The ultimate goal of cognitive neuroscience is to understand the mechanistic neural processes underlying the functional organization of the brain. Key to this study is understanding structure of both the structural and functional connectivity between anatomical regions. In this paper we follow previous work in developing a simple dynamical model of the brain by simulating its various regions as Kuramoto oscillators whose coupling structure is described by a complex network. However in our simulations rather than generating synthetic networks, we simulate our synthetic model but coupled by a real network of the anatomical brain regions which has been reconstructed from diffusion tensor imaging (DTI) data. By using an information theoretic approach that defines direct information flow in terms of causation entropy (CSE), we show that we can more accurately recover the true structural network than either of the popular correlation or LASSO regression techniques. We demonstrate the effectiveness of our method when applied to data simulated on the realistic DTI network, as well as on randomly generated small-world and Erdős-Rényi (ER) networks.

Knowledge of the structure of the complex networks helps to determine how they will respond to stresses, such as dynamical perturbations\cite{12}, or to structural network perturbations\cite{13,15}. However frequently the underlying network structure of a system is unknown. What may be available to us is time series data collected at each node. In the case of functional magnetic imaging (fMRI) data, time varying intensity is associated with each of the voxels, as a three-dimensional movie. The associated metabolism of an active region of the brain becomes apparent by increased oxygen levels resulting the the blood oxygenation level dependent (BOLD) signal that can be inferred by magnetic resonance imaging. These perturbations and interacting variations carry information about the underlying network connectivity of the brain. However, the inverse problem of inferring the underlying network connectivity from observations of state, is inherently an ill-posed problem that is sometimes called network tomography, and it remains a difficult problem today.

In this work we utilize a recently developed method named entropic regression\cite{17} to recover the underlying network structure of synthetic data which is generated from a network Kuramoto model from three different types of networks from synthetic models to a real structural brain network: an Erdős-Rényi\cite{19} (ER) network as a classical random graph model, a small-world network following\cite{21} with designed structure, and the experimentally observed DTI network by methods as discussed below. On these network models of successive challenge and realism, we will simulate a popular synthetic dynamic to challenge the accuracy entropic regression. We simulate the first order network Kuramoto model, that has popularly been used including for the generation of synthetic fMRI data\cite{22}. With the simulated data we demonstrate that entropic regression accurately outperforms for network recovery as compared to other leading methods, including correlation and LASSO regression utilizing the Bayesian information criterion (BIC).
II. MATERIALS AND METHODS

A. Complex Networks

For this work we will use the terms graph and complex network interchangeably. A graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is a set of nodes $\mathcal{V}$ and edges $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$. In context here where we will associate each anatomical region of the brain as a node, the goal is to infer interactions as information flow between these regions. The unweighted structure of a graph of $n$ nodes can be encoded by an $n \times n$ adjacency as follows:

$$[A]_{ij} = \begin{cases} 1 & \text{if } (i, j) \in \mathcal{E}, i \neq j, \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

A weighted directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{W})$ has a weight for each edge, and it can be encoded by a weight matrix, $[W]_{ij} = w_{ij}\text{sign}([A]_{ij})$, with weights $w_{ij}$. A simple adjacency matrix is defined by a graph where no self edges occur. A complete network is defined by the presence of all possible edges. We will generally allow for weighted directed graphs, which are well defined by a (possibly asymmetric) weight matrix $W$, which will be our goal to find from node-level observations only.

The degree distribution is often used to classify complex networks, even though it does not uniquely define its generative process. Recall that the degree of a node $d_i$ is defined by $d_i = \sum_{j=1}^{n} [A]_{ij}$, and the degree distribution is the discrete probability function of $d_i$ sampled across $i$. The Erdős-Rényi graph (ER) graph, for example, is the classical random graph wherein edges are assigned at random with probability $p$. It has been shown in the ER graph the degree distribution is Poisson in the limit of a large number of nodes. It became clear over time that real world networks tend to be highly clustered and yet have a small shortest path distance between any pair of nodes, and this led to the development of the Watts-Strogatz (WS) "small world" graph. ER graphs do not exhibit the clustering or short path length that the small world graphs do.

