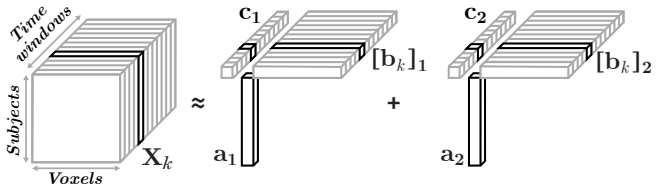

Fig. 1: Illustration of a two-component PARAFAC2 model.

sor $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$ is defined as $\mathcal{X} \approx \sum_{r=1}^{R} \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r$, where $\circ$ denotes the vector outer product, $\mathbf{a}_r, \mathbf{b}_r$, and $\mathbf{c}_r$ are the rth column of factor matrices $\mathbf{A} \in \mathbb{R}^{I \times R}, \mathbf{B} \in \mathbb{R}^{J \times R}$, and $\mathbf{C} \in \mathbb{R}^{K \times R}$, respectively. Using CP, each frontal slice ($\mathbf{X}_k \in \mathbb{R}^{I \times J}$) is formulated as:

$$\mathbf{X}_k \approx \mathbf{A}\text{diag}(\mathbf{c}_{k:})\mathbf{B}^\mathsf{T}, \tag{1}$$

where $\text{diag}(\mathbf{c}_{k:})$ is an $R \times R$ diagonal matrix with the kth row of $\mathbf{C}$ on the diagonal. If $\mathcal{X}$ represents a *subjects* by *voxels* by *time windows* tensor, each column of $\mathbf{B}$ may reveal a spatial network, $\mathbf{A}$ indicates which networks are present in each subject, and $\mathbf{C}$ contains the temporal profile of each network.

PARAFAC2: PARAFAC2 [21] is a more flexible model than the CP model. While CP assumes that each slice, $\mathbf{X}_k \in \mathbb{R}^{I \times J}$, has the same $\mathbf{A}$ and $\mathbf{B}$ matrices, PARAFAC2 allows each slice to have a different $\mathbf{B}$ matrix (Figure 1) as follows:

$$\mathbf{X}_k \approx \mathbf{A}\text{diag}(\mathbf{c}_{k:})\mathbf{B}_k^\mathsf{T}, \tag{2}$$

where $\mathbf{B}_k$s follow the *PARAFAC2 constraint*, i.e., $\mathbf{B}_{k_1}^\mathsf{T}\mathbf{B}_{k_1} = \mathbf{B}_{k_2}^\mathsf{T}\mathbf{B}_{k_2}$ for all $k_1, k_2 \leq K$. If $\mathbf{X}_k$ slices correspond to time windows, then networks captured by columns of $\mathbf{B}_k$ may change over time as long as they satisfy this constraint. We use the notation $[\mathbf{b}_k]_r$ to denote the rth column of matrix $\mathbf{B}_k$.

3. EXPERIMENTS

In this section, we assess the performance of PARAFAC2 in terms of extracting time-evolving networks using both simulated and real data. First, we generate data with time-evolving patterns, and demonstrate that the model can recover the underlying evolving patterns, performing much better than a CP model. Then we use the PARAFAC2 model to analyze fMRI signals from a group of subjects consisting of healthy controls and patients with schizophrenia and show the promise of the model in terms of revealing the difference in the evolution of task-related spatial networks across the two groups.

3.1. Simulated data

3.1.1. Generation of simulated data

We generated two types of simulated data: (i) data sets following (2) and the PARAFAC2 constraint, and (ii) data sets following (2), having time-evolving patterns $\mathbf{B}_k$s that do not necessarily follow the PARAFAC2 constraints.

