VI. Appendix 1: Theoretics and Differential Calculus

A. Epistemic Comparison of NMF & DNN

In this section, we discuss theoretical and epistemic differences between NMF-based and deep neural network (DNN)-based clustering, focusing on comparative strengths, interpretability, and underlying philosophies.

NMF models naturally facilitate unsupervised and supervised scenarios, well-suited for multi-layer optimization. Their fundamental assumption—objects perceived as additive combinations of meaningful parts—aligns closely with human perception mechanisms [1]. Nonnegative constraints further ensure interpretability; negative contributions are often meaningless in real-world tasks like face, image, or gene data analysis. Such nonnegative decompositions typically yield localized, semantically interpretable features (e.g., facial parts). Additionally, the inherent sparseness of NMF enhances representation, clearly distinguishing these models from purely distributed approaches.

While shallow NMFs are equivalent to single-layer perceptrons, differences emerge in multilayer architectures: NMFs remain linear reconstruction models, optimized by specialized multiplicative update rules [2], unlike nonlinear DNNs optimized via gradient-based chain rules. This distinction is not about determining a universally best method. Indeed, many successful DNN architectures utilize dot-product similarity mechanisms, alike factorization methods. Recent works indicate that NMF-based factorization models often provide interpretability and computational advantages in embedding-based tasks such as collaborative filtering [3], [4]. Consequently, emerging research increasingly blends NMF and DNN or extends NMF to deep hierarchical models to leverage complementary strengths and optimize application-specific trade-offs.

B. Derivation of update formula

To calculate the gradient of the objective function in Eq. (13), we first need to express the function as a trace expression. Then, we can solve Eq.(13) by introducing a Lagrangian multiplier matrix $\boldsymbol{\Theta}_i$ to ensure the nonnegativity constraints on $\boldsymbol{H}_i$. This results in an equivalent objective function as follows:

$$\begin{aligned}\min_{\boldsymbol{H}_i,\boldsymbol{\Theta}_i} \mathcal{L}(\boldsymbol{H}_i,\boldsymbol{\Theta}_i) &= \mathrm{Tr}(-2\boldsymbol{A}^\top\boldsymbol{\Psi}_i\boldsymbol{H}_i\boldsymbol{\Phi}_i\boldsymbol{W}_p\boldsymbol{\Phi}_i^\top\boldsymbol{H}_i^\top\boldsymbol{\Psi}_i^\top\\ &+ \boldsymbol{\Psi}_i\boldsymbol{H}_i\boldsymbol{\Phi}_i\boldsymbol{W}_p^\top\boldsymbol{\Phi}_i^\top\boldsymbol{H}_i^\top\boldsymbol{\Psi}_i^\top\boldsymbol{\Psi}_i\boldsymbol{H}_i\boldsymbol{\Phi}_i\boldsymbol{W}_p\boldsymbol{\Phi}_i^\top\boldsymbol{H}_i^\top\boldsymbol{\Psi}_i^\top)\\ &+ \lambda\mathrm{Tr}(\boldsymbol{\Phi}_i^\top\boldsymbol{H}_i^\top\boldsymbol{\Psi}_i^\top\boldsymbol{F}\boldsymbol{F}^\top\boldsymbol{\Psi}_i\boldsymbol{H}_i\boldsymbol{\Phi}_i) - \mathrm{Tr}(\boldsymbol{\Theta}_i\boldsymbol{H}_i^\top). \end{aligned} \tag{3}$$