B. Structural Brain Networks

Simulated time series data were generated using an actual structural brain network but a synthetic model, thus we call the data semi-synthetic, in addition to the synthetic complex networks with synthetic model dynamics, which we call fully synthetic data. The true brain networks were generated experimentally in previously published work using diffusion tensor imaging (DTI) data. The networks generated in this prior work are publicly available and "Subject1" through "Subject20" were used in the current work. Details of the image processing and network generation can be found in that manuscript but are briefly described below. The Diffusion Toolkit (FDT) in the FSL software package was utilized for image preprocessing and for diffusion tensor estimation. The cerebral gray matter was parcellated into 83 regions of interest (ROIs) based on the Lausanne anatomical atlas using Freesurfer software. There were 41 ROIs in each hemisphere and a single brainstem region. The regions were warped to the native space of each study participant based on the anatomical image and then transformed to the DTI space using FSL. Deterministic tractography was performed using the Diffusion Toolkit software. All white matter voxels were seeded and resulting fibers were assessed to determine if they connected two of the ROIs. The ROIs served as the nodes in the adjacency matrix. An edge was present between two nodes if there was at least one white matter fiber connecting the nodes. A weighted adjacency matrix was generated by counting the number of fibers connecting two nodes and normalizing this value using the length of the fibers and the area of the ROIs as previously proposed. The absence of a fiber connecting two ROIs was indicated in the adjacency matrix with a zero (0).

C. Regression, regularized regression, toward entropic regression

Suppose a parametric model, $y = x\beta$ is assumed for scalar real variable $y \in \mathbb{R}$ and $x \in \mathbb{R}^n$. Then $q$ samples, may be stated by

$$Y = X\beta + \epsilon, \quad (2)$$

with data matrices $X \in \mathbb{R}^{q \times n}$, $Y \in \mathbb{R}^q$ and here noise is assumed to be normal $\epsilon \sim \mathcal{N}(0, \sigma)$, gives $\beta \in \mathbb{R}^n$ unknown regressors. We assume enough samples so that the problem is over-determined, $q > n$. The goal in ordinary least squares (OLS) is to find the closest parametric fit to the data in the sense of square norm residual, so the minimizer of the loss function:

$$\min_{\beta \in \mathbb{R}^n} ||X\beta - Y||_2. \quad (3)$$

A flaw of OLS however, as noted in, is that while the OLS estimator has a low bias, it typically has a large variance. It tends to suffer from over-fitting if model complexity is not well chosen.

Tikhonov regularization is a strategy to mitigate overfitting of OLS. The regularization term introduces a penalty term to Eq. Let:

$$\min_{\beta \in \mathbb{R}^n} ||X\beta - Y||_2^2 + R(\beta), \quad (4)$$

where $R(\beta)$ is a regularization term on $\beta$. OLS paired with common types of regularization are often referred to as ridge regression and Least Absolute Shrinkage and Selection Operator (LASSO). In ridge regression, we choose

$$R(\beta) = \lambda ||\beta||_2, \quad (5)$$

emphasizing reduced variance. In LASSO regression, we have

$$R(\beta) = \lambda ||\beta||_1, \quad (6)$$

which emphasizes sparsity. Tikhonov regularization in general defines a selection of a favored solution amongst infinitely
many of an ill-posed problem as the philosophy of inverse problems, and in the ridge regression scenario, this sense can be thought of as filtering noise that is related to the small but nonzero singular values of the data matrix.\textsuperscript{38}

