Learning Sparse Potential Games in Polynomial Time and Sample Complexity

Appendix A Detailed Proofs

Proof of Lemma 1 (Minimum eigenvalue of population Hessian).
Fix any $\boldsymbol{\theta}^0, \boldsymbol{\theta}^1 \in \Theta^i$, with $\boldsymbol{\theta}^1 \neq \mathbf{0}$. For any $t \in (-\infty, \infty)$, let $F(t; x_i) \overset{\text{def}}{=} (\boldsymbol{\theta}^0 + t\boldsymbol{\theta}^1)^T \mathbf{f}(x_i, \mathbf{x}_{-i})$. Then for $\mathbf{x} \in \mathcal{A}$,

$$\ell(\mathbf{x}; \boldsymbol{\theta}^0 + t\boldsymbol{\theta}^1) = -F(t; x_i) + \log(\sum_{a \in \mathcal{A}_i} \exp(F(t; a))). \tag{18}$$

A little calculation shows that the double derivative of $\ell(\mathbf{x}; \boldsymbol{\theta}^0 + t\boldsymbol{\theta}^1)$ with respect to t is as follows:

$$\begin{aligned} \frac{\partial^2 \ell(\mathbf{x}; \boldsymbol{\theta}^0 + t\boldsymbol{\theta}^1)}{\partial t^2} &= \sum_{a \in \mathcal{A}_i} \sigma(t; a) F'(a)^2 - \Big(\sum_{a \in \mathcal{A}_i} \sigma(t; a) F'(a)\Big)^2, \\ \sigma(t; b) &= \frac{\exp(F(t; b))}{\sum_{a \in \mathcal{A}_i} \exp(F(t; a))}, \; (b \in \mathcal{A}_i) \end{aligned} \tag{19}$$

where $F'(a)$ is the derivative of $F(t; a)$ with respect to t. Since $F(t; a)$ is a linear function of t, $F'(a)$ is not a function of t. Also note that $\sum_{a \in \mathcal{A}_i} \sigma(t; a) = 1$. Since $\boldsymbol{\theta}^0, \boldsymbol{\theta}^1$ have bounded norm and $t \in (-\infty, \infty)$, we have that $\sigma(t; a) > 0, \forall a \in \mathcal{A}_i$. Therefore, from (19), the strict convexity of $(\cdot)^2$ and Jensen's inequality, we have:

$$\frac{\partial^2 \ell(\mathbf{x}; \boldsymbol{\theta}^0 + t\boldsymbol{\theta}^1)}{\partial t^2} > 0 \qquad\qquad (\forall t \in (-\infty, \infty)).$$

Thus we have that $\ell(\mathbf{x}, \boldsymbol{\theta})$ is strongly convex, i.e., $\lambda_{\min}(\mathbf{H}^i(\mathbf{x}; \boldsymbol{\theta})) > 0$, $\forall \boldsymbol{\theta} \in \Theta^i$. Finally, by concavity of $\lambda_{\min}(\cdot)$ [BV04] and the Jensen's inequality we have:

$$\lambda_{\min}(\mathbf{H}^i(\boldsymbol{\theta}^i)) = \lambda_{\min}(\mathbb{E}_{\mathbf{x}}\left[\mathbf{H}^i(\mathbf{x}; \boldsymbol{\theta}^i)\right]) \geq \mathbb{E}_{\mathbf{x}}\left[\lambda_{\min}(\mathbf{H}^i(\mathbf{x}; \boldsymbol{\theta}^i))\right] > 0.$$

□

Proof of Lemma 2 (Minimum eigenvalue of finite sample Hessian).
To simply notation in the proof we will denote S_i by S. The (j, k) block of $\mathbf{H}(\mathcal{D}; \boldsymbol{\theta}_S)$, where $j, k \in \{0\} \cup \mathcal{N}_i$, can be written as:

$$\mathbf{H}_{j,k}(\mathcal{D}; \boldsymbol{\theta}_S) = \underbrace{\sum_{l=1}^{n} \sum_{a \in \mathcal{A}_i} \sigma^i(a, \mathbf{x}_{-i}^{(l)}; \boldsymbol{\theta}_S) \mathbf{f}^{i,j}(a, x_j^{(l)})(\mathbf{f}^{i,k}(a, x_k^{(l)}))^T}_{\mathbf{B}_{j,k}(\mathcal{D}; \boldsymbol{\theta}_S)} - \underbrace{\sum_{l=1}^{n} \sum_{a,b \in \mathcal{A}_i} \sigma^i(a, \mathbf{x}_{-i}^{(l)}; \boldsymbol{\theta}_S) \mathbf{f}^{i,j}(a, x_j^{(l)}) \mathbf{f}^{i,k}(b, x_k^{(l)})^T}_{\mathbf{R}_{j,k}(\mathcal{D}; \boldsymbol{\theta}_S)},$$

where the matrices $\mathbf{B}$ and $\mathbf{R}$ have been defined above (blockwise). Since the matrix $\mathbf{R}$ is positive semi-definite $\lambda_{\max}(\mathbf{H}(\mathcal{D}; \boldsymbol{\theta}_S)) \leq \lambda_{\max}(\mathbf{B}(\mathcal{D}; \boldsymbol{\theta}_S))$. Further, since $\mathbf{B}$ is positive semi-definite, we have, from Lemma 7:

$$\begin{aligned} \lambda_{\max}(\mathbf{B}(\mathcal{D}; \boldsymbol{\theta}_S)) &\leq \sum_{j \in \{0\} \cup \mathcal{N}_i} \lambda_{\max}(\mathbf{B}_{j,j}(\mathcal{D}; \boldsymbol{\theta}_S)) \\ &\leq (d_i + 1) \max_{j \in \{0\} \cup \mathcal{N}_i} \lambda_{\max}\Big(\frac{1}{n} \sum_{l=1}^{n} \sum_{a \in \mathcal{A}_i} \sigma^i(a, \mathbf{x}_{-i}^{(l)}; \boldsymbol{\theta}_S) \mathbf{f}^{i,j}(a, x_j^{(l)})(\mathbf{f}^{i,j}(a, x_j^{(l)}))^T\Big) \\ &\leq \frac{(d_i + 1)}{n} \max_{j \in \{0\} \cup \mathcal{N}_i} \sum_{l=1}^{n} \sum_{a \in \mathcal{A}_i} \sigma^i(a, \mathbf{x}_{-i}^{(l)}; \boldsymbol{\theta}_S) \lambda_{\max}\left(\mathbf{f}^{i,j}(a, x_j^{(l)})(\mathbf{f}^{i,j}(a, x_j^{(l)}))^T\right) \\ &= d_i + 1. \end{aligned}$$

