Findings Report
Social Media Analytics

Exploring Social Media Impact & Community Engagement
*Answers to Part I & II*

Contributors
Meriem Mehri
Carlos Adrian Alarcon Delgado
Zhicheng Zhong

Winter 2024

Part 1: Finding Predictors of Influence

The analysis of the model's SHAP values and feature importances highlights that **the A_B_listed_ratio** and **A_B_followers_ratio** are the most telling predictors of influence on Twitter. Moreover, the significance of engagement-related variables, such as mentions and retweets received, provides deeper insight into the dynamics of influence across the network. These findings form the basis for a nuanced understanding of social media impact and its applications in strategic marketing.

Figure 1. SHAP Value Distribution for Feature Impact on Influencer Prediction Model
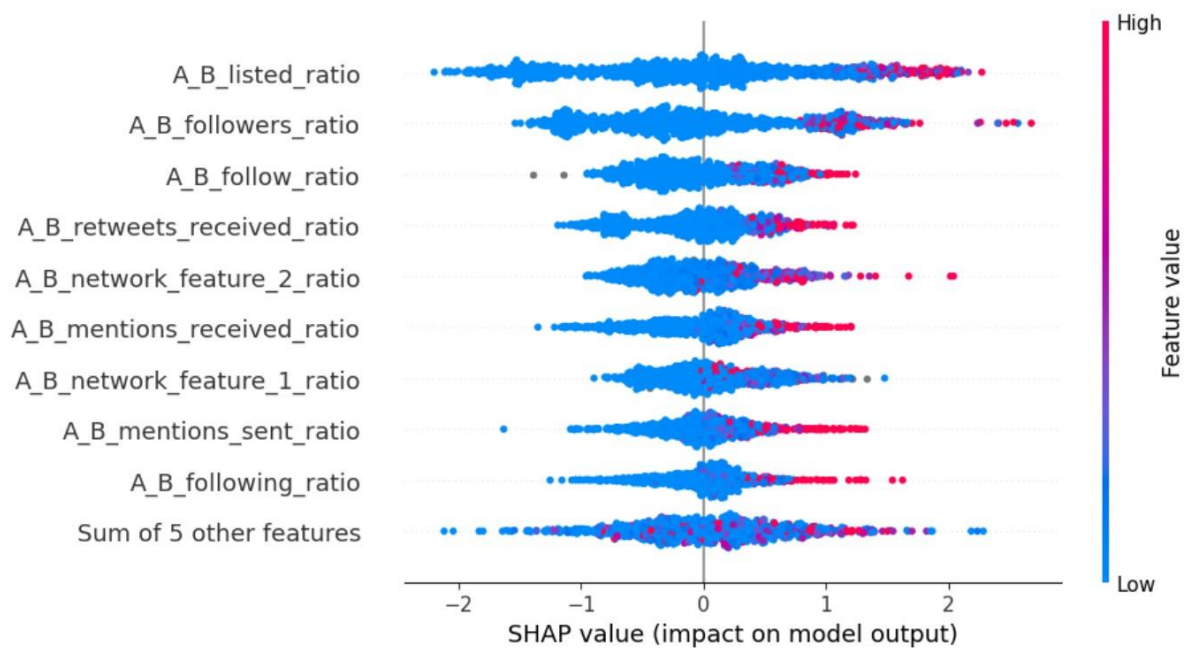
The influence of user interactions on the model's predictions underscores the value of engagement in social media influence. Mentions and retweets can be viewed as indicators of a user's active participation in the network and their ability to resonate with and mobilize their audience. This also aligns with recent trends in influencer marketing that prioritize engagement metrics over follower counts alone, acknowledging that a highly engaged following is often more valuable than a larger but passive audience.

The economic evaluation bridges the gap between these analytic insights and practical business applications. By leveraging the predictive power of the model to discern truly influential individuals, a retailer can maximize the impact of their marketing campaigns. The difference in expected revenue when using analytics versus not ($31,429 compared to $24,067) clearly demonstrates the model's utility in optimizing promotional strategies. It reflects a more targeted approach, where investment is directed towards those whose online behavior suggests they will yield a higher rate of return, as indicated by their interactions and standing within the network.

