HM-LDM:

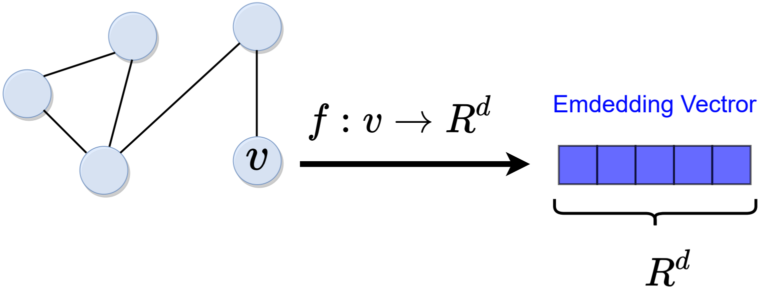
A Hybrid-Membership Latent Distance Model

Nikolaos Nakis Abdulkadir Çelikkanat Morten Mørup

Section for Cognitive Systems, Technical University of Denmark

Graph Representation Learning (*GRL*)

- **Network Embeddings**: Express and predict intrinsic structures of complex networks



DTU

Latent Space Models

Poisson Latent Distance Model[1,2]

- Poisson Rate: $\lambda_{ij} = \exp\left(\gamma_i + \gamma_j - ||\mathbf{z}_i - \mathbf{z}_j||_2\right)$

[1] Hoff, P.D.: Bilinear mixed-effects models for dyadic data. JASA 100(469), 286–295 (2005)
[2] Peter D. Hoff et al, Latent Space Approaches to Social Network Analysis. (2002)

DTU

Latent Space Models

Poisson Latent Distance Model[1,2]

- Poisson Rate: $\lambda_{ij} = \exp\left(\gamma_i + \gamma_j - ||\mathbf{z}_i - \mathbf{z}_j||_2\right)$
- Log-Likelihood: $\log P(Y|\boldsymbol{\lambda})) = \underbrace{\sum_{i<j:y_{ij}=1} \log(\lambda_{ij})}_{\text{Link Term } \mathcal{O}(N)} - \underbrace{\sum_{i<j} \lambda_{ij}}_{\text{Non-Link Term } \mathcal{O}(N^2)}$

[1] Hoff, P.D.: Bilinear mixed-effects models for dyadic data. JASA 100(469), 286–295 (2005)
[2] Peter D. Hoff et al, Latent Space Approaches to Social Network Analysis. (2002)

Latent Space Models

Poisson Latent Distance Model[1,2]

- Poisson Rate: $\lambda_{ij} = \exp\left(\gamma_i + \gamma_j - ||\mathbf{z}_i - \mathbf{z}_j||_2\right)$
- Log-Likelihood: $\log P(Y|\boldsymbol{\lambda})) = \underbrace{\sum_{i<j:y_{ij}=1} \log(\lambda_{ij})}_{\text{Link Term } \mathcal{O}(N)} - \underbrace{\sum_{i<j} \lambda_{ij}}_{\text{Non-Link Term } \mathcal{O}(N^2)}$

Homophily and Transitivity:

[1] Hoff, P.D.: Bilinear mixed-effects models for dyadic data. JASA 100(469), 286–295 (2005)
[2] Peter D. Hoff et al, Latent Space Approaches to Social Network Analysis. (2002)

Latent Space Models

Poisson Latent Distance Model[1,2]

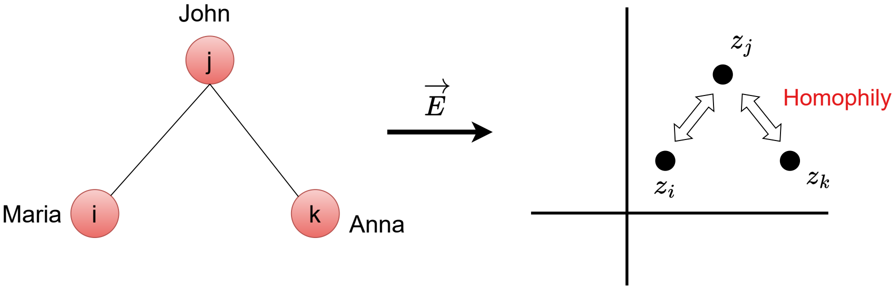
- Poisson Rate: $\lambda_{ij} = \exp\left(\gamma_i + \gamma_j - ||\mathbf{z}_i - \mathbf{z}_j||_2\right)$
- Log-Likelihood: $\log P(Y|\boldsymbol{\lambda})) = \underbrace{\sum_{i<j:y_{ij}=1} \log(\lambda_{ij})}_{\text{Link Term } \mathcal{O}(N)} - \underbrace{\sum_{i<j} \lambda_{ij}}_{\text{Non-Link Term } \mathcal{O}(N^2)}$