Thus we have that $\lambda_{\max}(\mathbf{H}(\mathcal{D}; \boldsymbol{\theta}_S)) \leq \lambda_{\max}(\mathbf{B}(\mathcal{D}; \boldsymbol{\theta}_S)) \leq d_i + 1 \overset{\text{def}}{=} R$. Also note that $\mathbf{H}(\mathcal{D}; \boldsymbol{\theta}_S) \in \mathbb{R}^{|S| \times |S|}$, with $|S| \leq m_i(1 + d_i m)$. Then using the matrix Chernoff bounds by [Tro12], we have:

$$\Pr\left\{\lambda_{\min}(\mathbf{H}(\mathcal{D}; \boldsymbol{\theta}_S)) \leq (1 - \delta)\lambda_{\min}\right\} \leq |S| \left(\frac{\exp(-\delta)}{(1 - \delta)^{(1-\delta)}}\right)^{(n\lambda_{\min}/R)}$$

Setting $\delta = {}^{1}\!/_{2}$ we get:

$$\Pr\left\{\lambda_{\min}(\mathbf{H}(\mathcal{D};\boldsymbol{\theta}_S)) \geq \frac{\lambda_{\min}}{2}\right\} \geq 1 - m_i(1+d_i m)\exp\left(-\frac{n\lambda_{\min}}{8(d_i+1)}\right)$$

Controlling the probability of error to be at most δ we obtain the lower bound on the number of samples. □

Proof of Lemma 3 (Gradient bound).
A simple calculation shows that

$$\frac{\partial \ell^i(\mathbf{x};\boldsymbol{\theta}^i)}{\partial \boldsymbol{\theta}^{i,j}} = -\mathbf{f}^{i,j}(x_i,x_j) + \sum_{a\in\mathcal{A}_i}\sigma^i(a,\mathbf{x}_{-i};\boldsymbol{\theta}^i)\mathbf{f}^{i,j}(a,x_j), \tag{20}$$

where $\sigma^i(\cdot)$ has been defined in (12). Let $\mathbf{g}(\mathbf{x}^{(1)},\ldots,\mathbf{x}^{(n)}) = (g_j(\mathbf{x}^{(1)},\ldots,\mathbf{x}^{(n)}))_{j\in\{0\}\cup\mathcal{N}_i}$, where $g_j(\cdot) = \|\frac{1}{n}\sum_{l=1}^{n}\frac{\partial \ell^i(\mathbf{x}^{(l)};\boldsymbol{\theta}^i)}{\partial\boldsymbol{\theta}^{i,j}}\|_2$. Then $\|\mathbf{g}(\cdot)\|_\infty = \|\nabla L^i(\mathcal{D};\boldsymbol{\theta}^i)\|_{\infty,2}$ and $\|\mathbb{E}_{\mathbf{x}}[\mathbf{g}(\cdot)]\|_\infty = \|\mathbb{E}_{\mathbf{x}}\left[\nabla\ell^i(\mathbf{x};\boldsymbol{\theta}^i)\right]\|_{\infty,2} = \nu$. Then, for any $\mathbf{x}^{(l)} \neq \mathbf{x}^{(l)'}$ we have that:

$$\begin{aligned}
&|g_j(\mathbf{x}^{(1)},\ldots,\mathbf{x}^{(l)},\ldots,\mathbf{x}^{(n)}) - g_j(\mathbf{x}^{(1)},\ldots,\mathbf{x}^{(l)'},\ldots,g_j(\mathbf{x}^{(n)})|\\
&= \frac{1}{n}\left\|\mathbf{f}^{i,j}(x_i^{(l)'},x_j^{(l)'}) - \mathbf{f}^{i,j}(x_i^{(l)},x_j^{(l)}) + \sum_{a\in\mathcal{A}_i}\sigma^i(a,\mathbf{x}_{-i}^{(l)};\boldsymbol{\theta}^i)\mathbf{f}^{i,j}(a,x_j^{(l)}) - \sigma^i(a,\mathbf{x}_{-i}^{(l)'};\boldsymbol{\theta}^i)\mathbf{f}^{i,j}(a,x_j^{(l)'})\right\|_2\\
&\leq \frac{1}{n}\Big(2 + \sum_{a\in\mathcal{A}_i}(\sigma^i(a,\mathbf{x}_{-i}^{(l)};\boldsymbol{\theta}^i))^2 + (\sigma^i(a,\mathbf{x}_{-i}^{(l)'};\boldsymbol{\theta}^i))^2\Big)^{1/2} \leq \frac{1}{n}(2+2)^{1/2} = {}^{2}\!/_{n},
\end{aligned}$$

where in the last line we used the fact that $\sum_a \sigma^i(a,\cdot) = 1$ along with the Cauchy-Schwartz inequality. Then using the McDiarmid's inequality we have:

$$\Pr\{|g_j(\cdot) - \mathbb{E}_{\mathbf{x}}[g_j(\cdot)]| \leq t\} \geq 1 - 2\exp\Big(\frac{-nt^2}{2}\Big).$$

Then using a union bound over all j we have:

$$\begin{aligned}
\Pr\left\{\max_j|g_j(\cdot) - \mathbb{E}_{\mathbf{x}}[g_j(\cdot)]| \leq t\right\} &\geq 1 - 2(d_i+1)\exp\Big(\frac{-nt^2}{2}\Big)\\
\implies \Pr\{\|\mathbf{g}(\cdot) - \mathbb{E}_{\mathbf{x}}[\mathbf{g}(\cdot)]\|_\infty \leq t\} &\geq 1 - 2(d_i+1)\exp\Big(\frac{-nt^2}{2}\Big)\\
\implies \Pr\{\|\mathbf{g}(\cdot)\|_\infty - \|\mathbb{E}_{\mathbf{x}}[\mathbf{g}(\cdot)]\|_\infty \leq t\} &\geq 1 - 2(d_i+1)\exp\Big(\frac{-nt^2}{2}\Big)\\
\implies \Pr\{\|\mathbf{g}(\cdot)\|_\infty \leq \nu + t\} &\geq 1 - 2(d_i+1)\exp\Big(\frac{-nt^2}{2}\Big),
\end{aligned}$$

where in the third line we used the reverse triangle inequality. Setting the probability of error to be δ and solving for t, we prove our claim. □

