The Enemy You Can't See: An Investigation of the Disruption of Dark Networks*

Michael McBride†
Department of Economics
University of California, Irvine

David Hewitt‡
Department of Economics
Whittier College

This version: August 2012

Abstract

We examine the optimal disruption of dark (covert and illegal) networks. Of central importance is that an interventionist will generally have incomplete information about the dark network's architecture. We derive the optimal disruption strategy in a stylized model of dark network intervention with incomplete information and show how it combines features of two types of disruption considered in the literature: random failure and targeted attacks. In particular, the optimal disruption strategy encourages greater risk as less of the architecture is observed. A laboratory experiment finds that subjects tasked with disrupting a dark network qualitatively mimic the theoretical predictions.

JEL Classifications: C91, D85, K42.

Keywords: terrorism, crime, intervention, defense.

*The material herein is based upon work supported by the Air Force Office of Scienctific Research Award No. FA9550-10-1-0569.

†Corresponding author. 3151 Social Science Plaza, Irvine, CA, 92697-5100, mcbride@uci.edu, 949-824-7417, 949-824-2182 (fax).

‡Hoover 123, Whittier, CA, 90608, dhewitt@whittier.edu, 562-907-4906.

1 Introduction

Social and technological networks enhance many forms of socially beneficial activities (Jackson 2008), but they can also foster harmful and destructive behaviors. Such is the case with covert and illegal networks, also called *dark networks* (Raab and Milward 2003, Milward and Raab 2005). Examples include terrorist organizations, youth gangs, drug-trafficking rings, and other criminal enterprises. Because these networks face a constant threat of disruption by law enforcement agencies and militaries, dark network members attempt to evade both detection and intervention. As Krebs (2002) demonstrates with the 9/11 hijackers, this evasion makes it difficult for the interventionist (or disrupter) to map the dark network's structure. An interventionist must thus intervene with very limited and incomplete information.

This paper examines dark network disruption by an interventionist with incomplete knowledge of the network structure. We ask: How should an interventionist choose her intervention when confronted with limited information of the network's architecture? And how does the interventionist's effectiveness depend on the amount of knowledge of the dark network and the dark network's actual, unobserved structure? These questions are both normative and positive, and they require both theoretical and empirical investigation. We first provide a formal theoretical examination of dark network disruption in a stylized dark network setting. The very nature of dark networks makes it impossible to test our predictions in the field, so for practical and ethics reasons we then implement a second best approach of conducting a laboratory experiment.

A main innovation of our paper is the development of a stylized model of dark network disruption under incomplete network information. We recognize that replicating all features of dark networks, including their subversive nature and intent, is impossible in an abstract model. However, our goal is not to create a model that perfectly represents actual dark networks but rather to create a simple model that, first, captures basic trade-offs faced by an interventionist having to choose a strategy with incomplete information and, second, is flexible enough to form the basis of future work. We examine the simplest setting to accomplish this goal: an interventionist observes only the links of a randomly selected subset of criminal nodes and then decides which single node to remove in order to have the largest expected reduction in crime.

With complete information about the network architecture, the optimal intervention in this setting is to remove the node with the most connections. However, we show that with incomplete observation, the interventionist must decide whether to remove the node with the most observed connections among those whose connections were monitored (the sure thing) or whether to arrest another node for whom some but possibly not all connections are observed (the risky choice). Under incomplete information the optimal intervention involves a making the risky choice with higher likelihood as less of the network architecture is observed. This optimal strategy is thus effectively a hybrid of the random failure and perfectly targeted attack considered in related literature (discussed below). The expected effectiveness of interventions also decreases as less of the network is known.

To test whether decision-makers tasked with disrupting dark networks act in accordance with these predictions or whether their disruptions systematically diverge from the optimal behavior, we conducted a controlled experiment with human subjects. Subjects were tasked with reducing the total level of crime generated by a "criminal network." In each round, each subject was presented with an incomplete image of a criminal network and asked to select a criminal to arrest. By exogenously varying the number of criminals monitored, we controlled the subject's level of observation of the network structure and were able to test the effect of changes in the observational level on disruption actions and outcomes. The results qualitatively support the predictions of the model: subjects make riskier interventions as the network darkness increases, and their effectiveness in deterring crime worsened despite their adaptation. However, we also find that the incidence of clearly inferior (payoff dominated) choices goes up when information is very limited.

The theoretical and experimental findings yield multiple lessons on dark network intervention more generally. When faced with very little information about the organization, an interventionist's optimal strategy generally entails disrupting parts of the network that have not been fully monitored. Moreover, the finding that untrained human subjects qualitatively mimic the optimal strategy implies that the basic intuition is comprehensible to decision-makers. Finally, we observe a striking deviation from the optimal strategy among some subjects in the extremely limited information setting which suggests that educating interventionists is most beneficial—indeed, most important—when information about the dark network is very incomplete.

2 Related Literature

Our work complements many existing literatures, both in and out of economics. One is the theoretic literature on network disruption. Borgatti (2003) defines the "key player problem" as identifying which node(s) to remove from the network to have the largest impact on network performance. Ballester et al (2006) examine this problem in a game theoretic setting in which criminals are located on an exogenous network and exert effort with either strategic complementarities or substitutabilities. They show that the key player in the Nash Equilibrium is shown to manifest a certain type of network centrality (Bonacich centrality, see Wasserman and Faust 1994 for a definition). Our disruption setting is simpler than Ballester et al in that the network is given and fixed. However, ours is the first to explicitly account for network darkness, and thus is the first to examine how the optimal disruption strategy changes as the darkness changes.[1]

Another strand of work examines the strategic interplay between the network members

[1] Though there is no economics literature on dark network disruption under incomplete network information, there is a literature in operations research that explicitly confronts the limited information aspect of dark networks. For example, Miffen et al (2004) use a random graph model to probabilistically infer an underlying terrrorist network from other noise; Atkinson and Wein (2010) leverage the fact that criminal and terrorist networks often overlap to identify a mathematical procedure for identifying how best to allocate scarce resources towared terrorist detection; and Kaplan (2010) uses Markov processes to model infiltration and interdiction in an ongoing terror network. Of course, real-world interventionists have many tactics and approaches that can be used in disrupting covert and illegal organizations. For example, eee Roberts and Everton 2011 for a typology of other kinds of disruption strategies.

and the disrupter. In Enders and Su (2007), Baccara and Bar-Isaac (2008), and Enders and Jindapon (2009), the disrupter wants to learn the information communicated through the network, and the criminals act to keep this information away from the disrupter. In Hoyer and Jaegher (2010), on the other hand, the disrupter wants to disrupt the information flow in the network. The network designer first selects the network, and the network disrupter then selects which nodes (or links) to remove. Goyal and Vigier (2010) and Hong (2008) consider similar settings but where the designer organizes the network to block important players. In Hong (2011), information is modeled as a flow on a network which the interventionist acts to block. Kovenock and Roberson (2010) consider strategic defense and attack in "weakest-link" and "best-shot" network settings where nodes are contested with costly resources. Colonel Blotto Games are a specific form of network-based conflict with simple (line or grid) architectures (e.g., see Roberson 2006). Franke and Öztürk (2009) show how conflict intensity can depend on the network structure. Hausken (2009) considers attack and defense of contested nodes for arbitrarily complex systems. This strategic interplay can also be modeled on a more abstract level. Arin, et al (2011) study the relationship between spending on terrorism and counter-terrorism without delving into the specific structure of the terror network. Likewise, Das (2008) studies possible mechanisms behind terror cycles. Again, our model differs in being decision theoretic rather than game theoretic, and our focus is on disruption with incomplete network information.

Though there are laboratory experiments studying various aspects of networks (see Kosfield 2004) and other experiments examining other forms of counterterrorist policies (Arce et al 2011), there are few experimental studies of network defense or disruption, with the experimental study of Blotto Games (e.g., Chowdhury et al 2009) a notable exception. Ours is the first experimental study of optimal network disruption with incomplete knowledge of the network's architecture.

A relevant literature outside of economics uses computational methods to examine the robustness of networks (not necessarily dark networks) to two types of disruption: random failures and targeted attacks (e.g., Albert, Hawoong, and Barabási 2000). An unexpected shut down at a power station on the electrical grid is a random failure. The inoculation of a well-connected citizen is a targeted attack intended to reduce the spread of disease.[2] Feinstein and Kaplan (2010) bridge the gap between computational and game theoretic studies in their use of simulations to study defense and attack on a terrorist organizations. Our paper links this computational research with the economics literature by showing how, in a dark network setting, the optimal disruption strategy is effectively a hybrid of random failure and perfectly targeted attacks: the interventionist targets her disruption conditionally on the realization of a random revelation of part of the network.

Of course, we acknowledge that the study of crime and crime deterrence (e.g., Becker 1968, Cameron 1988) and terrorism and counter-terrorism (e.g., Intriligator 2010, Llussá and Tavares 2007) has a tradition in economics. Overall, our paper contributes to these general literatures by providing a stylized dark network model, presenting new theoretical and experimental results on behavior in the stylized setting, and generating new lessons learned about dark network disruption.

3 A Theoretical Model

3.1 Model

Consider a set of known criminals $I = \{1, ..., n\}$ and an interventionist. Nature selects a criminal network (graph) g from the set of all possible criminal networks G. If a criminal collaboration exists between i and j in g, then $g_{ij} = g_{ji} = 1$; if no collaboration exists between i and j, then $g_{ij} = g_{ji} = 0$. To complete the network, assume $g_{ii} = 0$. Let $\pi_0 \in \Delta G$ represent the prior distribution of criminal networks according to which nature selects g. We elaborate on π_0 below.

[2] An important finding is that certain networks are more robust to one type of disruption than another. For example, scale-free network architectures like that of the World Wide Web are highly resilient to random failures but not targeted attacks (Albert, Hawoong, and Barabási 2000, Callaway, et al 2000, Cohen, et al 2001). Farley (2010) uses order theory to study robustness to attack when there are different criminal types (e.g., leaders and followers).

After nature selects the network g, nature then selects z criminals to be "monitored" by the interventionist, with $0 \leq z \leq n$. Being monitored means that all of a criminal's collaborations (to both other monitored criminals and to non-monitored criminals) are revealed to the interventionist. Let $I_z = \{i \in I | i \text{ is monitored}\}$ be the set of monitored criminals. Assume that monitored criminals are selected uniformly from I without replacement. Let $\overline{I}_z = I \backslash I_z = \{i \in I | i \notin I_z\}$ be the complement of I_z, i.e., the set of criminals who are not monitored.

After the interventionist monitors the criminals in I_z, she selects one criminal from I to arrest. She can arrest any single criminal from I, whether or not that criminal was monitored. Let $\widehat{g}$ be the post-arrest criminal network such that if criminal k is arrested then $\widehat{g}_{ij} = g_{ij}$ for all $i, j \in I$ with $i, j \neq k$ and $\widehat{g}_{ik} = \widehat{g}_{ki} = 0$ for all $i \in I$. As widely assumed for decisions under incomplete information, we assume that n, G, π_0, z, and all other facets of the setting are known by the interventionist when making her arrest decision. Of course, the interventionist does not know g (except in the trivial cases of $z \geq n-1$) when making her choice.[3]

After the interventionist makes her arrest decision, the criminals deterministically engage in crime according to $\widehat{g}$, and the interventionist's payoff is realized. We assume that crime is linear in the number of collaborations in the network. Letting $d_i(g')$ be the degree (number of collaborations) of criminal i in network $g' \in G$, we have crime function $C = c_0 + c_1 \sum_{j=1}^{n} d_j(\widehat{g})$, where $c_0 \geq 0$, and $c_1 > 0$. The interventionist's payoff is a strictly decreasing linear function of the total level of realized crime (or a strictly increasing function of the amount of crime deterred). Thus, the interventionist's optimal strategy is to choose the criminal with the highest expected degree. This setting is, in effect, a game played by a single player (the interventionist) against nature.

[3] A technical matter is whether information is *incomplete* or *imperfect* in this setting. Because we model the network as an apriori unknown state variable, it is appropriate to call this a setting of incomplete information. However, it is worth noting that if the network was the result of explicitly modeled unobserved collaboration choices by criminals, then the imperfect label is appropriate. We use the term incomplete here because it best reflects our simplest depiction.

