From Beats to Being

Using adolescents' listening data to identify developmental trajectories

Max Lauf
Supervisor(s): Sole Pera, Robin Ungruh
[1]EEMCS, Delft University of Technology, The Netherlands

A Thesis Submitted to EEMCS Faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering
June 22, 2025

Name of the student: Max Lauf
Final project course: CSE3000 Research Project
Thesis committee: Dr. Maria Soledad Pera, Robin Ungruh, Masoud Mansoury

Abstract

Adolescence is a pivotal period of identity formation, significantly influenced by media consumption, particularly music. While recommender sytems (RS) play a central role in shaping adolescents' music exposure, they are typically designed without considering developmental needs. We do not yet understand how identity development manifests in behavioral data over time. And without this understanding, we cannot design systems that meaningfully support it. This study addresses this critical gap by applying James Marcia's established Identity Status Theory to real-world listening behavior to get insight into how adolescents traverse and transition betweeen distinct developmental phases.

Utilizing the LFM2-b dataset, we operationalize James Marcia's dimensions of exploration and commitment through Shannon Entropy and Jensen-Shannon Divergence of genre distributions, respectively, to identify and characterize distinct behavioral phases (Diffusion, Foreclosure, Moratorium, Achievement).

Our results show that while most users briefly enter all four phases, sustained engagement is much rarer, particularly for Foreclosure and Moratorium, which appear less stable. In contrast, Diffusion and Achievement are both more frequently sustained, with Achievement emerging as the most common final phase (40.78%). A strong directional transition was observed from Moratorium to Achievement (0.67 probability), while reciprocal transitions between Diffusion and Foreclosure highlight re-evaluation among users. These patterns support the dynamic and recursive nature of adolescent identity development. By linking digital behavior to psychological theory, this work offers both developmental insights and practical implications. It lays the foundation for designing recommender systems that are not only personalized but also support healthy identity formation.

1 Introduction

Music has consistently played an important part in our lives. While past eras offered limited control over our listening experiences, recent technological advancements, particularly recommender systems (RS), have revolutionized music consumption, granting individuals access to vast musical catalogs and significantly influencing their listening choices.

However, current RS often focus on mainstream adult users, often overlooking minority groups such as children and adolescents [13]. Despite this, research highlights that "General recommender systems, even if not adapted to or designed especially for children, are widely used by children" [4]. This oversight is especially crucial during adolescence, a period during which a child undergoes important psychological changes. According to developmental psychology, adolescence is a period of intense identity formation, famously described by Erik Erikson as the "identity versus role confusion" crisis [3]. During this time, individuals actively explore their values, beliefs, social roles, and future paths.

Music plays an important role in adolescent development, serving as a medium for self-expression and exploration to create their own identity [1, 9]. Adolescents often leverage music preferences to define social boundaries, signal group affiliations, and differentiate themselves from others. As a result, RS plays a critical role during this process as it determines which music is recommended to users and thereby guides the exposure and potentially influences their identity development. If not carefully designed, these systems could limit the exposure to diverse genres, hindering crucial genre exploration. Research shows that current RS needs to include other points of evaluation besides accuracy, such as whether the tool scaffolds the child's well-being and development by prioritizing children's innate characteristics [4].

Despite the significant role music plays in adolescent development, a significant gap persists in the literature: no empirical research has yet applied established developmental theories, particularly those related to identity formation, to analyze listening data from recommender systems. This leaves a blind spot. We do not yet understand how identity development manifests in behavioral data over time. And without this understanding, we cannot design systems that meaningfully support it.
The research presented in this work aims to address this gap and deepen our understanding of adolescent development by investigating: **How do adolescents traverse and transition between distinct developmental phases as revealed through their longitudinal music listening patterns?** To answer this, we use the well-known identity theory of James Marcia, which categorizes identity formation into four distinct phasfes based on the presence or absence of exploration and commitment. By linking these phases to behavioral signals derived from listening data, we aim to identify the characteristic patterns of each phase and uncover how adolescents move between them, whether linearly or recursively. These insights contribute to our understanding of adolescent development in digital environments and lay the foundation for designing recommender systems that are not only personalized but developmentally supportive.

2 Related Work

To gain a deeper understanding of the problem and the existing research landscape, we review work on recommender systems and their impact on children. Furthermore, we explore the psychology of adolescence and identity development frameworks to establish a solid foundation for understanding user growth and change during this critical period. Finally, we examine information-theoretic tools that help track behavioral changes over time, providing insight into the available metrics that can be used to analyze evolving user preferences and get an understanding of what these metrics reveal.