Factor matrices in each mode are generated as follows (with $R = 4$): The factor matrix for the first mode (i.e. subjects) ($\mathbf{A}$) has a clustering structure as shown in Figure 2. For the second mode (i.e. voxels), ($\mathbf{B}_k$)s are generated in two different ways: (i) *Random:* random matrices following the PARAFAC2 constraint, (ii) *Network:* matrices with each column corresponding to either a shifting, a growing, a shrinking or a both shifting and growing network (Figure 3). Finally, the factor matrix in the third mode (i.e., time) ($\mathbf{C}$) is generated in two different ways: (i) *Random:* with all columns drawn from a uniform distribution randomly, (ii) *Trends:* with one random column, and other three columns following either a sinusoidal, an exponential or a sigmoidal curve (Figure 2).[1]

Once $\mathcal{X}$ is constructed based on (2) using the generated factor matrices, we add random noise: $\mathcal{X}_{\text{noisy}} = \mathcal{X} + \eta\mathcal{E}\frac{\|\mathcal{X}\|_F}{\|\mathcal{E}\|_F}$, where η is the noise level, $\|\cdot\|_F$ is the Frobenius norm, entries of $\mathcal{E}$ follow a standard normal distribution.

3.1.2. Performance evaluation

We generated twenty random data sets for all possible combinations of factor matrix generation schemes and used different noise levels; $\eta = 0$ and $\eta = 0.33$ ($\approx 10\%$ of the data being noise). We fit both CP and PARAFAC2 using alternating least squares [4, 18] with multiple random starts, and the start with the best fit score was chosen for further analysis (after validating the uniqueness of the model). Non-negativity constraints were imposed in the *time* mode when fitting PARAFAC2 to resolve the sign ambiguity [22], which is more critical for PARAFAC2 than CP. Model performance is assessed in terms of the following measures:

Model fit: One metric used to assess the quality of the data reconstruction, denoted by $\hat{\mathcal{X}}$, is the *model fit* defined as $\text{Fit}(\mathcal{X}, \hat{\mathcal{X}}) = 100 \times \left(1 - \frac{\|\mathcal{X} - \hat{\mathcal{X}}\|^2}{\|\mathcal{X}\|^2}\right)$. The fit tells us how well the model explains the data.

Factor Match Score (FMS): We measure the accuracy of the methods, i.e., how well the methods recover underlying factor vectors, using FMS defined, for each mode separately, as:

$$\text{FMS}_{\mathbf{U}} = \frac{1}{R}\sum_{r=1} \frac{|\mathbf{u}_r^\mathsf{T}\hat{\mathbf{u}}_r|}{\|\mathbf{u}_r\|\,\|\hat{\mathbf{u}}_r\|},$$

where $\mathbf{u}_r$ and $\hat{\mathbf{u}}_r$ are true and estimated rth column of factor matrix $\mathbf{U}$, respectively (after fixing the permutation ambiguity). For the evolving mode ($\mathbf{B}$), we concatenate all time steps to form a $(JK \times R)$ matrix, $\tilde{\mathbf{B}}$, and compute $\text{FMS}_{\mathbf{B}}$ using $\tilde{\mathbf{B}}$. For CP, the same $\mathbf{b}_r$ vector is repeated K times.

Clustering accuracy: The clustering performance is assessed using k-means clustering on the factor matrix extracted from the first mode, using all possible combinations of factor vectors, and the best performance is reported.

[1] Simulations are described in detail in the supplementary material: https://github.com/marieroald/ICASSP20, with links to the simulation code and Python implementations of CP and PARAFAC2.