We model crime as a linear function of degrees for two primary reasons. First, it greatly simplifies the interventionist's decision problem. Given our motivation to study the impact of imperfect information of the network structure on the effectiveness of the interventionist, we believe it prudent to keep the interventionist's underlying task simple in dimensions other than the imperfect monitoring. This simplicity is additionally important given our intent to study the interventionist's decision in the lab, where we are better able to induce preferences with an easy to understand decision problem. Second, although we acknowledge that real-life crime and other illegal or violent activities depend on more than just collaborations, it is still true that collaborations are important. Thus, our focus on collaborations is not without any merit. We also note that our assumption can be understood as implying a strong form of homogeneity among criminals. Criminals are not distinguished by some primitive type (thieves, enforcers, accountants, money launderers, etc.) but instead differ primarily by their collaborations. This simple setting is thus a viable benchmark with which future studies can be compared.

We initially make no assumption about the prior distribution π_0 other than a strict form of anonymity: if g' and g'' are isomorphic then $\pi_0(g') = \pi_0(g'')$. Networks g' and g'' are isomorphic if the collaborations are identical up to a renaming of criminals, so this assumption suggests no prior belief that a particular subject is likely to have more collaborations or fewer collaborations than another. This assumption simplifies the analysis but is also consistent with the homogeneity in criminal types assumed in the model.

3.2 Optimal Intervention

By her knowledge of the game, the interventionist uses Bayes Rule after monitoring z criminals to update her belief about all criminals' degrees. An important distinction arises between monitored criminals and non-monitored criminals. For a monitored criminal $i \in I_z$, Bayes Rule implies that the posterior beliefs π must assign probability 0 to any $g' \in G$ in which $d_i(g') \neq d_i(g)$. But the interventionist also observes some of the collaborations of the non-monitored criminals; i.e., she observes every collaboration between a monitored criminal and a non-monitored criminal. With $z < n-1$, the interventionist will in general not be sure if she has observed all of the collaborations of any of those non-monitored players. Yet, seeing some collaborations is very relevant, because for each non-monitored criminal $i \in \overline{I}_z$, Bayes Rule implies that the posterior beliefs π must assign probability 0 to any $g' \in G$ in which $d_i(g')$ is strictly less than the number of observed links for i.

Some additional notation will help us identify how the interventionist uses the information about both monitored and non-monitored criminals in making her decision. Let $d_z^{\max} \equiv \max_{i \in I_z} \{d_i(g)\}$ be the largest degree among the set of monitored criminals. Let $\widetilde{d}_z^{\max} \equiv \max_{i \in \overline{I}_z} \{d_i(\widetilde{g})\}$ be the largest observed degree among the set of non-monitored criminals, where $\widetilde{g}$ is that part of g that is observed. Let $\widetilde{I}_z^{\max} = \left\{ i \in \overline{I}_z | d_i(\widetilde{g}) = \widetilde{d}_z^{\max} \right\}$. Note that the largest observed degree is not necessarily the largest degree among the non-monitored criminals; it is only the largest degree observed.

We can now state our first proposition. The proof of this proposition and all other propositions are in the appendix.

Proposition 1 *Fix the number of monitored criminals z. The optimal arrest decision is to choose a monitored criminal with highest degree $d_z^{\max}$ among the set of monitored criminals I_z or to choose a non-monitored criminal with the highest observed degree $\widetilde{d}_z^{\max}$ among the set of non-monitored criminals $\overline{I}_z$.*

A monitored criminal known to have fewer collaborations than another monitored criminal should never be selected. The second part of the proposition is less trivial but still intuitive. Because criminals' *ex ante* expected degrees are identical (by anonymity), observing a non-monitored criminal to have more collaborations than another non-monitored criminal will, by applying Bayes Rule, lead the investigator to believe that the former is more likely to have more actual collaborations than the latter.

We turn now to the optimal arrest.

Proposition 2 *Fix the number of monitored criminals z. There exists a function*

$$m(z,\pi_0) \equiv \frac{1}{Q} \sum_{x=\widetilde{d}_z^{\max}}^{n-1-\left(z-\widetilde{d}_z^{\max}\right)} \sum_{g' \in \widehat{G}} x\pi_0(g')\, 1\left(d_i(g') = x\right) - \widetilde{d}_z^{\max} \geq 0$$

such that an optimal arrest is to arrest a highest degree monitored criminal when $d_z^{\max} \geq \widetilde{d}_z^{\max} + m(z,\pi_0)$, *but select a highest observed degree non-monitored criminal otherwise.*

According to Proposition 2, a highest degree monitored criminal is arrested only when he has sufficiently more collaborations than the highest observed degree non-monitored criminal. Playing it safe by selecting a monitored criminal is only worthwhile when the known crime reduction from selecting the monitored criminal is sufficiently better than the risky selection of a non-monitored criminal. We can interpret $m(z,\pi_0)$ as a premium that must be guaranteed to the interventionist in order to arrest a monitored criminal. If $d_z^{\max} \leq \widetilde{d}_z^{\max}$, then the interventionist cannot lose by arresting the non-monitored criminal with most collaborations. But when $d_z^{\max} > \widetilde{d}_z^{\max}$, the interventionist has a trade-off to consider: she can take the sure thing by arresting the monitored criminal or she can take a chance to do better (but also worse) by arresting the non-monitored criminal with most collaborations. Mathematically, $m(z,\pi_0)$ is exactly equal to the expected number of degrees of the highest observed degree non-monitored criminal minus the observed degree of that criminal.

How this premium $m(z,\pi_0)$ changes as z changes depends on the prior beliefs π_0. A case of particular interest is the Poisson random network. Fix the n criminals, and form a collaboration between criminals i and j with probability p, $0 \leq p \leq 1$, with these links independently formed across criminal pairs. Poisson networks have well-understood properties and form an important class of networks (see discussion in Jackson 2008). Any network is possible under this process, so G is the set of all possible networks with n nodes. With a prior belief that this process is used to select the network, then the premium $m(z,\pi_0)$ has a testable characteristic.

Proposition 3 *Suppose* π_0 *corresponds to a Poisson random network generation. Fix* $\widetilde{d}_z^{\max}$*.*

As z increases (decreases), then the premium $m(z, \pi_0)$ decreases (increases), and a monitored criminal is more (less) likely to be arrested.

With $\widetilde{d}_z^{\max}$ fixed in this setting, an increase in z implies that non-monitored criminals are now known to not have as many links as previously believed possible, and the expected crime reduction from arresting a non-monitored criminal decreases. We here see how the optimal strategy is a hybrid of the random failure and targeted attack mentioned earlier. With perfect information ($z \geq n-1$), the optimal disruption is equivalent to a perfectly targeted attack. With no information ($z = 0$), the optimal disruption is equivalent to a random failure in that the interventionist with the anonymous prior selects any node with equal chance. With partial but imperfect information ($0 < z < n-1$), the optimal disruption involves choosing the monitored node with the most links if that node has sufficiently many links, but it involves choosing a neighbor of a monitored node otherwise.

Our final proposition concerns expected crime reduction.

Proposition 4 *Fix g. As the number of monitored criminals z increases (decreases), the expected reduction in crime using the optimal disruption strategy increases (decreases).*

The interventionist's effectiveness at reducing crime worsens as her monitoring decreases. As she monitors fewer criminals, she has less information on which to base an informed arrest decision. She compensates by being more likely to select a non-monitored criminal, but this is a risky choice that is less and less likely to pay off as her monitoring decreases.

4 Experiment Design

4.1 Basics

We conducted multiple experiment sessions in a computer laboratory at a large public university with undergraduate students as human subjects. Students were told of the subject pool via classroom advertisements, and they registered to be in the pool via an online registration system. Days before each experiment session, an email was sent to the subject pool notifying them of our upcoming session. Interested students then signed up for a specific session on the subject pool web site. Those who signed up received a reminder email about the session the day before it was conducted. Subjects were not allowed to participate in more than one session.

To facilitate experiment management, instruction, and data collection, we used the z-Tree software package (Fischbacher 2007). Random draws by the computer combined with each subject's choices during the experiment resulted a reduction in crime, which we measured as "crime units" averted. The more crime units averted, the more U.S. currency the subject received at the experiment session's end, paid privately to each subject upon exiting the laboratory. Each crime unit corresponded to \$0.006 (US dollars), a point made clear during the instructional phase of the experiment. Each subject also received a separate \$5 payment for showing up. Final payments were rounded up to the nearest quarter. The average take-home amount was about \$28 for about 75 minutes of participation.

4.2 A Single Round

In each round (trial), the subject was shown a partial image of a criminal network comprised of 18 criminals and asked to select a single criminal in that network to arrest. The total crime produced by an individual criminal i, denoted c_i, was $c_i = 15 + 10l_i$, where l_i is the number of collaborations ("links") with other criminals. This formula was displayed on the screen during the duration of the experiment. Without any arrest made, the total (pre-arrest) crime produced by the underlying network would be $\sum_i c_i$, but with criminal j arrested, the total would become

$$\begin{aligned} & \sum_i c_i - c_j - 10l_j \\ = & \sum_i c_i - (15 + 10l_j) - 10l_j \\ = & \sum_i c_i - 15 - 20l_j. \end{aligned}$$

This formula was not displayed, but during the instructional phase the subjects were tested on their comprehension about how arrests reduce crime, and they received feedback on their answers.

At the start of a round, the computer generated a partial network image for each subject. It first selected which network would be the actual underlying criminal network (see below). It then randomly (with uniform probability) selected z of the 18 criminals to be monitored. The generated partial network image placed the 18 criminals on a circle and displayed only the collaborations of those z selected subjects. Criminals were represented by triangles, and revealed collaborations were displayed as links connecting triangles. The z monitored criminals were identified by green triangles, and the $18 - z$ other criminals were identified by blue triangles. Thus, the subject knew that she saw all of the collaborations of the greens, and she knew that the blues may have had links that were not displayed. A blue's collaboration with a green would be shown, but collaborations between blues were not shown. This color identification was explained to the subject during the instructional period.

The subject then arrested a criminal by mouse-clicking on a triangle. Any criminal (blue or green) could be selected, a fact explained during the instructional period. After arresting a criminal, the subject was presented with an image of the full underlying network, with the arrested criminal and her collaborations colored red. The pre-arrest level of crime, post-arrest level of crime, amount of crime prevented, and percent of crime reduced were displayed, as well as a running total of earnings up to that point in the session. Images of a pre-arrest and post-arrest screen for a particular network with $z = 7$ are displayed in Figure 1.

There are various ways to represent network structures. For example, rather than a network image, information about collaborations could have been revealed via text or a table. We chose a network image rather than text because it conveys a lot of information in an intuitive format. We chose the circle shape to minimize the bias in presentation. Future work is needed to determine to what extent the presentation of information about collaborations influences subjects' decisions in this setting.

4.3 Observation Levels

The key treatment variable is z, the number of monitored subjects. By exogenously changing this variable, we probabilistically control the amount of the network revealed to the subject. We also refer to z as the level of network observation.

We use three values of z: 12, 7, and 2. Treatment values 12 and 2 were chosen because they represent near extreme observation settings. With 12 criminals' collaborations revealed, the entire underlying network is almost fully revealed. With only 2 criminals' collaborations revealed, hardly any of the underlying network is revealed. 7 was chosen because it represents an intermediate setting where much of the underlying network is revealed but much is unseen.

4.4 Networks

The second important treatment variable is the set of underlying networks. Ideally, we could identify the properties of actual dark networks and create a set of underlying networks that have these properties. Researchers have attempted to construct images of dark networks using the limited information available. For example, various researchers have mapped out the known parts of certain terrorist networks post-attack (e.g., Sageman 2004, 2008; Pedahzur and Perliger 2006; and Xu and Chen 2008). We learn from these exercises that the revealed parts of these terrorist networks mimic other social networks in that they tend to have higher clustering and fatter-tailed degree distributions than uniformly-random generated networks. However, even these reconstructions are incomplete pictures; the amount of incompleteness is not known; and it is not known if and to what extent those partial mappings are biased.