2.1 Recommender systems and the impact on children

Current recommender systems have become a crucial part of our digital lives, fundamentally shaping how individuals discover and consume content across digital libraries of video, news, and music. These systems, powered by algorithms that learn from user interactions to personalize content delivery. Current RS design and research have predominantly focused on optimizing experiences for adult users, often prioritizing metrics such as click-through rates, watch times, and direct satisfaction [13]. However, current research has identified the need to include other points for evaluation, such as whether the tool scaffolds the child's well-being and development by prioritizing children's innate characteristics [4].

2.2 Adolescents as special users

Adolescence is a pivotal developmental phase characterized by a significant psychological and social transformation [3]. A crucial task during this period is the identity versus role confusion crisis, famously articulated by Erik Erikson, where individuals first explore before committing to an identity. This exploration process is nonlinear and involves continuous re-evaluation of personal commitments in response to new experiences and information.
Music, in particular, plays a central and profound role in the lives of adolescents. It serves as a tool of self-expression, social bonding, and the exploration of individual and collective

ideas [1, 9]. Adolescents actively utilize music preferences to define social boundaries, signal group affiliations, and differentiate themselves from others [10].

2.3 Marcia's Identity Status Theory

Building upon Erikson's work, James Marcia developed a more structured framework to study identity formation[7, 8]. His Identity Status Theory categorizes individuals based on two dimensions: Exploration, or Crisis, is described as the period during which an individual is actively engaged in choosing among alternatives in various life domains (e.g., occupation, ideology, values). And Commitment, the degree of personal investment in a chosen path, value, or belief. Based on the presence or absence of these two dimensions, Marcia identified four distinct identity statuses:

- Diffusion: No exploration or commitment; often characterized by disengagement or apathy.
- Foreclosure: Commitment without prior exploration; values and goals are adopted from authority figures.
- Moratorium: Active exploration without commitment; often associated with uncertainty or crisis.
- Achievement: Commitments formed after active exploration; represents a stable and self-directed identity.

It is important to note that identity development is not necessarily a linear progression from one status to the next. Individuals can move back and forth between statuses, a phenomenon described as MAMA-cycles (Moratorium-Achievement-Moratorium-Achievement) where identities may be re-explored in response to new challenges [5]. Furthermore, research suggests that the achievement of a stable, coherent identity often continues well into early adulthood. Only about half of individuals reach overall identity achievement by age 36.

2.4 Entropy and JSD

To observe identity exploration and commitment in practice, researchers often turn to behavioral data. Information metrics such as Shannon Entropy (Entropy) and Jensen-Shannon Divergence (JSD) offer useful tools for quantifying user preferences over time.
Entropy quantifies the unpredictability or diversity within a probability distribution [12]; in the context of user preferences, a higher entropy value within a user's content consumption indicates greater diversity. JSD is a symmetric and smoothed measure of the similarity between two probability distributions [6]. Unlike Kullback-Leibler Divergence, JSD is always finite and has a clear interpretation as a distance metric. JSD is frequently used to assess the stability of a users preferences over time by comparing distributions across different time points. A lower JSD value between two windows indicates a greater similarity in preferences. JSD has been previously used in a similar study to detect salient differences in genre consumption patterns of users within and between age groups [13].

3 Methodology

This study aims to investigate how adolescents traverse between the developmental phases. To achieve this, we use longitudinal music consumption data to quantify the theoretical dimensions of exploration and commitment, central to James Marcia's Identity Status Theory.

3.1 Data collection and Filtering

Our research utilizes the LFM-2b dataset, a comprehensive collection of longitudinal music events sourced from Last.fm. This dataset comprises over one billion individual listening events from more than 45 thousand distinct Last.fm users, spanning a significant period from February 2005 to March 2020. Each listening event typically includes a user ID, track ID, artist ID, and a timestamp. Additionally, the dataset is enriched with user-specific attributes such as age, gender, and country, as well as artist-specific metadata, including associated genres.

To ensure the relevance and quality of the data for studying the development of adolescent identity, we apply the following criteria:

- **Activity Threshold:** Each 3-month time window contains at least 50 listening events to be considered; windows falling short of this criterion are excluded from analysis. This threshold reduces the influence of highly variable behavior in periods of low activity, ensuring that included windows reflect more stable and representative listening patterns.
- **Age Range:** We only consider listening events when the user was between 12 and 36. To make sure that we focus on the prevalent period of identity development.
- **Period Start:** To get a full picture of the overall development, the first time window starts no later than age 14, as this is commonly seen as the beginning of middle adolescence [11].
- **Period End:** The valid time window extends to at least age 18, ensuring observation through the later phases of identity resolution in emerging adulthood [2].
- **Maximum Data Gap:** To maintain sequential integrity for analysis, the maximum gap between two consecutive time windows can not be more than one. Allowing for 1 inactive time window is a compromise to account for temporary breaks, without losing the sequential integrity of the data necessary for trajectory analysis. If a gap exceeds this, the user will be excluded.

