Local Aggregative Games

Vikas K. Garg
CSAIL, MIT
vgarg@csail.mit.edu

Tommi Jaakkola
CSAIL, MIT
tommi@csail.mit.edu

Aggregative games provide a rich abstraction to model strategic multi-agent interactions. We introduce local aggregative games, where the payoff of each player is a function of its own action and the aggregate behavior of its neighbors in a connected digraph. We show the existence of pure strategy ϵ-Nash equilibrium in such games when the payoff functions are convex or sub-modular. We prove an information theoretic lower bound, in a value oracle model, on approximating the structure of the digraph with non-negative monotone sub-modular cost functions on the edge set cardinality. We also define a new notion of structural stability, and introduce γ-aggregative games that generalize local aggregative games and admit ϵ-Nash equilibrium that are stable with respect to small changes in some specified graph property. Moreover, we provide algorithms for our models that can meaningfully estimate the game structure and the parameters of the aggregator function from real voting data.

1 Introduction

Structured prediction methods have been remarkably successful in learning mappings between input observations and output configurations [1; 2; 3]. The central guiding formulation involves learning a scoring function that recovers the configuration as the highest scoring assignment. In contrast, in a game theoretic setting, myopic strategic interactions among players lead to a Nash equilibrium or locally optimal configuration rather than highest scoring global configuration. Learning games therefore involves, at best, enforcement of local consistency constraints as recently advocated [4].

[4] introduced the notion of contextual potential games, and proposed a dual decomposition algorithm for learning these games from a set of pure strategy Nash equilibria. However, since their setting was restricted to learning undirected tree structured potential games, it cannot handle (a) asymmetries in the strategic interactions, and (b) higher order interactions. Moreover, a wide class of strategic games (e.g. anonymous games [5]) do not admit a potential function and thus locally optimal configurations do not coincide with pure strategy Nash equilibria. In such games, the existence of only (approximate) mixed strategy equilibria is guaranteed [6].

In this work, we focus on learning *local* aggregative games to address some of these issues. In an aggregative game [7; 8; 9], every player gets a payoff that depends only on its own strategy and the aggregate of all the other players' strategies. Aggregative games and their generalizations form a very rich class of strategic games that subsumes Cournot oligopoly, public goods, anonymous, mean field, and cost and surplus sharing games [10; 11; 12; 13]. In a local aggregative game, a player's payoff is a function of its own strategy and the aggregate strategy of its neighbors (i.e. only a subset of other players). We do not assume that the interactions are symmetric or confined to a tree structure, and therefore the game structure could, in general, be a spanning digraph, possibly with cycles.

We consider local aggregative games where each player's payoff is a convex or submodular Lipschitz function of the aggregate of its neighbors. We prove sufficient conditions under which such games admit some pure strategy ϵ-Nash equilibrium. We then prove an information theoretic lower bound that for a specified ϵ, approximating a game structure that minimizes a non-negative monotone submodular cost objective on the cardinality of the edge set may require exponentially many queries under a zero-order or value oracle model. Our result generalizes the approximability of the submodular minimum spanning tree problem to degree constrained spanning digraphs [14]. We argue that this lower bound might be averted with a dataset of multiple ϵ-Nash equilibrium configurations sampled

31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.

from the local aggregative game. We also introduce γ-aggregative games that generalize local aggregative games to accommodate the (relatively weaker) effect of players that are not neighbors. These games are shown to have a desirable stability property that makes their ϵ-Nash equilibria robust to small fluctuations in the aggregator input. We formulate learning these games as optimization problems that can be efficiently solved via branch and bound, outer approximation decomposition, or extended cutting plane methods [17; 18]. The information theoretic hardness results do not apply to our algorithms since they have access to the (sub)gradients as well, unlike the value oracle model where only the function values may be queried. Our experiments strongly corroborate the efficacy of the local aggregative and γ-aggregative games in estimating the game structure on two real voting datasets, namely, the US Supreme Court Rulings and the Congressional Votes.

2 Setting

We consider an n-player game where each player $i \in [n] \triangleq \{1, 2, \ldots, n\}$ plays a strategy (or action) from a finite set A_i. For any strategy profile a, a_i denotes the strategy of the i^{th} player, and a_{-i} the strategies of the other players. We are interested in local aggregative games that have the property that the payoff of each player i depends only on its own action and the aggregate action of its neighbors $N_G(i) = \{j \in V(G) : (j, i) \in E(G)\}$ in a connected digraph $G = (V, E)$, where $|V| = n$. Since, the graph is directed, the neighbors need not be symmetric, i.e., $(j, i) \in E$ does not imply $(i, j) \in E$.

For any strategy profile a, we will denote the strategy vector of neighbors of player i by $a_{N_G(i)}$. We assume that player i has a payoff function of the form $u_i(a_i, f_G(a, i))$, where $f_G(a, i) \triangleq f(a_{N_G(i)})$ is a local aggregator function, and u_i is convex and Lipschitz in the aggregate $f_G(a, i)$ for all $a_i \in A_i$. Since $f_G(a, i)$ may take only finitely many values, we will assume interpolation between these values such that they form a convex set. We can define the Lipschitz constant of G as

$$\delta(G) \triangleq \max_{i, a_i, a'_{-i}, a''_{-i}} \{u_i(a_i, f_G(a', i)) - u_i(a_i, f_G(a'', i))\}, \tag{1}$$

where the vectors a'_{-i} and a''_{-i} differ in exactly one coordinate. Clearly, the payoff of any player in the network does not change by more than $\delta(G)$ when one of the neighbors changes its strategy. We can now talk about a class of aggregative games characterized by the Lipschitz constant:

$$L(\Delta, n) = \{G : V(G) = n, \delta(G) \leq \Delta\}.$$

A strategy profile $a = (a_i, a_{-i})$ is said to be a *pure strategy ϵ-Nash equilibrium* (ϵ-PSNE) if no player can improve its payoff by more than ϵ by unilaterally switching its strategy a'_i. In other words, any player i cannot gain more than ϵ by playing an alternative strategy a'_i if the other players continue to play a_{-i}. More generally, instead of playing deterministic actions in response to the actions of others, each player can randomize its actions. Then, the distributions over players' actions constitute a *mixed strategy ϵ-Nash equilibrium* if any unilateral deviation could improve the expected payoff by at most ϵ. We will prove the existence of ϵ-PSNE in our setting. We will assume a training set $S = \{a^1, a^2, \ldots, a^M\}$, where each a^i is an ϵ-PSNE sampled from our game. Our objective is to recover the game digraph G and the payoff functions $u_i, i \in [n]$ from the set S.

The rest of the paper is organized as follows. We first establish some important theoretical paraphernalia on the local aggregative games in Section 3. In Section 4, we introduce γ-aggregative games and show that γ-aggregators are structurally stable. We formulate the learning problem in Section 5, and describe our experimental set up and results in Section 6. We state the theoretical results in the main text, and provide the detailed proofs in the Supplementary (Section 7) for improved readability.

3 Theoretical foundations

Any finite game is guaranteed to admit a mixed strategy ϵ-equilibrium due to a seminal result by Nash [6]. However, general games may not have any ϵ-PSNE (for small ϵ). We first prove a sufficient condition for the existence of ϵ-PSNE in local aggregative games with small Lipschitz constant. A similar result holds when the payoff functions $u_i(\cdot)$ are non-negative monotone submodular and Lipschitz (see the supplementary material for details).

Theorem 1. Any local aggregative game on a connected digraph G, where $G \in L(\Delta, n)$ and $\max_i |A_i| \leq m$, admits a $10\Delta\sqrt{\ln(8mn)}$-PSNE.

Proof. (Sketch.) The main idea behind the proof is to sample a random strategy profile from a mixed strategy Nash equilibrium of the game, and show that with high probability the sampled profile corresponds to an ϵ-PSNE when the Lipschitz constant is small. The proof is based on a novel application of the Talagrand's concentration inequality. $\square$

Theorem 1 implies the minimum degree d (which depends on number of players n, the local aggregator function A, Lipschitz constant Δ, and ϵ) of the game structure that ensures the existence of at least one ϵ-PSNE. One example is the following local generalization of binary summarization games [8]. Each player i plays $a_i \in \{0, 1\}$ and has access to an averaging aggregator that computes the fraction of its neighbors playing action 1. Then, the Lipschitz constant of G is $1/k$, where k is the minimum degree the underlying game digraph. Then, an ϵ-PSNE is guaranteed for $k = \Omega(\sqrt{\ln n}/\epsilon)$. In other words, k needs to grow slowly (i.e., only sub-logarithmically) in the number of players n.

