Appendix A Detailed Proofs

Proof of Proposition 1. Let $|\mathcal{NE}^*| > 2^{n-1}$. Then by the pigeon hole principle there are at least two joint actions $\mathbf{x}$ and $\mathbf{x}'$ in $\mathcal{NE}^*$ such that $\mathbf{x} = -\mathbf{x}'$. Since the payoff is strictly positive, it follows that the bias b_i for each player must be 0. If the bias for all players is 0, then for each $\mathbf{x} \in \mathcal{NE}^*$, $-\mathbf{x} \in \mathcal{NE}^*$. Therefore, $|\mathcal{NE}^*| = 2^n$. Since we have assumed that the game is non-trivial, we get a contradiction. □

Proof of Lemma 5. Define $\Delta \overset{\text{def}}{=} \widehat{\mathbf{v}} - \mathbf{v}^*$. Also for any vector $\mathbf{y}$ let the notation $\mathbf{y}_{\overline{S}}$ denote the vector $\mathbf{y}$ with the entries not in the support, S, set to zero, i.e.

$$[\mathbf{y}_{\overline{S}}]_i = \begin{cases} y_i & \text{if } i \in S, \\ 0 & \text{otherwise.} \end{cases}$$

Similarly, let the notation $\mathbf{y}_{\overline{S^c}}$ denote the vector $\mathbf{y}$ with the entries not in S^c set to zero, where S^c is the complement of S. Having introduced our notation and since, S is the support of the true vector $\mathbf{v}^*$, we have by definition that $\mathbf{v}^* = \mathbf{v}^*_{\overline{S}}$. We then have, using the reverse triangle inequality,

$$\begin{aligned} \|\widehat{\mathbf{v}}\|_1 &= \|\mathbf{v}^* + \Delta\|_1 = \left\|\mathbf{v}^*_{\overline{S}} + \Delta_{\overline{S}} + \Delta_{\overline{S^c}}\right\|_1 \\ &= \left\|\mathbf{v}^*_{\overline{S}} - (-\Delta_{\overline{S}})\right\|_1 + \|\Delta_{\overline{S^c}}\|_1 \\ &\geq \|\mathbf{v}^*\|_1 - \|\Delta_{\overline{S}}\|_1 + \|\Delta_{\overline{S^c}}\|_1 . \end{aligned} \tag{19}$$

Also, from the optimality of $\widehat{\mathbf{v}}$ for the ℓ_1-regularized problem we have that

$$\begin{aligned} &\ell(\mathbf{v}^*) + \lambda \|\mathbf{v}^*\|_1 \geq \ell(\widehat{\mathbf{v}}) + \lambda \|\widehat{\mathbf{v}}\|_1 \\ \Longrightarrow\ & \lambda(\|\mathbf{v}^*\|_1 - \|\widehat{\mathbf{v}}\|_1) \geq \ell(\widehat{\mathbf{v}}) - \ell(\mathbf{v}^*). \end{aligned} \tag{20}$$

Next, from convexity of $\ell(.)$ and using the Cauchy-Schwartz inequality we have that

$$\begin{aligned} \ell(\widehat{\mathbf{v}}) - \ell(\mathbf{v}^*) &\geq \nabla\ell(\mathbf{v}^*)^T(\widehat{\mathbf{v}} - \mathbf{v}^*) \\ &\geq - \|\nabla\ell(\mathbf{v}^*)\|_\infty \|\Delta\|_1 \\ &\geq -\frac{\lambda}{2} \|\Delta\|_1 , \end{aligned} \tag{21}$$

in the last line we used the fact that $\lambda \geq \|\nabla\ell(\mathbf{v}^*)\|_\infty$. Thus, we have from (19), (20) and (21) that

$$\begin{aligned} &\frac{1}{2} \|\Delta\|_1 \geq \|\widehat{\mathbf{v}}\|_1 - \|\mathbf{v}^*\|_1 \\ \Longrightarrow\ & \frac{1}{2} \|\Delta\|_1 \geq \|\Delta_{\overline{S^c}}\|_1 - \|\Delta_{\overline{S}}\|_1 \\ \Longrightarrow\ & \frac{1}{2} \|\Delta_{\overline{S^c}}\|_1 + \frac{1}{2} \|\Delta_{\overline{S}}\|_1 \geq \|\Delta_{\overline{S^c}}\|_1 - \|\Delta_{\overline{S}}\|_1 \\ \Longrightarrow\ & 3 \|\Delta_{\overline{S}}\|_1 \geq \|\Delta_{\overline{S^c}}\|_1 . \end{aligned} \tag{22}$$

Finally, from (22) and Lemma 4 we have that

$$\begin{aligned} \|\Delta\|_1 &= \|\Delta_{\overline{S}}\|_1 + \|\Delta_{\overline{S^c}}\|_1 \\ &\leq 4 \|\Delta_{\overline{S}}\|_1 \leq 4\sqrt{|S|}\, \|\Delta_{\overline{S}}\|_2 \\ &\leq \frac{5C_{\min}}{D_{\max}}. \end{aligned}$$

