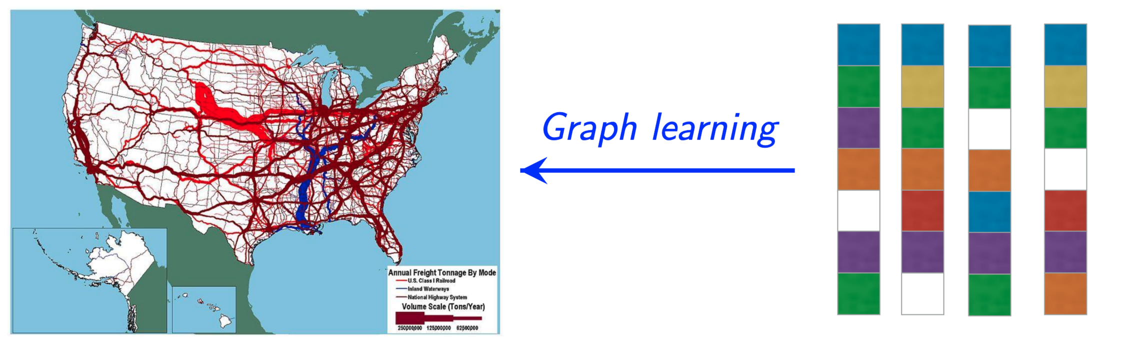
Network Games Induced Prior for Graph Topology Learning

Chenyue ZHANG (CUHK), Shangyuan LIU (CUHK), Hoi-To WAI (CUHK), Anthony Man-Cho SO (CUHK)

Paper ID: 3016

Motivation



- Given **network data**, a standard procedure is:

  data $\longrightarrow$ graph $\longrightarrow$ prediction/analysis/GSP

  Graph Learning make decision

- **Smooth-GL Model**: x_m = **smooth**, S = Adj. [1],

$$J(S;X)=\frac{1}{M}\sum_{i,j}S_{ij}\|x_i^{\rm row}-x_j^{\rm row}\|^2+\beta\|S\|_F^2,$$

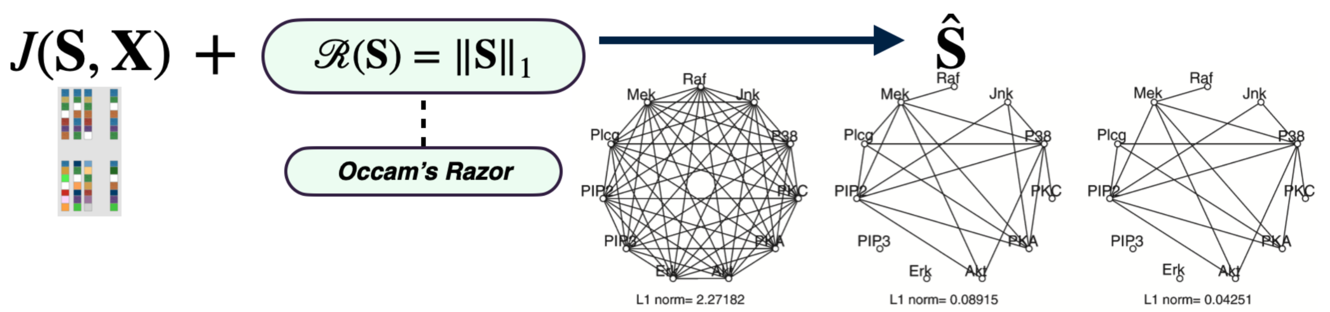
$$\mathcal{S}=\{S:S\mathbf{1}=c\mathbf{1},\ S\geq\mathbf{0},\ {\rm diag}(S)=0\},$$

Q: **Insufficient data** ($M \ll N$) to estimate S.