An important follow-up question is to determine the complexity of recovering the underlying game structure in a local aggregative game with an ϵ-PSNE. We will answer this question in a combinatorial setting with non-negative monotone submodular cost functions on the edge set cardinality. Specifically, we consider the following problem. Given a connected digraph $G(V, E)$, a degree parameter d, and a submodular cost function $h : 2^E \to \mathbb{R}^+$ that is normalized (i.e. $h(\emptyset) = 0$) and monotone (i.e. $h(S) \leq h(T)$ for all $S \subseteq T \in 2^E$), we would like to find a spanning directed subgraph[1] G_s of G such that $f(G_s)$ is minimized, the in-degree of each player is at least d, and G_s admits some ϵ-Nash equilibrium when players play to maximize their individual payoffs. We first establish a technical lemma that provides tight lower and upper bounds on the probability that a directed random graph is disconnected, and thus extends a similar result for Erdős-Rényi random graphs [25] to the directed setting. The lemma will be invoked while proving a bound for the recovery problem, and might be of independent interest beyond this work.

Lemma 2. Consider a directed random graph $DG(n, p)$ where $p \in (0, 1)$ is the probability of choosing any directed edge independently of others. Define $q = 1 - p$. Let P_n be the probability that DG is connected. Then, the probability that DG is disconnected is $1 - P_N = nq^{2(n-1)} + O\left(n^2 q^{3n}\right)$.

We will now prove an information theoretic lower bound for the recovery problem under the *value oracle* model [14]. A problem with an information theoretic lower bound of β has the property that any randomized algorithm that approximates the optimum to within a factor β with high probability needs to make superpolynomial number of queries under the specified oracle model. In the value oracle model, each query Q corresponds to obtaining the cost/value of any candidate set by issuing Q to the value oracle (which acts as a black-box). We invoke the Yao's minimax principle [28], which states the relation between distributional complexity and randomized complexity. Using Yao's principle, the performance of randomized algorithms can be lower bounded by proving that no deterministic algorithm can perform well on an appropriately defined distribution of hard inputs.

Theorem 3. Let $\epsilon > 0$, and $\alpha, \delta \in (0, 1)$. Let n be the number of players in a local aggregative game, where each player $i \in [n]$ is provided with some convex Δ-Lipschitz function u_i and an aggregator A. Let $D_n \triangleq D_n(\Delta, \epsilon, A, (u_i)_{i \in [n]})$ be the sufficient in-degree (number of incoming edges) of each player such that the game admits some ϵ-PSNE when the players play to maximize their individual payoffs u_i according to the local information provided by the aggregator A. Assume any non-negative monotone submodular cost function on the edge set cardinality. Then for any $d \geq \max\{D_n, n^\alpha \ln n\}/(1 - \alpha)$, any randomized algorithm that approximates the game structure to a factor $n^{1-\alpha}/(1 + \delta)d$ requires exponentially many queries under the value oracle model.

Proof. (Sketch.) The main idea is to construct a digraph that has exponentially many spanning directed subgraphs, and define two carefully designed submodular cost functions over the edges of the digraph, one of which is deterministic in query size while the other depends on a distribution. We make it hard for the deterministic algorithm to tell one cost function from the other. This can be accomplished by ensuring two conditions: (a) these cost functions map to the same value on *almost* all the queries, and (b) the discrepancy in the optimum value of the functions (on the optimum query) is massive. The proof invokes Lemma 2, exploits the degree constraint for ϵ-PSNE, argues about the optimal query size, and appeals to the Yao's minimax principle. $\square$

[1] A spanning directed graph spans all the vertices, and has the property that the (multi)graph obtained by replacing the directed edges with undirected edges is connected.

Theorem 3 might sound pessimistic from a practical perspective, however, a closer look reveals why the query complexity turned out to be prohibitive. The proof hinged on the fact that *all* spanning subgraphs with same edge cardinality that satisfied the sufficiency condition for existence of *any* ϵ-PSNE were equally good with respect to our deterministic submodular function, and we created an instance with exponentially such spanning subgraphs. However, we might be able to circumvent Theorem 3 by breaking the symmetry, e.g., by using data that specifies multiple distinct ϵ-Nash equilibria. Then, since the digraph instance would be required to satisfy these equilibria, fooling the deterministic algorithm would be more difficult. Thus data could, in principle, help us avoid the complexity result of Theorem 3. We will formulate optimization problems that would enforce margin separability on the equilibrium profiles, which will further limit the number of potential digraphs and thus facilitate learning the aggregative game. Moreover, the hardness result does not apply to our estimation algorithms that will have access to the (sub)gradients in addition to the function values.

4 γ-Aggregative Games

We now describe a generalization of the local aggregative games, which we call the γ-aggregative games. The main idea behind these games is that a player $i \in [n]$ may, often, be influenced not only by the aggregate behavior of its neighbors, but also to a lesser extent on the aggregate behavior of the other players, whose influence on the payoff of i decreases with increase in their distance to i. Let $d_G(i,j)$ be the number of intermediate nodes on a shortest path from j to i in the underlying digraph $G = (V, E)$. That is, $d_G(i,j) = 0$ if $(j,i) \in E$, and $1 + \min_{k \in V \setminus \{i,j\}} d_G(i,k) + d_G(k,j)$ otherwise. Let $W_G \triangleq \max_{i,j \in V} d_G(i,j)$ be the width of G. For any strategy profile $a \in \{0,1\}^n$ and $t \in \{0, 1, \ldots, W_G\}$, let $I_G^t(i) = \{j : d_G(i,j) = t\}$ be the set of nodes that have exactly t intermediaries on a shortest path to i, and let $a_{I_G^t(i)}$ be a strategy profile of the nodes in this set. We define aggregator functions $f_G^t(a,i) \triangleq f(a_{I_G^t(i)})$ that return the aggregate at level t with respect to player i. Let $\gamma \in (0,1)$ be a discount rate. Define the γ-aggregator function

$$g_G(a, \gamma, \ell, i) \triangleq \sum_{t=0}^{\ell} \gamma^t f_G^t(a,i) / \sum_{t=0}^{\ell} \gamma^t,$$

which discounts the aggregates based on the distance $\ell \in \{0, 1, \ldots, W_G\}$ to i. We assume that player $i \in [n]$ has a payoff function of the form $u_i(a_i, \cdot)$, which is convex and η-Lipschitz in its second argument for each fixed a_i. Finally, we define the Lipschitz constant of the γ-aggregative game as

$$\delta^{\gamma}(G) \triangleq \max_{i, a_i, a'_{-i}, a''_{-i}} \{u_i(a_i, g_G(a', \gamma, W_G, i)) - u_i(a_i, g_G(a'', \gamma, W_G, i))\},$$

where the vectors a'_{-i} and a''_{-i} differ in exactly one coordinate.

The main criticism of the concept of ϵ-Nash equilibrium concerns lack of stability: if any player deviates (due to ϵ-incentive), then in general, some other player may have a high incentive to deviate as well, resulting in a non-equilibrium profile. Worse, it may take exponentially many steps to reach an ϵ-equilibrium again. Thus, stability of ϵ-equilibrium is an important consideration. We will now introduce an appropriate notion of stability, and prove that γ-aggregative games admit stable pure strategy ϵ-equilibrium in that any deviation by a player does not affect the equilibrium much.

Structurally Stable Aggregator (SSA): Let $G = (E, V)$ be a connected digraph and $P_G(w)$ be a property of G, where w denotes the parameters of P_G. Let $\mathcal{A}$ be an aggregator function that depends on P_G. Suppose $M = (a_1, a_2, \ldots, a_n)$ be an ϵ-PSNE when $\mathcal{A}$ aggregates information according to $P_G(w)$, where a_i is the strategy of player $i \in V = [n]$. Suppose now $\mathcal{A}$ aggregates information according to $P_G(w')$. Then, $\mathcal{A}$ is a $(\alpha, \beta)_{P,w,w'}$-structurally stable aggregator (SSA) with respect to G, where α and β are functions of the gap between w, w', if it satisfies these conditions: (a) M is a $(\epsilon + \alpha)$-equilibrium under $P_G(w')$, and (b) the payoff of each player at the equilibrium profile M under $P_G(w')$ is at most $\beta = O(\alpha)$ worse than that under $P_G(w)$.

A SSA with small values of α and β with respect to a small change in w is desirable since that would discourage the players from deviating from their ϵ-equilibrium strategy, however, such an aggregator might not exist in general. The following result shows the γ-aggregator is a SSA.

Theorem 4. Let $\gamma \in (0, 1)$, and $g_G(\cdot, \cdot, \ell, \cdot)$ be the γ-aggregator defined above. Let $P_G(\ell)$ be the property "*the number of maximum permissible intermediaries in a shortest path of length l in G*". Then, g_G is a $(2\eta\kappa_G, \eta\kappa_G)_{P,W_G,L}$- SSA, where $L < W_G$ and κ_G depends on γ and $W_G - L$.

