

# Supporting Information

## Recovering lost and absent information in temporal networks

July 24, 2021

James P. Bagrow and Sune Lehmann

### Contents

<b>S1 Dataset descriptions</b>	<b>S-1</b>
S1.1 Time series of network sparsity . . . . .	S-2
<b>S2 Finding sparse solutions</b>	<b>S-2</b>
S2.1 Hyperparameter selection . . . . .	S-3
<b>S3 Multi-scale backbones</b>	<b>S-4</b>
<b>S4 Least-norm error bound</b>	<b>S-4</b>
<b>S5 Network structure and error bounds for sparse solutions</b>	<b>S-6</b>
S5.1 Comparison to network corpus . . . . .	S-8
S5.2 Comparison to network features . . . . .	S-8

### List of Figures

S1 Sparsity over time across network datasets . . . . .	S-3
S2 Hyperparameter selection using Bayes Information Criterion . . . . .	S-4
S3 Receiver operating characteristic (ROC) curves comparing multiscale backbones for full corpus . . . . .	S-5
S4 Minimum eigenvalue of $\mathbf{B}_t^T \mathbf{B}_t$ over time . . . . .	S-8
S5 The maximum diameter over the connected component(s) of $G_t$ over time . . . . .	S-8
S6 Empirical correlations between recovery error and different node and edge features. . . . .	S-10

### List of Tables

S1 Regularization hyperparameters selected with BIC. . . . .	S-4
S2 Correlations between $\ \mathbf{E}_t - \hat{\mathbf{E}}_t\ _2$ and the three measures . . . . .	S-9

## S1 Dataset descriptions

For this study we utilize five temporal network datasets taken across a variety of scientific domains:

**Copenhagen** A temporal physical proximity network tracking the smartphones of individuals participating in the Copenhagen Networks study [1, 2]. Edges were recorded as timestamped bluetooth interaction between participants. Interactions with devices outside the experiment were ignored. To process these data, edges with at least 200 interactions were retained and temporal interactions were binned from 5 minutes and then aggregated into 1h bins. Edge directionality was not used. After processing, the network consisted of  $n = 544$  nodes,  $m = 2389$  edges, and  $T = 672$  time periods. The original data are available at <https://doi.org/10.6084/m9.figshare.7267433>.

**Hospital** A temporal physical proximity network tracking patient-to-patient, patient-to-worker, and worker-to-worker contacts within a hospital ward in Lyon, France [3]. Health care workers and patients wore wearable sensors (RFID) that recorded close proximity and undirected edges were added to the temporal network when an close interaction was recorded (i.e., two wearables exchanged a packet) during a 20s window. The study covered the time period Monday, Dec. 6, 2010 1pm to Friday, Dec. 10 2pm. To process these data, edges with at least 20 interactions were retained and temporal interactions were aggregated into 1h bins. After processing, the network consisted of  $n = 68$  nodes,  $m = 329$  edges, and  $T = 96$  time periods. The original data are available at <http://www.sociopatterns.org/datasets/hospital-ward-dynamic-contact-network/>.

**Ant Colony** A temporal physical proximity network tracking interactions between ants (*Temnothorax rugatulus*) [4]. Edges were recorded between ants when one ant’s antenna made contact with another ant’s body. The colony was housed in a transparent nest and filmed; individual ants were distinguished by manual marks made with colored paint. Here we focused on Colony 1-1, the first filming period for colony 1. Filming sessions of all colonies covered periods of approximately 1800s. Data were provided as PDF table files which were extracted to plain text csv files using the Camelot Python package. To process these data, edges with at least 3 interactions were retained and temporal interactions were aggregated into 10s bins. Edge directionality was not used. After processing, the network consisted of  $n = 85$  nodes,  $m = 269$  edges, and  $T = 143$  time periods. The original data are available at <https://doi.org/10.1371/journal.pone.0020298.s008>.

**Manufacturing Email** A temporal social network tracking internal emails sent between employees of a mid-size manufacturing firm [5]. Edges were recorded when emails were sent between email addresses; emails with multiple recipients (including CC and BCC recipients) were recorded as multiple simultaneous edges. The data covered the period Jan. 1, 2010 to Sep. 30, 2010. To process these data, edges with at least 10 interactions were retained and temporal interactions were aggregated into 1 day bins. Edge directionality was not used. After processing, the network consisted of  $n = 140$  nodes,  $m = 1137$  edges, and  $T = 273$  time periods. The original data are available at <https://doi.org/10.7910/DVN/6Z3CGX>.

**College Message** An temporal social network originating from an online community of college students attending University of California, Irvine [6]. Timestamped links were recorded when messages were sent between students during the period April to October, 2004. To process these data, edges with at least 10 interactions were retained and temporal interactions were aggregated into 1 day bins. Edge directionality was not used. After processing, the network consisted of  $n = 609$  nodes,  $m = 1323$  edges, and  $T = 194$  time periods. The original data are available at [https://toreopsahl.com/datasets/#online\\_social\\_network](https://toreopsahl.com/datasets/#online_social_network).

### S1.1 Time series of network sparsity

Figure S1 shows the sparsity patterns of the temporal networks over time. The five representative networks all display different patterns or time scales. For instance, the interaction networks (Copenhagen and Hospital) display a daily periodicity, with Copenhagen covering four weeks compared to Hospital’s four days. In contrast, Ant Colony covers a short time period of 30 minutes while Manufacturing Email and College Message both cover several months of time. College Message also exhibits a non-stationary signature not present in the periodicity of Manufacturing Email with a burst of activity early on followed by a long period of low activity. All networks but Copenhagen exhibit a few time periods of zero activity (highlighted in Fig. S1 with red vertical bands).