Homophily and Transitivity:

Similar Nodes are positioned closer in space

[1] Hoff, P.D.: Bilinear mixed-effects models for dyadic data. JASA 100(469), 286–295 (2005)
[2] Peter D. Hoff et al, Latent Space Approaches to Social Network Analysis. (2002)

A small example



$$d_{z_i,z_k} \leq ||z_j - z_i||_2 + ||z_k - z_j||_2$$

Transitivity

Figure 1: Projecting a small social network

Latent Distance Models — Identifiability

The LDM embeddings are invariant under translation, rotation and reflection operations [3]

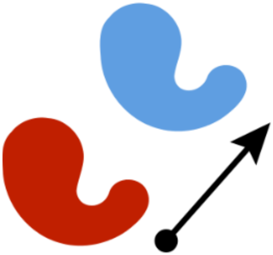


Figure 2: Translation

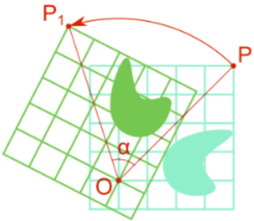


Figure 3: Rotation

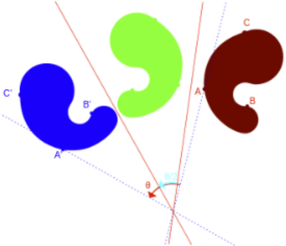


Figure 4: Reflection

[3] Images from Wikipedia

Background — Simplexes

- A D-simplex is a convex polytope in $\mathbb{R}^D$ and it is always the convex hull of $D+1$ affinely independent points
- Standard D-simplex:
  $\Delta^D = \left\{ (w_1, \ldots, w_{D+1} \,\middle|\, \sum_{i=1}^{D+1} w_i = 1 \text{ and } w_i \geq 0 \text{ for } i = 1, \ldots, D+1 \right\}$

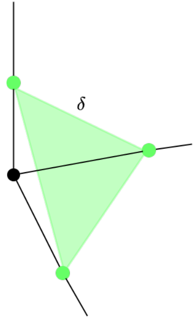


Figure 5: The standard 2-simplex in $\mathbb{R}^3$ for $\delta = 1$

Background — Non-Negative Matrix Factorization

- Non-Negative Matrix Factorization (NMF): Factorization of a non-negative matrix $V \in \mathbb{R}^{N \times K}$ into two (usually) matrices $W \in \mathbb{R}^{N \times D}$ and $H \in \mathbb{R}^{D \times K}$, also non-negative.
- NMF techniques: structure retrieval and interpretable part-based representations [4], clustering properties

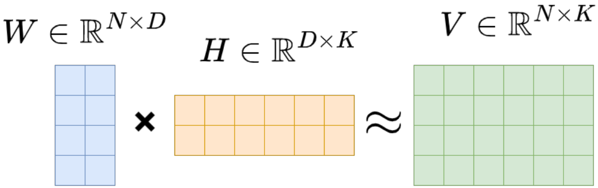


Figure 6: Non-symmetric NMF

[4] Lee, D.D., Seung, H.S.: Learning the parts of objects by nonnegative matrix factorization. Nature 401, 788–791 (1999)

Background — Non-Negative Matrix Factorization

- Symmetric Non-Negative Matrix Factorization

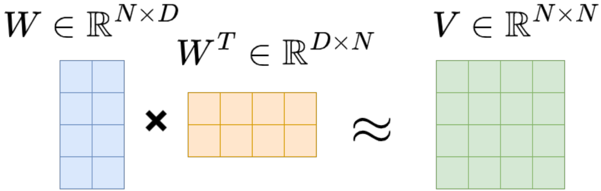


Figure 7: Symmetric NMF

Latent Space Models

Hybrid Membership-Latent Distance Model (HM-LDM)

