Political Honeymoon Effect on Social Media: Characterizing Social Media Reaction to the Changes of Prime Minister in Japan

Kunihiro Miyazaki
The University of Tokyo
Tokyo, Japan
kunihirom@acm.com

Taichi Murayama
SANKEN, Osaka University
Osaka, Japan
taichi@sanken.osaka-u.ac.jp

Akira Matsui
Yokohama National University
Kanagawa, Japan
matsui-akira-zr@ynu.ac.jp

Masaru Nishikawa
Tsuda University
Tokyo, Japan
nishikaw@tsuda.ac.jp

Takayuki Uchiba
Sugakubunka
Tokyo, Japan
takayuki.uchiba@sugakubunka.com

Haewoon Kwak, Jisun An
Singapore Management University
Singapore
{haewoon,jisun.an}@acm.org

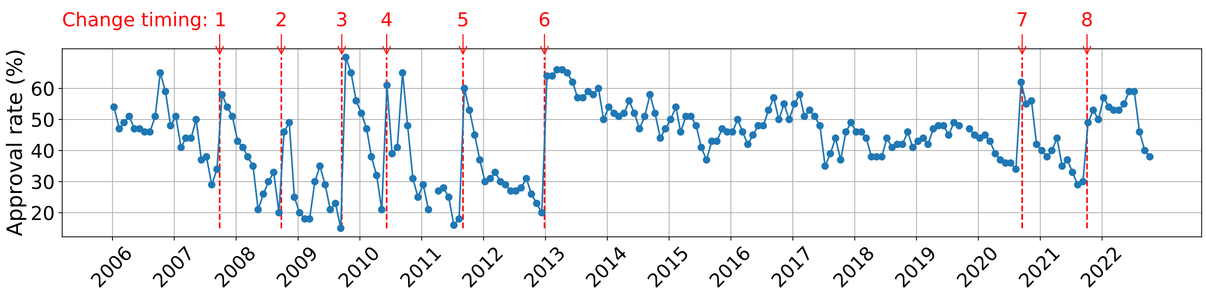

Figure 1: Monthly approval ratings for the Japanese prime ministers since 2006. The timing of the change of prime minister is annotated with a red dotted line. Data from [30].

ABSTRACT

New leaders in democratic countries typically enjoy high approval ratings immediately after taking office. This phenomenon is called the honeymoon effect and is regarded as a significant political phenomenon; however, its mechanism remains underexplored. Therefore, this study examines how social media users respond to changes in political leadership in order to better understand the honeymoon effect in politics. In particular, we constructed a 15-year Twitter dataset on eight change timings of Japanese prime ministers consisting of 6.6M tweets and analyzed them in terms of sentiments, topics, and users. We found that, while not always, social media tend to show a honeymoon effect at the change timings of prime minister. The study also revealed that sentiment about prime ministers differed by topic, indicating that public expectations vary from one prime minister to another. Furthermore, the user base was largely replaced before and after the change in the prime minister, and their sentiment was also significantly different. The implications of this study would be beneficial for administrative management.

Conference acronym 'XX, June 03–05, 2018, Woodstock, NY

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/XXXXXXX.XXXXXXX

CCS CONCEPTS

• **Applied computing** → *Sociology*; • **Social and professional topics**; • **Information systems** → *World Wide Web*;

KEYWORDS

twitter, honeymoon effect, prime minister, sentiment analysis, regression discontinuity design

ACM Reference Format:
Kunihiro Miyazaki, Taichi Murayama, Akira Matsui, Masaru Nishikawa, Takayuki Uchiba, and Haewoon Kwak, Jisun An. 2018. Political Honeymoon Effect on Social Media: Characterizing Social Media Reaction to the Changes of Prime Minister in Japan. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation emai (Conference acronym 'XX).* ACM, New York, NY, USA, 11 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Newly elected national leaders, such as presidents and parliamentary prime ministers (PMs), customarily enjoy high approval ratings immediately after their inaugurations. This phenomenon, known as the honeymoon effect, has been observed in democratic countries around the world [6, 27, 37, 40]. Figure 1 is a visualization of monthly approval ratings of the Japanese cabinet from 2006 with the timings of the PMs' change annotated, which indicates a significant jump in approval ratings at each change timing.

Since the honeymoon effect is an important phenomenon from the perspective of government management, politicians and the news media pay great attention to it. For example, new leaders tend to take advantage of this period of high approval ratings (i.e., the honeymoon period) as the best time to implement the most audacious (but potentially unpopular) policies that can leave their mark on the future economy and society [15, 32, 37]. Additionally, when the approval rating of the current leader declines, the administrative party attempts to preserve the people's support for the party by putting pressure on the current leader to step down, hoping for the next honeymoon effect [32, 47].

Despite its importance, the mechanism of the honeymoon effect is underexplored. There are many narratives for why the honeymoon effect occurs, most of which come from the news media and previous studies. For example, it is driven by the public expectation of a new leader [40], which makes the mainstream media more gentle to him/her [22]. On the other hand, such positivity could appear as the previous leader's great decision of resignation is praised [13], or appreciated [29], or treated as the virtue of integrity [24]. In a nutshell, these narratives see the honeymoon effect as a change in sentiment toward two persons—a new leader and a previous leader. However, to what extent these narratives reflect reality has not been sufficiently verified.

In this work, we aim to deepen our understanding of the honeymoon effect by analyzing social media reactions to changes in national leadership. To this aim, we first collect the tweets around the changes in the PMs of Japan, and construct the dataset that records the reactions to the eight changes of PMs on social media over fifteen years. We primarily use Japanese tweets for the analysis. Here, we would like to stress that Japan can provide a desirable case study for the analysis of the honeymoon effects on social media because 1) Japan has highly frequent changes of PMs compared to other countries since 2006 [39] when Twitter started its service, 2) Japan maintains a high degree of national and linguistic congruence [14], which is easy to detect the originating country of tweets, and moreover 3) Twitter is quite popular in Japan with approx. 60M users (almost half of its population) and roughly the same number of daily active users as U.S. [31], from which we can expect a high correlation between the opinions on Twitter and the real world.

We tackle the following research questions by leveraging the dataset:

RQ1: How does the honeymoon effect appear on social media? We examine whether there is a change in sentiment at an aggregated level on Twitter before and after PMs' change. Social media allows us to extract people's raw voices, and if we can observe the honeymoon effect on social media as well, we can get insights for government management from it.

RQ2: What aspects do people expect from the new prime ministers? We analyze the sentiments by topics to see what aspects people do expect from the new PMs. If the honeymoon effect arises from the expectations of the new PM, understanding which topics he/she is expected to address may help him/her avoid overpaying the cost of approval ratings when implementing bold policies [32].

RQ3: Who are changing the sentiments toward the prime ministers? We analyze the change in sentiment toward PMs at the user level. Knowing who changes the sentiments toward PM would be helpful for the communication strategies of new PMs. Also, to know whether the change of PM has successfully erased the negative image associated with the previous PM may have an impact on the actions that the governing party should take when the PM's approval rating declines again [27].

