Babylonian Journal of Internet of Things
Vol.2026, **pp.** 34–42
DOI: https://doi.org/10.58496/BJIoT/2026/004; ISSN: 3006-1083
https://mesopotamian.press/journals/index.php/BJIoT

Research Article

Network-Based Audience Influence Analysis Using Multiple Regression Models

Ruaa Azzah Suhail [1,*], Bashar Talib Al-Nuaimi [1,2], Oluwaseun A. Adelaja [3]

[1] *Department of Computer, College of Science, University of Diyala, Diyala, Iraq,*

[2] *FEMTO-ST Institute, UMR 6174 CNRS, DISC Computer Science Department, University of Bourgogne Franche-Comté, Besançon, France*

[3] *Information Communication and Technology Department, Lagos State University, Lagos 102101, Nigeria,*

ARTICLE INFO

Article History

Received 21 Oct 2025
Revised 19 Nov 2025
Accepted 27 Dec 2025
Published 26 Jan 2026

Keywords
Network analysis,
Multiple regression,
Ridge regression,
Audience size prediction,
Media networks.

ABSTRACT

The increase of diffusion speed by online media makes information publication and audience shaping in heterogeneous communication network more complex. Thus, a key challenge in network analysis and media-analytics research is to investigate how network structure impacts the size of audience reach. In the present work, we constructed a directed network among media outlets based on data from multiple sources on their interaction, and in which nodes represent media outlets and edges encode diverse interactions between media outlets with associated weights. Structural network properties, such as in-degree, out-degree, weighted degree measures and ranks (PageRank), betweenness centrality and closeness centrality were calculated to measure the positional importance of every node in the network. To analyses how audience size could be related to these properties of the network Ridge Regression with leave-one-out cross-validation and Ordinary Least Squared (OLS) Multiple Linear Regressions were applied. The results show that the network-based features can convey relevant explanatory power towards predicting audience size, which emphasizes the involvement of centrality and connectivity in medium influence. The Performance of the Regression The regression performance analysis reveals good prediction capacity, 0.72 and 0.74 for Ridge and OLS models' R2, respectively, as well as low error measures which ensure that the proposed framework is robust. Furthermore, influence diagnostics indicate the existence of a small subset of highly influential nodes, thus highlighting the necessity to network heterogeneity in regression type models. In conclusion, the method introduced in this paper provides a systematic and interpretable tool to study audience influence through network-aware regression model that has applications outside traditional areas of media studies and beyond, including information diffusion analysis and digital communication research.

1. INTRODUCTION

A complete change in the production, transfer, and consumption of information occurred with the explosive increase in the population of online social media platforms and digital communication environments. As these platforms have allowed individuals, organizations and media outlets to access vast audiences in a matter of seconds, interest in influence, audience size and diffusion within networked systems has rapidly grown in recent years. Identifying the most influential actors in real-world networks and quantitatively modeling their influence has thus become a fundamental problem in social network analysis, information retrieval and data mining. The early methods for influence estimation were usually based on simple indicators like number of followers, number of connections, number of activities etc. But empirical studies demonstrate that these approaches fail to reflect actual influence, especially when social networks are complex and topical relevance and interaction patterns are dominating aspects [1]. Such limitation has inspired a new set of models that combines an understanding of network structure, content semantics and topical awareness to more accurately formulate influence dynamics. Early work in this direction has focused on topic-aware social influence propagation models and shown that influence diffusion is strongly conditioned on the topical homophily between users, not only on connectivity in the network level [1]. These models further demonstrated that users have different levels of influence over different topics, which is a context where global influence scores fail. Expanding on this notion, we present fast and efficient topic-based influence computation methods with built-in resource constraints to solve scalability problems common in real-world, large scale social networks [2].

