Journal Homepage: www.ijarpr.com ISSN: 3049-0103 (Online)

International Journal of Advance Research Publication and Reviews

Vol 03, Issue 9, pp 626-630, September, 2026

Estimating Political Ideology on Twitter via Linear Inverse Modeling

[1]G. Joy Sangeeth Raj, [2] M Sudhakar,[3]Chaudhari Soham, [4]Kanjarla Vaishnavi, [5]P Akhila

[1,2,3,4] *Department of ECE, Narsimha Reddy Engineering College*

ABSTRACT:

Online Social Networks (OSNs) have evolved into central platforms for political discourse, opinion dissemination, and media commentary. However, public news coverage and political punditry in these networks are frequently marked by partisan bias. Estimating the political leaning of social media actors presents significant challenges in terms of quantification and computational scalability. In this paper, we formulate political leaning estimation as an ill-posed linear inverse problem based on a consistency principle between tweeting and retweeting behaviors. Our proposed convex optimization formulation directly infers numerical leaning scores without requiring explicit crawling of full network topologies or graph traversal. We evaluate our method using a dataset of 119 million U.S. election-related tweets collected over seven months. Empirical results demonstrate strong alignment with conventional media bias baselines while remaining resilient to API rate limits and computational overhead.

Keywords: Political leaning; Online Social Networks; Twitter; Retweet dynamics; Linear inverse optimization; Media bias; Sentiment analysis.

1. INTRODUCTION

Online Social Networks (OSNs) have become important channels for political communication, news dissemination and public discussion. Social media platforms like Twitter are utilized by political actors, media outlets, journalists, and citizens for the swift creation and distribution of information about elections. The structure of these interactions is also a useful signal of ideological alignment. Previous work on communication on Twitter has shown that retweet networks can be highly partisan, suggesting that information diffusion patterns can reflect substantive differences in political orientation that are not reducible to textual content.

But using social-media data to study electoral behavior calls for thoughtful thought about representativeness and limitations on inference. Twitter-based election prediction studies have indicated that perceived predictive performance can be influenced by sampling procedures, demographic differences between platform users and the general electorate, and methodological choices around data collection and evaluation . Thus, the activity on Twitter should not be used as a direct proxy for the population level voting behavior. Instead, an important complementary problem is to estimate the ideological or political positions of very active information sources, namely media organizations, political commentators and influential accounts.

Quantifying political leaning poses two closely related computational challenges. First, the representation should be a continuous and interpretable measure, rather than simply a categorization of accounts into discrete political categories. Previous work has shown that large corpora of Twitter interactions can be used to estimate ideological preferences . Bayesian ideal-point methods have also explored the use of social-network structure to estimate political positions . Empirical studies also suggest that political communication is often characterized by interactions between ideologically similar users, but the extent of ideological segregation appears to depend on the issue and communication context. Second, scalability is a must to handle millions of tweets and retweet relations. Approaches that require extensive network building, graph traversal, or advanced text processing can incur large computational and data-collection burdens, especially when operating under platform API limitations.

To address these issues, this work formulates political-leaning estimation as an ill-posed linear inverse problem based on the coherence of tweeting and retweeting behavior. The proposed framework estimates continuous leaning scores based on the observed interaction patterns using an optimization-based formulation, instead of explicitly reconstructing the full social-network topology. This view is closely connected to the framework proposed by Wong et al. who demonstrated that political leaning can be estimated from the correspondence between tweeting and retweeting behavior using a scalable

linear inverse formulation without the need for explicit knowledge of the network topology. In this spirit, the present study addresses a computationally efficient formulation for large-scale election-related Twitter data and highlights numerical quantification, scalability and robustness to data access constraints. This work provides a systematic framework for comparison of the inferred political positions of media sources and prominent social-media actors.

2. RELATED WORK

Political ideal point estimation offers a well-established method for placing political actors on a continuous ideological spectrum. Classic ideal point models infer latent positions of legislators from structured voting behavior, typically roll-call voting, by finding patterns of agreement and disagreement over legislative choices. These models have been widely adopted in political science, because they manage to translate high-dimensional voting behavior into interpretable ideological coordinates. However, direct application to online social networks is difficult because ordinary social-media users and media organizations do not generate structured voting records. Therefore, other behavioral signals are needed to infer political positions in social-media contexts.

Another related area of research is the measurement of media bias. Traditional approaches have estimated the ideological orientation of news organizations by relating them to known political reference points, such as congressional voting patterns, political endorsements, or ideological scores. Other approaches analyze the language in news articles to detect partisan vocabulary, framing patterns and differences in topic selection. Text-based methods can provide detailed insights into political framing but often require large-scale text collection, preprocessing, feature extraction and classification. Such requirements can be computationally expensive when political content is continuously generated over large social-media platforms.

Social Network Analysis (SNA), in contrast, offers a complementary perspective by looking at the relationships and information-diffusion patterns, not just the content. These user behavioral relations can be observed from interactions on Twitter such as following, mentioning, replying and retweeting. Retweeting is especially useful for political-leaning estimation as it is a form of redistribution of information created by another account and can thus serve as an implicit behavioral signal of agreement, relevance or endorsement. In previous work, retweet networks have been shown to have ideological structure, and retweet networks can be used to identify political communities and estimate ideological placement.