5 Learning formulation

We now formulate an optimization problem to recover the underlying graph structure, the parameters of the aggregator function, and the payoff functions. Let $S = \{a^1, a^2, \ldots, a^M\}$ be our training set, where each strategy profile $a^m \in \{0,1\}^n$ is an ϵ-PSNE, and a^m_i is the action of player i in example $m \in [M]$. Let f be a local aggregator function, and let $a^m_{N_i}$ be the actions of neighbors N_i of player $i \in [n]$ on training example m. We will also represent N as a 0-1 adjacency matrix, with the interpretation that $N_{ij} = 1$ implies that $j \in N_i$, and $N_{ij} = 0$ otherwise. We will use the notation $N_{i\cdot} \triangleq \{N_{ij} : j \neq i\}$. Note that since the underlying game structure is represented as a digraph, N_{ij} and N_{ji} need not be equal. Let h be a concave function such that $h(0) = 0$. Then $F_i(h) \triangleq h(|N_i|)$ is submodular since the concave transformation of the cardinality function results in a submodular function. Moreover $F(h) = \sum_{i \in [n]} F_i(h)$is submodular since it is a sum of submodular functions. We will use $F(h)$ as a sparsity-inducing prior. Several choices of h have been advocated in the literature, including suitably normalized geometric, log, smooth log and square root functions [15].

We would denote the parameters of the aggregator function f by θ_f. The payoff functions will depend on the choice of this parameterization. For a fixed aggregator f (such as the sum aggregator), linear parameterization is one possibility, where the payoff function for player $i \in [n]$ takes the form,

$$u_i^f(a^m, N_{i\cdot}) = a_i^m w_{i1}(w_f\, f(a^m_{N_i}) + b_f) + (1 - a_i^m) w_{i0}(w_f\, f(a^m_{N_i}) + b_f),$$

where $w_{i\cdot} = (w_{i0}, w_{i1})^\top$ and $N_{i\cdot}$ denote the independent parameters for player i and $\theta_f = (w_f, b_f)^\top$ are the shared parameters. Our setting is flexible, and we can easily accommodate more complex aggregators instead of the standard aggregators (e.g. sum). Exchangeable functions over sets [16] provide one such example. An interesting instantiation is a neural network comprising one hidden layer, an output sum layer, with tied weights. Specifically, let $W \in \mathbb{R}^{n \times (n-1)}$ where all entries of W are equal to w_{NN}. Let σ be an element-wise non-linearity (e.g. we used the ReLU function, $\sigma(x) = \max\{x, 0\}$ for our experiments). Then, using the element-wise multiplication operator $\odot$ and a vector $\mathbf{1}$ with all ones, u_i may be expressed as $u_i^{f_{NN}}(a^m, N_{i\cdot}) = a_i^m w_{i1} f_{NN}(a^m_{N_i}) + (1 - a_i^m) w_{i0} f_{NN}(a^m_{N_i})$, where the permutation invariant neural aggregator, parameterized by $\theta_{f_{NN}} = (w_{NN}, b_{NN})^\top$,

$$f_{NN}(a^m_{N_i}) = \mathbf{1}^\top \sigma(W\, a^m_{-i} \odot N_{i\cdot} + b_{NN}).$$

We could have more complex functions such as deeper neural nets, with parameter sharing, at the expense of increased computation. We believe this versatility makes local aggregative games particularly attractive, and provides a promising avenue for modeling structured strategic settings.

Each a^m is an ϵ-PSNE, so it ensures a locally (near) optimal reward for each player. We will impose a margin constraint on the difference in the payoffs when player i unilaterally deviates from a_i^m. Note that $N_i = \{j \in N_{i\cdot} : N_{ij} = 1\}$. Then, introducing slack variables ξ_i^m, and hyperparameters $C, C', C_f > 0$, we obtain the following optimization problem in $O(n^2)$ variables:

$$\begin{aligned}
&\min_{\theta_f, w_{1\cdot}, \ldots, w_{n\cdot}, N_{1\cdot}, \ldots, N_{n\cdot}} && \frac{1}{2}\sum_{i=1}^n ||w_{i\cdot}||^2 + \frac{C_f}{2M}||\theta_f||^2 + \frac{C'}{n}\sum_{i=1}^n F_i(h) + \frac{C}{M}\sum_{i=1}^n \sum_{m=1}^M \xi_i^m \\
&\text{s.t.} \quad \forall i \in [n], m \in [M]: && u_i^f(a^m, N_{i\cdot}) - u_i^f(1 - a^m, N_{i\cdot}) \geq e(a^m, a') - \xi_i^m \\
&\quad\quad\ \forall i \in [n], m \in [M]: && \xi_i^m \geq 0 \\
&\quad\quad\ \forall i \in [n]: && N_{i\cdot} \in \{0,1\}^{n-1},
\end{aligned}$$

where a^m and a' differ in exactly one coordinate, and e is a margin specific loss term, such as Hamming loss $e_H(a, \tilde{a}) = 1\{a \neq \tilde{a}\}$ or scaled 0-1 loss $e_s(a, \tilde{a}) = 1\{a \neq \tilde{a}\}/n$. From a game theoretic perspective, the scaled loss has a natural asymptotic interpretation: as the number of players $n \to \infty$, $e_s(a^m, a') \to 0$, and we get $\forall i \in [n], m \in [M] :\ u_i^f(a^m, N_{i\cdot}) \geq u_i^f(1 - a^m, N_{i\cdot}) - \xi_i^m$, i.e., each training example a^m is an ϵ-PSNE, where $\epsilon = \max_{i \in [n], m \in [M]} \xi_i^m$.

Once θ_f are fixed, the problem clearly becomes separable, i.e., each player i can solve an independent sub-problem in $O(n)$ variables. Each sub-problem includes both continuous and binary variables, and may be solved via branch and bound, outer approximation decomposition, or extended cutting plane methods (see [17; 18] for an overview of these techniques). The individual solutions can be forced to agree on θ_f via a standard dual decomposition procedure, and methods like alternating direction method of multipliers (ADMM) [19] could be leveraged to facilitate rapid agreement of the continuous parameters w_f and b_f. The extension to learning the γ-aggregative games is immediate.

We now describe some other optimization variants for the local aggregative games. Instead of constraining each player to a hard neighborhood, one might relax the constraints $N_{ij} \in \{0, 1\}$ to $N_{ij} \in [0, 1]$, where N_{ij} might be interpreted as the strength of the edge (j, i). The Lovász convex relaxation of F [20] is a natural prior for inducing sparsity in this case. Specifically, for an ordering of values $|N_{i(0)}| \geq |N_{i(1)}| \ldots \geq |N_{i(n-1)}|$, $i \in [n]$, this prior is given by

$$\Gamma_h(N) = \sum_{i=1}^{n} \Gamma_h(N, i), \text{ where } \Gamma_h(N, i) = \sum_{k=0}^{n-1} [h(k+1) - h(k)] |N_{i(k)}|.$$

Since the transformation h encodes the preference for each degree, $\Gamma_h(N)$ will act as a prior that encourages structured sparsity. One might also enforce other constraints on the structure of the local aggregative game. For instance, an undirected graph could be obtained by adding constraints $N_{ij} = N_{ji}$, for $i \in [n], j \neq i$. Likewise, a minimum in-degree constraint may be enforced on player i by requiring $\sum_j N_{ij} \geq d$. Both these constraints are linear in $N_{i\cdot}$, and thus do not add to the complexity of the problem. Finally, based on cues such as domain knowledge, one may wish to add a degree of freedom by not enforcing sharing of the parameters of the aggregator among the players.

6 Experiments

We now present strong empirical evidence to demonstrate the efficacy of local aggregative games in unraveling the aggregative game structure of two real voting datasets, namely, the US Supreme Court Rulings dataset and the Congressional Votes dataset. Our experiments span the different variants for recovering the structure of the aggregative games including settings where (a) parameters of the aggregator are learned along with the payoffs, (b) in-degree of each node is lower bounded, (c) γ-discounting is used, or (d) parameters of the aggregator are fixed. We will also demonstrate that our method compares favorably with the potential games method for tree structured games [4], even when we relax the digraph setting to let weights $N_{ij} \in [0, 1]$ instead of $\{0, 1\}$ or force the game structure to be undirected by adding the constraints $N_{ij} = N_{ji}$. For our purposes, we used the smoothed square-root concave function, $h(i) = \sqrt{i+1} - 1 + \alpha i$ parameterized by α, the sum and neural aggregators, and the scaled 0-1 loss function $e_s(a, \tilde{a}) = 1\{a \neq \tilde{a}\}/n$. We found our model to perform well across a very wide range of hyperparameters. All the experiments described below used the following setting of values: $\alpha = 1$, $C = 100$, and $C_f = 1$. C' was also set to 0.01 in all settings except when the parameters of the aggregator were fixed, when we set $C' = 0.01\sqrt{n}$.