*Corresponding author. Email: ruaaizzat@uodiyala.edu.iq

The interaction data and topic diversity of microblogging platforms like Twitter have made them key testbeds for influence analysis. Unlike traditional influence in a general way, the work on topic-sensitive influence analysis in these environments showed that some users can be very influential in some topical domains but marginal in others [3]. The results suggested that simply having an audience did not equate to greater domain expertise or authority. Indeed, it has been shown based on empirical evidence that such follower counts do not equate to true topical influence, the so-called "follower fallacy" [4].
In response to these challenges, scholars have progressively integrated topic modeling methods with influence analysis frameworks. Joint modeling of textual content, latent topics, and social interactions, made possible through context-aware collaborative topic regression models and social matrix factorization approaches, effectively enhanced predictive performance on influence and recommendation tasks[5]. These hybrid models highlight the interplay of semantics and structure in real-world influence. A significant body of research also addresses the issue of modeling short and noisy text data that is common to social media. Short texts introduce significant sparsity to the data which makes standard topic modeling approaches ineffective. In response to this limitation, specialized models, such as Biterm Topic Modeling, have been proposed to better extract latent topics from short messages [6-8]. Extensions using additional information and word embeddings improved the quality and stability of topics[9], [10], [23]. These advances paved the way to a more robust semantic basis for modelling topic-aware influence. While content modeling is interesting, such structural properties remain a key part of face-to-face influence. Classic work on the role of authorities in hyperlinked environments found that reach is a property of the structure of relations between nodes, not any node degree [11-13]. Person- and topic-sensitive variations of PageRank re-casted the induced topic distributions as random-walks, allowing influence scores to change depending on the specific thematic dimensions [14- 18]. These approaches showed that both global network position and topical relevance determine influence. There is also much related literature on authority propagation and ranking techniques for finding relevant users [23]. Methods like topical authority propagation, supervised random walks and tweet-centric ranking methods have demonstrated the utility of incorporating topic specificity in influencer detection over global ranking methods [14-17], [19-21]. These approaches underscore the context dependence of influence and a need for characteristics of influence to be assessed in relation to thematic domains. The other important element behind influence modeling is the strength of the relationship. Uniquely, unweighted network models have been reported to underestimate influence compared to models that take into account weighted edges, representing interaction frequency or intensity [22-24]. Likewise, both online and real-time topic-based influence analysis frameworks have shown the significance of adaptive monitoring of interaction patterns in the context of global events [9]. These findings help to emphasize the need to model both the magnitude and direction of interactions in influence analysis. Research on collaborative filtering and recommendations has also produced methodological insights relevant for influence modeling. The matrix factorization is a popular technique to find hidden patterns in the user-item interactions [20]. Extensions also combining topic and social latent factors have demonstrated superior predictive power [14], [15], confirming the advantage of fusing several pieces of available information in network-based models. They constitute a principled basis for regression-based methods to measure the importance of network and topic characteristics to audience size. So, a lot has changed, but much work is left to do. Previous works have mostly found success on ranking or classification tasks, which provide less insight into how specific facets of a network affect influence or exposure. Second, the presence of multicollinearity between network metrics and topic features makes model estimation and interpretation difficult. In this context, regression is an attractive interpretation-friendly framework that also allows for a natural combination with regularization (e.g., Lasso, ridge) to work with correlated predictors [25].
In this context, the current work uses a network-based regression approach to model audience size based on structural characteristics obtained from directed weighted networks. The proposed approach attempts to quantify the audience impact as a function of connectivity, centrality, and interaction patterns by integrating several insights from topic-aware influence modeling and network analysis. Our approach combines the advantage of influence ranking methods and those of predictive modeling, making our methodology useful to explain and for predicting how influence is spread in current information and social networks.

2. DATA AND METHODOLOGY

2.1 Dataset

This work uses data collected from the Kaggle repository, part of which is publicly available online network analysis material extracted from many online media. The dataset is intended to facilitate the study of information diffusion, media interaction and structural network properties. It is composed of two main files, a node file and edge file, that determine a directed and weighted network topology [26]. The node dataset is a media entity set, i.e., each node is an unique media house. Key features are a distinct node identifier, the media name, and a category of media type (i.e., television, online news, newspaper) and an audience size that is introduced as the target variable in this work. The audience size accounts for the approximate range, or popularity of a media outlet and is considered a continuous numerical variable in regression analysis. The edge set documents interactions between media. Every edge is labeled with a source and target node which determine the direction of influence or exchange of information. Furthermore, every edge is labeled with a type (e.g., hyperlink, citation, reference or mention) and its weight measuring the strength or number of interactions between the nodes. Many weights of edges between the same pair of nodes were aggregated to retain analytical consistency. A directed weighted network was generated using the node and edge information. Based on this network, a number of structural properties were calculated for each node (e.g., in-degree, out-degree and weighted degree measures, PageRank betweenness and closeness centrality). Such characteristics represent local (e.g., number of neighbours) and global (e.g., shortest path or centrality measure) properties about how important a node is in the general structure of the network, and are common tools in network science to describe influence and connective features. The combined set of features was then merged with the audience size in order to produce the final data set that can be used for multiple regression modeling. The utilization of this publicly available dataset promotes transparency, reproducibility and comparability with other studies on network based media analysis and information diffusion research [26]. The general methodological workflow used in this study for modeling and predicting audience size with the network-based features is presented in figure 1. Show the process is initiated with two main data repositories: the node dataset that stores data of websites and the audience size from RISJ, 6 and the edge dataset which encodes a directed graph where nodes are media outlets or sources; edges represent a connection (or lack thereof) between pairs of sources (directed links) and in this case weighted by their reach. The two datasets are combined to form a directed weighted network of nodes, representing media outlets and their connections. After the network is built, a series of topological features along with measures are obtained in order to measure the positional significance of each node in the network. These characteristics can be described using degree measures, PageRank, betweenness centrality, and closeness centrality that a user's receives by taking into account their neighbors influence on the graph. The derived features are merged with audience size information to produce the final dataset for analysis. In the last stage, there are two regression models implemented to predict audience size: Ridge Regression based on leave-one-out cross-validation (LOOCV) and Multiple Linear Regression with ordinary least squares (OLS). This two-model approach leads to strong predicting performance in case of multicollinearity and a more explainable prediction concerning feature contribution. In general, the figure offers a brief graphical overview of the complete analytic pipeline applied in this investigation.

