A Multi-Signal Framework for Measuring TrapIntensity in YouTube Recommendation Networks

Nitin Agarwal
COSMOS Research Center
University of Arkansas at Little Rock, USA
International Computer Science Institute
University of California, Berkeley, USA
nxagarwal@ualr.edu

Md Monoarul Islam Bhuiyan
COSMOS Research Center
University of Arkansas at Little Rock
Little Rock, USA
mbhuiyan@ualr.edu

***Abstract*—Algorithmic recommendation systems shape user exposure by structuring pathways through content and reinforcing sustained engagement within specific regions of the network. Prior work on content traps has primarily emphasized structural dynamics or isolated content signals, limiting the ability to measure how multiple forms of influence jointly contribute to algorithmic entrapment. In this study, we extend the TrapIntensity framework by integrating four components: structural entrapment, measured through attraction and retention dynamics in recommendation networks; persuasion intensity, derived from textual signals grounded in four persuasion theories; transcript-guided keyframe extraction to capture temporally relevant segments of video content; and social, cultural, economic, and political cues extracted from those keyframes. Rather than interpreting higher TrapIntensity scores as stronger trapping in themselves, we treat TrapIntensity as a measurement construct and evaluate whether the inclusion of these additional signals improves its validity. Across fourteen YouTube recommendation-network datasets, we find that the extended formulation improves alignment with engagement and more clearly distinguishes top focal structures from the rest of the network. These findings suggest that measuring content traps requires a joint consideration of structural exposure, persuasion, and visual-symbolic content cues.**

***Index Terms*—Persuasion, Symbols, TrapIntensity, YouTube Recommendation Network, Social Network Analysis, Social Media.**

I. Introduction

Algorithmic recommendation systems increasingly shape how users are exposed to and interact with information online. Platforms such as YouTube construct sequential pathways through content, influencing not only what users encounter but also how long they remain engaged. While these systems facilitate discovery, they also raise concerns about sustained interaction within specific regions of the network, where users may be repeatedly exposed to related content [1].

Recent work has conceptualized such phenomena through the notion of content traps, [1] defined as structurally cohesive regions of recommendation networks that attract users and retain their navigation over time. Existing approaches to measuring these dynamics have largely focused on structural properties of the network, such as connectivity and traversal behavior, or on textual signals derived from content. However, these approaches are typically considered in isolation and do not fully capture how different forms of influence jointly contribute to sustained engagement.

In practice, user exposure in video-based environments is shaped by multiple interacting factors. Structural pathways determine how users move through the network, persuasion signals embedded in textual content influence interpretation and engagement, and visual elements within videos provide additional cues that may reinforce attention. In particular, video content contains temporally localized segments and symbolic imagery that can contribute to sustained interaction in ways that are not captured by structural or textual signals alone.

To address this limitation, we extend the TrapIntensity framework by integrating multiple complementary components. First, we model structural entrapment through attraction and retention dynamics derived from recommendation networks [2]. Second, we incorporate persuasion signals grounded in four established theories namely Social Judgment Theory (SJT), the Elaboration Likelihood Model (ELM), Cognitive Dissonance Theory (CDT), and Narrative Paradigm to capture how textual content may influence user processing. Third, we extract temporally relevant segments of video content using transcript-guided keyframe extraction. Finally, we extract content-level signals from keyframes by identifying social, cultural, economic, and political cues present in the visual content. These cues capture abstract aspects of videos that may influence user attention and engagement beyond textual and structural factors.

A key challenge in extending TrapIntensity is how to interpret changes in the resulting scores. Importantly, the inclusion of additional signals does not imply that recommendation regions become inherently more trap-like. Rather, it reflects an expansion of the measurement space, allowing the framework to capture a broader set of characteristics associated with sustained engagement. We therefore treat TrapIntensity as a

measurement construct and evaluate whether incorporating additional signals improves its validity. Specifically, we assess whether the extended formulation (i) aligns more closely with observable behavioral outcomes and (ii) more effectively distinguishes structurally salient regions from the rest of the network. We evaluate this framework across twelve YouTube recommendation-network datasets and show that integrating structural, persuasive, and visual-symbolic signals leads to a more comprehensive measurement of content traps.

