Network Effects on Performative Prediction Games

Hoi-To Wai

Department of Systems Engineering and Engineering Management
The Chinese University of Hong Kong (CUHK)

(Joint work with **Xiaolu Wang** and **Chung-Yiu Yau**)

April 6, 2023
NUS-IMS Workshop on Games on Networks
Acknowledgement: HKRGC Project #24203520

Learning from Performative Data

Empirical risk minimization (ERM) for the loss function $\ell : \mathbb{R}^p \times \mathsf{Z} \to \mathbb{R}$

$$\min_{\boldsymbol{\theta}} \; \mathbb{E}_{\boldsymbol{Z} \sim \mathcal{D}} \left[\ell(\boldsymbol{\theta}; \boldsymbol{Z}) \right].$$

- Fixed data distribution $\mathcal{D}$; Example: static data (cats vs dogs).

Performative Prediction (PP)[1]: predictions support decisions that influence the outcome they aim to predict (data *react* to decision),

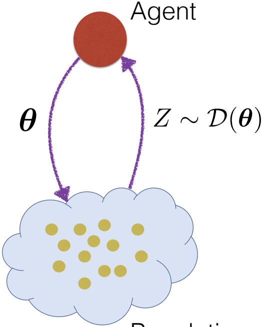
$$\min_{\boldsymbol{\theta}} \; \mathbb{E}_{\boldsymbol{Z} \sim \mathcal{D}(\boldsymbol{\theta})} \left[\ell(\boldsymbol{\theta}; \boldsymbol{Z}) \right].$$

- Decision-dependent distribution: $\mathcal{D}(\boldsymbol{\theta})$.
- Example: bank loan application – Individuals (data) may alter their profiles to increase the chance of success.
- Special case of PP: strategic classification[2].

[1][Perdomo et al., 2020] J. Perdomo, T. Zrnic, C. Mendler-Dunner, M. Hardt. Performative prediction. ICML 2020.

[2]By itself a **Stackelberg game** (agent = leader, population = follower), e.g., $Z \sim \mathcal{D}(\boldsymbol{\theta})$ satisfies $Z \in \arg\max_{\hat{Z}} U(\hat{Z}; \boldsymbol{\theta}, Z_0)$ with $Z_0 \sim \mathcal{D}_0$ (base distribution).

Two Solution Concepts for PP



Performative Optimal Solution (PO):

$$\boldsymbol{\theta}^{\mathsf{PO}} \in \arg\min_{\boldsymbol{\theta} \in \mathbb{R}^p} \; \mathbb{E}_{\boldsymbol{Z} \sim \mathcal{D}(\boldsymbol{\theta})} \left[\ell(\boldsymbol{\theta}; \boldsymbol{Z}) \right].$$

- Difficult $\because$ non-convexity, unknown $\mathcal{D}(\cdot)$, etc.

Performative Stable Solution (PS):

$$\boldsymbol{\theta}^{\mathsf{PS}} \in \arg\min_{\boldsymbol{\theta} \in \mathbb{R}^p} \; \mathbb{E}_{\boldsymbol{Z} \sim \mathcal{D}(\boldsymbol{\theta}^{\mathsf{PS}})} \left[\ell(\boldsymbol{\theta}; \boldsymbol{Z}) \right].$$

- In general $\boldsymbol{\theta}^{\mathsf{PS}} \neq \boldsymbol{\theta}^{\mathsf{PO}}$. Fixed point of repeated risk minimization (RRM)

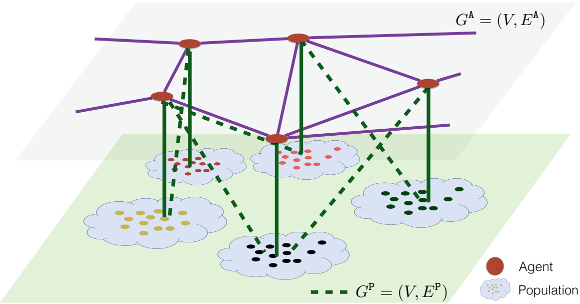
$$\boxed{\boldsymbol{\theta}^{+} \leftarrow \arg\min_{\tilde{\boldsymbol{\theta}} \in \mathbb{R}^p} \mathbb{E}_{\boldsymbol{Z} \sim \mathcal{D}(\boldsymbol{\theta})} \left[\ell(\tilde{\boldsymbol{\theta}}; \boldsymbol{Z}) \right].}$$

- RRM $=$ deployment-and-optimize where agents are **agnostic** to the performative effect.

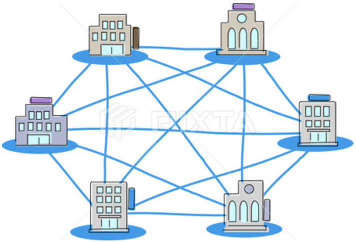
Multi-Agent Performative Prediction (Multi-PP)

This Talk: *multiplex network game* [Gómez-Gardenes et al., 2012] extension of PP with n agents, each agent i interacts with a local population $\mathcal{D}_i(\cdot)$.

