

Roles and Effects of Authenticity and Homophily in Influencer Marketing

Namita Goel

Pennsylvania State University

ARVIND RANGASWAMY

PennState Smeal College of Business

J. Andrew Petersen

Pennsylvania State University

Cite as:

Goel Namita, RANGASWAMY ARVIND, Petersen J. Andrew (2026), Roles and Effects of Authenticity and Homophily in Influencer Marketing. *Proceedings of the European Marketing Academy*, 55th, (134064)

Paper from the 55th Annual EMAC Conference, Bath, United Kingdom, June 2-5, 2026



Roles and Effects of Authenticity and Homophily in Influencer Marketing

Abstract: We assess the extent to which the volume and speed of followers' engagement with an influencer's post are influenced by: (1) the authenticity (to both the influencer and to the brand) of sponsored posts, and (2) the homophily of an influencer's network. Using Instagram data from Alo Yoga and Hello Fresh and doubly robust random forest causal estimation, we find that authenticity to brand drives volume and speed of engagement more than influencer self-authenticity. Greater follower homophily decreases both engagement volume and speed. Contrary to conventional wisdom, higher authenticity is not always better for driving engagement. We develop scalable text-based metrics using TF-IDF cosine similarity for authenticity and network density for homophily, overcoming limitations of survey-based measures.

Keywords: Influencer marketing; Authenticity; Homophily

Track: Digital and Social Media Marketing

1. Introduction and Research Questions

Consider two fitness influencers promoting the same brand of athletic wear. Both have similar follower counts and post types, yet the engagement their posts generate differ significantly. For one influencer, their post generates thousands of comments within hours; for the other, engagements trickle in slowly over weeks. What explains this difference? We propose that such performance variations stem from: (1) how posts balance the dual dimensions of authenticity (to the influencer and to the brand), and (2) how network homophily shapes follower receptivity. Traditional metrics used in industry, such as follower counts, fail to account for these performance differences. Also, due to this high variability in performance, brands struggle to find effective influencers.

The influencer marketing process involves three stakeholders: the brand collaborates with an influencer to promote products/services, the influencer incorporates the brand's products into content they post, and followers engage with that content. A key feature is that the brand's advertising message is only indirectly communicated through influencers who independently transform the advertiser's message (Hughes, Swaminathan, and Brooks 2019). The residual rights remain with the influencers—their participation and messaging are out of the advertiser's direct control (Peng and Van den Bulte 2023). This core tension between brands pursuing control and influencers retaining independence leads to highly varying engagement results. We examine how posts balance dual dimensions of authenticity (to influencer and to brand) and how network homophily shapes follower receptivity.

Research Questions: (RQ1) What are the effects of dual-dimensional authenticity and network homophily on engagement volume? (RQ2) What are the effects of these factors on engagement speed?

2. Key Contributions

This research makes five primary contributions that advance both theory and practice in influencer marketing:

2.1 Scalable measurement of authenticity

We develop a novel, scalable approach to measuring both dimensions of authenticity using TF-IDF (Term Frequency-Inverse Document Frequency) cosine similarity between text vectors. This computational method overcomes significant limitations of survey-based authenticity measures which are prone to bias. Our approach analyzes the actual language and

content of posts, comparing sponsored posts to: (1) the influencer's historical organic posts to measure authenticity of influencer to self, and (2) the brand's official content to measure authenticity to brand. The TF-IDF approach captures semantic content and stylistic elements (Gunawan et al., 2018).

2.2 Novel network-based measure of homophily

We propose an innovative measurement approach for homophily that combines affiliation-based links (who follows whom) with similarity-based links inferred from shared hashtag usage. This dual approach addresses a critical challenge in measuring homophily in influencer networks: direct follower-to-follower connections are often unobservable due to platform privacy restrictions. By treating shared hashtags as proxy links indicating latent similarity between influences and followers, we capture both observed and hidden relationships within each influencer's ego network. We calculate network density, where higher density indicates more homophilous networks.

2.3 Causal estimation minimizing bias

We employ doubly robust random forest estimation to address selection bias, with bias corrections averaging 41.76 units.

2.4 Counter-intuitive finding: greater authenticity is not always better

Contrary to belief and prior research suggesting greater authenticity enhances influencer marketing effectiveness, we find that greater authenticity is not always better for driving engagement. For instance, when authenticity to brand is low, increasing self-authenticity to influencer reduces engagement. These findings challenge the prevailing assumption that influencers should always maximize authenticity - to both self and to the brand. Calculated deviation from perfect authenticity can benefit both brands and influencers, depending on campaign objectives and network characteristics.