We pursued a different approach to selecting networks. We generated two sets of 16 networks each. For the first, Set A, we generated a sequence of networks such that certain nodes were more likely to have collaborations than others. The resulting set of networks has, on average, higher clustering and fatter-tailed degree distributions than uniformly-random generated networks. Thus, these networks have, on average, the properties common to many types of social networks, so we consider Set A as consisting of generically realistic social networks. To the extent that dark networks have these properties, the networks in Set A mimic actual dark networks. For example, Pedazhur and Perliger (2006) argue that Palestinian suicide bomber networks mapped post-attack do reflect underlying social relationships that have these properties.

The networks in Set B were also randomly generated, but they were selected by their individual degree distribution with more explicit control over the total number of degrees. There are four networks each of 22, 26, 34, and 38 degrees. For each of these degree levels, half of the networks have a single hub (a criminal with more than twice the average number of degrees) with a small number of other highly connected criminals, and the other half have multiple hubs.

We did not explain to the subjects the process used to generate the set of networks they experienced, and as such we will expect the subjects' choices to be qualitatively similar across both sets of networks. Having two sets of networks provides a way to test for this similarity.

4.5 Sessions

We ran four sessions with the Set A networks. We denote the first session by A-12-7-2. Each subject experienced 16 rounds of the Set A networks with $z = 12$. The sequence in which the subjects were shown the networks over the 16 rounds was independently random across subjects, and the selection of the z monitored criminals within the round was independent across subjects. After completing the 16 networks with $z = 12$, the subject then did 16 rounds with the same Set A networks in a new random order and with $z = 7$. Finally, the subject did another 16 rounds with the Set A networks in another new random order with $z = 2$. Before each set of 16 rounds, the subject did 2 practice rounds with the respective z. Upon completion of the session, each subject had made a choice for each of the 16 Set A networks under three different observation settings, for a total of 48 rounds, each with its own network-observation level combination.

Session A-2-7-12 is similar to the first session except the observation order is reversed. This "cross-over" in treatment variable order allows us to identify if exposure to a particular treatment variable value affects later decisions.

We also conducted two sessions with the Set A networks where the 48 network-observation level combinations were done in random order (i.e., a subject could have $z = 2$ in round t, $z = 12$ in round $t+1$, $z = 2$ again in round $t+2$, and so on). Subjects completed 6 practice rounds (2 per observation level) before starting the 48 real rounds. We denote these two sessions A-Random-1 and A-Random-2.

We conducted three sessions with the Set B networks: B-12-7-2, B-2-7-12, and B-Random-1. These are identical to the corresponding A counterparts in all ways except Set B networks were used.

After completing all 48 network-observation combinations, each subject answered a short questionnaire that asked for age, sex, major, year in school, number of economics courses taken, number of statistics courses taken, and more. The questionnaire concluded with an open-ended question that asked the subject to explain how she made her choices. We use this information to compare subjects across sessions.

Table 1 summarizes some information about the different sessions.

4.6 Hypotheses

We identify four hypotheses, each corresponding to a proposition in Section 3.

Hypothesis 1 *For each level of the number of monitored criminals z, subjects will be much more likely to arrest either a monitored criminal with the most collaborations or a non-monitored criminal with the most observed collaborations.*

Hypothesis 2 *For each level of the number of monitored criminals z, non-monitored criminals are more likely to be selected than monitored criminals when the highest observed collaboration non-monitored criminal has equal or more collaborations than all monitored criminals.*

The first hypothesis corresponds to a necessary condition for an optimal decision, while the second identifies a circumstance where the optimal decision is clearly identified. When a monitored criminal has more collaborations than what are observed among the non-monitored criminals, then, as explained in Section 3, the optimal decision will depend on the prior beliefs and how those are updated. However, because in this experiment we do not provide the subjects with any particular information about the distribution from which networks are selected, we conjecture that subjects have priors that mimic those arising from Poisson network generations. Our next prediction follows.

Hypothesis 3 *As the number of monitored criminals z increases holding* $\widetilde{d}_z^{\max}$ fixed*, a monitored criminal is more likely to be arrested.*

By this hypothesis, subjects will engage in the riskier option of arresting a non-monitored criminal as their level of observation decreases. Our last hypothesis concerns the effectiveness of intervention as observation changes.

Hypothesis 4 *As the number of monitored criminals z increases, the average reduction in crime increases.*

5 Results

5.1 Hypothesis 1

For each arrest decision, we identified the arrested criminal by whether he was monitored or not. When the arrested criminal was monitored, we further identified whether or not the criminal had the maximum number of collaborations among all monitored individuals. The arrested criminal was classified with the label "Max Monitored" if he was monitored and had the maximum number of collaborations among monitored criminals, and labeled the criminal "Other Monitored" if he was monitored but did not have the maximum number of collaborations. We made a similar classification for non-monitored criminals using the labels "Max Non-monitored and Other Non-monitored."

Table 2 reports the distribution of choices according to these classifications. As predicted by Hypothesis 1, subjects are much more likely to select a "Max Monitored" or "Max Non-monitored" criminal than other criminals. This pattern is robust; it holds when pooling the data or when separating by session or by observation level or by network set.

The only setting in which other criminals are selected in non-trivial amounts is when the observation setting is $z = 2$. With the Set A networks, the percent of Other Non-monitored is highest in Session A-12-7-2 when they encounter the $z = 2$ observation level last and lowest in Session A-2-7-12 when they encounter $z = 2$ first. Subjects may be more conservative in their play when facing $z = 2$ first, but be more accepting of the risk when facing $z = 2$ last. However, this pattern is reversed with the Set B networks, which suggests that something other than order of exposure to the observation levels leads subjects to choose Other Non-monitored criminals.

Whatever the explanation, the subjects are more likely to deviate from Bayes Rule when $z = 2$ as compared with other observation levels. Yet, even with these decision making errors in the $z = 2$ setting, the percent of Other Non-monitored criminals is always below, and usually well below, the percent of Max Monitored and Max Non-monitored selected. This can be seen visually in Figure 2, which displays the distribution of choices graphically. We interpret these results as largely supportive of Hypothesis 1.

5.2 Hypothesis 2

Table 3 reports the percent of times subjects arrested a monitored individual. This is reported for each observation level for both network set and the difference in (observed) degree between the maximum monitored and non-monitored criminal. As predicted by Hypothesis 2, when the difference is non-positive ($d_z^{\max} - \widetilde{d}_z^{\max} \leq 0$), the subjects overwhelmingly choose non-monitored criminals. There are many sub-optimal arrests of monitored criminals, yet these errors become less common as the difference $d_z^{\max} - \widetilde{d}_z^{\max}$ becomes more negative.

5.3 Hypothesis 3

Table 3 also reveals that monitored criminals are more likely to be selected as $d_z^{\max} - \widetilde{d}_z^{\max}$ increases as predicted by Hypothesis 3. Figure 3 provides an alternative depiction. Table 4 reports the coefficients from probit regressions, one for each network set, that predict the probability that a monitored criminal will be selected. The baseline observation level is $z = 12$. The likelihood that a monitored criminal is arrested drops when $z = 7$ and drops even more when $z = 2$. Statistical tests reject at very high significance levels the hypothesis that the probabilities are the same for $z = 2$ and $z = 7$. The coefficient on degree difference $(d_z^{\max} - \widetilde{d}_z^{\max})$ is positive and highly significant. The clear pattern in the tables and figure is that subjects are making riskier arrests as their level of observation decreases, and this pattern is highly significant statistically.

5.4 Hypothesis 4

Table 5 reports statistics about average crime reduction by session and observation level z. Figures 4 and 5 display the distribution of crime reduction and the mean crime reduction by observation level. We observe a clear worsening of crime reduction as the observation level decreases. Average crime reduction is similar across sessions within the same network set and observation level. Table 6 provides two different statistical tests that reveal the average crime reduction is statistically significant. According to both the Student t and Wilcoxon test statistics, the differences in means across observational levels are always highly significant, both within network set and within session.

6 Discussion

This paper presents the first theoretical and experimental study of intervention in dark networks with incomplete knowledge of network architecture. We show that the optimal disruption involves a risk-return trade-off. The interventionist must decide whether to remove a node with known connectivity (the sure thing) or to remove a node with imperfectly known connectivity (the risky choice). Experimental subjects tasked with making this choice act in a manner that qualitatively matches the predictions.

Multiple lessons emerge from this work: one theoretical and one experimental stand out. The first lesson is that when the network structure is not fully known, an interventionist with limited disruption resources should, in settings with very limited monitoring, consider focusing disruption efforts on actors in the network that are not monitored even though doing so is risky. Indeed, as less of the network is observed, focusing disruptions on these unobserved parts of the network is more likely to be optimal. In an incomplete information setting, the interventionist must be vigilant in identifying which parts of the network are unknown.

Another lesson emerges from the experimental results. The theory assumes that the interventionist uses Bayes Rule to update her prior beliefs, and a wide body of experimental work (e.g., see El-Gamal and Grether 1995) establishes that experimental subjects do not always follow Bayes Rule. That subjects in our study do not always follow Bayes Rule is evident in Figure 2 (where some choose "Other Non-monitored" with $z = 2$) and in Figure 3 (where monitored criminals are sometimes arrested even when there are non-monitored criminal with the same or more connections). However, we find that the subjects' behavior qualitatively supports the theoretical predictions overall despite the deviations from Bayes Rule, and this suggests that the basic logic behind the optimal disruption strategy is compelling enough even to imperfect Bayesians.

The primary setting where noticeable suboptimal behavior occurs is the least information ($z = 2$) setting. A number of subjects, though still a minority, arrest individuals that are not the most connected even among not-monitored criminals. Exactly why some subjects make this decision error is unclear, but it is striking that the mistakes only occur in non-trivial levels in the least information setting. Nonetheless, given that the basic intuition behind the optimal strategy appears to be understood, we expect that an interventionist taught the optimal strategy would do even better. Our results suggest that training is most likely to be beneficial in settings with a high degree of uncertainty about the network's true structure.

We see many directions for future research. One is to consider imperfect monitoring in a setting with strategic interdependence between network members and interventionists. Both theoretical work and experimental work along these lines are warranted. Another direction is to extend the same basic stylized setting studied here in other directions of interest. For example, interventionists often do not know which actors are criminals. The model could be easily amended to allow for some actors to not be criminals and where monitoring reveals collaborations and information about whether an actor is a criminal or not a criminal. In this setting, the interventionist must consider the possibility of inadvertently arresting a non-criminal, an error the interventionist seeks to avoid. New research along these lines will yield even more insights into the optimal disruption of dark networks.

APPENDIX

A Proofs

Proposition 1: *Fix the number of monitored criminals z. The optimal arrest decision is to choose a monitored criminal with highest degree $d_z^{\max}$ among the set of monitored criminals I_z or to choose a non-monitored criminal with the highest observed degree $\widetilde{d}_z^{\max}$ among the set of non-monitored criminals $\overline{I}_z$.*

Proof. *(Part I: Arresting a monitored criminal with degree less than $d_z^{\max}$ is not optimal.)* Consider a monitored criminal i. By Bayes Rule, the interventionist's posterior beliefs π must assign probability 0 to any network $g' \in G$ in which that monitored criminal has collaborations not equal to $d_i(g)$. Hence, the expected degrees of each $i \in I_z$ is $d_i(g)$. Clearly, arresting a monitored criminal with $d_i(g) < d_z^{\max}$ yields a strictly lower expected payoff (strictly higher expected post-arrest crime) than arresting a monitored criminal with $d_i(g) = d_z^{\max}$. Thus, the interventionist's optimal decision can never be to arrest a monitored criminal $i \in I_z$ with $d_i(g) < d_z^{\max}$.

(Part II: Arresting a non-monitored criminal with observed degree less than $\widetilde{d}_z^{\max}$ is not optimal.) (a) We first find the interventionist's posterior beliefs.

Consider a non-monitored criminal $i \in \overline{I}_z$. The prior probability that criminal i has degree x in g is

$$\Pr\left[d_i(g) = x\right] = \sum_{g'} \pi_0(g')\, 1\left(d_i(g') = x\right),$$

where $1(\cdot)$ is the indicator function. Let F_0 be the resulting c.d.f.