## S2 Finding sparse solutions

The system  $\mathbf{BE} = \mathbf{N}$  is underdetermined, but sparse solutions  $\hat{\mathbf{E}}$  for the unknown  $\mathbf{E}$  can be found by solving the following optimization problem (Eq. (S1)):

$$\min_{\mathbf{E}} \quad \frac{1}{2n} \|\mathbf{N} - \mathbf{BE}\|_F^2 + \alpha \|\mathbf{E}\|_{21} \quad (\text{S1a})$$

$$\text{subject to} \quad E_{ij} \geq 0, i = 1, \dots, m, j = 1, \dots, T, \quad (\text{S1b})$$

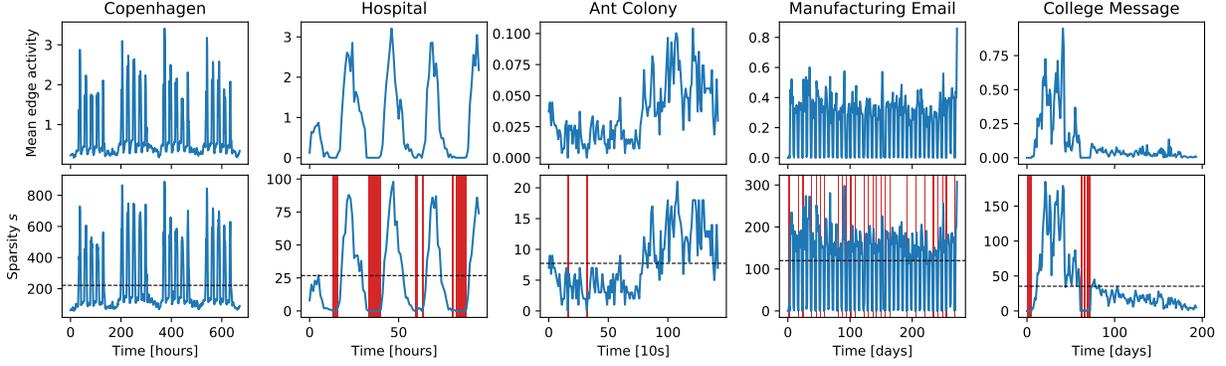


Figure S1: Sparsity over time across network datasets. (Top row) Mean edge activity  $\langle \mathbf{E}_t \rangle$  over time, where  $\mathbf{E}_t$  is the  $t$ -th column of the edge activity matrix  $\mathbf{E}$ . (Bottom row) Sparsity  $s_t$ , the number of active edges at time period  $t$ . Vertical red lines highlight  $t$  where  $s_t = 0$ ; cf. Fig. 2, second column.

where  $\|\cdot\|_F$  is the Frobenius norm,  $\|\mathbf{E}\|_{21} = \sum_i \sqrt{\sum_j E_{ij}^2}$  is a mixed  $L_1/L_2$  regularization term and  $\alpha$  serves as a hyperparameter balancing sparsity of  $\hat{\mathbf{E}}$  and solution accuracy (via the least-squares loss  $\|\mathbf{N} - \mathbf{B}\mathbf{E}\|_F$ ). The constraint  $\mathbf{E} > 0$  enforces that edge activity data are nonnegative. (The problem Eq. (S1) can be equivalently expressed with the regularization term as a constraint.)

Equation (S1) can be interpreted as a (multi-target) Lasso regression without an intercept term and with the added constraint enforcing nonnegative regression coefficients. The  $L_1/L_2$  regularization term is used to find sparse solutions for multiple regression problems (columns of  $\mathbf{N}$  and  $\hat{\mathbf{E}}$ ) jointly by treating each edge as a group over time [7, 8].

Solutions were found in Python version 3.8.8 using the Lasso implementation of Sci-Kit Learn version 0.24.1 [9]. Coordinate descent was used with a precomputed Gram matrix and random coefficient selection to speed convergence. Bayes Information Criterion was used to select  $\alpha$  (Sec. S2.1). All other parameters were left to their v0.24.1 defaults.

## S2.1 Hyperparameter selection

Solving Eq. (S1) requires first determining a value of  $\alpha$ , the regularization hyperparameter. Different values of  $\alpha$  can lead to very different solutions, so care must be taken to find the most appropriate value of  $\alpha$  possible. Generally, in a machine learning or statistical learning context, it is best to  $\alpha$  using additional held-out data, to help ensure your fitted model generalizes to new data. However, in our problem, we are performing inference, not learning, and while it is straightforward to use cross-validation to select  $\alpha$  and fit the model with different data subsets, we can safely use the same data for both selecting  $\alpha$  and estimating  $\hat{\mathbf{E}}$ .

Information criteria are an effective way to estimate hyperparameters [10, 11]. These balance the complexity of the model with its accuracy. For regularized regression, model complexity maps to the number of nonzero regression coefficients, which is controlled by the value of  $\alpha$ . This allows us to select the value of  $\alpha$  by minimizing the BIC: Here we use Bayes Information Criterion (BIC):

$$\text{BIC} = \frac{nT \left\langle (\mathbf{N} - \mathbf{B}\hat{\mathbf{E}})^2 \right\rangle}{\sigma^2} + \log(nT) \widehat{df} \quad (\text{S2})$$

where  $\langle (\mathbf{N} - \mathbf{B}\hat{\mathbf{E}})^2 \rangle$  is the mean squared error,  $\sigma^2 = \text{Var}(\mathbf{N})$  is the variance of elements of  $\mathbf{N}$ , and  $\widehat{df}$  is the number of nonzero coefficients, an unbiased estimator of the Lasso degrees of freedom [12]. Equation (S2) is a function of  $\alpha$  for a given temporal network dataset because the data are fixed and  $\hat{\mathbf{E}}$  depends on  $\alpha$ .