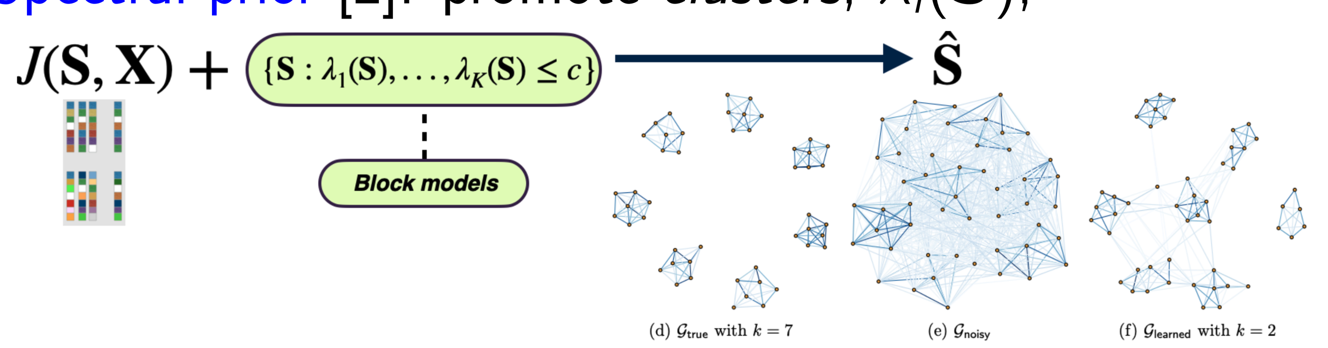
Existing Approach: structural priors on $\hat{S}$,

$$\min_S\ J(S,X)+\lambda R(S)\quad \text{s.t.}\quad S\in\mathcal{S} \qquad (1)$$

- Sparsity prior [1]: promote *sparseness*, $R(S)=\|S\|_1$.



- Spectral prior [2]: promote *clusters*, $\lambda_i(S)$;



Observation: *Structural priors are explicit on S and only capture* **macroscopic** *properties.*

Network Games in Social Network

- Generalized linear quadratic game [3] w/ utility function:

$$U(y_i;y_{-i};S):=-y_i^2/2+y_i\left(\sum_{j=1,j\neq i}^n S_{ij}f(y_j)-b_i\right)$$

  - b_i: marginal benefit, $f(\cdot)$ models interactions.
- Nash equilibrium (NE) = state where no agent has the intention to change,

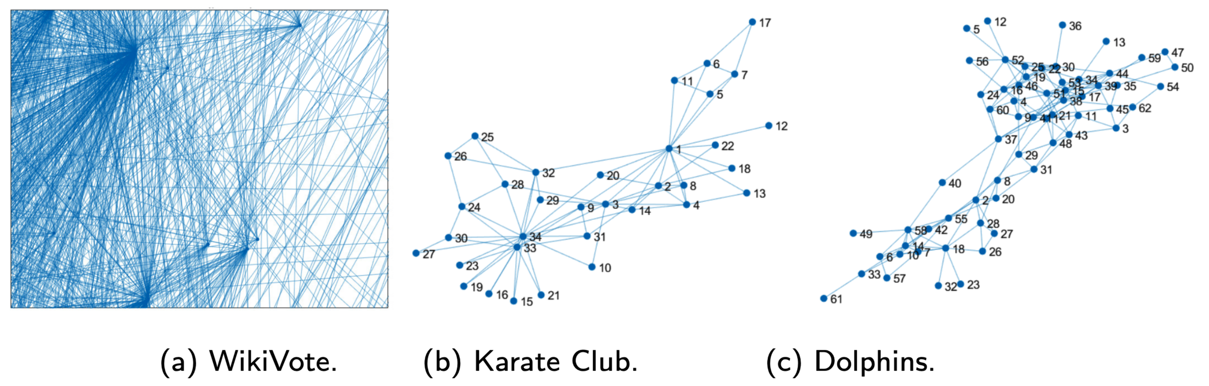
$$NE(S):=\{y\in\mathbb{R}^N: y_i=\arg\max_{y\geq 0}U(y_i;y_{-i};S),\ \forall\ i\}$$

- If $|NE(S)|=1$, $y^{NE}(S)$ denotes the unique NE.

Max-Welfare Induced Graph Learning

Suppose $|NE(S)|=1$, denote ${\rm Wel}(S):=\mathbf{1}^\top y^{\rm NE}(S)$ as the **social welfare.**

Empirical Study: Is intelligent network S locally optimized for max-welfare ?