- Poisson Rate: $\lambda_{ij} = \exp\left(\gamma_i + \gamma_j - \delta^p \cdot ||\mathbf{w}_i - \mathbf{w}_j||_2^p\right)$,

  where the node embeddings $\mathbf{w_i} \in [0,1]^{D+1}$ and $\sum_{d=1}^{D+1} w_{id} = 1$, $\delta \in \mathbb{R}_+$, $p \in \{1,2\}$ and $\gamma_i \in \mathbb{R}$ denotes the node-specific random-effects.

Latent Space Models

Hybrid Membership-Latent Distance Model (HM-LDM)

- Poisson Rate: $\lambda_{ij} = \exp\left(\gamma_i + \gamma_j - \delta^p \cdot ||\mathbf{w}_i - \mathbf{w}_j||_2^p\right)$,
  where the node embeddings $\mathbf{w_i} \in [0,1]^{D+1}$ and $\sum_{d=1}^{D+1} w_{id} = 1$, $\delta \in \mathbb{R}_+$, $p \in \{1,2\}$ and $\gamma_i \in \mathbb{R}$ denotes the node-specific random-effects.
- Log-Likelihood: $\log P(Y|\boldsymbol{\lambda})) = \underbrace{\sum_{i<j:y_{ij}=1} \log(\lambda_{ij})}_{\text{Link Term } \mathcal{O}(N)} - \underbrace{\sum_{i<j} \lambda_{ij}}_{\text{Non-Link Term } \mathcal{O}(N^2)}$

Latent Space Models

Hybrid Membership-Latent Distance Model (HM-LDM)

- Poisson Rate: $\lambda_{ij} = \exp\left(\gamma_i + \gamma_j - \delta^p \cdot ||\mathbf{w}_i - \mathbf{w}_j||_2^p\right),$

  where the node embeddings $\mathbf{w_i} \in [0,1]^{D+1}$ and $\sum_{d=1}^{D+1} w_{id} = 1$, $\delta \in \mathbb{R}_+$, $p \in \{1,2\}$ and $\gamma_i \in \mathbb{R}$ denotes the node-specific random-effects.
- Log-Likelihood: $\log P(Y|\boldsymbol{\lambda})) = \underbrace{\sum_{i<j:y_{ij}=1} \log(\lambda_{ij})}_{\text{Link Term } \mathcal{O}(N)} - \underbrace{\sum_{i<j} \lambda_{ij}}_{\text{Non-Link Term } \mathcal{O}(N^2)}$

Constrain the embeddings on the *D*-simplex

Latent Space Models

Hybrid Membership-Latent Distance Model (HM-LDM)

- Poisson Rate: $\lambda_{ij} = \exp\left(\gamma_i + \gamma_j - \delta^p \cdot ||\mathbf{w}_i - \mathbf{w}_j||_2^p\right),$

  where the node embeddings $\mathbf{w_i} \in [0,1]^{D+1}$ and $\sum_{d=1}^{D+1} w_{id} = 1$, $\delta \in \mathbb{R}_+$, $p \in \{1,2\}$ and $\gamma_i \in \mathbb{R}$ denotes the node-specific random-effects.
- Log-Likelihood: $\log P(Y|\boldsymbol{\lambda})) = \underbrace{\sum_{i<j:y_{ij}=1} \log(\lambda_{ij})}_{\text{Link Term } \mathcal{O}(N)} - \underbrace{\sum_{i<j} \lambda_{ij}}_{\text{Non-Link Term } \mathcal{O}(N^2)}$