The choice of the value of regularity parameter $\lambda$ is an important practical aspect of regularization, as a poor choice for this value can lead to poor estimation results. There exist several methods for its choice, but here we choose the Bayesian Information Criterion (BIC)\textsuperscript{39}, which is one of the popular choices for selecting the value of $\lambda$. The BIC solution is considered to be similar to maximum likelihood method for asserting the dimension of a model. The BIC estimation relies on the optimizing:\textsuperscript{40}

$$\text{BIC}(\beta, \lambda | X) = \kappa \ln(q) - 2 \ln(\mathcal{L}(\beta, \lambda | X)) \quad (7)$$

where $\kappa$ is the number of nonzero model parameters and $\mathcal{L}$ is the maximum likelihood solution of the given model with respect to the parameters $\beta$ for a given fixed value of the parameter $\lambda$ and given the data $X$. Then the optimal value ($\lambda_{\text{opt}}$) can be found by minimizing the BIC with respect to $\lambda$ for each $\beta$, that is:

$$\lambda_{\text{opt}} = \min_{(\beta, \lambda)} \text{BIC}(\beta, \lambda) \quad (8)$$

This formulation introduces a penalty term on making the dimension of the model too large as well as a penalty for having a model with too many parameters, which generally provides popularly pleasing estimates of the parameter $\lambda_{\text{opt}}$ when using LASSO.\textsuperscript{28,39,40}

D. Entropic Regression

Here we review entropic regression as developed in\textsuperscript{13} specializing the network discovery algorithm of optimal causation entropy\textsuperscript{32,33}, which is an efficient and accurate method to infer a parameterically defined model, as a system identification problem can be stated in matrix form:

$$\dot{Z} = \Phi(Z) \Gamma \quad (9)$$

where $Z \in \mathbb{R}^{q \times n}$ as before and is the measured state variables of the $n$-dimensional system with $q$ observations, $Z$ is the vector field estimated from $Z$, $\Phi: \mathbb{R}^{q \times n} \rightarrow \mathbb{R}^{q \times K}$ is a function that maps the state variables $Z$, to the expanded set of candidate functions (not necessarily linear), and $\Gamma \in \mathbb{R}^{K \times n}$ is the parameters matrix. Given a basis set of functions $\Phi = \{ \phi_i(Z) \}_{i=1}^K$, where a row vector $\phi_j$ is given by $\Phi_j = \{ \phi_1(Z_j), \phi_2(Z_j), \ldots, \phi_K(Z_j) \}$, and $\phi_i(Z_j): \mathbb{R}^n \rightarrow \mathbb{R}$, is a candidate function on the $n$-dimensional observation $Z_j$, that has a high flexibility of choice.

For example in this form, we can write Lorenz’s equations, with its three dimensional state vector $Z = (z_1, z_2, z_3)$, as $\dot{z}_1 = \sigma(z_2 - z_1), \dot{z}_2 = z_1(\rho - z_3) - z_2, \dot{z}_3 = z_1z_2 - \beta z_3$. Even though Lorenz is a nonlinear equation, it is a linear combination of nonlinear functions, and the linear equation implied by Eq. (9), using the second-order power polynomial functions, $\Phi(Z = (z_1, z_2, z_3)) = \{ 1, z_1, z_2, z_3, z_1^2, z_1z_2, z_1z_3, z_2^2, z_2z_3, z_3^2 \}$, we see that we have 10 candidate functions that contain linear and nonlinear terms (functions). However, the underlying dynamic of the Lorenz system is a linear combination of these candidate functions, and we see, for example, that $\dot{z}_1$ in the Lorenz system can be found as $\dot{z}_1 = \Phi(Z)[0, -\sigma, \sigma, 0, 0, 0, 0, 0, 0]^T$. The same applies for $\dot{z}_2$ and $\dot{z}_3$. This kind of linearization is generally known as Carleman linearization.\textsuperscript{44} The reconstructed vector field using the least squares solution may be written as:

$$V(Z, \Phi) = \Phi \Phi^\dagger Z = \Phi L(Z, \Phi) \quad (10)$$

where $\Phi^\dagger$ is the pseudoinverse of the matrix $\Phi$. The problem of regression for differential equation models $\Gamma$ has recently become especially relevant for data-driven science\textsuperscript{45–47} and recently popularized by sparse methods such as optimization by LASSO regression, and by information theoretic methods\textsuperscript{17,32,33}. For simplicity, we will write $\Phi(Z)$ as $\Phi$ in the following discussion. Our entropic regression has proved competitive improvements\textsuperscript{12} based on information criterion.