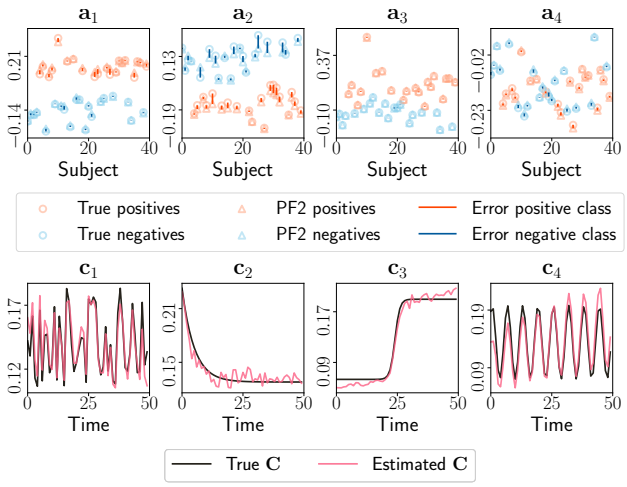

Fig. 2: Top: Scatter plot of true columns of $\mathbf{A}$ and the ones captured by PARAFAC2. Bottom: Line plot of true columns of $\mathbf{C}$ and the ones estimated by PARAFAC2 (noisy case).

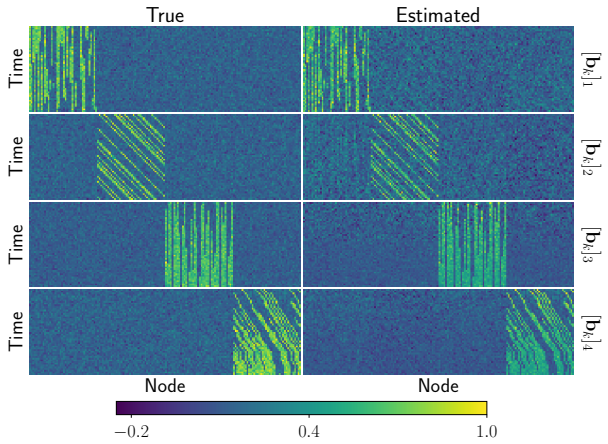

Fig. 3: True $\mathbf{B}_k$ factor vectors (left) and the ones captured by PARAFAC2 (right) (noisy case). Factor vectors are shown as heat maps, where brighter colors represent active nodes.

3.1.3. Results of simulated data analysis

Table 1 shows the average performance (for twenty random data sets) of CP and PARAFAC2 ($R = 4$) using different data generation schemes. Results demonstrate that PARAFAC2 performs well in terms of recovering the true patterns and their evolution with average FMS values above 0.90 for the evolving network mode, and with much higher FMS in other modes, consistently performing better than the CP model.

We observe that CP partially discovers $\mathbf{A}$ and $\mathbf{C}$ matrices for the network data, even in the noisy case, while completely failing for data with random $\mathbf{B}$s. Despite failing to capture the true $\mathbf{A}$, CP often has high clustering accuracy due to the clustering structure in the simulations designed to separate groups with several components, i.e., if a model recovers any informative components, the clustering accuracy will be high.

The results also demonstrate that there is room for improvement for PARAFAC2, especially when the PARAFAC2 constraint is not satisfied, i.e. the evolving network case. Note that PARAFAC2 does not reveal the true components perfectly when $\mathbf{B}_k$s have an evolving network structure, even in the noise-free case. This challenge stems from the PARAFAC2 constraint, which requires that the 2-norm of the network factor vectors are constant in time. If a network grows or shrinks in density, the resulting change in norm cannot be represented in $\mathbf{B}$, but rather in $\mathbf{C}$. In the noisy case, we observe similar performance for both random and time-evolving network data. Figure 2 and Figure 3 illustrate how well PARAFAC2 performs in terms of capturing the true patterns in the noisy data.

3.2. Real data: fMRI

Our motivation for exploring PARAFAC2 is to understand task-related dynamic connectivity in the brain and how that differs between controls and patients suffering from a psychiatric disorder. Here, we use PARAFAC2 to analyze fMRI data from the MCIC collection [23], which contains fMRI scans from patients with schizophrenia and healthy controls for different tasks. For this work, we used the sensory motor task (SM) data from the 3T fMRI scanners at the University of Iowa and Minnesota[2]. During the SM task, subjects were equipped with headphones that played sounds with increasing pitch, with a resting period between each sound. Subjects were instructed to push a button each time they heard a sound.

