Note: Urbanization and Literacy as factors in Politicians' Social Media Use in a largely Rural State: Evidence from Uttar Pradesh, India

Asmit Kumar Singh*
asmit18025@iiitd.ac.in
Indraprastha Institute of Information Technology, Delhi
Delhi, Delhi, India

Jivitesh Jain*
jivitesh.jain@students.iiit.ac.in
International Institute of Information Technology, Hyderabad
Hyderabad, Telangana, India

Lalitha Kameswari*
v.a.lalitha@research.iiit.ac.in
International Institute of Information Technology, Hyderabad
Hyderabad, Telangana, India

Ponnurangam Kumaraguru
pk.guru@iiit.ac.in
International Institute of Information Technology, Hyderabad
Hyderabad, Telangana, India

Joyojeet Pal
joyojeet@umich.edu
University of Michigan
Ann Arbor, Michigan, USA

ABSTRACT

With Twitter growing as a preferred channel for outreach among major politicians, there have been focused efforts on online communication, even in election campaigns in primarily rural regions. In this paper, we examine the relationship between politicians' use of social media and the level of urbanization and literacy by compiling a comprehensive list of Twitter handles of political party functionaries and election candidates in the run-up to the 2022 State Assembly elections in Uttar Pradesh, India. We find statistically significant relationships between political Twitter presence and levels of urbanization and with levels of literacy. We also find a strong correlation between vote share and Twitter presence in the winning party, a relationship that is even stronger in urban districts. This provides empirical evidence that social media is already a central part of electoral outreach processes in the Global South, but that this is still selectively more relevant to voters in, and politicians standing for elections from urban and higher-educated regions.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in collaborative and social computing**; • **Human-centered-computing** → Social Network Analysis.

KEYWORDS

Computational Social Science, Twitter, Election Campaign, Urbanization

*Authors contributed equally to this research.

COMPASS '22, June 29-July 1, 2022, Seattle, WA, USA

ACM ISBN 978-1-4503-9347-8/22/06…$15.00
https://doi.org/10.1145/3530190.3534845

ACM Reference Format:
Asmit Kumar Singh, Jivitesh Jain, Lalitha Kameswari, Ponnurangam Kumaraguru, and Joyojeet Pal. 2022. Note: Urbanization and Literacy as factors in Politicians' Social Media Use in a largely Rural State: Evidence from Uttar Pradesh, India. In *ACM SIGCAS/SIGCHI Conference on Computing and Sustainable Societies (COMPASS) (COMPASS '22), June 29-July 1, 2022, Seattle, WA, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3530190.3534845

1 INTRODUCTION

The past decade has seen an astronomical increase in access to the Internet and various social media platforms like Twitter, Facebook, Instagram, etc. This has resulted in the availability of a huge amount of data which allows insight into various aspects of communicative behavior in social and political settings. In politics, online outreach and propaganda, as well as campaigning during elections, has grown in size on account of the convenience of direct politician-to-citizen communications, bypassing professional journalists, as well as the temporal benefits of allowing political actors to communicate when and how they want to. While no research has shown definitive relationships between social media communication and electoral outcomes, a growing body of work has examined various aspects of political communication that have changed in recent years as a result of politicians engaging more aggressively on social media, especially Twitter [3, 4, 8].

In this study, we explore the relationship between presence of politicians on Twitter and level of urbanization and literacy in context of the 2022 Uttar Pradesh State Assembly elections. We do this by investigating the following research questions.

(1) Is the presence of politicians on Twitter related to the level of urbanization?
(2) Is the presence of politicians on Twitter related to the level of literacy?
(3) Is there a relationship between the electoral vote share and Twitter presence of the winning party, and how does this relationship relate to urbanization?

We address these questions, by taking the case of Uttar Pradesh, India's most populous state. With over 200 million residents, it has roughly the same number of people as all of the United Kingdom, France, and Italy combined. It is a largely rural state, with over-three fourths of its denizens living in villages. Literacy in the state is lower than the average for India; less than 60 per cent of females residents can read or write their own name. The state is divided into 75 administrative districts and 403 assembly constituencies, each of which elects one member to the state's legislative assembly.