By setting the partial derivative of $\mathcal{L}(\boldsymbol{H}_i,\boldsymbol{\Theta}_i)$ with respect to $\boldsymbol{H}_i$ to $\mathbf{0}$, we have:

$$\begin{aligned}\boldsymbol{\Theta}_i = &-2\boldsymbol{\Psi}_i^\top\boldsymbol{A}^\top\boldsymbol{\Psi}\boldsymbol{W}_p\boldsymbol{\Phi}_i^\top - 2\boldsymbol{\Psi}_i^\top\boldsymbol{A}\boldsymbol{\Psi}\boldsymbol{W}_p^\top\boldsymbol{\Phi}_i^\top\\ &+ 2\boldsymbol{\Psi}_i^\top\boldsymbol{\Psi}\left(\boldsymbol{W}_p^\top\boldsymbol{\Psi}^\top\boldsymbol{\Psi}\boldsymbol{W}_p + \boldsymbol{W}_p\boldsymbol{\Psi}^\top\boldsymbol{\Psi}\boldsymbol{W}_p^\top\right)\boldsymbol{\Phi}_i^\top\\ &+ 2\lambda\boldsymbol{\Psi}_i^\top\boldsymbol{F}\boldsymbol{F}^\top\boldsymbol{\Psi}\boldsymbol{\Phi}_i^\top. \end{aligned} \tag{4}$$

From the Karush-Kuhn-Tucker (KKT) complementary slackness conditions, we obtain $\boldsymbol{\Theta}_i \odot \boldsymbol{H}_i = \mathbf{0}$, which is the fixed point equation that the solution must satisfy at convergence. By solving it, we derive the following update rule for $\boldsymbol{H}_i$:

$$\boldsymbol{H}_i \leftarrow \boldsymbol{H}_i \odot \left[\frac{\boldsymbol{\Psi}_i^\top(\boldsymbol{A}^\top\boldsymbol{\Psi}\boldsymbol{W}_p + \boldsymbol{A}\boldsymbol{\Psi}\boldsymbol{W}_p^\top + \lambda[\boldsymbol{F}\boldsymbol{F}^\top\boldsymbol{\Psi}]^-)\boldsymbol{\Phi}_i^\top}{\boldsymbol{\Psi}_i^\top(\boldsymbol{\Psi}\boldsymbol{W}_p^\top\boldsymbol{\Psi}^\top\boldsymbol{\Psi}\boldsymbol{W}_p + \boldsymbol{\Psi}\boldsymbol{W}_p\boldsymbol{\Psi}^\top\boldsymbol{\Psi}\boldsymbol{W}_p^\top + \lambda[\boldsymbol{F}\boldsymbol{F}^\top\boldsymbol{\Psi}]^+)\boldsymbol{\Phi}_i^\top}\right]^{\frac{1}{4}} \tag{5}$$

where we separate the positive and negative parts of an arbitrary matrix $\boldsymbol{B}$ into $\boldsymbol{B}^+ = \max(\boldsymbol{B}, 0)$ and $\boldsymbol{B}^- = -\min(\boldsymbol{B}, 0)$, such that the main matrix is conveniently the addition of the negative and postive parts $\boldsymbol{B} = \boldsymbol{B}^+ - \boldsymbol{B}^-$.

VII. Appendix 2: Ablation Studies

A. Selecting $\lambda^\star$ and Bracketing Values

We sweep $\lambda \in \{10^{-3}, 10^{-2}, \ldots, 10^3\}$ and, for each dataset (and the k used in the main results), form the utility–fairness set $\{(Q(\lambda), \bar{B}(\lambda))\}$. We retain the Pareto front (undominated points), min–max scale $(Q, \bar{B})$ to $[0, 1]$ (per dataset and k), and select

$$\lambda^\star \;=\; \arg\min_{\lambda\in\Lambda_{\mathrm{P}}} \big\| \,(\tilde{Q}(\lambda), \tilde{B}(\lambda)) - (1, 1)\,\big\|_2,$$

with a tie–breaker preferring smaller $|\tilde{Q} - \tilde{B}|$ (closer to the identity guide). We also report a bracket, i.e., the nearest available grid values to $\{\lambda^\star/10,\, 10\lambda^\star\}$, as a transparent operating band. Table VII summarizes $\lambda^\star$, the selected layer sizes, and the achieved Q and $\bar{B}$. Moreover, it reports the bracketing values λ_{lo} and λ_{hi} related to each $\lambda^\star$ and the corresponding low and high Q and $\bar{B}$ values.