As a result, we found social media tend to have a significant jump in sentiment in PM change timings (4 out of 8 times), with one timing when sentiment significantly declined and no significant results in the remaining three times. The sentiment was often significantly lower for PM-specific topics than for topics common to all PMs. Also, comparisons among PMs using common topics showed that topics with high sentiment differed by PMs. Finally, we revealed that the majority of users were largely replaced between those who tweeted about the previous PM before the resignation and those who tweeted about the new PM after the inauguration. Furthermore, the sentiments of these different user groups were significantly different in many cases (7 out of 8 times). In addition, the sentiment of users who tweeted about the same PM after the inauguration and before the resignation was significantly different (5 out of 8 times), and in many cases, the sentiment was lower before the resignation (4 out of 5 times), confirming the honeymoon effect in a different way.

Our contribution is as follows:

1. To the best of our knowledge, this is the first study to examine the political honeymoon effect using social media. By looking into the past cases of the honeymoon effect, we can gain insights into how changes in administration can improve or exacerbate the quality of governance. Moreover, the implications from this study are expected to be helpful not only in future political communication but also in other contexts such as corporate management (i.e., CEO resignation).

2. We construct a longitudinal dataset on the change of PM in Japan. We take full advantage of Twitter data to create a 15-year tweet dataset on eight Japanese PMs consisting of 6.6M tweets. We will release this dataset to the public upon acceptance.

3. We demonstrate various data-analysis methodologies in this study. We use a combination of sentiment analysis and regression discontinuity design to capture the honeymoon effect. In addition, we propose an analytical flow to distinguish between unique and common topics for each PM in the topic analysis for multiple PMs, which allows for a cross-topic analysis of historical PMs. We believe these methods will be helpful for future research.

2 BACKGROUND AND RELATED WORKS

2.1 Political honeymoon effect

The honeymoon effect, in political terms, is the high approval ratings enjoyed by a new leader immediately after taking office [3, 5, 5]. Previous research on this political honeymoon effect has been concerned with examining the existence of the honeymoon effect itself, as well as the derived effects. Dominguez [15] found that U.S. presidents are more successful in passing legislation during their first 100 days in office than thereafter. However, it is said that the 45th president of the US, Donald Trump—and maybe the 46th president, Joe Biden as well— could not get the honeymoon period, and it has been debated whether a lack of honeymoon is a new normal or not [18]. Segatti et al. [40] quantitatively analyzed the honeymoon effect in Italy that led to Renzi's victory in the European elections immediately after his administration came to power. Masuyama [27], noting that a PM's approval rating gradually declines after the honeymoon effect, attempted to model the relationship between the probability of a PM's resignation and his approval rating by using survival time analysis. Qadan and Idilbi [37] analyzed the behavior of commodity markets during the U.S. honeymoon period and showed that price volatility was lower during this period. Additionally, studies have shown that the honeymoon effect is associated with the consolidation of democracy: the honeymoon effect is an important research topic in the field of comparative politics as well [7, 26].

In this study, for the first time to our knowledge, we analyze the factors that cause the honeymoon effect to occur from a social media perspective. In particular, we examine the change in sentiment toward a new leader and a previous leader, which has been at the heart of previous narratives on the political honeymoon effect. Also note that the focus of this study is on the honeymoon effect, i.e., the jump in approval ratings right after the change of PMs, and not on the honeymoon period, during which high approval ratings persist.

2.2 Resignations of Japanese Prime Ministers

In a parliamentary cabinet system, which the Japanese government adopts, PMs are often replaced when they lose public support, and their approval ratings become quite low [39]. In particular, Japan's PMs around 2010 was replaced at a higher frequency than in other countries, e.g., Germany and the UK, and most of them were in office for shorter terms such as about one year [39]. This is said to be because, since the late 1990s, changes in the electoral system have made the reputation of the *PM as the face of the party* more important. That is, even from the perspective of politicians belonging to the administrative party, the PM's approval rating has a significant impact on his or her own election results; thus, when the PM's approval rate is low, there is pressure from the party members to resign, and unpopular PMs tend to have shorter terms in office [32, 47]. We regarded this frequent turnover in Japan as valuable for an object of analysis because a change of national leaders is generally a rare event that is usually difficult to analyze quantitatively.

3 DATA COLLECTION

There have been eight PM changes in Japan since 2006, and we collected user reactions for each of those eight changes. There are two key events regarding the change of PM to evoke user reactions: the previous PM's resignation declaration and the new PM's inauguration. In this study, the period for obtaining tweets was set from 60 days prior to the previous PM's resignation declaration to 60 days after the new PM's inauguration. It should be noted that there is a lag between these two events, but in the analysis, we do not include the data from this lag, although we obtained the data in this lag. The tweets were acquired for the period for each change timing by using Twitter Academic API [34].

To retrieve tweets about all PMs involved in those timings, we set the search keywords to be formal as follows (English translation is associated with Japanese words):

- Surname + given name (i.e., full name, e.g., "岸田文雄" (Fumio Kishida)),
- Surname + "首相" (Prime Minister),
- Surname + "総理" (Prime Minister),
- Surname + "政権" (Administration),
- Surname + "内閣" (Cabinet).

In order to increase the precision of the detection of tweets about the PMs and cabinets, we did not include keywords of only surnames, surnames with casual titles (e.g., Kishida-san), and other nicknames.

We did not include retweets in the collection since our purpose was to capture user sentiment. Also, since Japanese is mostly spoken in Japan [14], we did not conduct additional filtering based on the country. As a result, we collected around 6,557,452 tweets written by 855,694 unique users.

As a post-processing of tweet acquisition, we conducted user-based removal of inappropriate tweets. First, since we were interested in the sentiment of general users, we excluded accounts marked as verified [17][1]. Next, to preserve ordinary users' accounts, we removed the following accounts:
News media: accounts that have "ニュース" (news), "news," "新聞" (newspaper), and "テレビ" (TV) in the account names, and
Organizations: accounts that have "公式アカウント" (official account) in the description.

We also used Botometer [11] to remove bots. Botometer provides scores from 0 to 1 to measure how the accounts are likely to be bots. There are two types of scores: one for English and the other for non-English (the language-independent score), among which we used the latter one. Since the prior studies suggested the threshold of Botometer score to classify bots and humans mainly targeted English texts and even the threshold of the score for English accounts varied to some extent [38], we decided to verify the accuracy with a bootstrap method to decide a threshold. Specifically, for four groups with scores of (0.8-0.85], (0.85-0.9], (0.9-0.95], and (0.95-1.00], we randomly sampled 30 accounts for each group, and three Japanese authors manually annotated whether the account was considered to be a bot or not As a result of majority voting, 13.3% (Fleiss' Kappa score: 0.24), 26.7% (0.19), 43.3% (0.49), and 63.6% (0.45) were determined to be bots, respectively. Therefore, in order to remove obvious bots, 0.95 was set as the threshold for this study, and accounts with scores higher than that were removed. We note we could not remove all bots in our data, which is a limitation of this study. Nonetheless, the ratio of (0.95-1.00] users is 0.5% and one of even (0.9-0.95] users is 2.2% and we believe there is no significant impact on the results. In addition, to remove unusual accounts, we removed accounts that had tweeted less than ten times (according to their profile information) at the time of data acquisition (i.e., Nov. 2022). At this point, the number of tweets was 6.1M, and the number of users was 838k.