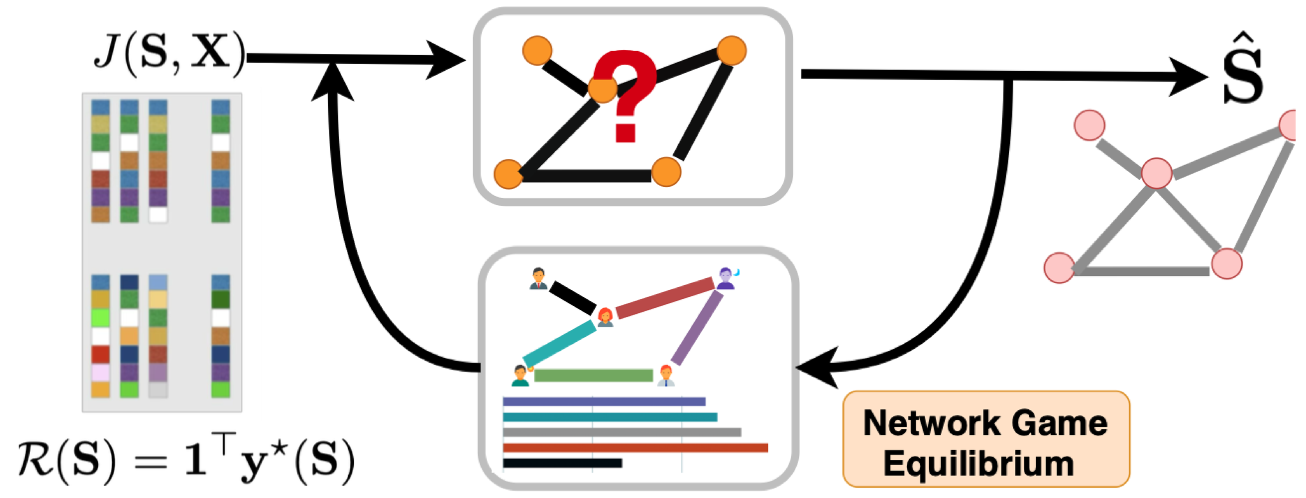
(a) WikiVote. (b) Karate Club. (c) Dolphins.

Rewiring	10%	20%	30%	40%	50%
WikiVote	96.29%	93.07%	90.33%	88.01%	86.11%
Karate	94.06%	88.86%	84.72%	81.17%	78.30%
Dolphins	98.15%	96.48%	95.13%	93.97%	93.08%

- *Random rewiring* and evaluate the relative % of welfare loss due to rewiring.
- Human-made networks (Karate & WikiVote) are sensitive to rewiring;
- non-human-made networks are not (Dolphins).

Conjecture: Human-made networks are **self-optimized** w.r.t. ${\rm Wel}(S)$.

$\Longrightarrow$ *Utilize* ${\rm Wel}(S)$ *to inform graph learning.*



Graph Learning with Functional Prior (GLFP):

$$\min_{S,y\in\mathbb{R}^N} J(S,X)-\lambda\mathbf{1}^\top y\ \text{ s.t. }\ y\in{\rm EQ}(S),\ \ S\in\mathcal{S}.$$

Q: What structural insights can be derived from (GLFP)?

Proposition: An *approximation* of (GLFP) admits the optimal solution:

$$S_{ij}^\star=\frac{1}{2\beta}\max\{0,\eta_i+\lambda b_j-D_{ij}\},\quad \text{for some } \eta_i\in\mathbb{R}.$$

- Implication: let $\lambda\gg 1$, optimal $S^\star$ admits a **'star' structure**.
- Empirical evidence: human-made network contains **few hub nodes**.

Two-timescales Gradient Algorithm for (GLFP)

Challenge: took network dynamics as constraint $\Rightarrow$ **Bilevel Optimization**:

- Upper level: regularized graph learning. Lower level: NE seeking.