a) What the table shows: (i) Datasets with strong community signal (e.g., SBM) select small $\lambda^\star$, preserving Q while still improving $\bar{B}$; The same pattern applies to NBA, and LastFM (ii) highly imbalanced or sparse social graphs (e.g., Diaries and Facebook) push $\lambda^\star$ higher to achieve parity; (iii) for medium-scale, noisy graphs (DrugNet, Friendship), $\lambda^\star$ sits near 10^{-1}, striking a balanced operating point. Brackets are tight where fronts are steep (clear trade-offs) and wider where fronts are flat (multiple near-equivalent settings). We use $\lambda^\star$ for main tables; Pareto plots in the paper illustrate the surrounding spectrum (including extreme settings within the bracket).

B. Sensitivity to the Number of Clusters (k)

Reporting & robustness. All entries are mean values over 10 random seeds (std. omitted for space); the same seeds are reused across λ for a given dataset and k. Although $\lambda^\star$ is selected using min–max–normalized $(Q, \bar{B})$ per dataset and k, a check with a simple linear scalarization $0.5\,Q + 0.5\,\bar{B}$ chose the same or bracket-adjacent setting in every case (i.e., within $\{\lambda^\star/10,\, 10\lambda^\star\}$). Width patterns follow fixed templates (e.g., $[64, k]$ or $[256, 64, k]$) across λ; the observed increase of $\lambda^\star$ with k persists under fixed-width variants, indicating the trend is driven by partition granularity rather than capacity. For *LastFM*, the steeper drop of $\bar{B}$ as k grows aligns with its high homophily/sparsity: finer partitions leave fewer cross-group ties to balance without larger λ. The dynamics of λ and k are illustrated in Fig 8.

Table VIII examines how the selected $\lambda^\star$ and the attained $(Q, \bar{B})$ evolve as k increases from 3 to 8. Three consistent trends emerge. (1) $\lambda^\star$ *generally rises with* k, reflecting that enforcing proportional representation becomes harder when clusters are smaller: Facebook shows a monotone increase

TABLE VII: Selected $\lambda^\star$ per dataset (with layers) and bracketing values. Q and $\bar{B}$ are raw scores at each λ; λ_{lo} and λ_{hi} are the nearest grid values to $\{\lambda^\star/10,\ 10\lambda^\star\}$.

Dataset	k	**Layers@** $\lambda^\star$	$\boldsymbol{\lambda^\star}$	$\boldsymbol{\lambda_{\text{lo}}}$	$\boldsymbol{\lambda_{\text{hi}}}$	$Q(\lambda^\star)$	$\bar{B}(\lambda^\star)$	$Q(\lambda_{\text{lo}})$	$\bar{B}(\lambda_{\text{lo}})$	$Q(\lambda_{\text{hi}})$	$\bar{B}(\lambda_{\text{hi}})$
NBA	2	[64, 2]	0.05	0.005	0.5	0.134	0.438	0.130	0.361	0.131	0.370
Pokec-n	2	[512, 16, 2]	10	10	10	0.164	0.177	0.164	0.177	0.164	0.177
Pokec-z	2	[512, 16, 2]	10	10	10	0.162	0.175	0.162	0.175	0.162	0.175
Diaries	5	[64, 16, 5]	50	5	500	0.716	0.787	0.745	0.688	0.517	0.802
DrugNet	5	[64, 5]	0.1	0.01	1	0.591	0.162	0.600	0.121	0.572	0.117
Facebook	5	[64, 5]	100	10	1000	0.503	0.768	0.512	0.638	0.461	0.752
Friendship	5	[64, 5]	0.1	0.01	1	0.666	0.614	0.670	0.641	0.561	0.663
LastFM	5	[256, 64, 5]	0.005	0.001	0.05	0.420	0.091	0.660	0.046	0.629	0.020
SBM	5	[64, 16, 5]	0.5	0.05	5	0.114	1.000	0.114	1.000	0.078	0.756

TABLE VIII: **Sensitivity of DFNMF to** k through the lens of $\lambda^\star$ on DrugNet, LastFM, and Facebook. Values reported at $\lambda^\star$ per k (Pareto-selected).

