Estimating Unobserved Networks from Heterogeneous Characteristics with an Application to the Swing Riots*

Kieran Marray[1]

[1]School of Business and Economics and Tinbergen Institute, Vrije Universiteit Amsterdam

Current version: July 2025

Abstract

Researchers often observe outcomes determined by economic networks, and characteristics that determine if agents form links, but not the network itself. Here we present an estimator for unobserved networks from panel data and characteristics that drive link formation. The estimator recovers the network by decomposing the covariance matrix of outcomes, penalising links more heavily the less likely they are given characteristics. We provide theoretical bounds on estimation error, and a fast coordinate descent algorithm that makes estimation tractable for large networks. As an application, we estimate patterns of coordinated uprisings during the Swing Riots of 1830–1831 among parishes distributed across space. We find evidence of a small core of coordinated unrest centred on known radical parishes. Exposure to coordinated unrest reduces elite preference for franchise expansion.

Keywords— Networks, graphical lasso, threat of revolution
JEL Codes: C31, D72, D85

1 Introduction

Interactions between economic agents shape many important outcomes. Classmates affect grades in school (Calvó-Armengol et al., 2009; Carrell et al., 2009; Jackson et al., 2022). Idiosyncratic shocks, such as severe weather shocks from climate change, propagate through supply links between firms (Carvalho et al., 2020; Barrot and Sauvagnat, 2016). Information spreads across villages through informal lending networks (Banerjee et al., 2013). But, often these networks are either partially observed, or completely unobserved (e.g see Chandrasekhar and Lewis, 2016; Griffith, 2022; Boucher and Houndetoungan, 2025, for a review). To study these networks, econometricians have begun to develop methods to estimate the unobserved links from panel data on outcomes generated by models of social interactions on the network (Manresa, 2013; Lam and Souza, 2019; Battaglini et al., 2021; Lewbel et al., 2023; De Paula et al., 2024; Griffith and Peng, 2024).

Simultaneously, there is a growing recognition of the importance of how social and economic networks are formed for their structure (e.g see Jackson and Wolinsky, 1996; Bala and Goyal, 2003; Carrell et al., 2013; Jackson et al., 2022; Jackson, 2025). Often individuals choose who to link to. Therefore things like strategic choice of partners, homophily over characteristics, and spatial search and matching frictions determine who is connected with whom in many observed networks (e.g see Currarini et al., 2009; Boucher,

*We are grateful to Ozan Candogan, Thomas de Graaff, Max Kasy, François Lafond, and seminar participants at the Tinbergen Institute for comments. Further thanks to Josha Dekker for advice optimising the code used in simulations. The usual disclaimer applies.

2015; Jackson et al., 2022; Arkolakis et al., 2023). Existing estimators that rely solely on outcomes do not account for these effects of heterogeneous individual characteristics on network structure.

Here, we introduce an estimator for unobserved networks from that uses both the outcomes generated from a process on the network and the characteristics of individuals that affect how they form links between themselves to estimate the unobserved links. As standard, outcomes are assumed to be generated from a canonical social interactions model. Individual outcomes depend on the weighted sum of others' outcomes through the network (Battaglini et al., 2021). But, we account for network formation by adding a hierarchical 'spike-and-slab' prior over possible links (Ročková and George, 2014; Gan et al., 2019). The first stage models individuals forming links at certain rates depending on their pairwise characteristics. As an example, consider a set of individuals in different villages located in space. Some villages are closer to each other; others are further away from each other. The villagers may form links at different rates with those in nearby villages than those in villages further away. The second stage is composed of a mixture of Laplace distributions that serves to regularise estimated links.

To estimate the network, we construct the posterior distribution of links given outcomes and the interaction rates of nodes with different characteristics, and take the maximum a-posteriori estimator. This maximum a-posteriori estimator corresponds to the penalised maximum-likelihood estimator of the canonical social interactions model in Battaglini et al. (2021). But, crucially, the penalisation rate varies by link depending on the characteristics of the individuals and the interaction rates between individuals with those characteristics. Links between types of agents with higher interaction rates are penalised less heavily. Links between types of agents with lower interaction rates are penalised more heavily. In our example, if villagers are more likely to interact with those in nearer-by villages, we would penalise potential links to further away villages more heavily than closer-by villages. We further show that the estimation problem is convex given a set of interaction rates. So, links are point identified given a set of interaction rates.

Of course, researchers do not know these interaction rates ex-ante. So, we show how the researcher can estimate both the network and these interaction rates simultaneously by solving a two-stage optimisation problem (Gan et al., 2019). Given each set of interaction rates, the researcher picks the maximum a-posteriori estimator for the links. To pick the interaction rates, the researcher minimises a Bayesian Information criterion over the set of optimal links (Yuan and Lin, 2007). Under correct choice of the tuning parameters, the estimated network has an optimal error rate of $\mathcal{O}_p\left(\sqrt{\frac{\ln N}{T}}\right)$, where the constant depends on the number of true links in the network. So, in cases where the network is sparse because of the structure of underlying characteristics, the error will be lower. Furthermore, this error rate extends to distributions with polynomial tails.

A benefit of our approach is that it turns the complex problem of sampling from the posterior into a tractable two-stage optimisation problem. The inner optimisation problem – estimating the network given interaction rates – involves maximising a non-linear objective function over $N(N-1)$ parameters. This is costly as N becomes moderately large. So, we derive a novel coordinate descent algorithm to similar to those in Friedman et al. (2007); Ročková and George (2014); Gan et al. (2019) that breaks the nonlinear optimisation problem down into a series of regularised linear regression problems that we can solve by soft-thresholding. Our algorithm allows us to tractably estimate large unobserved networks. In simulations, one coordinate descent cycle over a network with 1000 nodes takes only 2.25 seconds. Furthermore, we find that the estimator delivers relatively high precision and recall rates for links when $T << N$, and estimates interaction rates close to true interaction rates.

We apply our method to estimate underlying coordination networks in the $1830-1831$ Swing Riots in England, using geographical distance between parishes to predict coordination between labourers in the parishes. The Swing Riots were a series of spontaneous uprisings by agricultural labourers who, facing rapidly eroding standards of living during the expansion of capitalism into the English countryside and growing unemployment from automation of traditional tasks, rose up to expropriate lost wages from landlords and break machines that took their work (Hobsbawm and Rudé, 1973; Tilly, 1995; Holland, 2005). The uprisings, in combination with urban unrest and the 'Days of May', were instrumental to the expansion of the franchise in the UK through the Great Reform Act by raising the spectre of an English revolution (Aidt and Franck, 2015).

To separate out the effects networks of coordination from the well-documented slow spatial diffusion of

rioting north-west from Kent (e.g see Charlesworth, 1979; Aidt et al., 2022), we decompose the covariance of incidents of large-scale unrest (machine-breaking, wage riots, expropriation, prisoner rescues, and attacks on law-enforcement) within weeks, after removing common time-invariant factors that contributed to unrest (Caprettini and Voth, 2020). There is anecdotal evidence from the historical literature of organisation of these larger scale incidents by groups of labourers combining strength across parishes (Hobsbawm and Rudé, 1973).

We find sparse evidence of coordination between parishes, This supports the thesis of historians like Hobsbawm and Rudé (1973); Tilly (1995); Holland (2005) that the riots were largely uncoordinated responses to common conditions. But, we do find some local link concentrated in the Berkshire, Wiltshire border and focussed around parishes like Kintbury known as "great centres (sic) of millitancy in 1830" (Hobsbawm and Rudé, 1973). We find significantly more incidents of unrest in parishes that we estimate to be connected to at least one other parish than those that are unconnected. Examaining their characteristics, we see that connected parishes experienced enclosure at a higher rate, and have more exposure to newspapers.

Finally, our results allow us to separate out the effects of exposure to coordinated and uncoordinated rioting on support for the expansion of the franchise through the 1831 general election (widely considered as a referendum on franchise expansion). The 'threat of revolution' theory posits that Western elites agreed expand the franchise to lower orders to head off credible threats of revolutionary uprisings from those lower orders (Acemoglu and Robinson, 2000, 2001; Aidt and Jensen, 2013). Aidt and Franck (2015) examine this in the context of the Swing riots, and find that exposure to rioting increased vote share for politicians that supported expanding the franchise. We replicate their results, now controlling for exposure to coordinated rioting. We find that exposure to general unrest had a larger effect on the vote for politicians supporting franchise expansion than reported in Aidt and Franck (2015). Furthermore, we find a negative (but imprecisely estimated) effect of exposure to coordinated unrest on the vote share of pro-reform candidates conditional on levels of general unrest. This is suggestive evidence of an elite reaction against franchise expansion when exposed to the outcomes of labourers' political coordination.

Our paper contributes to the growing literature on the econometrics of unobserved or partially observed networks. An emerging literature on sampling and mis-specification in network models in econometrics shows the issues an applied researcher can face not observing a full network (e.g see Chandrasekhar and Lewis, 2016; Griffith and Peng, 2024; Boucher and Houndetoungan, 2025). Existing literature approaches the problem of estimating unobserved networks either using characteristics or outcomes generated by the network, and not both. One popular approach involves fitting a model of how economic actors form links based on characteristics to observed network data (e.g the gender of high-school students, or social status within villages). Researchers then imputes unobserved networks, or parts of this network, based on the model (e.g Breza et al., 2020). Taking the other approach, De Paula et al. (2024), Rose (2023), and Lewbel et al. (2023) estimate the matrix of reduced form coefficients in a system of linear equations corresponding to outcomes from the network, and then decompose matrix of reduced form coefficients to obtain an estimate for the network. Manresa (2013), Lam and Souza (2019) take similar approaches, using a lasso to directly select which economic actors influence each other. Our approach is closest to Battaglini et al. (2021) who set up a penalised maximum-likelihood estimator for a network from the covariance matrix of observed outcomes. Griffith and Peng (2024) expand this approach to incorporate the effect of common factors on outcomes. Our main contribution is to incorporate both information on individual characteristics and outcomes on links into the estimation problem in a way that is computationally tractable for realistically-sized networks (similar to Hardy et al. (2024), Lubold et al. (2023)). By using Empirical-Bayes methods to reduce a complex sampling problem to an optimisation problem, we follow the nascent literature in 'optimisation-conscious econometrics' and shrinkage methods in econometrics (Fessler and Kasy, 2019; Hansen, 2016). We further build on the large literature on graphical models in statistics and machine learning (Wainwright and Jordan, 2008; Janková and Van de Geer, 2018; Gan et al., 2019). Similarly to these papers, we estimate the structure of interactions between random variables by decomposing their covariance matrix. Particularly, our work extends that of Gan et al. (2019), who uses introduces a Bayesian estimator of links in Markov random fields that leads to unequal regularisation of links. Our coordinate-descent approach to solving the optimisation problem for large N builds on the graphical lasso of Friedman et al. (2007).

Finally, our paper relates to the literature in economics on history on the Swing riots and the impact of uprisings on franchise expansions. Aidt and Franck (2015) are the first to link exposure to the Swing riots to the passage of the Great Reform act, building on a large literature on the impact of threat of revolution on franchise expansions (e.g see Acemoglu and Robinson, 2000, 2001; Conley and Temimi, 2001; Aidt and Jensen, 2013). Caprettini and Voth (2020) link incidents of rioting to parish-level conditions. Our specification of distance as a predictor of contagion of unrest between parishes builds on Aidt et al. (2022), who study the slower diffusion of incidents of rioting across England over multiple weeks. Within the historical literature, Hobsbawm and Rudé (1973); Charlesworth (1979); Tilly (1995); Holland (2005) conduct detailed surveys of the causes, progress, and impact of the Swing riots.