Figure S2 plots the BIC as a function of  $\alpha$  for each network under study. Table S1 displays the final selected  $\alpha$  for each network. Solving Eq. (S1) many times can be costly; fortunately, the path of all solutions can be computed

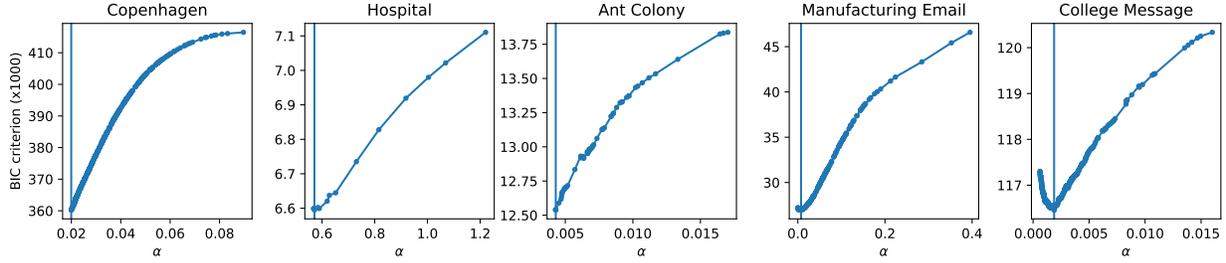


Figure S2: Hyperparameter selection using Bayes Information Criterion (BIC). For each network we plot the BIC curve as a function of the  $\alpha$  regularization parameter. The smallest value and our final choice for the regularization parameter is highlighted with a vertical line. BIC was computed along each piecewise-linear step of the LARS solution path and the  $\alpha$  minimizing BIC was selected. Note that the non-negativity constraint in Eq. (S1) means that the solution path does not converge to least-squares as  $\alpha \rightarrow 0$ . Here we report only those  $\alpha > 0$  that were reached by the step-wise LARS path as only those coefficients are in congruence with the coordinate descent Lasso.

efficiently using the LARS method [13]. Sci-kit Learn’s current LARS implementation does not support multi-target regression problems. To counter this, we reshaped the problem into a single regression by stacking the  $T$  columns of  $\mathbf{N}$  into a single vector, replicating  $\mathbf{B}$   $T$  times to match, and computing the LARS solution path using an  $L_1$  penalty. Then BIC was computed along the LARS solution path to select  $\alpha$ , which was then substituted back into Eq. (S1) and optimized to yield our final solution  $\hat{\mathbf{E}}$ .

Table S1: Regularization hyperparameters selected with BIC.

	$\alpha$ (BIC)
Copenhagen	0.019958537790598734
Hospital	0.5710784313725501
Ant Colony	0.004292537772684784
Manufacturing Email	0.007326007326005932
College Message	0.0018803877107223018

### S3 Multi-scale backbones

Supplementing main text Fig. 4, Fig. S3 shows the least-norm solution and sparse solution ROC curves for all networks in the corpus. The main text figure only showed the sparse solution ROC curve for one network (Copenhagen) and then used plots of the AUC to summarize performance across networks and both solutions.

### S4 Least-norm error bound

Here we derive the bound on estimation error for the classic, least-norm approach to solving the underdetermined system  $\mathbf{B}\mathbf{E} = \mathbf{N}$  discussed in the main text.

While  $\mathbf{B}^{-1}$  in general does not exist, the pseudoinverse  $\mathbf{B}^+$  can be efficiently computed with SVD and used to find the minimum or least-norm solution [14]:

$$\hat{\mathbf{E}} = \mathbf{B}^+\mathbf{N} = \mathbf{B}^+\mathbf{B}\mathbf{E} \neq \mathbf{E} \tag{S3}$$

where the last line follows from  $\mathbf{B}^+\mathbf{B} \neq \mathbf{I}$  since  $\mathbf{B}^+$  is a right-inverse, not a left-inverse, of  $\mathbf{B}$ . Since  $\|\mathbf{B}\hat{\mathbf{E}} - \mathbf{N}\|_F^2 = 0$

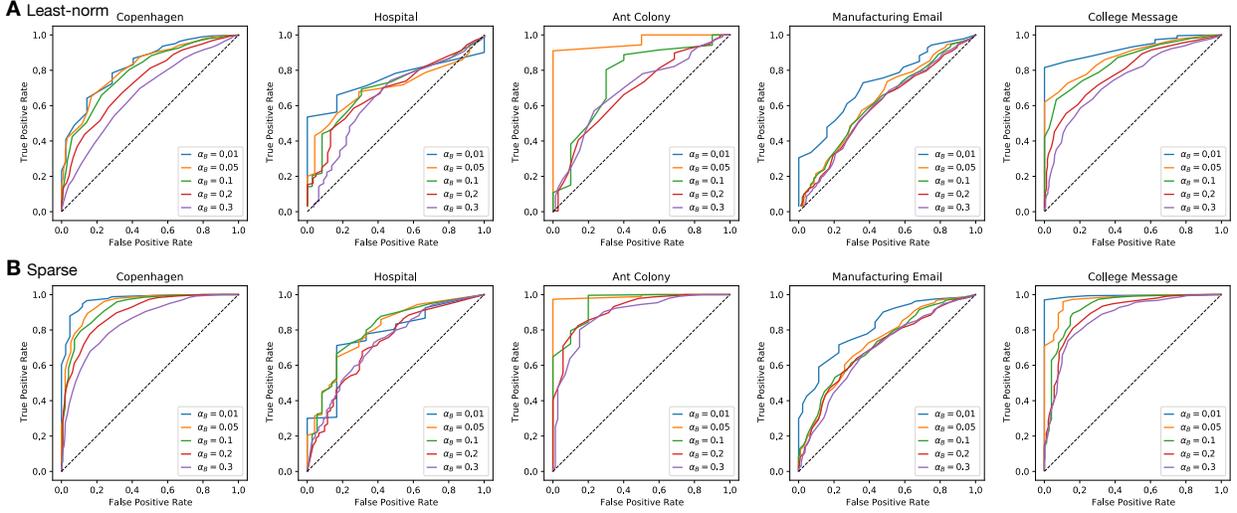


Figure S3: Receiver operating characteristic (ROC) curves comparing multiscale backbones for full corpus. This supporting plot supplements main text Fig. 4B which shows the ROC curve only for the Copenhagen network. **(A)** Backbones found with inferred edge weights recovered from the least-norm solution. **(B)** Backbones found with inferred edge weights recovered from the sparse solution. The diagonal dashed line corresponds to a classifier that is randomly guessing whether or not an edge is within the backbone; better-performing classifiers have ROC curves that push further into the upper-left corner of the ROC plot.

