POLITECNICO MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE E DELL'INFORMAZIONE

EXECUTIVE SUMMARY OF THE THESIS

Decoding The Trend: Trend Archetypes, Psychological Framing and Strategic Timing on TikTok and Instagram Reels

TESI MAGISTRALE IN MANAGEMENT ENGINEERING – INGEGNERIA GESTIONALE

AUTHOR: ALI FATTAHI, PARISA KHANIA

ADVISOR: LUCIO LAMBERTI

ACADEMIC YEAR: 2025-2026

1. Introduction

Digital platforms have fundamentally transformed how information spreads, shifting diffusion away from centralized broadcast institutions toward networked systems driven by interpersonal transmission, algorithmic amplification, and participatory reinterpretation. Within this environment, virality has become both a cultural force and a managerial aspiration, yet empirical research consistently shows that viral success is neither random nor fully controllable, but dependent on the interaction of network structure, psychological activation, and social amplification mechanisms.

This thesis operationalizes that insight. Rather than attempting to predict virality in a deterministic sense, a goal undermined by social influence and path dependence, it identifies systematic regularities in how content spreads, when it spreads, and under what structural and contextual conditions participation becomes strategically advantageous. In doing so, it bridges theoretical diffusion models with applied managerial decision-making, translating abstract constructs such as simple versus complex contagion, lifecycle acceleration, and structural diversity into observable indicators derived from large-scale social media data.

2. Literature Review

2.1. Network Structure

The Infrastructure of Diffusion: Information travels furthest through weak ties that bridge disconnected communities [1], while these bridges have been formalized as structural holes that create brokerage advantages [2]. However, behavioral adoption, as distinct from passive awareness, requires reinforcement from multiple independent sources [3]. This distinction between simple and complex contagion is foundational: different content types activate fundamentally different diffusion architectures and reach alone cannot predict adoption.

2.2. Content Psychology

Why Individuals Choose to Share: High-arousal emotions drive sharing more reliably than valence alone, content designed purely to be pleasant is not optimized for diffusion [4]. Beyond emotional activation, functional motivations such as social currency, practical utility, and identity signaling further stimulate transmission [5]. A critical boundary condition emerges in commercial contexts: audiences activate persuasion knowledge when exposed to branded content, raising the psychological threshold required for organic shareability [6].

2.3. Social Mechanisms

From Individual Sharing to Collective Momentum: Identical content launched under different initial visibility conditions can produce radically divergent outcomes, as early random advantages compound through cumulative amplification [7]. Path dependence, not content quality alone, shapes viral trajectories. Visible engagement signals function as social proof, lowering adoption thresholds and creating self-reinforcing feedback loops in which early visibility begets further visibility [8].

Behavioral similarity among connected users may reflect shared predispositions rather than genuine persuasive influence [9]. For this reason, the engagement models developed in this thesis are explicitly predictive rather than causal.

2.4. Trend Dynamics

Archetypes, Timing, and the Limits of Prediction: Large reach and genuine viral diffusion are structurally distinct phenomena, size and cascade depth correlate only weakly [10]. Temporal diffusion patterns reveal differences between endogenous reinforcement-driven growth and exogenous shock-driven spikes [11]. Timing therefore emerges as a primary strategic variable [12], yet viral outcomes remain inherently probabilistic rather than deterministically engineerable [7].

The managerial objective is thus probability management: identifying structural and temporal conditions under which participation is more likely to succeed.

Research Gaps

From reviewing the existing literature, relevant gaps were detected and addressed in this study:

1. Structural virality metrics and lifecycle dynamics are rarely integrated, leading large attention events to be labelled "viral" without examination of whether they reflect genuine cascading diffusion or simple broadcast reach.
2. Content-based and network-based explanations operate largely in isolation, emotional shareability research and structural diffusion research are rarely unified within a single empirical framework.
3. Lifecycle models identify trend phases in theory but offer limited evidence-based guidance on participation timing, particularly for the compressed, algorithmically accelerated lifecycles characteristic of short-form video platforms.
4. The homophily-influence confound complicates interpretation of engagement metrics: high adoption within clustered communities may reflect shared predispositions rather than genuine persuasive transmission, risking systematic over-attribution of campaign effectiveness.
5. Data-driven classification of trend archetypes remains underdeveloped, particularly in localized social media contexts such as the Italian ecosystem analyzed here.

