A Graphical Lasso Approach to Estimating Network Connections: The Case of U.S. Lawmakers*

Marco Battaglini, Cornell University and EIEF

Forrest W. Crawford, Yale University

Eleonora Patacchini, Cornell University

Sida Peng, Microsoft Research

Abstract

We develop a new approach to the estimation of social networks that relies only on the observable outcomes that they contribute to determine. We apply the method to the estimation of productivity spillovers in the U.S. Congress revealing new insights on the nature of the lawmakers' social interactions. We estimate a significant decrease over time in the importance of productivity spillovers among lawmakers, compensated by an increase in a party level common shock. This suggests that the rise of partisanship is not affecting only the ideological position of legislators when they vote, but more generally how lawmakers collaborate.
JEL Codes: C31, D85, D72

*We thank Laura Forastiere, Michael Konig, Yuan Liao, Xiaodong Liu, Julien Neves, Aureo de Paula, Sudipta Sarangi, Chen Qiu, Edoardo Rainone, Lingzhou Xue, and participants at the 2019 Network Econometrics Juniors' Conference at Northwestern University for constructive comments and discussions.

1 Introduction

In most applications in which social networks are important for economic outcomes, the social network itself is not observed. Social networks, for example, are considered to be key factors in the determination of risky behavior of adolescents, in the determination of politicians' effectiveness in legislatures, in the determination of R&D spillovers and in production networks, among other applications. In all these cases, we typically do not observe the true social or productive connections. At best we have proxies for these connections. Sometimes we observe whether the adolescents belong to the same classroom or legislators belong to the same alumni network; in other cases we observe trade or aggregate production data. The literature has therefore focused on studying the impact of these proxies for social connections, assuming that they are good approximations for the real social connections. This suggests the question: if we do not observe the real social connections, can we estimate them from the social outcome that these social connections contribute to determine?

Only recently has the literature started to address these questions, studying conditions under which the entire social network is identified and under which it can, at least in theory, be recovered from panel data on social outcomes. In other words, researchers seek conditions under which it is possible to predict social connections just by observing the students' risky behaviors, or the effectiveness of legislators, or the productivity of firms over time. The conditions under which this is possible naturally include the requirement that the time series of observed outcomes is sufficiently long, and/or the network is sufficiently sparse. A degree of sparsity and/or sufficiently long time series are necessary because, with the social outcomes of n agents observed for T periods, the number of possible network connections is on the order of n^2,while observations are on the order of $n \times T$ with, typically, $n \gg T$.[1]

The key question in the literature however is really not whether a network is identified under *some* conditions, but instead it is whether identification is possible for *feasible* and *realistic* data sets. While great advances have been made in the recent literature, four limitations still characterize existing approaches.[2] First, existing approaches appear to work well only for long panels on social observations with relatively small networks. The literature has indeed highlighted sufficient conditions for identification requiring the number of observed social outcomes T to grow at the same rate of the number of nodes n in the network. We however generally need a technique that works for relatively short panel data-sets and large networks: most networks are observed only for relatively short periods of time and few observations, and they comprise many nodes. For example, the U.S. Congress comprises over 400 members; more than 15% of the its members, moreover, is not reelected after 2 years. It is

[1] For instance, in legislatures there are typically over $n \geqslant 400$ legislators. The number of observations is however in the single digits or at best low two digits, depending on how frequently behavior is observed, since the same network of legislators cannot be observed for more than a few legislatures.

[2] A more exhaustive discussion of the related literature is presented in the next subsection.

difficult to imagine panel data-sets with hundreds of observations all generated by the same network: networks remain constant only for short periods of time. Second, the literature has focused on environments in which realistic forms of common shocks to the agents are ruled out. In many instances, however, we need to deal with environments with heterogeneous common shocks targeting specific sub-networks. In legislative environments, for example, it is natural to assume that Democrats and Republicans are affected by different shocks. Third, existing approaches rely on estimators for which no valid inference procedure is available.

In this paper, we propose a new approach to the estimation of networks using observable social outcomes that contributes to addressing these limitations. In our approach, we extend the Graphical Lasso model first proposed by Friedman et al. (2008) in order to incorporate the simultaneous equation framework typical of network models. By inducing a structure on the precision of the matrix, our extension leads to a new estimator, for which we prove identification, derive consistency, and we characterize the asymptotic distribution.

We contribute to addressing the questions highlighted above as follows. First, we show that we achieve consistency under the small T, rather than the requirement in the literature that $n/T \to 0$. More importantly, while we characterize the asymptotic relation between the node degree and network sparsity when $n/T \to \infty$; we can show with simulations that this improvement allows us to obtain much more accurate estimates than with existing methods in samples with fewer than 100 observations. Second, we allow for heterogeneous common shocks that, potentially, are different across sub-networks. Finally, we derive a valid inference procedure by deriving the asymptotic distribution for Graphical Lasso estimator under a simultaneous framework, which allows us to obtain confidence intervals on each link of the estimated network.

The cost for these extension is that, by embracing the Graphical Lasso framework, we add more structure to the environment by assuming sub-Gaussian noise. While our theoretical results rely on this structure, we however show in simulations that our approach works well even in more general environments.

We showcase our new estimator using Monte Carlo simulations in finite sample environments and, as a "real world" application, we apply our approach to the estimation of social links among U.S. representatives. The goal of the Monte Carlo simulations is to assess the performance of our estimator in finite samples, evaluate its robustness to the assumptions implicit in the Graphical Lasso and to compare its performance to previous work. A natural benchmark for us is the penalized Generalized Method of Moments (henceforth, penalized GMM) suggested in the seminal work by de Paula et al. (2019) and also earlier work from Caner et al. (2018). The simulations show that our approach works well starting from $T = 20$ observations and improves in accuracy as T increases even if n is big. Compared to de Paula et al. (2019)'s penalized GMM method, we observe a very significant improvement in accuracy with our method under small T and big n, as measured by the True Positive Rate

(TPR) and the False Positive Rate (FPR). Under small T, the penalized GMM almost lost predictability when $n \leqslant 200$ while our method can still detect true connections under a relative moderate trade-off between TPR and FPR .

We apply the techniques developed above to study production externalities among lawmakers in the U.S. Congress. For each congress starting from the 93rd to the 112th, we measure the legislative effectiveness of each legislator in the House of Representatives using the Legislative Effectiveness Score (LES) constructed by Volden and Wiseman (2018). This index collects information measuring a legislator's ability to move bills through the legislative process. Using the time series of the legislators' effectiveness in each congress, we estimate production externalities in each congress using exclusively data on the individual effectiveness. Heterogeneous party fixed effects allow us to capture the evolving importance of parties in shaping patterns of collaboration between congress members. Comparing the estimated social networks over time, we find two main results. First, we observe a significant decrease in importance in production spillovers among individual lawmakers over time, both between members of their own party, and across parties. Second, we show that this decrease in importance is compensated by an increase in the party level common shock over time. Combined, these two results suggest that the rise of partisanship is not affecting only the ideological position of legislators when they vote, but more generally how legislation is constructed in the U.S. Congress.

The organization of the remainder of the paper is as follows. We discuss related literature in the next subsection. Section 2 outlines the theoretical model linking social outcomes to social networks. In Section 3, we present our graphical Lasso estimator and theoretically derive its main properties. In Section 4, we explore the finite sample performance of the estimator. In Section 5 we apply our approach to investigate productivity spillovers among U.S. lawmakers.Section 6 discusses methodological extensions and Section 7 concludes.

1.1 Related literature

There is a recent but already significant literature studying the conditions under which social spillovers can be estimated using social outcomes despite being only partially or completely unobserved. The case with partially observed social spillovers has been studied by Souza (2014), Blume et al. (2015) and Peng (2019). Souza (2014) proposes a maximum likelihood estimator for environments in which the probability that pairs of individuals form connections in the social network depend on exogenous factors such as common gender. Blume et al. (2015) considers the case in which two individuals in the social network are known to not be directly connected. Peng (2019) proposes a new method to identify leaders and followers in a known network.[3]

[3]The fact that economic agents are engaged in many different relationoships has also been recently noted by Joshi et al. (2020). The solution proposed in that paper is a network formation model that allows for the

The focus of our work is on the case when the social network is not observed at all.[4]This case has been studied by de Paula et al. (2019) and Rose (2018).[5] de Paula et al. (2019) show that identification of the social network is obtained if the number T of observed social outcomes generated by the network is of the same order of the size n of the network, a requirement that would be too demanding in the environment we are interested in (such as, for example, the social networks in the U.S. Congress). Rose [2018] does not derive the statistical properties of the estimator. He presents an application of the STIV estimator characterized by Gautier and Tsybakov (2014) that is not designed for dealing with systems of simultanous equations. We show that the estimator in our paper achieve consistency if $s_n d^2 log\left(n\right)/(nT) \to 0$, and $d^2 log\left(n\right)/T \to 0$ where s_n denotes the size of non-sparse elements and d denotes the maximum degree centrality in the network. Differently from standard penalized estimators, the additional restiction on d is due to the simultanous nature of the equation system. However, our results still allow $n/T \to \infty$ which allow significantly shorter panels to be estimated in finite sample. As we mentioned above, moreover, the previous literature allow a common shocks that is the same across all individuals at every period. We instead adopt a random effect approach, which allows us to handle sub-network common shocks without row-sum normalization.

We contribute to overcoming these limitations proposing an extension to the Graphical Lasso model by incorporating the simultaneous equation framework. The Graphical Lasso method was first proposed by Friedman et al. (2008) to estimate sparse graphs as a competing method to Meinshausen and Bühlmann (2006). It is known for fast computational speed and good finite sample performance. Lam and Fan (2009), Xue et al. (2012) and Xue and Zou (2012) propose various extensions and derive statistical properties for Graphical Lasso type estimators. The simultaneous equation framework that is at the core of our extension essentially induces a structure on the precision matrix and leads to a novel estimator in our work. For this new estimator, we prove identification, derive the consistency for this estimator, and we characterize the asymptotic distribution.

Naturally, there is a cost for these additional properties of our estimator: we need to impose significant more structure on the error structure in our model. With Montecarlo

co-existence of multiple network relationships.

[4]If the network is partially observed, potential missing links can be recovered using the methodology presented in Chandrasekhar and Lewis (2016). In these papers, stochastic block models are adopted. The advantage of those methods is that they can be applied to very dense networks. The advantage of our approach is that we do not need any notion on the underlying network structure. In addition, our model can be augmented with partially observed network data. The known links are simply removed from the penalizating function and thus certaintly detected.

[5]Important recent work has also been done by Bonaldi et al. (2015) and Manresa (2013) for models where the social network affects individual decisions only through exogenous characteristics (exogenous social effect). We instead consider the case where own decisions are not only affected by the characteristics of the social connections but also by their own behavior (endogenous social effect) own behavior is affected. Such an extention boils down into a completely different model, with a simultaneous equation structure. As a result, a novel estimator needs to be developed.

simulations we illustrate the cost/benefits of our approach. We test the robustness of our approach with simulations and compare it to the performance of the penalized GMM estimator in de Paula et al. (2019). For long panels (i.e. $T = 500$ and $n = 10$), the two methods perform similarly with penalized GMM estimator doing better at true positive rate; but for $T \leqslant 100$ and $n \geqslant 50$, our estimator serves as a better classifier comparing with penalized GMM estimator. We show that our method performs well also if we deviate form the assumption of Gaussian errors.

A different approach to the estimation of unobserved networks is presented in Battaglini et al. (2020a), who present a model of social network formation and structurally estimate it by Bayesian methods. This approach requires shorter data-sets for the estimation and it does not require strong sparsity assumptions, but relies more on the underlying structural model of network formation.

The empirical exercise in our work should be distinguished from the literature attempting to estimate the determinants of network formation using network observations such as Graham (2016), Fowler and Christakis (2010), Mele (2017), Badev (2017), among others. In these works, it is assumed that at least a social network is observed and that these observations can be interpreted as realizations from an underlying probability distribution of social networks. In our problem, and the literature discussed above, we attempt to estimate the social connections using social outcomes generated by them.

2 Model

In Section 2.1 we present a simple theory of the effectiveness of lawmakers in a legislature in which the lawmakers' productivities depend on their social connections. The analysis will establish the following linear relationship between the levels of effectiveness and their social connections:

$$E_i = \xi_i + \delta \sum\nolimits_{j \in \mathcal{N}} g_{0,ij} E_j, \tag{1}$$

where E_i is the effectiveness in congress of legislator i, $g_{0,ij}$ is the link between legislator i and legislator j, δ represents the endogenous effect as in Manski (1993), and ξ_i is an individual fixed effect. In Section 3 we study under what condition this relationship can be used to estimate the (unobserved) social connections $g_{0,ij}$ in the U.S. Congress using only the (observed) levels of effectiveness $E = (E_1, ..., E_n)$.

While it is useful to ground our analysis in a specific application for which we have a clear microfoundation, we should note that the tools developed here and in Section 3 have wider applicability. First, the model can be used to study any other settings in which the individual performances of a team of players may be affected by positive productivity spillovers among the players. Secondly, the basic linear relationship derived in the model that is at the core of our estimation technique, can be derived in a variety of other contexts

and with alternative microfoundations (see Section 2.2 below). Finally, we observe that the linear relationship (1) is often directly assumed as a starting point of the empirical analysis of social networks without any microfoundation (see for instance de Paula et al. (2019) Rose (2018), and others).

2.1 A theory of legislative effectiveness

Consider a congress comprised of n legislators, where $\mathcal{N} = \{1, ..., n\}$ is the set of legislators. Each legislator has a pet legislative project that s/he cares to implement. The goal of each legislator is to maximize their legislative effectiveness. We assume that legislator i's legislative effectiveness E_i is a function of the legislator i's characteristics, the effort directly exerted by i and how socially connected the legislator is. The idea is that legislators who are well connected are able to benefit from their connections in pushing their pet legislative projects. We assume the following "production function" for legislative effectiveness:

$$E_i = \varphi \left(s_i\right)^{\alpha} \left(l_i\right)^{1-\alpha} + \xi_i \tag{2}$$

where l_i is legislator i's effort, s_i is the legislator's "social connectedness" and ξ_i is a factor idiosyncratic to i that contributes to i's efficacy independently from his/her connections or effort. The Cobb-Douglas in (2) describes the fact that legislator i's level of "social connectedness" s_i and his/her effort l_i are complementary inputs in lawmaking. Social connectedness is defined as:

$$s_i = \sum\nolimits_{j \in \mathcal{N}} g_{0,ij} E_j, \tag{3}$$

where the $n \times n$ matrix $G_0 = \left(g_{0,ij}\right)_{i,j \in N}$ measures of the social link between i and j. The idea behind (3) is that the higher is the effectiveness of the legislators socially connected to i, the higher is i's effectiveness; because of this, the effect of j's effectiveness on i's effectiveness is weighted in (3) by the degree of social connection of i to j. We assume that i's characteristics as captured by ξ_i are observed by the other lawmakers, but not by an econometrician studying the game.