To construct the data tensor, we first extracted the fractional amplitude of low-level fluctuations (fALFF) [24] signal from sliding time windows of the blood-oxygen-level-dependent signal. The fALFF is calculated in three steps: (1) Discard high and low frequency components to remove noise and the signal from the vascular system. (2) Sum the square root of the frequency components to get the amplitude of low-frequency fluctuation. (3) Divide by the total sum of frequencies in the window. This approach provides a time-evolving measure of brain activity within each voxel. The window size and stride length were chosen so that each time window, corresponding to 16 seconds, contains precisely one block of task or rest with no overlap between the windows. The data tensor was constructed using the fALFF values for the voxels corresponding to gray matter as a feature vector for each time window and each subject. The constructed tensor is in the form of 145 *subjects* (of which, 90 of them are healthy controls) by 63652 *voxels* by 14 *time windows*.

3.2.1. Results for fMRI data analysis

Before the analysis, the fMRI tensor was preprocessed by subtracting the mean fALFF signal across the *voxels* mode,

[2]We have excluded the data from other sites either due to scanner difference or observed site differences in our preliminary analysis.

Table 1: The mean performance of PARAFAC2 (PF2) and CP for different setups.

			Fit [%]		Clustering Acc [%]		FMS_A		FMS_B		FMS_C	
Noise	**C setup**	**B Setup**	CP	PF2	CP	PF2	CP	PF2	CP	PF2	CP	PF2
0	Random	Network	75.1	100.0	89.3	94.6	0.75	0.98	0.58	0.97	0.87	0.98
		Random	13.2	100.0	82.5	90.9	0.27	1.00	0.01	1.00	0.59	1.00
	Trends	Network	80.7	100.0	82.0	92.0	0.52	0.97	0.54	0.98	0.86	0.98
		Random	15.5	100.0	77.8	91.8	0.17	1.00	0.01	1.00	0.47	1.00
0.33	Random	Network	67.8	91.1	88.1	95.0	0.74	0.98	0.58	0.95	0.87	0.98
		Random	11.9	91.1	82.4	93.0	0.27	0.97	0.01	0.92	0.58	1.00
	Trends	Network	72.8	91.1	85.9	95.0	0.52	0.95	0.52	0.92	0.86	0.98
		Random	14.0	91.1	77.4	90.3	0.17	0.95	0.01	0.90	0.47	0.99

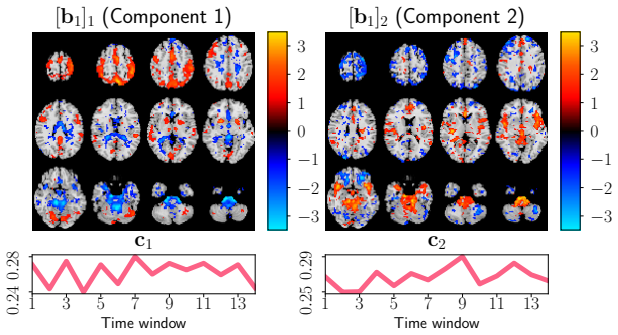

Fig. 4: Factor vectors in *voxels* and *time windows* modes of 2-component PARAFAC2 model. In *voxels* mode, only the first time window, i.e., $\mathbf{B}_1$, is visualized. The p-values are 2.8×10^{-4} and 4.1×10^{-2} for component 1 and 2, respectively.

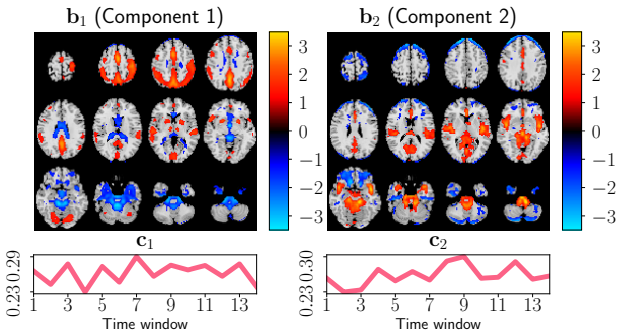

Fig. 5: Factor vectors in *voxels* and *time windows* mode of a 2-component CP model. The p-values are 6.2×10^{-4} and 3.7×10^{-1} for component 1 and 2, respectively.