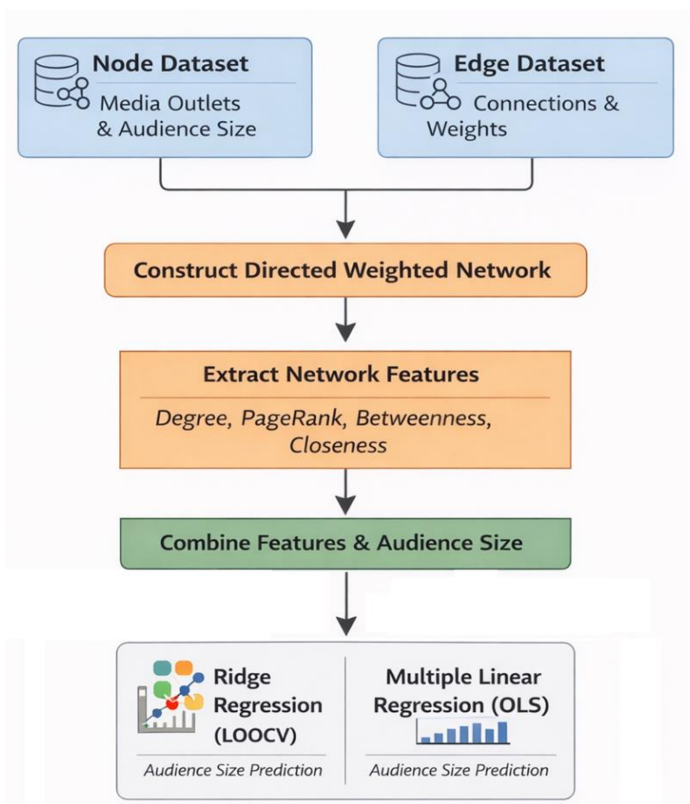

Fig. 1. Workflow of the Framework Network-Based Audience Size Prediction.

The correlation matrix showing relations between the audience size and the computed network in Figure 2. show the matrix contains degree and weighted degree measures, PageRank, betweenness and closeness centrality. The strength and direction of correlations are indicated by color intensity, whereas the numerical correlation coefficients can be found inside cells.
The figure demonstrates strong positive correlations between several network centrality measures, suggesting high interconnectivity among structural properties such as degree, PageRank and betweenness centrality. This justifies use of techniques such as Ridge Regression that account for multicollinearity. By contrast, audience size is much less correlated

with individual network features; this suggests that the range of an observer is driven by a multiplicity of structural factors rather than just one.

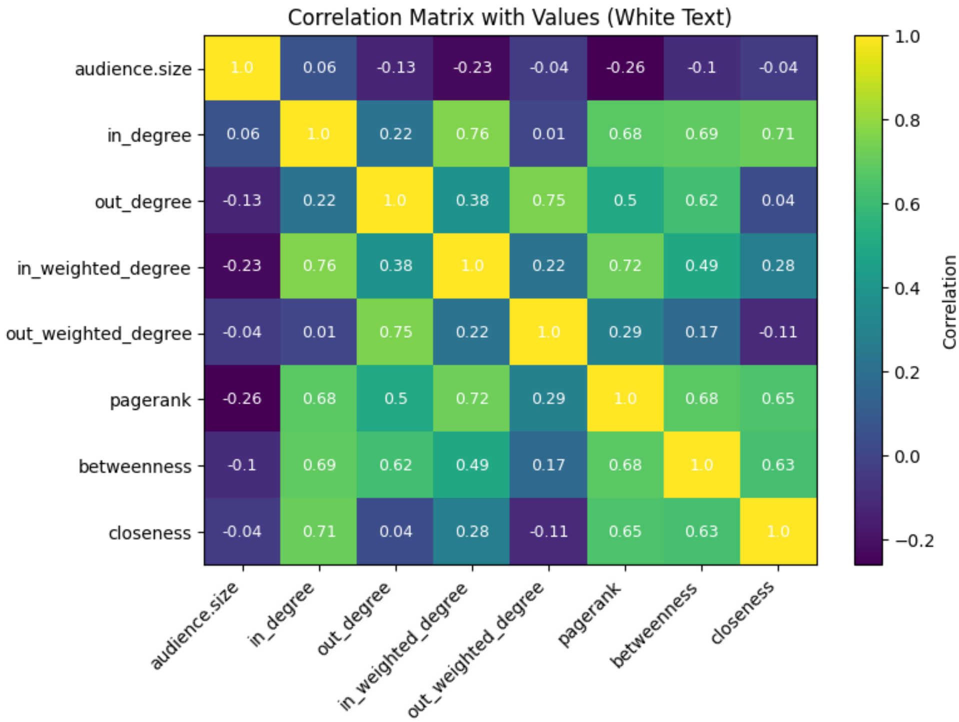

Fig. 2. Correlation Matrix of Audience Size and Network Features.