In conclusion, the model's interpretation provides compelling evidence for businesses to adopt sophisticated analytics in influencer marketing campaigns. The nuanced understanding of influence, as revealed by the **A_B_listed_ratio** and **A_B_followers_ratio**, along with interaction-based metrics, can significantly enhance the efficacy of social media marketing efforts. This approach not only elevates the effectiveness of individual campaigns but also offers a strategic edge in the competitive arena of digital marketing.

Insights Summary

- The most impactful predictors of Twitter influence in this model are the ratios of the listed count and follower count between two compared users.
- User interaction metrics, particularly those related to mentions and retweets received, are also critical and reflect the importance of engagement over mere follower counts.
- The economic evaluation suggests a substantial financial benefit from using analytics to identify influencers, with an increase in expected net profits when the model is used to select influencers for promotional activities.
- Our analytical approach shows an increase in expected net profit of approximately $31,429 versus $24,067 when analytics are not used, indicating a more efficient allocation of the promotional budget.

Part 2: Finding Influencers from Reddit

Our network graph was constructed to visualize and analyze the structure of interactions within the **r/sanitymemes** subreddit community on Reddit. Using NetworkX, a directed graph was created from preprocessed data, which included submissions and comments made by users. Each node in the graph represents an individual user, while directed edges signify the direction of communication, such as a comment on a submission or a reply to another comment, establishing a clear flow of information.

The visual network analysis revealed a densely connected core, with certain nodes appearing more central, indicating higher levels of engagement and potential influence within the community. This central core is surrounded by less connected peripheral nodes, illustrating the varying degrees of activity and influence among users.

To identify the top influencers, centrality metrics—specifically, degree, betweenness, and closeness—were calculated. Degree centrality measures the number of direct connections a user has, betweenness centrality captures the frequency with which a user acts as a bridge within the network paths, and closeness centrality reflects how close a user is to all other users in the network. These metrics collectively assess a user's prominence and potential to influence the subreddit community.

Figure 2. Table of Influencers Based on Network Centrality and Activity Metrics

Author	Degree	Betweness	Closeness	#Posts	#Comments
szechein	370	0.001787559	0.367683189	41	116
UnderpaidMook	128	0.000938742	0.335861261	6	131
Infinitale	168	0.001403612	0.351542694	12	106
Gwynvid	2	0	0.246338849		1
AshZE	52	0.000212014	0.290441652		41
SirKumference	474	0.005812954	0.384356292	62	237
PlotPlates	41	1.28E-05	0.318149535	4	8
UnakaPakala	1	0	0.00		1

(Refer to Centrality.csv in attachments)

A scoring system was applied, combining these centrality metrics with the users' activity data (number of submissions and comments), assigning subjective weights to each component, indicating their relative importance. This scoring framework allowed for the quantification of influence, ensuring that the top influencers were not merely the most active users but also the ones most central to the flow of information and engagement within the community.

The following would be the list of the **top 20 influencers** within the r/sanitymemes subreddit community, as presented in a formal report:

Influencer	Score
GlobelhHex	0.715628
ATalkingDoubleBarrel	0.630839
P0lskichomikv2	0.610161
JowettMcPepper	0.576677
Golb89	0.446969
Randomkidinthehood	0.433975
mad_harvest-6578	0.426779
WillaSato	0.401613
LegionLegacyDMK	0.39448
Rain_of_the_night-90	0.391507
DVG79	0.371795
Gapaot	0.369565
Giimasu	0.357749
SteeeelFieeld	0.355496
OneiceT	0.35151
StrawberryFloptart	0.332162
retarded-kancho	0.331301
DylanoRevs	0.319927
aratnagrid	0.318438
nerankori	0.31795

The final output is a ranked list of users, with the top 20 representing the subreddit's most influential members based on the composite score of their network position and activity level. These individuals are

considered the key players in spreading information, sparking discussions, and shaping community sentiment, making them valuable targets for outreach and engagement strategies. The methodology behind constructing the network and identifying influencers provides a robust foundation for understanding and leveraging the dynamics of online communities for real-world applications. If we select the top 100 influencers using the same ranking schema, the following directional network graph can be plotted, where we can see most influencers lie in the central cluster and highly connected with one another:

Top 100 Influencers Network