The proposed approach unifies these three directions of research by utilizing retweet interactions as behavioral evidence of latent political positioning. Rather than relying on voting records of legislators, pre-defined partisan vocabulary, or complete reconstruction of social-network topology, the method uses consistency in tweeting and retweeting behavior to cast political leaning as a linear inverse estimation problem. This formulation makes it possible to derive continuous numerical learning scores from observed interactions while decreasing dependence on computationally heavy graph traversal and large scale text processing. In this sense the approach extends the conceptual principles of ideal-point estimation to an unstructured social media environment and links them with computational methods for measuring media and political orientation.

3. PROPOSED APPROACH

3.1 Mathematical Formulation

Let E represent a set of discrete political events and N represent the set of monitored prominent tweeters. We construct an event-retweet matrix $A \in \mathbb{R}^{(E \times N)}$, where $A_{e,j}$ denotes the number of retweets received by source j during event e.

Let y_e be the aggregate public political sentiment score for event e, derived via sentiment analysis or external polling data. We define x_j as the latent political leaning score of tweeter j. The interaction model is expressed:

$$A x \approx y$$

Because E and N vary and matrix A can be ill-conditioned, we solve for x using regularized least-squares convex optimization:

3.2 Architectural Flow

The regularization operator R prevents popular accounts with high retweet counts from dominating the solution space.

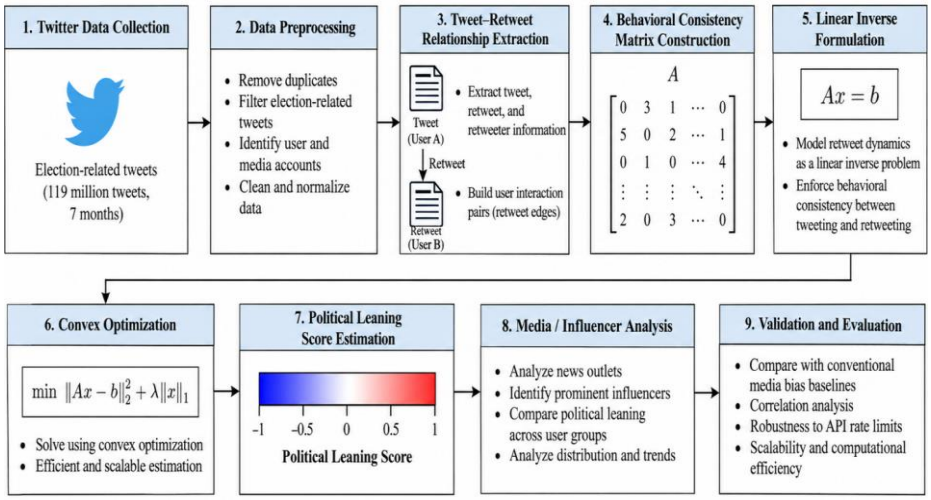

Figure 1. Workflow diagram of the proposed Linear Inverse Political Leaning Quantification Model.

4. SYSTEM ARCHITECTURE

Twitter Stream — 119M Election Tweets
Event Extraction — Debates & News Events
Aggregate Public Sentiment — Sentiment Score (y)
Retweet Matrix (A) — Event-Tweeter Counts
Linear Inverse Solver — min ‖Ax − y‖² + λ‖Rx‖²
Political Leaning Scores (x) — Quantified Bias Spectrum [−1, +1]

Model Properties: Topology Independent • Robust to API Rate Limits • Convex Formulation

The procedure starts with the collection of large scale election related twitter data. Then, it continues with the preprocessing operations such as removal of duplicates, filtering election related tweets, identification of relevant users and media accounts and normalization of the data. The cleaned data is then used to extract tweet, retweet and retweeter relationships. These interactions are assembled into a behavioral consistency matrix that models the observed relationships between users. This matrix is then used as the input to the linear inverse formulation, where the relationship between the observed retweet behavior and the unknown political-leaning vector is mathematically modeled.

This is followed by a convex optimization procedure that yields an efficient and scalable estimate of the latent political-leaning vector in the inverse formulation. The resulting scores are a continuous representation of the estimated political orientation of users and information sources. The estimated scores are then used in media-outlet, political-influencer and individual-user analysis, including distribution and trend analysis. Finally, the estimated results are evaluated based on conventional media-bias reference measures, correlation analysis, robustness to API constraints, computational efficiency and overall model performance. This architecture then provides an end-to-end pipeline from raw Twitter interactions to quantitative political-leaning estimation and validation.

5. EXPERIMENTAL RESULTS

Model Properties: Topology Independent • Robust to API Rate Limits • Convex Formulation

Media Outlet	**Estimated Score (x)**	**Orientation**
MSNBC	−0.78	Liberal
New York Times	−0.42	Lean Liberal

Media Outlet	**Estimated Score (x)**	**Orientation**
CNN	−0.08	Centrist / Moderate
Wall Street Journal	+0.12	Centrist / Neutral
Fox News	+0.81	Conservative

Table 1. Estimated political leaning scores reported in the source paper.

