UtilitasMathematica

ISSN 0315-3681 Volume 122, 2025

Understanding Instagram's Content Recommendation: A Probabilistic Approach using Markov Chains

Shamitha Rajesh (Corresponding author),

Department of Visual Communication, Faculty of Science & Humanities, SRM Institute of Science and Technology, Kattankulathur Campus, Chengalpattu 603203, Tamil Nadu, India.

ORCID - 0009-0005-0151-5828

Email: sr9108@srmist.edu.in Ph no: +91 9150935349

Divakar Aruna Kaliaperumal, Ph.D,

Department of Journalism and Mass Communication, Faculty of Science & Humanities, SRM Institute of Science and Technology, Kattankulathur Campus, Chengalpattu 603203, Tamil Nadu, India.

ORCID - 0000-0003-2523-026X

Email: divakara1@srmist.edu.in Ph no: +91 9952910190

Abstract

Social media platforms like Instagram personalise user experience by curating content feeds based on individual preferences and behaviour. This study explores the mathematical foundation behind Instagram's personalised suggestions using a Markov Chain-based probabilistic model. By treating each user interaction as a transition between content categories (states), the system calculates the probability of the next likely interaction, refining the content shown over time. This sequential modelling not only helps platforms maximise engagement by predicting the most relevant content but also raises concerns about selective and incidental exposure, where users are consistently shown similar viewpoints. The research presents a conceptual explanation of how Markov Chains guide content flow to highlight the recursive nature of content ranking. It also emphasises the importance of balancing personalisation with content diversity. Theoretical and practical implications of the study reveal that while algorithms optimise user satisfaction, they may also limit exposure variety. By understanding the probabilistic structures, platforms can adjust their recommendation logic to introduce greater diversity, mitigating the effects of filter bubbles and echo chambers. This framework can inform future development of algorithms that not only predict but also broaden user engagement in healthier, more balanced digital spaces.

Keywords – *Instagram, Algorithm, Content Recommendation System*

1. Introduction

1.1 Overview of Instagram's recommendation system

In the digital age, social media platforms have transformed the way individuals consume content, communicate, and form opinions. Among these platforms, Instagram stands out for its highly personalised and engaging user experience, powered by sophisticated content recommendation systems. As of 2024, Instagram boasts over 2.35 billion monthly active users worldwide with over 362 million users from India [1]. Unlike traditional media, which delivers the same content to a broad audience, Instagram curates a tailored feed for each user based on their individual preferences, behaviours, and interactions on the app.

At its core, Instagram's recommendation process involves collecting data on user activity—such as likes, shares, comments, time spent on posts, and the accounts followed—and using this information to suggest content that aligns with a user's interests. This dynamic personalisation is driven by machine learning algorithms that continuously adapt to user behaviour, ensuring a constant stream of relevant content. The system not only enhances user satisfaction by keeping them engaged but also plays a significant role in amplifying visibility for creators and influencers.

Instagram's content recommendation process precisely relies on machine learning algorithms that analyse a user's interactions to deliver personalised content across different sections of the app as well. Presently, as of 2024, for the Feed and Stories, it prioritises content from followed accounts based on recent engagement. On the Explore Page, it surfaces posts similar to those a user has interacted with, even from unfamiliar accounts. Reels, the exclusive Instagram's short video feature, uses signals like watch time, shares, and replays to predict what will entertain the user. These systems are continually updated and optimised to increase user engagement, retain attention, and promote content creators by predicting and meeting individual preferences in real time.

Understanding Instagram's content recommendation process is important because it directly influences what users see, how they engage, and what trends gain visibility. These algorithms don't just reflect preferences—they shape them, reinforcing existing interests and sometimes narrowing the scope of content exposure through filter bubbles or echo chambers. Social media becomes a primary source of information and influence—especially the younger demographics among whom Instagram has penetrated deeply—these algorithms carry significant implications for public discourse, consumer behaviour, mental health, and even democratic participation. Analysing how these systems work is therefore crucial for both users and researchers to critically engage with the digital environments shaping everyday life.