and dividing each *voxels* mode fiber by its norm. The preprocessed tensor was then modelled using PARAFAC2 and CP such that $\mathbf{A}$, $\mathbf{B}_k$(or $\mathbf{B}$ matrix for CP) and $\mathbf{C}$ correspond to *subjects*, *voxels* and *time windows*, respectively. To resolve the sign indeterminacy, we imposed non-negativity constraints on $\mathbf{C}$ for the PARAFAC2 model. Previously, PARAFAC2 was used for fMRI data analysis [20], where each slice $\mathbf{X}_k$ corresponded to raw signals from a single subject and patterns in time and/or voxels were allowed to change across subjects. Instead, we construct a tensor based on features capturing the dynamic activity within each window and model the tensor using PARAFAC2, letting the voxel patterns ($\mathbf{B}_k$) change in time, rather than across subjects.

Once the models were fit, we performed a two-sample t-test on each column of the *subjects* mode factor matrix. We achieved the lowest p-values using 2-component CP and 2-component PARAFAC2 models illustrated in Figures 4 and 5. Spatial maps are plotted using the patterns from the *voxels* modes as z-maps and thresholding at $|z| \geq 1.2$ such that red voxels indicate an increase in controls over patients and blue voxels indicate an increase in patients over controls.

Both CP and PARAFAC2 reveal activity in the sensorimotor cortex (component 1) and auditory cortex (component 2), two brain regions known to be engaged by the SM task [23]. The sensorimotor component has significantly lower activation for schizophrenic patients than healthy controls, i.e., entries of the corresponding *subjects* mode vector are positive for both groups but statistically significantly higher in controls than patients. That makes sense since it is difficult for patients to follow the task; hence the activation is lower. Temporal profiles, especially for the first component, reflect the on-off task pattern. While both models reveal similar spatial networks, PARAFAC2 captures the temporal evolution of these spatial networks (see supplementary material for videos showing the temporal evolution). We also observe that due to the flexibility of PARAFAC2, the PARAFAC2 spatial maps are noisier compared to the ones captured by the CP model.

4. CONCLUSION

In this work, we have demonstrated that the tensor factorization model PARAFAC2 can recover the underlying patterns and their evolution from dynamic data, even for data not strictly following the PARAFAC2 constraints. Moreover, PARAFAC2 shows promising performance for detecting task-related brain connectivity and its evolution from fMRI data. As future work, we plan to relax PARAFAC2 constraints [25] and incorporate prior information about the evolving network structure (e.g., temporal and spatial smoothness) by regularizing the model to improve the recovery of underlying patterns.

5. REFERENCES

[1] J. Yuan et al., "Spatio-temporal modeling of connectome-scale brain network interactions via time-evolving graphs," *Neuroimage*, vol. 180, pp. 350–369, 2018.

[2] S. Wong, C. Pastrello, M. Kotlyar, C. Faloutsos, and I. Jurisica, "SDREGION: Fast spoting of changing communities in biological networks," in *KDD'18*, 2018, pp. 867–875.

[3] E. Acar and B. Yener, "Unsupervised multiway data analysis: A literature survey," *IEEE T Knowl. Data En.*, vol. 21, no. 1, pp. 6–20, 2009.

[4] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, 2009.

[5] R. A. Harshman, "Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multimodal factor analysis," *UCLA working papers in phonetics*, vol. 16, pp. 1–84, 1970.

[6] J. D. Carroll and J. J. Chang, "Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition," *Psychometrika*, vol. 35, no. 3, pp. 283–319, 1970.

[7] E. Acar, D. M. Dunlavy, and T. G. Kolda, "Link prediction on evolving data using matrix and tensor factorizations," in *ICDM Workshops*, 2009, pp. 262–269.

