Popular Music (2025), **44:** 420–426
doi:10.1017/S0261143025101220

CAMBRIDGE UNIVERSITY PRESS

MIDDLE EIGHT

Using music streaming data for research: A relational approach

Nikita Braguinski

Independent Researcher
Email: n.braguinski.uni@gmail.com

The data foundations of music streaming research

Online music streaming has emerged as a central mode of consumption in Europe and many other parts of the world, representing a distinct break from older forms of music media such as compact discs or broadcast radio. Instead of the one-to-many models that have historically dominated music dissemination – where an individual's listening habits could only be guessed at through aggregate sales figures or broadcast audience estimates – streaming has introduced a one-to-one model. This shift, which gradually gained momentum during the last two decades (Drott 2024, p. 1), hinges on platforms and devices that also allow unprecedented observation of how individual listeners engage with recorded music. Consequently, hundreds of millions of people are now individually tracked in near-real time, generating vast amounts of granular data.

Among these troves of data, one section stands out because of a stronger concentration of listening activity compared to others. For the sake of simplicity, this area of music streaming can be called the Most Frequently Streamed Recordings (MFSRs). Within the narrow confines of this short comment, it would be impossible to discuss whether these MFSRs are coextensive with what is called 'mainstream' because the conceptual work I would like to do here is concerned with the correct way to use the data, but it is obvious that these two notions are intrinsically connected. As I will argue in more detail below, whatever one hopes to distil from the data should be, in any case, seen as a property not of a recording or a user, but of the overall system (or network, if discussed through the lens of actor-network theory).

By definition, what is, however, uniquely beneficial about MFSRs is that they are described by data that is far denser and more consistent than in other areas of streaming, making the data traces that are connected to them statistically more reliable due to higher playback numbers. This should prove especially useful for any kind of computational studies of the musical processes and practices called 'mainstream' – independently of the exact understanding of this term.

In my own computational exploration of streaming datasets that are available online, I have been confronted by the situation that, despite the enormous size of some of these collections, the vast majority of tracks in the dataset receive only sporadic plays. This extremely sparse data basis makes it very challenging to reliably deduce patterns of consumption for anything but the most successful tracks, in order to see, for example, how the popularity of a certain track develops over time. Because of its mass appeal, the mainstream represents here a logical and natural research object for statistical and

© The Author(s), 2026. Published by Cambridge University Press.

https://doi.org/10.1017/S0261143025101220 Published online by Cambridge University Press

computational methods, providing a much more reliable data foundation for analysis and interpretation than non-mainstream consumption.

The promise of computational approaches

At first glance, computational and statistical studies of MFSRs seem extremely promising. It is undeniable that, simply via its previously unimaginable volume of data on music consumption, especially the more popular tracks, data-driven streaming does create a new situation for music research, one that offers specific opportunities compared to previous eras.

This move towards online streaming has led to unprecedented granularity in how scholars can (theoretically) analyse music listening habits. Rather than relying on limited sales figures or self-reported behaviour, at least the industry-affiliated researchers now have the potential to access live, micro-level data on user-item interactions, such as skip rates or repeat listens, enabling overviews of real-time trends. This finer detail is significant because it can capture patterns that could easily go unnoticed when only aggregate statistics are available. Every user of a streaming platform, at least in the European Union, can get an impression of how the platform tracks their playlists, listening times, and skipping behaviours by using personal data download services – like the 'Understanding My Data' page offered by Spotify ('Understanding my data' n.d.) – highlighting the extent to which such data can illuminate larger-scale cultural trends when information about hundreds of millions of users is combined. If it were possible to motivate a sufficiently large and representative group of users to donate their data for research purposes, such a dataset would comprise a very valuable research asset (although the danger persists that companies might try to prevent such research through claims of data ownership).

Large datasets constructed from user-level information also have the potential to serve as resources for identifying and critically discussing the processes of contemporary musical practice. Observing streaming trends at a temporal scale of days and weeks might perhaps reveal how new commercially successful releases fluctuate in popularity, while simultaneously showing which older tracks persist in public consciousness. At a longer scale, these datasets might make larger patterns of stylistic change more clearly discernible. With respect to transnational and European contexts, these data streams might allow comparisons across nations and demographics that share certain cultural backgrounds but nevertheless exhibit distinct musical preferences (e.g. Nordic countries vs. Mediterranean cultures). In this regard, existing statistics about cultural industries and activities that are being collected and processed on a European level would offer additional data points for analysis, even if the cross-referencing data on such a scale can present considerable conceptual challenges (OpenMusE 2024 Chapter 4, 'Data availability and gaps'). What could emerge from such a data-based approach is an understanding of mainstream listening as a dynamic phenomenon, shaped by local processes as well as global trends.

