Network Security and Contagion

Daron Acemoglu, Azarakhsh Malekian, Asu Ozdaglar

Department of Economics
Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology

Banff Workshop on Asymptotics of Large-Scale Interacting Networks
February, 2013

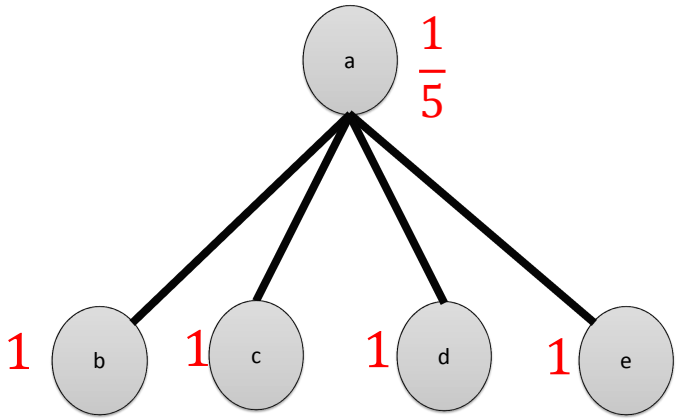
Motivation

- Computer, communication, transport and economic networks all depend on some degree of security for their operation.
- Almost all networks are protected with security investments.
  "Security failure is caused at least as often by bad incentives as by bad design" Anderson and Moore (2006, p. 610).
- An emerging literature at the boundary of economics and computer science → positive externality in security investments.
  - A domain that fails to protect itself adequately not only increases the probability of some type of disruption to its own operation, but also increases the likelihood that infection will spread to other domains.
- Based on this intuition, the literature has so far presumed that there will be underinvestment in security, at least in the case of random attacks [Anderson and Moore, 2006], [Goyal and Vigier, 2011], [Larson , 2011], [Bachrach, Draief and Goyal, 2012].
- But these are based on analysis of "symmetric networks"
  - Unrealistic and restrictive: no true network effects nor analysis of topology
    "Network topology can strongly influence conflict dynamics... Different topologies have different robustness properties with respect to various attacks" Anderson and Moore (2006, p. 613).

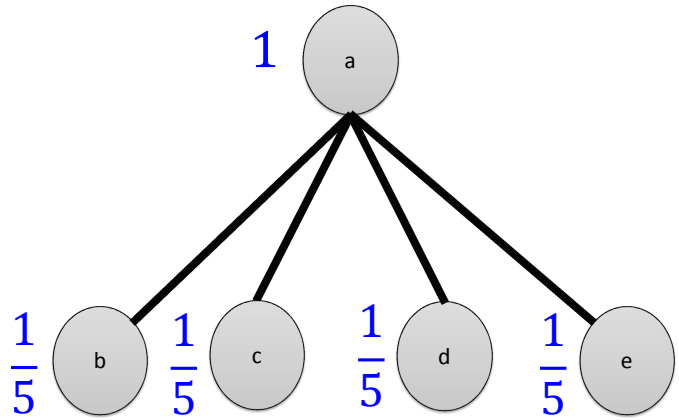
Overinvestment in Security

- For asymmetric networks, we may have overinvestment in security.

Equilibrium



Social Optimum



Security investment cost is $c(q) = \frac{q^2}{5}(2.9 - 1.33q)$.

- In fact, in this example, expected number of infections is greater in the social optimum than in equilibrium.

Why Overinvestment?

- Security decisions of different nodes not only create positive externalities but are typically also strategic substitutes, meaning that lower investment by a node increases the desired investment of others.
    - Positive externality $\rightarrow$ Node 1 underinvests $\rightarrow$ Through strategic substitutes effects, other nodes increase their investments, potential for overinvestment.
- This strategic substitutes property makes the analysis of asymmetric networks particularly important.

Model Overview

- Each node i is connected to a subset of other nodes and chooses a security investment q_i.
- A virus is probabilistically transmitted across connected nodes.
- The probability of successful infection of node i is $1 - q_i$, and the virus can only spread from node i to its neighbors if it successfully infects it.
- Tractable formulation, making positive externality from network investments particularly clear.
- We distinguish two types of attacks:
  1. Random attacks, which are likely to hit each node with equal probability (and in particular independent of their security investments);
  2. Strategic attacks, where the location of the attack is determined by an adversary wishing to maximize expected infection (e.g., [Bachrach, Draief and Goyal, 2012], [Goyal and Vigier, 2011].

This Paper - I

- We generalize network security models of both random and strategic attacks to general (random) networks.
- We show that the oft-presumed underinvestment in security investments is not generally true and overinvestment arises in a range of settings.
- We delineate conditions on the network structure and the attack model under which underinvestment or overinvestment incentives will dominate.

Our Results:

- We first provide a decomposition of individual payoffs into an own effect and an externality, a tractable decomposition that underpins the rest of our analysis and appears not to have been noticed so far in the literature.
- We show that symmetric equilibria of symmetric networks always involve underinvestment as presumed by the existing literature.
    - But not generally true in asymmetric networks, and also not true in asymmetric equilibria of symmetric networks.