3. Research Hypothesis

In light of these facts, the following hypotheses were devised:

Diffusion Structure and Archetype Differentiation:

H1: Distinct trend archetypes emerge from systematic differences in acceleration, peak magnitude, lifecycle duration, and recurrence.

H2: Trends exhibiting higher early acceleration display shorter overall lifecycle duration than gradually accelerating trends.
H3: Sudden-burst trends exhibit compressed, high-intensity primary episodes, whereas reinforcement-based and cyclical trends display broader and more sustained diffusion curves.

***Engagement Depth Within Diffusion Structures*:**

H4: Spike-based archetypes generate high visibility but lower engagement depth relative to reinforcement-based archetypes.
H5: Gradual-growth and recurrent archetypes exhibit broader creator participation and higher interaction intensity per participant than spike-based archetypes.

Content Psychology and Engagement Activation:

H6: High-arousal emotional content generates higher engagement efficiency than low-arousal or neutral content.
H7: Negative high-arousal content produces faster early engagement but greater volatility and shorter persistence than positive high-arousal content.
H8: Functionally or socially normative framed content is associated with stronger engagement persistence than purely entertainment-based content.

Organic vs. Sponsored Participation:

H9: Sponsored posts achieve higher immediate visibility but lower engagement efficiency than organic posts.
H10: The effect of sponsored participation on engagement efficiency varies across diffusion archetypes.
H11: Higher early sponsored participation weakens the relationship between early engagement and subsequent engagement growth.

Timing Strategy and Lifecycle Position:

H12: Early participation generates higher engagement efficiency when trend velocity is increasing, and content-trend alignment is high.
H13: Participation during peak saturation produces lower engagement efficiency than participation during acceleration phases.

4. Materials and Methods

The empirical foundation is a large-scale observational dataset of approximately 6.5 million posts from TikTok and Instagram Reels comprised mostly of Italian content creators, reduced to 1.9 million after structural filtering. Crucially, the dataset was not pre-filtered for virality, retaining the full activity distribution reflects the theoretical commitment that viral outcomes must be identified ex post rather than assumed. Trends were identified through a relative-to-baseline spike detection procedure, classifying hashtags into five archetypes: viral spike, sustained, loaded, recurrent, and seasonal. Each archetype decomposed into seven lifecycle phases. Post-level captions were further classified using a large language model pipeline into emotional and functional dimensions, with each post evaluated for alignment with its trend's dominant psychological profile, operationalizing narrative coherence as an analytical variable alongside structural and temporal measures.

5. Results

The empirical investigation was divided into two Blocks: preliminary hypothesis analysis and a regression model analysis.

5.1. Hypothesis Testing

Diffusion Structure and Archetype Validation: The five trend archetypes: viral spike, loaded, sustained, seasonal, and recurrent are empirically distinct diffusion structures, not merely labels for different levels of popularity. Loaded and viral spike trends exhibit the highest early velocity and shortest lifecycles, consistent with shock-driven diffusion where rapid activation quickly exhausts the available attention space. Seasonal and recurrent archetypes show the opposite pattern: slower acceleration, longer persistence, and broader community involvement. A statistically meaningful velocity-longevity tradeoff underpins this differentiation (Spearman $\rho = -0.52$): trends that accelerate fastest decay soonest, suggesting that rapid growth compresses the temporal window for sustained participation rather than expanding it. Figure 1 showcases the median lifecycle of each trend type along with the distribution of phases.

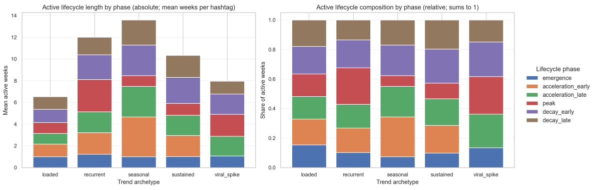

Figure 1: Trend archetype lifecycle distribution

Engagement Depth Across Archetypes: Structural form shapes not only how attention accumulates but how deeply participants engage within it. Seasonal and recurrent archetypes generate the broadest creator participation and highest cumulative engagement, building cultural scale through persistence and reinforcement across multiple waves. Spike archetypes, by contrast, concentrate engagement into narrow temporal windows with fewer unique creators and lower repetition per creator, consistent with a one-time participation model. Crucially, these differences are not explained by variation in creator follower counts, suggesting that diffusion structure itself, rather than creator scale, determines participation geometry. Figure 2 showcases the breadth, depth and efficiency of the engagement across different trend archetypes.

