A Controlled Study of Relevance–Diversity Trade-offs in News Recommendation with Embeddings

Ana Lilian Alfonso Toledo[1], Vitalio Alfonso Reguera[2], Gabriel Machado Lunardi[3]

[1]Computer Science Program – Universidade Federal de Santa Maria (UFSM)
Santa Maria – RS – Brazil

[2]Department of Information Technology – Universidad Tecnológica del Uruguay (UTEC)
Rivera – Uruguay

[3]Department of Electronics and Computing – Universidade Federal de Santa Maria (UFSM)
Santa Maria – RS – Brazil.

altoledo@inf.ufsm.br, vitalio.alfonso@utec.edu.uy, gabriel.lunardi@ufsm.br

Abstract. *News recommender systems reduce information overload, but personalization can reinforce the filter-bubble effect by repeatedly prioritizing similar content. Post-filtering diversification re-ranks top-N lists using item–item similarity, so outcomes depend on item representations. We study how replacing binary news features with TruncatedSVD embeddings affects the relevance–diversity balance in a controlled pipeline. Using a dataset of Brazilian political news, we run offline temporal replay with KNN-U, KNN-I, and SVD plus Maximal Marginal Relevance (MMR), sweeping embedding sizes. Results show embeddings boost neighborhood-based relevance but increase homogeneity, sharpening the relevance–diversity trade-off.*

1. Introduction

Recommender systems help users navigate information overload in digital environments. In news recommendation, ranking algorithms shape exposure to current events at scale, effectively acting as sociotechnical gatekeepers of public attention. A central element is the item representation: it determines how articles are encoded, influences similarity estimates, and ultimately shapes the diversity of exposure delivered to users [Ziegler et al. 2005].

The same personalization mechanisms that make recommenders effective can also produce the filter-bubble effect, repeatedly exposing users to similar content and narrowing perspectives [Liu et al. 2025]. Post-filtering (re-ranking) diversification mitigates this by reshaping a top-N list to reduce redundancy, with Maximal Marginal Relevance (MMR) as a widely adopted representative [Carbonell and Goldstein 1998]. Because these methods compute item–item similarity directly from item representations, the choice of representation is a key driver of diversification outcomes. Prior work introduced a filter-bubble metric and evaluated MMR-based re-ranking using explicit binary item features as the similarity signal [Lunardi et al. 2020], providing a reproducible baseline that we build upon here.

This work uses a news recommendation pipeline as a controlled testbed to isolate the role of item representation. We compare: (i) a binary-feature baseline computing

similarity from sparse feature vectors, and (ii) an embedding-based variant using Truncated Singular Value Decomposition (TruncatedSVD) [Martinsson and Tropp 2020] applied to the same feature matrix. Our contributions are: (1) a controlled comparison of binary features versus TruncatedSVD embeddings within the same pipeline, keeping all other components fixed; and (2) evidence that embeddings improve ranking relevance for neighborhood-based recommenders but simultaneously increase list homogeneity, making the relevance–diversity tension explicit.

2. Related Work

Embedding-aware diversification research operationalizes diversity through item–item relations in learned representation spaces. Wilhelm et al. (2018) integrate Determinantal Point Processes (DPPs) into YouTube's pipeline with a kernel whose off-diagonal entries encode pairwise embedding similarity and whose diagonal encodes quality, showing that representation design directly shapes what items are treated as redundant. Gao et al. (2022) propose TD-VAE-CF, which learns user and item embeddings via a Variational Autoencoder (VAE) and steers recommendations along a chosen conceptual axis (e.g., political leaning) to mitigate filter bubbles. These approaches generally train embeddings end-to-end or extract them from raw content, making it difficult to disentangle the contributions of representation and diversification strategy. Our work addresses this by holding the diversification method (MMR) fixed and replacing only the item representation (binary features versus TruncatedSVD embeddings from the same feature matrix) enabling a direct, interpretable assessment of how representation choice affects the relevance–diversity trade-off.

3. Proposed Method

3.1. Task Definition and Pipeline

Let $\mathcal{U}$ be the set of users, $\mathcal{I}$ the set of news items, and $\mathcal{R} = \{(u, i, r_{ui}, t_{ui})\}$ the observed interactions, where $r_{ui} \in \{-1, +1\}$ denotes feedback and t_{ui} its timestamp. A base recommender estimates relevance scores $\hat{r}_{uj}$ for candidates $j \in \mathcal{C}_u$; a post-filtering reranker then reorders the top-N items by combining predicted relevance with an item–item similarity term from the chosen representation. All experiments use an offline temporal replay protocol, iterating over $\mathcal{R}$ chronologically to yield a reproducible approximation of a feed-like setting.