a. Entropic Regression Algorithm The entropic regression involves an optimization method to associate data to a most informative and sparse set of basis functions. Most informative is interpreted in the sense of mutual information between basis function (observable set) and the measured data. The underlying optimization proceeds in two stages: forward greedy search (selection) and (backward) possible elimination of basis functions, and these are based on the conditional mutual information amongst competing observations through the selected basis functions. In the forward stage, our objective is to select the subset $s \subset \mathcal{S} = \{1, 2, ..., K\}$, which represent strong candidate functions. Starting from empty set $s_0 = \emptyset$, the forward selection stage can be written as:

$$u_k = \arg \max_{i \in \mathcal{S} \setminus s_{k-1}} I(\hat{Z}; \Phi_i | V(\hat{Z}, \Phi_{s_{k-1}}))$$

$$s_k = s_{k-1} \cup u_k \quad (11)$$

where $k = 1, ..., s_k$ is the iteration index, $u_k$ is the index with the maximum objective function value. Note that $s_0 = \emptyset \implies V(\hat{Z}, \Phi_{s_k}) = \emptyset$ which reduces the conditional $I(\cdot; \cdot)$ to the mutual information $I(\cdot; \cdot)$. During the forward stage, at each iteration $k$ and given the information $V(\hat{Z}, \Phi_{s_k})$ we already have from the set $s_{k-1}$, we are looking for the function that maximally add extra information to the model. The process terminates when either all basis functions are exhausted (with maximum number of iterations equal $K$), or the reward function $I(\hat{Z}; V(\Phi_j) | V(\hat{Z}, \Phi_{s_k})) = 0$ (or thresholded to $\leq \varepsilon$).
indicating that none of the remaining basis functions are relevant, in an information-theoretic sense. In other words, the process terminates when the strongest candidate has no further information beyond what we already have. Note that entropy, mutual information, and conditional mutual information can be estimated using any valid estimator. We choose the \( K \)-nearest neighbors estimator\textsuperscript{29,31}, for its accuracy, especially with relatively small sample sizes.

After the forward entropic regression, we have the set \( s \) that has the indices of the strong candidate functions. Eventually, \( s \) may have a few non-relevant functions that are selected due to a high degree of uncertainty and the rounding error at the end of forward entropic regression. Since we have reduced set \(|s| < K\), it will be computationally inexpensive to validate the accuracy of the model. This backward stage is an elimination stage, where the functions indexed by \( s \) are re-examined for their information-theoretic relevance so that redundancy may be removed. In particular, we label the set \( s \) as initial set \( s_0 = s \) for the backward stage, and we perform the following computations and updates,

\[
\begin{align*}
    u_k &= \arg\min_{i \in s_{k-1}} I(\hat{Z}; \mathcal{Y}(\hat{Z}, \Phi_i)|V(\hat{Z}, \Phi_i)), \\
    s_k &= s_{k-1} \setminus \{u_k\}. 
\end{align*}
\]

In particular, note the differences of this stage regarding the index set is different between Eq. (11) and Eq. (12). The backward stage includes a loss function, where at each iteration \( k \), we examine what information will be lost if we remove the index \( i \) from the set \( s_{k-1} \) and the process terminates when \( I(\hat{Z}; V(\hat{Z}, \Phi_i)|V(\hat{Z}, \Phi_{i_k-1}-i)) > 0 \). The result of the backward entropic regression is a set of indices \( s \). We emphasize a practical strength, that the forward entropic regression stage can substantially reduce the computational complexity of the backward stage, by limiting the elimination search space to a few candidate functions. However, the backward elimination is key in that in the special case of a low-dimension system, or we have efficient computational resources, we can plausibly skip the forward stage, to rely only on the backward stage directly, with initial set \( s = \{1, 2, \ldots, K\} \). Parameters \( \beta \in \mathbb{R}^K \) can be found by updating the vector of zeros \( \beta = \Theta_K \) such that:

\[
\beta_i = \mathcal{L}(\hat{X}, \Phi_i) \tag{13}
\]

where \( \beta_i \) are the entries of \( \beta \) indexed by the elements of \( s \). Thus, \(|\beta||_0 = |s|\). The primary role of entropic regression is to find the minimally optimal informative set of basis functions. Once that set is identified, the parameter values themselves are easily found, for example by ordinary least squares (OLS).