3.2 Preprocessing

Following data collection and initial filtering, all remaining listening events were aggregated into 3-month time windows for longitudinal analysis. This granularity was chosen as it is fine enough to capture observable behavioral changes over time, yet sufficiently broad to smooth out daily or weekly fluctuations, thereby providing a more stable and representative snapshot of genre preferences within a given period.
The raw LFM-2b dataset contains more than 2,000 distinct music genres, a level of granularity that is often too fine for meaningful macro-level calculations. To mitigate this, we reduced the genre space by retaining only the listening events whose associated artists fall

within the top 40 most prevalent genres. This reduction was performed after verifying that the vast majority of original listening events were still encompassed within our refined genre set, thus preserving data integrity.

3.3 Metrics

For the empirical operationalization of Marcia's identity status theory, we employed two key metrics as proxies for the dimensions of exploration and commitment, calculated on the preprocessed 3-month time windows. Genres serve as a central measurement in this analysis because they effectively allow us to model and compare the different types of music users prefer across age groups. This approach has a precedent in recent literature, which utilized genre distributions and JSD to analyze and compare user consumption patterns in music datasets [13].

3.3.1 Exploration Value: Shannon Entropy

We quantified a user's exploration value using Shannon Entropy[12]. This metric measures the diversity of a user's genre consumption within a time window. A higher Entropy value indicates a broader, more diverse range of genres. In the context of identity development, exploration involves actively searching for meaningful alternatives [7, 8]. When a user engages with a wide variety of music genres, it directly mirrors this searching behavior. Such diverse engagement reflects an exploratory stance towards their evolving musical identity, indicating they are actively seeking out different facets of their preferences rather than settling on a narrow set. This aligns perfectly with Marcia's concept of exploration, where individuals actively investigate various roles, beliefs, and values.

3.3.2 Commitment Value: JSD

For the Commitment value we use JSD, which is a metric that quantifies the similarity between two probability distributions [6]. In our approach, we calculate JSD to assess the change in a user's genre distribution between adjacent 3-month time windows. For our study, a lower JSD value between these consecutive windows signifies a greater similarity in genre consumption patterns. This stability directly serves as our proxy for commitment. Within identity theory, commitment refers to a sustained personal investment in a chosen path, belief, or identity. When a user consistently exhibits stable music preferences over time, it indicates a settled engagement with their musical taste, a clear investment in a particular musical identity, rather than a search for new alternatives.

3.4 Identity Status Classification

Identity statuses are assigned to each user for each time window based on combinations of their calculated Exploration (Entropy) and Commitment (JSD) values. Our classification methodology is detailed below:

3.4.1 Cutoff Point Derivation

To objectively differentiate between "Low" and "High" levels of Exploration and Commitment, cutoff points were derived from the distribution of metric values across all preprocessed

3-month time windows. To maintain objectivity and ensure a data-driven approach, the median (50th percentile) was used for each metric. The median is a robust measure of central tendency that is less sensitive to outliers than the mean, making it suitable for establishing a clear dividing line within the observed data distribution. This approach allows for a balanced partitioning of the data, ensuring that "low" and "high" categories represent distinct behavioral patterns based on the typical observed values.

- $E\_Threshold$: The median Exploration value derived from all 3-month time windows.
- $C\_Threshold$: The median Commitment value derived from all 3-month time windows

3.4.2 Phase Assignment Rules

For every time window, the identity status are classified according to the following rules, using the derived E_Threshold and C_Threshold values:

- Diffusion: $Exploration < E\_Threshold$ AND $Commitment < C\_Threshold$
- Foreclosure: $Exploration < E\_Threshold$ AND $Commitment \geq C\_Threshold$
- Moratorium: $Exploration \geq E\_Threshold$ AND $Commitment < C\_Threshold$
- Achievement: $Exploration \geq E\_Threshold$ AND $Commitment \geq C\_Threshold$

3.4.3 Analysis of Phase Duration and Transitions

To analyze the duration of distinct behavioral phases, a user is considered to be in a particular phase if they maintain the same identity status for at least three consecutive time windows. If this criterion is not met, the individual time window is classified as part of non-sustained phase. This criterion is chosen to ensure that the observed phases reflect a sustained pattern of behavior rather than momentary fluctuations. While a 3-month window provides a snapshot, psychological literature suggests that identity development involves processes unfolding over longer periods [5], making a longer consistent period a more robust indicator of a stable phase.