The evaluation proceeds in two stages. Stage 1 (Binary baseline): The pipeline runs using binary feature vectors $\mathbf{x}_i \in \{0, 1\}^d$, recording baseline metrics $(\mathrm{NDCG@10}_{\mathrm{bin}}, \mathrm{ILS}_{\mathrm{bin}})$. Stage 2 (Embedding sweep): For each size $k \in \mathcal{K}$ and seed $s \in \mathcal{S}$, embeddings $\mathbf{z}_i \in \mathbb{R}^k$ are derived via TruncatedSVD applied to $\mathbf{X}$; the same replay loop is re-run and metrics $(\mathrm{NDCG@10}_{k,s}, \mathrm{ILS}_{k,s})$ are recorded. Comparing both stages under fixed recommenders and re-ranking isolates the effect of representation.

3.2. Dataset

We use the publicly available news recommendation dataset obtained from the experiment conducted by Lunardi et al. (2020), summarized in Table 1. It contains 298 active users, 1,938 articles in Portuguese about Brazilian politics, and 12,363 explicit ratings collected between May and September 2019. Each article is described by 83 item-side

attributes (16 topic indicators, 2 sentiment scalars, and 65 binary attributes), defining a sparse item–feature matrix $\mathbf{X} \in \mathbb{R}^{1938\times 83}$ used for both similarity computation and embedding extraction.

Table 1. Dataset summary.

Characteristic	Value
Domain / Language	Brazilian politics / Portuguese
Active users	298
News articles	1,938
Interactions	12,363
Time span	2019-05-07 to 2019-09-10
Item features	83 (16 topics, 2 sentiment, 65 binary)
Item–feature matrix	$\mathbf{X} \in \mathbb{R}^{1938\times 83}$

3.3. Algorithms

We adopt a two-stage pipeline: a base recommender estimates relevance scores and MMR re-ranks the top-N list. The base recommenders are *user-based KNN* (KNN-U) [Herlocker et al. 1999], *item-based KNN* (KNN-I) [Sarwar et al. 2001], and matrix factorization via *SVD* [Koren et al. 2009], using Pearson baseline similarity with shrinkage regularization for the neighborhood models. KNN-U predicts:

$$\hat{r}_{ui} = \bar{r}_u + \frac{\sum_{v\in N_K(u)} s(u,v)(r_{vi} - \bar{r}_v)}{\sum_{v\in N_K(u)} |s(u,v)|}, \tag{1}$$

and KNN-I predicts $\hat{r}_{ui} = \sum_{j\in N_K(i)} s(i,j) r_{uj} \,/\, \sum_{j\in N_K(i)} |s(i,j)|$. SVD predicts $\hat{r}_{ui} = \mu + b_u + b_i + \mathbf{p}_u^\top \mathbf{q}_i$ via regularized optimization [Koren et al. 2009].

MMR re-ranking [Carbonell and Goldstein 1998] greedily builds a diversified list S_u by selecting:

$$i^\star = \arg\max_{i\in L_u \setminus S_u} \left[\lambda\, \hat{r}_{ui} - (1-\lambda) \max_{j\in S_u} \mathrm{sim}(i,j) \right], \tag{2}$$

where $\mathrm{sim}(i,j)$ is computed from the chosen item representation. We set $\lambda = 0.7$ to maintain a stronger weight on relevance while preserving a meaningful diversification effect; smaller values would sacrifice too much ranking quality, and $\lambda = 1$ reduces MMR to a pure relevance ranker.

In the embedding condition, $\mathbf{z}_i \in \mathbb{R}^k$ is obtained from the rank-k approximation $\mathbf{X} \approx \mathbf{U}_k \mathbf{\Sigma}_k \mathbf{V}_k^\top$, with $\mathbf{Z} = \mathbf{U}_k \mathbf{\Sigma}_k$ (L_2-normalized rows), so that $\mathrm{sim}(i,j)$ is cosine similarity in the embedding space. TruncatedSVD was chosen for its stable, interpretable linear reduction of the binary feature space: it preserves co-occurrence structure without the stochasticity and overfitting risk of nonlinear alternatives such as Autoencoders, which are particularly pronounced in small datasets.