1.1 Outline

First, in Section 2, we set up our econometric problem. Section 3, derives our estimator, its theoretical properties, and describes the algorithm we use to implement the estimator. Finally, in Section 4 we apply our estimator to the Swing Riots. All proofs are given in the appendix.

2 Setup

Consider a researcher who observes a set $\mathcal{N} = \{1, ...N\}$ of individuals across T time periods. The number of individuals N is fixed such that $N > T$. The individuals are connected on a weighted, simple network with adjacency matrix G. Individual outcomes y_{it} depend on weighted connections G, covariates x_{it}, and idiosyncratic shocks ϵ_{it} through the data generating process (Battaglini et al., 2021)

$$y_{it} = \sum_j G_{ij} y_{jt} + x_{it}\beta + \epsilon_{it} \tag{1}$$

– a common model in the literature on spatial and network econometrics (e.g see De Paula, 2017).

The researcher observes some individual-level outcome $\{y_{it}\}_{i\in\mathcal{N},t=1,...,T}$ and covariates x_{it}, but not how much each individual on the network affects each other, G. The researcher wants to estimate G.

We assume that G is formed in a process that we can describe in three steps. First, nature draws each individual i some position $z_i \in \mathbb{Z}^M$. Without loss of generality, we let these be discrete categories. Second, pairs of individuals form connections at rates that depend on their pairwise characteristics. Let A denote a binary matrix of connections where

$$G_{ij} \neq 0 \text{ only if } A_{ij} = 1. \tag{2}$$

Formally, A is one draw from a random graph generating process $\mathcal{A}(Z, \eta)$ with the vector of interaction rates $\eta \in [0, 1)^{M^2}$

$$\eta = \begin{pmatrix} \eta_{z_1,z_1} & \eta_{z_1,z_2} & ... & \eta_{Z_M,Z_M} \end{pmatrix}$$

such that

$$A_{ij}|(z_i, z_j) \sim \text{Bernoulli}(\eta_{z_i,z_j}). \tag{3}$$

We assume that η is not known to the researcher ex-ante. Third, given that two individuals are connected ($A_{ij} = 1$), they interact with some time-invariant intensity $G_{ij} > 0$.

This reduced-form network formation model is a plausible in many economic settings. For example, we may models individuals in villages forming links at rates that depend on which village each is from (Banerjee et al., 2013). We might model children in schools forming links at different rates depending on whether they have the same or different genders (Currarini et al., 2009). Alternatively, firms from different municipalities may search for suppliers across space at different rates depending on their own and the other municipality (Arkolakis et al., 2023). In the appendix, we show that this can be microfounded as the equilibrium outcome of a game on a network with linear quadratic utilities and individual-specific

consideration sets. If we consider the power set of possible $\mathbb{Z}^M$ as communities, we can think of our network formation model as a non-parametric approximation to a continuous underlying network formation process (Airoldi et al., 2013).[1]

Our specification imposes several noteworthy restrictions on our data. Firstly, links on the network do not change over time (similar to De Paula et al., 2024; Lewbel et al., 2023). This rules out endogenous effects where individuals adjust links endogenously over time to respond to spillovers (e.g see König et al., 2019). Second, all interactions have the same sign. Third, unlike De Paula et al. (2024) we do not allow for a direct effect of covariates on neighbours' outcomes through the network. The effect of one individual's covariates on others' outcomes only comes through the effect on the individual's own outcomes. This assumption is appropriate for our empirical example in section 4. We leave deriving this approach with covariates to further research.

We impose four following assumptions on this data generating process. To start with, we assume that all non-zero entries are 'large enough'.

Assumption 1. Minimum detectable entry size

$$min\{G_{ij}|G_{ij} \neq 0\} > c\sqrt{\frac{\ln N}{T}}$$

for some constant $c > 0$.

This assumption says that as the number of time periods gets smaller, or the number of individuals gets larger, the smallest present link must get stronger. In practice, this tells us the size of the smallest detectable link. As the number of possible links gets larger relative to time periods, it gets harder to distinguish the presence of the link from noise. Second, as is standard in the spatial and network econometric literature, the network must have a spectral radius less than 1 (Battaglini et al., 2021)

Assumption 2. Spectral radius – $\rho(G) < 1$.

Otherwise, outcomes are not well defined. Third, we assume for simplicity that any covariates are deterministic as in Battaglini et al. (2021)

Assumption 3. Deterministic covariates – x_{it} are uniformly-bounded, deterministic variables for all i.

Finally, as standard in the literature on graphical models (e.g see Wainwright and Jordan, 2008; Janková and Van de Geer, 2018; Gan et al., 2019) we assume that errors are independently and identically distributed from a Normal distribution.

Assumption 4. Distribution of errors – $\epsilon_t \sim N(0, \sigma^2 I)\ \forall t$.

This allows us to write down our likelihood function for links given. In practice, we can relax this assumption. Indeed, our theoretical guarantees on estimation accuracy applies to general distributions with exponential and polynomial tails. We discuss this more when we discuss the theoretical properties of the estimator.

For estimation, we will work with the component of y_{it} that varies orthogonally to x_{it}. Stacking each individual's outcomes y_{it} into $T \times 1$ vector y_t, we project out the effect of covariates

$$(I - P_x)y_t.$$

and the resulting outcomes into the $T \times N$ matrix Y. Under assumption 4, we can describe these outcomes as being normally distributed

$$Y \sim N(0, \sigma^2((I - G)(I - G)')^{-1}). \tag{4}$$

[1]Instead, we could let the network formation process be more general by instead parameterising η as a continuous function of underlying characteristics Z. An example of this might be exponential decay of interaction probabilities in distance.

The unobserved networks enters through the covariance of outcomes. Under assumption 2, we can rewrite G in invertible the expansion

$$(I - G)^{-1} = \sum_{k=0}^{\infty} G^k$$

– the matrix encoding all paths of any given length between any two individuals on the network. So the covariance of two individuals' outcomes depends on the variance of shocks and the number of paths of any length between them.[2]

3 Estimation

Consider the problem of jointly estimating links G, the variance of shocks σ^2 and interaction rates η from the posterior distribution of G, σ^2 given outcomes Y, characteristics Z and η. As data, we construct the sample covariance matrix of outcomes across individuals S such that

$$S_{ij} = \frac{1}{T} \sum_{t=1}^{T} Y_{it} Y_{jt}.$$

Our main result is that we can formulate this estimation problem as a two-stage optimisation problem, where the inner stage is convex conditional on an estimate of the interaction rates. This allows us to bound the error of the estimated network from the true network, and to derive an algorithm to tractably estimate moderately-sized networks.

3.1 Objective function

To construct the posterior distribution of links and the variance of shocks given interaction rates and our data, we need to formulate a likelihood for the network and variance given outcomes, and priors for both given characteristics and interaction rates. From the distribution of outcomes 4, it follows that the log-likelihood of outcomes Y given the network of connections G and variance σ^2 is

$$l(G, \sigma^2) = \log \det\Big(\frac{(I-G)(I-G)'}{\sigma^2}\Big) - \operatorname{trace}\Big(S\frac{(I-G)(I-G)'}{\sigma^2}\Big). \tag{5}$$

Given our network generating process, we can model the off-diagonal elements of the adjacency matrix using a 'spike-and-slab' prior (Ročková and George, 2018; Gan et al., 2019).

$$\begin{aligned} \pi(G_{ij}) &= p(A_{ij} = 1 | z_i, z_j) \frac{1}{2\nu_1} e^{\frac{-|G_{ij}|}{\nu_1}} + p(A_{ij} = 0 | z_i, z_j) \frac{1}{2\nu_0} e^{\frac{-|G_{ij}|}{\nu_0}} \\ &= \frac{\eta_{z_i,z_j}}{2\nu_1} e^{\frac{-|G_{ij}|}{\nu_1}} + \frac{1 - \eta_{z_i,z_j}}{2\nu_0} e^{\frac{-|G_{ij}|}{\nu_0}}. \end{aligned} \tag{6}$$

[2] An alternative way of stating this is that our outcomes can be described by a Markov random field parameterised by the precision matrix

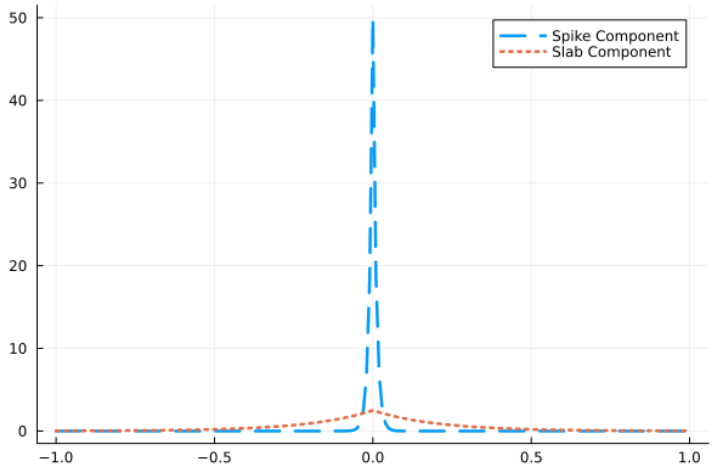
$$\begin{aligned} \Theta &= \frac{1}{\sigma^2}(I-G)(I-G)' \\ &= \frac{1}{\sigma^2}(I - 2G + GG') \end{aligned}$$

that encodes the conditional independence structure between the series of individual level outcomes Y (Janková and Van de Geer, 2018). Two individuals' outcomes are independent of each other conditional on all other outcomes if there are no links of up length two between them.

More intuitively, we can express this as the hierarchical prior

$$\pi(G_{ij}) = \begin{cases} p(G_{ij}|A_{ij}=0) & \sim \text{Laplace}(0,\nu_0) \\ p(G_{ij}|A_{ij}=1) & \sim \text{Laplace}(0,\nu_1) \end{cases}$$
$$A_{ij}|z_i,z_j = \text{Bernoulli}(\eta_{z_i,z_j}).$$

Figure 1: Example 'spike-and-slab' prior



Notes: Example for parameters $v_0 = 1/100, v_1 = 1/5$.

The first level of the prior models the presence of links between individuals at different rates η_{z_i,z_j} depending on individual characteristics z_i, z_j. The second level then models the entries of the adjacency matrix given the presence or not of a link. Here, we use a mixture of Laplace priors – one prior for individuals that are connected, and one for those that are not.[3] Figure 1. plots an example of these distributions. In our case, the penalty will force links that explain little of the covariance in outcomes to zero. In addition to η, which we will estimate, our spike-and-slab prior is parameterised by the scale parameters for the two Laplace distributions ν_0, ν_1. Setting $\nu_1 > \nu_0$ means that there is more likely to be a non-zero entry of the adjacency matrix if we predict that there is a link between individuals than if we do predict that there is a link between individuals. As a prior for σ^2, we use the uniform prior

$$f(\sigma^2) = U(0, \sigma^2_{\max}) \tag{7}$$

where $\sigma^2_{\max}$ is bounded below the observed variance of the data.

