Federal Reserve Bank of New York
Staff Reports

Insider Networks

Selman Erol
Michael Junho Lee

Staff Report No. 862
August 2018
Revised May 2019

Insider Networks
Selman Erol and Michael Junho Lee
Federal Reserve Bank of New York Staff Reports, no. 862
August 2018; revised May 2019
JEL classification: D85, G14, G20

Abstract

How do insiders respond to regulatory oversight? History suggests that they form sophisticated networks to share information and circumvent regulation. We develop a theory of the formation and regulation of information transmission networks. We show that agents with sufficiently complex networks bypass any given regulatory environment. In response, regulators employ broad regulatory boundaries to combat gaming, giving rise to regulatory ambiguity. Tighter regulation induces agents to migrate transmission activity from existing social networks to a core-periphery insider network. A small group of agents endogenously arise as intermediaries for the bulk of information. We provide centrality measures that identify intermediaries.

Key words: network formation, insider trading, regulatory ambiguity, endogenous intermediation

---

Lee: Federal Reserve Bank of New York (email: michael.j.lee@ny.frb.org). Erol: Carnegie Mellon University (email: erol@cmu.edu). Ali Polat provided excellent research assistance. The authors thank Nicola Cetorelli, David Childers, Marco Cipriani, Andrea Galeotti, Co-Pierre Georg, Burton Hollifield, Gabriele La Spada, Amir Kermani, Marco Di Maggio, Mihai Manea, Emre Ozdenoren, Bryan Routledge, Omer Tamuz, Chris Telmer, and participants in talks at LBS, Oxford, the New York Fed, CMU Tepper, the 2019 Pennsylvania Economic Theory Conference, the UCSB LAEF OTC Workshop, and the Vanderbilt Network Science and Economics Conference for helpful comments.

To view the authors' disclosure statements, visit
https://www.newyorkfed.org/research/staff_reports/sr862.html.

1 Introduction

From 1928 to 1932, Albert H. Wiggin, then president of the Chase National Bank, accumulated over $10,000,000 solely by trading Chase stock. $4,000,000 was made in the Crash of 1929, during which the stock market crashed, and with it, Chase as well. Wiggin had been shorting his own bank. Wiggin's trades were as legal as they were met with public outrage. In an effort to restore confidence in market integrity, the Securities Exchange Act of 1934 was passed which birthed the Security Exchange Commission (SEC). Section 16 of the act, also known as the "anti-Wiggin" proposal, was specifically included to root out abusive securities trading by people with insider information.[1]

In 2008, Mathew Martoma, portfolio manager at hedge fund SAC Capital Advisors, made a twenty minute phone call to owner Steven A. Cohen. Within a span of a week, Cohen reversed his long position in pharmaceutical firms Elan and Wyeth by nearly a billion dollars, which ultimately generated a profit of over $270,000,000. Martoma was later convicted of insider trading. Insider information was passed through a long chain of communication – from Elan to a doctor, to Martoma, who was introduced by an expert network firm. The conviction was the culmination of a painstaking six year investigation by the SEC.[2]

These two instances of insider trading, set apart by nearly a century, draw a striking contrast. In the first case, the insider legally traded directly with his information. In the latter, transmission was achieved through a complex network of connections that had adapted to greater regulatory sophistication. The sender and receiver were otherwise unrelated, with no overlapping social or professional networks, channels through which information typically diffuse. Instead, transmission was facilitated by an intermediary specializing in bringing together sources of information and those who seek it.

In this paper, we develop a model of endogenous network formation to analyze a cat and mouse game between regulators, who set and enforce the regulatory environment, and agents, who form links through which information can be transmitted. In the model, agents benefit from transmitting information to others that can utilize the information. The regulator's objective is to detect and punish agents for sharing information, which imposes negative

[1] http://scholarship.law.upenn.edu/cgi/viewcontent.cgi?article=6444&context=penn_law_review

[2] http://www.newyorker.com/magazine/2014/10/13/empire-edge

social externalities. The regulator is limited in his capacity to observe both agents' networks and their information transmissions. In particular, the regulator must incur an enforcement cost that is increasing in search intensity. Agents can form costly links with other agents, through which they may transmit information. While transmitting through a shorter chain increases expected gains, for instance, by improving the likelihood of successful transmission, agents may form complex networks that allow for information to be transmitted through a long chain of agents in order to circumvent regulation.

First, we show that regulatory ambiguity arises as an equilibrium phenomenon. For any given enforcement strategy potentially chosen by the regulator, agents with extensive networks can conceal their transmissions by using longer chains. Agents can game the system. Importantly, increasing the penalty from detection, or the search intensity of the regulatory environment does not generically hinder transmission. This is reminiscent of Tsebelis (1989). As a result, in equilibrium, the regulator mixes between low and high intensities of enforcement, effectively employing regulatory ambiguity. Doing so induces agents to engage in riskier transmission behavior, allowing the regulator to successfully catch the agent with a positive probability.

Our analysis rationalizes a long standing position taken by regulatory institutions that advocate for flexible, broad guidelines on what constitutes insider trading. Regulatory institutions have been criticized for only loosely defining what constitutes illegal activity pertaining to the use of insider information. With broad rules governing insider trading, courts have been relied upon to ultimately determine illegality. We show that a precise regulatory framework necessarily allows for more gaming, as information networks quickly adapt to the regulatory environment.[3]

Regulatory ambiguity impacts agents' network formation. In particular, we show that agents value being part of a network that enables flexible transmission of information, or which facilitates multiple path lengths of transmission between agents. A flexible network provides agents with the option of transmitting information either through a risky, direct path, or a safer, longer path. In equilibrium, agents form a flexible core-periphery network, where a small group of agents form the core and act as conduits of information on behalf of

[3]The need for deliberately imprecise regulation to effectively combat gaming is a sentiment that extends beyond the context of insider trading regulation. For instance, Greenwood et al. (2017) argues that one of the key benefits of bank stress tests is its flexibility.

the network.[4] This structure embeds the flexibility function into a highly connected core. In doing so, agents form a cost-effective network that ensures that all periphery members have access to a wide set of transmission paths.

Empirical studies have found that, historically, insider information has been shared within tight social networks.[5] We extend our model to consider when agents are endowed with existing social networks. We show that when regulatory enforcement becomes sufficiently strong, agents' information sharing shifts away from their respective social networks and instead prompts the formation of more centralized and complex insider networks. In this context, the model generates an endogenous rise of intermediaries as a reaction to greater regulatory sophistication. Intermediaries in the core are responsible for matching and transmitting information between a large mass of senders and receivers. Moreover, by extending its constituency, the core is able to adjust its flexibility to arbitrarily greater regulatory powers at a negligible cost. This suggests that in an environment where regulation becomes more stringent over time, the flexible core-periphery network structure offers a dynamic form of flexibility as well.[6]

Our results suggest that strengthened regulatory and legislative initiatives may trigger demand for, and therefore creation of, more sophisticated networks. Empirically, the formation of core-periphery networks to disseminate insider information can potentially explain the recent rise of expert networks that have become implicated in insider trading cases in the United States.[7] These firms are consulting intermediaries that specialize in connecting clients looking for experts in various fields – including but not limited to technology, media, medicine, law, and finance – reportedly at rates of up to $1,300 an hour.[8] By providing clients with access to experts that may own proprietary information, expert networks may offer a discrete channel through which information can be shared. These intermediaries have become a growing source of information to hedge funds, private equity firms, and other

[4]This provides a novel channel that may give rise to "law of the few," as discussed in Galeotti and Goyal (2010).

[5]For example, Ahern (2017) finds that insider trading cases predominantly involve familial, social, or geographic ties.

[6]In the US, enforcement indeed appears to have become stronger over time. For example, see Silvers (2016).

[7]See https://www.sec.gov/spotlight/insidertrading/cases.shtml for details on select cases involving expert networks.

[8]https://www.bloomberg.com/news/articles/2018-02-28/investors-are-paying-1-300-per-hour-for-expert-chats

investment firms. Furthermore, by providing an extra layer of information transmission, these firms can also insulate clients from legal trouble, as the current regulatory framework requires proof of knowledge that the nature of the information is, in fact, insider information.

Indeed, legislative and regulatory actions have been claimed to be at least partly responsible for the growth of the expert network industry (Jeng (2013)), which according to some industry estimates, has roughly doubled in size (by revenue) from 2012 to 2018. Regulation Fair Disclosure and the Global Analyst Research Settlements, enacted in the early 2000s, tightened governance on information disclosure and restricting insider information dissemination by financial intermediaries. In the political arena, the Stock Trading on Congressional Knowledge Act (STOCK Act), enacted in 2012, imposed restrictions on profiting from private information derived from Congressional activity. More recently, the European Union rolled out a legislative framework called MiFID II to improve market transparency, and among other things, explicitly bans banks from bundling transaction and research services. While these actions may succeed at displacing existing channels of information diffusion, they may prompt the formation of networks that undermine the main objective to improve market integrity.

The paper is organized as follows. We review related literature and our contribution in Section 2. In Section 3, we analyze the model given an exogenous network and lay out the strategic considerations that necessitate regulatory ambiguity. In Section 4, we solve the main model with network formation and show the equilibrium formation of flexible core-periphery networks. In Section 5, we extend the model by endowing agents with existing social networks, and demonstrate how strengthening regulation can trigger the rise of information intermediaries. In Section 6, we analyze the properties of flexible core-periphery networks and develop measures that can be used to identify intermediaries. We conclude in Section 7. All proofs are in the Appendix.

2 Contribution and Related Literature

This paper is the first to our knowledge to theoretically study the formation of insider networks. As in DeMarzo et al. (1998), our paper takes as given the objective of a regulator to detect and deter the sharing of insider information. We study the joint equilibrium determination of regulation, network formation, and information transmission. We depart from the conclusions of DeMarzo et al. (1998) – enforcement policy is necessarily ambiguous (i.e. subject to randomness) due to the possibility of gaming by agents.

An extensive literature examines the diffusion of information through social networks in financial markets. Cohen et al. (2010) finds strong evidence of information diffusion through educational ties. Maggio et al. (2017) finds extensive evidence of information diffusion through broker networks. Ahern (2017) shows that a majority of prosecuted insider trading cases involve insider information being transferred through geographical, family, and social networks. An implication of our paper is that as regulatory pressures increase, insider trading activity migrates from existing networks to those that better insulate agents from detection and prosecution.

Our paper is related to the literature on information transmission in endogenous networks. Acemoglu et al. (2014) studies how information aggregation occurs through communication on endogenous social networks. Bloch and Dutta (2009) studies how communication networks with endogenous link strength bring rise to star networks. We make a unique contribution by studying the formation of information networks and its interaction with the regulatory environment.[9] We show that a core-periphery structure arises endogenously in insider networks, and furthermore show that a small number of agents act as intermediaries to facilitate information transmission. Our result also relates to bottlenecks and essential intermediaries which Manea (2018) takes as given.

Our insights are applicable to other settings in which agents use networks as a strategic tool for the transmission of information or goods in a game against an adversary. Agents committing money laundering may utilize a long chain of financial intermediaries in order to obfuscate the source and destination of money transfers. The model is relevant for studying networks for organized crime or terrorism in which agents form networks to conceal communication and money transfers. Our results are consistent with empirical studies that document the use of long intermediation chains in terrorist networks intended to conceal relationships and preserve secrecy (Krebs (2002)), as well as the emergence of core-periphery structures in transnational criminal networks (Williams (2001)). Finally, our paper contributes broadly to the literature on attack and defense in networks.[10]

On the technical front, our paper develops an approach to modeling formation of net-

[9] A few papers have studied the impact of regulation on networks in other contexts. For example, see Erol and Ordoñez (2017) and Erol (2017).

[10] For example, see Acemoglu et al. (2016), Dziubiński and Goyal (2017), Hoyer (2012), Haller (2016), and Hoyer and Jaegher (2016).

works with a continuum of agents in which nodes are neither restricted to have a sparse neighborhood nor a dense neighborhood. A common challenge in network theory involves dealing with the discrete nature of network structure, which limits tractability. We provide a tractable method to incorporating a continuum of agents that form topologically non-trivial networks.

3 The Value of Regulatory Ambiguity

In this section, we take as given the agents' networks, and study the information transmission and regulation game. We extend the model to consider endogenous network formation in Section 4.

3.1 Model

Agents and the network. Consider a set of agents A. An edge is an element of A^2. $E \subset A^2$ is used to denote a set of edges. Given E, we use $N^-_{i,E} = \{j : (j,i) \in E\}$ is the in-neighbors of i. These agents can send information to i. $N^+_{i,E} = \{j : (i,j) \in E\}$ is the out-neighbors of i. i can send information to these agents. Denote $N_{i,E} = N^-_{i,E} \cup N^+_{i,E}$. A path from i to j in E is a finite sequence of agents $(n_0, n_1, ..., n_k)$ where $n_0 = i$, $n_k = j$, and $n_{u+1} \in N^+_{n_u,E}$ for all $u = 0, 1, ..., k-1$. We say that this path has length k.