Finally, in order to focus on the sentiment toward a single PM, tweets containing the surname of more than one PM were excluded. Finally, we got 5,590,028 tweets by users 818,043 users. We use this dataset for the rest of the analyses. Table 1 summarizes the name of the previous and new PMs, the date of declaring resignation, the date of inauguration, and the number of tweets for each change timing. We note that the administrative party changed twice in these eight change timings between the Liberal Democratic Party (LDP) and the Democratic Party of Japan (DPJ). Also, the dataset includes Abe's two appointments.

[1]Note that we finished our data collection before the Twitter acquisition by Elon Musk, which affects the quality of verified mark [16].

Change timing	Prev PM	Resignation declaration	New PM	Inauguration	Party of New PM	Tweet volume Prev	New	Total
1	Shinzo Abe	2007-09-12	Yasuo Fukuda	2007-09-26	LDP	391	183	574
2	Yasuo Fukuda	2008-09-01	Taro Aso	2008-09-24	LDP	1,538	2,201	3,739
3	Taro Aso	2009-08-30	Yukio Hatoyama	2009-09-16	DPJ	7,523	27,457	34,980
4	Yukio Hatoyama	2010-06-02	Naoto Kan	2010-06-08	DPJ	187,156	135,290	322,446
5	Naoto Kan	2011-06-02	Yoshihiko Noda	2011-09-02	DPJ	762,486	206,372	968,858
6	Yoshihiko Noda	2012-12-16	Shinzo Abe	2012-12-26	LDP	203,418	442,322	645,740
7	Shinzo Abe	2020-08-28	Yoshihide Suga	2020-09-16	LDP	1,490,975	662,597	2,153,572
8	Yoshihide Suga	2021-09-03	Fumio Kishida	2021-10-04	LDP	956,954	503,165	1,460,119

Table 1: The change timings of PMs, the names of the prev/new PMs, the dates of resignation declaration and inauguration, the party of new PMs, and the tweet volume of each change timing regarding prev/new PMs.

4 RQ1: HOW DOES THE HONEYMOON EFFECT APPEAR ON SOCIAL MEDIA?

In this section, we examine whether the honeymoon effect also occurs on social media. Specifically, we analyze the change in sentiment toward the PM before and after the change of PM using sentiment analysis and a regression discontinuity design (RDD).

4.1 Sentiment Analysis

Predictive Model: We first predict the sentiment of each tweet. As a predictive model, we use Asari, an open-source sentiment quantification model for Japanese sentences [19]. Asari is an SVM-based model that returns a sentiment value between 0 and 1 (the closer to 1, the more positive) when a Japanese sentence is inputted. We chose Asari because it is reported to perform with compelling accuracy as BERT-based models and is faster [20]. Although dictionary-based methods such as LIWC [33] (or J-LIWC for Japanese texts [23]) are frequently used for sentiment analysis, we did not use these methods because we found that some PMs had only a few tweets with sentiment words in the dictionary. As a bootstrap confirmation, we sampled 15 tweets each from the top 25th percentile and the bottom 25th percentile for the sentiment to see whether a Japanese author could classify them, resulting in an F1 score of 0.8.

Aggregation of sentiment scores: We aggregate sentiment on a daily basis because the number of tweets may vary by time of day. Also, to mitigate the influence of the more vocal users who tweet multiple times a day, we first averaged the sentiment of each user per day, and then averaged the sentiment of all users who tweeted that day. The sentiment S for a day t for a given PM p is depicted as follows:

$$S(t,p) = \sum_{u:user} \frac{\sum_{d \in D_{u,p,t}} s(d)/|D_{u,p,t}|}{|U_t|} \tag{1}$$

where t is the time (day), p is the PM, d is the document (tweet), $s(d)$ is the sentiment score of the tweet d by Asari, $D_{u,p,t}$ is the set of the user u's tweet about p, and U_t is the set of all users who tweeted on that day t.

4.2 Regression Discontinuity Design

To see the change in sentiment before and after the change of PM, we perform a regression discontinuity design (RDD). We employ a linear model:

$$S_t = \alpha_0 + \beta_0 t + \alpha i_t + \beta i_t t + \epsilon_t \tag{2}$$

where S_t is the sentiment toward a given PM on a given day t (eq 1), t is a date, which takes values from -60 to +60, starting 60 days before the previous PM's declaration of resignation and ending 60 days after the new PM's inauguration date, and i_t is an indicator variable, which is 1 after the change of PM (i.e., $t > 0$) and 0 otherwise. Here, the inauguration date ($t = 1$) is hypothetically set to be the next day of the date of the resignation declaration ($t = 0$). We assumed the lag between the actual resignation declaration and the inauguration date to be a "grace period" (following [21]), and the data for that period is omitted because PM's resignations are often sudden and tend to cause instability in sentiment. This model allows us to see not only the change in the quantity of sentiment before and after the change of PM, represented by α, but also the trend in sentiment before and after the change, represented by β_0 and β; thus, we can understand the change in sentiment more deeply than by simply comparing the average of sentiment before and after change timing. Lastly, ϵ_t is an error term at t.

In the introduced RDD model, we input the sentiment of the previous PM before the resignation declaration and the sentiment of the new PM after the inauguration into the model. This is in line with the traditional survey that changes the targets from the previous PM to the new PM before and after the change of PMs.

4.3 Results

Figure 2 shows the results of the RDD. The α of each subplot shows that sentiment increased significantly in four of the eight change timings (change timing: 1,3,7,8), while three did not show a significant change (2,4,5), and one saw a significant drop (6). In other words, we were able to observe that social media tends to show the honeymoon effect, but not at every change timing. While it has been known that the demographics of Twitter are not necessarily the same as in the real world [2], this result demonstrates a possibility of utilizing social media for the analysis of the honeymoon effect using sentiment toward the PM.

On the other hand, it is interesting that in the last two changes (Figure 2-7: Abe→Suga; Figure 2-8: Suga→Kishida), we clearly saw jumps in sentiment. Abe's second administration lasted about eight years, the longest among all the Japanese PMs in history, and the number of Twitter users increased during that time. In fact, the number of tweets jumped dramatically during the last two change timings (Table 1). In other words, the demographics of Twitter might have approached the actual population over the eight years, and the fact that the honeymoon effect was clearly confirmed twice in these circumstances may suggest that the correlation between changes in actual approval rating and sentiment has increased because the number of Twitter users increased.

5 RQ2: WHAT ASPECTS DO PEOPLE EXPECT FROM THE NEW PRIME MINISTERS?

In RQ1, we found that sentiments toward PMs significantly change at multiple change timings of PMs. Here we examine from which aspects the sentiment causes these changes by using topic modeling.

5.1 Topic Modeling

Model selection and data: We use the Biterm Topic Model (BTM) [48] for the topic modeling. This model is a derivative of LDA [9] and is known to be able to extract topics with high accuracy for short sentences. For this analysis, we focus on the tweets during the 60 days after the inauguration of PMs at each change timing. We combine the corpus for all new PMs for the later comparative analysis of topics across PMs. To reduce the imbalance in data volume, we limit the number of tweets for each PM to (randomly sampled) 30,000 for the model training. We then use BTM to label each tweet with the topic with the highest possibility calculated by BTM. After training the model, we predict the topic of tweets that were not used for training and include them for later analysis.