Computational practice of statistical estimators for information theoretic quantities must be dealt with to make a practical method. In theory, mutual information \( I(x;y|z) \) is always non-negative and \( I(x;y|z) = 0 \) if and only if \( x \) and \( y \) are statistically independent given \( z \). In practice, however, due to finite sample size and estimation inaccuracies, the estimated mutual information may be nonzero even when \( x \) and \( y \) are independent, and even worse, some otherwise useful favorite estimators may yield negative numbers\textsuperscript{30,35}. We require a way to deduce with statistical confidence that they differ from zero to confidently decide whether \( x \) and \( y \) should be deemed to be independent given the estimated value of \( I(x;y|z) \). In\textsuperscript{33,36}, a shuffle test with a “confidence” parameter \( \alpha \in [0, 1] \) for tolerance estimation was developed. The shuffle test involves random shuffling of one of the variables, repeating \( n_t \) times, to build a test statistic. In particular, for the \( i \)-th random shuffle, a random permutation \( \pi^{(i)} : [T] \rightarrow [T] \) is generated to shuffle one of the variables, say \( y \), which produces a new variable \( \tilde{y}^{(i)} \) where \( \tilde{y}^{(i)} = y_{\pi^{(i)}} \); \( x \) and \( z \) are kept the same. Then, we estimate the mutual information \( I(x;\tilde{y}^{(i)}|z) \) using the (partially) permuted variable \( (x, \tilde{y}^{(i)}, z) \), for each \( i = 1, \ldots, n_t \). For given \( \alpha \), we then compute a threshold value \( I_{\alpha}(x;y|z) \) as the \( \alpha \)-percentile from the values of \( I(x;\tilde{y}^{(i)}|z) \). If \( I(x;y|z) > I_{\alpha}(x;y|z) \), we determine \( x \) and \( y \) as dependent given \( z \); otherwise independent. This threshold (tolerance), of mutual independence is adopted in the forward and backward stages as the termination condition. Hence, the tolerance describes the minimum effective quantity of information. In this sense, in forward entropic regression we are selecting the functions which add a significant quantity of information to the model, while in the backward entropic regression, we discarding functions which are determined to be negligible.

### E. Kuramoto Oscillators

The Kuramoto model is a popular model of network coupled phase oscillators\textsuperscript{44}. While in the standard Kuramoto model the network is assumed to be a complete network\textsuperscript{45}, but this is easily generalized by\textsuperscript{32,41}:

\[
\dot{\theta}_i = \omega_i + k \sum_{j=1}^{n} a_{ij} \sin(\theta_j - \theta_i), i = 1, \ldots, n, \tag{14}
\]

where \( \theta_i(T) \) represents the phase angle of the \( i \)-th oscillator at time \( q \), and \( \omega_i \) represents the natural frequency and \( a_{ij} \in A \). The \( \omega_i \) are drawn from some distribution \( g(\omega) \) that is usually assumed to be unimodal and symmetric about the mean value \( \bar{\omega} \). Let \( \theta_i(T) = (\theta_1(T), \theta_2(T), \ldots, \theta_n(T)) \), for time \( T_i, i = 1, \ldots, q \) equally spaced times. The resulting trajectories data \( Z = (\theta_1, \theta_2, \ldots, \theta_q)^T \), is a \( q \times n \), matrix of vector states as columns. The first order Kuramoto model will eventually result in a synchronous state almost certainly\textsuperscript{32} if the value of coupling strength \( k \) is large enough. So, if \( k > k_c \), the result will most likely synchronize for any initial condition. In Fig. 1 we show three different scenarios around the critical coupling strength \( k_c \). As \( k \rightarrow \infty \), the system tends to synchronize faster making the system indistinguishable sooner.