Given our priors and likelihood, we can jointly estimate G and η from the posterior distribution by a form of empirical Bayes (Gan et al., 2019; Fessler and Kasy, 2019). First, consider estimation of G, σ^2 for fixed η^*. Using Eqs. 5, 6, and 7, we can write the logarithm of the posterior distribution as

$$\begin{aligned} L(G,\sigma^2|Y,Z,\eta^*) &= \ln \pi(G|Y,Z,\eta^*), \\ &= l(G,\sigma^2) + \sum_{i,j} \ln \pi(G_{ij}|z_i,z_j,\eta^*) + \text{ const}, \\ &= \log\det\Big(\frac{(I-G)(I-G)'}{\sigma^2}\Big) - \text{trace}\Big(S\frac{(I-G)(I-G)'}{\sigma^2}\Big) \\ &+ \sum_{i,j} \ln\Big(\frac{\eta^*_{z_i,z_j}}{2\nu_1} e^{-\frac{|G_{ij}|}{\nu_1}} + \frac{1-\eta^*_{z_i,z_j}}{2\nu_0} e^{-\frac{|G_{ij}|}{\nu_0}}\Big) + \text{ const}, . \end{aligned} \tag{8}$$

[3]Maximum a-posteriori estimation with a Laplace prior is equivalent to the standard l1 penalised maximum likelihood estimator or graphical lasso algorithm (Gan et al., 2019; Friedman et al., 2007). Including some form of penalty is necessary for identification of our subsequent estimator in the case where $N > T$ – see Battaglini et al. (2021).

We take the maximum a-posteriori estimators $\hat{G}(\eta^*), \sigma^2(\eta^*)$. Dropping the constant, the estimates maximise the objective function

$$L(G, \sigma^2 | Y, Z, \eta^*) = \log\det\Big(\frac{(I-G)(I-G)'}{\sigma^2}\Big) - \text{trace}\Big(S\frac{(I-G)(I-G)'}{\sigma^2}\Big) + \sum_{i,j} \ln\Big(\frac{\eta^*_{Z_i,Z_j}}{2\nu_1} e^{-\frac{|G_{ij}|}{\nu_1}} + \frac{1-\eta^*_{Z_i,Z_j}}{2\nu_0} e^{-\frac{|G_{ij}|}{\nu_0}}\Big)$$

This is a natural estimator for G, σ^2 as it corresponds to the usual penalised maximum-likelihood estimator (e.g see Friedman et al., 2007; Battaglini et al., 2021) except, instead of imposing a constant penalty on the occurrence of each link, it more heavily penalises links that less likely to occur based individual characteristics. To see this consider the subgradient of our penalty term above with respect to a link G_{ij}

$$\partial_{G_{ij}} \ln \pi(G_{ij} | z_i, z_j, \eta^*) = p(A_{ij} = 1 | \eta^*_{z_i,z_j}) \frac{1}{\nu_1} + (1 - p(A_{ij} = 1 | \eta^*_{z_i,z_j})) \frac{1}{\nu_0}.$$

The subgradient of the penalisation term is the average of two penalisation terms weighted by the conditional probabilities a link is present or not given characteristics and estimated interaction rates (Gan et al., 2019). Through setting $\nu_1 > \nu_0$, we penalise links more between individuals that are less likely connected based on characteristics (more heavily weighting the 'spike' of the mixture distribution), and less between those that we think are more likely to be present based in characteristics (more heavily weighting the 'slab' of the mixture distribution).

In practice, we need to impose our assumptions on the spectral radius of the adjacency matrix, and that there are no self loops. Adding these constraints alongside Eq. 8 gives the constrained optimisation problem

$$\begin{aligned}\hat{G}(\eta), \hat{\sigma}^2(\eta) = \max_{G_{ij}>0, \sigma^2 \in [0, \sigma^2_{\max}]} & \log\det\Big(\frac{(I-G)(I-G)'}{\sigma^2}\Big) - \text{trace}\Big(S\frac{(I-G)(I-G)'}{\sigma^2}\Big) \\ & + \sum_{i,j} \ln\Big(\frac{\eta^*_{Z_i,Z_j}}{2\nu_1} e^{-\frac{|G_{ij}|}{\nu_1}} + \frac{1-\eta^*_{Z_i,Z_j}}{2\nu_0} e^{-\frac{|G_{ij}|}{\nu_0}}\Big) \\ & \text{s.t } \rho(G) < 1, G_{ii} = 0 \ \forall i\end{aligned} \tag{9}$$

Given the inner problem 9, we can then solve for the interaction rates η given the estimates of $\hat{G}(\eta), \sigma^2(\eta)$. Following Gan et al. (2019), we pick η^* to minimise the Bayesian Information Criterion for high-dimensional covariance matrix estimators of Yuan and Lin (2007)

$$\begin{aligned}\eta^* = \min_{\eta} & - n\Big(\text{trace}\Big(S\frac{(I-\hat{G}(\eta))(I-\hat{G}(\eta))'}{\hat{\sigma}^2(\eta)}\Big) - \log\det\Big(\frac{(I-\hat{G}(\eta))(I-\hat{G}(\eta))'}{\hat{\sigma}^2(\eta)}\Big)\Big) \\ & + \ln(n) \times |\{\hat{G}_{ij}(\eta) : \hat{G}_{ij}(\eta) \neq 0\}|.\end{aligned}$$

3.2 Theoretical properties

First, we note that the inner optimisation problem Eq. 9 is convex.

Proposition 1. Assume that $\nu_1 > \nu_0$ and $\eta_{z_i,z_j} \in [0, 1) \ \forall i, j$. Then, the optimisation problem Eq. 9 is convex with respect to $G_{ij} \ \forall i \neq j$.

This tells us that, for a fixed set of interaction rates η, the adjacency matrix of the network G is identified.

Next, we want to bound the distance of the estimate from the true network. Let $\hat{G}$ denote the solution to the inner optimisation problem. We can show that the estimated network is close to the true network.

Theorem 1. Define $a = ||A||_\infty$, and denote $M_{B^0} = |||B^0|||_\infty$. Make assumptions 1, 2, 3 plus the assumptions that

$$\begin{aligned}\frac{1}{T\nu_0} &> C_4\sqrt{\frac{\ln N}{T}},\\ \frac{1}{T\nu_1} &> C_3\sqrt{\frac{\ln N}{T}}, \text{ and}\\ \sqrt{T} &\geq Ca\sqrt{N}\end{aligned}$$

where $C = 2(2C_1 + C_2)M_{\Gamma_0}\max(3M_{\Sigma_0}, 3M_{\Gamma_0}M_{\Sigma_0}^3)$.

If ϵ is from a distribution with exponential tails i.e there exist some $\gamma, c_1 > 0, N \leq c_1T^\gamma$, and for some $\delta_0 < 0$

$$E|\epsilon_i|^{4\gamma+4+\delta_0} \leq K \ \forall i = 1, ..., N,$$

then

$$||\hat{G} - G||_F \leq 2(2C_1 + C_3)M_{\Gamma_0}\sqrt{\frac{a\ln N}{T}}.$$

where

$$C_1 = a^{-1}(2 + \tau_0 + a^{-1}K^2)$$

with probability greater than $1 - 2N^{-\tau_0}$.

If we instead assume that ϵ is from a distribution with polynomial tails i.e there exist some $\gamma, c_1 > 0, N \leq c_1T^\gamma$, and for some $\delta_0 < 0$

$$E|\epsilon_i|^{4\gamma+4+\delta_0} \leq K \ \forall i = 1, ..., N,$$

then

$$||\hat{G} - G||_F \leq 2(2C_1 + C_3)M_{\Gamma_0}\sqrt{\frac{a\ln N}{T}}.$$

where

$$C_1 = \sqrt{(\Theta_{\max} + 1)(4 + \tau_0)}$$

with probability greater than $1 - O(T^{-\frac{\delta_0}{8}} + N^{-\frac{\tau_0}{2}})$.

Note that the necessary sample size depends on the maximum row sum of A, which itself depends on the characteristics Z. This theorem tells us that if number of possible connections for nodes in the network is sparse due to underlying characteristics – the rate $\eta_{ij} = 0$ for most i, j – then we can detect networks with a very small T relative to N. As the set of possible links gets larger and larger, and interaction rates get higher, we need larger T to detect the true network for a given N.

3.3 Implementation

To implement the estimator, we need to solve the constrained optimisation problem Eq. 9. This type of problem is hard to solve in practice (Ročková and George, 2014; Gan et al., 2019). So, here we derive an EM algorithm to efficiently solve 9 for large N. Recall that we can write the log of the prior (our regularisation term in 9) in terms of the matrix of connections A as

$$\sum_{ij} \ln\Big(p(A_{ij} = 1|z_i, z_j)\frac{1}{2\nu_1}e^{\frac{-|G_{ij}|}{\nu_1}} + p(A_{ij} = 0|z_i, z_j)\frac{1}{2\nu_0}e^{\frac{-|G_{ij}|}{\nu_0}}\Big).$$

To solve Eq. 9, we sample A as a latent variable alongside G, fixing σ^2 (Ročková and George, 2014). Then, we can solve for estimates of G maximising the expected likelihood given A (Gan et al., 2019).

Finally, we can solve for σ^2 by maximising the likelihood over the resulting estimates. First, note that we can rewrite the joint log-posterior distribution Eq. 8 as

$$L(G, \sigma^2, A|Y, Z, \eta) = l(G, \sigma^2) + \sum_{ij} \ln\left(P(A_{ij}|\eta, Z_i, Z_j)p(G_{ij}|A_{ij})\right) + \text{const.}$$

Let P denote the matrix such that

$$\begin{aligned} P_{ij} &= E(A_{ij}) \\ &= p(A_{ij} = 1|z_i, z_j). \end{aligned}$$

In the E step, we can sample P given estimated G^k as in Gan et al. (2019)

$$\ln \frac{P_{ij}}{1 - P_{ij}} = \ln \frac{\nu_0}{\nu_1} + \ln \frac{\eta_{z_i,z_j}}{1 - \eta_{z_i,z_j}} - \frac{|G^k_{ij}|}{\nu_1} + \frac{|G^k_{ij}|}{\nu_0}. \tag{10}$$

Then, we can write the expectation of log-posterior over A as

$$\begin{aligned} Q(G|G^k) = \log\,\det\Big(\frac{(I-G)(I-G)'}{\sigma^2}\Big) - \text{trace}\Big(S\frac{(I-G)(I-G)'}{\sigma^2}\Big) \\ + \sum_{i,j} P_{ij}(-\ln 2\nu_1 - \frac{|G_{ij}|}{\nu_1} + \ln \eta_{Z_i,Z_j}) \\ + \sum_{i,j} (1 - P_{ij})(-\ln 2\nu_0 - \frac{|G_{ij}|}{\nu_0} + \ln(1 - \eta_{Z_i,Z_j})). \end{aligned}$$

Maximising this gives us our estimate G^{k+1}. To solve for $\hat{G}, \hat{A}$, we can iterate until convergence.

