A Game Theory Model of Scapegoating in Social Networks

Opinion Dynamics and Graph Centrality

Qinyang Yu (MSEC' 26)

Department of Economics & Computer Science
Duke University

December 11, 2024

Motivation

- New interdiscipline: Graph Analysis + Game Theory $\Longrightarrow$ Network Economics
- Scapegoating: A community transfers the blames to an innocent person
- Significance:
  - Topic-wise: Social justice and humanitarianism
  - Method-wise: Network framework

Figure: Scapegoating

Figure: Witch Hunts

Literature Review and Research Gap

Why using networks?

- Political Science: 1. Strategies for political and social control
- Social Psychology:
  - 2. Groupthink, peer influence, and imitation behavior
  - 3. Targeting marginalized individuals or minority groups
- Economics:
  - Models with 1 and 3
  - No models with 2 $\Longrightarrow$ Opinion Dynamics and Social Learning!

Opinion Dynamics

- Firstly introduced by DeGroot (1974)
- Naive updating, not Bayesian updating

$$E.g., T = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \lim_{t\to\infty} T^t p(0) = \left(\lim_{t\to\infty} T^t\right) p(0) = \begin{pmatrix} \frac{2}{5} & \frac{2}{5} & \frac{1}{5} \\ \frac{2}{5} & \frac{2}{5} & \frac{1}{5} \\ \frac{2}{5} & \frac{2}{5} & \frac{1}{5} \end{pmatrix} p(0)$$

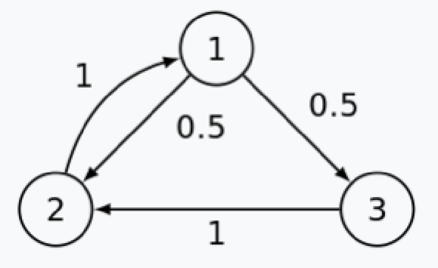


Figure: Degroot Learning

Opinion Dynamics

A game-theoretic version of Opinion Dynamics:

- Proposed by Bindel et al. (2015) and Ghaderi and Srikant (2014)
- Given internal belief b_i, agent i chooses a public opinion $x_i \in [0, 1]$

$$u_i(x_i, x_{-i}) = \underbrace{-(x_i - b_i)^2}_{\text{Penalty 1: Disagreement with true self}} \underbrace{- \sum_{j=1}^{n} A_{ij}(x_i - x_j)^2}_{\text{Penalty 2: Disagreement with others' opinions}}$$

The Networked Game Model

- A community: one leader; n agents forming a social network $G(V, E)$
- A problem or a crisis hit!
- Two possible states: 0-innocent, 1-guilty
- Agents are innocent in nature
- Agent i holds a prior belief that agent k is guilty with probability $\pi_i^{(k)}$
- Agents believe the leader may also prosecute an innocent agent with probability $p \in [0, 1] \implies$ social trust in leader

The Networked Game Model

Strategies:

- Phase 1: Scapegoat Selection
  - The leader will either self-blame (game ends)
  - or choose $k \in V$ to scapegoat (game proceeds)
- Phase 2: Opinion Dynamics
  - Agent i updates belief about agent k's guilt using Bayes' rule to form internal belief $b_i^{(k)} = \frac{\pi_i^{(k)}}{\pi_i^{(k)} + p(1-\pi_i^{(k)})}$
  - Agent i forms a public opinion $x_i^{(k)} \in [0, 1]$ about agent k's guilt through the network

Utilities:

- For agents: $u_i^{(k)}(x_i^{(k)}, x_{-i}^{(k)}) = -(x_i^{(k)} - b_i^{(k)})^2 - \sum_{j=1}^n A_{ij}(x_i^{(k)} - x_j^{(k)})^2$
- For the leader
  - Self-blame: fixed cost C
  - Scapegoat k: reputational cost $R^{(k)} = \sum_{i=1}^n (1 - x_i^{*(k)})$

Nash Equilibrium

Phase 2:

$$FOC : \frac{\partial u_i^{(k)}}{\partial x_i^{(k)}} = 0, \forall i \in V$$

$$(x_i^{(k)} - b_i^{(k)}) + \sum_{j=1}^{n} A_{ij}(x_i^{(k)} - x_j^{(k)}) = 0, \forall i \in V$$

$$(1 + \sum_{j=1}^{n} A_{ij})x_i^{(k)} - \sum_{j=1}^{n} A_{ij}x_j^{(k)} = b_i^{(k)}, \forall i \in V$$

$$(I + D)x^{(k)} - Ax^{(k)} = b^{(k)}$$

$$(I + L)x^{(k)} = b^{(k)}$$

$$x^{*(k)} = (I + L)^{-1} b^{(k)}$$

Nash Equilibrium

Lemma 1 (Conversation Principle)

Using $x^{*(k)} = (I + L)^{-1} b^{(k)}$, it follows that $\sum_{i=1}^{n} x_i^{*(k)} = \sum_{i=1}^{n} b_i^{(k)}$ for all $k \in V$.