2.5 Surprising dominance of authenticity to brand over influencer authenticity

We find that authenticity to brand is substantially more important than influencer authenticity for driving engagement volume. This finding is contrary to the common industry belief that influencers should prioritize staying true to their personal brand voice above all else. The findings suggest that while influencer authenticity affects engagement speed (how quickly followers respond), authenticity to brand dominates volume (accumulated total engagement). This has profound implications for influencer selection and content strategy.

3. Conceptual Framework

3.1 Dual-dimensional authenticity

We conceptualize authenticity as operating simultaneously along two dimensions. Authenticity of influencer to self reflects consistency with the influencer's historical content patterns, style, and positioning. When sponsored posts deviate significantly from an influencer's established persona, followers may perceive inauthenticity. Authenticity to brand captures how well sponsored content aligns with the brand's core values, messaging, and identity. High authenticity to brand ensures the sponsored content genuinely represents the brand rather than appearing as a superficial endorsement. These two dimensions can conflict, creating strategic tradeoffs. An influencer with an irreverent style partnering with a luxury brand faces tension: maintaining self-authenticity may compromise authenticity to brand, while maximizing authenticity to brand may feel inauthentic to followers. This creates nine possible configurations (low/medium/high on each dimension), each with distinct engagement implications.

3.2 Network homophily

Homophily represents the degree to which followers share commonalities. In highly homophilous networks, followers constitute a cohesive community with shared identities and norms. Homophily influences whether reactions are collectively amplified or dispersed. In homogeneous networks, shared norms lead followers to interpret others' reactions as social validation, reinforcing their own engagement (Festinger, 1954; Cialdini & Goldstein, 2004). Heterogeneous networks inhibit collective amplification because diverse expectations lead to varied reactions, fragmenting group responses into isolated reactions.

3.3 Engagement volume versus speed

Most research focuses on engagement volume such as total likes, comments, and shares. We argue that engagement speed - how quickly interactions occur - provides distinct strategic insights. Rapid initial responses signal content quality to platform algorithms, potentially increasing visibility and creating positive feedback loops (Bakshy et al., 2011). Engagement volume, even when slower to accumulate, provides insights into sustained audience interest (Goel et al., 2016). These metrics may respond differently to authenticity configurations, creating strategic tradeoffs between immediate buzz and sustained conversation.

4. Data and Methods

4.1 Data collection

We collected data from Instagram for Alo Yoga, a fitness and fashion apparel brand collaborating with diverse influencer types ranging from fitness professionals to fashion enthusiasts. We also obtained data from Hello Fresh (a meal subscription service) for robustness checks. This provides the variation in authenticity expression necessary for our research. We gathered comprehensive data from 160 influencers who have collaborated with Alo Yoga (400 posts per influencer covering both sponsored and organic content), 3,000 brand posts from Alo Yoga's official account, and follower networks. The most recent posts were scraped as of March 2024. For network structure, we randomly sampled 20 followers for each influencer and scraped their posts. We created a network consisting of the brand as a node, its influencers as connected nodes, with added nodes of their followers.

4.2 Measurement of key variables

Engagement speed. We collected timestamp data for all comments and computed hourly arrivals. We fit a lognormal distribution to hourly comment counts to capture the temporal pattern of each post. The skewness of the lognormal distribution provides an intuitive measure of engagement speed, capturing the entire distributional shape. Positive skewness implies comments arrive early and die down soon. Negative skewness implies comments take time but the arrival rate is sustained. *Engagement volume.* We measure volume as log of comments over a month, normalized by follower count.

Homophily. We measure homophily as network density of each influencer's ego network, including both affiliation-based links (who follows whom) and similarity-based links (who is similar to whom based on shared hashtags). We calculate network density as a fraction of all possible connections. Higher density values indicate more homophilous networks where followers share greater similarity. We treat homophily as exogenous because the follower network remains relatively stable over short periods and reflects accumulated follower decisions over the influencer's entire history.

Authenticity. We measure authenticity using cosine similarity between TF-IDF vectors to capture both semantic content and stylistic elements, providing similarity scores from -1 to 1, where higher values indicate greater textual alignment. Authenticity to brand is measured as similarity of a sponsored post with the brand's previous posts; authenticity of influencer to self measures similarity of a sponsored post with the influencer's non-sponsored posts. We construct TF-IDF vectors after standard preprocessing (lowercasing, stop word removal, stemming, emoji standardization). For each sponsored post, we calculate cosine similarity

against the aggregate TF-IDF vector of comparison posts. Authentic expressions naturally maintain linguistic coherence in vocabulary choice, topic focus, and communication style (Pennebaker et al., 2003). Importantly, authenticity differs from mere consistency because cosine similarity is rotation-invariant but not scale-invariant, allowing us to distinguish genuine linguistic alignment from superficial keyword matching.