Maximising the $Q()$ function over for a complete (simple) network G^k involves simultaneously estimating $N(N-1)$ parameters. As this scales quadratically in N, this becomes difficult to directly optimise for even small N. As a solution, we propose a network lasso algorithm inspired by the graphical lasso of (Friedman et al., 2007). This allows us to break this hard non-linear optimisation problem into a series of lasso problems that we can solve iteratively using a soft-thresholding algorithm. The soft-thresholding algorithm is relatively fast, allowing us to solve relatively large problems with thousands of individuals.

Taking subdifferentials of our $Q()$ function with respect to G gives the first-order condition

$$-\frac{1}{\sigma^2} S(I - G) + (I - G)^{-1} - Z = 0,$$

where the matrix Z encodes the subdifferentials of the penalty function with respect to each G_{ij}

$$Z_{ij} \begin{cases} = \in [-(\frac{1}{\nu_1}P_{ij} + \frac{1}{\nu_0}(1 - P_{ij})), (\frac{1}{\nu_1}P_{ij} + \frac{1}{\nu_0}(1 - P_{ij}))] & \text{if } G_{ij} = 0 \\ = (\frac{1}{\nu_1}P_{ij} + \frac{1}{\nu_0}(1 - P_{ij})) & \text{if } G_{ij} > 0. \end{cases}$$

for $i \neq j$. Rearranging, we have

$$S - \sigma^2(I - \gamma G)^{-1} - SG - \sigma^2 Z = 0.$$

Consider the i, jth entry of this matrix condition. We have

$$S_{ij} - \sigma^2(I - G)^{-1}_{ij} - S_{ii}G_{ij} - \sum_{k \neq j} S_{i,k}G_{k,j} - \sigma^2 Z_{ij} = 0.$$

– which we recognise as a Lasso problem (Tibshirani, 1996). Rearranging for G_{ij} gives

$$G_{ij} = \frac{1}{S_{ii}}(S_{ij} - \sum_{k \neq j} S_{i,k}G_{k,j} - \sigma^2(I - G)^{-1}_{ij} - \sigma^2 Z_{ij}).$$

Therefore we can solve for G_{ij} column-wise using an iterative soft-thresholding algorithm (Friedman et al., 2010). Denote

$$\rho_{ij} = S_{ij} - \sum_{k \neq j} S_{i,k} G_{k,j} - \sigma^2 (I - G)^{-1}_{ij}.$$

The updating rule for each G_{ij} is

$$G_{ij} = \begin{cases} \frac{\rho_{ij} - \sigma^2 Z_{ij}}{S_{ii}} & \text{if } \rho_{ij} > Z_{ij}, \\ 0 & \text{if } \rho_{ij} \leq Z_{ij}. \end{cases}$$

An additional complication is updating the inverse $(I - G)^{-1}$ after estimating each column, which would be computationally expensive. However, we can use that for a rank-one update to a matrix that we can write as the outer product of two vectors uv', the inverse is computable as (Sherman and Morrison, 1949)

$$((I - G) + uv')^{-1} = (I - G)^{-1} - \frac{(I - G)^{-1} uv' (I - G)^{-1}}{1 + v'(I - G)^{-1} u}.$$

Writing

$$\begin{aligned} u &= -(G_{12}^k - G_{12}^{k-1}) \\ v &= e_2 \text{ the unit basis vector,} \end{aligned}$$

we can write each new column estimate as a rank-one update to $I - G$, uv', and apply this rule within our algorithm.

The full algorithm we use to solve the inner optimisation problem is given in Alg. **??**. To test the speed of our algorithm, we time 100 single loops with a moderately sized network of $N = 1000$ (optimising over $999,000$ variables). The mean execution time is 2.55 seconds.

Using the results in Tseng (2001), we can show that the result of this algorithm converges to the true maximiser of the inner optimisation problem

Proposition 2. Make assumptions 1 and 2. Let $\hat{G}^k$ denote the k-th estimate from algorithm **??** and G^* denote the maximiser of the inner optimisation problem Eq. 9 fixing η, σ^2. As $k \to \infty$, $\hat{G}^k \to G^*$.

To solve for σ^2, we simply maximise the likelihood over the results of this algorithm. Having solved the inner optimisation problem fixing η, we have to determine the interaction rates η plus the regularisation parameters ν_1, ν_0. To estimate η, we minimise the BIC over the maximum a-posteriori estimates of G given η, ν_0, ν_1. Unfortunately, we have no theoretical guarantees that our outer problem is convex with a unique global minimum.[4]. Therefore, we recommend estimating η, ν_0, ν_1 by a form of global optimisation enforcing that $\nu_1 > 1$. In practice, we apply a local grid search strategy to sweep the parameter space. This works well in practice for low-dimensional η. Further research could explore more efficient strategies when the dimension of η becomes larger.

3.4 Simulations

In addition to our theoretical guarantees, we assess the finite sample performance of our estimates in simulation. As our unobserved networks, we simulate the interaction of individuals across villages. Nodes are partitioned into M blocks $Z_1, ..., Z_M$ (villages). The latent network of interations A is a drawn randomly such that nodes form links at a certain rate within blocks, and a certain rate between blocks.

$$\begin{aligned} A_{ij} | Z_i = Z_j &\sim \text{Bernoulli}(\eta_1), \\ A_{ij} | Z_i \neq Z_j &\sim \text{Bernoulli}(\eta_2). \end{aligned}$$

Given a sampled A, we set interaction intensities by row-normalising the matrix of connections

[4]Indeed, in simulations we can generate cases where there are multiple local minima.

Algorithm 1 Network lasso

procedure BLOCK COORDINATE DESCENT($S, G_0, \sigma^2, \nu_0, \nu_1, \delta$)
  $G \leftarrow G_0$
  $\Lambda = (I - G_0)^{-1}$
  while $||G^k - G^{k-1}||_F \geq \delta$ **do**
    $G^k \leftarrow G^{k-1}$.
    for $i, j \in 1, ..., N$ **do**
      $P_{ij} = \left(1 + \frac{\nu_1}{\nu_0} e^{-\frac{|G^k_{ij}|}{\nu_0} + \frac{|G^k_{ij}|}{\nu_1}} \frac{1-\eta_{Z_i,Z_j}}{\eta_{Z_i,Z_j}}\right)^{-1}$
    end for
    for $i \in 1, ..., N$ **do**
      $v = e_i$
      for $j \in 1, ..., N$ **do**
        $Z_{ij} = P_{ij}\frac{1}{\nu_1} + (1 - P_{ij})\frac{1}{\nu_0}$
        $\rho_{ij} = S_{ij} - \sum_{;\neq j} S_{i,l} G^k_{l,j} - \sigma^2 \Lambda_{ij}$.
        $G^k_{ij} = \begin{cases} 0 & \text{if } \rho_{ij} \leq \sigma^2 Z_{ij}, \\ \frac{\rho_{ij} - \sigma^2 Z_{ij}}{S_{ii}} & \text{if } \rho_{ij} > \sigma^2 Z_{ij}. \end{cases}$
      end for
      $u = -(G^k_{:,j} - G^{k-1}_{:,j})$
      $\Lambda = \Lambda - \frac{\Lambda u v' \Lambda}{1 + v' \Lambda u}$.
    end for
    $i \leftarrow i + 1$
  end while
  Return G^k.
end procedure

Algorithm 2 Full estimation algorithm

procedure GRID SEARCH($\mathcal{H}, \Sigma, G_0, S, \nu_0, \nu_1, \delta$)
  for $\eta \in \mathcal{H}$ **do**
    $\hat{G}(\eta), \hat{\sigma}^2(\eta) = \text{argmax}_{\sigma^2 \in \Sigma}\{L(\text{Network Lasso}(\eta, G_0, S, \delta), \sigma^2)\}$

$$\text{BIC}(\eta) = -n\Big(\text{trace}\Big(S\frac{(I - G^*(\eta))(I - G^*(\eta))'}{1}\Big) + \log\det\Big(\frac{(I - G^*(\eta))(I - G^*(\eta))'}{1}\Big)\Big) + \ln(n) \times |\{G^*_{ij}(\eta) : G^*_{ij}(\eta) \neq 0\}|$$

  end for
  $\hat{\eta} = \text{argmin}_{\eta \in \mathcal{H}}\{\text{BIC}(\eta)\}$
  $\hat{G}, \hat{\sigma}^2 = \hat{G}(\hat{\eta}), \hat{\sigma}^2(\hat{\eta})$.
end procedure

$$G_{ij} = \frac{A_{ij}}{\sum_j A_{ij}}$$

to generate the network of interaction intensities G. Given a sampled G, we then simulate outcomes from the model

$$Y_{it} = (I - G)^{-1}\epsilon_{it}$$

for $t = \{1, ..., T\}$ with $\epsilon_{it} \sim$ i.i.d $N(0, 1)$ for ease.[5] We set $T < N$ in each case, and vary the ratio $\frac{N}{T} \in \{2, 3, 6, 12\}$. To estimate η, we use a local grid search.

We consider four different measures of the quality of our estimates

1. Precision – the total number of true links detected divided by the total number of estimated links on the network
$$\text{Precision} = \frac{\sum_{ij} \mathbf{1}(\hat{G}_{ij} > 0)\mathbf{1}(G_{ij} > 0)}{\sum_{ij} \mathbf{1}(\hat{G}_{ij} > 0)}$$
2. Recall – the total number of true links detected divided by the total number of true links on the network
$$\text{Recall} = \frac{\sum_{ij} \mathbf{1}(\hat{G}_{ij} > 0)\mathbf{1}(G_{ij} > 0)}{\sum_{ij} \mathbf{1}(G_{ij} > 0)}$$
3. The distance between the estimated network and true network
$$||\hat{G} - G||_F.$$
4. The estimates $(\hat{\eta}_1, \hat{\eta}_2)$.

Outcomes are computed by averaging over 100 draws of $G, \{\epsilon_{it}\}_{i,t}$.

Table 1: Performance of estimator on simulated networks

N	T	Precision	Recall	Distance	$\hat{\eta_1}$	$\hat{\eta_2}$
300	150	0.75	0.90	7.07	0.28	0.05
300	100	0.71	0.850	7.40	0.32	0.05
300	50	0.63	0.76	8.02	0.28	0.041
300	25	0.45	0.63	8.93	0.28	0.065

Notes: Results averaged over 100 simulated networks and errors. 'Distance 'is Frobenius norm of the difference between the true and estimated network. The number of blocks is kept constant at $\frac{N}{10}$.

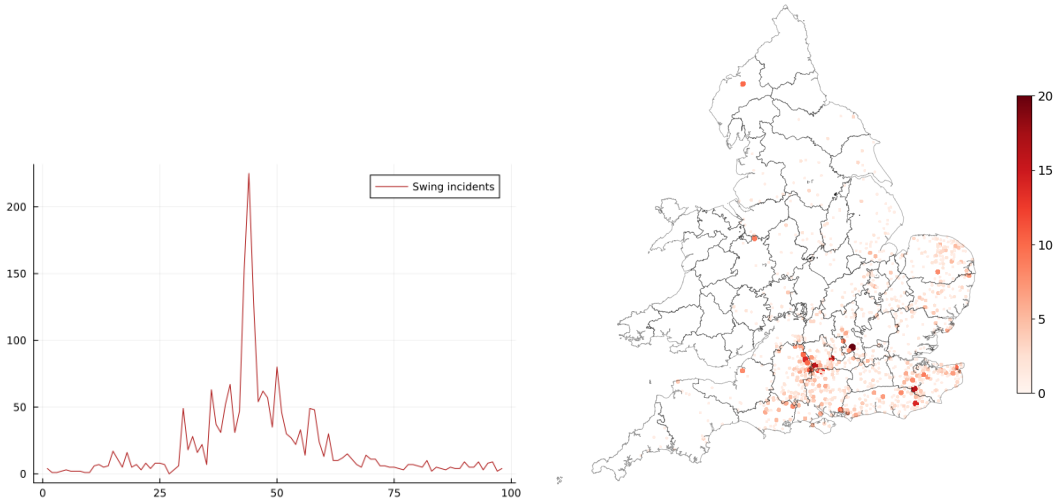
Precision and recall are high when T is relatively high compared to N. As T falls, performance falls as well. However with T very small relative to N, recall remains relatively high. In all cases, distance between estimated and true matrices are fairly low. Estimates of interaction rates also remain reasonably close to true interaction rates.