After monitoring the z criminals, the interventionist can update her beliefs about g. Define

$$\widehat{G} = \left\{ g' = G | \widetilde{g}_{ij} = g'_{ij} \text{ for all } i, j \right\},$$

which is the set of networks in G that are not ruled out by the monitoring. The posterior beliefs are

$$\pi'(g') = \begin{cases} \frac{1}{Q}\pi_0(g'), & \text{if } g' \in \widehat{G}, \\ 0, & \text{if } g' \notin \widehat{G}, \end{cases}$$

where $Q \equiv \sum_{g'' \in \widehat{G}} \pi_0(g'')$. The posterior probability that criminal i has degree x in g is

$$\Pr[d_i(g) = x] = \sum_{g' \in \widehat{G}} \pi'(g') 1(d_i(g') = x).$$

(b) We now show that with non-monitored criminals $i, j \in \overline{I}_Z$, arresting i is strictly better than arresting j when i's observed degree is strictly higher that j's $(d_i(\widetilde{g}) > d_j(\widetilde{g}))$.

Notice that for any network $g' \in \widehat{G}_z$, we must have $d_k(\widetilde{g}) \leq d_k(g') \leq n - 1 - (z - d_k(\widetilde{g}))$ for all $k \in \widetilde{I}_z$.

Case 1 is that $d_i(\widetilde{g}) > n-1-(z - d_j(\widetilde{g}))$. In this case, according to the posterior beliefs, the lowest possible degree for i must be strictly larger than the highest possible degree for j, and choosing i yields a strictly higher expected crime reduction than j.

Case 2 is $d_i(\widetilde{g}) \leq n - 1 - (z - d_j(\widetilde{g}))$. Suppose this case. Define

$$\widehat{G}_{ij} = \left\{ g' = \widehat{G} \middle| \begin{array}{l} \{d_i(\widetilde{g}) \leq d_i(g') \leq n - 1 - (z - d_j(\widetilde{g}))\} \cap \\ \{d_i(\widetilde{g}) \leq d_j(g') \leq n - 1 - (z - d_j(\widetilde{g}))\} \end{array} \right\},$$

which is the set of all $g' = \widehat{G}$ for which i and j have degree between $d_i(\widetilde{g})$ and $n - 1 - (z - d_j(\widetilde{g}))$. Define

$$\widehat{G}'_{ij} = \left\{ g' = \widehat{G} | g' \notin \widehat{G}_{ij} \right\},$$

which is the complement of $\widehat{G}_{ij}$ in $\widehat{G}$. As constructed, $d_i(g') > d_j(g')$ in any $g' \in \widehat{G}'_{ij}$ because either $d_j(g') < d_i(\widetilde{g})$ or $d_i(g') > n - 1 - (z - d_j(\widetilde{g}))$ or both.

Using the constructed posterior beliefs, the expected degrees for i and j are

$$\begin{aligned} \mathbf{E}d_i(g) &= \frac{1}{Q}\left\{ \sum_{x=d_i(\widetilde{g})}^{n-1-(z-d_j(\widetilde{g}))} x \sum_{g' \in \widehat{G}_{ij}} \pi_0(g') 1(d_i(g') = x) + \sum_{x=d_i(\widetilde{g})}^{n-1-(z-d_i(\widetilde{g}))} x \sum_{g' \in \widehat{G}'_{ij}} \pi_0(g') 1(d_i(g') = x) \right\}, \\ \mathbf{E}d_j(g) &= \frac{1}{Q}\left\{ \sum_{x=d_i(\widetilde{g})}^{n-1-(z-d_j(\widetilde{g}))} x \sum_{g' \in \widehat{G}_{ij}} \pi_0(g') 1(d_j(g') = x) + \sum_{x=d_j(\widetilde{g})}^{n-1-(z-d_j(\widetilde{g}))} x \sum_{g' \in \widehat{G}'_{ij}} \pi_0(g') 1(d_j(g') = x) \right\} \end{aligned}$$

By anonymity, the first term inside each bracket is identical. Hence,

$$\begin{aligned} \mathbf{E}d_i(g) &> \mathbf{E}d_i(g) \Leftrightarrow \\ \sum_{x=d_i(\widetilde{g})}^{n-1-(z-d_i(\widetilde{g}))} x \sum_{g'\in\widehat{G}'_{ij}} \pi_0(g')\,1(d_i(g')=x) &> \sum_{x=d_j(\widetilde{g})}^{n-1-(z-d_j(\widetilde{g}))} x \sum_{g'\in\widehat{G}'_{ij}} \pi_0(g')\,1(d_j(g')=x). \end{aligned}$$

Because $d_i(g') > d_j(g')$ for all $g' \in \widehat{G}'_{ij}$, the expected degree for i must be higher than the expected degree for j, and an optimal intervention would involve choosing to arrest i rather than j. This holds for all $i, j \in \overline{I}_Z$ with $d_i(\widetilde{g}) > d_j(\widetilde{g})$, so the optimal non-monitored criminal to arrest is one with observed degree equal to $\widetilde{d}_z^{\max}$.

Proposition 2: *Fix the number of monitored criminals z. There exists a function*

$$m(z,\pi_0) \equiv \frac{1}{Q} \sum_{x=\widetilde{d}_z^{\max}}^{n-1-\left(z-\widetilde{d}_z^{\max}\right)} \sum_{g'\in\widehat{G}} x\pi_0(g')\,1(d_i(g')=x) - \widetilde{d}_z^{\max} \geq 0$$

such that an optimal arrest is to arrest a highest degree monitored criminal when $d_z^{\max} \geq \widetilde{d}_z^{\max} + m(z,\pi_0)$, but select a highest observed degree non-monitored criminal otherwise.

Proof. (a) By Proposition 1, we can ignore non-monitored criminals with $d_i(\widetilde{g}) < \widetilde{d}_z^{\max}$ and monitored criminals with $d_i(g) < d_z^{\max}$. The expected degrees of a non-monitored criminal with $d_i(\widetilde{g}) = \widetilde{d}_z^{\max}$ is

$$\mathbf{E}d_i(g) = \frac{1}{Q} \sum_{x=\widetilde{d}_z^{\max}}^{n-1-\left(z-\widetilde{d}_z^{\max}\right)} \sum_{g'\in\widehat{G}} x\pi_0(g')\,1(d_i(g')=x).$$

If $d_z^{\max} \leq \widetilde{d}_z^{\max}$, then

$$\begin{aligned} d_z^{\max} &\leq \widetilde{d}_z^{\max} \\ &\leq \frac{1}{Q} \sum_{x=\widetilde{d}_z^{\max}}^{n-1-\left(z-\widetilde{d}_z^{\max}\right)} \sum_{g'\in\widehat{G}} x\pi_0(g')\,1(d_i(g')=x), \end{aligned}$$

in which case the optimal selection is to choose a non-monitored criminal with observed degree $\widetilde{d}_z^{\max}$.

Now suppose $d_z^{\max} > \widetilde{d}_z^{\max}$. Choosing a highest degree monitored criminal is optimal when

$$d_z^{\max} \geq \frac{1}{Q} \sum_{x=\widetilde{d}_z^{\max}}^{n-1-\left(z-\widetilde{d}_z^{\max}\right)} \sum_{g'\in\widehat{G}} x\pi_0(g')\,1(d_i(g')=x).$$

Define

$$m\left(z,\pi_0\right) \equiv \frac{1}{Q} \sum_{x=\widetilde{d}_z^{\max}}^{n-1-\left(z-\widetilde{d}_z^{\max}\right)} \sum_{g' \in \widehat{G}} x\pi_0\left(g'\right) 1\left(d_i\left(g'\right)=x\right) - \widetilde{d}_z^{\max}.$$

Then the maximum degree monitored criminal is optimally selected when

$$d_z^{\max} \geq \widetilde{d}_z^{\max} + m\left(z,\pi_0\right).$$

The first term of $m\left(z,\pi_0\right)$ is at least $\widetilde{d}_z^{\max}$, which implies $m\left(z,\pi_0\right) \geq 0$.

Proposition 3: *Suppose π_0 corresponds to a Poisson random network generation. Fix $\widetilde{d}_z^{\max}$. As z increases (decreases), then the premium $m\left(z,\pi_0\right)$ decreases (increases), and a monitored criminal is more (less) likely to be arrested.*

Proof: Consider $i \in \overline{I}_z$ with $d_i\left(\widetilde{g}\right) = \widetilde{d}_z^{\max}$. Because each link forms independently, we can calculate the posterior expected degree for i as the sum of $d_i\left(\widetilde{g}\right) = \widetilde{d}_z^{\max}$ plus the expected collaborations with all other non-monitored criminals in $\overline{I}_z$. By the Poisson process with link formation via probability p, the expected degree for i is

$$\begin{aligned} \mathbf{E}d_i\left(g\right) &= d_i\left(\widetilde{g}\right) + \left(n-1-z\right)p \\ &= \widetilde{d}_z^{\max} + \left(n-1-z\right)p. \end{aligned}$$

Inserting this expected degree into $m\left(z\right)$ yields

$$\begin{aligned} m\left(z,\pi_0\right) &= \mathbf{E}d_i\left(g\right) - \widetilde{d}_z^{\max} \\ &= \left(n-1-z\right)p. \end{aligned}$$

With $\widetilde{d}_z^{\max}$ fixed, an increase in z decreases $\mathbf{E}d_i\left(g\right)$. Because $d_z^{\max}$ is non-decreasing in z, it follows that a drop in $m\left(z,\pi_0\right)$ makes choosing an arrested criminal more likely.

Proposition 4: *Fix g. As the number of monitored criminals z increases (decreases), the expected reduction in crime using the optimal disruption strategy increases (decreases).*

Proof: Fix g. By Proposition 1, we need only consider arrests of monitored criminals with $d_z^{\max}$ and arrests of non-monitored criminals with $\widetilde{d}_z^{\max}$. We first show that $\max\left\{d_z^{\max}, \widetilde{d}_z^{\max}\right\}$ is weakly increasing in z after monitoring. We then show that this implies the claim.

Suppose monitoring has occurred with monitoring level z, which yields $\max\left\{d_z^{\max}, \widetilde{d}_z^{\max}\right\}$. Suppose z increases to $z' = z+1$. This implies that an $i \in \overline{I}_z$ leaves $\overline{I}_z$ and goes into I_z after being the $z+1$-th monitored criminal. It must be true that under z either (a) i is not a highest observed degree non-monitored criminal s.t. $d_i\left(\widetilde{g}\right) < \widetilde{d}_z^{\max}$ or (b) i is a highest observed degree non-monitored criminal s.t. $d_i\left(\widetilde{g}\right) = \widetilde{d}_z^{\max}$.

Suppose (a), then $\widetilde{d}_{z'}^{\max} \geq \widetilde{d}_{z}^{\max}$ because the high collaboration non-monitored criminal under z is still not monitored under z', and also $d_{z'}^{\max} \geq d_{z}^{\max}$ because the newly monitored criminal i may have more collaborations than $d_{z}^{\max}$, or both. Under each of these scenarios we are assured of having $\max\left\{d_{z'}^{\max}, \widetilde{d}_{z'}^{\max}\right\} \geq \max\left\{d_{z}^{\max}, \widetilde{d}_{z}^{\max}\right\}$.

Now suppose (b). It must be true that under z either (i) $\max\left\{d_{z}^{\max}, \widetilde{d}_{z}^{\max}\right\} = d_{z}^{\max}$, (ii) $\max\left\{d_{z}^{\max}, \widetilde{d}_{z}^{\max}\right\} = \widetilde{d}_{z}^{\max}$, or both (i) and (ii). If (i), then we must have $d_{z'}^{\max} \geq \max\left\{d_{z}^{\max}, \widetilde{d}_{z}^{\max}\right\} = d_{z}^{\max}$ as the highest degree monitored agent must still have at least as high degree as the highest observed degree non-monitored criminal. If (ii), then we must have $d_{z'}^{\max} \geq \max\left\{d_{z}^{\max}, \widetilde{d}_{z}^{\max}\right\} = \widetilde{d}_{z}^{\max}$ as i is now in $\overline{I}_z$. Either way, $\max\left\{d_{z'}^{\max}, \widetilde{d}_{z'}^{\max}\right\} \geq \max\left\{d_{z}^{\max}, \widetilde{d}_{z}^{\max}\right\}$, and this is also true if both (i) and (ii).