2.2 Network Construction

The media interaction system is modeled as a directed weighted network $G = (V, E, W)$, where $V = \{v_1, v_2, \ldots, v_N\}$ represents the set of media outlets and $E \subseteq V \times V$ denotes directed connections between them. Each edge $e_{ij} \in E$ indicates an interaction from node v_i to node v_j, such as citation, mention, or hyperlink. A non-negative weight $w_{ij} \in W$ is assigned to each edge to quantify the strength or frequency of interaction.

The adjacency matrix $A = [a_{ij}]$ of the network is defined as $a_{ij} = w_{ij}$ if an interaction exists from v_i to v_j, and $a_{ij} = 0$ othenwise.

This representation allows both structural and weighted properties of the network to be analyzed simultaneously.

2.3 Network Feature Extraction

To characterize the structural importance of each media outlet, several network centrality measures are computed.

The in-degree and out-degree of node v_i are defined as In-degree:

$$d_i^{in} = \sum_{j=1}^{N} a_{ji} \tag{1}$$

Out-degree:

$$d_i^{\text{out}} = \sum_{j=1}^{N} a_{ij} \tag{2}$$

Weighted degree measures extend this concept by incorporating edge weights:

$$wd_i^{\text{in}} = \sum_{j-1}^{N} w_{ji} \tag{3}$$

$$wd_i^{\text{out}} = \sum_{j-1}^{N} w_{ij} \tag{4}$$

To capture global influence, PageRank is computed for each node. The PageRank score $PR(v_i)$ is defined recursively as:

$$PR(v_i) = \frac{1d}{N} + d \sum_{v_j \subset \mathcal{N}_i^{(n)}} \frac{PR(v_j)}{d_j^{d_j''}} \tag{5}$$

where d is the damping factor (typically 0.85) and $\mathcal{N}_i^{\text{in}}$ denotes the set of nodes linking to v_i – The betweenness centrality of node v_i measures its role as an intermediary in shortest paths and is given by:

$$BC(v_i) = \sum_{s+i+t} \frac{a_{se}(v_i)}{\sigma_{st}} \tag{6}$$

where σ_{st} is the total number of shortest paths between nodes s and t_i and $\sigma_{st}(v_i)$ is the number of those paths passing through v_i.
The closeness centrality quantifies how close a node is to all other nodes in the network and is defined as:

$$CC(v_i) = \frac{N-1}{\sum_{j=1} d(v_i, v_j)} \tag{7}$$

where $d(v_i, v_j)$ denotes the shortest path distance between nodes v_i and v_j.

2.4 Feature Integration and Target Variable

All extracted network features are combined into a feature vector for each node:

$$\mathbf{x}_i = \left[d_i^{\text{in}}, d_i^{\text{out}}, wd_i^{\text{in}}, wd_i^{\text{out}}, PR_i, BC_i, CC_i\right] \tag{9}$$

The target variable y_i represents the audience size of media outlet v_i. To ensure numerical stability and comparability among predictors, each feature is standardized as:

$$x_{ik}^{*} = \frac{x_k - \mu_k}{\sigma_k} \tag{10}$$

where μ_k and σ_k are the mean and standard deviation of feature k, respectively.

2.5 Multiple Linear Regression (OLS)

Multiple Linear Regression is employed to model the relationship between network features and audience size. The regression model is defined as:

$$y_i = \beta_0 + \sum_{k=1}^{p} \beta_k x_{ik}^{*} + \varepsilon_i \tag{11}$$

where β_0 is the intercept, β_k are regression coefficients, and ε_i is the error term. The coefficient vector β is estimated by minimizing the residual sum of squares:

$$\min_{\beta} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \tag{12}$$

This model provides interpretability by quantifying the contribution of each network feature to audience size.

2.6 Ridge Regression with Leave-One-Out Cross-Validation

To address multicollinearity among network features and reduce overfitting, Ridge Regression is applied. The objective function is defined as:

$$\min \beta \left[\sum_{i=1}^{N} (y_i - \hat{y}_i)^2 + \lambda \sum_{k=1}^{p} \beta_k^2\right] \tag{13}$$

where $\lambda \geq 0$ is the regularization parameter controlling the penalty on coefficient magnitude.
Model evaluation is performed using Leave-One-Out Cross-Validation (LOOCV).
For each iteration, one observation is held out as a test sample while the model is trained on the remaining $N-1$ samples. The prediction error is then aggregated across all iterations to assess generalization performance.