The Bharatiya Janata Party (BJP), a right-wing party currently also in power at the national level, has been in power in the state since 2017, led by chief minister Yogi Adityanath, a Hindu religious leader turned politician. BJP again won the 2022 assembly elections. The party is by far the most dominant and well-funded, having won two general elections and running the state legislatures directly or in coalition in roughly two-thirds of the country's states. Its leader, Prime Minister Narendra Modi, is a widely popular figure. BJP's primary opposition in the state is the Samajwadi Party (SP), an ideologically-socialist party that derives its vote bank primarily from caste and religious alliances. The SP was in power in the state until 2017, led by Akhilesh Yadav, the son of the politician who founded the party. Three other major parties that were active and widely contesting in the state include the main national opposition party - the Indian National Congress (INC), the Bahujan Samaj Party (BSP), and the Aam Aadmi Party (AAP). In addition, there were a number of smaller, primarily caste-based parties that were fighting the elections mostly in alliance with one or the other major parties.

2 RELATED WORK

Computational techniques have significantly furthered our understanding of political communication on social media in the Global South. In particular, India has been an important space in this regard due to widespread adoption of social media by politicians. There have been several efforts in the past to study the elections campaigns, results, and the voter behavior as this has significant implications across social, political, economic, and technological spheres.

The increasing use of Twitter by politicians, journalists, political strategists and citizens has made it an important part of the networked sphere in which political issues are publicly negotiated. Previous efforts have been made to correlate Twitter campaigns and election outcomes in India for the Lok Sabha (National) elections [9, 14] as well as for the State elections in West Bengal [5] and Punjab [13]. They observed that the winning candidates tend to have a stronger Twitter presence, and the online activity of politicians can strongly contribute to the election outcome.

Similarly, there have been several interesting studies focusing on specific aspects of the Twitter behavior during elections, like location-based sentiment analysis of tweets to identify trends and patterns in online election campaigning [2, 12, 16]. Other work has looked at politicians' caste identity and how these play into their engagement with the party power structures when on social media [15]. Bot accounts and their activity during elections was studied in detail, and it was found that political parties aggressively used bot accounts to spread their preferred narrative on social media [7]. Hashtags have also been used to examine the semantics and rhetoric of political actors, especially as it relates to campaign strategies [6, 11].

To our knowledge, this work is the first of its kind which aims to correlate urbanization and literacy in a state with Twitter presence of the politicians and, subsequently, the election outcome.

3 DATASET AND METHODOLOGY

Our dataset consists of politicians and politically active[1] Twitter handles from the state of Uttar Pradesh. This information was supplemented with official election[2] and census data[3].

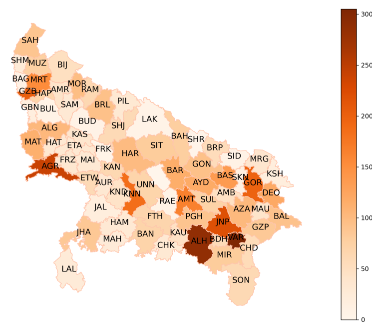


Figure 1: District map of the state of Uttar Pradesh indicating the number of politicians in our dataset from each district. Lucknow (LCK), the district which encompasses the state's capital city, has an abnormally high number of politicians (1,088) in our dataset, and has been omitted from this visualization.

Curating the user set We collected all Twitter accounts being followed by users from a seed set of ~3,900 Uttar Pradesh politicians collected using the NivaDuck pipeline [10]. Out of these ~850,000 collected accounts, we only retained those that mentioned at least one word from a list of curated political keywords in both English and Hindi (e.g. "MLA", "politician", party names and abbreviations) in their user bio, as well as a district or assembly constituency of the state of Uttar Pradesh. Further, we collected official lists of past and contesting assembly candidates with party and constituency information[4] and used the Google Search API[5] to identify their Twitter handles from query results. Finally, we obtained a set of 6,636 politicians and politically active accounts with associated district information (Figure 1).

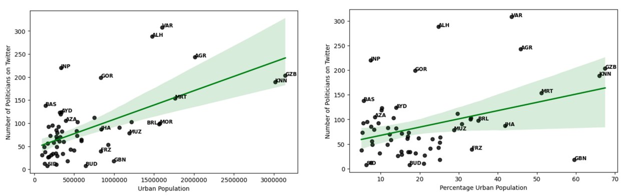
Adding party labels We could obtain party labels for ~60% of the politicians in our dataset from their user bios. For the remaining, we used majority-voting on the follower-friend network. A user for whom party annotation was not already available, was assigned the party in which they had the maximum connections, considering

[1]We consider as politically active any account that states itself as being affiliated to a particular party in its Twitter user bio.
[2]https://eci.gov.in
[3]https://censusindia.gov.in/
[4]http://uplegisassembly.gov.in/Members/main_members_en.aspx#/ElectedMembers
[5]https://developers.google.com/custom-search

only those with available party annotations. When tested on an 80-20 split of the portion of data with available party annotations, this approach yielded an accuracy upwards of 95%. Party- and district-wise statistics of our dataset are available in the supplementary materials.