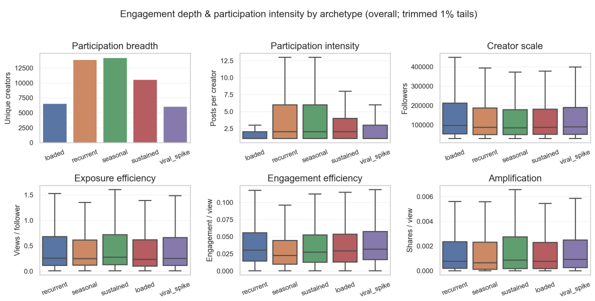

Figure 2: Engagement depth and participation intensity by trend archetype

Content Psychology and Engagement Activation: Emotional arousal increases engagement efficiency regardless of valence: high-arousal negative content (anger, anxiety) achieves the highest engagement per view, followed by high-arousal positive content, while neutral and low-arousal categories underperform. However, negative arousal comes at a cost, it produces sharper early engagement spikes followed by steeper declines, consistent with a volatility mechanism. Functional framing reveals a counterintuitive finding: social normative content achieves the highest engagement efficiency per view, while promotional content achieves the lowest. This is an apparent inversion of the post-level modelling results, explained by the difference between engagement depth and algorithmic amplification as outcome measures. Content psychology amplifies engagement within structural archetypes but does not override diffusion structure.

Organic vs. Sponsored Participation: Sponsored posts do not consistently outperform organic posts in visibility, median view counts are comparable or slightly lower, while engagement efficiency is systematically lower for sponsored content, with organic posts achieving 33% higher engagement per view and substantially higher amplification through shares. This pattern is consistent with a commercial intent discount, whereby audiences engage less intensely with explicitly branded content. The magnitude of this gap varies across archetypes: viral spike trends, characterized by strong organic momentum, show smaller sponsorship efficiency penalties, while spike dynamics with lower organic baseline show clearer gaps. Importantly, early sponsored participation does not weaken the persistence of engagement growth at the trend level as shown in figure 3, suggesting that commercial participation operates within established diffusion dynamics rather than disrupting them.

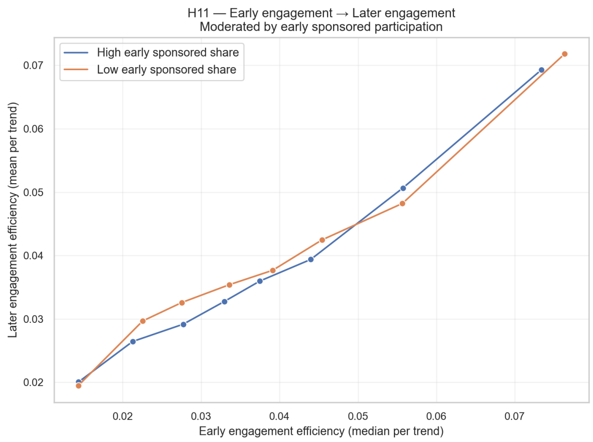

Figure 3: Sponsorship effect on engagement

Timing Strategy and Lifecycle Alignment: Timing interacts with both structural trend dynamics and narrative alignment to shape performance. During early and acceleration phases, posts that match the dominant emotional or functional framing of a trend consistently outperform misaligned posts in engagement efficiency. With emotionally aligned posts showing a 26.6% efficiency premium at peak relative to acceleration, compared to 14.6% for functionally aligned posts. Alignment enhances resonance rather than reach: misaligned posts can still achieve comparable visibility but convert that exposure into interaction less effectively. The alignment premium is most pronounced in high-velocity archetypes and weakens during decay, establishing that the early entry advantage is conditional rather than universal, it depends on both trend velocity and narrative coherence with the trend's dominant framing.

5.2. Regression Model

The modelling analysis proceeds in progressive stages: from a baseline OLS specification through quadratic, interaction, and piecewise threshold extensions, estimated separately for organic and sponsored content throughout. The central and most consequential finding is the engagement amplification gap. Pre-publication variables, including follower count, trend timing, content type, emotional framing, and lifecycle phase, collectively explain only 17% of organic view variance as shown in figure 4. Adding just two variables (historical likes and comments per post) raises that figure to 61% in the baseline and 73% in the final specification. This gap, produced by only two additional degrees of freedom, is the empirical signature of algorithmic distribution: reach on these platforms is not allocated based on who a creator is connected to, but dynamically in response to how early audiences engage with the content.