3.4. Metrics

Relevance is measured by NDCG@10; diversity by Intra-List Similarity (ILS), the average pairwise cosine similarity within a list (higher ILS means lower diversity). Both are

averaged over all replay checkpoints. The shared item–item similarity is:

$$\mathrm{sim}(i,j) = \mathbf{v}_i^\top \mathbf{v}_j \,/\, \|\mathbf{v}_i\|_2 \|\mathbf{v}_j\|_2, \tag{3}$$

where $\mathbf{v}_i \in \{\mathbf{x}_i, \mathbf{z}_i\}$. Binary relevance is $y_{ui} = \mathbf{1}[r_{ui} = +1]$, and:

$$\mathrm{NDCG@10}(u,t) = \frac{\sum_{a=1}^{10} y_{ui_a}/\log_2(a+1)}{\mathrm{IDCG@10}(u,t)}, \quad \mathrm{ILS}(L_{u,t}) = \frac{1}{\binom{N}{2}} \sum_{a<b} \mathrm{sim}(i_a, i_b). \tag{4}$$

To summarize the trade-off without a validation split, we select $k^{(\alpha)}$ via a convex combination of min–max normalized metrics:

$$k^{(\alpha)} = \arg\max_{k \in \mathcal{K}} \Big[(1-\alpha)\, \widetilde{\mathrm{NDCG@10}}(k) + \alpha\, (1 - \widetilde{\mathrm{ILS}}(k)) \Big], \tag{5}$$

where $\alpha = 0$ prioritizes relevance and $\alpha = 1$ minimizes homogeneity.

4. Experimental Results and Discussion

Table 2 compares the binary baseline against the best-relevance and best-diversity embedding configurations for each algorithm ($N = 10$, $\lambda = 0.7$). Replacing binary features with TruncatedSVD embeddings consistently improves relevance for neighborhood-based models: KNN-I+MMR gains NDCG@10 from 0.754 to 0.797 ($k = 25$), and KNN-U+MMR from 0.684 to 0.721 ($k = 19$). These gains arise because KNN methods couple their collaborative signal directly to item–item similarity, which embeddings reshape decisively. In contrast, SVD+MMR shows essentially no change ($0.596 \to 0.597$): the SVD model already learns a latent decomposition of the interaction space during training, so substituting the MMR similarity term with item-side embeddings does not alter what the model internally captures. The negligible sensitivity of SVD+MMR to the representation change suggests that its own decomposition saturates what the embedding can offer. The overall lower performance of SVD+MMR relative to KNN-based methods may further reflect insufficient interaction data for robust matrix factorization given the small, sparse dataset.

Embeddings also systematically increase list homogeneity across all algorithms (KNN-U: ILS $0.143 \to 0.162$; KNN-I: $0.144 \to 0.161$; SVD: $0.211 \to 0.222$), because the embedding space induces higher average cosine similarities than the sparse binary space, amplifying both MMR's redundancy penalty and ILS. Choosing k to minimize ILS partially recovers diversity (KNN-U: 0.150 at $k = 78$; KNN-I: 0.152 at $k = 76$) but lowers NDCG@10 (KNN-U: $0.721 \to 0.694$; KNN-I: $0.797 \to 0.754$).

Table 2. Binary baseline vs. TruncatedSVD embeddings: best-relevance (k_{best}) and best-diversity ($k_{\min \mathrm{ILS}}$) configurations under MMR re-ranking.

Algorithm	Binary (baseline)		Embeddings: best NDCG@10			Embeddings: min ILS		
	NDCG@10	ILS	k_{best}	NDCG@10	ILS	$k_{\min \mathrm{ILS}}$	NDCG@10	ILS
KNN-U + MMR	0.684	0.143	19	0.721	0.162	78	0.694	0.150
KNN-I + MMR	0.754	0.144	25	0.797	0.161	76	0.754	0.152
SVD + MMR	0.596	0.211	55	0.597	0.222	77	0.593	0.220

Table 3 makes this trade-off explicit via $k^{(\alpha)}$. At $\alpha = 0$, the selector recovers best-relevance configurations. As α grows, k shifts toward lower-ILS configurations,

more sharply for KNN-U than KNN-I, and negligibly for SVD. KNN-I offers a favorable intermediate regime ($\alpha \leq 0.75$): $k^{(\alpha)} = 52$ yields NDCG@10 ≈ 0.792 and ILS ≈ 0.154, recovering most of the diversity gain with a small relevance loss. At $\alpha = 1$, KNN-I switches to $k = 76$, where ILS reaches its minimum (0.152) but NDCG@10 returns to the binary-baseline level (0.754). These patterns confirm that embedding dimensionality acts as a tunable parameter for navigating the relevance–homogeneity balance.

