Call for Applications for a CNRS PhD fellowship (3 years), starting from October 1st, 2024 // Spatial analysis and modeling of music streaming dynamics

As part of the MELODYNAMICS project funded by the CNRS Interdisciplinary Mission (see detailed description of the project page 3), we offer a doctoral contract (3 years) starting from October 1st, 2024. The doctoral candidate will work on the spatial analysis and modeling of music streaming dynamics, applying methods from statistical physics of complex systems and machine learning. The doctoral student will be based at Géographie-cités and Deezer and will divide his or her working time between the two locations. He/She will be co-advised by Marc Barthelemy (Institute of Theoretical Physics), Thomas Louail (Géographie-cités), and Manuel Moussallam (Deezer).

Description of the thesis topic. Do new songs spread like epidemics, and do we observe spatial diffusion effects? Does the size of the city we live in influence the music we listen to? How do recommendations and network effects influence listening habits and the discoverability of songs from different geographical origins? The rise of streaming platforms, which are accessible in almost every country, offering standardized interfaces and collecting highly detailed anonymized streaming data, allows for addressing such questions empirically, and studying this important aspect of human behavior. The doctoral student will investigate the spatial dimension of listening habits and preferences, the influence and interactions of different discovery channels (platform recommendations, social networks of individuals, living environment), as well as the spatial propagation of new music on a global scale. Identifying stylized facts, building simplified models, and applying machine learning techniques will be the tools used.

Keywords: Spatial dynamics of music listening; Spatial propagation of digital content; Spatial diffusion models; Modeling complex social systems; Discoverability.

Supervision. The thesis will be hosted at the Doctoral School ‘Physique en Île-de-France’ (Physics in the Île-de-France region, PIF) and will be supervised by Marc Barthelemy (IPHT), with co-supervision from Thomas Louail (CNRS) and Manuel Moussallam (Deezer). The three co-supervisors are accustomed to conducting interdisciplinary research spanning social sciences, computer science, and statistical physics of complex systems.

Work Environment. It is expected that the PhD candidate will split her or his time between two research teams: Géographie-cités, a CNRS research unit located on the Condorcet Campus in Aubervilliers, and Deezer, whose offices are rue de Calais in Paris. He/she will have secure access to Deezer streaming databases, and scientific computing platforms at CNRS and Deezer. Géographie-cités comprises around sixty permanent members (researchers, teacher-researchers, and IT staff), as well as approximately 90 doctoral students, post-docs, and contractual IT staff. The Condorcet Campus is a social science research campus that opened in September 2019, hosting eleven research institutions and universities, over sixty research units, and several thousands of people. In addition to her/his individual research, the doctoral candidate will contribute, during the first two years of their PhD, to the organization of a monthly seminar dedicated to computational and data-driven approaches of music listening dynamics.

Expected qualifications and skills. We seek candidates holding a master's degree in physics, computer science, or applied mathematics, with a strong interest in computational social science and interdisciplinary work, along with a first experience of this type of collaboration (not mandatory). The position requires autonomy and skills in data analysis, programing, scientific computing, and modeling, established by previous personal projects or contributions to collective projects.

Timeline and remuneration. The CNRS doctoral contract will commence in October 2024, for a duration of thirty-six months. The salary for a CNRS doctoral contract is approximately 2200E gross monthly.

Application. Candidates should send a detailed CV, and a cover letter by **June 21st** to the following address: records@parisgeo.cnrs.fr
The cover letter should include a 1 up to 2-page research proposal, in line with the project ; along with a transcript of student records, and the names and email addresses of two references.

! Important !
To be considered eligible by CNRS HR, application files must also be submitted through the CNRS job portal, by responding to the offer available at the following link:
https://emploi.cnrs.fr/Offres/Doctorant/UMR8504-THOLOU-005/Default.aspx?lang=EN

Detailed Description of the MELODYNAMICS project

Streaming platforms have become the primary mode of access to recorded music and podcasts in many countries, particularly in France (IFPI Reports 2023). Individual streaming history data, which precisely record the actual music consumption of hundreds of millions of platform users over several years, are collected and stored by these platforms to calculate artist remuneration and sustain algorithmic recommendations. These data constitute a valuable resource for scientific research on various subjects related to music listening and more broadly digital content consumption: the diversity of platform usage and individual navigation behaviors within their almost illimited catalogs (Zhang et al. 2013; Louail and Barthelemy, 2017); the social diversity of listening habits and musical tastes (Renisio et al. 2024); the influence of recommendation systems on discoverability and the diversity of content users are exposed to (Beuscart, Coavoux, and Maillard 2019; Villermet et al., 2021).

