

---

# Neural Nonnegative CP Decomposition for Hierarchical Tensor Analysis

---

Joshua Vendrow, *Jamie Haddock*, Deanna Needell

JMM MSRI Special Session on Tensor Modeling and Optimization

April 9, 2022

appears in ACSSC 2021

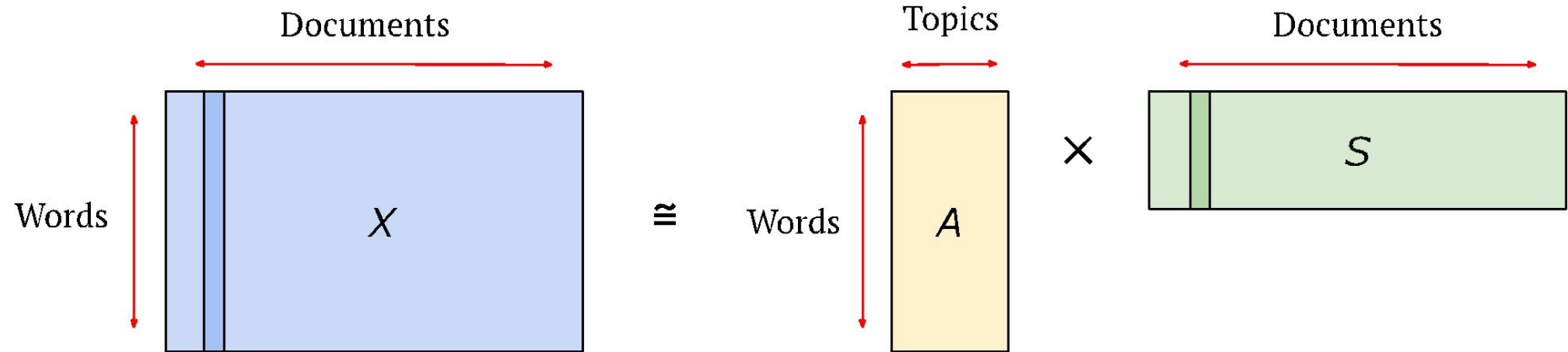
supported by NSF DMS #2211318

# Background

# Nonnegative Matrix Factorization (NMF)

Given a nonnegative matrix  $X$  ( $X \geq 0$ ), compute nonnegative  $A$  and  $S$  such that

$$X \approx AS$$

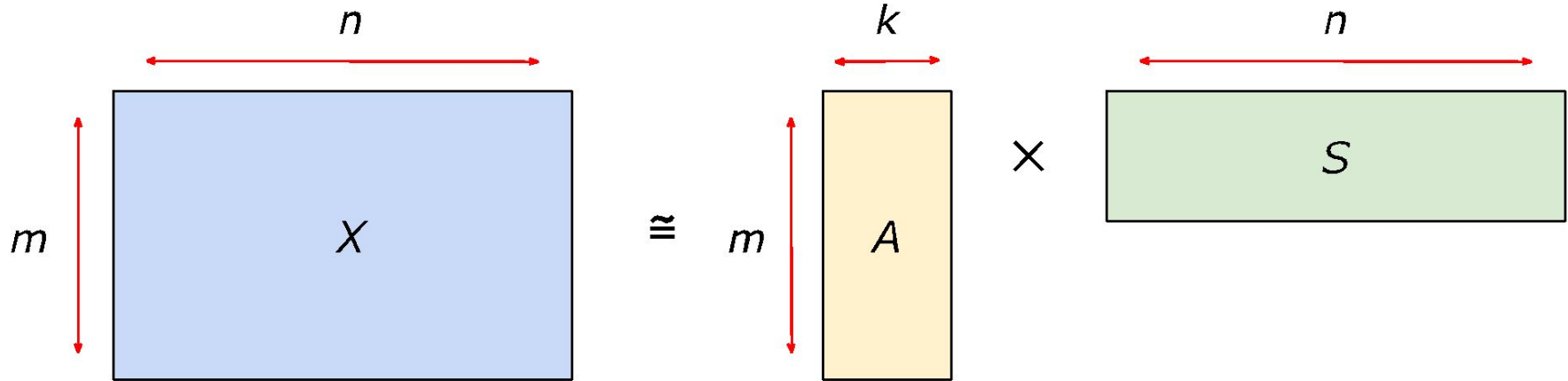


D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," Nature, vol. 401, no. 6755, pp. 788, 1999.

# Nonnegative Matrix Factorization (NMF)

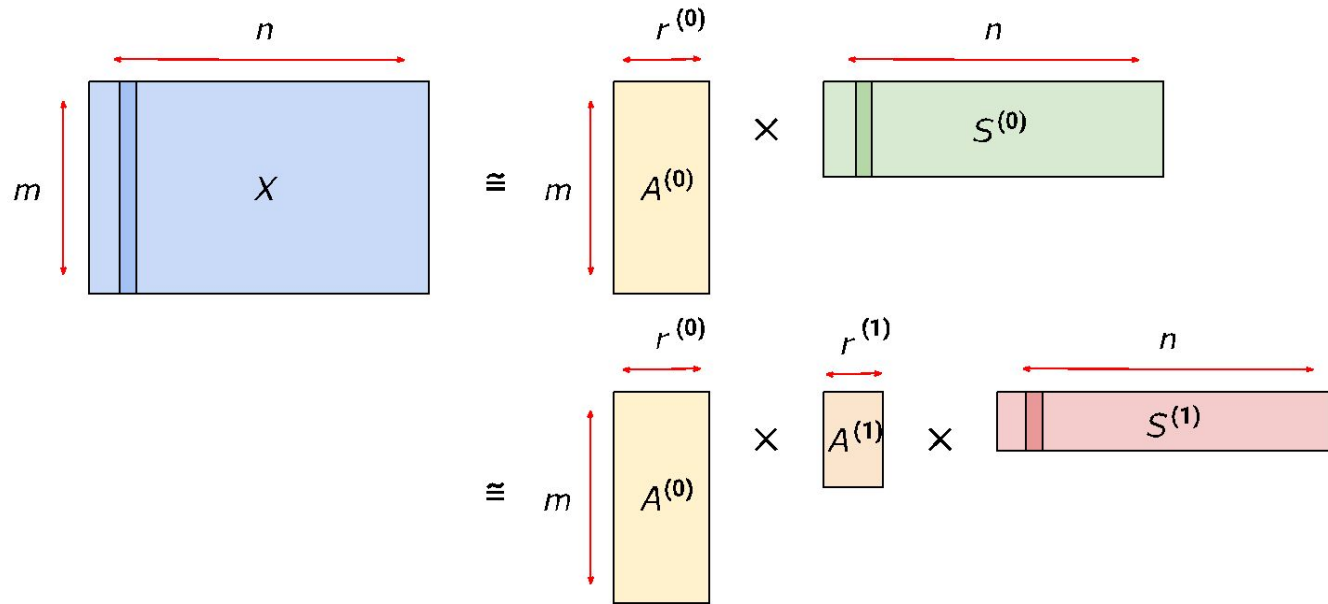
Formulated as the optimization task:

