



An Experimental Framework for Characterizing Adversarial Behavior in Social Networks

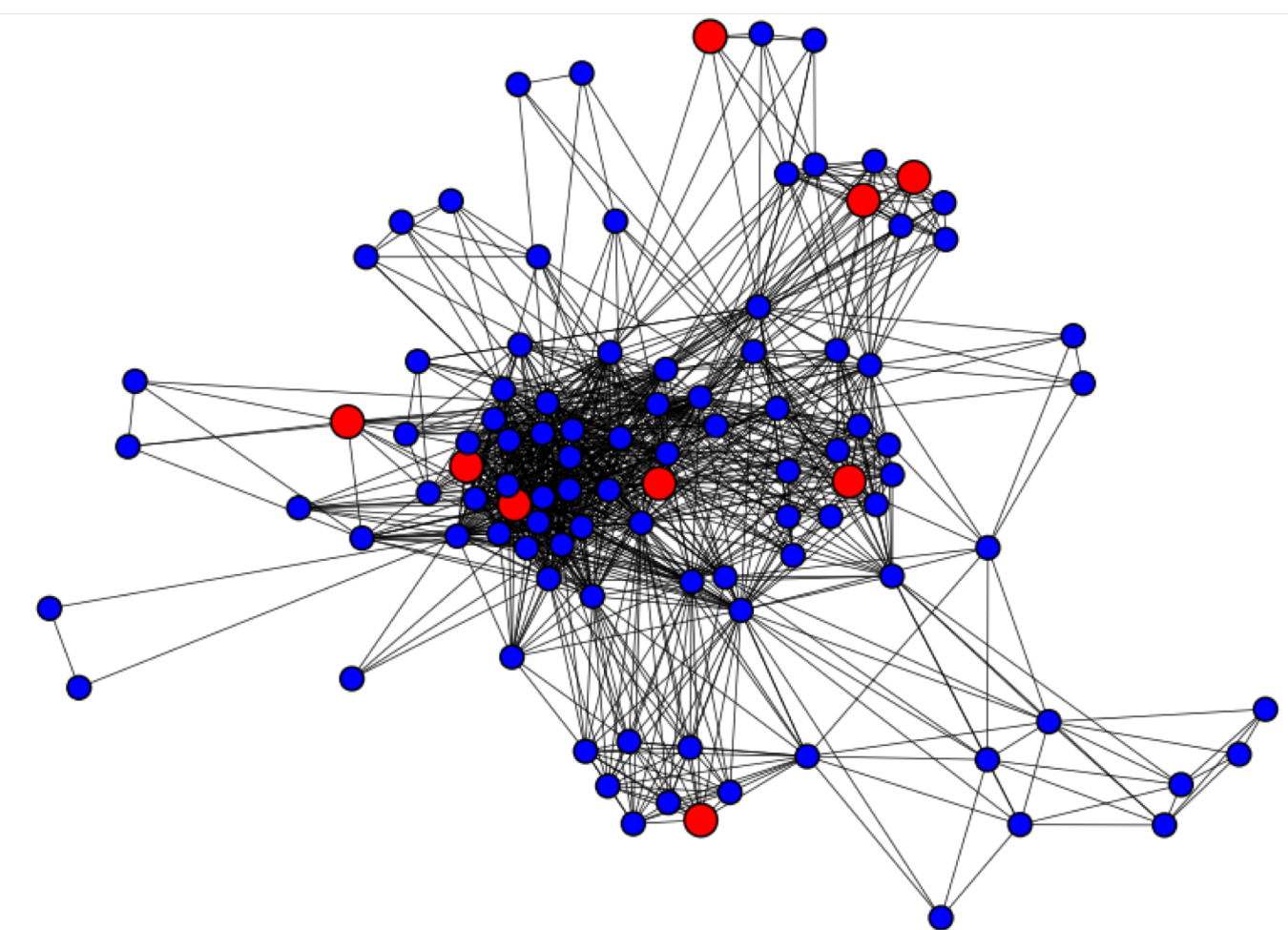
Pivithuru Wijegunawardana¹, Mario Taglic¹, Raluca Gera², Sucheta Soundarajan¹

¹Department of EECS, Syracuse University, Syracuse, NY

²Department of Applied Mathematics, Naval Postgraduate School

Challenge: Collecting Adversarial Network Data

- Covert groups hide themselves in civilian social networks
 - Criminal gangs, Terrorist cells
- Collecting data about such covert groups is challenging
 - They are adversarial against those who seek to collect or analyze their data.
- How can a data collector/investigator identify a so-called “dark” network?



Which factors motivate adversarial nodes to misreport their information to a data collector?

Our Approach

- **Basic Idea:** Use the Amazon Mechanical Turk crowdsourcing platform to hire “Turkers” to play the “Adversarial Social Network Analysis Game”
- **Big Questions:**
 - Are individuals more/less likely to report false data depending on the order in which the data collector investigates people?
 - What level of incentive should the data collector provide to obtain more useful, accurate data?
 - Does loyalty of individuals to the covert group play a role in deceptive behavior?