- Agent network $\mathcal{G}^{\mathtt{A}}$ described by $\boldsymbol{A}$, where agent i decision depends on $\boldsymbol{\theta}_j, j \in \mathcal{M}_i := \{j : A_{ij} \neq 0\}$.
- Population network $\mathcal{G}^{\mathtt{P}}$ described by $\boldsymbol{P}$, where $\mathcal{D}_i(\cdot)$ react to decisions $\boldsymbol{\theta}_i$ and $\boldsymbol{\theta}_j, j \in \mathcal{N}_i := \{j : P_{ij} \neq 0\}$.

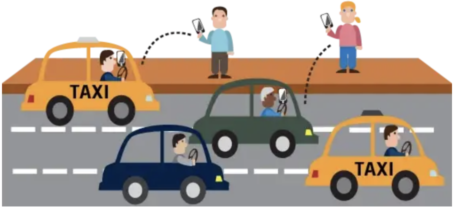


Example: Bank Loan Policy Learning



- Each bank trains a personalized classification model (policy) for predicting whether the loan applicants are creditworthy.
- Banks branches of the same corporate group share strategy to exploit more data $\Rightarrow$ Inter-bank cooperation network $\mathcal{G}^{\mathtt{A}}$.
- Applicants may be affected by local and neighbor branches' policies and manipulate their features to increase the chances of successfully applying for the loan $\Rightarrow$ Applicant influence network $\mathcal{G}^{\mathtt{P}}$.

Example: Ride-Sharing Market



- Multiple **platforms** (agents) forecast **supply-demand** ($\boldsymbol{Z}_i$) for rides at different locations in order to optimize their **revenue** (F_i) by using the forecasted demand to set **prices** ($\boldsymbol{\theta}_i$).
- Drivers/passengers participate in multiple platforms. Hence, the supply-demand vector $\boldsymbol{Z}_i \sim \mathcal{D}_i(\boldsymbol{\theta}_i, \boldsymbol{\theta}_{\mathcal{N}_i})$ for platform i depends on their own price $\boldsymbol{\theta}_i$ as well as their competitors' prices $\boldsymbol{\theta}_{\mathcal{N}_i}$.
- Typical setup: $\mathcal{G}^{\mathtt{A}} = n$-isolated nodes, $\mathcal{G}^{\mathrm{P}} =$ general graph.

Multi-Agent Performative Prediction (Multi-PP)

Setting: each agent minimizes its local risk F_i w.r.t. its own strategy $\boldsymbol{\theta}_i$,

$$\min_{\boldsymbol{\theta}_i \in \mathbb{R}^p} \; F_i(\boldsymbol{\theta}_i, [\boldsymbol{\theta}_j]_{j \in \mathcal{M}_i \cup \mathcal{N}_i}) := \underbrace{\mathbb{E}_{\boldsymbol{Z}_i \sim \mathcal{D}_i(\boldsymbol{\theta}_i, \boldsymbol{\theta}_{\mathcal{N}_i})} [f_i(\boldsymbol{\theta}_i, \boldsymbol{\theta}_{\mathcal{M}_i}; \boldsymbol{Z}_i)]}_{\text{Performative Risk}}, \qquad (1)$$

neighbors' strategies $[\boldsymbol{\theta}_j]_{j \in \mathcal{M}_i}$ are known and samples can be drawn from $\mathcal{D}_i(\boldsymbol{\theta}_i, \boldsymbol{\theta}_{\mathcal{N}_i})$.

- Focus on **personalized learning** [Bellet et al., 2018]:

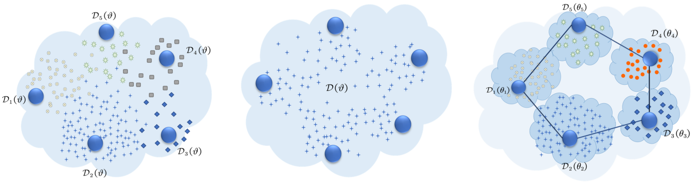
$$f_i(\boldsymbol{\theta}_i, \boldsymbol{\theta}_{\mathcal{M}_i}; \boldsymbol{Z}_i) := \underbrace{\ell_i(\boldsymbol{\theta}_i; \boldsymbol{Z}_i)}_{\text{Loss Function}} + \underbrace{\frac{\rho_i}{2} \sum_{j \in \mathcal{M}_i} A_{ij} \|\boldsymbol{\theta}_i - \boldsymbol{\theta}_j\|_2^2}_{\text{Graph Regularization}}.$$

  - local risk and partial (non)cooperation controlled by[3] $\rho_i \in \mathbb{R}$.

- Interested in equilibrium of the agents' strategies $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_n$ – *performative stable equilibrium (∼PS)* & Nash equilibrium (∼PO).

[3]WLOG, assume that $\boldsymbol{A}$ is normalized with $\sum_j A_{ij} = 1$.