Information transmission. Consider two agents, sender s and receiver r. If s successfully transmits the information to r, and r trades on the transmitted information, they derive private payoffs. First, s decides whether to transmit information to r, and if so, the path through which to transmit information.[11,12] Accordingly, an action for s regarding information transmission is $p \in \mathscr{P} \cup \{o\}$, where o denotes the decision not to send information. Conditional on receiving information, r trades based on the information.

Regulation. The use of this information is assumed to impose a negative social externality. There is a regulator whose objective is to minimize social costs that arise from the exploitation of information. In the event that r trades using information received from s, the identities of s and r are assumed to be revealed to the regulator, who monitors mar-

[11] For simplicity we assume that s chooses the path. This is akin to agents coordinating on how to transmit the information over the network from s to r. Alternatively, one can interpret it as s sending the information along with instructions on how to transmit it.

[12] Using a path dominates using a non-path trail. This becomes clear once the nature of the regulation is described.

kets for suspicious activity. In order to punish s and r, the regulator must provide direct evidence that s transmitted information to r. However, the regulator does not observe the transmission path directly, and instead must initiate costly search.[13] The regulator is able to punish the pair if and only if he can directly unravel the path through which information was indeed relayed from the sender to the receiver.

Formally, upon transmission ($p \neq o$), the regulator chooses a search intensity m. Given an intensity m, the regulator is able to search m steps along the transmission path starting from r. For example, suppose that information was transmitted via a path $s = n_0 \rightarrow n_1 \rightarrow ... \rightarrow n_{l-1} \rightarrow n_l = r$. The regulator inspects $r = n_l$, and finds who has sent the information to n_l. After finding out that n_{l-1} sent the information to n_l, regulator inspects n_{l-1} and finds out who sent the information to n_{l-1}. In the end, the regulator discovers the identities of $n_{l-1}, n_{l-2}, ..., n_{(l-m)^+}$.[14] If $m \geq l$, the regulator is able to identify the whole path through which information has been transmitted from s to r. Otherwise, the regulator fails to produce sufficient evidence for illegal insider trading.[15] If the regulator can map the path through which sender s transmitted information to receiver r, the agents are "caught." These are illustrated in Figures 1 and 2.

Timing and payoffs. First, s chooses an action in $p \in \mathscr{P} \cup \{o\}$. If $p = o$ the game ends. If $p \neq o$, the regulator observes s and r, and chooses search intensity m. The timing is summarized in Figure 3. If the information is transmitted by s via a path of length l, the exploitation of insider information yields a total benefit of $\Pi(l)$, where $\Pi(\cdot)$ is strictly decreasing. The payoff from taking the action o and not transmitting information is equal to 0. Given our focus on the interplay between agents and the regulator, we assume that agents fully cooperate to maximize their total expected payoff.

The regulator incurs a cost $\kappa(m)$ associated with search intensity m, where $\kappa(\underline{m}) = 0$ for some $\underline{m} \geq 2$ and κ is strictly increasing for all $m \geq \underline{m}$. Accordingly, we restrict the action space of the regulator to $\mathcal{M} := \{m \in \mathbb{N} : \underline{m} \leq m\}$ for simplicity. The regulator incurs a cost $B > 0$ if insider trading takes place and goes unpunished. Hence, B captures the social

[13] In this section, the regulator knows the network. In the next section with endogenous network formation, the regulator forms beliefs about the network.

[14] $x^+ = max\{x, 0\}$ denotes the positive part of x.

[15] Here, search is modeled as choosing m, which can be interpreted as inspecting m nodes and finding out who have sent them the information. One may consider alternative search protocols, e.g. searching through paths, searching through trails, searching from s towards r, searching from both s and r towards each other, etc. These protocols ultimately boil down to choosing m.

Figure 1: Successful search by the regulator

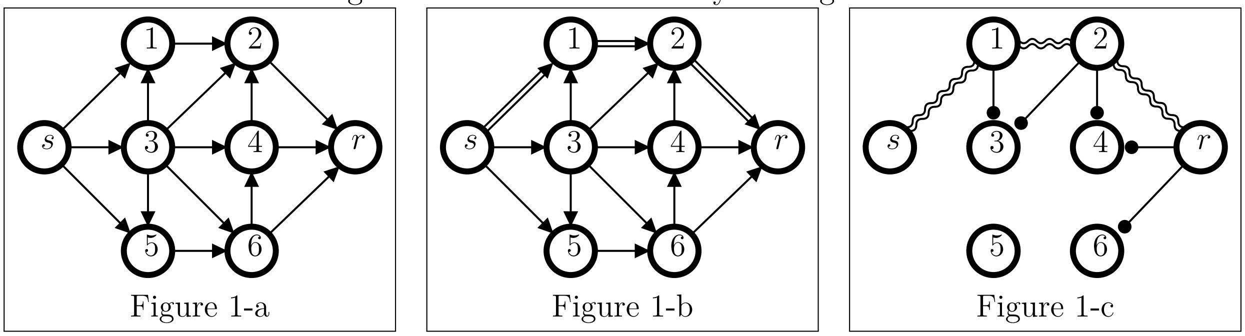


Figure 1 illustrates the game. The network available is shown in Figure 1-a. There are many paths that the agents can utilize. For example, $s \to 3 \to 4 \to r$ is a path of length 3; $s \to 3 \to 5 \to 6 \to 4 \to 2 \to r$ is a path of length 6. Figure 1-b illustrates the strategy of choosing the path $s \to 1 \to 2 \to r$. Figure 1-c illustrates the strategy of the regulator choosing some $m \geq 3$. Given m, the regulator can hold investigations for m steps until evidence is found. In the first round, r and all of its incoming links are inspected: $2 \to r$, $4 \to r$, and $6 \to r$. Upon the investigation, the regulator finds that the information has been sent by 2. In the second round, the regulator investigates 2 and all of its incoming links, $1 \to 2$, $3 \to 2$, and $4 \to 2$. The regulator discovers that information has been sent by 1. In the third round, the regulator inspects 1 and all of its incoming links: $s \to 1$ and $3 \to r$. The regulator finds definitive evidence that information was transmitted from s to r, and can inflict a punishment.

externality cost of insider trading. If agents get caught, they receive an expected punishment $K(l) > 0$. In order for enforcement to be credible, we assume that the cost $K(l)$ of getting caught by the regulator is sufficiently high such that no transmission is preferred to certainty of getting caught: $K(l) > \Pi(l)$ for all l. In Appendix A, we extend the model to allow for insider to explicitly trade on financial markets. Under this extension, both $\Pi(l)$ and $K(l)$ are endogenously determined, though the main insights of the model remain unchanged.

Let $\ell(p)$ be used to denote the path length associated with transmission path $p \in \mathscr{P}$. For an action $p \in \mathscr{P} \cup \{o\}$ and search intensity m, the (total sum of) agents' ex-post payoffs v_A are given by

$$v_A(p; m) = \begin{cases} \Pi(\ell(p)) & \text{if } m < \ell(p) \\ \Pi(\ell(p)) - K(\ell(p)) & \text{if } m \geq \ell(p) \end{cases}$$

for $p \in \mathscr{P}$ and 0 for $p = o$. The total ex-post payoff to the regulator is given by

$$v_R(m; p) = \begin{cases} -\kappa(m) & \text{if } m \geq \ell(p) \\ -\kappa(m) - B & \text{if } m < \ell(p) \end{cases}$$

for $p \in \mathscr{P}$ and 0 for $p = o$.

Figure 2: Failed search by the regulator

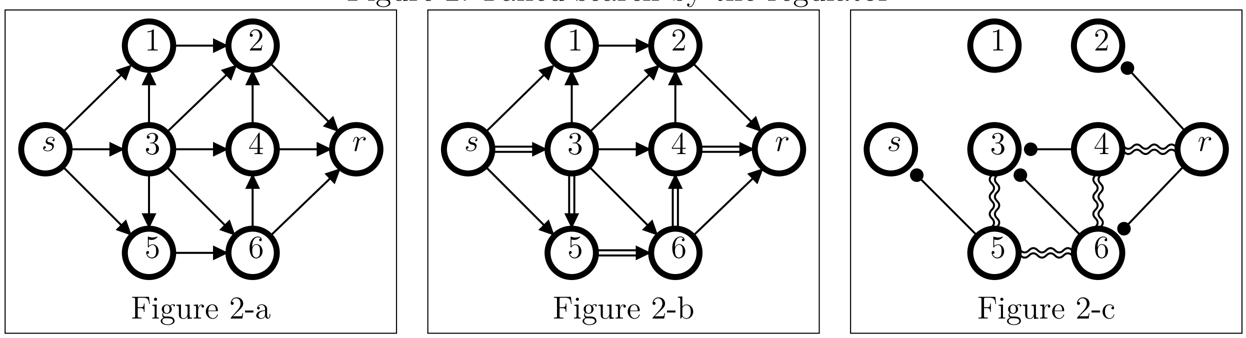
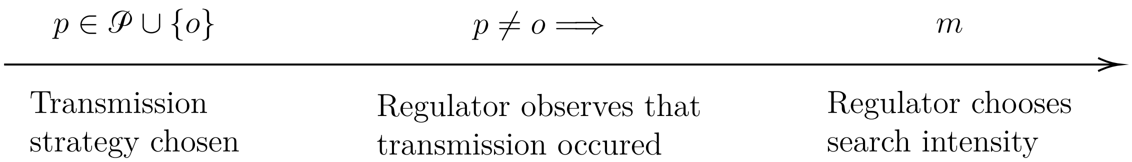


Figure 2 illustrates the game with failed search. Figure 2-a is the network. Figure 2-b illustrates the transmission strategy employed by the agents: $s \to 3 \to 5 \to 6 \to 4 \to r$. This is a path of length 5. Figure 2-c illustrates the search upon the regulator choosing $m = 4$, the maximum number of rounds for search. First, r and all of its incoming links are inspected: $2 \to r$, $4 \to r$, and $6 \to r$. Upon the investigation, the regulator finds that the information has been sent by 4. Then 4 and all of its incoming links are inspected: $3 \to 4$ and $6 \to 4$. Upon the investigation, the regulator finds that the information has been sent by 6. The search goes on this way. In the last round, upon the inspection of 5 and all of its links $r \to 5$ and $3 \to 5$, the regulator finds that the information has been relayed by 3. However, the search fails to identify the entire transmission path, and insiders are not prosecuted.

Figure 3: The timing of events for a given network and sender-receiver pair



A strategy for agents is $\sigma_A \in \Delta(\mathscr{P} \cup \{o\})$. A strategy for the regulator is $\sigma_R \in \Delta(\mathcal{M})$. A belief for the regulator, upon observing that agents have engaged in information transmission, is $\tilde{\sigma}_A \in \Delta(\mathscr{P})$. The expected payoffs of the regulator and agents are

$$V_R(\sigma_R; \tilde{\sigma}_A) = \sum_{p \in \mathscr{P}} \sum_{m} \tilde{\sigma}_A(p)\, \sigma_R(m)\, v_R(m; p)$$

$$V_A(\sigma_A; \sigma_R) = \sum_{p \in \mathscr{P}} \sum_{m} \sigma_A(p)\, \sigma_R(m)\, v_A(p; m)\,.$$

Equilibrium. An (Perfect Bayesian) equilibrium consists of strategies (σ_A^*, σ_R^*) and beliefs for the regulator $\tilde{\sigma}_A^*$ such that

- **(Sequential rationality)** Given his beliefs, the regulator's search strategy is optimal:

For all $m \geq \underline{m}$,

$$V_R\left(\sigma_R^*; \tilde{\sigma}_A^*\right) \geq V_R\left(m; \tilde{\sigma}_A^*\right). \tag{1}$$

- **(Consistency)** The beliefs of the regulator are derived from Bayesian updating whenever possible: If $\sigma_A^* \neq o$, for all $p \in \mathscr{P}$,

$$\tilde{\sigma}_A^*(p) = \frac{\sigma_A^*(p)}{1 - \sigma_A^*(o)}$$

- **(Optimal transmission)** For all $p \in \mathscr{P}$,

$$V_A\left(\sigma_A^*; \sigma_R^*\right) \geq max\left\{0, V_A\left(p; \sigma_R^*\right)\right\}. \tag{2}$$

3.2 Regulatory Ambiguity

In this section, we lay out key equilibrium properties in order to understand the regulator's equilibrium enforcement strategy. Suppose that the regulator chooses some intensity $m > 0$. For any given m, agents' best response is to transmit information through a path of length at least $m + 1$ if feasible or not to transmit information:

Lemma 1. *Consider the regulator's strategy of choosing m upon observing (s, r). The best response of agents is given as follows.*

- *Transmit along the shortest path with length greater than m, if there exists such path.*
- *Do not transmit, otherwise.*

Lemma 1 establishes that optimal transmission entails choosing a path that minimizes length conditional on no detection, or no transmission if detection occurs with probability one. Importantly, transmission along a path of at least length $m+1$ avoids detection by the regulator. Since avoiding prosecution is strictly preferred to being prosecuted, agents choose the shortest transmission path conditional on avoiding detection.