### III. RESULTS

To demonstrate the effectiveness of entropic regression, we perform analysis on synthetic and semi-synthetic time series data. We know the ground truth for the underlying structural
For the testing purposes we simulate data according to the first order network Kuramoto model as described in Eq. [14]. We simulate the Kuramoto model different types of networks with approximately 80 nodes (the DTI network is 83 nodes, the other two types are set to 80 to be close to the same number of nodes), including the directed ER network, a special type of small world network and the DTI network, examples of which are shown in Fig. 2. The small world network is a simple representation of the brain as seen in [18] and the DTI network as a realistic representation. We note that all methods perform poorly if the coupling coefficient $k$ is set too high as a result of near immediate synchronization, thus rendering the nodes indistinguishable. Thus the coupling strength is chosen to be nearby to its critical value, that is $k \approx k_c$.

We ran 50 trials for each of the randomly generated types of adjacency matrix. The parameters of the Kuramoto model where held fixed throughout all trials. The critical coupling of the system $k = k_c$ is defined as, $k_c = 2 \frac{\lambda_1}{\lambda_2}$. The function $g(\omega)$ denotes the pdf of $\omega$ across the network of oscillators, and $\lambda_1$ is the largest eigenvalue of the corresponding adjacency matrix $\hat{A}$. The initial conditions $\theta_i$ are randomly drawn from $\mathcal{N}[0, 2\pi]$, and the natural frequencies $\omega$ are drawn from $\mathcal{N}(0, 1)$. Each sample was integrated by an adaptive stepsize for target precision RK45 numerical solver, and solutions at 1001 equally spaced time steps were computed in the time range $0 \leq t \leq 1001$. Finally, $\theta$ is returned as modulated such that $0 \leq \theta < 2\pi$. The derivative $\dot{\theta_i}$ is estimated by finite differences, yielding 1000 estimated differences. For the Erdős-Rényi graph (Fig. 2(A)), the sparsity was chosen such that it was as close as possible to the DTI graph (0.2363).

This, along with choosing a coupling that defined in terms of the adjacency matrix structure are experimental controls available to eliminate sparsity as a contributing factor in the outcome of the estimation. The Community graph Fig. 2(B) is created with by adding edges randomly with equal weight as well, however groups are chosen such that each node has a community. We chose 80 nodes with 5 communities of 16 nodes each. Each node has some number of intra-community connections and inter-community connections. We chose 13 intra-community connections and 5 inter-community connections. To generate these graphs efficiently we only guaranty that these are the maximum number of edges and choose the edges with one forward pass. 13 and 5 were chosen so that the community graph has a maximum sparsity of 0.225 as close as possible to the DTI graph. The Community graph is undirected which increases the odds of not finding a configuration with the maximum number of edges, and as such they tend to be sparser than the Erdős-Rényi graphs.

In Fig. 3 we show receiver operator characteristic (ROC) curves for a single realization of the network coupled dynamics discussed above, in which true positive and false positive rates are displayed together. It is desirable in an ROC curve for a method to come as close as possible to the upper left hand corner, representing almost no false positives and almost all true positives. In the three examples, networks from the Erdos Renyi example, the community example show in Fig. 2 and the DTI derived example, we compare the accuracy of standard methods such as correlation and LASSO to that of our entropic regression for these three scenarios. Clearly the entropic regression out performs correlation method. In the case of LASSO, the ROC curves are calculated across a range of values of the regularization parameter $\lambda$. We see that for some $\lambda$, that entropic regression clearly outperforms Lasso, but for some $\lambda$ it is close to a tie, with Lasso slightly outperforming entropic regression at least for the Erdos Renyi example. However, without a method of model selection, or otherwise knowing the answer apriori, it is not possible to know what is
More likely to match the dynamics exhibited by the brain. In all 20 examples of simulated dynamics from different DTI networks, entropic regression had a lower false positive rate than any of the networks produced by LASSO. This allows for increased confidence in the edges which are inferred by entropic regression in the context of the brain. This highlights the utility of our approach over other existing methods in neurological applications.