□

Proof of Theorem 2. First, we construct a restricted ensemble of games $\widetilde{\mathfrak{G}} \subset \mathfrak{G}$ as follows. Each game $\mathcal{G} \in \widetilde{\mathfrak{G}}$ contains k, randomly chosen, *influential* players. The game graph for $\mathcal{G}$ is then chosen to be a complete directed bipartite graph from the set of k influential players to the set of $n - k$ non-influential players. The edge weights are all set to -1, the bias for the k influential players is set to $+1$, while the bias for the remaining $n - k$ players is set to 0. Then it is clear that *each game in* $\widetilde{\mathfrak{G}}$ *induces a distinct size-one PSNE set.* Specifically, for a game $\mathcal{G} \in \widetilde{\mathfrak{G}}$, a joint action $\mathbf{x} \in \mathcal{NE}(\mathcal{G})$ is such that $x_i = -1$ if player i is influential in $\mathcal{G}$, otherwise $x_i = +1$. Also, note that the minimum payoff in the PSNE set of each game in $\widetilde{\mathfrak{G}}$ is strictly positive, and is precisely 1. Finally, we assume that the data set is drawn according to the global noise model (3), with $q = 1/n$. Now let $\mathcal{G} \in \widetilde{\mathfrak{G}}$ be a uniformly-distributed random variable corresponding to the game that was picked by nature. From the Fano's inequality, we have that:

$$p_{\text{err}} \geq 1 - \frac{I(\mathcal{D}; \mathcal{G}) + \log 2}{H(\mathcal{G})}, \tag{23}$$

where $I(.)$ denotes mutual information and $H(.)$ denotes entropy. Since, $\mathcal{G}$ is uniformly distributed, we have that $H(\mathcal{G}) = \log \left|\widetilde{\mathfrak{G}}\right| = \log \binom{n}{k} \geq k(\log n - \log k)$. Let $\mathcal{P}_{\mathcal{D}|\mathcal{G}=\mathcal{G}_1}$ be the conditional distribution of the data set given a game $\mathcal{G}_1 \in \widetilde{\mathfrak{G}}$. We bound the mutual information $I(\mathcal{D}; \mathcal{G})$ by a pairwise KL-based bound from [16] as follows:

$$I(\mathcal{D}; \mathcal{G}) \leq \frac{1}{\left|\widetilde{\mathfrak{G}}\right|} \sum_{\mathcal{G}_1 \in \widetilde{\mathfrak{G}}} \sum_{\mathcal{G}_2 \in \widetilde{\mathfrak{G}}} \mathbb{KL}\left(\mathcal{P}_{\mathcal{D}|\mathcal{G}=\mathcal{G}_1} \middle\| \mathcal{P}_{\mathcal{D}|\mathcal{G}=\mathcal{G}_2}\right). \tag{24}$$

Now from the fact that data are sampled i.i.d , we get:

$$\begin{aligned}
&\mathbb{KL}\left(\mathcal{P}_{\mathcal{D}|\mathcal{G}=\mathcal{G}_1} \middle\| \mathcal{P}_{\mathcal{D}|\mathcal{G}=\mathcal{G}_2}\right) \\
&\quad = m \textstyle\sum_{\mathbf{x} \in \mathcal{X}} \mathcal{P}_{\mathcal{D}|\mathcal{G}=\mathcal{G}_1}(\mathbf{x}) \log \frac{\mathcal{P}_{\mathcal{D}|\mathcal{G}=\mathcal{G}_1}(\mathbf{x})}{\mathcal{P}_{\mathcal{D}|\mathcal{G}=\mathcal{G}_2}(\mathbf{x})} \\
&\quad = m \left\{ q \log \tfrac{q(2^n-1)}{1-q} + \tfrac{1-q}{2^n-1} \log \tfrac{1-q}{q(2^n-1)} \right\} \\
&\quad = \tfrac{m(2^n q - 1)}{2^n - 1} \left(\log q - \log\left(\tfrac{1-q}{2^n-1}\right)\right) \\
&\quad \leq m \log 2,
\end{aligned} \tag{25}$$

where the last line comes from the fact that $q = 1/n$. Putting together (23), (24) and (25), and setting $p_{\text{err}} = 1/2$, we get

$$m \leq \frac{k \log n - k \log k - 2 \log 2}{2 \log 2}.$$

By observing that learning the ensemble $\mathfrak{G}$ is at least as hard as learning a subset of the ensemble $\widetilde{\mathfrak{G}}$, we prove our main claim. □