Multi-PP Game: Existing Works



- Agent i optimizes & deploys $\boldsymbol{\theta}_i$, local distribution $\mathcal{D}_i(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_n)$. (left)
  - [Narang et al., 2022] A. Narang E. Faulkner, D. Drusvyatskiy, M. Fazel, L. Ratliff, Multiplayer performative prediction: Learning in decision-dependent games. JMLR, 2022.
- Similar model but identical distribution $\mathcal{D}(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_n)$. (middle)
  - [Piliouras and Yu, 2022] G. Piliouras, F.-Y. Yu. Multi-agent performative prediction: From global stability and optimality to chaos. arXiv, 2022.
- Agents deploy $\boldsymbol{\theta}_1 = \cdots = \boldsymbol{\theta}_n$; distribution $\mathcal{D}_i(\boldsymbol{\theta}_i)$. (right)
  - [Li et al., 2022] Q. Li, C.-Y. Yau, **HT**. Multi-agent performative prediction with greedy deployment and consensus seeking agents. NeurIPS 2022.

- Related works: multi-leader-follower game, multiplex network game, etc.

Questions & Our Results

- **How and when** can we find an (unique) equilibrium? How will the interaction between **topologies** affect the game's equilibrium?
    - we derive the conditions on sensitivity of $\mathcal{D}_i(\cdot)$, (non)cooperation strength ρ, for the existence/uniqueness of equilibriums.
    - symmetric vs asymmetric topology.

- If the data distribution at a local population/agent is perturbed, how will the perturbation affect the equilibrium solution at other agents on the network ($\approx$ **'butterfly effect'**)?
    - for a special case (quadratic loss), we derive closed form solution for the PSE.

Outline

Multi-PP Game: Assumptions

Recall the Multi-PP game: for $i \in [n]$,

$$\min_{\boldsymbol{\theta}_i \in \mathbb{R}^p} \mathbb{E}_{\boldsymbol{Z}_i \sim \mathcal{D}_i(\boldsymbol{\theta}_i, \boldsymbol{\theta}_{\mathcal{N}_i})} \left[\ell_i(\boldsymbol{\theta}_i; \boldsymbol{Z}_i) + \frac{\rho_i}{2} \sum_{j \in \mathcal{M}_i} A_{ij} \|\boldsymbol{\theta}_i - \boldsymbol{\theta}_j\|_2^2 \right].$$

Assumption 1: For $i \in [n]$, it holds

i) $\mathbb{E}_{\boldsymbol{Z}_i \sim \mathcal{D}_i(\boldsymbol{\theta}_i, \boldsymbol{\theta}_{\mathcal{N}_i})}[\ell_i(\cdot; \boldsymbol{Z}_i)]$ is μ_i-**strongly convex**.

ii) $\|\nabla \ell_i(\boldsymbol{\theta}_i; \boldsymbol{Z}_i) - \nabla \ell_i(\boldsymbol{\theta}_i'; \boldsymbol{Z}_i')\|_2 \leq L_i(\|\boldsymbol{\theta}_i - \boldsymbol{\theta}_i'\|_2 + \|\boldsymbol{Z}_i - \boldsymbol{Z}_i'\|_2)$.

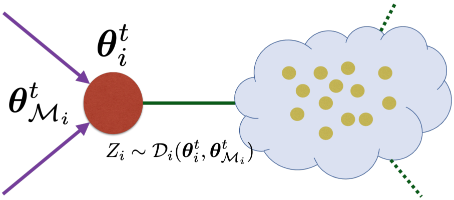
Assumption 2: For $i \in [n]$, there exists $\epsilon_i \geq 0$ such that

$$W_1(\mathcal{D}_i(\boldsymbol{\theta}_i, \boldsymbol{\theta}_{\mathcal{N}_i}), \mathcal{D}_i(\boldsymbol{\delta}_i, \boldsymbol{\delta}_{\mathcal{N}_i})) \leq \epsilon_i \|[\boldsymbol{\theta}_i; \boldsymbol{\theta}_{\mathcal{N}_i}] - [\boldsymbol{\delta}_i; \boldsymbol{\delta}_{\mathcal{N}_i}]\|_2,$$

where $W_1(\cdot, \cdot)$ is the Wasserstein-1 distance.

- Common assumptions for PP problems, see [Perdomo et al., 2020].
- ϵ_i bounds the **sensitivity** of the i-th population $\mathcal{D}_i(\cdot)$.

Repeated Risk Minimization Dynamics



- **Repeated Risk Minimization (RRM)**: In iteration t, agent i does

$$\begin{aligned}\boldsymbol{\theta}_i^{t+1} &= \mathcal{T}_i\left(\boldsymbol{\theta}_i^t, [\boldsymbol{\theta}_j^t]_{j \in \cup \mathcal{M}_i \cup \mathcal{N}_i}\right) \\ &:= \arg\min_{\boldsymbol{\theta}_i \in \mathbb{R}^{p_i}} \mathbb{E}_{\boldsymbol{Z}_i \sim \mathcal{D}_i(\boldsymbol{\theta}_i^t, \boldsymbol{\theta}_{\mathcal{N}_i}^t)} \left[f_i(\boldsymbol{\theta}_i, \boldsymbol{\theta}_{\mathcal{M}_i}^t; \boldsymbol{Z}_i) \right].\end{aligned}$$

- A natural setting for distributed learning (can be extended to SGD-like algorithm).
- Agent i does not need to know $\boldsymbol{\theta}_{\mathcal{N}_i}^t$, but need to know the neighbors' strategies / models $\boldsymbol{\theta}_{\mathcal{M}_i}^t$.