1.2 Algorithmic content personalisation

The algorithmic content personalisation carefully selects and recommends content based on each user's unique behaviour and preference rather than showing all randomly available content to users. Over time, the more a person uses Instagram, the more the platform learns about the user's preference—such as the kind of posts they interact with, accounts they follow, or topics they spend time on. Based on this information, Instagram shows more similar content, creating a highly personalised experience for each user. This process is driven by recommendation algorithms—mathematical systems that decide what content to show users. Commonly used types of recommendation algorithms in social media platforms are:

1.2.1 Collaborative Filtering: This method recommends content to a user based on what similar users have liked or interacted with. For example, if User A and User B both like the same posts, and User A engages with a new type of content, that content might be recommended to User B too. The suggestions are based on the content people similar to the user prefer.

1.2.2 Content-Based Filtering: This approach focuses on the characteristics of the content itself. If a user frequently interacts with videos about fitness or nature, the algorithm will suggest other content with similar tags, topics, or formats. It doesn't rely on what other users like, but instead on the specific features of the content the user has engaged with.

1.2.3 Hybrid Models: Many platforms, including Instagram, use a combination of both collaborative and content-based filtering. This allows the algorithm to make better predictions by using a wider range of data. For instance, it can consider both the user's own preferences and trends among similar users.

1.3 Effects of personalisation

Algorithms are generally known as the signals and data that govern a platform's operation, sorting and filtering content that appears on the user's feed, amplifying information that the user would possibly engage in [2]. While algorithmic personalisation on social media platforms like YouTube, Instagram appears user-friendly and engaging on the surface, it carries significant long-term effects that are concerning [3]. One of the key outcomes of this personalisation is the forming of filter bubbles—a situation where users are repeatedly exposed to a narrow range of content based on their past behaviour and preferences. In a filter bubble, opposing viewpoints or unfamiliar topics are algorithmically filtered out, meaning users are less likely to come across content that challenges their existing beliefs or broadens their understanding of a topic.

Scholarly debate surrounds the concept of social media as public sphere that enhances political participation enabling people to come together fostering public dialogue or as an echo chamber that fosters polarisation, as researches have evidences supporting both perspectives [4]. A popular view on social media focuses on echo chambers, attributing its dynamics to highly fragmented, customised, and niche-oriented aspects enabling space for polarisation [5]. Therefore, algorithms perform as echo chambers and enable selective exposure more than the user. Unlike filter bubbles, which are about content limitation, echo chambers amplify the same type of content and viewpoints. In these digital spaces, users constantly encounter posts, opinions, and information that align with their existing beliefs. This is primarily because the content is repeated and reinforced by other users with similar views, individuals begin to perceive their viewpoint as more valid or widespread than it actually is. This creates a distorted perception of social consensus. Early media theories from the 1960's have also discussed similar concepts through theories such as cultivation theory

[6]. Though the contexts have drastically changed with the growth and development of media and technology, the potential presence of effects have been observed among users of media from early media penetration as well.

The effects of filter bubbles and echo chambers are far-reaching. First, they contribute to opinion polarisation, a process in which people's attitudes become more extreme and less flexible over time [7]. Researchers emphasise on the role of social media in enabling higher levels of political polarisation by creating echo chambers by presenting the audience with patterns of information or content that reinforces their existing political beliefs and in turn limits exposure to opposing or contradictory views [8]. Since users are rarely exposed to counterarguments or diverse perspectives, they may develop a more one-sided and emotionally charged stance on issues. A person who consistently engages with content supporting a specific stance may become increasingly resistant to considering other viewpoints, not because they reject them intentionally, but because they rarely see them. The algorithm reinforces their existing stance by continually showing them similar content.