Number of topics and separation of unique and common topics: To determine the number of topics, we set the number of topics large at the beginning and then merge them to find the optimal number of common topics as in [4]. Here, it is possible that some topics are unique to some PMs, and some are common to all PMs. Topics common to all PMs allow for comparison of sentiment across PMs, but this is not the case for unique topics. Therefore, we leverage the topic number selection method to classify unique topics and common topics in the process of merging topics. Specifically, for each topic, we calculated $R(T, p)$, the ratio of the amount of text as for topic T assigned to each PM p. Then, if $R(T, p)$ of a PM p about a fixed topic T outstands among PMs, this topic can be considered a unique topic for PM p. We calculate $R(T, p)$ while addressing the imbalance in tweet volume among PMs. To calculate $R(T, p)$, we first compute the ratio $r(T, p)$: how dominant a topic T is from the tweets for a PM p.

$$r(T,p) = \frac{|\{\text{tweet} \in T \mid \text{tweet for a PM } p\}|}{\sum_{T':\text{topic}} |\{\text{tweet} \in T' \mid \text{tweet for a PM } p\}|}. \quad (3)$$

Next, we calculate R by modifying r so that the sum of r among PMs is 1:

$$R(T,p) = \frac{r(T,p)}{\sum_{p':\text{PM}} r(T,p')}, \quad (4)$$

i.e., $R(T, p)$ is the normalized version of $r(T, p)$. Here, assuming that tweets about each PM are equally assigned in terms of a given topic T, the expected value of this score R should be equally divided by 8 and be 0.125. We divide the topics generated by BTM into unique and common topics using a certain threshold of max R among the R scores for all PMs.

We tested 25, 50, 75, 100, and 125 topics and chose 75 as the initial number of topics by comparing the perplexity scores [49]. Then, we conducted BTM for all the tweets within the period of 60 days after all the inaugurations. In order to set the threshold for R to classify the unique and common topics for PMs, we sorted the topics by the max value of PMs' R scores. After the first author, who is fluent in Japanese, manually examined the top representative words of each topic in terms of the possibility of belonging assigned by BTM, we decided the threshold as 0.285, below which the topics are no longer unique for any of the PMs. In other words, for a topic T, the smaller the max R is, the flatter the distribution of R for all PMs becomes. Finally, the Japanese authors manually checked and replaced the labels of a total of 12 topics with the R scores around the threshold (7 from unique and 5 from common topics). As a result, we obtained 52 unique topics and 23 common topics for all the PMs. As for unique topics, we got 5 topics for Fukuda, 5 for Hatoyama, 4 for Kan, 8 for Noda, 3 for Aso, 10 for Abe, 7 for Suga, and 10 for Kishida. The details on common topics are shown in Table 3 in the appendices, while we omitted the details on unique topics for the reason of space.

Merging Common Topics: Some of the 23 common topics overlap in terms of their contents; thus, we merge them to find an optimal number of common topics. Here, we propose a clustering method to systematically merge similar topics. Specifically, we first connected 10 representative words of each topic into one pseudo-sentence and made 23 sentences in total. Then, we use the trained BTM to calculate the probability of each sentence belonging to each topic and calculate the Hellinger distance between the sentences using the probabilities [44]. Lastly, we make the distance matrix based on the Hellinger distance of the sentences and conduct hierarchical clustering with Ward's method [28] to aggregate similar sentences. We manually determined the threshold for aggregation by hierarchical clustering by looking at the dendrogram and the contents of resulting clusters (figure omitted for the reason of space). We note that we also tried clustering based on the average of the top word embedding by BERT model [12, 45] and Euclidean distance between topics, but the clusters made more sense when using the topic distribution by BTM and the Hellinger distance.

As a result, we obtained four clusters. We label them as *Diplomacy*, *Economy*, *Cabinet personnel (Personnel)*, and *Diet administration (Diet)* by examining the representative words of the topics that comprise the cluster (shown in Table 3 in the appendices).

5.2 Results

First, we compare the sentiments of common and unique topics. Expectations and criticisms of the PM can occur on both common and unique topics. For example, it is anticipated that specific topics may include scandals or topics that arise from time to time (e.g., COVID-19 or the Great East Japan earthquake), while the different PMs can have things in common, such as diplomatic or economic topics. Depending on which sentiment is greater, unique topics or common topics, the government would have different weights on whether it must focus on policy or personal image. Then, we will turn to compare the common topics among the different PMs. A comparison of sentiments on common topics across PMs can tell us what a new PM should learn from the administration of previous PMs.

Table 2 shows the average sentiments of common/unique topics for each PM with the results of the Mann-Whitney U test. We note that we use a user-based daily average of sentiments to mitigate the influence of extremely large users in the same way as eq (1).

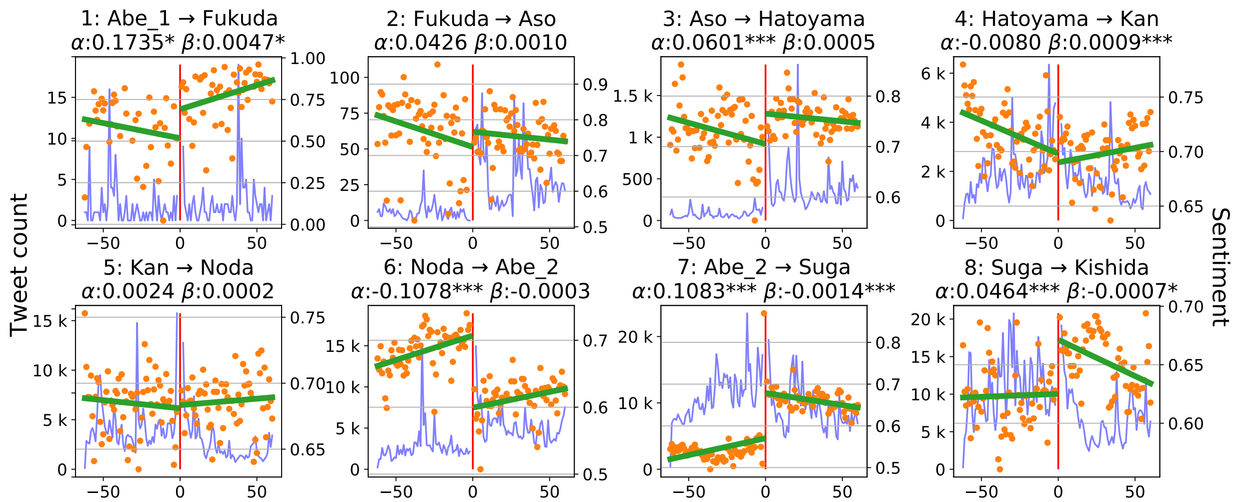

Figure 2: Change in daily sentiment before and after the change of PMs. The left side of the red line shows the period before the PM announced the resignation, and the right side shows the period after the new PM took office. The blue line shows the daily volume of tweets, and the orange dots show the average daily sentiment, with the left side about tweets about the former PM and the right side about tweets about the new PM. The green line depicts the model fitted to the sentiment in the regression discontinuity design (RDD). Above the subplots, we report coefficients associated with the PMs' change in the model (α and β). Coefficients for which p <0.001, 0.01, and 0.05 are marked with *, **, and *, respectively.**

The results show that sentiment is largely lower for unique topics: four PMs have significantly lower sentiment for their unique topics, two PMs have significantly higher sentiment for common topics, and two PMs do not have significant results. Here, we look at the unique topics with the lowest sentiment for four PMs who had lower sentiment for unique topics: Kan for consumption tax hike [8] (sentiment value: 0.643), Noda for nuclear power plant operation [46] (0.626), Suga for pressure on academia (i.e., the Science Council of Japan) [42] (0.640), and Kishida for unconditional cash-only handouts for COVID-19 [36] (0.542). The results actually show that sentiment was low on issues that were unique to the time or the PM's personal policy. Conversely, when the honeymoon effect occurs on social media, it can often be caused by a common topic.