The following analyses were performed to investigate how adolescents traverse and transition between these developmental phases. First, to understand the general overview of identity development, we analyze the overall prevalence of each identity status across all 3-month time windows for every user. This analysis provides an overall insight into which phases are most commonly observed within the adolescent population.
Next, to delve into the temporal dynamics, we focus on phase progression by identifying and analyzing the typical starting and ending phases for users within different sustained developmental periods. This aims to provide crucial insights into the common temporal sequences of developmental phases as manifested through music consumption.
Furthermore, to understand how individuals move between these states, we quantify the direct phase transitions between all identity statuses. This enables us to gain a better understanding of the common pathways users take as they move through different developmental phases, specifically identifying the most frequent shifts between Diffusion, Foreclosure, Moratorium, and Achievement.
Finally, to provide a comprehensive view of developmental paths, we identify and characterize the most common overall phase trajectories of identity development observed in the

dataset. This offers a deeper, more holistic understanding of the distinct ways different users progress through their adolescent musical identity formation.

4 Responsible Research

This research delves into the sensitive area of adolescent identity development, particularly as it manifests through music listening patterns. As such, careful consideration of ethical implications and research reproducibility has been paramount throughout this study.

4.1 Adolescents as Vulnerable Users

Our study focuses on adolescents, a pivotal developmental phase marked by significant psychological and social transformations. Music plays a crucial role during this period, influencing self-expression and identity formation. Recognizing adolescents as a vulnerable user group is critical, as RS can profoundly impact their exposure to diverse content and potentially guide their identity development. While this research does not directly interact with adolescents or collect personal data from them, the insights gained aim to inform the design of future RS that support healthy identity formation rather than hindering crucial exploration. Our work seeks to lay the foundation for systems that prioritize children's well-being and developmental needs, moving beyond mere personalization to foster growth.

4.2 Data Privacy and Anonymity

The LFM2-b dataset, utilized in this study, is a collection of longitudinal music events. While the dataset contains listening data and timestamps, it is crucial to emphasize that this research exclusively uses anonymized data. No personal identifying information of individual users was accessed or utilized. This adherence to anonymity ensures that individual users cannot be identified, mitigating privacy risks and aligning with ethical standards for data handling. Our focus remains on aggregated behavioral patterns within specified age ranges, rather than individual user profiles.

4.3 Reproducibility of Methods

It is important to note that the LFM2-b dataset, while a common resource in this field, is currently no longer publicly available online. However, many research groups in this domain are in possession of this dataset. Additionally, some researchers have shared smaller subsets of LFM-2b (e.g., for specific papers), and curated samples can be found in various git repositories. These smaller versions may suffice for replication or development purposes, even if they do not represent the full 2 billion-event dataset. Alternatively, LFM-1b, a smaller but still sizeable earlier dataset, remains publicly available and could serve as a valuable resource for similar research endeavors. By providing these detailed methodological descriptions and noting the availability of related datasets, we aim to enhance the transparency and trustworthiness of this research, facilitating verification and replication.

5 Results

This section details the characteristic behaviors observed within each identified developmental phase and gives insight into the transitions adolescents make as they navigate these

phases through their musical taste.

5.1 Overview of the phases

Our analysis contained a total of 1036 unique users, resulting in 33,598 valid 3-month observation windows after filtering. As shown in Table 1, 93.53% of the users reached Achievement in at least one (unsustained) time window while only 81.27% reached Diffusion. While having a closer look we can see that 64.47% of all the users reached all 4 phases at least once, and furthermore, all users reached at least two phases(Table 2). These findings provide insight into the diverse behavior most users have during their development.

Identity Status	**User Count**	**Percentage**
Achievement	969	93.53%
Diffusion	842	81.27%
Foreclosure	875	84.46%
Moratorium	947	91.41%

Table 1: Number and percentage of users who have entered each phase in at least one time window.

Raw (Unsustained) Phases			**Sustained Phases**		
Distinct Phases	**User Count**	**Percentage**	**Distinct Phases**	**User Count**	**Percentage**
4	668	64.47%	4	26	2.51%
3	225	21.72%	3	188	18.15%
2	143	13.80%	2	518	50.00%
1	0	0.00%	1	298	28.77%
0	0	0.00%	0	6	0.58%

Table 2: Comparison of number and percentage of users who reached different counts of distinct raw (unsustained) and sustained phases.

However, upon a closer examination of the total time windows that are actually in sustained phases it was observed that a large portion (38.91%) of these time windows did not meet the minimum three consecutive time window criterion to be classified as a sustained phase (Table 3). As a results, the number of users reaching each sustained phase drops significantly. This decline is particularly noticeable for Foreclosure and Moratorium, which drop down to only 7.00% and 11.37% respectively. The majority of users still reached Achievement and Diffusion, while Foreclosure and Moratorium are being reached significantly less (Table 4). Consequently, the number of users who reached at least three distinct sustained phases is reduced to 20.66% from 86.19% for the number of users who reached one of the raw unsustained phases (Table 2).