Computational approaches might even lead to a reexamination of established cultural theories. Big data from streaming platforms could offer a fresh lens through which to examine longstanding arguments and debates in the humanities and social sciences, such as how cultural hegemonies solidify or how canons are formed. Scholars interested in the tension between 'mainstream' and 'niche' or 'subcultural' music could draw on streaming logs to see, for example, whether niche categories indeed move into mainstream visibility over time. Researchers could also use datasets to question preconceived ideas about listener behaviour – such as whether user taste becomes narrower or less diverse through recommendation, as some cultural critics claim, or if it, conversely, potentially broadens because of the recommendation engine's goal of motivating the listener to stay on the platform.

https://doi.org/10.1017/S0261143025101220 Published online by Cambridge University Press

Beyond static entities: A relational perspective

Despite these potentially promising uses of statistical methods, I argue that the standard computational approaches applied today in research on streaming data suffer from a fundamental conceptual flaw: they treat recordings and listeners as if they would be the bearers of music-related properties that become visible in the process of listening, whereas I contend that these are the properties of the whole system, which includes the recording, the listener, the recommender engine, and many other human and non-human actors. In other words, statistical features such as the popularity of a recording do not describe the recording. They describe the interactions of all the actors in the system. Likewise, a computationally created taste profile of a listener does not, in fact, describe the listener. Rather, it describes past interactions between the listener and the rest of the system. In this sense, the property of 'mainstreamness' should also be seen as describing not recordings or users, but processes – an approach that is compatible with some of the existing scholarship on mainstream (Jost 2016). Using a different metaphor, one could also say that what one should be studying computationally should be not the products (recordings, recommendations, data traces) but the processes behind them (creating, recommending, listening, etc.).

To illustrate this, I want to use the following imaginary example: A researcher wants to gain a better understanding of the German musical mainstream. Suppose the researcher somehow gains full access to the latest data on what people in Germany listened to online – perhaps through a position at one of the streaming companies. The researcher examines playback statistics and finds that a certain musical act was the most frequently streamed artist in the past month. They then look at statistics of the self-reported age of the users who listened to this artist, and see that most of those users are between fifteen and seventeen years old. Finally, they review the names of the playlists featuring this artist and conclude that the artist belongs to a certain specific genre. Overall, they conclude that this genre represented the pinnacle of mainstream popular music in Germany during the previous month, driven by the listening activity of 15- to 17-year-olds. At first glance, it appears that the research question has been fully and objectively answered.

And yet, there are numerous strands of humanities theory that can be used to argue for a deep-rooted deficiency in such approaches. Some of these strands foreground relationality: the idea that musical practice, like other forms of activity or even reality itself, does not consist of stable, static entities but is fundamentally constituted by relations which are also constantly shifting. Actor-network theory has advanced this view since at least the 1980s, gaining wide influence in fields such as science and technology studies (Gertenbach and Laux 2019; Peuker 2010). Even earlier, Alfred Whitehead's process philosophy developed similar arguments, suggesting that it makes little sense to conceptualize the real world in terms of static entities (Ivakhiv 2018; Whitehead 1933, 1964). Scholars such as Hennion, Born, and others have actively sought to adapt these insights for the study of music (Born 2005; Hennion 2012; Piekut 2014). Likewise, the importance of seeing what happens in musical practice – and music itself – as a process rather than a static, stable object, has been emphasized by a number of influential theories, among which Christopher Small's notion of 'musicking' is perhaps the most widely known (Fuhrmann 2018; Small 1998).

Looking at the results obtained by the imaginary researcher above, it therefore becomes clear that they are deeply structured by tacit yet foundational assumptions: first, that the static and isolated audio recording itself can possess the intrinsic quality of mainstreamness; and second, that a static and isolated listener inherently possesses a certain musical taste. By contrast, a relational and processual approach insists that the static data on which the researcher relied – timestamped listening events, perhaps supplemented by automatically

https://doi.org/10.1017/S0261143025101220 Published online by Cambridge University Press

generated numerical descriptors of the audio – are not, in themselves, sufficient to reconstruct a full account of what occurs in the act of listening to a particular artist.