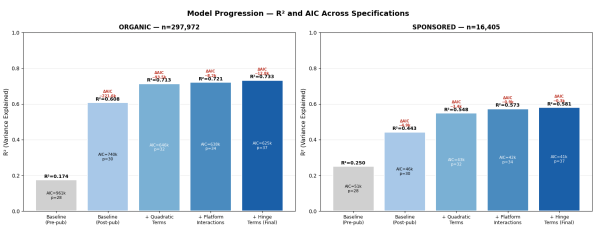

Figure 4: Organic and sponsored model progression.

The engagement-to-views relationship is not linear. Partial residual analysis reveals an S-shaped pattern for both likes and comments, with accelerating returns at high engagement levels. Critically, piecewise regression identifies a saturation threshold at the 90th percentile of the *likes* distribution, above which marginal returns sharply reverse: each additional unit of engagement is associated with a downward correction in predicted views rather than further amplification. This saturation effect holds across both organic and sponsored content, suggesting a platform-level mechanism rather than a content-specific one, and has direct implications for how paid amplification budgets should be allocated.

Platform architecture moderates these dynamics in structurally meaningful ways. According to figure 5, the follower elasticity of views is substantially weaker on TikTok than Instagram, while the like elasticity is stronger, meaning that on TikTok, engagement track record effectively displaces follower count as the primary reach driver. A creator with modest followers but consistently high engagement achieves nearly the same predicted reach as a large-audience creator with passive followers. This finding quantifies, at the coefficient level, TikTok's algorithmic meritocracy relative to Instagram's residual social-graph dependence.

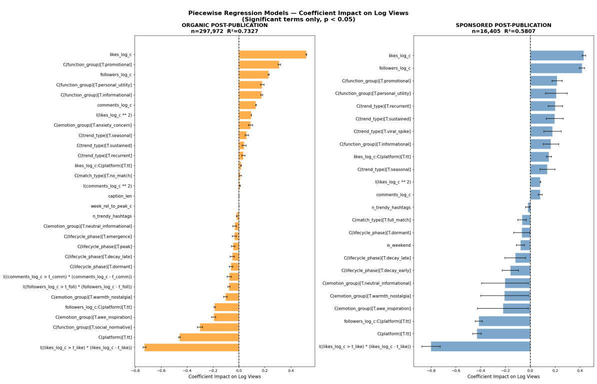

Figure 5: Final model coefficients for organic and sponsored datasets.

Among the categorical predictors, function group emerges as the strongest and most consistent driver of performance across both organic and sponsored content. Promotional content achieves the highest view coefficient of any categorical predictor in the organic model. A counterintuitive finding explained by the dense transactional engagement it generates, which the algorithm reads as a strong content-audience fit signal. Social normative content carries the largest negative coefficient in the entire model, consistent with its structural dependence on distributed social proof that no single post can supply. Emotion group effects are significant for organic content: anxiety and concern framing carries a positive premium, while awe and nostalgia carry substantial penalties, but are largely attenuated for sponsored content, suggesting that sponsorship disclosure alters the audience reception dynamics through which emotional framing normally operates.

Lifecycle timing further refines the picture. The acceleration phase immediately before peak is the algorithmically optimal posting window for organic content, outperforming both earlier emergence phases, where trends lack sufficient momentum; and the peak itself, where saturation dynamics suppress marginal performance. For sponsored content, the timing penalty structure is steeper in the decay phases, meaning that misjudging timing carries a disproportionately higher cost for brand-partnered content than for organic posts.

6. Discussion

The discussion situates the empirical results within three overarching interpretive frames. The first and most consequential is the paradigm shift from social graph to content graph. The engagement amplification gap is not simply evidence that likes and comments are useful predictors, it is evidence that the foundational assumption of influencer marketing, that reach flows from follower count and network position, is empirically inadequate for the algorithmic environments studied here. The pre-publication R^2 of 17.5% for organic content establishes a boundary of what any pre-publication decision can determine, with the remaining variance driven by stochastic early algorithmic audience assignment that no content or creator decision controls. This extends Salganik et al.'s (2006) laboratory findings on path dependence to a large-scale natural setting.