(assuming  $\hat{\mathbf{E}}$  exists), we cannot characterize the solution using the residual. Instead, at least theoretically, we consider the estimation error, the difference between  $\mathbf{E}$  and  $\hat{\mathbf{E}}$ :

$$\hat{\mathbf{E}} - \mathbf{E} = \mathbf{B}^+ \mathbf{B} \mathbf{E} - \mathbf{E} = (\mathbf{B}^+ \mathbf{B} - \mathbf{I}) \mathbf{E} \quad (\text{S4})$$

$$\|\hat{\mathbf{E}} - \mathbf{E}\|_F = \|(\mathbf{B}^+ \mathbf{B} - \mathbf{I}) \mathbf{E}\|_F \quad (\text{S5})$$

$$\leq \|(\mathbf{B}^+ \mathbf{B} - \mathbf{I})\|_F \|\mathbf{E}\|_F \quad (\text{S6})$$

$$\leq \sqrt{m - n} \|\mathbf{E}\|_F \quad (\text{S7})$$

where  $\|\cdot\|_F$  is the Frobenius norm, Eq. (S6) follows from the Cauchy-Schwarz inequality, and we now derive  $\|(\mathbf{B}^+ \mathbf{B} - \mathbf{I})\|_F \leq \sqrt{m - n}$ .

Recall the (unoriented) incidence matrix  $\mathbf{B}$  is an  $n \times m$  matrix with right-inverse  $\mathbf{B}^+$  such that  $\mathbf{B}^+ \mathbf{B}$  is  $m \times m$ . (We assume  $\mathbf{B}$  is full rank; otherwise, either the problem is not underdetermined and can be solved or the network is disconnected and we can find the solution individually for each connected component.) The Frobenius norm of a matrix  $\mathbf{A}$  is  $\|\mathbf{A}\|_F^2 = \text{Tr}(\mathbf{A} \mathbf{A}^*)$ . Therefore

$$\begin{aligned} \|\mathbf{B}^+ \mathbf{B} - \mathbf{I}\|_F^2 &= \text{Tr}[(\mathbf{B}^+ \mathbf{B} - \mathbf{I})(\mathbf{B}^+ \mathbf{B} - \mathbf{I})^*] \\ &= \text{Tr}[(\mathbf{B}^+ \mathbf{B} - \mathbf{I})((\mathbf{B}^+ \mathbf{B})^* - \mathbf{I})] && \text{(since } \mathbf{I}^* = \mathbf{I}) \\ &= \text{Tr}[(\mathbf{B}^+ \mathbf{B} - \mathbf{I})(\mathbf{B}^+ \mathbf{B} - \mathbf{I})] && \text{(}\mathbf{B}^+ \mathbf{B} \text{ is Hermitian so is } \mathbf{B}^+ \mathbf{B} - \mathbf{I}) \\ &= \text{Tr}[\mathbf{B}^+ \mathbf{B} - \mathbf{B}^+ \mathbf{B} - \mathbf{B}^+ \mathbf{B} + \mathbf{I}^2] && \text{(}\mathbf{B} \mathbf{B}^+ = \mathbf{I} \text{ since } \mathbf{B}^+ \text{ is a right-inverse of } \mathbf{B}) \\ &= \text{Tr}[\mathbf{I} - \mathbf{B}^+ \mathbf{B}] \\ &= \text{Tr} \mathbf{I} - \text{Tr}(\mathbf{B}^+ \mathbf{B}) \\ &\leq m - n. && \text{(}\mathbf{B}^+ \mathbf{B} \text{ is idempotent so } \text{Tr}(\mathbf{B}^+ \mathbf{B}) = \text{rank}(\mathbf{B}^+ \mathbf{B}) = \text{rank}(\mathbf{B}) \leq n, \text{ since } \text{rank}(\mathbf{B}) \leq n) \end{aligned}$$

*Remark.* In general, we can expect  $\text{rank}(\mathbf{B}) = n$  for most networks; however, this is not guaranteed. Indeed, a connected tree will have  $\text{rank}(\mathbf{B}) = n - 1$ . Likewise, the rank of the incidence matrix is also less than  $n$  for disconnected networks.

## S5 Network structure and error bounds for sparse solutions

Here we exploit theory developed for high-dimensional ( $m \gg n$ ) statistical learning and optimization problems to understand when the estimation error  $\|\mathbf{E}_t^* - \hat{\mathbf{E}}_t\|_2^2$  can be bounded. Problems in this area include basis pursuit, compressed sensing, and penalized regression [15, 16, 17]. We connect this theory with results from spectral graph theory to determine what network structure is necessary for the error of the recovery problem to be bounded.

To understand the performance of the optimization problem we need to investigate its loss function. If the loss function is strongly convex, then a single minima exists and, for our optimization problem, theoretical bounds can be put on the estimation error. If, however, the loss function is not strongly convex but only convex, then there will exist directions for which the loss is flat, and the optimization solution can become arbitrarily far away, preventing convergence.

For our optimization (Eq. (S1)), consider, at timestep  $t$ , i.e., column  $t$  of  $\mathbf{E}$  and  $\mathbf{N}$ , the least-squares loss  $f(\mathbf{E}_t) = \frac{1}{2n} \|\mathbf{N}_t - \mathbf{B}\mathbf{E}_t\|_2^2$  is always convex. It is strongly convex when the  $m \times m$  Hessian matrix for this loss,  $\nabla^2 f = \mathbf{B}^T \mathbf{B}/n$ , has eigenvalues bounded away from zero. Unfortunately, a Gram matrix of the form  $\mathbf{B}^T \mathbf{B}$  where  $\mathbf{B}$  is  $n \times m$  will have rank at most  $\min\{n, m\}$  and so will be rank-deficient in the high-dimensional setting ( $n < m$ ) [18].