Dataset	**k**	$\boldsymbol{\lambda^\star}$	**Layers**	**Q**	$\bar{\mathbf{B}}$
DrugNet	3	0.01	[64, 3]	0.482	0.170
	5	0.10	[64, 5]	0.591	0.162
	8	0.50	[64, 8]	0.456	0.158
LastFM	3	0.005	[256, 64, 3]	0.516	0.147
	5	0.005	[256, 64, 5]	0.420	0.091
	8	0.050	[256, 64, 8]	0.384	0.090
Facebook	3	60	[64, 3]	0.408	0.818
	5	100	[64, 5]	0.503	0.768
	8	150	[64, 8]	0.437	0.652

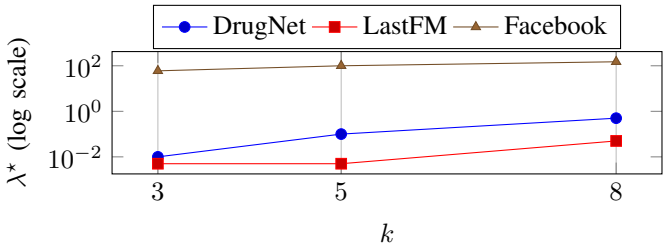

Fig. 8: Selected $\lambda^\star$ vs. k (log-scale on y).

$(60 \to 100 \to 150)$; DrugNet increases then plateaus $(0.01 \to 0.10 \to 0.50)$; LastFM is stable at small k and rises at k=8 $(0.005 \to 0.005 \to 0.05)$. (2) *Modularity often peaks at a moderate* k: DrugNet and Facebook achieve their best Q at k=5 (0.591 and 0.503, respectively) and then drop at k=8, while LastFM (sparser, more homophilous) shows a steady decline $(0.516 \to 0.420 \to 0.384)$ as partitions become finer. (3) *Balance degrades as* k *grows*: the drop is mild on DrugNet (0.170→0.162→0.158), pronounced on Facebook (0.818→0.768→0.652), and sharp early on for LastFM (0.147→0.091→0.090), consistent with tighter per-cluster parity constraints at smaller cluster sizes. Layer choices at $\lambda^\star$ remain compact and systematic—$[64, k]$ for DrugNet/Facebook and $[256, 64, k]$ for LastFM—supporting reproducible configurations across k.

a) Takeaway.: As k increases, maintaining per-cluster parity becomes more challenging: $\lambda^\star$ typically needs to grow, $\bar{B}$ tends to fall, and Q often peaks around a moderate k. Reporting $\lambda^\star$ alongside $(Q, \bar{B})$ at each k provides a transparent, reproducible operating point per dataset.

C. Extended Interpretability Rsults

This appendix complements Sec. V-E2 by reporting the *full* soft-membership matrices used in the 60-node example (Fig. 5). Recall the hierarchical mapping $\boldsymbol{\Psi} = \boldsymbol{H}_1\boldsymbol{H}_2$, where $\boldsymbol{H}_1$ encodes node→micro-cluster affinities and $\boldsymbol{H}_2$ maps micro-clusters→communities. For visual context in the main text, see the paired depiction of $\boldsymbol{H}_2$ and the trimmed $\boldsymbol{\Psi}$ in Fig. 6, and the compact slice of $\boldsymbol{H}_1$ in Table V. The full results are now illustrated in Tables IX, X, and XI