This Paper - II

- We show that when the network structure is represented by a tree and cost of investments are sufficiently convex, we always have underinvestment in security.
- We generalize the result to random networks with locally tree structures and symmetric random networks (such as Erdos-Renyi graphs).
- For symmetric random networks, we show that the expected number of infected people is higher in denser graphs and more "clustered" trees.
- For strategic attacks, we show that there is an additional reason for overinvestment, echoing an intuition going back to [de Meza and Gould, 1992]: preventive activities can create negative instead of positive externalities when they shift attacks to other nodes.
    - For a tree network with sufficiently convex cost functions, there can be overinvestment and limited spread of infection.
    - For symmetric random graphs and some additional conditions on cost functions, we show that the equilibrium always involves overinvestment.

Related Literature

- Models of infection on random graphs:
  [Molloy and Reed, 2000], [Newman, Strogatz, and Watts, 2001], [Chung and Lu, 2002]
- Models of infection with endogenous network formation:
  [Goyal, Vigier, 2010],][Larson, 2011], [Blume *et al.*, 2011]
- Models of strategic attacks:
  [Goyal, Vigier, 2010], [de Meza and Gould, 1992], [Bachrach, Dreif, and Goyal, 2012]

Model

- We consider a set $V = \{1, \ldots, n\}$ of agents interacting over random network A.
- We assume that A is drawn from a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where Ω is the set of (undirected) graphs with node set V.
- An attacker targets one of the agents and exposes him to an infection, which then spreads dynamically to the agents in the network.
  - Infection is transmitted on the edges of the realized graph.
  - Simple examples: Erdos-Renyi graphs (with parameter p) leading to transmission probability p of infection to each neighbor.
- Before A and the location of the attack is realized, each agent i invests in security level $q_i \in [0, 1]$.
- We use $\mathbf{q} = [q_j]_{j \in V}$ and $\mathbf{q}_{-i} = [q_j]_{j \in V, j \neq i}$ to denote the security profile of all agents and all agents other than i, respectively.

Model

- Upon being exposed to the virus for the first time, agent $i \in V$ with security level q_i gets infected with probability $1 - q_i$.
  - We adopt the natural assumption that the agent will not get infected in subsequent expositions to the virus.
    - This means that $1 - q_i$ is the probability of agent i being "susceptible".
- Given a network A and a security profile $\mathbf{q}$, we denote the probability of node i getting infected by $P_i(A, \mathbf{q})$.
- The utility function of agent i, $u_i : [0,1]^n \to \mathbb{R}$, is given by

$$u_i(A, \mathbf{q}) = (1 - P_i(A, \mathbf{q})) - c_i(q_i).$$

- $c_i(q_i)$ is the cost agent i incurs for investing in security level q_i.

Assumption 1 (Investment Cost)

For each i, the function $c_i : [0,1] \to \mathbb{R}$ is continuously differentiable, non-decreasing, and convex, and satisfies $c_i(0) = 0$.

Attack Model

- We study two types of attack:
  - Random Attack: The attack is likely to hit each agent with equal probability (and in particular independent of their security investments).
  - Strategic Attack: The location of the attack is determined by an adversary wishing to maximize the expected total number of infected people.

Random Attack Model - Infection Probability

- We first present a characterization that shows how the infection probability of an agent depends on his security investment.
- This enables us to provide a tractable decomposition of individual utility functions into an own effect and network effects of other individuals.

Proposition (Network Effect)

The infection probability of agent i is given by $P_i(A, \mathbf{q}) = (1 - q_i)\tilde{P}_i(A, \mathbf{q}_{-i})$, where $\tilde{P}_i(A, \mathbf{q}_{-i})$ is the probability of the infection reaching agent i.

- *Idea:* Agent i is susceptible with probability $1 - q_i$. A susceptible agent i gets infected only the first time he is exposed to the virus and the probability of the virus reaching i for the first time is independent of q_i.
- We refer to $\tilde{P}_i(A, \mathbf{q}_{-i})$ as the network effect of A on i.

Network Effect

- The network effect on an agent admits a simple recursive structure and can be computed by considering the network with one agent removed at a time.

Proposition (Decomposition)

The probability of the infection reaching agent j satisfies the following: for any $i \in V$,

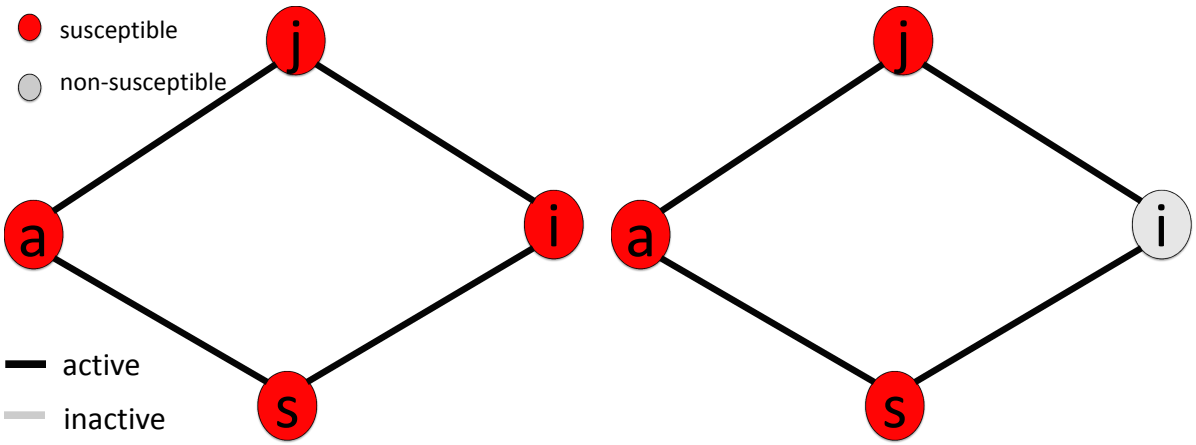
$$\tilde{P}_j(A, \mathbf{q}_{-j}) = \tilde{P}_j(A_{-i}, \mathbf{q}_{-\{j,i\}}) + (1 - q_i) Q_{ij}(A, \mathbf{q}_{-\{i,j\}}),$$