We have thus established that $\max\left\{d_{z}^{\max}, \widetilde{d}_{z}^{\max}\right\}$ weakly increases post monitoring for any g and z.

Step 2. Because $\max\left\{d_{z'}^{\max}, \widetilde{d}_{z'}^{\max}\right\} \geq \max\left\{d_{z}^{\max}, \widetilde{d}_{z}^{\max}\right\}$ holds for any g and for any z and $z' = z + 1$ after the monitoring occurs, it must be true that $\mathbf{E}\max\left\{d_{z'}^{\max}, \widetilde{d}_{z'}^{\max}\right\} \geq \mathbf{E}\max\left\{d_{z}^{\max}, \widetilde{d}_{z}^{\max}\right\}$ holds for any g and for any z and z' before monitoring occurs.

B Experiment Instructions

SCREEN 1: WELCOME

Welcome to this experiment at ______. Thank you for participating.

You are about to participate in a study of decision-making, and you will be paid for your participation in cash, privately at the end of this session. What you earn depends partly on your decisions and partly on chance.

Please turn off your cell phone.

The entire session consists of multiple practice rounds and real rounds. You will be paid for the real rounds only. The practice rounds are only for you to familiarize yourself with the decision-making environment.

All rounds will take place through the computer terminals. It is important that you do not communicate with any other participants during the session.

When you are ready, please click continue to go to the instructions.

SCREEN 2: Instructions 1

During each round, the computer will select a different criminal network comprised of 18 criminals.

The level of crime produced by each criminal depends in part on the number of that criminal's collaborations. Each criminal produces 15 units of crime plus an additional 10 units of crime for each collaboration. For example, if a criminal had 3 collaborations, then that criminal, if not arrested, would produce 15+(10*3) units of crime, which is 45 units of crime. The 15 units come from that criminal's own activities, while the 10*3 comes from the collaborations. To help you remember this formula, it will be displayed on your screen during the duration of the experiment.

After being shown an image of the network, you will choose which single criminal to arrest. Arresting a criminal results in a reduction in total overall crime because that criminal does not produce crime and because there are fewer collaborations among the surviving criminals.

For each unit of crime reduced by the arrest, you will earn $0.006. Thus, the more crime you reduce, the more money you earn during the experiment.

SCREEN 3: Test Screen 1

Before proceeding, you must answer some questions. These questions test your comprehension. Remember that a criminal with x collaborations produces 15+(10*x) units of crime.

Please select the answer.

1. What is the level of crime produced by a criminal with 7 collaborations?

SCREEN 4: Answer Screen 1

You are CORRECT!

Remember that a criminal with x collaborations produces 15+(10*x) units of crime.

1. What is the level of crime produced by a criminal with 7 collaborations?

The correct answer is (e) 85 because a criminal with 7 collaborations produces 15+(10*7) = 85 units of crime.

Now, please select the correct answer for the following question:

2. True or false: The correct answer above is also the amount of crime reduced by arresting a criminal with 7 collaborations.

Alternate Screen:

I am sorry to inform you that your answer is WRONG.

Remember that a criminal with x collaborations produces 15+(10*x) units of crime.

1. What is the level of crime produced by a criminal with 7 collaborations?

The correct answer would be (e) 85 because a criminal with 7 collaborations produces 15+(10*7) = 85 units of crime.

Now, please select the correct answer for the following question:

2. True or false: The correct answer above is also the amount of crime reduced by arresting a criminal with 7 collaborations.

SCREEN 5: Answer Screen 2

You are CORRECT!

Remember that a criminal with x collaborations produces 15+(10*x) units of crime.

2. True or false: The correct answer above is also the amount of crime reduced by arresting a criminal with 7 collaborations.

The correct answer is (b) false because, after the criminal is arrested, the former collaborators of the arrested criminal now have fewer collaborations and thus also produce less crime. Thus, the arrest directly reduces crime by removing a criminal but indirectly reduces even more crime because surviving criminals have fewer collaborations.

Click continue to proceed.

Alternate Screen:

I am sorry to inform you that your answer is WRONG.

Remember that a criminal with x collaborations produces 15+(10*x) units of crime.

2. True or false: The correct answer above is also the amount of crime reduced by arresting a criminal with 7 collaborations.

The correct answer is (b) false because, after the criminal is arrested, the former collaborators of the arrested criminal now have fewer collaborations and thus also produce less crime. Thus, the arrest directly reduces crime by removing a criminal but indirectly reduces even more crime because surviving criminals have fewer collaborations.

Click continue to proceed.

SCREEN 6: Instructions 2 (one version for each observation level):

Before making your choice in each round, you will be shown an image of part of the crime network.

Each criminal is represented by a small triangle, and the lines connecting criminals represent the collaborations between criminals.

You will view each criminal, but the computer will randomly select SOME of those criminals and show only the collaborations of those criminals. Any other collaborations in the crime network will not be shown. The criminals who have ALL of their collaborations revealed are green. The other criminals are blue - you may see all, some or none of the collaborations of each blue criminal. You will only see a blue criminal's collaboration if a partner's collaborations were revealed.

In each round, the computer will select ___ criminals and reveal their collaborations.

SCREEN 7: Instructions 3:

To arrest a particular criminal, please click on the triangle corresponding to that criminal.

NOTE: You can select any criminal to arrest, whether or not you saw that criminal's collaborations. That means you may click on any green OR blue triangle.

IMPORTANT! You make your choice by clicking on that triangle, so do not click on a triangle until you are ready to make your choice. Once you make your choice, you cannot change it.

Click continue to participate in the 2 practice rounds. After completing the practice rounds, you will participate in 16 real rounds. You will then receive further instructions.

References

[1] Albert, Réka, Hawoong Jeong, and Albert-László Barabási. 2000. "Error and Attack Tolerance of Complex Networks." *Nature* 406: 378-382.

[2] Arce, Daniel, et al. 2011. "Terrorism Experiments." *Journal of Peace Research* 48: 373-382.

[3] Arin, K. Peren, et al. 2011. "Exploring the Dynamics between Terrorism and Anti-Terror Spending: Theory and UK-Evidence." *Journal of Economic Behavior & Organization* 77: 189-202.

[4] Atkinson, Michael and Lawrence Wein. 2010. "An Overlapping Networks Approach to Resource Allocation for Domestic Counterterrorism." *Studies in Conflict and Terrorism* 33: 618-651.

[5] Baccara, Mariagiovanna and Heski Bar-Isaac. 2008. "How to Organize Crime." *Review of Economic Studies* 75: 1039-1067.

[6] Ballester, Coralio, et al. 2006. "Who's Who in Networks. Wanted: The Key Player." *Econometrica* 74: 1403-1417.

[7] Barabási, Albert-László and Réka Albert. 1999. "Emergence of Scaling in Random Networks." *Science* 286: 509-512.

[8] Becker, Gary. 1968. "Crime and Punishment: An Economic Approach." *Journal of Political Economy* 76: 169-217.

[9] Borgatti, S.P. 2003. "The Key Player Problem." Pp. 241-252 in R. Breiger, K. Carley, & P. Pattison, eds., *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers*. National Academy of Sciences Press.

[10] Callaway, Duncan, et al. 2000. "Network Robustness and Fragility: Percolation on Random Graphs." *Physical Review Letters* 85: 5468-5471.

[11] Cameron, Samuel. 1988. "The Economics of Crime Deterrence: A Survey of Theory and Evidence." *Kyklos* 41: 301-323.

[12] Chowdhury, Subhasish. 2009. "An Experimental Investigation of Colonel Blotto Games." CESIfo Working Paper No. 2688.

[13] Cohen, Rueven, el al. 2001. "Breakdown of the Internet under Intentional Attack." *Physical Review Letters* 86: 3682-3685.

[14] Das, Satya P. 2008. "Some Mechanisms of Terror Cycles." *Journal of Economic Behavior & Organization* 67: 644-656.

[15] El-Gamal, Mahmoud and David Grether. 1995. "Are People Bayesian? Uncovering Behavioral Strategies." *Journal of the American Statistical Association* 90: 1137-1145.

[16] Enders, Walter and Paan Jindapon. 2009. "Network Externalities and the Structure of Terror Networks." *Journal of Conflict Resolution* 54: 262-280.

[17] Enders, Walter and Xuejuan Su. 2007. "Rational Terrorists and Optimal Network Structure." *Journal of Conflict Resolution* 51: 33-57.

[18] Farley, Jonathan. 2010. "Breaking Al Qaeda Cells: A Mathematical Analysis of Counterterrorism Operations (A Guide for Risk Assessment and Decision Making)." *Studies in Conflict and Terrorism* 26: 399-411.

[19] Feinstein, Jonathan and Edward Kaplan. 2010. "Analysis of a Strategic Terror Organization." *Journal of Conflict Resolution* 54: 281-302.

[20] Fischbacher, Urs. 2007. "z-Tree: Zurich Toolbox for Read-made Economic Experiments." *Experimental Economics* 10: 171-178.

[21] Franke, Jörg and Tahir Öztürk. 2009. "Conflict Networks." Ruhr Economic Papers #116.

[22] Goyal, Sanjeev and Adrien Vigier. 2010. "Robust Networks." Working Paper.

[23] Hausken, Kjell. 2009. "Strategic Defense and Attack of Complex Networks." *International Journal of Permormability Engineering* 5: 13-30.

[24] Hong, Sunghoon. 2008. "Hacking-proofness and Stability in a Model of Information Security Networks." Working Paper.

[25] Hong, Sunghoon. 2011. "Strategic Network Interdiction." Working Paper.

[26] Hoyer, Britta and Kris De Jaegher. 2010. "Strategic Network Disruption and Defense." Tjalling C. Koopmans Research Instutite, Discussion Paper Series nr: 10-13.

[27] Intriligator, Michael. 2010. "The Economics of Terrorism." *Economic Inquiry* 48: 1-13.

[28] Jackson, Matthew. 2008. *Social and Economic Networks.* Princeton University Press.

[29] Kosfeld, Michael. 2006. "Economic Networks in the Laboratory: A Survey." *Review of Network Economics* 3: 20-42.

[30] Kovenock, Dan and Brian Roberson. 2010. "The Optimal Defense of Networks of Targets." Institute for Research in the Behavioral, Economic, and Management Sciences Paper No. 1251.

[31] Krebs, Valdis. 2002. "Mapping Networks of Terrorist Cells." *Connections* 24: 43-52.

[32] Llussá, Fernanda and José Tavares. 2007. "Economics and Terrorist: What We Know, What We Should Know and the Data We Need." CEPR Discussion Paper No. DP6509.

[33] Miffen, Thomas et al. 2004. A Random Graph Model for Terrorist Transactions. IEEE Aerospace Conference Proceedings 5: 3258-3264.

[34] Milward, H. Brinton and Joerg Raab. 2005. "Dark Networks as Problems Revisited: Adaptation and Transformation of Islamic Terror Organizations Since 9/11." Working Paper.

[35] Pedhahzur, Ami and Arie Perliger. 2006. "The Changing Nature of Suicide Attacks: A Social Network Perspective." *Social Forces* 84: 1987-2008.

[36] Raab, Joerg and H. Brinton Milward. 2003. "Dark Networks as Problems." *Journal of Public Administration Research and Theory* 13: 413-439.

[37] Roberson, Brian. 2006. "The Colonel Blotto Game." *Economic Theory* 29: 1-24.

[38] Roberts, Nancy and Sean Everton. 2011. "Strategies for Combating Dark Networks." *Journal of Social Structure* 12.

[39] Roth, Alvin. 1995. "Introduction To Experimental Economics." In John Kagel and Alvin Roth, eds., *The Handbook of Experimental Economics.* Princeton University Press.

[40] Sageman, Marc. 2004. *Understanding Terror Networks.* University of Pennsylvania Press.

[41] Sageman, Marc. 2008. *Leaderless Jihad: Terror Networks in the Twenty-first Century.* University of Pennsylvania Press.

[42] Starmer, Chris. 1999. "Experiments in Economics: Should we Trust the Dismal Scientists in White Coats?" *Journal of Economic Methodology* 6: 1-30.