LASSO in all cases averaged more than triple the FPR of entropic regression over the 20 networks examined. This could have severe implications for the network tomography. For example a future treatment relying on accurate knowledge of connectivity between the ROIs would suffer greatly from having a significant number of false inferred edges.

In Table 1 we average across 20 sample simulations, and for each with a new sampled randomly generated network and randomly selected initial condition. The average TPRs and FPRs as well as the standard deviation of the two best methods are reported. As can be seen, on average entropic regression and LASSO perform similarly in true positive rate (TPR), however entropic regression significantly outperforms, especially in terms of false positive rate (FPR). Across the different k Overall, entropic regression is capable of recovering the majority of the true networks while only generating very few false edges, while LASSO also recovers the majority of the true network but introduces many more false edges.

IV. CONCLUSIONS

In this work we have presented a simple model for brain dynamics using Kuramoto oscillators running on synthetic ER and DTI networks. Having access to accurate network structure is essential to understanding the dynamics of any network coupled system. Unfortunately in the brain the ground truth of the network structure is unknown and thus must be inferred. Common methods for inference of these networks from data include LASSO and correlation. We show that a new method, entropic regression, offers improvement upon both LASSO and correlation in terms of accuracy. Specifically entropic re-
FIG. 3. Here we show the ROC curves comparing the performance of LASSO, correlation and entropic regression on three different network types. The solid lines represent the ROC curves and the star represents the parameter chose by the entropic regression method, and the triangle represents the parameter $\lambda$ chosen by the BIC method for LASSO. In all three types of network including the small world network as discussed in Ref. 18, it is clear that entropic regression outperforms both correlation and LASSO. Furthermore both entropic regression and LASSO offer significant performance benefits over correlation. As a result the performance of entropic regression is much better than that of both LASSO and correlation.

FIG. 4. Twenty runs, on a different network for each run are shown above. In (A) there are 20 different random Erdős-Rényi networks with Kuramoto dynamics. The true positive rates and false positive rates are shown for entropic regression and LASSO choosing the appropriate value of $\lambda$ either with BIC or using 1 standard error. (B) Random community networks and (C) DTI networks are used in the place of ER networks from (A). The DTI networks are reconstructed from 20 different patients in this case. It is clear that entropic regression clusters closest to the upper left hand corner, showing its improved performance over the other methods. These results are summarized in Ref. [4].
entropy regression offers a similar true positive rate with a much improved false negative rate over the other methods.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ACKNOWLEDGMENTS

E.B. and A.A., and A. D. were supported by the Army Research Office (N68164-EG) and, J.F. and E. B. were supported by DARPA.