Proof of Lemma 4 (Minimum population eigenvalue at arbitrary parameter).
To simply notation in the proof we will denote S_i by S. The population Hessian matrix at $\mathbf{H}(\boldsymbol{\theta}_S)$ can also be written as $\mathbf{H}(\boldsymbol{\theta}_S^i + \boldsymbol{\Delta}_S)$, where $\boldsymbol{\Delta}_S = \boldsymbol{\theta}_S - \boldsymbol{\theta}_S^i$. Using the variational characterization of the minimum eigenvalue of $\mathbf{H}(\boldsymbol{\theta}_S^i + \boldsymbol{\Delta}_S)$ and the Taylor's theorem, we have:

$$\begin{aligned}
\lambda_{\min}(\mathbf{H}(\boldsymbol{\theta}_S^i + \boldsymbol{\Delta}_S)) &= \min_{\{\mathbf{y}\in\mathbb{R}^{|S|}\,|\,\|\mathbf{y}\|_2=1\}}\sum_{i,j\in S} y_i\{H_{i,j}(\boldsymbol{\theta}_S^i) + (\nabla H_{i,j}(\bar{\boldsymbol{\theta}}_S))^T\boldsymbol{\Delta}_S\}y_j\\
&\geq \lambda_{\min}(\mathbf{H}(\boldsymbol{\theta}_S^i)) - \max_{\{\mathbf{y}\in\mathbb{R}^{|S|}\,|\,\|\mathbf{y}\|_2=1\}}\sum_{i,j\in S} y_i\{(\nabla H_{i,j}(\bar{\boldsymbol{\theta}}_S))^T\boldsymbol{\Delta}_S\}y_j\\
&\geq \lambda_{\min}(\mathbf{H}(\boldsymbol{\theta}_S^i)) - \max_{\{\mathbf{y}\in\mathbb{R}^{|S|}\,|\,\|\mathbf{y}\|_2=1\}}\sum_{i,j\in S} y_i\{|(\nabla H_{i,j}(\bar{\boldsymbol{\theta}}_S))^T\boldsymbol{\Delta}_S|\}y_j,
\end{aligned} \tag{21}$$

where $\bar{\boldsymbol{\theta}} = t\boldsymbol{\theta}^i_S + (1-t)\boldsymbol{\theta}_S$ for some $t \in [0,1]$, and the third line follows from the monotonicity property of the spectral norm $|||\cdot|||_2$ [JN91]. For any vector $\boldsymbol{\theta} \in \Theta^i$, let $\mathbf{A}(\boldsymbol{\theta}_S) = (A_{i,j}(\boldsymbol{\theta}_S))$, where $A_{i,j}(\boldsymbol{\theta}_S) = |(\nabla H_{i,j}(\boldsymbol{\theta}_S))^T \boldsymbol{\Delta}_S|$. Then,

$$|||\mathbf{A}(\boldsymbol{\theta}_S)|||_2 = |||\mathbb{E}_{\mathbf{x}}\left[\mathbf{A}(\mathbf{x};\boldsymbol{\theta}_S)\right]|||_2 \le \max_{\mathbf{x}\in\mathcal{A}} |||\mathbf{A}(\mathbf{x};\boldsymbol{\theta}_S)|||_2. \tag{22}$$

Now consider the (j,k) block of $\mathbf{A}(\mathbf{x};\boldsymbol{\theta}_S)$ for any $\mathbf{x} \in \mathcal{A}$, where $j,k \in \{0\} \cup \mathcal{N}_i$. Then, from (11) we have that:

$$\begin{aligned}\mathbf{A}_{j,k}(\mathbf{x};\boldsymbol{\theta}) = &\underbrace{\sum_{a\in\mathcal{A}_i} |(\nabla\sigma^i(a,\mathbf{x}_{-i};\boldsymbol{\theta}))^T\boldsymbol{\Delta}_S|\mathbf{f}^{i,j}(a,x_j)(\mathbf{f}^{i,k}(a,x_k))^T}_{\mathbf{B}_{j,k}(\mathbf{x};\boldsymbol{\theta})} \\ &- \underbrace{\sum_{a,b\in\mathcal{A}_i} |\left\{\sigma^i(b,\mathbf{x}_{-i};\boldsymbol{\theta})\nabla\sigma^i(a,\mathbf{x}_{-i};\boldsymbol{\theta}) + \sigma^i(a,\mathbf{x}_{-i};\boldsymbol{\theta})\nabla\sigma^i(b,\mathbf{x}_{-i};\boldsymbol{\theta})\right\}^T\boldsymbol{\Delta}_S|\mathbf{f}^{i,j}(a,x_j)\mathbf{f}^{i,k}(a,x_k)^T}_{\mathbf{R}_{j,k}(\mathbf{x};\boldsymbol{\theta})}.\end{aligned}$$

Thus, $\mathbf{A}(\mathbf{x};\boldsymbol{\theta}) = \mathbf{B}(\mathbf{x};\boldsymbol{\theta}) - \mathbf{R}(\mathbf{x};\boldsymbol{\theta})$, where the matrices $\mathbf{B}$ and $\mathbf{R}$ have been defined above (block-wise). Observe that the matrix $\mathbf{R}$ is positive semi-definite. Therefore, $|||\mathbf{A}(\mathbf{x};\boldsymbol{\theta})|||_2 \le |||\mathbf{B}(\mathbf{x};\boldsymbol{\theta})|||_2$. Finally, since $\mathbf{B}$ is positive semi-definite, the spectral norm of $\mathbf{B}$ is at most the sum of the spectral norms of the diagonal blocks (c.f. Lemma 7). Therefore, we have

$$|||\mathbf{B}(\mathbf{x};\boldsymbol{\theta})|||_2 \le \sum_{j\in\{0\}\cup\mathcal{N}_i} |||\mathbf{B}_{j,j}(\mathbf{x};\boldsymbol{\theta})|||_2 \le (d_i+1)\Big(\max_{j\in\{0\}\cup\mathcal{N}_i} |||\mathbf{B}_{j,j}(\mathbf{x};\boldsymbol{\theta})|||_2\Big). \tag{23}$$

A little calculation shows that

$$\frac{\partial\sigma^i(a,\mathbf{x}_{-i};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_j} = \sigma^i(a,\mathbf{x}_{-i};\boldsymbol{\theta})\Big\{\mathbf{f}^{i,j}(a,x_j) - \sum_{a'\in\mathcal{A}_i}\sigma^i(a',\mathbf{x}_{-i};\boldsymbol{\theta})\mathbf{f}^{i,j}(a',x_j)\Big\},$$

and $\|{}^{\partial\sigma^i(a,\mathbf{x}_{-i};\boldsymbol{\theta})}/_{\partial\boldsymbol{\theta}_j}\|_\infty \le {}^1/_4$. Further, since for any given $a \in \mathcal{A}_i$, at most m_j+1 elements of the partial derivative vector above is non-zero, we have $\|{}^{\partial\sigma^i(a,\mathbf{x}_{-i};\boldsymbol{\theta})}/_{\partial\boldsymbol{\theta}_j}\|_2 \le {}^{(m_j+1)}/_4$ and $\|\nabla\sigma^i(a,\mathbf{x}_{-i};\boldsymbol{\theta})\|_{\infty,2} \le {}^{(m_j+1)}/_4 \le {}^{(m+1)}/_4$. Then using the Cauchy-Schwartz inequality and the monotonicity property of spectral norm [JN91] we have:

$$\begin{aligned}|||\mathbf{B}_{j,j}(\mathbf{x};\boldsymbol{\theta})|||_2 &\le \left|\left|\left|\sum_{a\in\mathcal{A}_i}\|\nabla\sigma^i(a,\mathbf{x}_{-i};\boldsymbol{\theta}))\|_{\infty,2}\|\boldsymbol{\Delta}_S\|_{1,2}\mathbf{f}^{i,j}(a,x_j)(\mathbf{f}^{i,j}(a,x_j))^T\right|\right|\right|_2 \\ &\le \frac{1}{4}m_i m\|\boldsymbol{\Delta}_S\|_{1,2}\end{aligned} \tag{24}$$