This forces us to seek weaker requirements than strong convexity for the loss function. Instead, we consider *restricted* strong convexity, where strong convexity is required only for a subset  $C \subset \mathbb{R}^m$  of estimation error vectors  $\mathbf{v} \equiv \mathbf{E}_t^* - \mathbf{E}_t \in C$ . A loss function  $f(\mathbf{E}_t)$  satisfies restricted strong convexity at  $\mathbf{E}_t^*$  if

$$\frac{\mathbf{v}^T \nabla^2 f(\mathbf{E}_t) \mathbf{v}}{\|\mathbf{v}\|_2^2} \geq \gamma \text{ for all nonzero } \mathbf{v} \in C, \quad (\text{S8})$$

and for all  $\mathbf{E}_t \in \mathbb{R}^m$  in the neighborhood of  $\mathbf{E}_t^*$ .

Specializing this to a linear model means that we seek to lower bound the *restricted* eigenvalues of  $\mathbf{B}^T \mathbf{B}$ :

**Definition S5.1** (Restricted eigenvalue (RE) condition for linear regression).

$$\frac{\frac{1}{n} \mathbf{v}^T \mathbf{B}^T \mathbf{B} \mathbf{v}}{\|\mathbf{v}\|_2^2} \geq \gamma \text{ for all nonzero } \mathbf{v} \in C. \quad (\text{S9})$$

Consider the regularized form of the Lasso<sup>1</sup>. The restricted set  $C$  for the regularized Lasso is defined based on the sparsity of the parameter  $\mathbf{E}_t^*$ . Suppose the “active set” of edges at time  $t$  is  $S_t(\mathbf{E}_t)$ , with  $|S_t| = s_t$ . For the Lasso error  $\mathbf{v}_t = \mathbf{E}_t^* - \hat{\mathbf{E}}_t$ , let  $\mathbf{v}_{t,S}$  be the subvector of  $\mathbf{v}_t$  with elements corresponding to index set  $S_t$ , the nonzero elements of  $\mathbf{E}_t$ . Define the complement likewise:  $\mathbf{v}_{t,S^c}$ . Then, the set of error vectors for regularized Lasso, depending on the regularization parameter [18], will satisfy a cone constraint:

$$C(S) \equiv \{ \mathbf{v} \in \mathbb{R}^m \mid \|\mathbf{v}_{S^c}\|_1 \leq 3 \|\mathbf{v}_S\|_1 \}. \quad (\text{S10})$$

Moving on to the network structure,  $\mathbf{B}$  is the  $n \times m$  unoriented incidence matrix of the graph  $G$  and this means our regression problem is related to the *line graph* of  $G$ :

**Definition S5.2** (Line graph). The line graph  $L(G)$  of a graph  $G$  is the graph where each node in  $L(G)$  corresponds to an edge in  $G$ , and an edge exists between two nodes in  $L(G)$  when the corresponding edges in  $G$  share an endpoint. The adjacency matrix  $\mathbf{A}(L(G))$  is of size  $m \times m$  for  $G$  with  $m$  edges.

Specifically, the regression problem is related to the line graph by

$$\mathbf{B}^T \mathbf{B} = 2\mathbf{I} + \mathbf{A}(L(G)). \quad (\text{S11})$$

Each eigenvalue of  $\mathbf{B}^T \mathbf{B}$  is  $2 + \lambda_j$ , where  $\lambda_j$  is an eigenvalue of  $\mathbf{A}(L(G))$ . Therefore, we seek to understand the

<sup>1</sup>For simplicity, here we consider the per-timestep Lasso, which is closely aligned with the group Lasso formulation we used. When analyzing each time period  $t$  separately (considering column  $\mathbf{E}_t$  of  $\mathbf{E}$ ), the optimization problem Eq. (S1) reduces to regularized Lasso (barring the non-negativity constraints).

eigenvalues of  $\mathbf{A}(L(G))$  for a given graph  $G$ , specifically when the minimum eigenvalue  $\lambda_{\min} \equiv \min_j \lambda_j > -2$ . Taking  $G = G_t$ , the graph of active edges at time  $t$ , we can then use  $\lambda_{\min}$  to determine if RE holds at time  $t$ .

To proceed, first we state two theorems due to Doob [19]:

**Theorem S5.1.** *The smallest eigenvalue  $\lambda_{\min}$  of the adjacency matrix of the line graph of  $G$  is*

$$\lambda_{\min}(\mathbf{A}(L(G))) = -2 \quad (\text{S12})$$

*unless every connected component of  $G$  is a tree or has one cycle only, of odd length.*

**Theorem S5.2.** *If  $G$  is a graph of diameter  $d$ , then*

$$-2 \leq \lambda_{\min}(\mathbf{A}(L(G))) \leq -2 \cos\left(\frac{\pi}{d+1}\right) \quad (\text{S13})$$

*Further, these bounds are the best possible.*

If there exists an eigenvector  $\mathbf{x}$  of  $\mathbf{B}^T \mathbf{B}$  with eigenvalue zero, and  $\mathbf{x} \in C$ , then the estimation error will not be bounded. In fact, we need only to consider the submatrix  $\mathbf{B}_t$  indexed by the active edge set  $S_t$ , i.e., the incidence matrix of  $G_t$ :