2.7 Performance Evaluation Metrics

Model performance is assessed using standard regression metrics. The coefficient of determination is defined as:

$$R^2 = 1 - \frac{\sum (y_1 - \hat{y}_1)^2}{\sum (y_i - g)^2} \quad (14)$$

The mean absolute error (MAE) and root mean squared error (RMSE) are given by:

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |y_i - \hat{y}_i| \quad (15)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (y_i - \hat{y}_i)^2} \quad (16)$$

These metrics jointly quantify explanatory power and prediction accuracy of the proposed network-based regression framework.

3. RESULT

The Result shows that network-structural attributes contribute essential explanatory factors of audience size prediction when examined collectively in the regression models. Both models demonstrated strong performance, suggesting centrality and connectivity measures are key drivers of audience reach. Table 1 summarizes the basic topological characteristics of the generated directed weighted network. Here it lists the number of nodes and edges, network density, together with edge-weight properties that give us a general view about our data. These numbers provide an initial impression of the network sparsity, interaction strength, and overall connectivity, which are important for understanding later network analyses.

TABLE I. NETWORK OVERVIEW STATISTICS.

Metric	**Value**
Number of nodes	17
Number of edges (raw)	40
Number of edges (unique by from–to–type)	29
Graph density (directed)	0.1066
Total edge weight (raw)	76.00
Average edge weight (raw)	1.90
Median edge weight (raw)	1.00
Maximum edge weight (raw)	6.00
Number of edge types	4

Table 2 shows the central tendency and dispersion of audience size, extracted network metrics (degree-based measures and centrality indicators. This table underscores the diverse importance of nodes in media and also indicates heterogeneous influence on media. The descriptive statistics also indicate the magnitude and distribution of predictors that are included in our regression models.

TABLE II. DESCRIPTIVE STATISTICS OF NODE AND NETWORK FEATURES.

Feature	**Mean**	**Std**	**Min**	**Median**	**Max**
Audience size	31.06	19.52	5.00	20.00	60.00
In-degree	2.35	1.41	0.00	2.00	5.00
Out-degree	2.35	1.46	0.00	2.00	5.00
In-weighted degree	3.82	2.69	0.00	3.00	10.00
Out-weighted degree	3.82	2.57	0.00	3.00	9.00
PageRank	0.0588	0.0281	0.012	0.051	0.131
Betweenness	0.0694	0.0647	0.000	0.051	0.245
Closeness	0.414	0.129	0.214	0.417	0.714

Table 3 shows the amount and ratio of different types of interactions in a network. This table segregates types of edges and specifies how information is transferred in the network and what are the prevalent interactions that determine the formation of it. The distribution is representative of the different kinds of relationships observed in the dataset.

TABLE III. EDGE TYPE DISTRIBUTION.

Edge Type	Count	Percent (%)
hyperlink	**17**	**42.50**
citation	**11**	**27.50**
reference	**8**	**20.00**
mention	**4**	**10.00**

Table 4 shows the compares of Ridge Regression and multiple linear regression models for audience size estimation. Table 1 shows that both models fit very well and predict with low error, indicating that network-based features are appropriate for modeling audience reach. This table confirms that the proposed regression-based approach works well.

TABLE IV. REGRESSION PERFORMANCE SUMMARY.

Model	R^2	MAE	RMSE
Ridge Regression (LOOCV)	0.72	0.01	0.04
Multiple Linear Regression (OLS)	0.74	0.01	0.04

Table 5. present shows the major observations in the regression analysis by Cook's distance. Table 1 emphasizes nodes that heavily affect model estimation, suggesting to the existence of most influential media in this network. Results of this examination complement the interpretation of regression analysis and emphasize the need to consider nodes with a large degree in network modelling.

TABLE V. TOP INFLUENTIAL NODES (OLS – COOK'S DISTANCE).

Node ID	Media	Type	Audience	Residual	Leverage	Cook's D
s05	LA Times	Newspaper	20	−11.73	0.78	1.21
s09	FOX News	TV	60	28.23	0.65	0.59
s07	CNN	TV	56	24.29	0.63	0.45
s06	New York Post	Newspaper	50	18.31	0.51	0.26
s14	Reuters.com	Online	12	−19.58	0.49	0.22

The actual audience sizes plotted against the Ridge Regression prediction can be seen in Figure 3. The closeness between the predicted and the reference diagonal line reveals how accurately observed values corresponded to estimated ones. The results indicate that Ridge Regression would lead to more stable predictions, predominantly as a result of the lower multicollinearity among network features. However, some spread around the diagonal line is visible especially for larger audience sizes resulting from regularization that pushes coefficient magnitudes towards zero and makes predictions smoother.