a) Reading the matrices.: Entries are nonnegative *soft memberships* (not necessarily normalized). Two patterns underpin interpretability: (i) **micro-sparsity**—rows of $\boldsymbol{H}_1$ are typically few-peaked, so most nodes participate in a small number of micro-clusters (but this highly depends on graph sructure and properties); and (ii) **structured aggregation**—columns of $\boldsymbol{H}_2$ are near one-hot for core micro-clusters (e.g., A→I, B→II, C→III), while boundary micro-clusters spread mass across communities. Consequently, $\boldsymbol{\Psi} = \boldsymbol{H}_1\boldsymbol{H}_2$ consolidates concentrated rows into near one-hot community affinities, while mixed rows remain soft. This traceability (node→micro→community) enables transparent auditing of how local structure (under fairness regularization) aggregates into global communities.

VIII. Appendix 3: Intersectional Multi-Attribute Fairness

A. Constructing the Intersectional Fairness Matrix

Suppose each node has A sensitive attributes. Attribute $a \in \{1, \ldots, A\}$ has m_a groups and one-hot indicator $G^{(a)} \in \mathbb{R}^{n \times m_a}$. To enforce *intersectional* (joint) parity across the Cartesian product of all attributes as previously done in [5], we build a joint one-hot matrix:

$$G_{\text{int}} \in \mathbb{R}^{n \times M}, \qquad M = \prod_{a=1}^{A} m_a, \tag{6}$$

whose columns correspond to all joint categories (e.g., *male*∧*Asian*, *female*∧*White*, ...). Concretely, each joint column

TABLE IX: Full node→micro-cluster soft memberships ($\boldsymbol{H}_1$, micro-clusters A–L).