$$\arg \min_{A \geq 0, S \geq 0} \|X - AS\|_F^2$$



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

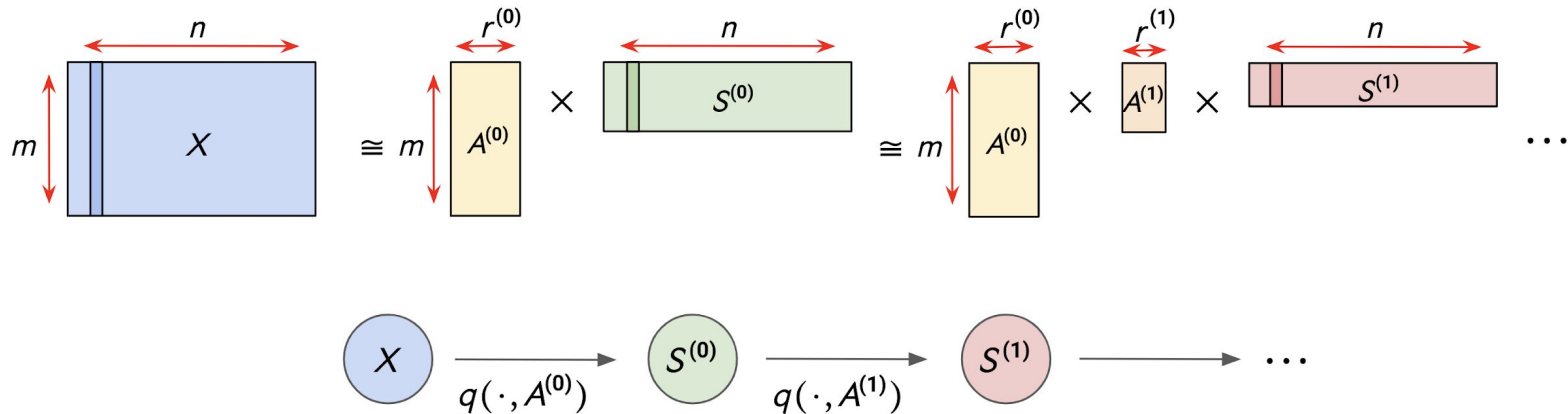
# Hierarchical NMF



# Neural NMF

Regard the  $A$  matrices as weights, and determine  $S$  matrices from  $A$  matrices, and define

$$q(X, A) := \operatorname{argmin}_{S \geq 0} \|X - AS\|_F^2$$

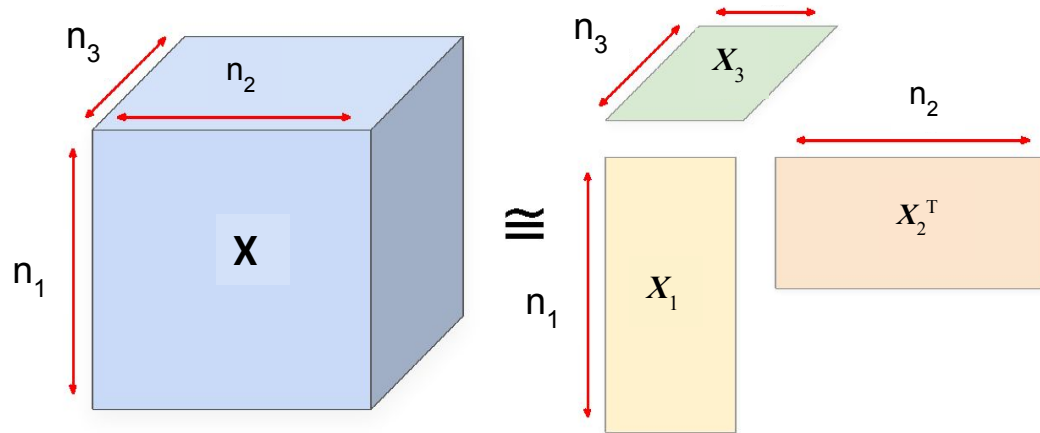


M. Gao et al. "Neural nonnegative matrix factorization for hierarchical multilayer topic modeling," in Proc. Int. Workshop on Comp. Adv. in Multi-Sensor Adaptive Process., 2019.

# Nonnegative CP Decomposition (NCPD)

Given a nonnegative tensor  $\mathbf{X}$  ( $\mathbf{X} \geq 0$ ), compute nonnegative  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_k$  such that

$$\mathbf{X} \approx \llbracket \mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_k \rrbracket \equiv \sum_{j=1}^r \mathbf{x}_j^{(1)} \otimes \mathbf{x}_j^{(2)} \otimes \dots \otimes \mathbf{x}_j^{(k)}$$