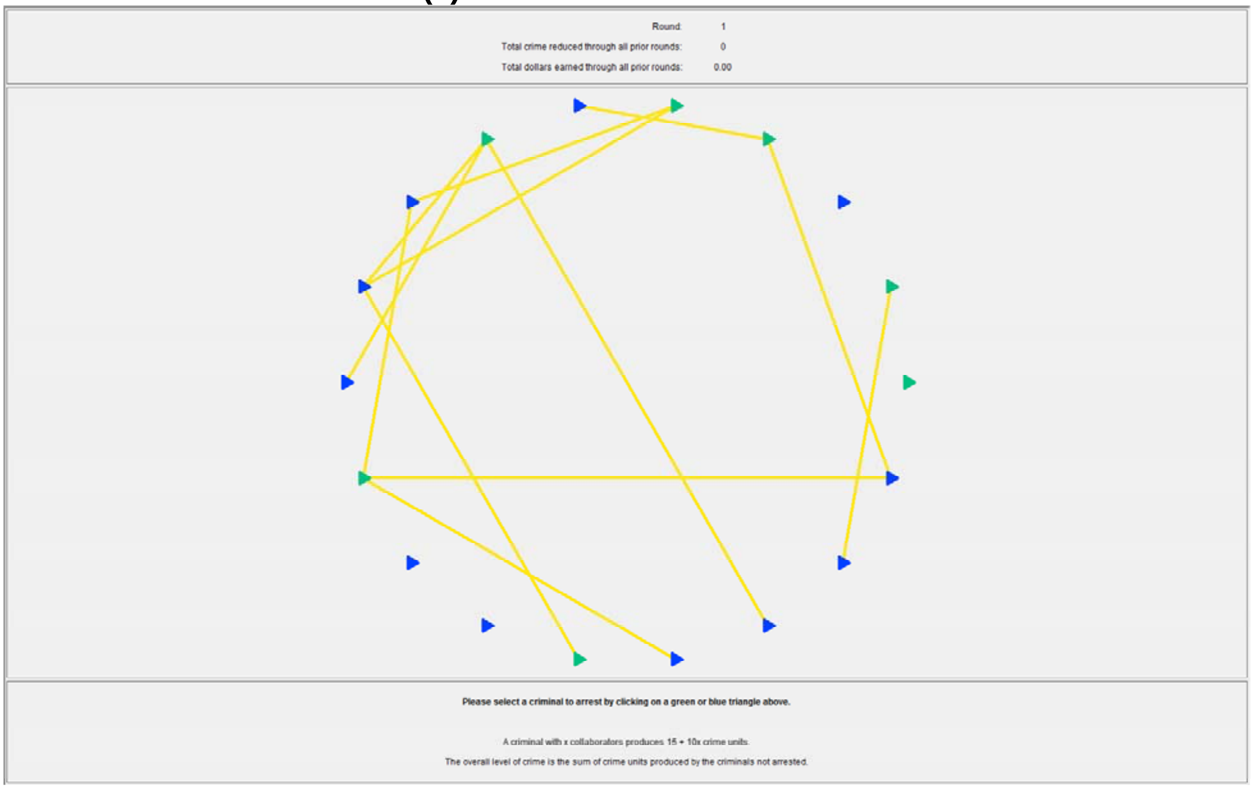
[43] Tanizawa, T., et al. 2005. "Optimization of Network Robustness to Waves of Targeted and Random Attacks." *Physical Review* 71.

[44] Wasserman, Stanley and Kathleen Faust. 1994. *Social Network Analysis: Methods and Applications.* Cambridge University Press.

[45] Xu, Jennifer and Hsinchun Chen. 2008. "The Topology of Dark Networks." *Communications of the ACM* 51: 58-65.

Figure 1: Experiment Display Screen-shots

(a) Pre-arrest Screen-shot



(b) Post-arrest Screen-shot

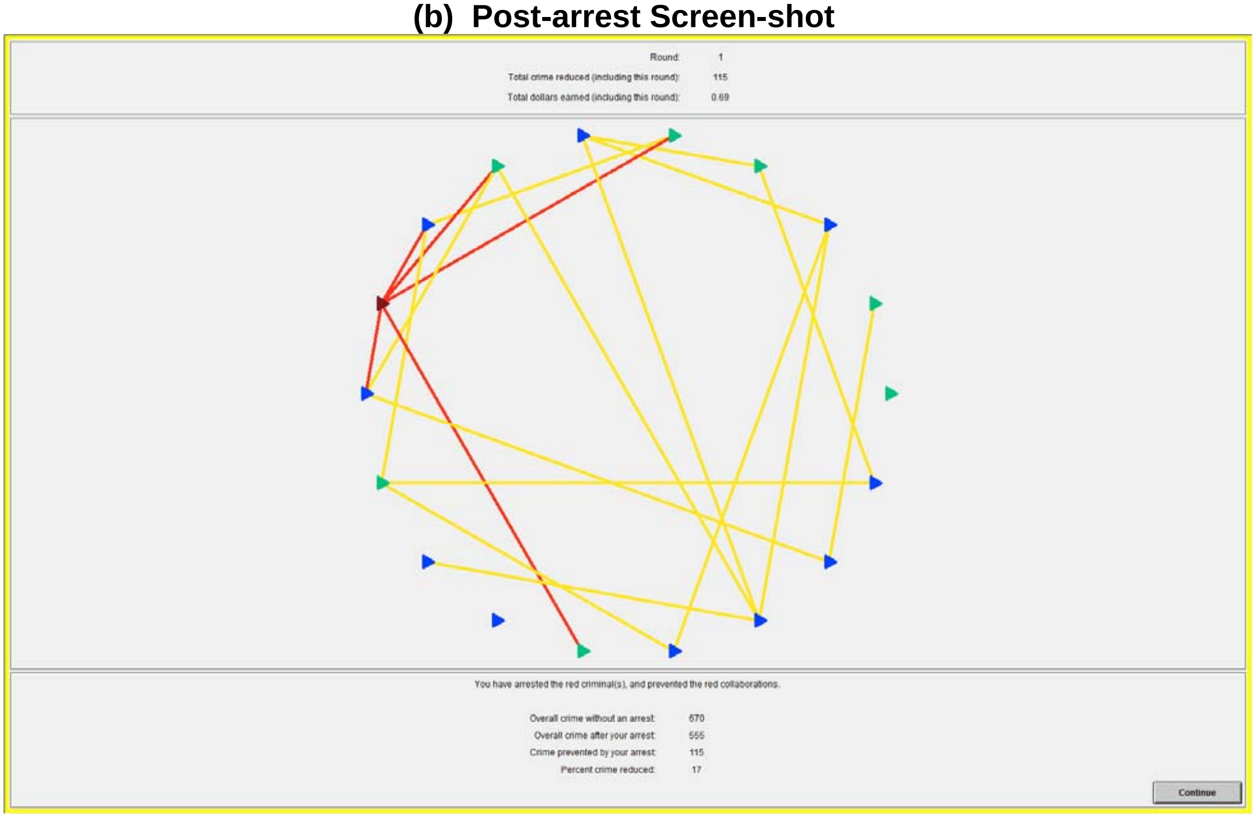


Table 1: Session Information

Session	Number of subjects	Percent male*	Percent with 1+ statistics courses*	Major*						Average take-home earnings**
				Percent Business or Economics	Percent Psychology or Cognitive Science	Percent Other Social Science, Sociology, Criminology	Percent Physical Science / Engineering	Percent Life Science / Biology / Public Health	Percent Humanities or Undecided/ undeclared	
All	105	30%	50%	17%	32%	16%	7%	19%	9%	$28.26
A-12-7-2	16	38%	63%	38%	25%	13%	0%	13%	13%	$27.56
A-2-7-12	18	11%	33%	11%	28%	22%	6%	28%	6%	$27.63
A-Random-1	11	36%	45%	9%	36%	9%	9%	18%	18%	$27.79
A-Random-2	18	39%	61%	28%	33%	17%	6%	6%	11%	$28.13
B-12-7-2	10	20%	50%	0%	40%	0%	10%	50%	0%	$29.08
B-2-7-12	17	41%	47%	12%	29%	24%	6%	24%	6%	$29.19
B-Random-1	15	20%	47%	13%	40%	20%	13%	7%	7%	$28.65

Notes: * denotes information obtained from the questionnaire. ** includes the $5 show-up payment.

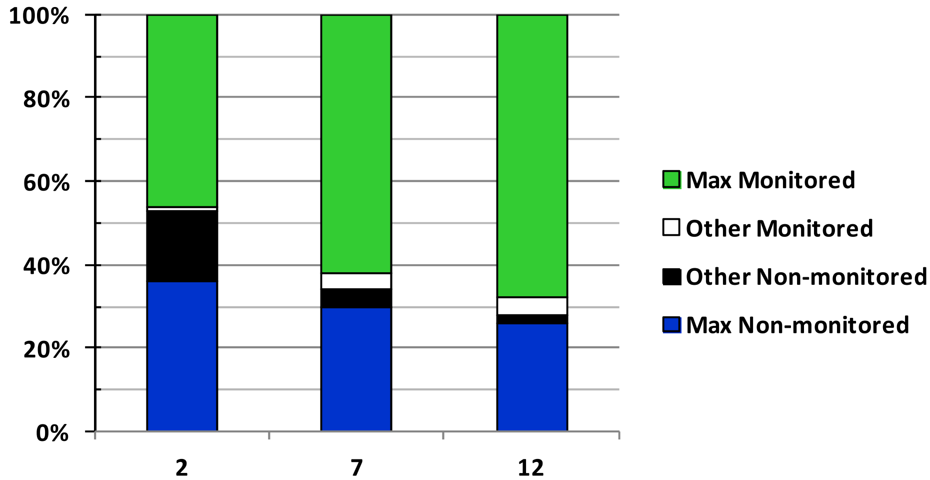
Table 2: Percent Criminal Choice by Network Set, Session, and Observation Level

		All	Observation Level 2	Observation Level 7	Observation Level 12
(a) Set A Networks					
All Set A Sessions	**Max Monitored**	**59%**	**46%**	**62%**	**68%**
	Other Monitored	3%	1%	4%	4%
	Max Non-monitored	**31%**	**36%**	**30%**	**26%**
	Other Non-monitored	8%	17%	4%	2%
A-12-7-2	**Max Monitored**	**55%**	**32%**	**55%**	**77%**
	Other Monitored	4%	1%	5%	6%
	Max Non-monitored	**30%**	**42%**	**34%**	**14%**
	Other Non-monitored	11%	26%	5%	3%
A-2-7-12	**Max Monitored**	**70%**	**60%**	**73%**	**76%**
	Other Monitored	3%	3%	3%	4%
	Max Non-monitored	**22%**	**26%**	**20%**	**19%**
	Other Non-monitored	5%	11%	4%	0%
A-Random-1	**Max Monitored**	**51%**	**43%**	**57%**	**54%**
	Other Monitored	3%	1%	5%	3%
	Max Non-monitored	**38%**	**40%**	**35%**	**39%**
	Other Non-monitored	8%	16%	4%	4%
A-Random-2	**Max Monitored**	**56%**	**47%**	**60%**	**60%**
	Other Monitored	2%	1%	3%	3%
	Max Non-monitored	**35%**	**37%**	**33%**	**34%**
	Other Non-monitored	7%	15%	4%	3%
(b) Set B Networks					
All Set B Sessions	**Max Monitored**	**52%**	**37%**	**54%**	**65%**
	Other Monitored	2%	1%	3%	3%
	Max Non-monitored	**36%**	**40%**	**38%**	**30%**
	Other Non-monitored	10%	22%	6%	2%
B-12-7-2	**Max Monitored**	**54%**	**28%**	**58%**	**75%**
	Other Monitored	2%	1%	1%	4%
	Max Non-monitored	**35%**	**48%**	**36%**	**20%**
	Other Non-monitored	10%	24%	5%	1%
B-2-7-12	**Max Monitored**	**51%**	**37%**	**51%**	**64%**
	Other Monitored	1%	2%	1%	0%
	Max Non-monitored	**35%**	**31%**	**41%**	**34%**
	Other Non-monitored	13%	30%	6%	2%
B-Random-1	**Max Monitored**	**52%**	**44%**	**53%**	**60%**
	Other Monitored	4%	2%	5%	5%
	Max Non-monitored	**37%**	**44%**	**36%**	**32%**
	Other Non-monitored	7%	10%	6%	4%

Notes: May add to >100% due to round off error. Monitored criminals are green. Non-monitored are blue.