where $Q_{ij}(A, \mathbf{q}_{-\{i,j\}})$ is the probability that the infection reaches agent j only through a path that contains agent i (conditional on i being susceptible).

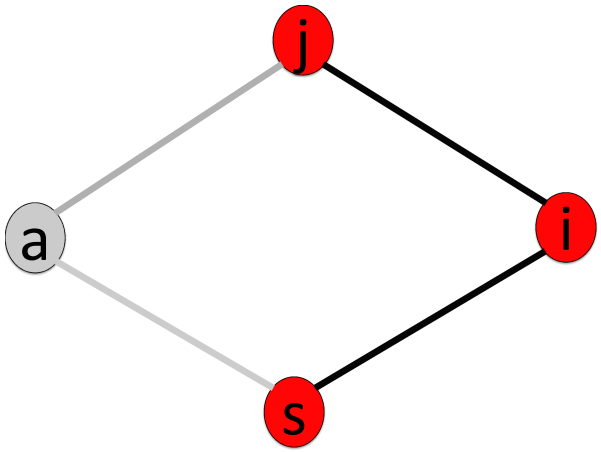
- We will see that $Q_{ij}\left(A, \mathbf{q}_{-\{i,j\}}\right)$ is the externality created by agent i on agent j.

Proof Idea

- Consider three possible types of realized graphs.



The sum of the probabilities of the first two events gives $\tilde{P}_j(A_{-i}, \mathbf{q}_{-\{j,i\}})$.



The sum of the probabilities of this event gives $Q_{ij}(A, \mathbf{q}_{-\{i,j\}})$.

Nash Equilibrium and Social Optimum

- We use these characterizations to express the utility function of agent i as

$$u_i(A, \mathbf{q}) = (1 - (1 - q_i)\tilde{P}_i(A, \mathbf{q}_{-i})) - c_i(q_i).$$

- Similarly, we can write the social welfare function as:

$$\begin{aligned} W(A, \mathbf{q}) &= \sum_{j \in V} u_j(A, \mathbf{q}) \\ &= \sum_{\substack{j \in V \\ j \neq i}} \left[1 - (1 - q_j)\left(\tilde{P}_j(A_{-i}, \mathbf{q}_{-\{i,j\}}) + (1 - q_i)Q_{ij}(A, \mathbf{q}_{-\{i,j\}})\right)\right] - c_j(q_j) \\ &\quad + (1 - (1 - q_i)\tilde{P}_i(A, \mathbf{q}_{-i})) - c_i(q_i). \end{aligned}$$

- We use $\mathbf{q}^*$ to denote the pure strategy Nash Equilibrium (security profile at which there exists no profitable unilateral deviations).
- We use $\mathbf{q}^s$ to denote the social optimum (global maximum of the social welfare function).

Theorem

There exists a pure-strategy Nash Equilibrium and a social optimum.

- Follows from the continuity of the utility function $u_i(A, \mathbf{q})$ in $\mathbf{q}$ and concavity in q_i (similarly, continuity of $W(A, \mathbf{q})$ in $\mathbf{q}$).

Best-response Characterizations

- We can characterize the "optimal" strategies of agents using the network effect representation.
- Let $B_i(A, \mathbf{q}_{-i})$ denote the best response strategy of agent i (strategy q_i that maximizes his utility function given $\mathbf{q}_{-i}$). The strategy $B_i(A, \mathbf{q}_{-i})$ satisfies

$$c_i'(B_i(A, \mathbf{q}_{-i})) = \tilde{P}_i(A, \mathbf{q}_{-i})).$$

- Similarly, let $S_i(A, \mathbf{q}_{-i})$ denote the welfare maximizing strategy of agent i (strategy q_i that maximizes the welfare function given $\mathbf{q}_{-i}$). The strategy $S_i(A, \mathbf{q}_{-i})$ satisfies

$$c_i'(S_i(A, \mathbf{q}_{-i})) = \tilde{P}_i(A, \mathbf{q}_{-i})) + \sum_{\substack{j \in V \\ j \neq i}} (1 - q_j) Q_{ij}(A, \mathbf{q}_{-\{i,j\}}).$$