Therefore, reconstructive interpretations are, in my view, indispensable for developing a clearer picture of musical practice. Since not all information about musical processes is retained in their traces, it is important – even if speculatively – to attempt to excavate and recreate the original processes because they are, as argued above, that which can ultimately have any musical properties.

Network-based methodologies and humanities perspectives

The humanities have developed numerous approaches that are conducive to such speculative reconstructions. The networked and relational character of musical practice – particularly in digital contexts – has been emphasized in research. Drawing on insights from science and technology studies and actor-network theory, such methodologies foreground the relational dynamics mentioned above that occur between listeners, social groups, platform algorithms, and other entities. This approach stresses how each actor continuously shapes – and is shaped by – the constantly changing network that constitutes musical practice (Born 2005, 2019; Piekut 2014). For example, this approach would avoid interpreting a user's listening log as a mirror of their tastes, instead trying to decipher how the playback activity emerged in the interplay of the user, the streaming service, and other actors.

Scholars in musicology, sociology, and other disciplines have also long argued that raw numbers, no matter how extensive, are insufficient on their own to capture the meanings and contexts of culture in daily life (Hagen 2016; Seaver 2022, p. 97). Along similar lines, the attempts at introducing corpus-based and computational methods into traditional humanities disciplines – the digital humanities – have been criticized as 'oblivious to cultural criticism' (Liu 2012). Another notable trend that needs to be taken into account in analysis is the increasing tendency among listeners to treat music – structured through playlists and recommendations – as a tool for achieving practical goals such as relaxation or emotional regulation, effectively turning them into a 'user' of music (Drott 2025; Prey 2025).

Such approaches call for conceptual clarity in dealing with listening data, meaning that critical reflection on cultural concepts is indispensable before data collection, or at least analysis. By carefully considering how terms and notions shape research goals and metrics, scholars can remain alert to structural biases embedded in the design and interpretation of computational metrics. As an example, the researcher could try to analyse which specific meaning of 'danceability' underlies this metric before using it in their own calculations (Steinbrecher 2021, p. 414). The same applies to such contestable and vaguely defined metrics as 'valence' or 'instrumentalness'.

Access, ownership, and the politics of data

Even outside of such considerations coming from a humanities-based critique, there are reasons to treat data about listening, and its computational analysis, if not with suspicion, then at least with some sort of cautious restraint.

The opaque ways in which platforms define and track musical properties – such as the frequently used but questionable company-specific metrics in Spotify's API (a set of metrics that were previously offered by the company but have since been removed from public access) – illustrate here a deeper concern: the potential for selective data politics. In the past, music streaming companies have been described by outside researchers as secretive and even confrontational (Bonini and Gandini 2020; Eriksson *et al.* 2018). By controlling which metrics are created or disclosed, platforms are in a position to shape the research

https://doi.org/10.1017/S0261143025101220 Published online by Cambridge University Press

agenda in subtle but significant ways, leaving scholars to wonder which aspects of listener behaviour remain invisible.

Given that streaming platforms produce and control the data, tensions inevitably arise concerning conflicts of interest, corporate partnerships, and nondisclosure agreements. Researchers who collaborate with industry may try to gain privileged access to datasets, but they also risk losing critical distance if they must align with company objectives or navigate confidentiality clauses. Conversely, scholars working without such partnerships may be left with restricted or outdated datasets that limit their analyses. In an academic culture that values methodological rigor and independence, these circumstances present a dilemma: can researchers maintain a working relationship with platform owners while still preserving the integrity of their scholarship?

The fragile nature of publicly available datasets complicates matters further. For instance, LFM-1b and LFM-2b (Schedl *et al.* 2022), which quickly gained prominence in recommender system research, were eventually withdrawn. LFM-2b, published in 2022 and removed in 2024, had already become a widely used source, cited in 77 publications by early November 2025, according to Google Scholar. Despite being derived from a relatively less popular music platform (Last.fm) and, like many comparable datasets, designed primarily for recommender system development rather than music research, such resources represent some of the only data on listening behaviour that are (or in this case, were) openly available to scholars. This sudden loss exemplifies how quickly vital data repositories can vanish, making it difficult for researchers to replicate or extend prior studies.