Supplementary information In addition to the Twitter dataset, we also obtained district-wise demographic data for Uttar Pradesh from the 2011 Census of India[6] and provisional vote-share and vote-margin information for the elections in question from the Lok Dhaba dataset [1].

4 OBSERVATIONS AND RESULTS



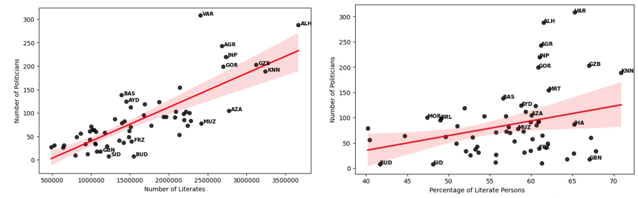
(a) Number **(b) Percentage**

Figure 2: Number and Percentage of urban residents in a district correlated with the number of politicians from that district active on Twitter.

4.1 Level of urbanization and Twitter presence of politicians

We examined the number of urban residents by location and found a statistically significant relationship between the number (and percentage) of urban residents and the likelihood that politicians in that area use Twitter. Taking the number of politicians on Twitter as the dependent variable and the urban population of the district as the independent variable, we found a strong correlation between the two with an R value of 0.61. The regression plot is shown in Figure 2 (a). The coefficient β_1 was found to be significantly different from 0 ($p<0.05$) (Table 1). Lucknow district, encompassing the state's capital city, has a large number of listed post holders for all parties. Hence, it is an outlier and was removed from this and subsequent regressions we ran (sections 4.2, 4.3). We see some variation – for instance, the districts with the highest urban populations, Ghaziabad (adjoining the country's capital, Delhi) and Kanpur, both have fewer politicians online than Varanasi, the Prime Minister's constituency, which has had a lot of mid- to lower-level politicians go online, as well as Allahabad/Prayagraj, which is a politically important district in the state, home to several key politicians from across parties.

When we ran the same regression to correlate the percentage of urban residents with the number of politicians active on Twitter, we again found a strong statistically significant ($p<0.05$) relationship between the two variables with an R value of 0.37, with some outliers as shown in Figure 2 (b). On the lower end, we see that Gautam Buddha Nagar/Noida has a smaller share of politicians active online (partly because of cross-listing with Delhi), whereas Gorakhpur, an overwhelmingly rural district, has a high number of politicians active on Twitter. This can be attributed to Chief Minister Yogi Adityanath, who contests from that district.

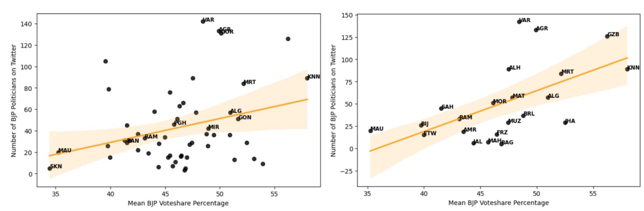


(a) Number **(b) Percentage**

Figure 3: Number and Percentage of literate residents in a district correlated with the number of politicians from that district active on Twitter.

4.2 Literacy rates and Twitter presence of politicians

We extend the earlier regression analysis by taking the literate population in the district as the dependent variable and the number of politicians on Twitter as the independent variable. The relationship here is the strongest, statistically significant ($p<0.05$) with an R value of 0.78 (Table 1). Again, Varanasi is an outlier, as seen in Figure 3 (a). As a sanity check, we ran the same regression on the percentage of literate persons in a district and found a statistically significant relationship ($p<0.05$) with an R value of 0.07 (Table 1). The regression plot is shown in Figure 3 (b).



(a) All districts **(b) Urban districts**

Figure 4: Mean vote share of BJP in winning districts correlated with the number of BJP politicians from that district active on Twitter.