The cost of effort is assumed to be represented by a linear function $L_i(l_i) = c \cdot l_i$, where c is a cost parameter. In the analysis below, we also assume $g_{0,ii} = 0$.

The complication of analyzing the game presented above is that the choice of effort l_i of a player i generates a complex cascade of externalities: effort affects i's effectiveness directly; but also the effectiveness of all players connected to i; and of those connected to those connected to i, etc. The complication is not dissimilar to the complication that arise in the analysis of a competitive equilibrium in an exchange economy where a change in an agent's demand has a direct obvious effect on an agent's utility and an indirect effect on equilibrium prices. The solution in general equilibrium analysis is to assume that agents are "price takers:" agents solve their optimization program taking prices as given. Prices,

however, must clear the market in equilibrium. Such analysis is motivated by the fact that, in many exchange economies, each agent has only a marginal impact on equilibrium prices, thus allowing us to ignore the indirect effects.

The same approach can be adopted in a network game with many players such as ours. Following Battaglini et al. [2020] and [2019], we define a *Network Competitive Equilibrium* as an allocation in which: 1. lawmakers choose their own optimal level of effort l_i taking the other legislator's effectiveness $(E_j)_{j\in\mathcal{N}}$ as given; 2. the expectations regarding the other player's effectiveness must be correct in equilibrium, satisfying (2)-(3). The idea is that legislators are individually small, so they are "price takers" with respect to the other legislators effectiveness: they select their individual efforts depending on how the expected equilibrium effectiveness of their socially connected legislators; but they assume that their effort levels are marginal and will not directly affect the levels of effectiveness of the other legislators. In equilibrium, however, the vector of effectivenesses must be consistent with individual optimal behavior.

Given the other legislator's effectiveness, the optimal level of effort l_i by an individual $i = 1, ..., n$ solves the problem:

$$\max_{l^i\in[0,\bar{l}]} \left\{\varphi\left(s_i\right)^{\alpha}\left(l_i\right)^{1-\alpha} - c\cdot l_i\right\}, \tag{4}$$

Substituting the solution to this maximization in (4), we obtain that the equilibrium levels of legislative efficiency are given by (1), where $\delta = \varphi\left[\left(1-\alpha\right)\varphi/c\right]^{\frac{1-\alpha}{\alpha}}$. These equations can be expressed in matrix form as:

$$\mathbf{E} = \xi + \delta\cdot G_0\cdot\mathbf{E} \tag{5}$$

where $\mathbf{E}$ is the vector of legislative effectiveness, and ξ is the vector of legislators' characteristics.

2.2 Alternative interpretations and microfoundations

The model presented above can be reinterpreted to describe other situations in which E represents a vectors of individual levels of performances in a team of players and G_0 describes productive spillovers among the team members. The linear relationship between E and G_0 in (5), moreover, can be obtained from other microfoundations. For example, Ballester et al. (2006) consider a model in which a level of output y_i is generated by the production function:

$$y_i = \alpha_i x_i + \frac{1}{2}\sigma_{ii}x_i^2 + \delta\sum_{j\neq i}\sigma_{i,j}x_i x_j,$$

where $x = (x_1, ..., x_n)$ is the vector of effort levels in the team. They show that the unique Nash equilibrium of the game in which each player chooses effort simultaneously is given by

the system $x_i = \alpha_i + \delta \cdot \sum_{j \in \mathcal{N}} \left(\frac{\sigma_{i,j}}{|\sigma_{i,i}|} \right) \cdot x_j$ for each i, which is the same as (5), where E is replaced with $x = (x_1, ..., x_n)$ and the matrix δG_0 is replaced by the $n \cdot n$ matrix Σ with i, j element equal to $\delta \cdot \sigma_{i,j} / |\sigma_{i,i}|$. This model has been used extensively to study peer effects in education and risky behavior, using evidence on x and observable proxies for $\sigma_{i,j}$ to estimate the magnitude of δ. The techniques developed in Section 3 for (5) can be used to estimate the entire Σ matrix using only information on x. The difference with the model of Section 2.1 is that while in Section 2.1 we relate a measure of performance (the levels of effectiveness) to the social matrix δG_0, in Ballester et al. (2006) the authors relate the individual levels of effort x to the social matrix Σ.

3 Estimation

Let $t = 1, 2, \cdots, T$ denote the indeces for the congresses we observe. Assume that $\xi_{i,t}$ can be expressed as a linear function of a congressman i's observed characteristics $X_{i,t}$, individuals fixed effect $\alpha = (\alpha_1, \cdots, \alpha_n)'$ and unobserved characteristics $\epsilon_t = (\varepsilon_1, ..., \varepsilon_n)'$ in congress t:

$$\mathbf{E}_t = \delta G_0 \cdot \mathbf{E}_t + X_t \beta + \alpha + \epsilon_t \tag{6}$$

where E_t is a n by 1 vector composed of the n legislators' effectiveness at congress t. In the standard analysis of (6), it makes sense to distinguish between δ and G_0: the social network G_0 is assumed to be observable, and δ is the object of interest. Here we are studying a more challenging problem in which G_0 is unobserved, so δ and G_0 cannot be separated.[6] This is irrelevant for us, since the object of interest is the entire matrix δG_0 in (6). As a result, we can normalize $\delta = 1$ and obtain the following data generating process:

$$\mathbf{E}_t = G_0 \cdot \mathbf{E}_t + X_t \beta + \alpha + \epsilon_t$$

To difference out the individual fixed effects, let $\mathbf{\Delta E}_t = \mathbf{E}_t - \mathbf{E}_{t-1}$ and $\Delta \epsilon_t = \epsilon_t - \epsilon_{t-1}$

$$\Delta \mathbf{E}_t = G_0 \cdot \Delta \mathbf{E}_t + \Delta X_t \beta + \Delta \epsilon_t \tag{7}$$

3.1 Assumptions

The identification of Equation (7) is well studied for the large T and fixed n case in the literature. In those cases, no sparsity assumption is required as the number of observations $n \times T$ is greater than the number of parameters $n \times (n-1)$. We consider the situation when

[6] Note that, because of this different approach, the commonly assumed row sum normalization in the traditional SAR literature (see Lee (2004)) identification of δ is not needed here.

$T \ll n$ and thus equation (7) is not directly estimable due to the coexistence of simultaneity and high-dimensionality. We impose the following assumption to achieve identification.

Assumption 3.1 (Sparsity)**.** *Let $S_n = \{(i,j) : g_{0,ij} \neq 0\}$ and $s_n = |S_n|$, which is the number of non-zero entries in the adjacency matrix G_0. Define d as the maximum of degree centrality in the graph. Then,*

$$s_n d^2 = o\left(nT/\log n\right) \quad \text{and} \quad d = o\left(\sqrt{T/\log n}\right) \quad \text{as } n \to \infty$$

There are $n(n-1)$ off-diagonal entries in an adjacency matrix with n legislators. Assumption 3.1 specify a tradeoff between the global sparsity s_n with the "individual sparsity" d. First, the maxium number of connections for any legislator d must be of order $\sqrt{T/\log(n)}$, thus allows star-shaped network to be detected when $n \gg T$, i.e. the number of legislators is greater than the time periods observed. Second, it requires the product $s_n d^2$ must be of order $nT/\log n$. For example, if individual sparsity d is growing at rate $(T/\log(n))^{1/3}$, then s_n can only grow at rate $O(n(T/\log(n))^{1/3})$. We requires the additional sparsity assumption on the maximum of degree centrality compare with standard Graphical Lasso model. This is because our sparsity assumption is not on the precision matrix $(I_n - G_0)^2/\sigma^2$, but on the adjacency matrix G_0 directly.

Assumption 3.2 (SAR restrictions)**.**

- *There exists $g_{\max} < 1$ such that $|g_{0,ij}| \leqslant g_{max}$ for all i and j.*
- *The $\Delta\epsilon_t$ are i.i.d sub-Gaussian random variable with 0 mean and variance σ^2*
- *The regressors $\Delta x_{t,i}$ in ΔX_t are non-stochastic and uniformly bounded for all n. $\lim_{n\to\infty} \Delta X_t' \Delta X_t/n$ exists and is nonsingular for all t.*

Assumption 3.2 is required to ensure the simultaneous framework is well defined. The first and second parts ensure the matrix $I_n - G_0$ is invertible. The assumption on the error term currently excludes common shocks from our model and we will discuss a relaxation in the next section. We also require the error term to be sub-Gaussian so that concentration inequalities can be derived to bound the empirical process. Sub-Gaussian processes are known to have "almost" bounded support due to the fast decay of its tails. This assumption is usually required for high dimensional estimators as in Belloni et al. (2014). We provide simulation evidence on the sensitivity of our estimator to this assumption. We assume the regressors are deterministic following the convention in the SAR literature (see Anselin (1988), Jin and Lee (2016)). This assumption simplifies the calculation of the variance-covariance matrix by assuming away the randomness from the exogenous effects. We leave the random design case as a topic for future research.

Assumption 3.3 (Graphical Lasso restrictions)**.** *There is a constant L such that*

$$0 < 1/L < \lambda_{min}(I_n - G_0) \leqslant \lambda_{max}(I_n - G_0) < L.$$

where λ_{min} and λ_{max} denotes the minimum and maxmium eigen-value function
The minimum eigenvalue condition is required to prevent the spatial errors from accumulating too fast. This is the same assumption as Assumption (A) in Lam and Fan (2009) and Jankvoá and van de Geer (2019).

3.2 Estimator

We assume G_0 is symmetric when designing our estimator (e.g. the network is undirected). However, this assumption is purely for notation convenience. [7]. We consider the following Graphical Lasso estimator for equation 7

$$\arg\max_{\sigma,G} \left(\log \left| \frac{(I_n - G)^2}{\sigma^2} \right| - \operatorname{tr} \left(S \frac{(I_n - G)^2}{\sigma^2} \right) - \lambda \|G\|_1 \right) \tag{8}$$

where S is the estimated variance-covariance matrix of $P_{\Delta X_t} \Delta \mathbf{E}_t$ where $P_{\Delta X_t}$ is the projection matrix onto the orthogonal space spanned by ΔX_t. Equation 8 is a combination of a loss function based on the maximum likelihood estimator for normal distribution as well as a penalty function where λ is the penalization parameter that induces sparsity in the estimated network (i.e. the smaller the lambda the more dense the network will be). As we are directly imposing sparsity restriction on the adjacency matrix instead of precision matrix, (8) differs from standard Graphical Lasso estimator by imposing additional structures on the precision matrix.

Theorem 3.1. *Under Assumptions 3.1, 3.2, and 3.3, equation 8 is concave. Further with $\lambda \gtrsim d \cdot \sqrt{\log(n)/T}$, we have*

$$\|\hat{G} - G_0\|_F^2 = O_P \left(\frac{(s_n + n) d^2 \log(n)}{T} \right) \quad \text{and} \quad |\hat{\sigma}^2 - \sigma^2|^2 = O_P \left(\frac{(s_n + n) d^2}{n} \cdot \frac{\log(n)}{T} \right)$$

where $\| \cdot \|_F$ is the Frobenius norm.

The identification result in Theorem 3.1 coincides with finding a consistent estimator under the sparsity assumption, which is standard in the Lasso literature. The convergence rate in our result also depends on the maximum of degree centrality. This is slower compared to the standard $(s_n + n)\log(n)/T$ rate in Graphical Lasso. This is again due to how we

[7] In the directed network case, the variance-covariance matrix is $(I_n - G_0)(I_n - G_0)^T$.

impose sparsity assumption in this simultaneous equation system. The results on Frobenius norm can also be stated as $\|\hat{G} - G_0\|_F^2/n = O_p\left((s_n + n)d^2 \log(n)/(nT)\right)$ which provides an approximation for the operator norm. Another practical issue is how to choose the tuning parameter λ. A theoretical guidance can be obtained by simulating the distribution for $n\|(S - \Sigma_0)(I_n - G_0)^2\|_\infty$ with an prior on Σ_0 and G_0. In practice, we recommend using cross-validation on the time dimension to choose λ. Penalized estimators suffer from shrinkage bias and thus have nonstandard distributions for inference. We propose the following de-biased estimator for our Graphical Lasso estimator that can achieve normality.

De-biased Glasso Estimator:

$$\tilde{G} = \hat{G} + \frac{1}{2}(I_n - \hat{G})^2 \left(-I_n + \frac{S}{\hat{\sigma}^2}(I_n - \hat{G})^2 \right) \tag{9}$$

Theorem 3.2. *Define*

$$\delta_{ij} = Var(((I_n - \tilde{G})^2)_{(i,\cdot)} \Delta \mathbf{E}_t \Delta \mathbf{E}_t' (I_n - \tilde{G})_{(\cdot,j)})/4,$$

where $(\cdot)_{(i,\cdot)}$ *and* $(\cdot)_{(\cdot,j)}$ *denote the* i*th row and* j*th column of a matrix. Assume that* $1/\delta_{ij} = O(1)$*. Under Assumptions 1-3, and further assume* $s_0 d^4 = o(\sqrt{T}/\log n \vee T/(n(\log n)^2))$ *and* $d = o\left((\sqrt{T}/(n \log n))^{1/4}\right)$ *, the de-sparse estimator*

$$\sqrt{T}(\tilde{g}_{ij} - g_{0,ij})/\delta_{ij} \xrightarrow{d} N(0, 1)$$

The normality condition is achieved at a cost of stronger sparsity assumption. Theorem 3.2 allows us to construct unbiased estimator for the strength of links and also to test $g_{i,j} = 0$ point-wise.