**Theorem S5.3.** *Every eigenvector of  $\mathbf{B}_t^T \mathbf{B}_t$  has a corresponding vector in  $C(S_t)$ .*

*Proof.* Let  $\mathbf{x} \in \mathbb{R}^{S_t}$  be an eigenvector of  $\mathbf{B}_t^T \mathbf{B}_t$ . Using  $S_t \subset [1..m]$ ,  $|S_t| = s_t$ , as an index set, define  $\mathbf{x}' \in \mathbb{R}^m$  to be the vector with elements

$$x'_j = \begin{cases} x_{\sigma(j)} & j \in S_t, \\ 0 & \text{otherwise,} \end{cases} \quad (\text{S14})$$

where  $\sigma : S_t \rightarrow [1..s_t]$  maps vector indices of  $\mathbb{R}^m$  to those of  $\mathbb{R}^{S_t}$ , as subsetted by  $S_t$ . In other words,  $\mathbf{x}'$  is the vector with active set  $\mathbf{x}'_{S_t} = \mathbf{x}$  and the complement  $\mathbf{x}'_{S_t^c} = \mathbf{0}$ . By the construction of  $S_t$ ,  $\|\mathbf{x}'_{S_t}\|_1 = \|\mathbf{x}\|_1 > 0$ , and so

$$\left\| \mathbf{x}'_{S_t^c} \right\|_1 < \delta \left\| \mathbf{x}'_{S_t} \right\|_1$$

for any  $\delta > 0$ . Therefore, the cone constraint of  $C(S_t)$  is satisfied (with  $\delta = 3$ ) and  $\mathbf{x}' \in C(S_t)$ .  $\square$

In summary, Theorem S5.3 tells us that we do not need to consider the eigenvector corresponding to  $\lambda_{\min}$ ; it suffices only to determine whether  $\lambda_{\min}(\mathbf{A}(L(G_t))) > -2$  per Thm. S5.1 and if so then the smallest eigenvalue of  $\mathbf{B}^T \mathbf{B} > 0$ , RE holds at time  $t$ , and the estimation error  $\|\mathbf{E}_t^* - \hat{\mathbf{E}}_t\|_2$  will be bounded.

*Remark.* While a lower bound on  $\lambda_{\min}$  tighter than  $\lambda_{\min} > -2$  would be useful when Thm. S5.1 holds, Thm. S5.2 can be used to specify an upper bound on  $\lambda_{\min}$ . Of course, given an observed  $G_t$ , one can compute the smallest eigenvalue of  $\mathbf{B}^T \mathbf{B}$  directly to determine  $\gamma$  and apply an appropriate bound. Bounds on estimation (and prediction) error in high-dimension settings are the subject of considerable research in many contexts [20, 17, 21, 22, 8, 23]. See [18, 24] for recent summaries.

*Remark.* If  $G_t$  is disconnected,  $L(G_t)$  will also be disconnected and  $\mathbf{A}(L(G_t))$  will be reducible. Because  $L(G_t)$  is undirected,  $\mathbf{A}(L(G_t))$  can further be put into block diagonal form when  $G_t$  is disconnected. Therefore, the spectrum of  $\mathbf{A}(L(G_t))$  will be the union of the spectra of each connected component of  $L(G_t)$  and Thm. S5.2 can be applied to upper-bound the smallest eigenvalue using

$$d_{\max} = \max \{d(L(g_t)) \mid g_t \in \text{components}(G_t)\}, \quad (\text{S15})$$

where  $d(L(g_t))$  is the diameter of the line graph of connected component  $g_t \in G_t$  and  $\text{components}(G)$  is the set of all connected component subgraphs of  $G$ .

## S5.1 Comparison to network corpus

Lastly, we contextualize our results above using the five networks in our corpus. For each network at each time  $t$ , we computed  $\lambda_{\min}(\mathbf{B}_t^T \mathbf{B}_t)$  in Fig. S4 and  $d_{\max}$  (Eq. (S15)) in Fig. S5. We see in Fig. S4 that the  $\lambda$  gives us a good indication of the difficulty of recovering  $\mathbf{E}$  from  $\mathbf{N}$  and  $\mathbf{B}$ . For Copenhagen, Hospital, and Manufacturing Email, we observe many periods where  $\lambda_{\min} = 0$ , indicating that the estimation error  $\|\mathbf{E}_t^* - \hat{\mathbf{E}}_t\|_2^2$  cannot be bounded. In comparison, Ant Colony and College Message should be far easier recovery problems, as  $\lambda_{\min} > 0$  for most  $t$ . Indeed, this comports well with the overall errors we report in the main text (Fig. 5).

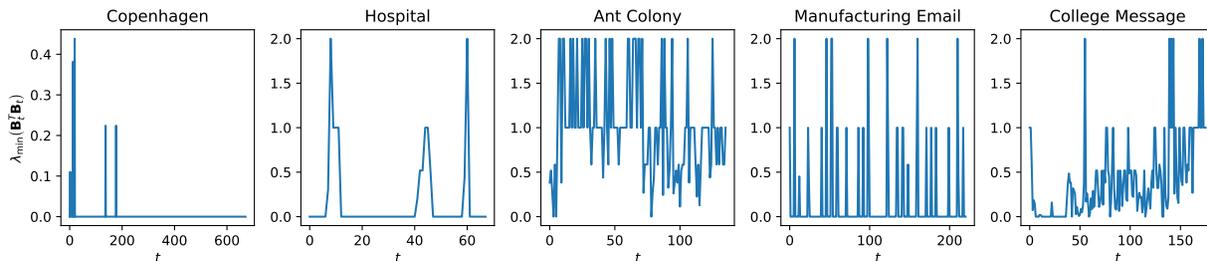


Figure S4: Minimum eigenvalue  $\lambda_{\min}$  of  $\mathbf{B}_t^T \mathbf{B}_t$  over time. Estimation error can only be bounded when  $\lambda_{\min} > 0$ . Copenhagen, Hospital, and Manufacturing Email networks are the most challenging recovery problems in our corpus, as  $\lambda_{\min}$  is frequently zero. Ant Colony and College Message, in contrast, are far easier recovery problems, as both frequently have  $\lambda_{\min} > 0$ . In Sec. S5 we derive the graph-theoretic properties governing whether  $\lambda_{\min} > 0$ .