The effects of this reinforcement can be observed on personal and societal levels. On a personal level, individuals may experience confirmation bias, where they seek out and give more weight to information that confirms their pre-existing views. Studies explained how individuals paid more attention to information that aligned with their beliefs and re-confirmed their initial beliefs. This can lead to increased certainty in their beliefs—even if those beliefs are based on misinformation or incomplete narratives [9]. Over time, such users may become distrustful of other sources of information, further deepening their isolation within the echo chamber. On a societal level, the consequences can be even more profound. Widespread polarisation can lead to divisions within communities, reduced tolerance for differing opinions, and challenges in democratic dialogue. Social media platforms, through algorithmic curation, can unintentionally foster environments where misinformation spreads more rapidly, critical thinking is discouraged, and users become more emotionally reactive and less reflective. Though algorithmic recommendations improve engagement and user satisfaction, they also carry unintended risks. By continuously reinforcing what users already believe, these systems can narrow people's worldviews, reduce exposure to diversity of thought, and contribute to social fragmentation. Recognising and understanding these effects is the first step toward building more responsible and transparent recommendation systems.

Behind the personalised experience that has alarming effects that users enjoy on Instagram lies a complex system powered by mathematics. As millions of users interact with billions of posts every day, it becomes impossible to manage content recommendations manually to recommend content to users based on their preferences. Instead, social media platforms depend on mathematical models to predict what each user is most likely to find interesting or engaging. These models help platforms make smart, data-driven decisions quickly and at scale. Probabilistic modelling is one of the most effective approaches for this purpose. In simple terms, a probabilistic model uses the idea of probability to make educated guesses about what might happen next. Instead of giving a fixed answer, it calculates the chances of different outcomes and picks the most likely one. For example, if a user often watches investment planning videos after liking business content, a probabilistic model might suggest investment content next, because there is a high chance the user will interact with it.

1.4 Role of probability

The role of probability in this system is crucial because user behaviour on platforms like Instagram is not always predictable in fixed terms—it is dynamic, influenced by mood, context, and evolving interests. Probabilistic models accommodate this uncertainty by allowing platforms to make flexible, adaptive predictions rather than rigid assumptions. These models work by assigning likelihoods to different actions—for instance, how likely a user is to click on a reel, share a story, or follow a new account—based on historical and real-time data. This helps the recommendation engine to not only personalise content more effectively but also to continuously learn and adjust its predictions as user behaviour changes.

Understanding how probability operates in these systems is important because it provides insight into the hidden mechanics shaping our digital experiences. While users may believe they are freely exploring content, the reality is that every click and scroll is being tracked and interpreted as part of a larger predictive model. These insights are then translated into recommendations that seem intuitive but are actually carefully calculated.

Understanding this process is essential to demystify how algorithms influence behaviour, preferences, and ultimately, public discourse.

This study focuses on explaining the underlying probabilistic logic used in Instagram's content recommendation process, particularly through the lens of the Markov Chain model to illustrate how the logic of probability underpins content recommendation on Instagram. By breaking down how probabilistic reasoning—through methods like Markov Chains—is used to guide user experience, the study seeks to shed light on the often-invisible framework behind digital personalisation. The study aims to conceptually break down how platforms use transition probabilities and optimisation strategies—such as maximisation of engagement likelihood and minimisation of irrelevant content—to personalise user feeds. By doing so, the research provides a theoretical understanding of how probabilistic systems guide user experience and exposure, ultimately shaping online behaviour. This explanatory approach is especially relevant in the context of growing concerns around algorithmic influence, filter bubbles, and opinion formation, and contributes to the broader discourse on transparency and accountability in digital media platforms while relying on mathematical methods.