One of the most important yet quantitatively poorly understood questions is the effect of geographical space on people's music listening and preferences, and on the global circulation of new music. Whether it's new songs, artists, or musical genres, in an era where vast repertoires of music (around 100 million tracks) are available online simultaneously and at the same cost for everyone, how important is the geographical space in the propagation of new music? The geography of music remains a relatively unstructured field, particularly its quantitative and statistical branch, despite attempts in various places and at different times to advocate for its development (Kong 1995, Carney 1998, Raibaud 2006). Music listening and production are highly geographically marked phenomena. This might explain why specialists in network science and/or spatial diffusion of innovation have occasionally studied the phenomenon of local music scenes (Klement and Strambach 2019). The resurgence offered by access to individual streaming data has generated a new wave of studies, mostly stemming from scientific communities working on web dynamics, social networks, and online behaviors (Lee and Cunningham 2012, Way et al. 2019 and 2020), and conducted in the research departments of streaming companies.

Several studies published in recent years use streaming data to measure listening differences between countries, confirming both national specificities (Bello and Garcia 2021) and the effect of geographical proximity on listening preferences (Way, Garcia-Gathright, and Cramer, 2020; Cura and Maisonobe 2023). Other preliminary and applied studies have been conducted by platforms to assess the interest of recommendation features that explicitly integrate space (for example, recommending local artists from the user's residential area). These works are primarily descriptive, and when they attempt to model, the approach followed is machine learning, which yields prediction gains but sacrifices a deeper understanding of underlying processes generating the observed dynamics. We believe analyzing streaming data using statistical physics methods would shed new light on these empirical results.

Methods. So far, the propagation of new digital content has been modeled as a complex diffusion phenomenon (Centola 2010), where the main vector of information circulation is peer-to-peer human interaction, where individuals discover a new content because other individual have shared it with them, and the adoption of this content depends on the frequency with which the individual is exposed to it. Inspired by epidemiology, these studies have examined the links between the reach and speed of content propagation on one hand, and the topology of the social network on the other, highlighting several diffusion classes based on the type of network in which the information circulates. The consumption data of

digital content and user interaction collected by internet platforms - especially streaming ones - increasingly allow precise measurement and empirical study of these contagion phenomena (Babul et al. 2024). We will model the propagation of musical novelties within this simplified framework of compartments (where each individual is in a state like "susceptible," "infected," etc., see for example Barrat et al., 2008), allowing us to better understand the effects of social network structure on the propagation of musical novelty, but also serving as a starting point for modeling the influence of geographical space on this propagation.

Furthermore, in the case of music, many discovery channels coexist at the individual level, and it is difficult to quantify and prioritize their importance: one can discover new music because a friend mentioned it, heard it on the radio, in a movie, an advertisement, or at a concert; it was recommended by a platform's algorithm, or by another media. Some of these discovery factors are geographic, and it is reasonable to think that the place of residence, for example, may influence content discovery: an individual living in a large city like London or Paris, which concentrate many more cultural facilities than other cities (concert halls), and where people from all over the world live (with varied musical tastes), will be more frequently exposed to diversified cultural information (via advertisements in public transportation, for example), and more inclined to discover new content. A musical novelty thus spreads through a (social) network, but many "off-network" effects exist, and it will be necessary to integrate this data into the models.

The goal of the PhD thesis will be to study a number of collective dynamics of music streaming using a typical approach of statistical physics when studying complex social systems, using for this purpose the large volumes of data collected by Deezer, especially in the RECORDS project (https://records.huma-num.fr). Close collaboration with Deezer's research team will allow for comparison of machine learning approaches and models applied to streaming data (Afchar et al. 2022) with more "traditional" and parsimonious modeling approaches, mainly derived from statistical physics or computer science. Existing methods will also be applied and extended, or ad-hoc methods developed, as has been done in previous works by the co-supervisors on space-population correlations in cities (Louail and Barthelemy 2022; Louail et al. 2015), thus enabling the synthesis of the vast amount of available data and extracting useful information to understand the phenomenon under study. Collaboration with Deezer will also allow for large-scale experiments, directly in the mobile application and involving tens to hundreds of thousands of participants, with, for example, A/B testing procedures to evaluate hypotheses or realistically measure propagation dynamics, in the spirit of pioneering "social" experiments conducted by Milgram (Milgram et al. 1992) such as the measurement of "social distance" between two individuals (A and D do not know each other, but A knows B who knows C who knows D, and the distance between A and D is therefore equal to 3). They may also draw inspiration from online experiments conducted by (Salganik, Dodds, and Watts 2006) in the case of music.