Symmetric Networks

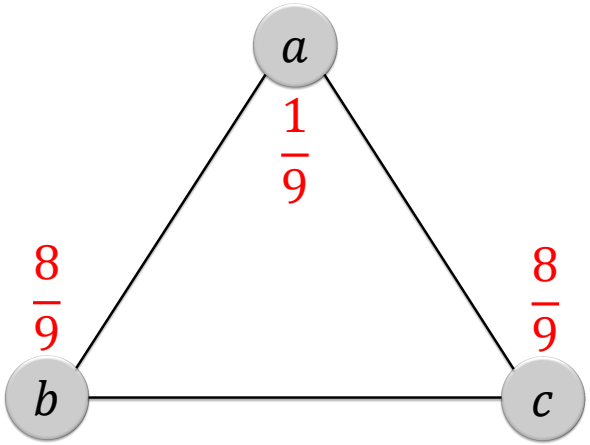
- Consider a symmetric environment:
  - Network is symmetric: for any permutation $\pi(\cdot) : V \mapsto V$ over nodes, $A' = \pi(A)$ has the same distribution as A.
  - All agents have the same cost function, i.e., $c_i(x) = c(x)$ for all i.
- Let q^e, q^s denote investments at the symmetric equilibrium and social optimum.

Proposition

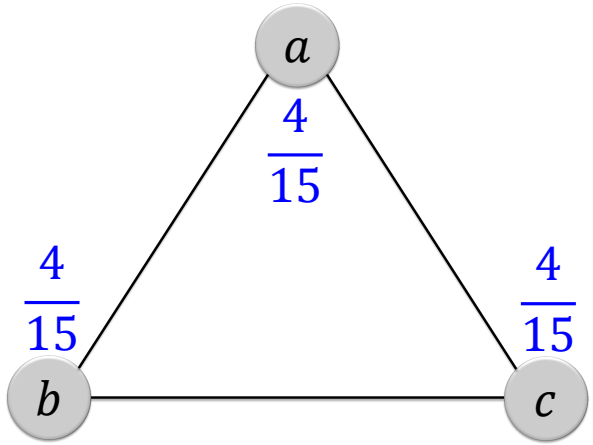
For a symmetric environment, a symmetric equilibrium exists and is unique. Moreover, we have $q^e \leq q^s$, i.e., the investment level at the symmetric equilibrium is less than or equal to that at the social optimum.

- Symmetric equilibria of symmetric networks always involve underinvestment.
- Intuitive since otherwise all agents would overinvest which would be inconsistent with positive externalities.
- Symmetric networks do not preclude asymmetric equilibria, which may still involve overinvestment.

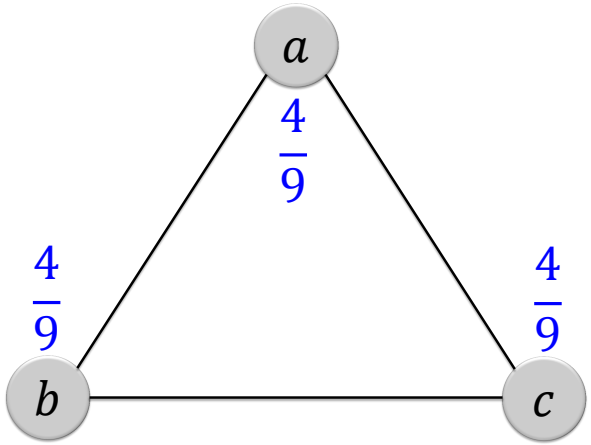
Asymmetric Equilibria of Symmetric Networks



Asymmetric Equilibrium.



Symmetric Equilibrium.



Social Optimum.

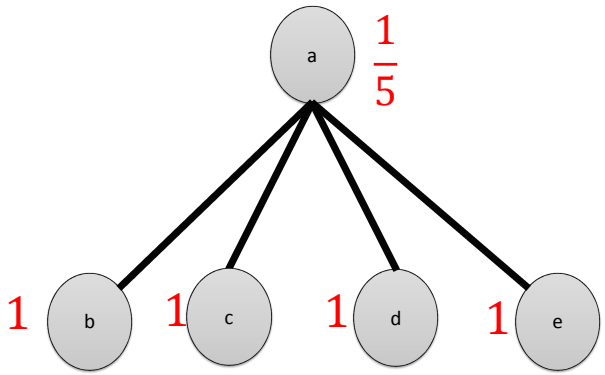
$$c(q) = q^2\left(\tfrac{49}{24} - \tfrac{5}{4}q\right)$$

- *Intuition:* Once the equilibrium is asymmetric, the underinvestment of one agent will trigger overinvestment by others.

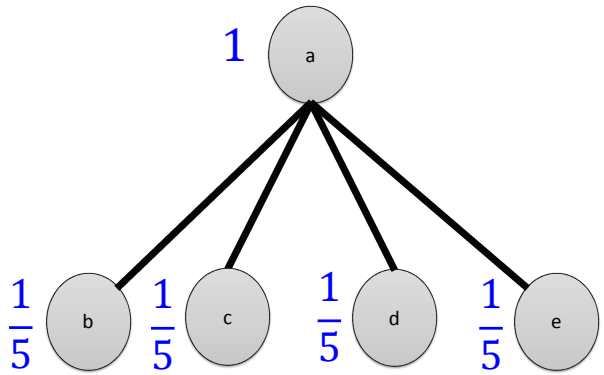
Expected Number of Infected People

- We can even show that the expected number of infected people at equilibrium is strictly lower than that in the social optimum.
- Recall Example 1:

Equilibrium



Social Optimum



Expected number of infected people at equilibrium is 0.16, at social optimum is 0.64.

- We next show that for tree network structures under stronger assumptions on the investment cost function, we always have underinvestment.

