Diffusion Models for Recommendation: Reproducibility and Conceptual Mismatch

Michael Benigni[1], Maurizio Ferrari Dacrema[1] and Dietmar Jannach[2]

[1]Politecnico di Milano, Italy
[2]University of Klagenfurt, Austria

Abstract

Recent studies have applied Denoising Diffusion Probabilistic Models (DDPMs) to recommender systems, reporting notable improvements. However, several reproducibility studies have shown that claims asserting the superiority of new methods are frequently not substantiated by rigorous evidence, as they often rely on non-reproducible experimental protocols, weak or untuned baselines, and questionable evaluation practices. This extended abstract presents key findings from the manuscript "Diffusion Recommender Models and the Illusion of Progress: A Concerning Study of Reproducibility and a Conceptual Mismatch" which investigates whether the reported advancements of diffusion-based models in recommendation are supported by rigorous and reproducible experimental evaluation.

The study re-executes the experiments of four DDPM-based models presented at SIGIR 2023 and 2024, revealing substantial methodological issues and limited reproducibility. In addition, it highlights a conceptual mismatch between the generative nature of DDPMs and the deterministic requirements of offline evaluation, underscoring the need for a reconsideration of evaluation procedures for generative models.

Keywords

Recommender Systems, Reproducibility, Diffusion Models, Evaluation

1. Introduction

With the emergence of advanced generative architectures, the recommender systems community has made significant efforts to apply such models to the field. In addition to transformer-based architectures [1, 2], which are state-of-the-art in natural language processing, Denoising Diffusion Probabilistic Models (DDPMs) [3, 4] have also gained attention in recommendation research as generative models. Originally developed to model and sample from complex distributions, DDPMs have shown remarkable results in image and video synthesis. Due to their strong modeling capacity and denoising properties [4], several works have adapted this architecture for collaborative filtering in recommender systems, claiming superior accuracy compared to traditional baselines [5, 6, 7, 8, 9]. Many of these contributions have appeared at top-tier venues such as ACM SIGIR 2023 and 2024, reinforcing the perception that DDPMs represent a promising direction for top-n recommendation.

However, a decade of research has repeatedly shown that many claimed improvements in recommendation effectiveness are often illusory, stemming from comparisons with weak or poorly tuned baselines and flawed evaluation protocols [10, 11, 12, 13, 14, 15]. In some cases, even simple models such as k-nearest neighbors [11] or matrix factorization [14, 15], when properly tuned, outperform modern deep learning architectures. These observations raise the critical question of whether recent advances in diffusion-based recommender systems truly reflect meaningful progress.

This extended abstract summarizes the work in [16], which addresses this question by examining the reproducibility and effectiveness of four recent diffusion-based recommendation models from SIGIR 2023 and 2024 [5, 6, 7, 8, 9]. The analysis is threefold: (i) assessing the reproducibility of reported results by re-executing experiments, (ii) comparing these models against a suite of strong, well-tuned baselines

IIR2025: 15th Italian Information Retrieval Workshop, 3th - 5th September 2025, Cagliari, Italy
michael.benigni@polimi.it (M. Benigni); maurizio.ferrari@polimi.it (M. Ferrari Dacrema); dietmar.jannach@aau.at (D. Jannach)
0009-0002-2810-5800 (M. Benigni); 0000-0001-7103-2788 (M. Ferrari Dacrema); 0000-0002-4698-8507 (D. Jannach)

from different model families, and (iii) reflecting on the conceptual suitability of DDPMs for top-n recommendation tasks.

The results of this analysis are concerning. Reproducibility remains elusive in many cases, often due to incomplete experimental descriptions and high variability in results. Furthermore, comparisons with well-tuned baselines reveal that the original experiments were not conducted under challenging conditions, casting doubt on the validity of the claimed improvements. Finally, a fundamental conceptual gap is highlighted between the probabilistic generative nature of diffusion models and the deterministic requirements of top-n evaluation. These observations motivate a critical reassessment of current evaluation practices and call for renewed scientific rigor and transparency in the field.