2. Theoretical background

2.1 Markov's Chain

A Markov Chain is a stochastic (random) process that models a sequence of events where the probability of each event depends only on the state attained in the previous event [10]. This is often referred to as the memoryless property or the Markov property that means the prediction of what comes next is based only on the current situation, not the entire history of how that situation was reached. This characteristic makes Markov Chains particularly useful for modelling dynamic systems where future outcomes are influenced by present states. Formally, a process $\{X_t\}$ is said to follow a Markov Chain if:

$$P\big(X_{t+1} = s_j \,\big|\, X_t = s_i, X_{t-1} = s_k, \dots, X_0 = s_0\big) = P(X_{t+1} = s_j \mid X_t = s_i$$

This formula depicts that the probability of moving to the next state s_j depends only on the current state s_i, and not on how that state was reached.

Instagram’s algorithm can use this model to maximise the likelihood of user engagement by recommending the content with the highest transition probability. At the same time, it minimises the chances of user disengagement by avoiding low-probability transitions. In recommendation systems, Markov Chains support both personalisation and scalability, allowing platforms to manage millions of users while providing each with a unique experience. They also offer a probabilistic logic that can be fine-tuned with additional parameters such as user preferences, content similarity, and engagement patterns, making them a flexible and powerful tool in algorithmic content curation. By using real-time engagement data, the system can adapt these probabilities dynamically, making them an ideal fit for social media recommendation engines.

2.2 Probability in recommendation algorithms

In recommendation systems like Instagram’s, probability is used to estimate how likely a user is to engage with each piece of content. These estimations are based on the user's previous actions—such as what they have liked, commented on, or spent time viewing. Each potential content item and category (like a reel, post, or story that is related to fitness, fashion or travel) is assigned a probability score that reflects its chances of getting a reaction from the user. The system then ranks content based on these scores, showing the most promising content first. This method of ranking helps platforms personalise the content feed efficiently, ensuring that users are shown the most engaging and relevant material at any given moment.

Conditional probability on the other end helps the algorithm understand how different types of user interactions are related. This means the system looks at how likely a user is to interact with a specific type of content given that they’ve already interacted with another type. For example, if data shows that users who watch business-related videos often go on to view motivational content, then Instagram can use this insight to suggest motivational posts after a user watches a business video. The relationship between past and future interactions helps the platform make smarter and more relevant predictions. This technique allows the system to go beyond simple preferences and understand patterns of behaviour, improving the flow of content suggestions and increasing the likelihood of continued engagement.

3. Mathematical foundation of Instagram of Instagram's suggestion model

3.1 Markov chain representation in Instagram's algorithm

A Markov Chain consists of a set of states and transition probabilities that define the likelihood of moving from one state to another. These probabilities are organized in a transition matrix, which provides a complete representation of the system's movement over time. Instagram's recommendation system can be modelled, where each state s_i might represent a specific content category (e.g., fashion, fitness, finance). When a user interacts with one type of content (e.g., liking a fitness reel), the algorithm estimates the probability of what the user is likely to engage with next showing that the transitions between the states are driven by user interactions. These estimations are transition probabilities, and they help the system decide what content to suggest following the current engagement.

For example:

- $p\ (fitness\ \rightarrow wellness) = 0.6$
- $p\ (fitness\ \rightarrow travel) = 0.3$
- $p\ (fitness\ \rightarrow fashion) = 0.1$

3.2 User interaction as states

Let the set of all content categories be represented as states: $S = \{s_1, s_2, \dots, s_n\}$

Each interaction a user makes translates into movement from one state to another. The transition from one state to another is not random, but based on past behaviour, which is captured in a transition matrix.

3.3 State Transition Matrix

A transition probability matrix P, where each element p_{ij} represents the probability of transitioning from state s_i to state s_j. The transition matrix P is represented as:

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix}$$

Each row in the matrix corresponds to a current state, and each column represents the next possible state. The values must satisfy:

$$\sum_{j=1}^{n} p_{ij} = 1 \quad \text{for all } i$$

This indicates that from any state s_i, the total probability of transitioning to any next state must equal 1.

3.4 Transition Probabilities Based on Engagement

These probabilities are derived from real-time engagement data. If users frequently watch travel content after engaging with wellness content, the corresponding transition probability $p_{wellness,travel}$ increases. The model keeps updating these probabilities as user behavior changes.