Performative Stable Equilibrium

Definition 1 (Performative Stable Equilibrium, PSE)

The strategy profile $\boldsymbol{\theta}^{\mathsf{pse}} = (\boldsymbol{\theta}_1^{\mathsf{pse}}, \ldots, \boldsymbol{\theta}_n^{\mathsf{pse}}) \in \mathbb{R}^{np}$ is a **performative stable equilibrium** of (1) if for all $i \in [n]$,

$$\boldsymbol{\theta}_i^{\mathsf{pse}} \in \arg\min_{\boldsymbol{\theta}_i \in \mathbb{R}^p} \left\{ \mathbb{E}_{\boldsymbol{Z}_i \sim \mathcal{D}_i(\boldsymbol{\theta}_i^{\mathsf{pse}}, \boldsymbol{\theta}_{\mathcal{N}_i}^{\mathsf{pse}})} \left[f_i(\boldsymbol{\theta}_i, \boldsymbol{\theta}_{\mathcal{M}_i}^{\mathsf{pse}}; \boldsymbol{Z}_i) \right] \right\}.$$

- At the PSE, agent i has no incentive to alter $\boldsymbol{\theta}_i^{\mathsf{pse}}$ based only on response $\mathcal{D}_i(\boldsymbol{\theta}_i^{\mathsf{pse}}, \boldsymbol{\theta}_{\mathcal{N}_i}^{\mathsf{pse}})$.
- Observation: PSE is a fixed point of RRM, but when does PSE exist and is unique? depends on the map $\mathcal{T}_i(\cdot)$...

Existence and Uniqueness of PSE

Theorem 2

Suppose that $\sum_{j=1}^{n} A_{ij} = 1$ *and* $\mu_i + \rho_i > 0$ *for all* $i \in [n]$*, Assumptions 1 and 2 hold. Let* $\boldsymbol{\mu} := [\mu_i]_{i=1}^{n}$ *and* $\boldsymbol{\rho} := [\rho_i]_{i=1}^{n}$*. Under the condition*

$$\sqrt{\max_{j\in[n]} \sum_{i=1}^{n} \left(\frac{P_{ij} L_i \epsilon_i}{\mu_i + \rho_i} \right)^2} + \left\| \mathrm{Diag}\left(\frac{\boldsymbol{\rho}}{\boldsymbol{\mu} + \boldsymbol{\rho}} \right) \boldsymbol{A} \right\|_2 < 1, \tag{2}$$

(i) *the Multi-PP game admits a unique PSE, and* (ii) *the RRM converges linearly to the PSE.*

- Eq. (2) gives **sufficient condition** for stability of RRM.
- *Stability* of RRM depends on $\mathcal{G}^{\mathtt{A}}, \mathcal{G}^{\mathtt{P}}$, ρ_i and ϵ_i; see next slides for elaboration.

Effects of Network Structure on PSE: Non Graph Regularized Cases ($\rho = 0$)

- If $n = 1$, $\rho_1 = 0$, and $\mu_1 > 0$, then

$$(2) \iff \epsilon_1 < \mu_1 / L_1,$$

which coincides with single-agent PP [Perdomo et al., 2020, Theorem 3.5].

- If $\boldsymbol{P} = \mathbf{1}\mathbf{1}^\top$ ($\mathcal{G}^{\mathsf{P}}$ is fully connected) and $\rho_i = 0$, then

$$(2) \iff \textstyle\sum_{i=1}^n L_i^2 \epsilon_i^2 / \mu_i^2 < 1$$

This coincides with [Narang et al., 2022, Theorem 2]. If further $\epsilon_i = \epsilon$, $L_i = L$, $\mu_i = \mu$, then

$$(2) \iff \epsilon < \mu / (\sqrt{n} L)$$

Effects of Network Structure on PSE: Graph Regularized Cases ($\rho \neq 0$)

Suppose that $\mu_i = \mu$, $\rho_i = \rho$, $L_i = L$. In this case, (2) can be implied by

$$L\sqrt{\|\boldsymbol{P}\|_\infty}\,\max_{i\in[n]} \epsilon_i < \mu - \rho(\|\boldsymbol{A}\|_2 - 1), \tag{3}$$

- Population network with less edges and small L can be beneficial for stability.
- If $\boldsymbol{A}$ is symmetric[4], then $\|\boldsymbol{A}\|_2 = 1$ and thus (3) is independent of ρ.
- If $\boldsymbol{A}$ is asymmetric, then $\|\boldsymbol{A}\|_2 > 1$ and increasing ρ may violate (3) (although this may have better generalization performance).

- Intuition? a possible reason is that RRM is no longer 'fair' for all agents (see the SG-GD algorithm).

[4] Recall that $\sum_j A_{ij} = 1$ still holds.