3.3 Common Shocks

The identification of (7) crucially relies on the independence assumption on the error term. However one of the biggest concerns in the network literature is the existence of homophily. While it is temping to conclude there is a connection between two legislators when we observe a strong correlation on their effectiveness, it could also be the case the two legislators are from the same party and it is the party level common shocks that is responsible for the observed correlation. In the linear-in-means model proposed by Manski (1993), homophily is modeled as additively separable from the individual idiosyncratic shocks . We adopt the same format and model the party effects as follows:

$$\mathbf{E}_t = G_0 \cdot \mathbf{E}_t + X_t\beta + \alpha + Z_R\delta_{Rt} + Z_D\delta_{Dt} + \epsilon_t,$$

where δ_{Rr} and δ_{Dr} are party random shocks that are the same for all party members in a given

year. Let R and D denote the set of Republicans and Democrats. Z_R and Z_D are dummy variables representing the membership for each individual, i.e. $Z_R = (\mathbf{1}(1 \in R), \cdots, \mathbf{1}(n \in R))'$. Difference out the individual fixed effects:

$$\Delta \mathbf{E}_t = G_0 \cdot \Delta \mathbf{E}_t + \Delta X_t \beta + Z_R \Delta \delta_{Rt} + Z_D \Delta \delta_{Dt} + \Delta \epsilon_t. \tag{10}$$

We consider a random effect approach to estimate (10).

Condition 3.1 (Correlated Effects)**.** *The random variable* $(\Delta\delta_{Rt}, \Delta\delta_{Dt})$ *is sub-Gaussian and independent of* $\Delta\epsilon_r$*, i.e.*

$$\begin{pmatrix} \Delta\delta_{Rt} \\ \Delta\delta_{Dt} \end{pmatrix} \sim subG\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_R^2 & \sigma_{RD} \\ \sigma_{RD} & \sigma_D^2 \end{pmatrix} \right)$$

Under (10) and Assumption 3.1, the variance-covariance matrix of $\Delta \mathbf{E}_t$ can be written as

$$\begin{aligned} \Sigma = {} & \sigma^2 (I_n - G_0)^{-2} + \sigma_D^2 (I_n - G_0)^{-1} Z_d Z_d' (I_n - G_0)^{-1} + \sigma_R^2 (I_n - G_0)^{-1} Z_r Z_r' (I_n - G_0)^{-1} \\ & + \sigma_{DR} (I_n - G_0)^{-1} Z_d Z_r' (I_n - G_0)^{-1} + \sigma_{DR} (I_n - G_0)^{-1} Z_r Z_d' (I_n - G_0)^{-1} \end{aligned}$$

The inverse of this matrix can be calculated by the generalized Sherman–Morrison formula given the following condition is satisfied:

$$\left(1 + \frac{\sigma_D^2}{\sigma^2} n_R\right)\left(1 + \frac{\sigma_R^2}{\sigma^2} n_R\right) - \frac{\sigma_{DR}^2}{\sigma_\epsilon^2} n_D n_R > 0$$

4 Simulations

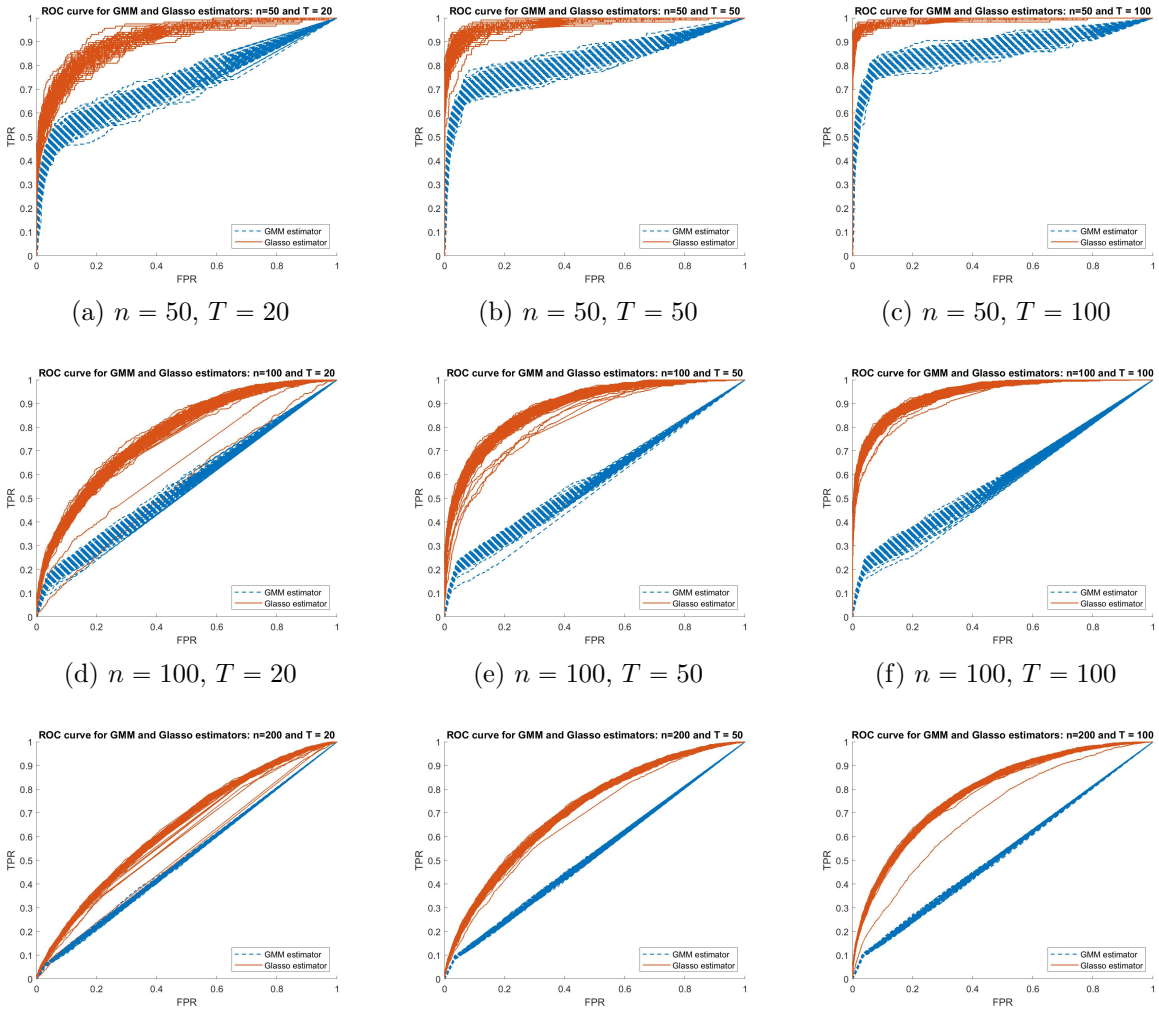
In this section, we present the finite sample performance of our estimator and compare it with the penalized GMM estimator under different network size n and panel lengths T . We simulate the data as in equation (7), e.g.,

$$\Delta \mathbf{E}_t = (I_n - G_0)^{-1} \cdot (\Delta X_t \beta + \Delta \epsilon_t)$$

We use the same simulation set up as de Paula et al. (2019).[8] The parameters are chosen as $\beta = 0.4$ $\varphi^2 = 0.7$ and G is generated by the Erdös Rényi algorithm Erdös and Rényi

[8]We thank Aureo de Paula, Imran Rasul and Pedro Souza for generously allowing us to use their original code to replicate their results in our simulations.

Figure 1: Comparing ROC under large n and small T



(a) $n = 50$, $T = 20$ (b) $n = 50$, $T = 50$ (c) $n = 50$, $T = 100$

(d) $n = 100$, $T = 20$ (e) $n = 100$, $T = 50$ (f) $n = 100$, $T = 100$

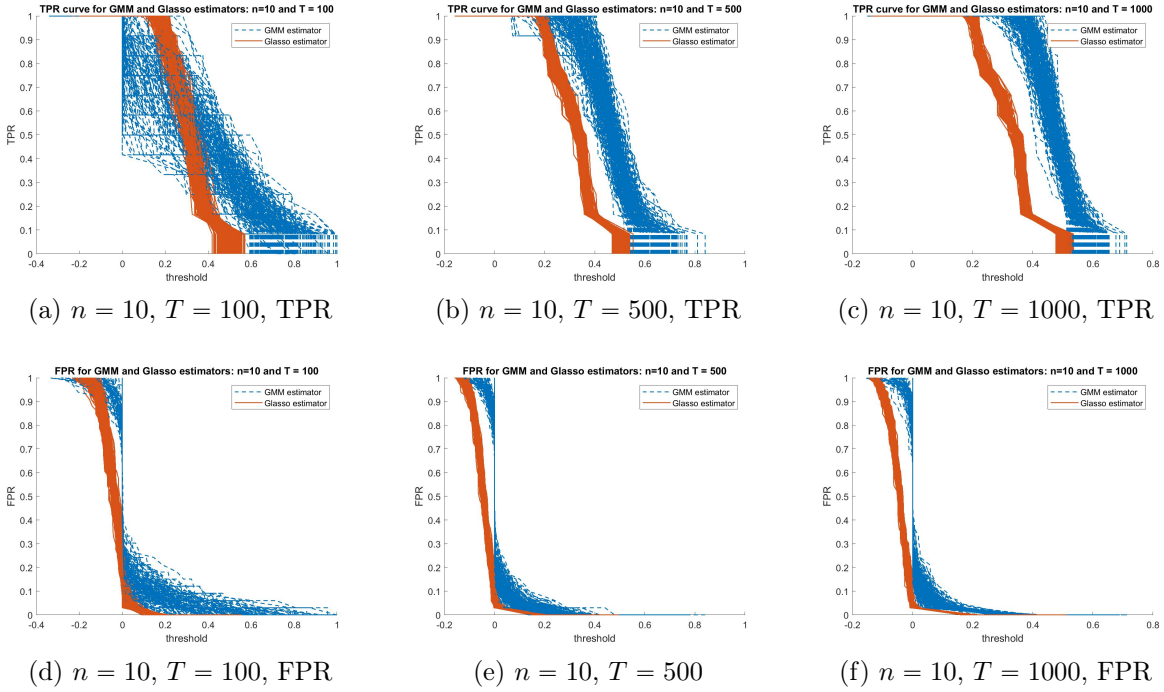
(g) $n = 200$, $T = 20$ (h) $n = 200$, $T = 50$ (i) $n = 200$, $T = 100$

For each scenario, we plot the ROC curves under 200 simulations. The blue dash line represents the ROC curves for penalized GMM method and red solid line represents ROC curves for Graphical Lasso method.

(1959). Each edge is included in the graph with probability $p = 0.1$ independent from every other edge. The variables ΔX_t and $\Delta \epsilon_t$ are generated from a standard normal distribution. To assess the performance of the estimated network with the true network, we consider the estimator as a binary classifier (a connection is detected or not) and compare two difference measures. The true positive rate (TPR) is calculated as

$$TPR(\tau) = \frac{\sum_{i \neq j} \mathbf{1}(g_{0,ij} \neq 0) \cdot \mathbf{1}(\hat{g}_{ij} > \tau)}{\sum_{i \neq j} \mathbf{1}(g_{0,ij} \neq 0)}.$$

Figure 2: Comparing TPR and FPR under small n and large T



(a) $n = 10$, $T = 100$, TPR (b) $n = 10$, $T = 500$, TPR (c) $n = 10$, $T = 1000$, TPR

(d) $n = 10$, $T = 100$, FPR (e) $n = 10$, $T = 500$ (f) $n = 10$, $T = 1000$, FPR

For each scenario, we plot the TPR and FPR separately under 200 simulations. The blue dash line represents the penalized GMM method and red solid line represents ROC curves for Graphical Lasso method.

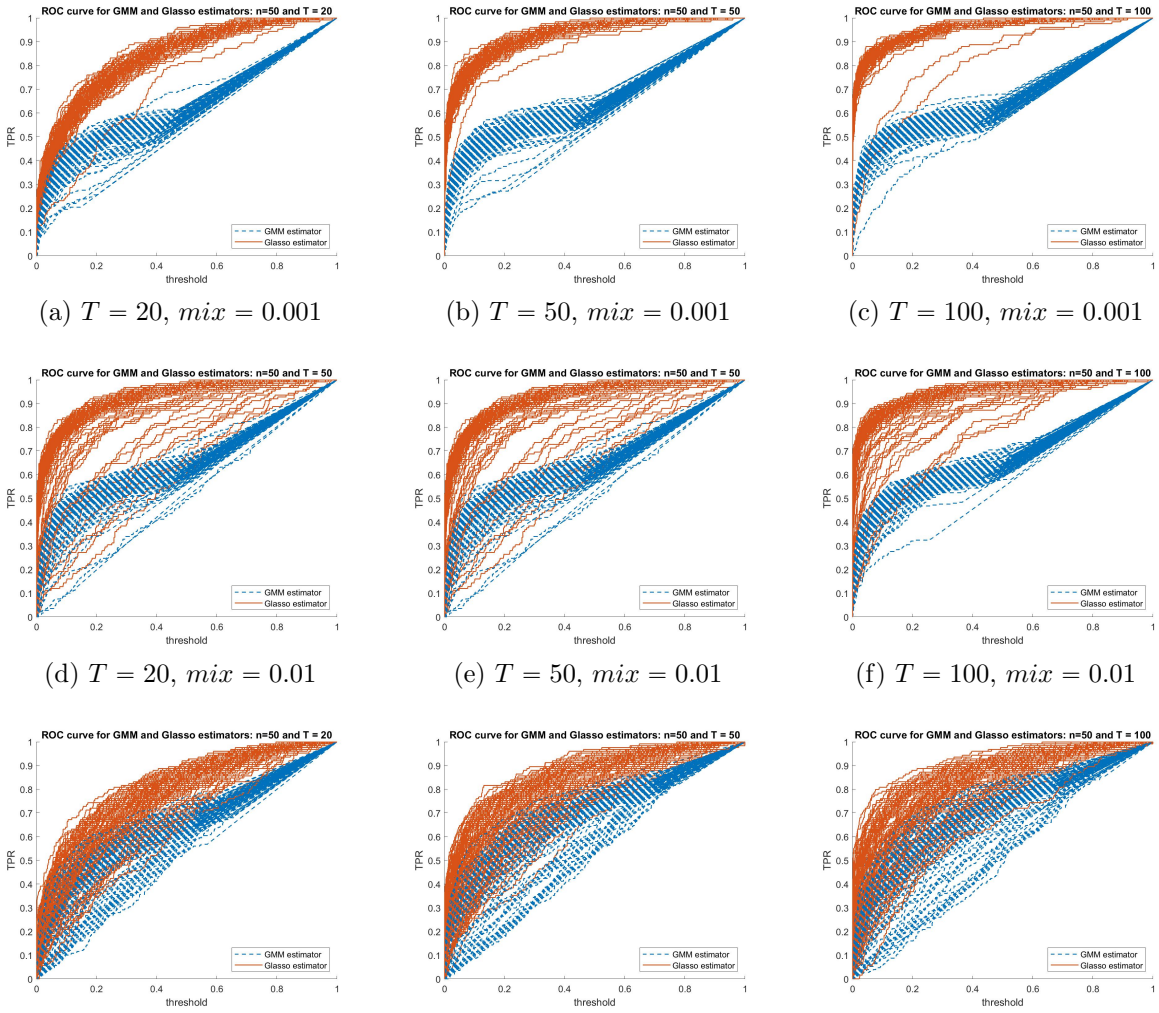
For a fixed threshold τ, the bigger the TPR, the better the classifier performs. The false positive rate (FPR) is calculated as

$$FPR(\tau) = \frac{\sum_{i \neq j} \mathbf{1}(g_{0,ij} = 0) \cdot \mathbf{1}(\hat{g}_{ij} > \tau)}{\sum_{i \neq j} \mathbf{1}(g_{0,ij} = 0)}.$$

For a fixed threshold τ, the smaller the FPR, the better the classifier performs. The Receiver operating characteristic (ROC) curve is a plot of $(TPR(\tau), FPR(\tau))$ with τ varying from 0 to ∞.