2. Methodology

2.1. Papers Selection

The analysis in [16] covers four articles, each introducing a different algorithm: DiffRec [6], CF-Diff [7], GiffCF [8], and DDRM [9]. These papers were selected based on three criteria: (i) they were presented in the *Diffusion in RecSys* session at SIGIR 2024, (ii) they propose a new algorithm for the top-k recommendation problem, (iii) the algorithm employs diffusion-based techniques. Additionally, DiffRec [6], published at SIGIR 2023, was included as it laid the foundation for the subsequent diffusion-based recommendation algorithms analyzed.

2.2. Reproducibility

The reproducibility protocol adopted in [16] consists of the following steps:

- **Artifact Verification:** The availability and consistency of required artifacts (source code, datasets, best hyperparameters values and experimental details) are checked. This step is essential for replicating the experiments under the original conditions.

- **Experimental Re-execution:** Once artifacts are collected, experiments are re-run. Although the original model code is used, it is integrated into the framework from [11] to ensure consistent evaluation and early-stopping execution across experiments. Each DDPM model is trained using the best hyperparameters values provided by the original artifacts, without any additional tuning.

- **Reproducibility Assessment:** In this extended abstract, reproducibility is intended as the ability to obtain numerical results that are sufficiently close to the original ones. However, due to the inherent stochasticity of diffusion models, a broader definition is adopted. For each experimental configuration, ten runs are performed to compute the mean μ and standard deviation σ of each evaluation metric. A metric is considered reproducible if: (i) the original value falls within the interval $[\mu - \sigma,\ \mu + \sigma]$, and (ii) the metric is *stable*, which in [16] means that $\sigma \leq \tau \cdot \mu$, with $\tau = 0.02$ be a chosen threshold. Stability is crucial for reproducibility: if a metric exhibits high variability, obtaining consistent results becomes inherently difficult.[1]

2.3. Benchmarking Against Baselines

In parallel with the reproducibility analysis, the diffusion models were benchmarked against 19 strong and widely adopted baseline methods, covering matrix factorization, neighborhood-based techniques, graph-based models, and neural architectures.[2] These baselines were carefully optimized using 50

[1]Notably, standard deviations are rarely reported in recommender system research. None of the papers analyzed in [16] reported variance measures, although the reproducibility analysis shows substantial variability.

[2]The selected baselines are: Random, TopPop, Global Effects, UserKNN [17, 18], ItemKNN [19, 18], $P^3\alpha$, $RP^3\beta$ [20], GF-CF [21], $EASE^R$ [22], SLIM-BPR [23], SLIM [24], MF-BPR [23], MF-WARP, SVDpp [25], PureSVD [26], iALS [27], MultVAE [28], and LightGCN [29].

Bayesian trials following the search space from [11, 12], ensuring near-optimal performance and offering a robust estimate of the current state-of-the-art in top-k recommendation.

While the diffusion models may not have undergone equally extensive tuning, the purpose of this comparison is not to penalize them, but to assess whether the original papers evaluated their proposals against sufficiently strong baselines to support their claims.

3. Key Findings

3.1. Reproducibility and Benchmarking Results

DiffRec DiffRec [6] applies unguided Gaussian diffusion to user profiles for collaborative filtering. It includes three variants: L-DiffRec (using profile partitioning and latent space diffusion), T-DiffRec (with temporal weighting), and LT-DiffRec (a hybrid approach). Experiments were conducted on MovieLens-1M, Yelp, and Amazon-Books datasets, with each dataset processed using three strategies: "clean," "natural noise," and "random noise." Minor inconsistencies in data statistics were observed, along with some overlap between training and test sets. The total number of configurations (i.e., dataset and DiffRec variant) potentially reproducible was 16, as not all DiffRec variants were tested on all dataset versions, and the "random noise" datasets splits were not shared.

Reproducibility experiments were only partially successful: results were fully or partially reproduced for 8 out of 16 configurations, with significant variance across runs. Methodological flaws were also noted, including a narrow hyperparameter search space and the use of fixed hyperparameters values without sufficient justification. It remains unclear whether the baselines were properly tuned, as the original paper omits key details and the provided code does not include baseline implementations. Additionally, no information is provided about how the models used to generate the pre-trained latent embeddings required by L-DiffRec and LT-DiffRec were trained.

In benchmarking, DiffRec consistently underperforms compared to well-established baselines. For instance, on MovieLens-1M and Amazon-Books, KNN-based methods, graph-based models and SLIM outperform all DiffRec variants. On Yelp, DiffRec is surpassed by graph-based models and iALS.