4.3 Identification strategy and doubly robust estimation

Estimating causal effects here is subject to selection bias: influencers strategically choose authenticity levels based on anticipated engagement. Influencers expecting strong response may create posts with high authenticity to brand content despite deviating from their typical style, while influencers who are uncertain stay closer to established patterns. Additionally, many factors correlate with both authenticity choices and engagement (popularity, posting time, seasonality). We use doubly robust estimation combining propensity scores and outcome regression via random forests. This provides consistent causal estimates if either model is correctly specified: if propensity scores are incorrect but outcome regression is correct, estimates remain consistent (and vice versa). We discretize authenticity scores into terciles (Low, Medium, High) for both dimensions, creating nine treatment combinations. We estimate propensity scores for all combinations using with covariates including influencer fixed effects, temporal variables, and post-specific features.

5. Results

Treatment combinations are coded as 0=Low, 1=Medium, 2=High for both authenticity to brand and authenticity of influencer to self.

5.1 Engagement speed patterns

The results (Fig1) show distinct patterns across authenticity combinations. For medium and high authenticity of influencer to self, engagement speed increases when authenticity to brand is also high. For low authenticity of influencer to self, the pattern reverses: the low-low combination generates relatively high speed, but speed decreases as authenticity to brand increases. Comments arrive fastest under two conditions: when posts are highly authentic to both brand and influencer, and when posts have low authenticity on both dimensions. However, these conditions differ in character. The low-low combination shows approximately 65% (Fig2) of comments arriving within the first ten hours, driven by

immediate negative reactions that quickly subside. The high-high combination shows rapid but sustained engagement.

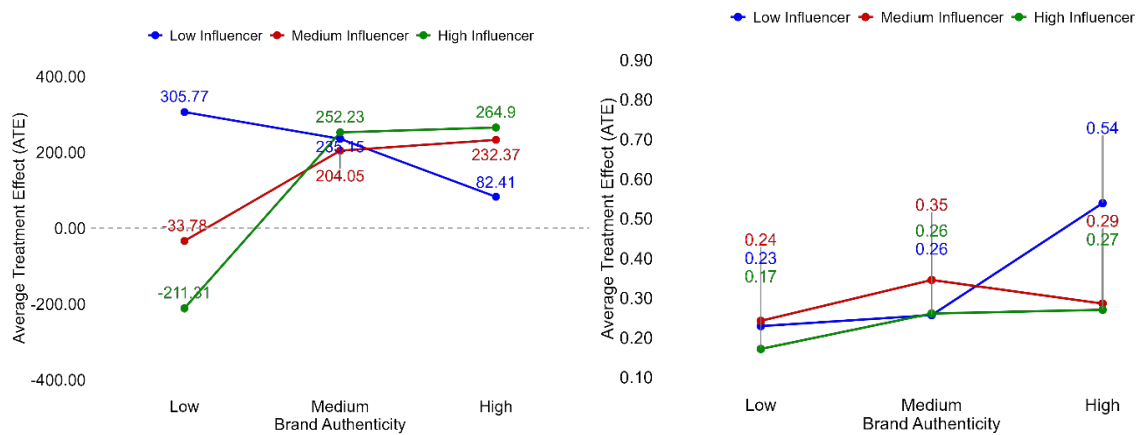


Figure 1: Left: Average Treatment Effect of Speed of Engagement (Skewness), Right: Average Treatment Effect of Volume of Engagement (log of comments ratio).

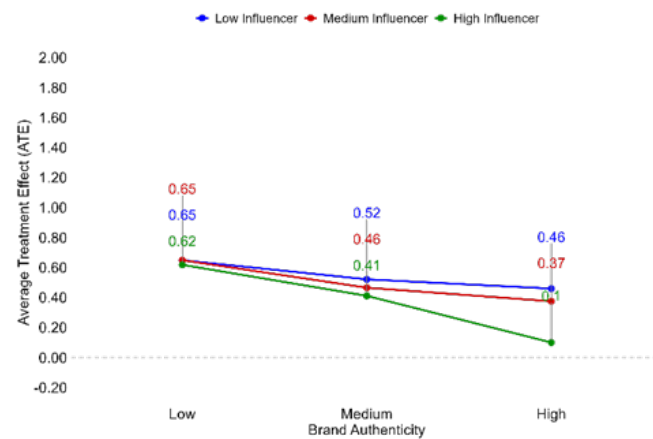


Figure 2: Percentage of total comments (based on lognormal distribution) that arrive in the first ten hours.