New PM	Common	Unique	Significance
Fukuda	0.848	0.849	-
Aso	0.781	0.885	<<<
Hatoyama	0.782	0.828	<<<
Kan	0.751	0.736	>>>
Noda	0.738	0.734	>
Abe_2	0.636	0.636	-
Suga	0.721	0.698	>>>
Kishida	0.704	0.679	>>>

Table 2: A comparison of sentiment on common and unique topics For each PM. Significance shows the results of the Mann-Whitney U test, where the number of brackets indicates 3: $p < 0.001$, 2: $p < 0.01$, 1: $p < 0.05$. The direction of the brackets indicates the greater values (e.g., for Aso, unique topics have higher sentiment than common topics).

Next, we conduct a comparison of sentiments between PMs about common topics. Here, since the average sentiment differs from topic to topic and from PM to PM, we correct the sentiment values in order to make the comparison. In particular, we create a PM-topic sentiment matrix, calculate the expected value $E(T, p)$ for each cell in the matrix, and then calculate the deviation of each sentiment value from that expected value (i.e., residuals) [41]. If this deviation is positive, it means that the sentiment is higher than expected. The expectation and residual can be calculated as follows:

$$E(T,p) = \sum_T \sum_p S_{T,p} \times \frac{\sum_p S_{T,p}}{\sum_T \sum_p S_{T,p}} \times \frac{\sum_T S_{T,p}}{\sum_T \sum_p S_{T,p}}, \quad (5)$$

$$Residual_{T,p} = (S_{T,p} - E_{T,p})/\sqrt{E_{T,p}}, \quad (6)$$

where $S_{T,p}$ means the average of tweets' sentiment for a topic T and a PM p.

Figure 3 is a heatmap visualizing the residuals, which shows that sentiment value varies by topic. First, we could observe the overall trend in the figure: in the last four PMs, the sentiment was basically high on Personnel and low on Economy. These topics may be the ones that should be anticipated first at the timing of the PM's inauguration.

Looking at the comparison of PMs, we could confirm the differences in sentiments by topics. For example, Kishida was said to be expected of high diplomatic skills [43], which corresponds to the result that there is high sentiment for him. In addition, Suga was known to have been favorably received about the appointment of cabinet ministers [25], which corresponds to his high sentiment toward Personnel. This suggests a correspondence between the actual political situation and the initial public expectations from the public.

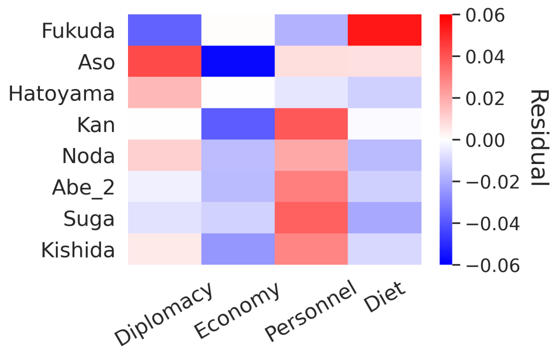

Figure 3: Average sentiments of the tweets with respect to each PM for each topic. The color of each cell corresponds to the value of residual (eq 6); the redder the color, the higher the sentiment. We clipped the values above an absolute value of 0.06 for visibility.

6 RQ3: WHO ARE CHANGING THE SENTIMENTS TOWARD THE PRIME MINISTERS?

For the administrative party, one of the intentions of the change of PM is to aim to calm users who have negative feelings toward the PM. However, the current hearing survey system for PM's approval rating does not track individual users, making it difficult to determine whether people with negative sentiment have decreased or people with positive sentiment have increased. Therefore, in this section, we conduct an analysis to capture changes in the sentiment at a user level by tracking individual users before and after the change of PMs.

Here, we examine whether there is an overlap in the user base of the previous and new PMs' tweets. Specifically, we first collected tweets for each change timing for two periods: 60 days before the resignation of the former PM (Before) and 60 days after the inauguration of the new PM (After). Then, we classified the authors of the tweets of each change timing into four groups: *Prev*, users who tweeted only about the previous PM; *New*, users who tweeted only about the new PM; *Both*, users who tweeted about both the previous and the new PM; *None*, users who did not tweet about either the previous and the new PM. Figure 4 shows the transition of these users Before and After.

We found the two types of users dominate the majority of users in each change timing: Prev⟶None, i.e., users who tweeted about the previous PM in Before but did not tweet about either the previous PM or the New PM in After; None⟶New, i.e., users who did not tweet about either the PM in Before but tweeted about only the New PM in After. To measure the presence of these users, we computed the following two indices:

$$F_{out} = 100 \times \frac{|\text{Before\_Prev} \cap \text{After\_None}|}{|\text{Before\_Prev}|}$$

$$F_{in} = 100 \times \frac{|\text{Before\_None} \cap \text{After\_New}|}{|\text{After\_New}|}$$

where A_B is the user group B in the period A. F_{out} indicates the percentage of users who stopped tweeting about either PM among those who tweeted about only the previous PM. F_{in} is the percentage of users who did not tweet about either PM among those who newly tweeted about only New PM. These two indicators are annotated in Figure 4.

The figure shows that F_{out} is mostly above 50% and F_{in} is all above 65%, indicating that the majority of users who mention the previous and new PMs are being replaced before and after a change timing of PMs. In other words, when there is a change in sentiment before and after a change timing of PMs, it is more likely that users with different perceptions have replaced the whole opinion rather than a change in the same individual's state of mind.

We also compare the sentiment among groups in user transition. We chose the groups for the comparison as the Prev⟶None and None⟶New groups, which are in the majority of the user transitions. In addition, we compare the sentiments of the same PM as New PM and as Prev PM, i.e., we compare the None⟶New groups and Prev⟶None for the same PMs. Figure 5 shows the result of the comparison.

First, comparing Prev⟶None and None⟶New at the same change timing, we find that there was a significant difference in sentiment in seven out of eight times. This means that the majority of tweets are substituted by a group with different sentiments. In five of those timings, the None⟶New group had higher sentiment, and in 2 timings, the None⟶New group had lower sentiment. This was in almost perfect agreement with the timing at which the jump in Figure 2 was confirmed. In other words, the change in sentiment that occurred around the time of change timings was confirmed to be largely due to the substitution of these groups of tweets.

Furthermore, a comparison of sentiments about the same PM at different timings shows that there was a significant difference in five out of eight cases. Moreover, four of the eight times, Prev⟶None sentiment was significantly lower, meaning that the PM's sentiment that was high when he took office was lower just before he resigned. This could be a case of the honeymoon effect being confirmed in another way.

Here, we should note that only Abe exhibits extremely low sentiment. Abe has apparently lower sentiment than other PMs at the end of his first administration, as well as at the beginning and end of his second administration. It seems that the low sentiment at the end of the first administration carried over into the second administration. While a PM's name recognition often jumps when they become PM, Abe's name recognition was already strong at the beginning of his second administration. This result may explain why the RDD results in Figure 2 indicate a drop in sentiment only when Abe started his administration.