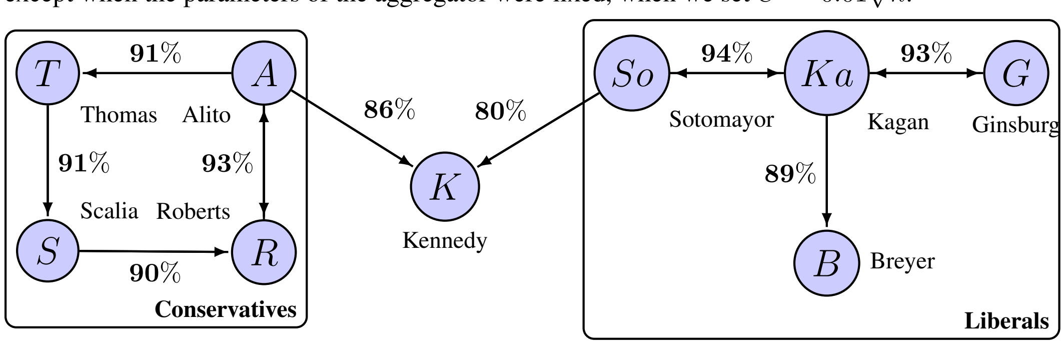

Figure 1: **Supreme Court Rulings (full bench):** The digraph recovered by the local aggregative and γ-aggregative games ($\ell \leq 2$, all γ) with the sum aggregator as well as the neural aggregator is consistent with the known behavior of the Justices: conservative and liberal sides of the bench are well segregated from each other, while the moderate Justice Kennedy is positioned near the center. Numbers on the arrows are taken from an independent study [21] on Justices' mutual voting patterns.

6.1 Dataset 1: Supreme Court Rulings

We experimented with a dataset containing all non-unanimous rulings by the US Supreme court bench during the year 2013. We denote the Justices of the bench by their last name initials, and add a second character to some names to avoid the conflicts in the initials: Alito (A), Breyer (B), Ginsburg(G), Kennedy (K), Kagan (Ka), Roberts (R), Scalia (S), Sotomayor (So), and Thomas (T). We obtained a binary dataset following the procedure described in [4].

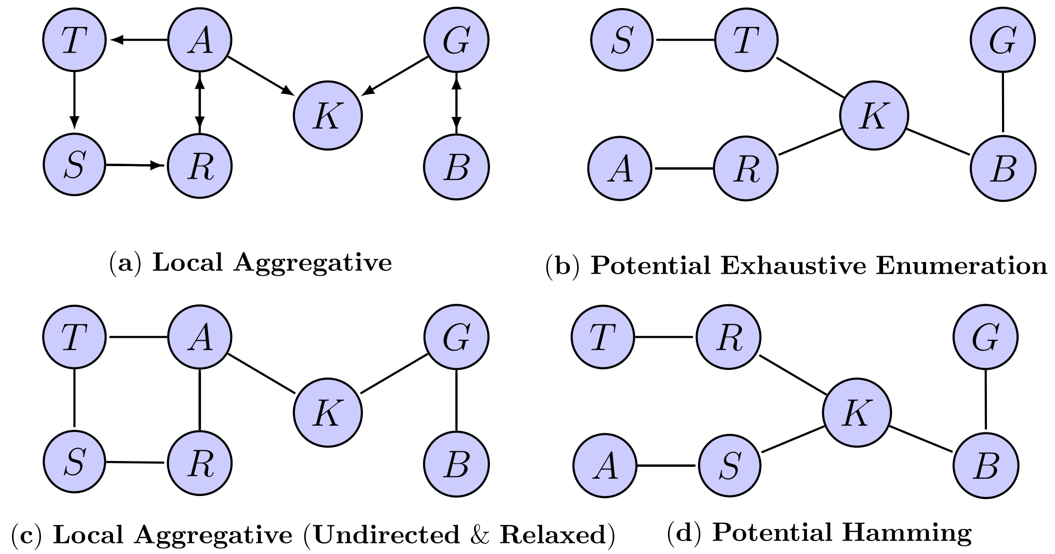

Figure 2: **Comparison with the potential games method [4]:** (a) The digraph produced by our method with the sum as well as the neural aggregator is consistent with the expected voting behavior of the Justices on the data used by [4] in their experiments. (c) Relaxing all $N_{ij} \in [0,1]$ and enforcing $N_{ij} = N_{ji}$ still resulted in a meaningful undirected structure. (b) & (d) The tree structures obtained by the brute force and the Hamming distance restricted methods [4] fail to capture higher order interactions, e.g., the strongly connected component between Justices A, T, S and R.

.

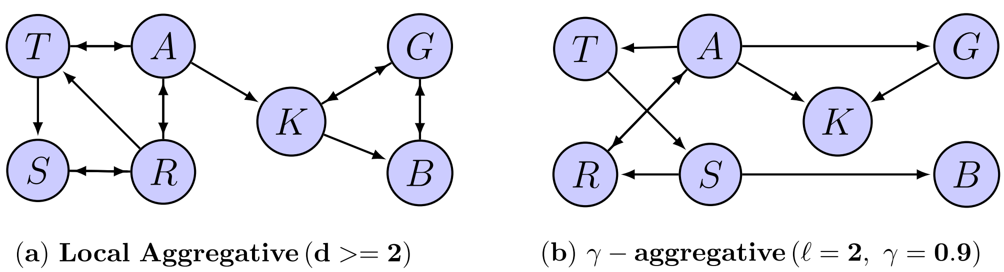

Figure 3: **Degree constrained and γ-aggregative games**: (a) Enforcing the degree of each node to be at least 2 reinforces the intra-republican and the intra-democrat affinity, reaffirming their respective jurisprudences, and (b) γ-aggregative games also support this observation: the same digraph as Fig. 2(a) is obtained unless ℓ and γ are set to high values (plot generated with $\ell = 2$, $\gamma = 0.9$), when the strong effect of one-hop and two-hop neighbors overpowers the direct connection between B and G.

Fig. 1 shows the structure recovered by the local aggregative method. The method was able to distinguish the conservative side of the court (Justices A, R, S, and T) from the left side (B, G, Ka, and So). Also, the structure places Justice Kennedy in between the two extremes, which is consistent with his moderate jurisprudence. To put our method in perspective, we also compare the result of applying our method on the same subset of the full bench data that was considered by [4] in their experiments. Fig. 2 demonstrates how the local aggregative approach estimated meaningful structures consistent with the full bench structure, and compared favorably with both the methods of [4]. Finally, Fig. 3(a) and 3(b) demonstrate the effect of enforcing minimum in-degree constraints in the local aggregative games, and increasing ℓ and γ in the γ-aggregative games respectively. As expected, the estimated γ-aggregative structure is stable unless γ and ℓ are set to high values when non-local effects kick in. We provide some additional results on the degree-constrained local aggregative games (Fig. 4) and the γ-aggregative games (Fig. 5). In particular, we see that the γ-aggregative games are indeed robust to small changes in the aggregator input as expected in the light of stability result of Theorem 4.

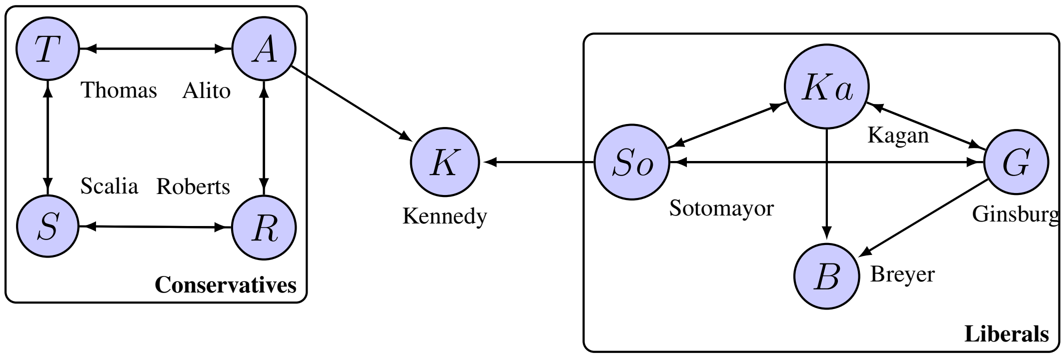

Figure 4: **Degree constrained local aggregative games (full bench):** The digraph recovered by the local aggregative method when the degree of each node was constrained to be at least 2. Clearly, the cohesion among the Justices on the conservative side got strengthened by the degree constraint (likewise for the liberal side of the bench). On the other hand, no additional edges were added between the two sides.