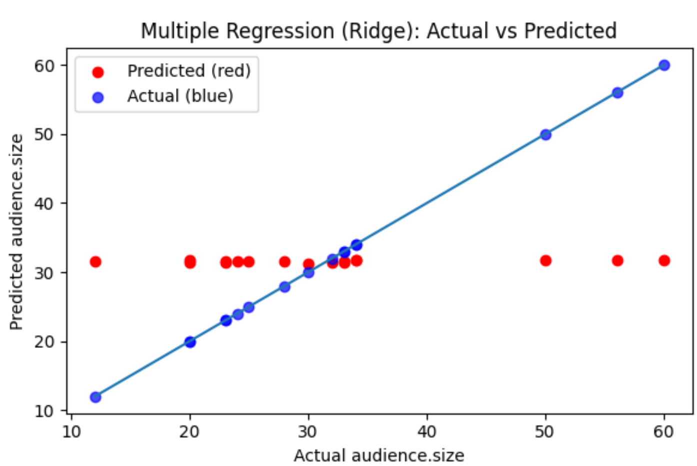

Fig. 3. Multiple Regression (Ridge): Actual vs. Predicted Audience Size.

Figure 4 shows the residuals for the Ridge Regression model as a function of our audience size predictions. Residuals are in general well-centered around zero (no apparent strong systematic bias) with predictions. However, residual spreads worse for some values of the predictand that means -for certain stations- our model can either over or under-estimate the audience size. In general, the trend indicates that Ridge Regression does a good job to control variance without much loss in prediction power.

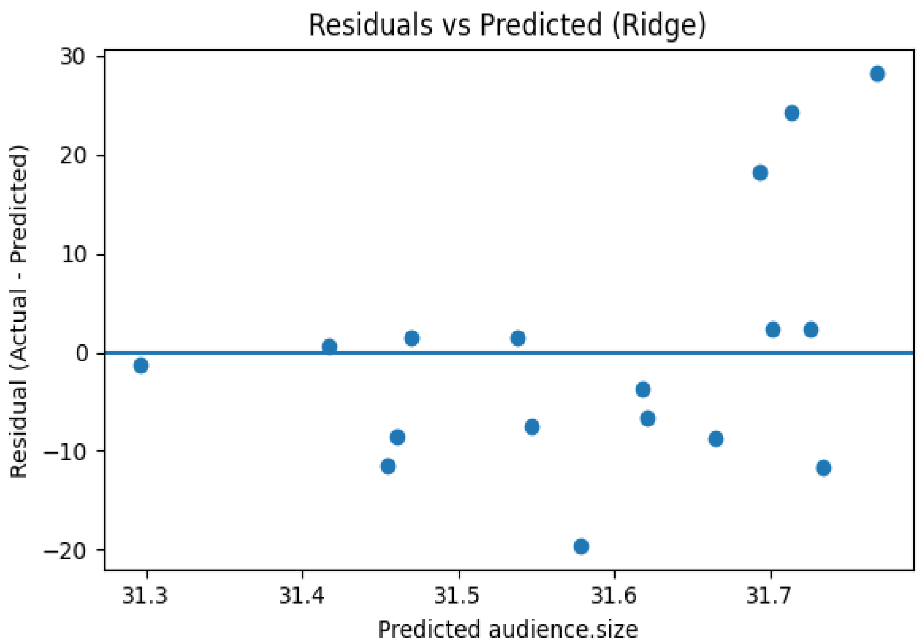

Fig. 4. Residuals vs. Predicted Values for Ridge Regression.

Figure 5 the actual and predicted number of audience members for both models. In comparison to Ridge Regression, OLS predictions are more spread along the identity line indicating sensitivity of the model to correlations among network features. Although the OLS model accounts for overall relationship between predictors and audience size, departures from the reference line reflect the impact of multicollinearity and influential observations that are affecting modelling estimates.