The rest of the paper is organized as follows. Section II presents the related work. Research methodology is explained in Section III. Section IV presents the results and discussions. Section V concludes the paper with future research directions.

II. Related Works

Understanding how recommendation systems shape user exposure and engagement has been a central focus in computational social science. Early work on networked systems emphasized structural mechanisms such as centrality, connectivity, and community formation, showing how network topology can influence information flow [3], [4]. In the context of online platforms, these ideas have been extended to study how recommendation systems guide user navigation and concentrate attention within specific regions of content networks.

Recent studies have examined problematic dynamics such as echo chambers, filter bubbles, and recommendation-driven polarization. Empirical evidence suggests that algorithmic curation can reinforce existing viewpoints and limit exposure to diverse information, particularly in politically salient domains [5], [6]. Prior work has also highlighted how recommendation pathways contribute to the amplification of specific narratives and sustained engagement patterns [7], [8]. These findings underscore the importance of analyzing both network structure and user behavior.

To operationalize such dynamics, recent research has introduced the concept of content traps, defined as structurally cohesive regions that attract and retain user navigation [2]. This approach builds on methods for identifying influential groups in networks, including focal structure analysis and collective influence–based techniques [9], [10]. Random-walk-based models further enable the estimation of behavioral dynamics such as attraction and retention, providing a quantitative framework for understanding how users traverse recommendation networks [11].

While structural approaches capture navigation patterns, they do not fully account for the role of content in shaping engagement. Prior work has therefore incorporated content-level signals, particularly through the analysis of textual persuasion using established theories such as SJT [12], the ELM [13], CDT [14], and Narrative Paradigm [15]. More recently, advances in video analysis have enabled the extraction of semantically meaningful segments and visual cues. Transcript-guided methods improve semantic coverage by identifying temporally relevant portions of video content, while vision–language models allow the detection of higher-level symbolic signals from images, capturing abstract concepts such as political or cultural cues [16].

Despite these advances, existing approaches typically treat structural and content-based signals separately. As a result, it remains unclear how different signals jointly contribute to engagement in recommendation environments. This study addresses this gap by extending the TrapIntensity framework to incorporate additional content-based signals alongside structural dynamics, enabling a more comprehensive measurement of content traps.

III. Methodology

This study develops an extended measurement framework for quantifying content trap intensity [1] by integrating structural, persuasive, and content-based signals within YouTube recommendation networks. The overall methodology consists of constructing recommendation graphs, identifying structurally significant groups, estimating attraction and retention dynamics, extracting persuasion cues from textual content, extracting higher-level cues from keyframes, and combining these components into a composite TrapIntensity measure.

We begin by constructing multi-hop recommendation networks using recursive crawling of YouTube's watch-next interface. Each video is represented as a node, and directed edges capture recommendation relationships [7]. To reduce sensitivity to initial conditions, multiple seed sets from stratified sampling [17] are used to generate independent network instances for each dataset.

To identify candidate trap regions, we apply collective influence focal structure analysis (CI-FSA) [18], which extracts cohesive groups of nodes that are structurally consequential within the network. Unlike node-centric approaches, CI-FSA emphasizes connected sets of videos whose collective position supports reachability and influence. These focal structures serve as the unit of analysis for subsequent measurements.

Structural entrapment is operationalized through attraction and retention dynamics estimated using hop-aware random walk simulations. Attraction measures how efficiently a random walk reaches a focal structure from nearby nodes, while retention measures how strongly the walk remains within the structure once entered [2]. These quantities are computed by initiating walks at varying hop distances and aggregating first-hitting times and residence behavior. The resulting scores are normalized and combined to produce a structural entrapment measure for each focal structure.