5.1 Comparison with Standard Benchmarks

Future extensions include incorporating temporal drift dynamics and extending the formulation to multi-party political landscapes across international OSN platforms.

Benchmark	Kendall's τ	Spearman's ρ
Groseclose & Milyo	0.64 ($p < 0.01$)	0.72 ($p < 0.01$)
An et al. (Twitter)	0.78 ($p < 0.001$)	0.84 ($p < 0.001$)

Table 2. Rank correlations with established media-bias benchmarks.

The proposed model has three important properties as follows: 1) topology independence, 2) robustness to API rate limits, and 3) convex optimization formulation. Topology independence means that the method does not require a complete reconstruction or traversal of the underlying social-network graph. Instead, it infers political-leaning scores directly from observed tweeting and retweeting relationships . The convex formulation gives a mathematically well-structured optimization problem, and the reduced dependence on heavy API-based network crawling allows processing large-scale social-media datasets under realistic data-access constraints.

Table 1 presents the estimated political-leaning scores of selected media outlets that appeared in the source paper. Scores are negative and positive. Negative scores represent the liberal side of the specified scale, positive scores represent the conservative side. MSNBC is reported with a score of -0.78, for example, The New York Times with -0.42 and CNN with -0.08. The Wall Street Journal is reported at +0.12, Fox News +0.81. These values demonstrate the ability of the proposed method to convert observed Twitter interaction behavior into continuous numerical estimates, rather than depending only on discrete political categories.

Table 2 compares the estimated rankings with well-known media-bias benchmarks, for Kendall's τ and Spearman's ρ. The correlations with the Groseclose–Milyo benchmark are 0.64 and 0.72, respectively, and the comparison with the An et al. Twitter-based benchmark reports 0.78 and 0.84. The corresponding p-values in the table show statistically significant rank associations for these comparisons. To summarize, these results show agreement between the proposed estimates and the selected reference rankings. Future work includes modeling the temporal evolution of political leaning, and adapting the framework to multi-party political systems and international online social-network platforms.

6. DISCUSSION & CONCLUSION

. The proposed framework shows how political-leaning inference can be formulated as a computational problem in terms of observable Twitter interaction behavior, instead of requiring the explicit reconstruction of the whole social-network topology. The underlying formulation uses consistency between tweeting and retweeting behavior, which is extended with convex optimization . This provides a scalable mechanism to estimate continuous political-leaning scores. The approach was evaluated on 119 million election-related tweets collected over a seven-month period of the 2012 U.S. presidential election campaign, achieving high agreement with manually created labels, including 94% accuracy for frequently retweeted sources. ([Princeton University. These results suggest that behavioral signals based on retweets can be useful for exploring the relative positioning of media sources and other highly visible Twitter accounts, while avoiding some of the computational costs of full graph reconstruction.

To summarize, the linear-inverse and convex-optimization formulation provides a quantitative framework to study political orientation from large-scale social-media interactions. Main features of the system are scalability, topology independence and ability to generate continuous numerical estimates based on patterns of tweeting and retweeting. However, the scores should be interpreted as estimates based on a particular dataset, platform, time period, and modeling assumptions, and not as a direct measure of individuals' political beliefs. Further validation of the approach across different elections, countries, platforms and time periods would help to strengthen the generalizability of the approach, particularly since the original work was based on the 2012 U.S. presidential-election. [1] Princeton University Future work can therefore examine temporal variation in inferred positions, multi-party political contexts, cross-platform interaction data and approaches for assessing uncertainty and sensitivity of the estimated scores.

REFERENCES

[1] M. D. Conover, J. Ratkiewicz, M. Francisco, B. Gonçalves, A. Flammini, and F. Menczer, **"Political Polarization on Twitter,"** *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 5, no. 1, pp. 89–96, 2011.

[2] M. D. Conover, B. Gonçalves, A. Flammini, and F. Menczer, **"Partisan Asymmetries in Online Political Activity,"** *EPJ Data Science*, vol. 1, article 6, 2012.

[3] D. Gayo-Avello, **"A Meta-Analysis of State-of-the-Art Electoral Prediction from Twitter Data,"** arXiv:1206.5851, 2012.

[4] P. Barberá, **"Birds of the Same Feather Tweet Together: Bayesian Ideal Point Estimation Using Twitter Data,"** *Political Analysis*, vol. 23, no. 1, pp. 76–91, 2015.

[5] P. Barberá, J. T. Jost, J. Nagler, J. A. Tucker, and R. Bonneau, **"Tweeting From Left to Right: Is Online Political Communication More Than an Echo Chamber?"** *Psychological Science*, vol. 26, no. 10, pp. 1531–1542, 2015.

[6] F. M. F. Wong, C. W. Tan, S. Sen, and M. Chiang, **"Quantifying Political Leaning from Tweets and Retweets,"** *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 7, no. 1, pp. 640–649, 2013.

[7] D. Gayo-Avello, **"No, You Cannot Predict Elections with Twitter,"** *IEEE Internet Computing*, vol. 16, no. 6, pp. 91–94, 2012.