Node	Micro-clusters											
	A	B	C	D	E	F	G	H	I	J	K	L
1	0.26	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.01	0.00	0.00	0.00
2	0.25	0.00	0.01	0.00	0.01	0.09	0.02	0.00	0.01	0.00	0.04	0.00
3	0.21	0.00	0.01	0.00	0.09	0.03	0.00	0.00	0.00	0.00	0.11	0.00
4	0.23	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00
5	0.31	0.00	0.02	0.00	0.01	0.15	0.05	0.10	0.05	0.02	0.09	0.00
6	0.05	0.00	0.00	0.00	0.00	0.26	0.61	0.00	0.00	0.00	0.00	0.00
7	0.03	0.00	0.00	0.00	0.00	0.17	0.81	0.00	0.00	0.00	0.00	0.00
8	0.01	0.00	0.00	0.00	0.00	0.04	0.81	0.00	0.00	0.00	0.00	0.00
9	0.19	0.00	0.00	0.00	0.00	0.33	0.23	0.00	0.01	0.00	0.00	0.00
10	0.11	0.00	0.03	0.00	0.00	0.16	0.52	0.00	0.00	0.00	0.07	0.00
11	0.10	0.02	0.00	0.00	0.00	0.05	0.00	0.50	0.03	0.00	0.00	0.00
12	0.06	0.00	0.00	0.00	0.00	0.03	0.00	0.63	0.08	0.00	0.00	0.00
13	0.16	0.10	0.03	0.00	0.00	0.06	0.00	0.54	0.23	0.04	0.15	0.03
14	0.07	0.06	0.00	0.00	0.00	0.03	0.00	0.56	0.05	0.00	0.00	0.00
15	0.14	0.04	0.00	0.00	0.02	0.18	0.07	0.64	0.05	0.01	0.00	0.01
16	0.13	0.00	0.00	0.00	0.00	0.05	0.00	0.00	0.03	0.00	0.00	0.00
17	0.06	0.00	0.00	0.00	0.00	0.03	0.10	0.00	0.04	0.00	0.00	0.07
18	0.19	0.00	0.00	0.00	0.00	0.09	0.01	0.00	0.01	0.00	0.00	0.00
19	0.20	0.00	0.00	0.00	0.00	0.22	0.02	0.00	0.02	0.00	0.00	0.00
20	0.22	0.00	0.03	0.00	0.02	0.04	0.00	0.00	0.11	0.00	0.06	0.65
21	0.00	0.06	0.00	0.35	0.00	0.00	0.00	0.00	0.00	0.31	0.00	0.00
22	0.03	0.00	0.00	0.28	0.00	0.05	0.00	0.00	0.01	0.50	0.00	0.05
23	0.00	0.00	0.00	0.31	0.00	0.00	0.00	0.00	0.00	0.57	0.00	0.00
24	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.00	0.57	0.00	0.00
25	0.00	0.00	0.01	0.40	0.04	0.00	0.00	0.00	0.00	0.86	0.00	0.00
26	0.00	0.22	0.00	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
27	0.00	0.21	0.00	0.05	0.00	0.02	0.00	0.08	0.01	0.00	0.00	0.00
28	0.00	0.23	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
29	0.00	0.28	0.01	0.02	0.06	0.00	0.00	0.02	0.00	0.00	0.12	0.00
30	0.00	0.20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
31	0.00	0.23	0.00	0.00	0.00	0.00	0.00	0.04	0.04	0.00	0.00	0.00
32	0.00	0.30	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
33	0.00	0.26	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
34	0.00	0.23	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
35	0.00	0.26	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
36	0.00	0.10	0.00	0.13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
37	0.00	0.00	0.00	0.84	0.01	0.00	0.10	0.00	0.00	0.07	0.01	0.00
38	0.00	0.01	0.00	0.54	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
39	0.00	0.05	0.00	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
40	0.00	0.09	0.00	0.09	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
41	0.00	0.00	0.18	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
42	0.00	0.00	0.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
43	0.01	0.00	0.24	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.24
44	0.00	0.00	0.30	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
45	0.00	0.00	0.26	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00
46	0.03	0.01	0.01	0.02	0.53	0.01	0.00	0.00	0.20	0.01	0.03	0.07
47	0.00	0.05	0.07	0.01	0.23	0.00	0.00	0.00	0.00	0.00	0.08	0.00
48	0.00	0.00	0.06	0.17	0.58	0.00	0.00	0.00	0.01	0.10	0.04	0.00
49	0.03	0.01	0.10	0.02	0.63	0.02	0.00	0.00	0.07	0.00	0.10	0.00
50	0.04	0.00	0.16	0.00	0.15	0.05	0.00	0.00	0.02	0.00	0.11	0.00
51	0.01	0.00	0.00	0.00	0.64	0.17	0.10	0.21	0.05	0.00	0.00	0.35
52	0.00	0.00	0.00	0.00	0.27	0.00	0.02	0.03	0.00	0.00	0.00	0.40
53	0.00	0.00	0.05	0.00	0.25	0.00	0.01	0.01	0.00	0.00	0.00	0.34
54	0.00	0.00	0.00	0.00	0.11	0.01	0.02	0.07	0.01	0.00	0.00	0.06
55	0.03	0.00	0.00	0.00	0.32	0.01	0.01	0.00	0.00	0.00	0.01	0.31
56	0.01	0.00	0.24	0.00	0.00	0.04	0.13	0.00	0.01	0.00	0.25	0.00
57	0.06	0.00	0.26	0.00	0.00	0.05	0.01	0.09	0.06	0.01	0.26	0.00
58	0.00	0.00	0.22	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.24	0.00
59	0.00	0.00	0.31	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.00
60	0.00	0.00	0.19	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00

TABLE X: Micro-cluster membership values of $\boldsymbol{H}_2$ across the three main clusters (I–III).