Appendix B Experiments

In order to verify that our results and assumptions indeed hold in practice, we performed various simulation experiments. We generated random LIGs for n players and exactly k neighbors by first creating a matrix $\mathbf{W}$ of all zeros and then setting k off-diagonal entries of each row, chosen uniformly at random, to -1. We set the bias for all players to 0. We found that any odd value of k produces games with strictly positive payoff in the PSNE set. Therefore, for each value of k in $\{1, 3, 5\}$, and n in $\{10, 12, 15, 20\}$, we generated 40 random LIGs. For experiments involving the local noise model, we only used $n \in \{10, 12, 15\}$. The parameter δ was set to the constant value of 0.01. For the global noise model, the parameters q_g was set to 0.01, while for the local noise model we used $q_1 = \ldots = q_n = 0.6$. The regularization parameter λ was set according to Theorem 1 as some constant multiple of $\sqrt{(2/m)\log(2n/\delta)}$. Figure 1 shows the probability of successful recovery of the PSNE, for various combinations of (n, k), where the probability was computed as the fraction of the 40 randomly sampled LIGs for which the learned PSNE set matched the true PSNE set exactly. For each experiment, the number of samples was computed as: $\lfloor (C)(10^c)(k^2 \log(6n^2/\delta)) \rfloor$, where c is the control parameter and the constant C is 10000 for $k = 1$ and 1000 for $k = 3$ and 5. Thus, from Figure 1 we see that, the sample complexity of $\mathcal{O}\left(k^2 \log n\right)$ as given by Theorem 1 indeed holds in practice, i.e., there exists constants c and c' such that if the number of samples is less than $ck^2 \log n$, we fail to recover the PSNE set exactly with high probability, while if the number of samples is greater than $c'k^2 \log n$ then we are able to recover the PSNE set exactly, with high probability. Further, the scaling remains consistent as the number of players n is changed from 10 to 20.

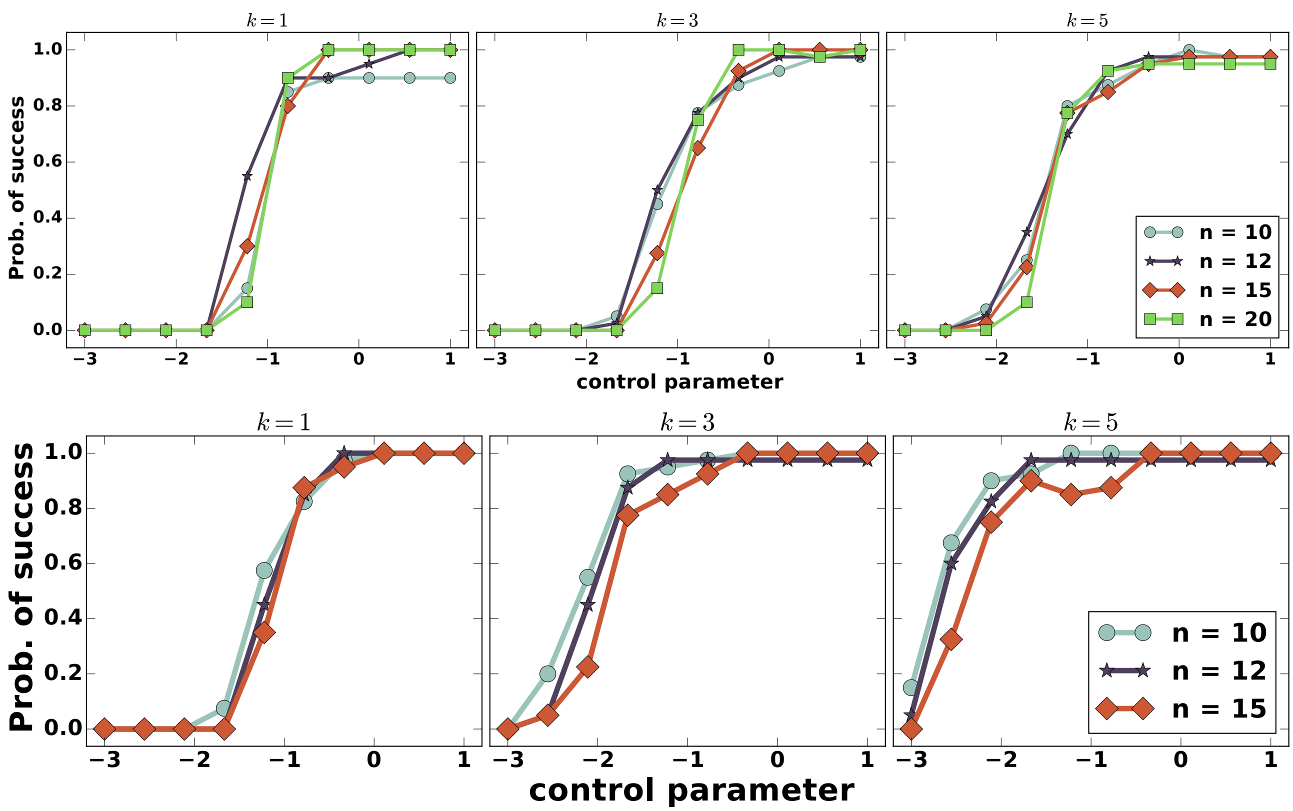

Figure 1: The probability of exact recovery of the PSNE set computed across 40 randomly sampled LIGs, using the global noise model (TOP) and local noise model (BOTTOM), as the number of samples is scaled as $\lfloor (C)(10^c)(k^2 \log(6n^2/\delta)) \rfloor$, where c is the control parameter and the constant C is 10000 for the $k = 1$ case and 1000 for the remaining two case. For the global noise model we set $q_g = 0.001$, while for the local noise model we used $q_1 = \ldots = q_n = 0.6$.