Figure 1 plots the ROC curves for both the Graphical Lasso and penalized GMM method with the network size varying from $n = 50$ to $n = 200$ and time period varying from $T = 20$ to $T = 100$. For each scenario, we plot the ROC curves under 200 simulations. The blue dashed line represents the ROC curves for the penalized GMM method and red solid line represents the ROC curves for the Graphical Lasso method. The accuracy for both classifiers improves as T increases and n decreases. The dominance of the solid red ROC curves versus the dashed blue ROC curves in all scenario indicates that Graphical Lasso performs better at detecting the true connections than the penalized GMM estimator. When $n = 200$, the

Figure 3: Comparing ROC under Non-subGaussian error



(a) $T = 20$, $mix = 0.001$ (b) $T = 50$, $mix = 0.001$ (c) $T = 100$, $mix = 0.001$

(d) $T = 20$, $mix = 0.01$ (e) $T = 50$, $mix = 0.01$ (f) $T = 100$, $mix = 0.01$

(g) $T = 20$, Cauchy (h) $T = 50$, Cauchy (i) $T = 100$, Cauchy

For each scenario, we plot the ROC curves under 200 simulations. The blue dash line represents the ROC curves for penalized GMM method and red solid line represents ROC curves for Graphical Lasso method.

TPR for the penalized GMM method increases at the same speed as FPR, which suggests the lack of predictability for the penalized GMM method at small T. The Graphical Lasso method instead can still serve as a reasonably good classifier.

Figure 2 plots the TPR and FPR separately for Graphical Lasso and penalized GMM method under small n but large T. The ROC curves for the two methods are very similar in such scenarios (both converge to the top right corner as T becomes large) so we present the TPR and FPR plots separately. The penalized GMM method outperforms the Graphical Lasso method in TPR as T becomes bigger. On the other hand, Graphical Lasso method is doing better at controlling the FPR.

Figure 3 plots the ROC curves for both the Graphical Lasso and penalized GMM method when the sub-Gaussian assumption is violated. We consider a two-dimensional mixture model for the error term between standard Normal and Cauchy distribution. The mixing weight varies between 1, 0.1 and 0.001. When the mixing weight is 1, the error term is drawn from a Cauchy distribution. When the mixing weight is 0.01, the error term is drawing from a Cauchy distribution with probability 0.01 and from a normal distribution with probability 0.99. This is similar for the case when mixing weight equals 0.001. While both methods suffer from the violation of the sub-Gaussian error, the Graphical Lasso method still outperforms the penalized GMM method in all cases.

5 Estimating production spillovers among U.S. lawmakers

An important literature in political science has highlighted an increase over time of both partisanship and polarization in the U.S. Congress.[9] This literature is based on the study of roll call votes and has focused on the question of whether parties exert pressure on congressional members in close votes (Snyder and Groseclose (2000)); or whether the increase in polarization determines changes in the congressional members' ideal points (McCarty et al. (2001)). An issue that has not been adequately studied, however, is whether the rise of partisanship is affecting how legislators collaborate in drafting legislation. There is a big difference between a legislature where bills are dropped by the leadership on the legislators for approval on partisan lines; and a system where legislation is build from the bottom up, as the result of active collaboration between individual legislators, both within and across parties. To study this question, we apply our technique to the estimation of social spillovers in the "production function" for bills in the U.S. Congress.

We apply our method to study the latent connections among legislators by using their Legislative Effectiveness Score (LES). This metric is based on the work of Volden and Wiseman (2018) and it is a summary metric of how successful a lawmaker is at moving his/her legislative agenda in the legislative process. For our analysis, we use the LES information for the U.S. House of Representative members in 20 Congresses (from the 93rd to the 112th Congress, spanning the years 1973-2013). We have reconstructed the LES at monthly level to allow us obtain enough variations from the time dimension in each congress. Even though we only have $T = 24$ for a two-year congress, the analysis of the properties of our estimator in finite samples in Section 4 shows that our estimator is suitable for this application when

[9] See Rohde (1991), James M. Snyder and Groseclose (2000), Cox and Poole (2002), McCarty et al. (2001), among others for partisanship; and McCarty et al. (2001) and Poole and Rosenthal (2007) on polarization.

T is small.[10]

We apply our method with and without correlated effects to estimate the latent network structure among legislators (equations (7) and (10)).[11] The estimated network can be interpreted as describing the network of production spillovers among lawmakers. We choose the tuning parameter λ proportional to $\sqrt{\log(n)/T}$ which guarantees the rate of convergence as shown in Theorem 3.1. While our methodology produces estimates of the link strength between any pair of Congress members, we present the detected network structures under different choice of λ. For the model with correlated effects, we assume independence between $\Delta\delta_{Rt}$ and $\Delta\delta_{Dt}$ to simplify the computational burden. This assumption does not affect our final results as one of the two correlated shocks are always estimated equal to 0.

Figure 4 illustrates the estimated social networks without correlated effects (e.g., under equations (7)) for two congresses at the beginning of the sample period, the 93rd and the 97th, and two congresses at the end, the 109th and 112th.[12] All Republicans are arranged at the right of the plot while Democrats are at left. The connections between two Republicans are colored in red while the connections between two Democrats are colored in blue. The color between blue and red represents a link between a Republican and a Democratic legislator. We plot all links with $\lambda = 30\sqrt{\log n/T}$. This picture reveals a surprising feature: we observe a significant reduction over time of the number of detected links between lawmakers, both within parties and between parties. We further investigate this finding in Table 1. We report in this table the detected number of links for all the 20 congresses and under different λ for the estimated link strength. Panel A show the baseline estimates (equation (7)), whereas in panel B we show the estimates that control for party level common shocks (equation ((10))). Each panel illustrates the estimated networks by reporting the number of links distinguishing between the Republican party (column R), the Democrat party (column D) and between parties (column RD). The majority party for each congress is indicated in the second column. In panel B , we also report the estimated level of common shocks by party

[10]Specifically, following Volden and Wiseman (2018), for each congress member i and month t we compute:

$$\begin{aligned} LES_{it} = \quad & \alpha(BILL^C_{it} + AIC^C_{it} + ABC^C_{it} + PASS^C_{it} + LAW^C_{it}) \\ & +\beta(BILL^S_{it} + AIC^S_{it} + ABC^S_{it} + PASS^S_{it} + LAW^S_{it}) \end{aligned}$$

where $\alpha = 1$ and $\beta = 5$. Bills are classified as "significant bills" or as "commemorative bills:" the sup-script S represents significant bill and sup-script C represents commemorative bills. For for $l = C, S$, the variable $BILL^l_{it}$ represents the number of bills introduced by i at time t, AIC^l_{it} represents the number of bills introduced by i at time t that received action in committee, ABC^l_{it} represents the number of bills introduced by i at time t that received action beyond committee, $PASS^l_{it}$ represents the number of bills introduced by i at time t that pass the Senate and LAW^l_{it} is the the number of bills introduced by i at time t that become law. Volden and Wiseman (2018) compute the formula below normalizing the LES scores by the total number of bills passed in each subcategory and in each time period. This transform the score into a percentage format which is relative easy to interpret. However, normalization may distort the variance-covariance matrix of the underlying metrics and thus we choose not to normalize.

[11]We take Δx_t as a costant since no other time varying attributes are observed for the legislators at monthly level.

[12]A figure with all the congresses from the 93rd to the 112th is presented in the Appendix.

when $\lambda = 30\sqrt{\log n/T}$. The common shocks estimated under the other two choice of λ are similar.

Three features appear clear from these results. First, the links in the majority party are more numerous than those in the minority party. This is to be expected since the party in the majority has an institutional advantage in lawmaking. Part of this advantage is naturally captured by the individual fixed effects and the party common shock. However, the pattern remains unchnaged even in Panel b), where the party common shock has been purged out. In Appendix Table B4, we show the estimates of the average link strength.[13] The table shows that links between members of the party in the majority are also stronger than between members of the party in the minority. It is therefore important to note that the party advantage does not only enter as a fixed effect, but it also significantly affects the marginal effect of social connections, making complementarities in effectiveness more numerous and significant.

The second, and perhaps the most surprising, observation is that the table confirms a significant reduction over time of the number of detected links between lawmakers, both within parties and between parties, irrespective of the threshold chosen. Looking at panel A, with $\lambda = 10\sqrt{\log n/T}$, in the first 5 congresses in our sample, we detect an average of over 228 links between lawmakers of different parties, over 100 links between lawmakers in the minority party, and over 700 between lawmakers in the majority party (that was the Democratic party). In the last 5 congresses in our sample, on the contrary, the links between lawmakers of different parties are on average only 52, the links between lawmakers in the minority are only 21, and between lawmakers in the majority are 200. It is important to observe that this reduction in connections does not reflect an overall reduction in the legislative activity in Congress. While the results in Table 1 show an important drop in the number of detected connections from the 97th Congress onward, Appendix Figure 6 documents that the average number of bills proposed in Congress over the period is stable.[14]

The third observation emerges when we introduce in the estimation the party level common shocks (panel B). On one hand, panel B confirms the decline in links shown in panel A. Using the most selective tunning, $\lambda = 30\sqrt{\log n/T}$, we always observe spillovers/links at the individual level (from a minimum of 8 to 36) in the first 5 congresses in our sample; using the same criterion, in the last 5 congresses in the sample, we observe less than 1 individually significant link in all congresses except the 11th, where we observe a grand total of 6 links. Using the least selective tunning, $\lambda = 10\sqrt{\log n/T}$, we observe from 164 to 268 significantly positive links in the first 5 congresses; we observe from a minimum of 40 to a maximum of 59 links in the last 5 congresses. Most importantly, however, on the other hand the table shows that the decrease in productivity links at the individual level is associated with an in-

[13] We present the de-biased estimates from Equation 9.

[14] Bills from house of representatives. The data is available at `https://thelawmakers.org/data-download`

creasing importance of the party common shock over time. As it can be seen from the table, the common shock is not detected in the first 5 congresses; it is detected, on the contrary, in all the last 5 congresses that we observe.

Combined, these observations suggest that, in the past 50 years, there has been a significant change in the way the U.S. Congress operates, with an increase in importance of the two leading parties, the Democratic and Republican parties; and a reduction in importance of direct personal connections between lawmakers.

5.1 Validation

We present here two alternative validation exercises for our approach. In the first we present a comparison between the social matrix estimated by our approach and other directly observed adjacency matrices commonly used in the literature to study social connections in the U.S. Congress. We show that our network, although based on different data, is correlated to these networks and thus it recovers information on the social graph captured by these alternative measures. In the second exercise, we present a placebo analysis in which we combine two randomly chosen congresses of different terms, and we show that our approach is significantly more likely to detect connections within terms than across terms.

5.1.1 Comparison with observed adjacency matrices used in the literature

Two main types of adjacency matrices have been used in previous work as proxies of social connections: the cosponsorship matrix, where two lawmakers are linked if they cosponsor each other's bills (see Fowler (2006)); and the alumni connection matrix, where two lawmakers are linked if they have attended the same educational institution within a given time window (see Cohen and Malloy (2014), Battaglini and Patacchini (2018), Battaglini et al. (2020b)).

In Table 2, we compare our estimated network with the consponsorship network. Following the seminal contribution by Fowler (2006), a large literature in political science has used the network of cosponsorships to measure social connections in Congress (see e.g. Kirkland (2011) and Kirkland and Gross (2014)). Cosponsorhip information is obtained using the Library of Congress data information system, THOMAS (`http://thomas.loc.gov`). Following the existing literature, we define two congress members as linked if they have cosponsored the same bills. Notice that our estimated network only uses legislators' effectiveness, which does not contain information on cosponsorships. We should however observe correlation between the two networks if cosponsorship activity is an important factor enhancing legislators' political agenda. For the comparison, we define different cosponsorship networks where a link between two legislators is established if they have cosponsored the same bills for more than $Q(\tau)$ times where $Q(\tau)$ is the τ's quantile of number of bills cosponsored between

any pair of legislators in that given congress. We consider (τ)=0.1, 0.3 or 0.5 (where 0.1 is the 10th percentile, 0.3 the 30th percentile and 0.5 is the median of the distribution). The different thresholds produce networks with different densities. Table 2 compares the *coverage probability* of each of those cosponsorship networks when using the Lasso-detected network and for a random network with the same density of the given cosponsorship network. The coverage probability when using the Lasso-predicted network is computed as the total number of Lasso-detected links which are also in the given cosponsorship network divided by the total number of links in the Lasso-predicted network. The normalized coverage probability reported in the columns denoted by *Lasso-detected* is the coverage probability divided by the coverage probability in a random network with the same density of the given cosponsorship network.[15] This ratio represents the power of predicting the cosponsorship network using the estimated network. The table shows values of the normalized coverage probability of our estimated network consistently higher than one, thus dominating the random guess. This suggests the estimated network contains information on cosponsorship activity in congress.