When posts have high authenticity to the brand and low authenticity to the influencer, the speed of engagement is, on average, moderate (82.41). A little less than 50 percent of the comments arrive within the first ten hours, indicating that comments take time to develop but remain sustained. Comment volume is high under these conditions, suggesting sustained audience interest. When posts have high influencer authenticity to self and low brand authenticity, then comments take longer to arrive (-211.31), and the volume is also low. Around 62% of the comments arrive within the first ten hours. The high skewness value (264.9) indicates a sharp, concentrated burst of comments that quickly dies down. Comment volume is also low (0.27), suggesting posts reach a peak with the receptive audience who respond rapidly before engagement reaches saturation.

5.3 Homophily moderation effects

We estimate treatment effects separately for high versus low homophily networks (Fig 3) (median split). The magnitudes of treatment effects diminish substantially under high homophily conditions across both speed and volume, suggesting the main effect of homophily is to reduce engagement. In high homophily networks characterized by shared perspectives, there may be redundancy in ideas, making individuals less likely to engage with sponsored content. Interestingly, in the high-homophily context, the combination of low authenticity of influencer to self and high authenticity to brand results in even lower engagement speed compared to low-homophily networks. This suggests that in homophilous networks, posts less authentic to the influencer are penalized more severely, possibly because cohesive communities have stronger consensus about what constitutes appropriate influencer behavior and collectively sanction perceived violations.