Recall that $\kappa(\underline{m}) = 0 < \kappa(\underline{m} + 1)$. Let $\overline{m}$ be the threshold intensity of regulation at which the cost exceeds the maximum potential benefit for the regulator, i.e.

$$\kappa(\overline{m}) \leq B < \kappa(\overline{m} + 1).$$

If $\kappa(m) \leq B$ for all $m > \underline{m}$, then we denote $\overline{m} = \infty$. A direct consequence is that the regulator never uses an enforcement strategy $m > \overline{m}$. Next, we characterize the regulator's best response to some transmission strategy l chosen by agents.

Lemma 2. *Consider the agents' strategy of using a path of length l. The best response of the regulator is given as follows.*

- *Use intensity* $m = l$*, if* $l \leq \overline{m}$*.*
- *Use intensity* $m = \underline{m}$*, if* $l > \overline{m}$*.*

Lemma 2 states that the regulator's best response is to choose enforcement intensity m equal to l if m is feasible (i.e. $l \leq \overline{m}$) or an intensity $\underline{m}$. Just as the agents prefer to transmit information through the shortest path possible that exceeds length m, the regulator prefers to extract the greatest payoff from choosing the lowest enforcement intensity that matches agents' transmission length. In other words, as long as the cost of search is not too large, the regulator opts to choose sufficiently high intensity to ensure punishing agents.

Together, Lemmas 1 and 2 offer a formal characterization of the cat and mouse game between agents and the regulator. For example, suppose that s and r have access to paths of length $l_1 < l_2 < ... < l_k$ where $l_1 > \underline{m}$ and $l_k \geq \overline{m} + 1$. If the regulator plays $\underline{m}$, agents want to play l_1. Then, the regulator wants to play l_1. In response, agents want to play l_2.... In the end, if the agents play l_k, the regulator finds it too costly to catch the agents, and plays $\underline{m}$. The best response cycle ensures that the equilibrium must be in mixed strategies.

Theorem 1 (Regulatory Ambiguity)**.** *Suppose that* (s, r) *has at least one path longer than* $\overline{m}$ *and at least one path that has length between* $\underline{m}$ *and* $\overline{m}$*. Then the regulator plays a mixed strategy in equilibrium.*

The above theorem formalizes the potential need for the regulator to employ regulatory ambiguity, in the form of a mixed enforcement strategy. Regulatory ambiguity arises when agents, in equilibrium, acquire access to a network that is able to successfully match senders to receivers through multiple paths of differing lengths.

Two things are worth noting. First, as long as agents have access to a transmission path that is sufficiently long, marginally increasing cutoff $\overline{m}$ (e.g. by shifting downward the enforcement cost function κ) does not in general improve the regulator's ability to deter agents from transmitting information. Second, while marginally increasing the penalty (e.g. by shifting upward the penalty function K) of getting caught conditional on transmitting information affects equilibrium mixing strategies, it does not in itself deter sharing of information by agents. Incurring a high enforcement cost is only justified conditional on detecting transmission or deterrence, but when agents can anticipate high regulatory oversight, senders transmit information on a path that allows circumvention with probability 1. At the same time, low regulatory oversight is justified if no transmission is expected or detection is not possible otherwise. In this case, however, agents send and receive information along a minimum path as shown in Lemma 1, which could be detected with high oversight. As a consequence, the regulator must employ a mixed strategy with respect to the enforcement intensity in equilibrium.

3.3 Legal Boundaries

It is worthwhile highlighting how our setting also offers a foundation for laws that may be deliberately set broadly so as to avoid gaming by agents. In our main setting, the regulator is able to punish agents as long as he can map the transmission path between the sender and the receiver. We relax this. Instead, suppose prior to information transmission, lawmakers select a boundary strategy b, which determines the maximum path length between a sender and receiver that constitutes illegal insider information if used for financial gains. For instance, if $l \leq b$, transmission may be regarded as a deliberate transfer of information intended for illegal profits; if $l > b$, the communication between the sender and receiver may be deemed too distant to constitute illegal activity.

Correspondingly, suppose that any given b is associated with a cost $\beta(b)$, where $\beta(b)$ is a strictly increasing function associated with the social cost of violating of investors' civil liberties and privacy. This reflects the idea that the legal boundary b confines the regulator's ability to explore whether illegal insider trading occurred. For example, a regulator may require authorization from a judge to search, confiscate, and analyze evidence. The scope of any particular investigation would then be limited to the legal boundary b. For simplicity, we suppose that $\kappa(m) = 0$ for any m, but the set of feasible m is bounded above by b. Accordingly, the regulator would set m to equal b.

It is straightforward to see how the arguments underpinning Theorem 1 may carry forward. As long as agents have sufficiently complex networks that facilitate long transmission paths, lawmakers' equilibrium boundary strategy must be a mixed strategy over a set $[\underline{b}, \overline{b}]$, for some thresholds $\underline{b}, \overline{b}$, where $\beta(\underline{b}) = 0$, and $\beta(\overline{b}) \leq B < \beta(\overline{b} + 1)$.

Given this interpretation, Theorem 1 rationalizes a common strategy implemented and advocated by regulators to maintain vagueness in what constitutes illegal insider trading activity. For instance, legal boundaries of insider trading in the US are ambiguous and often criticized for being unclear. As a consequence, insider trading prosecution cases ultimately depend on courts to determine whether the nature of the shared information is in fact insider information, i.e. material and non-public, and whether the transfer of information is illegal, e.g. a violation of fiduciary duty. This flexibility in what constitutes illegal insider information is often argued by enforcement officers of the SEC as what allows for successful prosecution and even deterrence. A quote by Arthur Levitt, former chairman of the SEC, captures this sentiment:[16]

> If the SEC had an option as to whether they wanted to have greater specificity and the Justice Department as well, they'd say 'Absolutely not' because greater specificity would give the legal fraternity various ways of getting around those specifics. They want these laws purposely vague to see to it they have the maximum leverage in terms of bringing cases.

4 The Formation and Regulation of Insider Networks

In the previous section, we showed that when agents are endowed with networks which provided access to both short and long paths of information transmission, the potential for gaming regulation pushes the regulator to employ regulatory ambiguity. We also took as given that a fixed pair of one sender and one receiver obtains an opportunity to exploit information.

Now, we assume that a random pair obtains such an opportunity, and that prior to the pair being realized, agents must form networks in advance to enable transmission of the information.[17] We will show that agents find it feasible and profitable to form a network that satisfies the conditions in Theorem 1, and further describe the equilibrium network structure.

4.1 Model with Endogenous Network Formation

Agents. Suppose that $A \subset \mathbb{R}$ and $\lambda(A) = \mu$, where λ denotes the Lebesgue measure. A fixed set $S \subset A$ are potential senders with $\lambda(S) = \mu_S$, and a fixed set $R \subset A$ are potential

[16]For a detailed discussion on SEC's approach of regulatory ambiguity, see http://knowledge.wharton.upenn.edu/article/insider-trading-2011-how-technology-and-social-networks-have-friended-access-to-confidential-information/

[17]This is consistent with the interpretation that the material insider information stochastically arrives, and relationships through which information is transmitted must be formed prior in order to take advantage in a timely manner.

receivers with $\lambda(R) = \mu_R$, where $S \cup R = A$ and $S \cap R = \emptyset$.[18] A random pair of a sender and a receiver $(s, r) \in S \times R$ obtain an opportunity to exploit information. Since the pair that is to receive information is not deterministic, agents must form networks in advance to enable transmission of the information. A model of a continuum of agents is used for tractability. Appendix B shows that equilibria in the continuum model are limits of equilibria in a discrete counterpart.

Network. Agents need to form directed bilateral links over which they can transmit information. For $E \subset A^2$, (A, E) is called a network if E is Borel in $\mathbb{R}^2$.[19] A potential network E is illustrated in Figure 4. Agents first collectively forge directed links to form a network through which information can be transmitted. Denote $\mathscr{P}(E, i, j)$ the set of paths from i to j in E.

Figure 4: Illustration of a network over a continuum

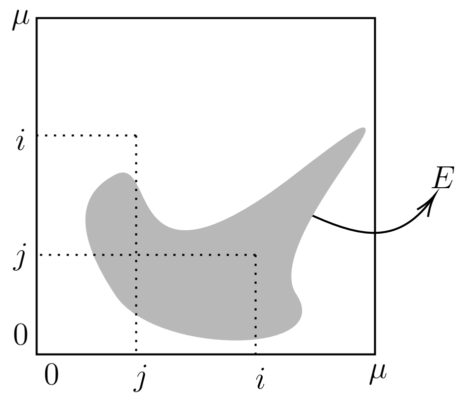


Figure 4 illustrates a network over a continuum of agents. The x-axis and y-axis both represent the set of agents. The shaded area represents the set of edges E. For example, since $(i, j) \in E$, there is a link from i to j. Since $(j, i) \notin E$, there is no link from j to i.

Network Formation. The cost structure for forming the network is given as follows. For any $s, s' \in S$, a directed link from s to s' costs η and 0, respectively. Symmetrically, for any $r, r' \in R$, a link from r to r' costs 0 and η, respectively. In this way, a sender (receiver) who wants to send (receive) incurs the cost associated with forming the link. For any link

[18]Our results remain intact even if there are also neutral agents in $A \setminus (S \cup R)$ that have neither information nor liquidity.

[19]By analyzing a network over a continuum of agents, our model is significantly more tractable relative to a discrete network counterpart. Our model should be viewed as an approximation of a discrete network with a large population of agents.

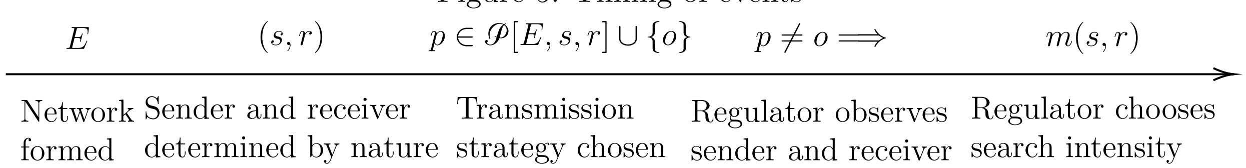
between a sender and a receiver, both incur a cost η. This reflects that agents may find it more difficult to search for and form a link with those of a different type. The resulting cost to agent i is denoted $c(i, E)$. Let ν be used to denote the counting measure. Then $c(s, E) = \eta(\nu(N^+_{s,E} \cap S) + \nu(N_{s,E} \cap R))$ for a sender s and $c(r, E) = \eta(\nu(N^-_{r,E} \cap R) + \nu(N_{r,E} \cap S))$ for a receiver r. The total cost of the network is given by

$$C(E) = \int_A c(\cdot, E)\, d\lambda.$$

We maintain our assumption that agents coordinate to maximize their expected total payoffs.[20]

Payoffs. Denote $\mathcal{E}$ the set of measurable subsets of A^2. After forming network $E \in \mathcal{E}$, a sender $s \in S$ obtains information and a receiver $r \in R$ obtains the means to trade, both determined independently and uniformly at random. Denote $\mathcal{P}(E, s, r)$ the set of paths from s to r in E. If agents choose $p = o$, the game ends. If agents choose $p \in \mathcal{P}(E, s, r)$, then the regulator observes s and r, and chooses a search intensity m, which depends on the identities of s and r. Note that the regulator neither observes the network nor the transmission path. The timing is summarized in Figure 5.

Figure 5: Timing of events



We focus on pure strategies for the network formation.[21] A strategy for agents is (E, σ_A) where $E \in \mathcal{E}$ is the network and $\sigma_A[E', s', r'] \in \Delta(\mathcal{P}(E', s', r') \cup \{o\})$ is the mixed transmission strategy for all networks and pairs $(E', s', r') \in \mathcal{E} \times S \times R$. A strategy for the regulator is σ_R where $\sigma_R[s', r'] \in \Delta(\mathcal{M})$ is the mixed search strategy for all pairs $(s', r') \in S \times R$. A belief for the regulator is $\left(\tilde{E}, \tilde{\sigma}_A\right)$ such that, upon observing a pair $(s', r') \in S \times R$, the regulator believes that the network is $\tilde{E}[s', r']$, and the transmission path used is $\tilde{\sigma}_A[s', r'] \in \Delta\left(\mathcal{P}\left(\tilde{E}[s', r'], s', r'\right)\right)$.

[20] By abstracting from the division of surplus between agents, this assumption allows us to focus on the conflict between the regulator and the agents. The formation of the efficient network can be implemented using an egalitarian sharing rule à la Jackson (2005).