4 Application: coordinated unrest in the Swing riots

As an application, we estimate patterns of coordination between labourers in different parishes in England during the Swing riots of 1830-1831. We then study the effects of elite exposure to coordinated and uncoordinated uprisings on votes to expand the electoral franchise in the United Kingdom.

[5] We get results that are quantitatively similar when jointly estimating σ^2.

Figure 2: Swing incidents in England 1830–31



(a) Incidents by day, June 1830–February 1831

(b) Distribution of incidents by parish – size/colour denotes number of riots

4.1 The Swing riots of 1830–1831

The Swing Riots were series of uprisings by landless labourers that spread mainly through the agricultural south-east of England between August 1830 and February 1831. They were the largest of any such risings in the 19th century England (Tilly, 1995). The English countryside at this times was mainly comprised of middle-sized and larger farms. A small fraction of landowners owned these farms, and employed a large mass of landless agricultural labourers in seasonal work. Changes in the economic system in England over the late 18th and early 19th centuries had eroded the traditional life and bonds of these country workers. Cottage gardens and traditional common lands were enclosed, removing the little land the labourers possessed. Employment contracts became more precarious. Local systems of poor relief, the last resort of those unable to get sufficient work to feed their families, began to break down (Hobsbawm and Rudé, 1973).

The spark for unrest is generally seen to be the increasing adoption of machines to replace seasonal labourers, combined with bad harvests in preceding years Hobsbawm and Rudé (1973); Holland (2005). Facing immiseration, desperate agricultural labourers began to rise up to demand liveable wages and employment. The risings started in Kent on June 28th 1830. Barns were burnt, threshing machines smashed, and labourers expropriated earnings that they had lost or felt entitled to from farm operators and landowners. Acts were often preceded by letters threatening action signed by a mysterious "Captain Swing", either named for the leader of the harvest gangs or the flails traditionally used in threshing (Holland, 2005).

Authorities and local militias were in general unable to prevent incidents or effectively suppress the labourers until after serious unrest (Hobsbawm and Rudé, 1973; Holland, 2005). Uprisings scared elites, raising the spectre of an English revolution similar to the contemporaneous revolutions France and Belgium. There is evidence that the riots led to a subsequent expansion of the franchise due to a fear of revolution (Aidt and Franck, 2015).

News of uprisings diffused across through the country through roads and local communication networks. This occurred relatively slowly, starting in Kent and spreading North and East (see Charlesworth, 1979; Aidt et al., 2022). Once rioting had begun in an area, in most cases it took the form of sporadic incendarism and machine breaking. But, in certain areas, there s evidence that local labourers coordinated actions across parishes to jointly agitate to improve their conditions. Labourers would organise wage-meetings where they would demand higher wages, and would unite and tour local landowners and farms to expropriate them. For example Hobsbawm and Rudé (1973) report that that, after the men of Hungerford and Kintbury in Berkshire were able to extract wage concessions from local justices of the peace

"Other villages, whose inhabitants had no doubt heard of the outcome of the Hungerford confrontation, sent a deputation that night to join the Kintbury labourers to invite them to join in combined operations. So the next day the riots continued ..." (Hobsbawm and Rudé, 1973, pg. 138)

Here, we focus on these links between parishes leading to coordinated uprisings as opposed to this longer process of spatial diffusion (already well documented by Charlesworth, 1979; Aidt et al., 2022). We do this by looking at the covariance in Swing incidents over the short term (within the week) as opposed to medium or longer term (over weeks, as in Aidt et al., 2022).

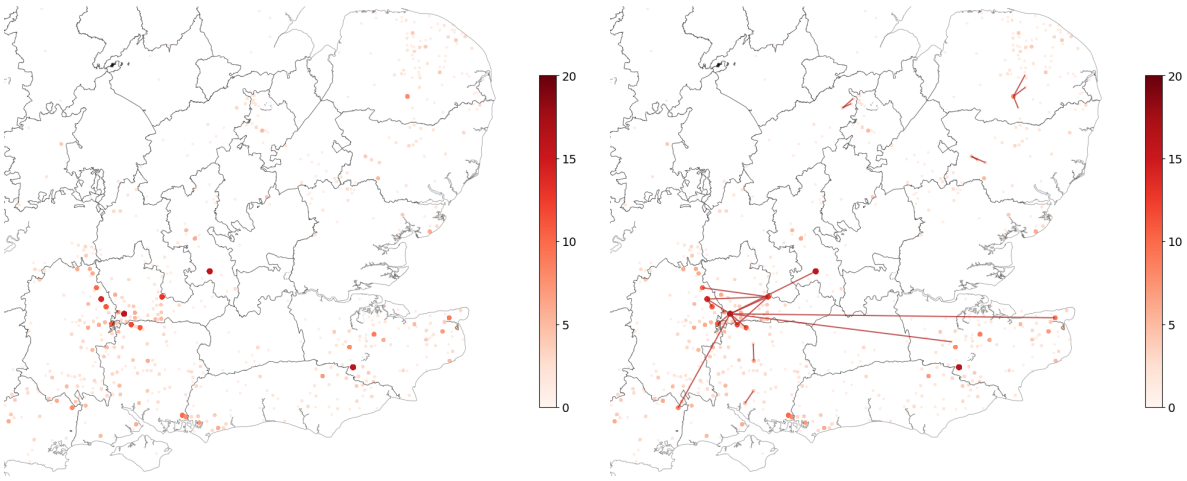
4.2 Data

Data on incidents associated with the Swing riots comes from Holland (2005). Our dataset contains the number of incidents associated with the Swing riots in each of 8597 parishes in England over 40 weeks (recorded on 98 individual days) between 28th of June 1830 and 3rd of April 1831, plus the latitude and longitude of each parish computed from historic maps by Aidt et al. (2022). We distinguish between incidents involving large mobs of labourers (wage riots, 'robbery', machine-breaking, prisoner rescues, and attacks on local authority figures) and other types of incidents (mainly incendarism, plus things like livestock maiming) (Tilly, 1995; Aidt et al., 2022). Only a small proportion (619) of these parishes experienced at least one large incident during the unrest.

Distances between each parish come from historic latitudes and longitudes by the Haversine formula.

To estimate coordination between parishes, we construct the covariance of 'large' incidents of unrest within each calendar week across parishes that experience at least one incident of 'large' unrest. Only a small proportion (619) of these parishes experienced at least one large incident during the unrest. As covariates that determine how labourers make connections between parishes, we use we use whether the parishes are within $10km$ of each other as in (Aidt et al., 2022).[6] The motivation for this specification documented incidents of radical cores of labourers travelling by foot between multiple parishes to gather others to press demands for wages on local elites or break machines documented in Hobsbawm and Rudé (1973); Holland (2005). This type of unrest is different from the sporadic incendarism carried out by small groups of anonymous labourers. Apparent coordination may come from spatial correlation of local conditions that are shown to influence the likelihood of uprisings in parishes (Caprettini and Voth, 2020). So, we demean incidents by parish before estimating the network. After experimentation, we use tuning parameters $\nu_0 = \frac{1}{30}, \nu_1 = \frac{1}{0.1}$.

Figure 3: Swing incidents in England 1830–31



(a) 'Large' incidents, Southeast of England (b) Estimated coordination patterns

[6]Distances between each parish come from the historic latitudes and longitudes reported in Aidt et al. (2022) by the Haversine formula.

4.3 Results

Figure 2 plots our estimated network. The estimated network is very sparse. We only detect 24 non-zero links between 25 distinct parishes. Estimated interaction rates are very small – at $\eta = 0.1$ for labourers within 10km – and the estimated variance of shocks is $\hat{\sigma}^2 = 0.09$ (compared to variance of outcomes of 0.189). This evidence suggests that coordination between parishes did not play a large role in the overall pattern of rioting across England. This result is in line with the historical literature, which emphasises a lack of cross-country coordination between labourers compared to later trade-union movements in England Hobsbawm and Rudé (1973).

The pattern of coordination we find differs across the counties affected by the Swing Riots. We only find evidence of local diffusion within 8 out of the 39 historic counties: Berkshire, Buckinghamshire, Dorset, Hampshire, Kent, Northamptonshire, Suffolk, and Wiltshire. But, these include five out of the six counties with the most intense Swing activity – Kent (441 incidents), Norfolk (301 incidents), Wiltshire (259 incidents), Hampshire (255 incidents), and Berkshire (169 incidents).[7] Where labourers appear to coordinate, we see that parishes with many large incidents are central within local interaction networks. The most connected parish is Kintbury in Berkshire, described by Hobsbawm and Rudé (1973), who detail of strong coordinated actions within Berkshire focussed on the men of Kintbury, as "a great centre of militancy in 1830".

Matching back to characteristics of parishes detailed in Caprettini and Voth (2020); Aidt et al. (2022) allows us to compute some descriptive evidence about these connected parishes compared to other parishes that experienced at least one Swing incident. Results are given in Tables 2 and 3.

Table 2: Difference in Swing incidents by estimated coordination

	Number of incidents
Coordinated	5.31
	(1.06)
Intercept	Yes
Observations	969
R^2	0.0222

Notes: Standard errors in parentheses are heteroskedasticity-robust standard errors computed using the HC3 estimator of MacKinnon and White (1985).

We find a strong positive relationship between the number of (both 'large' and 'small') incidents and coordination. Looking at characteristics that might influence coordination, we find that those where we estimate labourers coordinated with others are more likely to have had common land enclosed and have greater access to newspapers than parishes that experienced an uprising but did not coordinate action. Interestingly, we find no difference in the adoption of threshing machines, or access to markets, suggesting some difference in the causes of general levels unrest, spatial diffusion of information of unrest, and coordination (Caprettini and Voth, 2020; Aidt et al., 2022).

4.4 Local networks and threat of revolution

Finally, we use our results to separate out the effect on local elites of being exposed to coordinated or uncoordinated uprisings. To test if there is a different effect, we replicate the analysis in Aidt and Franck (2015) looking at the effect of the number of nearby Swing incidents on the vote share for the Whigs in the 1831 election. But, we also control for a dummy variable that encodes whether we find a parish that coordinated with others within at least $10km$ of the centroid of a constituency. Those 13 constituencies are not just exposed to general unrest, but to coordinated unrest. To address any

[7] The missing county is Sussex with (215 incidents).