Stochastic Algorithm for Computing PSE

Algorithm 1 Stochastic Gradient with Greedy Deployment (SG-GD)

for $t = 0, 1, \ldots$ **do**
  Deploy the models $\{\boldsymbol{\theta}_i^t\}_{i=1}^n$.
  for $i = 1$ **to** n **do** {`executed in parallel`}
    Sample $\boldsymbol{Z}_i^{t+1} \sim \mathcal{D}_i(\boldsymbol{\theta}_i^t, \boldsymbol{\theta}_{\mathcal{N}_i}^t)$
    $\boldsymbol{g}^t = \nabla \ell_i(\boldsymbol{\theta}_i^t; \boldsymbol{Z}_i^{t+1}) + \rho_i \sum_{j=1}^n A_{ij} \left(\boldsymbol{\theta}_i^t - \boldsymbol{\theta}_j^t\right)$ {`decen. opt.`}
    $\boldsymbol{\theta}_i^{t+1} = \boldsymbol{\theta}_i^t - \gamma_{t+1} \boldsymbol{g}^t$

Theorem 3

Suppose that $\mathbb{E}[\|\nabla \boldsymbol{\ell}(\boldsymbol{\theta}; \boldsymbol{Z}) - \mathbb{E}[\nabla \boldsymbol{\ell}(\boldsymbol{\theta}; \boldsymbol{Z})]\|_2^2] \leq \sigma_0^2 + \sigma_1^2 \|\boldsymbol{\theta} - \boldsymbol{\theta}^{pse}\|_2^2$ *and the same conditions as Theorem 2 holds, then for all* $t \geq 1$,

$$\mathbb{E}[\|\boldsymbol{\theta}^t - \boldsymbol{\theta}^{pse}\|_2^2] \leq \prod_{s=1}^t (1 - \gamma_s \widetilde{\mu}) \Delta^0 + \frac{2\sigma_0^2}{\widetilde{\mu}} \gamma_t. \tag{4}$$

where $\Delta_0 := \|\boldsymbol{\theta}^0 - \boldsymbol{\theta}^{pse}\|_2^2$, $\widetilde{\mu} = \mu + \rho(1 - \|\boldsymbol{A}\|_2) - L\epsilon\sqrt{\|\boldsymbol{P}\|_\infty}$, *and* $\widetilde{\sigma}^2 = \sigma_1^2 + 2\Big(L^2\epsilon^2 \|\boldsymbol{P}\|_\infty + (L + \rho\|\boldsymbol{I}_n - \boldsymbol{A}\|_2)^2\Big)$.

Case Study: Quadratic Loss Game

- Consider the loss function as

$$\ell_i(\boldsymbol{\theta}_i; \boldsymbol{Z}_i) = \frac{1}{2}\|\boldsymbol{\theta}_i - \boldsymbol{Z}_i\|_2^2. \quad (5)$$

  with the graph regularization parameter $\rho_i = \rho \geq 0$.

- The sample $\boldsymbol{Z}_i \sim \mathcal{D}_i(\boldsymbol{\theta}_i, \boldsymbol{\theta}_{\mathcal{N}_i})$ satisfies

$$\boldsymbol{Z}_i = \underbrace{\bar{\boldsymbol{Z}}_i}_{\text{'base' distribution}} + \varepsilon \underbrace{\sum_{j=1}^{n} P_{ij}\boldsymbol{\theta}_j}_{\text{influences from neighbors on } \mathcal{G}^{\text{P}}}, \quad (6)$$

  where $\varepsilon \in \mathbb{R}$ is a sensitivity parameter (can be negative!) and

$$\mathbb{E}[\bar{\boldsymbol{Z}}_i] = \boldsymbol{m}_i,\ \mathrm{Cov}(\bar{\boldsymbol{Z}}_i) = \sigma^2 \boldsymbol{I}_p.$$

- A 'toy' problem, both PSE and NE can be computed in closed form.

Existence and Uniqueness of PSE

Proposition 1

Consider the Multi-PP game with (5), (6). Suppose that $\sum_{j=1}^{n} A_{ij} = 1$. *Then, the RRM finds a unique PSE* **if and only if**

$$\max_{i \in [n]} \left| \lambda_i \left(\frac{\rho}{1+\rho} \boldsymbol{A} + \frac{\varepsilon}{1+\rho} \boldsymbol{P} \right) \right| < 1. \tag{7}$$

Moreover, the PSE admits the closed-form:

$$\boldsymbol{\theta}^{pse} = \left([(1+\rho)\boldsymbol{I}_n - \rho \boldsymbol{A} - \varepsilon \boldsymbol{P}] \otimes \boldsymbol{I}_{\bar{p}} \right)^{-1} \boldsymbol{m}. \tag{8}$$

- **Sufficient and necessary** condition for stability of RRM (extensible to SG-GD) with explicit dependence on the weighted graph:

$$\overline{\boldsymbol{A}}(\varepsilon, \rho) := \tfrac{\varepsilon}{1+\rho} \boldsymbol{P} + \tfrac{\rho}{1+\rho} \boldsymbol{A}$$

  see the next slide.

- Shows the combined effect of $G^{\mathtt{A}}, G^{\mathtt{P}}$.