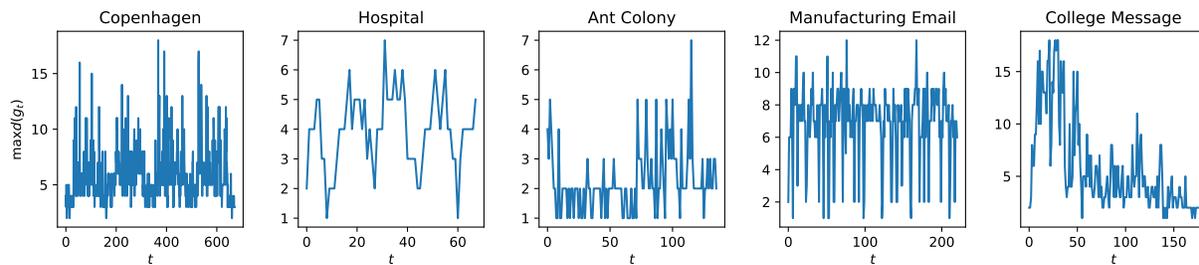


Figure S5: The maximum diameter  $d$  over the connected component(s) of  $G_t$  over time (see remark, Thm. S5.3).

Lastly, we support the observation in the main text that node activation fractions  $n_t/n$  correlate well with estimation error, nearly as well as edge activation fractions  $s_t/m$ , in Table S2 (in the main text we compared activation fractions with relative estimation error). Here we again see that estimation error correlates nearly as well with  $n_t/n$  as it does with  $s_t/m$ . While  $s_t/m$  provides better correlation in all networks,  $s_t$  will not be known in practice, while  $n_t$  will be, and so the closeness in correlation means  $n_t$  can be used as a suitable proxy for  $s_t$  when trying to judge the magnitude of recovery errors.

## S5.2 Comparison to network features

Expanding up main text Sec. 4 and Fig. 5, we further report relationships between estimation error and network properties in Fig. S6. Many simple measures for an edge centrality, such as edge betweenness or various combinations of the degrees of the nodes incident to the edge, actually show weak correlation with recovery error. For the latter measures,  $\min(k_i, k_j)$  tends to be most correlated with error for edge  $(i, j)$ —in all networks, this measure was more correlated with error than  $\max(k_i, k_j)$ ,  $(k_i + k_j)/2$ , and  $|k_i - k_j|$ . This measure was also more correlated with error

Table S2: Correlations between  $\|\mathbf{E}_t - \hat{\mathbf{E}}_t\|_2$  and the three measures. While the third measure,  $n_t/n$ , is the least correlated, it still gives correlation quite close in value to the other measures. This is important because any measure involving  $s_t$  cannot be computed in practice. We report both Pearson correlation  $r$  and Spearman correlation  $\rho$ .

	Corr.	$s_t/m$	$\sqrt{s_t \frac{\log m}{n}}$	$n_t/n$
Copenhagen	$r$	0.94350	0.93104	0.84295
	$\rho$	0.86529	0.86529	0.79140
Hospital	$r$	0.85576	0.84924	0.81097
	$\rho$	0.90346	0.90346	0.86602
Ant Colony	$r$	0.72979	0.73464	0.70718
	$\rho$	0.74810	0.74810	0.73497
Manufacturing Email	$r$	0.80367	0.81359	0.79153
	$\rho$	0.69007	0.69007	0.62962
College Message	$r$	0.89974	0.84597	0.84190
	$\rho$	0.92486	0.92486	0.90039

than edge betweenness. The sparsity of the edge, the proportion of times it was active, was also correlated with error, with more active edges tending to have higher recovery errors for most networks, with College Message being a notable exception.

## References

- [1] P. Sapiezynski, A. Stopczynski, D. D. Lassen, and S. Lehmann, “Interaction data from the Copenhagen Networks Study,” *Scientific Data*, vol. 6, no. 1, pp. 1–10, 2019.
- [2] A. Stopczynski, V. Sekara, P. Sapiezynski, A. Cuttone, M. M. Madsen, J. E. Larsen, and S. Lehmann, “Measuring large-scale social networks with high resolution,” *PLOS ONE*, vol. 9, no. 4, p. e95978, 2014.
- [3] P. Vanhems, A. Barrat, C. Cattuto, J.-F. Pinton, N. Khanafer, C. Régis, B.-a. Kim, B. Comte, and N. Voirin, “Estimating potential infection transmission routes in hospital wards using wearable proximity sensors,” *PLOS ONE*, vol. 8, no. 9, pp. 1–9, 2013.
- [4] B. Blonder and A. Dornhaus, “Time-ordered networks reveal limitations to information flow in ant colonies,” *PLOS ONE*, vol. 6, no. 5, pp. 1–8, 2011.
- [5] R. Michalski, S. Palus, and P. Kazienko, “Matching organizational structure and social network extracted from email communication,” in *Business Information Systems* (W. Abramowicz, ed.), (Berlin, Heidelberg), pp. 197–206, Springer Berlin Heidelberg, 2011.
- [6] P. Panzarasa, T. Opsahl, and K. M. Carley, “Patterns and dynamics of users’ behavior and interaction: Network analysis of an online community,” *Journal of the American Society for Information Science and Technology*, vol. 60, no. 5, pp. 911–932, 2009.
- [7] M. Yuan and Y. Lin, “Model selection and estimation in regression with grouped variables,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 68, no. 1, pp. 49–67, 2006.
- [8] K. Lounici, M. Pontil, S. Van De Geer, and A. B. Tsybakov, “Oracle inequalities and optimal inference under group sparsity,” *Annals of statistics*, vol. 39, no. 4, pp. 2164–2204, 2011.