Table 3. Trade-off selection via $\alpha \in [0, 1]$. $k^{(\alpha)}$ selected by Eq. (5); non-normalized NDCG@10 and ILS reported.

	KNN-U + MMR			KNN-I + MMR			SVD + MMR		
α	$k^{(\alpha)}$	NDCG@10	ILS	$k^{(\alpha)}$	NDCG@10	ILS	$k^{(\alpha)}$	NDCG@10	ILS
0.00	19	0.721	0.162	25	0.797	0.161	55	0.597	0.222
0.25	41	0.718	0.155	52	0.792	0.154	55	0.597	0.222
0.50	69	0.707	0.152	52	0.792	0.154	55	0.597	0.222
0.75	78	0.694	0.150	52	0.792	0.154	77	0.593	0.220
1.00	78	0.694	0.150	76	0.754	0.152	77	0.593	0.220

5. Conclusion

We examined how item representation affects the relevance–diversity balance in a news recommendation pipeline with post-filtering diversification. Comparing binary features against TruncatedSVD embeddings derived from the same item–feature matrix, while keeping base recommenders (KNN-U, KNN-I, SVD) and MMR re-ranking fixed, embeddings improved NDCG@10 for neighborhood-based methods but increased ILS across all algorithms. SVD showed minimal sensitivity to the representation change, consistent with its latent-factor structure already capturing the relevant variation; this behavior may also reflect limited interaction data for robust matrix factorization in the small, sparse dataset. The trade-off selector $k^{(\alpha)}$ provides a parameterized view of how embedding dimensionality controls the relevance–homogeneity balance. These results reinforce that item representation is not an implementation detail: changing the similarity signal reshapes both ranking quality and exposure homogeneity, even when all other pipeline components are fixed. Future work includes supervised or contrastive representation objectives, Autoencoder-based embeddings on larger corpora, and online evaluations to characterize downstream exposure effects.

Acknowledgements

The authors thank FAPERGS (grant 24/2551-0000645-1), CNPq Universal project 402086/2023-6, and CAPES.

References

Carbonell, J. and Goldstein, J. (1998). The use of mmr, diversity-based reranking for reordering documents and producing summaries. In *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 335–336.

Gao, Z., Shen, T., Mai, Z., Bouadjenek, M. R., Waller, I., Anderson, A., Bodkin, R., and Sanner, S. (2022). Mitigating the filter bubble while maintaining relevance: Targeted diversification with vae-based recommender systems. In *Proceedings of the 45th*

International ACM SIGIR Conference on Research and Development in Information Retrieval, pages 2524–2531.

Herlocker, J. L., Konstan, J. A., Borchers, A., and Riedl, J. (1999). An algorithmic framework for performing collaborative filtering. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, pages 230–237.

Koren, Y., Bell, R., and Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *Computer*, 42(8):30–37.

Liu, N., Hu, X. E., Savas, Y., Baum, M. A., Berinsky, A. J., Chaney, A. J. B., Lucas, C., Mariman, R., de Benedictis-Kessner, J., Guess, A. M., Knox, D., and Stewart, B. M. (2025). Short-term exposure to filter-bubble recommendation systems has limited polarization effects: Naturalistic experiments on youtube. *Proceedings of the National Academy of Sciences*, 122(8):e2318127122.

Lunardi, G. M., Machado, G. M., Maran, V., and de Oliveira, J. P. M. (2020). A metric for filter bubble measurement in recommender algorithms considering the news domain. *Applied Soft Computing*, 97:106771.

Martinsson, P.-G. and Tropp, J. A. (2020). Randomized numerical linear algebra: Foundations and algorithms. *Acta Numerica*, 29:403–572.

Sarwar, B., Karypis, G., Konstan, J., and Riedl, J. (2001). Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th international conference on World Wide Web*, pages 285–295.

Wilhelm, M., Ramanathan, A., Bonomo, A., Jain, S., Chi, E. H., and Gillenwater, J. (2018). Practical diversified recommendations on youtube with determinantal point processes. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, pages 2165–2173.

Ziegler, C.-N., McNee, S. M., Konstan, J. A., and Lausen, G. (2005). Improving recommendation lists through topic diversification. In *Proceedings of the 14th international conference on World Wide Web*, pages 22–32.