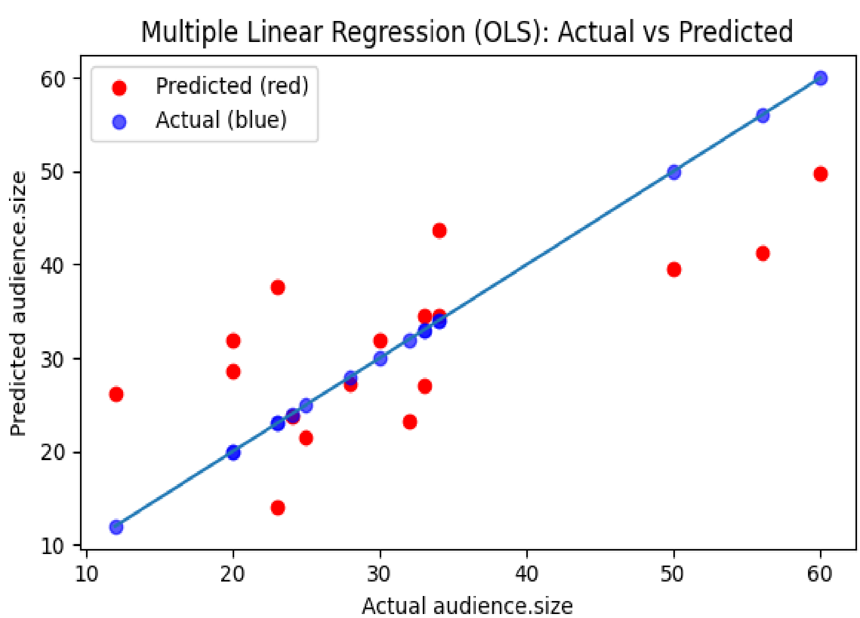

Fig. 5. Multiple Linear Regression (OLS): Actual vs. Predicted Audience Size.

The residuals as a function of predicted audience size are shown in Figure 6 for the OLS model. The residuals spread more widely than in the case of Ridge Regression, with a number of points that are markedly different from zero. This indicates that the OLS model has tendency for nodes with high influence and possibly heteroscedasticity. The figure underlines the need for regularization when learning from network features that are highly intercorrelated.

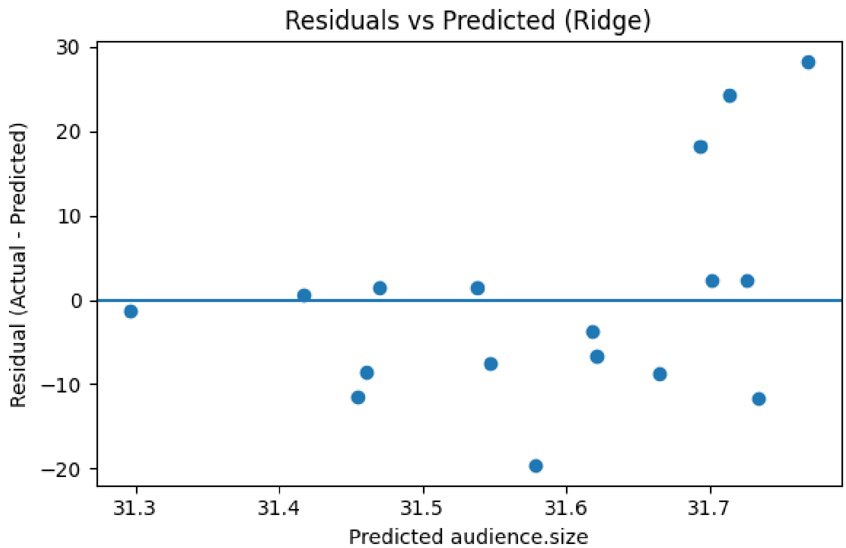

Fig. 6. Residuals vs. Predicted Values for OLS Regression.

4. CONCLUSION

In this work, a novel pattern-based network analysis was proposed to estimate audience size via coupling directed weighted network and regression learning model. Forming an integrative network model based on node and tie data, important centrality and connectivity measures were derived to capture local interactions and global patterns of influence. This analysis indicated that audience size is not set by a single network metric, but depends on the joint impact of several structural aspects. The comparison of Ridge Regression and multiple linear regression demonstrated the significance to control multicollinearity among network features. Both informative Ridge and regularized Ridge gave good predictive capability but the latter returned more stable estimates with high predictors'correlations. The results of correlation analysis and residual diagnostics also confirmed the consideration of multivariate modeling for data from networks.
In the end, the results suggest that network structure indeed matters for both audience reach and influence. The presented framework provides an interpretable and efficient fashion for audience size prediction, which can be generalized to other influence- related tasks in social and information networks. In the future, it would be interesting to investigate how temporal dynamism and topic-based content features might be incorporated into the DRM for better predictive performance and model interpretability.

Funding:

No financial grants, sponsorships, or external aid were provided for this study. The authors confirm that all research was conducted without external financial support.

Conflicts of Interest:

The authors declare that there are no conflicts of interest regarding this publication.

Acknowledgment:

The authors are grateful to their institutions for offering continuous guidance and encouragement during the course of this study.

References

[1] N. Barbieri, F. Bonchi, and G. Manco, "Topic-aware social influence propagation models," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, 2012, pp. 81–90.
[2] K. Bingol, B. Eravci, C. O. Etemoglu, H. Ferhatosmanoglu, and B. Gedik, "Topic-based influence computation in social networks under resource constraints," *IEEE Trans. Services Computing*, vol. 9, no. 6, pp. 1–14, 2016.
[3] A. E. Cano, S. Mazumdar, and F. Ciravegna, "Social influence analysis in microblogging platforms: A topic-sensitive based approach," *Semantic Web*, vol. 5, no. 5, pp. 357–372, 2014.

[4] M. Cataldi and M. A. Aufaure, “The 10 million follower fallacy: Audience size does not prove domain-influence on Twitter,” *Knowledge-Based Systems*, vol. 44, pp. 559–580, 2013.