For the comparison with the alumni network, we construct alumni networks by setting a link to 1 if two legislators attend the same educational institution within a 4-year, 8-year or 10-year time window (and 0 otherwise). This social network is obtained retriving information on high schools and higher education institutions attended for both undergraduate and graduate degrees from the Biographical Directory of the United States Congress (available online at `http://bioguide.Congress.gov/biosearch/biosearch.asp`). Although those social connections are formed many years before election to Congress, a recent literature documents that they are still relevant in explaining legislators' activities in Congress, such as roll call voting or campaign contributions (Cohen and Malloy, 2014, Battaglini and Patacchini 2018). Table 3 has a similar structure of Table 2. It compares the coverage probability of the different alumni networks when using the Lasso-detected networks and for a random network with the same density of the alumni network with the given graduation-time window. Here too we find evidence that the Lasso-predicted network show a good predictions of the alumni networks in congress 99, 104, 105 and 111. This is remarkable not only because we are not using any information from the alumni networks in our estimates, but also because the alumni networks are very sparse (random probability is at 1e-3 level) and only select around 50-100 links (as in Table 1). It is thus very likely that none of the detected connection is in the alumni network (the expect value is at 1e-1 level) even when our method is doing better than a random guess.

[15]The coverage probability in the case of the random network is computed as the total number of links in the given cosponsorship network divided by $(n^2 - n)$. This represents the probability that a random guess in the network coincides with a link in the cosponsorship network. This probability is reported in columns denoted by *Random Network*.

5.1.2 Placebo analysis

We perfrom a further validation check by pooling two congresses of different terms togther and estimate the connections among all legislators in both terms. We expect more connections will be detected from legislators in the same terms than legislators from different terms. In this simulation, two congresses are randomly drawed from the 93rd to 112nd congress. We estimate the network among all legislators using the graphical LASSO and latent graphical LASSO estimators. Figure 5 reportes the distribution of fraction of detected links that are among the legislators in the same term for 200 simulation. As shown in the figure, majority of the links detected in both graphical LASSO and latent graphical LASSO estimators are from legislators in the same term.

6 Extensions

We conclude our analysis by presenting here two methodological extension of the approach developed in the previous sections.

6.1 Endogenous and exogenous effects

A more general version of equation 7 includes spillovers from exogenous characteristics:

$$\Delta \mathbf{E}_t = G_{01} \cdot \Delta \mathbf{E}_t + G_{02} \cdot \Delta X_t + \Delta X_t \beta + \Delta \epsilon_t \tag{11}$$

The second term in (11) correspond to what Manski (1993) defined as exogenous effects, such as spillovers due to exogenous interventions. For example, consider ΔX_t as an indicator for the treatment status at time t. In this context, $\Delta X_t \beta$ represents the direct treatment effect whereas $G_{02} \cdot \Delta X_t$ represent the spillover effect of the treatment, that is the effect due to the treatment status of own social connections. The social network structure transmitting the treatment spillover effect G_{02} or its magnitude may or may not be the same as the transmission structure and magnitude of the endogenous spillover G_{01}. Observe that in this context our model allows to estimate endogenous and exogenous treatment spillovers without observing the social network.

The estimation for equation 11 can be considered as an extension of the procedure presented in Theorem 3.2. First, we can apply the same Graphical Lasso estimator to estimate G_{01}. Next, we can apply the standard Lasso estimator on the reduced form equation $(I_n - \hat{G}_1)\Delta \mathbf{E}_t = G_{02} \cdot \Delta X_t + \Delta X_t \beta + \Delta \epsilon_t$ to recover G_{02}.

6.2 Multidimensional common shocks

The composite structure of the error term in equation (10) (i.e., individual fixed effects plus common shocks) allows us to deal with issues related to endogenous network formation. For example, there may be an unobserved characteristic either at the individual or group level that may be responsible for the agglomeration of similar outcomes. Equation (10) is based on two-dimensional error structure. We focus on the two-dimenionsal case in order to be consistent with our empirical application (the two-partisan structure of the U.S. Congress). A multi-dimensional error structure is however straightforwad to include: our framework allows to include up to a finite number of shocks affecting different individual at the same time. While we assume that the structure of the shocks is known (i.e. we knwo who is Democrat and who is Republican), this is not a necessary condition for identification of equation (10). Along the line of the latent Graphical Lasso proposed by Chandrasekaran et al. (2012), one can separate common shocks from the estimation of G_0 even when the structure or dimension of the shocks is unknown. The derivation of the statistical properties of the latent Graphical Lasso estimator when extended to the case of simultaneous equations, however, is not trivial. We defer that as future research topic. [16]

7 Conclusion

In this paper, we propose a new approach to the estimation of networks using observable social outcomes and we apply it to the estimation of productivity spillovers in the U.S. Congress.

Our approach is designed to address two key problems associated with social networks among lawmakers, but also common with other social networks. First, social networks are not generally directly observed, their estimation therefore should be based on the effect that they induce on observable social outcomes. The cases in which social networks remain stable over long periods of time, moreover, are rare, thus their effects on observable outcomes are observed only for limited amounts of time, thus making the evaluation of their contribution and more generally their estimation challenging. Second, theoretical work on network formation and anecdotal evidence suggests social networks in the U.S. Congress and in other social settings are dense and prone to the formation of stars, i.e. nodes with a high density of connections. Identification, therefore must be guaranteed in environments with these challenging features.

[16] A similar framework can be found in factor models (see, Bai and Li (2012), for a review) where one does not need to assume the factor structure is known. In this setting, the variance-covariance matrix may exhibit certain pattern that can be explained by "low dimension" factors (e.g. block homogeneity: a small block that has similar within block correlation). The complication in our setting is that there exists an unknown high dimesional structure of interactions.

Our work extends the Graphical Lasso model to incorporate the simultaneous equation system arising from equilibrium conditions. Compared to alternative approaches in the literature, our analysis generates an estimator with two appealing properties. First, it is constructed for "small" asymptotics, thus requiring short panels of observations. Second, our estimator requires relatively unrestrictive sparsity assumptions for identification. The "cost" for these appealing features is that for our theoretical derivations, we add more structure to the environment, namely the assumption of sub-Gaussian noise. We present a battery of Monte Carlo simulations to show that our approach works well in small sample environments, both when the sub-Gaussian assumption is satisfied and when it is not.

Our application of these techniques to the study of the U.S. Congress gives us new insights about the nature and effect of social interactions among lawmakers and their evolution over time. Many political scientists have highlighted an increase over time of both partisanship and polarization in the U.S. Congress. This literature, based on the study of roll call votes, has focused on the question of whether parties exert pressure on Congress members in close votes; or whether the increase in polarization determines changes in the Congress members' ideal points. What has not been studied yet in the literaure is whether the rise of partisanship affects how legislators collaborate in drafting legislation. Our analysis suggests that, in the past 50 years, there has been a significant change in the way the U.S. Congress operates, with a reduction in importance of direct personal connections between lawmakers; and an increase in importance of the two leading parties, the Democratic and Republican parties. These results suggest that the rise of partisanship is not affecting only the ideological position of legislators when they vote, but more generally how legislation is constructed in the U.S. Congress.

References

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Boston: Kluwer.

Badev, A. (2017). Discrete games in endogenous networks: Equilibria and policy. `https://arxiv.org/abs/1705.03137`.

Bai, J. and Li, K. (2012). Statistical analysis of factor models of high dimension. *Annals of Statistics*, 40(1):436–465.

Bala, V. and Goyal, S. (2000). A noncooperative model of network formation. *Econometrica*, 68(5):1181–1229.

Ballester, C., Calvó-Armengol, A., and Zenou, Y. (2006). Who's who in networks. wanted: The key player. *Econometrica*, 74(5):1403–1417.

Battaglini, M. and Patacchini, E. (2018). Influencing connected legislators. *Journal of Political Economy*, 126(6):2277–2322.

Battaglini, M., Patacchini, E., and Rainone, E. (2020a). Endogenous social connections in legislatures.

Battaglini, M., Sciabolazza, V. L., and Patacchini, E. (2020b). Effectiveness of connected legislators. *American Journal of Political Science*, 64(4):735–1049.

Belloni, A., Chernozhukov, V., and Hansen, C. (2014). Inference on treatment effects after selection amongst high-dimensional controls. *The Review of Economic Studies*, 81(2):608–650.

Blume, L. E., Brock, W. A., Durlauf, S. N., and Jayaraman, R. (2015). Linear social interactions models. *Journal of Political Economy*, 123(2):444–496.

Bonaldi, P., Hortaçsu, A., and Kastl, J. (2015). An empirical analysis of funding costs spillovers in the euro-zone with application to systemic risk.

Bühlmann, P. and van de Geer, S. (2011). *Statistics for high-dimensional data*. Springer.

Caner, M., Han, X., and Lee, Y. (2018). Adaptive elastic net gmm estimation with many invalid moment conditions: Simultaneous model and moment selection. *Journal of Business & Economic Statistics*, 36(1):24–46.

Chandrasekaran, V., Parrilo, P. A., and Willsky, A. S. (2012). Latent variable graphical model selection via convex optimization. *The Annals of Statistics*, 40(4):The Annals of Statistics.

Chandrasekhar, A. G. and Lewis, R. (2016). Econometrics of sampled networks. `https://web.stanford.edu/~arungc/CL.pdf`.

Cohen, L. and Malloy, C. J. (2014). Friends in high places. *American Economic Journal: Economic Policy*, 6(3):63–91.

Cox, G. W. and Poole, K. T. (2002). On measuring partisanship in roll-call voting: The u.s. house of representatives, 1877-1999. *American Journal of Political Science*, 46(3):477–489.

Currarini, S. and Morelli, M. (2000). Network formation with sequential demands. *Review of Economic Design*, 5(3):229–249.

de Paula, A., Rasul, I., and Souza, P. C. (2019). Recovering social networks from panel data: identification, simulations and an application.

Erdös, P. and Rényi, A. (1959). On random graphs. i. *Publicationes Mathematicae*, 6:290–297.

Falk, A. and Kosfeld, M. (2012). It's all about connections: Evidence on network formation. *Review of Network Economics*, 11(3):Article 2.

Fowler, J. H. (2006). Connecting the congress: A study of cosponsorship networks. *Political Analysis*, 14(4):456–487.

Fowler, J. H. and Christakis, N. A. (2010). Cooperative behavior cascades in human social networks. *Proceedings of the National Academy of Sciences*, 107(12):5334–5338.

Friedman, J., Hastie, T., and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441.

Goeree, J., Riedl, A., and Ule, A. (2009). In search of stars: Network formation among heterogeneous agents. *Games and Economic Behavior*, 67(2):445–466.

Goyal, S., van der Leij, M. J., and Moraga-González, J. L. (2006). Economics: An emerging small world. *Journal of Political Economy*, 114(2):403–412.

Goyal, S. and Vega-Redondo, F. (2007). Structural holes in social networks. *Journal of Economic Theory*, 137(1):460–492.

Graham, B. S. (2016). Homophily and transitivity in dynamic network formation. revision requested by the Review of Economic Studies.

Jackson, M. O. and Wolinsky, A. (1996). A strategic model of social and economic networks. *Journal of Economic Theory*, 71(1):44–74.

James M. Snyder, J. and Groseclose, T. (2000). Estimating party influence in congressional roll-call voting. *American Journal of Political Science*, 44(2):193–211.

Jankvoá and van de Geer (2019). Inference in high-dimensional graphical models.

Jin, F. and Lee, L.-F. (2016). Lasso maximum likelihood estimation of parametric models with singular information matrices.

Joshi, S., Mahmud, A. S., and Sarangi, S. (2020). Network formation with multigraphs and strategic complementarities. *Journal of Economic Theory*, 188.

Kirkland, J. H. (2011). The relational determinants of legislative outcomes: Strong and weak ties between legislators. *The Journal of Politics*, 73(3):887–898.

Kirkland, J. H. and Gross, J. H. (2014). Measurement and theory in legislative networks: The evolving topology of congressional collaboration. *Social Networks*, 36:97– 109.

Lam, C. and Fan, J. (2009). Sparsistency and rates of convergence in large covariance matrix estimation. *The Annals of Statistics*, 37(6B):4254–4278.

Lee, L. (2004). Asymptotic distributions of quasi-maximum likelihood estimators for spatial econometric models. *Econometrica*, 72:1899–1926.

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

Manski, C. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies*, 60(3):531–542.

McCarty, N., Poole, K. T., and Rosenthal, H. (2001). The hunt for party discipline in congress. *The American Political Science Review*, 95(3):673–687.

Meinshausen, N. and Bühlmann, P. (2006). High-dimensional graphs and variable selection with the lasso. *The Annals of Statistics*, 34(1436-1462).

Mele, A. (2017). A structural model of dense network formation. *Econometrica*, 85(3):825–850.

Peng, S. (2019). Heterogeneous endogenous effects in networks.

Poole, K. T. and Rosenthal, H. (2007). On party polarization in congress. *Daedalus*, 136(3):104–107.

Rohde, D. W. (1991). *Parties and Leaders in the Postreform House*. University of Chicago Press.

Rose, C. (2018). Identification of spillover effects using panel data.

Rosenthal, H. and Poole, K. T. (1997). *Congress: a political-economic history of roll call voting*. New York: Oxford University Press.

Sant'Anna, P. H. C. and Zhao, J. B. (forthcoming). Doubly robust difference-in-differences estimators. *Journal of Econometrics*.

Snyder, J. M. and Groseclose, T. (2000). Estimating party influence in congressional roll-call voting. *American Journal of Political Science*, 44(2):193–211.

Souza, P. C. (2014). Estimating network effects without network data. mimeo.

Volden, C. and Wiseman, A. E. (2018). Legislative effectiveness in the united states senate. *Journal of Politics*, 80(2):731–735.