Micro cluster	Clusters		
	I	II	III
A	1.04	0.00	0.00
B	0.00	1.00	0.00
C	0.00	0.00	1.00
D	0.00	0.05	0.01
E	0.02	0.00	0.19
F	0.00	0.00	0.00
G	0.08	0.00	0.00
H	0.06	0.08	0.00
I	0.00	0.00	0.00
J	0.00	0.02	0.06
K	0.00	0.00	0.01
L	0.02	0.00	0.01

is the elementwise AND (Hadamard product) of one columns taken from each $G^{(a)}$:

$$G_{\text{int}} \;=\; G^{(1)} \odot G^{(2)} \odot \cdots \odot G^{(A)}. \tag{7}$$

We then apply proportional centering and drop one redundant column to avoid linear dependence:

$$F_{\text{int}} \;=\; G_{\text{int}} \;-\; \tfrac{1}{n}\,\mathbf{1}_n\big(\mathbf{1}_n^\top G_{\text{int}}\big), \qquad F_{\text{int}} \in \mathbb{R}^{n\times(M-1)}. \tag{8}$$

DFNMF's fairness penalty becomes

$$\mathcal{R}_{\text{int}}(H) \;=\; \big\|F_{\text{int}}^\top H\big\|_F^2, \tag{9}$$

which enforces demographic parity *jointly* over all intersections. Complexity-wise, this adds $O(nk\,M)$ per iteration (for forming $F_{\text{int}}^\top H$). In our SBM study with two attributes (gender: $m_1{=}2$, ethnicity: $m_2{=}5$), $M{=}10$ is modest.

a) Tiny schematic (two attributes).: For gender $\{\mathrm{M},\mathrm{F}\}$ and ethnicity $\{\mathrm{A},\mathrm{W},\mathrm{B},\mathrm{C},\mathrm{D}\}$, the $M{=}10$ joint groups are

$$\{(\mathrm{M},\mathrm{A}),(\mathrm{M},\mathrm{W}),(\mathrm{M},\mathrm{B}),(\mathrm{M},\mathrm{C}),(\mathrm{M},\mathrm{D}),$$
$$(\mathrm{F},\mathrm{A}),(\mathrm{F},\mathrm{W}),(\mathrm{F},\mathrm{B}),(\mathrm{F},\mathrm{C}),(\mathrm{F},\mathrm{D})\}.$$

Each column of G_{int} is the indicator of one joint group; F_{int} is its centered version (with one dropped column).

B. Metrics under Intersectional Fairness

We evaluate standard utility and fairness metrics:

- **Modularity** Q (higher is better).
- **Intersectional balance** $\bar{B}_{\text{int}}$ computed over the M joint groups: $\bar{B}_{\text{int}} = \frac{1}{k}\sum_{l=1}^{k} \min_{g\neq g'} \frac{|V_g \cap C_l|}{|V_{g'} \cap C_l|}$, where g, g' range over joint categories.

C. SBM at Scale: Enforcing Intersectional Parity

We compare *single-attribute* fairness (gender-only; ethnicity-only) against *intersectional* fairness (gender×ethnicity) on SBMs with $k{=}10$ clusters and $n \in \{2\mathrm{K}, 5\mathrm{K}, 10\mathrm{K}\}$ nodes. We fix $\lambda{=}100$ to isolate the effect of the constraint. For single-attribute runs, $\bar{B}$ is computed on that attribute; for the intersectional run, $\bar{B}_{\text{int}}$ is computed on the $M{=}10$ joint groups. The results are illustrated in Table XII.

TABLE XI: Full node→community soft memberships ($\boldsymbol{\Psi} = \boldsymbol{H}_1\boldsymbol{H}_2$, communities I–III).