[21] We do not allow for mixed strategies at the network formation stage for simplicity. Mixtures over pure strategy networks also constitute equilibria.

Given his strategy and beliefs, conditional on a realized pair (s,r), the interim expected payoff of the regulator is

$$U_R\left(\sigma_R[s,r];\left(\tilde{E}[s,r],\tilde{\sigma}_A[s,r]\right)\right)=\sum_{p\in\mathcal{P}\left(\tilde{E}[s,r],s,r\right)}\sum_m \tilde{\sigma}_A[s,r](p)\,\sigma_R[s,r](m)\,v_R(m;p)\,.$$

Given the strategy of the regulator, the interim and ex-ante payoffs of the agents are

$$U_A\left((E,\sigma_A[E,s,r]);\sigma_R[s,r]\right)=\sum_{p\in\mathcal{P}(E,s,r)}\sum_m \sigma_A[E,s,r](p)\,\sigma_R[s,r](m)\,v_A(p;m)$$

$$U_A\left((E,\sigma_A);\sigma_R\right)=-C(E)+\mathbb{E}_{s,r}\left[U_A\left((E,\sigma_A[E,s,r]);\sigma_R[s,r]\right)\right].$$

Equilibrium. An (Perfect Bayesian) equilibrium consists of strategies $(E^*,\sigma_A^*,\sigma_R^*)$ and beliefs of the regulator $(\tilde{E}^*,\tilde{\sigma}_A^*)$ such that

- **(Sequential rationality**) Given his beliefs, the regulator's search strategy is optimal: For all $(s,r)\in S\times R$ and $m\geq \underline{m}$,
$$U_R\left(\sigma_R^*[s,r];\left(\tilde{E}^*[s,r],\tilde{\sigma}_A^*[s,r]\right)\right)\geq U_R\left(m;\left(\tilde{E}^*[s,r],\tilde{\sigma}_A^*[s,r]\right)\right)$$
- **(Consistency)** The beliefs of the regulator are derived from Bayesian updating whenever possible: For all $(s,r)\in S\times R$ such that $\sigma_A^*(E^*,s,r)\neq o$, $\tilde{E}[s,r]=E^*$ and for all $p\in\mathcal{P}(E^*,s,r)$,
$$\tilde{\sigma}_A^*[s,r](p)=\frac{\sigma_A^*[E^*,s,r](p)}{1-\sigma_A^*[E^*,s,r](o)}.$$
- **(Optimal transmission)** For any network and any realized pair, agents' transmission strategies maximize the expected payoff given the equilibrium strategy of the regulator: For all $(E,s,r)\in\mathcal{E}\times S\times R$ and for all $p\in\mathcal{P}(E,s,r)$,
$$U_A\left((E,\sigma_A^*[E,s,r]);\sigma_R^*[s,r]\right)\geq max\left\{0,U_A\left((E,p);\sigma_R^*[s,r]\right)\right\}.$$
- **(Optimal network)** E^* maximizes expected payoff given the continuation play: For all $E\in\mathcal{E}$,
$$U_A\left((E^*,\sigma_A^*);\sigma_R^*\right)\geq U_A\left((E,\sigma_A^*);\sigma_R^*\right)$$

Our equilibrium notion is a refinement of Perfect Bayesian equilibrium by allowing for joint deviations on the side of the agents. This allows us to study strongly stable networks in a Perfect Bayesian equilibrium. The other restriction we impose on the Perfect Bayesian equilibrium is that the choice of the network must be a pure strategy. This, along with Bayesian updating, implies that the regulator correctly infers the network E^* in equilibrium on the path of play.

Regarding beliefs about transmission, on-the-path, these beliefs are accurately given by the strategy of the agents. Notice that we do not impose any restriction on off-the-path beliefs of the regulator regarding the transmission strategies. If the regulator observes a pair (s, r) such that $\sigma_A^* (E^*, s, r) = o$, he would correctly identify a deviation and be free to assign any belief regarding the network and path employed for transmission. In this case, the regulator can play any strategy that is not weakly dominated with the appropriate off-the-path belief.

4.2 Regulation and Formation of Insider Networks

We begin by noting several useful observations that help determine the properties of the equilibrium network. Recall, the realization of the pair of sender and receiver occurs after the formation of the network. If the network were costless to form, agents would want a.e. sender-receiver pair to have a path between them. We call such a network bi-connected.

Definition 1. A network is said to be *bi-connected* if for a.e. sender-receiver pair in the network, there is a path from the sender to the receiver.

Such a desirable network is not free to form, however. For example, if there were a link every sender to every receiver, the cost of the network would be ∞. The smallest cost possible for a bi-connected network is given as follows.

Lemma 3. *Any bi-connected network costs at least $\mu\eta$.*

Agents, in addition to forming a bi-connected network at a small cost, must also anticipate regulation. Given the strategic use of the network to facilitate information transmission, an important property of a network is the extent to which it can accommodate various paths of transmission, which endogenously determine the transmission strategy set of agents. Since $m = \underline{m}$ is costless, the regulator always chooses $m \geq \underline{m}$; as such, agents do not use paths shorter than $\underline{m} + 1$. Any $m \geq \overline{m} + 1$ is strictly dominated for the regulator, and so agents would not use path lengths longer than $\overline{m} + 1$.

Definition 2. A network is said to be *flexible* if a.e. sender-receiver pair in the network is connected with at least one path of length $\underline{m} + 1, \underline{m} + 2, ..., \overline{m} + 1$.[22]

[22]For $\overline{m} = \infty$, this means that all path lengths $l \geq \underline{m} + 1$ are available to a.e. sender-receiver pair.

The equilibrium strategies regarding information transmission depend on what lengths of paths are available to agents. An important consideration is whether agents find it profitable to form a network that enables transmissions that extend beyond the scope of regulation. Flexible networks are by definition bi-connected, which implies that the smallest possible cost for a flexible network is $\eta\mu$. We first show that there does in fact exist a network that costs $\eta\mu$ and allows for arbitrary levels of flexibility.

Lemma 4. *Take any set $L \subset \mathbb{N}$ of path lengths that are larger than $\underline{m}$. There exists a network with cost $\mu\eta$ such that for all $l \in L$ and for almost every sender receiver pair (s, r), there is a path from s to r that has length equal to l.*

Lemma 4 establishes that agents can in principle form complex networks that facilitate arbitrarily long chains between any given sender and receiver at a cost $\eta\mu$. An implication is that there exists a flexible network that costs $\eta\mu$. We show that such networks characterize all equilibrium networks:

Proposition 1. *For $\overline{m} < \infty$, equilibrium networks are given by flexible networks that cost $\mu\eta$. Furthermore, the regulator mixes totally between search intensities m such that $\underline{m} \leq m \leq \overline{m}$. The agents mix totally between path lengths l such that $\overline{m} + 1 \leq l \leq \overline{m} + 1$. For $\overline{m} = \infty$, equilibrium networks are empty networks.*[23]

What network topologies are both flexible and cost $\mu\eta$? For the remainder of the paper, we focus on $\overline{m} < \infty$.

Definition 3. A network E is a *core-periphery* if there exists a set of agents C, called the *core*, and a set of agents P, called the *periphery*, such that

- Almost all agents are in the periphery: $\lambda(P) = \mu$.
- Core and periphery are disjoint: $P \cap C = \emptyset$.
- C is at most countable and weakly connected.[24]
- Each sender in the periphery has exactly one link, which is to a sender in the core. Formally, $\forall s \in P \cap S$, $\exists s' \in C \cap S \, s.t. \, N^{+}_{s,E} = \{s'\}$ and $N^{-}_{s,E} = \emptyset$.
- Each receiver in the periphery has exactly one link, which is from a receiver in the core. Formally, $\forall r \in P \cap R$, $\exists r' \in C \cap R \, s.t. \, N^{-}_{r,E} = \{r'\}$ and $N^{+}_{r,E} = \emptyset$.

Figure 6: A core-periphery network

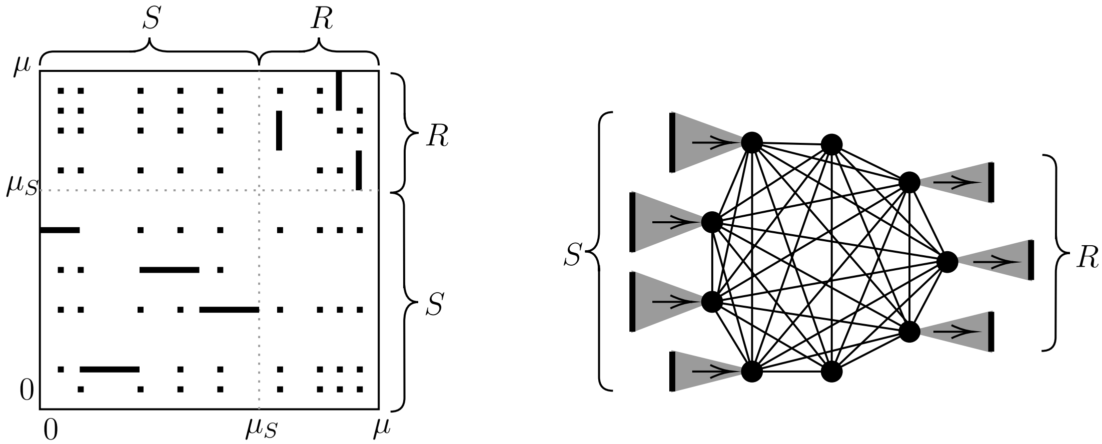


Figure 6 illustrates two alternative representations of the a core-periphery network. On the left panel, we show a representation of a core-periphery as a subset of $\mathbb{R}^2$, in line with the representation in Figure 5. On the right panel, we have a more standard visual representation that illustrates the senders on the left, receivers on the right, and the core in the middle, along with links between them.

Figure 6 provides visual examples of a core-periphery network. It is straightforward to show the following:

Lemma 5. *Core-periphery networks cost $\mu\eta$.*

By Lemma 5 and Proposition 1, if there exists a core-periphery network that is flexible, then it is an equilibrium network. Intuitively, a core-periphery structure has features conducive to flexibility. By forming a network in which intermediate edges in the core are shared, agents can collectively minimize the cost of links required to form a flexible network. Then, a core-periphery network can achieve flexibility simply by making its core "flexible." Then, it suffices for each peripheral agent to attach herself to the core with a single link. We formally confirm this in the below theorem:

Theorem 2. *For any $k \geq \overline{m} + 1$, there exists a flexible core-periphery with k agents in the core. Furthermore, all such networks are equilibrium networks.*

A core-periphery structure decouples the role of the network to facilitate various transmission paths to a select core members who form a dense subnetwork, with the need for

[23]Empty network in this context means a.e. agent has 0 links with respect to counting measure ν.

[24]Weakly connected means that between every pair of nodes, there is an undirected path. It is different from strong connectivity which requires there to be a directed path between every pair.

connectivity between many senders and receivers, by having the entire mass μ of periphery members include at least one core member within their neighborhood.

5 Emergence of Information Intermediaries

As evident by the equilibrium network structure, a.e. sender in the periphery relies on the core to relay information to the receiver using a desired path length. An important characterization to consider is the extent to which agents endogenously act as intermediaries in order to facilitate transmission. We offer a formal definition of an intermediary for a given equilibrium network:

Definition 4. An agent is said to be an intermediary if he transmits information with positive probability.

While passing information on behalf of other agents is an act of intermediation, in order for an agent to be considered an intermediary, he must act as a middleman for a positive mass of other agents in the network. This requires that an agent must provide intermediation for other agents beyond "scratching each other's back." This naturally arises in our setting in which middlemen are necessary to avoid punishment. Given the characterization of intermediaries outlined in Definition 4, this implies that (a subset of) the core constitutes intermediaries:

Proposition 2. *For a flexible core-periphery network, all intermediaries are in the core. If non-intermediaries are taken out of the core along with their links, the resulting network is still a flexible core-periphery network, and hence an equilibrium network.*

An implication of the model is the emergence of intermediaries within a setting in which agent are ex-ante identical (up to a sender, receiver type). Specifically, the core structure results in the endogenous rise of information intermediaries, as defined under Definition 4. A small set of agents specialize in providing flexible channels of information transmission and ultimately intermediate a.e. information transmission between the set of potential senders and receivers in the model.

The rise of intermediaries. Interestingly, the emergence of information intermediaries draws striking parallel to the rise of consultancy firms that have played an outsized role in recent years. While information intermediaries are not and should not be viewed as an illegal entity, these consultancy firms have been implicated in a number of insider trading cases in the past decade. A large fraction of these firms is commonly referred to as expert network firms, which specialize in connecting clients to experts in various fields ranging from technology, medicine, healthcare, energy, and even economics.[25] What triggers the rise of such intermediaries?

We shed light on this question by extending our setting to allow for preexisting social networks that can be used for information transmission. As we will show, the key takeaway is that as regulation gets tighter, agents migrate their information transmission activities from social networks to complex insider networks, as regulation strains the flexibility offered by existing networks.