Putting together (21), (22), (23) and (24) we get

$$\begin{aligned}\lambda_{\min}(\mathbf{H}(\boldsymbol{\theta}^i_S + \boldsymbol{\Delta}_S)) &\ge \lambda_{\min}(\mathbf{H}(\boldsymbol{\theta}^i_S)) - |||\mathbf{A}(\boldsymbol{\theta}_S)|||_2 \\ &\ge \lambda_{\min}(\mathbf{H}(\boldsymbol{\theta}^i_S)) - (d_i+1)\Big(\max_{\mathbf{x}\in\mathcal{A}}\max_{j\in\{0\}\cup\mathcal{N}_i}|||\mathbf{B}_{j,j}(\mathbf{x};\boldsymbol{\theta})|||_2\Big) \\ &\ge \lambda_{\min}(\mathbf{H}(\boldsymbol{\theta}^i_S)) - \frac{1}{4}(d_i+1)m_i m\|\boldsymbol{\Delta}_S\|_{1,2}.\end{aligned}$$

□

Proof of Lemma 5 (Error of the i-th estimator on the support set).
To simplify notation in the proof, we will write S instead of S_i. Recall that $L^i(\mathcal{D};\boldsymbol{\theta})$ is the empirical loss for the i-th player for parameter $\boldsymbol{\theta}$. For the purpose of the proof we will often write $L(\boldsymbol{\theta})$ instead of $L^i(\mathcal{D};\boldsymbol{\theta})$. Let $F(\boldsymbol{\theta}) = L(\boldsymbol{\theta}) + \lambda\|\boldsymbol{\theta}\|_{1,2}$. For any $\boldsymbol{\theta} \in \Theta^i$, let $\boldsymbol{\Delta}_S = \boldsymbol{\theta}_S - \boldsymbol{\theta}^i_S$ denote the difference between $\boldsymbol{\theta}$ and the true parameter $\boldsymbol{\theta}^i$ on the true support set S. We introduce the following shifted and reparameterized regularized loss function:

$$\widetilde{F}(\boldsymbol{\Delta}_S) = \underbrace{L(\boldsymbol{\theta}^i_S + \boldsymbol{\Delta}_S) - L(\boldsymbol{\theta}^i_S)}_{\text{term 1}} + \underbrace{\lambda(\|\boldsymbol{\theta}^i_S + \boldsymbol{\Delta}_S\|_{1,2} - \|\boldsymbol{\theta}^i_S\|_{1,2})}_{\text{term 2}}, \tag{25}$$

which takes the value 0 at the true parameter $\boldsymbol{\theta}^i$, i.e., $\widetilde{F}(\mathbf{0}) = 0$. Let $\widehat{\boldsymbol{\Delta}}_S = \widehat{\boldsymbol{\theta}}^i_S - \boldsymbol{\theta}^i_S$, where $\widehat{\boldsymbol{\theta}}^i$ minimizes $F(\boldsymbol{\theta})$. Since $\widehat{\boldsymbol{\theta}}^i$ minimizes $F(\boldsymbol{\theta})$, we must have that $\widetilde{F}(\widehat{\boldsymbol{\Delta}}_S) \leq 0$. Thus, in order to upper bound $\|\widehat{\boldsymbol{\Delta}}_S\|_{1,2} = \|\widehat{\boldsymbol{\theta}}_S - \boldsymbol{\theta}^i_S\|_{1,2} \leq b$, we show that there exists an $\ell_{1,2}$ ball of radius b such that function $\widetilde{F}(\boldsymbol{\Delta}_S)$ is strictly positive on the surface of the ball. To see this, assume the contrary, i.e, $\forall \boldsymbol{\Delta} \in \Theta^i \wedge \|\boldsymbol{\Delta}_S\|_{1,2} = b$, $\widetilde{F}(\boldsymbol{\Delta}_S) > 0$, but $\widehat{\boldsymbol{\Delta}}_S$ lies outside the ball, i.e., $\|\widehat{\boldsymbol{\Delta}}_S\|_{1,2} > b$. Then, there exists a $t \in (0, 1)$ such that $(1-t)\mathbf{0} + t\widehat{\boldsymbol{\Delta}}_S$ lies on the surface of the ball, i.e., $\|(1-t)\mathbf{0} + t\widehat{\boldsymbol{\Delta}}_S\|_{1,2} = b$. However, by convexity of $\widetilde{F}$ we have that

$$0 < \widetilde{F}((1-t)\mathbf{0} + t\widehat{\boldsymbol{\Delta}}_S) \leq (1-t)\widetilde{F}(\mathbf{0}) + t\widetilde{F}(\widehat{\boldsymbol{\Delta}}_S) = t\widetilde{F}(\widehat{\boldsymbol{\Delta}}_S),$$

which implies that $\widetilde{F}(\widehat{\boldsymbol{\Delta}}_S) > 0$ and therefore is a contradiction to the fact that $\widetilde{F}(\widehat{\boldsymbol{\Delta}}_S) \leq 0$. Going forward, our strategy would be to lower bound $\widetilde{F}(\boldsymbol{\Delta}_S)$ in terms of $\|\boldsymbol{\Delta}_S\|_{1,2} = b$. We then set the lower bound to 0 and solve for b, to obtain the radius of the $\ell_{1,2}$ ball on which the function is non-negative. Towards that end we first lower bound the first term of (25).

Using the Taylor's theorem and the Cauchy-Schwartz inequality, for some $t \in [0, 1]$, we have:

$$\begin{aligned}
&L(\boldsymbol{\theta}^i_S + \boldsymbol{\Delta}_S) - L(\boldsymbol{\theta}^i_S) \\
&\quad = \nabla L(\boldsymbol{\theta}^i_S)^T \boldsymbol{\Delta}_S + \boldsymbol{\Delta}_S^T \nabla^2 L(\boldsymbol{\theta}^i_S + t\boldsymbol{\Delta}_S)\boldsymbol{\Delta}_S, \\
&\quad \geq -\|\nabla L(\boldsymbol{\theta}^i_S)\|_{\infty,2} \|\boldsymbol{\Delta}_S\|_{1,2} + \|\boldsymbol{\Delta}_S\|_2^2 \lambda_{\min}(\mathbf{H}(\mathcal{D}; \boldsymbol{\theta}^i_S + t\boldsymbol{\Delta}_S)) \\
&\quad \geq -\frac{b\lambda}{2} + \frac{\|\boldsymbol{\Delta}_S\|_{1,2}^2}{d_i + 1} \lambda_{\min}(\mathbf{H}(\mathcal{D}; \boldsymbol{\theta}^i_S + t\boldsymbol{\Delta}_S)) \\
&\quad \geq -\frac{b\lambda}{2} + \frac{b^2}{2(d_i + 1)} \left(C_{\min} - \frac{m^2 b(d_i + 1)}{4} \right) \\
&\quad \geq -\frac{b\lambda}{2} + \frac{b^2 C_{\min}}{4(d_i + 1)},
\end{aligned} \tag{26}$$