Node	Clusters		
	I	II	III
1	0.27	0.00	0.00
2	0.26	0.00	0.01
3	0.22	0.00	0.03
4	0.24	0.00	0.00
5	0.33	0.01	0.02
6	0.10	0.00	0.00
7	0.10	0.00	0.00
8	0.08	0.00	0.00
9	0.22	0.00	0.00
10	0.16	0.00	0.03
11	0.13	0.06	0.00
12	0.10	0.05	0.00
13	0.20	0.14	0.03
14	0.11	0.10	0.00
15	0.19	0.09	0.00
16	0.14	0.00	0.00
17	0.07	0.00	0.00
18	0.20	0.00	0.00
19	0.21	0.00	0.00
20	0.24	0.00	0.04
21	0.00	0.08	0.02
22	0.03	0.02	0.03
23	0.00	0.03	0.04
24	0.00	0.01	0.03
25	0.00	0.04	0.07
26	0.00	0.22	0.00
27	0.00	0.22	0.00
28	0.00	0.23	0.00
29	0.00	0.28	0.02
30	0.00	0.20	0.00
31	0.00	0.23	0.00
32	0.00	0.30	0.00
33	0.00	0.26	0.00
34	0.00	0.23	0.00
35	0.00	0.26	0.00
36	0.00	0.11	0.00
37	0.01	0.04	0.01
38	0.00	0.04	0.01
39	0.00	0.05	0.00
40	0.00	0.09	0.00
41	0.00	0.00	0.18
42	0.00	0.00	0.33
43	0.02	0.00	0.24
44	0.00	0.00	0.30
45	0.00	0.00	0.26
46	0.04	0.01	0.11
47	0.00	0.05	0.11
48	0.01	0.01	0.18
49	0.04	0.01	0.22
50	0.04	0.00	0.19
51	0.05	0.02	0.13
52	0.02	0.00	0.06
53	0.01	0.00	0.10
54	0.01	0.01	0.02
55	0.04	0.00	0.06
56	0.02	0.00	0.24
57	0.07	0.01	0.26
58	0.00	0.00	0.22
59	0.00	0.00	0.31
60	0.00	0.00	0.19

TABLE XII: **SBM, intersectional multi-attribute fairness** (k=10, λ=100). Intersectional uses M=10 joint groups (gender×ethnicity).

Attribute	n	M (#groups)	$\mathbf{Q}$	$\bar{\mathbf{B}}$
Gender only	2K	2	1.000	1.000
	5K	2	1.000	1.000
	10K	2	1.000	1.000
Ethnicity only	2K	5	0.119	0.8985
	5K	5	0.124	0.9999
	10K	5	0.124	1.0000
Intersectional (G×E)	2K	10	0.110	0.3522
	5K	10	0.112	0.4263
	10K	10	0.115	0.4507

a) Discussion:

- **Single-attribute parity**. Gender-only aligns with the planted structure in this SBM (both Q and $\bar{B}$ near 1). Ethnicity-only parity achieves very high balance with modest $Q{\approx}0.12$, improving or stabilizing as n grows.
- **Intersectional parity**. Enforcing joint parity across M=10 groups is stricter: $\bar{B}_{\text{int}}$ is lower than single-attribute balances at small n, but *increases with scale* (from 0.35 at 2K to 0.45 at 10K). The utility cost is controlled (slight Q increase with n from 0.110 to 0.115), indicating that larger graphs make intersectional constraints more feasible.
- **Takeaway**. DFNMF can efficiently enforce intersectional demographic parity directly via F_{int}; The soft balanced fairness encoding introduced in Section IV-A allows this extension efficiently. The expected trade-off (tighter fairness $\Rightarrow$ lower Q) is observed, while scaling the graph improves joint feasibility without tuning λ.

REFERENCES

[1] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.

[2] C. Lin, "On the convergence of multiplicative update algorithms for nonnegative matrix factorization," *IEEE Trans. Neural Networks*, vol. 18, no. 6, pp. 1589–1596, 2007.

[3] D. Xu, C. Ruan, E. Körpeoglu, S. Kumar, and K. Achan, "Rethinking neural vs. matrix-factorization collaborative filtering: the theoretical perspectives," in *ICML*, vol. 139, 2021, pp. 11 514–11 524.

[4] S. Rendle, W. Krichene, L. Zhang, and J. R. Anderson, "Neural collaborative filtering vs. matrix factorization revisited," in *RecSys*. ACM, 2020, pp. 240–248.

[5] A. Roy, J. Horstmann, and E. Ntoutsi, "Multi-dimensional discrimination in law and machine learning - A comparative overview," in *FAccT*. ACM, 2023, pp. 89–100.