Network Data Collection Game

• “Werewolves and Villagers”

- A pack of **werewolves (Red nodes)**, representing members of an adversarial group, is hiding among innocent **civilian villagers (Blue nodes)** in a village social network
- Turkers play as werewolves in the game
- A sheriff has come to investigate the werewolves, and will pay for honest answers
- But if a werewolf betrays her pack, they will punish her!

Potions Left: 4

View an animation of the data collector's order of past choices

Budget the data collector is left with

Node 8

If you say --	And if they are	Total Reward / Penalty
Werewolf	Investigated	1
Werewolf	Not Investigated	-2
Villager	Investigated	-1
Villager	Not Investigated	2

Your answer is

1/3 COMPLETED

EXIT

Cumulative Payment: \$0

FILTERING GAME: 1 / 3

ANIMATION: ON

RESET ANSWERS

OPEN TUTORIAL

Village Social Network

• Data Collector (Sheriff) : Identify as many werewolves as possible

- Can inquire nodes whether their neighbors in the network are werewolves or not
- Gives a **reward** for providing **true information** and **penalty** for **false information**
- Has a budget b on how many nodes she can investigate

• Adversarial Group (Werewolf pack) : Protect identities of group members

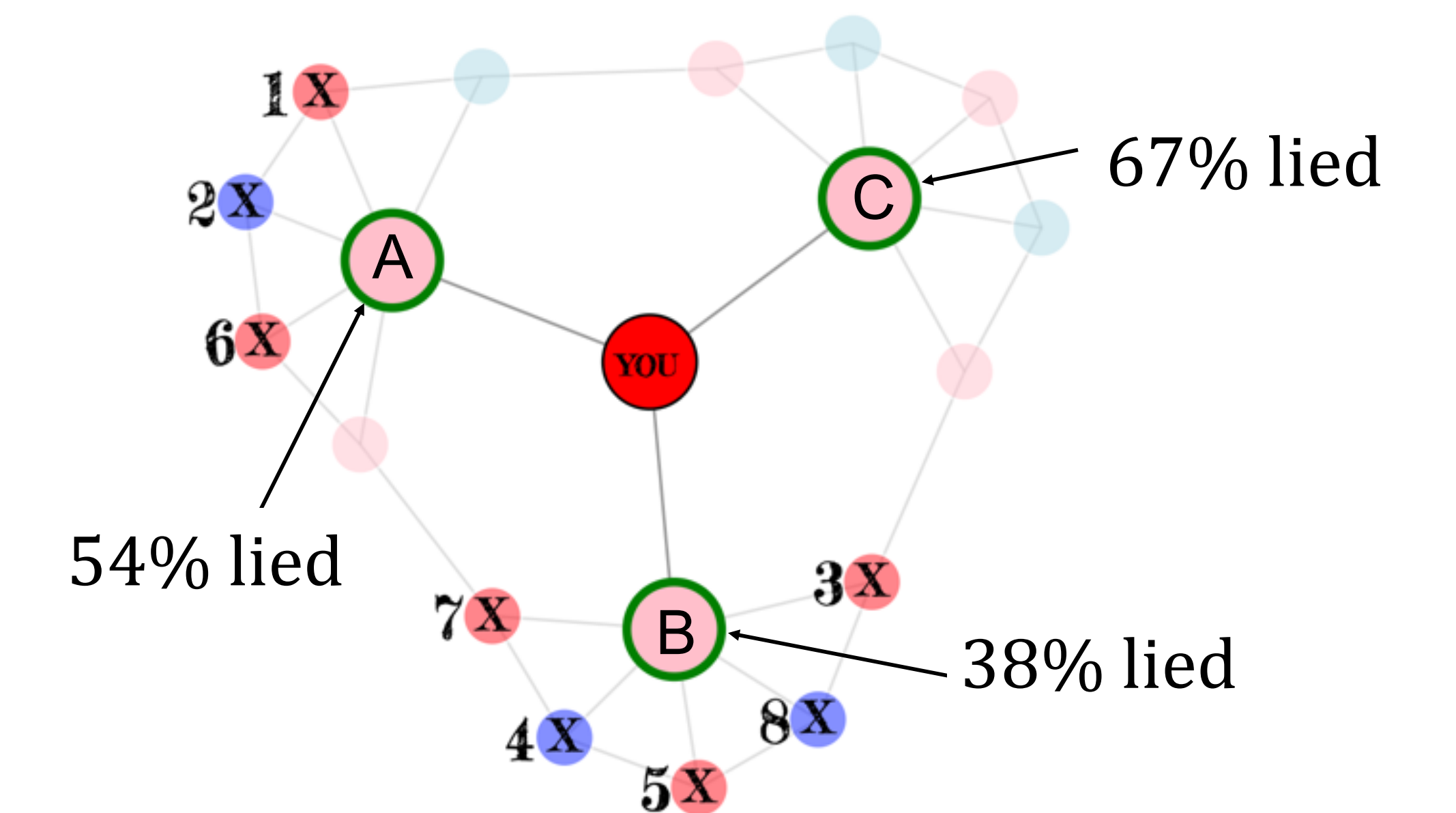
- Gives a **reward** to werewolves who provides **false information** and **penalty** for **true information**