- L: row stochastic
- Opinion Dynamics redistributes and adjusts individual opinions
- But the overall "mass" of opinions remains unchanged!
- Intuitions?

Phase 1:

- Scapegoat $\iff$ reputational cost $\leq$ fixed cost $\iff R^{(k)} \leq C \iff$ $\sum_{i=1}^{n}(1 - x_i^{*(k)}) \leq C \iff \sum_{i=1}^{n}(1 - b_i^{(k)}) \leq C \iff \sum_{i=1}^{n}(1 - \frac{\pi_i^{(k)}}{\pi_i^{(k)} + p(1-\pi_i^{(k)})}) \leq C$
- Choose to scapegoat k with smallest $R^{(k)}$

Discrete Network Effects

Assume

$$\pi_i^{(k)} = \begin{cases} 0 & \text{if } i \in N(k), \\ \frac{1}{2} & \text{if } i \notin N(k). \end{cases}$$

Then,

$$b_i^{(k)} = \begin{cases} 0 & \text{if } i \in N(k), \\ \frac{1}{1+p} & \text{if } i \notin N(k). \end{cases}$$

$$\begin{aligned} R^{(k)} &= \sum_{i \notin N(k)} \left(1 - \frac{1}{1+p}\right) + \sum_{i \in N(k)} (1 - 0) \\ &= \left(1 - \frac{1}{1+p}\right) + \frac{\deg(k)}{n(1+p)} \end{aligned}$$

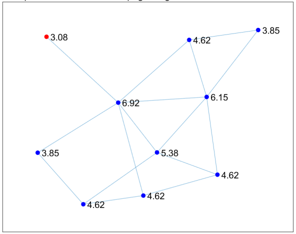
$\Longrightarrow$ The leader scapegoats the agent with the lowest degree centrality!

Discrete Network Effects

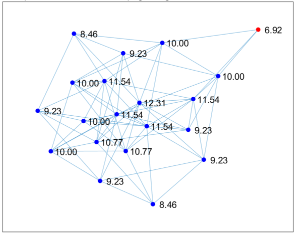
Numerical Simulations:

- $n = 10, 20, 30, 50$
- $p = 0.3$
- Red nodes: selected scapegoats

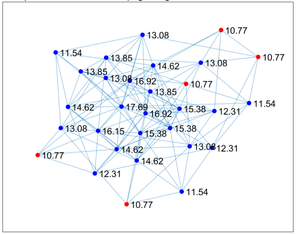
Reputational Cost For Scapegoating - Discrete Distance Effect



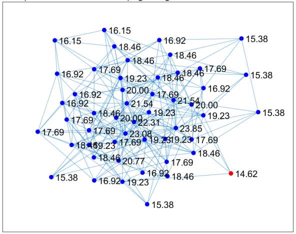
Reputational Cost For Scapegoating - Discrete Distance Effect



Reputational Cost For Scapegoating - Discrete Distance Effect



Reputational Cost For Scapegoating - Discrete Distance Effect



Discrete Network Effects

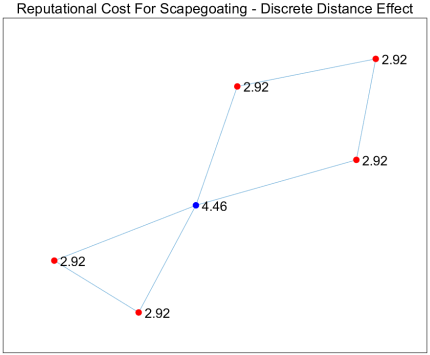


Figure: A poor scapegoat prediction with $n = 6$, $p = 0.3$, and discrete priors (red nodes as selected scapegoats)

Decay Network Effects

Let l_{ik} denote the geodesic/shortest path length between i and k. Assume

$$\pi_i^{(k)} = \frac{1}{2} - \left(\frac{1}{2}\right)^{l_{ik}} = \begin{cases} 0 & \text{if } l_{ik} = 1, \\ \frac{1}{2} & \text{if } l_{ik} \to \infty. \end{cases}$$