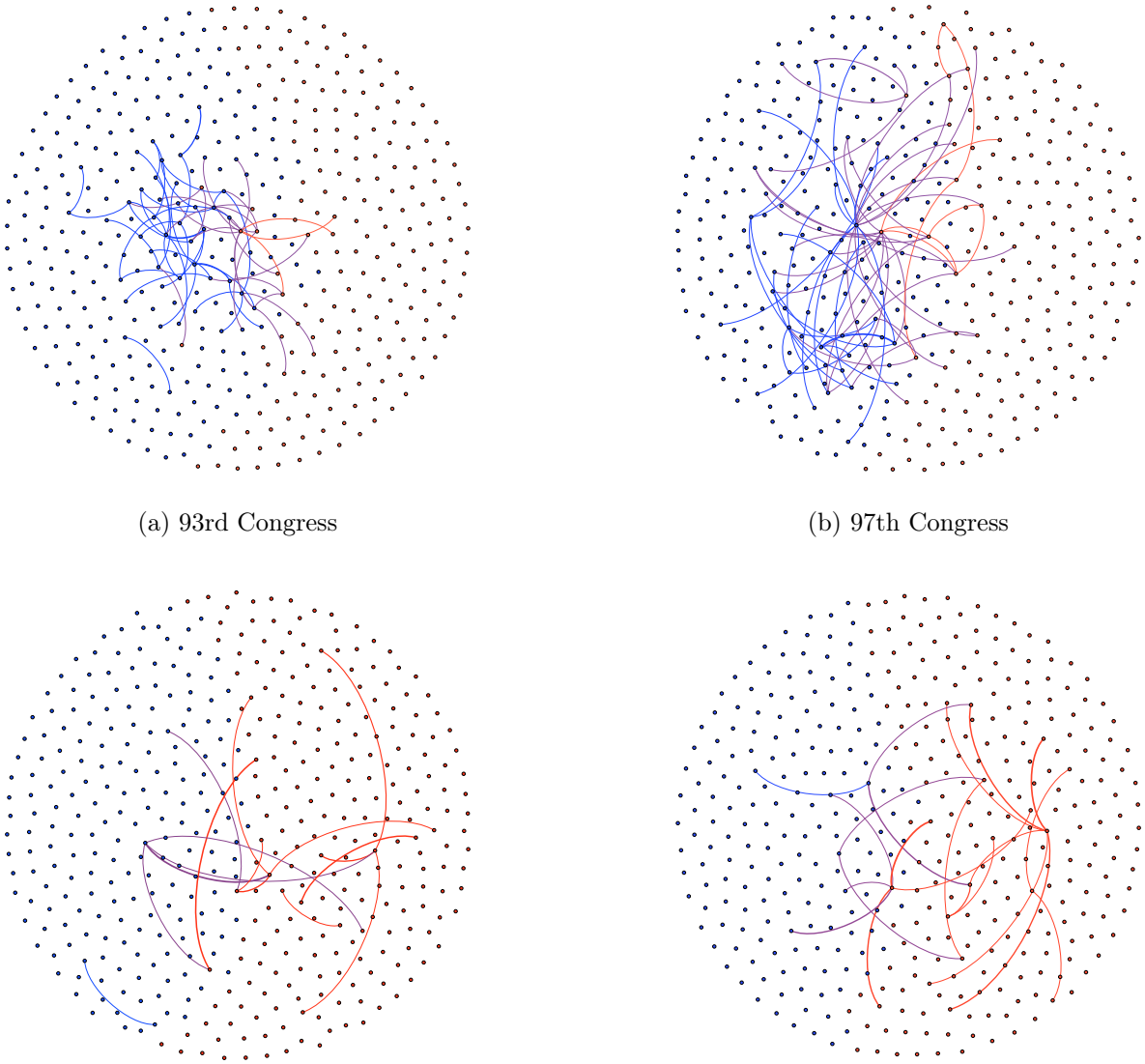
Xue, L., Ma, S., , and Zou, H. (2012). Positive-definite 1-penalized estimation of large covariance matrices. *Journal of the American Statistical Association*, 107:1480–1491.

Xue, L. and Zou, H. (2012). Regularized rank-based estimation of high-dimensional non-paranormal graphical models. *Annals of Statistics*, 40(5):2541–2571.

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

Tables and Figures

Figure 4: Estimated Network Plot



(a) 93rd Congress

(b) 97th Congress

(c) 109th Congress

(d) 112nd Congress

This figure plots the estimated networks with $\lambda = 30\sqrt{\log n/T}$. The red dots represent the Republic party and blue dots represent the Democratic party. Red edges represent connections within the Republican party, blue edges represent the connections within the Democratic party and purple edges represent the connections between the two parties. The stronger the connections are the thicker the line is depicted.

Table 1: Link Distribution

		Panel A									Panel B										
	λ	$50\sqrt{\log n/T}$			$30\sqrt{\log n/T}$			$10\sqrt{\log n/T}$			$50\sqrt{\log n/T}$			$30\sqrt{\log n/T}$			$10\sqrt{\log n/T}$			Common Shock	
Cong	Maj	D	R	DR	D	R	DR	D	R	DR	D	R	DR	D	R	DR	D	R	DR	D	R
93	D	108	2	24	202	6	62	858	30	268	54	0	3	148	4	43	520	14	168	0.00	0.00
94	D	70	0	13	162	0	26	962	22	224	30	0	5	50	0	11	588	10	135	0.00	0.00
95	D	132	0	24	234	0	47	806	8	164	54	0	8	128	0	21	582	6	124	0.00	0.00
96	D	38	40	8	88	76	55	588	260	220	20	2	1	32	42	7	682	306	294	0.00	0.00
97	D	24	16	36	40	64	65	316	220	262	10	4	13	26	36	49	226	230	252	0.00	0.00
98	D	22	0	4	38	0	8	436	14	95	22	0	4	34	0	6	1390	72	315	0.00	0.00
99	D	12	0	2	28	2	7	366	10	60	10	0	2	34	2	9	404	12	69	0.00	0.00
100	D	0	6	6	14	10	9	262	36	88	0	6	6	18	10	10	320	38	104	0.04	0.00
101	D	4	0	2	20	0	4	284	16	70	8	0	3	16	0	4	510	22	124	0.01	0.00
102	D	0	2	3	8	6	5	360	26	108	0	0	2	10	6	7	178	14	65	0.02	0.00
103	D	0	0	0	14	0	4	230	20	63	0	0	1	20	0	6	430	40	115	0.01	0.00
104	R	0	12	5	4	24	8	20	398	85	0	8	5	2	14	8	18	266	60	0.00	0.01
105	R	2	0	4	4	24	9	16	272	71	2	0	2	4	10	9	54	536	143	0.00	0.04
106	R	0	0	0	0	6	0	6	250	60	0	0	0	0	2	0	8	322	83	0.00	0.05
107	R	0	0	2	0	10	3	14	146	34	0	0	2	2	26	5	34	330	92	0.00	0.03
108	R	0	0	1	0	6	3	16	184	40	0	0	0	0	4	2	18	296	70	0.00	0.06
109	R	0	0	1	0	10	3	10	164	41	0	2	0	0	8	3	18	306	93	0.00	0.09
110	D	0	0	1	12	4	6	244	48	58	0	0	0	2	2	3	124	40	45	0.08	0.00
111	D	2	0	3	14	0	4	278	22	59	0	0	1	14	0	7	294	38	63	0.39	0.00
112	R	0	4	0	0	18	5	10	132	42	0	2	0	0	16	5	44	354	136	0.00	0.03

This table presents the detected network structures under different choice of tunning parameters. Panel A shows the baseline estimates (equations 7), whereas in panel B we show the estimates that control for party level common shocks (equations 10). Each panel illustrates the estimated networks by reporting the number of links distinguishing between the Republican party (column R), the Democrat party (column D) and between parties (column RD). The majority party for each congress is indicated in the second column.

Table 2: Comparison with Cosponsorship Network

	cosponsor $> Q(0.1)$			cosponsor $> Q(0.3)$			cosponsor $> Q(0.5)$		
	$\lambda = \sqrt{\log n/T}\times$			$\lambda = \sqrt{\log n/T}\times$			$\lambda = \sqrt{\log n/T}\times$		
Cong	50	30	10	50	30	10	50	30	10
93	1.24	1.23	1.11	1.43	1.40	1.21	1.68	1.59	1.33
94	1.15	1.11	1.06	1.20	1.25	1.12	1.68	1.67	1.31
95	1.12	1.12	1.08	1.46	1.45	1.32	1.82	1.82	1.58
96	1.07	1.05	1.05	1.26	1.13	1.14	1.62	1.39	1.33
97	0.80	0.78	0.88	0.74	0.70	0.81	0.82	0.70	0.89
98	1.04	1.03	0.98	1.26	1.21	1.08	1.45	1.57	1.32
99	1.11	1.11	0.99	1.47	1.37	1.14	1.82	1.54	1.35
100	0.86	0.96	0.98	0.33	0.70	0.98	0.23	0.66	1.13
101	1.12	1.03	1.06	1.46	1.25	1.13	1.91	1.74	1.45
102	1.11	1.12	1.02	1.48	1.31	1.20	2.07	1.84	1.50
103	N/A	1.10	1.06	N/A	1.45	1.12	N/A	1.80	1.37
104	1.12	1.08	1.06	1.32	1.26	1.20	1.71	1.56	1.48
105	1.14	1.08	1.08	1.46	1.31	1.15	2.18	1.58	1.22
106	N/A	0.38	0.77	N/A	0.00	0.70	N/A	0.00	0.72
107	1.12	1.15	1.04	1.41	1.24	1.01	2.23	0.96	0.60
108	1.05	0.94	1.02	0.00	0.46	1.05	0.00	0.69	1.01
109	0.00	1.11	1.57	0.00	1.08	1.57	0.00	1.05	1.57
110	1.27	1.51	1.73	1.12	1.56	1.74	1.23	1.53	1.73
111	1.33	1.54	1.49	1.29	1.51	1.49	1.39	1.58	1.49
112	2.27	1.87	1.67	2.21	1.84	1.65	2.33	1.85	1.64

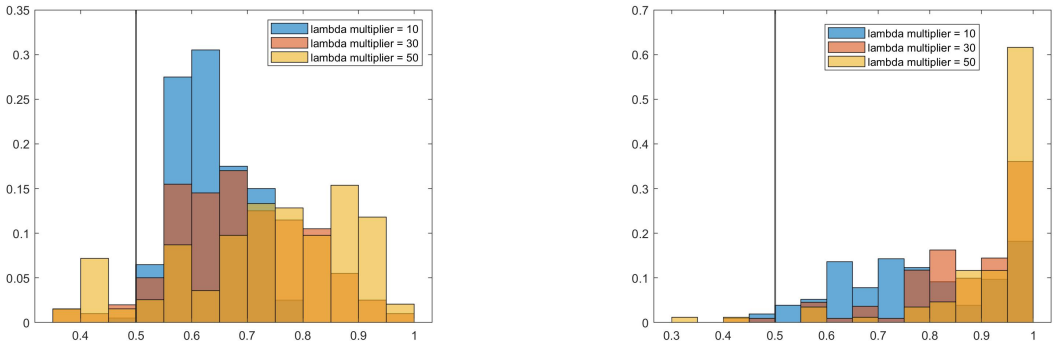
This table display the odds ratio between the event a cosponsorships connection is detected by Graphical Lasso method and the event a cosponsorships connection is detected by a randomly generated network with the same sparsity as the Graphical Lasso detected network. We define two congress members as linked in a cosponsorship network if they have cosponsored the same bills for more than $Q(\tau)$ times where $Q(\tau)$ is the τ's quantile of number of bills cosponsored between any pair of legislators in that given congress. The N/A in the table is due to the empty network detected by Lasso method as shown in table 1.

Table 3: Comparison with Alumni Network

	4 years			8 years			10 years		
	$\lambda = \sqrt{\log n/T}\times$			$\lambda = \sqrt{\log n/T}\times$			$\lambda = \sqrt{\log n/T}\times$		
Cong	50	30	10	50	30	10	50	30	10
93	0.00	0.00	0.00	0.00	0.00	0.57	0.00	0.00	0.49
94	0.00	0.00	0.58	0.00	0.00	0.69	0.00	0.00	0.59
95	0.00	0.00	0.00	0.00	0.00	0.24	0.00	0.00	0.20
96	0.00	0.00	0.00	0.00	0.00	0.42	0.00	0.00	0.36
97	0.00	0.00	0.45	0.00	0.00	0.57	0.00	0.00	0.71
98	0.00	0.00	0.00	0.00	0.00	0.88	0.00	0.00	0.75
99	0.00	0.00	1.99	0.00	0.00	1.15	0.00	0.00	1.49
100	0.00	0.00	1.06	0.00	0.00	0.63	0.00	0.00	0.55
101	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
102	0.00	0.00	0.76	0.00	0.00	0.93	0.00	0.00	0.79
103	N/A	0.00	0.00	N/A	0.00	0.67	N/A	0.00	0.57
104	0.00	0.00	3.59	0.00	0.00	2.11	0.00	0.00	1.84
105	0.00	0.00	1.25	0.00	0.00	0.77	0.00	0.00	0.67
106	N/A	0.00	0.00	N/A	0.00	0.00	N/A	0.00	0.00
107	0.00	0.00	2.41	0.00	0.00	1.42	0.00	0.00	1.17
108	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.99
109	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
110	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
111	0.00	0.00	2.53	0.00	0.00	2.39	0.00	0.00	2.03
112	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

This table display the odds ratio between the event a cosponsorships connection is detected by Graphical Lasso method and the event an alumni connection is detected by a randomly generated network with the same sparsity as the Graphical Lasso detected network. We define two congress members as alumni if they overlapped in a 4-year, 8-year and 10-year window when attending colleges. The N/A in the table is due to the empty network detected by Lasso method as shown in table 1.

Figure 5: Distribution of fraction of links from the same congress term



(a) graphical LASSO

(b) latent graphical LASSO

This figure plots the distribution of fraction of links detected among legislators from the same term in 200 simulation. The distribution are robust to the tunning parameter choosen as a mutiplier to a base rate $\sqrt{\log n/T}$.