CF-Diff CF-Diff [7] employs Gaussian diffusion guided by a multi-hop graph random walk. It was evaluated on MovieLens-1M, Yelp, and Anime. However, inconsistencies were found in dataset statistics, data split ratios, and guidance construction. Preprocessing steps were not documented. Moreover, discrepancies between the implementation and the paper description were frequent, with model components present in the code but missing or misdescribed in the paper.

Reproducibility was largely unsuccessful: only 1 out of 12 metrics was reproduced, with deviations as high as 40% and standard deviations up to 15% of the mean. Methodological issues such as inadequate hyperparameter tuning and the use of fixed values were present. Baseline optimization is again not sufficiently described, as the shared code omits the corresponding implementation.

Benchmarking results show that CF-Diff is outperformed on all datasets and all metrics by at least four and up to ten baselines. In many cases, simpler models such as UserKNN, RP$^3\beta$, and SLIM perform significantly better.

GiffCF GiffCF [8] uses graph smoothing as the forward process and corrupted user profiles as guidance. It relies on the same datasets and preprocessing as DiffRec, inheriting its inconsistencies. Reproducibility attempts were unsuccessful, with one metric matched out of 18 and substantial instability observed. For instance, on MovieLens-1M, the variance of GiffCF's results ranged from 14% to 18% on different evaluation metrics. Further methodological flaws include limited hyperparameter tuning, reliance on default values, and unclear baseline optimization. The most concerning issue is that hyperparameters were selected based on test performance, introducing data leakage and compromising the validity of the reported results. Benchmarking shows that GiffCF is outperformed on all datasets and all metrics by at least one baseline, including simple models such as UserKNN, RP$^3\beta$, and SLIM. On MovieLens-1M in particular, most baselines outperform GiffCF.

DDRM DDRM [9] applies diffusion for denoising pre-trained user and item embeddings, using user embeddings to guide item denoising and vice versa. It is evaluated on the "natural noise" and "random noise" versions of MovieLens-1M, Yelp, and Amazon-Books, inheriting the same inconsistencies noted for DiffRec. The "random noise" version of the datasets was not shared. Reproducibility was limited, with only 3 out of 36 configurations showing results close to the original. Interestingly, DDRM exhibited very low variance, in contrast to other diffusion-based models. Methodological issues include the use of fixed or default hyperparameter values and a lack of clarity around baseline tuning. Again, the shared code does not provide implementations for the baselines.

In benchmarking, DDRM is outperformed by simple models such as ItemKNN and SLIM on Amazon-Books, EASE^R on MovieLens-1M, and MultVAE and iALS on Yelp, often by a significant margin.

3.2. Theoretical Reflection and Outlook

The study in [16] also provides a conceptual analysis of the suitability of DDPMs for collaborative filtering. Several foundational issues are highlighted.

One central concern is the mismatch between the generative nature of DDPMs and the deterministic requirements of offline top-k recommendation evaluation. While DDPMs are designed to generate diverse samples, offline evaluation instead requires identifying the most relevant items from a fixed set, favoring deterministic outputs. In practice, DDPMs for recommendation are used more like multi-step denoising autoencoders than true generative models. This is evidenced by the limited corruption of input data (i.e., low number of diffusion steps and low noise levels), which restricts their generative capacity, since complete deconstruction of input data is a key aspect of DDPMs. As already pointed out by Yang et al. [30], in the context of recommendation tasks, a *"diffusion model is mostly used for adding noise in the training samples for robustness, and the learning objectives are largely categorized as classification instead of generation"*. Moreover, the recommendation systems field differs in several ways from domains where DDPMs have been successfully applied, i.e., image and video generation, for example, due to the lack of ground truth and the limited information structure [16].

These design choices and domain-specific constraints prevent DDPMs from fully exploiting their intended functionality and raise questions about their suitability for current offline evaluation frameworks. Going forward, research should better align DDPMs with the objectives of recommendation, possibly by revisiting the guidance mechanism and inference procedure. Additionally, reconciling the probabilistic outputs of DDPMs with deterministic evaluation protocols will likely require new evaluation paradigms capable of fairly assessing the performance of generative models in recommendation settings.