Identity Status	Time Window Count	Percentage
Achievement	6415	19.09%
Diffusion	7939	23.63%
Foreclosure	2352	7.00%
Moratorium	3819	11.37%
Unsustained phase	13073	38.91%

Table 3: Distribution of identity statuses across all time windows, including users who did not sustain any phase for 3 or more windows (labeled as "Unsustained phase").

Identity Status	User Count	Percentage
Achievement	614	59.27%
Diffusion	601	58.01%
Foreclosure	385	37.16%
Moratorium	402	38.80%

Table 4: Number and percentage of users who have reached the Sustained Phase.

5.2 Phase progression

To better understand how identity development unfolds over time, we analyzed the distribution of users' first and last sustained identity phases, comparing them to the overall frequency of each sustained phase across all time windows (Table 5 and Table 6).

The distribution of users' first sustained identity phases reveals important insights into the early phases of identity development. Notably, Moratorium was the first sustained phase for 26.21% of users, significantly higher than its overall frequency of 18.59% across all sustained time windows. This suggests that many users initially enter a period of active exploration. Similarly, Diffusion, was also overrepresented early in users' trajectories. In contrast, Achievement was underrepresented as a first sustained phase (26.50%) compared to its overall occurrence (31.26%), indicating that it is less common as an entry point and more likely to be reached later in the process. This interpretation is reinforced when examining users' last sustained identity phases. While Diffusion was the most common sustained status overall (38.69%), it accounted for only 30.39% of users' final phases, suggesting that users tend to move beyond it over time. Conversely, Achievement made up 40.78% of users' last sustained phases, notably higher than its overall frequency, supporting the idea that it often represented a stable or endpoint identity. Foreclosure also showed a slight increase in the final position, suggesting that for some, premature identity commitments may remain unresolved.

First Sustained Phase	User Count	Percentage	Overall Proportion
Achievement	273	26.50%	31.26%
Diffusion	381	37.00%	38.69%
Foreclosure	106	10.29%	11.46%
Moratorium	270	26.21%	18.59%
Total	1030	100.00%	100.00%

Table 5: Comparison of users' first sustained identity phase with overall phase distribution (excluding unsustained windows).

Last Sustained Phase	User Count	Percentage	Overall Proportion
Achievement	420	40.78%	31.26%
Diffusion	313	30.39%	38.69%
Foreclosure	189	18.35%	11.46%
Moratorium	108	10.49%	18.59%
Total	1030	100.00%	100.00%

Table 6: Comparison of users' last sustained identity phase with overall phase distribution (excluding unsustained windows).

This pattern is further supported by Figure 1 and Figure 2, which visualize the ages at which users first and last enter sustained identity phases. These figures show that users who first enter Diffusion or Moratorium tend to do so at earlier ages, consistent with the idea that these statuses often represent initial phases of identity development. In contrast, the upper 25th percentile for the age of last entry into a sustained phase is later for Achievement and Foreclosure, suggesting that these statuses are more commonly reached after a longer period of development and reflection.

Together, these comparisons provided a more nuanced view of identity trajectories. Users appeared to often begin with low commitment phases (Moratorium or Diffusion) and increasingly move toward the commitment ones, such as Achievement and Foreclosure.