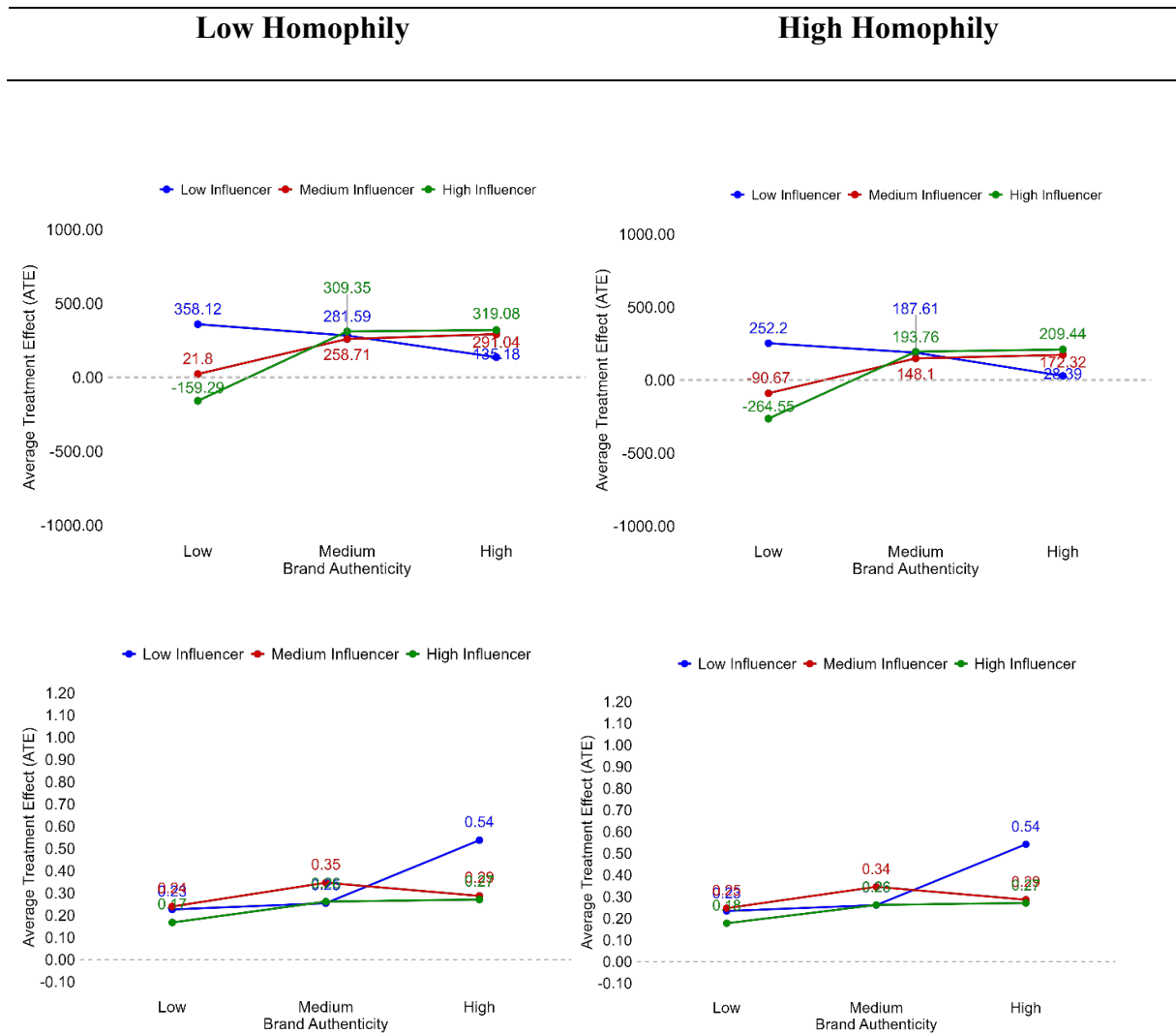


Figure 3: Average Treatment Effect of Speed and Volume of Engagement conditioned on homophily. 1. Top left: ATE of skewness for low homophily, 2. Top right: ATE of skewness for high homophily, 3. Bottom left:

ATE of log of comments ratio (volume) for low homophily, 4. Bottom right: ATE of log of comments ratio (volume) for high homophily.

6. Discussion and Implications

6.1 Theoretical implications

Our findings challenge the prevailing assumption that authenticity always enhances influencer marketing effectiveness. The surprising dominance of authenticity to brand over influencer authenticity suggests that sponsored content effectiveness depends more on how well it represents brand values than on maintaining influencer voice consistency. This shifts theoretical focus from influencer-centric to brand-message-centric models of influencer marketing effectiveness. Our finding reveals that more authenticity is not always better, suggests careful optimization rather than authenticity maximization strategies. The moderating role of homophily demonstrates that network composition fundamentally shapes content reception. Contrary to beliefs that homophilous networks amplify influence, we find they suppress engagement with sponsored content, likely reflecting heightened collective skepticism in consensus-driven communities.

6.2 Managerial implications

Our findings help managers think differently about influencer marketing. Our results and conceptualization of authenticity provide support to move away from the view that “more authenticity is better”. Managers do not have to chase after authenticity alignments, a post which is high on one dimension of authenticity and low on another can still generate engagement. Rather than seeking perfect brand-influencer alignment, managers should identify strategic opportunities where controlled authenticity tensions can generate superior engagement outcomes. Also, our findings suggest moving beyond traditional influencer selection metrics focused on follower count, toward evaluating network composition and authenticity alignment patterns. For influencers, these findings provide flexibility: when maintaining personal authenticity is challenging, strategic alignment with brand values can be effective in generating compensation, without sacrificing future engagement.

6.3 Limitations and future research

We measure observable engagement rather than downstream outcomes like purchases or brand perception changes. Our measures rely on text analysis, potentially missing visual authenticity cues. Future research should also examine how effects vary across product

categories, and investigate whether repeated sponsored posts affect authenticity perceptions over time, explore platform-specific dynamics, and connect engagement metrics to business outcomes. Longitudinal studies could track how influencer-brand partnerships evolve and whether audiences adapt to or increasingly resist commercial content over time.

References

- Audrezet, A., de Kerviler, G., & Moulard, J. G. (2020). Authenticity under threat: When social media influencers need to go beyond self-presentation. *Journal of Business Research*, 117, 557-569.
- Bakshy, E., Hofman, J. M., Mason, W. A., & Watts, D. J. (2011). Everyone's an influencer: Quantifying influence on Twitter. *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*, 65-74.
- Cialdini, R. B., & Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annual Review of Psychology*, 55, 591-621.
- Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, 7(2), 117-140.
- Goel, S., Anderson, A., Hofman, J., & Watts, D. J. (2016). The structural virality of online diffusion. *Management Science*, 62(1), 180-196.
- Gunawan, D., Sembiring, C. A., & Budiman, M. A. (2018). The implementation of cosine similarity to calculate text relevance between two documents. *Journal of Physics: Conference Series*, 978, 012120.
- Hughes, C., Swaminathan, V., & Brooks, G. (2019). Driving brand engagement through online social influencers: An empirical investigation of sponsored blogging campaigns. *Journal of Marketing*, 83(5), 78-96.
- Lee, J. A., & Eastin, M. S. (2021). Perceived authenticity of social media influencers: Scale development and validation. *Journal of Research in Interactive Marketing*, 15(4), 822-841.
- Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology*, 54, 547-577.
- Peng, C., & Van den Bulte, C. (2023). Participation vs. effectiveness in sponsored tweet campaigns: A quality-quantity conundrum. *Management Science*, 69(11), 6838-6859.