A Proofs

Lemma A.1. *Let A be a symmetric, positive semi-definite matrix. Then*

$$\frac{1}{4}(I_n \otimes A^2 + A \otimes A) - (I_n \otimes A^{-2} + A^{-1} \otimes A^{-1})^{-1} \geqslant 0$$

Proof. Denote $A = L\Lambda L'$ as the eigenvalue decomposition, where $\Lambda = diag(\lambda_1, \lambda_2, \cdots, \lambda_n)$ and $\lambda_i > 0$. Thus,

$$\begin{aligned}(I_n \otimes A^{-2} + A^{-1} \otimes A^{-1})^{-1} &= (L \otimes L) \cdot (I_n \otimes \Lambda^{-2} + \Lambda^{-1} \otimes \Lambda^{-1})^{-1} \cdot (L' \otimes L') \\ \frac{1}{4}(I_n \otimes A^2 + A \otimes A) &= \frac{1}{4}(L \otimes L) \cdot (I_n \otimes \Lambda^2 + \Lambda \otimes \Lambda) \cdot (L' \otimes L')\end{aligned}$$

Since

$$\begin{aligned}(I_n \otimes \Lambda^{-2} + \Lambda^{-1} \otimes \Lambda^{-1})^{-1} = diag(&\lambda_1^2/2, (\lambda_2^{-2} + \lambda_1^{-1}\lambda_2^{-2})^{-1}, \cdots, (\lambda_n^{-2} + \lambda_1^{-1}\lambda_n^{-2})^{-1}, \\ &(\lambda_1^{-2} + \lambda_2^{-1}\lambda_1^{-2})^{-1}, \lambda_2^2/2, \cdots, (\lambda_n^{-2} + \lambda_2^{-1}\lambda_n^{-2})^{-1}, \\ &\cdots \\ &(\lambda_1^{-2} + \lambda_n^{-1}\lambda_1^{-2})^{-1}, (\lambda_2^{-2} + \lambda_n^{-1}\lambda_2^{-2})^{-1}, \cdots, \lambda_n^2/2)\end{aligned}$$

and

$$\begin{aligned}\frac{1}{4}(I_n \otimes \Lambda^2 + \Lambda \otimes \Lambda) = diag(&\lambda_1^2/2, (\lambda_2^2 + \lambda_1\lambda_2)/4, \cdots, (\lambda_n^2 + \lambda_1\lambda_n)/4, \\ &(\lambda_1^2 + \lambda_2\lambda_1)/4, \lambda_2^2/2, \cdots, (\lambda_n^2 + \lambda_2\lambda_n)/4, \\ &\cdots \\ &(\lambda_1^2 + \lambda_n\lambda_1)/4, (\lambda_2^2 + \lambda_n\lambda_2)/4, \cdots, \lambda_n^2/2)\end{aligned}$$

Notice that $(\lambda_i^2 + \lambda_j\lambda_i)/4 - (\lambda_i^{-2} + \lambda_j^{-1}\lambda_i^{-2})^{-1} = (\lambda_i\lambda_j^{-1} + \lambda_i^{-1}\lambda_j - 2)/(4(\lambda_i^{-2} + \lambda_j^{-1}\lambda_i^{-2})) \geqslant 0$, thus $\frac{1}{4}(I_n \otimes \Lambda^2 + \Lambda \otimes \Lambda) - (I_n \otimes A^{-2} + A^{-1} \otimes A^{-1})^{-1}$ is postive semi-definite. As a results, the lemma holds.

Lemma A.2. *The following minimization problem is convex.*

$$\arg\min_{G,\sigma^2} -\log\left|\frac{(I_n - G)^2}{\sigma^2}\right| + tr\left(S\frac{(I_n - G)^2}{\sigma^2}\right) + \lambda\|G\|_1 \tag{12}$$

Proof. First, define

$$\begin{aligned} q(\sigma^2, G) &:= -\log\left|\frac{(I_n - G)^2}{\sigma^2}\right| + tr\left(S\frac{(I_n - G)^2}{\sigma^2}\right) \\ &= n\log(\sigma^2) + \frac{1}{\sigma^2}tr\left(S(I_n - G)^2\right) - 2\log|(I_n - G)| \end{aligned}$$

Taking the derivative and Hessian:

$$\frac{\partial q(\sigma, G)}{\partial G} = -\frac{2}{\sigma^2}S(I_n - G) + 2(I_n - G)^{-1}$$

$$\frac{\partial q(\sigma, G)}{\partial \sigma^2} = \frac{n}{\sigma^2} - \frac{1}{\sigma^4}tr\left(S(I_n - G)^2\right)$$

and

$$H_n = \begin{pmatrix} 2\left(I_n \otimes \frac{S}{\sigma^2}\right) + 2(I_n - G)^{-1} \otimes (I_n - G)^{-1} & vec\left(2\left(\frac{S(I_n - G)}{\sigma^4}\right)\right) \\ vec\left(2\left(\frac{S(I_n - G)}{\sigma^4}\right)\right)' & -\left(\frac{n}{\sigma^4}\right) + 2\frac{tr(S(I_n - G)^2)}{\sigma^6} \end{pmatrix}$$

The first order equation implies $\sigma^2 = \frac{1}{n}tr\left(S(I_n - G)^2\right)$. As $T \to \infty$, $S \to \sigma^2(I_n - G)^{-2}$, together with Lemma A.1, the determinant of H_n can be bounded as

$$\det(H_n) \geqslant \det\left(I_n \otimes \frac{S}{\sigma^2}\right) \cdot \left[-\left(\frac{n}{\sigma^4}\right) + \frac{tr(S(I_n - G)^2)}{\sigma^6}\right] = 0$$

Lemma A.3. *Let S be the sample variance-covariance matrix of a sub-Gaussian process and Σ_0 is the population variance-covariance matrix. For non-random matrices A, $B \in \mathbb{R}^{n\times n}$, it holds that $\|A_i\|_2 \leqslant M$ and $\|B_i\|_2 \leqslant M$ for all $i \in 1, \cdots, p$. Then for all $t > 0$, with probability at least $1 - e^{-nt}$ it holds that*

$$\|A'(S - \Sigma_0)B\|_\infty/(2M^2K^2) \leqslant t + \sqrt{2t} + \sqrt{\frac{2\log(2n^2)}{T}} + \frac{\log(2n^2)}{T}$$

This is Lemma 14.13 in Bühlmann and van de Geer (2011). The proof is omitted.

Lemma A.4. *Define the function $q(\sigma_0^2, G_0) = n\log(\sigma_0^2) - \frac{1}{\sigma_0^2}tr\left(S(I_n - G_0)^2\right)$. Define $\Xi_G(\Delta\sigma^2)$ as*

$$\begin{aligned} \Xi_G(\Delta\sigma^2) &= q(\sigma_0^2 + \Delta\sigma^2, G_0) - q(\sigma_0^2, G_0) \\ &= n\log(\sigma_0^2 + \Delta\sigma^2) - n\log(\sigma_0^2) + \left(\frac{1}{\sigma_0^2 + \Delta\sigma^2} - \frac{1}{\sigma_0^2}\right)tr\left(S(I_n - G_0)^2\right) \end{aligned}$$

when $\lambda_0 \gtrsim \sqrt{\log n/T}$ and $d = o\left(T/(\sqrt{n}\log n)\right)$, there exist constant $c_1 = -1/\tilde{\sigma}_1^4 + 2\sigma_0^2/\tilde{\sigma}_1^6$

and $c_2 > 0$ *such that*

$$\Xi_G(\Delta\sigma^2) \geqslant nc_1(\Delta\sigma^2)^2 - c_2 n(1+d)\lambda_0 \left|\Delta\sigma^2\right|$$

Proof. First, notice that

$$tr\left(S(I_n - G_0)^2\right) = n\sigma_0^2 + tr\left((S-\Sigma_0)(I_n - G_0)^2\right)$$

Since $\|(I_n - G_0)_i\|_2 \leqslant \sqrt{1+d}$, from Lemma A.3, with probability approach to 1, there exist $c_2 > 0$, such that

$$|tr\left((S-\Sigma_0)(I_n - G_0)^2\right)| \leqslant n\|(S-\Sigma_0)(I_n - G_0)^2\|_\infty \leqslant c_2 n(1+d)\sqrt{\frac{\log n}{T}}$$

When $d = o_P\left(\sqrt{T/(n\log n)}\right)$, $|tr\left((S-\Sigma_0)(I_n - G_0)^2\right)| = o_P(n)$. By Taylor expansion and mean value theorem, for some $\tilde{\sigma}_1^2$ between σ_0^2 and $\sigma_0^2 + \Delta\sigma^2$,

$$\Xi_G(\Delta\sigma^2) = \frac{n}{\sigma_0^2}\Delta\sigma^2 - \frac{n}{\tilde{\sigma}_1^4}(\Delta\sigma^2)^2 + \left(-\frac{1}{\sigma_0^4}\Delta\sigma^2 + \frac{2}{\tilde{\sigma}_1^6}(\Delta\sigma^2)^2\right)\cdot\left(n\sigma_0^2 + tr\left((S-\Sigma_0)(I_n - G_0)^2\right)\right)$$

and for constant $c_1 = -1/\tilde{\sigma}_1^4 + 2\sigma_0^2/\tilde{\sigma}_1^6$,

$$\Xi_G(\Delta\sigma^2) \geqslant nc_1(\Delta\sigma^2)^2 - c_2 n(1+d)\sqrt{\frac{\log n}{T}}\left|\Delta\sigma^2\right|$$

Lemma A.5. *Define* $\Xi_\sigma(\Delta)$ *as*

$$\begin{aligned}\Xi_\sigma(\Delta) = q(\sigma^2, G_0+\Delta) - q(\sigma^2, G_0) = \frac{1}{\sigma^2}tr\left(S(I_n - G_0 + \Delta)^2 - S(I_n - G_0)^2\right)\\ - 2(\log|(I_n - G_0 + \Delta)| - \log|(I_n - G_0)|)\end{aligned}$$

For all $\|\Delta\|_F \leqslant 1/(2L)$*, there exists constant* $c_3 > 0$ *and* $c_4 = 1/2(L + 1/(2L))^2 + 1/L^2$*, such that*

$$\Xi_\sigma(\Delta) > c_4\|\Delta\|_F^2 - c_3(1+d)\lambda_0\cdot\|\Delta\|_1 - \frac{2}{\sigma^2 L}|\Delta\sigma^2|\cdot\|\Delta\|_1$$

Proof. The first component is

$$tr\left(S(I_n - G_0 + \Delta)^2 - S(I_n - G_0)^2\right) = 2tr(\Delta S(I_n - G_0)) + vec(\Delta)'\cdot(I_n \otimes S)\cdot vec(\Delta)$$

A second order Taylor expansion with remainder in integral form yields

$$\log|(I_n - G_0 + \Delta)| - \log|(I_n - G_0)| = tr(\Delta(I_n - G_0)^{-1})$$
$$+ vec(\Delta)' \cdot \int_0^1 (1-v)(I_n - G_0 + v\Delta)^{-1} \otimes (I_n - G_0 + v\Delta)^{-1} dv \cdot vec(\Delta)$$

Thus

$$\Xi_\sigma(\Delta) = 2tr\left(\Delta\left(\frac{S(I_n - G_0)}{\sigma^2} - (I_n - G_0)^{-1}\right)\right) \tag{13}$$
$$+ vec(\Delta)' \cdot \left(I_n \otimes \frac{S}{\sigma^2}\right) \cdot vec(\Delta) \tag{14}$$
$$+ vec(\Delta)' \cdot \left(2\int_0^1 (1-v)(I_n - G_0 + v\Delta)^{-1} \otimes (I_n - G_0 + v\Delta)^{-1} dv\right) \cdot vec(\Delta) \tag{15}$$

First consider equation 14:

$$vec(\Delta)' \cdot \left(I_n \otimes \frac{S}{\sigma^2}\right) \cdot vec(\Delta) = vec(\Delta)' \cdot \left(I_n \otimes (I_n - G_0)^{-2}\right) \cdot vec(\Delta) \tag{16}$$
$$+ vec(\Delta)' \cdot \left(I_n \otimes \frac{S - \Sigma_0}{\sigma^2}\right) \cdot vec(\Delta) \tag{17}$$

On the event $\{\|S - \Sigma\|_\infty \leqslant \lambda\}$, equation 17 is in a smaller order than other term. As $\Lambda_{min}\left(I_n \otimes (I_n - G_0)^{-2}\right) \geqslant \Lambda_{max}^{-2}\left(I_n - G_0\right) \geqslant 1/L^2$, thus equation 14 is greater than $\|\Delta\|_F^2/L^2$. Next consider equation 15, for all $\|\Delta\|_F \leqslant 1/(2L)$, a standard argument as in Theorem 1 in Lam and Fan (2009) or Lemma 2 in Jankvoá and van de Geer (2019) yield

$$\Lambda_{min}\left(2\int_0^1 (1-v)(I_n - G_0 + v\Delta)^{-1} \otimes (I_n - G_0 + v\Delta)^{-1} dv\right) \geqslant \frac{1}{(L + 1/(2L))^2}$$

Finally, consider equation 13,

$$2tr\left(\Delta\left(\frac{S(I_n - G_0)}{\sigma^2} - (I_n - G_0)^{-1}\right)\right) = 2tr\left(\Delta\left(\frac{S}{\sigma^2} - \frac{\Sigma_0}{\sigma_0^2}\right) \cdot (I_n - G_0)\right)$$
$$= \frac{1}{\sigma^2} 2tr(\Delta(S - \Sigma_0) \cdot (I_n - G_0)) \tag{18}$$
$$+ \left(\frac{1}{\sigma^2} - \frac{1}{\sigma_0^2}\right) 2tr(\Delta\Sigma_0 \cdot (I_n - G_0)) \tag{19}$$

From Lemma A.3 for a constant $c_3 > 0$

$$\frac{2}{\sigma^2}|tr(\Delta(S - \Sigma_0) \cdot (I_n - G_0))| \leqslant \frac{1}{\sigma^2}\|(S - \Sigma_0) \cdot (I_n - G_0)\|_\infty \|\Delta\|_1 \leqslant c_3(1+d)\lambda_0$$

Given $\Sigma_0 \cdot (I_n - G_0) = \sigma_0^2 (I_n - G_0)^{-1}$, thus $\|\Sigma_0 \cdot (I_n - G_0)\|_\infty \leqslant \sigma_0^2 \Lambda_{max}((I_n - G_0)^{-1}) = \sigma_0^2/L$.