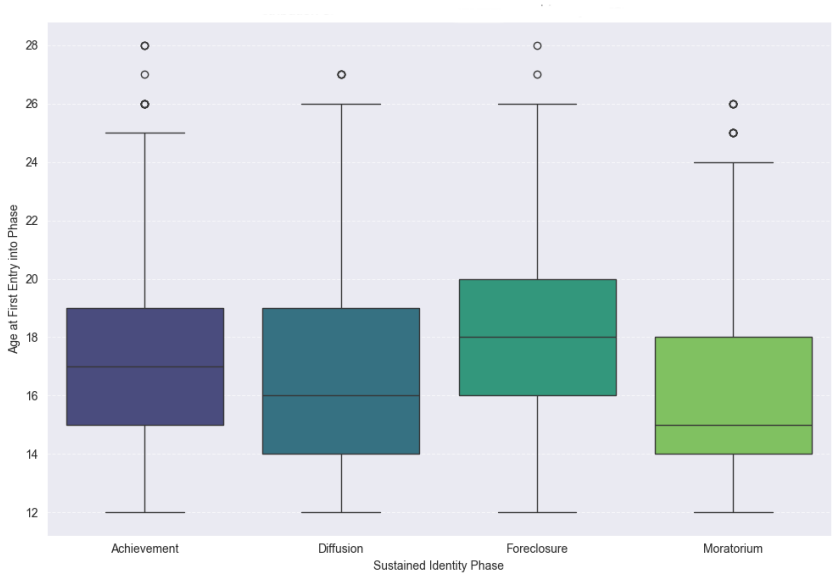

Figure 1: Distribution of user age at first entry into phase

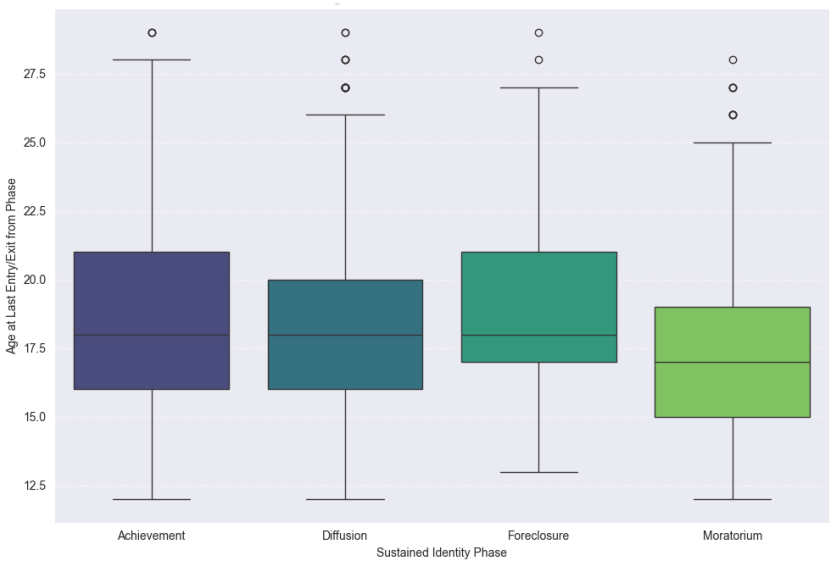

Figure 2: Distribution of user age at last entry into sustained phase

5.3 Transitions between developmental phases

To gain deeper insight into the pathways adolescents follow in their identity development, we analyzed the full transition sequences between sustained phases. This revealed several prominent patterns in how users move through different identity statuses over time.
A key finding was the predominant pattern of transitions from Moratorium to Achievement, which occurred 86 times among the 1036 users (Table 7). Not only was this transition the most frequent, but it was also supported by a high transition probability of 0.67 (Figure 3), indicating that users in a state of active exploration (Moratorium) often progressed toward a status of commitment and resolution (Achievement). This aligns with established developmental theories suggesting that exploration frequently precedes identity consolidation.

Trajectory	**Count**
Moratorium → Achievement	86
Diffusion → Foreclosure	61
Diffusion → Achievement	38
Moratorium → Diffusion	32
Diffusion → Foreclosure → Diffusion	31
Achievement → Moratorium → Achievement	26
Achievement → Diffusion	26
Foreclosure → Diffusion	22
Moratorium → Achievement → Moratorium → Achievement	20
Moratorium → Achievement → Moratorium	19

Table 7: Top 10 most common sustained phase trajectories among users.

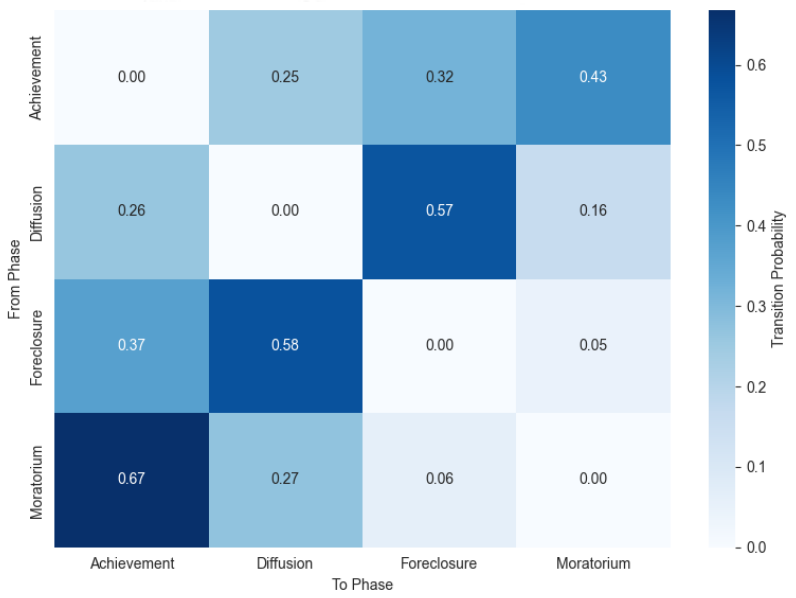

Figure 3: Transition probabilities between sustained phases

In addition to linear progressions, we observed reciprocal transitions, particularly between Diffusion and Foreclosure. The transition from Diffusion to Foreclosure occurred 61 times, with a high probability of 0.57, while the reverse transition, from Foreclosure back to Diffusion, was nearly as likely, with a probability of 0.58. This bidirectional pattern suggests a degree of instability or re-evaluation among users shifting between confusion and premature commitment.