4.3 Vote Share of the winning party and Twitter presence of politicians

We examined the relationship between the BJP's (winning party) vote share in districts and the number of politicians active on Twitter from the party and found a statistically significant correlation between the two. The correlation is observed to be much stronger when only urban districts are considered. The number of politicians from BJP was taken as the dependent variable and the mean vote share of BJP (mean of vote share in the winning constituencies for each district) was taken as the independent variable. The relationship was significant ($p<0.05$) with an R value of 0.29 (Table 2). The

[6] https://censusindia.gov.in/2011-common/censusdata2011.html

S.No	Regression Variables (X, Y)	Slope (β_1)	Intercept	Pearson's Correlation (R)	R Squared	P Value	Standard Error
1	Politicians on Twitter, Urban Population	6.20e-05	46.16	0.611	0.374	2.089e-07	1.054e-05
2	Politicians on Twitter, Percentage of Urban Population	1.67	50.72	0.37	0.14	0.003	0.54
3	Politicians on Twitter, Literate Population	7.24e-05	-32.78	0.78	0.61	1.19e-13	7.52e-06
4	Politicians on Twitter, Percentage of Literate Population	2.51	-84.66	0.07	0.61	0.03	1.17

Table 1: Results of Regression performed over 4 different variables to determine their correlation with the number of politicians on Twitter from a district.

S.No	Regression Variables (X, Y)	Slope (β_1)	Intercept	Pearson's Correlation (R)	R Squared	P Value	Standard Error
1	Mean BJP vote share in winning districts, Number of BJP Politicians in the district	2.22	-60.08	0.29	0.15	0.03	1.01
2	Mean BJP vote share in districts with >20% urban population, Number of BJP Politicians in the district	4.59	-164.72	0.58	0.23	0.01	1.43

Table 2: Results of regression performed over BJP (winning party) politicians correlating their mean vote share with the number of politicians.

result points towards a relation between a strong Twitter presence and a better on-ground election outcome.

In order to see the effect of urbanization on the relation, we considered the districts with at least 20% urban population. As seen from Table 2, the relationship becomes much stronger with an R value of 0.58, with the coefficient β_1 significantly different from 0 ($p<0.05$). The exceptions to the pattern are similar to those observed in sections 4.2, 4.1 – districts like Varanasi and Agra, as shown in Figure 4. The strong relationship shows that the winning party reaped the benefits of a stronger Twitter presence significantly more in urban districts, where the general population would tend to be more active online.

5 CONCLUSION

Our paper offers two important contributions to the literature. First, we hope that it lays to rest the consistent refrain against scholarship on social media in the global south as being an elite phenomenon. While we do show that the relatively better off residents of a state are more likely to find their politicians online, the larger fact is that not only is social media here to stay in Indian elections, but that it may indeed become a primary, if not moving, factor in urban elections. Our finding that a strong presence on Twitter and the party's vote share are related underlines that parties themselves recognize the importance of social media. We also find that while the margin of success for winning being strongly correlated to their Twitter presence could be a factor of several other elements, such as star politicians being more active on Twitter, that this finding is stronger in Urban districts suggests that the electorate is likewise engaging this trend. A decade ago, social media as an engine of campaigning was unthinkable – the current Chief Minister, now one of India's most followed politicians, was not even on social media, and the few that were had a largely cursory presence. What we are seeing now is that the Global South is already at the vanguard of how social media is impacting elections, and holding strong. A longitudinal study of the presence and activity of politicians on Twitter could quantify these effects, and follows in the natural course of future work.

REFERENCES

[1] Ananay Agarwal, Neelesh Agrawal, Saloni Bhogale, Sudheendra Hangal, Francesca Refsum Jensenius, Mohit Kumar, Chinmay Narayan, Basim U Nissa, Priyamvada Trivedi, and Gilles Verniers. 2022. *TCPD Indian Elections Data v2.0.* Trivedi Centre for Political Data, Ashoka University. https://lokdhaba.ashoka.edu.in/

[2] Omaima Almatrafi, Suhem Parack, and Bravim Chavan. 2015. Application of location-based sentiment analysis using Twitter for identifying trends towards Indian general elections 2014. In *Proceedings of the 9th international conference on ubiquitous information management and communication.* 1–5.

[3] Linh Dang-Xuan, Stefan Stieglitz, Jennifer Wladarsch, and Christoph Neuberger. 2013. An investigation of influentials and the role of sentiment in political communication on Twitter during election periods. *Information, communication & society* 16, 5 (2013), 795–825.

