 2019. Predicting Economic Development using Geolocated Wikipedia Articles. In *KDD*.

Shenoy, K.; Ilievski, F.; Garijo, D.; Schwabe, D.; and Szekely, P. 2021. A Study of the Quality of Wikidata. *Journal of Web Semantics*.

Shi, F.; Teplitskiy, M.; Duede, E.; and Evans, J. A. 2019. The wisdom of polarized crowds. *Nature Human Behaviour*.

Slattery, S. P. 2009. "edit this page". In *Proceedings of the 27th ACM international conference on Design of communication*.

Stanley, L. D. 2003. Beyond access: Psychosocial barriers to computer literacy special issue: ICTs and community networking. *The Information Society*.

Swarts, J. 2009. The collaborative construction of" fact" on Wikipedia. In *Proceedings of the ACM International Conference on Design of Communication*.

Tasnim Huq, K.; and Ciampaglia, G. L. 2021. Characterizing Opinion Dynamics and Group Decision Making in Wikipedia Content Discussions. In *Proceedings of the World Wide Web Conference*, 632–639. ACM.

Umarova, K.; and Mustafaraj, E. 2019. How partisanship and perceived political bias affect wikipedia entries of news sources. In *Companion of WebConference*.

Van Deursen, A.; and Van Dijk, J. 2011. Internet skills and the digital divide. *New media & society*.

Van Dijk, J.; and Hacker, K. 2003. The digital divide as a complex and dynamic phenomenon. *The Information Society*.

Van Dijk, J. A. 2006. Digital divide research, achievements and shortcomings. *Poetics*.

Viegas, F. B.; Wattenberg, M.; Kriss, J.; and van Ham, F. 2007. Talk Before You Type: Coordination in Wikipedia. In *Proceedings of the Annual Hawaii International Conference on System Sciences*.

Waller, I.; and Anderson, A. 2021. Quantifying social organization and political polarization in online platforms. *Nature*.

Wang, G.; Gill, K.; Mohanlal, M.; Zheng, H.; and Zhao, B. Y. 2013. Wisdom in the social crowd. In *WWW*.

WikimediaFoundation. 2023. Page history. https://en.wikipedia.org/wiki/Help:Page_history. Accessed: 2023-05-15.

Wikipedia(jp). 2023. Wikipedia:Notability (Person) in Japanese. https://ja.wikipedia.org/wiki/Wikipedia:%E7%89%B9%E7%AD%86%E6%80%A7_(%E4%BA%BA%E7%89%A9). Accessed: 2023-05-15.

Wooldridge, J. M. 2015. *Introductory econometrics: A modern approach 5th Edition*. Cengage learning.

Yang, D.; Halfaker, A.; Kraut, R.; and Hovy, E. 2016. Who did what: Editor role identification in Wikipedia. In *ICWSM*.

9 Paper checklist

1. For most authors...
   (a) Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socio-economic divide, or implying disrespect to societies or cultures? Yes
   (b) Do your main claims in the abstract and introduction accurately reflect the paper's contributions and scope? Yes
   (c) Do you clarify how the proposed methodological approach is appropriate for the claims made? Yes
   (d) Do you clarify what are possible artifacts in the data used, given population-specific distributions? Yes
   (e) Did you describe the limitations of your work? Yes
   (f) Did you discuss any potential negative societal impacts of your work? Yes
   (g) Did you discuss any potential misuse of your work? Yes
   (h) Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? Yes

(i) Have you read the ethics review guidelines and ensured that your paper conforms to them? Yes

2. Additionally, if your study involves hypotheses testing...
   (a) Did you clearly state the assumptions underlying all theoretical results? Yes
   (b) Have you provided justifications for all theoretical results? Yes
   (c) Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? Yes
   (d) Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? Yes
   (e) Did you address potential biases or limitations in your theoretical framework? Yes
   (f) Have you related your theoretical results to the existing literature in social science? Yes
   (g) Did you discuss the implications of your theoretical results for policy, practice, or further research in the social science domain? Yes
3. Additionally, if you are including theoretical proofs...
   (a) Did you state the full set of assumptions of all theoretical results? NA
   (b) Did you include complete proofs of all theoretical results? NA
4. Additionally, if you ran machine learning experiments...**without compromising anonymity**...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? NA
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? NA
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? NA
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? NA
   (e) Do you justify how the proposed evaluation is sufficient and appropriate to the claims made? NA
   (f) Do you discuss what is "the cost" of misclassification and fault (in)tolerance? NA
5. Additionally, if you are using existing assets (e.g., code, data, models) or curating/releasing new assets, **without compromising anonymity**...
   (a) If your work uses existing assets, did you cite the creators? Yes
   (b) Did you mention the license of the assets? Yes
   (c) Did you include any new assets in the supplemental material or as a URL? NA
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? Yes
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? Yes
   (f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR? NA
   (g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset ? NA
6. Additionally, if you used crowdsourcing or conducted research with human subjects, **without compromising anonymity**...
   (a) Did you include the full text of instructions given to participants and screenshots? NA
   (b) Did you describe any potential participant risks, with mentions of Institutional Review Board (IRB) approvals? NA
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? NA
   (d) Did you discuss how data is stored, shared, and deidentified? NA

A Appendix

This appendix provides supplementary information to augment the discussion in the main body of the paper.