To quantify persuasive intensity, we use a Large Language Model *(Gemma3-27B-IT, Modality: Text)* procedure grounded in four persuasion theories: SJT, the ELM, CDT, and Narrative Paradigm. The model evaluates each video's title, description, and transcript separately in order to detect theory-grounded persuasion cues. To improve consistency and auditability, the output is constrained to a structured format requiring binary indicators for the presence of each theoretical cue, supporting textual evidence, and a brief rationale. We use low-temperature inference to reduce stochastic variation across runs. Video-level persuasion scores are computed by aggregating detected

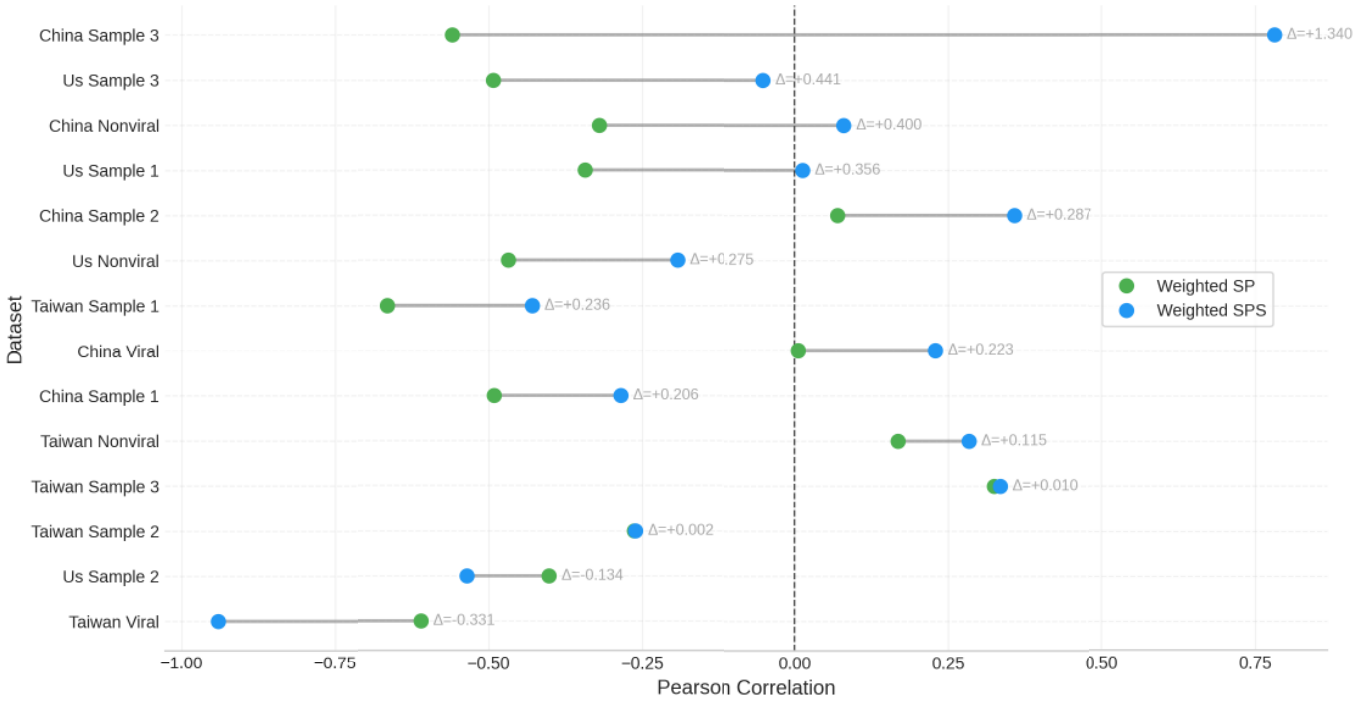

Fig. 1. Relationship between TrapIntensity and engagement across datasets for structural and keyframe-augmented formulations. Incorporating keyframe-based signals improves alignment with engagement, highlighting the role of temporally salient content segments beyond structural dynamics.

cues across fields, and focal-structure-level persuasion intensity is then obtained by averaging these scores across the videos belonging to each focal structure.