Suppose that the agents are endowed with pre-existing social networks described as follows. There are $\mu_S = \mu/2$ mass of senders and $\mu_R = \mu/2$ mass of receivers. Senders $s_i \in S$ are indexed by $i \in [0, \mu/2]$ and receivers $r_i \in R$ are also indexed by $i \in [0, \mu/2]$. Each pair (s_i, r_i) has access to an exclusive social network with which s_i can transmit to r_i using paths ranging from lengths 1 to m_i. These social networks provide an alternative to forming costly insider networks.

As before, one sender-receiver pair is realized randomly, say $(s, r) = (s_i, r_j)$.[26] If sender s_i transmits the information to the receiver r_j via a path of length l, this generates $\Pi(l)$ profit. As a departure from the baseline model, all receivers have some means to profit from sender s's information. Accordingly, sender s_i can also transmit the information via her social network to the receiver r_i. This generates $\chi\Pi(l)$ profit if the path length used is l, where $\chi < 1$. This reflects that a generic receiver is typically not able to trade based on the information to the maximum potential.[27]

This extension presents a clear tradeoff to agents: by relying on existing social networks, agents can transmit information "locally." Alternatively, they can forge new, costly links to form an insider network. The insider network not only acts a tool to cope with regulation, but potentially enables agents to transmit information to receivers who can generate the

[25]https://www.bloomberg.com/news/articles/2018-02-28/investors-are-paying-1-300-per-hour-for-expert-chats

[26]Note that $i = j$ has probability 0.

[27]For instance, r may have information regarding the underlying asset or market conditions that complements the use of s's insider information. Alternatively, differences in profitability could be driven by heterogeneity with respect to liquidity available to execute trades in a timely manner.

greatest profits.

For a given $\overline{m}$, if $m_i \leq \overline{m}$, the social network of the pair i does not provide them sufficiently long transmission paths. The social network does not constitute a lucrative outside option for such pairs. Then it is in the interest of agents to include such pairs to an insider network. It is not, however, necessarily optimal to form an insider network with all such agents. The mass of pairs whose social networks are not competitive with the regulation needs to be large enough to make it worth forming an insider network. Denote $\mu_{S,\overline{m}} = |\{s_i \in S : m_i \leq \overline{m}\}|$ the mass of pairs that can not make use of their social network due to stringent regulation. Note that $\mu_{S,\overline{m}}$ is increasing in $\overline{m}$. Denote S' and R' the senders and receivers that form links and join an equilibrium insider network (flexible core-periphery among S' and R'). Denote $\mu'_S = \lambda(S')$ and $\mu'_R = \lambda(R')$. Finally, denote $\delta = \frac{\eta\mu}{\Pi(\overline{m}+1)} < 1$ and $\bar{\mu}_S = \left(1 - \frac{1-\delta}{\chi}\right)\mu_S$.

Theorem 3. *Suppose that* $\chi < \frac{\delta}{2}$.[28] *If* $\mu_{S,\overline{m}} < \bar{\mu}_S$ *(i.e. small* $\overline{m}$*) no one joins the insider network and the equilibrium has* $\mu'_S = 0, \mu'_R = 0$. *Social networks are exclusively used for information transmission. If* $\mu_{S,\overline{m}} > \bar{\mu}_S$ *(i.e. high* $\overline{m}$*), everyone joins the insider network and the equilibrium has* $\mu'_S = \mu_S, \mu'_R = \mu_R$. *Insider networks are exclusively used for information transmission.*

By Theorem 3, if regulation renders sufficiently many agents' social networks ineffective, captured by $\mu_{S,\overline{m}}$, agents collectively switch from using their external networks to forming a core-periphery insider network. This gives rise to information intermediaries.

Recall that by Section 3.3, $\overline{m}$ can be interpreted as a legal boundary for any given regulatory framework. Accordingly, Theorem 3 forms the basis for the potential link between tightening regulation and the rise of information intermediaries. Major shifts in the regulatory framework in the early 2000s developed through Regulation Fair Disclosure (Reg FD), which was promoted by the SEC in 2000, and the Global Analyst Research Settlements, which was an enforcement agreement reached between the SEC, other regulatory agencies, and the ten largest investment firms in the US. Together, regulation focused on tightening governance on information disclosure by public companies, and imposing controls on the leakage of material non-public information through financial intermediaries, such as research

[28]Under $\chi < \frac{\delta}{2}$, the insider network enhances total expected profits not only by providing senders with increased flexibility relative to their social networks, but also by improving the allocation of information.

analysts and broker-dealers. What followed was dramatic growth in the expert network industry.

This relation between tighter regulatory control and the rise of information intermediation is also observed in the official sector. In 2012, the US Congress passed the Stock Trading on Congressional Knowledge Act (STOCK Act). The general intent of the law was to prevent government officials and employees from exploiting privileged access to non-public information that could potentially be used for financial gains. Following the passage of the STOCK Act, information intermediaries emerged in the form of political intelligence firms, which specialize in connecting clients to experts in areas of policy, law, and regulation.

Intermediation in a Changing Regulatory Environment. Theorem 3 highlights how tightening regulation can prompt the formation of insider networks in which bulk of information is intermediated by a small core. Once an insider network is established, how does it adjust to further tightening of the regulatory environment? Our model presents one more stark insight. As a corollary of Theorem 2, the core of the network can adapt to rising regulation by adding more intermediaries to the core at negligible (zero) cost. New members of the core do not need to be linked to any periphery agents – rather, their primary function is to increase the flexibility of the core, and hence of the whole network. In other words, the flexibility of the existing insider network can be increased arbitrarily at zero cost, and there is no level of regulation that the agents can not adjust to.

6 Identifying Intermediaries with Centrality Measures

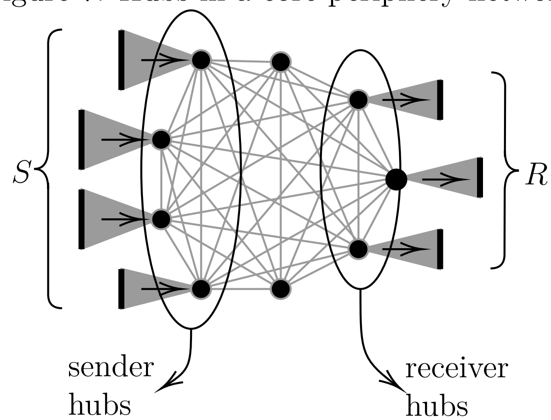
In this section, we explain some properties of a flexible core-periphery networks and introduce a network centrality measures that identifies intermediaries. We start by defining "hubs."

Definition 5. Given a core-periphery E with core C and periphery P, $s \in C \cap S$ is a *sender-hub* if $\lambda \left(N_{s,E}\right) > 0$ and $r \in C \cap R$ is a *receiver-hub* if $\lambda \left(N_{r,E}\right) > 0$.

By the definition of the core-periphery, a sender-hub's a.e. neighbor is a sender and a receiver hubs a.e. neighbor is a receiver. Moreover, a.e. sender is the neighbor of a sender-hub and a.e. receiver is a neighbor of a receiver-hub. Furthermore, neighbors of hubs can not have positive measure intersection. Hubs are illustrated in Figure 7.

Hubs are the agents in the core that connect the periphery to the core. There can be other nodes in the core in order to provide flexibility, but all transmission must go through

Figure 7: Hubs in a core-periphery network



hubs to enter the core. What else can be said about the hubs? An important observation is that hubs have positive betweenness centrality (defined appropriately for the continuum below). Moreover, a.e. (directed) geodesic[29] must go through hubs. In order to formalize this, denote $g(s,r)$ the number of geodesics from sender s to receiver r. Denote $g_i(s,r)$ the number of geodesics from s to r that go through i. Then between centrality of agent i as

$$\int_S \int_R \frac{g_i(s,r)}{g(s,r)} d\lambda(r)\, d\lambda(s).$$

Proposition 3. *Betweenness centrality of a sender-hub s is at least $\lambda(N_{s,E})\,\mu_R$ and that of a receiver-hub r is at least $\lambda(N_{r,E})\,\mu_S$. Periphery agents have 0 betweenness centrality.*

This illuminates the fact that hubs are bottle-necks for information transmission. Notice that hubs transmits information with positive probability on behalf of pairs of senders and receivers, and so they are intermediaries with respect to Definition 4. But hubs are not the only intermediaries. Hubs have positive betweenness centrality but there is no direct link between being an intermediary and having positive betweenness centrality. Being an actor of information transmission requires that the agent is, for a positive measure of sender-receiver pairs, is on at least one path of lengths from $\underline{m}+1$ to $\overline{m}+1$ from the sender to the receiver. We define flexible-betweenness centrality for this purpose.

Definition 6. Denote $g(s,r,k)$ the number of paths of length k from s to r, and $g_i(s,r,k)$ the number of such paths that go through i. The flexible-betweenness centrality of agent i

[29]A path between two nodes is called a geodesic if it is one the shortest between the two nodes.

is given by

$$\sum_{k=\underline{m}+1}^{\overline{m}+1} \int_S \int_R \frac{g_i(s,r,k)}{g(s,r,k)} d\lambda(r)\, d\lambda(s).$$

Flexible betweenness centrality is a network centric measure that can serve as a proxy for intermediation. This expanded notion of centrality takes into account that intermediaries are positioned on large set of possible transmission paths. An empirical implication is that flexible-betweenness centrality is a more precise measure to identify intermediaries in a network when flexibility is a valuable feature of transmission.

Theorem 4. *For any flexible core-periphery network E, there exists an equilibrium in which the network formed is E, and an agent is an intermediary if and only if it has positive flexible-betweenness centrality.*

7 Conclusion

In this paper, we study a model of endogenous formation of networks over which agents transmit information under regulation. We show how a cat and mouse game arises between a regulator, who sets and enforces a regulatory environment, and agents, who form networks to disseminate insider information. In equilibrium, the regulator implements regulatory ambiguity that induces agents to take greater risks in information transmission. Agents adapt to regulation by forming a flexible network with a core-periphery structure, which endows agents with the option to transmit information through various paths of differing length.

We show how the core represents the endogenous rise of information intermediaries. A small set of agents that form the core of the network intermediate information between potential senders, i.e. insiders, and receivers, i.e. those that seek to exploit information. In an extension, we show that tightening regulation can trigger agents to migrate transmission activity from social networks to an insider network. We draw parallels to the recent growth of the expert network and political intelligence industry following stricter regulation regarding disclosure and insider trading. The surge of information intermediaries suggest that rather than curtailing insider trading, market participants may have adapted by developing alternative and more complex channels through which insider information is shared and exploited.

As a final note, we believe that our setting is applicable to a broader set of problems. In particular, the model can be used to understand network design problems, in which agents want to transmit messages or goods, but must combat a strategic actor (as in our case) or exogenous risks. Many networks involving communication or information sharing require achieving a sufficient level of security and privacy. An efficient network entails safeguarding the anonymity of messages from a malicious attacker while economizing on the cost of building and using the network. The model can be extended to study trading networks, in which agents prefer trading in proximity, but face counterparty risk. In particular, we highlight potential benefits of having a core-periphery structure that allows for intermediaries to flexibly re-direct flow between counterparties. We leave these applications for future research.

References

Acemoglu, Daron, Azarakhsh Malekian, and Asu Ozdaglar, "Network security and contagion," *Journal of Economic Theory*, 2016, *166*, 536–585.

_ , Kostas Bimpikis, and Asuman Ozdaglar, "Dynamics of information exchange in endogenous social networks," *Theoretical Economics*, 2014, *9* (1), 41–97.

Ahern, Kenneth R, "Information networks: Evidence from illegal insider trading tips," *Journal of Financial Economics*, 2017, *125* (1), 26–47.

Bloch, Francis and Bhaskar Dutta, "Communication networks with endogenous link strength," *Games and Economic Behavior*, 2009, *66* (1), 39–56.

Chakraborty, Archishman and Bilge Yılmaz, "Informed manipulation," *Journal of Economic Theory*, 2004, *114* (1), 132–152.

Cohen, Lauren, Andrea Frazzini, and Christopher Malloy, "Sell-side school ties," *The Journal of Finance*, 2010, *65* (4), 1409–1437.

DeMarzo, Peter M, Michael J Fishman, and Kathleen M Hagerty, "The optimal enforcement of insider trading regulations," *Journal of Political Economy*, 1998, *106* (3), 602–632.

Dziubiński, Marcin and Sanjeev Goyal, "How do you defend a network?," *Theoretical Economics*, 2017, *12* (1), 331–376.

Erol, Selman, "Network hazard and bailouts," 2017.

_ **and Guillermo Ordoñez**, "Network reactions to banking regulations," *Journal of Monetary Economics*, 2017, *89*, 51–67.

Galeotti, Andrea and Sanjeev Goyal, "The law of the few," *American Economic Review*, 2010, *100* (4), 1468–92.