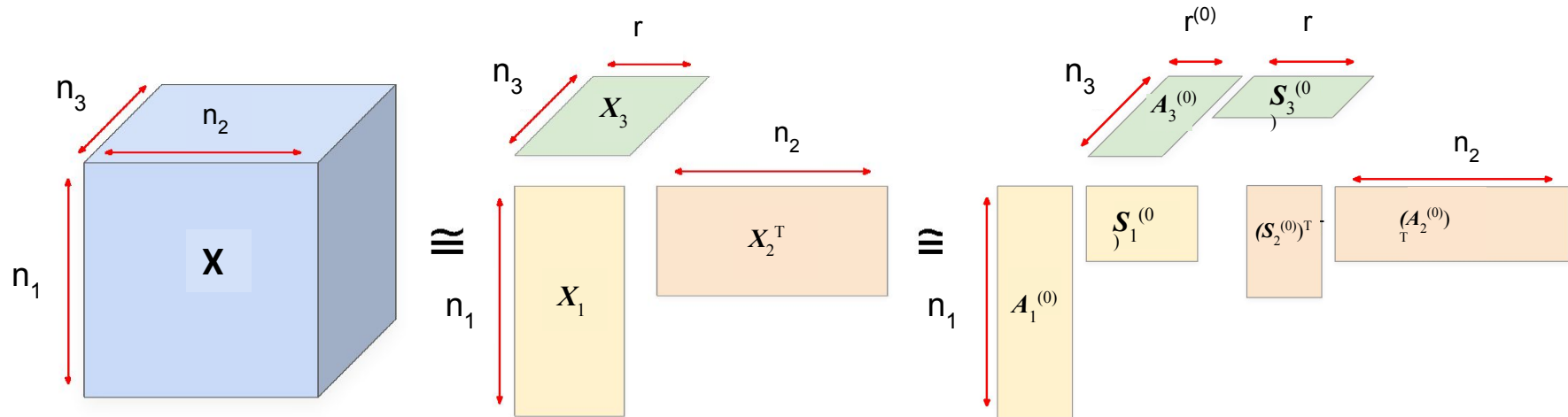
# Our Contributions



# Hierarchical Nonnegative CP Decomposition (HNCPD)

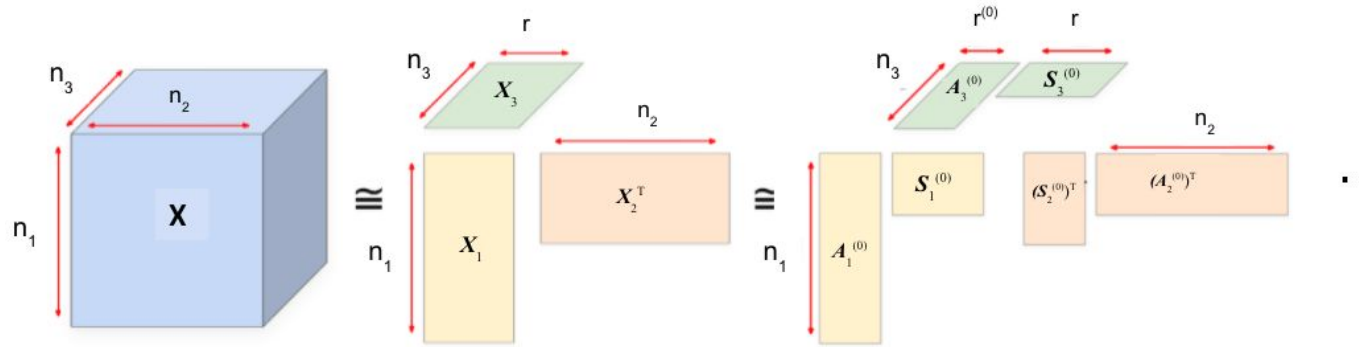
Apply a hierarchical NMF onto each factor matrix:

$$\mathbf{X}_i \approx \widetilde{\mathbf{X}}_i \equiv \mathbf{A}_i^{(0)} \mathbf{A}_i^{(1)} \dots \mathbf{A}_i^{(\ell-2)} \mathbf{S}_i^{(\ell-2)}$$

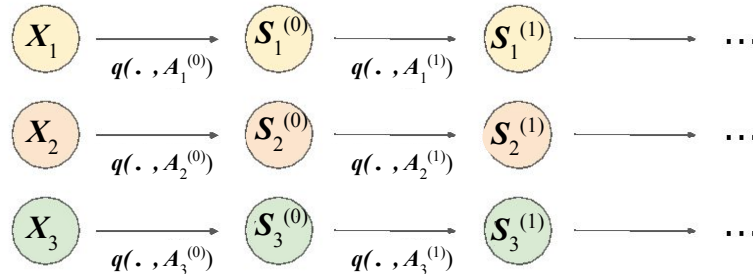


# Neural Nonnegative CP Decomposition (Neural NCPD)

## Hierarchical NCPD Model



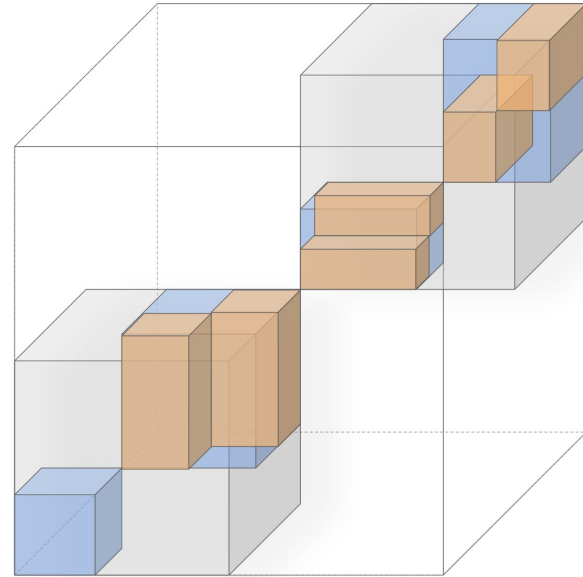
## Neural NMF Branches



# Experiments

# Synthetic Tensor Data Set Experiments

- CP rank seven tensor of size  $40 \times 40 \times 40$
- Overlapping and non-overlapping blocks of varying size and intensity to form a hierarchical structure
- Positive random noise added to each entry



# Synthetic Tensor Data Set Experiments

TABLE I: Topic modeling loss and relative reconstruction loss,  $C_{\text{rel}}$ , on the synthetic dataset for Neural NCPD, Standard HNCPD, HNTF, Neural NMF, and Standard HNMF with two levels of noise over 10 trials. For HNTF we report runs on three re-orderings of the modes the tensor, and for matrix methods we report results for flattening along each mode of the tensor.

Method	Mode	Topic Modeling Loss						Relative Reconstruction Loss					
		$\sigma^2 = 0.1$			$\sigma^2 = 0.4$			$\sigma^2 = 0.1$			$\sigma^2 = 0.4$		
		7 - 2	7 - 4	4 - 2	7 - 2	7 - 4	4 - 2	$r = 7$	$r^{(0)} = 4$	$r^{(1)} = 2$	$r = 7$	$r^{(0)} = 4$	$r^{(1)} = 2$
Neural HNCPD		<b>0.043</b>	0.042	<b>0.042</b>	<b>0.087</b>	<b>0.087</b>	<b>0.081</b>	0.119	0.252	0.563	<b>0.454</b>	0.508	0.714
Standard HNCPD		0.106	0.101	0.189	0.145	0.193	0.204	0.119	0.494	0.828	<b>0.454</b>	0.612	0.892
HNTF [12]	1	0.163	0.236	0.182	0.171	0.144	0.170	0.119	0.502	0.795	<b>0.454</b>	0.576	0.781
	2	0.087	<b>0.040</b>	0.101	0.090	0.116	0.142	0.119	0.309	0.665	<b>0.454</b>	0.587	0.765
	3	0.078	0.122	0.106	0.084	0.111	0.161	0.119	0.417	0.713	<b>0.454</b>	0.560	0.747
Neural NMF [9]	1	0.151	0.192	0.105	0.169	0.219	0.127	0.146	0.268	0.593	0.478	0.521	0.705
	2	0.075	0.244	0.146	0.153	0.190	0.160	0.141	0.289	0.585	0.475	0.513	0.710
	3	0.119	0.164	0.110	0.158	0.197	0.140	0.151	0.236	0.576	0.477	0.512	<b>0.693</b>
Standard HNMF	1	0.098	0.182	0.052	0.164	0.219	0.139	<b>0.118</b>	0.235	<b>0.558</b>	0.472	0.524	0.707
	2	0.080	0.199	0.090	0.151	0.213	0.088	<b>0.118</b>	0.245	0.566	0.472	0.505	0.709
	3	0.060	0.165	0.085	0.137	0.193	0.114	<b>0.118</b>	<b>0.233</b>	0.563	0.472	<b>0.503</b>	0.717