where the third follows from our assumption that $\|\nabla L(\boldsymbol{\theta}^i)\|_{\infty,2} \leq \lambda/2$ and the fact for any vector $\mathbf{x}$, $\|\mathbf{x}\|_2 \geq (1/\sqrt{g})\|\mathbf{x}\|_{1,2}$ where the $\ell_{1,2}$ norm is evaluated over g groups. The fourth line follows from Lemma 4 with $t = 1$ and Lemma 2. Finally, in the last line we assumed that $b \leq 2C_{\min}/(m^2(d_i+1))$ — an assumption that we will verify momentarily. The second term of (25) is easily lower bounded using the reverse triangle inequality as follows:

$$\lambda(\|\boldsymbol{\theta}^i_S + \boldsymbol{\Delta}_S\|_{1,2} - \|\boldsymbol{\theta}^i_S\|_{1,2}) \geq -\lambda\|\boldsymbol{\Delta}_S\|_{1,2} = -b\lambda \tag{27}$$

Putting together (25), (26) and (27) we get:

$$\widetilde{F}(\boldsymbol{\Delta}_S) \geq -\frac{b\lambda}{2} + \frac{b^2 C_{\min}}{4(d_i + 1)} - b\lambda.$$

Setting the above to zero and solving for b we get:

$$b = \frac{6\lambda(d_i + 1)}{C_{\min}}.$$

Finally, coming back to our assumption that $b \leq 2C_{\min}/(m^2(d_i+1))$, it is easy to show that the assumption holds if the regularization parameter λ satisfies:

$$\lambda \leq \frac{C_{\min}^2}{3m^2(d_i + 1)^2},$$

The lower bound on the number of samples is obtained by ensuring that the lower bound on λ is less than the upper bound. The final claim follows from using the high probability bound on $\|\nabla L(\boldsymbol{\theta}^i)\|_{\infty,2}$ from Lemma 3. □

Proof of Lemma 6 (Error of the i-th parameter estimator).
$\boldsymbol{\Delta} \overset{\text{def}}{=} \widehat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}^i$. We will denote the true support of $\boldsymbol{\theta}^i$ by S, and the complement of S by S^c. We will also simply

write $L(\boldsymbol{\theta})$ instead of $L^i(\mathcal{D};\boldsymbol{\theta})$. For any vector $\mathbf{y}$, let $\mathbf{y}_{\bar{S}}$ denote the vector $\mathbf{y}$ with elements not in the support set S zeroed out, i.e.,

$$(\mathbf{y}_{\bar{S}})_j = \begin{cases} y_j & j \in S, \\ 0 & \text{otherwise} \end{cases}$$

Then by definition of S, we have that $\|\boldsymbol{\theta}^i_{\bar{S}}\|_{1,2} = \|\boldsymbol{\theta}^i\|_{1,2}$.

$$\begin{aligned} \|\widehat{\boldsymbol{\theta}}^i\|_{1,2} = \|\boldsymbol{\theta}^i + \boldsymbol{\Delta}\|_{1,2} &= \|\boldsymbol{\theta}^i_{\bar{S}} + \boldsymbol{\Delta}_{\bar{S}} + \boldsymbol{\Delta}_{\bar{S}^c}\|_{1,2} \\ &= \|\boldsymbol{\theta}^i_{\bar{S}} + \boldsymbol{\Delta}_{\bar{S}}\|_{1,2} + \|\boldsymbol{\Delta}_{\bar{S}^c}\|_{1,2} \\ &\geq \|\boldsymbol{\theta}^i_{\bar{S}}\|_{1,2} - \|\boldsymbol{\Delta}_{\bar{S}}\|_{1,2} + \|\boldsymbol{\Delta}_{\bar{S}^c}\|_{1,2}, \end{aligned}$$

where in the second line follows from the fact that the index sets S and S^c have non-overlapping groups, and in the last line we used the reverse triangle inequality. Rearranging the terms of the previous equation, and from the fact that $\|\boldsymbol{\theta}^i_{\bar{S}}\|_{1,2} = \|\boldsymbol{\theta}^i\|_{1,2}$, we get:

$$\|\boldsymbol{\theta}^i\|_{1,2} - \|\widehat{\boldsymbol{\theta}}^i\|_{1,2} \leq \|\boldsymbol{\Delta}_{\bar{S}}\|_{1,2} - \|\boldsymbol{\Delta}_{\bar{S}^c}\|_{1,2} \tag{28}$$

Next, by optimality of $\widehat{\boldsymbol{\theta}}^i$ we have that $L(\boldsymbol{\theta}^i) + \lambda\|\boldsymbol{\theta}^i\|_{1,2} \geq L(\widehat{\boldsymbol{\theta}}^i) + \lambda\|\widehat{\boldsymbol{\theta}}^i\|_{1,2}$. Rearranging the terms and continuing, we get

$$\begin{aligned} \lambda(\|\boldsymbol{\theta}^i\|_{1,2} - \|\widehat{\boldsymbol{\theta}}^i\|_{1,2}) &\geq L(\widehat{\boldsymbol{\theta}}^i) - L(\boldsymbol{\theta}^i) \\ &\geq (\nabla L(\widehat{\boldsymbol{\theta}}^i))^T(\widehat{\boldsymbol{\theta}}^i - \boldsymbol{\theta}^i) \\ &\geq -\|\nabla L(\widehat{\boldsymbol{\theta}}^i))^T\|_{\infty,2}\|\boldsymbol{\Delta}\|_{1,2} \\ &\geq -\frac{\lambda}{2}\|\boldsymbol{\Delta}\|_{1,2}, \end{aligned} \tag{29}$$

where the third line follows from the convexity of $L(\cdot)$, the fourth line follows from the Cauchy-Schwartz inequality and the last line follows from our assumption that $\lambda \geq 2\|\nabla L(\boldsymbol{\theta}^i)\|_{\infty,2}$. Thus, from (28) and (29) we have that

$$\begin{aligned} &\frac{1}{2}\|\boldsymbol{\Delta}\|_{1,2} \geq \|\boldsymbol{\Delta}_{\bar{S}^c}\|_{1,2} - \|\boldsymbol{\Delta}_{\bar{S}}\|_{1,2} \\ \implies &\frac{1}{2}\|\boldsymbol{\Delta}_{\bar{S}}\|_{1,2} + \frac{1}{2}\|\boldsymbol{\Delta}_{\bar{S}^c}\|_{1,2} \geq \|\boldsymbol{\Delta}_{\bar{S}^c}\|_{1,2} - \|\boldsymbol{\Delta}_{\bar{S}}\|_{1,2} \\ \implies &3\|\boldsymbol{\Delta}_{\bar{S}}\|_{1,2} \geq \|\boldsymbol{\Delta}_{\bar{S}^c}\|_{1,2}. \end{aligned}$$