Greenwood, Robin, Jeremy C Stein, Samuel G Hanson, and Adi Sunderam, "Strengthening and streamlining bank capital regulation," *Brookings Papers on Economic Activity*, 2017, *2017* (2), 479–565.

Haller, Hans, "Network vulnerability: a designer-disruptor game," 2016.

Hoyer, Britta, "Network disruption and the common enemy effect," *Discussion Paper Series/Tjalling C. Koopmans Research Institute*, 2012, *12* (06).

_ **and Kris De Jaegher**, "Strategic network disruption and defense," *Journal of Public Economic Theory*, 2016, *18* (5), 802–830.

Jackson, Matthew O, "A survey of network formation models: stability and efficiency," *Group Formation in Economics: Networks, Clubs, and Coalitions*, 2005, pp. 11–49.

Jeng, Daniel, "Expert networks and insider trading: An introduction and recommendation," *Review of Banking and Financial Law*, 2013, *32* (2), 245–264.

Krebs, Valdis, "Uncloaking terrorist networks," *First Monday*, 2002, *7* (4).

Lovász, László, *Large networks and graph limits*, Vol. 60, American Mathematical Soc., 2012.

Maggio, Marco Di, Francesco Franzoni, Amir Kermani, and Carlo Sommavilla, "The relevance of broker networks for information diffusion in the stock market," 2017.

Manea, Mihai, "Bottleneck links, essential intermediaries, and competing paths of diffusion in networks," 2018.

Parise, Francesca and Asuman Ozdaglar, "Graphon games," *arXiv preprint arXiv:1802.00080*, 2018.

Silvers, Roger, "The valuation impact of SEC enforcement actions on nontarget foreign firms," *Journal of Accounting Research*, 2016, *54* (1), 187–234.

Tsebelis, George, "The abuse of probability in political analysis: The Robinson Crusoe fallacy," *American Political Science Review*, 1989, *83* (1), 77–91.

Williams, Phil, "Transnational criminal networks," *Networks and netwars: The future of terror, crime, and militancy*, 2001, *1382*, 61.

A Informed Trading in Financial Markets

In this section, we offer a simple extension to the model to allow for insiders to trade directly in financial markets, in effect endogenizing $\Pi(l)$ and $K(l)$.

Insider Information and Transmission. There is a financial market where agents can trade an asset with some fundamental value θ, where θ takes a value 1 or 0 with equal probability. Let the sender s be perfectly informed about θ. Suppose that the sender s transmits information regarding θ to the receiver r along a path of length l. We assume that longer transmission poses a higher risk that s's information becomes public, i.e. nonmaterial, before trading is executed. One interpretation is that information is more likely to leak along the transmission path as information is passed through more agents. Alternatively, transferring information may take time, and a longer path increases the likelihood that transmission does not occur in time for profitable opportunities. Then, the likelihood that transmission provides r with an informational advantage is given by $\rho(l)$, where $\rho(\cdot) \in [0, 1]$ and $\rho(\cdot)$ decreases in l. With $1 - \rho(l)$ probability, then θ becomes common knowledge, i.e. the market-maker becomes informed.

Financial Market. The market is populated with the receiver, a market-maker, and noise traders. If the receiver gains an informational advantage, the receiver makes a market order $x \in \{1, -1\}$. Noise traders' demand is stochastically determined $\tilde{\zeta}$ drawn from uniform distribution $U[-\zeta, \zeta]$, where $\zeta > 1$. The market-maker observes total demand $X = x + \tilde{\zeta}$, where $x = 0$ if the receiver chooses not to trade. The market-maker sets price $P = E[\theta|X]$. Given price P, the payoff from trade for the receiver is given by $x(\theta - P)$. Finally, the regulator is assumed to observe individual order flow, i.e. x. This implies that as long as $x \neq 0$, the regulator initiates search. If the regulator initiates a search and catches agents, a punishment K is imposed.

The solution to the trading game is characterized below:

Proposition 4. *In the trading game equilibrium, the expected profits from transmitting on a path of length l are given by $\Pi(l) = \rho(l) \cdot \frac{\zeta - 1}{2\zeta}$.*

Proof. We conjecture and verify that if if informed makes order $x = 1$ if $\theta = 1$, and $x = -1$ otherwise. Given this, the market-maker observes total order flow $X = 1 + \tilde{\zeta}$ if $\theta = 1$ or $X = -1 + \tilde{\zeta}$ if $\theta = 0$ if transmission successfully occurs. Given this, prices are given by:

$$E[\theta|X] = \begin{cases} 1 & \text{if } X > -1 + \zeta \\ \frac{1}{2} & \text{if } X \in [1 - \zeta, -1 + \zeta) \\ 0 & \text{if } X < 1 - \zeta \end{cases}$$

The profits of the receiver if informed is $\frac{\zeta - 1}{2\zeta}$. It is straightforward to see that deviating is not profitable. Given a transmission over a path of length l, the receiver's expected payoff is given by $\Pi(l) = \rho(l) \cdot \frac{\zeta - 1}{2\zeta}$, and $K(l) = \rho(l) \cdot K$. □

The trading game provides a microfoundation for how insider trading profits are inversely related to the length of the transmission path. In equilibrium, the market-maker is less informed than the receiver with probability $\rho(l)$. This in turn affects the ex-ante expected payoff from transmitting through a path of length l. Imbedding this result in the rest of the model, we can see that the core intuition from the main model follows. The sender and receiver face a tradeoff between transmitting information with higher expected value (i.e. $\Pi(l)$ decreases in l) and the higher likelihood of prosecution (i.e. $Prob(m \geq l)$ decreases in l).

Given the objective of this section, we deliberately chose a simple extension that abstracts from other potentially interesting aspects. For one, there is little scope for strategic interaction between the regulator's decision to initiate a search and the informed receiver to conceal his identity. If after trade, the regulator's beliefs regarding the probability that the receiver traded on insider information were dispersed, the regulator could also choose a search intensity contingent on market prices. With commitment, agents could potentially devise more complicated transmission and trading strategies.[30] We leave this for future work.

[30] For example, uncertainty regarding the existence of an informed trader can substantially increase the scope for market manipulation as in Chakraborty and Yılmaz (2004).

B Continuum as the Limit Equilibrium

In this section, we show that the approximate equilibria of large populations converge to the equilibrium we have identified in our continuum model as the population size grows. The main takeaway is that for a sufficiently large population of agents, the optimal network takes form of a core-periphery structure, which imbues flexibility in its core.

We start with defining the finite model. Take any $n \in \mathbb{N}$. $S(n)$ is a set of $\lfloor \mu_S n \rfloor$ many potential senders and $R(n)$ is a set of $\lfloor \mu_R n \rfloor$ many potential receivers. Denote $A(n) = S(n) \cup R(n)$. The sender and the receiver are drawn uniformly at random. A network $E(n)$ is an element of $A(n)^2$. The costs of forming links are normalized by n to $\frac{\eta}{n}$. Senders incur $\frac{\eta}{n}$ cost for outgoing links and receiver incur $\frac{\eta}{n}$ cost for incoming links. The rest is identical to the original description. Call this model $m(n)$.

The equilibrium in the finite model $m(n)$ is defined similarly to the continuum case in Section 4. For any $\varepsilon > 0$, an ε-equilibrium satisfies sequential rationality, consistency, and optimal transmission. The optimal network condition is replaced by ε-optimal network condition: For all $E \in \mathcal{E}$,

$$U_A((E^*, \sigma_A^*); \sigma_R^*) \geq U_A((E, \sigma_A^*); \sigma_R^*) - \varepsilon.$$

Convergence of finite networks to networks over continuum requires state-of-the art tools. We use the graphon approach (see Parise and Ozdaglar (2018), Lovász (2012)). The idea is to discretize a given continuum equilibrium network into n agents and take n to infinity. Without loss of generality, suppose that $S = [0, \mu_S)$ and $R = [\mu_S, \mu)$. Take any $n \in \mathbb{N}$. Divide S into $\lfloor \mu_S n \rfloor$ many intervals of length $\Delta_S = \frac{\mu_S}{\lfloor \mu_S n \rfloor}$: $S_n = \left\{ s_{1,n}, s_{2,n}, ..., s_{\lfloor \mu_S n \rfloor, n} \right\}$ where $s_{t,n} = [(t-1)\Delta_S, t\Delta_S)$. Elements of S_n represent the senders in the discretized network. Similarly, divide R into $\lfloor \mu_R n \rfloor$ many intervals of length $\Delta_R = \frac{\mu_R}{\lfloor \mu_R n \rfloor}$: $R_n = \left\{ r_{1,n}, r_{2,n}, ..., r_{\lfloor \mu_R n \rfloor, n} \right\}$ where $r_{t,n} = [(t-1)\Delta_R, t\Delta_R)$. Elements of R_n represent the receivers in the discretized network. Denote $A_n = S_n \cup R_n$. Elements of A_n^2 represent the potential links in the discretized network. Given a network E in the continuum, the discrete counterpart of E in a population of size n, denoted $E_n \subset A_n^2$, is defined as follows: $(I, J) \in E_n$ if there exists $i \in I$ and $j \in J$ such that $(i, j) \in E$.

Given any network E in the continuum, (A_n, E_n) defines the corresponding discrete network in a population of size n. Elements of S_n correspond to senders $S(n)$ in $m(n)$ and elements of R_n correspond to receivers $R(n)$ in $m(n)$. This way, (A_n, E_n) lives in $m(n)$.

Theorem 5. *For any flexible core-periphery network E with a finite core and for any $\varepsilon > 0$, there exists $\underline{n}(E, \varepsilon)$ such that for all $n > \underline{n}(E, \varepsilon)$, (A_n, E_n) is an ε-equilibrium network in $m(n)$.*

Proof. (Outline) Call a network in $m(n)$ δ-flexible if at least $n(1-\delta)$ agents have path lengths ranging from $\underline{m}+1$ to $\overline{m}+1$. Denote k the number of agents in the core of E. Since k is finite, for large n, all elements in the core of E fall into disjoint intervals described by elements of $A_n = S_n \cup R_n$. Hence, all links in the core in E exists in A_n. Therefore, A_n in $m(n)$ provides flexibility to all agents outside the core. In other words, A_n is $O\left(\frac{1}{n}\right)$-flexible. Moreover, the cost of (A_n, E_n) is between $\frac{\eta}{n}(\lfloor \mu_S n \rfloor + \lfloor \mu_R n \rfloor)$ and $\frac{\eta}{n}(\lfloor \mu_S n \rfloor + \lfloor \mu_R n \rfloor) + \frac{\eta k^2}{n}$. That is, the cost is $\eta\mu + O\left(\frac{1}{n}\right)$. We will show that, for any ε, for large n, in $m(n)$, $O\left(\frac{1}{n}\right)$-flexible networks that cost $\eta\mu + O\left(\frac{1}{n}\right)$ are ε-equilibria. The proof of this mainly follows that of Proposition 1. The ideas can be summarized as follows. Given the candidate network, which is flexible, the regulator mixes totally between $\underline{m}$ to $\overline{m}$ and a realized pair that has the flexibility mixes totally between $\underline{m}+1$ and $\overline{m}+1$. This generates payoff $\Pi(\overline{m}+1)$ for the agents. Then a (unobserved) deviation from this network, upon the realization of the pair, will be met with the same strategy of the regulator. Then any transmission strategy, conditional on the pair, generates at most $\Pi(\overline{m}+1)$ payoff. So agents can improve their payoffs by at most $O\left(\frac{1}{n}\right)$ by potentially providing paths to pairs that do not have the flexibility (which must be in the core). Then the only potential profitable deviations are those that reduce the cost of links by reducing the number of pairs that have flexibility. Following equation (3), payoffs can not be improved this way. Then the only possible profitable deviation involves reducing the number of links inside the core, which gives a benefit of at most $O\left(\frac{1}{n}\right)$. So, there are no deviations that increase the payoffs by more than $O\left(\frac{1}{n}\right)$. □

C Proofs

Proof. **(Lemma 1), (Lemma 2)**, **(Theorem 1)** QED. □

Proof. **(Lemma 3)** In a flexible network, a.e. sender has an outgoing link and a.e. receiver has an incoming link. Then, for a.e. agent i, $c(i, E) \geq \eta$. Then $C(E) \geq \eta\mu$. □

Proof. **(Lemma 4)** Let $k = sup\, L$. Note that $k \geq 3$. Take $k-1$ agents from A, say C, with at least one sender and at least one receiver. Form a complete network among C. Partition

C into C_s, senders, and C_r, receivers. Connect each sender in $A \backslash C$ to exactly one agent in C_s, connect each receiver in $A \backslash C$ to exactly one receiver in C_r. In the network formed, $c(i, E) = \eta$ for all $i \in A \backslash C$. C has zero measure and $A \backslash C$ has μ measure, and so $C(E) = \mu\eta$. Notice that every pair in C have paths of length $1, 2, ..., k-2$ between them since C is a complete subnetwork. Then all agents in $A \backslash C$ have path lengths $3, 4, ..., k$. $\square$

Proof. **(Proposition 1)** Take any equilibrium: say E^* for the network, $\sigma_A^*(E^*, s, r)$ for transmission strategy upon the realization of s and r, $\sigma_R^*(s, r)$ for the search intensity upon observing s and r. With slight abuse of notation, denote $\sigma_A^*(E^*, s, r)[l]$ the total probability that $\sigma_A^*(E^*, s, r)$ puts on paths of length l.