Sufficiently Convex Cost Functions

Assumption 2 (Sufficiently Convex Cost)

For each i, the function $c_i : [0,1] \to \mathbb{R}$ satisfies Assumption 1 and is sufficiently convex, i.e., $c_i'(q)(1-q)$ is strictly increasing over [0,1].

- Example: $c_i(q) = -q - \log(1-q)$ is a sufficiently convex cost function.
- Denote the best response strategy of i, $B_i(A, \mathbf{q}_{-i}) = q_i$ and recall that this satisfies

$$c_i'(q_i) = \tilde{P}_i(A, \mathbf{q}_{-i}),$$

implying

$$P_i(A, q_i, \mathbf{q}_{-i}) = c_i'(q_i)(1-q_i).$$

Lemma

Suppose Assumption 2 holds. Let $\mathbf{q}$ *and* $\bar{\mathbf{q}}$ *be such that* $\tilde{P}_i(A, \mathbf{q}_{-i}) \geq \tilde{P}_i(A, \bar{\mathbf{q}}_{-i})$. *Then, we have* $P_i(A, B_i(A, \mathbf{q}_{-i}), \mathbf{q}_{-i}) \geq P_i(A, B_i(A, \bar{\mathbf{q}}_{-i}), \bar{\mathbf{q}}_{-i})$.

- Even though underinvestment by others triggers overinvestment by agent i, sufficiently convex cost functions ensure that i's overall infection probability increases when this is the case—and thus bounding how much i's investment can increase.

Uniqueness with Tree Structure

Theorem

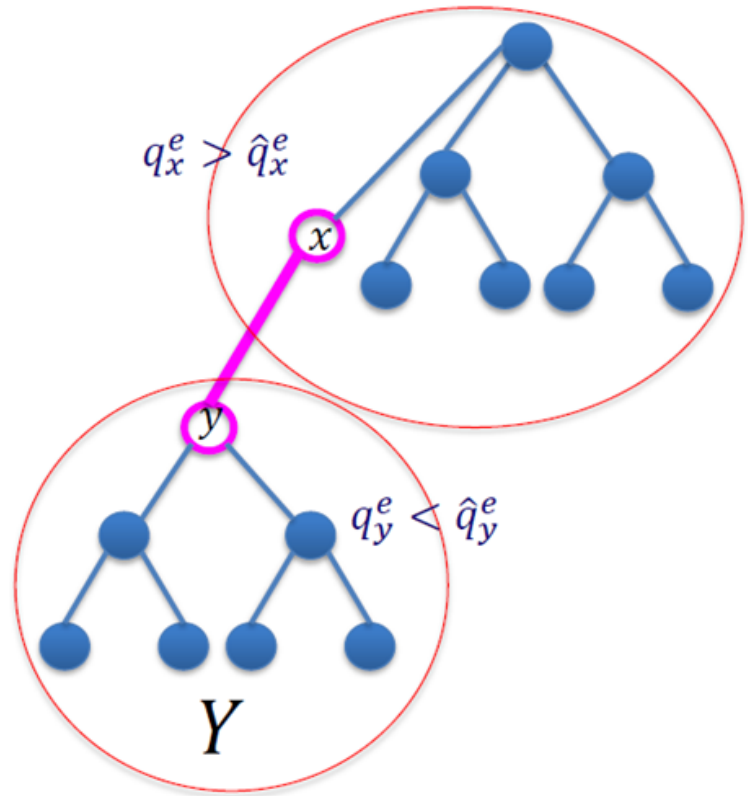
Suppose Assumption 2 holds. For any tree network structure, there exists a unique pure-strategy Nash equilibrium.

Proof Idea: Assume there exists two equilibria: $\mathbf{q}^e, \hat{\mathbf{q}}^e \in [0,1]^n$.

- There exists $(x, y) \in E$ such that
  $q_x^e > \hat{q}_x^e, \quad q_y^e < \hat{q}_y^e, \quad q_v^e \leq \hat{q}_v^e$ for all $v \in Y$.
- By sufficient convexity, we have
  $\mathbf{P}_x(A, \mathbf{q}^e, \Phi) > \mathbf{P}_x(A, \hat{\mathbf{q}}^e, \Phi), \quad \mathbf{P}_y(A, \mathbf{q}^e, \Phi) < \mathbf{P}_y(A, \hat{\mathbf{q}}^e, \Phi)$.
- Using the tree network structure,

$$\mathbf{P}_x(A, \mathbf{q}^{\mathbf{e}}, \Phi) = \hat{P}_x(\mathbf{q}^{\mathbf{e}}) + (1 - q_x^e)\hat{P}_y(\mathbf{q}^{\mathbf{e}}),$$
$$\mathbf{P}_y(A, \mathbf{q}^{\mathbf{e}}, \Phi) = \hat{P}_y(\mathbf{q}^{\mathbf{e}}) + (1 - q_y^e)\hat{P}_x(\mathbf{q}^{\mathbf{e}}),$$

  $\hat{P}_x$ and $\hat{P}_y$: infection probabilities of agents x and y when y and x are removed from the network.
- Since $q_v^e < \widehat{q}_v^e$ for all $v \in Y$, we have $\hat{P}_y(\mathbf{q}^{\mathbf{e}}) \geq \hat{P}_y(\widehat{\mathbf{q}}^{\mathbf{e}})$.
- We use this with $q_x^e > \hat{q}_x^e$ to show
  $\mathbf{P}_y(A, \mathbf{q}^e, \Phi) > \mathbf{P}_y(A, \hat{\mathbf{q}}^e, \Phi)$, contradiction.