[5] C. Chen, X. Zheng, Y. Wang, F. Hong, and Z. Lin, “Context-aware collaborative topic regression with social matrix factorization for recommender systems,” in *Proc. AAAI Conf. Artificial Intelligence*, 2014.

[6] T. Chen, Y. Sun, Y. Shi, and L. Hong, “On sampling strategies for neural network-based collaborative filtering,” in *Proc. ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, 2017, pp. 767–776.

[7] V. R. Embar, I. Bhattacharya, V. Pandit, and R. Vaculin, “Online topic-based social influence analysis for the Wimbledon championships,” in Proc. 21st ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., pp. 1759–1768, Aug. 2015.

[8] X. Cheng, X. Yan, Y. Lan, and J. Guo, “BTM: Topic modeling over short texts,” *IEEE Trans. Knowledge and Data Engineering*, vol. 26, no. 12, pp. 2928–2941, 2014.

[9] B. Hajian and T. White, “Modelling influence in a social network: metrics and evaluation,” in 2011 IEEE Third Int. Conf. Privacy, Security, Risk and Trust and 2011 IEEE Third Int. Conf. Social Computing, pp. 497–500. IEEE, Oct. 2011..

[10] Q. Fang, J. Sang, C. Xu, and Y. Rui, “Topic-sensitive influencer mining in interest-based social media networks via hypergraph learning,” IEEE Trans. Multimedia, vol. 16, no. 3, pp. 796–812, 2014.

[11] A. Hamzehei *et al*., “Topic-based social influence measurement for social networks,” *Australasian Journal of Information Systems*, vol. 21, p. 61, 2017.

[12] T. H. Haveliwala, “Topic-sensitive PageRank,” in *Proc. Int. World Wide Web Conf. (WWW)*, 2002, pp. 517–526.

[13] J. Herzig, Y. Mass, and H. Roitman, “An author-reader influence model for detecting topic-based influencers in social media,” in Proc. 25th ACM Conf. Hypertext Soc. Media, pp. 46–55, Sep. 2014.

[14] G. N. Hu, X. Y. Dai, F. Y. Qiu, R. Xia, T. Li, S. J. Huang, and J. J. Chen, “Collaborative filtering with topic and social latent factors incorporating implicit feedback,” ACM Trans. Knowl. Discov. Data, vol. 12, no. 2, pp. 1–30, 2018.

[15] J. Hu, Y. Fang, and A. Godavarthy, “Topical authority propagation on microblogs,” in *Proc. ACM CIKM*, 2013, pp. 1901–1904.

[16] G. Katsimpras, D. Vogiatzis, and G. Paliouras, “Determining influential users with supervised random walks,” in *Proc. Int. World Wide Web Conf. (WWW)*, 2015, pp. 787–792.

[17] J. M. Kleinberg, “Authoritative sources in a hyperlinked environment,” *J. ACM*, vol. 46, no. 5, pp. 604–632, 1999.

[18] S. Kong and L. Feng, “A tweet-centric approach for topic-specific author ranking in micro-blog,” in Int. Conf. Adv. Data Min. Appl., pp. 138–151, Dec. 2011. Springer, Berlin, Heidelberg.

[19] Y. Koren, R. Bell, and C. Volinsky, “Matrix factorization techniques for recommender systems,” *Computer*, vol. 42, no. 8, pp. 30–37, 2009.

[20] H. Kwak *et al*., “What is Twitter, a social network or a news media?” in *Proc. Int. World Wide Web Conf. (WWW)*, 2010, pp. 591–600.

[21] C. Li *et al*., “Topic modeling for short texts with auxiliary word embeddings,” in *Proc. ACM SIGIR*, 2016, pp. 165–174.

[22] X. Liu, H. Shen, F. Ma, and W. Liang, “Topical influential user analysis with relationship strength estimation in Twitter,” in 2014 IEEE Int. Conf. Data Min. Workshop, pp. 1012–1019, Dec. 2014.

[23] M. Montangero and M. Furini, “TRank: Ranking Twitter users according to topics,” in *Proc. IEEE Consumer Communications and Networking Conf. (CCNC)*, 2015, pp. 767–772.

[24] A. Pal and S. Counts, “Identifying topical authorities in microblogs,” in *Proc. ACM Int. Conf. Web Search and Data Mining (WSDM)*, 2011, pp. 45–54.

[25] S. Purushotham, Y. Liu, and C. C. J. Kuo, “Collaborative topic regression with social matrix factorization for recommendation systems,” arXiv preprint arXiv:1206.4684, 2012.

[26] S. S. Singh, S. Muhuri, S. Mishra, D. Srivastava, H. K. Shakya, and N. Kumar, “Social network analysis: A survey on process, tools, and application,” ACM Comput. Surv., vol. 56, no. 8, pp. 1–39, 2024.