Table 3: Coordination by parish-level covariates

	Coordinate	
Enclosure	0.0212	0.0624
	(0.0118)	(0.0347)
Market within 10km	0.0011	−0.0258
	(0.0157)	(0.0298)
Coach stop within 10km	0.0122	0.0437
	(0.0109)	(0.0225)
Newspaper within 10km	0.0319	0.0974
	(0.0214)	(0.0760)
Distance to nearest newspaper	0.0003	0.0052
	(0.0006)	(0.00306)
Threshing machine adoption	0.0092	−0.0296
	(0.0223)	(0.0195)
Demographic controls	No	Yes
Intercept	Yes	Yes
Observations	969	969
R^2	0.0107	0.107

Notes: Standard errors in parentheses are heteroskedasticity-robust standard errors computed using the HC3 estimator of MacKinnon and White (1985).

concerns about endogeneity of the number of nearby Swing incidents, we replicate both the regular and instrumental variables estimators, using their instrument of distance to the starting point of the first riot in Kent.

Table 4 compares the baseline and preferred specifications in Aidt and Franck (2015) to ones also accounting for the effect of coordination. When we account for the exposure of elites to coordinated action, we find a larger effect of exposure to rioting on vote share for the Whigs (0.596 per riot as opposed to 0.439 in the main ordinary least-squares specification, 3.19 per riot as opposed to 2.53 per riot in the instrumental variables specification). The change is explained by a large (imprecisely estimated in the ordinary least-squares specifications) negative effect of being exposed to coordinated rioting on vote share for the Whigs. In the preferred specification of Aidt and Franck (2015), the vote share for the Whigs is on average 16.2% lower in constituencies exposed to coordinated rioting compared to those constituencies with an equivalent amount of uncoordinated rioting. When instrumenting, this effect gets much larger and more precisely estimated. Alongside the direct effect of general revolution, this is suggestive evidence of a negative reaction of elites to experiencing political action and demands of labourers first hand.

Table 4: Exposure to coordinated uprisings and 1831 election results

	Whig share 1831 (%)					
	OLS estimates				IV estimates	
	(1)	(2)	(3)	(4)	(5)	(6)
Riots within 10km	0.569	0.837	0.439	0.596	2.53	3.19
	(0.268)	(0.219)	(0.192)	(0.185)	(0.825)	(1.03)
Coordination within 10km		−27.8		−16.2		−82.8
		(15.4)		(12.5)		(29.1)
Whig share 1826			0.380	0.380	0.143	0.196
			(0.078)	(0.077)	(0.223)	(0.218)
Reform support 1830			12.6	12.1	13.5	10.6
			(5.36)	(5.22)	(6.15)	(6.07)
County constituency			31.6	31.7	40.2	38.2
			(4.97)	(5.04)	(8.80)	(8.72)
University constituency			−61.8	−63.2	−56.3	−52.1
			(20.3)	(21.2)	(13.1)	(13.1)
Patronage index			−15.3	−15.2	−9.06	−8.04
			(3.67)	(3.67)	(4.64)	(4.81)
Declining economy			−10.3	−9.60	−12.8	−16.3
			(6.15)	(6.23)	(6.76)	(6.94)
Spatial controls	No	No	No	No	Yes	Yes
Intercept	Yes	Yes	Yes	Yes	Yes	Yes
Observations	244	244	244	244	244	244
R^2	0.0253	0.0393	0.467	0.472	0.491	0.492

Notes: Controls included to match preferred specifications in Aidt and Franck (2015). In the final two columns, the number of riots within 10km is instrumented by distance to Sevenoaks. For details, see Aidt and Franck (2015). Standard errors in parentheses are heteroskedasticity-robust standard errors computed using the HC3 estimator of MacKinnon and White (1985).

5 Conclusion

Here, we introduce an estimator for unobserved networks from panel data in a canonical social interactions model that also includes the effect of covariates on network formation. Covariates lead to unequal penalisation across links, leading to more links where covariates suggest individuals are more likely to interact and fewer links where covariates suggest individuals are less likely to interact. We apply it to estimate patterns of coordination in the Swing riots of 1830–1831. We evidence of small patterns of local coordination, and suggestive evidence that this affected support for the subsequent expansion of the franchise in the United Kingdom. Our specification excludes direct effects of covariates on others' outcomes through the network, and changes in networks over time (as in De Paula et al., 2024). Further econometric work could extend the estimator to this type of setting.

References

Abaluck, J. and Adams-Prassl, A. (2021). What do Consumers Consider Before They Choose? Identification from Asymmetric Demand Responses*. *The Quarterly Journal of Economics*, 136(3):1611–1663.

Acemoglu, D. and Robinson, J. A. (2000). Why did the west extend the franchose? democracy, inequality, and growth in historical perspective. *Quarterly Journal of Economics*, 115(4):1167–1199.

Acemoglu, D. and Robinson, J. A. (2001). A theory of political transitions. *American Economic Review*, 91(4):938–963.

Aidt, T., Leon-Ablan, G., and Satchell, M. (2022). The social dynamics of collective action: Evidence from the diffusion of the swing riots, 1830–1831. *The Journal of Politics*, 84(1):209–225.

Aidt, T. S. and Franck, R. (2015). Democratization under the threat of revolution: Evidence from the great reform act of 1832. *Econometrica*, 83(2):505–547.

Aidt, T. S. and Jensen, P. S. (2013). Workers of the world unite! franchise extensions and the threat of revolution in europe 1820–1938. *European Economic Review*, 72:52–75.

Airoldi, E. M., Costa, T. B., and Chan, S. B. (2013). Stochastic blockmodel approximation of a graphon: Theory and consistent estimation. *Advances in Neural Information Processing Systems*, 26:1–9.

Arkolakis, C., Huneeus, F., and Miyauchi, Y. (2023). Spatial production networks. Working Paper 30954, National Bureau of Economic Research.

Bala, V. and Goyal, S. (2003). A noncooperative model of network formation. *Econometrica*, 68:1181–1229.

Banerjee, A., Chandrasekhar, A., Duflo, E., and Jackson, M. (2013). The Diffusion of Microfinance. *Science*, 341(1236498):363–341.

Barrot, J.-N. and Sauvagnat, J. (2016). Input Specificity and the Propagation of Idiosyncratic Shocks in Production Networks. *The Quarterly Journal of Economics*, 131(3):1543–1592.

Battaglini, M., Crawford, F., Patacchini, E., and Peng, S. (2021). A graphical lasso approach to estimating network connections: the case of us lawmakers. *Mimeo*.

Boucher, V. (2015). Structural homophily. *International Economic Review*, 56(1):235–264.

Boucher, V. and Houndetoungan, E. A. (2025). Estimating peer effects using partial network data. *Mimeo*.

Breza, E., Chandrasekhar, A. G., McCormick, T. H., and Pan, M. (2020). Using aggregated relational data to feasibly identify network structure without network data. *American Economic Review*, 110(8):2454–84.

Cai, T., Liu, W., and Luo, X. (2011). A constrained l1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association*, 106(494):594–607.

Calvó-Armengol, A., Patacchini, E., and Zenou, Y. (2009). Peer effects and social networks in education. *The Review of Economic Studies*, 76(4):1239–1267.

Caprettini, B. and Voth, H.-J. (2020). Rage against the machines: Labor-saving technology and unrest in industrializing england. *American Economic Review: Insights*, 2(3):305–20.

Carrell, S. E., Fullerton, R. L., and West, J. E. (2009). Does your cohort matter? measuring peer effects in college achievement. *Journal of Labor Economics*, 27(3):439–464.

Carrell, S. E., Sacerdote, B. I., and West, J. E. (2013). From natural variation to optimal policy? the importance of endogenous peer group formation. *Econometrica*, 81(3):855–882.

Carvalho, V. M., Nirei, M., Saito, Y. U., and Tahbaz-Salehi, A. (2020). Supply Chain Disruptions: Evidence from the Great East Japan Earthquake*. *The Quarterly Journal of Economics*, 136(2):1255–1321.

Chandrasekhar, A. and Lewis, R. (2016). Econometrics of sampled networks. *Mimeo*.

Charlesworth, A. (1979). *Social Protest in Rural Society*. Norwich University Press.

Conley, J. P. and Temimi, A. (2001). Endogenous enfranchisement when groups' preferences conflict. *Journal of Political Economy*, 109(1):79–102.

Currarini, S., Jackson, M. O., and Pin, P. (2009). An economic model of friendship: Homophily, minorities, and segregation. *Econometrica*, 77(4):1003–1045.

De Paula, A. (2017). Econometrics of network models. In B. Honore, A. Pakes, M. P. and Samuelson, L., editors, *Advances in Economics and Econometrics: Theory and Applications: Eleventh World Congress (Econometric Society Monographs,*, Econometric Society Monographs, pages 268–323. Cambridge University Press, Cambridge.

De Paula, A., Rasul, I., and Souza, P. (2024). Identifying network ties from panel data. *Review of Economic Studies*, 00:1–39.

Fessler, P. and Kasy, M. (2019). How to use economic theory to improve estimators: Shrinking toward theoretical restrictions. *The Review of Economics and Statistics*, 101(4):681–698.

Friedman, J., Hastie, T., and Tibshirani, R. (2007). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441.

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, 33:1–22.

Gan, L., Narisetty, N. N., and Liang, F. (2019). Bayesian Regularisation for Graphical Models with Unequal Shrinkage. *Journal of the American Statistical Association*, 114(527):1218–1231.

Griffith, A. (2022). Name your friends, but only five? the importance of censoring in peer effects estimates using social network data. *Journal of Labour Economics*, 40(4):779–805.

Griffith, A. and Peng, S. (2024). Identification of network structure in the presence of latent, unobserved factors: a new result using turán's theorem. *Mimeo*.

Hansen, B. E. (2016). Efficient shrinkage in parametric models. *Journal of Econometrics*, 190:115–132.

Hardy, M., Heath, R., Lee, W., and McCormick, T. (2024). Estimating spillovers using imprecisely measured networks. *Arxiv*, `https://arxiv.org/pdf/1904.00136.pdf`.

Hobsbawm, E. J. and Rudé, G. (1973). *Captain Swing*. Penguin Books.

Holland, M. (2005). *Swing Unmasked: The Agricultural Riots of 1830 to 1832 and Their Wider Implications*. Milton Keynes: FACHRS Publications.

Jackson, M. and Wolinsky, A. (1996). A strategic model of social and economic networks. *Journal of Economic Theory*, 71:44–74.

Jackson, M. O. (2025). Inequality's economic and social roots: the role of social networks and homophily.

Jackson, M. O., Nei, S. M., Snowberg, E., and Yariv, L. (2022). The dynamics of networks and homophily. Working Paper 30815, National Bureau of Economic Research.

Janková, J. and Van de Geer, S. (2018). Inference in high-dimensional graphical models. In Maathuis, M., Drton, M., Lauritzen, S., and Wainwright, M., editors, *Handbook of Graphical Models*, chapter 14, pages 325–351. CRC Press, 1 edition.

König, M. D., Liu, X., and Zenou, Y. (2019). R&D networks: Theory, empirics and policy implications. *Review of Economics and Statistics*, 101(3):476–491.

Lam, C. and Souza, P. (2019). Estimation and selection of spatial weight matrix in a spatial lag model. *Journal of Business and Economic Statistics*, 38(3):693–710.

Lewbel, A., Qu, X., and Tang, X. (2023). Social networks with unobserved links. *Journal of Political Economy*, 131(4):898–946.