Beyond individual transitions, we also identified repeating developmental cycles that reflect the non-linear nature of identity formation. Among the most common trajectories were: Diffusion → Foreclosure → Diffusion (31 users), Achievement → Moratorium → Achievement (26 users), and a longer loop of Moratorium → Achievement → Moratorium → Achievement, which was observered for 20 users. These cycles provide insight into the non-linear developmental progression, where users can move back and forth between statuses such as the well-known The MAMA-cycles [5].

Together, these transition patterns provide compelling evidence for both forward progression and recursive movement within identity development.

6 Discussion

The research demonstrates that longitudinal music listening patterns can be meaningfully interpreted through the lens of identity development theory. By operationalizing the concepts of exploration and commitment through Entropy and JSD respectively, we were able to identify behavioral patterns that correspond to the four identity phases described by Marcia. These patterns offer insight into how adolescents explore, commit to, and cycle through different phases of identity development in a real-world, digital environment.

6.1 Interpretation of key findings

Our findings offer a more nuanced understanding of adolescent developmental behavior through music consumption. Most users began their journeys in either the Diffusion or Moratorium phases, characterized by low commitment. This aligns with developmental psychology's view of early adolescence as a time of uncertainty and experimentation. The Moratorium phase, in particular, appeared to serve as a transitional platform: it was not only common as a first sustained phase but also had the strongest transition toward Achievement, reflecting the developmental move from searching to resolution

Achievement, in turn, emerged as the most frequent final phase, suggesting that many adolescents eventually arrive at a stable musical identity. Interestingly, while Foreclosure appeared as a potential endpoint for some users, the transitions into and out of this phase, especially the frequent back-and-forth with Diffusion, suggest that premature commitments without prior exploration may often be re-evaluated. These bidirectional patterns illustrate the non-linear, recursive nature of identity development, such as in the MAMA cycles, where users revisit exploration after initially committing.

A striking contrast emerges when comparing unsustained and sustained phases: while the majority of users pass through all four identity statuses in an unsustained manner, far fewer maintain these phases over time. This effect is particularly noticeable for Foreclosure and Moratorium, which see steep drops when applying the sustained-phase criterion of three consecutive time windows-dropping to just 7.00% and 11.37% of total windows, respectively. This suggests that these phases are less stable or more fleeting in nature. The data aligns with identity theory, which describes Moratorium as a temporary state of exploration before commitment, and Foreclosure as potentially unstable due to its lack of prior exploration. Together, the findings point to these phases being common but short-lived, reinforcing the dynamic and often recursive nature of adolescent identity development [5].

By identifying these sustained phases and their transition patterns, we uncover more than just which identities exist; we also see how they form, shift, and stabilize. The finding that Achievement is most commonly reached after Moratorium reinforces the idea that healthy identity development involves exploration before commitment. Conversely, the instability between Diffusion and Foreclosure suggests tension between uncertainty and externally imposed identities, with some users struggling to maintain these states over time.

6.2 Implications of the Research

This research provides a novel behavioral approach to studying identity development in adolescents, translating abstract developmental theory into observable patterns in real-world digital behavior. It not only complements traditional self-report methods but also opens new possibilities for developmental insights using passive data.

Practically, this work has important implications for the design of recommender systems. Understanding a user's identity phase can inform how systems should adapt their behavior. For example, a user in a Moratorium-like phase may benefit from genre-expanding recommendations to support active exploration. In contrast, a user in an Achievement-like phase may prefer content that reinforces and deepens their established preferences. Additionally, the presence of cyclical transitions; such as re-entries into Moratorium, emphasizes that RS should remain adaptive over time, supporting re-exploration rather than locking users into static patterns based on prior behavior.

By integrating awareness of identity status, recommender systems can move beyond surface-level personalization and instead foster developmental growth, helping adolescents not only

consume content but also construct identity in the process.

6.3 Limitations of the Study

Despite its contributions, this study has limitations. Our approach relies on music listening behavior as a proxy for complex psychological constructs like exploration and commitment. While our entropy and stability metrics are grounded in theory and are aligned with prior findings, they remain indirect measures of identity. Thus, we cannot directly assess users' internal self-concepts or motivations.
Furthermore, several methodological decisions may have influenced the patterns we observed. These include the choice of a 3-month time window, the 50-listening-event threshold, and the use of median-based cutoffs for phase classification. While each choice is grounded in prior work or logical reasoning, alternative parameterizations might yield different results and warrant future investigation.