Work Program and Expected Results. Specifically, we propose to study the influence of space at different scales, both in individual streaming choices and in the geographical circulation of works.

Understanding the role of listening context. Does the listening context (location, activity performed while listening, people present) determine the choice of music to be listened to? Although intuitive, this hypothesis has been little corroborated by quantitative studies based on detailed and continuous individual listening data, over several weeks or months.

Automatic recommendation systems have so far integrated very little of this contextual information to choose the content to recommend. Data collected by platforms contain low-level information (timestamp, IP address, hardware, platform features used) that may be useful for inferring, through statistical learning, higher-level listening contexts (for example, "at home," "on the go," "at a party", etc.). This information could be used to compare intra-individual and inter-individual variability of streaming content according to contexts, and suggest useful models to link context, music listened to, and listener satisfaction.

Understanding the role of place of residence. When streaming data is spatially aggregated and the most listened-to musical genres and artists are calculated in a given area, regional differences appear on the map (Way et al. 2021, Bello and Garcia 2021). In addition to the relation between social position and music listened to (Renisio et al. 2024) - the social organization of space inducing a spatial organization of listening - other factors can explain the observed spatial differences, including the existence of local music genres and scenes, and the presence of cultural infrastructures. At which spatial scales is location an explanatory factor for listening? To what extent do local musical traditions and genres and subgenres remain influential on platforms that give access to global catalogs?

Understanding the spatial propagation of new music. It is easy to imagine that a new album by Taylor Swift does not spread in the same way as the new album by a local band that is mostly know in its local area. But between these two extreme cases, what are the different modes of propagation? Among the many factors influencing how users discover content, what is the importance of their place of residence, the size of the city they live in? Recent results seem to indicate that musical diversity consumed is greater in large cities (Lee, Hennequin, and Moussallam 2023), especially at the level of collective diversity. But these studies do not propose models that allow us to link these spatially aggregated diversity measures to underlying processes of music propagation. We will attempt to link these aggregated measures of listening diversity in each spatial unit to propagation and adoption processes, that explicitly take spatial factors into account.

References

(Afchar et al. 2022) Afchar D, Melchiorre A, Schedl M, Hennequin R, Epure E et Moussallam M (2022) 'Explainability in Music Recommender Systems', *AI Magazine*, 43(2), pp. 190–208. Available at: https://doi.org/10.1002/aaai.12056

(Babul et al. 2024) Babul, S. et al. (2024) 'Link Me Baby One More Time: Social Music Discovery on Spotify'. *arXiv*. Available at: https://doi.org/10.48550/arXiv.2401.08818

(Barrat et al. 2008) Barrat A, Barthelemy M, Vespignani A. Dynamical processes on complex networks. Cambridge university press; 2008 Oct 23.

(Bello et Garcia 2021) Bello, P. and Garcia, D. (2021) 'Cultural Divergence in popular music: the increasing diversity of music consumption on Spotify across countries', *Humanities and Social Sciences Communications*, 8(1), pp. 1–8. Available at: https://doi.org/10.1057/s41599-021-00855-1

(Beuscart, Coavoux, et Maillard, 2019) 'Les algorithmes de recommandation musicale et

l'autonomie de l'auditeur', *Réseaux*, 1(213), pp. 17–47. Available at https://hal.science/halshs-01639788/

(Carney 1998) Carney, George. 1998. « Music Geography ». *Journal of Cultural Geography* 18 (1): 1-10. https://doi.org/10.1080/08873639809478309

(Centola, 2010) 'The Spread of Behavior in an Online Social Network Experiment', *Science*, 329(5996), pp. 1194–1197. Available at: https://doi.org/10.1126/science.1185231

(Cura et Maisonobe 2023) Cura R et Maisonobe M (2023). 'Geography of Musical Listening. Variation in Space and Time'. In *Proceedings of ECTQG 2023*. Porto, Portugal.
Available at https://ucpages.uc.pt/site/assets/files/1249198/ectqg_2023_proceedings_final.pdf

(Klement et Strambach 2019) Klement B and Strambach S (2019) 'Innovation in Creative Industries: Does (Related) Variety Matter for the Creativity of Urban Music Scenes?', *Economic Geography*, 95(4), pp. 385–417. Available at: https://doi.org/10.1080/00130095.2018.1549944