Navigating available data

Having outlined the methodological challenges of working with online listening data, let us now briefly look at the pragmatic practicalities of doing so. Which resources are currently available for researchers not affiliated with streaming companies?

Several publicly available large datasets of listening data and company-provided APIs, such as MLHD and the Spotify API (Vigliensoni and Fujinaga 2017; 'Web API. Spotify for Developers' n.d.), offer some possibilities for examining music trends in Europe and elsewhere, especially for MFSRs. By comparing multiple sources and treating metrics with a critical eye, scholars can attempt to reconstruct the ways in which musical behaviours evolve across user groups and time periods. All of these datasets do have notable constraints. Data often spans limited periods and is processed or anonymized in ways that render cross-dataset comparisons challenging. More recent releases, such as Deezer and Yandex.Music datasets, even go so far as to anonymize tracks and artists (making content- and style-based analysis impossible), a shift driven by platforms' increasing focus on user-item interaction data in recommendation engines (Ploshkin *et al.* 2025; Tran *et al.* 2024). This necessitates a strategic balancing of old, but more complete data, such as in the case of MLHD, with newer, but more limited sources, like Deezer and Yandex.Music.

Conclusion: Towards methodological pluralism

To conclude, I want to stress that the main call here is to adopt a methodologically diverse approach that harnesses the potential of big data but remains grounded in nuanced theoretical frameworks about the dynamic processes in music, technology, culture, and society. Rather than seeking to delegitimize or oppose computational paradigms, my aim is to broaden the analyst's toolkit by emphasizing the added value of relational, processual, interpretive, and other humanities-inspired approaches. An approach like this neither dismisses online streaming data outright nor accepts the statistical results uncritically,

https://doi.org/10.1017/S0261143025101220 Published online by Cambridge University Press

but rather treats them as components in a broader toolbox for understanding music as a process.

Actor-network theorist Annemarie Mol offers an argument about research methods that is particularly relevant here. She argues that different methods enact – literally bring into being – different versions of the research object (Mol 1999). Looking at musical data, we can see how different methods construct such different objects of study. A purely statistical method constructs a musical practice made up of objects with fixed features (recordings), consumed by listeners imagined as having objective and stable profiles. While this can be useful for addressing certain questions, it also risks overstating both the interpretive value and the meaningfulness of the results.

By contrast, a processual, relational, and interpretive method – which, admittedly, still awaits fuller development and is currently more a possibility than a practice – would construct a musical world composed of interacting actors and temporally unfolding processes, with constantly changing features attributed both to the music and the listeners. In my view, such a method holds considerable promise, as it offers the potential to more fully capture the dynamic realities of contemporary musical life.

References

Bonini, T. and Gandini, A. 2020. 'The field as a black box: ethnographic research in the age of platforms', *Social Media + Society*, (October–December), pp. 1–10.

Born, G. 2005. 'On musical mediation: ontology, technology and creativity', *Twentieth-Century Music*, 2/1, pp. 7–36.

Born, G. 2019. 'On nonhuman sound – sound as relation', in *Sound Objects*, ed. J. A. Steintrager and R. Chow (Durham, Duke University Press), pp. 185–207.

Drott, E. 2024. *Streaming Music, Streaming Capital* (Durham, Duke University Press).

Drott, E. 2025. 'Historicizing the music user', paper presented at *Audiences*, Publics, Experience*:* Rethinking Music Reception, London.

Eriksson, M., Fleischer, R., Johansson, A., Snickars, P., and Vonderau, P. 2018. *Spotify Teardown: Inside the Black Box of Streaming Music* (Cambridge, MA, MIT Press).

Fuhrmann, W. 2018. 'Musicking', in *Handbuch Sound: Geschichte – Begriffe – Ansätze*, ed. D. Morat and H. Ziemer (Stuttgart, J.B. Metzler), pp. 63–6.

Gertenbach, L. and Laux, H. 2019. 'Die Akteur-Netzwerk-Theorie als Entwurf einer neuen Soziologie', in *Zur Aktualität von Bruno Latour* (Wiesbaden, Springer Fachmedien Wiesbaden), pp. 87–142.

Hagen, A. N. 2016. 'Music streaming the everyday life', in *Networked Music Cultures*, ed. R. Nowak and A. Whelan (London, Palgrave Macmillan), pp. 227–45.