Finally, we compare the same PM at the same change timing. Here, we compare Prev⟶None and None⟶Prev for the same PM. The PMs resign as a form of taking responsibility for their low approval ratings, and we test whether this eases the harsh atmosphere against them. We got the result that three out of eight times, the sentiment of None⟶Prev is significantly higher than that of Prev⟶None (Hatoyama, Abe_2, and Suga). Conversely, on one timing, None⟶Prev was lower (Kan). This indicates that there is a tendency for some users to emerge to appreciate a previous PM's labor, although not always.

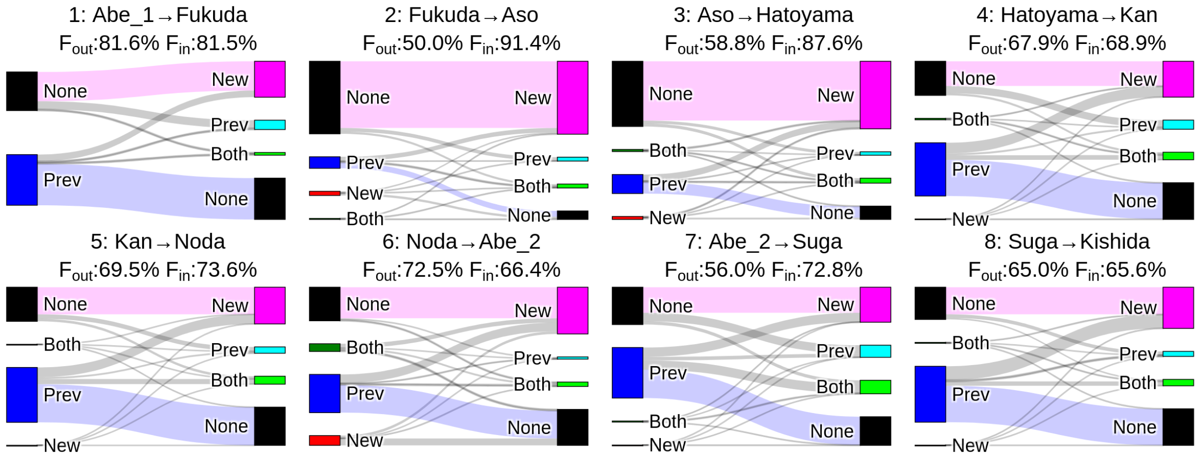

Figure 4: Sankey diagrams indicating user group transitions at each change timing. Refer to the main text for the annotations.

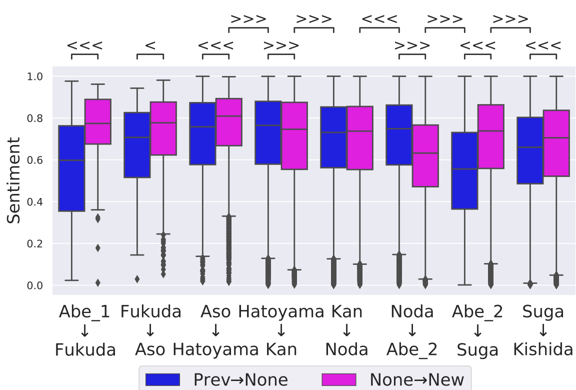

Figure 5: Boxplot of sentiment for groups of user transitions at each change timing. The sentiment of the groups of Prev→None and None→New are compared in each change timing. Also, the sentiments for the same PM in different change timings are also compared. The number of brackets indicates the result of the Mann-Whitney U test: 3: p < 0.001, 2: p < 0.01, 1: p < 0.05. The direction of the brackets is greater for the direction of the open side of the brackets. Visualization by [10].

7 DISCUSSION AND CONCLUSION

7.1 Main Findings

In this study, we attempted to explore and characterize the honeymoon effect on social media by analyzing Japanese Twitter users' reactions to eight changes in the PM in Japan. The results showed that in the RDD for the sentiment of the previous and new PMs before and after their change (RQ1), we confirmed a significant jump in sentiment in four out of eight times, with one timing when sentiment significantly declined and no significant results in the remaining three times. The results suggest that there is an overall tendency to have the honeymoon effect on social media, although not in all cases.

In the analysis of sentiment by topics (RQ2), the sentiment was often significantly lower for PM-specific topics than for common topics to all PMs (4 out of 8 times). Comparisons among PMs using topics common to all PMs showed that topics with high sentiment differed by the PM, demonstrating the effectiveness of social media in understanding the relative position of each PM and the issues expected of each PM.

Finally, the user-based analysis (RQ3) revealed that the majority of users were replaced between those who tweeted about the previous PM before the resignation and those who tweeted about the new PM after the inauguration. Furthermore, the sentiments of these replaced user bases were significantly different in many cases (7 out of 8 times), indicating that the user groups which the administration should appeal to on social media may differ depending on whether it is immediately after the inauguration or when the approval rating is declining. In addition, the sentiment of users who tweeted about the same PM after the inauguration and before the resignation was significantly different (5 out of 8 times), and in many cases, the sentiment was lower before the resignation (4 out of 5 times), confirming the honeymoon effect in a different way.

7.2 Implications for practitioners

The honeymoon effect was observed in social media at the PM change timings, although not always, suggesting that if the honeymoon effect emerges, social media could be used as a gauge of public response to policies that are implemented during the honeymoon period. For example, the pressure Suga put on academia soon after taking office received the lowest sentiment, which is claimed to have actually influenced the unnecessary drop in his approval rating [42]. We confirmed that social media had captured this impact, and considering that the typical approval surveys are conducted on a monthly basis, monitoring social media may have enabled a quicker reaction to the drop in sentiment.

When a country's leadership changes, the new PM must draw lessons from the cases of the previous leaders in managing the government, but each politician also has unique characteristics and expectations, making it challenging to apply the lessons. It could be easier to communicate with the public if we could identify the features that are popular for each PM, as we have just confirmed. For example, Suga is particularly expected in cabinet appointments, while Kishida has high sentiment on foreign affairs. It would be important to recognize the strengths of each PM and take them into account in their communication strategies.

PMs have greater control over who they interact with on social media than in traditional media. In this regard, it would be important to determine whether users are continuously vocal or one-time users in terms of different change timings and PMs. In fact, as we confirmed, new users tend to generate higher sentiment right after the inauguration of a new PM. In this regard, an essential tactic for enhancing general attitude is likely to be how to involve one-time posters in communication.

7.3 Limitation and future work

RDD setting and more elaborate model: In conducting the RDD, we used sentiments toward the *different* PMs before and after the threshold. We made this decision in order to be consistent with traditional surveys on approval ratings, but we could not fully take into account the impact on the potential impact of individual PM on sentiment. Modeling the potential impact on the PM's personal sentiment is challenging; for example, using sentiment prior to the PM's inauguration would result in a different nature of sentiment because a PM's name recognition surges only after becoming PM in many cases. However, the result of this study showed that only Abe, who was appointed to PM twice, received anomaly sentiments; thus, we thought the same model could basically be applied to figures who served as PM for the first time. Nevertheless, there is still room for improvement in the modeling of the elaboration of the impact of each individual PM, including the change of administrative party (LDP and DPJ), for example.