3.5 Probability Distribution and Recursive Update

If $\pi^{(t)}$ is the vector representing the probability distribution over states at time t, the distribution at the next time step is given by:

$$\pi^{(t+1)} = \pi^{(t)}.P$$

This recursive formula allows the platform to update the expected next state based on user activity continuously.

3.6 Steady-State Probability Distribution

Over time, the system can reach a steady state where the probability distribution stabilises and doesn't change significantly with further iterations. Let this be

$$\pi = \pi\,.P$$

This steady-state vector helps determine long-term user preferences and informs how often certain content types should appear in a user's feed.

3.7 Maximisation and Minimisation in Content Suggestion

Maximisation: The intent of the platform is to show content with the highest likelihood of engagement. The system chooses the next content category s_i with the highest transition probability from the current state s_j:

$$arg\ max_j\ p_{ij}$$

This ensures that the most relevant and engaging content is ranked at the top.

Minimisation: To avoid user dissatisfaction, the system avoids transitions with low probabilities. i.e., states where p_{ij} is low, to ensure higher relevance and engagement. These are less likely to be relevant and are therefore deprioritized in the feed.

4. Implications

4.1 Theoretical implications

The study contributes to the theoretical understanding of algorithm-driven content recommendation systems by framing user interaction as a Markovian process, where future states (content recommendations) depend solely on the present state (user engagement) and not on the sequence of events that preceded it. This offers a simplified but powerful way to model and predict digital behaviour using mathematical constructs.

From a theoretical perspective, the study extends the application of probabilistic modelling to digital media studies, especially in the context of incidental and selective exposure. It connects behavioural science and algorithmic design by showing how user preferences, consciously or subconsciously formed, influence content exposure on platforms like Instagram. By doing so, it provides a framework that bridges traditional communication theories—like cultivation theory, agenda-setting and uses and gratifications—with computational approaches rooted in probability and system dynamics.

Additionally, this research rethinks the concept of exposure diversity and filter bubble theory within digital environments. The use of Markov Chains demonstrates how repeated transitions between similar content types (e.g., fitness → wellness → motivation) can reinforce certain ideologies or interests while gradually filtering out divergent viewpoints. The steady-state analysis suggests that content exposure may eventually converge around dominant themes or categories, shaped not by user intention alone but also by probabilistic maximisation and minimisation principles coded into recommendation algorithms. The study discusses the theoretical discourse by embedding probability, state transitions, and mathematical modelling into how user experience and exposure online is conceptualised. It sets the stage for future research on algorithmic bias, ideological polarisation, and engagement optimisation in social platforms, opening new interdisciplinary pathways between media studies, mathematics, and human-computer interaction.

4.2 Practical implications

On a practical level, the study offers significant insights for platform developers, marketers, and content strategists. By modelling Instagram's content recommendation as a Markov process, it becomes possible to understand the sequential nature of user behaviour, where one type of engagement often leads predictably to another. Understanding this sequence of user actions can help designers and developers optimise their algorithms for better user retention and satisfaction. By applying concepts like transition probability matrices and steady-state probabilities, platforms can ensure that content flows logically and contextually from one piece to another, creating a smoother and more engaging user journey.

Data scientists working with large datasets can leverage this model to analyse user trends over time without needing to store extensive historical sequences while keeping the effects of the exposure into account and designing user experiences that ensure diversified content exposure. Markov Chains require only the current state and transition probabilities, making them efficient tools for managing massive volumes of data while maintaining real-time adaptability. This can significantly reduce computational complexity and storage requirements, particularly important in platforms with millions of users and billions of interactions.

For marketers and content creators, understanding which transitions are most likely allows for better targeted content planning. It becomes possible to position content strategically within the algorithm's "flow" so that it appears in places where user engagement is maximised. This improves reach, visibility, and conversion rates in a more intelligent and data-driven manner.