Finally, from the above inequality, we have $\|\boldsymbol{\Delta}\|_{1,2} = \|\boldsymbol{\Delta}_{\bar{S}}\|_{1,2} + \|\boldsymbol{\Delta}_{\bar{S}^c}\|_{1,2} \leq 4\|\boldsymbol{\Delta}_S\|_{1,2}$. The final result follows from the upper bound on $\|\boldsymbol{\Delta}_S\|_{1,2}$ derived in Lemma 5. □

Lemma 7 (Maximum eigenvalue of block positive semi-definite matrix)**.** *Let* $\mathbf{X} \in \mathbb{R}^{n \times n} \succeq \mathbf{0}$ *be any positive semi-definite matrix, with* $\mathbf{X}_{i,i}$ *being the* i*-th diagonal block of* $\mathbf{X}$*. Then*

$$\lambda_{\max}(\mathbf{X}) \leq \sum_i \lambda_{\max}(\mathbf{X}_{i,i})$$

Proof. We will prove the result by decomposing $\mathbf{X}$ into two blocks as follows:

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_{1,1} & \mathbf{X}_{1,2} \\ \mathbf{X}_{2,1} & \mathbf{X}_{2,2} \end{bmatrix},$$

where $\mathbf{X}_{1,1} \in \mathbb{R}^{n_1 \times n_1}$, $\mathbf{X}_{2,2} \in \mathbb{R}^{n_2 \times n_2}$ and $n_1 + n_2 = 1$. The general result for multiple diagonal blocks is obtained by recursively decomposing the blocks $\mathbf{X}_{1,1}$ and $\mathbf{X}_{2,2}$. Any unit vector $\mathbf{x}$ can be written as $\mathbf{x} = c_1(\mathbf{x})\mathbf{x}_1(\mathbf{x}) + c_2(\mathbf{x})\mathbf{x}_2(\mathbf{x})$, with $\mathbf{x}_1(\mathbf{x}) = (x_1/\|\mathbf{x}_1(\mathbf{x})\|_2, \ldots, x_{n_1}/\|\mathbf{x}_1(\mathbf{x})\|_2, \mathbf{0})$, $\mathbf{x}_2(\mathbf{x}) = (\mathbf{0}, x_{n_2}/\|\mathbf{x}_2(\mathbf{x})\|_2, \ldots, x_n/\|\mathbf{x}_2(\mathbf{x})\|_2)$, and $c_1(\mathbf{x}) = \|\mathbf{x}_1(\mathbf{x})\|_2$ (similarly $c_2(\mathbf{x})$). For notational simplicity we will drop the $(\mathbf{x})s$. Note that $c_1^2 + c_2^2 = 1$, thus $\mathbf{c} = (c_1, c_2)$ is also a unit vector. Further, for any unit vector $\mathbf{x}$, we have $\mathbf{x}^T\mathbf{X}\mathbf{x} = \mathbf{c}^T\mathbf{Y}\mathbf{c}$, where

$$\mathbf{Y} = \begin{bmatrix} \mathbf{x}_1^T\mathbf{X}\mathbf{x}_1 & \mathbf{x}_1^T\mathbf{X}\mathbf{x}_2 \\ \mathbf{x}_2^T\mathbf{X}\mathbf{x}_1 & \mathbf{x}_2^T\mathbf{X}\mathbf{x}_2 \end{bmatrix} \in \mathbb{R}^{2 \times 2}.$$

Note that $\mathbf{x}_1^T \mathbf{X} \mathbf{x}_1 \leq \lambda_{\max}(\mathbf{X}_{1,1})$ and $\mathbf{x}_2^T \mathbf{X} \mathbf{x}_2 \leq \lambda_{\max}(\mathbf{X}_{2,2})$ for all $\mathbf{x}$. Thus, using the variational characterization of the maximum eigenvalue of $\mathbf{X}$ we get:

$$\begin{aligned}
\lambda_{\max}(\mathbf{X}) &= \max_{\|\mathbf{x}\|_2=1} \mathbf{x}^T \mathbf{X} \mathbf{x} \\
&= \max_{\{\mathbf{c}=(\|\mathbf{x}_1(\mathbf{x})\|_2, \|\mathbf{x}_2(\mathbf{x})\|_2) \,:\, \|\mathbf{x}\|_2=1\}} \mathbf{c}^T \mathbf{Y} \mathbf{c} \\
&\leq \max_{\|\mathbf{c}\|_2=1} \mathbf{c}^T \mathbf{Y} \mathbf{c} = \lambda_{\max}(\mathbf{Y}) \leq \operatorname{Tr}(\mathbf{Y}) \qquad \text{(since } \mathbf{Y} \text{ is positive semi-definite)} \\
&\leq \lambda_{\max}(\mathbf{X}_{1,1}) + \lambda_{\max}(\mathbf{X}_{2,2}),
\end{aligned}$$

where the third line follows from the fact that the maximization is over a superset of the set $\{\mathbf{c} = (\|\mathbf{x}_1(\mathbf{x})\|_2, \|\mathbf{x}_2(\mathbf{x})\|_2) \,:\, \|\mathbf{x}\|_2 = 1\}$. □

Appendix B Details of Synthetic Experiments

We generated random polymatrix games $\mathcal{G}$ by first generating random graphs over p players with degree exactly d, and number of pure strategies $m = 3$ per player. For each edge (i, j) in the graph, we set the payoffs as follows:

$$\begin{aligned}
u^{i,i}(a) &= 0 && (\forall a \in [3]) \\
u^{i,j}(a, b) &\sim \mathcal{N}(0, 2) && (\forall a \in [2] \wedge b \in [3]) \\
u^{i,j}(3, b) &= 0 && (\forall b \in [3])
\end{aligned}$$

We then generated a data set $\mathcal{D}$ from the game using the local noise model (5), with the noise parameter $q_i = 0.6$ for all $i \in [p]$. We then used our method to learn a game $\widehat{\mathcal{G}}$ from the data set $\mathcal{D}$, and computed $\mathbf{1}\left[\mathcal{NE}(\widehat{\mathcal{G}}) = \mathcal{NE}(\mathcal{G})\right]$. We then estimated the probability of successful PSNE recovery, $\Pr\left\{\mathcal{NE}(\widehat{\mathcal{G}}) = \mathcal{NE}(\mathcal{G})\right\}$, across 40 randomly sampled polymatrix games. Figure 1 plots the probability of successful PSNE recovery as the number of samples is varied as $n = 10^c (d+1)^2 \log(2p(d+1)/\delta)$ and for various values of $d \in \{1, 3, 5\}$, with c being the control parameter and $\delta = 0.01$.