Regulator's strategy:

Step 1: (For finite $\overline{m}$) Note that for the regulator, $m > \overline{m}$ are strictly dominated. So, $\sigma_R^*(s, r)[m] = 0$ for $m > \overline{m}$.

Step 2: Consider a pair (s, r). Suppose that there is some $\tilde{l} \geq \underline{m} + 1$ such that $\sigma_A^*(E^*, s, r)\left[\tilde{l}\right] = 0$. Then for the regulator, playing $m = \tilde{l} - 1$ has the same benefit with playing $m = \tilde{l}$ in terms of catching s and r, whereas the playing $m = \tilde{l}$ is strictly more costly in terms of search costs as κ is strictly increasing. Therefore, the regulator does not put positive probability on $m = \tilde{l}$. That is, if $\tilde{l} \geq \underline{m} + 1$ and $\sigma_A^*(E^*, s, r)\left[\tilde{l}\right] = 0$ then $\sigma_R^*(s, r)\left[\tilde{l}\right] = 0$.

Step 3: By Step 2, if (s, r) does not transmit information, $\sigma_R^*(s, r)[\underline{m}] = 1$.

Agents' strategy:

Step 4: Consider any pair (s, r). Regulator plays $m \geq \underline{m}$. Using a path of length $l \leq \underline{m}$ is strictly dominated by the strategy of not sending information. So, $\sigma_A^*(E^*, s, r)[l] = 0$ for all $l \leq \underline{m}$.

Step 5: (For finite $\overline{m}$) Consider any pair (s, r). Suppose that there are at least two paths in $\mathcal{P}(E^*, s, r)$ with path lengths larger than $\overline{m}$, say $l' > l > \overline{m}$. By Step 1, $m > \overline{m}$ is strictly dominated. Then using a path with length l' is iteratively strictly dominated by using one with length l. Therefore, (s, r) puts 0 probability on all paths in $\mathcal{P}(E^*, s, r)$ longer than $\overline{m}$ except for the shortest such paths.

Equilibrium and deviations:

Step 6: Denote μ'_S the mass of senders S' and μ'_R the mass of receivers R' that have links in E^*. Denote $\mu' = \mu'_S + \mu'_R$. Note that the cost of forming links that has been incurred is at least $\mu'\eta$ since each agent in $A' = S' \cup R'$ have at least one link.

Pick a flexible core-periphery network among A' and denote it E'. Note that the cost of E' is $\mu'\eta$ which less than or equal to the cost of E^*. That is, E' is not more expensive than than E^*.

Step 7: Consider all pairs $(s,r) \in S' \times R'$ such that that there is some $\tilde{l}(s,r) \in \mathbb{N}$ such that $\sigma_A^*(E^*,s,r)\left[\tilde{l}(s,r)\right] = 0 < \sigma_A^*(E^*,s,r)\left[\tilde{l}(s,r)+1\right]$. Denote this set of pairs X.

By Step 2, for all $(s,r) \in X$, $\sigma_R^*(E^*,s,r)\left[\tilde{l}(s,r)\right] = 0$.

Consider all pairs $(s,r) \in S' \times R'$ that do not transmit information with probability 1. Denote this set Y. By Step 3, for all $(s,r) \in Y$, $\sigma_R^*(s,r)[\underline{m}] = 1$.

Suppose that $\lambda(X \cup Y) > 0$. Then consider the following deviation. Agents deviate to network E'.

Each $(s,r) \notin X$, for each $l \in \mathbb{N}$, (s,r) puts probability $\sigma_A^*(E^*,s,r)[l]$ on a path in $\mathscr{P}(E',s,r)$ that has length $min\{l, \overline{m}+1\}$.

For each $(s,r) \in X$, for each $l \notin \left\{\tilde{l}(s,r), \tilde{l}(s,r)+1\right\}$, (s,r) puts probability $\sigma_A^*(E^*,s,r)[l]$ on a path in $\mathscr{P}(E',s,r)$ that has length $min\{l, \overline{m}+1\}$.

For each $(s,r) \in X$, (s,r) puts $\sigma_A^*(E^*,s,r)\left[\tilde{l}(s,r)\right] + \sigma_A^*(E^*,s,r)\left[\tilde{l}(s,r)+1\right]$ on a path in $\mathscr{P}(E',s,r)$ that has length $\tilde{l}$, and 0 probability on paths in $\mathscr{P}(E',s,r)$ that have length $\tilde{l}(s,r)+1$.

For each $(s,r) \in Y$, (s,r) transmits information using a path of length $\underline{m}+1$.

For pairs in X, this deviation does not change the expected cost of getting caught for agents, but it increases the benefits from transmission because a positive measures of pairs use shorter paths with positive probability.

For pairs in Y, by Step 3, there is no cost of getting caught so the benefit from transmission is improved.

Therefore, since $\lambda(X \cup Y) > 0$, expected payoff of agents is improved whereas the network is not more expensive. Contradiction. Hence $\lambda(X) = \lambda(Y) = 0$, and the cost of the network E^* is $\eta\mu'$ (i.e. $e(E^*) = \mu'$).

That is, in equilibrium, almost every pair transmits information with positive probability (i.e. $\lambda(Y) = 0$), and for almost every pair (s,r) and any $l \geq \underline{m}+1$, if $\sigma_A^*(E^*,s,r)[l] = 0$ then $\sigma_A^*(E^*,s,r)[l] = 0$. (i.e. $\lambda(X) = 0$.)

For each l, denote $\bar{l}(s,r) = sup\{l : \sigma_A^*(E^*,s,r)[l] > 0\}$, which is well defined for almost every pair (s,r) since $\lambda(Y) = 0$. By Step 4, agents never use path lengths $l \leq \underline{m}$, and so $\bar{l}(s,r) \geq \underline{m}+1$.

By combining $\lambda(X) = 0$ and Step 4, we find that for a.e. pair (s, r), $\sigma_A^*(E^*, s, r)[l] > 0$ if and only if $\underline{m} + 1 \leq l \leq \bar{l}(s, r)$.

Step 8: Suppose that $\sigma_R^*(s, r)[\underline{m}] = 0$. Then the regulator puts all probability on $m \geq \underline{m} + 1$. Then using a path of length $\underline{m} + 1$ means certainty of getting caught. Since $K(\underline{m} + 1) > \Pi(\underline{m} + 1)$, (s, r) would deviate to shifting the probability of $\sigma_A^*(E^*, s, r)[\underline{m} + 1]$ to the strategy of not transmitting information, which would be a profitable deviation. Contradiction. Hence, $\sigma_R^*(s, r)[\underline{m}] > 0$.

Step 9: (Finite $\overline{m}$) If $\bar{l}(s, r) \leq \overline{m}$, playing $\overline{m}$ then gives certainty of catching (s, r) so the payoff the to regulator is $-\kappa(\overline{m})$. Since regulator puts positive probability on $\underline{m}$, the equilibrium payoff is equal to the payoff from playing $\underline{m}$, which $-\kappa(\underline{m}) - B = -B$. But $-B < -\kappa(\overline{m})$ by definition, meaning that the regulator has a profitable deviation. Therefore, $\bar{l}(s, r) > \overline{m}$. Also, by Step 5 we find $\bar{l}(s, r) \leq \overline{m} + 1$. Therefore, $\bar{l}(s, r) = \overline{m} + 1$.

Step 10: Suppose that $\sigma_R^*(s, r)[l] = 0$ for some l such that $\underline{m} + 1 \leq l \leq \bar{l}(s, r) - 1 = \overline{m}$. Then (s, r) would shift the probability of using paths of length $l+1$ on to using paths of length l. (Note that by Step 7 $\sigma_A^*(E^*, s, r)[l] > 0$, meaning that paths of length l are available in E^*.) This does not change the probability of getting caught, but increases the benefits from transmission. Contradiction. Thus, $\sigma_R^*(s, r)[l] > 0$ for all l such that $\underline{m} + 1 \leq l \leq \overline{m}$.

Step 11: By combining Step 9 and Step 10, we find that, in equilibrium, a.e. pair totally randomizes over all path lengths from $\underline{m} + 1$ to $\overline{m} + 1$ (which implies that all such paths are available for almost every pair) and the regulator totally randomizes between all search intensities from $\underline{m}$ to $\overline{m}$. The probabilities are given by indifference conditions.

In particular, since all path lengths from $\underline{m} + 1$ to $\overline{m} + 1$ are available, the network E^* must be flexible. Moreover, by Step 7, $e(E^*) = \mu'$.

Step 12: (Finite $\overline{m}$) The indifference conditions yield that the payoff of agents from information transmission is equal to the payoff of agents from using the long path $\overline{m} + 1$. This payoff is $\Pi(\overline{m} + 1)$ because regulator never uses $m > \overline{m}$ and so there is no expected cost of getting caught. Note that since $\Pi(\overline{m} + 1) > 0$, agents do not put any probability on not sending information. Then, the expected payoff is $\frac{\mu'_S \mu'_R}{\mu_S \mu_R} \Pi(\overline{m} + 1)$ from transmission because there is $\frac{\mu'_S}{\mu_S}$ probability that a sender in S' has information and $\frac{\mu'_R}{\mu_R}$ probability that a receiver in R' has liquidity. Then the expected ex-post payoff of agents is $V' := \frac{\mu'_S \mu'_R}{\mu_S \mu_R} \Pi(\overline{m} + 1) - (\mu'_S + \mu'_R)\eta$.

Then consider the following deviation. Suppose that all agent in $S \cup R$ come together and form a flexible core-periphery. If a pair in $S' \times R'$ realizes, it uses the equilibrium transmission strategy in the deviation as well. If pairs in $(S \times R) \setminus (S' \times R')$ realize, they use a path of length $\overline{m}+1$. (Note that regulator never plays $m > \overline{m}$.) Therefore, at the deviation, when a pair in $(S \times R) \setminus (S' \times R')$ realizes, which have used a path longer than $\overline{m}$, they do not get caught. Then the expected payoff from this deviation is $V := \Pi(\overline{m}+1) - (\mu_S + \mu_R)\eta$.

Suppose that $\mu'_S \neq \mu_S$ or $\mu'_R \neq \mu_R$. Then by $\mu'_S \leq \mu_S$, and $\mu'_R \leq \mu_R$ (at least one strict), we have $(\mu'_S)^{-1} + (\mu'_R)^{-1} > \mu_S^{-1} + \mu_R^{-1}$. Then $(\mu'_S + \mu'_R)\frac{\mu_S\mu_R}{\mu'_S\mu'_R}\eta > (\mu_S + \mu_R)\eta \implies$

$$V = \Pi(\overline{m}+1) - (\mu_S + \mu_R)\eta > \Pi(\overline{m}+1) - (\mu'_S + \mu'_R)\frac{\mu_S\mu_R}{\mu'_S\mu'_R}\eta = \frac{\mu_S\mu_R}{\mu'_S\mu'_R}V'. \tag{3}$$

Recall that $\Pi(l) > \eta\mu$ for all l. Then $V > 0$. If $V' \leq 0$, then $V > V'$. If $V' > 0$, then by (3), $V > \frac{\mu_S\mu_R}{\mu'_S\mu'_R}V' > V'$, and so $V > V'$. In both cases, this is a profitable deviation. Contradiction. Then the equilibrium network must have $\mu'_S = \mu_S$ and $\mu'_R = \mu_R$.

That is, the equilibrium network must be flexible and it must cost $\mu\eta$. As shown in Step 1, realized pairs must mix totally between paths of length $\underline{m}+1$ and $\overline{m}+1$ and regulator must mix totally between $\underline{m}$ and $\overline{m}$.

Step 13: ($\overline{m} = \infty$) For the regulator, the payoff of using m satisfies

$$lim_{m\to\infty} - \kappa(m) - \left(\sum_{m\leq l-1} \sigma_A^*(E^*, s, r)[l]\right) B = -lim_{m\to\infty}\kappa(m) = \kappa(\overline{m}+1).$$

The regulator mixes between all $m \geq \underline{m}$. Therefore, the payoff of the regulator is equal to his payoff from playing $\underline{m}$. Therefore,

$$-\kappa(\underline{m}) - (1 - \sigma_A^*(E^*, s, r)[o])B = \kappa(\overline{m}+1).$$

Thus,

$$\sigma_A^*(E^*, s, r)[o] = 1 - \frac{\kappa(\infty)}{B} > 0.$$

Then Agents' payoff from transmission is equal to 0. Then agents would deviate to not forming any links to save the cost. Thus, $\mu' = 0$.