• Adversarial Node (A werewolf) : Maximize the reward they can earn

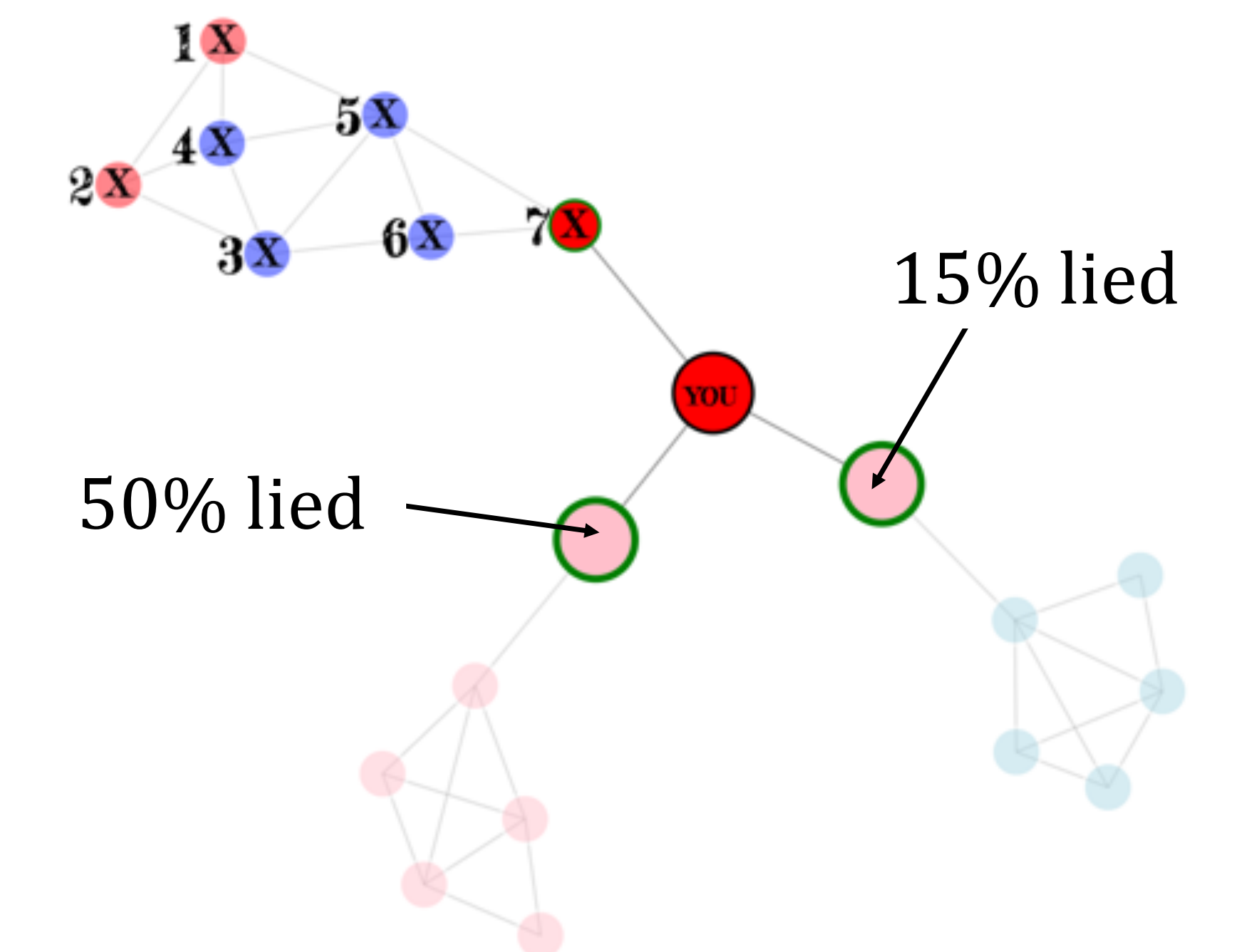
- mTurkers will play as members of the adversarial group
- **Has full information about** : network structure, who else have been investigated, how much budget the investigator is left with
- Presented with possible rewards and penalties from the data collector and adversarial group
 - **Decide whether to lie to the data collector or tell the truth about a neighbor**

Experiments

Participants are more likely to report false information if they think that the data collector will not discover the truth.



Participants are loyal to the Red group



Future Work

- How do adversarial nodes work as a group to protect as many members as possible?
- Are there network structures that would motivate deceptive behavior more often than others?