# Synthetic Tensor Data Set Experiments

TABLE I: Topic modeling loss and relative reconstruction loss,  $C_{\text{rel}}$ , on the synthetic dataset for Neural NCPD, Standard HNCPD, HNTF, Neural NMF, and Standard HNMF with two levels of noise over 10 trials. For HNTF we report runs on three re-orderings of the modes the tensor, and for matrix methods we report results for flattening along each mode of the tensor.

Method	Mode	Topic Modeling Loss						Relative Reconstruction Loss					
		$\sigma^2 = 0.1$			$\sigma^2 = 0.4$			$\sigma^2 = 0.1$			$\sigma^2 = 0.4$		
		7 - 2	7 - 4	4 - 2	7 - 2	7 - 4	4 - 2	$r = 7$	$r^{(0)} = 4$	$r^{(1)} = 2$	$r = 7$	$r^{(0)} = 4$	$r^{(1)} = 2$
Neural HNCPD		<b>0.043</b>	0.042	<b>0.042</b>	<b>0.087</b>	<b>0.087</b>	<b>0.081</b>	0.119	0.252	0.563	<b>0.454</b>	0.508	0.714
Standard HNCPD		0.106	0.101	0.189	0.145	0.193	0.204	0.119	0.494	0.828	<b>0.454</b>	0.612	0.892
HNTF [12]	1	0.163	0.236	0.182	0.171	0.144	0.170	0.119	0.502	0.795	<b>0.454</b>	0.576	0.781
	2	0.087	<b>0.040</b>	0.101	0.090	0.116	0.142	0.119	0.309	0.665	<b>0.454</b>	0.587	0.765
	3	0.078	0.122	0.106	0.084	0.111	0.161	0.119	0.417	0.713	<b>0.454</b>	0.560	0.747
Neural NMF [9]	1	0.151	0.192	0.105	0.169	0.219	0.127	0.146	0.268	0.593	0.478	0.521	0.705
	2	0.075	0.244	0.146	0.153	0.190	0.160	0.141	0.289	0.585	0.475	0.513	0.710
	3	0.119	0.164	0.110	0.158	0.197	0.140	0.151	0.236	0.576	0.477	0.512	<b>0.693</b>
Standard HNMF	1	0.098	0.182	0.052	0.164	0.219	0.139	<b>0.118</b>	0.235	<b>0.558</b>	0.472	0.524	0.707
	2	0.080	0.199	0.090	0.151	0.213	0.088	<b>0.118</b>	0.245	0.566	0.472	0.505	0.709
	3	0.060	0.165	0.085	0.137	0.193	0.114	<b>0.118</b>	<b>0.233</b>	0.563	0.472	<b>0.503</b>	0.717

# Synthetic Tensor Data Set Experiments

TABLE I: Topic modeling loss and relative reconstruction loss,  $C_{\text{rel}}$ , on the synthetic dataset for Neural NCPD, Standard HNCPD, HNTF, Neural NMF, and Standard HNMF with two levels of noise over 10 trials. For HNTF we report runs on three re-orderings of the modes the tensor, and for matrix methods we report results for flattening along each mode of the tensor.

Method	Mode	Topic Modeling Loss						Relative Reconstruction Loss					
		$\sigma^2 = 0.1$			$\sigma^2 = 0.4$			$\sigma^2 = 0.1$			$\sigma^2 = 0.4$		
		7 - 2	7 - 4	4 - 2	7 - 2	7 - 4	4 - 2	$r = 7$	$r^{(0)} = 4$	$r^{(1)} = 2$	$r = 7$	$r^{(0)} = 4$	$r^{(1)} = 2$
Neural HNCPD		<b>0.043</b>	0.042	<b>0.042</b>	<b>0.087</b>	<b>0.087</b>	<b>0.081</b>	0.119	0.252	0.563	<b>0.454</b>	0.508	0.714
Standard HNCPD		0.106	0.101	0.189	0.145	0.193	0.204	0.119	0.494	0.828	<b>0.454</b>	0.612	0.892
HNTF [12]	1	0.163	0.236	0.182	0.171	0.144	0.170	0.119	0.502	0.795	<b>0.454</b>	0.576	0.781
	2	0.087	<b>0.040</b>	0.101	0.090	0.116	0.142	0.119	0.309	0.665	<b>0.454</b>	0.587	0.765
	3	0.078	0.122	0.106	0.084	0.111	0.161	0.119	0.417	0.713	<b>0.454</b>	0.560	0.747
Neural NMF [9]	1	0.151	0.192	0.105	0.169	0.219	0.127	0.146	0.268	0.593	0.478	0.521	0.705
	2	0.075	0.244	0.146	0.153	0.190	0.160	0.141	0.289	0.585	0.475	0.513	0.710
	3	0.119	0.164	0.110	0.158	0.197	0.140	0.151	0.236	0.576	0.477	0.512	<b>0.693</b>
Standard HNMF	1	0.098	0.182	0.052	0.164	0.219	0.139	<b>0.118</b>	0.235	<b>0.558</b>	0.472	0.524	0.707
	2	0.080	0.199	0.090	0.151	0.213	0.088	<b>0.118</b>	0.245	0.566	0.472	0.505	0.709
	3	0.060	0.165	0.085	0.137	0.193	0.114	<b>0.118</b>	<b>0.233</b>	0.563	0.472	<b>0.503</b>	0.717