Structure of the PSE Solution

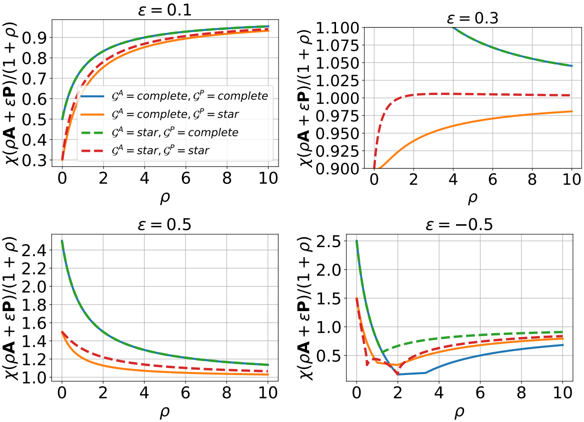
If $p = 1$, then $\boldsymbol{\theta}^{\mathsf{pse}} = ((1+\rho)\boldsymbol{I}_n - \varepsilon\boldsymbol{P} - \rho\boldsymbol{A})^{-1}\boldsymbol{m}$.

- If $\varepsilon > 0$, then $\boldsymbol{\theta}^{\mathsf{pse}}$ is the *weighted Katz-Bonacich centrality* vector [Jackson, 2010] for the weighted adjacency matrix $\overline{\boldsymbol{A}}(\varepsilon, \rho)$.
- Suppose that the j-th mean $\boldsymbol{m}_j$ is perturbed by κ and let $\bar{\boldsymbol{\theta}}^{\mathsf{pse}}(j) \in \mathbb{R}^n$ be the new PSE. Then, the changes in the PSE solution at agent i after perturbing the jth population is

$$\Delta_{ij} \coloneqq \bar{\theta}_i^{\mathsf{pse}}(j) - \theta_i^{\mathsf{pse}} = \tfrac{\kappa}{1+\rho} \textstyle\sum_{k=1}^{\infty} [(\overline{\boldsymbol{A}}(\varepsilon, \rho))^k]_{ij}.$$

  If $\varepsilon > 0$ and $\rho = 0$, then $\overline{\boldsymbol{A}}(\varepsilon, \rho) = \varepsilon\boldsymbol{P}$ and Δ_{ij} is proportional to the total number of walks from i to j in $\mathcal{G}^{\mathsf{P}}$.

Effects of Cooperation on the Stability of PSE



- For small (resp. large) sensitivity, $\varepsilon = 0.1$ (resp. $\varepsilon = 0.5$), (7) is always satisfied (resp. violated) irrespective of the value of ρ.
- For $\varepsilon = 0.3$, increasing ρ lead to violation of (7) for the case when both $\mathcal{G}^{\mathtt{A}}, \mathcal{G}^{\mathtt{P}}$ are star graphs. This coincides with the previous observation that $\rho \gg 1$ can destabilize the PSE when $\boldsymbol{A}$ is asymmetric.
- For $\varepsilon = -0.5$, increasing ρ can stabilize the PSE, i.e., satisfying (7).

Structure of the PSE Solution

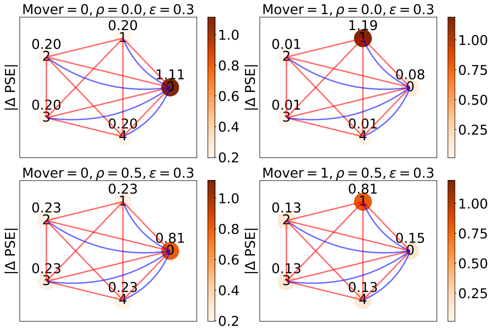


Figure 1: Illustrating $|\Delta_{ij}|$ for the PSE of mean estimation problem when the mean of one of the local populations ('Mover') is perturbed. ($\mathcal{G}^{\mathtt{A}}$: red, $\mathcal{G}^{\mathtt{P}}$: blue.)

- $|\Delta_{ij}|$ increases if agent i is closer to agent j on the combined graph.
- Increasing ρ makes the variations of $|\Delta_{ij}|$ more uniform across the network.

Case Study: Logistic Regression Game

- Each agent trains a personalized logistic regression model with

$$\ell_i(\boldsymbol{\theta}_i; \boldsymbol{Z}_i) = -y_i \boldsymbol{\theta}_i^\top \boldsymbol{x}_i + \log\left(1 + e^{\boldsymbol{\theta}_i^\top \boldsymbol{x}_i}\right), \tag{9}$$