  $\rightarrow$ **Two-timescales algo.:** η **(Lower-level)** $>$ γ **(Upper-Level)**

$$y_i^{k+1}=y_i^k+\eta\nabla_{y_i}U(y_i^k,y^k;S^k),\ \forall i\in[N]$$
$$S^{k+1}={\rm Proj}_{\mathcal{S}}(S^k-\gamma\widehat{\nabla}_S\Phi(S^k,y^k)), \qquad (2)$$

$\widehat{\nabla}_S\Phi(S^k,y^k)$ estimates *hypergradient* — from derivatives of $U(\cdot)$

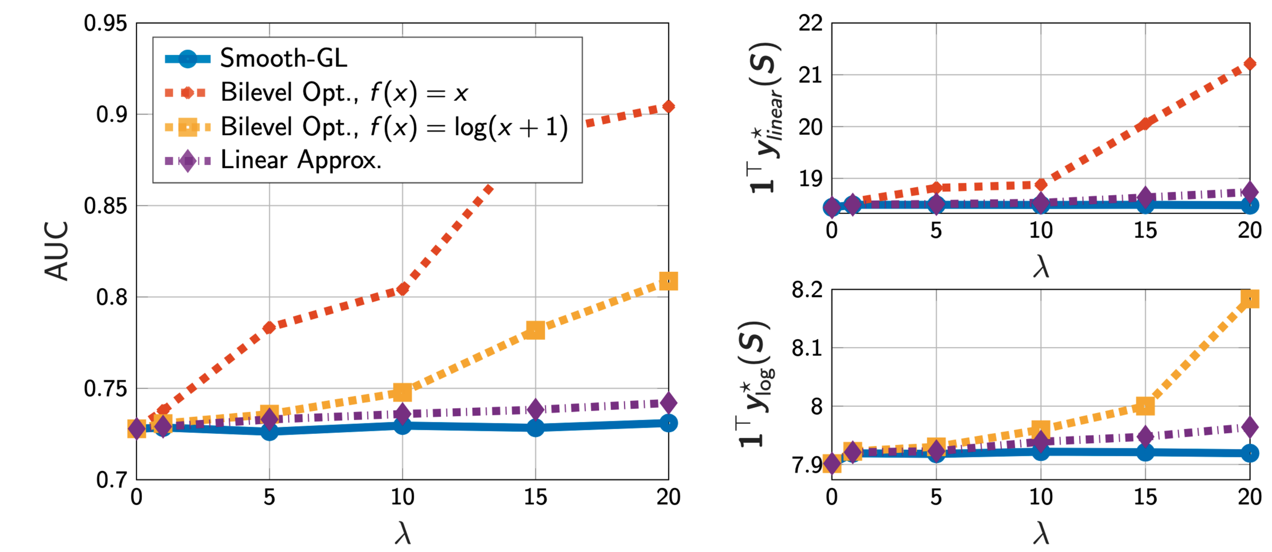
- **Theorem**: With $\eta=\frac{1-c}{(1+c)^2}$, $\gamma\ll\eta$ (+ additional assumptions),

$$\min_{k=1,\ldots,K}\|\gamma^{-1}(S^k-{\rm Proj}_{\mathcal{S}}(S^k-\gamma\nabla\Phi(S^k,y^\star(S^k))))\|^2=\mathcal{O}(K^{-1}).$$

$\Longrightarrow$ finds **a stationary point** of (GLFP).