# Synthetic Tensor Data Set Experiments

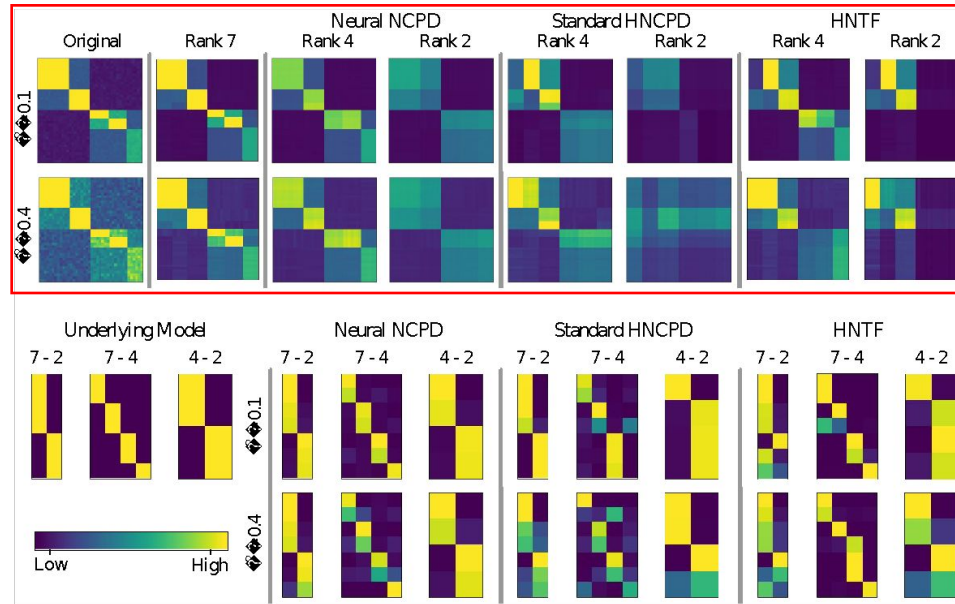


Fig. 2: (Top left) Data tensor  $\mathbf{X}$  with two levels of noise. (Top right) ranks 7, 5, and 3 Neural NCPD, Standard HNCPCD, and HNTF approximations of  $\mathbf{X}$ . (Bottom left) Underlying topic modelling matrix. (Bottom right) topic modelling matrices discovered by Neural NCPD, Standard HNCPCD, and HNTF, with rows and columns permuted to minimize topic error.



# Synthetic Tensor Data Set Experiments

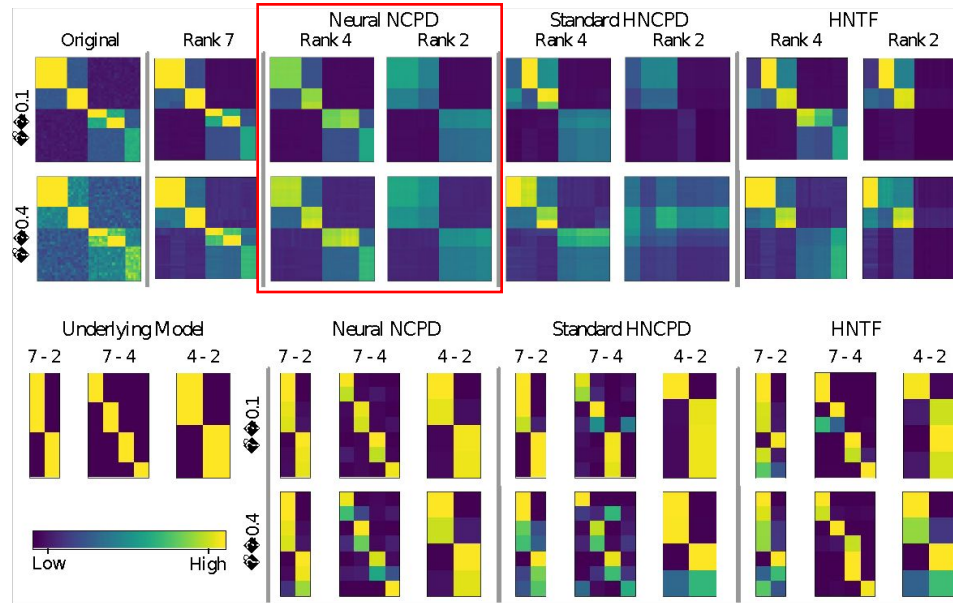


Fig. 2: (Top left) Data tensor  $\mathbf{X}$  with two levels of noise. (Top right) ranks 7, 5, and 3 Neural NCPD, Standard HNCPCD, and HNTF approximations of  $\mathbf{X}$ . (Bottom left) Underlying topic modelling matrix. (Bottom right) topic modelling matrices discovered by Neural NCPD, Standard HNCPCD, and HNTF, with rows and columns permuted to minimize topic error.

# Synthetic Tensor Data Set Experiments

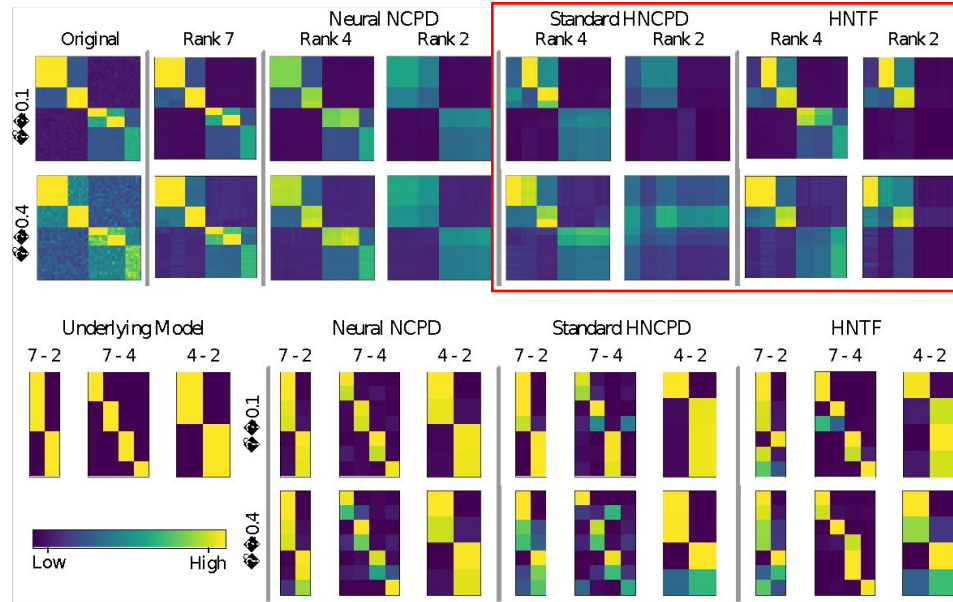


Fig. 2: (Top left) Data tensor  $\mathbf{X}$  with two levels of noise. (Top right) ranks 7, 5, and 3 Neural NCPD, Standard HNCPCD, and HNTF approximations of  $\mathbf{X}$ . (Bottom left) Underlying topic modelling matrix. (Bottom right) topic modelling matrices discovered by Neural NCPD, Standard HNCPCD, and HNTF, with rows and columns permuted to minimize topic error.