Appendix C Experiments on real-world data

We validated our method on three publicly available real-world data sets containing (a) U.S. supreme court justices rulings, (b) voting records of senators from the 114th U.S. congress, and (c) roll-call votes in the U.N. General Assembly. We present evaluations of our method for each of the data set below.

C.1 Supreme court voting records

We analyzed two data sets of supreme court rulings: the first data set contains rulings of 9 justices across 512 cases spanning years 2010 to 2014, while the second data set contains rulings of 8 justices across 75 cases from year 2015 onwards [3]. We pre-processed the data, according the available code book, to map the vote of each justice, which was originally an integer between 1 to 8, to an integer between 1 to 3. Votes $\{1, 3, 4, 5\}$ were mapped to 1 and was interpreted as "voting with majority", votes $\{6, 7, 8\}$ were mapped to 2 and was interpreted as "not participating in the decision" , while vote 3 was mapped to 2 and was interpreted as "dissent". Thus, after pre-processing, each justice's vote was an integer between 1 to 3, with 1 corresponding to majority, 2 corresponding to abstention, and 3 corresponding to dissent.

After pre-processing the data, we learned a polymatrix game over supreme court justices using our algorithm. The regularization parameter λ was set according to Theorem 1 with reasonable values for different unknown population parameters. A more principled way to chose the regularization parameter λ is to assume a specific observation model, for instance, the global or local noise model, and then using crossvalidation to maximize the log-likelihood. The game graphs are shown in Figure 2 and the PSNE sets are shown in Table 1 for the two supreme court rulings data sets (years 2010-2014 and year 2015 onwards).

[3] All the data sets are publicly available at `http://scdb.wustl.edu`.

From 2 it is clear that our method recovers the well-established ideologies of the supreme court justices. This is especially evident for the graph learned from the first data set — there are two strongly connected components corresponding to the conservative and liberal bloc within the supreme court. The PSNE set recovered by our algorithm is also quite revealing. In both the data sets, a unanimous vote of 1 is a Nash equilibrium. Justice Kennedy, who has a moderate jurisprudence, always votes with the majority in the PSNE set. Further, strategy profiles where the conservative blocs and liberal blocs vote unanimously but dissent against each other are also in the PSNE set. In the second data set, there is a strongly connected component between the justice Kagan, Kennedy, and Breyer — this also bears out in the corresponding PSNE set where the strategies of the three justices are identical.

To compute the price of anarchy (PoA), we shifted all the payoff matrices by a constant to make the payoffs non-negative. Note that this does not change the PSNE set of the game. The price of anarchy was computed to be the ratio between the maximum welfare across all strategy profiles and the minimum welfare across all strategy profiles in the PSNE set. The PoA for the two data sets were, respectively, 1.9104 and 1.6115.

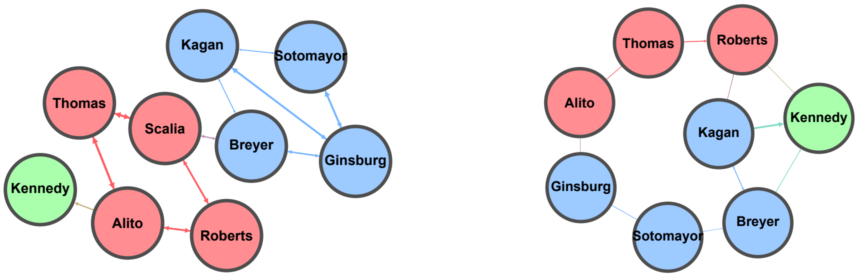

Figure 2: The graphical game recovered from supreme court rulings data set 1 (years 2010-2014) on the left, and data set 2 (year 2015 onwards) on the right. Justice Thomas, Scalia, Roberts and Alito are widely known to be conservative and are denoted by the color 🟥, while Justice Breyer, Kagan, Sotomayor and Ginsburg, who are known to have a more liberal jurisprudence, are denoted by color 🟦. Justice Kennedy, who has a reputation of being moderate, is denoted by the color 🟩. The game graph was generated by adding all edges (i, j) if the corresponding payoff matrix $u^{i,j}$ was not all zeros. The average "influence" from j to i was calculated as the mean absolute payoff, i.e., $\frac{1}{6}\sum_{a=2}^{3}\sum_{b=1}^{3}|u^{i,j}(a,b)|$. The thickness of the edge denotes this influence of player j on i. Only the top 50% of the edges, in terms of influence, are shown.

Thomas	Scalia	Alito	Roberts	Kennedy	Breyer	Kagan	Ginsburg	Sotomayor
1	1	1	1	1	1	1	1	1
1	1	1	1	1	3	3	3	3
2	2	2	2	1	2	2	2	2
3	3	3	3	1	1	1	1	1
3	3	3	3	1	3	3	3	3

Thomas	Alito	Roberts	Kennedy	Breyer	Kagan	Ginsburg	Sotomayor
1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2
2	2	2	2	2	2	3	3
3	3	3	2	2	2	3	3

Table 1: The PSNE set learned from supreme court rulings data sets 1704 (top) and 1705 (bottom) respectively. Colors represent **conservative**, **liberal**, and **neutral** justices respectively. The price of anarchy for the two data sets was computed to be 1.9 and 1.6 respectively.

C.2 Senate voting records

We analyzed U.S. congressional voting records for the second session of the 114th congress (January 4, 2016 to January 3, 2017) [4]. The data set comprised of the votes of 100 senators on 63 bills. The votes were pre-processed to take one of the three values: 1 (“yes”), 2 (“abstention”), and 3 (“no”). After pre-processing the data set we ran our algorithm to recover a polymatrix game from congressional voting records. Figure 3 shows the recovered game graph. Once again our method recovers the connected components corresponding the republicans and democrats. Interestingly, the connected components also have a nice geographic interpretation, for instance, the graph groups senators from Idaho, New Mexico, New York and midwestern states in their respective connected components. Strategy profiles where the overwhelming majority of senators in a connected component vote “yes” are in equilibria.