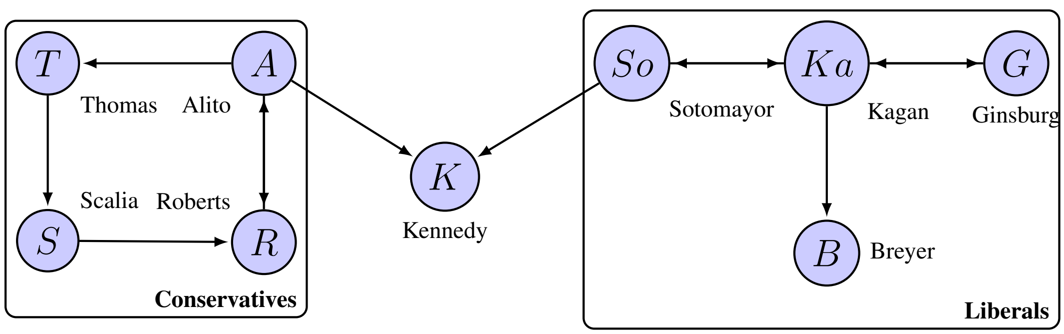

Figure 5: **γ-Aggregative Games (full bench):** The digraph estimated by the γ-aggregative method for $\ell = 2, \gamma = 0.9$, and lower values of γ and/or ℓ. Note that an identical structure was obtained by the local aggregative method (Fig. 1). This indicates that despite heavily weighting the effect of the nodes on a shortest path with one or two intermediary hops, the structure in Fig. 1 is very stable. Also, this substantiates our theoretical result about the stability of the γ-aggregative games.

6.2 Dataset 2: Congressional Votes

We also experimented with the Congressional Votes data [22], that contains the votes by the US Senators on all the bills of the 110 US Congress, Session 2. Each of the 100 Senators voted in favor of (treated as 1) or against each bill (treated as 0). Fig. 6 shows that the local aggregative method provides meaningful insights into the voting patterns of the Senators as well. In particular, few connections exist between the nodes in red and those in blue, making the bipartisan structure quite apparent. In some cases, the intra-party connections might be bolstered due to same state affiliations, e.g. Senators Corker (28) and Alexander (2) represent Tennessee. The cross connections may also capture some interesting collaborations or influences, e.g., Senators Allard (3) and Clinton (22) introduced the Autism Act. Likewise, Collins (26) and Carper (19) reintroduced the Fire Grants Reauthorization Act. The potential methods [4] failed to estimate some of these strategic interactions. Likewise, Fig. 7 provides some interesting insights regarding the ideologies of some Senators that follow a more centrist ideology than their respective political affiliations would suggest.

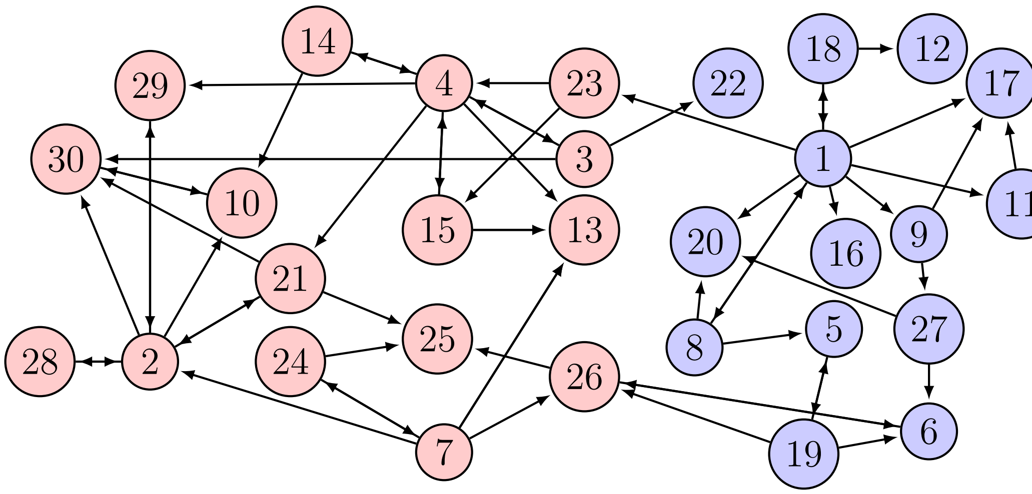

Figure 6: **Comparison with [4] on the Congressional Votes data:** The digraph recovered by local aggregative method, on the data used by [4], when the parameters of the sum aggregator were fixed ($w_f = 1$, $b_f = 0$). The segregation between the Republicans (shown in red) and the Democrats (shown in blue) strongly suggests that they are aligned according to their party policies.

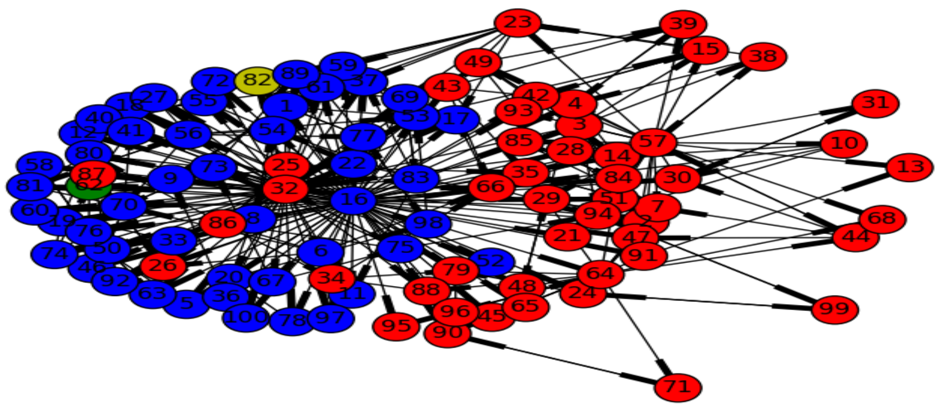

Figure 7: **Complete Congressional Votes data:** The digraph recovered on fixing parameters, relaxing N_{ij} to $[0, 1]$, and thresholding at 0.05. The estimated structure not only separates majority of the reds from the blues, but also associates closely the then independent Senators Sanders (82) and Lieberman (62) with the Democrats. Moreover, the few reds among the blues generally identify with a more centrist ideology - Collins (26) and Snowe (87) are two prominent examples.

Conclusion

An overwhelming majority of literature on machine learning is restricted to modeling non-strategic settings. Strategic interactions in several real world systems such as decision/voting often exhibit local structure in terms of how players are guided by or respond to each other. In other words, different agents make rational moves in response to their neighboring agents leading to locally stable configurations such as Nash equilibria. Another challenge with modeling the strategic settings is that they are invariably unsupervised. Consequently, standard learning techniques such as structured prediction that enforce global consistency constraints fall short in such settings (cf. [4]). As substantiated by our experiments, local aggregative games nicely encapsulate various strategic applications, and could be leveraged as a tool to glean important insights from voting data. Furthermore, the stability of approximate equilibria is a primary consideration from a conceptual viewpoint, and the γ-aggregative games introduced in this work add a fresh perspective by achieving structural stability.

References

[1] J. Lafferty, A. McCallum, and F. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data, *ICML*, 2001.

[2] B. Taskar, C. Guestrin, and D. Koller. Max-margin Markov networks, *NIPS*, 2003.

[3] I. Tsochantaridis, T. Joachims, T. Hofmann, and Y. Altun. Large margin methods for structured and interdependent output variables, *JMLR*, 6(2), pp. 1453-1484, 2005.

[4] V. K. Garg and T. Jaakkola. Learning Tree Structured Potential Games, *NIPS*, 2016.

[5] C. Daskalakis and C. H. Papadimitriou. Approximate Nash equilibria in anonymous games, *Journal of Economic Theory*, 156, pp. 207-245, 2015.

[6] J. Nash. Non-Cooperative Games, *Annals of Mathematics*, 54(2), pp. 286-295, 1951.

[7] R. Selten. *Preispolitik der Mehrproduktenunternehmung in der Statischen Theorie*, Springer-Verlag, 1970.

[8] M. Kearns and Y. Mansour. Efficient Nash computation in large population games with bounded influence, *UAI*, 2002.

[9] R. Cummings, M. Kearns, A. Roth, and Z. S. Wu. Privacy and truthful equilibrium selection for aggregative games, *WINE*, 2015.

[10] R. Cornes and R. Harley. Fully Aggregative Games, *Economic Letters*, 116, pp. 631-633, 2012.

[11] W. Novshek. On the Existence of Cournot Equilibrium, *Review of Economic Studies*, 52, pp. 86-98, 1985.

[12] M. K. Jensen. Aggregative Games and Best-Reply Potentials, *Economic Theory*, 43, pp. 45-66, 2010.

[13] J. M. Lasry and P. L. Lions. Mean field games, *Japanese Journal of Mathematics*, 2(1), pp. 229-260, 2007.

[14] G. Goel, C. Karande, P. Tripathi, and L. Wang. Approximability of Combinatorial Problems with Multi-agent Submodular Cost Functions, *FOCS*, 2009.

[15] A. J. Defazio and T. S. Caetano. A convex formulation for learning scale-free networks via submodular relaxation, *NIPS*, 2012.

[16] M. Zaheer, S. Kottur, S. Ravanbakhsh, B. Poczos, R. Salakhutdinov, and A. Smola. Deep Sets, *arXiv:1703.06114*, 2017.