4. Conclusions and Implications

The analysis in [16] shows that, despite the perceived potential of Denoising Diffusion Probabilistic Models (DDPMs), their effectiveness for top-k recommendation is not convincingly demonstrated. The experimental evaluations in the original papers were not conducted under sufficiently challenging conditions, as highlighted by the benchmarking results, and the experiments are very often not reproducible. While DDPMs may still hold promise for recommender systems, their current application requires significant refinement.

Future research must focus on three critical areas. First, a more rigorous experimental methodology is needed, including the use of strong, well-tuned baselines and clear reporting of variability in results. Second, a better alignment between the generative nature of DDPMs and the deterministic nature of top-k evaluation is essential, potentially requiring a rethinking of how these models are assessed. Third, reproducibility must be prioritized. This includes the provision of complete artifacts, detailed experimental protocols, and transparent reporting practices. Ultimately, ensuring scientific rigor and methodological transparency will not only allow researchers to more reliably assess the contributions of generative models, but also facilitate meaningful progress in the field of recommender systems.

Acknowledgments

We acknowledge ISCRA for awarding this project access to the LEONARDO supercomputer, owned by the EuroHPC Joint Undertaking, hosted by CINECA (Italy).

Declaration on Generative AI

During the preparation of this work, the author used GPT-4 in order to: Grammar and spelling check. After using these tool(s)/service(s), the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

References

[1] W.-C. Kang, J. McAuley, Self-attentive sequential recommendation, in: ICDM '18, 2018, pp. 197–206.

[2] F. Sun, J. Liu, J. Wu, C. Pei, X. Lin, W. Ou, P. Jiang, Bert4rec: sequential recommendation with bidirectional encoder representations from transformer, in: Proceedings of the 28th ACM international conference on information and knowledge management, 2019, pp. 1441–1450.

[3] J. Sohl-Dickstein, E. A. Weiss, N. Maheswaranathan, S. Ganguli, Deep unsupervised learning using nonequilibrium thermodynamics, in: Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37, ICML'15, JMLR.org, 2015, p. 2256–2265.

[4] J. Ho, A. Jain, P. Abbeel, Denoising diffusion probabilistic models, in: Proceedings of the 34th International Conference on Neural Information Processing Systems, NIPS '20, 2020.

[5] J. Walker, T. Zhong, F. Zhang, Q. Gao, F. Zhou, Recommendation via collaborative diffusion generative model, in: Knowledge Science, Engineering and Management: 15th International Conference, KSEM 2022, 2022, p. 593–605. URL: https://doi.org/10.1007/978-3-031-10989-8_47. doi:10.1007/978-3-031-10989-8_47.

[6] W. Wang, Y. Xu, F. Feng, X. Lin, X. He, T. Chua, Diffusion recommender model, in: H. Chen, W. E. Duh, H. Huang, M. P. Kato, J. Mothe, B. Poblete (Eds.), Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2023, Taipei, Taiwan, July 23-27, 2023, ACM, 2023, pp. 832–841. URL: https://doi.org/10.1145/3539618.3591663. doi:10.1145/3539618.3591663.

[7] Y. Hou, J. Park, W. Shin, Collaborative filtering based on diffusion models: Unveiling the potential of high-order connectivity, in: G. H. Yang, H. Wang, S. Han, C. Hauff, G. Zuccon, Y. Zhang (Eds.), Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2024, Washington DC, USA, July 14-18, 2024, ACM, 2024, pp. 1360–1369. URL: https://doi.org/10.1145/3626772.3657742. doi:10.1145/3626772.3657742.

[8] Y. Zhu, C. Wang, Q. Zhang, H. Xiong, Graph signal diffusion model for collaborative filtering, in: G. H. Yang, H. Wang, S. Han, C. Hauff, G. Zuccon, Y. Zhang (Eds.), Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2024, Washington DC, USA, July 14-18, 2024, ACM, 2024, pp. 1380–1390. URL: https://doi.org/10.1145/3626772.3657759. doi:10.1145/3626772.3657759.

[9] J. Zhao, W. Wang, Y. Xu, T. Sun, F. Feng, T. Chua, Denoising diffusion recommender model, in: G. H. Yang, H. Wang, S. Han, C. Hauff, G. Zuccon, Y. Zhang (Eds.), Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2024, Washington DC, USA, July 14-18, 2024, ACM, 2024, pp. 1370–1379. URL: https://doi.org/10.1145/3626772.3657825. doi:10.1145/3626772.3657825.