# Synthetic Tensor Data Set Experiments

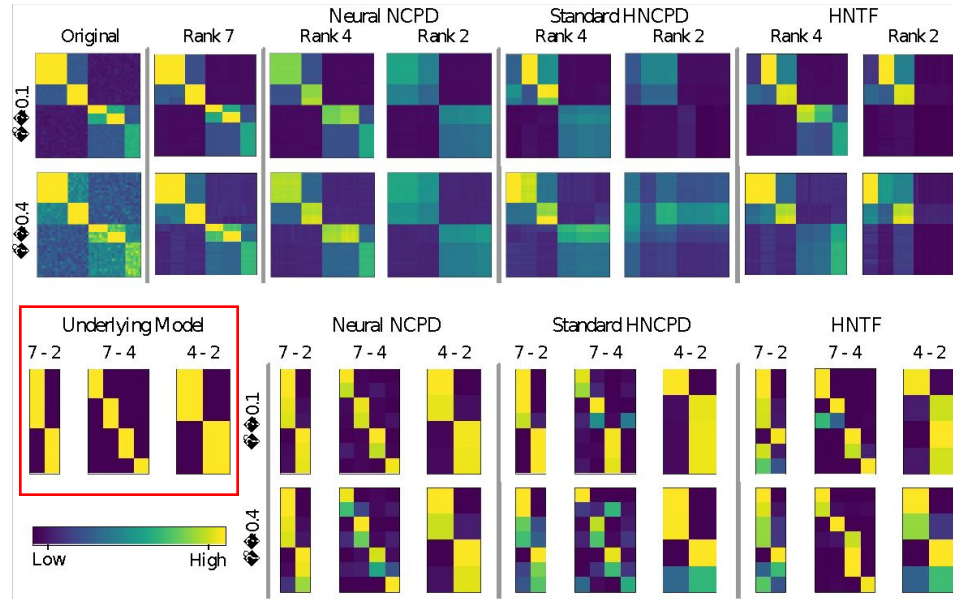


Fig. 2: (Top left) Data tensor  $\mathbf{X}$  with two levels of noise. (Top right) ranks 7, 5, and 3 Neural NCPD, Standard HNCPCD, and HNTF approximations of  $\mathbf{X}$ . (Bottom left) Underlying topic modelling matrix. (Bottom right) topic modelling matrices discovered by Neural NCPD, Standard HNCPCD, and HNTF, with rows and columns permuted to minimize topic error.

# Synthetic Tensor Data Set Experiments

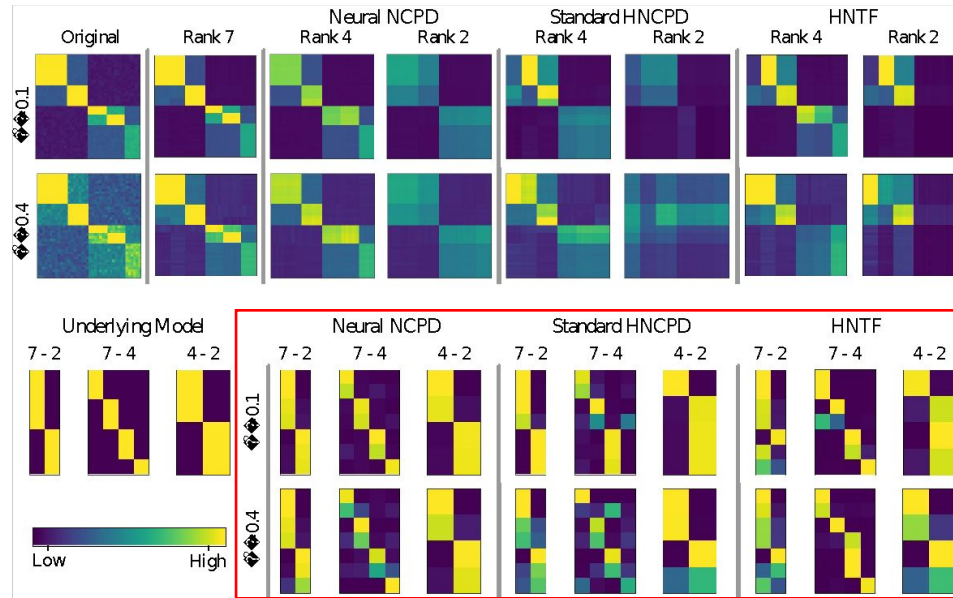
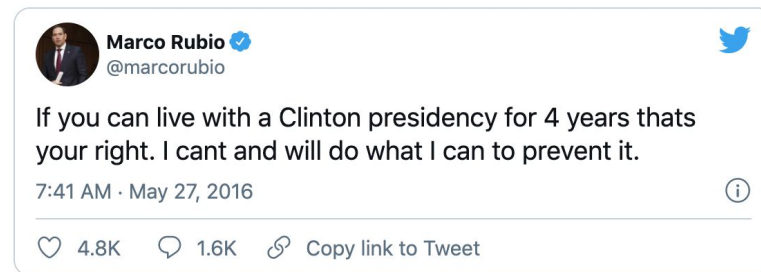
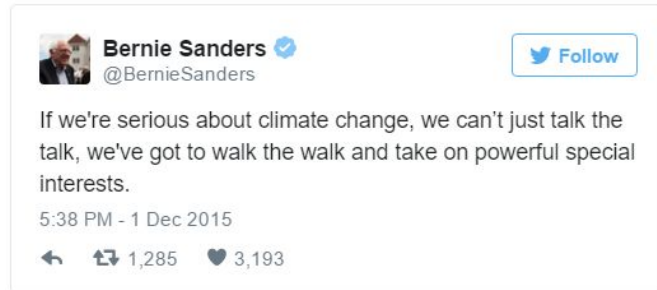


Fig. 2: (Top left) Data tensor  $\mathbf{X}$  with two levels of noise. (Top right) ranks 7, 5, and 3 Neural NCPD, Standard HNCPPD, and HNTF approximations of  $\mathbf{X}$ . (Bottom left) Underlying topic modelling matrix. (Bottom right) topic modelling matrices discovered by Neural NCPD, Standard HNCPPD, and HNTF, with rows and columns permuted to minimize topic error.