Step 14: For finite $\overline{m}$, a.e. agent has a link, and a.e. pair engages in information transmission. So observing a.e. pair (s, r) is on-the-path for the regulator. So there are no profitable deviations following the earlier arguments.

For $\overline{m} = \infty$, By Step 12, the only candidate network is almost empty. Therefore, a.e. pair (s, r) is off-the-path. For each off-path pair (s, r), suppose that regulator plays the strategy by given by Step 11, and regulator's off-path beliefs are given by the the strategy of (s, r) in Step 11. This constitutes an equilibrium by the earlier steps. $\square$

Proof. **(Lemma 5)** By definition, P has μ measure. Also, $c(i, E) = \eta$ for all $i \in P$. C has zero measure. Then $C(E) = \mu\eta$. $\square$

Proof. **(Theorem 2)** The construction in the proof of Lemma (4) for is a flexible core-periphery network. $\square$

Proof. **(Proposition 2)** The core is at most countable and information transmission always occurs through finite path lengths all bounded above by $\overline{m} + 1$. By the countable additivity property of the measure, if one takes the non-intermediaries from the core, a.e. pair of senders and receivers still have access to paths of lengths ranging from $\underline{m} + 1$ to $\overline{m} + 1$. the resulting network is still a flexible core-periphery network. $\square$

Proof. **(Theorem 3)** The cost of the insider network $-\eta(\mu'_S + \mu'_R)$. The expected profit generated by the insider network by sending the information to high liquidity receivers is $\frac{\mu'_S \mu'_R}{\mu_S \mu_R} \Pi(\overline{m} + 1)$. The expected profit generated by insiders μ'_S sending the information via the social network to low liquidity receivers is $\frac{\mu'_S}{\mu_S}\left(1 - \frac{\mu'_R}{\mu_R}\right) \chi \Pi(\overline{m} + 1)$. Denote $x_{\overline{m}} \leq \mu_S - \mu'_S$ the mass of senders outside the insider network who have $m_i \leq \overline{m}$. These senders do not have sufficiently long paths to overcome regulation with positive probability, and their expected profit from the continuation game is 0. Then the expected profit generated by sending information via the social network is $\left(1 - \frac{\mu'_S}{\mu_S} - \frac{x_{\overline{m}}}{\mu_S}\right) \chi \Pi(\overline{m} + 1)$. Then the total expected payoff is

$$U = -\eta(\mu'_S + \mu'_R) + \frac{\mu'_S \mu'_R}{\mu_S \mu_R} \Pi(\overline{m} + 1) + \left(\frac{\mu'_S}{\mu_S}\left(1 - \frac{\mu'_R}{\mu_R}\right) + \left(1 - \frac{\mu'_S}{\mu_S} - \frac{x_{\overline{m}}}{\mu_S}\right)\right) \chi \Pi(\overline{m} + 1)$$

$$-\eta(\mu'_S + \mu'_R) + \Pi(\overline{m} + 1)(1 - \chi) \frac{\mu'_S \mu'_R}{\mu_S \mu_R} + \Pi(\overline{m} + 1) \chi \left(1 - \frac{x_{\overline{m}}}{\mu_S}\right)$$

All else fixed, agents prefer to have $x_{\overline{m}}$ as small possible. Denote $\mu_{S,\overline{m}}$ is the the total mass of senders s_i who have their m_i less than or equal to $\overline{m}$. If they can all be fit into μ'_S, then it would be possible to make $x_{\overline{m}} = 0$. The optimal network requires putting as much of $\mu_{S,\overline{m}}$ as possible into the insider network so their low m_i does not impede transmission. Then $min\{\mu'_S, \mu_{S,\overline{m}}\}$ mass of senders in $\mu_{S,\overline{m}}$ should belong to the insider network. Then $\mu_{S,\overline{m}} - min\{\mu'_S, \mu_{S,\overline{m}}\} = max\{0, \mu_{S,\overline{m}} - \mu'_S\}$ is the smallest amount of senders in $\mu_{S,\overline{m}}$ that must stay outside the insider network. So $x_{\overline{m}} = max\left\{0, \mu_{S,\overline{m}} - \mu'_S\right\}$ is optimal, all else fixed. Then

$$U = -\eta\left(\mu'_S + \mu'_R\right) + \Pi\left(\overline{m}+1\right)\left(1-\chi\right)\frac{\mu'_S\mu'_R}{\mu_S\mu_R} + \Pi\left(\overline{m}+1\right)\chi\left(1 - \frac{max\left\{0, \mu_{S,\overline{m}} - \mu'_S\right\}}{\mu_S}\right)$$

As $\overline{m}$ goes up, $\mu_{S,\overline{m}}$ goes up from 0 to μ_S. Notice that U is linear in μ'_R. We analyze the various cases as follows.

Case 1: $\Pi\left(\overline{m}+1\right)\left(1-\chi\right)\frac{\mu'_S}{\mu_S\mu_R} - \eta > 0$. In this case, optimally, $\mu'_R = \mu_R$. Then

$$U = -\eta\left(\mu'_S + \mu_R\right) + \Pi\left(\overline{m}+1\right)\left(1-\chi\right)\frac{\mu'_S}{\mu_S} + \Pi\left(\overline{m}+1\right)\chi\left(1 - \frac{max\left\{0, \mu_{S,\overline{m}} - \mu'_S\right\}}{\mu_S}\right)$$

Case 1.1: Consider $\mu'_S \geq \mu_{S,\overline{m}}$. Then

$$U = -\eta\left(\mu'_S + \mu_R\right) + \Pi\left(\overline{m}+1\right)\left(1-\chi\right)\frac{\mu'_S}{\mu_S} + \Pi\left(\overline{m}+1\right)\chi$$

Since $\mu_S = \mu_R$ and $\Pi\left(\overline{m}+1\right)\left(1-\chi\right)\frac{1}{\mu_S} - \eta > 0$, U is maximized at $\mu'_S = \mu_S$ within the constraint $\mu'_S \geq \mu_{S,\overline{m}}$. Then

$$\begin{aligned} U &= -\eta\left(\mu_S + \mu_R\right) + \Pi\left(\overline{m}+1\right)\left(1-\chi\right) + \Pi\left(\overline{m}+1\right)\chi \\ &= -\eta\mu + \Pi\left(\overline{m}+1\right). \end{aligned}$$

The solution of Case 1.1. is then $\mu'_S = \mu_S, \mu'_R = \mu_R$ with resulting value $-\eta\mu + \Pi\left(\overline{m}+1\right)$.

Case 1.2: Consider $\mu'_S \leq \mu_{S,\overline{m}}$. Then

$$U = -\eta\left(\mu'_S + \mu_R\right) + \Pi\left(\overline{m}+1\right)\left(1-\chi\right)\frac{\mu'_S}{\mu_S} + \Pi\left(\overline{m}+1\right)\chi\left(1 - \frac{\mu_{S,\overline{m}} - \mu'_S}{\mu_S}\right).$$

This is increasing in μ'_S. Then the constrained optimum is $\mu'_S = \mu_{S,\overline{m}}$. Then

$$U = -\eta\left(\mu_{S,\overline{m}} + \mu_R\right) + \Pi\left(\overline{m}+1\right)\left(1-\chi\right)\frac{\mu_{S,\overline{m}}}{\mu_S} + \Pi\left(\overline{m}+1\right)\chi.$$

This is smaller than the value in Case 1.1.

Therefore, the optimum under $\Pi\left(\overline{m}+1\right)\left(1-\chi\right)\frac{\mu'_S}{\mu_S\mu_R} - \eta > 0$ (Case 1) is $\mu'_S = \mu_S, \mu'_R = \mu_R$, which gives $U = -\eta\mu + \Pi\left(\overline{m}+1\right)$.

Case 2: $\Pi\left(\overline{m}+1\right)\left(1-\chi\right)\frac{\mu'_S}{\mu_S\mu_R} - \eta \leq 0$. Then $\mu'_R = 0$ is optimal.[31] Note that this does not mean there are no receivers in the insider network. This means there is a zero measure of receivers in the insider network. Then

$$U = -\eta\mu'_S + \Pi\left(\overline{m}+1\right)\chi\left(1 - \frac{max\left\{0, \mu_{S,\overline{m}} - \mu'_S\right\}}{\mu_S}\right).$$

Case 2.1: Consider $\mu'_S \geq \mu_{S,\overline{m}}$. (Note that this is possible only if $\Pi\left(\overline{m}+1\right)\left(1-\chi\right)\frac{\mu_{S,\overline{m}}}{\mu_S\mu_R} -$

[31]If $\Pi\left(\overline{m}+1\right)\left(1-\chi\right)\frac{\mu'_S}{\mu_S\mu_R} - \eta \leq 0$, any μ'_R is optimal. Otherwise, $\mu'_R = 0$ is the unique optimum.

$\eta \leq 0$ because of the premise of Case 2.) Then

$$U = -\eta \mu'_S + \Pi\left(\overline{m}+1\right)\chi.$$

This is decreasing and maximized at $\mu'_S = \mu_{S,\overline{m}}$, which gives value

$$U = -\eta \mu_{S,\overline{m}} + \Pi\left(\overline{m}+1\right)\chi.$$

Case 2.2: Consider $\mu'_S \leq \mu_{S,\overline{m}}$. Then

$$U = -\eta \mu'_S + \Pi\left(\overline{m}+1\right)\chi \frac{\mu'_S}{\mu_S} + \Pi\left(\overline{m}+1\right)\chi\left(1 - \frac{\mu_{S,\overline{m}}}{\mu_S}\right).$$

Note that $\mu'_S = \mu_{S,\overline{m}}$ gives the optimal value in Case 2.1. So we can ignore Case 2.1.

Since $-\eta + \Pi\left(\overline{m}+1\right)\chi \frac{1}{\mu_S} < 0$ (because $\chi < \frac{\delta}{2}$), U is decreasing and maximized at $\mu'_S = 0$ which gives value $U = \Pi\left(\overline{m}+1\right)\chi\left(1 - \frac{\mu_{S,\overline{m}}}{\mu_S}\right)$.

All in all, first option is $U_1 = -\eta\mu + \Pi\left(\overline{m}+1\right)$ with $\mu'_S = \mu'_R = \mu_S = \mu_R = \mu/2$ (the optimal in Case 1). The second option is $U_2 = \Pi\left(\overline{m}+1\right)\chi\left(1 - \frac{\mu_{S,\overline{m}}}{\mu_S}\right)$ with $\mu'_S = \mu'_R = 0$ (the optimal in Case 2).

$$U_1 > U_2 \iff -\eta\mu + \Pi\left(\overline{m}+1\right) > \Pi\left(\overline{m}+1\right)\chi\left(1 - \frac{\mu_{S,\overline{m}}}{\mu_S}\right)$$

$$\iff \left(\frac{\eta\mu}{\Pi\left(\overline{m}+1\right)\chi} - \frac{1}{\chi} + 1\right)\mu_S < \mu_{S,\overline{m}}.$$

At the cutoff $\left(\frac{\eta\mu}{\Pi(\overline{m}+1)\chi} - \frac{1}{\chi} + 1\right)\mu_S$, agents switch from social networks to insider networks. For $\mu_{S,\overline{m}} < \left(\frac{\eta\mu}{\Pi(\overline{m}+1)\chi} - \frac{1}{\chi} + 1\right)\mu_S$, the optimal is the social networks with $\mu'_S = \mu'_R = 0$. For the reverse, the optimal is $\mu'_S = \mu_S, \mu'_R = \mu_R$. □

Proof. **(Proposition 3)** Take any $i \in H_S$. All geodesics from $s \in N^-_{s,E} \cap S \cap P$ to all receivers go through i. Then $g_i\left(s,r\right) = g\left(s,r\right)$ for all $s \in N^-_{s,E} \cap S \cap P$ and $r \in R$. Then i has at least $\lambda\left(N^-_{s,E} \cap S \cap P\right)\mu_R = \lambda\left(N_{s,E}\right)\mu_R$ betweenness centrality. Same argument works for H_R. □

Proof. **(Theorem 4)** Since $\overline{m}$ is finite, an agent i has positive flexible-betweenness centrality if and only if there exists l between $\underline{m}+1$ and $\overline{m}+1$ such that for a positive measure of sender receiver pairs i is on a path of length l from the sender to the receiver.

Notice that in the proof of Proposition (1), when a path of length l is to be utilized by a pair, it does not matter which path of length l is used. Any distribution over all paths of length l give rise to an equilibrium. For example, given any equilibrium, uniformly distributing the probability of using path length l over over all paths of length l is also an equilibrium. This can be done for all l together. Since there are countably many agents in the core and $\overline{m}$ is finite, by the countable additivity property, the newly constructed equilibrium makes all agents with positive flexible-betweenness centrality an intermediary. $\square$