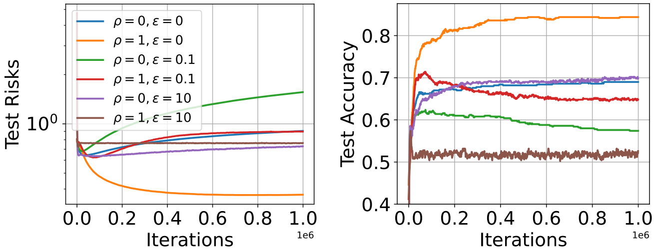
  where $\boldsymbol{Z}_i = (\boldsymbol{x}_i, y_i) \in \mathbb{R}^{p_i} \times \{0, 1\}$ is the feature-label pair.
- Features are generated according to:

$$\boldsymbol{x}_i = \begin{cases} \bar{\boldsymbol{x}}_i^0 + \varepsilon \sum_{j=1}^n P_{ij} \boldsymbol{\theta}_j, & \text{if } y_i = 0, \\ \bar{\boldsymbol{x}}_i^1, & \text{if } y_i = 1, \end{cases} \tag{10}$$

  where $\bar{\boldsymbol{x}}_i^0$ and $\bar{\boldsymbol{x}}_i^0$ follow some base distributions with $\mathbb{E}[\bar{\boldsymbol{x}}_i^0] = \boldsymbol{m}_i^0 \in \mathbb{R}^p$ and $\mathbb{E}[\bar{\boldsymbol{x}}_i^1] = \boldsymbol{m}_i^1 \in \mathbb{R}^p$, and $\varepsilon_i \in \mathbb{R}$.

When $n = 1$, this setting reduces to the *strategic classification* problem that has been studied in the literature [Hardt et al., 2016, Dong et al., 2018, Perdomo et al., 2020, Zrnic et al., 2021].

Case Study: Logistic Regression Game



($\mathcal{G}^{\mathtt{A}}$: complete, $\mathcal{G}^{\mathtt{P}}$: star)

- Enabling graph regularization (with $\rho = 1$) allows the agents to maintain a high accuracy in classification for small distribution shifts $\varepsilon \in \{0, 0.1\}$.
- But setting $\rho = 1$ under large distribution shifts ($\varepsilon = 10$) may lead to degraded performance.

Nash Equilibrium

Definition 4 (Nash Equilibrium, NE)

A vector $\boldsymbol{\theta}^{\rm ne} = [\boldsymbol{\theta}_1^{\rm ne}; \dots; \boldsymbol{\theta}_n^{\rm ne}] \in \mathbb{R}^p$ is called a **Nash equilibrium (NE)** of the game (1) if it holds for all $i \in [n]$ that

$$\boldsymbol{\theta}_i^{\rm ne} \in \mathop{\arg\min}_{\boldsymbol{\theta}_i \in \mathbb{R}^{p_i}} \left\{ \mathbb{E}_{\boldsymbol{Z}_i \sim \mathcal{D}_i(\boldsymbol{\theta}_i, \boldsymbol{\theta}_{\mathcal{N}_i}^{\rm ne})} \left[f_i(\boldsymbol{\theta}_i, \boldsymbol{\theta}_{\mathcal{M}_i}^{\rm ne}; \boldsymbol{Z}_i) \right] \right\}.$$

- Recall that PSE was defined as:

$$\boldsymbol{\theta}_i^{\sf pse} \in \mathop{\arg\min}_{\boldsymbol{\theta}_i \in \mathbb{R}^{p}} \left\{ \mathbb{E}_{\boldsymbol{Z}_i \sim \mathcal{D}_i(\boldsymbol{\theta}_i^{\sf pse}, \boldsymbol{\theta}_{\mathcal{N}_i}^{\sf pse})} \left[f_i(\boldsymbol{\theta}_i, \boldsymbol{\theta}_{\mathcal{M}_i}^{\sf pse}; \boldsymbol{Z}_i) \right] \right\}$$

- The NE can be found with the *best response (BR) dynamics*,

$$\boldsymbol{\theta}_i^{t+1} = \mathcal{B}_i \left(\left[\boldsymbol{\theta}_j^t \right]_{j \in \mathcal{M}_i \cup \mathcal{N}_i} \right) := \mathop{\arg\min}_{\boldsymbol{\theta}_i \in \mathbb{R}^{p_i}} \mathbb{E}_{\boldsymbol{Z}_i \sim \mathcal{D}_i(\boldsymbol{\theta}_i, \boldsymbol{\theta}_{\mathcal{N}_i}^t)} \left[f_i(\boldsymbol{\theta}_i, \boldsymbol{\theta}_{\mathcal{M}_i}^t; \boldsymbol{Z}_i) \right],$$

  for all $i \in [n]$ $\leftarrow$ can be difficult!

Existence and Uniqueness of NE: Assumptions

Assumption 3: For any $i \in [n]$, the map $\mathbb{E}_{\boldsymbol{Z}_i \sim \mathcal{D}_i(\cdot, \boldsymbol{\theta}_{\mathcal{N}_i})} [f_i(\boldsymbol{\theta}_i, \boldsymbol{\theta}_{\mathcal{M}_i}; \boldsymbol{Z}_i)]$ is differentiable at $\boldsymbol{\theta}_i$ and its derivative is continuous in $[\boldsymbol{\theta}_i; \boldsymbol{\theta}_{\mathcal{N}_i}]$.