[17] P. Bonami et al. An algorithmic framework for convex mixed integer nonlinear programs, *Discrete Optimization*, 5(2), pp. 186-204, 2008.

[18] P. Bonami, M. Kilinç, and J. Linderoth J. Algorithms and Software for Convex Mixed Integer Nonlinear Programs, *Mixed Integer Nonlinear Programming*, The IMA Volumes in Mathematics and its Applications, 154, Springer, 2012.

[19] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3, 2011.

[20] F. Bach. Structured sparsity-inducing norms through submodular functions, *NIPS*, 2010.

[21] J. Bowers, A. Liptak and D. Willis. Which Supreme Court Justices Vote Together Most and Least Often, *The New York Times*, 2014.

[22] J. Honorio and L. Ortiz. Learning the Structure and Parameters of Large-Population Graphical Games from Behavioral Data, *JMLR*, 16, pp. 1157-1210, 2015.

[23] Y. Azrieli and E. Shmaya. Lipschitz Games, *Mathematics of Operations Research*, 38(2), pp. 350-357, 2013.

[24] E. Kalai. Large robust games. *Econometrica*, 72(6), pp. 1631-1665, 2004.

[25] E. N. Gilbert. Random Graphs, *The Annals of Mathematical Statistics*, 30(4), pp. 1141-1144,1959.

[26] W. Feller. *An Introduction to Probability Theory and its Applications*, Vol. 1, Second edition, Wiley, 1957.

[27] M.-F. Balcan and N. J. A. Harvey. Learning Submodular Functions, *STOC*, 2011.

[28] A. Yao. Probabilistic computations: Toward a unified measure of complexity, *FOCS*, 1977.

[29] U. Feige, V. S. Mirrokni, and J. Vondrak. Maximizing non-monotone submodular functions, *FOCS*, 2007.

[30] M. X. Goemans, N. J. A. Harvey, S. Iwata, and V. S. Mirrokni. Approximating submodular functions everywhere, *SODA*, 2009.

[31] Z. Svitkina and L. Fleischer. Submodular approximation: Sampling based algorithms and lower bounds, *FOCS*, 2008.

7 Supplementary (Local Aggregative Games)

We now provide detailed proofs of all the theoretical results stated in the main text.

7.1 Existence of ϵ-PSNE in local aggregative games

We will use the following version of Talagrand's inequality that gives a concentration of convex Lipschitz functions around the median.

Lemma (Talagrand's inequality). Let $P = \mu_1 \otimes \mu_2 \ldots \otimes \mu_n$ be a product probability measure on the Cartesian product $A = A_1 \times A_2 \ldots \times A_n$ of metric spaces (A_i, d_i) equipped with the l^1 metric $d = \sum_{i=1}^{n} d_i$ where each d_i is bounded. Let $a = (a_1, \ldots, a_n)$ be a point in this product space and $F : A \to \mathbb{R}$ be a convex 1-Lipschitz function on (A, d). Then, for any $r > 0$,

$$P(|F - M_F| \geq r) \leq 4e^{-r^2/4},$$

where M_F is the median of F.

Theorem 1. Any local aggregative game on a connected digraph G, where $G \in L(\Delta, n)$ and $\max_i |A_i| \leq m$, admits a $10\Delta\sqrt{\ln(8mn)}$-PSNE.

Proof. The main idea behind the proof is to sample a random strategy profile from a mixed strategy Nash equilibrium of the game (any finite game is guaranteed to have a mixed strategy Nash equilibrium). Specifically, using a concentration argument, we show a *self-purification* [24; 23; 9] result that a profile sampled from a mixed strategy equilibrium is likely to be an approximate pure strategy equilibrium when the Lipschitz constant is small. This which would typically be the case in large population aggregative games, where each player will have a sufficiently large number of neighbors. We follow the probabilistic approach outlined in [23], who introduced Lipschitz games and proved a remarkable high probability result for general games with a small Lipschitz constant. We exploit the convexity of the payoff functions to obtain a tighter bound in the number of players n.

The proof involves three steps. First, using an application of the Talagrand's inequality, we will bound the deviation of each individual payoff function around their median value with high probability. Then, we will obtain a deviation around the mean by bounding the discrepancy between the median and the mean. Finally, we will take a union bound over all players to obtain an approximate pure strategy equilibrium by sampling from the mixed strategy equilibrium.

Let $\mu = (\mu_1, \ldots, \mu_n)$ be a mixed strategy Nash equilibrium of G, where each μ_i is a probability distribution on A_i. Then, the support of each μ_i contains all those pure strategies in A_i that maximize the expected payoff of player i when the other players play according to μ_{-i}. Since the players play their mixed strategies independently, we obtain a product probability measure $P = \prod_i \mu_i$ over the strategy space $A = \prod_i A_i$.

Fix $\epsilon' > 0$ and let $\Delta = \epsilon'/5\sqrt{\ln(8mn)}$. We define event $E_{i,h,\epsilon'}$ to be the set of all strategy profiles a such that by playing strategy h against strategy profile a_{-i} of her neighbors, player $i \in [n]$ receives nearly the same payoff as the median payoff $M_i(h)$ when it plays h and others play their Nash equilibrium strategy:

$$E_{i,h,\epsilon'} = A_i \times \{a_{-i} \in A_{-i} : |u_i(h, f_G(a, i)) - M_i(h)| \leq \epsilon'\}.$$

Likewise we define event $E'_{i,h,\epsilon'}$ where we instead consider the mean payoff $\mathbb{E}_i(h)$ to player i instead of the median payoff, when others play their Nash equilibrium strategy:

$$E'_{i,h,\epsilon'} = A_i \times \left\{ a_{-i} \in A_{-i} : \left| u_i(h, f_G(a, i)) - \underbrace{\int u_i(h, z)\mu_{-i}(dz)}_{\mathbb{E}_i(h)} \right| \leq \epsilon' \right\}.$$

We denote the complement of event E by E^c. Since the payoff function $u_i(\cdot,\cdot)$ is convex and δ_i-Lipschitz in the second argument (where $\delta_i \leq \Delta$) for every fixed instantiation of the first argument, we get from the Talagrand's inequality (following a scaling of u_i by Δ):

$$P(E^c_{i,h,\epsilon'}) \leq 4e^{-\epsilon'^2/4\Delta^2}. \tag{2}$$

Now, we can bound the deviation between the median $M_i(h)$ and the expected value $\mathbb{E}_i(h)$ using a standard result to obtain

$$P(E^{'c}_{i,h,\epsilon'}) \leq 8e^{-\epsilon'^2/16\Delta^2} < \frac{1}{mn},$$

for every $i \in [n]$ and $h \in A_i$. Since there are only n players with at most m strategies each, we immediately note on taking a union bound over players i and their strategies $h \in A_i$, that there is some non-zero probability that the players play a mixed strategy ϵ'-equilibrium. Therefore, we can sample a pure-strategy profile a^* from the support of μ such that $a^* \in \cap_{i\in[n],h\in A_i} E^{'}_{i,h,\epsilon'}$.

The arguments in [23] can be recycled to show that a^* is a pure $2\epsilon'$-equilibrium. We reproduce these arguments for sake of completeness. Consider any player i and strategy $h'_i \in A_i \setminus \{a^*_i\}$. Since $a^* \in E^{'}_{i,h'_i,\epsilon'}$ and the support of μ_i contains only those pure strategies in A_i that maximize the expected payoff of player i when others play according to μ_{-i}, we note that

$$\begin{aligned}
u_i(h'_i, f_G(a^*,i)) &\leq \int u_i(h'_i,z)\mu_{-i}(dz) + \epsilon' && \left(\text{since } a^* \in E^{'}_{i,h'_i,\epsilon'}\right)\\
&\leq \int u_i(a^*_i,z)\mu_{-i}(dz) + \epsilon' && (\text{since } a^*_i \text{ is in the support of } \mu_i)\\
&\leq u_i(a^*_i, f_G(a^*,i)) + 2\epsilon' && \left(\text{since } a^* \in E^{'}_{i,a^*_i,\epsilon'}\right)
\end{aligned}$$

The proof is complete since our choice of i and h'_i was arbitrary, and $\Delta = \epsilon'/5\sqrt{\ln(8mn)}$.