To incorporate visual content signals, we first extract representative frames from videos using a transcript-guided keyframe selection approach [19]. This method treats transcripts as a temporal index of semantic relevance, enabling the identification of segments that correspond to meaningful content rather than relying solely on visual changes. Transcripts are partitioned into overlapping temporal windows and scored using a hybrid relevance function that combines semantic similarity and keyword matching. High-relevance segments are aggregated into intervals, from which representative frames are extracted. This process preserves semantically important portions of videos while reducing redundancy, allowing for efficient downstream analysis.

We then analyze the extracted keyframes to quantify higher-level content signals by identifying social, cultural, economic, and political cues present in the visual content. Rather than focusing on object-level detection, we use a vision-language model *(Gemma3-27B-IT, Modality: Image + Text)* to interpret frames at a conceptual level, guided by a domain-informed prompting framework. The model produces structured binary indicators for each cue category, enabling consistent and interpretable measurement across videos. To ensure reproducibility, inference is performed under low-temperature settings, reducing variability across runs. These cue-based signals are computed at the video level and aggregated across focal structures.

The final TrapIntensity measure is constructed by combining normalized structural and content-based components. We evaluate incremental formulations that progressively incorporate persuasion and keyframe-derived cues alongside structural entrapment. In addition to equal-weight combinations, we use Cliff's delta [20] to derive context-sensitive weights for the structural, persuasion, and symbol-based cue components. Specifically, we compute separability between high-intensity focal structures and the remaining focal structures for each component, transform the absolute Cliff's delta values into normalized weights, and use these weights to construct weighted TrapIntensity variants. This allows the contribution of each component to vary across datasets according to its empirical discriminative strength.

IV. Results and Analysis

We evaluate the extended TrapIntensity formulation by examining how progressively incorporating additional signals affects its relationship with engagement and its ability to distinguish structurally salient regions within the network. Across the fourteen datasets, the structural and persuasion formulation (SP) provides a baseline level of association with engagement, reflecting how network topology and textual persuasion jointly shape user navigation and attention. This baseline captures how users are guided into and retained within specific regions of the recommendation graph.

The inclusion of keyframe-derived content signals leads to consistent improvements in alignment with engagement metrics, as illustrated in Figure 1. Compared to the structure-and-