Assumption 4: For any $i \in [n], \boldsymbol{\delta}, \boldsymbol{\theta}$, the map

$$H^i_{\boldsymbol{\delta}}(\boldsymbol{\theta}) := \frac{\partial}{\partial \boldsymbol{u}_i} \mathbb{E}_{\boldsymbol{Z}_i \sim \mathcal{D}_i(\boldsymbol{u}_i, \boldsymbol{\delta}_{\mathcal{N}_i})} [f_i(\boldsymbol{\theta}_i, \boldsymbol{\theta}_{\mathcal{M}_i}; \boldsymbol{Z}_i)] \Big|_{\boldsymbol{u}_i = \boldsymbol{\delta}_i}$$

is monotone w.r.t. $\boldsymbol{\delta}$.

- Standard for guaranteeing strong monotonicity.
- Our focus is on the network effects on NE.

Existence and Uniqueness of NE

Theorem 5

Suppose that $\sum_{j=1}^{n} A_{ij} = 1$ *for all* $i \in [n]$ *and Assumptions 1-4 hold. Let* $\mu_{\min} := \min_{i\in[n]}\{\mu_i\}$ *and* $\rho_{\min} := \min_{i\in[n]}\{\rho_i\}$*. If it holds that*

$$\sqrt{\max_{j\in[n]}\left\{\sum_{i=1}^{n}\left(\frac{P_{ij}L_i\epsilon_i}{\mu_{\min}+\rho_{\min}}\right)^2\right\}}+\left\|\mathrm{Diag}\left(\frac{\boldsymbol{\rho}}{\mu_{\min}+\rho_{\min}}\right)\boldsymbol{A}\right\|_2 < 1-\frac{\max_{i\in[n]}\{L_i\epsilon_i\}}{\mu_{\min}+\rho_{\min}},$$

then (1) *is strongly monotone, and admits a unique NE (Facchinei and Pang [2003, Theorem 2.3.3(b)]).*

- If $\mu_i = \mu > 0$ for all $i \in [n]$, then the condition in Theorem 5 is equivalent to $\sqrt{\sum_{i=1}^{n} L_i^2\epsilon_i^2} + \max_{i\in[n]}\{L_i\epsilon_i\} \le \mu$.
- Strictly *weaker* than the condition $2\sqrt{\sum_{i=1}^{n} L_i^2\epsilon_i^2} \le \mu$ required by [Narang et al., 2022, Theorem 5].

Conclusions & Perspectives

- Multi-PP game is a new class of game at the intersection of machine learning and game theory.
- We characterize the equilibriums (PSE and NE) of Multi-PP, highlighting on the effects of sensitivity of population, strength of cooperation, graph topology.
- Perturbation analysis (with quadratic loss) reveals how network centrality affects equilibrium.

Open Problems

- Fine-grained analysis on the general case beyond quadratic loss.
- Algorithms for reaching the equilibrium(s) in the general setting (with non-convex loss, imperfect signaling, etc.).
- Inverse problem for learning the graph topologies from PSEs.

Thank you! Pre-print available soon (or email me: `htwai@cuhk.edu.hk`)

References I

Aurélien Bellet, Rachid Guerraoui, Mahsa Taziki, and Marc Tommasi. Personalized and private peer-to-peer machine learning. In *Proceedings of the 21st International Conference on Artificial Intelligence and Statistics*, pages 473–481. PMLR, 2018.

Jinshuo Dong, Aaron Roth, Zachary Schutzman, Bo Waggoner, and Zhiwei Steven Wu. Strategic classification from revealed preferences. In *Proceedings of the 2018 ACM Conference on Economics and Computation*, pages 55–70, 2018.

Francisco Facchinei and Jong-Shi Pang. *Finite-Dimensional Variational Inequalities and Complementarity Problems*. Springer, 2003.

Jesús Gómez-Gardenes, Irene Reinares, Alex Arenas, and Luis Mario Floría. Evolution of cooperation in multiplex networks. *Scientific Reports*, 2(1):1–6, 2012.

Moritz Hardt, Nimrod Megiddo, Christos Papadimitriou, and Mary Wootters. Strategic classification. In *Proceedings of the 2016 ACM conference on Innovations in Theoretical Computer Science*, pages 111–122, 2016.

References II

Matthew O Jackson. *Social and Economic Networks*. Princeton University Press, 2010.

Qiang Li, Chung-Yiu Yau, and Hoi To Wai. Multi-agent performative prediction with greedy deployment and consensus seeking agents. In *Advances in Neural Information Processing Systems 35*, 2022.

Adhyyan Narang, Evan Faulkner, Dmitriy Drusvyatskiy, Maryam Fazel, and Lillian J Ratliff. Multiplayer performative prediction: Learning in decision-dependent games. *arXiv preprint arXiv:2201.03398*, 2022.

Juan Perdomo, Tijana Zrnic, Celestine Mendler-Dünner, and Moritz Hardt. Performative prediction. In *Proceedings of the 37th International Conference on Machine Learning*, pages 7599–7609. PMLR, 2020.

Georgios Piliouras and Fang-Yi Yu. Multi-agent performative prediction: From global stability and optimality to chaos. *arXiv preprint arXiv:2201.10483*, 2022.

Tijana Zrnic, Eric Mazumdar, Shankar Sastry, and Michael Jordan. Who leads and who follows in strategic classification? *Advances in Neural Information Processing Systems 34*, 34:15257–15269, 2021.