Figure 2: Distribution of Arrests by Monitored

(a) Set A Networks



(b) Set B Networks

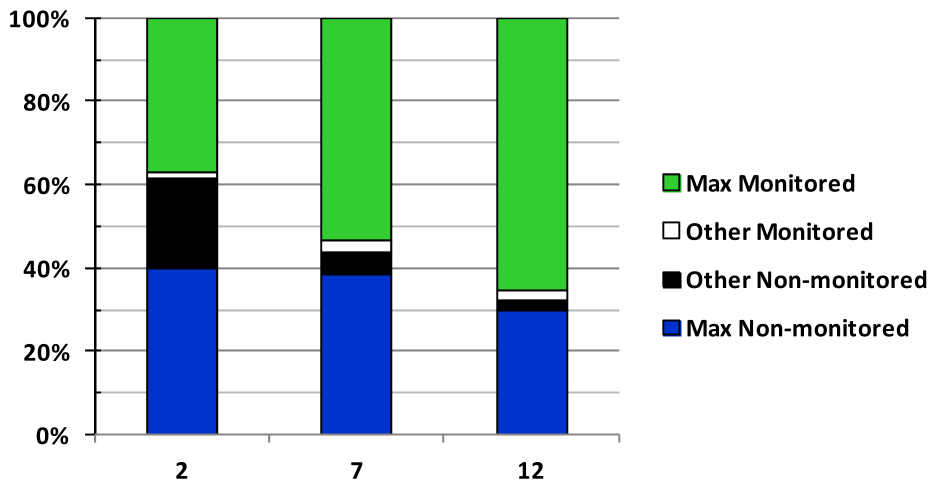
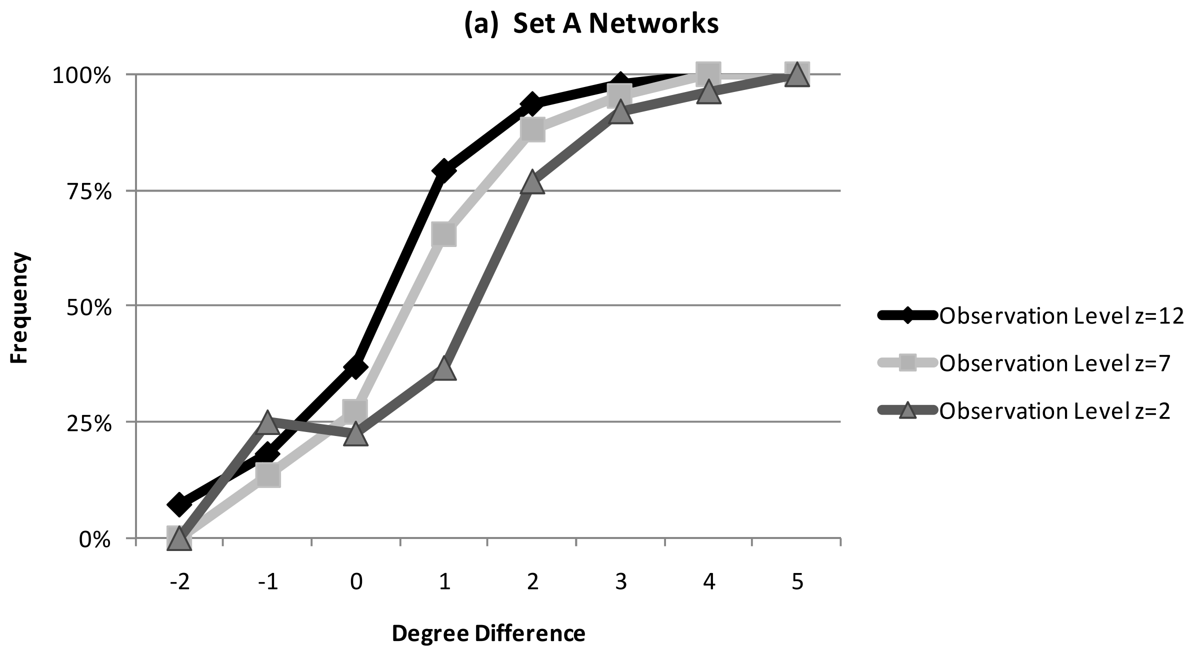


Table 3: Percent Arrested Monitored Criminal by Network Set and Observation Level

		Degree Difference between Max Monitored and Max Non-monitored								
		-3	-2	-1	0	1	2	3	4	5
(a) Set A Networks										
All	**%**		6%	17%	28%	60%	86%	95%	99%	100%
	Obs.		16	113	680	1121	754	258	72	10
Observation Level z=12	**%**		7%	18%	37%	79%	94%	98%	100%	100%
	Obs.		14	72	182	384	234	96	24	2
Observation Level z=7	**%**		0%	14%	27%	65%	88%	95%	100%	100%
	Obs.		2	37	210	353	295	87	21	3
Observation Level z=2	**%**			25%	23%	37%	77%	92%	96%	100%
	Obs.			4	288	384	225	75	27	5
(b) Set B Networks										
All	**%**	0%	8%	9%	17%	47%	83%	93%	97%	100%
	Obs.	2	26	77	416	739	454	194	92	16
Observation Level z=12	**%**	0%	4%	10%	31%	73%	94%	96%	97%	100%
	Obs.	2	23	50	112	204	160	83	36	2
Observation Level z=7	**%**		33%	8%	19%	47%	86%	95%	96%	
	Obs.		3	24	135	252	176	37	45	
Observation Level z=2	**%**			0%	7%	29%	64%	88%	100%	100%
	Obs.			3	169	283	118	74	11	14

Figure 3: Percent Arrested Monitored Criminal by Degree Difference and Observation Level



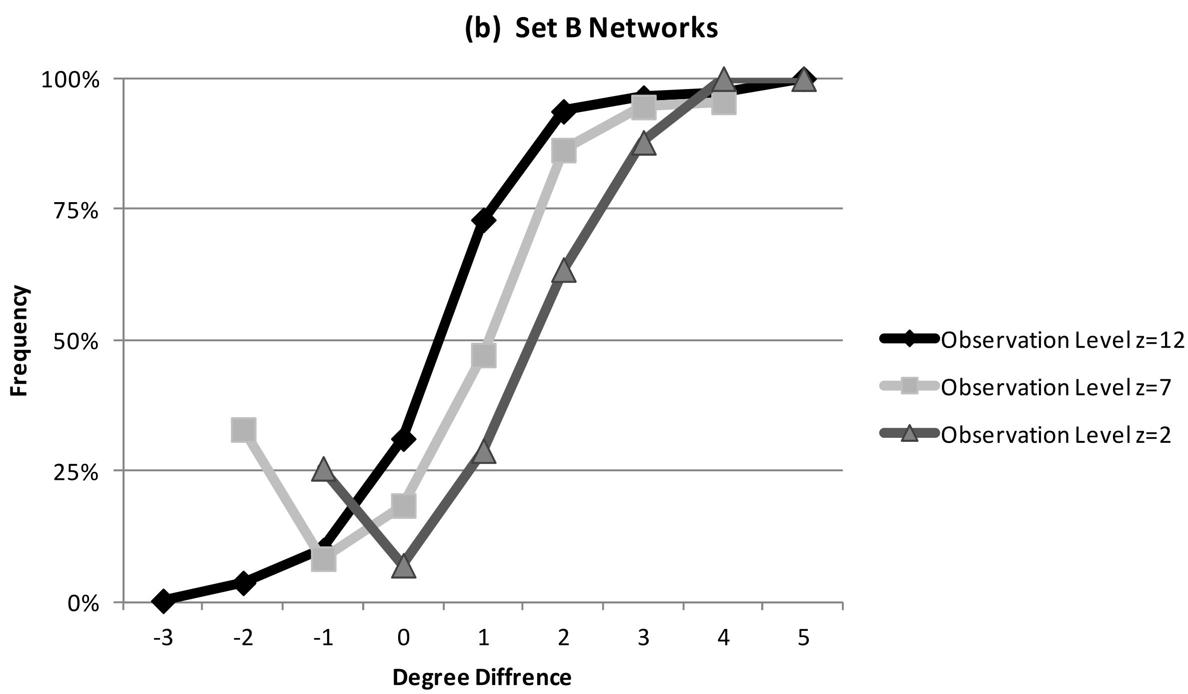


Table 4: Probit Regressions Predicting Arrest of Monitored Criminal, by Network Set

	Set A Networks	Set B Networks
Intercept	-0.11**	-0.33***
	(0.053)	(0.069)
Dummy Observation Level z=7	-0.35***	-0.50***
	(0.067)	(0.083)
Dummy Observation Level z=2	-0.85***	-1.06***
	(0.066)	(0.85)
Degree Difference	0.80***	0.85***
	(0.031)	(0.038)
Observations	3024	2016
Percent Concordant	0.798	0.835
Percent Discordant	0.137	0.108
Percent Tied	0.065	0.057
Pairs	2,162,240	1,008,839
Test: Dummy Obs z=7 equals z=2	62.10***	50.04***

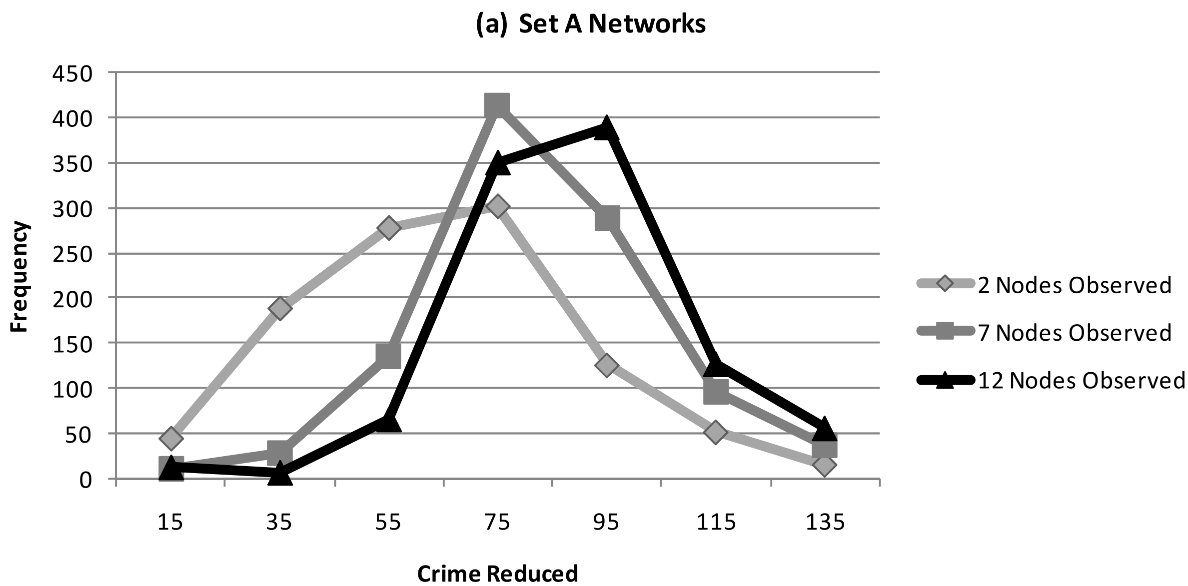
Notes: Standard errors in parentheses. *, **, and *** denote significance at 10%, 5%, and 1% levels, respectively.