Thus equation 13 is boudned by

$$c_3(1+d)\lambda_0 \cdot \|\Delta\|_1 + \frac{2}{\sigma^2 L}|\Delta\sigma^2| \cdot \|\Delta\|_1$$

As a result, define $c_4 = 1/2(L + 1/(2L))^2 + 1/L^2$

$$\Xi_\sigma(\Delta) > c_4\|\Delta\|_F^2 - c_3(1+d)\lambda_0 \cdot \|\Delta\|_1 - \frac{2}{\sigma^2 L}|\Delta\sigma^2| \cdot \|\Delta\|_1$$

A.1 Proof of Theorem 3.1

Proof. From Lemma A.2, define $\hat{G}$ and $\hat{\sigma}^2$ as the solution to the minimization problem in equation 12. Then with the definition of $q(\sigma^2, G)$, $\Xi_G(\Delta\sigma^2)$ and $\Xi_\sigma(\Delta)$ in Lemma A.4 and A.5, we have

$$q(\hat{\sigma}^2, \hat{G}) + \lambda\|\hat{G}\|_1 \leqslant q(\sigma_0^2, G_0) + \lambda\|G_0\|_1$$

Thus,

$$\Xi_{\hat{\sigma}^2}(\Delta) + \Xi_{G_0}(\Delta\sigma^2) + \lambda\|\hat{G}\|_1 \leqslant +\lambda\|G_0\|_1$$

Therefore,

$$\begin{aligned} nc_1(\Delta\sigma^2)^2 + c_4\|\Delta\|_F^2 + \lambda\|\hat{G}\|_1 \leqslant{}& c_2 n(1+d)\lambda_0\left|\Delta\sigma^2\right| + c_3(1+d)\lambda_0 \cdot \|\Delta\|_1 \\ &+ \frac{2}{\hat{\sigma}^2 L}|\Delta\sigma^2| \cdot \|\Delta\|_1 + \lambda\|G_0\|_1 \end{aligned}$$

First notice that

$$\frac{2}{\hat{\sigma}^2 L}|\Delta\sigma^2| \cdot \|\Delta\|_1 \leqslant \frac{\alpha^2 n}{\hat{\sigma}^2}(\Delta\sigma^2)^2 + \frac{1}{\alpha^2 nL}\|\Delta\|_1^2 \leqslant \frac{\alpha^2 n}{\hat{\sigma}^4}(\Delta\sigma^2)^2 + \frac{1}{\alpha^2 L^2}\|\Delta\|_F^2$$

Since $c_1 = -1/\tilde{\sigma}_1^4 + 2\sigma_0^2/\tilde{\sigma}_1^6$ and $c_4 = 1/2(L + 1/(2L))^2 + 1/L^2 > 1/(\alpha^2 L^2)$ for $\alpha^2 < 1$, for all $\Delta\sigma^2 \leqslant \min\{(1-\alpha)\sigma_0^2, (1-\alpha^2)/(1+\alpha^2)\sigma_0^2\}$, we have

$$\tilde{c}_1 = c_1 - \frac{\alpha^2}{\hat{\sigma}^4} > 0 \text{ and } \tilde{c}_4 = c_4 - \frac{1}{\alpha^2 L^2} > 0$$

by triangular inequality and take $\lambda > 2c_3(1+d)\lambda_0$

$$2n\tilde{c}_1(\Delta\sigma^2)^2 + 2\tilde{c}_4\|\Delta\|_F^2 + \lambda\|\hat{G}_{S^c}\|_1 \leqslant 2c_2 n(1+d)\lambda_0\left|\Delta\sigma^2\right| + 3\lambda\|\Delta_S\|_1$$

Add and subtract both side with $\lambda\|\Delta_S\|_1$ and

$$\begin{aligned} 2n\tilde{c}_1(\Delta\sigma^2)^2 + 2\tilde{c}_4\|\Delta\|_F^2 + \lambda\|\Delta\|_1 &\leqslant 4\lambda\sqrt{s_n}\|\Delta\|_F + 2c_2 n(1+d)\lambda_0\left|\Delta\sigma^2\right| \\ &\leqslant 4\frac{s_n\lambda^2}{\tilde{c}_4} + \tilde{c}_4\|\Delta\|_F^2 + \frac{4c_2^2 n(1+d)^2\lambda_0^2}{\tilde{c}_1} + n\tilde{c}_1(\Delta\sigma^2)^2 \end{aligned}$$

which implies

$$n\tilde{c}_1(\Delta\sigma^2)^2 + \tilde{c}_4\|\Delta\|_F^2 + \lambda\|\Delta\|_1 \leqslant 4\frac{s_n\lambda^2}{\tilde{c}_4} + \frac{4c_2^2 n(1+d)^2\lambda_0^2}{\tilde{c}_1}$$

Thus $|\Delta\sigma^2| = O_P(\sqrt{(s_0+n)/n}\cdot d\lambda_0)$, $\|\Delta\|_F = O_P(\sqrt{(s_0+n)}\cdot d\lambda_0)$ and $\|\Delta\|_1 = O_P((s_0 + nd)\cdot\lambda_0)$

A.2 Proof of Theorem 3.2

The KKT condition on equation 12 gives

$$(I_n - \hat{G})^{-1} - \frac{S}{\hat{\sigma}^2}(I_n - \hat{G}) + \frac{\lambda}{2}\hat{Z} = 0$$

$$\frac{n}{\hat{\sigma}^2} - \frac{1}{\hat{\sigma}^4}tr\left(S(I_n - \hat{G})^2\right) = 0$$

where $\hat{Z}_{ij} = sign(\hat{G}_{ij})$. Since $\Sigma_0(I - G_0)^2/\sigma_0^2 = I_n$,

$$(I_n - \hat{G})^{-1} - (I_n - G_0)^{-1} + \frac{\Sigma_0}{\sigma_0^2}(I_n - G_0) - \frac{S}{\hat{\sigma}^2}(I_n - \hat{G}) + \frac{\lambda}{2}\hat{Z} = 0$$

which implies

$$\begin{aligned}(I_n - \hat{G})^{-1} - (I_n - G_0)^{-1} + \frac{\Sigma_0}{\sigma_0^2}(\hat{G} - G_0) \\ = -\left(\frac{\Sigma_0}{\sigma_0^2} - \frac{S}{\hat{\sigma}^2}\right)(I_n - G_0) - \frac{\lambda}{2}\hat{Z} - \left(\frac{\Sigma_0}{\sigma_0^2} - \frac{S}{\hat{\sigma}^2}\right)(G_0 - \hat{G})\end{aligned} \tag{20}$$

By the symmetric of $(I_n - G_0)^{-1}$, we have $(I_n - \hat{G})^{-1} - (I_n - G_0)^{-1} = (I_n - G_0)^{-1}\cdot(I_n - \hat{G})^{-1}\cdot(\hat{G} - G_0)$. Equation (20) can be written as

$$\begin{aligned}(\hat{G} - G_0) = &-\frac{1}{2}(I_n - \hat{G})(I_n - G_0)\left(\frac{\Sigma_0}{\sigma_0^2} - \frac{S}{\hat{\sigma}^2}\right)(I_n - G_0) \\ &- \frac{\lambda}{4}(I_n - \hat{G})(I_n - G_0)\hat{Z} - \frac{1}{2}(I_n - \hat{G})(I_n - G_0)\left(\frac{\Sigma_0}{\sigma_0^2} - \frac{S}{\hat{\sigma}^2}\right)(G_0 - \hat{G})\end{aligned}$$

Thus we can define the debias estimator as

$$\tilde{G} = \hat{G} + \frac{1}{2}(I_n - \hat{G})^2\left(-I_n + \frac{S}{\hat{\sigma}^2}(I_n - \hat{G})^2\right)$$

A decomposition of $\tilde{G} - G_0$ yields

$$(\tilde{G} - G_0) = -\frac{1}{2}(I_n - G_0)^2 \left(\frac{\Sigma_0}{\sigma_0^2} - \frac{S}{\hat{\sigma}^2}\right)(I_n - G_0) + \underbrace{\frac{1}{2}(\hat{G} - G_0)(I_n - G_0)\left(\frac{\Sigma_0}{\sigma_0^2} - \frac{S}{\hat{\sigma}^2}\right)(I_n - G_0)}_{R1}$$
$$- \underbrace{\frac{\lambda}{4}(I_n - \hat{G})\Big(\hat{G} - G_0\Big)\hat{Z}}_{R2} - \underbrace{\frac{1}{2}(I_n - \hat{G})(I_n - G_0)\left(\frac{\Sigma_0}{\sigma_0^2} - \frac{S}{\hat{\sigma}^2}\right)(G_0 - \hat{G})}_{R3}$$

From Lemma A.4 and theorem 3.1

$$\left\|\frac{\Sigma_0}{\sigma_0^2} - \frac{S}{\hat{\sigma}^2}\right\|_\infty \leqslant \left\|\frac{\Sigma_0}{\sigma_0^2} - \frac{S}{\sigma_0^2}\right\|_\infty + \left\|\frac{S}{\sigma_0^2} - \frac{S}{\hat{\sigma}^2}\right\|_\infty \lesssim \sqrt{(s_n + nd^2)/n} \cdot \lambda$$

Further define $\|\cdot\|_{op1}$ as the operator norm from $l_\infty \to l_\infty$, which is the maximum l_1 norm of a row. Since $\|\hat{G} - G_0\|_{op1} \leqslant \sqrt{n}\|\hat{G} - G_0\|_F$ and $\|I_n - G_0\|_{op} \leqslant (1 + d)$

$$\|R_1\|_\infty \leqslant \frac{1}{2}\|\hat{G} - G_0\|_{op1} \cdot \|I_n - G_0\|_{op1}^2 \cdot \left\|\frac{\Sigma_0}{\sigma_0^2} - \frac{S}{\hat{\sigma}^2}\right\|_\infty = O_P((s_0 + n)d^4\lambda_0^2)$$

$$\begin{aligned}
\|R_2\|_\infty &\leqslant \frac{\lambda}{4}\|(I_n - G_0)(\hat{G} - G_0)\hat{Z}\|_\infty + \frac{\lambda}{4}\|(\hat{G} - G_0)^2\hat{Z}\|_\infty \\
&\leqslant \frac{\lambda}{4}\|I_n - \hat{G}\|_{op1}\|\hat{G} - G_0\|_{op1}\|\hat{Z}\|_\infty + \frac{\lambda}{4}\|\hat{G} - G_0\|_{op1}^2\|\hat{Z}\|_\infty \\
&= O_P(\sqrt{n(s_0 + n)}d^2\lambda_0^2 + n(s_0 + n)d^3\lambda_0^3)
\end{aligned}$$

$$\begin{aligned}
\|R_3\|_\infty \leqslant& \frac{1}{2}\left\|(I_n - G_0)^2\left(\frac{\Sigma_0}{\sigma_0^2} - \frac{S}{\hat{\sigma}^2}\right)(G_0 - \hat{G})\right\|_\infty + \frac{1}{2}\left\|(G_0 - \hat{G})(I_n - G_0)\left(\frac{\Sigma_0}{\sigma_0^2} - \frac{S}{\hat{\sigma}^2}\right)(G_0 - \hat{G})\right\|_\infty \\
&\leqslant \frac{1}{2}\|(I_n - G_0)\|_{op1}^2\left\|\frac{\Sigma_0}{\sigma_0^2} - \frac{S}{\hat{\sigma}^2}\right\|_\infty\|G_0 - \hat{G}\|_{op1} + \frac{1}{2}\|(I_n - G_0)\|_{op1}\left\|\frac{\Sigma_0}{\sigma_0^2} - \frac{S}{\hat{\sigma}^2}\right\|_\infty\|G_0 - \hat{G}\|_{op1}^2 \\
&= O_P((s_0 + n)d^4\lambda_0^2 + \sqrt{n}(s_0 + n)^{3/2}d^3\lambda_0^3)
\end{aligned}$$

Thus when taking $\lambda_0 \gtrsim \sqrt{\log n/T}$ and $s_0 d^4 = o_P(\sqrt{T}/\log n \vee T/(n(\log n)^2))$ and $d = o_P\left((\sqrt{T}/(n\log n))^{1/4}\right)$

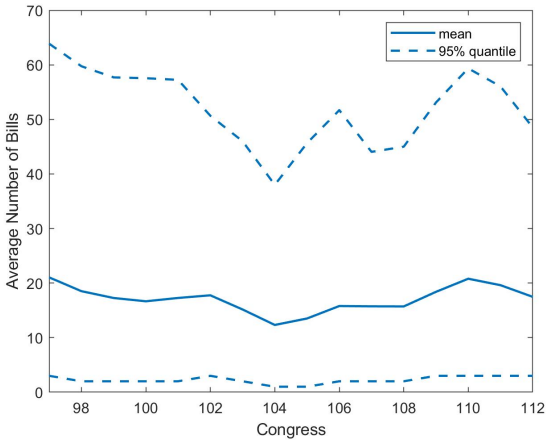
$$(\tilde{G} - G_0) = -\frac{1}{2}(I - G_0)^2\left(\frac{\Sigma_0}{\sigma_0^2} - \frac{S}{\hat{\sigma}^2}\right)(I - G_0) + o_P(1/\sqrt{T})$$

B Additional Results

Table B4: Average Link Strength (×0.01)

Congress	Majority	R	D	between
93	D	2.89	2.44	2.40
94	D	3.75	4.36	3.99
95	D	0.90	1.08	0.86
96	D	1.12	1.39	0.98
97	D	0.73	1.76	0.88
98	D	1.17	2.20	1.34
99	D	0.51	1.19	0.67
100	D	1.32	2.37	1.61
101	D	1.21	1.47	1.09
102	D	0.91	2.37	1.40
103	D	1.23	1.42	0.90
104	R	2.97	1.25	1.82
105	R	2.99	1.15	1.54
106	R	3.36	1.27	2.13
107	R	2.82	0.82	1.45
108	R	3.50	1.26	1.93
109	R	4.08	1.87	2.76
110	D	2.02	6.00	3.19
111	D	1.79	6.81	2.96
112	R	1.69	0.73	1.13

Figure 6: Bills in the U.S. Congress



This figure plots the average number of bills proposed in the House of Representatives, Congress 97th-112nd. Source: `https://thelawmakers.org/data-download`