Constrain the embeddings on the D-simplex

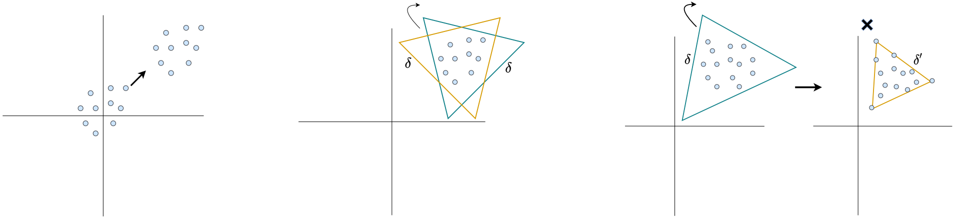
Control the simplex volume by the δ value

Identifiability of the embedding solution

Definition (Identifiabilty)

An embedding matrix $\mathbf{W}$ whose rows indicating the corresponding node representations is called an *identifiable solution up to a permutation* if it holds $\widetilde{\mathbf{W}} = \mathbf{WP}$ for a permutation $\mathbf{P}$ and a solution $\widetilde{\mathbf{W}} \neq \mathbf{W}$.

Road to identifiability



(a) Translation invariances.

(b) Rotation invariances.

(c) Decreased simplex volume ensuring identifiability.

Figure 8: A 2-dimensional latent space with the 2-simplex given as the green and yellow triangles, the blue points denote embedding positions of the LDM and δ is the simplex size.

Nodes residing in the simplex corners

Definition (Community champion)

A node for a latent community is called *champion* if it belongs to the community (simplex corner) while forming a binary unit vector.

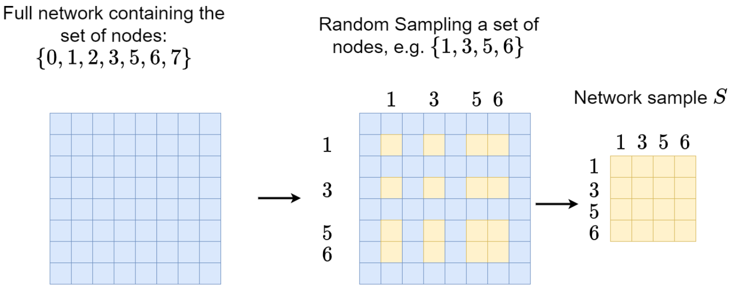
Relation to NMF — HM-LDM ($p = 2$)

- Re-parameterization of log rate by
  $\gamma_i + \gamma_j - \delta^2 \cdot ||\mathbf{w}_i - \mathbf{w}_j||_2^2 = \tilde{\gamma}_i + \tilde{\gamma}_j + 2\delta^2 \cdot (\mathbf{w}_i \mathbf{w}_j^\top)$
- $\mathbf{W}\mathbf{W}^\top$ defines a symmetric NMF problem
  - identifiable and unique factorization (up to permutation invariance) when **W** is full-rank and at least one node resides solely in each simplex corner, ensuring separability [5] [6].

[5] Huang, K. et al : Non-negative matrix factorization revis- ited: Uniqueness and algorithm for symmetric decomposition. IEEE Trans. Signal Process 62(1), 211–224 (2014)

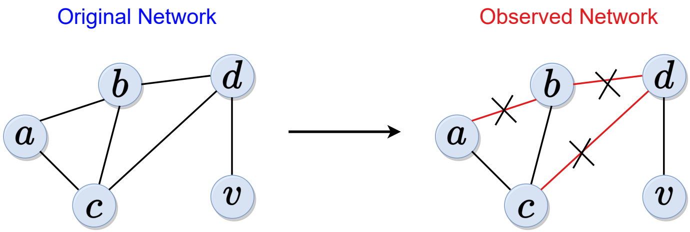
[6] Mao, X. et al: On mixed memberships and symmetric non- negative matrix factorizations. In: ICML. vol. 70 (2017)

Scalable inference — Unbiased likelihood estimators

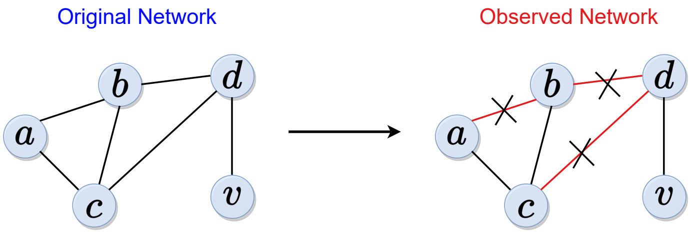


$$\log P(Y|\boldsymbol{\lambda})) = \underbrace{\sum_{i<j:y_{ij}=1} \log(\lambda_{ij})}_{\text{Link Term } \mathcal{O}(S)} - \underbrace{\sum_{i<j} \lambda_{ij}}_{\text{Non-Link Term } \mathcal{O}(S^2)} \quad \text{with } i,j \in S$$