□

7.1.1 Extending the result to submodular functions

The result of Theorem 1 can be extended to submodular functions. Balcan and Harvey [27] proved the following result for a certain class of submodular functions. Specifically,

Theorem (Balcan & Harvey [27]). Let $F : 2^{[n]} \to \mathbb{R}^+$ be a non-negative, monotone, submodular, 1-Lipschitz function, and let $X \in [n]$ have a product distribution. Then for any $b, t \geq 0$,

$$P(F(X) \leq b - t\sqrt{b}) \cdot P(F(X) \geq b) \leq e^{-t^2/4}.$$

Defining $m = b - t\sqrt{b}$, and setting $b = 1$, this immediately yields the following for all $t \geq 0$:

$$P(F(X) \leq m)P(F(X) \geq m + t) \leq e^{-t^2/4}.$$

Let M_F be the median of F. Since $P(F \geq M_F) = 1/2 = P(F \leq M_F)$, invoking this inequality twice with $m = M_F$ and $m = M_F - t$, where $t \geq 0$, we immediately get the following concentration inequality which is of identical form as the Talagrand concentration result in Lemma 7.1:

$$P(|F - M_F| \geq t) \leq 4e^{-t^2/4}.$$

Therefore, the result regarding existence of pure strategy Nash equilibrium under convex Lipchitz assumption on the individual payoff functions in Theorem 1 carries over nicely to non-negative, monotone, submodular, Lipschitz functions as well.

7.2 Information theoretic lower bounds

Lemma 2. Consider a directed random graph $DG(n,p)$ where $p \in (0,1)$ is the probability of choosing any directed edge independently of others. Define $q = 1-p$. Let P_n be the probability that DG is connected. Then, the probability that DG is disconnected is $1 - P_N = nq^{2(n-1)} + O\left(n^2 q^{3n}\right)$.

Proof. Gilbert [25] proved an elegant result for bounding the probability that an Erdős-Rényi graph is disconnected. We will adapt his proof for our setting. Consider any node i. For i to be in a component of size k, there must be exactly $(k-1)$ other nodes in the component. Moreover, there should not be any edge between this component and any of the other $(n-k)$ nodes. The number of these missing edges is exactly $2k(n-k)$ when we account for the direction. Therefore, we must have the following recurrence relation:

$$1 - P_n = \sum_{k=1}^{n-1} \binom{n-1}{k-1} P_k q^{2k(n-k)} \tag{3}$$

First we bound this quantity from above. [25] noted that since $x(N-x)$ is convex, we have:

$$2k(n-k) \geq \begin{cases} (n-2)k + n & \text{if } 1 \leq k \leq n/2 \\ (n-2)(n-k) + n & \text{if } n/2 \leq k \leq n-1. \end{cases} \tag{4}$$

Since $q < 1$, we can decompose the sum on right side of (3) into two sums to obtain

$$1 - P_n \leq \sum_{k=1}^{n-1} \binom{n-1}{k-1} q^{2k(n-k)} \leq \underbrace{\sum_{k=1}^{n/2} \binom{n-1}{k-1} q^{2k(n-k)}}_{(A)} + \underbrace{\sum_{k=n/2+1}^{n-1} \binom{n-1}{k-1} q^{2k(n-k)}}_{(B)}.$$

We will bound these two sums (A) and (B) separately. Note that, using (4),

$$\begin{aligned} (A) &= \sum_{k=1}^{n/2} \binom{n-1}{k-1} q^{2k(n-k)} \leq \sum_{k=1}^{n/2} \binom{n-1}{k-1} q^{(n-2)k+n} = \sum_{j=0}^{n/2-1} \binom{n-1}{j} q^{(n-2)(j+1)+n} \\ &= q^{2(n-1)} \sum_{j=0}^{n/2-1} \binom{n-1}{j} q^{(n-2)j} \leq q^{2(n-1)} \left[\sum_{j=0}^{n-1} \binom{n-1}{j} q^{(n-2)j} - \binom{n-1}{n-1} q^{(n-2)(n-1)} \right] \\ &= q^{2(n-1)} \left[\left(1 + q^{(n-2)}\right)^{n-1} - q^{(n-2)(n-1)} \right]. \end{aligned}$$

Moreover, since $\binom{n-1}{k-1} = \binom{n-1}{n-k}$, using (4),

$$\begin{aligned} (B) &= \sum_{k=n/2+1}^{n-1} \binom{n-1}{k-1} q^{2k(n-k)} \leq \sum_{k=n/2+1}^{n-1} \binom{n-1}{n-k} q^{(n-2)(n-k)+n} \\ &= q^n \sum_{k=n/2+1}^{n-1} \binom{n-1}{n-k} q^{(n-2)(n-k)} = q^n \sum_{j=1}^{n/2-1} \binom{n-1}{j} q^{(n-2)j} \\ &= q^n \left[\sum_{j=0}^{n/2-1} \binom{n-1}{j} q^{(n-2)j} - 1 \right] \leq q^n \left[\left(1 + q^{(n-2)}\right)^{n-1} - 1 \right]. \end{aligned}$$

Adding (A) and (B), we get an upper bound on the probability that the graph is disconnected:

$$1 - P_n \leq q^{2(n-1)} \left[\left(1 + q^{(n-2)}\right)^{n-1} - q^{(n-2)(n-1)} \right] + q^n \left[\left(1 + q^{(n-2)}\right)^{n-1} - 1 \right]. \quad (5)$$

On the other hand, the lower bound denotes the probability that some node is *isolated*, i.e. it does not have any incoming or outgoing edge. In other words, the lower bound L is dictated by the union of the individual events E_i that the node i is isolated. Feller attributes the following inequality to Bonferroni [26]:

$$L \geq \sum_i P(E_i) - \sum_{i<j} P(E_i E_j).$$

Clearly, $P(E_i) = q^{2(n-1)}$ since i does not share any edges in either direction with the remaining $(n-1)$ vertices. Now, the event $E_i \cap E_j$ happens when nodes i and j do not have any edge between them, and with any of the other $(n-2)$ vertices. Therefore, the total number of missing edges is $2 + 2 * 2(n-2) = 2(2n-3)$. Since there are $\binom{n}{2}$ such pairs (i, j), we immediately get the lower bound:

$$1 - P_n \geq L \geq nq^{2(n-1)} - \binom{n}{2} q^{2(2n-3)} = nq^{2(n-1)} \left[1 - \binom{n-1}{2} q^{2(n-2)} \right]. \quad (6)$$

The statement of the lemma follows by combining the bounds from (5) and (6). □

Theorem 3. Let $\epsilon > 0$, and α, $\delta \in (0, 1)$. Let n be the number of players in a local aggregative game, where each player $i \in [n]$ is provided with some convex Δ-Lipschitz function u_i and an aggregator A. Let $D_n \triangleq D_n(\Delta, \epsilon, A, (u_i)_{i \in [n]})$ be the sufficient in-degree (number of incoming edges) of each player such that the game admits some ϵ-PSNE when the players play to maximize their individual payoffs u_i according to the local information provided by the aggregator A. Assume any non-negative monotone submodular cost function on the edge set cardinality. Then for any $d \geq \max\{D_n, n^\alpha \ln n\}/(1-\alpha)$, any randomized algorithm that approximates the game structure to a factor $n^{1-\alpha}/(1+\delta)d$ requires exponentially many queries under the value oracle model.

Proof. The main idea is to construct a directed graph that has exponentially many spanning directed subgraphs, and define two carefully designed submodular cost functions over the edges of the graph, one of which is deterministic in query size while the other depends on a distribution. We will make it hard for a deterministic algorithm to tell one cost function from the other. This general paradigm [29; 30; 31] can be accomplished by ensuring two conditions: (a) these cost functions map to the same value on *almost* all the queries, and (b) the discrepancy in the optimum value of the functions (on the optimum query) is massive. Thus, since the total number of queries is exponential, we would make it difficult for the non-optimal function to figure out the optimal query when the optimal query would be chosen from a distribution over a large collection of the spanning subgraphs that satisfy the degree constraint with high probability (and thus, in turn, guarantee a pure strategy ϵ-Nash equilibrium). Our analysis falls under the general framework introduced in [14], where lower bounds on some well known combinatorial problems were proved.

We first construct an good graph instance for our setting. Fix α. Specifically, we consider n^α cliques $DG_1, \ldots, DG_{n^\alpha}$, each with $n^{1-\alpha}$ vertices. We form a spanning graph DG(V, E) by choosing an arbitrary vertex from each clique and then joining these vertices together via edges of arbitrary orientation. We now construct a random subset of edges $DR = \bigcup_{i \in [n]} DR_i$, where each DR_i is obtained by randomly sampling every edge in DG_i independently with probability $p = d/n^{1-\alpha}$. Since each DG_i is a clique on $n'_i \triangleq n^{1-\alpha}$ vertices, and each edge is sampled independently with probability p, we can invoke Lemma 2 on each subset DG_i separately. Then, taking a union bound, it is easy to see that the probability that DR is connected is at least $1 - n \exp\left(-\Omega(n^\alpha \ln n\,(1 - o(1))\right)$.