Table 5: Crime Reduction Statistics by Network Set, Session, and Observation Level

	Statistic	All	Observation Level: 2	7	12
(a) Set A Networks					
All Set A Sessions	**Mean**	**78.61**	**64.82**	**82.28**	**88.73**
	<Median>	<75>	<55>	<75>	<95>
	(Std. dev.)	(25.12)	(25.65)	(21.88)	(21.25)
	[Obs.]	[3024]	[1008]	[1008]	[1008]
A-12-7-2	**Mean**	**77.81**	**65.63**	**80.08**	**87.73**
	<Median>	<75>	<75>	<75>	<95>
	(Std. dev.)	(25.78)	(26.68)	(23.11)	(22.38)
	[Obs.]	[768]	[256]	[256]	[256]
A-2-7-12	**Mean**	**78.13**	**61.74**	**82.71**	**89.93**
	<Median>	<75>	<55>	<75>	<95>
	(Std. dev.)	(25.45)	(26.41)	(21.52)	(18.91)
	[Obs.]	[864]	[288]	[288]	[288]
A-Random-1	**Mean**	**78.56**	**64.55**	**83.30**	**87.84**
	<Median>	<75>	<75>	<75>	<95>
	(Std. dev.)	(24.85)	(24.65)	(21.20)	(22.24)
	[Obs.]	[528]	[176]	[176]	[176]
A-Random-2	**Mean**	**79.84**	**67.36**	**83.19**	**88.96**
	<Median>	<75>	<75>	<75>	<95>
	(Std. dev.)	(24.34)	(24.31)	(21.50)	(21.86)
	[Obs.]	[864]	[288]	[288]	[288]
(b) Set B Networks					
All Set B Sessions	**Mean**	**82.87**	**67.20**	**86.28**	**95.12**
	<Median>	<75>	<55>	<95>	<95>
	(Std. dev.)	(29.17)	(28.16)	(27.03)	(24.97)
	[Obs.]	[2016]	[672]	[672]	[672]
B-12-7-2	**Mean**	**83.29**	**67.25**	**87.50**	**95.13**
	<Median>	<95>	<75>	<95>	<95>
	(Std. dev.)	(29.45)	(30.01)	(26.13)	(24.72)
	[Obs.]	[480]	[160]	[160]	[160]
B-2-7-12	**Mean**	**83.70**	**65.96**	**87.87**	**97.28**
	<Median>	<95>	<55>	<95>	<95>
	(Std. dev.)	(29.89)	(28.53)	(27.72)	(24.22)
	[Obs.]	[816]	[272]	[272]	[272]
B-Random-1	**Mean**	**81.64**	**68.58**	**83.67**	**92.67**
	<Median>	<75>	<55>	<75>	<95>
	(Std. dev.)	(28.13)	(26.46)	(26.73)	(25.83)
	[Obs.]	[720]	[240]	[240]	[240]

Notes: Crime reduction equals 15+20d, where d is the degree of the arrested criminal.

Figure 4: Frequency of Crime Reduction Level by Observation Level



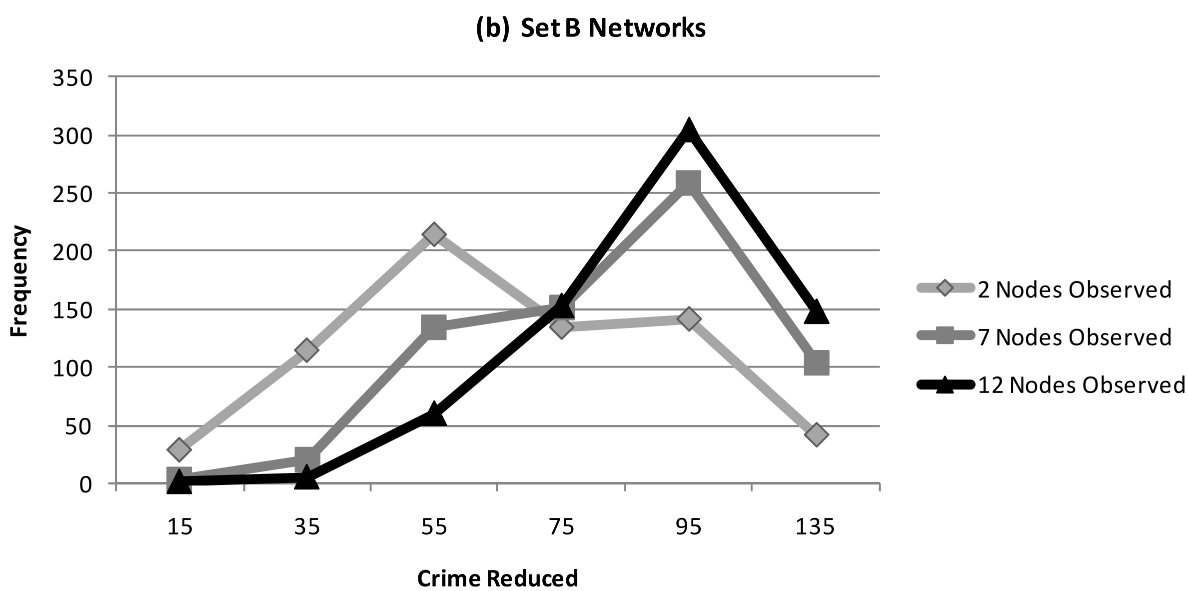
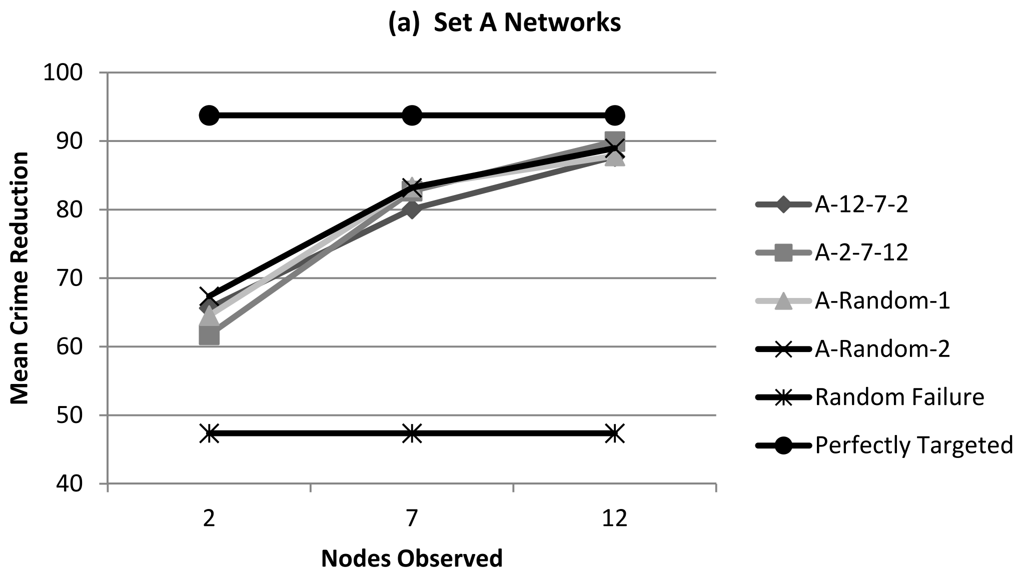


Figure 5: Mean Crime Reduction by Observation Level and Session



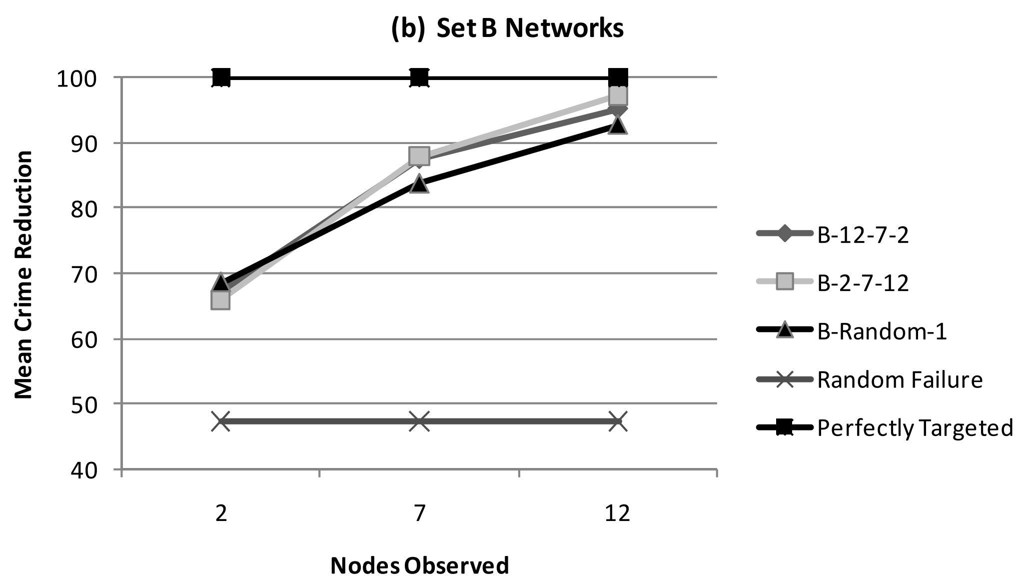


Table 6: Crime Reduction Matched Pair Test Statistics by Session and Observation Level

	Statistic	Observation Levels Matched 12 - 7	12 - 2	7 - 2
(a) Set A Networks				
All Set A Sessions	Student t (p-value)	**10.0 (<0.001)**	**27.1 (<0.001)**	**18.7 (<0.001)**
	Wilcoxon (p-value)	21137 (<0.001)	108387 (<0.001)	82214 (<0.001)
	Counts {+,=,-}	{296,392,96}	{641,698,57}	{562,687,125}
	[Obs.]	[1008]	[1008]	[1008]
A-12-7-2	Student t (p-value)	**5.1 (<0.001)**	**12.0 (<0.001)**	**7.1 (<0.001)**
	Wilcoxon (p-value)	1634 (<0.001)	6299 (<0.001)	4170 (<0.001)
	Counts {+,=,-}	{83,111,28}	{156,172,16}	{129,172,43}
	[Obs.]	[256]	[256]	[256]
A-2-7-12	Student t (p-value)	**6.7 (<0.001)**	**17.0 (<0.001)**	**12.0 (<0.001)**
	Wilcoxon (p-value)	2078 (<0.001)	10517 (<0.001)	8182 (<0.001)
	Counts {+,=,-}	{88,112,24}	{199,210,11}	{175,204,29}
	[Obs.]	[288]	[288]	[288]
A-Random-1	Student t (p-value)	**3.20 (0.002)**	**11.1 (<0.001)**	**9.4 (<0.001)**
	Wilcoxon (p-value)	526 (0.001)	3456 (<0.001)	2948 (<0.001)
	Counts {+,=,-}	{50,71,21}	{113,126,13}	{104,121,17}
	[Obs.]	[176]	[176]	[176]
A-Random-2	Student t (p-value)	**4.9 (<0.001)**	**14.0 (<0.001)**	**9.4 (<0.001)**
	Wilcoxon (p-value)	1331 (<0.001)	7899 (<0.001)	6061 (<0.001)
	Counts {+,=,-}	{75,98,23}	{173,190,17}	{154,190,36}
	[Obs.]	[288]	[288]	[288]
(b) Set B Networks				
All Set B Sessions	Student t (p-value)	**9.5 (<0.001)**	**23.2 (<0.001)**	**15.3 (<0.001)**
	Wilcoxon (p-value)	9634 (<0.001)	46450 (<0.001)	33735 (<0.001)
	Counts {+,=,-}	{189,244,55}	{415,448,33}	{360,437,77}
	[Obs.]	[672]	[672]	[672]
B-12-7-2	Student t (p-value)	**4.2 (<0.001)**	**10.5 (<0.001)**	**7.6 (<0.001)**
	Wilcoxon (p-value)	543 (<0.001)	2672 (<0.001)	2023 (<0.001)
	Counts {+,=,-}	{46,62,16}	{98,110,12}	{89,107,18}
	[Obs.]	[160]	[160]	[160]
B-2-7-12	Student t (p-value)	**6.7 (<0.001)**	**16.8 (<0.001)**	**10.7 (<0.001)**
	Wilcoxon (p-value)	1584 (<0.001)	8250 (<0.001)	6450 (<0.001)
	Counts {+,=,-}	{75,92,17}	{176,184,8}	{156,186,30}
	[Obs.]	[272]	[272]	[272]
B-Random-1	Student t (p-value)	**5.4 (<0.001)**	**12.6 (<0.001)**	**7.9 (<0.001)**
	Wilcoxon (p-value)	1235 (<0.001)	5349 (<0.001)	3414 (<0.001)
	Counts {+,=,-}	{68,90,22}	{141,154,13}	{115,144,29}
	[Obs.]	[240]	[240]	[240]

Notes: p-values in parentheses.