Downstream tasks: link prediction

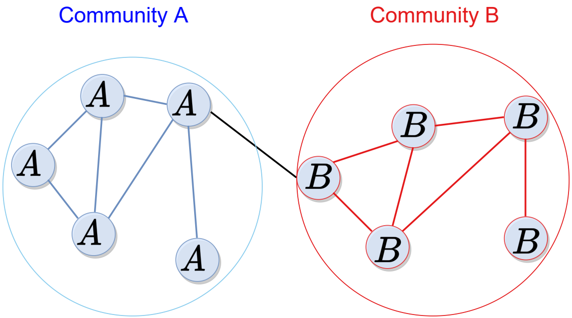


Downstream tasks: link prediction

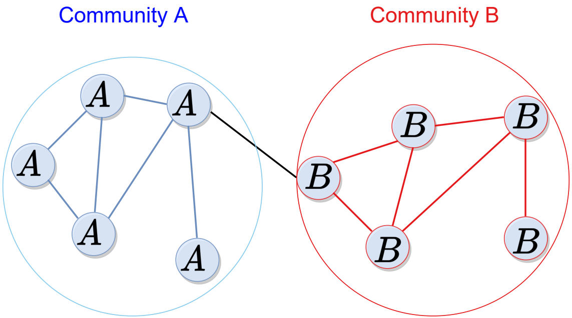


Can we predict the missing (removed) links?

Downstream tasks: community detection



Downstream tasks: community detection



Can we infer the community labels?

Modeling — Real networks

- For our experiments we consider eight networks [7,8]

Table 1: Network statistics; $|\mathcal{V}|$: # Nodes, $|\mathcal{E}|$: # Edges, $|\mathcal{K}|$: # Communities.

	AstroPh	*GrQc*	*Facebook*	*HepTh*	*Hamilton*	*Amherst*	*Rochester*	*Mich*
$\|\mathcal{V}\|$	17,903	5,242	4,039	8,638	2,118	2,021	4,145	2,933
$\|\mathcal{E}\|$	197,031	14,496	88,234	24,827	87,486	87,496	145,305	54,903
$\|\mathcal{K}\|$	-	-	-	-	15	15	19	13

[7] Leskovec, J., Krevl, A.: SNAP Datasets: Stanford large network dataset collection (2014)
[8] Mucha, P., Porter, M.: Social structure of facebook networks. Physica A: Statistical Mechanics and its Applications 391, 4165–4180 (2012)

Experiments — Community champions and δ values

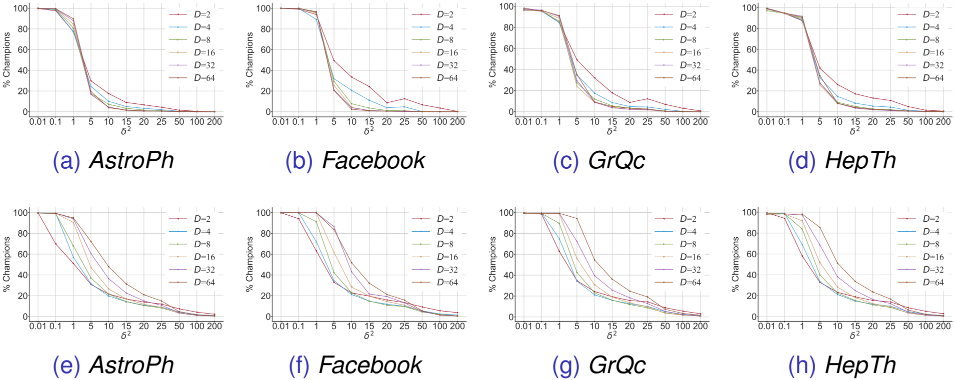


Figure 9: Total community champions (%) in terms of δ^2 across dimensions for HM-LDM. Top row: $p = 2$. Bottom row $p = 1$.