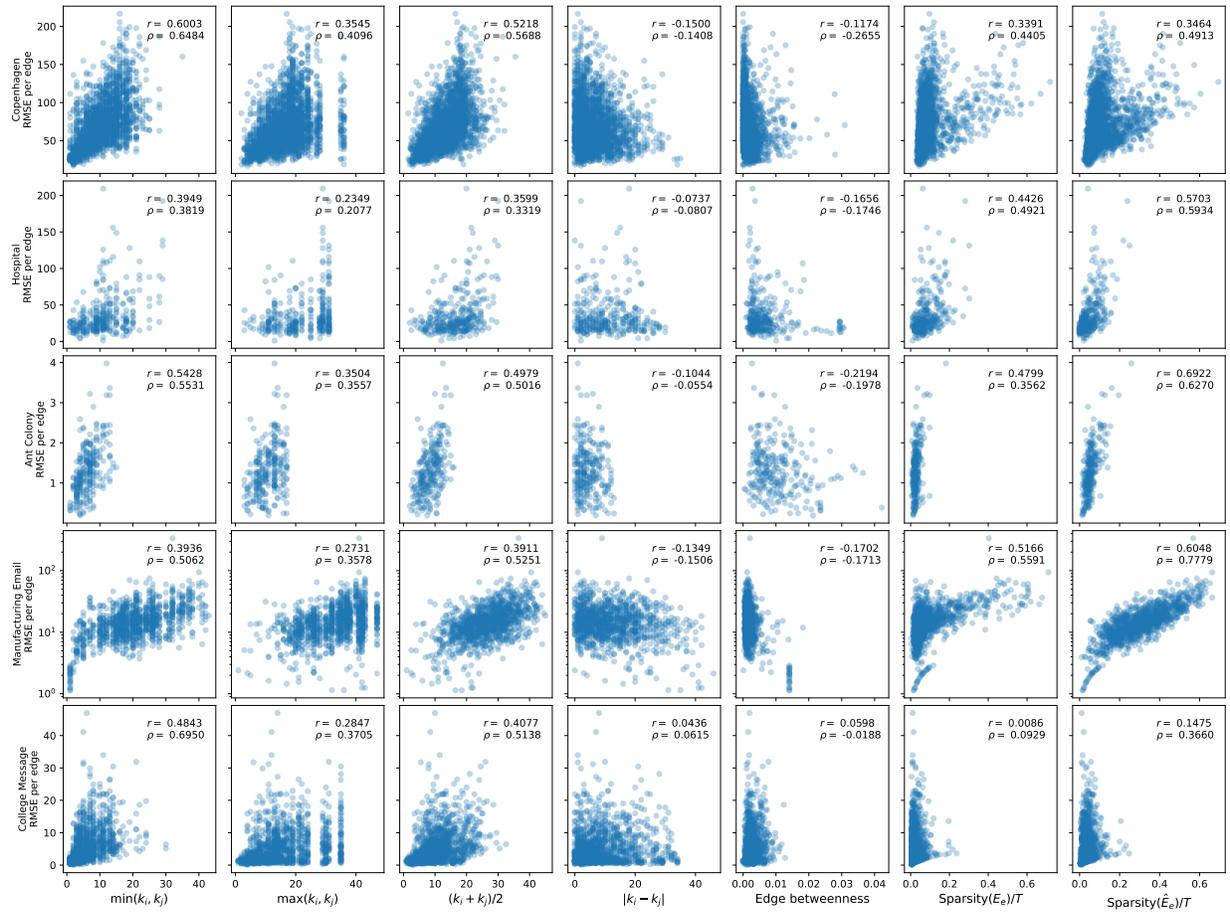


Figure S6: Empirical correlations between recovery error and edge features.

- [9] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [10] H. Akaike, “A new look at the statistical model identification,” *IEEE transactions on automatic control*, vol. 19, no. 6, pp. 716–723, 1974.
- [11] T. Sawa, “Information criteria for discriminating among alternative regression models,” *Econometrica: Journal of the Econometric Society*, pp. 1273–1291, 1978.
- [12] H. Zou, T. Hastie, and R. Tibshirani, “On the “degrees of freedom” of the lasso,” *The Annals of Statistics*, vol. 35, no. 5, pp. 2173–2192, 2007.
- [13] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, “Least angle regression,” *Annals of statistics*, vol. 32, no. 2, pp. 407–499, 2004.
- [14] R. Penrose, “On best approximate solutions of linear matrix equations,” *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 52, no. 1, pp. 17–19, 1956.
- [15] S. Chen and D. Donoho, “Basis pursuit,” in *Proceedings of 1994 28th Asilomar Conference on Signals, Systems and Computers*, vol. 1, pp. 41–44, IEEE, 1994.
- [16] R. Tibshirani, “Regression shrinkage and selection via the lasso,” *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.
- [17] E. Candes and T. Tao, “The Dantzig selector: Statistical estimation when  $p$  is much larger than  $n$ ,” *Annals of statistics*, vol. 35, no. 6, pp. 2313–2351, 2007.
- [18] T. Hastie, R. Tibshirani, and M. Wainwright, *Statistical learning with sparsity: the Lasso and generalizations*. Chapman and Hall/CRC, 2015.
- [19] M. Doob, “An interrelation between line graphs, eigenvalues, and matroids,” *Journal of Combinatorial Theory, Series B*, vol. 15, no. 1, pp. 40–50, 1973.
- [20] K. Knight and W. Fu, “Asymptotics for lasso-type estimators,” *Annals of statistics*, pp. 1356–1378, 2000.
- [21] T. Zhang, “Some sharp performance bounds for least squares regression with  $L_1$  regularization,” *The Annals of Statistics*, vol. 37, no. 5A, pp. 2109–2144, 2009.
- [22] M. Hebiri and S. Van De Geer, “The smooth-lasso and other  $\ell_1 + \ell_2$ -penalized methods,” *Electronic Journal of Statistics*, vol. 5, pp. 1184–1226, 2011.
- [23] Y. Zhang, M. J. Wainwright, and M. I. Jordan, “Optimal prediction for sparse linear models? lower bounds for coordinate-separable M-estimators,” *Electronic Journal of Statistics*, vol. 11, no. 1, pp. 752–799, 2017.
- [24] M. J. Wainwright, *High-dimensional statistics: A non-asymptotic viewpoint*, vol. 48. Cambridge University Press, 2019.