Expected Infection with Tree Structure

- Let $I(A, \mathbf{q})$ denote the expected number of infected people given a network A and security profile $\mathbf{q}$.
- Define the contribution of i to infections in sub-graph $A_{-\bar{V}}$ as:

$$C_I\left(i, A_{-\bar{V}}\right) = (1-q_i)\left(\tilde{P}_i(A_{-\bar{V}}, \mathbf{q}_{-\bar{V}\cup\{i\}}) + \sum_{j\in V-\bar{V}\cup\{i\}} (1-q_j) Q_{ij}(A_{-\bar{V}}, \mathbf{q}_{-\bar{V}\cup\{i\}\cup\{j\}})\right).$$

- By Decomposition result, for any given set $\bar{V} \subset V$ and for $i \notin \bar{V}$ we have,

$$I(A_{-\bar{V}}, \mathbf{q}_{-\bar{V}}) = C_I(i, A_{-\bar{V}}) + I(A_{-\bar{V}\cup\{i\}}, \mathbf{q}_{-\bar{V}\cup\{i\}}),$$

- Recall that $\mathbf{q}^e$ denotes the Nash equilibrium and $\mathbf{q}^s$ denotes the social optimum.

Theorem

Suppose Assumption 2 holds. In any tree network structure, we have $I(A, \mathbf{q}^e) \geq I(A, \mathbf{q}^s)$.

- This theorem also holds for any random graph where the realizations correspond to a set of potentially disconnected trees.

Proof Idea I

- If $\mathbf{q}^s \geq \mathbf{q}^e$, we are done.
- Otherwise, let $V_1 = \{i \in V \mid q_i^s < q_i^e\} = \{1, \ldots, k\}$ for some $k \leq n$.

Lemma

For a tree structure, for any $\bar{V} \subset V$,

$$\sum_{i \in \bar{V}} C_I(i, A_{-\bar{V}}) \leq \sum_{i \in \bar{V}} C_I(i, A)$$

- Proof idea: for tree structures, if node k is removed, then the network effect of agent i on agent j either remains constant (if the path between the two did not include k) or decreases to zero (if the path did include k. (This is not necessarily true for other graphs).

Proof Idea II

- Recall that in social optimum solution,

$$c'(q_i^s) = \tilde{P}_i(A, \mathbf{q^s}_{-i}) + \sum_{\substack{j \in V \\ j \neq i}} (1 - q_j^s) Q_{ij}(A, \mathbf{q^s}_{-\{i,j\}}) = \frac{C_I(i, A)}{1 - q_i^s}.$$

- We further have for any given set $\bar{V} \subset V$,

$$\begin{aligned} I(A_{-\bar{V}}, \mathbf{q}^s_{-\bar{V}}) &= C_I(i, A_{-\bar{V}}) + I(A_{-\bar{V}\cup\{i\}}, \mathbf{q}^s_{-\bar{V}\cup\{i\}}) \\ &\leq C_I(i, A) + I(A_{-\bar{V}\cup\{i\}}, \mathbf{q}^s_{-\bar{V}\cup\{i\}}) = c'(q_i^s)(1 - q_i^s) + I(A_{-\bar{V}\cup\{i\}}, \mathbf{q}^s_{-\bar{V}\cup\{i\}}) \end{aligned}$$

- Applying to the agents in V_1 recursively, we have,

$$I(A, \mathbf{q}^s) \leq I(A_{-V_1}, \mathbf{q}^s_{-V_1}) + \sum_{i \in V_1} c'(q_i^s)(1 - q_i^s).$$

- In the equilibrium we have, $c'(q_i^e) = \tilde{P}_i(A, \mathbf{q}_{-i})$. Moreover,
  - $I(A, \mathbf{q}^e) \geq \sum_{i \in V_1} \mathbf{P}_i(A, \mathbf{q}^e) + I(A_{-V_1}, \mathbf{q}^e_{-V_1})$.
  - $I(A_{-V_1}, \mathbf{q}^e_{-V_1}) \geq I(A_{-V_1}, \mathbf{q}^s_{-V_1})$ since $q_i^s \geq q_i^e$ for all $i \in V - V_1$,
  - And $\sum_{i \in V_1} \mathbf{P}_i(A, \mathbf{q}^e) = \sum_{i \in V_1} c_i'(q_i^e)(1 - q_i^e) \geq \sum_{i \in V_1} c_i'(q_i^s)(1 - q_i^s)$ by the sufficiently convex cost assumption and the fact that $q_i^s < q_i^e$ for $i \in V_1$ by hypothesis.

Local Tree Network Structures I

- This result generalizes to random networks with local tree structures.

Definition (h-Local Tree Structure)

A random network has h-local tree structure if the connected component attached to each agent is acyclic with probability at least h.

Theorem

Suppose Assumption 2 holds. In any $(1-\epsilon)$*−local tree network structure,* $I(A, \mathbf{q}^s) \leq I(A, \mathbf{q}^e) + \epsilon\, n$.