The second frame concerns the empirical validation of diffusion archetypes as distinct mechanisms rather than labels. The structural differentiation findings confirm that loaded and viral spike trends operate through shock-driven simple contagion, while sustained, seasonal, and recurrent trends operate through reinforcement-based complex contagion; a distinction with direct consequences for what strategic intervention can achieve within each archetype. Notably, the non-finding on H11 is theoretically significant: early sponsored participation does not suppress organic diffusion momentum at the trend level, suggesting that commercial presence and organic engagement can coexist rather than compete, which revises the conservative assumption underlying much influencer authenticity guidance.

The third frame is the hierarchy of explanatory power across the three analytical layers. Structure defines the geometry of available engagement. Content activates engagement within that geometry. Engagement track record determines whether the algorithm amplifies the result beyond the initial audience. No single layer is sufficient, but they are not equally important, and the engagement track record dominates all others for organic content.

6.1. Research and Managerial Implications

The theoretical contributions of the thesis cluster around four advances. First, the content graph framework reframes the pre-cascade algorithmic threshold, the decision point at which the algorithm shifts content from limited test distribution to broad amplification, as the primary object of diffusion research on these platforms, displacing cascade topology post-amplification as the dominant focus. Second, the archetype framework provides a methodology for disaggregating trend populations before analysis, resolving the conflation of structurally heterogeneous mechanisms that has produced averaged and often misleading estimates in prior viral content research. Third, the pre-publication R^2 provides a measurement of the predictability boundary in social media performance. Fourth, the platform interaction results quantify TikTok's algorithmic meritocracy at the coefficient level, the first observational evidence of platform architecture effects on diffusion elasticities independently of content characteristics.

The managerial implications are equally concrete. The most immediately actionable is the replacement of follower-based CPM (Cost Per Mille) pricing on TikTok with engagement-adjusted models: the follower advantage on TikTok is empirically negligible after controlling for engagement track record, meaning brands using standard follower-based pricing are systematically paying for a structural advantage the algorithm does not provide. Engagement track record, the historical central tendency of likes and comments per post across multiple campaigns, should replace follower count as the primary influencer selection criterion for organic campaigns, while follower count retains relevance for campaigns targeting the creator's established social graph directly.

The lifecycle findings sharpen the timing recommendation from "enter early" to "enter at the inflection point": the *acceleration_early* phase is the specific window combining sufficient trend momentum with remaining amplification headroom. The saturation threshold identified by the piecewise regression, above which marginal engagement returns reverse, provides a computable criterion for paid amplification allocation, shifting the media buying logic from purchasing reach directly to pushing early engagement signals across the algorithm's amplification trigger. Finally, the benchmark-relative ROI framework offers a practically implementable solution to the over attribution problem: distinguishing decision quality from outcome quality and rewarding process optimization rather than outcome luck in a system where the algorithm resolves the majority of variance after the post goes live.

7. Conclusions

This thesis set out to identify the structural determinants of content performance on short-form video platforms, examining how trend archetype, lifecycle timing, platform architecture, content characteristics, and creator engagement history jointly shape view outcomes. The central finding is the engagement amplification gap: pre-publication variables collectively explain less than 18% of organic view variance, while the addition of historical likes and comments raises that figure to 73% in the final specification. This is not a modelling deficiency; it is the empirical signature of algorithmic distribution. Reach on TikTok and Instagram Reels is allocated dynamically in response to early audience engagement, not assigned based on creator scale or content decisions made before posting. The content graph has structurally displaced the social graph as the primary diffusion infrastructure.

Beyond the amplification gap, the thesis establishes that trends are structurally distinct diffusion mechanisms, not points on a single popularity continuum. Each archetype carries a different risk-reward profile, participation window, and optimal content strategy. The acceleration phase immediately before peak is the algorithmically optimal posting window. TikTok and Instagram operate fundamentally different elasticity structures, with direct implications for how influencer partnerships should be priced. And a saturation threshold exists above which marginal engagement returns reverse, a practically computable criterion for amplification decisions.

Taken together, the findings reframe influencer campaign planning around a fundamental empirical boundary: the majority of outcome variance is determined after publication by algorithmic dynamics that no pre-publication decision fully controls. Recognizing this boundary is not a counsel of pessimism, it is the prerequisite for building forecasting frameworks, pricing models, and evaluation criteria that are honest about what strategy can and cannot achieve.