Lubold, S., Chandrasekhar, A. G., and McCormick, T. H. (2023). Identifying the latent space geometry of network models through analysis of curvature. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 85(2):240–292.

MacKinnon, J. G. and White, H. (1985). Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics*, 29(3):305–325.

Manresa, E. (2013). Estimating the structure of social interactions using panel data. *Mimeo*.

Ravikumar, P., Wainwright, M. J., Raskutti, G., and Yu, B. (2011). High-dimensional covariance estimation by minimizing l1-penalized log-determinant divergence. *Electronic Journal of Statistics*, 5:935–980.

Rose, C. (2023). Identification of spillover effects using panel data. *Mimeo*.

Ročková, V. and George, E. (2018). The spike-and-slab lasso. *Journal of the American Statistical Association, Theory and Methods*, 113:431–444.

Ročková, V. and George, E. I. (2014). Emvs: The em approach to bayesian variable selection. *Journal of the American Statistical Association*, 109(506):828–846.

Sherman, J. and Morrison, W. (1949). Adjustment of an inverse matrix corresponding to a change in one element of the given matrix. *Annals of Mathematical Statistics*, 20(4):620–624.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288.

Tilly, C. (1995). *Contention in Great Britain, 1758–1834.* Cambridge: Harvard University Press.

Tseng, P. (2001). Convergence of a block coordinate descent method for nondifferentiable minimization. *Journal of Optimization Theory and Applications*, 109.

Wainwright, M. J. and Jordan, M. I. (2008). *Graphical models, exponential families, and variational inference*. Now Publishers Inc.

Yuan, M. and Lin, Y. (2007). Model selection and estimation in the gaussian graphical model. *Biometrika*, 94(1):19–35.

Appendix

A1 Example microfoundation

Our data-generating process corresponds to equilibrium outcomes of a game on a network with linear-quadratic utilities and individual-specific consideration sets determined by characteristics (Abaluck and Adams-Prassl, 2021).

Imagine each of $i = 1, ..., N$ individuals on a weighted network G_{ij} who play an action $y_i \in \mathbb{R}_+$. An individual's payoffs depend on their own action and the actions of others through the linear-quadratic utility function

$$u_i = (\beta x_i + \epsilon_i) y_i + \sum_j \gamma_{ij} a_{ij} y_i y_j - \frac{1}{2} y_i^2.$$

There are $T + 1$ periods. In the first period, nature draws a vector-valued position Z_i in some vector space $\mathbb{R}_+^M$ for each individual i, covariates x_{it}, and time-varying shocks to marginal benefits ϵ_{it}. Then, each individual forms links with the other at rates η_{Z_i, Z_j} with interaction intensities γ_{ij}. In the subsequent periods, individuals adjust their actions to maximise their payoffs for the fixed network G. The first-order conditions of this game are

$$(\beta x_{it} + \epsilon_{it}) + \lambda \sum_j \gamma_{ij} a_{ij} y_j - y_i = 0$$

Rearranging

$$y_i^* = (\beta x_i + \epsilon_i) + \lambda \sum_j \gamma_{ij} a_{ij} y_j^*.$$

Stacking this into matrix form and inverting gives

$$y^* = (I - G)^{-1} (X\beta + \epsilon)$$

where $G_{ij} = \gamma_{ij} A_{ij}$ encodes the intensity of interactions between each individual.

A2 Proofs

We start with the following lemma

Lemma 2.

$$\frac{d \log \det(XX')}{dX} = \frac{d \log \det(X'X)}{dX}$$

By determinant rules

$$\begin{aligned} \det(XX') &= \det(X) \det(X') \\ &= \det(X') \det(X) \\ &= \det(X'X). \end{aligned}$$

as $\det(A)$ is a constant. Therefore

$$\frac{d \log \det(XX')}{dX} = \frac{d \log \det(X'X)}{dX}.$$

A2.1 Proposition 1

Proof. Our objective function (fixing $\sigma^2 = 1$ without loss of generality) is

$$l(G) = \log\det\Big(\frac{(I-\gamma G)(I-\gamma G)'}{\sigma^2}\Big) - \operatorname{trace}\Big(S\frac{(I-\gamma G)(I-\gamma G)'}{\sigma^2}\Big) + \sum_{i,j}\ln\Big(\frac{\eta^*_{Z_i,Z_j}}{2\nu_1}e^{-\frac{|G_{ij}|}{\nu_1}} + \frac{1-\eta^*_{Z_i,Z_j}}{2\nu_0}e^{-\frac{|G_{ij}|}{\nu_0}}\Big).$$

To show that this is convex, we inspect the second-order subgradients of each term. The second-order subgradient of the first term is

$$2I \otimes S + 2(I-G)^{-1} \otimes (I-G)^{-1}.$$

By construction, $S \succ 0$, and $G \succ 0$. Considering the Neumann series expansion

$$(I-G)^{-1} = \sum_{k=0}^{\infty} G^K$$

as $\rho(G) < 1$. So, $G \succ 0 \implies (I-G)^{-1} \succ 0$. Therefore

$$2I \otimes S + 2(I-G)^{-1} \otimes (I-G)^{-1}$$

is positive semidefinite, which implies that the first term is strictly convex. Now, consider the penalty term. The second-order subdifferential for each ij is (Gan et al., 2019)

$$\frac{\Big(\frac{1}{\nu_0}-\frac{1}{\nu_1}\Big)\frac{\eta_{Z_i,Z_j}\nu_0}{(1-\eta_{Z_i,Z_j})\nu_1}e^{(\frac{1}{\nu_0}-\frac{1}{\nu_1})G_{ij}}}{\Big(\frac{\eta_{Z_i,Z_j}\nu_0}{(1-\eta_{Z_i,Z_j})\nu_1}e^{(\frac{1}{\nu_0}-\frac{1}{\nu_1})G_{ij}}+1\Big)^2}$$

We can write the subgradient expression as

$$\Big(\frac{1}{\nu_0}-\frac{1}{\nu_1}\Big)\frac{x}{(1+x)^2} \text{ for } x = \frac{\eta_{Z_i,Z_j}\nu_0}{(1-\eta_{Z_i,Z_j})\nu_1}e^{(\frac{1}{\nu_0}-\frac{1}{\nu_1})G_{ij}}.$$

For $x > 0$,

$$0 \le \frac{x}{(x+1)^2} \le \frac{1}{4}.$$

Assume that $\nu_1 > \nu_0$ and $\eta_{Z_i,Z_j} \in [0,1)$. Then $x > 0$, and it follows that

$$0 \ge \frac{\Big(\frac{1}{\nu_0}-\frac{1}{\nu_1}\Big)\frac{\eta_{Z_i,Z_j}\nu_0}{(1-\eta_{Z_i,Z_j})\nu_1}e^{(\frac{1}{\nu_0}-\frac{1}{\nu_1})G_{ij}}}{\Big(\frac{\eta_{Z_i,Z_j}\nu_0}{(1-\eta_{Z_i,Z_j})\nu_1}e^{(\frac{1}{\nu_0}-\frac{1}{\nu_1})G_{ij}}+1\Big)^2} \le \frac{1}{4}\Big(\frac{1}{\nu_0}-\frac{1}{\nu_1}\Big).$$

These terms are weakly positive. Therefore, if we stack these terms into a $N \times N$ matrix (with zeros on the diagonal as we impose a simple matrix), the result is positive semidefinite. Therefore, the second term is also convex. The sum of convex functions over the same domain is also convex. Therefore, the whole problem is convex.

□

A2.2 Theorem 1

Our proof follows the logic of the proof of the main theorem in Gan et al. (2019). Throughout, we work with the case of $\sigma^2 = 1$ without loss of generality, to lighten notation. We start by proving the following lemma

Lemma 3. Define

$$r = \max\Big\{2M_{\Gamma^0}(||\tilde{W}||_\infty + \max(\frac{1}{2}\text{pen}(\delta), \tau), 2(2C_1 + C_3)M_{\Gamma^0}\sqrt{\frac{\ln N}{T}}\Big\}.$$

If

$$\begin{aligned} r &\leq \min\Big\{\frac{1}{3aM_{\Sigma_0}}, \frac{1}{3aM_{\Gamma_0}M_{\Sigma_0}^3}\Big\} \\ \min|G^0_{\mathcal{B}}| &\geq r + \delta \\ \rho(G_0) &\leq 1 \end{aligned}$$

then $\mathcal{A}$ is non-empty and there exists $\hat{G} \in \mathcal{A}$ s.t $||\Delta||_\infty := ||\hat{G} - G_0||_\infty \leq r$.

Proof. To prove this, define the set of true links that are 'large enough'

$$\mathcal{B} = \{(i,j) : |G^0_{ij}| > 2(2C_1 + C_3)M_{\Gamma_0}\sqrt{\frac{\ln N}{T}}\},$$

the set of diagonal entries as $\mathcal{D}$, and

$$\Delta = \hat{G} - G^0.$$

We want to bound $||\Delta||_\infty \leq 2(2C_1 + C_3)M_{\Gamma_0}\sqrt{\frac{\ln N}{T}} \leq r$. From the triangle inequality

$$||\Delta||_\infty = ||\Delta_{\mathcal{B}} + \Delta_{\mathcal{B}^c}||_\infty \leq ||\Delta_{\mathcal{B}}||_\infty + ||\Delta_{\mathcal{B}^c}||_\infty.$$

As the entries of the matrices $\Delta_{\mathcal{B}}, \Delta_{\mathcal{B}^c}$ are in different locations by the definition of the index set,

$$||\Delta_{\mathcal{B}}||_\infty + ||\Delta_{\mathcal{B}^c}||_\infty = \max(||\Delta_{\mathcal{B}}||_\infty, ||\Delta_{\mathcal{B}^c}||_\infty)$$

So, we can sharpen our bound as

$$||\Delta||_\infty \leq \max(||\Delta_{\mathcal{B}}||_\infty, ||\Delta_{\mathcal{B}^c}||_\infty).$$

So we bound $||\Delta||_\infty$ by bounding $||\Delta_{\mathcal{B}}||_\infty$ and $||\Delta_{\mathcal{B}^c}||_\infty$ individually. By definition,

$$\begin{aligned} ||\Delta_{\mathcal{B}^c}|| &= \max(\{G^0_{ij} : (i,j) \in \mathcal{B} \cap \mathcal{D}\}) \\ &\leq 2(2C_1 + C_3)M_{\Gamma_0}\sqrt{\frac{\ln N}{T}}. \end{aligned}$$

So, we have to prove that

$$||\Delta_{\mathcal{B}}||_\infty \leq 2(2C_1 + C_3)M_{\Gamma_0}\sqrt{\frac{\ln N}{T}}.$$

Let $\Gamma^0_{\mathcal{B}\mathcal{B}}$ be the $\mathcal{B}, \mathcal{B}$ block of the Hessian of $\text{logdet}((I-G)'(I-G))$, and Z be the matrix of penalisation terms. Define the mapping

$$F(\text{vec}(\Delta_{\mathcal{B}})) = -\Gamma^{-1}_{0,\mathcal{B}\mathcal{B}}\,\text{vec}(H(G_{\prime,\mathcal{B}} + \Delta_{\mathcal{B}})) + \text{vec}(\Delta_{\mathcal{B}}).$$

The function F is continuous. Now $F(\text{vec}(\Delta_{\mathcal{B}})) = \text{vec}(\Delta_{\mathcal{B}})$ at the point where $H(G_{\prime,\mathcal{B}} + \Delta_{\mathcal{B}}) = H(G_{\prime,\mathcal{B}}) = 0$. So, we will prove the result by showing that $F(\mathbb{B}(r)) \subseteq \mathbb{B}(r)$ for the convex and compact l_∞ ball in $\mathbb{R}^{|\mathcal{B}|}$. Then, by Brouwer's fixed point theorem, there exists a fixed point $\text{vec}(\Delta_{\mathcal{B}}) \in \mathbb{B}^r$ i.e $||\Delta_{\mathcal{B}}||_\infty \leq r$.