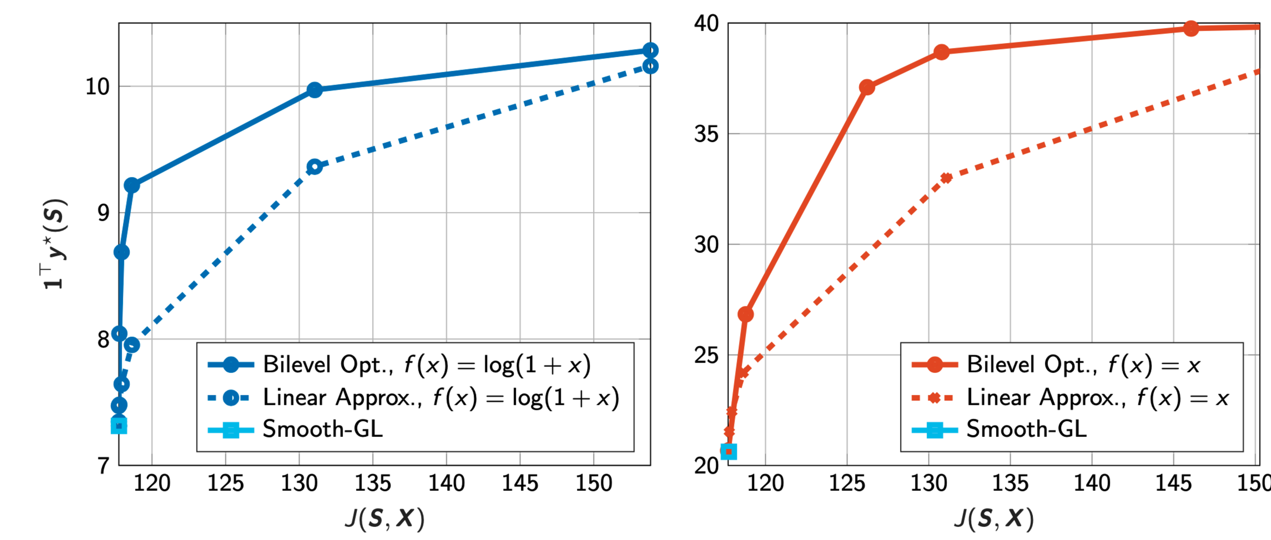
Numerical Results

(1) Synthetic Data:



- **Setting**: $S\sim$ PA graph with $N=50$ nodes & $M=10$ stationary + smooth graph signals $\Rightarrow$ **limited amount of data ($M\ll N$)**.
- **Benchmark**: (i) linear approximation of (GLFP) with $\mathbf{1}^\top y^\star(S)\approx\mathbf{1}^\top Sb\Rightarrow$ single-level problem & (ii) Smooth-GL [4].
- For the network dynamics & (GLFP), we take $b={\rm TopEV}(D)$.
- Proposed methods has better AUROC than vanilla Smooth-GL.

(2) Karate Club:



- $N=34$ from $M=50$ smooth graph signals.
- *Pareto Front* by adjusting λ to get tradeoffs for $h(y)$ vs $J(S,X)$.
- Bilevel OPT by TT algorithm achieves better *Pareto optimality*.

(3) Real Data: Indian village [5]

AUROC	Maximum	Average	Minimum
Bilevel $-\ f(y)=y$	0.6345	0.5788	0.5271
Bilevel $-\ f(y)=\log(1+y)$	**0.6570**	**0.5937**	**0.5490**
Smooth-GL	0.5075	0.4888	0.4777

${\rm Welfare}(\hat{S};f)-{\rm Welfare}(S^{true};f)$	Maximum	Average	Minimum
Bilevel $-\ f(y)=y$	**4.2754**	**3.2077**	**2.1693**
Smooth-GL $-\ f(y)=y$	-2.3288	-5.8141	-8.6115
Bilevel $-\ f(y)=\log(1+y)$	**1.4464**	**1.1892**	**0.8461**
Smooth-GL $-\ f(y)=\log(1+y)$	-0.5440	-2.8292	-5.0419

- **Data**: network sizes $N=77$ to 330, each w/ $M=16$ samples $\Rightarrow$ **limited amount of data ($M\ll N$)**.
- Graph learning with *functional prior* maximizing ${\rm Welfare}(S)$ improves AUROC relative to ground truths .

[1] Friedman J, Hastie T, Tibshirani R. Sparse inverse covariance estimation with the graphical lasso. Biostatistics. 2008 Jul;9(3):432-41.
[2] Kumar, Sandeep, et al. "A unified framework for structured graph learning via spectral constraints." JMLR 21.22 (2020): 1-60.
[3] Cai J, Zhang C, Wai HT. Optimal pricing for linear-quadratic games with nonlinear interaction between agents. IEEE L-CSS. 2024 Jun 6.
[4] Candogan O, Bimpikis K, Ozdaglar A. Optimal pricing in networks with externalities. Operations Research. 2012 Aug;60(4):883-905.
[5] Kalofolias V. How to learn a graph from smooth signals. In AISTATS 2016 (pp. 920-929). PMLR.
[6] Banerjee A, Chandrasekhar AG, Duflo E, Jackson MO. The diffusion of microfinance. Science. 2013 Jul 26;341(6144):1236498.