[10] T. G. Armstrong, A. Moffat, W. Webber, J. Zobel, Improvements that don't add up: Ad-hoc retrieval results since 1998, in: CIKM '09, CIKM '09, 2009, pp. 601–610.

[11] M. Ferrari Dacrema, P. Cremonesi, D. Jannach, Are we really making much progress? a worrying analysis of recent neural recommendation approaches, in: Proceedings of the 2019 ACM Conference on Recommender Systems (RecSys 2019), Copenhagen, 2019.

[12] M. Ferrari Dacrema, S. Boglio, P. Cremonesi, D. Jannach, A troubling analysis of reproducibility and progress in recommender systems research, ACM Transactions on Information Systems 39 (2021).

[13] S. Rendle, W. Krichene, L. Zhang, J. R. Anderson, Neural collaborative filtering vs. matrix factorization revisited, in: RecSys 2020: Fourteenth ACM Conference on Recommender Systems, 2020, pp. 240–248. URL: https://doi.org/10.1145/3383313.3412488. doi:10.1145/3383313.3412488.

[14] A. Milogradskii, O. Lashinin, A. P, M. Ananyeva, S. Kolesnikov, Revisiting bpr: A replicability study of a common recommender system baseline, in: Proceedings of the 18th ACM Conference on Recommender Systems, RecSys '24, 2024, p. 267–277. URL: https://doi.org/10.1145/3640457.3688073. doi:10.1145/3640457.3688073.

[15] S. Rendle, W. Krichene, L. Zhang, Y. Koren, Revisiting the performance of ials on item recommendation benchmarks, in: RecSys '22: Sixteenth ACM Conference on Recommender Systems, ACM, 2022, pp. 427–435. URL: https://doi.org/10.1145/3523227.3548486. doi:10.1145/3523227.3548486.

[16] M. Benigni, M. F. Dacrema, D. Jannach, Diffusion recommender models and the illusion of progress: A concerning study of reproducibility and a conceptual mismatch, CoRR abs/2505.09364 (2025). URL: https://doi.org/10.48550/arXiv.2505.09364. doi:10.48550/ARXIV.2505.09364. arXiv:2505.09364.

[17] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, J. Riedl, Grouplens: An open architecture for collaborative filtering of netnews, in: J. B. Smith, F. D. Smith, T. W. Malone (Eds.), CSCW '94, Proceedings of the Conference on Computer Supported Cooperative Work, Chapel Hill, NC, USA, October 22-26, 1994, ACM, 1994, pp. 175–186. URL: https://doi.org/10.1145/192844.192905. doi:10.1145/192844.192905.

[18] R. M. Bell, Y. Koren, Improved neighborhood-based collaborative filtering, in: KDD Cup and Workshop at the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '07), 2007, pp. 7–14.

[19] B. M. Sarwar, G. Karypis, J. A. Konstan, J. Riedl, Item-based collaborative filtering recommendation algorithms, in: V. Y. Shen, N. Saito, M. R. Lyu, M. E. Zurko (Eds.), Proceedings of the Tenth International World Wide Web Conference, WWW 10, Hong Kong, China, May 1-5, 2001, ACM, 2001, pp. 285–295. URL: https://doi.org/10.1145/371920.372071. doi:10.1145/371920.372071.

[20] B. Paudel, F. Christoffel, C. Newell, A. Bernstein, Updatable, accurate, diverse, and scalable recommendations for interactive applications, ACM Trans. Interact. Intell. Syst. 7 (2017) 1:1–1:34. URL: https://doi.org/10.1145/2955101. doi:10.1145/2955101.

[21] Y. Shen, Y. Wu, Y. Zhang, C. Shan, J. Zhang, K. B. Letaief, D. Li, How powerful is graph convolution for recommendation?, in: G. Demartini, G. Zuccon, J. S. Culpepper, Z. Huang, H. Tong (Eds.), CIKM '21: The 30th ACM International Conference on Information and Knowledge Management, Virtual Event, Queensland, Australia, November 1 - 5, 2021, ACM, 2021, pp. 1619–1629. URL: https://doi.org/10.1145/3459637.3482264. doi:10.1145/3459637.3482264.