Hennion, A. 2012. 'Music and mediation: toward a new sociology of music', in *The Cultural Study of Music: A Critical Introduction*, 2nd edition, ed. M. Clayton, R. Middleton and T. Herbert (New York, Routledge), pp. 249–60.

Ivakhiv, A. J. 2018. 'Contemporary process-relational thought: a primer', in *Shadowing the Anthropocene: Eco-Realism for Turbulent Times* (Earth, Milky Way, Punctum Books), pp. 231–40.

Jost, C. 2016. 'Musikalischer Mainstream. Aufgaben, Konzepte und Methoden zu seiner Erforschung', *POP. Kultur und Kritik*, 5, pp. 152–72.

Liu, A. 2012. 'The state of the digital humanities: a report and a critique', *Arts and Humanities in Higher Education*, 11/1-2, pp. 8–41.

Mol, A. 1999. 'Ontological politics. A word and some questions', *The Sociological Review*, 47/S1, pp. 74–89.

OpenMusE. 2024. *D2.1 v2.0 – Music Diversity and Circulation: Novel Data Collection Methods and Indicators, OpenMusE. An open, scalable data-to-policy pipeline for European music ecosystems.*

Peuker, B. 2010. 'Akteur-Netzwerk-Theorie (ANT)', in *Handbuch Netzwerkforschung*, 1st edition, ed. C. Stegbauer and R. Häussling (Wiesbaden, VS Verlag für Sozialwissenschaften), pp. 325–35.

Piekut, B. 2014. 'Actor-networks in music history: clarifications and critiques', *Twentieth-Century Music*, 11/2, pp. 191–215.

Ploshkin, A., Tytskiy, V., Pismenny, A., ... Savushkin, N. 2025. 'Yambda-5B – a large-scale multi-modal dataset for ranking and retrieval', arXiv, doi:10.48550/arXiv.2505.22238

Prey, R. 2025. 'Aspirational self-labour and listening practices amongst young Brazilians', paper presented at *Audiences*, Publics, Experience*:* Rethinking Music Reception, London.

https://doi.org/10.1017/S0261143025101220 Published online by Cambridge University Press

Schedl, M., Brandl, S., Lesota, O., Parada-Cabaleiro, E., Penz, D., and Rekabsaz, N. 2022. 'LFM-2b: a dataset of enriched music listening events for recommender systems research and fairness analysis', in *ACM SIGIR Conference on Human Information Interaction and Retrieval* (Regensburg, ACM), pp. 337–41.

Seaver, N. 2022. *Computing Taste: Algorithms and the Makers of Music Recommendation* (Chicago, University of Chicago Press).

Small, C. 1998. *Musicking: The Meanings of Performing and Listening* (Hanover, University Press of New England).

Steinbrecher, B. 2021. 'Mainstream popular music research: a musical update', *Popular Music*, 40/3-4, pp. 406–27.

Tran, V.-A., Salha-Galvan, G., Sguerra, B., and Hennequin, R. 2024. 'Transformers meet ACT-R: repeat-aware and sequential listening session recommendation', in *18th ACM Conference on Recommender Systems* (Bari, ACM), pp. 486–96.

Understanding My Data. n.d. Retrieved 18 February 2025, from https://support.spotify.com/us/article/understanding-my-data/

Vigliensoni, G. and Fujinaga, I. 2017. 'The music listening histories dataset', in *Proceedings of the 18th International Society for Music Information Retrieval Conference*, pp. 96–102.

Web API. Spotify for Developers. n.d. Retrieved 18 February 2025, from https://developer.spotify.com/documentation/web-api

Whitehead, A. N. 1933. 'Relativity', in *Science and the Modern World* (Cambridge, Cambridge University Press), pp. 142–60.

Whitehead, A. N. 1964. 'Theories of the bifurcation of nature', in *The Concept of Nature. The Tarner Lectures Delivered in Trinity College, November*, 1919 (Cambridge, Cambridge University Press), pp. 26–48.

Cite this article: Braguinski, N. 2025. 'Using music streaming data for research: A relational approach'. *Popular Music* 44, no. 4: 420–426. https://doi.org/10.1017/S0261143025101220

https://doi.org/10.1017/S0261143025101220 Published online by Cambridge University Press