- Follows by considering expected number of infected people in cyclic and acyclic components of each realization.
- Expected number of infected people in cyclic components bounded above by expected number of agents belonging to cyclic components.

Local Tree Network Structures II

- In the previous theorem, it can be ensured that $\epsilon(n)\, n$ goes to zero as $n \to \infty$ (i.e., for large networks).

Proposition

Suppose that each edge is active independently with probability p. If the size of the largest connected component of the activated graph is bounded by C, the connected component attached to an agent v is acyclic with probability $(1-p)^{C^2-C}$.

- This probability is computed recursively by activating agents one at a time in the realized network.

Local Tree Network Structures III

- For large graphs, we can provide conditions on the graph and p under which $\epsilon(n)\, n$ goes to zero as $n \to \infty$.

Proposition

A $d-$ regular pseudo-random graph with $d \gg \sqrt{n}$ and $p < \frac{1}{d}$ is a $1 - \epsilon(n)$-local tree network structure where $\lim_{n\to\infty} \epsilon(n) = 0$. Also for any d, if $p < \frac{1}{n \log^2(n)}$, then $\lim_{n\to\infty} \epsilon(n)n = 0$.

- This relies on a result from [Frieze, Krivelevich, Martin, 2003], which states that in a $d-$regular pseudo-random graph if $p < \frac{1}{d}$, then with high probability, the maximum component size is $\log(n)$.

Symmetric Random Networks I

Theorem

Suppose Assumption 2 holds. In a symmetric random network, there exists a unique pure-strategy Nash equilibrium.

- Proof idea: Let q^e denote the symmetric equilibrium security level. If there exists an asymmetric equilibrium $\mathbf{q}^*$, we can find two agents i, j such that $q_j^* < q^e < q_i^*$.
  - By symmetry and decomposition, infection probability of node i linear in $(1 - q_i) \rightarrow$ infection probability of i less than j.
  - Assuming $c(q)$ is sufficiently convex, in $\mathbf{q}^*$ infection probability of i should be higher than j, which is a contradiction.

- This shows for such networks, there exists no asymmetric equilibrium.

Symmetric Random Networks II

- Recall that q^e, q^s denote the investment levels at the symmetric equilibrium and the social optimum.

Theorem

Suppose Assumption 2 holds. In any random symmetric network, we have $I(A, q^e) \geq I(A, q^s)$.

Ranking Symmetric Random Networks

- A symmetric random network $\hat{G}$ can be represented by a fixed base graph G and the uniform allocation of each agent to one of the nodes of this graph.

Proposition

Suppose Assumption 2 holds. For two base graphs G_1 and G_2, if $G_2 \subset G_1$ (i.e., G_1 has additional links relative to G_2), then $I(\hat{G}_2, q_2^e) \leq I(\hat{G}_1, q_1^e)$.

- Follows from the characterization:

Lemma

For two symmetric random networks $\hat{G}_1$ and $\hat{G}_2$, $I(\hat{G}_1, q_1^e) \geq I(\hat{G}_2, q_2^e)$ if and only if $I(\hat{G}_1, q) \geq I(\hat{G}_2, q)$ for all $q \in [0, 1]$.

- At the same security profile, more connections clearly create more infection.
- *Intuition:*
  - With symmetric random networks, the only thing that matters is the probability of infection transmitted to me from the rest of the network.
  - Sufficiently convex cost functions ensure that when this probability is higher, my investment goes up, but not enough to reduce my overall probability of infection.

Ranking Symmetric Random Trees

Definition (Distance vector)

For a given tree graph T, $d_T = (d_1, \ldots, d_n)$ is the distance vector of T, where d_i is the probability that two randomly selected vertices(with replacement) are at distance $i-1$ from each other.

- *Example*: For a star with n nodes, $d = (\frac{1}{n}, \frac{2\cdot(n-1)}{n^2}, \frac{(n-1)(n-2)}{n^2}, 0, \ldots, 0)$.

Definition (Domination, $\preccurlyeq$)

For two tree graphs T and T', let $d_T = (d_1, \ldots, d_n)$ and $d_{T'} = (d'_1, \ldots, d'_n)$ represent their distance vector. Graph T dominates T', denoted by $d' \preccurlyeq d$, if and only if for all $1 \leq i \leq n$, $\sum_{j=1}^{i} d'_j \leq \sum_{j=1}^{i} d_j$.

- A star dominates all trees and a path will be dominated by all trees.

Proposition

Suppose Assumption 2 holds. For two symmetric random trees with base graphs T_1 and T_2, if $T_2 \preccurlyeq T_1$, then $I(T_2, q_2^e) \leq I(T_1, q_1^e)$.

Ranking Symmetric Networks

- For a given base graph G, let $\varrho_G(\frac{k}{n})$ denote the expected size of the connected component attached to a random agent v in the induced subgraph of G over set V_k where V_k is a set of k randomly selected agents from G.

Proposition

Suppose Assumption 2 holds. For two symmetric random networks with base graphs G_1,G_2, if $\varrho_{G_2}(x) \leq \varrho_{G_1}(x)$ for all $x \in \{\frac{1}{n}, \ldots, 1\}$, then $I(\hat{G}_1, \mathbf{q}_1^e) \geq I(\hat{G}_2, \mathbf{q}_1^e)$.