To show this, let Δ denote the $N \times N$ zero-padded matrix that equals $\Delta_{\mathcal{B}}$ on $\mathcal{B}$ and zero on $\mathcal{B}^c$. Then, we can write

$$\begin{aligned} F(\text{vec}(\Delta_{\mathcal{B}})) &= -\Gamma^{-1}_{0,\mathcal{B}\mathcal{B}}\Big(-(I - G + \Delta)^{-1}_{\mathcal{B}} - (S(I - G + \Delta))_{\mathcal{B}} + Z_{\mathcal{B}}\Big) + \text{vec}(\Delta_{\mathcal{B}}). \\ &= \Gamma^{-1}_{0,\mathcal{B}\mathcal{B}}\Big(-(I - G + \Delta)^{-1}_{\mathcal{B}} + (I - G_0)^{-1} - (I - G_0)^{-1} - (S(I - G + \Delta))_{\mathcal{B}} + Z_{\mathcal{B}}\Big) + \text{vec}(\Delta_{\mathcal{B}}). \end{aligned}$$

As in Gan et al. (2019), we can separate out this expression into two terms

$$F(\text{vec}(\Delta_{\mathcal{B}})) \leq ||\mathbf{I}||_\infty + ||\mathbf{II}||_\infty$$

First, lets bound

$$\begin{aligned} ||\mathbf{I}||_\infty &= ||\Gamma^{-1}_{0\mathcal{B}\mathcal{B}} R(\Delta)||_\infty \\ &= M_{\Gamma_0}||R(\Delta)||_\infty. \end{aligned}$$

To bound $||R(\Delta)||_\infty$, note that as Σ_0 is positive semi-definite by definition, it has a unique matrix square root. Therefore, we can bound

$$\begin{aligned} ||\Sigma_0^{\frac{1}{2}}||_\infty &= ||\Sigma_0 \Sigma_0^{-\frac{1}{2}}||_\infty \\ &\leq ||\Sigma_0||_\infty ||\Sigma_0^{-\frac{1}{2}}||_\infty \\ &= ||\Sigma_0||_\infty ||(I - G_0)||_\infty \\ &=\leq 2||\Sigma_0||_\infty. \end{aligned}$$

Then, following the steps in the proofs of lemmas 4 and 5 in Ravikumar et al. (2011) exactly, we can bound

$$||M_{\Gamma_0} R(\Delta)||_\infty \leq 3a||\Delta||^2_\infty K^3_{\Sigma_0}$$

where a is the maximum row sum of the matrix of connections A.

Now, by assumption, $\min|G_{\mathcal{B}}| \geq r + \delta$. Therefore, $||\Delta||_\infty \leq r$, $\min|G_{\mathcal{B}}| \geq \delta$, and since $\text{pen}(G)$ is monotonically decreasing we have that $||Z_{\mathcal{B}}||_\infty \leq \text{pen}(\delta)$. It follows that we can write

$$||\mathbf{II}||_\infty \leq M_{\Gamma_0}\Big(||W(I - G_0)||_\infty + \max(\frac{1}{2}\text{pen}(\delta), \tau)\Big)$$

Now, we know from Cai et al. (2011).

$$||W||_\infty \leq C_1\sqrt{\frac{\ln N}{T}}.$$

As $G \succ 0, \rho(G) \leq 1$ we know that $||G_0||_\infty \leq 1$. As G_0 is a simple network by assumption, we have that

$$||I - G_0||_\infty = 1 + ||G_0||_\infty \leq 2.$$

As $||.||_\infty$ is an operator norm, for any A, B

$$||AB||_\infty \leq ||A||_\infty ||B||_\infty.$$

Therefore

$$||W(I-G_0)||_\infty \leq ||W||_\infty ||(I-G_0)||_\infty \leq 2C_1\sqrt{\frac{\ln N}{T}}.$$

Applying the steps in the proof of Theorem A in Gan et al. (2019) under our assumptions on $\frac{1}{T\nu_0}, \frac{1}{T\nu_1}$ gives

$$\frac{|\text{pen}'_{SS}(\delta)|}{T} < C_3\sqrt{\frac{\ln N}{T}}.$$

Therefore

$$\begin{aligned} 2M_{\Gamma_0}(||W(I-G_0)||_\infty + \max\Big(\frac{|\text{pen}'_{SS}(\delta)|}{T}, \frac{2}{T}\tau\Big) &\leq 2(2C_1+C_3)M_{\Gamma_0}\sqrt{\frac{\ln N}{T}} \\ &\leq \min\Big\{\frac{1}{3aM_{\Sigma_0}}, \frac{1}{3aM_{\Gamma_0}M_{\Sigma_0}^3}\Big\}. \end{aligned}$$

Therefore, there exists a $\hat{G}$ such that $||\hat{G}-G_0|| \leq r$.

□

Proof. We structure the proof similar to the proof of Theorem A in Gan et al. (2019), itself based on Ravikumar et al. (2011). Define the set of true links that are 'large enough'

$$\mathcal{B} = \{(i,j) : |G^0_{ij}| > 2(2C_1+C_3)M_{\Gamma_0}\sqrt{\frac{\ln N}{T}}\},$$

and the set of diagonal entries as $\mathcal{D}$. Consider the problem

$$\underset{G>0, \rho(G)\leq 1, G_{\mathcal{B}^c}=0}{\arg\min} L(G).$$

Define the solution set as

$$\mathcal{A} = \{G : (-(I-G)^{-1} + S(I-G))_{\mathcal{B}} + Z_{\mathcal{B}} = 0, G > 0, \rho(G) \leq 1\},$$

and the function

$$H(G_{\mathcal{B}}) = -(I-G_{\mathcal{B}})^{-1} + S(I-G_{\mathcal{B}}) + Z_{\mathcal{B}}.$$

$\mathcal{A}$ is the set such that $H(G_{\mathcal{B}}) = 0$.

Define $\min(G^0_{ij}) = 2(2C_1+C_3)M_{\Gamma_0}\sqrt{\frac{\ln N}{T}}$. $G^0_{ij} \geq 2(2C_1+C_3)M_{\Gamma_0}\sqrt{\frac{\ln N}{T}}$ if $G^0_{ij} \in \mathcal{B}$, and $G^0_{ij} \leq 2(2C_1+C_3)M_{\Gamma_0}\sqrt{\frac{\ln N}{T}}$ if $G^0_{ij} \in \mathcal{B}^c \cap \mathcal{D}$.

Now, we want to prove that the set $\mathcal{A}$ is not empty. Define

$$\begin{aligned} \tilde{W} &= S - \Sigma_0 \\ \Delta &= \hat{G} - G, \\ a &= ||A||_\infty \text{ maximum degree on the binary network} \\ R(\Delta) &= (I-\hat{G})^{-1} - \Sigma_0^{\frac{1}{2}} + \Sigma_0^{\frac{1}{2}}\Delta\Sigma_0^{\frac{1}{2}}, \end{aligned}$$

the difference between the gradient of $\text{logdet}((I-\hat{G})(I-\hat{G}))$ and the first-order Taylor expansion of the gradient about G_0. Now, we apply our lemma. To apply this, we must verify the conditions for the lemma. Consider

$$r = 2(2C_1+C_3)M_{\Gamma_0}\sqrt{\frac{\ln N}{T}}.$$

Therefore $\min|G^0_{ij}|> r+\delta$. Furthermore, given our conditions on the sample size, the condition on r also holds. Therefore

$$||\hat{G}-G_0||_\infty \leq 2(2C_1+C_3)M_{\Gamma_0}\sqrt{\frac{\ln N}{T}}.$$

To convert this into a Frobenius norm bound, we note that there are at most a non-zero entries

$$||\hat{G}-G_0||_F \leq 2\sqrt{a}(2C_1+C_3)M_{\Gamma_0}\sqrt{\frac{\ln N}{T}}.$$

The result follows. Finally, we need to verify that $\hat{G}$ is a local minimiser of the loss function. We do thus by showing that

$$L(\hat{G}+\Delta)-L(\hat{G})\geq 0$$

for some $||\Delta||_\infty \leq \epsilon$. To do this, split our loss function into

$$(1) = \log\det\Big(\frac{(I-\gamma G)(I-\gamma G)'}{\sigma^2}\Big) - \text{trace}\Big(S\frac{(I-\gamma G)(I-\gamma G)'}{\sigma^2}\Big)$$
$$(2) = \sum_{i,j}\ln\Big(\frac{\eta^*_{Z_i,Z_j}}{2\nu_1}e^{-\frac{|G_{ij}|}{\nu_1}} + \frac{1-\eta^*_{Z_i,Z_j}}{2\nu_0}e^{-\frac{|G_{ij}|}{\nu_0}}\Big).$$

Applying lemma A.5 in Battaglini et al. (2021) gives that this is greater than zero. Now we need to bound the deviation in the penalty term (B). The same argument as in the bound of the fraction in (B) step 3 in Theorem A. in Gan et al. (2019) is sufficient to bound this term as greater than zero. So, our result follows.

Finally, we need to show the results for the values of C_1 given the different tail conditions on the distribution of ϵ. We can take these directly from the results bounding the deviation of $||W||_\infty$ in Cai et al. (2011).

□

A2.3 Proposition 2

Proof. Our algorithm is a block-coordinate descent algorithm over a function that is separable into a differentiable and non-differentiable part. To start with, we repeat a result from Tseng (2001). Consider a function

$$f(x) = g(x) + \sum_i h(x_i)$$

where $g(x)$ is a convex, differentiable, and single-valued function over its domain, and $h(x_i)$ is convex and non-differentiable for each x_i. Then a block-coordinate descent algorithm converges to the true optimum of $f()$.

Split our optimisation problem Eq. 9 into two parts

$$\log\det\Big(\frac{(I-\gamma G)(I-\gamma G)'}{\sigma^2}\Big) - \text{trace}\Big(S\frac{(I-\gamma G)(I-\gamma G)'}{\sigma^2}\Big), \text{ and}$$
$$\sum_{i,j}\ln\Big(\frac{\eta^*_{Z_i,Z_j}}{2\nu_1}e^{-\frac{|G_{ij}|}{\nu_1}} + \frac{1-\eta^*_{Z_i,Z_j}}{2\nu_0}e^{-\frac{|G_{ij}|}{\nu_0}}\Big).$$

From the proof of proposition 1, we see that the first part is a convex, differentiable, and single-valued function over the domain $\rho(G)<1$. Furthermore, we see that the second part is convex and non-differentiable over the domain $\rho(G)<1$. So, we can apply the result from Tseng (2001) to show that a block-coordinate descent algorithm will converge.

□