[4] Purva Grover, Arpan Kumar Kar, Yogesh K Dwivedi, and Marijn Janssen. 2019. Polarization and acculturation in US Election 2016 outcomes–Can twitter analytics predict changes in voting preferences. *Technological Forecasting and Social Change* 145 (2019), 438–460.

[5] Rynaa Grover, Gazal Shekhawat, and Joyojeet Pal. 2021. Twitter superstars don't win elections: A Poster on Twitter Campaigning and Electoral Realities in the 2021 West Bengal Assembly Elections. In *ACM SIGCAS Conference on Computing and Sustainable Societies.* 428–431.

[6] Saurabh Gupta, Asmit Kumar Singh, Arun Balaji Buduru, and Ponnurangam Kumaraguru. 2020. Hashtags are (not) judgemental: The untold story of Lok Sabha elections 2019. In *2020 IEEE Sixth International Conference on Multimedia Big Data (BigMM).* IEEE, 216–220.

[7] Hitkul Hitkul, Omkar Gurjar, Aanshul Sadaria, Kanay Gupta, Shashank Srikanth, Rajiv Ratn Shah, and Ponnurangam Kumaraguru. 2020. Are Bots Humans? Analysis of Bot Accounts in 2019 Indian Lok Sabha Elections (Workshop Paper). In *2020 IEEE Sixth International Conference on Multimedia Big Data (BigMM).* IEEE, 441–450.

[8] Kawaljeet Kaur Kapoor and Yogesh K Dwivedi. 2015. Metamorphosis of Indian electoral campaigns: Modi's social media experiment. *International Journal of*

Indian Culture and Business Management 11, 4 (2015), 496–516.

[9] Devansh Manu, Radhika Krishnan, and Ponnurangam Kumaraguru. 2020. Analysing How the Shift in Discourses on Social Media Affected the Narrative Around the Indian General Election 2019. *Journal of Advanced Research in Social Sciences* 3 (05 2020), 21–31. https://doi.org/10.33422/jarss.v3i1.296

[10] Anmol Panda, A'ndre Gonawela, Sreangsu Acharyya, Dibyendu Mishra, Mugdha Mohapatra, Ramgopal Chandrasekaran, and Joyojeet Pal. 2020. NivaDuck - A Scalable Pipeline to Build a Database of Political Twitter Handles for India and the United States. In *SMSociety'20: International Conference on Social Media and Society*. Association for Computing Machinery, Association for Computing Machinery, 200–209. https://www.microsoft.com/en-us/research/publication/nivaduck-a-scalable-pipeline-to-build-a-database-of-political-twitter-handles-for-india-and-the-united-states/

[11] Anmol Panda, Ramaravind Kommiya Mothilal, Monojit Choudhury, Kalika Bali, and Joyojeet Pal. 2020. Topical Focus of Political Campaigns and its Impact: Findings from Politicians' Hashtag Use during the 2019 Indian Elections. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW1 (2020), 1–14.

[12] Jyoti Ramteke, Samarth Shah, Darshan Godhia, and Aadil Shaikh. 2016. Election result prediction using Twitter sentiment analysis. In *2016 International Conference on Inventive Computation Technologies (ICICT)*, Vol. 1. 1–5. https://doi.org/10.1109/INVENTIVE.2016.7823280

[13] Prabhsimran Singh, Yogesh K Dwivedi, Karanjeet Singh Kahlon, Annie Pathania, and Ravinder Singh Sawhney. 2020. Can twitter analytics predict election outcome? An insight from 2017 Punjab assembly elections. *Government Information Quarterly* 37, 2 (2020), 101444.

[14] Avinash Tulasi, Kanay Gupta, Omkar Gurjar, S Buggana, Paras Mehan, Arun Balaji Buduru, and Ponnurangam Kumaraguru. 2020. Who Won it Online? A comparative study of 2019 Indian General Elections on Twitter.

[15] Palashi Vaghela, Ramaravind K Mothilal, and Joyojeet Pal. 2021. Birds of a Caste - How Caste Hierarchies Manifest in Retweet Behavior of Indian Politicians. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW3, Article 212 (jan 2021), 24 pages. https://doi.org/10.1145/3432911

[16] Ussama Yaqub, Nitesh Sharma, Rachit Pabreja, Soon Ae Chun, Vijayalakshmi Atluri, and Jaideep Vaidya. 2020. Location-based sentiment analyses and visualization of Twitter election data. *Digital Government: Research and Practice* 1, 2 (2020), 1–19.