Consideration of user scale: The second Abe administration lasted eight years, during which time the number of Twitter users increased dramatically, which might affect the behavior of sentiment at an aggregated level. In fact, for the two PMs after Abe (i.e., Suga and Kishida), the behavior of sentiment toward them was specifically similar, including RDD results; thus, it is reasonable to anticipate that the analysis of PMs will be continued in the future in order to confirm this tendency with a sufficiently large number of users.

Other countries and domains: The honeymoon effect in actual approval ratings has also been observed in other countries besides Japan; thus, the analysis of other countries (e.g., the United Kingdom, which has a similar parliamentary system) is also expected to be conducted. Also, in addition to politics, companies have also used executive resignations to recover from scandals and business downturns; thus, it would be interesting to analyze the effect of such actions, including the linkage with stock prices.

Stance detection: Sentiment analysis is used in this study to examine the honeymoon effect. An alternative approach would be stance detection. Since stance and sentiment are known to differ from one another [1], undertaking stance analysis is also useful for gaining a more multifaceted understanding of the phenomena.

The end of the honeymoon period: We focused on the honeymoon effect and analyzed it at the timing of PM changes. On the other hand, it would also be important to analyze the timing of the end of the honeymoon period, i.e., when the high approval ratings of PMs start to fall.

More sophisticated flow: In this study, we propose a data-processing flow to classify unique topics and common topics for PMs. However, this flow requires manual adjustment at the end, and it would be desirable to be able to efficiently extract unique/common topics using a more sophisticated method.

Aspect-based sentiment analysis: We measured the sentiments for topics by calculating the average sentiments of tweets of a topic. An alternative approach would be to use aspect-based sentiment analysis [35], which trains the sentiments of texts with certain aspects.

7.4 Ethical Considerations

We pay the utmost attention to the privacy of individuals in this study. We did not include personal names except for PM or account names in our analysis. When sharing our tweet data, we will publish only a list of tweet IDs according to Twitter's guidelines.

8 APPENDICES

REFERENCES

[1] Abeer AlDayel and Walid Magdy. 2021. Stance detection on social media: State of the art and trends. *Information Processing & Management* 58, 4 (2021), 102597.

[2] Jisun An and Ingmar Weber. 2015. Whom should we sense in "social sensing"-analyzing which users work best for social media now-casting. *EPJ Data Science* 4 (2015), 1–22.

[3] Julia Azari. 2017. A President's First 100 Days Really Do Matter | FiveThirtyEight. https://fivethirtyeight.com/features/a-presidents-first-100-days-really-do-matter/. (Accessed on 11/28/2022).

[4] Siva K Balasubramanian, Mustafa Bilgic, Aron Culotta, Libby Hemphill, Anita Nikolich, and Matthew A Shapiro. 2022. Leaders or Followers? A Temporal Analysis of Tweets from IRA Trolls. In *ICWSM*, Vol. 16. 2–11.

[5] Matthew N Beckmann and Joseph Godfrey. 2007. The policy opportunities in presidential honeymoons. *Political Research Quarterly* 60, 2 (2007), 250–262.

[6] Philip Begley and Sally Sheard. 2021. From "Honeymoon Period" to "Stable Marriage":: The Rise of Management Consultants in British Health Policymaking. *Bulletin of the History of Medicine* 95, 2 (2021), 227.

[7] Michael Bernhard, Christopher Reenock, and Timothy Nordstrom. 2003. Economic Performance And Survival In New Democracies: Is There a Honeymoon Effect? *Comparative Political Studies* 36, 4 (2003), 404–431.

[8] Gavin Blair. 2010. Japan's new prime minister stumbles over consumption tax - CSMonitor.com. https://www.csmonitor.com/World/Asia-Pacific/2010/0712/Japan-s-new-prime-minister-stumbles-over-consumption-tax. (Accessed on 12/01/2022).

[9] David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research* 3, Jan (2003), 993–1022.

[10] Florian Charlier, Marc Weber, Dariusz Izak, Emerson Harkin, Marcin Magnus, Joseph Lalli, Louison Fresnais, Matt Chan, Nikolay Markov, Oren Amsalem, Sebastian Proost, Agamemnon Krasoulis, getzze, and Stefan Repplinger. 2022. *Statannotations*. https://doi.org/10.5281/zenodo.7213391

[11] Clayton Allen Davis, Onur Varol, Emilio Ferrara, Alessandro Flammini, and Filippo Menczer. 2016. Botornot: A system to evaluate social bots. In *WWW*. 273–274.

[12] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *NAACL-HLT*. 4171–4186.

[13] Torun Dewan and Keith Dowding. 2005. The corrective effect of ministerial resignations on government popularity. *American Journal of Political Science* 49, 1 (2005), 46–56.

[14] Worldwide distribution. 2022. Japanese - Worldwide distribution. https://www.worlddata.info/languages/japanese.php. (Accessed on 11/24/2022).

[15] Casey Byrne Knudsen Dominguez. 2005. Is it a honeymoon? An empirical investigation of the president's first hundred days. In *Congress & the Presidency: A Journal of Capital Studies*, Vol. 32. Taylor & Francis, 63–78.

[16] Clare Duffy. 2022. Elon Musk wants Twitter users to pay to be verified. It could create a new set of headaches for the company | CNN Business. https://edition.cnn.com/2022/11/03/tech/elon-musk-twitter-verification-plans/index.html. (Accessed on 11/24/2022).

Merged topic	Topic id	Top words	Tweet num	Total
Diplomacy	24	Korea, Japan, Japan-Korea, President, summit meeting, teleconference	111,824	505,899
	33	Sign, North Korea, abduction, release, kidnapping	50,269	
	41	summit meeting, China, visit, diplomacy, minister	90,324	
	45	worship, Yasukuni shrine, Yasukuni, Yasukuni visit, China	38,771	
	52	Okinawa, Futenma, relocation, governor, meeting	59,881	
	54	China, diplomacy, Senkaku, U.S., Biden	123,616	
	70	speech, announcement, UN, reduction, summit	31,214	
Economy	22	bureaucrats, finance ministry, people, tax hike, Ozawa	94,702	481,475
	25	enterprise, wage increase, request, economy, professor	30,518	
	32	tax hike, consumption tax, income, finance, taxation	67,136	
	43	budget, government, direction, politics, meeting	72,676	
	61	people, policy, bureaucracy, tax hike, LDP	68,554	
	66	yen weakening, economy, policy, stock price, expectation	98,173	
	72	economy, recovery, measures, employment, challenges	49,716	
Cabinet personnel	12	bureaucrats, personnel, appointment, LDP, charge	46,961	170,082
	30	inauguration, cabinet ministers, Tokyo, economy, press conference	80,378	
	38	minister, Renho, cabinet minister, minister in charge, allegations	42,743	
National diet	3	policy speech, speech, concept, LDP, Osaka	50,677	532,439
	4	LDP, people, debate, opposition party, diet	66,684	
	23	approval rating, support, poll, polls, survey, cabinet approval rating	146,087	
	29	Diet, opposition parties, questions from representatives, policy speeches, extraordinary diet session	115,450	
	53	question, press, pressure, incident, problem	43,924	
	73	people, language, politics, society, unhappiness	109,617	

Table 3: Common topics that are aggregated by clustering. "Merged topic" indicates the name of the cluster as labeled by the authors. "Topic id" is a random id assigned to a topic by BTM. "Top words" are the five most possible words belonging to each topic. The number of tweets for each topic and cluster is also shown.