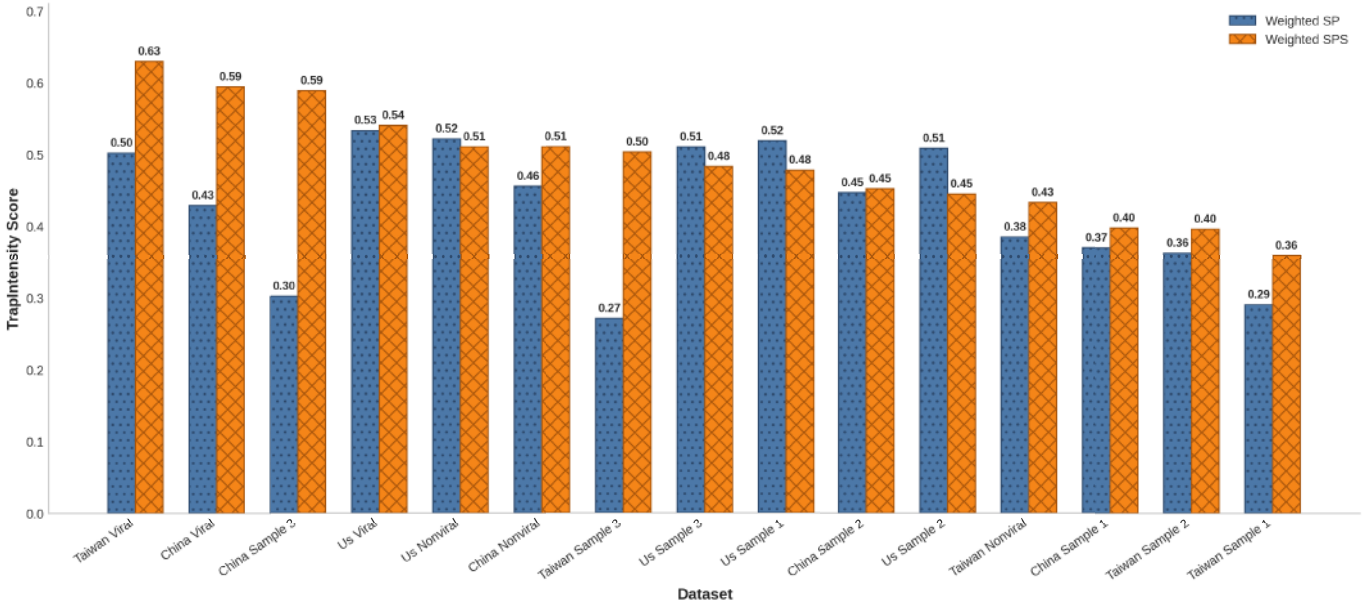

Fig. 2. Comparison of TrapIntensity formulations with the inclusion of content-level cues extracted from keyframes, including social, cultural, economic, and political signals. The extended formulation shows the strongest alignment with engagement, indicating that these cues enhance the measurement of trap-relevant characteristics.

persuasion formulation, the extended variant exhibits stronger and more consistent trends across datasets. By focusing on transcript-aligned segments of video content, the keyframe representation captures temporally salient portions of videos that are more likely to influence user interaction. This suggests that engagement is not uniformly distributed across entire videos, but is often driven by semantically meaningful segments. As a result, incorporating keyframe-based content signals provides a more precise representation of user exposure than structure-and-persuasion measures alone.

Further improvements are observed when social, cultural, economic, and political cues extracted from keyframes are incorporated. As shown in Figure 2, this formulation yields the strongest alignment with engagement across most datasets. These higher-level cues capture properties of the content itself, such as politically or culturally salient imagery, that are not reflected in structural dynamics or textual persuasion alone. Their contribution suggests that visual content plays an important role in sustaining user attention within recommendation pathways.

Importantly, these improvements should not be interpreted as evidence that the network becomes more trap-like with the inclusion of additional signals. Rather, incorporating content-level features expands the measurement space, allowing TrapIntensity to capture a broader set of characteristics associated with sustained engagement. From this perspective, increases in TrapIntensity reflect improvements in measurement sensitivity rather than direct increases in trapping.

To further examine how different signals contribute to the extended formulation, we analyze the dominant weight assigned to each component across datasets. Table I summarizes

TABLE I
DOMINANT SIGNAL CONTRIBUTING TO TRAPINTENSITY ACROSS DATASETS.

Dataset	Dominant Signal
China Sample 3	Symbols
US Sample 3	Structural
China Nonviral	Structural
US Sample 1	Structural
China Sample 2	Structural
US Nonviral	Structural
Taiwan Sample 1	Persuasion
China Viral	Symbols
China Sample 1	Persuasion
Taiwan Nonviral	Persuasion
Taiwan Sample 3	Persuasion
Taiwan Sample 2	Structural

whether the dominant signal is structural, persuasion-based, or symbolic. Structural signals dominate in a majority of datasets, indicating that network topology remains a primary driver of entrapment dynamics. However, persuasion signals emerge as dominant in several cases, particularly in Taiwan-related datasets, suggesting that textual framing and narrative cues can play a significant role in shaping engagement. Notably, symbolic signals dominate in specific datasets such as China Sample 3 and China Viral, indicating that visual-symbolic content can substantially contribute to sustained attention in certain contexts.

This distribution highlights that content trap dynamics are not governed by a single factor, but instead emerge from the interplay between structural exposure, persuasive framing, and symbolic content. The variation across datasets further indicates that the relative importance of these signals is context-dependent. In some cases, structural cohesion dominates, while

in others, persuasion or symbolic cues play a more prominent role in sustaining engagement.