[8] B. W. Bader, M. W. Berry, and M. Browne, *In Survey of Text Mining II: Clustering, Classification, and Retrieval*, chapter : Discussion Tracking in Enron Email Using PARAFAC, pp. 147–163, Springer, 2008.

[9] D. Bruns-Smith, M. M. Baskaran, J. Ezick, T. Henretty, and R. Lethin, "Cyber security through multidimensional data decompositions," in *Cybersecurity Symposium*, 2016, pp. 59–67.

[10] M. Mørup, L. K. Hansen, S. M. Arnfred, L. H. Lim, and K. H. Madsen, "Shift-invariant multilinear decomposition of neuroimaging data," *NeuroImage*, vol. 42, no. 4, pp. 1439–1450, 2008.

[11] M. Mørup, L. K. Hansen, and K. H. Madsen, "Modeling latency and shape changes in trial based neuroimaging data.," in *ACSCC*, 2011, pp. 439–443.

[12] A. Afshar, I. Perros, E. E. Papalexakis, E. Searles, J. Ho, and J. Sun, "COPA: Constrained PARAFAC2 for sparse & large datasets," in *CIKM'18*, 2018, pp. 793–802.

[13] R. Bro, C. A. Andersson, and H. A. L. Kiers, "PARAFAC2 - Part II. Modeling chromatographic data with retention time shifts," *J. Chemom.*, vol. 13, no. 3-4, pp. 295–309, 1999.

[14] J. Sun, D. Tao, S. Papadimitriou, P. S. Yu, and C. Faloutsos, "Incremental tensor analysis: Theory and applications," *ACM Trans. Knowl. Discovery Data*, vol. 2, no. 3, pp. 11, 2008.

[15] S. Fernandes, H. Fanaee-T, and J. Gama, "Dynamic graph summarization: A tensor decomposition approach," *Data Min. Knowl. Discovery*, vol. 32, no. 5, pp. 1397–1420, 2018.

[16] W. Yu, C. C. Aggarwal, and W. Wang, "Temporally factorized network modeling for evolutionary network analysis," in *WSDM'17*, 2017, pp. 455–464.

[17] Y. Koren, "Collaborative filtering with temporal dynamics," in *KDD'09*, 2009, pp. 447–456.

[18] H. A. L. Kiers, J. M. F. Ten Berge, and R. Bro, "PARAFAC2 - Part I. A direct fitting algorithm for the PARAFAC2 model," *J. Chemom.*, vol. 13, no. 3-4, pp. 275–294, 1999.

[19] M. E. Timmerman and H. A. L. Kiers, "Four simultaneous component models for the analysis of multivariate time series from more than one subject to model intraindividual and interindividual differences," *Psychometrika*, vol. 68, pp. 105–121, 2003.

[20] K. H. Madsen, N. W. Churchill, and M. Mørup, "Quantifying functional connectivity in multi-subject fmri data using component models," *Hum. Brain Mapp*, vol. 38, no. 2, pp. 882–899, 2017.

[21] R. A. Harshman, "PARAFAC2: Mathematical and technical notes," *UCLA working papers in phonetics*, vol. 22, pp. 30–44, 1972.

[22] R. Bro, R. Leardi, and L. G. Johnsen, "Solving the sign indeterminacy for multiway models," *J. Chemom.*, vol. 27, no. 3-4, pp. 70–75, 2013.

[23] R. L. Gollub et al., "The MCIC collection: a shared repository of multi-modal, multi-site brain image data from a clinical investigation of schizophrenia," *Neuroinformatics*, vol. 11, no. 3, pp. 367–388, 2013.

[24] Q.-H. Zou et al., "An improved approach to detection of amplitude of low-frequency fluctuation (ALFF) for resting-state fMRI: Fractional ALFF," *J. Neurosci. Methods*, vol. 172, no. 1, pp. 137 – 141, 2008.

[25] J. E. Cohen and R. Bro, "Nonnegative PARAFAC2: A flexible coupling approach," in *LVA/ICA'18*, 2018, pp. 89–98.