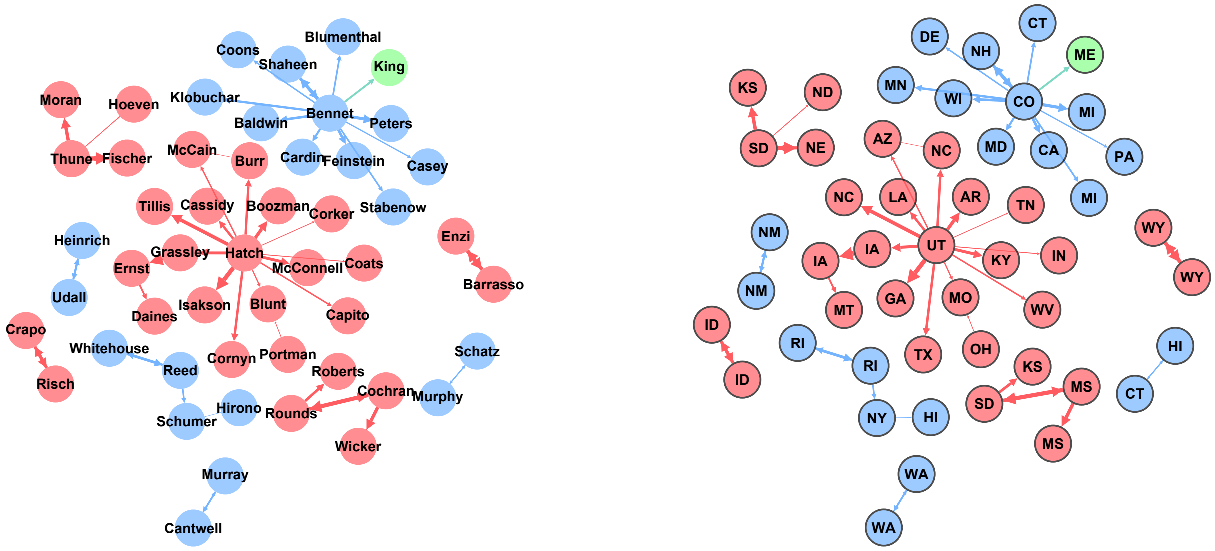

Figure 3: The game graph learned from 114th U.S. congressional voting records. Only nodes with degree greater than one are shown. Colors represent the following: **Democrat** , **Republican** , **Independent** . The graph on the right shows the states that the senators belong to. The thickness of the edges denote the amount of influence, computed as the mean absolute payoff, between the senators. Only nodes with degree at least 1 are shown.

Baldwin	Bennet	Blumenthal	Cardin	Casey	Coons	Feinstein	King	Klobuchar	Peters	Shaheen	Stabenow
1	1	1	1	1	1	1	1	1	1	1	1

Cochran	Roberts	Rounds	Wicker
1	1	1	1

Fischer	Hoeven	Moran	Thune
1	1	1	1
3	3	3	3

Hirono	Reed	Schumer	Whitehouse
1	1	1	1
3	3	3	3

Blunt	Boozman	Burr	Capito	Cassidy	Coats	Corker	Cornyn	Daines	Ernst	Grassley	Hatch	Isakson	McCain	McConnell	Portman	Tillis
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	3	1
1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	2	1
1	3	3	3	1	3	3	3	1	1	3	3	1	3	3	1	3
3	1	1	1	1	1	1	1	1	1	1	2	1	1	1	3	1
3	3	3	3	1	3	3	3	1	1	3	3	1	3	3	2	3
3	3	3	3	1	3	3	3	1	1	3	3	1	3	3	3	3

Table 2: The PSNE set for the major connected components in the game graph learned from congressional voting records. The combined number of Nash equilibria computed across senators with degree at least 1 was 144 and the price of anarchy was computed to be 2.6297.

C.3 United Nations voting data

In our final real-world experiment we analyzed roll-call votes in the U.N. General Assembly. The data set

[4] The data set is publicly available at `http://www.senate.gov/legislative/votes.htm`

contained votes of 193 countries for 847 U.N. resolutions [5]. Each vote could take one of the three values in $\{1, 2, 3\}$, with 1 denoting "yes", 2 denoting "abstention", and 3 denoting "no". The game graph learned from the data set is shown in Figure 4 while the PSNE set is shown in Table 3. As evident from Figure 4 our method recovered two major connected components: the first consisting of members of the Arab League, and the second consisting of majorly Southeast Asian countries and a few other Caribbean islands. The PSNE set once again comprised of strategy profiles where the overwhelming members of a connected component voted "yes". Within the component corresponding to the Arab league, Saudi Arabia, U.A.E., and Bahrain made up a small coalition of countries that voted identically in the PSNE set.

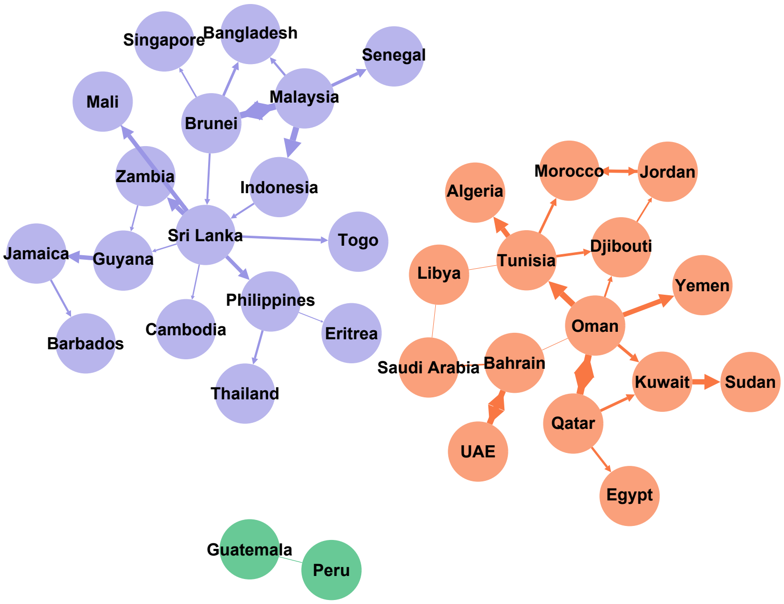

Figure 4: The game graph learned from United Nations voting data set. Nodes belonging to the same connected component have the same color. Only countries with degree at least 1 are shown.

Algeria	Bahrain	Djibouti	Egypt	Jordan	Kuwait	Libya	Morocco	Oman	Qatar	Saudi Arabia	Sudan	Tunisia	UAE	Yemen
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	2	1	1	1	1	1	1	1	1	2	1	1	2	1

Barbados	Bangladesh	Brunei	Cambodia	Eritrea	Guyana	Indonesia	Jamaica	Malaysia	Mali	Philippines	Senegal	Singapore	Sri Lanka	Thailand	Togo	Zambia
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	2	2	1	1	1	2	1	2	1	1	1	2	1	1	1	1
1	2	2	2	2	1	2	1	2	2	2	1	2	2	2	2	2
2	2	2	2	2	2	2	2	2	2	2	1	2	2	2	2	2

Table 3: The PSNE set for the two major connected components in the game graph learned from United Nations voting data set. The total number of PSNE was 24 and the price of anarchy was computed to be 3.07.

[5] The data set can be downloaded from `https://dataverse.harvard.edu/dataset.xhtml?persistentId=hdl:1902.1/12379`.