# Twitter Political Data Set



- A data set of tweets sent by political candidates during the 2016 election season
- We subset the tweets from eight politicians, four Republicans and four Democrats:  
Hillary Clinton, Tim Kaine, Martin O'Malley, Bernie Sanders, Ted Cruz, John Kasich, Marco Rubio, and Donald Trump.



J. Littman, L. Wrubel, and D. Kerchner, "2016 United States Presidential Election Tweet Ids," 2016.

# Twitter Political Data Set



- We use a bag-of-words (12,721 words in corpus) representation of all tweets made by a candidate
- We bin all tweets made by a candidate each 30 days (from Feb to Dec 2016)
- Resulting tensor is size

$$\begin{array}{ccccc} 8 & \times & 10 & \times & 12721 \\ \uparrow & & \uparrow & & \uparrow \\ \text{\# Candidates} & & \text{\# Months} & & \text{\# Words} \end{array}$$

# Neural NCPD in Twitter Political Data Set

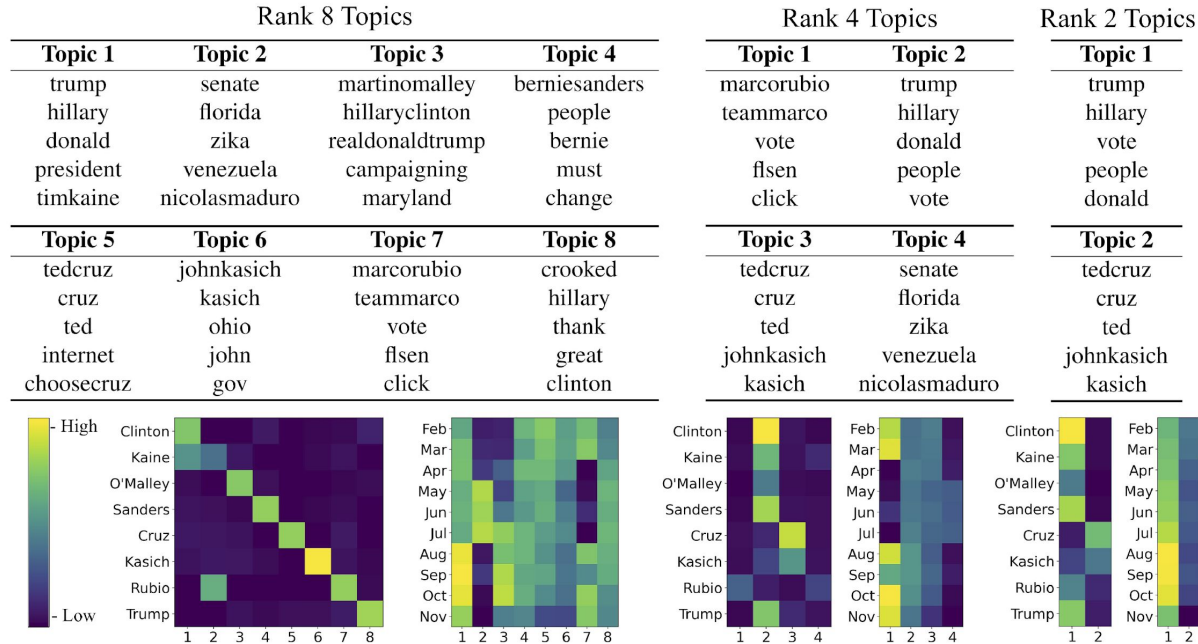


Fig. 3: A three-layer Neural NCPD on the Twitter dataset at ranks  $r = 8$ ,  $r^{(0)} = 4$  and  $r^{(1)} = 2$ . At each rank, we display the top keywords and topic heatmaps for candidate and temporal modes.

# Neural NCPD in Twitter Political Data Set

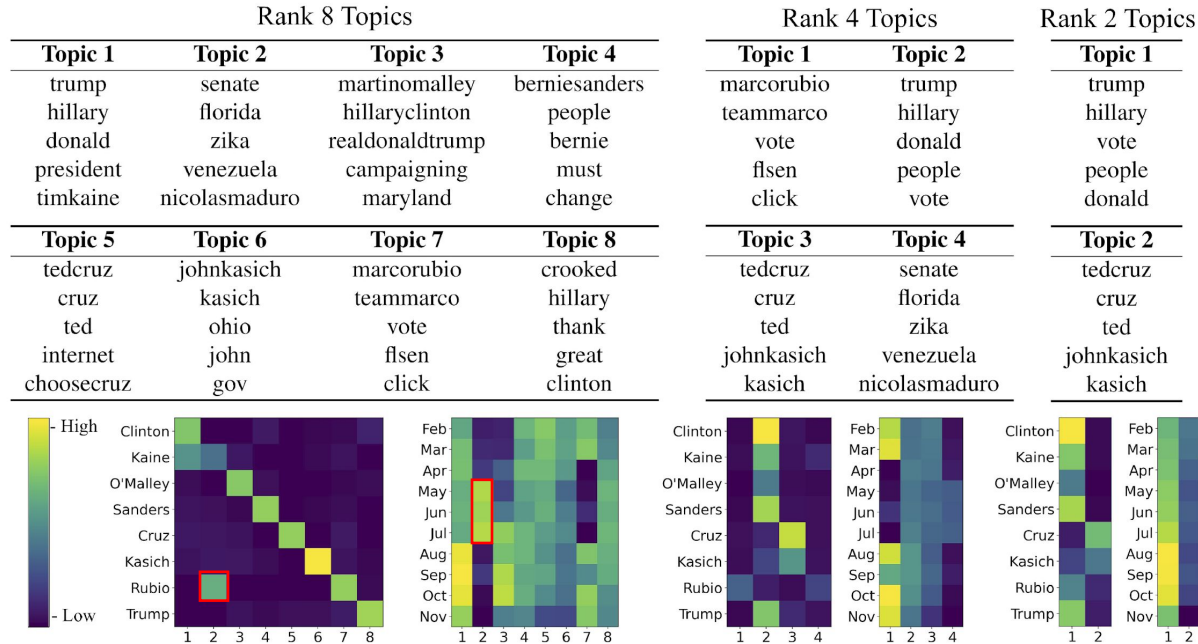


Fig. 3: A three-layer Neural NCPD on the Twitter dataset at ranks  $r = 8$ ,  $r^{(0)} = 4$  and  $r^{(1)} = 2$ . At each rank, we display the top keywords and topic heatmaps for candidate and temporal modes.