- In a given graph G and with a given security $q \in [0,1]$,

$$I(G,q) = \sum_{k=1}^{n} \binom{n}{k} (1-q)^k \cdot q^{n-k} \cdot \varrho_G(\frac{k}{n})$$

- Can establish necessity of (a version of) this condition using Bernstein theorem.

Strategic Attack

- In some security domains, such as wars or terror attacks, the origin of attacks is not a random event, but rather the decision of a strategic adversary.
- Key assumption: The attacker observes the security levels of the agents.
- He selects one of the agents to attack with the goal of maximizing the expected number of infected people.
  - Attacker decision is a probability vector $\Phi = (\rho_1, \ldots, \rho_n)$, where ρ_i is the probability of attacking agent i.
  - His payoff is given by the expected number of infected people minus the cost of the attack given by $\xi(\Phi) = \sum_{i=1}^{n} \xi(\rho_i)$ where ξ is a convex function.
- We analyze the Stackelberg equilibrium of the resulting game:
  - The agents select their security levels anticipating the decision of the attacker and the attacker optimizes his attack strategy given the security choices.

Nonexistence of a Pure Strategy Nash Equilibrium

- To understand the role of cost, consider the case when attack decisions are costless, which will lead to nonexistence of a pure Nash equilibrium.

Example

Consider a network with 2 singleton agents.

- For any security profile $\mathbf{q}$, the attacker selects the agents with minimum security level to attack.
- The following list considers candidate equilibria and profitable unilateral deviations, establishing nonexistence of a Nash equilibrium:

1. $q_1 < q_2$: Agent 2 has an incentive to decrease q_2 since this will not change the attack strategy of the attacker.
2. $q_1 = q_2 < 1$: Agent 2 has an incentive to slightly increase q_2. This reduces his attack probability from 1/2 to 0 while slightly increasing his cost.

Strategic Attack Model: Infection Probability

- Expected number of infected people when agent i is targeted can be expressed in terms of the infection probability of agent i in the random attack model.

Lemma

Given network A and security profile $\mathbf{q}$, we have

$$I(A, \mathbf{q}, e_i) = |V| P_i(A, \mathbf{q}) = |V| \cdot (1 - q_i) \tilde{P}_i(A, \mathbf{q}),$$

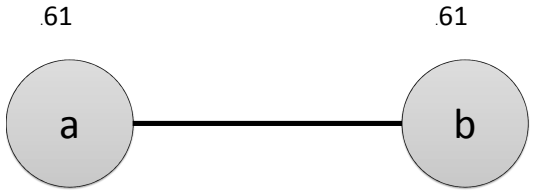
where $I(A, \mathbf{q}, e_i)$ denotes the expected number of infected people when i is attacked.

- *Intuition:*
  - Infection probability of agent i under the random attack model is the probability of having a path between i and a randomly selected agent.
  - Similarly, the expected number of infected people when i is attacked is the sum (over all j) of the probability of having a path between i and j.
- Allows us to use the recursive characterizations of the random attack model in expressing the utility function of the attacker.

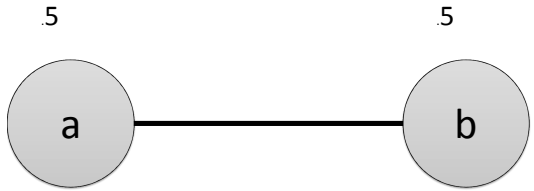
Costly Strategic Attack May Lead to Overinvestment

- For costly strategic attack, we establish existence of a pure strategy Nash Equilibrium under some assumptions on the agent and attacker cost functions.
- This equilibrium may involve overinvestment even for tree networks with sufficiently convex investment cost functions.

Equilibrium



Social Optimum



$$c'(q) = \frac{1}{2(1-q)}, \xi(\rho) = \frac{\rho^2}{20}$$

- *Intuition:* Preventive activities can create negative instead of positive externalities when they shift attacks to other nodes.

Overinvestment in Symmetric Random Networks

- For symmetric random networks and some additional conditions on cost functions, we show that there exists a pure strategy symmetric Nash equilibrium. Moreover, equilibrium may involve overinvestment.

Theorem

In the strategic attack model, for any symmetric random network, in the symmetric equilibrium security profile compared to the social optimum,

- *agents will overinvest if* $\frac{n-1}{n^2} c'^{-1}(\frac{1}{n})(1 - c'^{-1}(n)) \geq \xi''(\frac{1}{n})$.
- *agents will underinvest if* $\frac{1}{n} \leq \xi''(\frac{1}{n})$.

- If $\xi''\left(\frac{1}{n}\right)$, the second derivative of the attacker's cost in the symmetric equilibrium, is not too large, then the attacker will change his attack plan as a function of the investment profile and this encourages overinvestment.
- If it is sufficiently large, then we are close to a situation of random attack and hence underinvestment.

Conclusions and Future Work

- We provided a systematic analysis of the equilibrium and optimal security investments in general random networks subject to an attack.
- We show how new economic forces arise in the setting that were absent in symmetric equilibria.
- We establish that overinvestment arises in a range of settings for well-defined economic reasons in contrast to the underinvestment presumed by the existing literature.

Future Work:

- More detailed analysis of network structure.
- Environments that feature both strategic substitutes and complements.
- Intervention mechanisms (subsidies, taxation) that will improve performance in equilibrium.