[17] Stephanie Edgerly and Emily K Vraga. 2019. The blue check of credibility: Does account verification matter when evaluating news on Twitter? *Cyberpsychology, behavior, and social networking* 22, 4 (2019), 283–287.

[18] FiveThirtyEight. 2021. Is The Honeymoon Period In The Presidency Over? | FiveThirtyEight. https://fivethirtyeight.com/live-blog/biden-inauguration/301893/?amp. (Accessed on 11/28/2022).

[19] Hirosan. 2019. Hironsan/asari: Japanese sentiment analyzer implemented in Python. https://github.com/Hironsan/asari. (Accessed on 11/25/2022).

[20] Hirosan. 2019. Japanese Sentiment Analyzer was created and packaged. - Ahogrammer. https://hironsan.hatenablog.com/entry/japanese-sentiment-analyzer. (Accessed on 11/25/2022).

[21] Manoel Horta Ribeiro, Shagun Jhaver, Savvas Zannettou, Jeremy Blackburn, Gianluca Stringhini, Emiliano De Cristofaro, and Robert West. 2021. Do platform migrations compromise content moderation? evidence from r/the_donald and r/incels. *CSCW* 5 (2021), 1–24.

[22] William J Hughes. 1995. The "not-so-genial" conspiracy: The New York Times and six presidential "honeymoons," 1953–1993. *Journalism & Mass Communication Quarterly* 72, 4 (1995), 841–850.

[23] Tasuku Igarashi, Shimpei Okuda, and Kazutoshi Sasahara. 2021. Development of the Japanese Version of the Linguistic Inquiry and Word Count Dictionary 2015 (J-LIWC2015). *Frontiers in Psychology* (2021), 665.

[24] Naoki Kambe. 2017. Representing Disaster with Resignation and Nostalgia: Japanese Men's Responses to the 2011 Earthquake. *RCC Perspectives* 4 (2017), 15–22.

[25] Elaine Lies. 2021. New appointment by Japan's Suga on vaccinations may boost support, neutralise rival | Reuters. https://www.reuters.com/article/health-coronavirus-japan-suga-idINKBN29O0K2. (Accessed on 12/01/2022).

[26] Ko Maeda. 2015. Honeymoon or consolidation, or both?: Time dependence of democratic durability. *Democratization* 23, 4 (2015), 575–591.

[27] Mikitaka Masuyama. 2007. The Survival of Prime Ministers and the House of Councillors. *Social Science Japan Journal* 10, 1 (2007), 81–93.

[28] Fionn Murtagh and Pierre Legendre. 2014. Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion? *Journal of classification* 31, 3 (2014), 274–295.

[29] Jcast news. 2020. Why did the Abe administration and the LDP's approval ratings "reverse" from the slump? Comparing the "exceptional results" of each company's polls. https://www.j-cast.com/2020/09/10394047.html Accessed on 11/20/2022.

[30] NHK. 2022. Monthly Survey of Political Attitudes. https://www.nhk.or.jp/bunken/research/yoron/political/2020.html Accessed on 11/22/2022.

[31] NHK. 2022. Musk calls Twitter 'Japan-centric' given high number of users | NHK WORLD-JAPAN News. https://www3.nhk.or.jp/nhkworld/en/news/20221123_10/. (Accessed on 12/01/2022).

[32] Benjamin Nyblade. 2011. The 21st century Japanese prime minister: an unusually precarious perch. *The Journal of Social Science* 62, 1 (2011), 195–209.

[33] James W Pennebaker, Ryan L Boyd, Kayla Jordan, and Kate Blackburn. 2015. *The development and psychometric properties of LIWC2015*. Technical Report.

[34] Juergen Pfeffer, Angelina Mooseder, Luca Hammer, Oliver Stritzel, and David Garcia. 2023. This Sample seems to be good enough! Assessing Coverage and Temporal Reliability of Twitter's Academic API. *ICWSM* (2023). to appear.

[35] Maria Pontiki, Dimitrios Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammad Al-Smadi, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, Orphée De Clercq, et al. 2016. Semeval-2016 task 5: Aspect based sentiment analysis. In *International workshop on semantic evaluation*. 19–30.

[36] Jiji Press. 2021. Kishida Faces Major Task of Fleshing Out Handout Pledge | Nippon.com. https://www.nippon.com/en/news/yjj2021110401019/. (Accessed on 12/01/2022).

[37] Mahmoud Qadan and Yasmeen Idilbi. 2022. Presidential honeymoons, political cycles and the commodity market. *Resources Policy* 77 (2022), 102631.

[38] Adrian Rauchfleisch and Jonas Kaiser. 2020. The false positive problem of automatic bot detection in social science research. *PloS one* 15, 10 (2020), e0241045.

[39] Osamu Ryoichi. 2021. POLITICAL CHANGING FOR PRIME MINISTER OF JAPAN. *International Journal of Law Reconstruction* 5, 1 (2021), 75–92.

[40] Paolo Segatti, Monica Poletti, and Cristiano Vezzoni. 2015. Renzi's honeymoon effect: The 2014 European election in Italy. *South European Society and Politics* 20, 3 (2015), 311–331.

[41] Donald Sharpe. 2015. Chi-square test is statistically significant: Now what? *Practical Assessment, Research, and Evaluation* 20, 1 (2015), 8.

[42] Asahi Shimbun. 2020. EDITORIAL: Suga's Science Council meddling puts academic freedom at risk | The Asahi Shimbun: Breaking News, Japan News and Analysis. https://www.asahi.com/ajw/articles/13784232. (Accessed on 12/01/2022).

[43] Asahi Shinbun. 2021. Who is Fumio Kishida? Foreign observers expect his diplomatic skills and stability in his administration | Asahi Shinbun. https://www.asahi.com/articles/ASP9Y6T7PP9YUHBI032.html. (in Japanese) (Accessed on 12/01/2022).

[44] Ashok N Srivastava and Mehran Sahami. 2009. *Text mining: Classification, clustering, and applications*. CRC press.

[45] Masatoshi Suzuki. 2019. cl-tohoku/bert-japanese: BERT models for Japanese text. https://github.com/cl-tohoku/bert-japanese. (Accessed on 11/25/2022).
[46] Osamu Tsukimori and Rebekah Kebede. 2011. Analysis: Reactor restarts first energy hurdle for Japan's Noda | Reuters. https://www.reuters.com/article/us-japan-energy-noda-idUSTRE7801LS20110901. (Accessed on 12/01/2022).
[47] Yu Uchiyama. 2022. Japanese prime ministers and party leadership. *Asian Journal of Comparative Politics* (2022), 20578911221114509.
[48] Xiaohui Yan, Jiafeng Guo, Yanyan Lan, and Xueqi Cheng. 2013. A biterm topic model for short texts. In *WWW*. 1445–1456.
[49] Weizhong Zhao, James J Chen, Roger Perkins, Zhichao Liu, Weigong Ge, Yijun Ding, and Wen Zou. 2015. A heuristic approach to determine an appropriate number of topics in topic modeling. In *BMC bioinformatics*, Vol. 16. Springer, 1–10.