	Election
Variable	**Description**
#of Votes	The number of vote
Win	Win or not
	Population
Variable	**Description**
Old pop.	Population over 65 years old
Young pop.	Population below 15 years old
Daytime pop.	The number of people working or schooling of individuals in an area.
	Population by Industory
Variable	**Description**
Farm	Primary sector(agriculture, forestry, and fisheries)
Mining	Mining
Construction	Construction
Manufacturing	Manufacturing
Elec/water supply	Elecric, Water supply
Information	Information
Transportation	Transportation
Wholesale	Wholesale
Finance	Finance
Real estate	Real estate
Research	Academic research, professional and technical services
Accommodation	Accommodation and food services
Education	Education
Medical care	Medical care
Service	Service
Public servant	Public servant

Table A.1: Variables and Descriptions

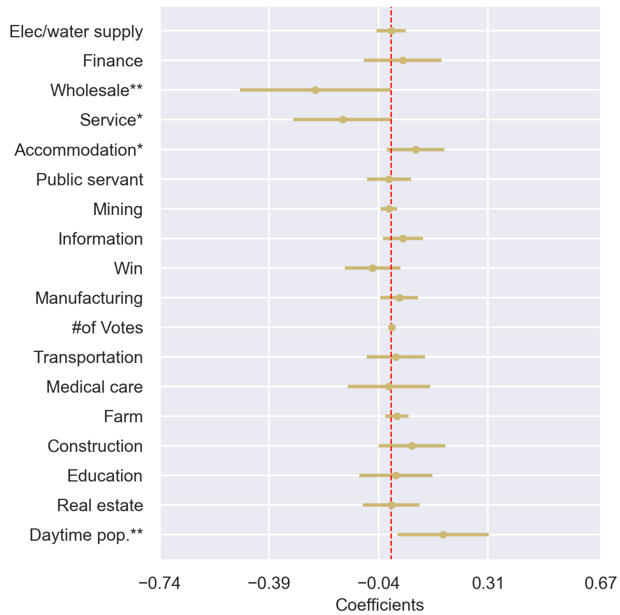


Figure A.1: The coefficient values estimated by the fixed-effect models presented in Section 3.1, excluding the population variables (”Old pop.” and ”Young pop.”). Other details are the same as in Figure 1. Stars represent p-values: *p-val$<$ 0.1; ** p-val$<$ 0.05; *** p-val$<$ 0.01; **** p-val$<$ 0.001.

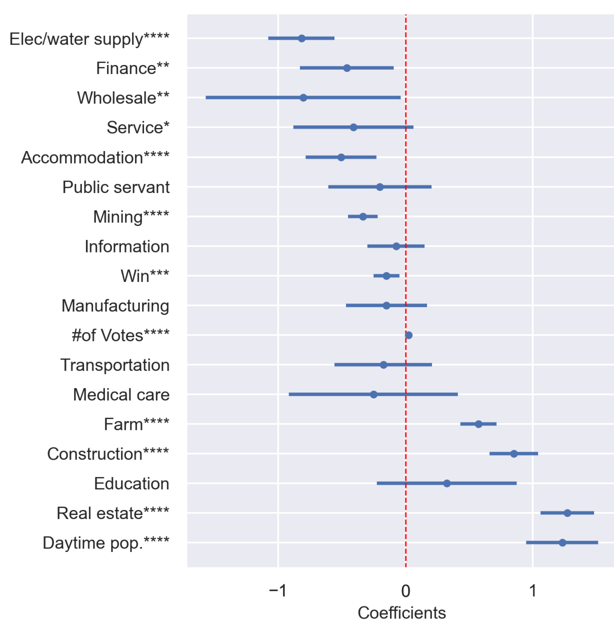


Figure A.2: The coefficient values estimated by the fixed-effect models presented in Section 3.1, excluding the population variables (”Old pop.” and ”Young pop.”). Other details are the same as in Figure 1. Stars represent p-values: *p-val$<$ 0.1; ** p-val$<$ 0.05; *** p-val$<$ 0.01; **** p-val$<$ 0.001.