7 Conclusions and Future Work

This study set out to investigate how adolescents traverse and transition between developmental phases as revealed through their longitudinal music listening data. By operationalizing James Marcia's dimensions of exploration and commitment using Shannon Entropy and Jensen-Shannon Divergence applied to the music genre distribution. We identified and characterized the behavioral phases that align with Diffusion, Foreclosure, Moratorium and Achievement.
Our results show that identity development in adolescence is both dynamic and non-linear. While most users pass through all four identity phases in some form, sustained engagement with each phase is much less common, especially for Moratorium and Foreclosure. These two phases, though frequently visited, drop sharply when applying a stricter criterion for sustained engagement (three consecutive time windows). This suggests they are inherently less stable, an insight that aligns with developmental theory. Moratorium reflects active, transitional searching, while Foreclosure may involve premature commitments that are later re-evaluated. In contrast, Achievement emerges as a common endpoint and the most stable identity phase, often reached after a sustained period of exploration.
These patterns answer our central research question by revealing clear behavioral trajectories: adolescents tend to start in lower commitment phases like Diffusion or Moratorium and, over time, transition toward more stable phases like Achievement. However, these paths are rarely linear. Reciprocal transitions and cycles-such as re-entering Moratorium after Achievement, reflect the fluid nature of identity development. Not all users move in one direction; instead, they revisit earlier phases as part of an ongoing developmental process.
This study highlights the potential of using passive digital data to gain insights into human development. It contributes to both developmental psychology and the field of recommender systems by providing a method to infer identity dynamics from behavior, rather than relying on self-report.

A key direction for future work is to design and evaluate recommender systems that leverage these findings. Systems could dynamically adapt recommendations based on the inferred identity phase of the user. For example, during periods of exploration (Moratorium), systems might emphasize genre diversity and novelty. During more committed phases (such as Achievement), systems might prioritize deepening engagement with established preferences.

Additionally, systems should remain flexible, supporting cyclical development rather than locking users into static profiles.
To further strengthen the validity of our behavioral proxies, future work could integrate self-report measures of identity status with longitudinal data. This would allow for a direct comparison and correlation between the inferred phases and the psychologically assessed identity statuses.
Finally, the methods and insights developed in this work could be applied to other digital domains such as video streaming or social media. This opens the door to developmentally-aware recommender systems across platforms, helping adolescents not just consume content, but grow through it.

References

[1] Jeffrey Jensen Arnett. Adolescents' uses of media for self-socialization. *Journal of Youth and Adolescence*, 24(5):519–533, 1995.

[2] Jeffrey Jensen Arnett. Emerging adulthood: A theory of development from the late teens through the twenties. *American Psychologist*, 55(5):469, 2000.

[3] Erik H Erikson. *Identity: Youth and Crisis*, volume 68. Norton, 1968.

[4] Emilia Gomez, Vicky Charisi, and Stephane Chaudron. Evaluating recommender systems with and for children: Towards a multi-perspective framework (2021), 2021.

[5] Jane Kroger, Monica Martinussen, and James E Marcia. Identity status change during adolescence and young adulthood: A meta-analysis. *Journal of Adolescence*, 33(5):683–698, 2010.

[6] Jianhua Lin. Divergence measures based on the shannon entropy. *IEEE Transactions on Information Theory*, 37(1):145–151, 2002.

[7] James E Marcia. Development and validation of ego-identity status. *Journal of Personality and Social Psychology*, 3(5):551, 1966.

[8] James E Marcia et al. Identity in adolescence. In *Handbook of Adolescent Psychology*, volume 9, pages 159–187. John Wiley Sons Inc, New York, 1980.

[9] Katrina McFerran. Music and adolescents. In *Life-long Engagement in Music: Benefits for Health and Well-being*, pages 95–106. Nova Science Publishers, New York, 2012.

[10] Peter J Rentfrow and Samuel D Gosling. The do re mi's of everyday life: The structure and personality correlates of music preferences. *Journal of Personality and Social Psychology*, 84(6):1236, 2003.

[11] Susan M. Sawyer et al. The age of adolescence. *Lancet Child Adolescent Health*, 2(3):223–228, 2018.

[12] Claude E Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, 1948.

[13] Robin Ungruh, Alejandro BellogÃn, and Maria Soledad Pera. The impact of mainstream-driven algorithms on recommendations for children. In *European Conference on Information Retrieval*, ECIR, pages 67–84, Cham, 2025. Springer.