Experiments — Link prediction and δ values

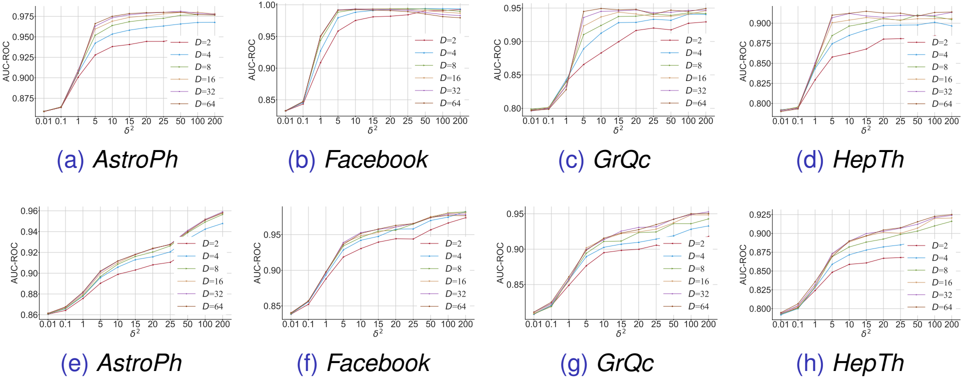


Figure 10: AUC-ROC scores as a function of δ^2 across dimensions for HM-LDM. Top row: $p = 2$. Bottom row $p = 1$.

Experiments — Identifiable HM-LDM link prediction vs baselines

Table 2: AUC-ROC scores for varying representation sizes.

	AstroPh			*GrQc*			*Facebook*			*HepTh*		
Dimension (D)	8	16	32	8	16	32	8	16	32	8	16	32
DeepWalk	.945	.950	.952	.919	.916	.929	.986	.986	.984	.874	.867	.873
Node2Vec	.950	.962	.957	.897	.913	.930	.988	.988	.987	.881	.882	.881
LINE	.909	.938	.947	.920	.925	.919	.981	.987	.983	.873	.886	.882
NetMF	.813	.823	.839	.860	.866	.877	.935	.963	.971	.792	.806	.821
NetSMF	.891	.901	.919	.837	.858	.886	.975	.981	.985	.809	.822	.836
LouvainNE	.813	.811	.819	.868	.875	.873	.958	.961	.963	.874	.867	.873
ProNE	.907	.929	.947	.885	.911	.921	.971	.982	.987	.827	.846	.859
NNSED	.861	.882	.891	.792	.808	.828	.908	.927	.935	.756	.779	.796
MNMF	.893	.925	.943	.911	.928	.937	.965	.978	.982	.857	.880	.891
BigClam	.500	.723	.810	.752	.769	.780	.744	.722	.647	.776	.700	.748
SymmNMF	.767	.779	.800	.729	.772	.835	.933	.942	.951	.696	.727	.766
HM-LDM ($p = 1$)	.956	.952	.952	**.944**	**.948**	**.951**	.982	.979	.974	**.916**	**.921**	**.924**
HM-LDM ($p = 2$)	**.972**	**.973**	**.963**	.940	.942	.946	**.992**	**.993**	**.993**	.908	.910	.911

Experiments — Community detection vs baselines

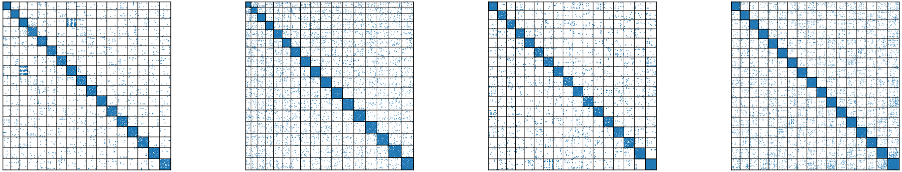
Table 3: Normalized Mutual Information (NMI) and Adjusted Rand Index (ARI) scores for networks with ground-truth communities, setting $\delta = 1$.