The model's use of maximisation and minimisation principles offers a clear route to improving recommendation logic: maximising the probability of showing relevant content and minimising the likelihood of irrelevant or disengaging suggestions. This results in better content ranking, less noise in user feeds, and ultimately a more meaningful social media experience by also taking the effects of the exposure into consideration while designing the feeds.

5. Conclusion and future directions

The study provides a foundational understanding of how Instagram's content recommendation system can be modelled using probability and Markov Chains. By representing user interaction as a sequence of transitions between content types, it shows how algorithms optimise relevance through mathematical calculations—maximising engaging content and minimising irrelevant posts. While this enhances personalisation and engagement, it also raises concerns about narrowing content exposure. The theoretical and practical insights drawn from this research highlight the need for awareness of algorithmic influence on user experience. Importantly, the implications of this model extend to the design of future recommendation systems. By considering the effects of selective exposure, content prediction models can be refined to allow room for greater diversity in user feeds. This ensures not just efficiency, but also balance—supporting a healthier, more informed digital environment.

Future work can expand on this by testing real-world datasets and designing simulations that align with this framework. The limitation of this study is that the Markov Chain model assumes user behaviour follows a fixed probability structure, which may not capture long-term dependencies or evolving preferences over time. Additionally, it does not account for external factors influencing user choices. Future research can explore more dynamic models and apply simulations to better mirror real-time interactions on Instagram. Developing tools based on Markov Chains for Instagram's algorithm could help evaluate how different exposure patterns affect user experience. Such applications may also guide the creation of systems that balance personalisation with exposure to diverse content categories.

REFERENCES

[1] Statista, "Countries with the most Instagram users 2024," Statista, Feb. 15, 2024. https://www.statista.com/statistics/578364/countries-with-most-instagram-users/

[2] J. A. Tucker et al., “Social Media, Political polarization, and Political Disinformation: A review of the Scientific literature,” SSRN Electronic Journal, Jan. 2018, doi: 10.2139/ssrn.3144139.

[3] A. Banks, E. Calvo, D. Karol, and S. Telhami, “#PolarizedFeeds: Three experiments on polarization, framing, and social media,” The International Journal of Press/Politics, vol. 26, no. 3, pp. 609–634, Jul. 2020, doi: 10.1177/1940161220940964.

[4] K. Chang and J. Park, “Social media use and participation in dueling protests: the case of the 2016–2017 presidential corruption scandal in South Korea,” The International Journal of Press/Politics, vol. 26, no. 3, pp. 547–567, Jul. 2020, doi: 10.1177/1940161220940962.

[5] S. Hong and S. H. Kim, “Political polarization on twitter: Implications for the use of social media in digital governments,” Government Information Quarterly, vol. 33, no. 4, pp. 777–782, May 2016, doi: 10.1016/j.giq.2016.04.007.

[6] G. Gerbner, “Cultivation Analysis: An Overview,” Mass Communication & Society, vol. 1, no. 3–4, pp. 175–194, Jun. 1998, doi: 10.1080/15205436.1998.9677855.

[7] S. C. Rhodes, “Filter bubbles, echo chambers, and fake news: How social media conditions individuals to be less critical of political misinformation,” Political Communication, vol. 39, no. 1, pp. 1–22, May 2021, doi: 10.1080/10584609.2021.1910887.

[8] C. A. Bail et al., “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, Aug. 2018, doi: 10.1073/pnas.1804840115.

[9] M. Moussaïd, J. E. Kämmer, P. P. Analytis, and H. Neth, “Social influence and the collective dynamics of opinion formation,” PLoS ONE, vol. 8, no. 11, p. e78433, Nov. 2013, doi: 10.1371/journal.pone.0078433.

[10] W. Li and C. Zhang, “Markov Chain analysis,” in Elsevier eBooks, 2009, pp. 407–412. doi: 10.1016/b978-0-08-102295-5.10403-2