[22] H. Steck, Embarrassingly shallow autoencoders for sparse data, in: L. Liu, R. W. White, A. Mantrach, F. Silvestri, J. J. McAuley, R. Baeza-Yates, L. Zia (Eds.), The World Wide Web Conference, WWW 2019, San Francisco, CA, USA, May 13-17, 2019, ACM, 2019, pp. 3251–3257. URL: https://doi.org/10.1145/3308558.3313710. doi:10.1145/3308558.3313710.

[23] S. Rendle, C. Freudenthaler, Z. Gantner, L. Schmidt-Thieme, BPR: bayesian personalized ranking from implicit feedback, in: J. A. Bilmes, A. Y. Ng (Eds.), UAI 2009, Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence, Montreal, QC, Canada, June 18-21, 2009, AUAI Press, 2009, pp. 452–461. URL: https://www.auai.org/uai2009/papers/UAI2009_0139_48141db02b9f0b02bc7158819ebfa2c7.pdf.

[24] X. Ning, G. Karypis, SLIM: sparse linear methods for top-n recommender systems, in: D. J. Cook, J. Pei, W. Wang, O. R. Zaïane, X. Wu (Eds.), 11th IEEE International Conference on Data Mining, ICDM 2011, Vancouver, BC, Canada, December 11-14, 2011, IEEE Computer Society, 2011, pp.

497–506. URL: https://doi.org/10.1109/ICDM.2011.134. doi:10.1109/ICDM.2011.134.

[25] L. Lerche, D. Jannach, Using graded implicit feedback for bayesian personalized ranking, in: A. Kobsa, M. X. Zhou, M. Ester, Y. Koren (Eds.), Eighth ACM Conference on Recommender Systems, RecSys '14, Foster City, Silicon Valley, CA, USA - October 06 - 10, 2014, ACM, 2014, pp. 353–356. URL: https://doi.org/10.1145/2645710.2645759. doi:10.1145/2645710.2645759.

[26] P. Cremonesi, Y. Koren, R. Turrin, Performance of recommender algorithms on top-n recommendation tasks, in: X. Amatriain, M. Torrens, P. Resnick, M. Zanker (Eds.), Proceedings of the 2010 ACM Conference on Recommender Systems, RecSys 2010, Barcelona, Spain, September 26-30, 2010, ACM, 2010, pp. 39–46. URL: https://doi.org/10.1145/1864708.1864721. doi:10.1145/1864708.1864721.

[27] Y. Hu, Y. Koren, C. Volinsky, Collaborative filtering for implicit feedback datasets, in: Proceedings of the 8th IEEE International Conference on Data Mining (ICDM 2008), December 15-19, 2008, Pisa, Italy, IEEE Computer Society, 2008, pp. 263–272. URL: https://doi.org/10.1109/ICDM.2008.22. doi:10.1109/ICDM.2008.22.

[28] D. Liang, R. G. Krishnan, M. D. Hoffman, T. Jebara, Variational autoencoders for collaborative filtering, in: P. Champin, F. Gandon, M. Lalmas, P. G. Ipeirotis (Eds.), Proceedings of the 2018 World Wide Web Conference on World Wide Web, WWW 2018, Lyon, France, April 23-27, 2018, ACM, 2018, pp. 689–698. URL: https://doi.org/10.1145/3178876.3186150. doi:10.1145/3178876.3186150.

[29] X. He, K. Deng, X. Wang, Y. Li, Y. Zhang, M. Wang, Lightgcn: Simplifying and powering graph convolution network for recommendation, in: J. X. Huang, Y. Chang, X. Cheng, J. Kamps, V. Murdock, J. Wen, Y. Liu (Eds.), Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval, SIGIR 2020, Virtual Event, China, July 25-30, 2020, ACM, 2020, pp. 639–648. URL: https://doi.org/10.1145/3397271.3401063. doi:10.1145/3397271.3401063.

[30] Z. Yang, J. Wu, Z. Wang, X. Wang, Y. Yuan, X. He, Generate what you prefer: Reshaping sequential recommendation via guided diffusion, in: A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, S. Levine (Eds.), Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023, 2023. URL: http://papers.nips.cc/paper_files/paper/2023/hash/4c5e2bcbf21bdf40d75fddad0bd43dc9-Abstract-Conference.html.