(Kong 1995) Kong, Lily. 1995. « Popular music in geographical analyses ». *Progress in Human Geography* 19 (2): 183-98. https://doi.org/10.1177/030913259501900202

(Lee et Cunningham 2012) Lee C and Cunningham P (2012) 'The Geographic Flow of Music', in *Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012)*. Washington, DC, USA: IEEE Computer Society (ASONAM '12), pp. 691–695.
Available at: https://doi.org/10.1109/ASONAM.2012.237

(Lee, Hennequin et Moussallam 2023) 'Understanding Individual and Collective Diversity of Cultural Consumption through Large-Scale Music Listening Events', in *CEUR Workshop Proceedings. CHR 2023: Computational Humanities Research Conference*, Paris, France. Available at https://ceur-ws.org/Vol-3558/paper3539.pdf

(Louail et al. 2015) Louail, T. et al. (2015) 'Uncovering the spatial structure of mobility networks', *Nature Communications*, 6, p. 6007. Available at: https://doi.org/10.1038/ncomms7007

(Louail et Barthelemy 2017) Headphones on the wire – Statistical patterns of music listening practices. *arXiv*:1704.05815 [physics]. Available at https://cea.hal.science/cea-01626088/document

(Louail et Barthelemy 2022) Louail T et Barthelemy M. (2022) A dominance tree approach to systems of cities, *Computers, Environment and Urban Systems*, 97, p. 101856.
Available at: https://doi.org/10.1016/j.compenvurbsys.2022.101856

(Matrosova et al. 2024) Matrosova K, Moussallam M, Louail T and Bodini O. (2024) 'Depict or Discern? Fingerprinting Musical Taste from Explicit Preferences', *Transactions of the*

International Society for Music Information Retrieval (TISMIR), 7(1), pp. 15–29. Available at: https://doi.org/10.5334/tismir.158

(Milgram et al. 1992) Milgram S, Sabini JE, Silver ME. The individual in a social world: Essays and experiments. Mcgraw-Hill Book Company; 1992.

(Raibaud 2006) Raibaud Y (2006) « « Comment la musique vient-elle au territoire ? » ». *Volume !. La revue des musiques populaires*, no 5 : 2 (septembre): 205-9. https://doi.org/10.4000/volume.586

(Rapports IFPI 2023) IFPI's Global Music Report 2023. Available for download at https://globalmusicreport.ifpi.org/

(Renisio et al. 2024) Renisio Y, Beaumont A, Beuscart JS, Coavoux S, Coulangeon P, Cura R, Le Bigot B, Moussallam M, Roth C et Louail T (2024) Integrating digital data into mixed methods designs. An application to the study of online music listening using stream, survey and interview data collected from the same people. Actuellement en cours d'évaluation à la *Revue Française de Sociologie* ; manuscrit disponible sur demande.

(Rosati et al. 2021) Rosati D.P. et al. (2021) 'Modelling song popularity as a contagious process', *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 477(2253), p. 20210457.
Available at: https://doi.org/10.1098/rspa.2021.0457

(Salganik, Dodds et Watts 2006) Salganik MJ, Dodds PS and Watts DJ (2006) 'Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market', *Science*, 311(5762), pp. 854–856.
Available at: https://doi.org/10.1126/science.1121066

(Villermet et al. 2021) Quentin Villermet, Jérémie Poiroux, Manuel Moussallam, Thomas Louail and Camille Roth (2021) Follow the guides: disentangling human and algorithmic curation in online music consumption. *Proc. 15th ACM Conference on Recommender Systems, RecSys'21*, pp. 380-389. https://dl.acm.org/doi/10.1145/3460231.3474269

(Way et al. 2019) Way, Samuel F., Santiago Gil, Ian Anderson, et Aaron Clauset. 2019. « Environmental changes and the dynamics of musical identity ». In *Proceedings of the International AAAI Conference on Web and Social Media*, 13:527-36. https://ojs.aaai.org/index.php/ICWSM/article/view/3250.

(Way, Garcia-Gathright et Cramer, 2020) Way, S.F., Garcia-Gathright, J. and Cramer, H. (2020) 'Local Trends in Global Music Streaming', *Proceedings of the International AAAI Conference on Web and Social Media*, 14, pp. 705–714. Available for download at https://ojs.aaai.org/index.php/ICWSM/article/view/7336

(Zhang et al. 2013) 'Understanding user behavior in Spotify', in *2013 Proceedings IEEE INFOCOM*. 2013 Proceedings IEEE INFOCOM, pp. 220–224. Available at: https://doi.org/10.1109/INFCOM.2013.6566767.