In addition to alignment with engagement, we evaluate the ability of the extended formulation to distinguish between top focal structures and the rest of the network. A valid measure of content traps should assign systematically higher scores to structurally salient regions. Across datasets, the extended formulation produces clearer separation between these groups compared to simpler variants, indicating improved discriminative power.

Taken together, these results suggest that content trap dynamics cannot be fully explained by structural properties alone. While network topology determines how users navigate between videos, the characteristics of the content encountered along these pathways also play a crucial role in sustaining engagement. The contribution of keyframe-based signals highlights the importance of temporal concentration within videos, while the contribution of symbolic cues underscores the role of abstract visual content in reinforcing attention.

Overall, the extended TrapIntensity formulation should be interpreted as a more comprehensive measurement of trap-relevant characteristics. A richer formulation is preferable not because it yields larger scores, but because it better aligns with observable indicators of sustained attention and more effectively distinguishes salient regions within the network.

V. Conclusion and Future Work

In this work, we extend the TrapIntensity framework by incorporating additional content-level signals derived from video data, including keyframe-based representations and symbolic cues extracted from visual content. Our objective is not to increase the magnitude of TrapIntensity scores, but to improve the measurement of content traps by capturing a broader set of characteristics associated with sustained user engagement. A central takeaway of this study is that the inclusion of additional signals should not be interpreted as making recommendation regions inherently more trap-like. Instead, these signals expand the measurement space, enabling the framework to capture aspects of user exposure not reflected in structural dynamics alone. From this perspective, improvements in TrapIntensity indicate enhanced measurement sensitivity rather than increased trapping.

We evaluate the extended formulation along two complementary dimensions. First, we observe stronger alignment between TrapIntensity and engagement metrics across datasets, suggesting that the additional signals capture factors associated with sustained user attention. Second, we find that the extended formulation produces clearer separation between top focal structures and the rest of the network, indicating improved discriminative power. Together, these results support improved construct validity.

Despite these findings, several limitations remain. Engagement metrics such as views and likes serve as observational proxies and do not directly capture user-level exposure trajectories. The content-level cues extracted from keyframes, including social, cultural, economic, and political signals, depend on model assumptions and prompting strategies, which may introduce bias. In addition, our analysis operates at the focal-structure level and does not explicitly model individual user behavior or temporal viewing sequences, limiting causal interpretation. Future work may address these limitations by incorporating user-level interaction data, validating symbolic signals through human annotation or alternative extraction methods, and exploring additional modalities and adaptive weighting schemes.

More broadly, this work highlights the importance of treating content traps as a sociotechnical phenomenon shaped by both network structure and content characteristics. A richer TrapIntensity formulation should therefore be considered more valid, not because it yields larger scores, but because it better aligns with observable behavioral patterns and more effectively distinguishes salient regions within the network.

Acknowledgements

This research is funded in part by the U.S. National Science Foundation (OIA-1946391, OIA-1920920), U.S. Office of the Under Secretary of Defense for Research and Engineering (FA9550-22-1-0332), U.S. Army Research Office (W911NF-23-1-0011, W911NF-24-1-0078, W911NF-25-1-0147), U.S. Office of Naval Research (N00014-21-1-2121, N00014-21-1-2765, N00014-22-1-2318), U.S. Air Force Research Laboratory, U.S. Defense Advanced Research Projects Agency, the Australian Department of Defense Strategic Policy Grants Program, Arkansas Research Alliance, the Jerry L. Maulden/Entergy Endowment, and the Donaghey Foundation at the University of Arkansas at Little Rock. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding organizations. The researchers gratefully acknowledge the support.

References

[1] N. Agarwal and M. M. I. Bhuiyan, "Attraction and retention dynamics in recommendation graphs: a cross-dataset analysis using uniform and degree-biased random walks," *Social Network Analysis and Mining*, 2026.

[2] M. M. I. Bhuiyan and N. Agarwal, "Evaluating structural attractors and retainers in youtube recommendation networks," in *Proceedings of the 17th International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2025)*. Niagara Falls, Ontario, Canada: IEEE/ACM, August 2025.