We now claim that with high probability the degree of each vertex in DG_i restricted to edges in DR_i is at least D_n. Let $deg_i(v)$ be the in-degree of any node v restricted to set DR_i. Invoking the Chernoff's bound on each DG_i, we have

$$P(\exists v \in DG_i : |deg_i(v) - n'_i p| \geq \alpha n'_i p) \leq \underbrace{2n' \exp(-\alpha^2 n'_i p/3)}_{\delta_1}.$$

Equivalently, with probability at least 1-δ_1, we have for all $v \in DG_i$:

$$deg_i(v) \in [(1-\alpha)n_i'p, (1+\alpha)n_i'p] = [(1-\alpha)d, (1+\alpha)d].$$

Since $d \geq \dfrac{D_n}{1-\alpha}$, this immediately implies that with high probability, $deg_i(v) \geq D_n$, $\forall v$ in DG_i. Therefore taking a union bound over $DG_i, \ldots, DG_{n^\alpha}$, with probability at least 1 - $n^\alpha \delta_1$, or equivalently $1 - 2n \exp(-\alpha^2 d/3)$, we have that degree of each vertex restricted to DR is at least D_n. Since $d = \Omega(n^\alpha \ln n)$, this bound holds with high probability. This would ensure the existence of an ϵ-Nash equilibrium in the underlying local aggregative game.

Thus far, we have shown a high probability result that DG(V, E), when restricted to DR, simultaneously satisfies both the spanning and degree constraints. Fix δ. We denote the complement of a subset $S \subseteq E$ by $S^c = E \setminus S$. We now define two submodular functions $f_{DR}, g : 2^E \to \mathbb{R}^+$ that score any query $Q \in 2^E$. The cost of optimal solution in f_{DR} is

$$\begin{aligned} f_{DR}(Q) &= \min\{|Q \cap DR^c| + \min\{|Q \cap DR|, (1+\delta)npd\}, nd\} \\ g(Q) &= \min\{|Q|, nd\}. \end{aligned}$$

Since $|Q| = |Q \cap DR| + |Q \cap DR^c|$, we have $f_{DR}(Q) \leq g(Q)$ for all Q. Moreover, since with high probability DR is connected, the cost of optimal spanning graph under f_{DR} is $(1+\delta)npd$. On the other hand, the optimal cost under g is nd. Therefore, we have with high probability that the ratio of the optimal cost in g and that in f_{DR} is at least $\dfrac{1}{(1+\delta)p} = \dfrac{n^{1-\alpha}}{(1+\delta)d}$.

Note that since $f_{DR}(Q) \leq g(Q)$ for all Q, we have $P(f_{DR}(Q) \neq g(Q)) = P(f_{DR}(Q) < g(Q))$. Now we claim that the size of optimal query Q^* is nd. To see this, consider first the case $|Q| \geq nd$. We have $g(Q) = nd$. Therefore,

$$P(f_{DR}(Q) < g(Q)) = P(\min\{|Q \cap DR^c| + \min\{|Q \cap DR|, (1+\delta)npd\}, nd\} < nd).$$

Clearly, this probability increases when we reduce the size of Q. Since $|Q| \geq nd$, we must have $Q^* = nd$. Now consider the other side, i.e. $|Q| \leq nd$. In this case, $g(Q) = |Q|$ and

$$f_{DR}(Q) = |Q \cap DR^c| + \min\{|Q \cap DR|, (1+\delta)npd\}.$$

Therefore, since we sampled the edges randomly, we see via an application of Chernoff's bound that

$$\begin{aligned} P(f_{DR}(Q) < g(Q)) &= P(|Q \cap DR^c| + \min\{|Q \cap DR|, (1+\delta)npd\} < |Q|) \\ &= P(\min\{|Q \cap DR|, (1+\delta)npd\} < |Q \cap DR|) \\ &= P((1+\delta)npd < |Q \cap DR|), \end{aligned}$$

increases when $|Q \cap DR|$ increases which happens when $|Q|$ increases. Therefore, we must have $|Q^*| = nd$ in this case as well. Also,

$$\begin{aligned} P(f_{DR}(Q) < g(Q)) &= P((1+\delta)\mathbb{E}|Q \cap DR| < |Q \cap DR|) \\ &\leq \exp(-\delta^2 npd/3), \end{aligned}$$

which is exponentially small. In other words, the probability that f_{DR} and g can be distinguished by an arbitrary query is exponentially small. The result stated in the theorem then follows immediately from the Yao's minimax principle. □

7.3 Stability in γ-aggregative games

Theorem 4. Let $\gamma \in (0,1)$, and $g_G(\cdot,\cdot,\ell,\cdot)$ be the γ-aggregator defined above. Let $P_G(\ell)$ be the property "*the number of maximum permissible intermediaries in a shortest path of length l in G*". Then, g_G is a $(2\eta\kappa_G, \eta\kappa_G)_{P, W_G, L}$- SSA, where $L < W_G$ and κ_G depends on γ and $W_G - L$.

Proof. The proof proceeds in three steps. First step is to show the existence of an approximate pure strategy Nash equilibrium under $P_G(D_G)$. The proof from Theorem 1 carries over directly while noting that we now instead need to use $\delta^\gamma(G)$ as the Lipschitz constant. The second step is to show that the players are in an approximate pure strategy equilibrium when the aggregator now aggregates using $P_G(L)$. To see this, note that since under $P_G(W_G)$ we have an approximate pure ϵ'-equilibrium profile a^*, and so for any strategy $h'_i \in A_i$ (note that Theorem 1 proved the following result for any deviation from a^*_i, however, the result holds trivially for $h'_i = a^*_i$ as well, and so we can consider an arbitrary action in A_i):

$$u_i(h'_i, g_G(a^*, \gamma, W_G, i)) \leq u_i(a^*_i, g_G(a^*, \gamma, W_G, i)) + \epsilon'. \tag{7}$$

Define $B(\gamma, \ell) \triangleq \sum_{t=0}^{\ell} \gamma^t$, and $C(a, \gamma, \ell, i) = \sum_{t=0}^{\ell} \gamma^t f_G^t(a, i)$. Then, since the payoff functions u_i are η-Lipschitz in the second argument, it is easy to show that for every a'_i,

$$|u_i(a'_i, g_G(a^*, \gamma, \ell, i)) - u_i(a'_i, g_G(a^*, \gamma, W_G, i))| \leq \eta \kappa_G(\gamma, \ell),$$

where

$$\kappa_G(\gamma, \ell) \triangleq \max_i \left| C(a^*, \gamma, \ell, i) \left(\frac{B(\gamma, W_G) - B(\gamma, \ell)}{B(\gamma, W_G) B(\gamma, \ell)} \right) - \frac{C(a^*, \gamma, W_G, i) - C(a^*, \gamma, \ell, i)}{B(\gamma, W_G)} \right|.$$

In particular, substituting $a'_i = h'_i$, we have

$$u_i(h'_i, g_G(a^*, \gamma, L, i)) \leq u_i(h'_i, g_G(a^*, \gamma, W_G, i)) + \eta \kappa_G(\gamma, L). \tag{8}$$

Also, substituting $a'_i = a^*_i$ and using the other direction, we have

$$u_i(a^*_i, g_G(a^*, \gamma, W_G, i)) \leq u_i(a^*_i, g_G(a^*, \gamma, L, i)) + \eta \kappa_G(\gamma, L). \tag{9}$$

Combining (7), (8) and (9), we get that players are playing an $(\epsilon' + 2\eta\kappa_G(\gamma, L))$-PSNE under $P_G(L)$ by sticking to the profile a^*:

$$\begin{aligned} u_i(h'_i, g_G(a^*, \gamma, L, i)) &\leq u_i(h'_i, g_G(a^*, \gamma, W_G, i)) + \eta \kappa_G(\gamma, L) \\ &\leq u_i(a^*_i, g_G(a^*, \gamma, W_G, i)) + \epsilon' + \eta \kappa_G(\gamma, L) \\ &\leq u_i(a^*_i, g_G(a^*, \gamma, L, i)) + \epsilon' + 2\eta \kappa_G(\gamma, L). \end{aligned} \tag{10}$$

Finally, exploiting the η-Lipschitz property again, we immediately get that the payoff of each player does not decrease too much under $P_G(L)$:

$$u_i(a^*_i, g_G(a^*, \gamma, L, i)) \geq u_i(a^*_i, g_G(a^*, \gamma, W_G, i)) - \eta \kappa_G(\gamma, L).$$

Therefore, g_G is $(2\eta\kappa_G(\gamma, \ell), \eta\kappa_G(\gamma, \ell))_{P, W_G, L}$ − structurally stable. □

Note that $\kappa_G(\gamma, \ell)$ is often small when $W_G - L$ is small. For instance, in large population games with the average aggregator $f_G^t(\cdot, \cdot)$, for a fixed γ, $C(\gamma, \ell, i)$ would typically be small compared to $B(\gamma, \ell)$ since the $f_G^t(a, i)$ would generally be much smaller than 1. Then, κ_G is largely determined by the gap between $B(\gamma, \ell)$ and $B(\gamma, W_G)$ which would be small when ℓ is close to W_G. In particular, κ_G would be small when either γ is small (in which case the aggregator behaves increasingly like a local aggregator), or when ℓ is close to W_G. Thus, by controlling ℓ and γ, we can ensure κ_G is small.