1 B. F. Zhan, C. E. Noon Shortest path algorithms: an evaluation using real road networks Transportation Science 32, 65-73 (1998)
2 P. Morrell, C. Lu The environmental cost implication of hub–hub versus hub by-pass flight networks Transportation Research Part D: Transport and Environment 12, 143-157 (2007)
3 K. Lewis et. al. Tastes, ties, and time: A new social network dataset using Facebook.com Social Networks 30, 330-342 (2008)
4 B. Suh, L. Hong, P. Pirolli, E. H. Chi Want to be retweeted? Large scale analytics on factors impacting retweet in twitter network 2010 IEEE Second International Conference on Social Computing, 177-184
5 N. Tompkins, M. C. Cambria, A. L. Wang, M. Heymann, S. Fraden Functional brain mapping by blood oxygenation level dependent contrast magnetic resonance imaging. A comparison of signal derived from Diffusion Tensor Imaging with nonlinear parameter estimation with Kroncker covariance models arXiv, (2014)
6 A. Arenas, A. Díaz-Guilera, J. Kurths, Y. Moreno, C. Zhou Synchronization in complex networks 2003, Physica D: Nonlinear Phenomena, 227, 1, 78-99,
7 E. I. Pas, S. L. Principio 2013, Biophysical Journal 64, 803-812
8 E. D. Bullmore Small-world brain networks Nature 404, 295–307, (2000)
9 L. W ang, et. al. Dynamic reconfiguration of human brain networks during transfer entropy PRE 86, 061121 (2012)
10 T. Schreiber Measuring information transfer PRL, 85, 461 (2000)
11 W. M. Lord, J. Sun, E. M. Bollt Geometric k-nearest neighbor estimation of entropy and mutual information Chaos 28, 033141 (2018)
12 J. Runge, J. Heitzig, N. Marwan, J. Kurths Quantifying causal coupling strength: A lag-specific measure for multivariate time series related to transfer entropy 2010, Phys. Rev. Lett. 105, 034102 (2010)
13 E. B. and A.A, and A. D. were supported by the Army Research Office (N68164-EG) and, J.F. and E. B. were supported by DARPA.
14 M. E. J. Newman Random graphs as models of networks Handbook of graphs and networks 1, 36-68 (2003)
15 D. J. Watts, S. H. Strogatz Collective dynamics of ‘small-world’ networks Nature 393, 440-442 (1998)
16 L. Bonilha, et. al. Reproducibility of the Structural Brain Connectivity Derived from Diffusion Tensor Imaging PLoS One 10, e0135247 (2015)
17 T. E. Behrens, et. al. Characterization and propagation of uncertainty in diffusion-weighted MR imaging Magn Reson Med 50, 1077-1088 (2003)
18 B. Fischl, et. al. Whole brain segmentation: automated labeling of neuroanatomical structures in teh human brain Neuron 33, 341-355 (2002)
19 P. Hagemann, et. al. Mapping the Structural Core of Human Cerebral Cortex PLoS Biol 6, e159 (2008)
20 A. E. Raftery, D. Madigan, J. A. Hoeting Bayesian model averaging for linear regression models Journal of the American Statistical Association 92, 772–782 (1997)
21 G. H. Golub and C. F. Van Loan Matrix Computations Johns Hopkins University Press, (2013)
22 R. Tibshirani Regression shrinkage and selection via the lasso Journal of the Royal Statistical Society: Series B 58, 267-288 (1996)
23 E. B. and A.A, and A. D. were supported by the Army Research Office (N68164-EG) and, J.F. and E. B. were supported by DARPA.
24 B. J. Stoltz, H. A. Harrington, M. A. Porter Persistent homology of time-dependent functional networks constructed from coupled time series Chaos: An Interdisciplinary Journal of Nonlinear Science 27, 047410 (2017)
25 P. Erdős, A. Rényi On the evolution of random graphs Publ. Math. Inst. Hung. Acad. Sci. 5, 17-60 (1960)
26 M. A. R. AlMomani, J. Sun, E. Bollt How entropic regression beats the outliers problem in nonlinear system identification Chaos: An Interdisciplinary Journal of Nonlinear Science 30, 013107 (2020)
27 J. G. Stoltz, H. A. Harrington, M. A. Porter Persistent homology of time-dependent functional networks constructed from coupled time series Chaos: An Interdisciplinary Journal of Nonlinear Science 27, 047410 (2017)
### Table I

|                      | ER Network | Small World Network | DTI Network |
|----------------------|------------|---------------------|-------------|
| Mean TPR Ent. Reg.   | 0.8938 (0.1035) | 0.8858 (0.0892)   | 0.7983 (0.0940) |
| Mean TPR LASSO (BIC)| 0.9538 (0.0571) | 0.9150 (0.086)    | 0.8994 (0.0743) |
| Mean FPR Ent. Reg.  | 0.0535 (0.0301) | 0.0551 (0.0305)   | 0.0718 (0.0268) |
| Mean FPR LASSO (BIC)| 0.2844 (0.0410) | 0.1647 (0.0247)   | 0.2350 (0.0440) |

TABLE I. True and False Positive rates (TPR and FPR) of entropic regression and LASSO averaged over 20 synthetic network realizations of either Erdos-Renyi or Small World Network types, and 20 DTI networks constructed from different patients. On average both LASSO and entropic regression return similar TPR, however entropic regression returns a much lower FPR than LASSO. In the parenthesis the standard deviation is reported.
This figure "fig_1.png" is available in "png" format from:

http://arxiv.org/ps/2010.04247v2