[3] S. Brin and L. Page, "The anatomy of a large-scale hypertextual web search engine," *Computer networks and ISDN systems*, vol. 30, no. 1-7, pp. 107–117, 1998.

[4] J. M. Kleinberg, "Authoritative sources in a hyperlinked environment," *Journal of the ACM (JACM)*, vol. 46, no. 5, pp. 604–632, 1999.

[5] C. A. Bail, L. P. Argyle, T. W. Brown, J. P. Bumpus, H. Chen, M. F. Hunzaker, J. Lee, M. Mann, F. Merhout, and A. Volfovsky, "Exposure to opposing views on social media can increase political polarization," *Proceedings of the National Academy of Sciences*, vol. 115, no. 37, pp. 9216–9221, 2018.

[6] M. Cinelli, G. D. F. Morales, A. Galeazzi, W. Quattrociocchi, and M. Starnini, "Echo chambers on social media: A comparative analysis," *arXiv preprint arXiv:2004.09603*, 2020.

[7] M. C. Cakmak, N. Agarwal, and R. Oni, "The bias beneath: analyzing drift in youtube's algorithmic recommendations," *Social Network Analysis and Mining*, vol. 14, no. 1, p. 171, 2024.

[8] M. H. Ribeiro, R. Ottoni, R. West, V. A. Almeida, and W. Meira Jr, "Auditing radicalization pathways on youtube," in *Proceedings of the 2020 conference on fairness, accountability, and transparency*, 2020, pp. 131–141.

[9] F. Sen, R. T. Wigand, N. Agarwal, D. Mahata, and H. Bisgin, "Identifying focal patterns in social networks," in *2012 Fourth International Conference on Computational Aspects of Social Networks (CASoN)*. IEEE, 2012, pp. 105–108.

[10] M. Alassad and N. Agarwal, "Contextualizing focal structure analysis in social networks," *Social Network Analysis and Mining*, vol. 12, no. 103, 2022.

[11] F. W. Takes and W. A. Kosters, "Identifying prominent actors in online social networks using biased random walks," in *Proceedings of the 23rd benelux conference on artificial intelligence (BNAIC)*. Citeseer, 2011, pp. 215–222.

[12] M. Sherif and C. I. Hovland, *Social Judgment: Assimilation and Contrast Effects in Communication and Attitude Change*. New Haven, CT: Yale University Press, 1961.

[13] R. E. Petty and J. T. Cacioppo, *Communication and Persuasion: Central and Peripheral Routes to Attitude Change*. New York, NY: Springer, 1986.

[14] L. Festinger, *A Theory of Cognitive Dissonance*. Stanford, CA: Stanford University Press, 1957.

[15] W. R. Fisher, "Narration as a human communication paradigm: The case of public moral argument," *Communication Monographs*, vol. 51, no. 1, pp. 1–22, 1984.

[16] M. I. Gurung, N. Agarwal, M. M. I. Bhuiyan, and D. Poudel, "Symbolic signals on instagram: how visual media shapes engagement, emotion, trust, and diffusion," *Social Network Analysis and Mining*, vol. 15, no. 1, pp. 1–16, 2025.

[17] J. Neyman, "On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection," in *Breakthroughs in statistics: Methodology and distribution*. Springer, 1992, pp. 123–150.

[18] R. Amure and N. Agarwal, "CI-FSA: Toward scalable discovery of influential groups in social networks," in *Complex Networks & Their Applications XIV*. New York, USA: Springer Nature, Dec. 2025.

[19] D. Poudel and N. Agarwal, "Keys: Keyframe extraction and yielding summaries," in *Proceedings of the 37th IEEE International Conference on Tools with Artificial Intelligence (ICTAI)*. IEEE, 2025, pp. 1–6.

[20] N. Cliff, "Dominance statistics: Ordinal analyses to answer ordinal questions." *Psychological bulletin*, vol. 114, no. 3, p. 494, 1993.