Algorithm 1 Scapegoat Selection Algorithm with Decay Distance Effect

Require: Graph $G(V, E)$, parameter p

Ensure: Selected nodes with the lowest D_k

```
 1: for each node k ∈ V do
 2:     Initialize vector π_k ∈ ℝ^|V|
 3:     for each node i ∈ V, i ≠ k do
 4:         Calculate the shortest path length l_ki between k and i using Breadth-First
            Search (BFS) for unweighted G and Dijkstra's or Bellman-Ford Algorithm for
            weighted G
 5:         π_i^(k) ← 0.5 − 0.5^(l_ki)
 6:     end for
 7:     π_k^(k) ← 0
 8:     Generate vector b_k by b_i^(k) = π_i^(k) / (π_i^(k) + p·(1 − π_i^(k))) where ∀i ∈ V.
 9:     Calculate D_k = Σ_{i∈V} (1 − b_i^(k))
10: end for
11: Select the nodes with the lowest D_k values
12: Output the selected nodes
```

$\Longrightarrow$ The leader scapegoats the agent with the lowest D_k!

Decay Network Effects

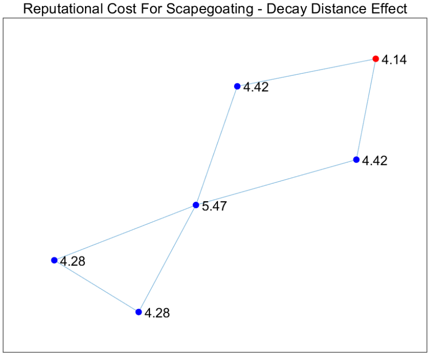
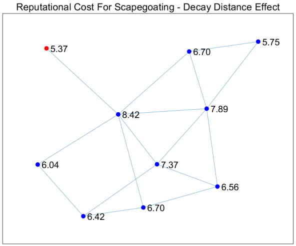


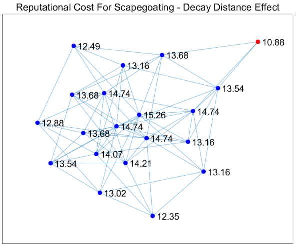
Figure: An improved scapegoat selection with $n = 6$, $p = 0.3$, and decay priors (red nodes as selected scapegoats)

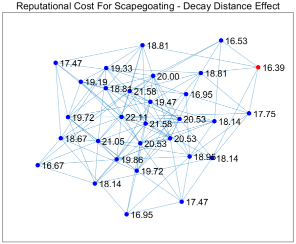
Decay Network Effects

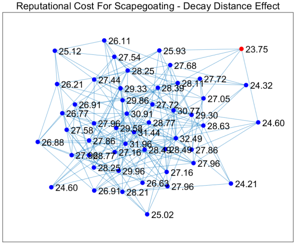
Numerical Simulations:

- $n = 10, 20, 30, 50$
- $p = 0.3$
- Red nodes: selected scapegoats









Decay Network Effects

About the metric D_k:

- A component-wise convex transformation of decay centrality $C_k = \sum_{i=1}^{n} \alpha^{l_{ik}}$ with $\alpha = \frac{1}{2} \in (0, 1)$
- Specifically, when $p = 1$, $D_k \sim C_k$.
- Advantage: global structure instead of local connectivity
- Disadvantage:
  - Harder to interpret
  - Computationally inefficient
  - Unweighted graph: $O(nm + n^2)$; Weighted graph: $O(nm + n^2 \log(n))$

Extensions

Main Conclusion:
Marginalized individuals $\Longrightarrow$ lower centrality $\Longrightarrow$ lower information exposure in networks $\Longrightarrow$ more likely to become scapegoats

Centrality Measures:

- Degree: local connectivity
- Decay, Closeness: global connectivity, ease of reaching other nodes
- Betweenness: importance as an intermediary connector
- Eigenvector, Bonacich: influence, prestige, power

The End

"The best way to bring folks together is to give them a real good enemy."

— The Wonderful Wizard of Oz, in *Wicked*

Reference

- Bindel, D., Kleinberg, J., and Oren, S. (2015). How bad is forming your own opinion? *Games and Economic Behavior*, 92:248–265.
- DeGroot, M. H. (1974). Reaching a consensus. *Journal of the American Statistical association*, 69(345):118–121.
- Ghaderi, J. and Srikant, R. (2014). Opinion dynamics in social networks with stubborn agents: Equilibrium and convergence rate. *Automatica*, 50(12):3209–3215.