	Amherst		*Rochester*		*Mich*		*Hamilton*	
Metric	NMI	ARI	NMI	ARI	NMI	ARI	NMI	ARI
DEEPWALK	.498	.347	.348	.205	.207	.157	.447	.303
NODE2VEC	.535	.375	.364	.223	.217	.161	.481	.348
LINE	.549	.452	.365	.217	**.249**	.192	.499	.411
NETMF	.491	.330	.377	.243	.237	.136	.456	.297
NETSMF	<u>.562</u>	.408	<u>.381</u>	.228	<u>.242</u>	.169	.494	.391
LOUVAINNE	<u>.562</u>	.395	.347	.204	.175	.114	.475	.334
PRONE	.536	.443	.356	.312	.229	<u>.200</u>	.478	.396
NNSED	.295	.243	.168	.116	.064	.035	.335	.285
MNMF	.542	.362	.324	.171	.188	.102	.466	.287
BIGCLAM	.091	.066	.028	.022	.024	.015	.053	.041
SYMMNMF	**.596**	.397	.308	.175	.207	.088	.437	.341
HM-LDM($p = 1$)	<u>.562</u>	<u>.502</u>	**.400**	**.392**	.228	**.205**	**.527**	<u>.485</u>
HM-LDM($p = 2$)	.539	**.506**	<u>.384</u>	<u>.373</u>	.217	.183	<u>.507</u>	**.504**

The extracted identifiable node embeddings convey information about (latent) community memberships

DTU

Experiments — Latent community structures



(a) *GrQc* ($p = 2$) (b) *HepTh* ($p = 2$) (c) *GrQc* ($p = 1$) (d) *HepTh* ($p = 1$)

Figure 11: Ordered adjacency matrices based on the memberships of a $D = 16$ dimensional HM-LDM with δ values ensuring identifiability.

Comparison with the vanilla LDM

Table 4: AUC-ROC HM-LDM and LDM-RE comparison for the link prediction task.

	AstroPh			GrQc			Facebook			HepTh		
Dimension (D)	8	16	32	8	16	32	8	16	32	8	16	32
LDM-RE	.973	.974	.979	.949	.952	.954	.993	.994	.992	.920	.923	.923
HM-LDM($p=1, \delta^2=$ identifiable)	.956	.952	.952	.944	.948	.951	.982	.979	.974	.916	.921	.924
HM-LDM($p=1, \delta^2=10^3$)	.967	.967	.965	.956	.955	.951	.985	.986	.987	.932	.931	.926
LDM-RE-$(\ell^2)^2$	.979	.978	.976	.944	.944	.945	.990	.990	.991	.913	.912	.909
HM-LDM($p=2, \delta^2=$ identifiable)	.972	.973	.963	.940	.942	.946	.992	.993	.993	.908	.910	.911
HM-LDM($p=2, \delta^2=10^3$)	.984	.983	.980	.948	.946	.946	.991	.991	.992	.920	.918	.913

Table 5: HM-LDM and LDM-RE comparison for the clustering task.

	Amherst		Rochester		Mich		Hamilton	
Metric	NMI	ARI	NMI	ARI	NMI	ARI	NMI	ARI
LDM-RE	.548	.366	.391	.212	.230	.132	.491	.320
HM-LDM($p=1, \delta^2=$ identifiable)	.562	.502	.400	.392	.228	.205	.527	.485
HM-LDM($p=1, \delta^2=10^3$)	.439	.386	.308	.303	.176	.133	.405	.377
LDM-RE-$(\ell^2)^2$	.546	.370	.393	.211	.231	.137	.497	.327
HM-LDM($p=2, \delta^2=$ identifiable)	.539	.506	.384	.373	.217	.183	.507	.504
HM-LDM($p=2, \delta^2=10^3$)	.240	.133	.206	.119	.116	.056	.232	.209

Conclusion

- Constrain LDMs to the simplex without loss of expressive power
- Reduced simplex: unique representations, ultimately resulting in hard clustering of nodes to communities
- Combination of the important network characteristics of homophily and transitivity with latent community detection
- Enabling explicit control of soft and hard assignment through the volume of the induced simplex

Thank you!