# Neural NCPD in Twitter Political Data Set

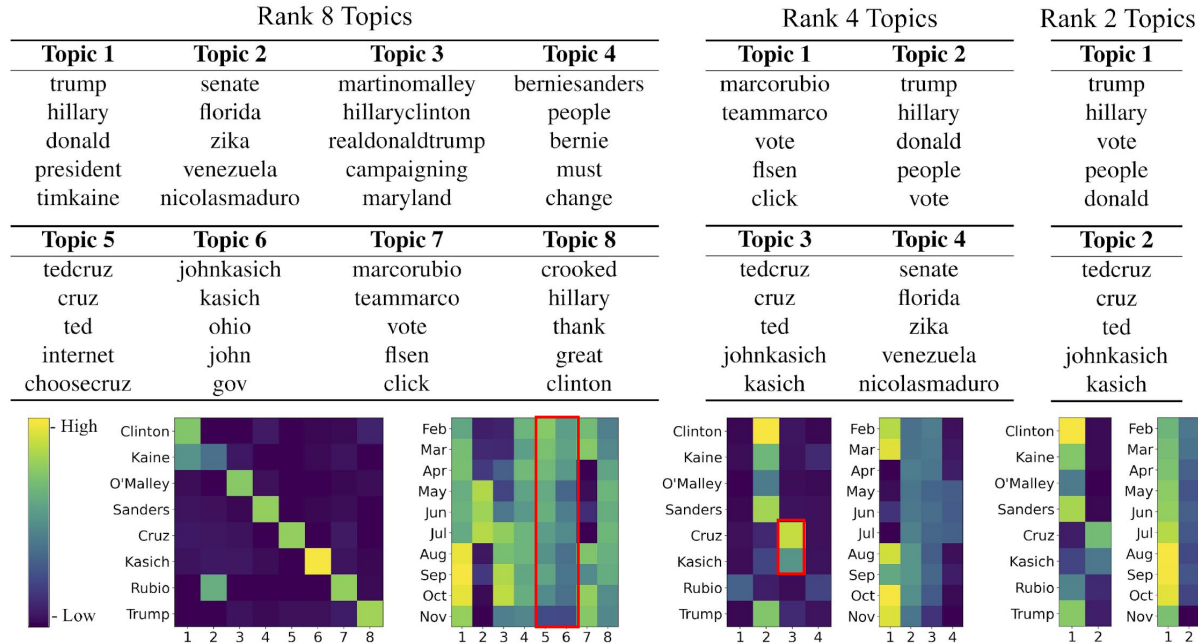


Fig. 3: A three-layer Neural NCPD on the Twitter dataset at ranks  $r = 8$ ,  $r^{(0)} = 4$  and  $r^{(1)} = 2$ . At each rank, we display the top keywords and topic heatmaps for candidate and temporal modes.

# Standard NCPD in Twitter Political Data Set

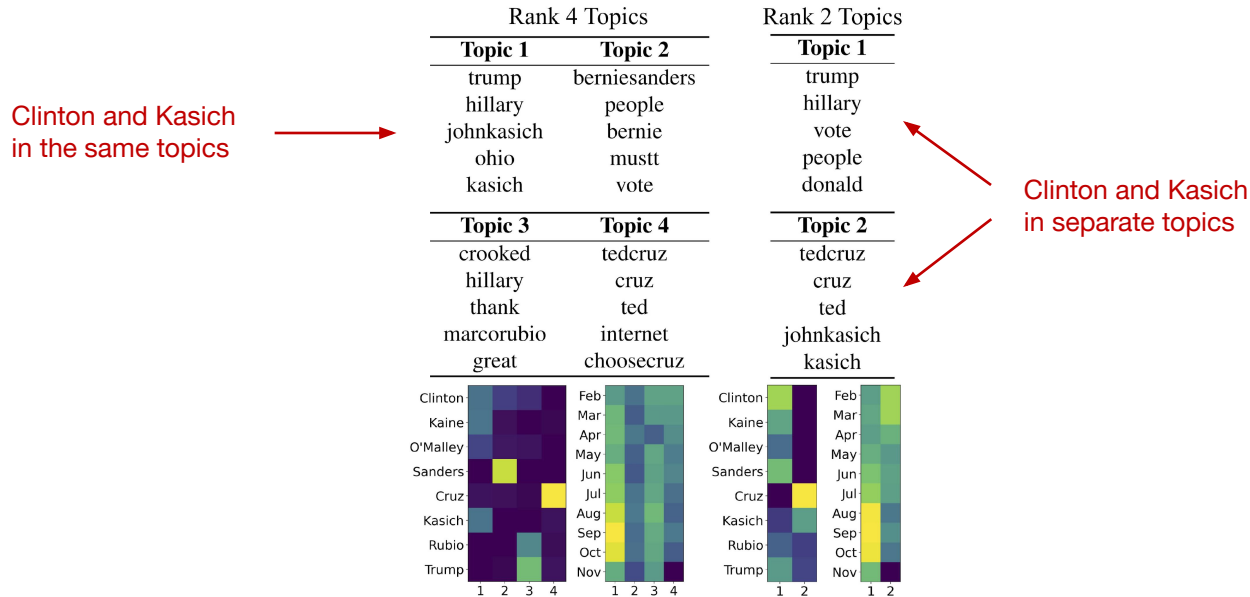


Fig. 5: Ranks 4 and 2 Standard NCPD of the Twitter dataset. At each rank, we display the top five keywords and candidate and temporal mode heatmaps.

# Twitter Political Data Set Numerical Results

TABLE II: Relative reconstruction loss,  $C_{\text{rel}}$ , on the Twitter political dataset for Neural NCPD, Standard NCPD, Standard HNCPD, and HNTF at ranks  $r = 8$ ,  $r^{(0)} = 4$ , and  $r^{(1)} = 8$ . For HNTF we display the loss given the three possible arrangements of the tensor.

Method	$r = 8$	$r^{(0)} = 4$	$r^{(1)} = 2$
Neural NCPD	0.834	0.883	0.918
Standard NCPD	0.834	0.889	0.919
Standard HNCPD	0.834	0.931	0.950
HNTF-1 [12]	0.834	0.890	0.927
HNTF-2 [12]	0.834	0.909	0.956
HNTF-3 [12]	0.834	0.895	0.942

# **Thank you!**

---

## Questions?