7.1. Limitations and Future Research

Several boundaries warrant acknowledgement. Trend identification relies exclusively on hashtags from captions, excluding audio and format driven participation central to TikTok; multimodal frameworks represent the most consequential methodological extension. The *is_sponsored* classification captures disclosure rather than confirmed paid distribution, meaning sponsorship and media effects cannot be disentangled. Engagement metrics are predictive proxies rather than causal estimates, and experimental designs would be required to establish causal direction. The sponsored subsample represents only 5.2% of observations, limiting statistical power for several sponsored-specific effects. Trend archetypes are assigned retrospectively, making them unknowable at the point of posting; early-signal classification models would make the framework operationally usable in real time. Finally, engagement track record stability is assumed rather than demonstrated, as no longitudinal creator-level panel data was available.

8. Bibliography

[1] Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology, 78*(6), 1360–1380.

[2] Burt, R. S. (1992). *Structural holes: The social structure of competition.* Harvard University Press.

[3] Centola, D., & Macy, M. (2007). Complex contagions and the weakness of long ties. *American Journal of Sociology, 113*(3), 702–734.

[4] Berger, J., & Milkman, K. L. (2012). What makes online content viral? *Journal of Marketing Research, 49*(2), 192–205.

[5] Berger, J. (2014). *Contagious: Why things catch on.* Simon & Schuster.

[6] Campbell, M. C., & Kirmani, A. (2000). Consumers' use of persuasion knowledge. *Journal of Consumer Research, 27*(1), 69–83.

[7] Salganik, M. J., Dodds, P. S., & Watts, D. J. (2006). Experimental study of inequality and unpredictability in an artificial cultural market. *Science, 311*(5762), 854–856.

[8] Cialdini, R. B. (2009). *Influence: Science and practice* (5th ed.). Pearson.

[9] Aral, S., Muchnik, L., & Sundararajan, A. (2009). Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *PNAS, 106*(51), 21544–21549.

[10] Goel, S., Anderson, A., Hofman, J., & Watts, D. J. (2016). The structural virality of online diffusion. *Management Science, 62*(1), 180–196.

[11] Crane, R., & Sornette, D. (2008). Robust dynamic classes revealed by measuring the response function of a social system. *PNAS, 105*(41), 15649–15653.

[12] Wu, F., & Huberman, B. A. (2007). Novelty and collective attention. *PNAS, 104*(45), 17599–17601.

Appendix

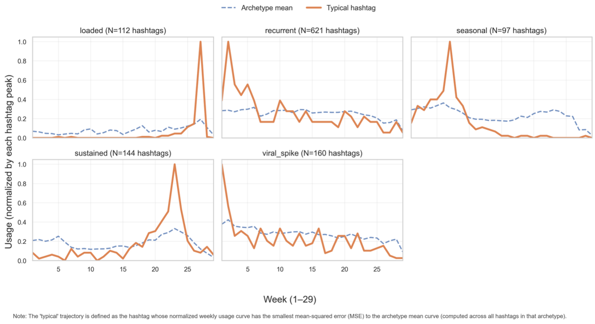

Figure 6: Closest to median trend shape per trend archetype

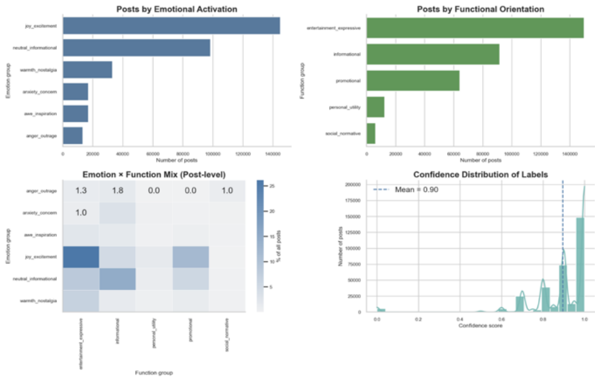

Figure 7: The Emotion - Function distribution on post level

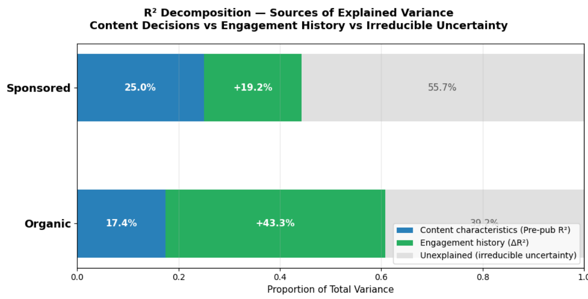

Figure 8: Baseline models explained variance